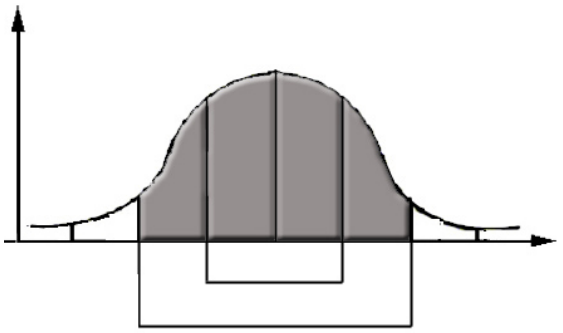
الجمهورية العربية السورية
وزارة التعليم العالي
جامعة تشرين
كلية الاقتصاد

الإحصاء التطبيقي



الدكتور
ابراهيم محمد العلي
أستاذ في قسم الإحصاء

الدكتور
محمد عكروش
مدرس في قسم الإحصاء

السنة الثانية
مديرية الكتب والمطبوعات

2004 −2005 م

الإحصاء التطبيقي

الفهرس

العنوان	الصفحة
المقدمة	**11**
الفصل التمهيدي: التحليل التوافقي	**13**
1 . المفهوم الأساسي للتحليل التوافقي.	15
2 . شروط تشكيل المجموعات.	16
3 . المتبادلات.	17
4 . المرتبات.	22
5 . المتوافقات.	28
6 . تجزئة مجموعة.	37
7 . ثنائي الحد.	40
8 . كثير الحدود.	43
9 . تطبيقات ثنائي الحد	44
10 . تمرينات.	50
الفصل الأول: احتمالات الحوادث.	**53**
1 . أنواع الحوادث.	55
2 . تعاريف الاحتمال.	58
3 . قواعد جمع الاحتمالات.	64
4 . قواعد ضرب الاحتمالات.	66
5 . قاعدة الاحتمال الكلي وقانون (بيز).	71
6 . تمرينات.	75

الفصل الثاني: المتحولات العشوائية وقوانين توزيعها. **77**
1 . تعاريف عامة. 79
2 . قوانين التوزيع الاحتمالية. 80
3 . خواص قانون التوزيع. 88
4 . تابع التوزيع. 91
5 . خواص تابع التوزيع. 98
6 . التوقع الرياضي. 104
7 . الوسيط. 107
8 . المنوال. 107
9 . مقاييس التشتت (التباين والانحراف المعياري). 108
10 . تمرينات. 111
الفصل الثالث: بعض قوانين التوزيع الاحتمالية **113**
1 . قانون التوزيع الاحتمالي المنتظم لمتحول منقطع. 115
2 . قانون التوزيع الثنائي للمتحول المنقطع 117
3 . قانون توزيع بواسون للمتحول المنقطع x. 130
4 . قانون التوزيع المنتظم العام (المتحولات مستمرة). 141
5 . التوزيع الطبيعي العام. 145
6 . قانون التوزيع الطبيعي المعياري. 152
7 . قانون توزيع χ^2 (كاي مربع). 155
8 . قانون (توزيع ستودينت) أو T. 159
9 . قانون التوزيع F (توزيع فيشر . سينديكور). 162
10 . قانون التوزيع الاحتمالي الإحصائي. 167

11 . تمرينات. 170

الفصل الرابع: العينات وقضايا التقدير. 175
1 . تمهيد. 175
2 . أنواع العينات. 177
3 . التقدير النقطي لمؤشرات المجتمع. 184
4 . مجالات الثقة (للمتوسط والإجمالي والنسبة والفروقات). 198
5 . تقدير حجم العينة (للمتوسط والإجمالي والنسبة). 214
6 . تمرينات. 220
الفصل الخامس: اختبار الفرضيات البسيطة بواسطة Z و t. 223
1 . تمهيد. 225
2 . أنواع الفرضيات الإحصائية. 227
3 . عناصر الاختيار. 228
4 . اختبارات حول متوسط المجتمع. 235
5 . اختبارات حول إجمالي المجتمع. 239
6 . اختبارات حول النسبة في المجتمع. 241
7 . اختبارات حول الفرق بين مجتمعين طبيعيين. 243
8 . اختبارات حول الفرق بين نسبتين في مجتمعين طبيعيين. 247
9 . اختبار حول القياسات الشاذة لحذفها. 251
10 . تمرينات. 254
الفصل السادس: اختبار كاي مربع χ^2 وتطبيقاته. 257
1 . مقدمة. 259
2 . اختبار مطابقة التكرارات بواسطة χ^2 . 261

3 . اختبار مطابقة النسب بواسطة χ^2 . 263
4 . اختبار مطابقة توزيع تجريبي مع توزيع نظري. 267
5 . اختبار استقلالية متحولين. 281
6 . شروط تطبيق المؤشر χ^2 . 284
7 . حالة خاصة اختبار χ^2 عندما يكون لدينا جدول 2×2. 289
8 . معامل التوافق C. 292
9 . معامل التوافق T^2 . 292
10 . اختبار تباين مجتمع طبيعي بواسطة χ^2 . 293
11 . اختبار تناظر الجداول المربعة بواسطة χ^2 . 296
12 . تمرينات. 299
الفصل السابع: الاختبارات اللامعلمية من عينة واحدة. 301
1 . مقدمة. 303
2 . الاختبارات اللامعلمية المعرفة على المتحولات الأسمية. 308
3 . الاختبارات المعرفة على المتحولات الرتبية والمجالية. 327
4 . اختبارات الارتباط بين متحولين رتيبيين (x,y) . 339
5 . اختبارات الارتباط بين متحول اسمي ومتحول رتبي. 347
6 . تمرينات. 351
الفصل الثامن: مراقبة جودة الإنتاج. 355
1 . مقدمة. 357
2 . أهمية الرقابة على جودة الإنتاج. 358
3 . الرقابة الإحصائية على جودة الإنتاج. 359

4 . الأساليب الإحصائية في مراقبة الجودة. 360
5 . حالات أخرى من مراقبة الجودة. 394
6 . تمرينات. 404

الفصل التاسع: تحليل التباين باتجاه واحد. 409
1 . مقدمة. 411
2 . تصميم تجربة على تحليل التباين باتجاه واحد. 412
3 . المقارنات المتعددة للمتوسطات. 432
4 . شروط تطبيق تحليل التباين 442
5 . علاقة تطبيق التباين باختبار ستودينت وبمعامل الارتباط 444
5 . تمرينات. 445
الفصل العاشر: تحليل التباين باتجاهين. 449
1 . مقدمة. 451
2 . تصميم تجربة تحليل التباين باتجاهين. 451
3 . تمرينات. 474
الفصل الحادي عشر: الانحدار الخطي البسيط والارتباط. 477
1 . مقدمة. 477
2 . تحليل الانحدار البسيط: 478
. طريقة المربعات الصغرى. 483
. تحليل تباينات خط الانحدار. 490
. اختبار معنوية المعاملات. 497
3 . حساب معامل التحديد. 501
4 . معامل الارتباط الخطي. 511
5 . اختبار دلالة معامل الارتباط ρ . 517
6 . حالات خاصة لمعامل الارتباط. 518

7 . تمرينات. 522

الفصل الثاني عشر: الانحدار الخطي المتعدد 525

1 . مقدمة. 527

2 . إيجاد معادلة الانحدار الخطي المتعدد باستخدام المعادلات الطبيعية 527

3 . الاختبارات الإحصائية المتعلقة بالانحدار الخطي المتعدد 541

4 . إيجاد معادلة الانحدار الخطي المتعدد باستخدام المصفوفات 550

5 . تمرينات. 555

قائمة المراجع العربية والأجنبية. 557

دليل بالمصطلحات العلمية. 563

ملاحق الجداول الإحصائية. 571

الثنائي.

بواسون.

الطبيعي المعياري.

توزيع ستودينت.

توزيع χ^2.

توزيع F.

التوزيعات اللامعلمية.

المقدمة

يعد هذا الكتاب مكملاً لكتابنا السابق (مبادئ علم الإحصاء) المقرر على طلاب السنة الأولى، ويتناول عدداً من موضوعات الإحصاء التطبيقي التي تستخدم بكثرة في الدراسات التطبيقية والأبحاث العلمية في المجالات الاقتصادية والاجتماعية والسكانية والتجارية. لذلك يفترض المؤلفان أن القارئ لهذا الكتاب ملم بأساسيات علم الإحصاء: كالمتوسطات والتباين والانحراف المعياري والوسيط والمنوال والارتباط والسلاسل الزمنية والأرقام القياسية... إلخ، ولهذا جاءت موضوعاته متعمدة على تلك الأساسيات ومكملة لها.

يبدأ الكتاب بفصل تمهيدي عن التحليل التوافقي كتقدمة للفصل الأول، الذي يتناول احتمالات الحوادث وقواعد جمع وضرب الاحتمالات، ثم يتدرج بدراسة المتحولات العشوائية وقوانين توزيعها الاحتمالية، ويستعرض أهم خواص تلك القوانين ومميزاتها العددية. وبعدها ينطلق لدراسة أهم قوانين التوزيع الاحتمالية المنقطعة والمستمرة، كالتوزيع المنتظم والتوزيع الثنائي، وتوزيع بواسون، والتوزيع المنتظم المستمر، والتوزيع الطبيعي العام، والتوزيع الطبيعي المعياري، وتوزيع (كاي مربع) وتوزيع (ستودينت) وتوزيع (فيشر. سينديكور). ثم التوزيع الاحتمالي الإحصائي (التجريبي).

ويتضمن فصلاً خاصاً عن العينات وأنواعها وعن قضايا التقدير النقطي والتقدير المجالي لكل من المتوسط والإجمالي والنسبة والفرق المتوسطين والنسبتين لمجتمعين طبيعيين، وكيفية تقدير حجم العينة اللازم لتحقيق الدقة المطلوبة. ويتابع دراسة اختبارات الفرضيات البسيطة بواسطة المتحول الطبيعي المعياري Z أو بواسطة متحول توزيع (ستودينت) t لكل من المتوسط والإجمالي والنسبة والفرق بين المتوسطين والنسبتين لمجتمعين. ويخصص فصلاً كاملاً لاختبار (كاي مربع) وتطبيقاته على التكرارات والنسب والتوزيعات التجريبية لمتحول واحد، وعلى الاستقلال والتناظر والتوافق لمتحولين نوعيين. ثم يستمر في استعراض أهم الاختبارات اللامعلمية المعرفة على عينة واحدة. كاختبار ثنائي الحد والتعاقب والاقتران والتوافق والتطابق والتناظر حول

الوسيط والارتباط، ثم ينعطف لدراسة قضايا مراقبة الجودة وأساليبها المختلفة. وبعدها ينتقل إلى تحليل التباين باتجاه واحد ثم تحليل التباين باتجاهين، ويخصص الفصلين الأخيرين لمعالجة قضايا الانحدار الخطي البسيط وقضايا الانحدار الخطي المتعدد.

ويلحق بالكتاب دليل للمصطلحات العلمية وقائمة للمراجع العربية والأجنبية، وعدد من الجداول الإحصائية الضرورية لإجراء الاختبارات الواردة فيه.

ومع أن التأليف المشترك هو مسؤولية مشتركة، إلا أننا نود أن نشير إلى أن مساهمة كل من المؤلفيّن كانت كمايلي:

إن الفصول التمهيدي والأول والثاني والثالث والرابع والخامس والسادس والسابع، كانت من تأليف الدكتور ابراهيم العلي.

وإن الفصول الثامن والتاسع والعاشر والحادي عشر والثاني عشر كانت من تأليف الدكتور محمد عكروش.

وإننا إذ نقدم هذا الكتاب إلى طلابنا الأعزاء والى زملائنا الكرام، نأمل أن نكون قد وفقنا في تأليفه وعرض موضوعاته، ونحن لا ندعي الكمال في عرضه ولا الشمولية في موضوعاته، ونعتذر مسبقاً لجميع القراء الكرام عن كل خطأ لم نلاحظه، أو عن كل اختصار لم نشرحه، وعن كل إسهاب لم نقصده، ونرجوهم جميعاً أن يرسلوا إلينا بملاحظاتهم حول الكتاب وموضوعاته، وحول الأخطاء المطبعية وغير المطبعية، التي لم نستطع التخلص منها رغم كل الجهود التي بذلت لتداركها وتصحيحها عند الطباعة.

اللاذقية في 2004/10/16

والله ولي التوفيق **المؤلفان**

الفصل التمهيدي

التحليل التوافقي

Comparnatics Analysis

ويتضمن الفقرات التالية:

1 . المفهوم الأساسي للتحليل التوافقي.
2 . شروط تشكيل المجموعات.
3 . المتبادلات.
4 . المرتبات.
5 . المتوافقات.
6 . تجزئة مجموعة.
7 . ثنائي الحدين.
8 . كثير الحدود.
9 . تطبيقات ثنائي الحد.
10 . تمرينات.

الفصل التمهيدي

التحليل التوافقي

عندما ندرس البدائل المحتملة لحادث نهتم بمعرفة عدد الطرق التي يظهر بها ذلك الحادث، أو عدد الطرق الممكنة لترتيب عناصر مجموعة، مثلاً، بكم طريقة يمكن ترتيب ستة أشخاص حول طاولة مستديرة؟ أو بكم طريقة يمكن تشكيل لجنة من أربعة أعضاء من مجموعة مكونة من عشرين شخصاً ... الخ. كما أننا نسعى إلى تشكيل مجموعات جزئية من مجموعة كلية. ونقصد بتشكيل مجموعة A تحوي k عنصراً من مجموعة كلية Ω تحوي n عنصراً هو اختيار k عنصراً من n عنصراً واعتبار هذه العناصر مجموعة جديدة.

ولكننا ندرك مباشرة أن عملية الاختيار هذه يمكن أن تنفَّذ بطرق متعددة ومختلفة، لذلك فإننا سنتناول في هذا الفصل عملية تشكيل جميع المجموعات الممكنة وعدد طرق تشكيلها وأساليب تشكيلها. ولقد وضعناها تحت عنوان (التحليل التوافقي) Comparnatics Analysis.

0-1 : المفهوم الأساسي للتحليل التوافقي:

لنفترض أنه لدينا مجموعة A_1 تحوي n_1 عنصراً، و مجموعة A_2 تحوي n_2 عنصراً، ..... ومجموعة A_m تحوي n_m عنصراً. عندئذ يكون عدد الطرق الممكنة لتشكيل مجموعة مؤلفة من عنصر واحد من كل مجموعة من هذه المجموعات التي عددها m مجموعة مساوياً:

$$n_1 \cdot n_2 \cdot n_3 \ldots n_m$$

وإذا لم تكن الأعداد $n_1\ n_2\ n_m$ كبيرة جداً، فمن الممكن عندئذ حساب عدد هذه الطرق، لتوضيح هذا المفهوم نقدم الأمثلة التالية:

مثال 1:

في منطقة ما، يوجد 14 نوع من الذباب و17 نوع من الفراشات و13 نوع من

البعوض، بكم طريقة يمكن تشكيل مجموعة مؤلفة من حشرة واحدة من كل نوع من هذه الأنواع؟

لدينا $m=3$ مجموعة $n_1=14$ طريقة لاختيار حشرة من الذباب، و $n_2=17$ طريقة لاختيار فراشة من الفراشات، و $n_3=13$ طريقة لاختيار بعوضة من البعوض. وبالاستناد إلى المفهوم الأساسي في التحليل المزجي، فإنه يوجد:

$$41\cdot 17\cdot 13=3094$$

طريقة مختلفة لتشكيل مجموعة مؤلفة من حشرة واحدة من كل نوع من هذه الأنواع الثلاثة.

مثال 2:

تتكون مدرسة من أربعة صفوف، أعداد طلابها على الترتيب *47, 51, 54, 55* طالباً، ونريد أن نختار عريفاً لكل صف؟ فبكم طريقة يمكن اختيار مجموعة عرفاء هذه الصفوف؟

يمكن اختيار عريف الصف الأول بـ 47 طريقة.

وعريف الصف الثاني بـ 51 طريقة.

وعريف الصف الثالث بـ 54 طريقة.

وأخيرا عريف الصف الرابع بـ 55 طريقة.

وبالتالي، وبالاستناد إلى المفهوم الأساسي في التحليل التوافقي، يوجد:

$$47\cdot 51\cdot 54\cdot 55=7119090$$

طريقة مختلفة لاختيار مجموعة عرفاء هذه الصفوف.

0-2: شروط تشكيل المجموعات:

عند تشكيل المجموعات بحجم n، لابد وأن نأخذ بعين الاعتبار مختلف الشروط الممكنة التي تميّز تلك المجموعات بعضها عن بعض، وسنميز بين أنواع المجموعات التالية:

المتبادلات Permutations:

وهي المجموعات التي تحوي نفس العناصر ولكنها تختلف عن بعضها باختلاف ترتيب أحد هذه العناصر داخل المجموعة نفسها.

المرتبات Arrangements:

وهي المجموعات التي تختلف عن بعضها إذا اختلفت بترتيب أحد عناصرها على الأقل ضمن المجموعة، أو إذا اختلفت باختلاف أحد عناصرها على الأقل، أو بكلاهما معاً.

المتوافقات Combinations:

وهي المجموعات التي تعتبر مختلفة عن بعضها باختلاف أحد عناصرها على الأقل، وذلك بغض النظر عن ترتيب تلك العناصر داخل المجموعة.

0-3: المتبادلات:

لنفترض أنه لدينا مجموعة تحوي n عنصراً، وإننا نهتم بترتيب هذه العناصر وفق نظام معين، من اجل ذلك، نختار أحد هذه العناصر ليكون "الأول"، وبعدها نأخذ عنصراً آخر ليكون " الثاني" وهكذا... ونسمي كل تنظيم من هذا القبيل "متبادلة"، ومنه نتوصل إلى التعريف التالي:

تعريف المتبادلة: المتبادلة لـ n عنصراً هي أي ترتيب لهذه العناصر وفق تنظيم محدد.

ومتبادلات مجموعة ما ولتكن A (بفرض أن $A \subseteq \Omega$) هي جملة المجموعات التي يمكن تشكيلها من عناصر A نفسها بحيث تختلف كل مجموعة منها عن الأخرى بترتيب عنصر واحد على الأقل.

مثال 3 :

اكتب جميع متبادلات الأحرف الأربعة w,z,y,x.

إن جميع المتبادلات الممكنة هي:

wxyz	wxzy	wyxz	wyzx	wzxy	wzyx
xwyz	xwzy	xywz	xyzw	xzwy	xzyw

ywxz	ywzx	yxwz	yxzw	yzwx	yzxw
zwyx	zwxy	zxwy	zxyw	zywx	zyxw

وبذلك نجد أن عدد هذه المتبادلات هو 24 متبادلة، ونحصل على هذا العدد من الجداء: $4 \cdot 3 \cdot 2 \cdot 1 = 24$

مثال 4:

ليكن لدينا ثلاثة مواد كيميائية C_3, C_2, C_1 محفوظة في ثلاثة أنابيب اختبار، وهذه الأنابيب الثلاثة موضوعة على حامل واحد. اكتب جميع الإمكانات الممكنة لترتيب هذه المواد على الحامل.

إن جميع الإمكانات الممكنة لترتيب هذه المواد الثلاث على الحامل هي المتبادلات التالية:

$$C_1C_2C_3,\ C_1C_3C_2,\ C_2C_1C_3,\ C_2C_3C_1,\ C_3C_1C_2\ ,C_3\ C_2\ C_1$$

ونرى أن عدد متبادلات ثلاثة عناصر هو ستة متبادلات، وتحصل عليه من الجداء: $3 \cdot 2 \cdot 1 = 6$

نظرية 1:

إن عدد متبادلات n عنصراً هو $n!$ ونرمز له بـ P_n ونكتب ذلك كمايلي:

$$P_n = n!$$

البرهان:

لترتيب n عنصراً وفق ترتيب محدد، نختار أحد العناصر ليكون الأول، ويوجد n طريقة لنقوم بهذا الاختيار، ومن الـ $(n-1)$ عنصراً المتبقية نختار العنصر الثاني، أي يوجد $(n-1)$ طريقة للقيام بهذا الاختيار الثاني. وبتكرار هذه المحاكمة نرى أن العنصر الذي ترتيبه k يتم اختياره بـ $n-k+1$ طريقة، حيث $(k = 1,2....n)$.

وبالاعتماد على المفهوم الأساسي في التحليل التوافقي، نجد أن عدد المتبادلات الممكنة يساوي:

$$P_n = n \cdot (n-1) \cdot \ldots\ldots \cdot (n-k+1) \ldots\ldots 3 \cdot 2 \cdot 1 = n! \qquad (1-0)$$

متبادلة لـ n عنصراً.

مثال 5:

بكم طريقة يمكن لستة أشخاص أن يرتبوا أنفسهم:

أ . وفق رتل.

ب . حول طاولة مستديرة.

الحل:

أ . إن عدد طرق الترتيب وفق رتل هو عدد متبادلات ستة عناصر أي:

$$P_6 = 6! = 720$$

ب . لترتيب ستة أشخاص حول طاولة دائرية، فإننا نختار شخصاً ما بطريقة افتراضية، ونضعه على أحد المقاعد ثم نرتب الأشخاص الخمسة الباقين بالنسبة لذلك الشخص، وهذا يمكن أن يتم بـ:

$P_5 = 5! = 120$ طريقة

مثال 6:

إذا كانت A هي المجموعة المؤلفة من العناصر $A = \{a,b,c,d,e\}$ ، فبكم طريقة يمكن أن نرتب عناصر هذه المجموعة.

الحل:

إن عدد الطرق هو عدد متبادلات خمسة عناصر أي:

$$P_5 = 5! = 120$$

مثال 7:

لتكن لدينا الأحرف (أ . ب . ج . د). ماهو عدد الكلمات (دون النظر إلى المعنى) التي يمكن تشكيلها من هذه الأحرف الأربعة، بشرط أن لايتكرر الحرف في الكلمة أكثر من مرة.

الحل:

إن عدد الكلمات التي يمكن تشكيلها دون تكرار أي حرف فيها هو عدد المتبادلات لأربعة عناصر أي:

$$P_4 = 4! = 24$$

ويمكننا أن نعالج تحت هذه الفقرة قضايا أكثر تعقيداً لذلك نأخذ المثال التالي:

مثال 8:

مقعد يتسع لخمسة أشخاص: ولدينا 3 أشخاص من الذكور و2 من الإناث. بكم طريقة مختلفة يمكن ترتيب الذكور والإناث على هذا المقعد؟ علماً بأننا لانميّز بين الذكور. كما لانميزّ بين الإناث.

إن إحدى الطرق لحل هذه المسألة هي أن نكتب جميع الإمكانات. ولنرمز للذكر بـ M وللإنثى بـ F ، فنجد إن إمكانات الترتيب هي:

FFMMM	FMMMF	MFMMF	MMFMF	MMMFF
MMFFM	MFMFM	MFFMM	FMMFM	FMFMM

أي أنه لدينا عشر إمكانيات، أو عشر طرق مختلفة لترتيب الذكور والإناث على هذا المقعد دون التمييز بين الإناث ودون التمييز بين الذكور.

هنالك طريقة عامة لحل المسائل من هذا النوع: نحن نعلم بأنه لدينا $5! = 120$ متبادلة من أجل خمسة عناصر، وذلك إذا كانت جميع هذه العناصر مختلفة عن بعضها. أما إذا كانت هنالك ثلاثة عناصر متشابهة (لايمكن تمييزها)، عندئذ نرى أنه يوجد $3! = 6$ متبادلة لايمكن تمييز أية متبادلة من هذه المتبادلات الستة عن بعضها. لتبيان هذا نأخذ المرتبة FFMMM ، ولنرمز للذكور عوضاً عن M بـ $M_3, M_2 M_1$ ، عندها تكون هذه المتبادلات الست المتشابهة هي:

$$FFM_1M_2M_3 - FFM_2M_3M_1 - FFM_3M_1M_2$$
$$FFM_1M_3M_2 - FFM_2M_1M_3 - FFM_3M_2M_1$$

وهكذا نرى أن هذه المتبادلات تكون مختلفة لو كان الذكور متميزين عن بعضهم، ولكن

بما أنه ورد في نص المسألة بأننا لانميّز بين الذكور، فإن هذه المتبادلات الست تعبر عن إمكانية واحدة، وبالتالي فإن عدد المتبادلات المختلفة يتناقص بمقدار $3!=6$ مرات، والتحليل نفسه يطبق بالنسبة للإناث أي أن عدد المتبادلات للإناث ينقص بمقدار $2!=2$ مرة، ومنه نجد أن عدد الطرق التي يمكن أن نرتب بها ثلاثة ذكور دون تمييز واثنتين من الإناث دون تمييز على مقعد ما هو:

$$\frac{5!}{3!\cdot 2!}=10$$

وهذا الجواب يتطابق تماماً مع الجواب السابق.

يمكن تعميم المثال السابق بسهولة، على مجموعة تحوي n عنصراً بينهم:

n_1 عنصراً مكرراً من النمط "1" ولايمكن تمييزها عن بعضها.

n_2 عنصراً مكرراً من النمط "2" ولايمكن تمييزها عن بعضها.

.....

n_k عنصراً مكرراً من النمط "k" ولايمكن تمييزها عن بعضها.

نظرية 2:

إن عدد متبادلات n عنصراً منها n_1 عنصراً مكرراً و n_2 عنصراً مكرراً .... و n_K عنصراً مكرراً هو:

$$P_n^{n_1,n_2,....n_k}=\frac{n!}{n_1!\cdot n_2!....n_K!} \qquad (2-0)$$

إن البرهان على هذه النظرية هو تعميم للمثال السابق.

مثال 9:

كم كلمة متميزة يمكن أن نكتب باستعمال جميع أحرف الكلمة Repetition.

الحل:

لدينا في هذه الكلمة عشر عناصر، بينها 3 عناصر غير متميزة ومكررة مثنى مثنى وهي الأحرف "e" "t" "i" ، وبذلك نجد أن عدد المتبادلات المتميزة الممكنة هو:

$$P_{10}^{2,2,2} = \frac{10!}{2!\cdot 2!\cdot 2!} = \frac{3628800}{2\cdot 2\cdot 2} = 453600 \text{ كلمة}$$

مثال 10 :

زرعنا ثلاثة أنواع مختلفة من الجراثيم في تسعة أنابيب اختبار. تحوي ثلاثة أنابيب جراثيم من النوع الأول، وأربعة من النوع الثاني، واثنين من النوع الثالث. بكم طريقة يمكن ترتيب أنابيب الاختبار هذه في صف واحد، علماً بأننا نهتم فقط بنوعية الجراثيم ولانهتم بترتيب الأنابيب في كل نوع.

الحل:

إن عدد الطرق يساوي:

$$P_9^{3,4,2} = \frac{9!}{3!\cdot 4!\cdot 2!} = \frac{362880}{6\cdot 24\cdot 2} = 12600 \text{ طريقة}$$

0-4: المرتبات:

تتعلق عدة مسائل في التحليل التوافقي، وبصورة خاصة في علم الاحتمالات بكيفية اختيار عنصر من مجموعة تضم n عنصراً (مثلاً كيفية اختيار كرة من صندوق يحوي n كرة، أو كيفية اختيار ورقة من مجموعة ورق لعب يضم 52 ورقة، أو كيفية اختيار رئيس لجنة لصف يضم عدداً معيناً من الطلبة). وقد تتعدى عملية الاختيار العنصر الواحد، وتشمل عدة عناصر أي لتشمل كيفية اختيار k عنصراً من مجموعة تحتوي n عنصراً (بشرط أن: $k<n$). وقد يتم هذا الاختيار بإعادة العنصر أو بدون إعادته إلى المجموعة الأصلية.

فمثلاً عند اختيار 3 كرات من صندوق يحوي 10 كرات، فإن اختيارنا للكرة الأولى يتم من أصل 10 كرات، أما اختيارنا للكرة الثانية، فقد يتم إما بدون إعادة الكرة الأولى إلى الصندوق أي من أصل 9 كرات، أو بعد إعادة الكرة الأولى إلى الصندوق أي من أصل 10 كرات.

بصورة عامة، نقول إن:

إن الاختيار بدون إعادة يشترط علينا عند تشكيل أي مجموعة من المجموعات الممكنة

بألاّ نختار العنصر الواحد من المجموعة Ω أكثر من مرة واحدة ليكون عنصراً في المجموعة المشكلة. أما الاختيار مع الإعادة فيسمح لنا باختيار العنصر الواحد من المجموعة Ω أكثر من مرة ليكون عنصراً من المجموعة المراد تشكيلها. لذا سنميّز بين نوعين من المرتبات.

مرتبات بدون إعادة.

مرتبات مع الإعادة.

*** المرتبات بدون إعادة:**

لتكن لدينا المجموعة $\Omega = \{a,b,c,d,e\}$، فإذا أردنا أن نشكل منها جميع المجموعات الممكنة والمكونة من عنصر واحد، (k=1) ، فإننا سنحصل على المجموعات التالي:

$$\{a\},\{b\},\{c\},\{d\},\{e\}$$

والتي عددها يساوي 5 مجموعات.

وإذا حاولنا أن نشكل من المجموعة Ω نفسها جميع المجموعات الممكنة والمكونة من عنصرين (k=2)، فإننا سنحصل على المجموعات التالية:

$ab \quad ac \quad ad \quad ae$

$ba \quad bc \quad bd \quad be$

$ca \quad cb \quad cd \quad ce$

$da \quad db \quad dc \quad de$

$ea \quad eb \quad ec \quad ed$

ونلاحظ أن عددها يساوي 20 مجموعة، وقد حصلنا عليها بإضافة عنصر إلى يمين كل عنصر من عناصر Ω على الترتيب، ولمرة واحدة فقط، وذلك حتى لايزيد عدد عناصر المجموعة عن اثنين. وبصورة عامة، نلاحظ أنه قد حصلنا مقابل كل عنصر من Ω (التي تحوي n عنصراً) على عدد من المجموعات الثنائية يساوي n-1 مجموعة، وبما أن Ω تتضمن n عنصراً، فإن عدد المجموعات التي يمكن تشكيلها من عنصرين من Ω يساوي:

$n(n-1)$

لذلك كان عددها مثالنا السابق مساوياً: $5 \times 4 = 20$.

أما إذا أردنا تشكيل المجموعات الممكنة التي تحوي ثلاثة عناصر من Ω، فإنه يجب علينا أن نضيف إلى يمين كل من المجموعات الثنائية السابقة أحد عناصر Ω المتبقية وغير الموجودة (لأن الاختيار دون إعادة) في المجموعة الثنائية التي نعالجها، بحيث نشمل جميع العناصر المتبقية، أي نحصل مقابل كل مجموعة ثنائية على عدد من المجموعات الثلاثية يساوي عدد العناصر المتبقية في Ω أي يساوي $(n-2)$ مجموعة، وبما أنه لدينا $n(n-1)$ مجموعة ثنائية، فإن عدد المجموعات التي يمكن تشكيلها من ثلاثة عناصر غير مكررة يساوي:

$n(n-1)(n-2)$ مجموعة

بصورة عامة، يمكننا تعريف المتبادلات بدون إعادة كمايلي:

تعريف المرتبات بدون إعادة: تدعى المجموعات التي يمكن تشكيلها k ف k عنصراً مختلفاً من n عنصراً بمرتبات k عنصراً منn عنصراً، إذا تميّزت كل مجموعة عن الأخريات باختلاف أحد العناصر التي تتكون منها أو باختلاف ترتيبها ضمن المجموعة.

نظرية 3:

إن عدد المرتبات k ف k عنصراً مختلفاً من n عنصراً، والذي سنرمز له بالرمز A_n^k يساوي:

$$A_n^k = n(n-1)(n-2).....(n-k+1) = \frac{n!}{(n-k)!} \qquad (3-0)$$

البرهان:

يمكن اختيار العنصر الأول بـ n طريقة، والعنصر الثاني بـ $(n-1)$ طريقة... والعنصر الذي ترتيبه k بـ $(n-k+1)$ طريقة، وللبرهان على النظرية السابقة نعتمد على المفهوم الأساسي للتحليل المزجي، فنجد أن عدد الطرق الممكنة يساوي:

$$A_n^k = n(n-1)(n-2).....(n-k+1)$$

ثم نضرب الطرف الثاني بالكسر $\frac{(n-k)!}{(n-k)!}$ فنجد أن:

$$A_n^k = n \cdot (n-1)(n-1)...(n-k+1) \cdot \frac{(n-k)!}{(n-k)!} = \frac{n!}{(n-k)!}$$

وبذلك نحصل على أن عدد هذه المرتبات يساوي:

$$A_n^k = \frac{n!}{(n-k)!}$$

مثال 11:

ماهو عدد الكلمات المؤلفة من أربعة حروف، والتي يمكن تشكيلها من أحرف كلمة "مزهرية" (الكلمات قد لايكون لها معنى).

الحل:

إن عدد الكلمات من أربعة حروف الممكن تشكيلها من مجموعة الحروف { م، ز، هـ ، ر، ي ، ة} هو عبارة عن عدد المرتبات مأخوذة أربعة فأربعة من ستة عناصر ويساوي:

$$A_6^4 = \frac{6!}{(6-4)!} = 360 \text{ كلمة}$$

مثال 12:

ماهو عدد الأرقام الأصغر من 100 والتي يمكن تشكيلها من عناصر المجموعة Ω ، وذلك بشرط ألايتكرر العنصر الواحد أكثر من مرة، حيث أن:

$$\Omega = \{0,1,2,3,4,5,6,7,8,9\}$$

الحل:

نلاحظ أن الأرقام التي يطلب تشكيلها تتألف من رقمين مختلفين فقط (k=2). وبذلك نجد أن عددها يساوي:

$$A_n^k = A_{10}^2 = \frac{10!}{(10-2)!} = 10 \cdot 9 = 90 \text{ (رقماً)}$$

ملاحظة:

إن هذه المرتبات تشمل الأرقام 02.01,.....09 ، وبالمقابل أنها لاتشمل الأرقام: 00,11,22,.....99، لوجود شرط عدم تكرار العنصر الواحد أكثر من مرة. ولذلك كان عددها مساوياً لـ 90 رقماً.

*** المرتبات مع الإعادة:**

لنأخذ كمثال وكتمهيد المجموعة Ω السابقة $\Omega = \{a,b,c,d,e\}$ ولنحاول أن نشكل منها المرتبات التي تضم عنصراً واحداً (أي $k=1$) من Ω فنجد أنها تساوي 5 ومحددة كالتالي: $\{a\},\{b\},\{c\},\{d\},\{e\}$

وإذا حاولنا أن نشكل المرتبات التي تحوي عنصرين (أي $k=2$) من Ω مع الأخذ بعين الاعتبار أن الاختيار يتم مع الإعادة فنجد أن هذه المرتبات هي كمايلي:

$aa \quad ab \quad ac \quad ad \quad ae$
$ba \quad bb \quad bc \quad bd \quad be$
$ca \quad cb \quad cc \quad cd \quad ce$
$da \quad db \quad dc \quad dd \quad de$
$ea \quad eb \quad ec \quad ed \quad ee$

وعددها يساوي25 لأنها أصبحت تتضمن المتبادلات التي تسمح بتكرار العنصر نفسه فيها، وهي المتبادلات aa, bb, cc, dd, ee.

تعريف المرتبات مع الإعادة:

تدعى المجموعات التي يمكن تشكيلها بحجم k مع الإعادة من n عنصراً، بمرتبات n مع الإعادة مأخوذة k ف k، وذلك إذا تميزت المجموعات عن بعضها باختلاف العناصر فيها أو بترتيبها ضمن المجموعة، وبشرط أن تتم إعادة العنصر المختار إلى المجموعة الأصلية ليدخل في التشكيل عند اختيار العنصر اللاحق، وهكذا...(أي مع إمكانية تكرار العنصر نفسه عدداً من المرات).

نظرية 4:

إن عدد المرتبات k ف k عنصراً مع الإعادة من n عنصراً مختلفاً والذي سنرمز له بـ B_n^k يساوي:

$$B_n^k = n^k$$

البرهان:

يمكن اختيار العنصر الأول بـ n طريقة، وبما أننا نعيد العنصر إلى المجموعة الأصلية، فإن اختيار العنصر الثاني يتم أيضاً بـ n طريقة... واختيار العنصر k بـ n طريقة أيضاً، ومنه واستناداً إلى المفهوم الأساسي للتحليل التوافقي نستطيع أن نكتب بأن عدد الطرق لاختيار k عنصراً من n عنصراً يساوي:

$$B_n^k = \overbrace{n \cdot n \cdot n \cdot .....n}^{k \text{ مرة}} = n^k$$

مثال 13:

ليكن لدينا المجموعة التالية:

$$\Omega = \{0,1,2,3,4,5,6,7,8,9\}$$

ماهو عدد الأرقام التي يمكن تشكيلها من هذه المجموعة والمؤلفة من ثلاثة أعداد $k=3$ على أن يكون الاختيار مع الإعادة.

الحل:

إن اختيار العنصر الأول يتم بـ 10 طرق واختيار العنصر الثاني بعد إعادة الأول يتم أيضاً بـ 10 طرق، واختيار العنصر الثالث بعد إعادة العنصر الثاني يتم أيضاً بـ 10 طرق، وبالتالي فإن عدد الأرقام التي يمكن تشكيلها يحدد بـ:

$$B_n^k = B_{10}^3 = 10^3 = 1000$$

وهي مجموعة الأعداد من 000 إلى 999، وتساوي 1000 عدداً مستقلاً.

مثال 14:

بكم طريقة يمكننا اختيار ثلاثة أوراق من مجموعة ورق لعب تحوي /52/ ورقة، وذلك حسب الحالتين التاليتين.

أ . مع الإعادة.

ب . بدون إعادة.

أ . إذا أعدنا كل ورقة إلى مجموعة الورق قبل أن نقوم بسحب الورقة التالية، فإن عدد الطرق لاختيار أية ورقة هو 52 طريقة مختلفة، وبالتالي عدد الطرق الممكنة لسحب الورقات الثلاث يساوي:

$$B_n^k = (52)^3 = 140608 \text{ طريقة}$$

ب . أما إذا لم نعد الورقة المسحوبة إلى مجموعة الورق، فإن عدد الطرق لاختيار الورقة الأولى هو 52 طريقة، وعدد الطرق لاختيار الورقة الثانية هو 51 طريقة، وعدد الطرق لاختيار الورقة الثالثة هو 50 طريقة، وبالتالي:

$$A_{52}^3 = \frac{52!}{(52-3)!} = 52 \cdot 51 \cdot 50 = 1326000 \text{ طريقة}$$

0-5: المتوافقات:

لقد عالجنا فيما سبق عدداً من مسائل التحليل التوافقي، وكان لترتيب العناصر فيها حساب خاص، لتمييز المجموعات عن بعضها البعض، ولكن يوجد نوع آخر من المسائل لايكون لترتيب عناصر المجموعة أي تأثير، فمثلاً، إذا أردنا أن نختار 100 شخص من مجموعة مؤلفة من 1000 شخص للقيام بدراسة تجريبية، نرى أن ترتيب الأشخاص المختارين ليس مهماً في عملية الاختيار، ولكن المهم هو طبيعة هؤلاء الأشخاص. لفهم هذه الفكرة بشكل أوضح لنحاول دراسة المثال التالي:

مثال 15:

يود وكيل إحدى الشركات زيارة ثلاثة فقط من زبائنه الخمسة (دون أن يكون لترتيب الأشخاص في هذه الزيارات أية أهمية). فبكم طريقة يمكن أن يتم هذا الاختيار؟

الحل:

لو كان للترتيب أثر في الاختيار، فإننا نرى أن عدد الطرق الممكنة لهذا الاختيار هو عدد المرتبات مأخوذة ثلاثة فثلاثة من خمسة عناصر دون إعادة، ولكان عددها يساوي:

$$A_5^3 = \frac{5!}{(5-3)!} = \frac{5!}{2!} = 60 \text{ طريقة}$$

ولكن في معظم هذه المرتبات، تم اختيار نفس الزبائن الثلاثة بتبديل المواقع. فإذا رمزنا بـ M_1 , M_2, M_3, M_4, M_5، لهؤلاء الزبائن الخمسة على الترتيب، فإننا نجد عندها أن المتبادلات:

$$M_3M_2M_1 , M_3M_1M_2 , M_2M_3M_1 , M_2M_1M_3, M_1M_3M_2 , M_1M_2M_3$$

تعني أن الاختيار قد وقع على نفس الأشخاص، وبالتالي لحساب عدد الاختيارات المختلفة فإنه يجب تقسيم عدد المرتبات على عدد متبادلات ثلاثة عناصر، أي أن عدد الطرق المتوفرة لزيارة ثلاثة زبائن من أصل خمس بدون أهمية للترتيب يساوي:

$$\frac{A_5^3}{3!} = \frac{5!}{(5-3)!\cdot 3!} = 10$$

وهذه البدائل هي:

$$M_1M_2M_3 \quad M_1M_2M_4 \quad M_1M_2M_5 \quad M_1M_3M_4 \quad M_1M_3M_5$$
$$M_1M_4M_5 \quad M_2M_5M_4 \quad M_2M_3M_5 \quad M_2M_4M_5 \quad M_3M_4M_5$$

تعريف المتوافقات: هي المجموعات التي يمكن تشكيلها بحجم k فـ k عنصراً من أصل n عنصرا دون أن يكون لترتيب العناصر داخل المجموعة أي أهمية أو تأثير.

وسنميز بين نوعين من المتوافقات هما: المتوافقات بدون إعادة، والمتوافقات مع الإعادة.

*** المتوافقات بدون إعادة:**

وهي المتوافقات المشكلة k فـ k عنصراً من n عنصراً دون النظر للترتيب، وتكون عناصرها مختلفة عن بعضها (لايوجد تكرار فيها)، أي يتم التمييز بين المتوافقات (بين متوافقة وأخرى) باختلاف أحد العناصر في المتوافقة عن عناصر

المتوافقة الأخرى.

ونرمز لعدد المتوافقات التي يمكن تشكيلها بحجم k ف k عنصراً من n عنصراً بدون إعادة بالرمز C_n^k . وبذلك يمكننا تقديم النظرية التالية:

نظرية :

إن عدد المتوافقات المأخوذة k ف k عنصراً من n عنصراً، وبدون إعادة يساوي:

$$C_n^k = \frac{n!}{k!(n-k)!}$$

البرهان:

إن البرهان على هذه النظرية هو تعميم للمحاكمة التي أجريناها على المثال السابق وإن عدد المتوافقات يساوي عدد المرتبات مقسوماً على عدد المتبادلات أي أن:

$$C_n^k = \frac{A_n^k}{k!} = \frac{\frac{n!}{(n-k)!}}{k!}$$

وبذلك نجد أن:

$$C_n^k = \frac{n!}{k!(n-k)!} \qquad (5-0)$$

مثال 16:

احسب عدد المتوافقات التي يمكن تشكيلها بدون إعادة 3 ف 3 من الأحرف $\{a,b,c,d,e\}$. ثم شكل هذه المتوافقات في جدول خاص.

الحل: إن عدد المتوافقات المطلوب يساوي:

$$C_5^3 = \frac{5!}{3!(5-2)!} = \frac{5!}{3!\cdot 2!} = 10 \text{ متوافقات}$$

والمتوافقات الممكنة هي:

abc abd abe acd ace ade
bcd bce bde cde

مثال 17:

لتحديد نوع المرض الذي أصاب مجموعة مكونة من ستة شباب وإحدى عشرة فتاة، قررنا أخذ عينات دم من اثنين من الشباب ومن اثنتين من الفتيات. فبكم طريقة يمكن أن يتم هذا الاختيار؟

الحل:

1 . إن عدد الطرق لاختيار الشباب:

$$C_6^2 = \frac{6!}{2!(6-2)!} = \frac{6!}{2!4!} = 15$$

2 . إن عدد الطرق لاختيار الفتيات:

$$C_{11}^2 = \frac{11!}{2!(11-2)!} = \frac{11!}{2!\cdot 9!} = 55$$

ومن المفهوم الأساسي في التحليل التوافقي، نجد أن هنالك:

$15 \times 55 = 825$ طريقة لاختيار الشابين والفتاتين.

مثال 18:

تتألف لجنة من 12 عضواً. وحتى تجتمع هذه اللجنة، يجب أن يتوفر الحد الأدنى من نصابها القانوني والبالغ 8 أعضاء.

1 . بكم طريقة يمكن تأمين الحد الأدنى لنصابها القانوني؟

2 . بكم طريقة يمكن تأمين أي نصاب قانوني؟

الحل:

بما أنه ليس للترتيب أية أهمية، فإن هذه المسألة هي مسألة متوافقات، وبذلك نجد أن:

1 . إن عدد الطرق لاختيار 8 أعضاء كحد أدنى يساوي:

$$C_{12}^{8} = \frac{12!}{8!(12-8)!} = \frac{12!}{8!4!} = 495 \text{ طريقة}$$

أي أنه هنالك 495 طريقة لتأمين الحد الأدنى للنصاب القانوني.

2 . يمكن أن يتم تأمين أي نصاب قانوني بحضور 8 أعضاء أو 9 أو 10 أو 11 أو 12 عضواً. وبالتالي فإن عدد طرق تأمين أي نصاب قانوني هو:

$$C_{12}^{8} + C_{12}^{9} + C_{12}^{10} + C_{12}^{11} + C_{12}^{12} =$$
$$495 + 220 + 66 + 12 + 1 = 794 \text{ طريقة}$$

مثال 19:

يضم صف ثمانية طلاب وست طالبات. أوجد عدد طرق تكوين لجنة من خمسة طلاب إذا اشترط أن:

1 . تضم اللجنة خمسة عناصر دون تحديد الجنس.

2 . تضم اللجنة ثلاثة طلاب وطالبتان

3 . تضم اللجنة خمسة عناصر من الجنس نفسه (5 ذكور أو 5 إناث).

الحل:

1 . في الطلب الأول ليس للجنس أية أهمية وبالتالي فإن عدد الطرق لتشكيل تلك اللجنة يساوي عدد المتوافقات بحجم 5 من 14 ويساوي:

$$C_{14}^{5} = \frac{14!}{5! \cdot 9!} = 2002$$

أي هناك 2002 طريقة لاختيار لجنة مكونة من خمس طلبة دون تحديد الجنس.

2 . للإجابة على الطلب الثاني نحسب عدد المتوافقات، للذكور فقط بحجم 3 من 8 فنجد أن:

$$C_8^3 = \frac{8!}{3! \cdot 5!} = 56$$

طريقة لاختيار ثلاثة طلاب.

ثم نحسب عدد المتوافقات للإناث فقط بحجم 2 من 6 فنجد أن:

$$C_6^2 = \frac{6!}{2! \cdot 4!} = 15$$

طريقة لاختيار طالبتين فقط، وبالتالي يوجد:

$$C_8^3 C_2^6 = 56 \times 15 = 840$$

أي هناك 840 طريقة لاختيار ثلاثة طلاب وطالبتين لعضوية تلك اللجنة.

3 . للإجابة على الطلب الثالث نحسب عدد المتوافقات للجنة من الذكور فقط فنجد أن:

طريقة لاختيار خمسة طلاب ذكور: $C_8^5 = \frac{8!}{5! \cdot 3!} = 56$

وبالمقابل نحسب عدد المتوافقات للجنة من الإناث فقط، فنجد أن:

طريقة لاختيار خمس طالبات: $C_6^5 = \frac{6!}{5! \cdot 1!} = 6$

ومنهما نجد أنه يوجد لدينا:

$$C_8^5 + C_6^5 = 56 + 6 = 62$$

طريقة لاختيار خمسة عناصر من الجنس نفسه.

*** المتوافقات مع الإعادة:**

إن عملية تشكيل متوافقات k ف k عنصراً لـ n عنصراً مختلفاً في حالة الاختيار مع الإعادة، تشترط علينا عندما نقوم باختيار أي عنصر من عناصر المجموعة التي تحوي n عنصراً ليكون عنصراً في المتوافقة المطلوبة، أن نعيد ذلك العنصر إلى المجموعة الأصلية. وذلك لتعاد إمكانية اختياره في المتوافقات التالية:

فمثلاً لإيجاد جميع المتوافقات التي تحوي ($k=2$) عنصراً من مجموعة Ω تحوي 5 عناصر حيث أن $\Omega = \{a,b,c,d,e\}$، نقوم باختيار عنصر ما من Ω وليكن a ثم

نعيده إلى Ω نفسها ونختار عنصراً ثانياً لنضعه على يمين العنصر المسحوب سابقاً.
وبما أن a أعيد إلى Ω إذن يمكن أن نحصل على المتوافقات التالية المتعلقة بالسحب الأول للعنصر a :

$$aa, ab, ac, ad, ae$$

أما المتوافقات المتعلقة بالسحب الأول للعنصر b فهي:

$$bb, bc, bd, be$$

ملاحظة: بما أن الترتيب ليس له أهمية في المتوافقات لذلك حذفنا المتوافقة ba لأنها ووردت سابقاً على الشكل ab.

وهكذا ويمكننا أن نحدد جميع المتوافقات التي تحوي عنصرين ($k=2$) من ($n = 5$) في حالة الاختيار مع الإعادة كالتالي:

$$\begin{array}{ccccc} aa & ab & ac & ad & ae \\ & bb & bc & bd & be \\ & & cc & cd & ce \\ & & & dd & de \\ & & & & ee \end{array}$$

وعددها يساوي 15 متوافقة.

وبالأسلوب نفسه يمكننا أن نشكل جميع المتوافقات لثلاثة عناصر من مجموعة تحوي n عنصراً.

لذلك نرمز بـ C_{n+k-1}^{k} إلى عدد المتوافقات مع الإعادة، وتقدم النظرية التالية:

نظرية 6:

إن عدد المتوافقات k من k لـ n عنصراً مختلفاً مع إمكانية تكرار العنصر نفسه (أي مع الإعادة) يعطى بالدستور التالي:

$$C_{n+k-1}^{k} = \frac{(n+k-1)!}{k!(n-1)!} \qquad (6-0)$$

البرهان:

لحساب عدد المتوافقات k ف k عنصراً من n عنصراً مختلفاً في حالة الاختيار مع الإعادة، فإننا نستعين بعملية تشكيل المتوافقات من n عنصراً مختلفاً دون إعادة، وذلك لأنه يمكن اعتبار عملية تشكيل متوافقات k من k لـ n عنصراً مع الإعادة مطابقة لعملية تشكيل متوافقات k ف k لـ $(n+k-1)$ عنصراً بدون إعادة وسنوضح ذلك على الشكل التالي:

عندما نختار العنصر الأول من عناصر المجموعة التي تحوي n عنصراً ليكن العنصر i حيث إن $i=1,2,.....n$ فإن فكرة الإعادة تتطلب منا، وضع هذا العنصر مرةً أخرى ضمن العناصر n، ومن أجل هذا نجرده من رقمه السابق ونعيده إلى المجموعة بعد أن نعطيه رقماً جديداً هو الرقم $n+1$، وهذا يعنى أن العنصر ذا الرقم i الذي اختير سابقاً لم يعد موجوداً في المجموعة التي تحوي n عنصراً تحت الرقم i وإنما أصبح موجوداً فيها تحت الرقم $n+1$.

وعند اختيار العنصر الثاني نكون أمام مجموعة جديدة مؤلفة من $n+1$ عنصراً كان قد اختير منها العنصر ذو الرقم i ولم يعاد إليها؛ أي من مجموعة تحوي العناصر المتبقية فقط والتي أصبح عددها n عنصراً، وهكذا يمكننا أن نتابع هذا الأمر ونقول: إنناعند اختيار العنصر k نكون أمام مجموعة جديدة مؤلفة من $(n+k-1)$ عنصر كان قد اختير منها $(k-1)$ عنصراً ولم يعادوا إليها؛ أي تحوي العناصر المتبقية فقط والتي عددها n.

فإذا أمعنا النظر نلاحظ بأن عملية تشكيل متوافقات k من k لـ n عنصراً مختلفاً مع الإعادة ما هي إلا عبارة عن تشكيل متوافقات k ف k من $(n+k-1)$ عنصراً مختلفاً دون إعادة، ومنه نجد أن عدد هذه المتوافقات يساوي:

$$C_{n+k-1}^{k}=\frac{(n+k-1)!}{k!(n-1)!} \qquad (7-0)$$

مثال 20:

ماهو عدد العينات المكونة من خمسة طلاب والتي يمكن سحبها مع الإعادة من مجموعة مؤلفة من 10 طلاب:

إن العدد المطلوب هو عدد المتوافقات مع الإعادة ويساوي:

$$C_{10+5-1}^{5}=\frac{(10+5-1)!}{5!(10-1)!}=2002 \text{ عينة}$$

مثال 21:

ماهو عدد الجداءات المؤلفة من ثلاثة عوامل إذا كان كل عامل هو أحد الأرقام $(1,2,.....9)$، وكان بالإمكان تكرار العامل نفسه بالجداء الواحد:

إن ذلك العدد يساوي عدد المتوافقات مع الإعادة ويساوي:

$$C_{9+3-1}^{3}=\frac{(9+3-1)!}{3!(9-1)!}=165$$

ملاحظة:

يمكن أن ننظر إلى عدد المتوافقات بدون إعادة المأخوذة k فـ k من n عنصراً على أنه عدد الطرق التي يمكن بها أن نجزئ مجموعة تحوي n عنصراً إلى مجموعات جزئية تحوي k عنصراً و $(n-k)$ عنصراً.

وأخيراً لابد لنا أن تشير إلى أن هنالك مسألة أعم مما سبق وهي عدد المجموعات الجزئية لمجموعة معطاة A تحوي n عنصراً. فهل يمكن تعداد الطرق الممكنة للقيام بهذا العمل؟ فمثلاً ، بكم طريقة يمكن لتسعة أشخاص أن ينقسموا إلى ثلاث لجان علماً بأن هذه اللجان مؤلفة من أربعة وثلاثة واثنان على الترتيب. للإجابة على هذا السؤال، لابد لنا من تعريف تجزئة مجموعة ما. وهذا ماسندرسه في الفقرة التالية:

0-6: تجزئة مجموعة:

نعرف تجزئة مجموعة A تحوي n عنصراً بأنه تقسيمها إلى k مجموعة جزئية كل منها تحوي $n_1, n_2, ...n_k$ عنصراً على الترتيب، وبحيث يكون:

$$n_1 + n_2 + ... + n_k = n$$

ولنحسب الآن عدد الطرق الممكنة لتجزئة لمجموعة A تضم n عنصراً إلى مجموعة جزئية $A_1 , A_2 , ...A_k$ ، حيث إن A_1 تحوي n_1 عنصراً، و A_2 تحوي n_2 عنصراً، ..... و A_k تحوي n_k عنصراً، وبحيث يكون لدينا:

$$n_1 + n_2 + ... + n_k = n$$

فنجد أنه لدينا أولاً: C_n^{n1} طريقة لاختيار n_1 عنصر في A_1 من أصل الـ n عنصراً الموجودة في A، وعندما يتم اختيار A_1 ، فإنه يبقى لدينا (n-n) عنصراً في A ، وعندها يكون لدينا: $C_{n-n_1}^{n_2}$ طريقة لاختيار n_2 عنصراً في A_2 من أصل الـ$(n - n_1)$ عنصراً المتبقية.

وبشكل مشابه يكون لدينا: $C_{n-n_1-n_2}^{n_3}$ طريقة لاختيار n_3 عنصراً للمجموعة A_3 من $(n - n_1 - n_2)$ عنصراً المتبقية ..... وهكذا.

ونستمر بهذه المحاكمة حتى يتم اختيار جميع المجموعات الجزئية.

وأخيراً نستنتج النتيجة العامة من المفهوم الأساسي في التحليل التوافقي التي تلخص بالنظرية التالية:

نظرية 7:

لتكن A مجموعة ما تحوي n عنصراً، وليكن $n_k,.....n_2,n_1$ أعداداً صحيحة موجبة وبحيث:

$$n_1 + n_2 + ... + n_k = n$$

عندئذ يكون لدينا:

$$C_n^{n_1} \cdot C_{n-n_1}^{n_2} \cdot C_{n-n_1-n_2}^{n_3} .....C_{n-n_1-n_2.....n_{k-1}}^{n_k} = P \quad (8-0)$$

طريقة لتجزئة المجموعة A إلى مجموعات $A_1 , A_2 , ...A_k$ حيث أن:

A_1 مجموعة جزئية من A تحوي n_1 عنصراً.

A_2 مجموعة جزئية من A تحوي n_2 عنصراً.

.....

A_k مجموعة جزئية من A تحوي n_m عنصراً.

إن البرهان على هذه النظرية هو تعميم للمحاكمة السابقة.

ونشير أخيراً إلى أنه يمكن تبسيط العلاقة الواردة في النظرية السابقة بأن نبرهن بسهولة على أن:

$$C_n^{n_1} \cdot C_{n-n_1}^{n_2} \cdot C_{n-n_1-n_2}^{n_3} .....C_{n-n_1-n_{2-n_{k-1}}}^{n_k} = P_n^{n_1,n_2.....n_k}$$

$$P_n^{n_1,n_2.....n_k} = \frac{n!}{n_1!n_2!n_3!....n_k!} \qquad (9-0)$$

حيث أن: $n_1 + n_2 + n_3 + ... + n_k = n$

مثال 22:

يحتوي صندوق سبع قطع خشبية مرقمة من 1 إلى 7. بكم طريقة يمكننا اختيار، أولاً قطعتين من الصندوق، ومن ثم ثلاث قطع وأخيراً قطعتين.

الحل:

يجب أن نحسب عدد تجزئة المجموعة A والتي تحتوي $n=7$ قطع خشبية إلى ثلاث مجموعات جزئية A_1 والتي تحتوي على قطعتين $n_1 = 2$ ومن ثم A_2 والتي تحتوي على ثلاث قطع $n_2 = 3$ ، وأخيراً A_3 والتي تحتوي على قطعتين $n_3 = 2$، وبما أن مجموع الأعداد $n_1 + n_2 + n_3 = n$ فإنه:

لتحديد متوافقات المجموعة الجزئية الأولى A_1 نجد أن $C_n^{n1} = C_7^2$

لتحديد متوافقات المجموعة الجزئية الأولى A_2 نجد أن $C_n^{n2} = C_5^3$

لتحديد متوافقات المجموعة الجزئية الأولى A_3 نجد أن $C_n^{n3} = C_2^2$

ومنه نجد أن عدد الطرق لتحديد المجموعات الجزئية الثلاث هو:

$$C_7^2 \cdot C_5^3 \cdot C_2^2 = \frac{7!}{2!\,5!} \cdot \frac{5!}{3!\,2!} \cdot \frac{2!}{2!\,0!} = \frac{7!}{2!\,3!\,2!} = 210 \quad \text{طريقة}$$

كما نلاحظ أنه يمكننا حساب عدد طرق التجزئة من العلاقة (0–9) التالية:

$$P_n^{n1 \cdot n2 \cdot n3} = \frac{n!}{n_1!n_2!n_3} = \frac{7!}{2\,!3\,!2\,!} = 210 \quad \text{طريقة}$$

مثال 23:

بكم طريقة يمكن تشكيل ثلاث لجان مؤلفة من 4، 3، 2 أشخاص على الترتيب، وذلك من مجموعة مؤلفة من تسعة أشخاص.

الحل:

إن عدد الطرق لتشكيل اللجان الثلاث هو عدد طرق تجزئة المجموعة A، لأن عدد عناصر A يساوي مجموع عناصر المجموعات الجزئية المطلوبة، وبما أنه يراد تشكيل ثلاث مجموعات جزئية هي A_1 , A_2 , A_3 تحوي الأولى 4 عناصر، والثانية 3 عناصر، والثالثة عنصرين فإن عدد الطرق الممكنة يساوي عدد طرق التجزئة ويساوي:

$$P_n^{n1 \cdot n2 \cdot n3} = \frac{9\,!}{4\,!3\,!2\,!} = 1260 \quad \text{طريقة}$$

مثال 24:

إذا كان لدى فرقة الكشافة 9 أعلام: 3 منها حمراء، 2 منها سوداء، 4 خضراء. فما هو عدد الإشارات التي يمكن تركيبها من استعمال كل هذه الأعلام.

$$P_9^{2,3,4} = \frac{9\,!}{4!\,3!\,2!} = 1260 \quad \text{إشارة}$$

0–7: ثنائي الحد:

نصادف في الرياضيات، تعابير من الشكل $(a+b)^n$ أو من $(a+b+c)^n$ حيث إن a, b, c أعداد حقيقية وحيث إن n عدد صحيح موجب. وبما أن هذه التعابير تلعب دوراً أساسياً في دراسة الاحتمالات. لذلك سنحاول كتابتها على شكل مجموع حدود كما يلي.

$$\sum_{k=0}^{n} f_k = f_0 + f_1 + f_2 + ... + f_n$$

ويقرأ الجانب اليساري كما يلي " مجموع f_k من $k = 0$ إلى $k = n$ ".

نظرية 8:

إذا كان a , b عددان حقيقيان، وكان n عدداً صحيحاً موجباً، عندئذ يمكن أن نكتب ذي الحدين على الشكل التالي:

$$(a+b)^n = \sum_{k=0}^{n} C_n^k a^{n-k} b^k =$$

$$C_n^0 a^n b^0 + C_n^1 a^{n-1} b + ... + C_n^k \cdot a^{n-k} \cdot b^k + ... + C_n^n a^0 b^n \qquad (10-0)$$

$$(a+b)^n = a^n + na^{n-1}b + \frac{n(n-1)}{1 \cdot 2} a^{n-2} b^2 + ...$$

$$+ \frac{n(n-1)....(n-k+1)}{k!} a^{n-k} b^k + .. + b^n$$

البرهان:

لنكتب الجداء السابق كما يلي:

$$(a+b)^n = (a+b)(a+b).....(a+b)$$

حيث يوجد n جداء في الطرف الأيمن، وكل حد فيه هو $(a+b)$. لحساب الجداء السابق نقوم بضرب كل حد (وهو هنا $a+b$) بجميع الحدود الباقية ثم نجمع جميع الجداءات الناتجة فنجد مثلاً أن:

$$(a+b)^2 = a^2 + ab + ba + b^2 = a^2 + 2ab + b^2$$

$$(a+b)^3 = (a+b)(a+b)^2 =$$

$$= a^3 + a^2 b + aba + ab^2 + ba^2 + bab + b^2 a + b^3 =$$

$$= a^3 + 3a^2 b + 3b^2 + b^3$$

وهنا نلاحظ أن عناصر الحد العام في المجموع الأيمن يمكن أن تكتب على الشكل $a^{n-k} \cdot b^k$. وإن عدد مرات تكرار كل حد من هذا النوع يساوي عدد المرات التي يمكن أن يظهر فيها b تماماً k مرة من أصل n جداءاً (قوساً). وهو يساوي عدد

المتوافقات بدون إعادة C_n^k. وبذلك يمكننا كتابة الحد العام في المجموع السابق كمايلي: $C_n^k \cdot a^{n-k} \cdot b^k$

ونكتب ثنائي الحد على الشكل التالي:

$$(a+b)^n = \sum_{k=0}^{n} C_n^k \cdot a^{n-k} \cdot b^k \qquad (11-0)$$

وأخيراً نلاحظ أن منشور ثنائي الحد $(a+b)^n$ يتصف بمايلي:

1 . إن عدد الحدود يساوي $(n+1)$ حداً.

2 . إن مجموع قوى a,b في كل حد يساوي n.

3 . إن أس a يتناقص حداً بعد حد من n إلى 0، في حين أن أس b يتزايد حداً بعد حد من 0 إلى n.

4 . إن معامل أي حد هو C_n^k حيث أن k تكون أس b.

5 . إن معاملات الحدود المتساوية البعد عن طرفي المنشور (الحدين الأول والأخير) تكون متساوية. وذلك لأن:

$$C_n^k = \frac{n!}{k!(n-k)} = \frac{n!}{(n-k)![n-(n-k)]!} = C_n^{n-k}$$

لذلك يمكننا كتابة الحد العام على الشكل $C_n^{n-k} \cdot a^k \cdot b^{n-k}$

مثال 25:

احسب مفكوك التركيب التالي:

$$(x+y)^5$$

الحل:

من منشور ثنائي الحد، نجد:

$$(x+y)^5 = \sum_{k=0}^{5} C_5^k x^{5-k} y^k$$

$$C_5^0 x^5 y^0 + C_5^1 x^4 y^1 + C_5^2 x^3 y^2 + C_5^3 x^2 y^3 + C_5^4 x^1 y^5 + C_5^5 x^0 y^5$$

$$= x^5 + 5x^4 y + 10x^3 y^2 + 10x^2 y^3 + 5xy^4 + y^5$$

مثال 26:

ما هو معامل الحد $x^2 y^4$ في منشور $(x+y)^7$؟

إن الحد العام في المنشور هو:

$$C_7^k x^{7-k} y^k$$

يظهر الحد $x^2 y^4$ عندما $k=4$ والمعامل هو:

$$C_7^4 = \frac{7!}{4!(7-3)!} = 35$$

مثال 27:

لدينا، كحالة خاصة من منشور ثنائي الحد، النتيجة التالية:

$$2^n = (1+1)^n = \sum_{k=0}^{n} C_n^k 1^{n-k} 1^k$$

$$2^n = \sum_{k=0}^{n} C_n^k = C_n^0 + C_n^1 + C_n^2 + ... + C_n^k + ...C_n^n \quad (12-0)$$

يمكن الاستعانة بهذا القانون في المجموعات. إذ نعلم أن C_n^k هو عدد المجموعات الجزئية التي يمكن تشكيلها من k عنصراً من مجموعة تحوي n عنصراً، وبذلك نجد مجموع أعداد هذه المجموعات الجزئية الممكنة التشكيل بجميع الأحجام من 0 إلى n يساوي:

$$C_n^0 + C_n^1 + C_n^2 + ...C_n^n = 2^n \qquad (13-0)$$

وهو يمثل العدد الكلي للمجموعات الجزئية التي يمكن تشكيلها من مجموعة عدد عناصرها n. وهذا العدد الكلي يساوي 2^n.

0-8: كثير الحدود:

سنعمم منشور ثنائي الحد لحساب التركيب الذي من الشكل $(a+b+c)^n$. أو من الشكل $(a_1+a_2+a_3+.....+a_m)^n$.

نظرية 9:

إذا كانت $a_1.a_2.....a_m$ أعداداً حقيقية وكان n عدداً صحيحاً موجباً، عندئذ يكون لدينا:

$$(a_1+a_2+.....+a_m)^n=\sum p_n^{n1,n2,.....nm}\,a_1^{n1}a_2^{n2}.....a_m^{nm} \qquad (14-0)$$

حيث إن: $n_1+n_2+n_3+....+n_m=n$

البرهان:

نلاحظ أنه يوجد n جداءاً في الطرف الأيمن وكل حد فيه هو $(a_1+a_2+.....+a_m)$، لحساب ذلك الجداء نضرب كل حد بجميع الحدود الباقية ثم نجمع جميع الجداءات الناتجة فنحصل على المطلوب، وبذلك نجد أن الحد العام في الجداء هو من الشكل:

$$a_1^{n1}a_2^{n2}.....a_m^{nm}$$

حيث أن:

$$n_1+n_2+.....+n_m=n$$

وأن معامل هذا الحد هو عدد طرق تجزئة مجموعة تحوي n عنصراً إلى مجموعات جزئية تحوي على الترتيب $n_1,n_2,...,n_m$ عنصراً. ونحن نعلم، أن هذا المعامل هو:

$$p_n^{n1,n2,.....nm}=\frac{n!}{n_1!n_2!....n_m!}$$

وبذلك يكون قد تم البرهان.

مثال 28:

ماهو معامل الحد ab^2cd^2 في منشور التركيب التالي:

$$(a+b+c+d)^6$$

الحل:

في هذا المثال لدينا $n=6$. ونحن نتساءل عن معامل الحد الموافق لـ $n_1=1$ $n_2=2$, $n_3=1$, $n_4=2$. إن هذا المعامل هو إذن:

$$P_6^{1,2,1,2}=\frac{6!}{1!2!1!2!}=180$$

0-9: تطبيقات ثنائي الحد:

يمكن تعميم منشور ثنائي الحد إلى الحالة التي لايكون فيها n عدداً صحيحاً موجباً. نعلم أنه عندما تكون n عدداً صحيحاً موجباً ويكون a,b أعداداً حقيقية، فإن منشور $(a+b)^n$ هو:

$$(a+b)^n=a^n+na^{n-1}b+\frac{n(n-1)}{2!}a^{n-2}b^2+$$
$$+\frac{n(n-1)(n-2)}{3!}a^{n-3}b^3+.....+b^n$$

إن الطرف الأيمن هو مجموع منتهٍ ويحوي $(n+1)$ حداً، ويلاحظ أنه يمكننا كتابة ثنائي الحد كمايلي:

$$a+b=a\left(1+\frac{b}{a}\right)$$

وبالتالي:

$$(a+b)^n=a^n\left(1+\frac{b}{a}\right)^n$$

وبالتالي فإنه لمعرفة منشور $(a+b)^n$ يكفي معرفة منشور $(1+b/a)^n$ وضرب الناتج بـ a^n.

للتبسيط نضع $b/a=x$ ونأخذ منشور ثنائي الحد التالي:

$$(1+x)^n = 1 + nx + \frac{n(n-1)}{2!}x^2 +$$
$$+ \frac{n(n-1)(n-2)}{3!}x^3 + ... + x^n \qquad (15-0)$$

ولقد برهنا على صحة هذه النتيجة من أجل n عدد صحيح موجب. ولنتساءل ماذا يحصل لو كانت n عدداً حقيقياً أو غير موجب؟ فهل يبقى لمنشور ثنائي الحد معنى، عندما $n = 1/2$ أو $n = -1$.

من أجل $n = 1/2$، يكون منشور ثنائي الحد $(1+x)^n$ هو:

$$(1+x)^{1/2} = 1 + 1/2\,x + \frac{(1/2)(-1/2)}{2!}x^2 + \frac{1/2(-1/2)(-3/2)}{3!}x^3 + ...$$

ومن أجل $n = -1$ يكون منشور ثنائي الحد $(1+x)^{-1}$ هو:

$$(1+x)^{-1} = 1 + (-1)x + \frac{(-1)(-2)}{2!}x^2 + \frac{(-1)(-2)(-3)}{3!}x^3 + ...$$
$$= 1 - x + x^2 - x^3 + ... \qquad (16-0)$$

وهكذا نرى أن منشور ثنائي الحد $(1+x)^{1/2}$ و $(1+x)^{-1}$ يأخذ شكل سلسلة لانهائية ولذلك سنسميه منشوراً لثنائي الحد $(1+x)^{1/2}$ أو غيره. لن نسهب هنا في ذكر خواص مثل هذه السلاسل غير المنتهية ولكننا نشير إلى أنه إذا كانت $-1 < x < +1$ فإن السلسلة التي تعطي منشور $(1+x)^n$ تكون متقاربة ويكون لها مجموع محدّد (حتى لو كانت n عدداً صحيحاً غير موجب) وهذا يعطينا وسيلة مفيدة لحساب منشور $(1+x)^n$.

إن الشرط الذي نضعه وهو أن تكون $-1 < x < +1$ يضمن لنا بأن الحدود في المجموع غير المنتهي والتي من الشكل:

$$\frac{n(n-1).....(n-k+1)}{k!}x^k$$

هي صغيرة عندما تكون k كبيرة.

مثال 29:

1 . احسب $(1{,}10)^{1/2}$

2 . احسب $(1{,}05)^{1/5}$

3 . احسب $\left(\sqrt{5}\right)$

الحل:

1 . لحساب $(1{,}10)^{1/2}$. نضع $x = 1/10$ ونحسب حدود منشور $(1+x)^{1/2}$.

$$\left(1+\frac{1}{10}\right)^{1/2} = 1+\frac{1}{2}\left(\frac{1}{10}\right)+\frac{1}{2!}\left(\frac{1}{2}\right)\left(-\frac{1}{2}\right)\left(\frac{1}{10}\right)^2 +$$

$$+\frac{1}{3!}\left(\frac{1}{2}\right)\left(-\frac{1}{2}\right)\left(\frac{-3}{2}\right)\left(\frac{1}{10}\right)^3 + \ldots.$$

$$= 1 + 0{,}05 - 0{,}00125 + 0{,}0000625 - \ldots.$$

$$= 1{,}04881$$

ومن الواضح أنه يمكن حساب $(1+1/10)^{1/2}$ حتى أرقام عشرية كثيرة بعد الفاصلة.

2 . لحساب $(1.05)^{\frac{1}{5}}$ تضع $x = \frac{1}{20} = 0{,}05$ فنجد أن:

$$(1{,}05)^{1/5} = \left(1+\frac{1}{20}\right)^{1/5} = 1+\frac{1}{5}\left(\frac{1}{20}\right)+\frac{1}{2!}\left(\frac{1}{5}\right)\left(-\frac{4}{5}\right)\left(\frac{1}{20}\right)^2 + \ldots.$$

$$= 1 + 0{,}01 - 0{,}0002 + \ldots.$$

$$= 1{,}0098$$

3 . لحساب $\sqrt{5}$ نكتب $5 = 4 + 1$ وبالتالي:

$$\sqrt{5} = \sqrt{4+1} = \sqrt{4(1+1/4)}$$

وبالتالي:

$$\sqrt{5} = 2\sqrt{1+1/4}$$

ويمكن حساب $(1+1/4)^{1/2}$ من منشور ثنائي الحد كمايلي:

$$\left(1+\frac{1}{4}\right)^{1/2}=1+\frac{1}{2}\left(\frac{1}{4}\right)+1+\frac{1}{2!}\left(\frac{1}{4}\right)\left(-\frac{1}{2}\right)\left(\frac{1}{4}\right)2+$$

$$+\frac{1}{3!}\left(\frac{1}{2}\right)\left(-\frac{1}{2}\right)\left(-\frac{3}{2}\right)\left(\frac{1}{4}\right)^3+....$$

$$=1+0{,}125-0{,}007801+0{,}000975-....$$

$$=1{,}118174$$

وبالتالي: $\sqrt{5}=2(1{,}118174)=2{,}236...$

مثال 30:

السلسلة الهندسية.

إن منشور ثنائي الحد $(1-x)^{-1}$ هو:

$$(1-x)^{-1}=1+(-1)(-x)+\frac{(-1)(-2)}{2\,!}(-x)^2+$$

$$+\frac{(-1)(-2)(-3)}{3\,!}(-x)^3+....$$

$$=1+x+x^2+x^3+....$$

إن هذه السلسلة التي تظهر في كثير من التطبيقات تأخذ اسماً خاصاً هو السلسلة الهندسية وهي متقاربة عندما تكون $-1<x<+1$ وإن مجموعها هو:

$$\frac{1}{1-x}=(1-x)^{-1}$$

وبذلك يمكننا أن نكتب:

$$1+x+x^2+x^3+x^4+.....=\frac{1}{1-x}$$

مثال 31:

احسب مجموع السلسلة الهندسية:

$$1+x+x^2+x^3+....$$

وذلك عندما:

$$x=\frac{2}{3}, \qquad x=\frac{1}{3}, \qquad x=\frac{1}{2}$$

الحل:

1 . نجد عندما $x=\frac{1}{2}$ أن:

$$1+1/2+(1/2)^2+(1/2)^3+....=1+0,5+0,025+....$$

ولكننا نعلم أن مجموع هذه السلسلة يساوي:

$$\frac{1}{1-1/2}=2$$

وبالتالي فإن مجموع هذه السلسلة عندما $x=1/2$ هو 2.

2 . عندما $x=1/3$ يكون المجموع:

$$\frac{1}{1-1/3}=3/2$$

3 . عندما $x=2/3$ يكون المجموع:

$$\frac{1}{1-2/3}=3$$

وأخيراً نورد منشور ثنائي الحد ذو الأس السالب والذي يعطى بالعلاقة التالية:

$$(1-x)^{-n}=\sum_{r=0}^{\infty}C_{n+r-1}^{n-1}x^r \qquad (17-0)$$

كما إن:

$$\lim_{x\to\infty}\left(1+\frac{1}{x}\right)^x=e\approx 2,718..... \qquad (18-0)$$

تمرينات

1 . ليكن لدينا مجموعة الأرقام $\{9,8,6,4,3,1\}$ ، بكم طريقة يمكن تشكيل أعداداً من ثلاثة أرقام ضمن الشروط الآتية ؟

أ . عدم وجود أي رقم مكرر ضمن التشكيلة.

ب . إذا كانت الأعداد أصغر من العدد 500.

جـ . إذا كانت الأعداد فردية.

2 . لدينا ثمانية رجال وسبع نساء. ونريد أن نشكل لجنة مؤلفة من 7 أعضاء، 4 أعضاء من الرجال و3 من النساء . فما هي عدد الطرق الممكن تشكيل اللجنة منها.

3 . بفرض أنه لدينا المجموعة $A=\{a,b,c,d,e\}$. يراد تشكيل مجموعات مؤلفة من $k=2$ علماً بأن كل عنصر يعاد إلى المجموعة الأصلية (الترتيب ليس له أهمية).

4 . ليكن لدينا حجري نرد، إذا رميناهما معاً، ماهو عدد النتائج التي يمكن ان تظهر خلال سير العملية.

5 . بكم طريقة يمكن أن نختار 13 ورقة لعب من مجموعة الأوراق 52، باعتبار الأوراق تختلف بأنواعها فقط.

6 . بكم طريقة يمكن توزيع 8 أشخاص على الوظائف التالية: رئيس دائرة . رئيس ديوان . سكرتير مدير . سكرتير مجلس . أمين صندوق.

7 . ماهو عدد المتبادلات التي يمكن تكوينها من الأحرف $a,b,c,\alpha,\beta,\gamma$ والتي تبدأ وتنتهي بحرفين يونانيين.

8 . لدينا 20 شخصاً، نريد تقسيمهم إلى ثلاث مجموعات: الأولى مكونة من عشرة أشخاص ومجموعتين ثانية وثالثة كلاً منهما مكونة من /5/ أشخاص. بكم طريقة يمكن أن يتم هذا؟

9 . لنفترض أنه لدينا خمس طرق تصل المدينتين A, B، وثلاث طرق تصل المدينتين B, C فبكم طريق مختلف يمكن أن نصل المدينتين A, C؟

10 . برهن صحة مايلي:

1 . $C_n^k = C_n^{n-k}$

2 . $n^2 = C_n^2 + C_{n+1}^2$

4 . $C_{n+2}^{k+2} = C_n^{k+2} + 2C_n^{k+1} + C_n^k$

5 . $C_n^k \cdot C_n^r = C_n^r \cdot C_{n-r}^{k-r}$

حيث أن $r \le k \le n$

11 . سنختار عشرة أشخاص من مجموعة مكونة من عشرة رجال وعشر نساء.

أ . ماهو عدد طرق اختيار العشرة أشخاص ؟

ب . ماهو عدد الطرق إذا اشترط أن يكون عدد الرجال أكثر من عدد النساء؟

ج . إذا اشترط أن يكون 8 على الأقل من النساء، فبكم طريقة يمكن اختيار العشرة أشخاص ؟

12 . أوجد معامل x^3y^4 في منشور:

(1) $(x+y)^7$

(2) $(x^3+y^2)^3$

(3) $(x-y)^7$

(4) $(x+y^4)^4$

الفصل الأول

احتمالات الحوادث

ويشمل الفقرات التالية:

1 . أنواع الحوادث.
2 . تعاريف الاحتمال.
3 . قواعد جمع الاحتمالات.
4 . قواعد ضرب الاحتمالات.
5 . قاعدة الاحتمال الكلي وقانون بييز.
6 . تمرينات.

الفصل الأول

احتمالات الحوادث

إن الاحتمالات وثيقة الصلة بالإحصاء، وحتى نظهر بعض خصائص هذه الصلة نقوم بدراسة احتمالات الحوادث، وذلك بأسلوب رياضي وبطريقة إحصائية. وقبل أن نتعرض إلى الحوادث وأنواعها نقوم بتعريف تعريف الاحتمال بما يلي:

المفهوم العام للاحتمال: إن الاحتمال هو عدد حقيقي يعبر عن إمكانية تحقق حادث ما (أو عدم تحققه).

إذن الاحتمال ليس بديلاً من البدائل الممكنة كما يفهمه عامة الناس، بل هو عدد حقيقي يعبر عن إمكانية تحقق الحادث المدروس، أو البديل المقصود.

وبما أن الاحتمالات مرتبطة بالحوادث فلا بد من التعرض للحوادث وأنواعها، لأن ذلك يؤثر على الاحتمالات نفسها.

1-1: أنواع الحوادث:

عند إجراء تجربة فإن نتائجها تشكل مجموعة مؤلفة من عدة حوادث أو بدائل ممكنة، ولكل منها احتمال مقابل. فمثلاً، عند إلقاء حجر النرد مرة واحدة فإنه قد يظهر لنا أحد الحوادث التالية:

الحادث الأول: ظهور العدد 1 ونرمز له بـ A_1

الحادث الثاني: ظهور العدد 2 ونرمز له بـ A_2

الحادث الثالث: ظهور العدد 3 ونرمز له بـ A_3

الحادث الرابع: ظهور العدد 4 ونرمز له بـ A_4

الحادث الخامس: ظهور العدد 5 ونرمز له بـ A_5

الحادث السادس: ظهور العدد 6 ونرمز له بـ A_6

الحادث السابع: ظهور أي عدد فردي ونرمز له بـ A_7

الحادث الثامن: ظهور أي عدد زوجي ونرمز له بـ A_8

الحادث التاسع: ظهور أي عدد أقل من 5 ونرمز له بـ A_9

.....

وانطلاقاً من هذا المثال البسيط يمكننا أن نميز أنواع الحوادث التالية:

* **الحوادث البسيطة:** هي الحوادث غير القابلة للتجزئة، ويمثل كل منها إحدى إمكانيات التجربة أو الظاهرة المدروسة. ونذكر كمثال على ذلك الحوادث الست الأولى في إلقاء حجر النرد في المثال السابق.

* **الحوادث المركبة:** وهي الحوادث التي يتألف كل منها من حوادث بسيطة، مثل حادث حصولنا على أي عدد فردي أو على أي عدد زوجي أو على عدد أقل من 3 ...إلخ.

الحوادث المتكافئة: وهي الحوادث (البسيطة أو المركبة) التي تكون إمكانية تحققها متساوية (أي يكون احتمال تحققها متساوٍ). ونذكر كمثال على ذلك أوجه النرد، أو وجهي قطعة نقود صحيحة، أو نتيجة الولادة، أو ظهور أي رقم طبيعي على عجلات الرهان (اليانصيب).

وكمثال على الحوادث غير المتكافئة نذكر حادث اختيار سيارة لشرائها أو استئجارها، والحادثتين السابقتين A_7 و A_9. حيث يتأثر احتمال اختيار السيارة بنوعها وسعرها وشكلها، كما يتأثر احتمال الحادثين A_7 و A_9 بعدد الحوادث البسيطة الداخلة فيه. لذلك فهي حوادث غير متكافئة.

الحوادث المتنافية: هي الحوادث التي إذا تحقق أحدها نفى أو منع تحقق الحوادث الأخرى، وكمثال على الحوادث المتنافية نذكر أيضاً أوجه النرد، ووجهي القطعة النقدية... إلخ. وكمثال على الحوادث غير المتنافية نذكر الحادثين A_8 و A_9، فظهور أي عدد أقل من 5، (أي تحقق A_9) لا ينفي تحقق A_8 الذي يعني ظهور أي عدد زوجي.

الحوادث المستقلة: هي الحوادث التي إذا تحقق أحدها لا يؤثر على احتمال تحقق الحوادث الأخرى. وكمثال على ذلك نذكر الحوادث الناتجة عن سحب كرة معينة من

صندوق ثم إعادتها إلى عدة مرات لتتعرض للسحب ثانيةً في كل مرة. وهنا من الواضح أن إمكانية احتمال سحب الكرة المعنية يبقى ثابتاً؛ أي أنه لا يتأثر بنتيجة السحب ثم الإعادة. ولكن الأمر على عكس ذلك إذا كان السحب بدون إعادة. فإذا كان في الصندوق 5 كرات، وكان السحب بدون إعادة فإن احتمال سحب كرة معينة في المرة الأولى يساوي $\frac{1}{5}$ ، وفي المرة الثانية (إذا لم تسحب) يساوي $\frac{1}{4}$. وفي المرة الثالثة (إذا لم تسحب) يساوي $\frac{1}{3}$ ...إلخ.

أي أن احتمالات الحوادث المتتالية في حالة السحب بدون إعادة تتغير وتتأثر بنتيجة الحادث أو الحوادث التي تسبقها، لذلك فإن هذه الحوادث هي حوادث غير مستقلة أو هي حوادث مترابطة.

الحوادث المتكاملة: هي جملة الحوادث المتنافية التي تشكل مجموعة الحوادث الممكنة للتجربة. وكمثال على ذلك نذكر:

. الأوجه المختلفة لحجر النرد.

. وجهي قطعة النقود.

. الأرقام من 0 إلى 9 على عجلة الرهان.

. عدد أفراد الأسرة من 1 إلى 15 شخصاً أو أكثر.

.....

ونرمز لمجموعة الحوادث المتكاملة المؤلفة من n حادثاً متنافياً بالرمز Ω ونكتب:

$$\Omega = \{A_1, A_2, A_3, ..., A_n\}$$

الحادث المستحيل: هو الحادث الذي لا يمكن له أن يتحقق ونرمز له بالرمز ϕ.

الحادث الأكيد: هو الحادث (البسيط أو المركب) الذي لابد أن يتحقق عند إجراء التجربة. وخير مثال على ذلك الحادث التالي:

. تحقق أي من حوادث المجموعة المتكاملة.

لذلك نرمز للحادث الأكيد أيضاً بالرمز Ω.

الحادث المتمم: هو الحادث الذي يتحقق إذا لم يتحقق الحادث الأصلي. وإذا كان A حادثاً ما فإننا نرمز لمتممه بالرمز $\overline{A}$. ويكون لدينا:

$$\overline{A} = \Omega - A$$

اجتماع حادثين *A* و*B*: ونرمز له بـ $A \cup B$ وهو عبارة عن حادث ثالث يتحقق إذا تحقق أي من الحادثين A أو B.

تقاطع حادثين *A* و*B*: ونرمز له بـ $A \cap B$ وهو عبارة عن حادث ثالث D يتحقق إذا تحقق كلا الحادثين A و B معاً.

فرق حادثين *A* و*B*: هو حادث ثالث ونرمز له بـ $A - B$. وهو يتحقق إذا تحقق A ولم يتحقق B.

1-2: تعاريف الاحتمال:

للاحتمال عدة تعاريف أهمها وأبسطها التعاريف التالية:

1 . التعريف النظري: يعرف احتمال تحقق حادث A بأنه نسبة عدد الحالات الملائمة لـ A على عدد الحالات الممكنة للتجربة؛ أي أن:

$$\text{احتمال تحقق } A = \frac{\text{عدد الحالات الملائمة لـ } A}{\text{عدد الحالات الممكنة}}$$

ويشترط هنا أن تكون الحالات الممكنة والملائمة حوادث متكافئة ومتنافية، كما يشترط أن تكون الحالات الملائمة مجموعة جزئية من الحالات الممكنة، فإذا رمزنا لعدد الحالات الملائمة لـ A بالرمز M، ولعدد الحالات الممكنة والمتكافئة بالرمز N، ولاحتمال تحقق الحادث A بالرمز $P(A)$، فإنه يكون لدينا العلاقة الحسابية التالية:

$$P(A) = \frac{M}{N} \qquad (1-1)$$

وبما أن $0 \le M \le N$ ، فإنه إذا قسمنا ذلك على N فإننا نجد أن الاحتمال $P(A)$ يحقق العلاقة التالية:

$$0 \le P(A) \le 1 \qquad (2-1)$$

وكأمثلة على ذلك نذكر الحوادث التالية:

. إن احتمال ظهور العدد 2 في حجر النرد يساوي:

$$P(2) = \frac{1}{6}$$

لأن هناك ست حالات ممكنة بينما توجد حالة واحدة ملائمة.

. إن احتمال ظهور عدد أقل من 5 في حجر النرد يساوي:

$$P(A_9) = \frac{4}{6}$$

لأن عدد الحالات الممكنة هو ست حالات، بينما عدد الحالات الملائمة هو 4 حالات محددة بالحالات 1 و 2 و 3 و 4.

. احتمال ظهور العدد 7 أو أحد الرقمين الصحيحين المجاورين مباشرةً له، على عجلة الرهان يساوي:

$$P(A) = \frac{3}{10}$$

لأن عدد الحالات الممكنة هي 10 حالات بينما عدد الحالات الملائمة هي 3 حالات محددة بالحالات 6 و 7 و 8.

ـ إذا كان لدينا في صندوق 5 كرات، منها 3 كرات بيضاء وكرتين سوداوتين، فإن احتمال ظهور كرة بيضاء عند إجراء عملية السحب الأولى يساوي:

$$P(A) = \frac{3}{5}$$

وإن احتمال ظهور كرة سوداء عند إجراء عملية السحب الأولى يساوي:

$$P(B) = \frac{2}{5}$$

وأن احتمال ظهور كرتين بيضاوين عند سحبهما مباشرةً (بدون إعادة) يساوي:

$$P = \frac{C_3^2}{C_5^2} = \frac{\frac{3!}{2!(3-2)!}}{\frac{5!}{2!(5-2)!}} = \frac{3}{10} \quad = \quad \frac{\text{عدد الأزواج الملائمة}}{\text{عدد الأزواج الممكنة}}$$

وذلك لأن عدد الحالات الممكنة (الأزواج الممكنة) يساوي C_5^2، وعدد الحالات الملائمة (الأزواج البيضاء) يساوي C_3^2.

ملاحظة:

إن التعريف النظري للاحتمال هو تعريف بسيط وسهل الفهم، ولكنه يكون متحيزاً، أو قاصراً، أو عديم الفائدة في كثير من الأحيان؛ وخاصةً في الحالات التالية:

. الحالات التي تكون فيها الحوادث الممكنة غير متكافئة؛ مثل اختيار السيارات بأنواعها أو أعمارها، ظهور شخص ذي دخل كبير وآخر فقير ... إلخ.

. الحالات التي يكون فيها عدد الحوادث الممكنة غير معلوم؛ مثل عدد سكان مدينة ما أو عدد المسافرين منها أو إليها، أو عدد الأعداد في مجال ما ... إلخ.

. الحالات التي يكون فيها عدد الحوادث الملائمة غير معلوم؛ مثل عدد المتعلمين في مدينة ما، أو عدد المدخنين فيها، أو عدد العاطلين عن العمل ... إلخ.

وبصورةٍ عامة فإن التعريف النظري للاحتمال مبني على أساس أن الحوادث الممكنة للتجربة هي حوادث متكافئة، وإذا أُخِلَّ بهذا الشرط فإن التعريف المذكور يفقد معناه. إن هذه الأسباب أو النواقص هي التي دعت إلى ظهور تعريف آخر للاحتمال هو التعريف الإحصائي أو التجريبي.

2 . التعريف الإحصائي (التجريبي):

يعرف احتمال تحقق حادث A بأنه يساوي نسبة عدد مرات تحقق ذلك الحادث على عدد المشاهدات أو التجارب التي أجريت عليه. وإذا فرضنا أن عدد مرات تحققه يساوي m، وعدد المشاهدات أو التجارب التي أجريت عليه يساوي n، فإننا نجد أن الاحتمال

الإحصائي التجريبي لتحقق ذلك الحادث يساوي التكرار النسبي لظهوره خلال هذه المشاهدات أو التجارب؛ أي أن $P(A)$ يساوي:

$$P(A)=\frac{m}{n} \qquad (3-1)$$

وبعبارة أخرى فإن الاحتمال الإحصائي هو عبارة عن نسبة ظهور أو وجود الحادث في المجتمع المدروس. فمثلاً: إذا كان عدد الولادات خلال عام ما مساوياً 7500 ولادة، وكان عدد الذكور فيهم 3900 ذكراً، فإن احتمال ولادة الذكر إحصائياً يساوي:

$$P(A)=\frac{3900}{7500}=0,52 \qquad (52\%)$$

وهي قيمة تختلف عن القيمة النظرية للاحتمال والتي نعتبرها مساوية لـ $0,50$.

ملاحظة:

إن هذا التعريف لا يشترط أن تكون الحوادث الممكنة متكافئة وذلك لأن كل حادث يفرض شروطه حسب ثقله ويعطينا من خلال ذلك الاحتمال المناسب له.

كما أنه ليس من الضروري أن يكون عدد الحوادث الممكنة أو الملائمة معلوماً، لأن حساب هذا الاحتمال يعتمد أساساً على عدد المشاهدات أو التجارب n وعلى عدد مرات التحقق m. ويمكن النظر إلى المشاهدات أو التجارب التي أجريت على أنها عينة من جملة المشاهدات أو التجارب الممكنة.

ملاحظة:

إن الاحتمال الإحصائي $\frac{m}{n}$ قد يختلف من عينة إلى أخرى (لها نفس الحجم n) ولكن يبرهن في الإحصاء الرياضي على أن النسبة $\frac{m}{n}$ تنتهي إلى نهاية محددة P عندما تنتهي n إلى اللانهاية؛ أي أن:

$$\lim_{n\to\infty}\frac{m}{n}=P \qquad (4-1)$$

حيث P هو الاحتمال الحقيقي لتحقق الحادث A وهو قيمة قد تكون غير معلومة. ومن الواضح أيضاً أن $0 \le m \le n$ وهذا يعني أن الاحتمال الإحصائي يحقق المتراجحة التالية:

$$0 \le p \le 1 \qquad (5-1)$$

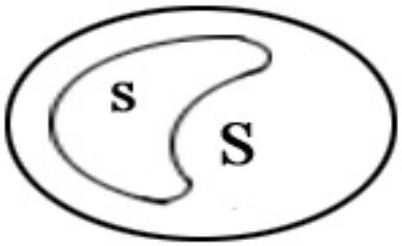
3 . التعريف الهندسي:

يعرف الاحتمال هندسياً بأنه نسبة المساحة الملائمة على المساحة الكلية وذلك بشرط أن يكون التوزيع منتشراً بانتظام فوق المساحة الكلية.

أي أن الاحتمال الهندسي يساوي:

$$P = \frac{s}{S} \qquad (6-1)$$

حيث s هي المساحة الملائمة، وS هي المساحة الكلية المبينة في الشكل رقم (1−1).

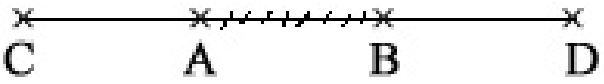


الشكل رقم (1−1)

وكمثال على ذلك نجد أن احتمال وقوع حادث على القطعة AB من الطريق CD يساوي (إذا كان التوزيع منتظماً):

$$P = \frac{AB}{CD}$$

والشكل رقم (1−2) يوضح ذلك:



الشكل رقم (1-2)

*** احتمال الحادث الأكيد:**

إذا كان الحادث أكيداً فإننا نرمز له بالرمز Ω، وبما أن عدد حالاته الملائمة M يساوي عدد الحالات الممكنة N، فإن احتمال تحققه يساوي الواحد؛ أي أن:

$$P(\Omega)=\frac{M}{N}=\frac{N}{N}=1 \qquad (7-1)$$

مثال 1:

إذا كان لدينا 200 بيضة طازجة، وإذا سحبنا بصورة عشوائية بيضة واحدة منها، فإن احتمال أن تكون تلك البيضة طازجة يساوي احتمال الحادث الأكيد ويساوي:

$$P(\Omega)=\frac{200}{200}=1$$

*** احتمال الحادث المستحيل:**

إذا كان الحادث مستحيلاً فإننا نرمز له بالرمز ϕ وأن عدد حالاته الملائمة M يساوي الصفر، وبذلك يكون احتمال تحققه يساوي الصفر أيضاً؛ أي أن:

$$P(\phi)=\frac{0}{N}=0 \qquad (8-1)$$

مثال 2:

إذا كان لدينا 200 مصباحاً معيباً، وسحبنا بصورة عشوائية واحداً منها، فإن احتمال أن يكون ذلك المصباح سليماً يساوي احتمال الحادث المستحيل ويساوي:

$$P(\phi)=\frac{0}{200}=0$$

1-3: قواعد جمع الاحتمالات:

1 . إذا كان A و B حادثين متنافيين فإن احتمال اجتماعهما (تحقق A أو B) يساوي مجموع احتماليهما؛ أي أن:

$$P(A\cup B)=P(A)+P(B) \qquad (9-1)$$

مثال 3:

يوجد في بحيرة 200 سمكة، منها 150 سمكة من النوع A، و30 سمكة من النوع B، و20 سمكة من النوع C. فما هو احتمال اصطياد سمكة من النوع A أو النوع B.

الحل:

إن السمكة المصادة تكون: إما من النوع A أو النوع B أو النوع C وهي حوادث متنافية. وبالتالي فإن الاحتمال المطلوب يساوي:

$$P(A \cup B) = P(A) + P(B)$$

ومن معطيات المسألة نجد أن:

$$P(A) = \frac{150}{200} \quad , \qquad P(B) = \frac{30}{200}$$

وبذلك نجد أن:

$$P(A \cup B) = \frac{150}{200} + \frac{30}{200} = \frac{180}{200} = 0{,}9$$

2 . يمكن تعميم هذه القاعدة على n حادثاً متنافياً فيكون لدينا:

$$P\left(\bigcup_{i=1}^{n} A_i\right) = P(A_1) + P(A_2) + P(A_3) + ... + P(A_n) \qquad (10-1)$$

3 . إذا كانت الحوادث $A_1, A_2, A_3, ..., A_n$ متنافية وتشكل مجموعة كلية Ω، فإن احتمال اجتماعها يساوي الواحد؛ أي إذا كانت الحوادث السابقة تحقق العلاقة:

$$A_1 \cup A_2 \cup A_3 \cup ..... \cup A_n = \Omega$$

وإن الاحتمال المطلوب يساوي:

$$P(\cup A_i) = P(\Omega) = 1$$
$$P(A_1) + P(A_2) + P(A_3) + ... + P(A_n) = 1 \qquad (11-1)$$

أي إن مجموع احتمال الحوادث المتنافية والمتكاملة يساوي الواحد.

مثال 4:

مـن المعطيـات فـي المثـال (3) أوجـد احتمـال أن تكـون السـمكة مـن النـوع A أو النوع B أو النوع C.

الحل:

بما أن الأنواع تشكل حوادث متنافية ومتكاملة فإن:

$$P(A \cup B \cup C) = P(A) + P(B) + P(C)$$

$$P(A \cup B \cup C) = \frac{150}{200} + \frac{20}{200} + \frac{20}{200} = \frac{200}{200} = 1$$

4 . بما أن A و $\overline{A}$ حادثين متنافيين ويشكلان لوحدهما مجموعة كلية Ω؛ (أي أن A و $\overline{A}$ متكاملين ومتنافيين) فإن احتمال اجتماعها يساوي الواحد؛ أي أن:

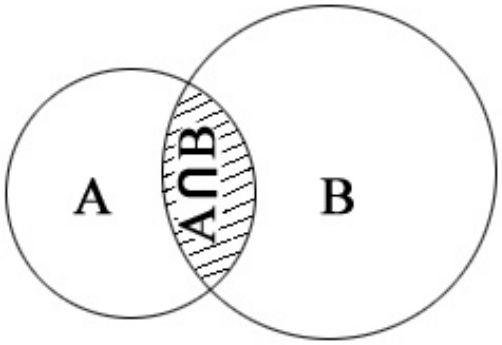
$$P(A \cup \overline{A}) = P(\Omega) = 1 \qquad (12-1)$$

$$P(A) + P(\overline{A}) = 1 \qquad (13-1)$$

5 . إذا كـان A وB حـادثين غيـر متنـافيين؛ أي يمكـن أن يتحققـا معـاً. فـإن احتمـال اجتماعهمـا يسـاوي مجمـوع احتماليهمـا مطروحـاً منـه احتمـال تقاطعهمـا؛ أي أن:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \qquad (14-1)$$

ويبين الشكل رقم (1−3) كيفية استخراج هذه العلاقة، والذي يعتمد على طرح التقاطع $(A \cap B)$ من الاجتماع $(A \cup B)$ حتى لا يتكرر حسابه مرتين.



الشكل رقم (1-3)

ويمكننا استنتاج العلاقة (1-9) من العلاقة (1-10) إذا لاحظنا أنه عندما يكون الحادثان A وB متنافيان فإن $(A \cap B) = \phi$، وأن $P(A \cap B) = 0$.

مثال 5:

إذا كان A هو حادث حصولنا على رقم لا يتجاوز 5 أي $(X \leq 5)$ في عجلة الرهان، وكان B هو حادث حصولنا على رقم فردي، فما هو احتمال تحقق A أو B.

الحل:

نلاحظ مما سبق أن A وB غير متنافيين لأنهما يمكن أن يحدثا معاً فيأخذا معاً أحد الأرقام 1 أو 3 أو 5. وبذلك نجد أن:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A \cup B) = \frac{6}{10} + \frac{5}{10} - \frac{3}{10} = \frac{8}{10} = 0{,}8$$

وذلك لأن:

$$P(A \cap B) = \frac{3}{10}$$

1-4: قواعد ضرب الاحتمالات:

1 . إذا كان A وB حادثين مستقلين، فإن احتمال تقاطعهما (حدوثهما معاً) يساوي جداء احتماليهما؛ أي أن:

$$P(A \cap B) = P(A) \cdot P(B) \qquad (15-1)$$

2 . يمكن تعميم العلاقة (1-15) على n حادثاً مستقلاً، فإذا كانت $A_1, A_2, A_3, ..., A_n$ حوادث مستقلة؛ فإن:

$$P\left(\bigcap_{i=1}^{n} A_i\right) = P(A_1) \cdot P(A_2) \cdot P(A_3) \cdot \ldots \cdot P(A_n) \qquad (16-1)$$

مثال 6:

يوجد في صندوق ثلاثة أنواع من الصابون A وB وC. ويوجد 8 قطع من النوع A، و3 قطع من النوع B، وقطعة واحدة من النوع C. وإذا سحبنا قطعتين (مع الإعادة) من الصندوق، فما هو احتمال أن نحصل على:

. قطعة من النوع A وأخرى من النوع B.

. قطعة من النوع B وأخرى من النوع C.

. قطعة من النوع A وأخرى من النوع C.

. قطعتين من النوع A.

. قطعتين من النوع B.

وإذا سحبنا ثلاث قطع مع الإعادة، فما هو احتمال حصولنا على قطعة من A وأخرى من B وثالثة من C.

الحل:

قبل كل شيء نلاحظ أنه بما أن السحب يجري مع الإعادة، فإن سحب القطعة الأولى لا يؤثر على احتمال سحب القطعة الثانية؛ أي أن سحب القطعتين عمليتان أو حادثان مستقلان، وبما أنه لدينا:

$$P(A)=\frac{8}{12} \quad , \qquad P(B)=\frac{3}{12} \quad , \qquad P(C)=\frac{1}{12}$$

فإننا نجد أن احتمالات تحقق الحالات المطلوبة تساوي على الترتيب:

$$P(A \cap B) = P(A) \cdot P(B) = \frac{8}{12} \cdot \frac{3}{12} = \frac{24}{144}$$

$$P(B \cap C) = P(B) \cdot P(C) = \frac{3}{12} \cdot \frac{1}{12} = \frac{3}{144}$$

$$P(A \cap C) = P(A) \cdot P(C) = \frac{8}{12} \cdot \frac{1}{12} = \frac{8}{144}$$

$$P(A \cap A) = P(A) \cdot P(A) = \frac{8}{12} \cdot \frac{8}{12} = \frac{64}{144}$$

$$P(B \cap B) = P(B) \cdot P(B) = \frac{3}{12} \cdot \frac{3}{12} = \frac{9}{144}$$

$$P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C) = \frac{8}{12} \cdot \frac{3}{12} \cdot \frac{1}{12} = \frac{24}{1728}$$

وبنفس الطريقة يمكننا معالجة الحالات المشابهة الأخرى والتي نتركها للطالب على سبيل التمرين.

3 . إذا كان الحادثان A وB غير مستقلين فإن مسالة حساب احتمال تقاطعهما تعتمد على مفهوم الاحتمال الشرطي والذي سنوضحه من خلال المثال التالي:

إذا أخذنا نفس معطيات المثال (6) السابق وقمنا بسحب قطعة أولى ولم نعدها إلى الصندوق، ثم قمنا بسحب القطعة الأخرى، فإن احتمالات السحب في المرة الثانية تتغير، وذلك حسبما تكون نتيجة السحبة الأولى، وتصبح الحوادث غير مستقلة (السحبة الثانية غير مستقلة عن السحبة الأولى لأن احتمالاتها تتغير).

فإذا كانت نتيجة السحبة الأولى قطعة من النوع A، فإن احتمالات السحب في المرة الثانية والتي سنرمز لها بالرموز التالية تساوي:

$$P(A/A) = \frac{7}{11} \quad , \quad P(B/A) = \frac{3}{11} \quad , \quad P(C/A) = \frac{1}{11}$$

حيث أن الرمز B/A يعني تحقق الحادث B بعد شرط تحقق A، وكذلك الأمر بالنسبة لـ C/A الذي تحقق C بعد تحقق A وكذلك A/A الذي تحقق A ثانية بعد تحقق A أولاً... إلخ.

أما إذا كانت نتيجة السحبة الأولى قطعة من النوع B، فإن احتمالات السحب في المرة الثانية والتي سنرمز لها بالرموز التالية تساوي:

$$P(A/B)=\frac{8}{11}\quad,\quad P(B/B)=\frac{2}{11}\quad,\quad P(C/B)=\frac{1}{11}$$

إن الرمز (A/B) يعني تحقق الحادث A بعد شرط تحقق B. وكذلك الأمر بالنسبة للرموز الأخرى.

أما إذا كانت نتيجة السحبة الأولى قطعة من النوع C، فإن احتمالات السحب في المرة الثانية والتي سنرمز لها بالرموز التالية تساوي:

$$P(A/C)=\frac{8}{11}\quad,\quad P(B/C)=\frac{3}{11}\quad,\quad P(C/C)=\frac{0}{11}=0$$

وكذلك نجد أن احتمال حصولنا على قطعة من النوع A في السحبة الأولى وقطعة من النوع B في السحبة الثانية يساوي:

$$P(A\cap B)=P(A)\cdot P(B/A)=\frac{8}{12}\cdot\frac{3}{11}=\frac{24}{132}=\frac{2}{11}$$

واحتمال حصولنا على قطعة من النوع B في السحبة الأولى وقطعة من النوع A في السحبة الثانية يساوي:

$$P(B\cap A)=P(B)\cdot P(A/B)=\frac{3}{12}\cdot\frac{8}{11}=\frac{24}{132}=\frac{2}{11}$$

واحتمال حصولنا على قطعة من النوع A في السحبة الأولى وقطعة من النوع C في السحبة الثانية يساوي:

$$P(A\cap C)=P(A)\cdot P(C/A)=\frac{8}{12}\cdot\frac{1}{11}=\frac{8}{132}$$

وبصورة عامة نجد أنه إذا كان A وB حادثين غير مستقلين فإن احتمال تقاطعهما يساوي احتمال الأول في الاحتمال الشرطي للثاني؛ أي أن:

$$P(A\cap B)=P(A)\cdot P(B/A)\qquad(17-1)$$

وبما أن:

$$P(A \cap B) = P(B \cap A) = P(B) \cdot P(A/B)$$

إذن نجد أن:

$$P(A \cap B) = P(A) \cdot P(B/A) = P(B) \cdot P(A/B) \qquad (18-1)$$

مثال 7:

أوجد الاحتمالات للأزواج المختلفة لمعطيات المثال (6) السابقة، وذلك في الحالات التالية (السحب بدون إعادة).

1 . إذا كانت نتيجة السحبة الأولى من النوع A.

2 . إذا كانت نتيجة السحبة الأولى من النوع B.

3 . إذا كانت نتيجة السحبة الأولى من النوع C.

الحل:

إذا كانت نتيجة السحبة الأولى من النوع A، فإن:

$$P(A \cap A) = P(A) \cdot P(A/A) = \frac{8}{12} \cdot \frac{7}{11} = \frac{14}{33}$$

$$P(A \cap B) = P(A) \cdot P(B/A) = \frac{8}{12} \cdot \frac{3}{11} = \frac{6}{33}$$

$$P(A \cap C) = P(A) \cdot P(C/A) = \frac{8}{12} \cdot \frac{1}{11} = \frac{2}{33}$$

وإذا كانت نتيجة السحبة الأولى من النوع B، فإن:

$$P(B \cap A) = P(B) \cdot P(A/B) = \frac{3}{12} \cdot \frac{8}{11} = \frac{2}{11}$$

$$P(B \cap B) = P(B) \cdot P(B/B) = \frac{3}{12} \cdot \frac{2}{11} = \frac{1}{22}$$

$$P(B \cap C) = P(B) \cdot P(C/B) = \frac{3}{12} \cdot \frac{1}{11} = \frac{1}{44}$$

وإذا كانت نتيجة السحبة الأولى من النوع C، فإن:

$$P(C \cap A) = P(C) \cdot P(A/C) = \frac{1}{12} \cdot \frac{8}{11} = \frac{2}{33}$$

$$P(C \cap B) = P(C) \cdot P(B/C) = \frac{1}{12} \cdot \frac{3}{11} = \frac{1}{44}$$

$$P(C \cap C) = P(C) \cdot P(C/C) = \frac{1}{12} \cdot \frac{0}{11} = 0$$

وهي تمثل جميع الأزواج المختلفة التي يمكن أن تظهر عند سحب قطعتين بدون إعادة. وهنا نشير إلى أن مجموع احتمالاتها يساوي الواحد (تأكد من ذلك) لأنها تشكل مجموعة متكاملة من الحوادث المؤلفة من الأزواج المختلفة.

1-5: قاعدة الاحتمال الكلي وقانون (بييز Bayes):

تعالج قاعدة الاحتمال الكلي مسألة احتمال تحقق حادث A ، لا يتحقق إلا بعد تحقق أحد الحوادث المتنافية والمتكاملة $H_1, H_2, H_3, ..., H_n$ ، والتي تسمى بالفرضيات الممكنة التي يمكن للحادث A أن يتحقق بعد تحقق أحدها.

بما أن A لا يتحقق إلا بعد تحقق أحد الحوادث $H_1, H_2, H_3, ..., H_n$ ، فإنه يمكن أن يتحقق في الحالة $(H_1 \cap A)$، أو في الحالة $(H_2 \cap A)$، أو في الحالة $(H_3 \cap A)$..... أو في الحالة $(H_n \cap A)$. لذلك يمكننا أن نكتب الحادث A على الشكل التالي:

$$A = (H_1 \cap A) \cup (H_2 \cap A) \cup (H_3 \cap A) \cup ..... \cup (H_n \cap A)$$

وبما أن الحوادث $(H_i \cap A)$ متنافية (لأن H_i متنافية) فإننا نجد أن احتمال تحقق A يساوي:

$$P(A) = P(H_1 \cap A) + P(H_2 \cap A) + P(H_3 \cap A) + ... + P(H_n \cap A)$$

واعتماداً على العلاقة (1-17) للحوادث غير المستقلة نجد أن:

$$P(A) = P(H_1) \cdot P(A/H_1) + P(H_2) \cdot P(A/H_2) + ... + P(H_n) \cdot P(A/H_n)$$

ويمكن كتابة هذه المعادلة على الشكل التالي:

$$P(A)=\sum_{i=1}^{n}P(H_i)\cdot P(A/H_i) \qquad (19-1)$$

وتسمى هذه القاعدة بقاعدة الاحتمال الكلي وهي تعطينا الاحتمال الكلي $P(A)$ بدلالة الاحتمالات الفردية $P(H_i)$ والاحتمالات الشرطية $P(A/H_i)$.

أما قانون (بييز) فيعالج مسألة حساب الاحتمالات الشرطية المعاكسة $P(H_i/A)$ والتي تعني احتمال أن يكون الحادث H_i قد رافق تحقق الحادث A. فإذا أخذنا العلاقة الآتية (انظر العلاقة (1−18)):

$$P(A\cap H_i)=P(A)\cdot P(H_i/A)=P(H_i)\cdot P(A/H_i)$$

ومن الطرف الأيمن نجد أن:

$$P(H_i/A)=\frac{P(H_i)\cdot P(A/H_i)}{P(A)} \qquad (20-1)$$

وإذا أخذنا العلاقة (1−19) بعين الاعتبار نجد أنه يمكننا كتابة العلاقة السابقة على الشكل التالي:

$$P(H_i/A)=\frac{P(H_i)\cdot P(A/H_i)}{\sum P(H_i)\cdot P(A/H_i)} \qquad (21-1)$$

وتسمى هذه العلاقة بقانون (بييز) أو بقاعدة قسمة الاحتمالات.

مثال 8:

لنفترض أن رسوب الطالب في أحد المقررات يعود إلى أحد الأسباب المتنافية والمتكاملة الآتية:

H_1 . حادث غيابه عن الامتحان لأي سببٍ كان.

H_2 . حادث عدم تحضيره للمقرر لأي سببٍ كان مع حضوره الامتحان.

H_3 . حادث صعوبة الأسئلة مع حضوره الامتحان.

ولنفترض أن:

$P(H_1)=0,2$ وهو احتمال غيابه عن الامتحان.

$P(H_2)=0,5$ وهو احتمال عدم تحضيره للمقرر

$P(H_3)=0,3$ وهو احتمال صعوبة الأسئلة

كما نفترض أن A هو حادث الرسوم وأن الاحتمالات الشرطية له تساوي:

$P(A/H_1)=1$ وهو احتمال رسوبه بسبب غيابه عن الامتحان.

$P(A/H_2)=0,90$ وهو احتمال رسوبه بسبب عدم تحضيره للمقرر

$P(A/H_3)=0,60$ وهو احتمال رسوبه بسبب صعوبة الأسئلة

فما هو احتمال رسوب الطالب بشكل عام؟ وما هي احتمالات أسباب رسوبه؟

الحل:

إن حادث الرسوب A يمكن أن يحدث مع أي من الحوادث H_3,H_2,H_1

ويمكن كتابته على الشكل التالي:

$$A=(H_1\cap A)\cup(H_2\cap A)\cup(H_3\cap A)$$

وبذلك نجد أن احتمال رسوب الطالب بشكل عام يساوي:

$$P(A)=P(H_1)\cdot P(A/H_1)+P(H_2)\cdot P(A/H_2)+P(H_3)\cdot P(A/H_3)$$

$$P(A)=(0,2)\cdot(1)+(0,5)\cdot(0,90)+(0,3)\cdot(0,60)=0,83$$

وإن احتمالات أسباب رسوبه تحسب من العلاقة:

$$P(H_i/A)=\frac{P(H_i)\cdot P(A/H_i)}{P(A)}$$

وهي تساوي على الترتيب:

$$P(H_1/A)=\frac{P(H_1)\cdot P(A/H_1)}{P(A)}=\frac{(0,20)\cdot 1}{0,83}=\frac{20}{83}$$ (بسبب الغياب)

وهو احتمال أن يكون سبب رسوبه هو الغياب عن الامتحان، وذلك بفرض أننا لا نعلم إن كان الطالب قد تقدم للامتحان أم لا.

$$P(H_2/A)=\frac{P(H_2)\cdot P(A/H_2)}{P(A)}=\frac{(0,5)\cdot(0,9)}{0,83}=\frac{45}{83}$$ (بسبب عدم التحضير)

وهو احتمال أن يكون سبب الرسوب هو عدم التحضير.

$$P(H_3/A)=\frac{P(H_3)\cdot P(A/H_3)}{P(A)}=\frac{(0,3)\cdot(0,60)}{0,83}=\frac{18}{83}$$

وهو احتمال أن يكون سبب الرسوب هو صعوبة الأسئلة.

وهنا نلاحظ أن مجموع الاحتمالات يساوي الواحد.

تمرينات

1 . لنفترض أننا خلطنا ورق اللعب المؤلف من 52 ورقة خلطاً جيداً وسحبنا منه ورقة، فما هو:

1 . احتمال أن تكون هذه الورقة 7.

2 . احتمال أن تكون هذه الورقة صورة.

3 . احتمال أن تكون هذه الورقة (آص).

4 . احتمال أن تكون هذه الورقة من النوع الديناري.

5 . احتمال أن تكون هذه الورقة ذات عدد زوجي.

6 . احتمال أن تكون هذه الورقة من النوع الأحمر.

2 . لنفترض أن X هو نتيجة الحجر الأول للنرد، وY هو نتيجة الحجر الثاني فماهو:

1 . احتمال أن يكون $X+Y=10$.

2 . احتمال أن يكون $X+Y>5$.

3 . احتمال أن يكون $(X=5),(Y=6)$ معاً.

4 . احتمال أن يكون $X\cdot Y=16$.

5 . احتمال أن يكون $X+Y>X\cdot Y$.

3 . يحتوي صندوق على 10 كرات مرقمة من 1 حتى 10، ومخلوطة جيداً والمطلوب حساب احتمال ظهور هذه الكرات بشكل متسلسل في كل من الحالتين التاليتين:

1 . حالة السحب مع الإعادة.

2 . حالة السحب بدون إعادة لجميع هذه الكرات بشكل متتالي.

الجواب: $\frac{1}{10^{10}}-1$ $\frac{1}{10!}-2$

4 . إذا سحبنا 6 كرات من الصندوق المذكور في المسألة السابقة، فما هو احتمال أن تكون هذه الكرات مختلفة الأرقام علماً بأن السحب جرى مع الإعادة.

الجواب:

$$\frac{10(10-1).....(10-4)}{10^6}$$

5 . نرمي قطعتين نقديتين فما هو احتمال حصولنا على:

. الحادث A الذي هو ظهور الشعار على القطعة الأولى.

. الحادث B الذي هو ظهور الشعار على القطعة الثانية.

. الحادث $A \cap B$ الذي هو ظهور الشعار على القطعتين معاً.

. الحادث $A \cup B$ الذي هو ظهور الشعار على أحد القطعتين.

6 . نسحب بدون إعادة بطاقتين من ورق اللعب الذي يتضمن 52 بطاقة فماهي احتمالات أن:

. أن تكون واحدة دينارية.

. أن تكون واحدة كبة.

. أن تكون واحدة كبة والأخرى دينارية.

. أن تكون الاثنتين دينارية.

. أن تكون الاثنتان كبة.

. أن تكون الاثنتان دينارية أو الاثنتان كبة.

الفصل الثاني

المتحولات العشوائية وقوانين توزيعها

ويشمل الفقرات التالية:

1 . تعاريف عامة.

2 . قوانين التوزيع الاحتمالية.

3 . خواص قوانين التوزيع الاحتمالية.

4 . تابع التوزيع.

5 . خواص تابع التوزيع.

6 . التوقع الرياضي.

7 . الوسيط.

8 . المنوال.

9 . مقاييس التشتت (التباين والانحراف المعياري).

10 . تمرينات.

الفصل الثاني

المتحولات العشوائية وقوانين توزيعها

يحتل مفهوم المتحول العشوائي في نظرية الاحتمالات مكاناً أساسياً وهو عبارة عن تعميم لمفهوم الحوادث. فما هو المتحول العشوائي؟

2-1- تعاريف عامة:

تعريف المتحول العشوائي: إن المتحول العشوائي هو متحول يأخذ قيمه الممكنة بشكل عشوائي (غير معروف مسبقاً) وبتكرارات معينة لكل منها،؛ أي أنه يمكن أن يأخذ أياً من قيمه عدة مرات وبدون نظام معين.

من التعريف السابق نلاحظ أن المتحول العشوائي يتميز بما يلي:

ـ إن مجموعة قيمه تشكل جملة من القيم (الحوادث) المتنافية والمتكاملة ويرمز لها بـ Ω وتسمى بفضاء الاحتمالات.

ـ إن تحولاته تتم بشكل عشوائي وتأخذ أياً من قيمه مراراً وبدون نظام معين. ولكن يمكننا أن نحسب أو نقدر كيفية توزع التكرارات على القيم الممكنة.

وكنماذج عن المتحولات العشوائية نذكر الأمثلة التالية:

. عدد الطلاب الذين يحضرون المحاضرة: وهو متحول عشوائي يأخذ قيمه من 0،1، 2، ... إلى n بشكل عشوائي (لا يمكننا معرفة عدد الذين سيحضرون المحاضرة بشكل مسبق). وقد يبلغ عددهم قيمة ما، ثم يعود ويبلغ نفس القيمة السابقة ...إلخ.

. عدد المكالمات الهاتفية الواردة إلى المقسم في الساعة: وهو متحول عشوائي له نفس خصائص المتحول السابق.

. عدد أفراد الأسرة: وهو متحول عشوائي يأخذ قيمه من 1، 2، ....، 25 وهو متحول قد يأخذ أية قيمة من قيمه عدة مرات.

. البعد بين مكان الإصابة ومركز الهدف عند إطلاق عدد كبير من الطلقات على هدف تدريبي. فهذا البعد هو متحول عشوائي يختلف من طلقة إلى أخرى، ويأخذ في المجال [a ,b] ويمكن لأية قيمة (أي بعد) أن تتكرر عدة مرات.

. كمية الأمطار التي تهطل في منطقة ما: هي متحول عشوائي تأخذ قيمتها الممكنة في المجال [a ,b] ويمكن لأية قيمة أن تتكرر عدة مرات.

ـ درجة الحرارة عند ظهر كل يوم: هي متحول عشوائي تأخذ قيمتها الممكنة في المجال [a ,b] ويمكن لأية قيمة أن تتكرر عدة مرات في الأيام المتتالية.

وباختصار إذا نظرنا إلى جميع المتحولات في الطبيعة والمجتمع لوجدناها متحولات عشوائية. وأن المتحولات العادية (التي تأخذ قيمها مرة واحدة وباتجاه معروف) ما هي إلا حالة خاصة من المتحولات العشوائية.

ومن الأمثلة السابقة يمكننا التمييز بين نوعين من المتحولات العشوائية هي:

. المتحولات العشوائية المنقطعة: وهي المتحولات العشوائية التي تكون قيمها الممكنة منقطعة، كالأمثلة الثلاثة الأولى السابقة.

. المتحولات العشوائية المستمرة: وهي المتحولات العشوائية التي تكون قيمها الممكنة مستمرة ضمن مجال معين، كالأمثلة الثلاثة الأخيرة السابقة.

وسنرمز للمتحول العشوائي بحرف لاتيني كبير مثل X أو Y ولقيمه بحرف صغير مثل x أو y.

2-2: قوانين التوزيع الاحتمالية:

عند تعريف المتحولات العشوائية ذكرنا أن المتحول يمكن أن يأخذ أية قيمة من قيمه بتكرارات معينة. وهذا يعني أن هناك تكرارات تقابل القيم الممكنة. وإذا حولنا هذه التكرارات إلى شكلها النسبي، فإننا نحصل على سلسلة الأعداد النسبية التي تعبر عن الاحتمالات التي تقابل تلك القيم، وهي ما يسمى بسلسلة التوزيع الاحتمالي: وبذلك يمكننا أن نعرّف قانون التوزيع الاحتمالي كما يلي:

. تعريف قانون التوزيع الاحتمالي:

إن قانون التوزيع الاحتمالي لمتحول عشوائي X هو العلاقة التي تربط بين القيم الممكنة لـ X وبين الاحتمالات المقابلة لها، أو هو سلسلة الاحتمالات التي تقابل تلك القيم.

وتكون قوانين التوزيع الاحتمالية منقطعة أو مستمرة، وذلك حسبما يكون المتحول العشوائي X منقطعاً أو مستمراً. كما يمكن أن تكون قوانين نظرية أو إحصائية وذلك حسب طبيعة المتحول X وإمكانية إخضاعه لقانون نظري معروف أو عدمها.

2-2-1: قانون التوزيع الاحتمالي للمتحول المنقطع:

ليكن لدينا متحولاً عشوائياً منقطعاً X ، فضاء إمكاناته S محدد كمايلي:

$$X(S) = X(x_1, x_2, ..., x_n)$$

وعليه يمكننا أن نحدد الاحتمالات المقابلة لكل قيمة من القيم $x_1, x_2, ..., x_n$ ؛ أي يمكننا حساب $P(X = x_i)$ من خلال تابع $f(x_i)$. كمايلي:

$$f(x_i) = P(X = x_i) \qquad (1-2)$$

يدعى $f(x_i)$ بقانون توزيع X، والذي يمكن أن نعرضه وفق الجدول الآتي:

x_1	x_2	.....	x_n
$f(x_1)$	$f(x_2)$	.....	$f(x_n)$

إن التوزيع $f(x_i)$ يجب أن يحقق الشرطين الآتيين:

$$f(x_i) \geq 0 \quad , \quad \sum_{i=1}^{n} f(x_i) = 1 \qquad (2-2)$$

وقد يتم التعبير عن قانون توزيع المتحول العشوائي X المنقطع بوساطة سلسلة من الأعداد كمايلي:

x_1	x_2	x_3	.....	x_n
P_1	P_2	P_3	.....	P_n

وبالتالي يمكننا التعبير عن الشرطين السابقين وفق ما يلي:

$$P_i \geq 0 \qquad , \qquad \sum_{i=1}^{n} P_i = 1$$

مثال 1:

رمينا حجري نرد. لنفرض أن X ترمز إلى العدد الأكبر من بين عددي حجري النرد. حدد قانون توزيع X.

الحل:

إن فضاء إمكانات تجربة إلقاء كل نرد تساوي:

$$\Omega = \{1,2,3,4,5,6\}$$

لتحديد قانون توزيع X ندرس كل حالة على حدى فنجد أن:

$$f(1) = P(X=1) = P[\{(1,1)\}] = \frac{1}{36}$$

$$f(2) = P(X=2) = P[\{(2,1),(2,2),(1,2)\}] = \frac{3}{36}$$

$$f(3) = P(X=3) = P[\{(3,1),(3,2),(3,3),(2,3),(1,3)\}] = \frac{5}{36}$$

$$f(4) = P(X=4) = P[\{(4,1),(4,2),(4,3),(4,4),(3,4)(2,4),(1,4)\}] = \frac{7}{36}$$

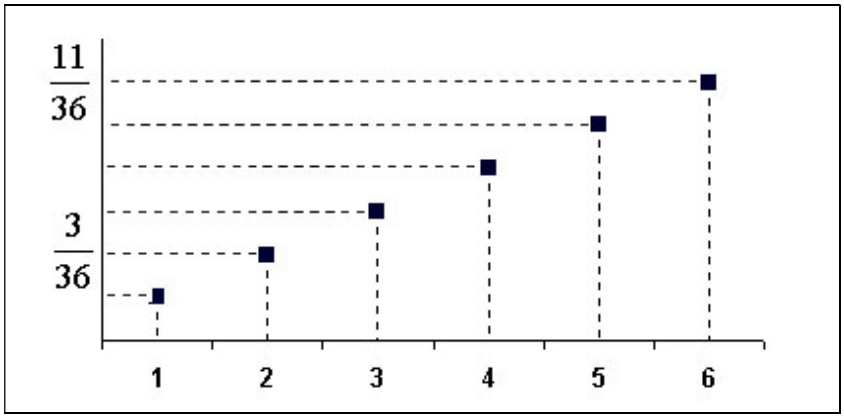
$$f(5) = P(X=5) = P[\{(5,1),(5,2),(5,3),(5,4),$$
$$(5,5),(4,5),(3,5)(2,5),(1,5)\}] = \frac{9}{36}$$

$$f(6) = P(X=6) = P[\{(6,1),(6,2),(6,3),(6,4),$$
$$(6,5),(6,6),(5,6),(4,6),(3,6)(2,6),(1,6)\}] = \frac{11}{36}$$

ثم نقوم بوضع هذه الاحتمالات ضمن جدول كالتالي:

x_i	1	2	3	4	5	6	$\sum$
$f(x_i)=p_i$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$	1

ويمكن تمثيل هذا القانون كما في الشكل التالي:



الشكل رقم (2-1)

مثال 2:

لنفرض الآن أنه لدينا متحولاً عشوائياً منقطعاً وليكن Y، يرمز إلى مجموع عددي النردين، فنجد أن فضاء إمكانات Y هي الإمكانيات المتوقعة لمجموع الرقمين التي تأخذ القيم التالية:

$$Y:\{2,3,4,5,6,7,8,9,10,11,12\}$$

ولحساب الاحتمالات المقابلة لهذه الحالات نقوم بإعداد جدول المجاميع المختلفة لهذين العددين فنحصل على الجدول التالي:

الحجر الثاني / الحجر الأول	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9

4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

$$g(Y_i) = P(Y=2) = P\{(1,1)\} = \frac{1}{36}$$

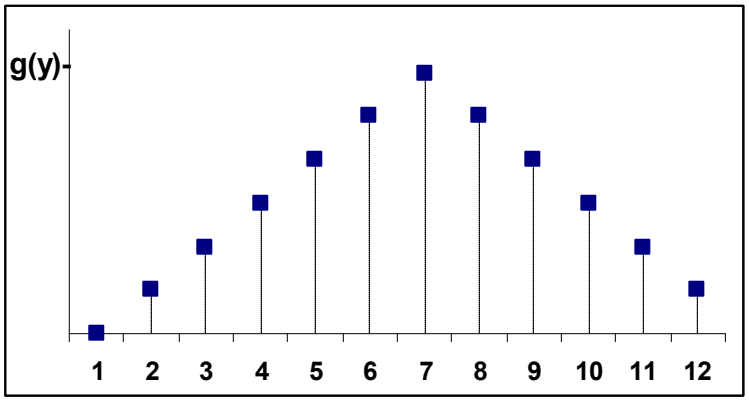
$$g(Y_2) = P(Y=3) = P\{(1,2),(2,1)\} = \frac{2}{36}$$

$$g(Y_3) = P(Y=4) = P\{(1,3),(2,2),(3,1)\} = \frac{3}{36}$$

وهكذا نحصل على بقية الاحتمالات، وبالتالي يمكننا الحصول على قانون التوزيع الاحتمالي لـ Y المبين في الجدول التالي:

y_i	2	3	4	5	6	7	8	9	10	11	12	$\sum$
$g(y_i)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	1

وأخيراً يمكننا أن نمثل قانون التوزيع السابق بيانياً وفق الشكل الآتي:



شكل رقم (2-2)

ومن هنا نلاحظ أن طول الأعمدة المقامة على المحور الأفقي يتناسب واحتمالات القيم الموافقة.

2-2-2: قانون التوزيع الاحتمالي للمتحول المستمر:

وسنوضحه من خلال المثال الآتي:

مثال 3:

لنفترض أن توزع سكان أحد البلدان حسب الفئات العمرية كان كما في الجدول الآتي:

الفئات العمرية	[0-10[	[10-20[	[20-30[	[30-40[	[40-50[	[50-60[	[60-70[	[70-80[	[80-90[	[90-100[	Σ
عدد السكان بالآلاف	7	14	10	22	15	12	9	6	3	1	100

المصدر: فرضي.

لنرمز بـX إلى عمر شخص اخترناه بشكل عشوائي من سكان ذلك البلد. قبل كل شيء نلاحظ إن المتحول X هو متحول عشوائي مستمر لأنه يستطيع أن يأخذ أية قيمة بين الحد الأدنى للعمر 0، والحد الأعلى للعمر 100 عاماً.

إنّ احتمال أن ينتمي الشخص إلى فئة عمرية محددة هو عبارة عن النسبة بين عدد أفراد هذه الفئة على مجموع عدد سكان هذه المنطقة؛ مثلاً: إن احتمال انتماء الشخص إلى الفئة العمرية]40-30] يساوي نسبة تكرار هذه الفئة على مجموع التكرارات؛ أي 22/100.

وإذا رمزنا للتكرارات المقابلة لفئة i بالرمز n_i ولمجموع التكرارات بالرمز $n = \sum n_i$ فإننا نجد أنه يمكننا حساب الاحتمالات المقابلة لكل فئة من الفئات العمرية السابقة من العلاقة $P(X \in Ii) = \frac{n_i}{n}$، ويمكننا بالتالي تحديد قانون التوزيع الاحتمالي للمتغير X بحساب التكرارات النسبية في كل الفئات المختلفة كالتالي:

I_i	[0-10[	[10-20[	[20-30[	[30-40[	[40-50[	[50-60[	[60-70[	[70-80[	[80-90[	[90-100[	Σ
$f(x_i)$	0,07	0,14	0,10	0,22	0,15	0,12	0,09	0,06	0,03	0,01	1

وهكذا يكون لدينا قانون التوزيع الاحتمالي للعمر X كمايلي:

$$f(x_i) = P(X \in I_i) = \frac{n_i}{n} = P_i$$

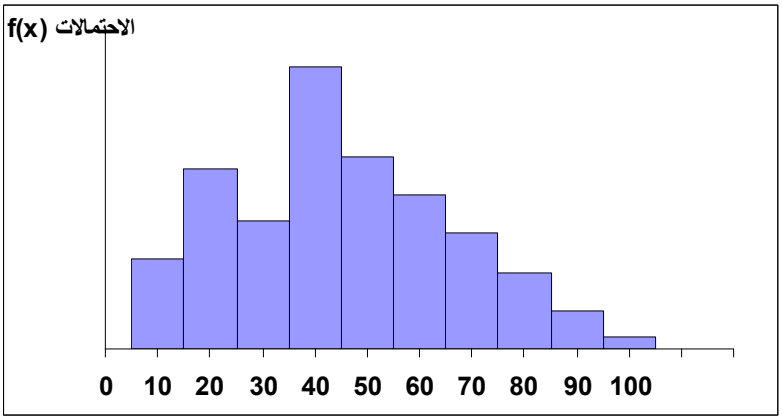
وإن التوزيع f يحقق الشرطين الآتيين الواجب تحققهما في قانون التوزيع الاحتمالي:

$$P_i \geq 0 \qquad , \qquad \sum_{i=1}^{n} P_i = 1$$

أو اختصاراً:

$$P_i \geq 0 \qquad , \qquad \sum P_i = 1$$

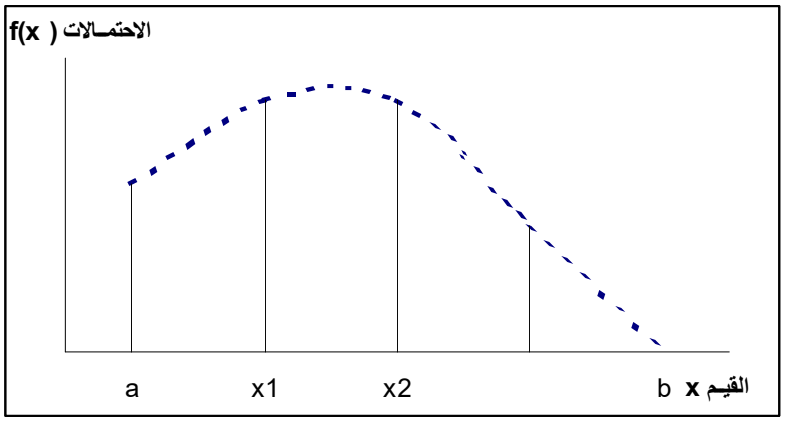
ويمكننا أن نمثل قانون التوزيع الاحتمالي السابق بيانياً وفق الشكل التالي:



شكل رقم (2-3)

لنفترض الآن أن عدد القيم التي يأخذها المتحول العشوائي X كبير جداً، كما يحدث دائماً في الحالات العملية التطبيقية، حيث يكون المتحول العشوائي متعلقاً بمجتمع إحصائي عدد مفرداته كبير جداً وينتهي إلى اللانهاية، وبالتالي فإن عدد المجالات المحصورة بين الحد الأدنى والحد الأعلى يتزايد ويتناهى أيضاً إلى اللانهاية، بحيث يصبح مدى كل فئة لا متناهياً في الصغر. وتبعاً لذلك، تتحول الأعمدة الممثلة لقانون التوزيع الاحتمالي إلى خطوط متلاصقة إلى حد كبير، وإذا وصلنا بين رؤوس

المستقيمات العمودية، التي تظهر كما لو كانت الواحدة متصلة بالأخرى، نحصل على خطٍ منحنٍ يسمى بمنحنى التوزيع الاحتمالي كما يظهر في الشكل الآتي.



شكل رقم (2-4)

من الواضح أن احتمال تحقق المتراجحة $x_1 < X < x_2$ حسب الرسم البياني يساوي مجموع المستقيمات العمودية المرسومة ضمن المجال (x_1, x_2). كما نلاحظ بأنه عندما أصبح عدد القيم الممكنة التي يأخذها المتحول العشوائي المستمر لا نهائياً فلم يعد باستطاعتنا أن نضع هذا القانون على شكل جدول كالسابق، وإنما نكتبه على شكل تابع رياضي لقيمه المتحولة x، وذلك على الشكل التالي:

$$p(x)=\begin{cases} f(x) & \forall \quad a \le x \le b \\ 0 & \text{فيما عدا ذلك} \end{cases}$$

ويشترط هنا أيضاً أن يكون $f(x) \ge 0$ [*]، وأن يكون $\int_{x_1}^{x_2} f(x) \cdot dx = 1$

يمكننا انطلاقاً من قانون التوزيع هذا المتمثل بتابع رياضي أن نقوم بحساب قيمة الاحتمال المقابل لأي مجال جزئي ضمن المدى العام لتغير المتحول.

[*] يدعى التابع f(x) أيضاً بتابع الكثافة أو بكثافة الاحتمال.

فمثلاً نجد أن احتمال أن ينتمي المتحول العشوائي المستمر X إلى المجال (x_1, x_2) كما يظهر في الشكل (2-4) يساوي المساحة الواقعة تحت المنحني $f(x)$، والمحصورة في المجال (x_1, x_2)، وهي تساوي التكامل التالي:

$$P(x_1 < X < x_2) = \int_{x_1}^{x_2} f(x) \cdot dx \qquad (4-2)$$

حيث $f(x) \cdot dx$ فيمثل احتمال أن ينتمي المتحول العشوائي X إلى المجال الصغير $(x, x + dx)$.

يمكننا بنتيجة هذا التحليل أن نحصل إلى تعريف وخواص قوانين التوزيع مهما كان نوع المتحول العشوائي مستمراً أو منقطعاً في مجال محدود أو غير محدود.

2-3: خواص قانون التوزيع:

. إن الاحتمالات المقابلة للقيم أو المجالات الجزئية لا يمكن أن تكون سالبة؛ أي إن:

$f(x_i) \geq 0 \quad \forall \quad x_i \in \Omega \qquad (5-2)$ للمتحول المنقطع

$f(x) \geq 0 \quad \forall x \in [a,b]$ للمتحول المستمر

. إن المجموع الكلي للاحتمالات يساوي الواحد؛ أي إن:

$$\sum_{i=1}^{n} f(x_i) = \sum_{i=1}^{n} p_i = 1 \qquad (6-2)$$ للمتحول المنقطع

$$\int_{-\infty}^{+\infty} f(x) \cdot dx = \int_{D} f(x) \cdot dx = 1$$ للمتحول المستمر

حيث D تمثل المجال أو المدى العام لتغير المتحول المستمر.

وتنتج هذه الخاصة الأخيرة من كون مجموعة القيم الممكنة التي يستطيع أن يأخذها المتحول X ضمن مجال تغيره، تشكل مجموعة كلية من الحوادث، ومن أن احتمال تحقق الحادث الأكيد يساوي الواحد.

مثال 4:

ليكن f(x) تابعاً معيناً بالعلاقة:

$$f(x)=\begin{cases}\frac{1}{6}x+\frac{1}{12} & \forall \quad 0\le x\le 3\\ 0 & \text{فيما عدا ذلك}\end{cases}$$

أثبت أن f(x) يمكن أن يكون قانون توزيع لمتحول عشوائي نرمز إليه بـ X، ثم احسب الاحتمال $P(1\le x\le 2)$ وقم بتمثيل ذلك بيانياً:

الحل:

لكي يكون f(x) قانون توزيع للمتحول العشوائي X، يجب أن يتحقق الشرطان:

1. $f(x)\ge 0$
2. $\int_D f(x)\cdot dx=1$

بما أن قيمة $0\le x\le 3$ فإن $f(x)>0$ دوماً؛ أي أن الشرط الأول محقق.

أما بالنسبة للشرط الثاني، فيجب أن يتحقق معنا:

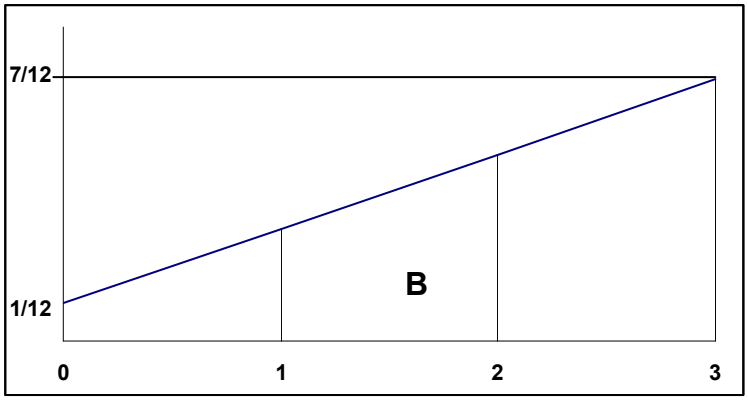
$$\int_D f(x)\cdot dx=\int_0^3\left(\frac{x}{6}+\frac{1}{2}\right)\cdot dx=1$$

وللبرهان على ذلك نجد أن:

$$\int_0^3\left(\frac{x}{6}+\frac{1}{12}\right)dx=\left[\frac{x^2}{12}+\frac{x}{12}\right]_0^3=\frac{1}{12}\left[x(x+1)\right]_0^3=\frac{1}{12}\left[3(3+1)\right]-0=1$$

أي أن المساحة تحت الخط المستقيم f(x) تساوي الواحد، وهي المساحة المبينة في الشكل الآتي.

بما أن الشرطين محققان، فيمكننا أن نقول أن التابع f(x) هو قانون توزيع للمتحول العشوائي X.



شكل رقم (2-5)

أما الاحتمال المطلوب فيتمثل بـ:

$$\int_{1}^{2} f(x)\cdot dx$$

أي لدينا:

$$p(1\le X\le 2)=\int_{1}^{2}\left(\frac{x}{6}+\frac{1}{12}\right)=\frac{1}{12}\left[x(x+1)\right]_{1}^{2}=\frac{1}{3}$$

وهذا الاحتمال ممثل بيانياً بالمساحة B ، التي يمكن حسابها بحساب مساحة شبه المنحرف B المحدد بالقيمتين x=1 وx=2 ، اللتين تفيداننا في حساب قاعدتي شبه المنحرف كمايلي:

عندما يأخذ X القيم x=1 وx=2، يكون لدينا:

$$f(x=1)=\frac{1}{6}+\frac{1}{12}=\frac{3}{12}$$

$$f(x=2)=\frac{1}{8}+\frac{1}{12}=\frac{5}{12}$$

وتمثل القيمة $\frac{3}{12}$ طول القاعدة الصغرى لشبه المنحرف، و $\frac{5}{12}$ تمثل طول القاعدة الكبرى، وبملاحظة أن طول الارتفاع الجانبي يساوي الواحد، نجد أن مساحة شبه المنحرف تساوي:

$$\frac{\frac{3}{12}+\frac{5}{12}}{2}\cdot 1=\frac{1}{3}$$

ومنه نستطيع أن نكتب:

$p(1\le X\le 2)$ =B مساحة

$$p(1\le X\le 2)=\frac{1}{3}$$

2-4- تابع التوزيع*:

إن حصولنا على قانون التوزيع الاحتمالي لـ X يعتبر مهماً جداً ويفيدنا كثيراً عند دراسة خصائص ذلك المتحول. وأن أهم أمر قد نهتم به عند دراسة متغيرات المتحول X هو حساب الاحتمالات المقابلة للقيم المختلفة. قد تكون هذه القيم محدودة بنقطة واحدة أو مجال معين ضمن المجال أو المدى العام. ويتم ذلك باستخدام قانون التوزيع لاستخراج ما يسمى بتابع التوزيع.

تعريف تابع التوزيع:

لنفرض أن X متحولاً عشوائياً (منقطعاً أو مستمراً). نعرف تابع التوزيع X بأنه التابع F الذي يعطينا احتمال أن يأخذ المتحول العشوائي X أية قيمة أصغر أو تساوي قيمة معينة a، ونكتب ذلك على الشكل التالي:

$$F(a)=P(X\le a) \qquad (7-2)$$

ويمكن أن نميز حالتين: المتحول العشوائي منقطع.

المتحول العشوائي مستمر.

تابع التوزيع للمتحول العشوائي منقطع:

إذا كان المتحول العشوائي X منقطعاً، وكانت قيمه x_i مرتبة ترتيباً تصاعديا، وكان

* يسمى هذا التابع بتابع الاحتمالات، كما يسميه البعض بالتابع التكاملي للتوزيع، أو القانون التكاملي للتوزيع، أو أيضاً تابع التوزيع التجميعي لأنه يسعى إلى تحديد الاحتمال التجميعي للمتحول العشوائي X.

قانون توزيعه $f(x_i)$ محدداً. فيمكننا أن نحدد تابع التوزيع F كما يلي:

$$F(x_i)=P(x\le x_i)=\sum_{x\le x_i} f(x_i)=\sum_{x\le x_i} P_i \qquad (8-2)$$

ويساعدنا هذا التابع في تحديد الاحتمالات التجميعية المقابلة لأية قيمة من قيم المتحول العشوائي المنقطع X.

مثال 5:

لدينا قانون توزيع المتحول العشوائي المنقطع X كما يلي:

x_i	x_1	x_2	x_3	.....	x_n
$f(x_i)=P_i$	$f(x_1)=P_1$	$f(x_2)=P_2$	$f(x_3)=P_3$	.....	$f(x_n)=P_n$

وبفرض أن القيم x_i مرتبة ترتيباً تصاعدياً نجد أن:

. احتمال أن يأخذ المتحول العشوائي X قيمة أقل أو تساوي x_s هو:

$$F(x_s)=P(X\le x_s)=\sum_{i=1}^{s} P_i \qquad (9-2)$$

. احتمال أن يأخذ المتحول العشوائي X قيمة أقل أو تساوي x_4 هو:

$$F(x_4)=P(X\le x_4)=\sum_{i=1}^{4} P_i = P_1+P_2+P_3+P_4$$

. احتمال أن يأخذ المتحول العشوائي X أكبر من x_s هو:

$$P(X>x_s)=\sum_{i=s+1}^{n} P_i = 1-F(x_s) \qquad (10-2)$$

أي أن هذا الاحتمال هو عبارة عن احتمال الحادث المتمم لحادث أن يأخذ المتحول العشوائي القيم $X\le x_s$ (متمم الحالة السابقة).

. فمثلاً: احتمال أن يأخذ المتحول العشوائي X أكبر من x_4 هو:

$$P(X > x_4) = P_5 + P_6 + P_7 + ... + P_n = 1 - F(x_4)$$

. احتمال أن يأخذ المتحول العشوائي X القيمة x_5 على الأقل هو:

$$P(X \geq x_s) = \sum_{i=s}^{n} P_i = 1 - F(x_{s-1}) \qquad (12-2)$$

أي أن هذا الاحتمال هو عبارة عن احتمال الحادث المتمم لحادث أن يأخذ المتحول العشوائي القيمة $X < x_s$ (متمم الحالة السابقة).

. فمثلاً: احتمال أن يأخذ المتحول X القيمة x_4 على الأقل هو:

$$P(X \geq x_4) = \sum_{i=4}^{n} P_i = P_4 + P_5 + P_6 + ... + P_n$$
$$= 1 - F(x_{4-1}) = 1 - F(x_3)$$
$$= 1 - (P_1 + P_2 + P_3)$$

. احتمال أن يأخذ المتحول العشوائي X قيمة محصورة بين قيمتين هما x_t , x_s حيث إنّ $x_t < x_s$ هو:

$$P(x_t < X \leq x_s) = \sum_{i=t+1}^{s} P_i = F(x_s) - F(x_t) \qquad (13-2)$$

. فمثلاً: احتمال أن يأخذ المتحول العشوائي قيمة X بين x_7 , x_3 هو:

$$P(x_3 < X \leq x_7) = \sum_{i=4}^{7} P_i = P_4 + P_5 + P_6 + P_7 = F(x_7) - F(x_3)$$

. احتمال أن يأخذ المتحول العشوائي X قيمة تحقق المتراجحة $x_t < X < x_s$ هو:

$$P(x_i < X < x_s) = \sum_{i=t+1}^{s-1} P_i = F(x_{s-1}) - F(x_t) \qquad (14-2)$$

. فمثلاً: نجد أن:

$$P(x_3 < X < x_7) = \sum_{i=4}^{6} P_i = P_4 + P_5 + P_6$$

. احتمال أن يأخذ المتحول العشوائي X قيمة $x_t \leq X \leq x_s$ هو:

$$P(x_t \le X \le x_s) = \sum_{i=t}^{s} P_i = F(x_s) - F(x_{t-1}) \qquad (15-2)$$

. فمثلاً احتمال أن يأخذ المتحول العشوائي X قيمة $x_3 \le X \le x_7$ هو:

$$P(x_3 \le X \le x_7) = \sum_{i=3}^{7} P_i = P_3 + P_4 + P_5 + P_6 + P_7$$
$$= F(x_7) - F(x_2)$$

مثال 5:

لنعد إلى معطيات المثال (1) من الفقرة (1−2) حيث تحدد معنا قانون توزيع المتحول العشوائي X كما يلي:

x_i	1	2	3	4	5	6
$f(x_i) = p_i$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$

ومنه يمكننا إيجاد الاحتمالات التجميعية كالتالي:

x_i	1	2	3	4	5	6
الاحتمالات التجميعية $f(x_i)$	$\frac{1}{36}$	$\frac{4}{36}$	$\frac{9}{36}$	$\frac{16}{36}$	$\frac{25}{36}$	$\frac{36}{36}$

من الواضح أنه يمكننا تحديد تابع التوزيع لهذا لمتحول العشوائي X كما يلي:

x_s	1	2	3	4	5	6
$f(x_s) = p(X \le x_s)$	$\frac{1}{36}$	$\frac{4}{36}$	$\frac{9}{36}$	$\frac{16}{36}$	$\frac{25}{36}$	$\frac{36}{36}$

واعتماداً على هذا الجدول يمكننا حساب الاحتمالات التالية:

$$P(X \le 4) = F(4) = \frac{16}{36}$$

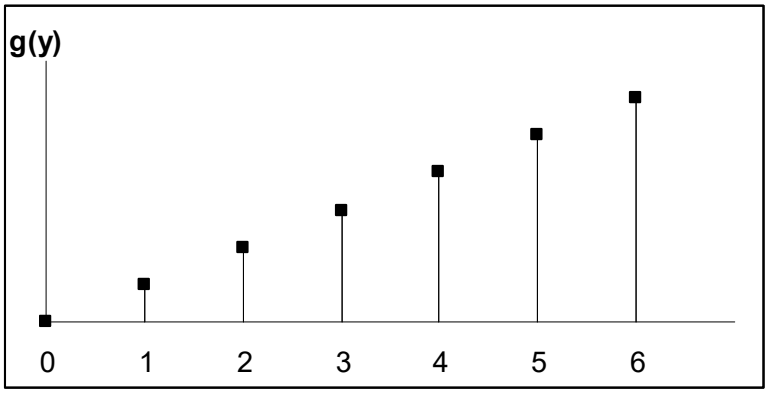
$$P(X > 2) = 1 - F(2) = 1 - \frac{4}{36} = \frac{32}{36}$$

$$P(2 < X \le 5) = F(5) - F(2) = \frac{25}{36} - \frac{4}{36} = \frac{21}{36}$$

$$P(2 \le X \le 5) = F(5) - F(1) = \frac{25}{36} - \frac{1}{36} = \frac{24}{36}$$

$$P(2 < X < 5) = F(4) - F(2) = \frac{16}{36} - \frac{4}{36} = \frac{12}{36}$$

إن التمثيل البياني لتابع التوزيع هو كما مبين على الشكل التالي:



شكل رقم (2–6)

تابع التوزيع للمتحول العشوائي مستمر:

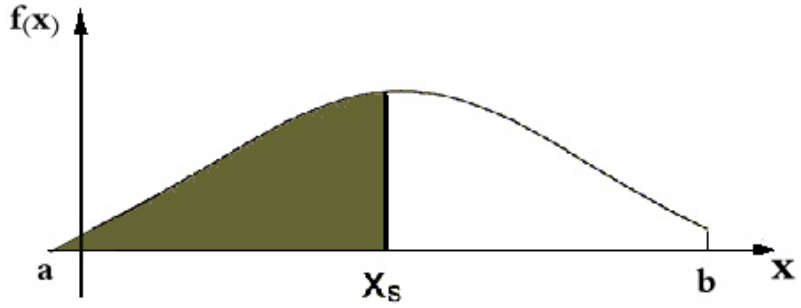
إذا كان المتحول العشوائي X مستمراً ومحدداً بقانون التوزيع f(x) فيمكننا أن نحدد تابع التوزيع F كما يلي:

$$F(x_S) = \int_{-\infty}^{x_S} f(x) \cdot dx \qquad (16-2)$$

وهـذا التـابع يحـدد لنـا احتمـال أن يأخـذ المتحـول العشـوائي X قيمـة أصـغر أو تساوي x_s؛ أي:

$$F(x_s) = P(X \le x_s) = \int_{-\infty}^{x_s} f(x) \cdot dx \qquad (17-2)$$

وهو يمثل المساحة المحصورة تحت المنحني الممثل لقانون التوزيع f(x) والواقعة على يسار النقطة x_s كما في الشكل التالي:



الشكل رقم (2-7)

ويساعدنا التابع F(x) على حساب احتمال أن يأخذ المتحول العشوائي المستمر X أية قيمة ضمن مجال جزئي من مجال تغيره.

مثلاً: لو أردنا أن يأخذ X قيمة محصورة بين قيمتين هما: x_s و x_t

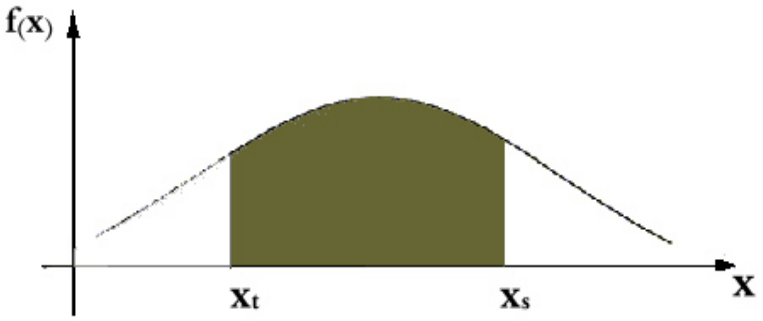
حيث أن $(x_t \le x_s)$، فإن ذلك يتم بالشكل التالي:

$$P(x_t \le X \le x_s) = \int_{x_t}^{x_s} f(x) \cdot dx$$

$$= \int_{-\infty}^{x_s} f(x) \cdot dx - \int_{-\infty}^{x_t} f(x) \cdot dx$$

$$P(x_t \le X \le x_s) = F(x_s) - F(x_t) \qquad (18-2)$$

وهو يمثل بيانياً بالمساحة الواقعة تحت المنحني الذي يرسمه f(x) والمحصورة بين النقطتين x_s و x_t كما في الشكل التالي:



شكل رقم (2–8)

ويمكننا الاستفادة من العلاقة (2–18) للبرهان على أنه إذا كان x مستمراً، فإن احتمال أن يأخذ قيمة وحيدة معينة x_s يساوي الصفر؛ أي أن:

$$P(X = x_t) = 0$$

حيث أن x_s أية قيمة من قيم X.

وللبرهان نلاحظ أنه لو أعطينا X تزايداً مقداره dx، فإن:

$$P(x < X < x + dx) = F(x + dx) - F(x) = f(x') \cdot dx$$

حيث أن x' هي قيمة تحقق المتراجحة $x < x' \leq x + dx$.

وعندما $dx \to 0$ فإن عرض المستطيل الذي قاعدته dx ينتهي إلى الصفر، وبالتالي فإن مساحته تنتهي إلى الصفر. وعندها نجد أن:

$$\lim_{dx \to 0} P(x < X \leq x + dx) = P(X = x')$$

وبما أن: F(x) مستمر، نجد من جهة أخرى أن:

$$\lim_{dx \to 0} P(x \leq X \leq dx) = \lim_{dx \to 0} [F(x + dx) - F(x)] = \lim_{dx \to 0} f(x') \cdot dx = 0$$

وبالمقارنة نجد أن:

$$P(X = x') = 0$$

نتيجة هامة: هذا يعني أن احتمال أن يأخذ المتحول العشوائي المستمر X اية قيمة محددة x من مجال تحوله يساوي الصفر، وذلك لأن عدد قيمه في ذلك المجل يكون لا نهائياً.

. أما إذا أردنا أن يأخذ X قيمة لا تقل عن x_s، فالاحتمال يساوي:

$$P(X \geq x_s) = \int_{x_S}^{+\infty} f(x) \cdot dx$$

$$= \int_{-\infty}^{+\infty} f(x) \cdot dx - \int_{-\infty}^{x_S} f(x) \cdot dx$$

$$P(X \geq x_s) = 1 - F(x_s) \qquad (20-2)$$

2-5: خواص تابع التوزيع:

. إن تابع التوزيع يحقق العلاقة $F(x) \geq 0$.

- إن تابع التوزيع هو تابع غير متناقص؛ أي أنه إذا كان x_s و x_t عددين حقيقيين، وكان $x_1 < x_2$ فإن تابع التوزيع يحقق العلاقة التالية:

$$x_1 < x_2$$

$$F(x_1) \leq F(x_2)$$

وذلك لأن تابع التوزيع هو عبارة عن مجموع (تكامل) الاحتمالات المطلوبة، وبما أن جميع قيم قانون التوزيع لأنه يحقق الشرط التالي:

$$f(x) \geq 0$$

فإن أية زيادة في القيم تؤدي إلى زيادته أو إبقاءه ثابتاً على الأقل.

. إن قيمة تابع التوزيع عند نقطة $(-\infty)$ تساوي الصفر؛ أي إنّ:

$$F(-\infty) = 0$$

وذلك لأن:

$$F(-\infty) = \lim_{x \to -\infty} F(x) = \lim_{x \to -\infty} P(X \leq x) = P(x < -\infty)$$

أي إن $F(-\infty)$ يمثل احتمال أن يأخذ المتحول X قيماً أصغر من أي عدد حقيقي معطى، وهو بالطبع أمر مستحيل؛ أي أنه يساوي احتمال حدوث حادث مستحيل ومنه نجد أن:

$$F(-\infty)=0 .$$

. إن قيمة تابع التوزيع عند نقطة $(+\infty)$ تساوي الواحد؛ أي أن:

$$F(+\infty)=\int_{-\infty}^{+\infty} f(x)\cdot dx=1$$

وذلك لأن F(x) يكون قد شمل جميع القيم الممكنة لأن:

$$F(+\infty)=\lim_{x\to\infty} F(x)=\lim_{x\to\infty} P(X\le x)=P(x<\infty)$$

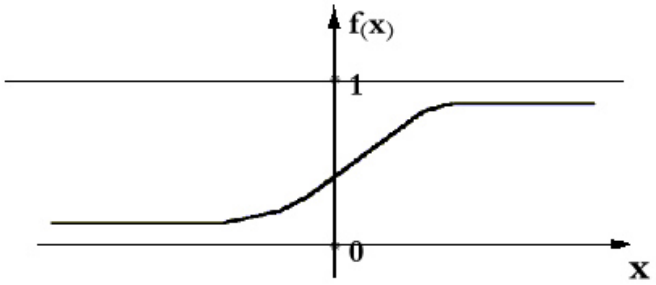
أي أن $F(+\infty)$ هو احتمال أن يأخذ المتحول X أية قيمة محدودة، وبالتالي فهو يمثل احتمال حدوث حادث أكيد، ومنه فإننا نجد أن:

$$F(+\infty)=1 .$$

ومن الخاصتين الأولى والأخيرة، يمكننا الاستنتاج أن:

$$0\le F(x)\le 1 \qquad (21-2)$$

وبما أن تابع التوزيع غير متناقص، فيمكننا أن نتصور الشكل العام للمنحني الذي يرسمه تابع التوزيع، وهو كما مبين في الشكل التالي، أو أي شكل يحقق الخواص السابقة:



شكل (2-9)

. إذا كان F(x) قابلاً للاشتقاق في مجال تعريفه، فإن مشتقه بالنسبة لـ x يساوي قانون التوزيع الاحتمالي f(x)، وذلك لأن:

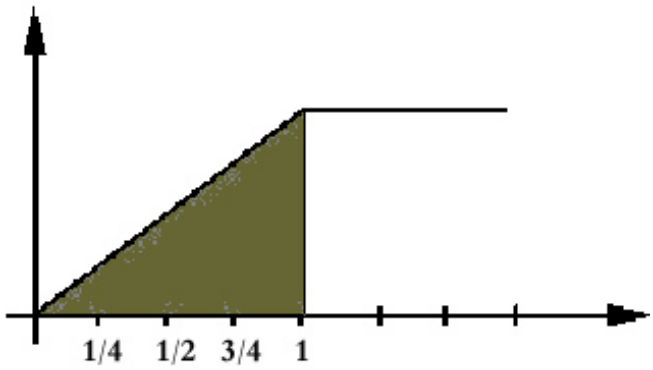
$$F'(x)=\frac{d}{dx}\int_{-\infty}^{x} f(x)\cdot dx=f(x)$$

ويستفاد كثيراً من هذه الخاصة في إيجاد قانون التوزيع إذا عُلِمَ تابع التوزيع. ولكي نتفهم بشكل واضح ودقيق مفهوم كل من قانون التوزيع وتابع التوزيع لمتحول عشوائي مستمر X، نقدم المثالين التاليين:

مثال 7:

لنفترض أنه لدينا تابع التوزيع الممثل بيانياً بالشكل التالي والمحدد رياضياً بالمعادلة:

$$F(x)_{\text{„}} \begin{cases} 0 & \forall \quad x \le 0 \\ x & 0 < x \le 1 \\ 1 & x > 1 \end{cases}$$



شكل رقم (2−10)

نلاحظ بأننا نستطيع اعتبار هذا التابع تابع توزيع لأنه يحقق الشروط التالية:

$$F(-\infty)=0$$

$$F(+\infty)=1$$

وإذا كان $x_1 < x_2$ فإن:

$$F(x_1) \le F(x_2)$$

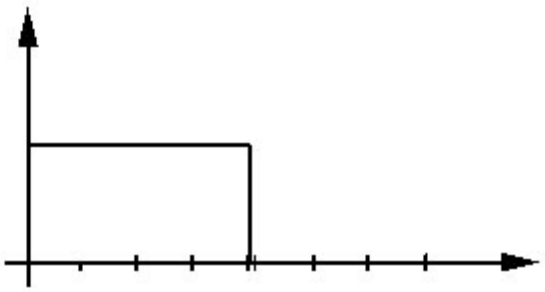
وإذا اشتققنا $F(x)$ بالنسبة لـ X نحصل على قانون التوزيع التالي:

$$f(x)=F'(x)_{\text{„}} \begin{cases} 0 & \forall \quad x \le 0 \\ 1 & 0 < x \le 1 \\ 0 & x < 1 \end{cases}$$

الذي يمكن كتابته على الشكل التالي:

$$f(x)\begin{cases}1 & \forall \quad 0 < x \le 1 \\ 0 & \text{فيما عدا ذلك}\end{cases}$$

وإن شكله البياني يأخذ الشكل التالي:



شكل رقم (2−11)

وباستخدام قانون التوزيع هذا، أو تابع التوزيع الذي بدأنا به، نستطيع أن نحدد احتمال أن يأخذ المتحول العشوائي أقل من قيمة ما أصغر من الواحد.

مثلاً: ما هو احتمال أن يأخذ المتحول العشوائي:

1 . قيمة أقل من 0,3.

2 . قيمة بين 0,2 و 0,8.

إن احتمال أن يأخذ المتحول العشوائي قيمة أقل من 0,3 يساوي:

$$P(X < 0,3) = F(0,3) = \int_0^{0,3} f(x) \cdot dx$$

$$= \int_0^{0,3} 1 \cdot dx = [x]_0^{0,3} = 0,3$$

أو بطريقة مختصرة:

$$P(X < 0,3) = F(0,3) = 0,3$$

واحتمال أن يأخذ قيمة بين 0,2 و 0,8 يساوي:

$$P(0,2 < X < 0,8) = \int_{0,2}^{0,8} f(x) \cdot dx$$

$$= \int_{0,2}^{0,8} 1 \cdot dx = [x]_{0,2}^{0,8} = 0,6$$

أو بطريقة مختصرة:

$$P(0,2 < X < 0,8) = F(0,8) - F(0.2) = 0,8 - 0,2 = 0,6$$

لقد حددنا في المثال السابق، تابع التوزيع واشتققنا منه قانون التوزيع، ومن ثم عمدنا إلى حساب بعض الاحتمالات. لنعد الآن إلى مثال آخر نحدد فيه قانون التوزيع ونستنبط تابع التوزيع.

مثال 8:

ليكن لدينا التابع الرياضي التالي:

$$f(x)_{"} \begin{cases} 2e^{-2x} & \forall \quad x > 0 \\ 0 & \text{عدا ذلك} \end{cases}$$

برهن على أن f(x) هو قانون توزيع لمتحول عشوائي X ثم أوجد تابع توزيعه، ثم أوجد احتمال أن يقع X في المجال [2,4].

الحل:

لنتأكد من أن هذا التابع يمثل بالفعل قانون توزيع المتحول العشوائي X، يجب أن نتحقق من أنه يحقق الشرطين التاليين:

$f(x) \geq 0$ لكل قيم x

$$\int_{-\infty}^{+\infty} f(x) \cdot dx = 1$$

فيما يتعلق بالشرط الأول، نرى أنه محقق، لأنه مهما كانت قيمة x، يكون لدينا:

$$f(x) \geq 0$$

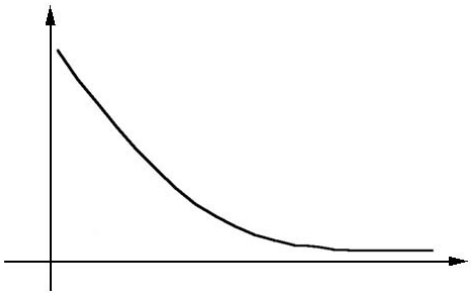
أما فيما يتعلق بالشرط الثاني، فلدينا:

$$\int_{-\infty}^{\infty} f(x)\cdot dx = \int_{-\infty}^{\infty} 2e^{-2x}\cdot dx = 2\left[\frac{e^{-2x}}{-2}\right]_0^{\infty} = \left[-e^{-2x}\right]_0^{\infty} = 1$$

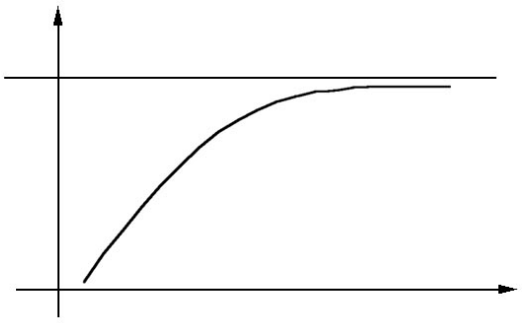
لنحدد الآن تابع التوزيع F(x)، لدينا:

$$F(x) = \int_{-\infty}^{x} 2e^{-2x} = \begin{cases} 1-e^{-2x} & \text{عندما} \quad x > 0 \\ 0 & \text{فيما عدا ذلك} \end{cases}$$

وإن تمثيل كل من قانون التوزيع السابق وتابع التوزيع المتعلق به يأخذان الشكلين التاليين:



شكل (2–12) قانون التوزيع



شكل (2–13) تابع التوزيع

لتحديد الاحتمالات المقابلة لمختلف قيم المتحول العشوائي X، نستطيع أن نستخدم f(x) أو F(x) (بالتأكيد في المسائل العملية نختار أحدهما) لكننا الآن سنسعى إلى تحديد احتمال أن يأخذ المتحول العشوائي قيمة في المجال [2-4] بالاعتماد على الاثنين.

1 . بواسطة f(x)؛ لدينا:

$$P(2 \le X \le 4) = \int_2^4 2e^{2x} \cdot dx = \left[-e^{-2x} \right]_2^4 = e^{-4} - e^{-8} \approx 0{,}018$$

2 . بواسطة F(x)، لدينا:

$$P(2 \le X \le 4) = F(4) - F(2) = \left(1 - e^{-8}\right) - \left(1 - e^{-4}\right) = \\ = e^{-4} - e^{-8} \approx 0{,}018$$

2-6: التوقع الرياضي:

إن التوقع الرياضي للمتحول العشوائي X هو القيمة المتوسطة لقيمه الممكنة ونرمز له بالرمز $E(X)$. ولكن طريقة حسابها هنا تختلف عما ورد في مبادئ الإحصاء لأنها أصبحت تتأثر بنوع المتحول (منقطع أو مستمر)، وبنوع قانون التوزيع الاحتمالي الذي يخضع له المتحول X. ولهذا فإننا سنعرف هذا التوقع حسب نوعية المتحولين كمايلي:

1 . التوقع الرياضي لمتحول عشوائي منقطع:

يعرف التوقع الرياضي لمتحول منقطع، عرف قانون توزيعه، بالعلاقة التالية:

$$E(X) = \sum_{i=1}^{n} x_i f(x_i) = \sum_{i=-1}^{n} x_i P_i \qquad (20-2)$$

من الواضح أن التوقع الرياضي هو عبارة عن قيمة عددية وتمثل القيمة المتوسطة لقيم X ، ونلاحظ أن المجموع السابق يشمل جميع قيم X من 1 حتى n مضروبةً باحتمالاتها، وعندما تكون مجموعة قيم X غير محدودة فإن:

$$E(X) = \sum_{=1i}^{\infty} x_i f(x_i) = \sum_{i=-1}^{\infty} x_i P_i \qquad (21-2)$$

ويجب الإشارة هنا إلى أن المجموع اللانهائي الوارد في (2-21) قد يكون متقارباً أو متباعداً. فإذا كان متباعداً، فإن التوقع الرياضي الموافق يكون غير موجود.

مثال 9:

إذا ألقينا حجر نرد عدد أكبر من المرات، فما هي القيمة المتوقعة للمتحول العشوائي X الذي يمثل العدد الذي يظهر على وجه ذلك النرد؟

الحل:

نعلم أن قانون التوزيع لـ X هو:

X قيم $= x_i$	1	2	3	4	5	6
$f(x_i) = p_i$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

ومنه نجد أن القيمة المتوقعة تساوي:

$$E(X) = 1\left(\frac{1}{6}\right) + 2\left(\frac{1}{6}\right) + 3\left(\frac{1}{6}\right) + 4\left(\frac{1}{6}\right) + 5\left(\frac{1}{6}\right) + 6\left(\frac{1}{6}\right) = \frac{7}{2} = 3{,}5$$

هذا يعني أنه إذا قمنا بإلقاء حجر نرد عدداً كبيراً من المرات فإن القيمة المتوسطة لـ X هي 3,5، (وذلك بشرط أن يكون ذلك النرد متوازناً).

2 . التوقع الرياضي لمتحول عشوائي مستمر:

يعرف التوقع الرياضي لمتحول عشوائي مستمر، عرف قانون توزيعه f(x)، بالعلاقة التالية:

$$E(X) = \int_{-\infty}^{+\infty} x \cdot f(x) \cdot dx \qquad (22-2)$$

من الواضح أن التوقع الرياضي لـ X هو عبارة عن قيمة عددية وتمثل القيمة المتوسط لقيم X، كما ظهر في المثال أعلاه.

خواص التوقع الرياضي:

إن التوقع الرياضي لأي متحول عشوائي مستمر أو منقطع يتمتع بالخواص التالية:

1 . إن التوقع الرياضي لعدد ثابت يساوي نفسه؛ أي أنه لو كان X يساوي عدداً ثابتاً C فإن:

$$E(C) = C \qquad (23-2)$$

للبرهان على ذلك نلاحظ أن:

$$E(C) = \sum C.P_i = C\sum P_i = C.1 = C$$

2 . إن التوقع الرياضي لـ $C \cdot X$ يساوي $E(C \cdot X)$ ؛ ويساوي $C \cdot E(X)$ ، أي إن:

$$E(C \cdot X) = C \cdot E(X) \qquad (24-2)$$

ويمكن البرهان على ذلك كما يلي:

. إذا كان X منقطعاً فإن:

$$E(C \cdot X) = \sum_{=1i}^{n} f(x_i) \cdot Cx_i = C \sum_{i-=1}^{n} x_i f(x_i) = C \sum_{i-=1}^{n} x_i p_i = C \cdot E(X)$$

. أما إذا كان X مستمراً، فإن:

$$E(C \cdot X) = \int_{-\infty}^{+\infty} C \cdot x \cdot f(x) \cdot dx = C \int_{-\infty}^{+\infty} x \cdot f(x) \cdot dx = C \cdot E(X)$$

3 . إن التوقع الرياضي لـ X+C يساوي E(X)+C؛ أي إن:

$$E(X + C) = E(X) + C \qquad (25-2)$$

ويمكن البرهان على ذلك كما يلي:

. إذا كان X منقطعاً فإن:

$$E(X + C) = \sum_{i=1}^{n} f(x_i) \cdot (x_i + C) = \sum_{i-=1}^{n} x_i f(x_i) + \sum_{i-=1}^{n} Cf(x_i)$$

$$= \sum_{i=1}^{n} x_i p_i + \sum_{i-=1}^{n} Cp_i = E(X) + C \sum_{i-=1}^{n} p_i = E(X) + C$$

. إذا كان X مستمراً، فإن:

$$E(X + C) = \int_{-\infty}^{+\infty} (x + C) \cdot f(x) \cdot dx$$

$$= \int_{-\infty}^{+\infty} x \cdot f(x) \cdot dx + C \int_{-\infty}^{+\infty} f(x) \cdot dx$$

$$= E(X) + C$$

وبصورة عامة، يكون لدينا:

$$E(aX + b) = aE(X) + b \qquad (26-2)$$

2-7: الوسيط:

إذا لم يكن ممكناً حساب التوقع الرياضي نظراً لعدم وجوده أو لعدم تقارب هذه القيمة، فإننا نسعى إلى قياس التمركز بمقياس آخر نسميه الوسيط. ويعرف الوسيط لمتحول عشوائي X ، والذي نرمز له بالرمز M_e ، بأنه القيمة من قيم X التي تحقق:

$$P(X < M_e) = P(X > M_e) \qquad (28-2)$$

أي إن الوسيط هو القيمة التي يتساوى مجموع الاحتمالات التي على يسارها مع مجموع الاحتمالات التي على يمينها. وفي حالة كون التوزيعات مستمرة فإن العلاقة السابقة تكتب على الشكل التالي:

$$P(X < M_e) = P(X > M_e) = \tfrac{1}{2} \qquad (28-2)$$

أي إن هذا الوسيط يتحدد في التوزيعات المستمرة لقيمة X التي تكون حلاً للمعادلة:

$$F(X) = \frac{1}{2}$$

2-8: المنوال:

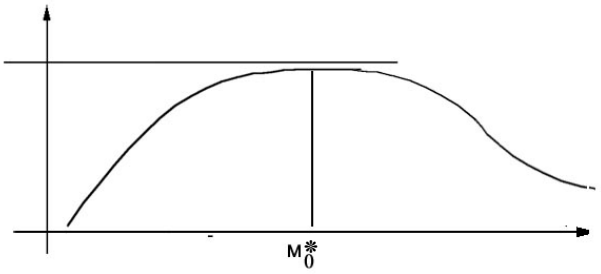
وهو عبارة عن القيمة من قيم X المقابلة لأكبر احتمال. فبالنسبة للمتحولات العشوائية المنقطعة، يتحدد المنوال بالقيمة M_0 التي تقابل x_t حيث:

$$\underset{i=1}{\overset{n}{M}} ax\ p_i = P_t$$

ونكتب ذلك على الشكل التالي:

$$. M_0 = x_t : \underset{i=1}{\overset{n}{M}} ax\ pi = p_t \qquad (29-2)$$

أما بالنسبة للمتحولات العشوائية المستمرة، فإن المنوال يكون مساوياً للقيمة التي تقابل النهاية العظمى لقانون التوزيع الاحتمالي f(x)، كما هو موضحاً على الشكل التالي:

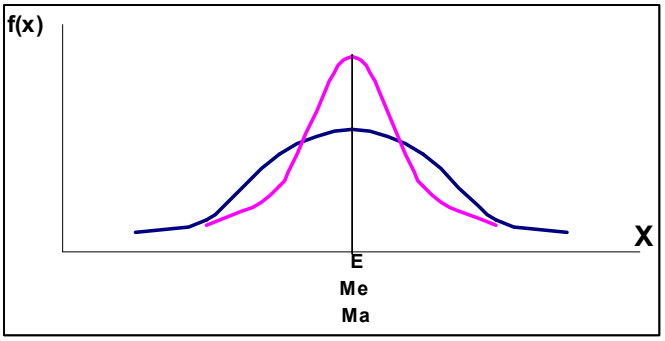


الشكل رقم (2-14)

ويتم الحصول على المنوال باشتقاق f(x) ووضع ذلك المشتق مساوياً الصفر وحساب جذر المعادلة التالية: $f'(x)=0$

2-9: مقياس التشتت:

إن القيم المركزية قد لا تكون كافية في بعض الأحيان لإعطاء فكرة واضحة ودقيقة عن المتحول العشوائي. وقد يحصل أن يكون لمتحولين نفس المقاييس المركزية رغم اختلافهما، كما يظهر في الشكل التالي الذي يمثل توزيعين لمتحولين مختلفين:



شكل رقم (2-15)

حيث نلاحظ بأن المتحولين في هذين التوزيعين يتميزان بأن التوقع الرياضي نفسه والقيمة الوسيطية نفسها والقيمة المنوالية نفسها أيضاً، وذلك على الرغم من أنهما محتولان مختلفان. ولإظهار هذا الاختلاف الذي يعبر عن تشتت القيم عن القيمة المركزية، كان لابد من استخدام مقاييس تقيس هذا التشتت وأهمها:

2-8-1: التباين:

. يعرف التباين لمتحول عشوائي X بالعلاقة العامة التالية:

$$\sigma^2(X) = E[X - E(X)]^2 \qquad (30-2)$$

فإذا كان X منقطعاً، فإن التباين يأخذ الشكل التالي:

$$\sigma^2(X) = \sum_{i=1}^{n} f(x_i) \cdot [x_i - E(X)]^2 = \sum_{i=-=1}^{n} p_i [x_i - E(X)]^2 \qquad (31-2)$$

أما إذا كان X مستمراً، فإن التباين يأخذ الشكل التالي:

$$\sigma^2(X) = \int_{-\infty}^{+\infty} f(x) \cdot [x - E(X)]^2 \cdot dx \qquad (32-2)$$

خواص التباين:

يتمتع التباين بالخواص التالية:

1 . إذا كان يأخذ عدداً ثابتاً C فإن:

$$\sigma^2(C) = 0 \qquad (33-2)$$

2 . إنّ تباين جداء X بعدد ثابت C يساوي جدار C^2 في التباين الأصلي، أي يساوي:

$$\sigma^2(C \cdot X) = C^2 \sigma^2(X) \qquad (34-2)$$

3 . إن X المضاف إليه عدد ثابت يساوي نفسه، أي يساوي:

$$\sigma^2(C + X) = \sigma^2(X) \qquad (35-2)$$

4 . إن تباين أي تركيب خطي لـ X يساوي :

$$\sigma^2(aX + b) = a^2 \sigma^2(X) \qquad (36-2)$$

ونترك البرهان على هذه الخواص للقارئ على سبيل التمرين.

2-9-2: الانحراف المعياري:

وهو عبارة عن الجذر الموجب للتباين، لذلك فإننا سنرمز إليه بالرمز σ، فيكون:

$$\sigma(X) = +\sqrt{\sigma^2(X)} \qquad (37-2)$$

ويعتبر الانحراف المعياري مؤشراً هاماً عند دراسة المتحولات العشوائية، إذ يعبر عن المتوسط التربيعي للخطأ بالنسبة للتوقع الرياضي E(X).

ومن خواص الانحراف المعياري ما يلي:

$$\sigma(C)=0 \qquad (38-2)$$

$$\sigma(C\cdot X)=C\sigma(X) \qquad (39-2)$$

$$\sigma(X+C)=\sigma(X) \qquad (40-2)$$

$$\sigma(aX+b)=a\sigma(X) \qquad (41-2)$$

تمرينات

1 . إذا كان X متحولاً عشوائياً مستمراً وخاضعاً لقانون التوزيع التالي:

$$f(x)=\begin{cases}\frac{1}{2}x & :0\le x\le 2\\ 0\end{cases}\Bigg\}$$

والمطلوب:

. إيجاد تابع التوزيع $F(x)$.

. حساب الاحتمال $P(1\le x\le 1{,}5)$.

. تمثيل $F(x)$ و $f(x)$ بيانياً.

. حساب التوقع الرياضي والتباين والانحراف المعياري للمتحول X.

. حساب الوسيط والمنوال والتباين.

2 . إذا كان تابع التوزيع للمتحول العشوائي X محدد بالعلاقة:

$$F(x)=\begin{cases}1-\frac{1}{x} & :x\ge 1\\ 0\end{cases}\Bigg\}$$

والمطلوب:

. إيجاد تابع قانون التوزيع الاحتمالي $f(x)$ ، وتحديد النقاط التي يكون فيها غير محدد.

. تمثيل $F(x)$ و $f(x)$ بيانياً.

. حساب التوقع الرياضي والتباين والانحراف المعياري لـ X.

. حساب الوسيط والمنوال.

3 . يطلق أحد الرماه 3 طلقات على هدف واحد، فإذا كان احتمال الإصابة في كل مرة ثابتاً، ويساوي 0.4، وكان X هو عدد إصابات الهدف نتيجة هذه الطلقات، فالمطلوب:

. إيجاد قانون توزيع X المنقطع ثم حساب احتمال $P(X\le 2)$.

. حساب قيمة التوقع الرياضي والانحراف المعياريين والوسيط والمنوال.

4 . إذا كان $f(x)$ تابعاً للمتحول X ومحدداً بالعلاقة:

$$f(x)=\begin{cases} k\cdot x(x-1) & :0\le x\le 1 \\ 0 & \end{cases}$$

فالمطلوب:

. تحديد قيمة الثابت k حتى يصبح التابع السابق قانون توزيع لـ X.

. إيجاد قيمة التوقع الرياضي والانحراف المعياري والوسيط والمنوال لـ X.

. حساب الاحتمال $P(0{,}5\le X\le 0{,}75)$.

الفصل الثالث

بعض قوانين التوزيع الاحتمالية

ويشمل الفقرات التالية:

1- قانون التوزيع الاحتمالي المنتظم لمتحول منقطع.

2- قانون التوزيع الثنائي للمتحول المنقطع.

3 . قانون توزيع بواسون للمتحول المنقطع X.

4 . قانون التوزيع المنتظم المستمر.

5 . قانون التوزيع الطبيعي.

6 . قانون التوزيع الطبيعي المعياري.

7 . قانون توزيع χ^2 (كاي مربع).

8 . قانون توزيع T (توزيع ستودينت).

9 . قانون التوزيع F (توزيع فيشر . سينديكور).

10 . قانون التوزيع الاحتمالي الإحصائي.

11 . تمرينات.

الفصل الثالث

بعض قوانين التوزيع الاحتمالية

لقد أشرنا إلى أن عدد المتحولات العشوائية لانهائياً، وإلى أنه لكل متحول عشوائي قانون توزيع خاص به، ونظراً لصعوبة حصر هذه المتحولات وقوانينها، تم تصنيف تلك المتحولات حسب طبيعة ونوعية القانون الذي تخضع له، وأمكن دمج العديد من المتحولات ضمن قوانين توزيع متشابهة، وأهم هذه القوانين هي:

3-1: قانون التوزيع الاحتمالي المنتظم لمتحول منقطع:

وهو القانون الذي تكون الاحتمالات المقابلة للقيم الممكنة لـX متساوية وتساوي $\frac{1}{n}$.

فإذا كان المتحول العشوائي X يأخذ القيم الممكنة والمرتبة التالية:

$$X : X_1 \ , \ X_2 \ , \ X_3 \ , ..., X_n$$

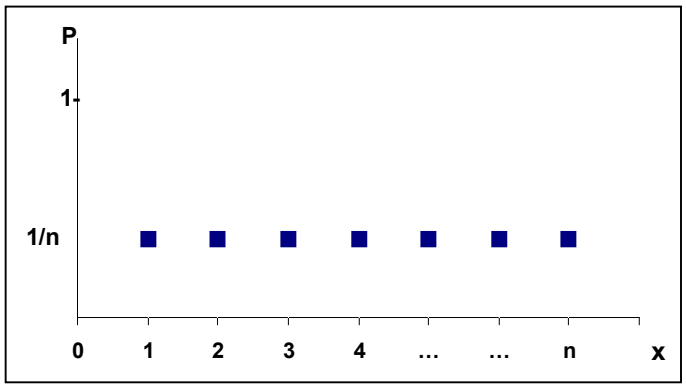
فإن الاحتمالات المنتظمة والمتساوية التي تقابلها تساوي:

$$P_i : \frac{1}{n} \ , \ \frac{1}{n} \ , \ \frac{1}{n} \ ... \ \frac{1}{n} \ ... \ \frac{1}{n} \qquad (1-3)$$

وهنا نشير إلى أن هذه الاحتمالات تحقق الشرطين التاليين:

$$\sum_{i=1}^{n} P_i = 1 \quad \text{وأن} \quad 0 \le P_i \le 1$$

وهو يرسم الشكل البياني التالي:



الشكل رقم (3-1)

أن متوسط هذه القيم يسمى التوقع الرياضي، ويعطى بالعلاقة التالية:

$$E(X)=\sum_{i=1}^{n}P_i x_i=\frac{1}{n}\sum_{i=1}^{n}x_i=\bar{x} \qquad (2-3)$$

وأن تباينها يعطى بالعلاقة التالية:

$$\sigma^2(X)=\sum_{i=1}^{n}P_i[x_i-E(X)]^2=\sum_{i=1}^{n}\frac{1}{n}[x_i-\bar{x}]^2=\sigma^2 \qquad (3-3)$$

وكأمثلة على المتحولات الخاضعة لهذا التوزيع المنتظم نذكر المتحولات التالية:

1 . نتيجة توقف عجلة الرهان: إذ إن نتيجة توقف العجلة عند الدورات المختلفة هي متحول عشوائي X يأخذ القيم الممكنة التالية:

$$x_i: 0\ ,\ 1\ ,\ 2\ ,\ 3\ ,\ 4\ ,\ 5\ ,\ 6\ ,\ 7\ ,\ 8\ ,\ 9$$

أما الاحتمالات المقابلة لهذه القيم فهي متساوية وتساوي $\frac{1}{10}$؛ أي إن:

$$P_i:\frac{1}{10}\ ,\ \frac{1}{10}\ ,\ \frac{1}{10}\ ,\ \frac{1}{10}\ ,\ \frac{1}{10}\ ,\ \frac{1}{10}\ ,\ \frac{1}{10}\ ,\ \frac{1}{10}\ ,\ \frac{1}{10}\ ,\ \frac{1}{10}$$

وإنّ متوسطها أو توقعها الرياضي يساوي:

$$E(X)=\sum_{i=0}^{10}P_i x_i=\frac{1}{10}\cdot 45=4{,}5$$

وإنّ تباينها يساوي:

$$\sigma^2(X)=\sum_{i=0}^{10}P_i(x_i-4{,}5)^2=8{,}25$$

2 . نتيجة رمي حجر النرد: وهو متحول عشوائي X يأخذ القيم والاحتمالات الممكنة التالية:

$$x_i: 1\ ,\ 2\ ,\ 3\ ,\ 4\ ,\ 5\ ,\ 6$$

$$P_i:\frac{1}{6}\ ,\ \frac{1}{6}\ ,\ \frac{1}{6}\ ,\ \frac{1}{6}\ ,\ \frac{1}{6}\ ,\ \frac{1}{6}$$

وإنّ توقعها الرياضي يساوي:

$$E(X)=\sum_{i=1}^{6}P_i x_i=\frac{1}{6}\cdot 21=3{,}5$$

وإن تباينها يساوي:

$$\sigma^2(X)=\sum_{i=1}^{6}P_i(x_i-3{,}5)^2=\frac{1}{6}\sum(x_i-3{,}5)^2=2{,}92$$

3 . نتيجة الولادة: هي متحول عشوائي X له الحالتان والاحتمالان التاليان:

X:	ذكر	أنثى
P_i	$\frac{1}{2}$	$\frac{1}{2}$

وهناك أمثلة كثيرة خاضعة للتوزيع المنتظم، ونترك أمر البحث عن بعضٍ منها للقارئ على سبيل التمرين.

3- 2: قانون التوزيع الثنائي للمتحول المنقطع:

وهو يعالج مسألة حساب الاحتمالات التي تقابل عدد مرات تحقق ظاهرة معينة عند إجراء n تجربة مستقلة عليها، وذلك بفرض أن احتمال تحقق الظاهرة المدروسة في كل تجربة معلوم ويساوي مقداراً ثابتاً P.

وهنا يفترض أن الظاهرة المدروسة لها وجهان فقط: أحدهما هو الوجه الذي نبحث عن تحققه (النجاح) واحتماله P، والآخر هو الوجه المعاكس (الإخفاق) واحتماله يساوي $q=1-P$.

وحتى نستخرج هذا القانون نقدم له بالمثال التالي:

مثال 1:

لنفترض أن نسبة المدخنين بين طلاب الجامعة تساوي P=0,20، فما هي الاحتمالات المقابلة لعدد المدخنين عند اختيار ثلاثة طلاب بشكل عشوائي؟

لنرمز لعدد المدخنين بين عناصر العينة بالرمز X، وبذلك نجد أن X يمكن أن تأخذ القيم التالية:

$X=0$ عدم وجود أي مدخن بينهم.

$X=1$ وجود مدخن واحد.

$X=2$ وجود مدخنين.

$X=3$ وجود ثلاثة مدخنين.

ولنرمز إلى حادث كون الطالب i مدخناً بالرمز A_i ، ولكونه غير مدخن بالرمز $\overline{A}_i$ ، وبذلك نجد أن:

احتمال أن يكون $X=0$ يساوي احتمال أن تتحقق الحوادث المستقلة $\overline{A}_1$ و $\overline{A}_2$ و $\overline{A}_3$ معاً؛ أي أن:

$$\begin{aligned} P(X=0) &= P(\overline{A}_1 \cap \overline{A}_2 \cap \overline{A}_3) \\ &= P(\overline{A}_1)\cdot P(\overline{A}_2)\cdot P(\overline{A}_3) \\ &= q\cdot q\cdot q = q^3 = (0{,}80)^3 = 0{,}512 \end{aligned}$$

وأن احتمال أن يكون $X=1$ يساوي احتمال أن يكون أحد الثلاثة مدخناً والاثنين الآخرين غير مدخنين، وإن ذلك يمكن أن يحدث بأشكال مختلفة كما يلي:

$$P(X=1) = P\left[(A_1 \cap \overline{A}_2 \cap \overline{A}_3)\cup(\overline{A}_1 \cap A_2 \cap \overline{A}_3)\cup(\overline{A}_1 \cap \overline{A}_2 \cap A_3)\right]$$

وبما أن الحالات متنافية، إذن يكون لدينا:

$$P(X=1) = P\left[(A_1 \cap \overline{A}_2 \cap \overline{A}_3) + P(\overline{A}_1 \cap A_2 \cap \overline{A}_3) + P(\overline{A}_1 \cap \overline{A}_2 \cap A_3)\right]$$

وبما أن الحوادث مستقلة، إذن يكون لدينا:

$$\begin{aligned} P(X=1) &= P(A_1)P(\overline{A}_2)P(\overline{A}_3) + P(\overline{A}_1)P(A_2)P(\overline{A}_3) + P(\overline{A}_1)P(\overline{A}_2)P(A_3) \\ &= p\cdot q\cdot q + q\cdot p\cdot q + q\cdot q\cdot p \\ P(X=1) &= 3pq^2 = 3(0{,}20)(0{,}80)^2 = 0{,}384 \end{aligned}$$

وأن احتمال أن يكون $X=2$ يساوي احتمال أن يكون اثنان من الثلاثة مدخنين وواحد غير مدخن، وأن ذلك ممكن أن يحدث بأشكال مختلفة كما يلي:

$$P(X=2)=P\left[\left(A_1\cap A_2\cap \overline{A}_3\right)\cup\left(\overline{A}_1\cap \overline{A}_2\cap A_3\right)\cup\left(\overline{A}_1\cap A_2\cap A_3\right)\right]$$
$$P(X=2)=P\left(A_1\cap A_2\cap \overline{A}_3\right)+P\left(A_1\cap \overline{A}_2\cap A_3\right)+P\left(\overline{A}_1\cap A_2\cap A_3\right)$$
$$P(X=2)=P(A_1)P(A_2)P\left(\overline{A}_3\right)+P(A_1)P\left(\overline{A}_2\right)P(A_3)+P\left(\overline{A}_1\right)P(A_2)P(A_3)$$
$$=p\cdot p\cdot q+p\cdot q\cdot p+q\cdot p\cdot p$$
$$P(X=2)=3p^2q=3(0{,}20)^2(0{,}80)=0{,}096$$

وأنّ احتمال أن يكون $X=3$ يساوي احتمال أن يكون الثلاثة مدخنين. ويتم ذلك إذا تحققت الحوادث الثلاثة المستقلة A_1 و A_2 و A_3 معاً؛ أي أن:

$$P(X=3)=P(A_1\cap A_2\cap A_3)=P(A_1)P(A_2)P(A_3)=P\cdot P\cdot P$$
$$P(X=3)=p^3=(0{,}20)^3=0{,}008$$

ويمكننا بالتالي أن نضع النتائج في جدول على الشكل التالي:

X	0	1	2	3	Σ
P(X)	q^3	$3pq^2$	$3p^2q$	p^3	1
القيم العددية	0,512	0,384	0,096	0,008	1

وهنا نلاحظ أن العدد 3 الذي يسبق الحالين $(X=1)$ أو $(X=2)$ ناتج عن الأشكال المختلفة لكل منهما، ويمكن حسابه من عدد المتوافقات؛ إذ إن:

$$C_3^1=\frac{3!}{1!(3-1)!}=\frac{3\cdot 2\cdot 1}{1\cdot 2\cdot 1}=3$$
$$C_3^2=\frac{3!}{2!(3-2)!}=\frac{3\cdot 2\cdot 1}{2\cdot 1\cdot 1}=3$$

كما يمكن حساب أمثال q^3 و p^3 من المتوافقات نفسها لأن:

$$C_3^0=\frac{3!}{0!\cdot 3!}=1 \qquad , \qquad C_3^3=\frac{3!}{3!\cdot 0!}=1$$

كما يمكن الحصول على هذه الاحتمالات من مفكوك $(q+p)^3$ حيث نجد أن:

$$(q+p)^3=C_3^0q^3+C_3^1pq^2+C_3^2p^2q+C_3^3p^3$$

ومما تقدم يمكننا صياغة القانون الذي يعطينا احتمالات أن يأخذ المتحول X القيمة k عند إجراء n تجربة مستقلة على الظاهرة المدروسة، والتي يكون فيها احتمال التحقق p معلوماً وثابتاً بالعلاقة التالية:

$$P(X=k)=C_n^k \cdot P^k \cdot q^{n-k} \qquad (4-3)$$

إذ إن: X هو عدد مرات التحقق خلال n تجربة وإن:

$$(k=0\ ,\ 1\ ,\ 2\ ,\ 3\ ,\ldots\ldots,\ n)$$

ويسمى هذا القانون بقانون التوزيع الثنائي.

ويبرهن في نظرية الاحتمالات على إن توقعه الرياضي (متوسط قيمه) يساوي:

$$E(X)=\sum_{k=0}^{n} P_k \cdot x_k = n \cdot p \qquad (5-3)$$

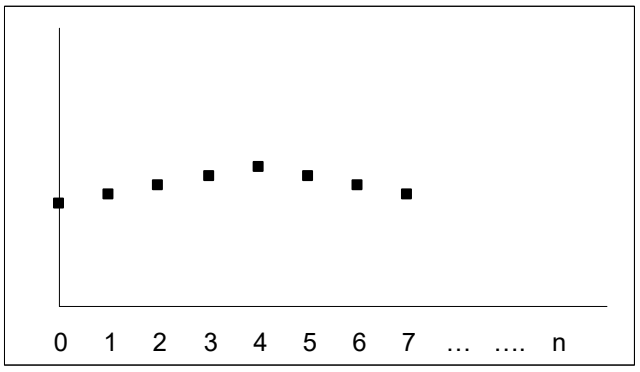
وإنّ تباينه يساوي:

$$\sigma^2(X)=\sum_{k=0}^{n} P_k \cdot (x_k - n \cdot p)^2 = n \cdot p \cdot q \qquad (6-3)$$

ويمكن وضع قانون التوزيع الثنائي في جدول كالتالي:

قيم X	0	1	2	3	·	k..	n	Σ
P(X = k)	q^n	$C_n^1 pq^{n-1}$	$C_n^2 p^2 q^{n-2}$	$C_n^3 p^3 q^{n-3}$		$C_n^k p^k q^{n-k}$	p^n	1

وبشكل عام هو يرسم الشكل البياني التالي:



شكل رقم (3-2)

ملاحظة:

يجب على كل طالب أن يلاحظ أن قيم X تبدأ من القيمة $x=0$ وتنتهي بالقيمة $X=n$ ، وبذلك يكون عدد القيم الممكنة لـ X مساوياً لـ $n+1$ قيمة.

مثال 2:

لنفترض أنه لدينا 200 قطعة من الخبز وفيها 150 قطعة من النوع الأول، و50 قطعة من النوع الثاني. وإذا سحبنا منها مع الإعادة عينة عشوائية من 5 قطع (السحب مع الإعادة حتى لايتغير الاحتمال P), فاحسب الاحتمالات المقابلة لعدد قطع النوع الأول التي يمكن أن تظهر في العينة.

الحل:

بما أن السحب جرى مع الإعادة فإن احتمال ظهور قطع النوع الأول في كل سحبه (تجربة) لايتغير ويبقى ثابتاً ويساوي:

$$P=\frac{150}{200}=\frac{3}{4}$$

وبذلك نجد أنّ:

$$q=1-p=\frac{1}{4}$$

وإذا رمزنا لعدد قطع النوع الأول التي يمكن أن تظهر في العينة بالرمز X، فإننا نجد أن الاحتمالات المطلوبة تحسب من قانون التوزيع الثنائي لأن الظاهرة ثنائية ومنها وجهان (نوع أول ونوع ثان) ولأن احتمال تحقق النوع الأول معلوم وثابت. وبذلك يأخذ المتحول العشوائي X القيم الممكنة التالية:

$$X: 0,\ 1,\ 2,\ 3,\ 4,\ 5$$

وتحسب الاحتمالات المقابلة لهذه القيم كما يلي:

إنّ احتمال أن لا تظهر أية قطعة من النوع الأول يساوي:

$$P(X=0)=C_5^0P^0q^5=1\cdot q^5=\left(\frac{1}{4}\right)^5=\frac{1}{1024}$$

وإنّ احتمال أن تظهر قطعة واحدة من النوع الأول يساوي:

$$P(X=1)=C_5^1P^1q^4=5\cdot P\cdot q^4=5\left(\frac{3}{4}\right)\left(\frac{1}{4}\right)^4=\frac{15}{1024}$$

وإنّ احتمال أن تظهر قطعتان من النوع الأول يساوي:

$$P(X=2)=C_5^2P^2q^3=10\cdot P^2\cdot q^3=10\left(\frac{3}{4}\right)^2\left(\frac{1}{4}\right)^3=\frac{90}{1024}$$

وإنّ احتمال أن تظهر ثلاث قطع فقط من النوع الأول يساوي:

$$P(X=3)=C_5^3P^3q^2=10\cdot P^3\cdot q^2=10\left(\frac{3}{4}\right)^3\left(\frac{1}{4}\right)^2=\frac{270}{1024}$$

وإنّ احتمال أن تظهر أربع قطع من النوع الأول يساوي:

$$P(X=4)=C_5^4P^4q\ =10\cdot P^4\cdot q^3=5\left(\frac{3}{4}\right)^4\left(\frac{1}{4}\right)=\frac{405}{1024}$$

وإنّ احتمال أن تظهر خمس قطع من النوع الأول يساوي:

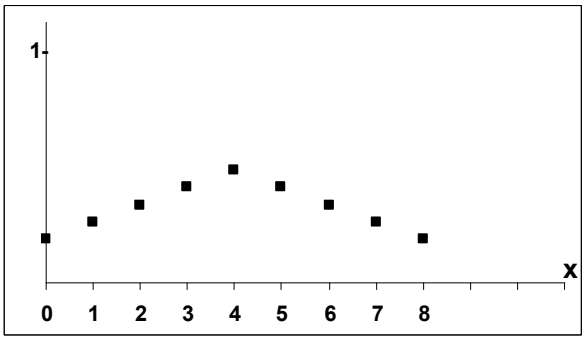
$$P(X=5)=C_5^5P^5q^0=1\cdot P^5=\left(\frac{3}{4}\right)^5=\frac{243}{1024}$$

وأخيراً نضع هذه الاحتمالات في جدول كالتالي:

X	0	1	2	3	4	5	$\sum$
p_i	$\frac{1}{1024}$	$\frac{15}{1024}$	$\frac{90}{1024}$	$\frac{270}{1024}$	$\frac{405}{1024}$	$\frac{243}{1024}$	1

*** خواص التوزيع الثنائي:**

1 . إن التوزيع الثنائي هو توزيع منقطع وترسم قيم الاحتمالات فيه نقطاً منقطعة، ويتوقف شكل التوزيع على مقدار الاحتمال P وعلى عدد التجارب n. فإذا كان الاحتمال $p=q=\frac{1}{2}$ فإن التوزيع يكون متناظراً حول القيمة الوسطى مهما كانت n كما في الشكل التالي:



الشكل رقم (4-3)

2 . إنّ قيم الاحتمال تحقق الشرطين التاليين:

$$0 \le P_k \le 1 \qquad (7-3)$$

$$\sum_{k=0}^{n} P_k = 1 \qquad (8-3)$$

3 . يمكن حساب الاحتمالات الناتجة عنه بواسطة العلاقة التتابعية التالية:

$$P_{k+1} = P_k \cdot \frac{n-k}{k+1} \cdot \frac{p}{q} \qquad (9-3)$$ لحساب كل احتمال من سابقه

أو بواسطة العلاقة:

$$P_{k-1} = P_k \cdot \frac{k}{n-k+1} \cdot \frac{q}{p} \qquad (10-3)$$ لحساب كل احتمال من لاحقه

ولتطبيق هاتين العلاقتين يتم حساب أحد الاحتمالات P_k بواسطة العلاقة (3-4) الأساسية ثم نقوم بحساب الاحتمالات المتتابعة من إحدى العلاقتين بحسب موقعها من P_k (سابقة أم لاحقة).

مثال 2:

أوجد الاحتمالات المقابلة لعدد الذكور في حالة 11 ولادة.

الحل:

نعلم أن ظاهرة الولادة هي ظاهرة ثنائية، وأن احتمال ولادة الذكر ثابت ويساوي

$p=\frac{1}{2}$ ،وبما أن n=11 ، كبيرة. لذلك نقوم بحساب أحد الاحتمالات المطلوبة من العلاقة (3–4) وليكن الاحتمال p_5 ا لمقابل لـ k=5 فنجد أن:

$$P(X=5)=P_5=C_{11}^5\cdot p^5\cdot q^6=\frac{11!}{5!6!}\cdot\left(\frac{1}{2}\right)^5\cdot\left(\frac{1}{2}\right)^6$$

$$P_5=\frac{462}{2048}=0{,}22559$$

وباستخدام العلاقة (3–9) نحسب الاحتمالات اللاحقة فنجد أن:

$$P_6=P_5\cdot\frac{n-5}{k+1}\cdot\frac{\frac{1}{2}}{\frac{1}{2}}=\frac{462}{2048}\cdot\frac{6}{6}=\frac{462}{2048}=0{,}22559$$

$$P_7=P_6\cdot\frac{n-6}{6+1}\cdot\frac{\frac{1}{2}}{\frac{1}{2}}=\frac{462}{2048}\cdot\frac{5}{7}=\frac{330}{2048}=0{,}16113$$

$$P_8=P_7\cdot\frac{n-7}{7+1}\cdot\frac{\frac{1}{2}}{\frac{1}{2}}=\frac{330}{2048}\cdot\frac{4}{8}=\frac{165}{2048}=0{,}08057$$

$$P_9=P_8\cdot\frac{n-8}{8+1}\cdot\frac{\frac{1}{2}}{\frac{1}{2}}=\frac{165}{2048}\cdot\frac{3}{9}=\frac{55}{2048}=0{,}02685$$

$$P_{10}=P_9\cdot\frac{n-9}{9+1}\cdot\frac{\frac{1}{2}}{\frac{1}{2}}=\frac{55}{2048}\cdot\frac{2}{10}=\frac{11}{2048}=0{,}00537$$

$$P_{11} = P_{10} \cdot \frac{n-10}{10+1} \cdot \frac{\frac{1}{2}}{\frac{1}{2}} = \frac{11}{2048} \cdot \frac{1}{11} = \frac{1}{2048} = 0{,}00049$$

أما الاحتمالات السابقة لـ P_5 فنحسبها من العلاقة (3−10) فنجد أن:

$$P_4 = P_5 \cdot \frac{5}{11-5+1} \cdot \frac{\frac{1}{2}}{\frac{1}{2}} = \frac{462}{2048} \cdot \frac{5}{7} = \frac{330}{2048} = 0{,}16119$$

$$P_3 = P_4 \cdot \frac{4}{11-4+1} \cdot \frac{\frac{1}{2}}{\frac{1}{2}} = \frac{330}{2048} \cdot \frac{4}{8} = \frac{165}{2048} = 0{,}08057$$

$$P_2 = P_3 \cdot \frac{3}{11-3+1} \cdot \frac{\frac{1}{2}}{\frac{1}{2}} = \frac{165}{2048} \cdot \frac{3}{9} = \frac{55}{2048} = 0{,}02685$$

$$P_1 = P_2 \cdot \frac{2}{11-2+1} \cdot \frac{\frac{1}{2}}{\frac{1}{2}} = \frac{55}{2048} \cdot \frac{2}{10} = \frac{11}{2048} = 0{,}00537$$

$$P_0 = P_1 \cdot \frac{1}{11-1+1} \cdot \frac{\frac{1}{2}}{\frac{1}{2}} = \frac{11}{2048} \cdot \frac{1}{11} = \frac{1}{2048} = 0{,}00049$$

ومن الملاحظ أن قيم الاحتمالات في هذا المثال متناظرة حول القيمتين p_5 و P_6 اللتين تقابلان المنوالين $X = 5, X = 6$ (القيمة إن الأكثر احتمالاً).

4 . من الملاحظ أن الحسابات تزداد تعقيداً كلما كبر عدد التجارب n، ولهذا فإنه يمكن حساب قيمة احتمالات التوزيع الثنائي بشكل تقريبي وخاصةً عندما يكون n كبيراً من العلاقة التالية:

$$P_k \approx \frac{1}{\sqrt{2\pi np \cdot q}} e^{\frac{-(k-np)^2}{2np \cdot q}} \qquad (11-3)$$

وتسمى هذه العلاقة بصيغة (لابلاس) للتوزع الطبيعي الذي سندرسه لاحقاً.

5 . يمكن حساب قيمة المنوال للتوزيع الثنائي (القيمة الصحيحة الأكثر احتمالاً) بواسطة المتراجحة:

$$np - q \le Mo \le np + p \qquad (12-3)$$

وحساب الاحتمال المقابل له بواسطة العلاقة التقريبية التالية:

$$P_{Mo} \approx \frac{1}{\sqrt{2\pi np \cdot q}} \qquad (13-3)$$

وبذلك تأخذ العلاقة (3−11) الشكل التالي:

$$P_k \approx P_{\dot{M}o} e^{\frac{-(k-np)^2}{\dot{2}np \cdot q}} \qquad (14-3)$$

مثال 3:

أوجد الاحتمالات التقريبية للمثال (2) السابق.

الحل:

أولاً: نقوم بحساب الاحتمال المنوالي P_{Mo} فنجد أن:

$$P_{Mo} = \frac{1}{\sqrt{2\pi \cdot n \cdot p \cdot q}} = \frac{1}{\sqrt{2\pi \cdot 11 \cdot \frac{1}{2} \cdot \frac{1}{2}}} = 0{,}24057$$

وبذلك نجد أن:

$$P_0 = P_{\dot{M}o} e - \frac{\left(0 - 11 \cdot \frac{1}{2}\right)^2}{2 \cdot 11 \cdot \frac{1}{2} \cdot \frac{1}{2}} = 0{,}24057 \cdot (0{,}004086) = 0{,}00098$$

$$P_1 = P_{\dot{M}o} e^{-\frac{\left(1 - 11 \cdot \frac{1}{2}\right)^2}{5{,}5}} = 0{,}24057 \cdot (0{,}02518) = 0{,}00657$$

$$P_2 = P_{\dot{M}o} e^{-\frac{\left(2 - 11 \cdot \frac{1}{2}\right)^2}{5{,}5}} = 0{,}24057 \cdot (0{,}10782) = 0{,}02594$$

$$P_3 = P_{\dot{M}o} e^{-\frac{\left(3 - 11 \cdot \frac{1}{2}\right)^2}{5{,}5}} = 0{,}24057 \cdot (0{,}32098) = 0{,}07722$$

$$P_4 = P_{\dot{M}o} e^{-\frac{\left(4 - 11 \cdot \frac{1}{2}\right)^2}{5{,}5}} = 0{,}24057 \cdot (0{,}66425) = 0{,}159800$$

$$P_5 = P_{\dot{M}o} e^{-\frac{\left(5 - 11 \cdot \frac{1}{2}\right)^2}{5{,}5}} = 0{,}24057 \cdot (0{,}9555) = 0{,}22988$$

$$P_6 = P_{\dot{M}o} e^{-\frac{\left(6 - 11 \cdot \frac{1}{2}\right)^2}{5{,}5}} = 0{,}24057 \cdot (0{,}9555) = 0{,}22988$$

وبما أن العلاقة (3−14) متناظرة، إذن نستنتج أن:

$$P_7 = 0{,}1598$$
$$P_8 = 0{,}07722$$
$$P_9 = 0{,}02594$$
$$P_{10} = 0{,}00606$$
$$P_{11} = 0{,}00098$$

ومن الواضح أن هذه القيم قريبة جداً من القيم التي حسبناها في المثال السابق. وأن هذا الاقتراب يزداد كلما كبر حجم عدد التجارب n.

وبصورة عامة ينصح باستخدام العلاقة (3−14) عندما تكون $n > 30$، أما عندما تكون $n < 30$، فيمكن الاستعانة بالجدول (I) الخاص باحتمالات هذا التوزيع والملحق في آخر الكتاب.

مثال 4:

صندوق يحتوي على 5 كرات بيضاء و3 كرات حمراء. نسحب كرة من الصندوق ثم نعيدها إليه 6 مرات. ليكن المتحول العشوائي X الذي يمثل عدد الكرات الحمر التي يمكن أن تظهر نتيجة هذا السحب. والمطلوب:

1 . حساب عدد القيم الممكنة للمتحول العشوائي X.

2 . حساب احتمال أن يأخذ X كل القيم الممكنة.

3 . حساب التوقع الرياضي والانحراف المعياري لـ X.

الحل:

1 . إن المتحول X يمثل عدد الكرات الحمراء المسحوبة خلال التجارب الست، وهو يأخذ القيم $X : 0 , 1 , 2 , 3 , 4 , 5 , 6$.

2 . بما أنّ الظاهرة ثنائية (حمراء أو بيضاء)، وأنّ السحب جرب مع الإعادة، فإن احتمال سحب كرة حمراء يساوي مقداراً ثابتاً هو $P = \frac{3}{8}$، وإن الاحتمالات المطلوبة تحسب من قانون التوزيع الثنائي، الذي يأخذ الصيغة التالية:

$$P\,(X=k)=C_6^k\left(\frac{3}{8}\right)^k\cdot\left(\frac{5}{8}\right)^{6-k}$$

وبإعطاء k القيم المتسلسلة من 0 حتى 6 نجد أن:

$$P\,(X=0)=C_6^0\left(\frac{3}{8}\right)^0\cdot\left(\frac{5}{8}\right)^6=0{,}0596$$

$$P\,(X=1)=C_6^1\left(\frac{3}{8}\right)^1\cdot\left(\frac{5}{8}\right)^5=0{,}2146$$

$$P\,(X=2)=C_6^2\left(\frac{3}{8}\right)^2\cdot\left(\frac{5}{8}\right)^4=0{,}3219$$

$$P\,(X=3)=C_6^3\left(\frac{3}{8}\right)^3\cdot\left(\frac{5}{8}\right)^3=0{,}2575$$

$$P\,(X=4)=C_6^4\left(\frac{3}{8}\right)^4\cdot\left(\frac{5}{8}\right)^2=0{,}1159$$

$$P\,(X=5)=C_6^5\left(\frac{3}{8}\right)^5\cdot\left(\frac{5}{8}\right)^1=0{,}0278$$

$$P\,(X=6)=C_6^6\left(\frac{3}{8}\right)^6\cdot\left(\frac{5}{8}\right)^0=0{,}0028$$

ثم نقوم بوضع هذه الاحتمالات في جدول كالتالي:

X	**0**	**1**	**2**	**3**	**4**	**5**	**6**	$\sum$
p_k	**0.0596**	**0.2146**	**0.3219**	**0.2575**	**0.1159**	**0.0278**	**0.0028**	**1**

وأخيراً نشير إلى أن مجموع هذه الاحتمالات يساوي الواحد.

ولحساب التوقع الرياضي نطبق العلاقة:

$$E(x)=n\cdot P=6\cdot\frac{3}{8}=2{,}25$$

ولحساب الانحراف المعياري نحسب التباين أولاً فنجد أنّ:

$$\sigma^2(x) = n \cdot p \cdot q = 6 \cdot \frac{3}{8} \cdot \frac{5}{8} = 1{,}40625$$

وبذلك نجد أنّ الانحراف المعياري:

$$\sigma = \sqrt{n \cdot p \cdot q} = \sqrt{1{,}40625} = 1{,}1858$$

3 -3: قانون توزيع بواسون للمتحول المنقطع X:

عندما يكون لدينا تجربة ثنائية يتولد عنها . كما في التوزيع الثنائي . نتيجتان متنافيتان، وعندما يكون احتمال تحقق حادث النجاح المطلوب في التجربة الواحدة P صغيراً جداً ويكون عدد التجارب n كبيراً جداً، بحيث تكون نهاية الجداء $(n \cdot p)$ مساوية لعدد ثابت موجب λ ، أي عندما يكون لدينا:

$$\lim_{n \to \infty} n.p = \lambda \qquad (15-3)$$

حيث إن: $\lambda > 0$ فإنه لحساب احتمال تحقق الحادث المطلوب k مرة تماماً عند تكرار التجربة n مرة، نستخدم عوضاً عن التوزيع الثنائي توزيعاً جديداً هو توزيع (بواسون).

لنرمز بـ X إلى المتحول الذي يمثل عدد مرات تحقق الحادث الملائم فتكون القيم الممكنة لـ X هي:

$$X = 0\ ,1\ ,2\ ,\ldots\ldots\ ,k, k+1\ ,\ldots . n(\infty)$$

إن احتمال تحقق الحادث الملائم k مرة من أصل n مرة يمكن حسابه من قانون التوزيع الثنائي التالي:

$$P_k = C_n^k p^k \cdot q^{n-k}$$

ولكن بما أن عدد التجارب n كبير جداً فإن عملية الحساب ستكون طويلة ومعقدة. لذلك سنتبع طريقة أخرى ونقوم بتعويض قيمة p بما يعادلها في المعادلة السابقة؛ أي نضع $P = \lambda / n$ وكذلك نضع $q = 1 - \frac{\lambda}{n}$، فإننا نجد أن:

$$P_k = C_n^k \left(\frac{\lambda}{n}\right)^k \cdot \left(1-\frac{\lambda}{n}\right)^{n-k}$$

$$P_k = \frac{n!}{k!(n-k)!} \frac{\lambda^k}{n^k} \frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^k}$$

$$P_k = \frac{n(n-1)(n-2)....(n-k+1)}{k!} \frac{\lambda^k}{n^k} \frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^k}$$

وهذه العلاقة يمكن كتابتها على الشكل التالي:

$$P_k = \frac{n(n-1)(n-2)....(n-k+1)}{n^k} \frac{\lambda^k}{n^k} \frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^k}$$

لاستخراج قانون بواسون نحسب نهاية P_k، عندما $n \to \infty$. لذلك نقوم بحساب نهاية كل كسر على حده فنجد أن:

. نهاية الكسر الأول على اليسار تساوي الواحد، لأن صورته ومخرجه من مرتبة واحدة n^k. وإن أمثال أكبر قوة فيهما تساوي الواحد.

. نهاية $\lambda^k / k!$ تساوي نفسها لأنها لاتتعلق بـ n.

ـ نهاية مقام الكسر الأخير تساوي الواحد لأن $\frac{\lambda}{n} \underset{n\to\infty}{\to} 0$ ولأن أسها k لاعلاقة له بـ n.

. أما نهاية المقدار $\left(1-\frac{\lambda}{n}\right)^n$ فتساوي:

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{n}=\lim_{n\to\infty}\left[\left(1-\frac{\lambda}{n}\right)^{n/\lambda}\right]^{\lambda}=e^{-\lambda}$$

وذلك لأنه لدينا من التحليل الرياضي أن:

$$\lim\left(1+\frac{1}{x}\right)^{x}=e \quad , \quad \lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{x}=e^{-1}$$

وبذلك نجد أن P_k تساوي:

$$P_k=e^{-\lambda}\frac{\lambda^k}{k!} \qquad (16-3)$$

وهو قانون توزيع بواسون.

ولنبرهن على أن هذا القانون يحقق الشرطين المطلوبين، نلاحظ أولاً أن جميع $P_k\geq 0$، وللبرهان على أن مجموع الاحتمالات فيه يساوي الواحد نأخذ المجموع اللانهائي التالي:

$$\begin{aligned}\sum_{k=0}^{+\infty}P(X=k)&=e^{-\lambda}+e^{-\lambda}\frac{\lambda^1}{1!}+e^{-\lambda}\frac{\lambda^2}{2!}+.....+e^{-\lambda}\frac{\lambda^k}{k!}+....\\&=e^{-\lambda}\left(1+\frac{\lambda}{1!}+\frac{\lambda^2}{2!}+.....+\frac{\lambda^k}{k!}+.....\right)\\&=e^{-\lambda}\cdot e^{+\lambda}=1\end{aligned}$$

ملاحظة هامة:

نلاحظ أنّ هذا القانون (3-16) يتعلق بقيمة الثابت λ، ولذلك فإنه يعرف بالرمز $X\approx P(\lambda)$ ليدل على أن المتحول العشوائي X يخضع لقانون بواسون. وبما أنه يشترط أن يكون P صغيراً، فلذلك يسمى أيضاً بقانون الظواهر النادرة. ولكن يمكن تطبيقه على الظواهر شبه المؤكدة أيضاً، وذلك لأنه عندما تكون $p\approx 1$، فإن احتمال تحقق نقوم بتطبيقها على الظاهرة المتممة $q\approx 0$، وعندها تكون $\lambda=n\cdot q$ ، ولكن

عندما نقوم بتطبيقه على الظاهرة المتممة باعتبارها ظاهرة نادرة، يجب الانتباه إلى أن λ تحسب من العلاقة $\lambda = n \cdot q$.

خواص قانون بواسون:

إن خواص قانون بواسون الرئيسية تتعلق بالمنوال والأمل الرياضي والتباين، لذلك سوف نشرح كل خاصة من هذه الخواص بشكل موجز.

ـ المنوال: إن منوال قانون بواسون هو القيمة الصحيحة الموجبة من قيم X الموجودة بين العددين $\lambda - 1$ و λ.

البرهان:

لنفترض أن المنوال هو القيمة k وعندها نجد أن P_k يجب أن يحقق المتراجحتين التاليتين (لأن الاحتمال P_k المقابل للمنوال أكبر من القيمتين المجاورتين له):

$$\frac{P_{k-1}}{P_k} < 1 \qquad , \qquad \frac{P_{k+1}}{P_k} < 1$$

من المتراجحة اليسرى نجد أن:

$$\frac{P_{k-1}}{P_k} = e^{-\lambda} \frac{\lambda^{k-1}}{(k-1)!} \cdot \frac{k!}{\lambda^k \cdot e^{-\lambda}} = \frac{k}{\lambda} < 1 \qquad (17-3)$$

ومن المتراجحة اليمنى نجد أن:

$$\frac{P_{k+1}}{P_k} = e^{-\lambda} \frac{\lambda^{k+1}}{(k+1)!} \cdot \frac{k!}{e^{-\lambda} \cdot \lambda^k} = \frac{\lambda}{k+1} < 1 \qquad (18-3)$$

أي حتى تكون k منوالاً، يجب أن يحقق المتراجحتين:

$$\frac{k}{\lambda} < 1 \quad , \qquad \frac{\lambda}{k+1} < 1$$

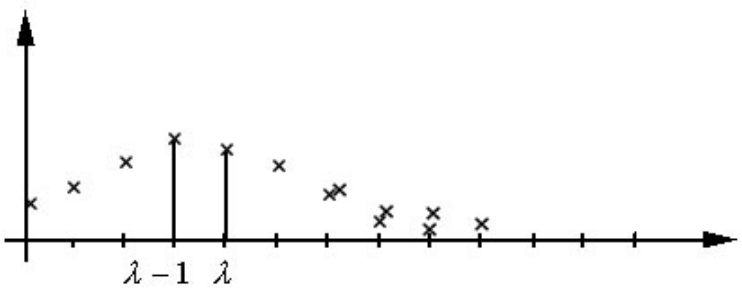
أي أن:

$$\lambda - 1 < k < \lambda \qquad (19-3)$$

وهو المطلوب.

ملاحظة:

عندما تكون λ عدداً صحيحاً، فإن $\lambda-1$ يكون عدداً صحيحاً، وبالتالي يكون للتوزيع قيمتان منواليتان هما λ و $\lambda-1$ كما في الشكل التالي:



شكل رقم (3-4)

التوقع الرياضي والتباين:

إن التوقع الرياضي والتباين لقانون توزيع بواسون يساويا λ_0 ؛ أي أن:

$$E(X)=\lambda \qquad , \qquad \sigma^2_{(x)}=\lambda$$

تابع التوزيع:

إذا كان X خاضعاً لتوزيع بواسون، فإن تابع التوزيع هو التابع الذي يعطينا احتمال أن يأخذ قيمة X قيمة أقل أو تساوي قيمة معينة m؛ أي أن:

$$F(X)=P(X\le m)=\sum_{k=0}^{m}P_k=\sum_{k=0}^{m}e^{-\lambda}\frac{\lambda^k}{k!} \qquad (18-3)$$

وكذلك يكون لدينا الاحتمال المتمم لتابع التوزيع هو:

$$P(X>m)=\sum_{k=m+1}^{\infty}P_k e^{-\lambda}\frac{\lambda^k}{k!} \qquad (19-3)$$

ولكن بما أن حساب مثل هذا المجموع يعتبر أمراً متعذراً، لذلك من الأفضل حساب الاحتمال الأخير عن طريق الحادث المتمم له حيث نجد:

$$F(X>m)=1-F(X)=1-P(X\le m)=1-\sum_{k=0}^{m}e^{-\lambda}\frac{\lambda^k}{k!} \qquad '(19-3)$$

ولحساب احتمال أن يأخذ قيمة محصورة بين قيمتين t و m حيث $t < m$ نطبق العلاقة:

$$F(t < X \le m) = \sum_{k=t}^{m} e^{-\lambda} \frac{\lambda^k}{k!} \qquad (20-3)$$

حساب التوزيع الاحتمالي لقانون بواسون بطريقة عملية:

إن حساب القيم العددية للتوزيع الاحتمالي لقانون بواسون يعتبر أكثر سهولة من حساب التوزيع الاحتمالي لقانون التوزيع الثنائي وخاصةً عندما نستخدم العلاقة الرياضية التي تربط الاحتمالات المتتالية وهي (انظر العلاقة (3-15)):

$$\frac{P_{k+1}}{P_k} = \frac{\lambda}{k+1}$$

والتي تعطينا أن:

$$\frac{P_{k+1}}{P_k} = \frac{\lambda}{k+1} \qquad \text{كل احتمال من سابقه} \qquad (21-3)$$

كما نلاحظ (انظر العلاقة (3-15)) أن:

$$\frac{P_{k-1}}{P_k} = \frac{k}{\lambda}$$

والتي تعطينا أن:

$$P_{k-1} = P_k \cdot \frac{k}{\lambda} \qquad (21-3)'$$

مثال 5:

ليكن X متحولاً عشوائياً خاضعاً لقانون توزيع بواسون الذي فيه $\lambda = 1{,}2$ ، والمطلوب: حساب الاحتمالات المقابلة لقيم X المختلفة.

الحل:

نلاحظ أن قيم المتحول يمكن أن تأخذ القيم التالية:

$$X: 0\ ,\ 1\ ,\ 2\ ,\ 3\ ,\ 4\ ,\ 5\ ,\ 6\ ,\ 7\ ,\ 8$$

وبما أن $\lambda = 1{,}2$ فإنه يمكننا حساب الاحتمالات المطلوبة من القانون مباشرةً كما يلي:

$$P_{(x=0)} = \frac{\lambda^0}{0!} \cdot e^{-1,2} = e^{-1,2} = 0{,}30119$$

$$P_{(x=1)} = \frac{\lambda^1}{1!} \cdot e^{-1,2} = (1{,}2)e^{-1,2} = 0{,}36143$$

$$P_{(x=2)} = \frac{\lambda^2}{2!} \cdot e^{-1,2} = \frac{(1{,}2)^2}{2} e^{-1,2} = 0{,}21686$$

$$P_{(x=3)} = \frac{\lambda^3}{3!} \cdot e^{-1,2} = \frac{(1{,}2)^3}{6} e^{-1,2} = 0{,}08674$$

$$P_{(x=4)} = \frac{\lambda^4}{4!} \cdot e^{-1,2} = \frac{(1{,}2)^4}{24} e^{-1,2} = 0{,}02602$$

$$P_{(x=5)} = \frac{\lambda^5}{5!} \cdot e^{-1,2} = \frac{(1{,}2)^5}{120} e^{-1,2} = 0{,}00625$$

$$P_{(x=6)} = \frac{\lambda^6}{6!} \cdot e^{-1,2} = \frac{(1{,}2)^6}{720} e^{-1,2} = 0{,}00125$$

$$P_{(x=7)} = \frac{\lambda^7}{7!} \cdot e^{-1,2} = \frac{(1{,}2)^7}{5040} e^{-1,2} = 0{,}00021$$

وهكذا يمكن أن نتابع إلى اللانهاية.

ولكننا نلاحظ أن الاحتمالات تصبح صغيرة جداً بعد حد معين من قيم X.

كما يمكننا إجراء هذه الحسابات اعتماداً على العلاقة (3−20) واستخلاص الاحتمالات بشكلٍ متتالٍ.

لذلك نقوم أولاً بحساب الاحتمال الأول P_0 فنجد أن $P_0 = 0{,}30119$.

وبعدها نحسب الاحتمالات المتتالية من العلاقة:

$$P_{k+1} = P_k \frac{\lambda}{k+1}$$

أي أنه للحصول على الاحتمال P_{k+1} نضرب الاحتمال السابق P_k بالنسبة $\frac{\lambda}{k+1}$ وذلك كما هو موضح في الجدول التالي:

جدول (4-2): حساب احتمالات قانون توزيع بواسون حيث $\lambda = 1,2$

قيمة المتحول العشوائي الخاضع لقانون بواسون k	النسبة $\frac{P_{k+1}}{P_k} = \frac{\lambda}{k+1}$	الاحتمال P_k
0	1,2	0,30119
1	0,6	0,36143
2	0,4	0,21686
3	0,3	0,08674
4	0,24	0,02602
5	0,20	0,00625
6	0,1714	0,00125
7	0,15	0,00021
8	-	0,00005
المجموع		≈ 1

مثال 6:

لـيكن W متحـولاً عشـوائياً يخضـع لقـانون بواسـون، الـذي فيـه $\lambda = 1$، ولـيكن لـدينا متحولات أخرى t, x, y, z.... خاضعة لقوانين التوزيع الثنائي التي فيها:

$$P_1 = \frac{1}{5},\ n_1 = 5 \Rightarrow n_1 \cdot P_1 = 1$$

$$P_2 = \frac{1}{10},\ n_2 = 10 \Rightarrow n_2 \cdot P_2 = 1$$

$$P_3 = \frac{1}{20},\ n_3 = 20 \Rightarrow n_3 \cdot P_3 = 1$$

$$P_4 = \frac{1}{40},\ n_4 = 40 \Rightarrow n_4 \cdot P_4 = 1$$

$$P_5 = \frac{1}{80},\ n_5 = 80 \Rightarrow n_5 \cdot P_5 = 0$$

علـى الترتيـب، والتـي جميعهـا تحقـق العلاقـة الشـرطية فـي قـانون بواسـون السـابق: $n \cdot P = 1$.

وبنـاءً علـى مـا تقدم يمكننـا حساب الاحتمـالات المطلوبـة لهذه المتحولات من التوزيـع الثنائي:

$$P_k = C_n^k \cdot P_k \cdot q^{n-k}$$

كمـا يمكـن حسـابها مـن قـانون توزيـع بواسـون الـذي فيـه $\lambda = 1$ باعتبـاره كتعريـب للتوزيعات الثنائية السابقة والذي يعطى بالعلاقة:

$$P_k = e^{-\lambda}\frac{\lambda^k}{k!} = e^{-1} \cdot \frac{1}{k!}$$

لذلك نقوم بإجراء هذه الحسابات بكلتا الطريقتين، ثم نقارن النتائج التي نحصل عليها في حالة مع الحالات الأخرى فنحصل على الجدول التالي:

جدول (4–3): مقارنة الاحتمالات للتوزيعات الثنائية وتوزيع بواسون

توزيع بواسون	التوزيعات الثنائية للمتحولات t, x, y, z						
المتحول	H	....	z	y	x	t	المتحول
$\frac{W}{\lambda=1}$	$n=80$ $P=\frac{1}{80}$		$n=40$ $P=\frac{1}{40}$	$n=20$ $P=\frac{1}{20}$	$n=10$ $P=\frac{1}{10}$	$n=5$ $P=\frac{1}{5}$	
0.3679	0.3656	---	0.3632	0.3585	0.3487	0.3277	0
0.3679	0.3702	---	0.3726	0. 3774	0.3874	0.4096	1
0.1839	0.1851	---	0.1863	0.1887	0.1937	0.2048	2
0.0613	0.0609	---	0.0605	0.0596	0.0574	0.0512	3
0.0153	0.0148	---	0.0143	0.0133	0.0112	0.0064	4
0.0031	0.0029	---	0.0026	0.0022	0.0015	0.0003	5
0.0005	0.0004	---	0.0004	0.0003	0.0001	-	6
0.0001	0.0001	---	0.0001	-	-	-	7
-	---	---	---	-	-	-	8
-	-	-	-	-	-	-	-
1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	المجموع

وبمقارنة هذه الحسابات نلاحظ أن قيم الاحتمالات في كل سطر (الذي يقابل قيمة معينة للمتحول)، تقترب من قيم توزيع بواسون الذي فيه $\lambda = 1$ ، أي أن توزيع بواسون يصبح تقريباً مقبولاً للتوزيع الثنائي عند تصبح n أكثر من 50.

مثال 7:

ليكن متوسط عدد المخابرات التي تلقاها مقسم مابين الساعة العاشرة والساعة الحادية عشرة هو 1,8 مخابرة بالدقيقة. والمطلوب حساب احتمال أن يكون لدينا خلال الدقيقة بين الساعة 10,53 و10,54، أحد الحالات التالية.

1 . عدم وجود أي مخابرة.

2 . مخابرة واحدة.

3 . مخابرتين.

4 . على الأقل مخابرتين.

5 . أكثر من مخابرتين.

الحل:

إن عدد المخابرات في الدقيقة يمثل متحولاً عشوائياً X يخضع لقانون بواسون. وإن ثابت هذا القانون يساوي متوسط عدد المخابرات في الدقيقة. أي: $\lambda = 1,8$

وإن قانون التوزيع يأخذ الشكل التالي:

$$P(X=k) = e^{-1,8}\frac{(1,8)^k}{k!} \qquad k = 0,1,\ 2,3....$$

1 . احتمال عدم استقبال أي مخابرة:

$$P(X=0) = e^{-1,8}\frac{(1,8)^0}{0!} = e^{-1,8} = 0,16529$$

2 . احتمال استقبال مخابرة واحدة:

$$P(X=1) = e^{-1,8}\frac{(1,8)^1}{1!} = 0,29752$$

3 . احتمال استقبال مخابرتين:

$$P(X=2) = e^{-1,8}\frac{(1,8)^2}{2!} = 0,26776$$

4 . احتمال استقبال مخابرتين على الأقل:

$$P(X \geq 2) = 1 - \left[P(X=0) + P(X=1)\right]$$
$$= 1 - \left[0,16529 + 0,29752\right] = 0,53719$$

5 . احتمال استقبال أكثر من مخابرتين:

$$P(X > 2) = 1 - \left[P(X=0) + P(X=1) + P(X=2)\right]$$
$$= 1 - \left[0,16529 + 0,29752 + 0,26776\right] = 0,27123$$

مثال 8:

طريق جبلي تجتازه 50 سيارة في الساعة، فإذا كان احتمال وقوع حادث عليه كل ساعة يساوي 0,04، والمطلوب حساب مايلي:

. احتمال عدم وقوع أي حادث.

. احتمال وقوع حادث واحد فقط.

. احتمال وقوع حادثين فقط.

. احتمال وقوع حادثين على الأقل.

. التوقع الرياضي لعدد الحوادث ثم تباينه.

الحل:

لنرمز بـ X لعدد الحوادث التي يمكن أن تقع خلال ساعة على الطريق فيكون:

$$X : 0, 1, 2, 3.....50$$

ولإيجاد الاحتمالات المطلوبة نطبق قانون بواسون وذلك لأن P صغيرة جداً و N=50 تعتبر كبيرة نسبياً، فنجد أن قيمة الوسيط λ تساوي:

$$\lambda = n \cdot P = 50 \cdot 0{,}04 = 2$$

وبذلك يمكننا حساب الاحتمالات المطلوبة من قانون بواسون مباشرة كمايلي:

$$P_0 = e^{-2} \frac{2^0}{0!} = 0{,}135$$

$$P_1 = e^{-2} \frac{2}{1!} = 0{,}271$$

$$P_2 = e^{-2} \frac{2^2}{2!} = 0{,}271$$

$$P(X \geq 2) = 1 - P(X < 2) = 1 - (0{,}135 + 0{,}271) = 0{,}594$$

$$E(X) = \lambda = 2$$

$$\sigma^2(x) = \lambda = 2$$

3-4: قانون التوزيع المنتظم العام (للمتحولات المستمرة):

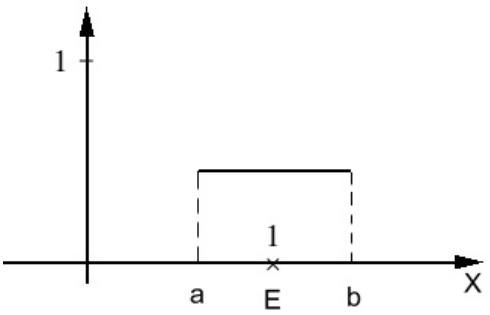
عندما يكون المتحول العشوائي X مستمراً، فإن عدد قيمه الممكنة يصبح لا نهائياً، وعندها يتعذر علينا إعداد جدول يتضمن تلك القيم واحتمالاتها المقابلة. لذلك فإن توزع الاحتمالات على قيم المتحولات يعطى بواسطة علاقة رياضية مناسبة وبدلالة المتحول العشوائي X وتأخذ الشكل العام f(x). وهناك عدد من هذه التوزيعات وأبسطها التوزيع المنتظم.

يعرف قانون التوزيع المنتظم للمتحول العشوائي المستمر X، والمعرف على المجال $[a,b]$ بالعلاقة التالية:

$$f(x)=\begin{cases}\dfrac{1}{b-a} & : \quad a\le x\le b\\ 0 & \text{عدا ذلك}\end{cases} \qquad (27-3)$$

ومن خواصه أنه يحقق الشرطين $f(x)\ge 0$, $\int_{-\infty}^{+\infty} f(x)dx=1$

وهو يرسم الشكل البياني التالي (مستقيم موازي لـ X 0).



الشكل رقم (3–5) قانون التوزيع المنتظم

وإن توقعه الرياضي يساوي:

$$E(x)=\int_{-\infty}^{+\infty} f(x)dx=\int_{-\infty}^{+\infty} x\cdot\frac{1}{b-a}dx=\frac{1}{b-a}\left[\frac{x^2}{2}\right]_a^b$$

$$=\frac{1}{b-a}\left[\frac{b^2-a^2}{2}\right]=\frac{a+b}{2}$$

$$E(x)=\frac{a+b}{2} \qquad \text{(يقع في منتصف المجال)} \qquad (28-3)$$

وإن تباينه يساوي:

$$\sigma^2_{(x)} = \int_{-\infty}^{+\infty} (x-E)^2 \cdot f(x)dx = \int_{-\infty}^{+\infty} \left(x - \frac{a+b}{2}\right)^2 \cdot \frac{1}{b-a} \cdot dx$$

$$= \frac{1}{b-a} \int_{-\infty}^{+\infty} (x - \frac{a+b}{2})^2 dx = \frac{1}{b-a} \cdot \left[\frac{\left(x - \frac{a+b}{2}\right)^3}{3} \right]_a^b$$

$$= \frac{1}{b-a} \left[\frac{\left(b - \frac{a+b}{2}\right)^3}{3} - \frac{\left(a - \frac{a+b}{2}\right)^3}{3} \right]$$

$$= \frac{1}{b-a} \left[\frac{(b-a)^3}{24} - \frac{(a-b)^3}{24} \right]$$

$$= \frac{1}{b-a} \left[\frac{2(b-a)^3}{24} \right] = \frac{(b-a)^2}{12}$$

$$\sigma^2_{(x)} = \frac{(b-a)^2}{12} \qquad (29-3)$$

وبذلك نجد أن الانحراف المعياري لـX المنتظم يساوي:

$$\sigma(x) = \sqrt{\frac{(b-a)^2}{12}} = \frac{b-a}{\sqrt{12}} \qquad (30-3)$$

أما تابع التوزيع الاحتمالي فيحسب من العلاقة:

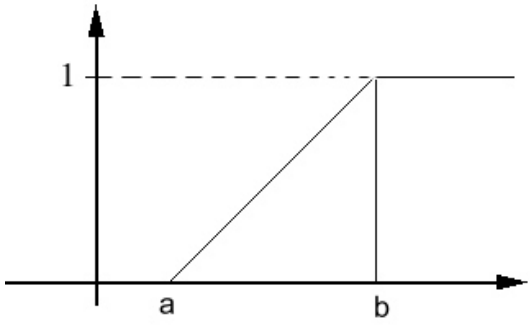
$$F(x) = p(X \le x) = \int_{-\infty}^{x} f(x)dx =$$

$$\int_{-\infty}^{x} \frac{1}{b-a} \cdot dx = \frac{1}{b-a}[x]_a^x = \frac{x-a}{b-a}$$

ويمكن كتابة تابع التوزيع على الشكل التالي:

$$F(x)=1\begin{cases}0 & x<a\\ \dfrac{x-a}{b-a} & a\le x\le b\\ 0 & x>b\end{cases} \qquad (31-3)$$

وهو يرسم الشكل البياني التالي (مستقيم ميله $\left(\dfrac{1}{b-a}\right)$.



الشكل رقم (3–6) تابع التوزيع المنتظم

مثال 9:

لنفترض أن المدة التي نستغرقها رحلة الحافلة من اللاذقية إلى دمشق بالساعات هي متحول عشوائي X ، يأخذ قيمه في المجال $[3,4]$. ويخضع للتوزيع المنتظم. والمطلوب:

. كتابة الصيغة الرياضية لقانون توزيع X، وحساب توقعه وتباينه وانحرافه المعياري.

. حساب الاحتمالات $P(X\le 3.5)$ $p(3.35\le X\le 3.75)$

الحل: إن الصيغة الرياضية لقانون التوزيع المنتظم في هذه الحال هي:

$$f(x)=\begin{cases}\dfrac{1}{4-3} & : \quad 3\le x\le 4\\ 0 & \text{عدا ذلك}\end{cases} \qquad (27-3)$$

وإن توقعه الرياضي يساوي:

$$E(x)=\frac{a+b}{2}=\frac{3+4}{2}=3.5 \quad \text{ساعة}$$

وتباينه يساوي:

$$\sigma^2_{(x)}=\frac{(a+b)^2}{12}=\frac{(3+4)^2}{12}=\frac{1}{12}$$

وانحرافه المعياري:

$$\sigma_{(x)}=\sqrt{\frac{1}{12}}=\frac{1}{\sqrt{12}}$$

أما الاحتمالات المطلوبة فتساوي:

$$p(x\le 3.5)=F(3.5)=\frac{3.5-3}{4-3}=0.50$$

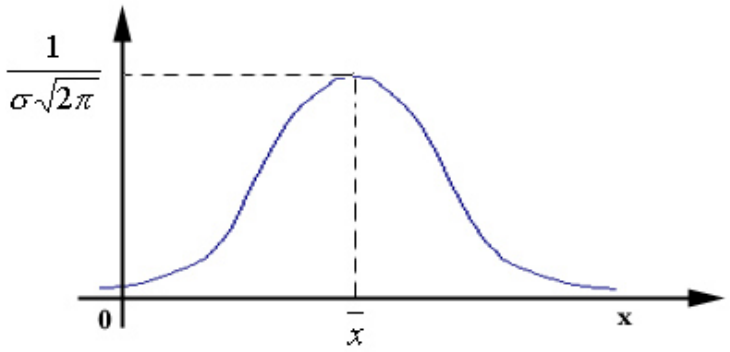
$$p(3.35\le x\le 3.75)=F(3.75)=\frac{3.75-3}{4-1}-\frac{3.35-3}{41}=0.40$$

3-5: التوزيع الطبيعي العام:

يعتبر التوزيع الطبيعي من أشهر وأهم التوزيعات الاحتمالية المستمرة وهو يعرف بالعلاقة التالية:

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}.e^{-\frac{1}{2}\left(\frac{x-\bar{x}}{\sigma}\right)^2} \qquad (32-3)$$

إذ إن $-\infty<X<+\infty$، وإن $\bar{x}$ هو المتوسط وإن σ الانحراف المعياري لـx ، ولذلك يرمز له بالرمز $N(\bar{x},\sigma^2)$، كما يسمى أحياناً بقانون (غوص . لابلاس). وهو يرسم منحنياً متناظراً حول $\bar{x}$ كما هو مبين في الشكل التالي:



الشكل رقم (3-7)

خواص التوزيع الطبيعي العام:

1 . إن قيم التوزيع غير سالبة، أي أن $f(x) > 0$ ، وهذا واضح في الشكل البياني.

2 . إن المساحة تحت المنحني وفق المحور Ox يساوي الواحد، أي أن:

$$\int_{-\infty}^{+\infty} f(x)dx = 1 \qquad (33-3)$$

3 . التوقع الرياضي للمتحول المستمر يعرف بالعلاقة:

$$E(x) = \int_{-\infty}^{+\infty} x \cdot f(x)dx \qquad (34-3)$$

وهو يساوي في حالة التوزيع الطبيعي ما يلي:

$$E(x) = \bar{x} \qquad (35-3)$$

4 . التباين:

إن تباين المتحول المستمر يعرف بالعلاقة:

$$\sigma^2(x) = \int_{-\infty}^{+\infty} (x-\bar{x})^2 \cdot f(x) \cdot dx \qquad (36-3)$$

وهو يساوي في حالة التوزيع الطبيعي ما يلي:

$$\sigma^2(x) = \sigma^2 \qquad (37-3)$$

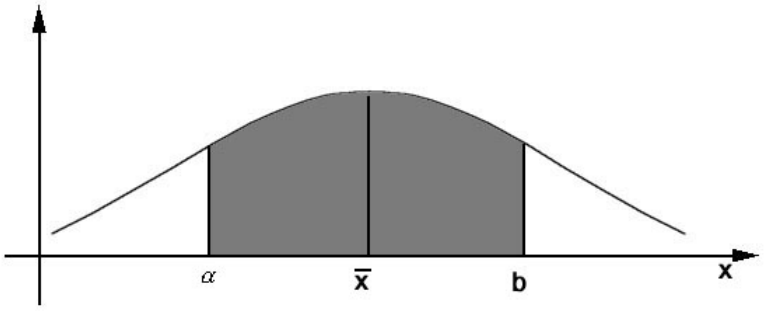
وبذلك نجد أن انحرافه المعياري يساوي:

$$\sigma(x)=\sigma$$

5 . إن احتمال أن يقع X في المجال $[a\ ,\ b]$ يساوي:

$$P(a\le X\le b)=\int_a^b f(x)dx=\int_b^b \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\bar{x}}{\sigma}\right)^2}\cdot dx \qquad (38-3)$$

وهو يساوي المساحة الواقعة تحت المنحني والمحصورة وبين النقطتين a وb. كما هو موضح على الشكل التالي:



الشكل رقم (3-8)

ولحساب قيمة هذا الاحتمال نجري التحويل التالي:

$$t=\frac{x-\bar{x}}{\sigma}\Rightarrow x=\sigma t+\bar{x}$$

لإيجاد الحدود الجديدة للتكامل السابق نجد أنّ:

عندما $x=a$ فإن $t=\dfrac{a-\bar{x}}{\sigma}$

عندما $x=b$ فإن $t=\dfrac{b-\bar{x}}{\sigma}$

كما نجد أن: $dx=\sigma dt$.

وبالتعويض في العلاقة (3-28) نجد أن:

$$P(a \le X \le b) = \int_{\frac{a-\bar{x}}{\sigma}}^{\frac{b-\bar{x}}{\sigma}} \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-1}{2}\cdot t^2} \cdot \sigma dt$$

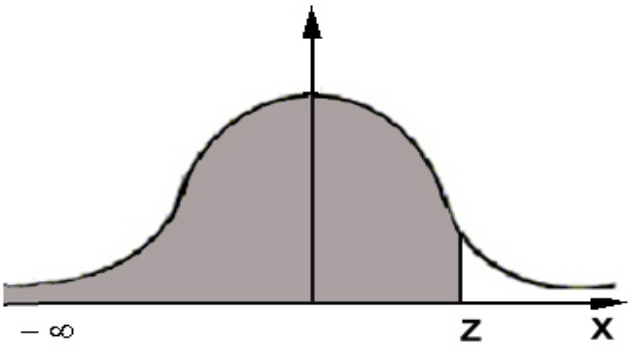
ولحد

$$P(a \le X \le b) = \frac{1}{\sqrt{2\pi}} \cdot \int_{\frac{a-\bar{x}}{\sigma}}^{\frac{b-\bar{x}}{\sigma}} \cdot e^{\frac{-1}{2}\cdot t^2} \cdot \sigma dt \qquad (39-3)$$

ساب التكامل الأخير تم وضع جداول خاصة تعطينا قيمة التكامل التالي:

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{z} \cdot e^{\frac{-t^2}{2}} \cdot dx = p(x \le z) \qquad (40-3)$$

وإن قيمة هذا التكامل تساوي المساحة الواقعة تحت المنحني من $-\infty$ حتى النقطة Z كما في الشكل التالي:



الشكل رقم (3-9)

ويسمى هذا التكامل بتابع الاحتمالات وأن قيم $\Phi(z)$ تؤخذ من الجدول رقم (III) في آخر الكتاب، وذلك حسب قيم Z. وهنا نشير إلى أن:

$$\Phi(-z) = 1 - \Phi(z) \qquad (41-3)$$

وبالعودة إلى التكامل (3-39) نجد أنه من الممكن تحويله إلى تكاملين من الشكل (3-40) كما يلي:

$$P(a \le X \le b) = \frac{1}{\sqrt{2\pi}} \cdot \int_{\frac{a-\bar{x}}{\sigma}}^{\frac{b-\bar{x}}{\sigma}} \cdot e^{\frac{-t^2}{2}} \cdot dt$$

$$= \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{\frac{b-\bar{x}}{\sigma}} \cdot e^{\frac{-t^2}{2}} \cdot dt - \int_{-\infty}^{\frac{a-\bar{x}}{\sigma}} \cdot e^{-t^2/2} \cdot dt$$

إذن يكون لدينا:

$$P(a \le X \le b) = \Phi\left(\frac{b-\bar{x}}{\sigma}\right) - \Phi\left(\frac{a-\bar{x}}{\sigma}\right) \qquad (42-3)$$

وهي العلاقة التي تعطينا احتمال أن يقع X في المجال $[a\ ,\ b]$ بواسطة قيم التابع $\Phi(z)$ عند النقطتين:

$$Z_1 = \frac{a-\bar{x}}{\sigma} \quad , \quad Z_{2=} \frac{b-\bar{x}}{\sigma}$$

أما إذا أردنا حساب احتمال أن يكون X أصغر من قيمة C فإننا نجد أن:

$$P(X \le C) = \int_{-\infty}^{C} f(x)dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{C-\bar{x}}{\sigma}} e^{-\frac{t^2}{2}} \cdot dt$$

$$= \Phi\left(\frac{C-\bar{x}}{\sigma}\right) \qquad (43-3)$$

وكذلك نجد أن:

$$P(X \ge d) = 1 - P(X \le d) = 1 - \Phi\left(\frac{d-\bar{x}}{\sigma}\right) \qquad (44-3)$$

مثال 10:

لنفترض أن طول الشخص هو متحول عشوائي X. وأنه خاضع للتوزيع الطبيعي الذي متوسطه $\bar{x} = 170\,\text{cm}$، وانحرافه المعياري $\sigma = 5$. فما هي الصيغة المحددة للتوزيع

الطبيعي لـ X؟ وما هو احتمال أن يقع X في المجال [160 , 185]؟ واحتمال أن يكون أكبر من 170؟ واحتمال أن يكون أصغر من 160؟

الحل:

إن الصيغة الرياضية للتوزيع الطبيعي المطلوب هي:

$$f(x)=\frac{1}{5\sqrt{2\pi}}\cdot e^{\frac{-1}{2}\cdot\left(\frac{x-170}{5}\right)^2}$$

واحتمال أن يقع X في المجال [160 , 185] يساوي اعتماداً على (3-42):

$$P(160\le X\le 185)=\Phi\left(\frac{185-170}{5}\right)-\Phi\left(\frac{160-170}{5}\right)=$$
$$=\Phi(3)-\Phi(-2)=\Phi(3)-[1-\Phi(2)]$$
$$=\Phi(3)+\Phi(2)-1$$

ومن الجدول الملحق رقم (III) نجد أن:

$\Phi(3)=0{,}99865$ وأن $\Phi(2)=0{,}9772$.

وبذلك مكون لدينا:

$$P(160\le X\le 185)=0{,}99865+0{,}9772-1=0{,}97585$$

وأن احتمال أن يكون X أكبر من 170، ثم أصغر من 160 يساويان:

$$P(X\ge 170)=1-P(X\le 170)=1-\Phi\left(\frac{170-170}{\sigma}\right)=$$
$$=1-\Phi(0)=1-0{,}5=0{,}5$$
$$P(X\le 160)=\Phi\left(\frac{160-170}{\sigma}\right)=1-\Phi(-2)=$$
$$=\Phi(2)=1-0{,}9772=0{,}0228$$

ملاحظة:

إن احتمال أن يأخذ المتحول المستمر X قيمة معينة مثل x_0 يساوي الصفر لأن:

$$P(X = x_0) = P(x_0 \le X \le x_0) = \int_{x_0}^{x_0} f(x) \cdot dx = 0$$

أي أن:

$$P(X = x_0) = 0 \qquad \forall x_0 \qquad (45-3)$$

6 . إن احتمال أن يقع المتحول X في مجال متناظر حول المتوسط $\bar{x}$ ونصف طوله من مضاعفات الانحراف المعياري يساوي ما يلي:

$$P(\bar{x} - k\sigma \le X \le \bar{x} + k\sigma) = \left\{\frac{\bar{x} + k\sigma - \bar{x}}{\sigma}\right\} - \Phi\left\{\frac{\bar{x} - k\sigma - \bar{x}}{\sigma}\right\}$$

$$P(\bar{x} - k\sigma \le X \le \bar{x} + k\sigma) = \Phi(k) - \Phi(-k) = \Phi(k) - [1 - \Phi(k)]$$

$$P(\bar{x} - k\sigma \le X \le \bar{x} + k\sigma) = 2\Phi(k) - 1 \qquad (36-3)$$

وكتطبيق على ذلك نعطي k القيم 1 ثم 2 ثم 3 فنجد أنه عندما تكون k=1 فإننا نحصل على المجال الأول $[\bar{x} - \sigma \ , \ \bar{x} + \sigma]$ وبالتالي يكون لدينا:

$$P(\bar{x} - \sigma \le X \le x + \sigma) = 2\Phi(1) - 1 = 2(0{,}8413) - 1 = 0{,}6826$$

وعندما تكون $k=2$ فإننا نحصل على المجال $[\bar{x} - 2\sigma \ , \ \bar{x} + 2\sigma]$ وبالتالي يكون لدينا:

أما $P(\bar{x} - 2\sigma \le X \le \bar{x} + 2\sigma) = 2\Phi(2) - 1 = 2(0{,}9772) - 1 = 0{,}9544$

ا عندما نجعل $k = 2{,}58$ فإننا نحصل على المجال:

$[\bar{x} - 2{,}58\sigma \ , \bar{x} + 2{,}58\sigma]$ وبالتالي يكون لدينا:

$$P(\bar{x} - 2{,}58\sigma \le X \le \bar{x} + 2{,}58\sigma) = 2\Phi(2{,}58) - 1 =$$

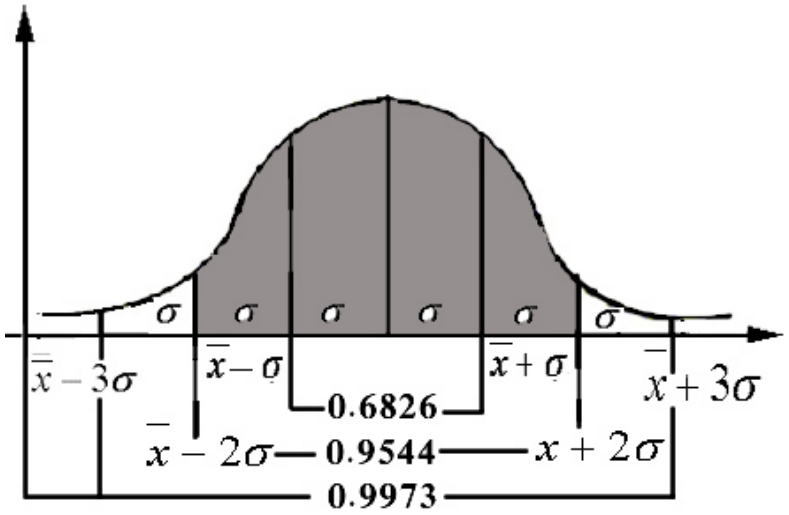
$$= 2(0{,}9951) - 1 = 0{,}9902$$

وكذلك نجد أنه عندما نجعل $k = 3$، فإننا نحصل على:

$$P\left(\bar{x} - 3\sigma \le X \le \bar{x} - 3\sigma\right) = 2\Phi(3) - 1 = 2(0{,}99865) - 1 = 0{,}9973$$

وهكذا نلاحظ أنه إذا جعلنا نصف طول المجال مساوياً لـ σ فإن الاحتمال المقابل يساوي 0,6826، وإذا جعلنا نصف طول المجال مساوياً لـ 2σ فإن الاحتمال المقابل يساوي 0,9544، وإذا جعلناه مساوياً لـ $2{,}58\sigma$ فإن الاحتمال المقابل له يساوي 0,9902، وإذا جعلناه مساوياً لـ 3σ فإن الاحتمال المقابل له يساوي 0,9973؛ أي أن معظم القيم تصبح واقعة في المجال الأخير.

والشكل رقم (3–10) يوضح لنا مواقع هذه المجالات والاحتمالات المقابلة لها.



الشكل رقم (3–10)

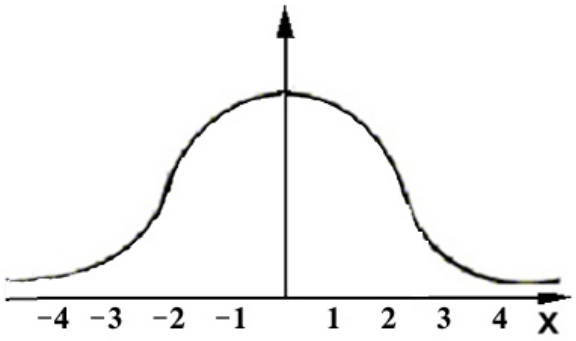
3–6 : قانون التوزيع الطبيعي المعياري:

وهو حالة خاصة من التوزيع الطبيعي العام. يكون فيها $\bar{x} = 0$ و $\sigma = 1$ ولذلك يرمز له بالرمز $N(0,1)$، ويعطى بالعلاقة التالية:

$$\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{\frac{-t^2}{2}} \qquad (47-3)$$

حيث أن: $-\infty < t < \infty$

وهو يرسم على المحاور الإحداثية منحنياً متناظراً بالنسبة للمحور الشاقولي كما هو مبين على الشكل التالي:



الشكل رقم (3−11)

خواصه:

1 . إن $\varphi(t) \geq 0$.

2 . إن $\int_{-\infty}^{+\infty} \varphi(t)dt = 1$ $\qquad (48-3)$

3 . لقد ذكرنا أن متوسطه $\bar{x} = 0$ وتباينه $\sigma^2 = 1$.

4 . إنه متناظر بالنسبة للمحور الشاقولي.

5 . يبرهن في نظرية الاحتمالات على أنه:

إذا كان X خاضعاً للتوزيع الطبيعي العام الذي متوسطه $\bar{x}$ وتباينه σ^2 فإن المتحول العشوائي t المعرف بالعلاقة:

$$t = \frac{x - \bar{x}}{\sigma} \qquad (49-3)$$

يكون خاضعاً للتوزيع الطبيعي المعياري $N(0,1)$ ، لذلك يسمى المتحول t بالمتحول المعياري لـ X.

5 . إن حساب الاحتمالات للمتحولات الخاضعة لـ $N(0,1)$ تتم كمايلي:

$$P(t_1 \le t \le t_2) = \frac{1}{\sqrt{2\pi}} e \int_{t_1}^{t_2} e^{-\frac{t^2}{2}} \cdot dt$$

$$= \frac{1}{\sqrt{2\pi}} e \int_{-\infty}^{t_2} e^{-\frac{t^2}{2}} \cdot dt - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t_1} e^{-\frac{t^2}{2}} \cdot dt$$

وبالاستناد إلى العلاقة (3−40) نجد أن:

$$P(t_1 \le t \le t_2) = \Phi(t_2) - \Phi(t_1) \qquad (50-3)$$

وإذا أخذنا نقطتين متناظرتين مثل $t_2 = k$, $t_1 = -k$ فإننا نجد أن:

$$P(-k \le t \le k) = \Phi(k) - \Phi(-k) = \Phi(k) - [1 - \Phi(k)]$$

$$P(-k \le t \le k) = 2\Phi(k) - 1 \qquad (51-3)$$

ويتم استخراج قيم $\Phi(k)$ من جدول تابع الاحتمالات المعياري، هو الجدول رقم III الوارد في الملحق.

مثال 11 :

إذا كان X هو مقدار إنتاجية الهكتار من القمح في مختلف المزارع وكان متوسطه $\text{Ton}\,\bar{x} = 4000$ وتباينه $\sigma^2 = 490000$، وكان X خاضعاً للتوزيع الطبيعي. فما هو احتمال أن تكون إنتاجية الهكتار محصورة بين 3000 و 6000 طن؟

يمكننا حساب هذا الاحتمال اعتماداً على العلاقة (3− 42)، ولكننا هنا سنستخدم أسلوباً آخر يعتمد على قانون التوزيع الطبيعي المعياري: لذلك نحسب قيمتي المتحول المعياري t المقابلتين لقيمتي X الحديتين فنجد أن:

$$t_1 = \frac{x_1 - \bar{x}}{\sigma} = \frac{3000 - 4000}{700} = -1{,}43$$

$$t_2 = \frac{x_2 - \bar{x}}{\sigma} = \frac{6000 - 4000}{700} = 2{,}85$$

وبذلك نجد أن:

$$P(3000 \le X \le 6000) = P(-1{,}43 \le t \le 2{,}85)$$
$$= \Phi(2{,}85) - \Phi(-1{,}43) = \Phi(2{,}85) - [1 - \Phi(1{,}43)]$$
$$= \Phi(2{,}86) + \Phi(1{,}43) - 1$$
$$= 0{,}99788 + 0{,}92364 - 1 = 0{,}92152$$
$$P(3000 \le X \le 6000) = 0{,}92152$$

مثال 12:

إذا كان t خاضعاً لـ $N(0,1)$ فاحسب أن يقع t في كل من المجالات التالية:

$[-1,1]$, $[-2,2]$, $[-2{,}58 \cdot 2{,}58]$, $[3,3]$.

من العلاقة (3−51) نجد أن:

$$P(-1 \le t \le 1) = 2\Phi(1) - 1 = 2(0{,}8413) - 1 = 0{,}6826$$
$$P(-2 \le t \le 2) = 2\Phi(2) - 1 = 2(0{,}97725) - 1 = 0{,}9545$$
$$P(-2{,}58 \le t \le 2{,}58) = 2\Phi(2{,}58) - 1 = 2(0{,}995) - 1 = 0{,}9902$$
$$P(-3 \le t \le 3) = 2\Phi(3) - 1 = 2(0{,}99865) - 1 = 0{,}9973$$

3−7: قانون التوزيع χ^2 (كاي مربع):

يعرف هذا القانون بالعلاقة التالية:

$$\chi_n^2(x) = \frac{x^{\frac{n}{2}-1} \cdot e^{\frac{-x}{2}}}{2^{\frac{n}{2}} \cdot \Gamma\left(\frac{n}{2}\right)} \qquad (52-3)$$

حيث أن $x > 0$، وأن $\Gamma(\frac{n}{2})$ هو قيمة التكامل $\Gamma(\frac{n}{2})$ المعرف بالعلاقة:

$$\Gamma(\frac{n}{2}) = \int_0^{\infty} x^{\frac{n}{2}-1} e^{-x} \cdot dx = \text{قيمة عددية}$$

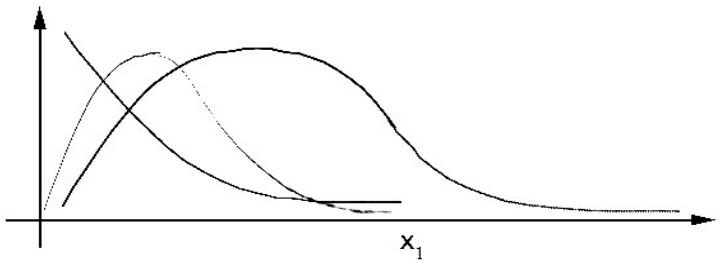
ومن خواص قيمة التكامل $\Gamma(\frac{n}{2})$ أنها مرتبطة بالقيم الأدنى منها بالعلاقة:

$$\Gamma(P) = (P-1) \cdot \Gamma(P-1)$$

أو بالعلاقة:

$$\Gamma(P+1)=P\cdot\Gamma(P)$$

حيث أن n عدد صحيح موجب، يطلق عليه اسم عدد درجات الحرية، إن هذا التوزيع يرسم منحنيات مختلفة وذلك تبعاً للعدد n، كما في الشكل التالي:



الشكل رقم (3-12)

خواصه:

إن $\chi^2(x)$ يحقق الخاصيتين التاليتين:

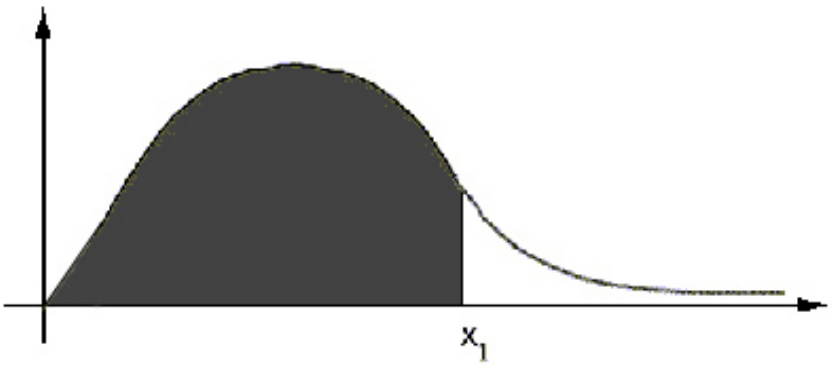
1 . من الواضح أن $\chi_n^2(x)\geq 0$.

2 . إن المساحة التي تحت المنحني تساوي الواحد.

أما تابع التوزيع فيساوي:

$$P(X\leq x_1)=\frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}\int_0^{x_1}x^{\frac{n-2}{2}}\cdot e^{-\frac{x}{2}}\cdot dx \qquad (53-3)$$

وذلك بشرط أن تكون $x_1>0$ ، ويمثل هذا الاحتمال القيمة المظللة التالية المبينة على الشكل التالي:

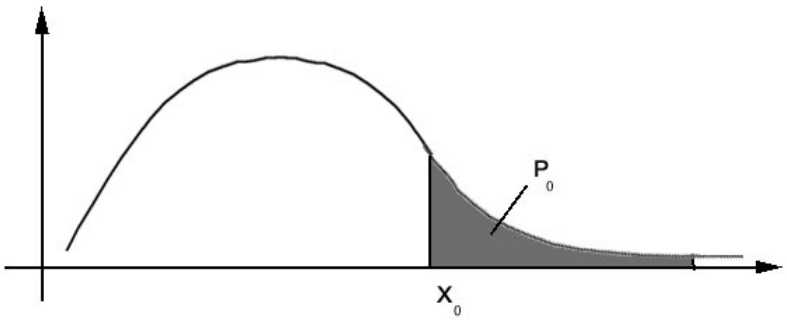


الشكل رقم (3-13)

إلا أن الاحتمال الأكثر أهميةً من تابع التوزيع، هو احتمال أن تأخذ X الخاضع لـ $\chi^2(x)$ قيمة أكبر من قيمة x_0 (بشرط أن تكون $x_0 > 0$)، وهو يساوي:

$$P(X \geq x_0) = \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} \int_{x_0}^{+\infty} x^{\frac{n-2}{2}} \cdot e^{-\frac{x}{2}} \cdot dx = P_0 \qquad (54-3)$$

وبطريقة معاكسة يمكننا اعتماداً على هذه العلاقة إيجاد قيمة x_0 التي تقابل قيمة معينة للاحتمال P_0. ونسمى الاحتمال P_0 باحتمال الدلالة وذلك ضمن درجات الحرية المحددة، ويمثل هذا الاحتمال بالمساحة المبينة على الشكل التالي:



الشكل رقم (3-14)

وللعلاقة (3–54) تطبيقات كثيرة وخاصة في مجال اختبار الفرضيات ذات الجانب الواحد. لذلك فإن القيم الجدولية لـ x_0 تحسب وتوضع ضمن جداول خاصة بدلالة n و P_0، مثل الجدول IV الملحق بآخر الكتاب.

أما الاحتمال المهم الآخر فهو احتمال أن يأخذ X الخاضع لـ $\chi^2(x)$ قيمة محصورة في مجال معين مثل المجال $[x_1 \ , \ x_2]$ والذي يساوي:

$$P(x_1 \le X \le x_2) = \frac{1}{2^{\frac{n}{2}} \cdot \Gamma\left(\frac{n}{2}\right)} \int_{x_1}^{x_2} x^{\frac{n-2}{2}} \cdot e^{\frac{-x}{2}} \cdot dx = P_0 \qquad (55-3)$$

وتستخدم هذه العلاقة في اختبار الفرضيات ذات الاتجاهين لتحديد حدي منطقة القبول.

ومن المتحولات العشوائية المشهورة والخاضعة لتوزيع $\chi_n^2(x)$ نذكر ما يلي:

3 . إذا كان لدينا n متحولاً عشوائياً $X_1 \ , \ X_2 \ , \ X_3 ,\ldots\ldots, X_n$ وكان كل منها يخضع للتوزيع الطبيعي العام وتوقعها وتباينها يساويان:

$$E(X_i) = 0 \quad \forall i$$
$$\sigma^2(x_i) = \sigma_i^2 \ \ (i : 1 , 2 , 2 ,\ldots\ldots, n)$$

فعندها يكون المقدار χ^2 المعرف بالعلاقة:

$$X^2 = \sum_{i=1}^{n} \frac{x^2}{\sigma_i^2}$$

خاضعاً للتوزيع $\chi_n^2(x)$ ذي n درجة حرية. ومن خلال هذا المثال تم استنباط التوزيع $\chi_n^2(x)$.

2 . وكحالة خاصة للمثال السابق، نفرض أن كلاً من المتحولات $X_1 , X_2 , X_3 ,.., X_n$ خاضع للتوزيع الطبيعي المعياري $N(0,1)$؛ أي:

$$E(X_i) = 0 \quad \forall i$$
$$\sigma^2(x_i) = 1$$

فعندها يكون المقدار χ^2 المعرف بالعلاقة:

$$X^2 = \sum_{i=1}^{n} x_i^2$$

ويكون خاضعاً للتوزيع $\chi_n^2(x)$ ذي n درجة حرية.

4 . من الطبيعي أن ننظر إلى الفروقات $(x_i - \bar{x})$ لمتحول عشوائي على أنها هي الأخرى متحولات عشوائية توقعها الرياضي يساوي الصفر، وتباينها يساوي σ^2 لذلك فإننا نجد اعتماداً على المثال الأول أن المقدار:

$$X^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{\sigma^2}$$

هو الآخر خاضع للتوزيع $\chi_{n-1}^2(x)$ ؛ أي ذو (n-1) درجة حرية، لأن المتحولات مرتبطة مع بعضها بعلاقة حساب $\bar{x}$.

5 . ونشير أخيراً إلى أنه عندما يكون n كبيراً $(n > 30)$ فإن $\sqrt{2\chi^2(x)}$ ينتهي إلى التوزيع الطبيعي العام الذي توقعه $E(X) = \sqrt{2n-1}$ وتباينه $\sigma^2 = 1$.

وتستخدم هذه الخاصية عندما يكون n كبير في حساب الاحتمالات للمتحولات الخاضعة للتوزيع $\chi_n^2(x)$.

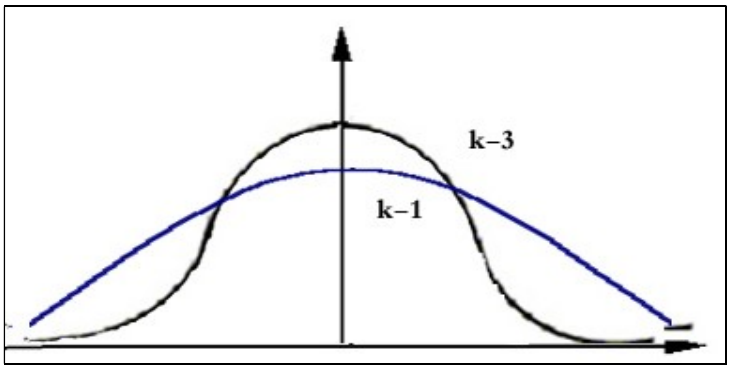
3-8: قانون توزيع (ستودينت) أو T:

يعرف هذا القانون بالعلاقة التالية:

$$T_k(\chi) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{k\pi}\,\Gamma\left(\frac{k}{2}\right)} \cdot \left(1 + \frac{x^2}{k}\right)^{-\left(\frac{k+1}{2}\right)} \qquad (56-3)$$

حيث أن $-\infty < \chi < +\infty$ وأن k عدد صحيح موجب يسمى بعدد درجات الحرية.

يرسم هذا القانون منحنيات مختلفة تبعاً لقيمة k وفيما يلي بعض منها.



الشكل رقم (3–15)

خواصه:

إن $T_k(x)$ يحقق الشرطين التاليين:

1 . إن $T_k(x) \geq 0$.

2 . إن المساحة تحت المنحني تساوي الواحد؛ أي أن:

$$\int_{-\infty}^{+\infty} T_k(X) \cdot dx = 1$$

3. إن قيمه المميزة تساوي:

$$E(X) = 0 \qquad (57-3)$$

$$\sigma^2(X) = \frac{k}{k-2} \quad (K > 2) \quad \text{حيث} \qquad (58-3)$$

إن أكثر العلاقات الاحتمالية تطبيقاً هي العلاقة:

$$P(X \geq t_0) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{k\pi}\,\Gamma\left(\frac{k}{2}\right)} \int_{t_0}^{\infty} \left(1 + \frac{\chi^2}{k}\right)^{-\left(\frac{k+1}{2}\right)} \cdot dx = P_0 \qquad (59-3)$$

وبواسطة هذه العلاقة يمكننا إيجاد قيمة t_0 المقابلة لاحتمال محدد P_0 ، وذلك ضمن عدد درجات الحرية المحددة n. وإن القيم الحرجة t_0 معروضة في (الجدول رقم V الملحق في آخر الكتاب.

ملاحظة:

عندما تكون $k=1$ فإننا نحصل على التوزيع التالي:

$$f(x)=\frac{1}{\pi}\cdot\frac{1}{1+x^2} \qquad -\infty<x<+\infty \quad \text{حيث إن:} \qquad (3-59)'$$

وهو التوزيع الذي يسمى بتوزيع كوشي، والذي له التوقع نفسه، ولكن تباينه غير معرف كما هو واضح من العلاقة (3–58).

وللتوزيع T تطبيقات عديدة في المجالات العلمية، ونذكر منها بشكل خاص مجال العينات الصغيرة.

فعندما نسحب عينة بحجم n من مجتمع إحصائي طبيعي نقوم بتقدير متوسط ذلك المجتمع $\bar{y}$ بواسطة متوسط العينة $\bar{x}$.

كما نقوم بتقدير تباينه σ^2 بواسطة تباين العينة المصحح والمعرف بالعلاقة:

$$S^2=\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}$$

أما خطأ التقدير في تقدير متوسط المجتمع $\bar{y}$ (في حالة السحب مع الإعادة) فيقدر بالكمية $\frac{S}{\sqrt{n}}$، وعندما نريد إجراء اختبار حول المتوسط $\bar{y}$ نحتاج إلى حساب المقدار t المعرف بالعلاقة التالية:

$$t=\frac{\bar{x}-\bar{y}}{S/\sqrt{n}}$$

وهنا نشير بدون برهان إلى أن المتحول الأخير t يخضع للتوزيع $T_k(x)$ ذي $k=n-1$ درجة حرية.

ونشير أخيراً إلى أن المتحول العشوائي t هو الذي أثار الباحث (غوسيه) على استنباط قانون التوزيع $T_k(x)$ والذي نشره تحت الاسم المستعار (ستودينت)، لأنه كان حينها طالباً في مرحلة الدراسة العليا، ولم يتجرأ على نشر اسمه الحقيقي.

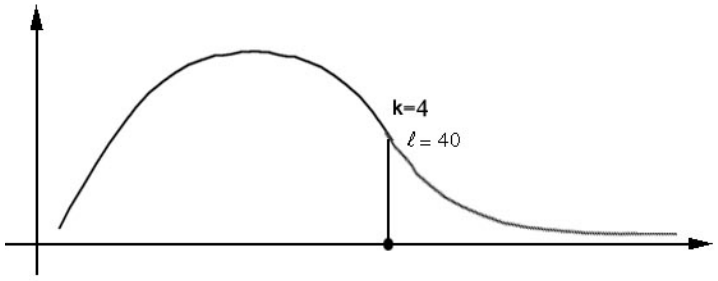
ولكن التوزيع $T_k(x)$ يصبح متقارباً من التوزيع الطبيعي المعياري $\varphi(x)$ عندما تصبح k أكبر من 30– ، ويتطابق معه عندما تنتهي k إلى اللانهاية.

3–9: قانون التوزيع F (توزيع فيشر . سينديكور):

يعرف هذا القانون بالعلاقة التالية:

$$F_{k\cdot\ell}(X)=\frac{k^{k/2}\cdot\ell^{\ell/2}\cdot\Gamma\left(\frac{k+\ell}{2}\right)}{\Gamma\left(\frac{k}{2}\right)\cdot\Gamma\left(\frac{\ell}{2}\right)}\cdot\frac{x^{\frac{k-2}{2}}}{(\ell+kx)^{\frac{k+\ell}{2}}} \qquad (60-3)$$

حيث أن $0\le x<+\infty$ وأن k,ℓ عددان صحيحان موجبان يسميان بعددي درجاتي الحرية. وهذا التابع يرسم أشكالاً متعددة من المنحنيات، والشكل التالي يبين إحدى هذه المنحنيات.



الشكل رقم (3–16)

خواصه:

إن $F_{k,l}(x)$ يحقق الشرطين التاليين:

1. $F_{k,l}(x)\ge 0$.

2. $\int_{-\infty}^{+\infty}F_{k,l}(X)=1$

أما قيمه المميزة فهي:

$$E(X)=\frac{\ell}{\ell-2} \qquad (\ell>2) \qquad (61-3)$$

$$\sigma^2(X)=\frac{2\ell^2(k+\ell-2)}{k(\ell-2)^2(\ell-4)} \qquad (\ell>4) \qquad (62-3)$$

وإن أشهر متحول عشوائي يخضع لتوزيع F هو نسبة تباييني عينتين مستقلتين ومسحوبتين من مجتمع واحد. ولتوضيح ذلك نفترض أننا سحبنا من مجتمع عينتين مستقلتين بحجمين n_1 و n_2، وكان متوسطاهما x_1 و x_2 وتباينها S_1^2 و S_2^2 على الترتيب. وبما أن كل من S_1^2 و S_2^2 يعتبر تقديراً لتباين المجتمع σ^2. وبما أنهما مختلفان، فنحن نتساءل هل يعتبر اختلافهما اختلافاً حقيقياً أم عرضياً؟

وعندما نأخذ النسبة:

$$F=\frac{S_1^2}{S_2^2} \qquad \left(S_1^2>S_2^2\right)$$

فإن المقدار F يمثل متحولاً عشوائياً خاضعاً للتوزيع $F_{k,\ell}$ ذي درجتي الحرية $k=n_1-1$ و $\ell=n_2-1$ المأخوذتين من درجات الحرية في حساب S_1^2 و S_2^2 على الترتيب. ولحساب الاحتمالات نستخدم العلاقة الأكثر تطبيقاً التالية:

$$F_k(X)=\frac{k^{\frac{k}{2}}\cdot\ell^{\frac{k}{2}}\cdot F\left(\frac{k+\ell}{2}\right)}{\Gamma\left(\frac{k}{2}\right)\cdot\Gamma\left(\frac{\ell}{2}\right)}\cdot\int_{F_0}^{\infty}\frac{x^{\frac{k-2}{2}}}{(\ell+kx)^{\frac{k+\ell}{2}}}\cdot dx=P_0 \qquad (63-3)$$

واعتماداً على هذه العلاقة يمكننا تحديد قيمة F_0 المقابلة للاحتمال P_0. وذلك من الجداول الملحقة في آخر الكتاب (الجدول رقم VI)، وحسب درجتي الحرية k و ℓ.

أمثلة تطبيقية:

مثال 13:

ليكن لدينا ثلاثة متحولات X و Y و Z خاضعة بانفراد للقانون χ_n^2 والذي فيه درجات الحرية تساوي:

$n_3 = 50$ ، $n_2 = 21$ ، $n_1 = 15$ على الترتيب.

والمطلوب:

1 . إيجاد القيم x_0 ، y_0 ، z_0 المقابلة للاحتمال $P_0 = 0{,}05$ لكل من المتحولات الثلاثة والموافقة للعلاقة (3−54).

2 . إيجاد وسيط كل من هذه المتحولات.

الحل:

1 . إن قيم x_0 ، y_0 ، z_0 المقابلة لـ $P_0 = 0{,}05$ والموافقة للعلاقة (3−54) لكل من المتحولات تؤخذ من جداول χ^2_n ، وهي القيم الموجودة في العمود الذي فيه $\alpha = P_0 = 0{,}05$ والمقابلة لدرجة الحرية المطلوبة.

فمن أجل $n_1 = 15$ نجد أن $x_0 = 24{,}996$ ،

وهذا يعني أن: $P(X > 24{,}996) = 0{,}05$

ومن أجل $n_2 = 21$ نجد أن $y_0 = 36{,}671$ ،

وهذا يعني أن: $P(Y > 36{,}671) = 0{,}05$

أما من أجل $n_3 = 50$ فإن قيمة χ_0 لا تظهر في جداول χ^2_n . وفي مثل هذه الحالات نشكل المتحول المعياري التالي:

$$T = \frac{\sqrt{2\chi} - \sqrt{2n-1}}{1} = \sqrt{2z} - \sqrt{99} \qquad (A)$$

والذي يخضع للتوزيع الطبيعي المعياري. ثم نبحث الآن في جداول احتمالات التوزيع الطبيعي المعياري $\Phi(Z)$ عن القيمة T_0 التي تحقق العلاقة:

$$P(> T_0) = P_0 = 0{,}05$$

وحتى نتمكن من استخدام جدول $\Phi(Z)$ نحول هذا الاحتمال على الشكل التالي:

$$P(T > T_0) = 1 - P(T \le T_0) = 0{,}05$$

$$P(T \le T_0) = 0{,}95$$

ومن جدول تابع التوزيع الطبيعي المعياري نجد أن قيمة T_0 التي تحقق هذه العلاقة تساوي $T_0 = 1{,}65$

وبوضع هذه النتيجة في العلاقة (A) السابقة نجد أن:

$$1{,}65 = \sqrt{2\chi_0} - \sqrt{99}$$

ومنها نجد أن:

$$\chi_0 = 67{,}278$$

وهذا يعني أنه عندما n= 50 فإن الاحتمال:

$$P(Z > 67{,}278) = 0{,}05$$

B – ولحساب وسيط كل من هذه المتحولات، نعلم أن الوسيط M_0 لمتحول مستمر يحقق العلاقة التالية:

$$P(X < M_e) = 0{,}05$$

فمن أجل $n_1 = 15$ للمتحول X نجد أن $M_e(X) = 14{,}339$.

ومن أجل $n_2 = 21$ للمتحول Y نجد أن $M_e(Y) = 20{,}337$.

أما من أجل $n_3 = 50$ للمتحول Z فنعتمد على العلاقة (A) السابقة ونبحث عن T_0 التي تحقق العلاقة:

$$P(T > T_0) = 0{,}05 = P(T \le T^0)$$

وهذا يعطينا أن T_0 لأن التوزيع الطبيعي المعياري متناظر بالنسبة للعمود الشاقولي.

ومن العلاقة (A) نجد أن:

$$0 = \sqrt{2M_e} - \sqrt{99} \Rightarrow M_e(Z) = 49{,}5$$

أي أنه في حالة $n_3 = 50$ فإن وسيط المتحول Z يساوي:

$$M_e(Z) = 49{,}5$$

وأخيراً نشير إلى أن توقعات هذه المتحولات تساوي:

$$E(X) = n_1 = 15 \ , \ E(Y) = n_2 = 21 \ , \ E(Z) = n_3 = 50$$

وهي تختلف عن قيم الوسيط لكل منهم.

مثال 14:

لـيكن لـدينا ثلاثـة متحـولات X، Y، Z خاضـعة لتوزيـع سـتودينت بـدرجات حريـة $n_1 = 16$, $n_2 = 21$, $n_3 = 100$ على الترتيب والمطلوب: إيجاد القيمة الحرجة t_0 لكـل مـن هـذه المتحـولات والمقابلـة للاحتمـال $P_0 = 0,05$ ، ثـم المقابلـة للاحتمـال $P_0 = 0,10$ ، والموافقتين لصيغة العلاقة (3−59) السابقة.

الحل:

إن القيمـة الحرجـة لـ t_0 تؤخذ من جدول ستودينت وهـي القيمـة الموجـودة فـي العمـود الذي فيه $P_0 = 0,05$ والمقابلة لعدد درجات الحرية المطلوبة.

فمن أجل X ودرجات حريته $n_1 = 16$ نجد أن $t_0(X) = 1,746$.

ومن أجل Y ودرجات حريته $n_2 = 21$ نجد أن $t_0(Y) = 1,721$.

ومن أجل Z ودرجات حريته $n_3 = 100$ فلا نجد قيمة لـ t_0 في جدول ستودينت، ولكن بما أن توزيع ستودينت ينتهي إلى التوزيع الطبيعي المعياري فإننا نبحث عن t_0 من جدول التوزيع الطبيعي المعياري فنجد أن:

$$P(Z > t_0) = 1 - P(Z \le t_0) = 1 - \Phi(t_0) = 0,05$$
$$\Phi(t_0) = 0,05$$

ومن جدول قيم التوزيع المعياري نجد أن:

$$t_0 = 1,65$$

أما عندما تكون $P_0 = 0,01$ فإننا نجد مباشرةً أنه:

من أجل X ودرجات حريته $n_1 = 16$ نجد أن $t_0 = 1,34$.

ومن أجل Y ودرجات حريته $n_2 = 21$ نجد أن $t_0 = 1,32$.

ومن أجل Z ودرجات حريته $n_3 = 100$ فنبحث عن t_0 في جدول التوزيع الطبيعي المعياري وهي التي تحقق العلاقة:

$$\Phi(t_0) = 0,90 \Rightarrow t_0 = 1,29$$

3-10: قانون التوزيع الاحتمالي الإحصائي:

عندما لا نستطيع معرفة قانون التوزيع الاحتمالي الذي يخضع له المتحول العشوائي X فإننا نقوم بجمع معلومات إحصائية عن سلوكه، ثم نقوم بترتيبها إذا كان منقطعاً، أو بتبويبها إذا كان مستمراً.

فإذا كان المتحول العشوائي X منقطعاً وكانت قيمه الممكنة والمرتبة هي:

$$X : X_1 \ , \ X_2 \ , \ X_3 \quad X_i \ , \ X_m$$

وكانت التكرارات المطلقة المقابلة لهذه القيم هي:

$$n_1 \ , \ n_2 \ , \ n_3 \ , \ n_i \ , \ n_m$$

فإن قانون التوزيع الاحتمالي الإحصائي أو التجريبي هو عبارة عن التكرارات النسبية المقابلة للقيم الممكنة؛ أي أن قانون التوزيع الاحتمالي الإحصائي هو سلسلة التكرارات النسبية المعروضة في الجدول التالي:

X قيم	$X_1 \ , \ X_2 \ , \ X_3 \ \ldots\ldots \ X_i \ \ldots\ldots \ X_m$
$P_i :$	$\frac{n_1}{n} \ , \ \frac{n_2}{n} \ , \ \frac{n_3}{n} \ \ldots\ldots \ \frac{n_i}{n} \ \ldots\ldots \ \frac{n_m}{n}$

ويتم حساب الاحتمالات في هذه الحالة من خلال العلاقتين التاليتين:

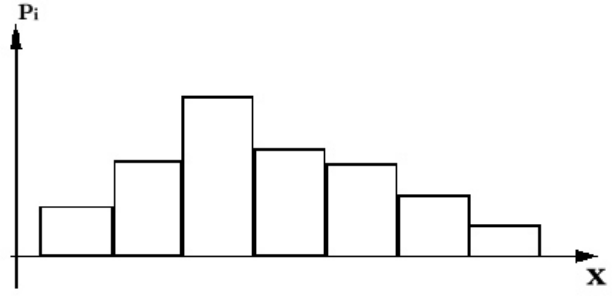
$$P(X = x_k) = \frac{n_k}{n} \qquad (64-3)$$

$$P(x_k \le X \le x_{k+s}) = \sum_{i=k}^{k+s} \frac{n_i}{n} \qquad (65-3)$$

إن مضمون التوزيع الإحصائي ينسجم تماماً مع التعريف الإحصائي للاحتمال، إذ أن كل قيمة من قيم X تقابل احتمالاً يساوي عدد مرات تحققه على عدد التجارب n. أما إذا كان المتحول العشوائي X مستمراً فإننا نقوم بتبويب المعلومات الإحصائية المتجمعة عنه على شكل مجالات مرتبة ومتجاورة وغير متقاطعة، فنحصل على الجدول التالي:

مجالات X	$[x_0 - x_1[$	$[x_1 - x_2[$	$[x_2 - x_3[$	$[x_3 - x_4[$	...	$[x_{m-1} - x_m[$	Σ
التكرارات المطلقة	n_1	n_2	n_3	n_4	...	n_m	n
الاحتمالات P_i	$\frac{n_1}{n}$	$\frac{n_2}{n}$	$\frac{n_3}{n}$	$\frac{n_4}{n}$	...	$\frac{n_m}{n}$	1

وبذلك يكون التوزيع الإحصائي لـ X المستمر هو سلسلة التكرارات النسبية المقابلة للمجالات المختلفة السابقة. ويمكننا أن نوضح ذلك بيانياً كما هو في الشكل التالي:



الشكل رقم (3-17)

وإن حساب الاحتمالات في هذه الحالة يتم من خلال العلاقات التالية:

$$P(x_{k-1} \le X \le x_k) = \frac{n_k}{n} \qquad (66-3)$$

$$P(x_{k-1} \le X \le x_{k+s}) = \sum_{i=k}^{k+s} \frac{n_k}{n} \qquad (67-3)$$

وذلك بشرط أن يكون x_{k+s} يمثل حداً من حدود أحد المجالات السابقة. وإذا لم يكن كذلك فيجب إجراء بعض التصحيح بشكل تناسبي.

ويحسب التوقع الرياضي لـ X من خلال العلاقة التالية:

$$E(X) = \bar{x} = \begin{cases} \sum_{i=1}^{m} p_i x_i = \sum \frac{n_i}{n} x_i & \text{للمتحول المنقطع} \\ \sum_{i=1}^{m} p_i x'_i = \sum \frac{n_i}{n} x'_i & \text{للمتحول المستمر} \end{cases} \qquad (68-3)$$

إذ أن x_i' هو مركز المجال i في التبويب السابق.

كما أن تباين المتحول X يحسب من العلاقة التالية:

$$\sigma^2(X)=\sigma^2=\begin{cases}\sum_{i=1}^{m} p_i(x_i-\overline{x})^2=\sum_{i=1}^{m}\frac{n_i}{n}(x_i-\overline{x})^2\\ \sum p_i(x_i'-\overline{x})^2=\sum_{i=1}^{m}\frac{n_i}{n}(x_i'-\overline{x})^2\end{cases}\qquad(69-3)$$

مثال 14:

إذا كانت لدينا المعلومات المبوبة التالية عن علامات الطلاب في مقرر الإحصاء:

مجالات العلامات	0-20	20-30	30-40	40-50	50-60	60-70	70-80	80-100	$\sum$
التكرارات	5	10	15	20	25	15	10	5	105

فإن التوزيع الاحتمالي الإحصائي لهذه العلاقات هو التوزيع التالي:

مجالات العلامات	0-20	20-30	30-40	40-50	50-60	60-70	70-80	80-100	$\sum$
التكرارات النسبية	$\frac{5}{105}$	$\frac{10}{105}$	$\frac{15}{105}$	$\frac{20}{105}$	$\frac{25}{105}$	$\frac{15}{105}$	$\frac{10}{105}$	$\frac{5}{105}$	1

وأن احتمال أن تكون علامة الطالب بين 50 و80 يساوي:

$$p(50<X<80)=\frac{25}{105}+\frac{15}{105}+\frac{10}{105}=\frac{50}{105}$$

وأن احتمال أن تكون علامة الطالب بين 30 و73 يساوي:

$$p(30<X<73)=\frac{15}{105}+\frac{20}{105}+\frac{25}{105}+\frac{15}{105}+\frac{10}{105}\cdot\frac{3}{10}=\frac{78}{105}$$

ولقد ضربنا الاحتمال الأخير بالنسبة $\frac{3}{10}$ لنعبر عن الاحتمال المقابل لجزء المجال المقابل الداخل في الاحتمال المطلوب.

وأن القيمة المتوقعة للعلامة تساوي:

$$E(X)=\sum p_i x_i=50{,}24$$

أما التباين فنترك أمر حسابه للقارئ على سبيل التمرين.

تمرينات

1 . أطلقت أربع طلقات على هدف ما. فإذا كان احتمال الإصابة لكلٍ منها ثابتاً ويساوي $p = \frac{1}{4}$ ، فما هو:

. احتمال عدم إصابة الهدف؟

. احتمال إصابة الهدف بطلقة واحدة؟

. احتمال إصابة الهدف بطلقتين فقط؟

. احتمال إصابة الهدف بثلاث طلقات؟

. احتمال إصابة الهدف بالطلقات الأربع؟

. احتمال إصابة الهدف بطلقتين على الأقل؟

. احتمال إصابة الهدف بثلاث طلقات على الأكثر؟

. ما هو متوسط عدد الإصابات وتباينه؟

2 . تبيع إحدى الوكالات سياراتها على أساس عدم تعطلها خلال عامين من وضعها في الاستخدام يساوي $(0,08)$.

فإذا اشترينا منها 5 سيارات، فما هو:

. احتمال عدم تعطل أي منها خلال عامين؟

. احتمال تعطل واحدة فقط خلال عامين؟

. احتمال تعطل اثنتين خلال عامين؟

. احتمال تعطل اثنتين على الأقل خلال عامين؟

. احتمال تعطل اثنتين على الأكثر خلال عامين؟

. متوسط عدد السيارات التي يمكن أن تتعطل منها، وتباينه؟

3 . إذا كان متوسط وزن الطفل عند الولادة X يخضع للتوزيع الطبيعي، وكان متوسطه $\bar{x} = 3$، وتباينه $\sigma^2 = 0,64$. فما هي الصيغة الرياضية المحددة لقانون التوزيع X ؟ وما هي الاحتمالات التالية:

. احتمال أن يكون الوزن في المجال [4 , 2].

. احتمال أن يكون أكبر من 3 .

. احتمال أن يكون أقل من 3 .

4 . إذا كان الإنتاج اليومي لإحدى المؤسسات X خاضع للتوزيع الطبيعي، وكان متوسطه $\bar{x} = 3000$، وتباينه $\sigma^2 = 14900$. وكان الإنتاج المخطط خلال شهر يساوي 100000 قطعة. فما هو احتمال تنفيذ هذه الخطة؟ وما هو احتمال تجاوزها بمقدار 2% منها؟

الفصل الرابـع

العينـات وقضـايا التقديـر

ويشمل الفقرات التالية:

1 . تمهيد.
2 . أنواع العينات.
3 . التقدير النقطي لمؤشرات المجتمع.
4 . مجالات الثقة (للمتوسط والإجمالي والنسبة والفروقات).
5 . تقدير حجم العينة (للمتوسط والإجمالي والنسبة).
6 . تمرينات.

الفصل الرابع

العينـات وقضـايا التقديـر

4-1- تمهيد:

يمكن أن تتم عملية جمع المعلومات الإحصائية عن المجتمع الإحصائي المدروس بطريقتين أساسيتين هما:

. طريقة الإحصاء الشامل:وتشمل دراسة جميع عناصر المجتمع الإحصائي المدروس.

. طريقة العينات: وتتناول بالدراسة عناصر جزء معين ومسحوب عشوائياً من المجتمع الإحصائي المدروس.

* **تعريف المجتمع الإحصائي:** هو أي مجموعة من العناصر تشترك بخاصة معينة أو أكثر وتكون خاضعة للدراسة الإحصائية. وبذلك نجد أن المجتمع الإحصائي يمكن أن يكون مجموعة من البشر، أو الشجر، أو الحجر؛ من الأشياء، أو الحيوانات؛ من المنتجات، أو المبيعات...إلخ. ونرمز لحجم المجتمع الإحصائي بالرمز N. وهكذا نجد أن مفهوم المجتمع الإحصائي يختلف عن المجتمع الإنساني وله معنى أوسع من المجتمع الإنساني.

* **تعريف العينة:** إن العينة هي جزء من واحدات المجتمع الإحصائي المدروس، نقوم بسحبها عشوائياً ودراسة خصائص وحداتها، لتعميم مميزات تلك الخصائص على المجتمع الإحصائي الكلي، وسنرمز لحجم العينة بالرمز n.

وتعتبر طريقة العينات من الطرائق الهامة في عمليات جمع المعلومات الإحصائية وتحليلها، وذلك لأنها توفر كثيراً من المال والجهد والوقت في الحصول على المعلومات اللازمة.

وتكون طريقة العينات في كثير من الأحيان هي الأسلوب الإحصائي الوحيد، الذي يمكن تطبيقه في الحصول على المعلومات الإحصائية؛ وذلك كما في حالات تحليل

الدم، أو تحليل التربة، أو تحليل المياه، أو تقدير كميات السمك في البحر، أو تقدير عمر المصابيح الكهربائية، أو تقدير مدة صلاحية الأغذية والأدوية...إلخ.
أما خطوات البحث المتبعة في طريقة العينات فهي نفسها المتبعة في حالة الإحصاء الشامل. غير أن طريقة العينات تحتاج إلى خطوة إضافية هي خطوة سحب العينة.

*** سحب العينة:**

نشترط في سحب العينة أن يكون السحب سحباً عشوائياً، وحتى يكون ذلك السحب عشوائياً، يجب أن نقوم بما يلي:

1 . ترقيم عناصر المجتمع الإحصائي المدروس بواسطة أرقام طبيعية من 1 حتى N أو بواسطة رموز أخرى يتفق عليها.

2 . تشكيل الإطار اللازم: والإطار هو عبارة عن نسخة ثانية عن أرقام أو رموز المجتمع السابق، يكون لها الحجم أو اللون نفسه، وتوضع في كيس أو صندوق وتخلط جيداً.

3 . سحب عدد من أرقام الإطار، ويتم خلط الأرقام في الإطار بعد كل عملية سحب.
أما السحب فيمكن أن يكون على شكلين هما:

ـ السحب مع الإعادة: وفيه يعاد الرقم المسحوب إلى الإطار ليتعرض للسحب مرةً أخرى.

ـ السحب بدون الإعادة: وفيه لا يعاد الرقم المسحوب إلى الإطار، وبذلك لن يتعرض للسحب مرةً أخرى.

ويمكن أن نستعيض عن الإطار في عملية السحب باستخدام جداول الأرقام العشوائية، أو عجلات الرهان، أو بواسطة علاقات رياضية مبرمجة على الآلات الحاسبة والحواسيب المختلفة.

4 . نعتبر العناصر ذات الأرقام المسحوبة في الخطوة السابقة هي عناصر العينة، ونبحث عنها في المجتمع الإحصائي المدروس لاستقصاء المعلومات الإحصائية اللازمة عن كل منها.

إن عملية السحب العشوائي لعناصر العينة تضمن لنا عدم التحيز في الانتقاء، كما تؤمن تمثيلاً جيداً لجميع عناصر المجتمع في العينة المسحوبة. أما حجم العينة n اللازم سحبه في دراسة معينة فيتوقف على مقدار تجانس عناصر المجتمع من حيث الخاصة المدروسة، وعلى مقدار الدقة المطلوبة، وعلى حجم المجتمع نفسه، وعلى مقادير أخرى سنتعرض لها فيما بعد.

4-2- أنواع العينات:

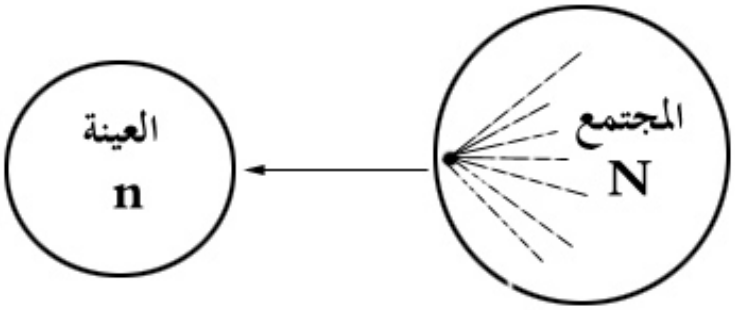
يمكن أن يتم تصميم سحب العينة بعدة طرائق، ونحصل منه على عينة أو عينات ذات صفات خاصة، وحسب ذلك التصميم نحصل على أنواع مختلفة للعينات المسحوبة، وأهم هذه الأنواع هي:

1 . العينة العشوائية البسيطة:

وهي العينة التي تسحب عشوائياً من المجتمع الإحصائي مباشرةً، وبحيث يكون احتمال سحب أي عنصر فيه يساوي احتمال سحب عنصر آخر كما يجب أن يكون احتمال سحب أي عينة بحجم n يساوي احتمال سحب أي عينة أخرى بالحجم n نفسه.

فإذا كان السحب مع الإعادة فإن احتمال سحب أي عنصر يساوي $\frac{1}{N}$ لجميع العناصر. أما إذا كان السحب بدون الإعادة فإن احتمال سحب عنصر ما يساوي $\frac{1}{N}$ في السحبة الأولى، ويساوي $\frac{1}{N-1}$ في السحبة الثانية. ويساوي $\frac{1}{N-2}$ في السحبة الثالثة .....إلخ.

ويمكن تمثيل عملية سحب العينة العشوائية البسيطة من المجتمع الإحصائي بالشكل الآتي.



الشكل رقم (4-1)

2 . العينة المنتظمة:

وهي عينة تسحب من المجتمعات الإحصائية المتحركة: كرواد متحف أو مسرح أو مكتبة أو مطعم .....إلخ.

ويتم سحب العينة المنتظمة باختيار عنصر البداية بشكل عشوائي وليكن x_0، ثم نأخذ العناصر التي يأتي ترتيبها بعد كل k عنصراً، فيأخذ ترتيب العناصر الأرقام الآتية:

x_0 ثم $x_0 + k$ ثم $x_0 + 2k$ ثم $x_0 + 3k$ ...إلخ.

ويتم تحديد المدى k بحيث يكون $n \cdot k = N$، أو من العلاقة $\frac{N}{n} = k$، ويشترط أن يكون k عدداً صحيحاً لذلك يتم تقريبه إلى أقرب عدد صحيح.

3 . العينة العشوائية الطبقية:

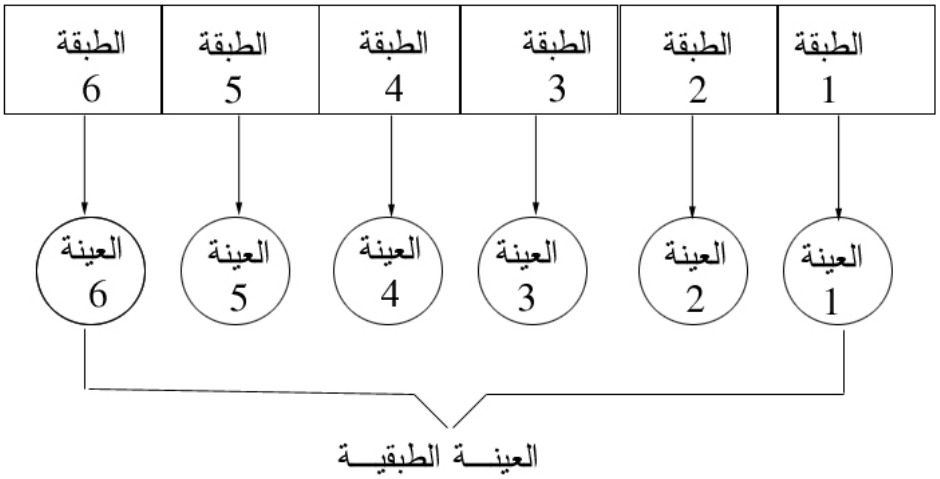
عندما يكون المجتمع الإحصائي غير متجانس من حيث الخاصة المدروسة فإننا نقسمه إلى طبقات متجانسة، ونسحب من كل طبقة عينةً عشوائيةً بسيطة حسب حجم تلك الطبقة، وبذلك نحصل على عدة عينات بسيطة مأخوذة من الطبقات المختلفة.

وعندما نعالج هذه العينات نتعامل معها كعينات مستقلة، ويمكن النظر إليها كعينة واحدة نسميها بالعينة الطبقية.

العينة الطبقية: وهي العينة المؤلفة من مختلف عينات الطبقات الجزئية التي يتألف منها المجتمع الكلي.

وكمثال على العينة الطبقية: نفترض أننا قسمنا طلاب الجامعة حسب كلياتهم (طلاب كل كلية هم مجتمع متجانس من طلاب الجامعة) وسحبنا من طلاب كل كلية عينة

عشوائية بسيطة فإننا سنحصل على عدة عينات بسيطة، وأن جملة هذه العينات البسيطة والمسحوبة من كل كلية تشكل عينة طبقية. ويمكن تمثيل عملية سحب العينة الطبقية كما في الشكل الآتي:



الشكل رقم (4-2)

4 . العينة العنقودية البسيطة:

وهي عينات تسحب من المجتمعات التي يمكن تجزئتها (أو تكون مجزأة بطبيعتها) إلى مجتمعات جزئية كثيرة أو إلى فئات متعددة. وإن عملية سحب هذه العينة هي عملية مركبة من عمليتي سحب هما:

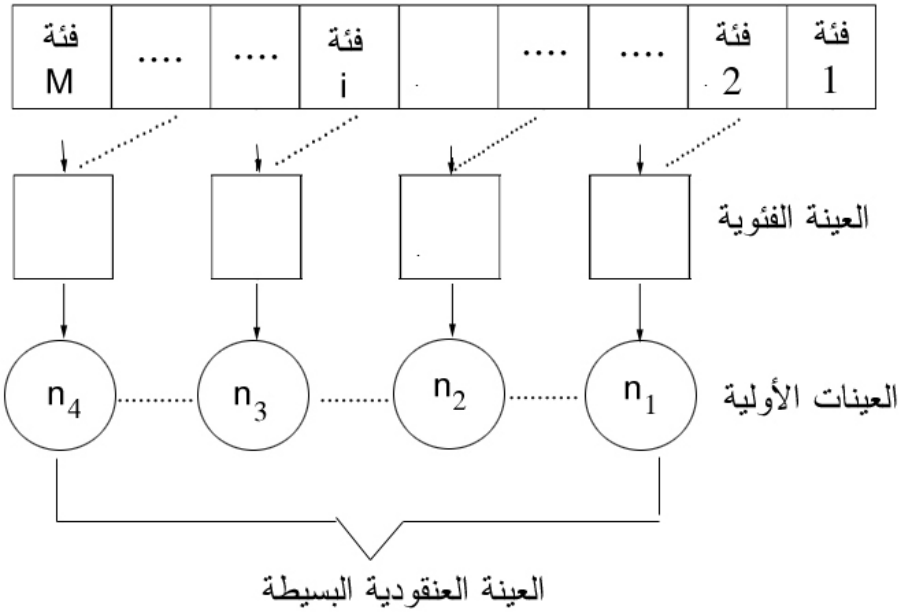
. سحب عينة من الفئات التي يتألف منها المجتمع الكلي بطريقة عشوائية بسيطة تسمى العينة الفئوية.

. سحب عينة من العناصر من كل فئة تم سحبها في العملية الأولى فنحصل على عدد من العينات نسميها العينات الأولية.

وإن جملة هذه العينات الأولية تشكل عينة كلية تسمى العينة العنقودية البسيطة.

ويمكن تمثيل المعاينة العنقودية البسيطة كما هو مبين في الشكل الآتي.

المجتمع المؤلف من M فئة



الشكل رقم (4-3)

وهنا نشير إلى أنه توجد أنواع أخرى من العينات العنقودية مثل:

. العينة العنقودية الطبقية.

. العينة العنقودية متعددة المراحل.

. العينة العنقودية المتناسبة مع الحجم.

وتماشياً مع منهاج هذا المقرر، فإننا سنقصر دراستنا في هذا الكتاب على العينات العشوائية البسيطة فقط.

وسنبحث في الفقرة التالية مسألة إيجاد تقديرات لعدد من مؤشرات المجتمع المتعلقة بالخاصة المدروسة فيه.

فإذا رمزنا لقيم الخاصة المدروسة في المجتمع المدروس بالرموز التالية:

y_1 , y_2 , y_3 ,....., y_i ....., y_N (وهي قيم مجهولة حتى الآن)

وبما أن هذه القيم مجهولة، فإن جميع معالم المجتمع (كالمتوسط والتباين والانحراف المعياري) تكون مجهولة أيضاً. لذلك فإننا سنبحث عن تقديرات لها من خلال سحب عينة عشوائية من المجتمع المدروس، ثم دراسة خصائصها وحساب مؤشراتها المختلفة،

ثم الاستفادة من هذه المعلومات في تقدير معالم المجتمع المدروس. وسنقوم هنا بإيجاد تقديرات لمعالم المجتمع المجهولة التالية:

1 . تقدير متوسط الخاصة المدروسة في المجتمع والذي يساوي:

$$\bar{y} = \frac{\sum_{i=1}^{N} y_i}{N}$$

2 . تقدير إجمالي الخاصة المدروسة في المجتمع والذي يساوي:

$$Y = \sum_{i=1}^{N} y_i$$

3 . تقدير تباين الخاصة في المجتمع والذي يساوي:

$$\sigma^2 = \frac{1}{N} \cdot \sum_{i=1}^{N} (y_i - \bar{y})^2$$

4 . تقدير نسبة لخاصة في المجتمع والتي تساوي:

$$R = \frac{M}{N}$$

5 . تقدير تباين تقدير المتوسط.

6 . تقدير تباين تقدير الإجمالي.

7 . تقدير تباين نسبة خاصة ما في المجتمع.

8 . تقدير حجم العينة اللازم لتحقيق دقة معينة وباحتمال محدد.

وقبل أن نعالج كل هذه التقديرات نشير إلى أن جميع تلك التقديرات تحسب من معطيات العينة المسحوبة.

ولكن بما أن العينة نسحب بشكل عشوائي فإننا لانستطيع تحديد أية عينة ستكون بين أيدينا، وذلك لأن العينة المسحوبة ماهي إلا واحدة من العينات التي يمكن تشكيلها بحجم واحد n من المجتمع المؤلف من N عنصراً. وبفرض أن عدد تلك العينات الممكنة يساوي M عينة فإن M يساوي:

$$M = \begin{cases} C_{N+n-1}^{n} & \text{في حالة السحب مع الإعادة} \\ C_{N}^{n} & \text{في حالة السحب بدون إعادة} \end{cases} \qquad (1-4)$$

إذ أن:

$$C_N^n = \frac{N!}{n!(N-n)!}$$

وبما أن كل عينة تعطينا تقديراً معيناً للمؤشر المدروس، فإنه يمكننا القول أنه يوجد لدينا M تقديراً لكل مؤشر من مؤشرات المجتمع المدروس، مع إننا لن نستخدم إلا تقديراً واحداً من هذه التقديرات وهو التقدير المأخوذ من العينة المحسوبة، ولكننا لا نستطيع أن نتجاهل التقديرات الأخرى المأخوذة من العينات الممكنة الأخرى.

وإذا افترضنا أن حجم العينة هو n عنصراً وأن القياسات العددية المتعلقة بالخاصة المدروسة كانت كما يلي:

$$x_1 \ , \ x_2 \ , \ x_3 , \quad , x_i \quad , x_N$$

وعلينا أن نستخدم هذه القياسات بالشكل المناسب للحصول على تقديرات مؤشرات المجتمع المطلوبة.

فإذا كان هدفنا تقدير متوسط الخاصة في المجتمع $\overline{y}$، فإننا نأخذ المتوسط الحسابي لتلك القيم في العينة ونعتبره تقديراً لذلك المتوسط $\overline{y}$. ولكن يمكننا أن نقدر المتوسط بطرائق أخرى كالوسيط، أو المنوال، أو بمتوسط القيمتين الصغرى والعظمى، أو بمتوسط عدد معين من القيم السابقة ....الخ.

وهكذا يتضح لنا أن تقدير أي مؤشر إحصائي في المجتمع يمكن أن يأخذ قيماً متعددة، وذلك بسبب تعدد العينات الممكنة؛ كما يمكن أن يتم بصيغ متعددة، وذلك بسبب تعدد الصيغ الرياضية والحسابية.

ولهذا فإنه يمكننا اعتبار أي تقدير من التقديرات السابقة متحولاً عشوائياً وخاضعاً لتوزيع احتمالي معين، يختلف من تقدير لآخر.

وإن جملة التقديرات الممكنة لمؤشر معين تخضع لما يسمى معايير جودة التقدير وهي:

***عدم التحيز:** نقول عن التقدير $\widetilde{\theta}$ أنه تقدير غير متحيز للمؤشر θ المجهول إذا كان يحقق العلاقة التالية:

$$E(\widetilde{\theta}) = \theta \qquad (2-4)$$

إذ أن التوقع $E(\widetilde{\theta})$ مأخوذ على جميع العينات الممكنة، وهذا يعني أن يكون التوقع الرياضي لجميع التقديرات الممكنة من النوع $\widetilde{\theta}$ والمأخوذة من جميع العينات الممكنة مساوياً للقيمة المجهولة θ.

* **التماسك (الاتساق):** نقول عن التقدير $\widetilde{\theta}$ أنه تقدير متماسك للمؤشر θ، إذا كانت قيم $\widetilde{\theta}$ تنتهي احتمالياً إلى θ عندما تنتهي n إلى $N(\infty)$، أي إذا كان:

$$P\left(\widetilde{\theta} \underset{n\to N}{\to} \theta\right) = 1 \qquad (3-4)$$

ويتعلق هذا المعيار بالصيغة الرياضية التي ستحسب منها $\widetilde{\theta}$، وهذا يعني أن الصيغة الرياضية المستخدمة في حساب التقدير $\widetilde{\theta}$ تنتهي إلى الصيغة الرياضية نفسها المستخدمة في حساب القيمة المجهولة θ عندما تنتهي n إلى N أو إلى ∞.

* **الفعالية:** نقول عن التقدير $\widetilde{\theta}$ أنه تقدير فعال للمؤشر θ إذا كان التباين الناتج عنه أصغر من جميع التباينات الناتجة عن التقديرات الأخرى، أي إذا كان:

$$\sigma^2_{\widetilde{\theta}} = Min\left[\sigma^2_{\widetilde{\theta}_k}\right] \qquad (4-4)$$

ويتعلق هذا المعيار أيضاً بالصيغة الرياضية التي سنحسب منها $\widetilde{\theta}$ وبطريقة الحصول عليها.

* **الكفايــة:** نقول عن التقدير $\widetilde{\theta}$ أنه تقدير كافٍ للمؤشر θ، إذا استخدم جميع المعلومات التي توفرها العينة عن θ . فتقدير متوسط المجتمع $\overline{y}$ من خلال متوسط القيمتين الصغرى والعظمى فقط، لا يعتبر تقديراً كافياً لأنه لا يستخدم القياسات الأخرى التي توفرها العينة، بل يقتصر على قيمتين منها فقط، هما القيمة الصغرى والقيمة العظمى. أما تقدير $\overline{y}$ من خلال متوسط العينة $\overline{x}$ فهو تقدير كافٍ لأنه يستخدم جميع قياسات العينة.

4-3: التقدير النقطي لمؤشرات المجتمع:

ويتناول إيجاد تقديرات عددية لمؤشرات المجتمع المطلوبة وهي:

4-3-1: تقدير متوسط الخاصة في المجتمع:

لقد رمزنا لمتوسط الخاصة في المجتمع (أو باختصار متوسط المجتمع) بالرمز $\bar{y}$، وعرفناه بالعلاقة:

$$\bar{y} = \frac{\sum_{i}^{N} y_i}{N} \qquad (5-4)$$

إذ أنّ y_i هي قيم الخاصة في المجتمع، وهي مقادير مجهولة كما أن متوسطها $\bar{y}$ مجهول. ولإيجاد تقدير المتوسط $\bar{y}$ نستخدم الوسط الحسابي لقياسات العينة المسحوبة والمعرف بالعلاقة:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad (6-4)$$

حيث إنّ x_i هي قيم الخاصة نفسها في العينة ذات الحجم n، ونكتب ذلك التقدير على الشكل التالي:

$$\tilde{\bar{y}} = \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad (7-4)$$

وسنبرهن على أن هذا التقدير $\bar{x}$ هو تقدير غير متحيز لـ $\bar{y}$ وذلك كما يلي:

نلاحظ أنه حسب تعريف عدم التحيز فإن $\bar{x}$ يجب أن يحقق العلاقة التالية:

$$E(\bar{x}) = \bar{y}$$

وبما أن $\bar{x}$ هو متوسط العينة المسحوبة، والذي يمكن أن يأخذ قيماً مختلفة حسب العينات الممكنة والتي عددها M عينة، لذلك سنرمز له بـ $\bar{x}_k$. وبما أن احتمال سحب كل عينة من العينات المختلفة يساوي $\frac{1}{M}$ فإننا نجد أن:

$$E(\bar{x}) = \sum_{k=1}^{M} \frac{1}{M} \cdot \bar{x}_k = \frac{1}{M} \cdot \sum_{k=1}^{M} \frac{(x_1 + x_2 + x_3 + ... + x_n)_k}{n}$$

$$E(\bar{x}) = \frac{1}{n \cdot M} \sum_{k=1}^{M} (x_1 + x_2 + x_3 + ... + x_n)_k$$

وبملاحظة أن المجموع الأخير ما هو إلا عبارة عن مجموع القيم في جميع العينات الممكنة. وهو يشمل جميع عناصر المجتمع مع تكرار كل منها بعدد متساو من المرات وذلك بسبب وجوده في عينات مختلفة. وبملاحظة أن عدد القيم الداخلة في المجموع :

$$\sum_{k=1}^{M} (x_1 + x_2 + x_3 + ... + x_n)_k$$

يساوي n قيمة في كل عينة، ويساوي $n \cdot M$ قيمة في جميع العينات. لذلك فإن عدد مرات تكرار كل عنصر من المجتمع في ذلك المجموع يساوي $\frac{n \cdot M}{N}$ مرة. وبذلك نجد أن المجموع السابق يساوي مجموع القيم في المجتمع مضروباً بمقدار تكراراتها؛ أي أنه يساوي:

$$\sum_{k=1}^{M} (x_1 + x_2 + x_3 + ... + x_n)_k = \frac{n \cdot M}{N} \cdot (y_1 + y_2 + y_3 + ... + y_N)$$

بالتعويض في العبارة السابقة نجد أن:

$$E(\bar{x}) = \frac{1}{n \cdot M} \cdot \frac{n \cdot M}{N} \cdot \sum_{i=1}^{N} (y_i)$$

$$E(\bar{x}) = \frac{1}{N} \cdot \sum_{i=1}^{M} y_i = \bar{y}$$

وهذا يعني أن التقدير $\bar{x}$ هو تقدير غير متحيز لـ $\bar{y}$.

وكذلك يمكن البرهان بسهولة على أن هذا التقدير متماسك لأن الصيغة الرياضية لحساب $\bar{x}$ تنتهي إلى الصيغة الرياضية لحساب $\bar{y}$، لأنه يكون لدينا:

$$\bar{x} = \frac{\sum_{}^{M} x_i}{n} \xrightarrow[n \to N]{} \frac{\sum_{}^{M} y_i}{N} = \bar{y}$$

وكذلك يمكننا البرهان على أن $\bar{x}$ هو تقدير فعال للمتوسط $\bar{y}$ وذلك اعتماداً على الخصائص التي عالجناها في مبادئ الإحصاء وبرهنّا على أن تباين القيم عن وسطها الحسابي هو اصغر ما يمكن.

إذن نستنتج أن تباين قيم المتوسطات $\bar{x}$ عن وسطها (أو توقعها) $\bar{y}$ هو أصغر ما يمكن، وهذا يعني أن التقدير $\bar{x}$ هو تقدير فعال. وبما أنه يستخدم جميع معلومات العينة فهو تقدير كاف.

ومما تقدم يمكننا الاعتماد على متوسط العينة $\bar{x}$ المعرف بالعلاقة (4–6) في تقدير متوسط المجتمع $\bar{y}$ بدون تحفظ.

مثال 1:

سحبنا عينة عشوائية بحجم $n = 8$ أشخاص من مجتمع ما وسألناهم عن عدد أفراد أسرهم فكانت الإجابات كما يلي:

$\bar{x}_i$: , 4 , 5 , 6 , 3 , 10 , 12 , 7 , 9

أوجد تقديراً لمتوسط عدد أفراد الأسرة في ذلك المجتمع.

الحل:

إن متوسط عدد أفراد الأسرة في المجتمع المذكور يبقى مجهولاً، ولكن العينة المسحوبة تساعدنا على إيجاد تقدير له.

وبواسطة العلاقة (4–6) نجد أن:

$$\widetilde{\bar{y}} = \bar{x} = \frac{\sum_{n}^{n} x_i}{n} = \frac{4+5+6+3+10+12+7+9}{8} = 7$$

أي أن متوسط عدد أفراد الأسرة في المجتمع يقدر بـ 7 أفراد.

مثال 2:

من بين 10 طلاب سحبنا عشوائياً عينة بحجم $n = 4$، وسألناهم عن علاماتهم في مقرر الرياضيات، فكانت إجاباتهم كما يلي:

$$\bar{x}_i : \ , 55 \ , \ 60 \ , \ 75 \ , \ 40$$

فما هو تقدير متوسط علامات هؤلاء الطلاب العشرة.

الحل:

إن متوسط علامة الطالب يقدر بواسطة المتوسط الحسابي لقياسات العينة، فنجد أنه:

$$\widetilde{\bar{y}} = \bar{x} = \frac{55+60+75+40}{4} = 57{,}5$$

أي أن متوسط علامة الطالب في المجتمع المذكور يساوي 57,5.

4-3-2: تقدير إجمالي المجتمع:

إن إجمالي المجتمع Y يعرف بالعلاقة:

$$Y = \sum y_i = N \cdot \bar{y}$$

ولتقديره نستفيد مما توصلنا إليه في تقدير المتوسط $\bar{y}$، فنجد أنه عندما يكون العدد N معلوماً، فإن تقدير الإجمالي يعطى بالعلاقة:

$$\widetilde{Y} = N\widetilde{\bar{y}} = N \cdot \bar{x} \qquad (7-4)$$

ويمكننا البرهان بسهولة على أن هذا التقدير هو تقدير غير متحيز ومتماسك وفعال وكافٍ للإجمالي المجهول y.

4-3-3: تقدير تباين المجتمع:

يعرف تباين المجتمع بالعلاقة التالية:

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})^2$$

وسنحاول تقدير هذا التباين بواسطة تباين العينة D^2 المعرف بالعلاقة:

$$D^2 = \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2$$

وحتى يكون التباين D^2 تقديراً غير متحيز للتباين σ^2، فإنه يجب أن يكون:

$$E(D^2) = \sigma^2$$

وللبرهان على ذلك، أو على عدم ذلك نقوم بحساب التوقع الرياضي لـ D^2 فنجد أن:

$$E(D^2) = E\left[\frac{1}{n} \sum (x_i - \bar{x})^2 \right]$$

$$= \frac{1}{n} \cdot E \sum_{i=1}^{n} (x_i - \bar{x})^2$$

نضيف ونطرح المتوسط $\bar{y}$ إلى ما داخل القوس فنجد أن:

$$E(D^2) = \frac{1}{n} \cdot E \sum_{i=1}^{n} \left[(x_i - \bar{y}) - (\bar{x} - \bar{y}) \right]^2$$

$$= \frac{1}{n} \cdot E \sum_{i=1}^{n} \left[(x_i - \bar{y})^2 - 2(x_i - \bar{y}) \cdot (\bar{x} - \bar{y}) + (\bar{x} - \bar{y})^2 \right]$$

$$= \frac{1}{n} \cdot E \left[\sum_{i=1}^{n} (x_i - \bar{y})^2 - 2 \sum_{i=1}^{n} (x_i - \bar{y}) \cdot (\bar{x} - \bar{y}) + \sum_{i=1}^{n} (\bar{x} - \bar{y})^2 \right]$$

وبملاحظة أن:

$$\sum_{i=1}^{n} (x_i - \bar{y}) \cdot (\bar{x} - \bar{y}) = (\bar{x} - \bar{y}) \cdot \left(\sum_{i=1}^{n} x_i - n\bar{y} \right)$$

$$= (\bar{x} - \bar{y})(n\bar{x} - n\bar{y}) = n(\bar{x} - \bar{y})^2$$

وكذلك أيضاً:

$$\sum_{i=1}^{n} (\bar{x} - \bar{y})^2 = n(\bar{x} - \bar{y})^2$$

وبذلك نجد أن:

$$E\left(D^2\right)=\frac{1}{n}\cdot E\left[\sum_{i=1}^{n}\left(x-\bar{y}\right)^2-n\left(\bar{x}-\bar{y}\right)^2\right]$$

$$E\left(D^2\right)=\frac{1}{n}\cdot\sum_{i=1}^{n}E\left(x_i-\bar{y}\right)^2-\frac{n}{n}E\left(\bar{x}-\bar{y}\right)^2$$

$$E\left(D^2\right)=\frac{1}{n}\cdot\sum\sigma^2-\sigma_{\bar{x}}^2=\sigma^2-\sigma_{\bar{x}}^2 \qquad (8-4)$$

وذلك لأن التوقع $E\left(x_i-y\right)^2$ يساوي تباين المجتمع σ^2.ولأن التوقع $E\left(\bar{x}-\bar{y}\right)^2$ يساوي تباين المتوسطات $\bar{x}$ عن $\bar{y}$ ورمزنا له بـ $\sigma_{\bar{x}}^2$.

ومن العلاقة الأخيرة نلاحظ أن $E\left(D\right)^2\neq\sigma^2$؛ أي أن تباين العينة D^2 لا يمثل تقديراً غير متحيز لتباين المجتمع σ^2.

ولإيجاد التقدير غير المتحيز لـ σ^2 يجب حساب قيمة $\sigma_{\bar{x}}^2$ بدلالة σ^2. ومعلوم في الإحصاء الرياضي أن قيمة $\sigma_{\bar{x}}^2$ بدلالة σ^2 تساوي:

$$\sigma_x^2=\begin{cases}\dfrac{\sigma^2}{n} & \text{في حالة السحب مع الإعادة}\\ \dfrac{N-n}{N-1}\cdot\dfrac{\sigma^2}{n} & \text{في حالة السحب بدون إعادة}\end{cases} \qquad (9-4)$$

وبالتعويض في (4-8) نجد أنه في حالة السحب مع الإعادة أن:

$$E\left(D^2\right)=\sigma^2-\frac{\sigma^2}{n}=\frac{n-1}{n}\cdot\sigma^2\neq\sigma^2$$

وهو مايزال لايمثل تقديراً غير متحيز للتباين σ^2. ولكن يمكننا الاستفادة من العلاقة الأخيرة للحصول على تقدير غير متحيز لـ σ^2، وذلك بضرب D^2 بالمقدار $\frac{n}{n-1}$ فنجد أن:

$$E\left(\frac{n}{n-1}D^2\right)=\frac{n}{n-1}E\left(D^2\right)=\frac{n}{n-1}\cdot\frac{n-1}{n}\sigma^2=\sigma^2$$

أي أنه في حالة السحب مع الإعادة يكون المقدار $\left(\frac{n}{n-1}D^2\right)$ تقديراً غير متحيز للتباين σ^2. فإذا رمزنا له بـ s^2 نجد أن:

$$s^2=\frac{n}{n-1}D^2=\frac{n}{n-1}\cdot\frac{1}{n}\cdot\sum_{i=1}^{n}(x_i-\bar{x})^2=\frac{1}{n-1}\cdot\sum_{i=1}^{n}(x_i-\bar{x})^2$$

وبذلك نجد أن التقدير غير المتحيز لـ σ^2 هو التقدير التالي (في حالة السحب مع الإعادة):

$$\tilde{\sigma}^2=s^2=\frac{1}{n-1}\cdot\sum_{i}^{n}(x_i-\bar{x})^2 \qquad (10-4)$$

أما في حالة السحب بدون إعادة، فنجد أنه لو عوضنا $\sigma_{\bar{x}}^2$ في العلاقة (4-8) نحصل على:

$$E(D^2)=\sigma^2-\frac{N-n}{N-1}\cdot\frac{\sigma^2}{n}$$

$$E(D^2)=\frac{\sigma^2}{n}\cdot\left[\frac{n(N-1)-(N-n)}{N-1}\right]$$

$$E(D^2)=\frac{n-1}{n}\cdot\frac{N}{N-1}\cdot\sigma^2$$

ولإيجاد تقدير غير متحيز σ^2 ، نأخذ المقدار السابق $\frac{n}{n-1}D^2$ نفسه، فنجد أن:

$$E\left(\frac{n}{n-1}D^2\right)=E(s^2)=\frac{N}{N-1}\cdot\sigma^2$$

وإذا قمنا بتعديل آخر وأخذنا المقدار $\frac{N-1}{N}S^2$ فإننا نجد أن:

$$E\left(\frac{N-1}{N}s^2\right)=\sigma^2$$

وبذلك نجد أن تقدير التباين σ^2 يساوي (في حالة السحب بدون إعادة):

$$\widetilde{\sigma}^2 = \frac{N-1}{N} s^2 = \frac{N-1}{N(n-1)} \cdot \sum_i^n (x_i - \bar{x})^2 \qquad (11-4)$$

وبما أن المقدار $\frac{N-1}{N}$ قريب جداً من العدد (1) ويمكننا اعتباره مساوياً للواحد؛ أي أنه يمكننا اعتبار تباين العينة s^2 تقديراً لـ σ^2 في كلتا حالتي السحب، ونكتب:

$$\widetilde{\sigma}^2 = s^2 = \frac{1}{n-1} \cdot \sum_i^n (x_i - \bar{x})^2 \qquad (12-4)$$

مثال 3:

سحبنا عينة بحجم $n=12$ عاملاً من عمال إحدى الشركات التي تضم 200 عاملاً ودرسنا مقادير أجورهم فوجدنا أنها كانت تساوي:

$\bar{x}_i$: 1500 1600 1700 1800 1900 2000
2100 1850 1910 1740 1350 1670

فما هو تقدير كل من متوسط وإجمالي الأجور وتباينها ؟

الحل:

إن تقدير متوسط الأجر يعطى بواسطة العلاقة:

$$\widetilde{\bar{y}} = \bar{x} = \frac{\sum x_i}{n}$$

$$\widetilde{\bar{y}} = \bar{x} = \frac{1500+1600+1700+1800+1900+.....+1740+1350+1760}{12}$$

$$\widetilde{\bar{y}} = 1760\ L.S$$

وإنّ تقدير إجمالي الأجور في الشركة يعطى بالعلاقة:

$$\widetilde{Y} = N\widetilde{\bar{y}} = 200 \cdot 1760 = 352000 L.S$$

وإنّ تقدير تباين الأجور يعطى بالعلاقة:

$$\tilde{\sigma}^2 = s^2 = \frac{1}{n-1} \cdot \sum_i^n (x_i - \bar{x})^2$$

$$\tilde{\sigma}^2 = s^2 = \frac{1}{11}\left[(1500-1760)^2 + (1600-1760)^2 + ... + (1670-1760)^2\right]$$

$$\tilde{\sigma}^2 = s^2 = 45309{,}1$$

ويمكن الحصول على هذا التقدير من الآلة الحاسبة مباشرة باستخدام المفتاح σ_{n-1} ثم تربيعه فنحصل على المطلوب. وهنا نشير إلى أنه لا علاقة لكيفية السحب بالتقديرات السابقة لأنها تقديرات موحدة في كلتا حالتي السحب.

4-3-4: تقدير تباين تقدير المتوسط $\sigma_{\bar{x}}^2$:

نعلم أن $\sigma_{\bar{x}}^2$ هو عبارة عن تباين متوسطات العينات الممكنة $\bar{x}$ عن متوسط المجتمع $\bar{y}$ ويساوي:

$$\sigma_{\bar{x}}^2 = E(\bar{x} - \bar{y})^2$$

وبما أن $\sigma_{\bar{x}}^2$ معطى بدلالة التباين σ^2 ، وفق العلاقة (4-9) وحسب حالتي السحب، فإن تقديره إذن يساوي لـ:

$$\tilde{\sigma}_x^2 = \begin{cases} \dfrac{\tilde{\sigma}^2}{n} & \text{في حالة السحب مع الإعادة} \\ \dfrac{N-n}{N} \cdot \dfrac{\tilde{\sigma}^2}{n} & \text{في حالة السحب بدون إعادة} \end{cases}$$

وباستبدال التقدير $\tilde{\sigma}^2$ من العلاقتين (4-10) و(4-11) نحصل على التقديرين التاليين للمقدار $\tilde{\sigma}_{\bar{x}}^2$ ، وذلك حسب حالتي السحب:

$$\tilde{\sigma}_{\bar{x}}^2 = \begin{cases} \dfrac{s^2}{n} & \text{في حالة السحب مع الإعادة} \\ \dfrac{N-n}{N} \cdot \dfrac{s^2}{n} & \text{في حالة السحب بدون إعادة} \end{cases} \quad (4-12)$$

وهنا نشير إلى أنه إذا كان حجم المجتمع N كبيراً، فإن $\frac{N-n}{N-1} \approx 1$، وبذلك يمكننا اعتبار المقدار $\frac{s^2}{n}$، في هذه الحالة وفي كلتا حالتي السحب، تقديراً للتباين $\widetilde{\sigma}_{\bar{x}}^2$.

4-3-5: تقدير تباين تقدير إجمالي المجتمع:

بما أن تقدير الإجمالي يعطى بالعلاقة:

$$\widetilde{Y} = N \cdot \bar{x}$$

فإن تباين التقدير $\widetilde{Y}$ (حسب خواص التباين) يعطى بالعلاقة التالية:

$$\sigma_{\widetilde{Y}}^2 = N^2 \sigma_{\bar{x}}^2$$

وإن تقديره يعطى بالعلاقة التالية:

$$\widetilde{\sigma}_{\widetilde{Y}}^2 = N^2 \widetilde{\sigma}_{\bar{x}}^2 = \begin{cases} N^2 \dfrac{s^2}{n} & \text{في حالة السحب مع الإعادة} \\ N^2 \dfrac{N-n}{N} \cdot \dfrac{s^2}{n} & \text{في حالة السحب بدون إعادة} \end{cases} \qquad (13\text{-}4)$$

مثال 4:

بفرض أن العينة الواردة في المثال 3 السابق، قد سحبت بدون إعادة، فما هو تقدير تباين متوسط المجتمع $\sigma_{\bar{x}}^2$؟ وما هو تقدير تباين تقدير إجمالي الأجور؟

الحل:

من معطيات المسألة ونتائج الحل في المثال 3 ، نجد أن تقدير تباين تقدير متوسط المجتمع يساوي:

$$\widetilde{\sigma}_{\bar{x}}^2 = \frac{N-n}{N} \cdot \frac{s^2}{n} = \frac{200-12}{200} \cdot \frac{45309,1}{12} = 3549,21$$

ومنه نجد أن تقدير تباين تقدير إجمالي الأجور يساوي:

$$\widetilde{\sigma}_{\bar{x}}^2 = N^2 \widetilde{\sigma}_{\bar{x}}^2 = (200)^2 \cdot (3549,21) = 141968513,3$$

4-3-6: تقدير نسبة خاصة ما في المجتمع (نسبة مؤشر نوعي):

لتقدير نسبة خاصة ما في المجتمع لابد من تحديد عدد عناصر المجتمع التي تتصف بها، والذي سنرمز له بـ M. لذلك فإننا نفترض متحولاً Y يأخذ القيمة (1) إذا كان العنصر يتصف بالخاصة المطلوبة، والقيمة (0) إذا كان العنصر لا يتصف بها. وبذلك يكون عدد العناصر التي تتصف بالخاصة مساوياً لمجموع قيم المتحول Y؛ أي أن:

$$M = \sum_{i=1}^{N} y_i$$

حيث إن: 1 أو $y_i = 0$.

وبذلك تأخذ النسبة R في المجتمع الشكل التالي:

$$R = \frac{M}{N} = \frac{\sum_{i=1}^{N} y_i}{N} = \bar{y} \qquad (14-4)$$

أي أنه يمكننا النظر إلى النسبة في المجتمع على أنها متوسط قيم متحول يأخذ القيمتين (1) و(0). وكذلك الأمر في العينة حيث نرمز بـ X للمتحول الذي يرصد لنا تواجد الخاصة في عناصرها، ونعطيه القيمة (1) إذا كان العنصر متصفاً بتلك الخاصة، والقيمة (0) إذا لم يتصف بها.

وبذلك يكون عدد عناصر العينة التي تتصف بالخاصة المدروسة يساوي m ويساوي مجموع قيم المتحول X؛ أي أن:

$$m = \sum_{i=1}^{n} x_i$$

وبذلك تأخذ النسبة r في العينة الشكل التالي:

$$r = \frac{m}{n} = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x} \qquad (15-4)$$

وبما أن $\bar{x}$ هو تقدير لـ $\bar{y}$ وبالقياس على ما برهناه في حالة تقدير متوسط المجتمع، فإننا نجد أن تقدير النسبة R في المجتمع يعطى بواسطة النسبة r في العينة؛ أي أن:

$$\widetilde{R} = r = \frac{m}{n} \qquad (16-4)$$

وإن هذا التقدير هو تقدير غير متحيز ومتماسك وفعال وكاف.

كما يقدر تباينه بواسطة العلاقة (4–12) السابقة. وذلك بعد استبدال s^2 بما تساويه حيث نجد أن في هذه الحالة يساوي:

$$s^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{1}{n-1} \cdot \left[\sum x_i^2 - n\bar{x}^2 \right]$$

وبما أن:

$$\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i = m = n \cdot r \qquad \text{وكذلك} \qquad \bar{x} = r$$

إذن يكون لدينا:

$$s^2 = \frac{1}{n-1}\left[nr - nr^2 \right] = \frac{n}{n-1} \cdot nr(1-r)$$

$$s^2 = \frac{n}{n-1} \cdot r(1-r) = \frac{n}{n-1} rq \qquad (17-4)$$

حيث: $q = 1 - r$

وإذا كان حجم العينة n كبيراً، فإن $\frac{n}{n-1} \approx 1$ ، وبذلك يمكن أن نعتمد على العلاقة التالية:

$$s^2 \approx rq \qquad (18-4)$$

وبالقياس على العلاقة (4–12) نجد أن تقدير تباين التقدير r يعطى بالعلاقة التالية:

$$\widetilde{\sigma}_r^2 = \begin{cases} \dfrac{rq}{n} & \text{في حالة السحب مع الإعادة} \end{cases} \qquad (19-4)$$

$$\frac{N-n}{N}\cdot\frac{rq}{n}$$

في حالة السحب بدون إعادة

مثال 5:

سحبنا عينة عشوائية من طلاب الجامعة بحجم $n=28$ طالباً، فوجدنا أن 7 طلاب يدخنون. فما هو تقدير نسبة المدخنين في الجامعة. وما هو تقدير تباينه إذا علمت أن السحب كان مع الإعادة؟.

إن تقدير نسبة المدخنين في الجامعة اعتماداً على العينة المسحوبة يساوي:

$$\widetilde{R}=r=\frac{7}{28}=\frac{1}{4}=0{,}25$$

وإن تقدير تباين هذا التقدير في حالة السحب مع الإعادة يساوي:

$$\widetilde{\sigma}_r^2=\frac{rq}{n}=\frac{(0{,}25)(0{,}75)}{28}=0{,}0067$$

4-3-7: تقدير الفرق بين متوسطي مجتمعين وتباينه:

لنفترض أنه لدينا مجتمعان مستقلان، وأن متوسط الخاصة المدروسة في المجتمع الأول هو $\bar{y}_1$ ومتوسطها في المجتمع الثاني هو $\bar{y}_2$. فإذا سحبنا عينة بحجم n_1 من المجتمع الأول، وكان متوسط الخاصة في هذه العينة هو $\bar{x}_1$، فإنه يكون فإن هذا المتوسط يكون تقديراً لمتوسط المجتمع الأول، أي يكون لدينا:

$$\widetilde{\bar{y}}_1=\bar{x}_1$$

وإذا سحبنا عينة بحجم n_2 من المجتمع الثاني، وكان متوسط الخاصة في هذه العينة هو $\bar{x}_2$، فإن هذا المتوسط يكون تقديراً لمتوسط المجتمع الثاني، أي يكون لدينا:

$$\widetilde{\bar{y}}_2=\bar{x}_2$$

وبذلك يمكننا أن نستنتج أن تقدير الفرق بين متوسطي المجتمعين $(\bar{y}_1-\bar{y}_2)$ هو الفرق بين متوسطي العينتين $(\bar{x}_1-\bar{x}_2)$ ، وبذلك يكون لدينا:

$$(\bar{y}_1-\bar{y}_2)=\bar{x}_1-\bar{x}_2 \qquad (20-4)$$

وبما أن العينتين مستقلتان فإن تباين الفرق $\left(\bar{x}_1 - \bar{x}_2\right)$ يساوي مجموع تباينهما؛ أي أن:

$$\sigma^2_{\left(\bar{x}_1 - \bar{x}_2\right)} = \sigma^2_{\bar{x}_1} + \sigma^2_{\bar{x}_1} \qquad (21-4)$$

وبذلك نجد أن تقدير تباين الفرق $\left(\bar{x}_1 - \bar{x}_2\right)$ يساوي:

$$\tilde{\sigma}^2_{\left(\bar{x}_1 - \bar{x}_2\right)} = \tilde{\sigma}^2_{\bar{x}_1} + \tilde{\sigma}^2_{\bar{x}_1} \qquad (22-4)$$

واعتماداً على العلاقة (4-12) فإننا نجد أن تقدير التباين $\sigma^2_{\left(\bar{x}_1 - \bar{x}_2\right)}$ يساوي حسب حالتي السحب مايلي:

$$\tilde{\sigma}^2_{\left(\bar{x}_1 - \bar{x}_2\right)} = \begin{cases} \text{في حالة السحب مع الإعادة} \\ \dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2} \\ \text{في حالة السحب بدون إعادة} \\ \dfrac{N_1 - n_1}{N_1} \cdot \dfrac{s_1^2}{n_1} + \dfrac{N_2 - n_2}{N_2} \cdot \dfrac{s_2^2}{n_2} \end{cases} \qquad (22-4)$$

حيث أن: s_1^2 و s_2^2 هما تباينا العينتين الأولى والثانية على الترتيب.

وبطريقة مشابهة نجد أن الفرق بين النسبتين R_1 و R_2 لخاصةً في المجتمعين المذكورين يقدر بواسطة الفرق بين النسبتين في العينتين r_1 و r_2؛ أي أن:

$$\left(R_1 - R_2\right) = \left(r_1 - r_2\right) \qquad (23-4)$$

وإن تباين تقدير الفرق $\left(R_1 - R_2\right)$ يقدر بواسطة العلاقة المزدوجة التالية:

$$\tilde{\sigma}^2_{(r_1 - r_2)} = \tilde{\sigma}^2_{r_1} + \tilde{\sigma}^2_{r_2} \begin{cases} \text{في حالة السحب مع الإعادة} \\ \dfrac{r_1 q_1}{n_1} + \dfrac{r_2 q_2}{n_2} \\ \text{في حالة السحب بدون إعادة} \\ \dfrac{N_1 - r_1}{N_1 - 1} \cdot \dfrac{r_1 q_1}{n_1} + \dfrac{N_2 - n_2}{N_2 - 1} \cdot \dfrac{r_2 q_2}{n_2} \end{cases} \qquad (24-4)$$

مثال 5:

لنفترض أننا سحبنا عينة من تجار دمشق بحجم $n_1 = 30$ تاجراً، و عينة من تجار حلب بحجم $n_2 = 20$ تاجراً. وكان متوسط دخل عينة تجار دمشق $\bar{x}_1 = 20000$ ل.س شهرياً، وكان متوسط دخل عينة تجار حلب $\bar{x}_2 = 15000$ ل.س. شهرياً.
فما هو تقدير الفرق بين متوسطي الدخل للتجار في كل من المدينتين وتقدير تباينه؟
إذا علمت أن تباين العينة الأولى $s_1^2 = 3000$ و تباين العينة الثانية $s_2^2 = 2500$.

الحل:

إن تقدير الفرق بين المتوسطين يساوي:

$$\left(\bar{y}_1 - \bar{y}_2\right) = \left(\bar{x}_1 - \bar{x}_2\right) = (20000 - 15000) = 5000\, L.S$$

ولتقدير تباينه نلاحظ أولاً أن مجتمع التجار في كل من المدينتين يعد مجتمعاً كبيراً. لذلك نعتبر أن العينتين قد سحبتا بطريقة السحب مع الإعادة ونجد أن:

$$\tilde{\sigma}^2_{(\bar{x}_1 - \bar{x}_2)} = \frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} = \frac{3000}{30} + \frac{2500}{20} = 225$$

4-4: مجالات الثقة:

بعد أن درسنا بعض مسائل التقدير النقطية (العددية) لبعض مؤشرات المجتمع، وتوصلنا إلى علاقات رياضية لحساب كل منها، ننتقل إلى إيجاد مجال ثقة لكل من تلك التقديرات، وذلك لأنه على الرغم من أن معظم التقديرات النقطية غير متحيزة وفعالة ومتماسكة وكافية للمؤشر المجهول، إلا أن أياً منها لا يعطينا أية درجة من الثقة فيه. فنحن لا نعرف فيما إذا كانت قيمة التقدير النقطي قريبة من القيمة الحقيقية المجهولة أو بعيدة عنها. ومن هنا كان لابد من إيجاد وسيلة تضمن لنا وباحتمال كبير أن تكون القيمة الحقيقية المجهولة واقعة في مجال معين حول القيمة التقديرية.
فإذا كان θ مؤشراً في المجتمع (كالمتوسط أو الإجمالي ....) وكان تقديره $\tilde{\theta}$ من العينة، فكيف يمكننا أن ننشئ مجالاً يحوي القيمة الحقيقية θ باحتمال قدره β؟ أو بمستوى دلالة قدره α حيث $\alpha = 1 - \beta$؟

يمكن إنشاء مجال الثقة المطلوب إذا استطعنا تحديد مركزه C، وتحديد نصف طوله ε، فيكون المجال المطلوب لـ θ هو المجال الذي يحقق العلاقة التالية:

$$P(c-\varepsilon<\theta<c+\varepsilon)=1-\alpha=\beta \qquad (25-4)$$

وحتى نستطيع تحديد كل من C و ε، فإننا نفترض أن التقديرات الممكنة $\widetilde{\theta}$ خاضعة للتوزيع الطبيعي الذي توقعه $E(\widetilde{\theta})=\theta$ وانحرافه المعياري يقدر بـ $\widetilde{\sigma}_{\widetilde{\theta}}$. فإذا كان حجم العينة n كبيراً بشكل كافٍ $(n\geq 30)$، فإننا نجد أن المتحول المعياري t إذ:

$$t=\frac{\widetilde{\theta}-\theta}{\widetilde{\sigma}_{\widetilde{\theta}}} \qquad (26-4)$$

يكون خاضعاً للتوزيع الطبيعي المعياري N(0,1).

وبما أننا نريد أن ننشئ مجالاً يحوي القيمة الحقيقية المجهولة θ باحتمال قدره β، فإنه استناداً إلى العلاقة (3−41) نجد أن ذلك الاحتمال المأخوذ على مجال متناظر يساوي:

$$P(-Z\leq t\leq Z)=2\Phi(Z_{\beta})-1=\beta \qquad (27-4)$$

حيث إن Z هو نصف طول المجال المعياري الذي يجعل ذلك الاحتمال مساوياً للمقدار β.

فإذا كان الاحتمال β معلوماً (وغالباً ما نجعله مساوياً لـ 0,95 أو 0,99) فإنه يمكننا حساب Z المقابلة له وذلك باستخدام جدول تابع التوزيع الطبيعي المعياري وهو الجدول رقم (III) الملحق بآخر الكتاب، وذلك بطريقة معاكسة، فنجد أن:

$$Z=\Phi^{-1}\left(\frac{1+\beta}{2}\right)=\Phi^{-1}\left(1-\frac{\alpha}{2}\right) \qquad (28-4)$$

حيث أن $\Phi^{-1}\left(1-\frac{\alpha}{2}\right)$ هو قيمة المتحول Z المقابلة للاحتمال $\left(1-\frac{\alpha}{2}\right)$، لذلك نستبدل الرمز Z بالرمز $Z_{\left(1-\frac{\alpha}{2}\right)}$ للدلالة على قيمة Z المقابلة للاحتمال $\left(1-\frac{\alpha}{2}\right)$.

وبتعويض t في قيمتها من العلاقة (4–26) نحصل على العلاقة التالية:

$$P\left(-Z_{1-\frac{\alpha}{2}} \leq \frac{\tilde{\theta}-\theta}{\tilde{\sigma}_{\tilde{\theta}}} \leq Z_{1-\frac{\alpha}{2}}\right)=\beta=1-\alpha \qquad (29\text{-}4)$$

وبما أننا نريد إنشاء مجال ثقة يحوي القيمة الحقيقية θ فإننا نضرب أطراف المتراجحة بـ $\tilde{\sigma}_{\tilde{\theta}}$، ثم نطرح منها $\tilde{\theta}$، ثم نضرب الأطراف الناتجة بـ (–1) فنحصل على العلاقة التالية:

$$P\left(\tilde{\theta}-Z_{1-\frac{\alpha}{2}}\tilde{\sigma}_{\tilde{\theta}} \leq \theta \leq \tilde{\theta}+Z_{1-\frac{\alpha}{2}}\tilde{\sigma}_{\tilde{\theta}}\right)=1-\alpha=\beta \qquad (30-4)$$

ومن هذه العلاقة يتضح لنا أننا توصلنا إلى مجال حقيقي يحوي θ باحتمال β، أو باحتمال $(1-\alpha)$. وإن مركزه هو التقدير $\tilde{\theta}$ ونصف طوله $\varepsilon = Z_{1-\frac{\alpha}{2}}\tilde{\sigma}_{\beta}$.

وأخيراً نشير إلى أنه إذا كان حجم العينة n صغيراً $(n<30)$، فإن المقدار:

$$t=\frac{\tilde{\theta}-\theta}{\tilde{\sigma}_{\tilde{\theta}}}$$

يكون خاضعاً لتوزيع (ستودينت) ذي $(n-1)$ درجة حرية وتأخذ العلاقة (4–29) الشكل التالي:

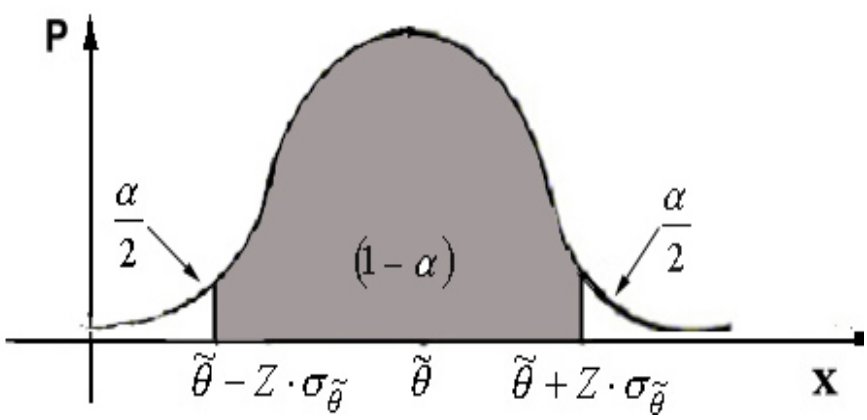
$$P\left(-t_{\beta} \leq \frac{\tilde{\theta}-\theta}{\tilde{\sigma}_{\tilde{\theta}}} \leq t_{\beta}\right)=\beta=1-\alpha \qquad (31-4)$$

إذ أن t هي قيمة متحول (ستودينت) المقابل للاحتمال $1-\frac{\alpha}{2}$ ولـ $(n-1)$ درجة حرية، لذلك نرمز لها بالرمز $t_{1-\frac{\alpha}{2}}$، وتؤخذ من الجدول IV الملحق في آخر الكتاب.

وبإصلاح العلاقة (4–31) نحصل على مجال الثقة التالي:

$$P\left(\tilde{\theta}-t_{1-\frac{\alpha}{2}}\tilde{\sigma}_{\tilde{\theta}} \leq \theta \leq \tilde{\theta}+t_{1-\frac{\alpha}{2}}\tilde{\sigma}_{\tilde{\theta}}\right) \qquad (32-4)$$

وأن الشكل (4-4) يوضح لنا مجال الثقة واحتماله $(1-\alpha)$.



الشكل رقم (4-4)

وسنقوم بتطبيق هذه العلاقات عند إنشاء مجالات ثقة لكل من المتوسط والإجمالي والنسبة وغيرها.

4-4-1: إنشاء مجال ثقة لمتوسط مجتمع طبيعي:

لقد رمزنا لمتوسط المجتمع بـ $\bar{y}$ ، ولتقديره من خلال متوسط العينة بـ $\bar{x}$ ولانحرافه المعياري بـ $\sigma_{\bar{x}}$.

وإن تقدير $\sigma_{\bar{x}}$ يحسب من تقدير التباين $\sigma^2_{\bar{x}}$ المعطى بالعلاقة (4-12) كمايلي:

$$\tilde{\sigma}_{\bar{x}}\begin{cases}\sqrt{\dfrac{s^2}{n}} & \text{في حالة السحب مع الإعادة}\\[2ex] \sqrt{\dfrac{N-n}{N}\cdot\dfrac{s^2}{n}} & \text{في حالة السحب بدون إعادة}\end{cases}$$

مع العلم أنه إذا كان حجم المجتمع كبيراً، فإن $\dfrac{N-n}{N-1}\approx 1$، وعندها نعتبر السحب بدون إعادة وكأنه سحب مع الإعادة، ويكون $\tilde{\sigma}_{\bar{x}}=\dfrac{s}{\sqrt{n}}$ في كلتا حالتي السحب.

وبالاعتماد على العلاقة (4–30) نجد أن مجال الثقة الذي يحوي المتوسط $\bar{y}$ باحتمال قدره β هو المجال التالي:

$$P\left(\bar{x}-Z_{1-\alpha/2}\cdot\frac{s}{\sqrt{n}}\le\bar{y}\le\bar{x}+Z_{1-\alpha/2}\cdot\frac{s}{\sqrt{n}}\right)=\beta=1-\alpha \qquad (33-4)$$

وعندما يكون حجم العينة n صغيراً، فإننا نستبدل $Z_{\frac{1-\alpha}{2}}$ بـ $t_{\frac{1-\alpha}{2}}$ المقابلة لـ $(n-1)$ درجة حرية في توزيع ستودينت.

مثال 6:

سحبنا عينة (بدون إعادة) من طلاب جامعة تشرين بحجم $n=50$ طالباً، وسألناهم عن علاماتهم في مقرر الثقافة. فوجدنا أن متوسط علاماتهم كان يساوي $\bar{x}=55$، وتباينها $s^2=100$. فما هو مجال الثقة الذي يحوي متوسط هذه العلامة باحتمال قدره $\beta=0{,}95$ ؟

الحل:

إن المجال المطلوب هو:

$$P\left(\bar{x}-Z_{1-\alpha/2}\cdot\frac{s}{\sqrt{n}}\le\bar{y}\le\widetilde{\theta}+\bar{x}+Z_{1-\alpha/2}\cdot\frac{s}{\sqrt{n}}\right)=1-\alpha$$

وذلك لأن حجم المجتمع N كبيراً، ولهذا اعتبرنا السحب بدون إعادة وكأنه سحب مع الإعادة، وهذا يعني أننا جعلنا: $\frac{N-n}{N}\cdot\frac{s^2}{n}\approx\frac{s^2}{n}$، وبالتعويض نجد أن:

$$P\left(55-Z_{1-\alpha/2}\cdot\frac{10}{\sqrt{50}}\le\bar{y}\le55+Z_{1-\alpha/2}\cdot\frac{10}{\sqrt{50}}\right)=\beta$$

ولحساب $Z_{1-\alpha/2}$ نعود إلى جدول تابع التوزيع الطبيعي رقم (III) الملحق في آخر الكتاب (وذلك لأن حجم العينة n يعد كبيراً)، وبما $Z_{1-\alpha/2}$ هي قيمة Z المقابلة

للاحتمال $1-\frac{\alpha}{2}$، فإننا نجد أن الاحتمال المقابل هو $1-\frac{0,05}{2}=0,975$،

وإن الجدول يعطينا أن قيمة Z المقابلة للاحتمال 0,975 تساوي:

$$Z_{1-\alpha/2}=Z_{0,975}=1,96$$

وبذلك نجد أن المجال المطلوب يساوي:

$$P\left(55-(1,96)\cdot\frac{10}{\sqrt{50}}\le\bar{y}\le 55+(1,96)\cdot\frac{10}{\sqrt{50}}\right)=0,95$$

$$P(52,23\le\bar{y}\le 57,77)=0,95$$

أي أن متوسط العلامة يكون محصوراً بين 52,23 و57,77 باحتمال قدره 0,95.

ملاحظة:

يجب الانتباه دائماً عند استخراج قيمة $Z_{\frac{1-\alpha}{2}}$ و $t_{\frac{1-\alpha}{2}}$ بأنهما القيمتان اللتان تقابلان الاحتمال $\left(1-\frac{\alpha}{2}\right)$ وليس الاحتمال الوارد في مجال الثقة $\beta=1-\alpha$.

مثال 7:

إذا كانت لدينا القياسات التالية عن درجات الحرارة خلال أسبوع ما (مأخوذة يومياً في الساعة 12 تماماً):

15 , 20 , 17 , 18 , 22 , 16 , 19

فما هو تقدير متوسط درجة الحرارة خلال ذلك الأسبوع؟ وما هو مجال الثقة المقابل للاحتمال $\beta=0,95$ ؟

الحل:

بما أن حجم العينة $n=7$ صغير، فإننا نعتمد على توزيع (ستودينت) ونستخدم $t_{\frac{1-\alpha}{2}}$ في العلاقة (4-34). وبذلك يأخذ مجال الثقة للمتوسط الشكل التالي:

$$P\left(\bar{x}-t_{(1-\alpha/2)}\cdot\frac{s}{\sqrt{n}}\le\bar{y}\le\bar{x}+t_{(1-\alpha/2)}\cdot\frac{s}{\sqrt{n}}\right)=\beta$$

ثم نقوم بحساب المقادير الداخلة في طرفي المجال، فنجد أن:

$$\widetilde{\bar{y}} = \bar{x} = 18{,}14$$

$$s = \sqrt{\frac{1}{n-1}\sum\left(x_i - \bar{x}\right)^2} = 2{,}41$$

ونحصل على $t_{1-\alpha/2}$ من جدول قيم t لتوزيع ستودينت، وهي القيمة t_0 المقابلة للاحتمال $1-\frac{\alpha}{2} = 0{,}975$ ولدرجات الحرية $K = 7-1 = 6$؛ أي إنها القيمة الواقعة على تقاطع عمود الاحتمال 0,975 وسطر درجات الحرية 6 وتساوي:

$$t_{(6,975)} = 2{,}447$$

وبذلك نجد أن مجال الثقة المطلوب هو:

$$P\left(18{,}14 - (2{,}447)\cdot\frac{2{,}41}{\sqrt{7}} \le \bar{y} \le 18{,}14 + (2{,}447)\cdot\frac{2{,}41}{\sqrt{7}}\right) = 0{,}95$$

$$P\left(15{,}91 \le \bar{y} \le 20{,}37\right) = 0{,}95$$

أي أن درجة الحرارة خلال هذا الأسبوع كانت ضمن 15,91 و 20,37 باحتمال 0,95.

4-4-2- إنشاء مجال ثقة لإجمالي مجتمع طبيعي:

لقد رمزنا لإجمالي الخاصة في المجتمع بالرمز Y ولتقديره بـ $\widetilde{Y}$ ولانحرافه المعياري بـ $\sigma_{\widetilde{Y}}$ ووجدنا أن تقديره يعطى بواسطة جذر تباينه الذي يحسب من العلاقة (4-13)، وبذلك نجد أن:

$$\widetilde{\sigma}_{\widetilde{Y}} = N\widetilde{\sigma}_{\bar{X}} = \begin{cases} N\dfrac{s}{\sqrt{n}} \\ \\ \sqrt{\dfrac{N-n}{N}\cdot\dfrac{s}{n}} \end{cases} \qquad (34-4)$$

وبالاعتماد على العلاقة (4-31) نجد أن مجال الثقة الذي يحوي الإجمالي Y باحتمال قدره β هو المجال التالي:

$$P\left(\widetilde{Y} - Z_{1-\alpha/2} \cdot \widetilde{\sigma}_{\widetilde{Y}} \le \overline{Y} \le \widetilde{\theta} + \widetilde{Y} + Z_{1-\alpha/2} \cdot \widetilde{\sigma}_{\widetilde{Y}}\right) = \beta \qquad (35-4)$$

وعندما يكون حجم العينة n صغيراً، فإننا نستبدل $Z_{1-\alpha/2}$ بـ $t_{1-\alpha/2}$ المقابلة لـ (n - 1) درجة حرية.

مثال 9:

أوجد مجال الثقة الذي يقابل احتمالاً $\beta = 0,99$ لإجمالي أجور 200 عاملاً. إذا علمت أن عينة بحجم 12 عاملاً سحبت منهم بدون إعادة، أعطتنا تقديراً لمتوسط الأجر مساوياً $\bar{x} = 1760$ وتقديراً لتباين الأجر مساوياً $s^2 = 45309$.

الحل:

قبل كل شئ نلاحظ أن حجم المجتمع ليس بكبير. وبما أن السحب جرى بدون إعادة. لذلك فإن تقدير الانحراف المعياري $\widetilde{\sigma}_{\widetilde{Y}}$ يعطى بالعلاقة التالية:

$$\widetilde{\sigma}_{\widetilde{Y}} = N \cdot \sqrt{\frac{N-n}{N}} . \frac{s}{\sqrt{n}} = 200 \sqrt{\frac{200-12}{200}} \cdot \frac{\sqrt{45309}}{\sqrt{12}} = 11915$$

وأن تقدير الإجمالي نفسه يساوي:

$$\widetilde{Y} = N \cdot \bar{x} = 200 \cdot 1760 = 352000$$

وبما أن حجم العينة صغير فإننا نستبدل $Z_{1-\alpha/2}$ بـ $t_{1-\alpha/2}$ ونحسب من جدول توزيع (ستودينت) قيمة t التي تقابل الاحتمال $1 - \frac{\alpha}{2} = 1 - \frac{0,01}{2} = 0,995$ ودرجة الحرية $k = 12 - 1 = 11$ فنجد أن:

$$t_\beta = t_{(0,995,11)} = 3,106$$

وبذلك نجد أن المجال المطلوب هو:

$$P\left((Y - t_{1-\alpha/2} \cdot \widetilde{\sigma}_{\widetilde{Y}} \le Y \le \widetilde{Y} + t_{1-\alpha/2} \cdot \widetilde{\sigma}_{\widetilde{Y}})\right) = \beta \qquad (37-4)$$

$$P\left[352000 - (3{,}106)(11915)\right] \le Y \le \left[352000 + (3{,}106)(11915)\right]$$

$$= 0{,}99$$

$$P[314992 \le Y \le 389008] = 0{,}99$$

مثال 10:

سحبنا بدون إعادة عينة بحجم n=40 أسرة، من مجتمع فيه N=2500 أسرة فوجدنا أن تقدير متوسط عدد أفراد الأسرة $\bar{x} = 6$ وأن تقدير تباينه يساوي $s^2 = 50$. أوجد تقديراً لإجمالي عدد الأفراد في ذلك المجتمع، ثم أوجد مجال ثقة له باحتمال قدره 0,96، أو بمستوى دلالة $\alpha = 0{,}04$.

الحل:

من المعروف أن تقدير الإجمالي يعطى بالعلاقة:

$$\widetilde{Y} = N \cdot \bar{x} = 2500 \cdot 6 = 15000$$

وأن مجال الثقة المطلوب هو:

$$P\left(\widetilde{Y} - Z_{1-\alpha/2} \cdot \widetilde{\sigma}_{\widetilde{Y}} \le \widetilde{Y} \le \widetilde{Y} + Z_{1-\alpha/2} \cdot \widetilde{\sigma}_{\widetilde{Y}}\right) = \beta$$

وبما أن السحب جرى بدون إعادة فإن:

$$\widetilde{\sigma}_{\widetilde{Y}} = N \cdot \sqrt{\frac{N-n}{N}} . \frac{S}{\sqrt{n}} = 2500 \cdot (0{,}98) \frac{\sqrt{50}}{\sqrt{40}} = 2739{,}1$$

وإن قيمة Z المقابلة للاحتمال $1 - \frac{\alpha}{2} = 0{,}98$ تساوي $Z = 2{,}06$، وبذلك نجد أن المجال المطلوب هو:

$$P\left[15000 - (2{,}06)(2739)\right] \le Y \le \left[15000 + (2{,}06)(2739)\right] = 0{,}96$$

$$P[9357 \le Y \le 20642] = 0{,}96$$

4-4-3: إنشاء مجال ثقة للنسبة R في مجتمع طبيعي:

إن إنشاء مجال ثقة للنسبة R لايختلف كثيراً عن إنشاء مجال ثقة للمتوسط. وبما أن تقدير R يعطى بواسطة r في العينة وأن تباينها يقدر بـ $\widetilde{\sigma}_r^2 = \frac{rq}{n}$ أو $\widetilde{\sigma}_r^2 = \frac{N-n}{N} \cdot \frac{rq}{n}$، وذلك حسب حالتي السحب.

وبالاعتماد على العلاقة (4-32) نجد أن مجال الثقة الذي يحوي النسبة R باحتمال قدره β هو المجال التالي:

$$P\left(r - Z_{1-\frac{\alpha}{2}} \widetilde{\sigma}_r \le R \le r + Z_{1-\frac{\alpha}{2}} \widetilde{\sigma}_r) \right) = \beta \quad (38-4)$$

وعندما يكون حجم العينة صغيراً فإننا نستبدل $Z_{1-\alpha/2}$ بـ $t_{1-\alpha/2}$ والمقابلة لـ $(n-1)$ درجة حرية.

مثال 11:

لتقدير نسبة الأسر التي تسكن بـSecالإيجار سحبنا بدون إعادة عينة بحجم $n = 100$ أسرة من مجتمع مؤلف من 8000 أسرة، فوجدنا أن 40 منها تسكن بالإيجار. أوجد تقديراً لنسبة المستأجرين، ثم أوجد مجال ثقة له باحتمال قدره 0,95.

الحل:

نعلم أن تقدير النسبة في المجتمع R يساوي:

$$\widetilde{R} = r = \frac{40}{100} = 0{,}40$$

وبذلك نجد أن تقدير تباينها:

$$\widetilde{\sigma}_r^2 = \frac{N-n}{N} \cdot \frac{rq}{n} = \frac{8000-100}{8000} \cdot \frac{(0{,}40)(0{,}60)}{100} = 0{,}00237$$

وهكذا نجد أن:

$$\widetilde{\sigma}_r = 0{,}0487$$

ولتحديد مجال الثقة المطلوب يجب حساب $Z_{1-\alpha/2}$، حيث نجد أن قيمة Z المقابلة للاحتمال $1-\frac{\alpha}{2}=0,975$ تساوي $Z=1,96$، وبذلك نجد أن مجال الثقة المطلوب هو:

$$P[0,40-(1,96)(0,0487)\le R\le 0,40+(1,96)(0,0487)]=0,95$$
$$P[0,305\le R\le 0,495]=0,95$$

مثال 12:

لتقدير نسبة العطب في مصابيح الكهرباء سحبنا عينة بدون إعادة بحجم $n=25$ مصباحاً، فوجدنا أن نسبة العطب في العينة كانت $r=0,20$، أوجد مجال ثقة لنسبة العطب باحتمال قدره $0,90$.

الحل:

بما أن حجم المجتمع غير محدد لأنه كبير جداً فإننا نعتبر السحب بدون إعادة وكأنه سحب مع الإعادة، ونجد أن تقدير التباين $\tilde{\sigma}_r^2$ يساوي:

$$\tilde{\sigma}_r^2=\frac{rq}{n}=\frac{(0,20)(0,80)}{25}=0,0064$$

والانحراف المعياري يساوي:

$$\tilde{\sigma}_r=0,08$$

وبما أن حجم العينة صغير لذلك يجب استبدال $Z_{1-\frac{\alpha}{2}}$ بـ $t_{1-\frac{\alpha}{2}}$، ولحساب $t_{1-\frac{\alpha}{2}}$ نبحث في الجدول رقم (IV) عن قيمة t التي تقابل الاحتمال $1-\frac{\alpha}{2}=1-\frac{0,10}{2}=0,95$ ودرجة الحرية $k=25-1=24$ فنجد أن:

$$t(0,95\ ،\ 24)=1,711$$

وبذلك نجد أن مجال الثقة المطلوب يساوي:

$$P[0,20-(1,711)(0,08)\le R\le 0,20+(1,711)(0,08)]=0,90$$
$$P[0,06\le R\le 0,495]337=0,90$$

4-4-4: إنشاء مجال ثقة للفرق بين متوسطي مجتمعين طبيعيين:

لقد وجدنا أن تقدير الفرق بين متوسطي المجتمعين $(\bar{y}_1 - \bar{y}_2)$ يعطى بوساطة الفرق بين متوسطي العينتين $(\bar{x}_1 - \bar{x}_2)$ ، وكتبنا ذلك . أنظر العلاقة (4-20) . كمايلي:

$$(\bar{y}_1 - \bar{y}_2) = (\bar{x}_1 - \bar{x}_2)$$

كما وجدنا أن تباين هذا الفرق يقدر في حالة السحب مع الإعادة أو عندما يكون مجموع حجمي العينتين $(n_1 + n_2)$ كبيراً بوساطة العلاقة:

$$P\left(-Z_{1-\frac{\alpha}{2}} \le \frac{(\bar{x}_1 - \bar{x}_2) - (\bar{y}_1 - \bar{y}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} + Z_{1-\frac{\alpha}{2}} \right) = 1 - \alpha \qquad (39-4)$$

وقياساً على مافعلناه عند إنشاء مجال الثقة للمتوسط $\bar{y}$ نجد أن مجال الثقة المعياري للفرق $(\bar{y}_1 - \bar{y}_2)$ يعطى بالعلاقة:

$$P\left((\bar{x}_1 - \bar{x}_2) - Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \le (\bar{y}_1 - \bar{y}_2) \le (\bar{x}_1 - \bar{x}_2) + Z_{1-\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}) \right)$$
$$= 1 - \alpha \qquad (40-4)$$

مثال :

لنفترض إننا نريد دراسة الفرق بين متوسطي درجات الطلاب في مقرر الرياضيات في مدرستين متنافستين. فقمنا بسحب عينتين منهما مع الإعادة بحجمين $n_2 = 22, n_1 = 20$ فوجدنا أن متوسطي الدرجات فيهما كانا يساويان.

$\bar{x}_2 = 70, \bar{x}_1 = 60$ ، كما وجدنا أن تباينيهما كانا $s_2^2 = 200, s_1^2 = 100$.

والمطلوب: إيجاد مجال الثقة للفرق بين المتوسطين الحقيقيين $(\bar{y}_1 - \bar{y}_2)$ لدرجة الرياضيات. وذلك بمستوى دلالة أو باحتمال ثقة قدره *0,95*.

الحل:

بما أن مجموع حجمي العينتين كبير نسبياً $(n = 20 + 22 = 42)$، لذلك نحسب قيمة $Z_{1-\frac{\alpha}{2}}$ فنجد من الجدول أن: $Z_{1-\frac{\alpha}{2}} = Z_{1-\frac{0,05}{2}} = Z0,975 = 1,96$

ثم نقوم بتعويض البيانات السابقة في العلاقة (4-) السابقة فنحصل على أن:

$$P\left(\begin{array}{l}(60-70)-(1,96)\cdot\sqrt{\frac{100}{20}+\frac{200}{22}} \le \left(\overline{y}_1-\overline{y}_2\right) \\ \le (60-70)+(1,96)\cdot\sqrt{\frac{100}{20}+\frac{200}{22}}\end{array}\right)=0,95$$

$$P\left(-10-7,36 \le \left(\overline{y}_1-\overline{y}_2\right) \le -10-7,36\right)=0,95$$

$$P\left(-17,36 \le \left(\overline{y}_1-\overline{y}_2\right) \le -17,36\right)=0,95$$

ونستنتج من هذا المجال أن الفرق بين متوسطي درجتي الرياضيات في المدرستين يأخذ دائماً قيمة سالبة لأنه محصور بين قيمتين سالبتين. أي يوجد فرق بين المتوسطين في المدرستين.

حالة خاصة:

إذا كان المجتمعان متجانسين من حيث تباين قيم الخاصة المدروسة فإن ذلك يعني أن التباينين متساويين، وفي هذه الحالة يكون لدينا:

$$\sigma_1^2 = \sigma_2^2 = \sigma^2$$

وعندها فإن تقدير القيمة المشتركة للتباين σ^2 تعطى بواسطة تركيب غير متحيز من التباينين s_1^2, s_2^2 هو:

$$S^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2} \qquad (41-4)$$

وهكذا يمكننا أن نكتب ذلك التقدير على الشكل التالي: $\widetilde{\sigma}^2 = S^2$

أما تباين الفرق بين المتوسطين $\left(\overline{y}_1-\overline{y}_2\right)$ فيأخذ الشكل التالي:

$$\tilde{\sigma}^2_{(\bar{x}_1-\bar{x}_2)}=\frac{\sigma_1^2}{n_1}+\frac{\sigma_1^2}{n_2}=\sigma^2\left[\frac{1}{n_1}+\frac{1}{n_2}\right]$$

وإن تقديره في هذه الحالة يأخذ الشكل التالي:

$$\tilde{\sigma}^2_{(\bar{x}_1-\bar{x}_2)}=S^2\left[\frac{1}{n_1}+\frac{1}{n_2}\right]=$$

$$=\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}\cdot\left[\frac{1}{n_1}+\frac{1}{n_2}\right] \qquad (42-4)$$

وعندها فإن مجال الثقة المعياري للفرق $(\bar{y}_1-\bar{y}_2)$ يأخذ الشكل التالي:

$$P\left(-Z_{1-\frac{\alpha}{2}}\le\frac{(\bar{x}_1-\bar{x}_2)-(\bar{y}_1-\bar{y}_2)}{S\sqrt{\frac{1}{n_1}+\frac{2}{n_2}}}\le Z_{1-\frac{\alpha}{2}})\right)=1-\alpha \qquad (43-4)$$

وبعد إصلاحه يأخذ الشكل العادي التالي:

$$P\left((\bar{x}_1-\bar{x}_2)-Z_{1-\frac{\alpha}{2}}\cdot S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\le(\bar{y}_1-\bar{y}_2)\le(\bar{x}_1-\bar{x}_2)+Z_{1-\frac{\alpha}{2}}\cdot S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right)=$$

$$=1-\alpha \qquad (44-4)$$

ملاحظة:

لايجوز تطبيق هذه الحالة إلا بعد التأكد من أن: $\sigma_1^2=\sigma_2^2$.

مثال:

لنفترض أننا سحبنا عينتين من مجتمعين فيهما $\sigma_1^2=\sigma_2^2$ بحجمين $n_1=10$, $n_2=15$ فوجدنا أن: $\bar{x}_1=50,\bar{x}_2=40$ وأن: $s_1^2=60,s_2^2=80$.

والمطلوب إنشاء مجال الثقة للفرق $(\bar{y}_1-\bar{y}_2)$ باحتمال ثقة $0,95$.

الحل:

بما أن التباينين متساويان $\sigma_1^2=\sigma_2^2$ فرضاً. فإن هذه الحالة تنطبق على الحالة الخاصة السابقة، وبذلك نجد أن تقدير القيمة المشتركة للتباين يساوي:

$$S^2 = \frac{9 \cdot 60 + 14 \cdot 80}{10 + 15 - 2} = 72{,}17$$

وإن مجال الثقة المطلوب مع ملاحظة أن $Z_{1-\frac{\alpha}{2}} = 1{,}96$ هو:

$$P\left(\begin{array}{l} (50 - 40) - (1{,}96) \cdot \sqrt{72{,}17} \sqrt{\frac{1}{10} + \frac{1}{15}} \le (\overline{y}_1 - \overline{y}_2) \\ \le (50 - 40) + (1{,}96) \cdot \sqrt{72{,}17} \sqrt{\frac{1}{10} + \frac{1}{15}} \end{array} \right) = 0{,}95$$

$$P(-10 - 6{,}796 \le (\overline{y}_1 - \overline{y}_2) \le -10 + 6{,}796) = 0{,}95$$

$$P(3{,}204 \le (\overline{y}_1 - \overline{y}_2) \le 16{,}796) = 0{,}95$$

4-4-5: إنشاء مجال ثقة للفرق بين نسبتين في مجتمعين طبيعيين:

لقد وجدنا أن تقدير الفرق بين نسبتين يعطى بدلالة الفرق بين النسبتين في العينتين، أي أن:

$$(R_1 - R_2) = r_1 - r_2$$

وإن تباينه في حالة السحب مع الإعادة أو عندما يكون مجموع حجمي العينتين كبيراً يقدر كمايلي:

$$\sigma^2_{(r_1 - r_2)} = \frac{r_1 q_1}{n_1} + \frac{r_2 q_2}{n_2}$$

وبطريقة مشابهة لما سبق نجد أن مجال الثقة المعياري للفرق $(R_1 - R_2)$ يعطى بالعلاقة:

$$P\left(-Z_{1-\frac{\alpha}{2}} \le \frac{(r_1 - r_2) - (R_1 - R_2)}{\sqrt{\frac{r_1 q_1}{n_1} + \frac{r_2 q_2}{n_2}}} \le Z_{1-\frac{\alpha}{2}}) \right) = 1 - \alpha \qquad (45-4)$$

وبإجراء بعض الإصلاحات على العلاقة السابقة نحصل على مجال الثقة العادي التالي:

$$P\left(\begin{array}{l}(r_1-r_2)-Z_{1-\frac{\alpha}{2}}\cdot\sqrt{\frac{r_1q_1}{n_1}+\frac{r_2q_2}{n_2}}\le(R_1-R_2)\le\\(r_1-r_2)+Z_{1-\frac{\alpha}{2}}\cdot\sqrt{\frac{r_1q_1}{n_1}+\frac{r_2q_2}{n_2}})\end{array}\right)=1-\alpha \qquad (46-4)$$

أما عندما يكون التباينان في المجتمعين متساويين $\left(\sigma_1^2=\sigma_2^2=\sigma^2\right)$ فإن مجال الثقة السابق يأخذ الشكل التالي:

$$P\left(\begin{array}{l}(r_1-r_2)-Z_{1-\frac{\alpha}{2}}\cdot S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\le(R_1-R_2)\le\\(r_1-r_2)+Z_{1-\frac{\alpha}{2}}\cdot S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}})\end{array}\right)=1-\alpha \qquad (47-4)$$

حيث أن S يحسب من العلاقة:

$$S=\sqrt{\frac{n_1r_1q_1+n_2r_2q_2}{n_1+n_2-2}}$$

مثال:

لدراسة نسبتي العطب في إنتاج معملين للمصابيح الكهربائية. سحبنا عينتين بحجمين $n_1=30$, $n_2=40$ فوجدنا أن نسبة العطب فيها كانت كمايلي: $r_1=0,07, r_2=0,09$ والمطلوب إنشاء مجال الثقة للفرق $\left(R_1-R_2\right)$ باحتمال ثقة $0,95$.

الحل:

نقوم بتطبيق العلاقة (4–46) السابقة مباشرة مع ملاحظة أن $Z_{1-\frac{\alpha}{2}}=1,96$ فنجد أن:

$$P\left(\begin{array}{l}(0,07-0,09)-(1,96)\cdot\sqrt{\frac{(0,07)(0,93)}{30}+\frac{(0,09)(0,91)}{40}}\le(R_1-R_2)\\ \le(0,07-0,09)+(1,96)\cdot\sqrt{\frac{(0,07)(0,93)}{30}+\frac{(0,09)(0,91)}{40}}\end{array}\right)=0,95$$

$$P\left[(-0,02)-(1,96)(0,0649)\le(R_1-R_2)\le(-0,02)-(1,96)(0,0649)\right]=0,95$$

وبذلك نحصل على المجال المطلوب التالي:

$$P\left[-0,127\le(R_1-R_2)\le 0,087\right]=0,95$$

4-5: تقدير حجم العينة في مجتمع طبيعي:

إن حجم العينة اللازم سحبه من المجتمع مرتبط بمقدار الدقة والثقة التي نرغب في تحقيقها.

وبما أن الدقة هي الحد الأعلى للخطأ الذي يمكن أن يرتكب في عمليات التقدير، لذلك فإن الخطأ المرتكب في عمليات التقدير يجب أن يكون أقل من أو يساوي تلك الدقة.

فإذا رمزنا للدقة المطلوبة لـ d ولاحظنا أن الخطأ المرتكب في تقدير المؤشر θ هو المقدار $Z_{1-\frac{\alpha}{2}}\widetilde{\sigma}_{\widetilde{\theta}}$ أو $\left(t_{1-\frac{\alpha}{2}}\widetilde{\sigma}_{\widetilde{\theta}}\right)$ الوارد في المجال المعرف بالعلاقتين (4-30) و(4-32)، فإننا نجد أن d يجب أن تحقق المتراجحة التالية:

$$Z_{1-\frac{\alpha}{2}}\widetilde{\sigma}_{\widetilde{\theta}}\le d \qquad (48-4)$$

واعتماداً على هذه المتراجحة سنستخرج حجم العينة n اللازم لتحقيق الدقة d في كل من حالات تقدير المتوسط $\bar{y}$ والنسبة في المجتمع.

1-5-4: تقدير حجم العينة لتقدير المتوسط:

عندما نريد تقدير متوسط المجتمع $\bar{y}$ بدقة قدرها d وباحتمال ثقة قدره β، فإننا نجد اعتماداً على العلاقة (4-48) أن:

$$Z_{1-\frac{\alpha}{2}}\widetilde{\sigma}_{\bar{x}}\le d$$

وبتربيع الطرفين نجد أن:

$$Z^2_{1-\frac{\alpha}{2}} \widetilde{\sigma}^2_{\overline{x}} \le d^2 \qquad (49-4)$$

وبما أن قيمة $\widetilde{\sigma}^2_{\overline{X}}$ ترتبط بحالة السحب فإننا سنميز بين حالتي السحب فنجد أنه في:

1 . حالة السحب مع الإعادة حيث يكون لدينا:

$$\widetilde{\sigma}^2_{\overline{X}} = \frac{s^2}{n}$$

وبالتعويض في المتراجحة السابقة نحصل على:

$$Z^2_{1-\frac{\alpha}{2}} \cdot \widetilde{\sigma} \cdot \frac{s^2}{n} \le d^2$$

ومنها نجد أن حجم العينة n يجب أن يحقق في هذه الحالة المتراجحة التالية:

$$n \ge \frac{Z^2_{1-\frac{\alpha}{2}} \cdot s^2}{d^2} \qquad (50-4)$$

ومنها يمكننا حساب الحد الأدنى لحجم العينة من العلاقة:

$$n_0 = \frac{Z^2 \cdot s^2}{d^2} \qquad (51-4)$$

2 . حالة السحب بدون إعادة: في هذه الحالة يكون لدينا:

$$\widetilde{\sigma}^2_{\overline{X}} = \frac{N-n}{N} \cdot \frac{s^2}{n}$$

وبالتعويض في (4-49) نحصل على أن:

$$Z^2 \cdot \frac{N-n}{N} \cdot \frac{s^2}{n} \le d^2$$

وبإصلاح هذه المتراجحة نجد أن حجم العينة n في حالة السحب بدون إعادة يحقق العلاقة التالية:

$$n \geq \frac{N \cdot Z^2 \cdot s^2}{Nd^2 + Z^2 s^2} \qquad (52-4)$$

ومنها يمكننا حساب الحد الأدنى لحجم العينة في حالة السحب بدون إعادة من العلاقة:

$$n_0 = \frac{N \cdot Z^2 \cdot s^2}{Nd^2 + Z^2 \cdot s^2} \qquad (53-4)$$

وبذلك نكون قد توصلنا إلى علاقتين تعطياننا حجم العينة n في كلتا حالتي السحب. إلا أن s^2 الداخل فيها يحسب من تباين أي عينة اختبارية تسحب على سبيل التجربة.

ملاحظة:

يمكن استخدام العلاقتين (4−51) و(4−53) عند تقدير حجم العينة لإجمالي المجتمع Y ، مع الانتباه إلى ضرب دقة الإجمالي بـ N.

مثال 13:

أوجد حجم العينة اللازم لتقدير متوسط دخل الأسرة بدقة $d = 200$ ل.س. وباحتمال قدره 0,95 ، علماً بأن عينة اختبارية أعطتنا $s^2 = 300000$ وأن السحب سيجري مع الإعادة.

الحل:

لحساب حجم العينة n نطبق من العلاقة (4−51) فنجد أن قيمة Z المقابلة للاحتمال $1 - \frac{\alpha}{2} = 0,975$ تساوي 1,96 هكذا نجد أن:

$$n \geq \frac{Z^2 \cdot s^2}{d^2} = \frac{(1,96)^2 (300000)}{(200)^2} = 28,8 \approx 29$$

أي أنه يكفي أن تكون $n = 29$ لنحقق الدقة المطلوبة وباحتمال الثقة المطلوب.

مثال 14:

أوجد حجم العينة اللازم لتقدير متوسط استهلاك الأسرة من المازوت في الشهر وبدقة $d = 20$ لتـراً، وباحتمـال قـدره 0,90 ، علمـاً بـأن عينـة اختباريـة أعطتنـا أن: $s^2 = 16000$ ، وأن السحب سيجري بدون إعادة، وأن حجم المجتمع المدروس $N = 1000$ أسرة.

الحل:

بما أن السحب يجرى بدون إعادة فإننا نطبق العلاقة (4−53)، ومن الجدول رقم (III) نجد أن قيمة Z المقابلة للاحتمال $1 - \frac{\alpha}{2} = 0,95$ تساوي $Z = 1,65$ ، وبذلك نجد أن:

$$n \geq \frac{N \cdot Z^2 \cdot s^2}{N \cdot d^2 + Z^2 s^2} = \frac{1000(1,65)^2(16000)}{100(20)^2 + (1,65)^2(16000)} = 98,2 \approx 99`$$

أي أنه يكفي أن يكون $n = 99$ أسرة لنحقق الدقة المطلوبة وبالاحتمال المطلوب.

ملاحظة:

إن حجم العينة الذي نحصل عليه هو الحد الأدنى لحجم العينة n، ويمكن أن يكون حجمها أكبر من ذلك.

ونشير هنا إلى أن حجم العينة لابد وأن يكون عدداً صحيحاً. لذلك يجب تقريبه إلى الأعلى دائماً.

4−5−2: تقدير حجم العينة لتقدير النسبة:

عندما نريد تقدير نسبة خاصة ما في مجتمع R وبدقة δ وباحتمال ثقة قدره $(1-\alpha)$، فإننا نجد اعتماداً على العلاقة (4−48)، إنه إذا كانت δ هي الدقة المطلوبة فإن:

$$Z_{1-\alpha/2}\tilde{\sigma}_r \leq \delta$$

$$Z^2_{1-\alpha/2}\tilde{\sigma}^2_r \leq \delta^2$$

وبما أن $\widetilde{\sigma}_r^2$ ترتبط بحالة السحب فإننا نجد أنه:

1 . في حالة السحب مع الإعادة أن:

$$Z^2_{1-\alpha/2} \cdot \frac{rq}{n} \le \delta^2$$

ومنها نحصل على أن:

$$n \ge \frac{Z^2_{1-\alpha/2} \cdot rq}{\delta^2} \qquad (54-4)$$

2 . في حالة السحب بدون إعادة أن:

$$Z^2_{1-\alpha/2} \cdot \frac{N-n}{N} \cdot \frac{S^2}{n} \le \delta^2 \qquad (55-4)$$

زمنها نحصل على أن:

$$n \ge \frac{NZ^2_{1-\alpha/2} rq}{N\delta^2 + Z^2_{1-\alpha/2} rq} \qquad (56-4)$$

علماً بأن r و $(q = 1-r)$ تحسب بشكل مبدئي من أية عينة اختبارية سريعة.

مثال 15:

أوجد حجم العينة اللازم سحبها لتقدير نسبة المدخنين بدقة $\delta = 0{,}10$. وباحتمال قدره $0{,}95$ ، علماً بأن السحب سيجري مع الإعادة، وأن عينة اختبارية سريعة أعطتنا أن نسبتهم $r = 0{,}4$.

الحل:

من العلاقة (4–54) نجد أن:

$$n \ge \frac{(1{,}96)^2 (0{,}4)(0{,}6)}{(0{,}01)^2} = 256{,}1 \approx 257$$

أي أنه يكفي أن يكون $n = 257$ لنحقق الدقة المطلوبة وبالاحتمال المطلوب.

تمرينات

1 . لنفترض أن مجتمعاً مؤلفاً من العناصر المعلومة التالية:

$$Yi : 12 \ , 14 \ , 16 \ , 18 \ , 15 \ , 17$$

. احسب المتوسط الحسابي لهذه الأرقام $\overline{y}$. ثم احسب مجموعها Y.

. اسحب بشكل عشوائي وبدون إعادة عينة بحجم $n = 3$.

. أوجد تقديراً متوسطاً المجتمع وقارنه مع المتوسط الحقيقي $\overline{y}$.

. أوجد تقديراً لإجمالي وتباين ذلك المجتمع.

. أوجد مجال ثقة لمتوسط المجتمع وباحتمال قدره 0,95 .

. حاول أن تسحب عينة أخرى وبنفس الحجم ثم أوجد مختلف التقديرات السابقة.

2 . سحبنا عينة بدون إعادة وبحجم $n = 3$ من مجتمع مؤلف من 50 أسرة فوجدنا أن دخولها الشهرية تساوي (مئة ل.س).

$$X_i : 12 \ , 10 \ , 14 \ , 18 \ , 17 \ , 11 \ , 19 \ , 20 \ , 13$$

والمطلوب:

. إيجاد تقدير لمتوسط الدخل في المجتمع $\overline{y}$.

. إيجاد تقدير لإجمالي الدخل فيه Y .

. إيجاد تقدير لتباين الدخل فيه σ^2 .

. إيجاد تقدير للتباينين $\sigma_{\overline{x}}^2$ و $\sigma_{\overline{y}}^2$.

. إيجاد تقدير نسبة الدخول التي هي أقل من 15 وتقدير تباينه.

. إيجاد مجال الثقة المقابل للاحتمال 0,95 لكل من المتوسط والإجمالي والنسبة.

2 . لنفترض أن مجتمعاً مؤلفاً من 500 عنصراً، ونريد سحب عينة مع الإعادة لتقدير متوسطه بدقة $d = 5$ ، وباحتمال قدره 0,95 . فماهو حجم العينة اللازم تحديده لتأمين الدقة السابقة وبالاحتمال المفروض؟، علماً بأن عينة اختبارية أعطتنا أن تباينها يساوي: $s^2 = 250$.

. احسب حجم العينة إذا كان السحب بدون إعادة، وافترض أرقاماً لقيم عناصر العينة ثم قم بتقدير متوسط وإجمالي المجتمع، وأوجد مجال ثقة لكل منها، باحتمال قدره 0,990.

4 . برهن نظرياً وعملياً على أن حجم العينة اللازم لتحقيق نفس الدقة وبنفس الاحتمال في حالة السحب مع الإعادة أكبر منه في حالة السحب بدون إعادة، وعلل سبب ذلك.

5 . أوجد مجال الثقة للفرق بين نسبتين $(R_1 - R_2)$.

6 . أوجد تقديراً للفرق بين إجمالي مجتمعين وتباينه ومجال ثقة له.

7 . سحبنا من بين الطالبات عينة بحجم $n = 15$ طالبة، ومن بين الطلاب عينة بحجم $n = 20$ طالباً، ودرسنا متوسط نفقاتهم في الشهر فوجدنا أن متوسط نفقات الطالبة في العينة $x_1 = 1200$ ل.س ، وأن متوسط نفقات الطالب في العينة $x_1 = 1000$ ل.س ، وأن تباين النفقات في عينة الطالبات كان يساوي $s_1^2 = 600$ ل.س ، وأن تباين النفقات في عينة الطلاب كان يساوي $s_1^2 = 500$ ل.س ، وأن نسبة ماتنفقه الطالبات في العينة على الكتب كانت تساوي $r_1 = 0,10$ ، بينما كانت نسبة ماينفقه الطلاب في العينة على الكتب كانت تساوي $r_1 = 0,20$ ، والمطلوب: (يعتبر السحب مع الإعادة).

. تقدير الفرق بين متوسطي النفقات للطالبات والطلاب ثم إيجاد مجال ثقة له باحتمال قدره 0,99.

. تقدير الفرق بين نسبتي النفقات للطالبات والطلاب على الكتب ثم إيجاد مجال ثقة له باحتمال قدره 0,95.

الفصل الخامس

اختبار الفرضيات البسيطة بواسطة z و t

1 . تمهيد.

2 . أنواع الفرضيات الإحصائية.

3 . عناصر الاختبار.

4 . اختبارات حول متوسط المجتمع.

5 . اختبارات حول إجمالي المجتمع.

6 . اختبارات حول النسبة في المجتمع.

7 . اختبارات حول الفرق بين متوسطي مجتمعين طبيعيين.

8 . اختبارات حول الفرق بين نسبتين في مجتمعين طبيعيين.

9 . اختبار القياسات الشاذة.

10 . تمرينات.

الفصل الخامس

اختبار الفرضيات البسيطة بواسطة z و t

5-1: تمهيد:

يتناول اختبار الفرضيات مسألة اتخاذ قرار احتمالي حول قبول أو رفض تلك الفرضيات. أما الفرضيات فقد تتناول أي مؤشر من مؤشرات المجتمع (كالمتوسط والإجمالي والنسبة ..إلخ) ، وذلك بافتراض أنه يساوي قيمة معينة معروفة مسبقاً، أو تتناول نوع التوزيعات التكرارية بافتراض أن لها شكلاً منتظماً أو طبيعياً ...إلخ.

فمثلاً، قد يدّعي أحدهم أن متوسط استهلاك الأسرة من الماء يساوي 50 لتراً يومياً، أو متوسط الدخل في بلده يساوي 5000 ل.س. شهرياً، أو أن عدد أيام العمل في شركته يساوي 150 يوماً في السنة، أو أن نسبة المتعلمين في مدينته تساوي 60%....إلخ.

وعندما نسمع أو نقرأ هذه الادعاءات لا نستطيع التسليم بها، لذلك فإننا نعتبر هذه الادعاءات افتراضات حتى نتحقق من صحتها، وبناءً على ذلك التحقق يتم قبولها أو رفضها.

وللتحقق من صحة أية فرضية، نقوم بسحب عينة عشوائية من عناصر المجتمع المدروس، ثم نقوم بإجراء الحسابات اللازمة لتقدير قيمة المؤشر الذي تتناوله الفرضية، ثم مقارنة القيمة التقديرية التي حصلنا عليها من العينة مع القيمة المفترضة في الفرضية. فإذا كان الفرق بين القيمة التقديرية والقيمة المفترضة قليلاً ومقبولاً من الناحية الإحصائية، فإننا نقبل الفرضية الموضوعة. وإذا لم يكن ذلك الفرق مقبولاً فإننا نرفض الفرضية الموضوعة، ولكن متى يكون الفرق بين القيمتين التقديرية والمفترضة مقبولاً من الناحية الإحصائية؟

نذكّر هنا بأن الفرق بين القيمتين يكون مقبولاً إذا كان مقداره لا يتجاوز نصف طول مجال الثقة الذي يمكن إنشاؤه للمؤشر المدروس.

وبتعبير آخر، لاختبار أية فرضية حول أي مؤشر من مؤشرات المجتمع، نقوم بسحب عينة عشوائية من عناصره، ثم نقوم بإجراء الحسابات اللازمة لتقدير قيمة ذلك المؤشر ولتقدير تباين تقديره، وبعدها ننشئ مجال الثقة له المقابل لاحتمال معين. فإذا كانت القيمة المفترضة تنتمي إلى مجال الثقة السابق، فإننا نعتبر الفرق مقبولاً من الناحية الإحصائية ونقبل الفرضية الموضوعة، وإلا فإننا نرفضها. وهكذا نجد أن اختبار الفرضيات يعتمد اعتماداً كلياً على العينات وعلى مجالات الثقة، وبالتحديد يعتمد على العلاقتين (4–29) و(4–31) الواردتين في الفصل السابق.

لنفترض أننا نريد اختبار فرضية حول المؤشر θ في المجتمع، فإذا كانت الفرضية تقول بأن $\theta = \theta_0$ ، وكان $\widetilde{\theta}$ هو تقدير θ من العينة، وكان $\widetilde{\sigma}_{\widetilde{\theta}}$ الانحراف المعياري للتقدير $\widetilde{\theta}$، فإن عملية الاختبار تقتضي أن تقوم بحساب قيمة المتحول t من العلاقة (4–26) التالية:

$$t = \frac{\widetilde{\theta} - \theta_0}{\widetilde{\sigma}_{\widetilde{\theta}}} \qquad (1-5)$$

ثم مقارنة تلك القيمة مع مجال الثقة المعرف بأحد العلاقتين (4–29) أو (4–31)، وهنا سنميز بين حالتين هما:

1 . عندما يكون حجم العينة n كبيراً: فإن المقدار t المعرف بالعلاقة (5–1) يكون خاضعاً للتوزيع الطبيعي المعياري $N(0\,,1)$. لذلك نقارن قيمة t مع قيمة Z الحرجة. فإذا كانت $|t| \le Z_{1-\frac{\alpha}{2}}$، فهذا يعني أن θ_0 تقع في المجال المحدد في العلاقة (4–29)، وأن الفرضية الموضوعة تكون مقبولة من الناحية الإحصائية، حيث أن $Z_{1-\frac{\alpha}{2}}$ هي قيمة المتحول الطبيعي المعياري المقابل للاحتمال $\left(1-\frac{\alpha}{2}\right)$. وتؤخذ من جدول التوزيع الطبيعي في الجدول (III) الملحق في آخر الكتاب.

وإذا كانت $|t| > Z_{1-\frac{\alpha}{2}}$ فإن هذا يعني أن θ_0 تقع خارج المجال المحدد في العلاقة (4–29) ولذلك فإننا نرفض الفرضية الموضوعة.

يسمى المقدار t المعرف بالعلاقة (5-1) بمؤشر الاختبار ويطبق على كثير من المؤشرات الإحصائية للمجتمع.

2 . أما عندما يكون حجم العينة n صغيراً: فإن المتحول t المعرف بالعلاقة (5-1) يكون خاضعاً لتوزيع ستودينت ذي $(n-1)$ درجة حرية. لذلك نقارن قيمة t مع قيمة متحول ستودينت الحرجة t، المقابلة للاحتمال $\left(1-\frac{\alpha}{2}\right)$ و لدرجة حرية $(n-1)$.فإذا كانت $|t| < t_{(1-\alpha/2\,,\,n-1)}$، فهذا يعني أن θ_0 تقع في مجال الثقة المحدد بالعلاقة (4-31)، وأن الفرضية الموضوعة مقبولة من الناحية الإحصائية، أما إذا كانت $\geq t_{(1-\alpha/2\,,\,n-1)}$ فإننا نعتبر الفرضية الموضوعة مرفوضة، حيث إن الكمية $t_{(1-\alpha/2\,,\,n-1)}$ ، هي قيمة متحول ستودينت المقابلة للاحتمال $\left(1-\frac{\alpha}{2}\right)$ و لدرجة حرية $(n-1)$.

وتؤخذ من جدول تابع توزيع ستودينت في الجدول رقم (V) الملحق في آخر الكتاب.

5-2- أنواع الفرضيات الإحصائية:

يمكن تقسيم الفرضيات الإحصائية إلى الأنواع التالية:

. فرضيات تتناول مؤشرات المجتمع مثل المتوسط أو الإجمالي أو النسبة أو التباين...إلخ.

. فرضيات تتناول شكل أو نوع التوزيع التكراري، إذ يمكن أن يفترض أن له الشكل المنتظم أو الطبيعي...إلخ.

كما يمكن تقسيم الفرضيات الإحصائية إلى فرضيات بسيطة ومعقدة:

. الفرضيات البسيطة: وهي الفرضيات التي تتناول جميع مؤشرات المجتمع الواردة في قانون التوزيع الاحتمالي.

أما عندما نتناول أحدها فإنها تفترض أن المؤشرات الأخرى معلومة أو مقدرة، وكمثال على ذلك نأخذ الفرضية التي تتناول متوسط متحول X خاضع للتوزيع الطبيعي،

$N(\bar{x}, \sigma^2)$ ، والتي تقول أن $\bar{y} = \bar{y}_0$، فإنها تفترض أن التباين σ^2 معلوم أو مقدر، وغير خاضع للاختبار.

. الفرضيات المعقدة: وهي الفضيات التي تتناول بعض مؤشرات المجتمع وتترك المؤشرات الأخرى دون تحديد.

5-3- عناصر الاختبار:

لاختيار أية فرضية حول مؤشر من مؤشرات المجتمع لابد لنا من أن نقوم بتحديد العناصر التالية:

1 . مستوى الدلالة α: وهو عبارة عن الاحتمال المتمم لاحتمال الثقة β، وهو يعبر عن احتمال اتخاذ قرار خاطئ حول فرضية صحيحة.

2 . الفرضية الابتدائية (أو فرضية العدم): وهي الفرضية التي يجري اختبارها، وتتضمن قيمة افتراضية محددة لمؤشر المجتمع المدروس، ونرمز لها بالرمز H_0، كأن نضع الفرصة الابتدائية حول المؤشر θ في المجتمع كما يلي: $\theta = \theta_0$، ونكتب ذلك على الشكل التالي:

$$H_0 : \theta = \theta_0$$

أي إننا نفترض (أو ندعي) بأن قيمة المؤشر θ في المجتمع تساوي θ_0؛ أي نفترض عدم وجود فرق بين قيمته الحقيقية والقيمة المفترضة θ.

أو كأن نضع الفرضية التالية حول متوسط المجتمع:

$$H_0 : \bar{y} = 300$$

أي إننا نفترض أن متوسط المجتمع يساوي 300؛ أي نفترض عدم وجود فرق بين متوسط المجتمع $\bar{y}$ والقيمة المفترض 300.

ملاحظة:

يجب وضع فرضية العدم على شكل مساواة للاستفادة منها في حساب قيمة مؤشر الاختبار.

3 ـالفرضية البديلة: وهي فرضية مخالفة للفرضية الابتدائية، نرمز لها بالرمز H_1، ويمكن لهذه الفرضية أن تأخذ أحد الأشكال التالية:

a . $H_1 : \theta \neq \theta_0$، وهي تعني أن قيمة المتحول θ لا تساوي القيمة المفترضة θ_0 والواردة في الفرضية H_0، وهي تختلف عنها اختلافاً كبيراً بالزيادة أو النقصان لدرجة تجعل الفرضية H_0 غير مقبولة من الناحية الإحصائية.

b . $H_1 : \overline{\theta} > \theta_0$، وهي تعني أن قيمة المتحول θ أكبر من القيمة المفترضة θ_0 بمقدار كبير، بحيث تعد غير مقبولة من الناحية الإحصائية.

c . $H_1 : \overline{\theta} < \theta_0$، وهي تعني أن قيمة المتحول θ أصغر من القيمة المفترضة θ_0 بمقدار كبير، بحيث تعد غير مقبولة من الناحية الإحصائية.

وعند إجراء الاختبار يجب أن نختار أحد هذه الأشكال فقط، وذلك حسبما نريد أن يكون الاختبار ثنائي الجانب (في حالة $H_1 : \theta \neq \theta_0$) أو أحادي الجانب في الحالتين الأخريتين.($H_1 : \overline{\theta} > \theta_0$، أو $H_1 : \overline{\theta} < \theta_0$)

*** مؤشر الاختبار:** وهو علاقة رياضية تتضمن معطيات الفرضية H_0 ومعلومات العينة المسحوبة من المؤشر المدروس. وإن أشهر مؤشرات الاختبار هو المؤشر الذي عرفناه في العلاقة (5−1) التالي:

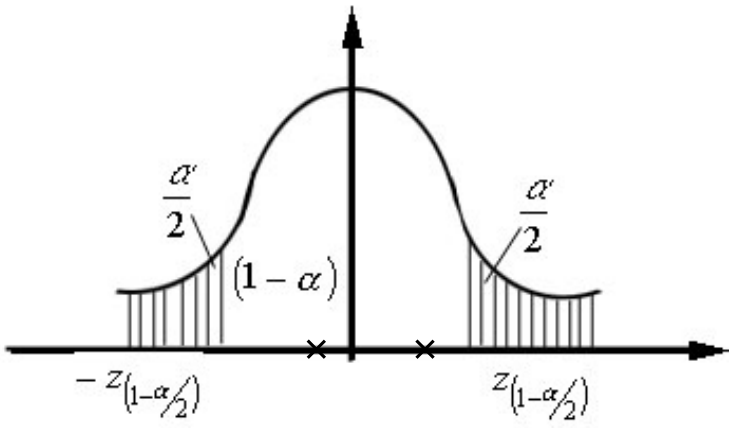
$$t = \frac{\widetilde{\theta} - \theta_0}{\widetilde{\sigma}_{\widetilde{\theta}}} \qquad (2-5)$$

إن هذه العلاقة تصلح لإجراء عدد كبير من الاختبارات الإحصائية، مثل الاختبارات التي توضع حول متوسط المجتمع وإجماليه، وكذلك نسبة خاصة فيه، أو الاختبارات المختلفة حول الفرق بين متوسطي مجتمعين أو بين نسبتين فيهما ...إلخ.

4 . منطقتي القبول والرفض:

تعبر منطقة القبول عن المجال المعرف على المحور الحقيقي OX، الذي يجعل الفرضية الابتدائية (فرضية العدم) مقبولة أو محققة في جميع نقاطه. ويتم تحديد حدود هذا المجال بحيث يكون الاحتمال المقابل له مساوياً لاحتمال الثقة $(1-\alpha)$. ويمكن أن يأخذ ثلاثة أشكال مختلفة وذلك حسب نوعية وشكل الفرضية البديلة H_1 كما يلي:

* إذا كانت الفرضية البديلة H_1 من الشكل $H_1 : \theta \neq \theta_0$، فإن هذا الاحتمال يعني أن منطقة القبول تكون واقعة على طرفي قيمة التقدير $\widetilde{\theta}$. وهي تقابل تماماً مجال الثقة المعياري المقابل للاحتمال $(1-\alpha)$ للمؤشر المدروس θ كما هو موضح على الشكل التالي:



الشكل (5-1) منطقتي القبول والرفض لاختبار ثنائي الجانب

أما منطقة الرفض . في هذه الحالة . فتكون واقعة على طرفي ذلك المجال (على طرفي منطقة القبول)، ويخصص لكل طرفٍ منهما احتمالاً يساوي $\frac{\alpha}{2}$ نصف مستوى الدلالة α، ولذلك يسمى الاختبار في هذه الحالة بالاختبار ثنائي الجانب، (أو ذي الطرفين، أو الذيلين).

ويتم تحديد النقطتين الفاصلتين بين منطقتي الرفض والقبول بواسطة حساب القيمة الحرجة لمتحول تابع التوزيع الطبيعي المعياري $Z_{1-\frac{\alpha}{2}}$ المقابلة للاحتمال التكاملي $\left(1-\frac{\alpha}{2}\right)$. وبعد الحصول على هذه القيمة يمكننا تشكيل المجال الممثل لمنطقة القبول بالمجال التالي:

مجال منطقة القبول $= \left[-Z_{1-\frac{\alpha}{2}} \text{ ، } +Z_{1-\frac{\alpha}{2}} \right]$

كما يمكننا تشكيل المجالين المقابلين لطرفي منطقة الرفض بالمجالين التاليين:

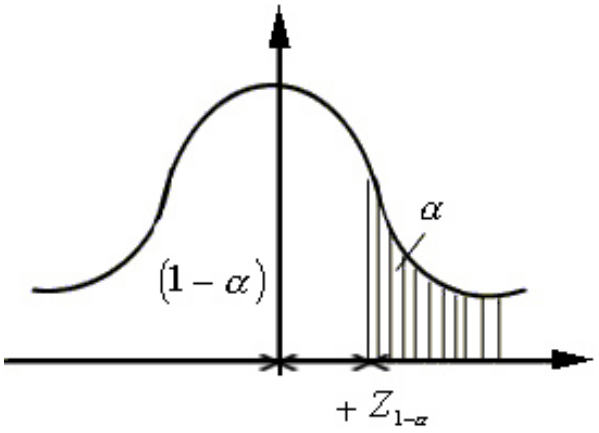
الطرف الأيسر لمنطقة الرفض $= \left[-\infty, -Z_{1-\frac{\alpha}{2}} \right]$

الطرف الأيمن لمنطقة الرفض $= \left[+Z_{1-\frac{\alpha}{2}}, +\infty \right]$

ملاحظة:

عندما يكون حجم العينة صغيراً $(n<30)$ فإن مؤشر الاختبار t يصبح خاضعاً لتوزيع ستودينت ذي $(n-1)$ درجة حرية. لذلك نقوم باستبدال قيمة المتحول $Z_{1-\frac{\alpha}{2}}$ المعياري بالمتحول $t_{1-\frac{\alpha}{2}}$ الخاضع لتوزيع ستودينت. دون إجراء أي تغيير في المناقشة السابقة.

* إذا كانت الفرضية البديلة H_1 من الشكل $H_1 : \theta > \theta_0$، فإن هذا الاحتمال يعني أن منطقة الرفض تقع على جهة واحدة، وهي الجهة الواقعة على يمين منطقة القبول. وهي تقابل كامل مستوى الدلالة α كما هو موضح على الشكل التالي:



الشكل رقم (5-2)

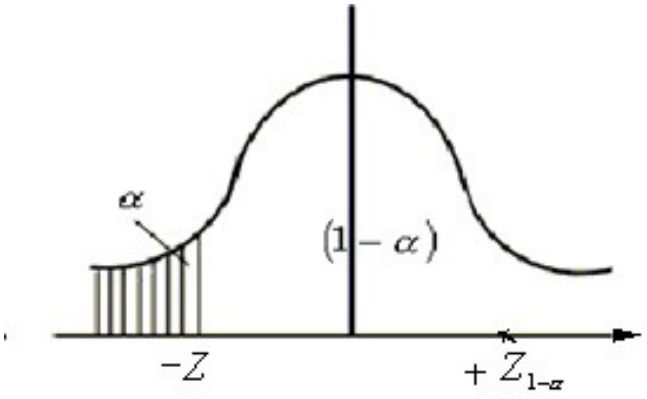
ولتحديد النقطة الفاصلة بين منطقتي القبول والرفض نبحث عن القيمة الحرجة لمتحول تابع التوزيع الطبيعي المعياري $Z_{1-\alpha}$ المقابلة للاحتمال التكاملي $(1-\alpha)$. وبعد ذلك نقوم بتشكيل منطقتي القبول والرفض كما يلي:

منطقة القبول $=]\ -\infty, +Z_{1-\alpha}\ ]$ وتقابل الاحتمال $(1-\alpha)$.

منطقة الرفض $=]\ +Z_{1-\alpha},\ -\infty\ ]$ وتقابل كامل (α).

ولذلك يسمى الاختبار في هذه الحالة باختبار أحادي الجانب يميني.

* إذا كانت الفرضية البديلة H_1 من الشكل $H_1 : \theta < \theta_0$، فإن هذا الاحتمال يعني أن منطقة الرفض تقع في الجهة الواقعة على يسار منطقة القبول. وهي تقابل كامل مستوى الدلالة α كما هو موضح على الشكل التالي:



الشكل رقم (5-3)

ولتحديد النقطة الفاصلة بين منطقتي القبول والرفض نبحث عن القيمة الحرجة لمتحول تابع التوزيع الطبيعي المعياري $Z_{1-\alpha}$ ونضربها بـ (1-) (لأنها موجبة) لنحصل على القيمة $(Z_{1-\alpha})$ والتي تساوي (Z_α) السالبة الموجودة في الجهة اليسرى وغير الواردة في الجداول الملحقة، لأنها تسحب من العلاقة:

$$Z_\alpha = -Z_{1-\alpha}$$

وبعد ذلك نقوم بتشكيل منطقتي القبول والرفض كما يلي:

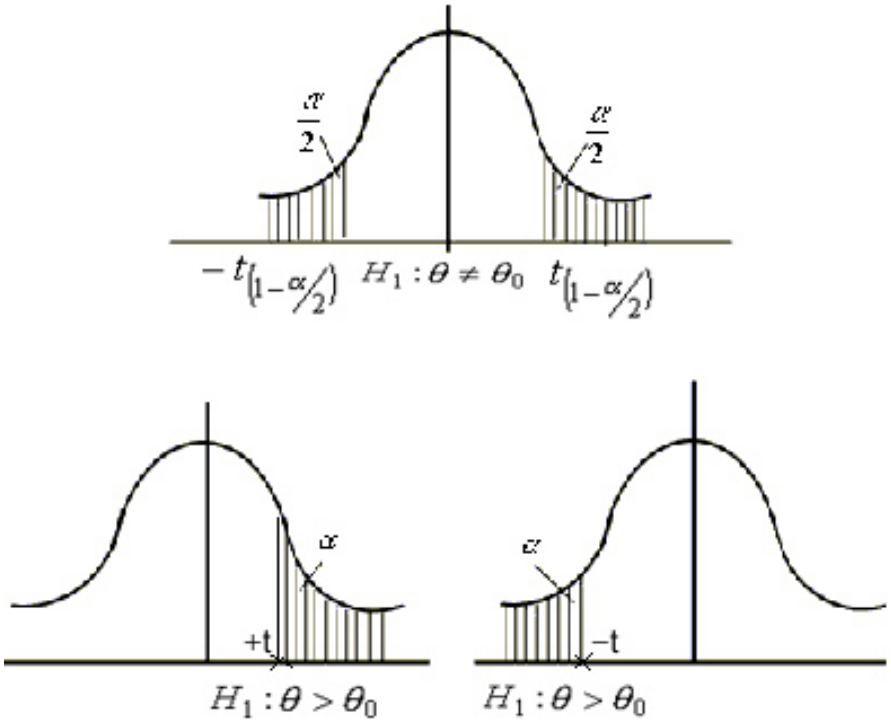
منطقة القبول $=[-Z_{1-\alpha},\ +\infty]$ وتقابل الاحتمال $(1-\alpha)$.

منطقة الرفض $=[-\infty, -Z_{1-\alpha}]$ وتقابل كامل (α).

ولذلك يسمى الاختبار في هذه الحالة باختبار أحادي الجانب يساري.

ملاحظة:

عندما يكون حجم العينة n صغيراً فإننا نستبدل $(Z_{1-\alpha})$ بمتحول ستودينت $(t_{1-\alpha})$ وذي $(n-1)$ درجة حرية وذلك للحالتين الأخيرتين دون أي تغيير آخر.



الشكل رقم (5-4)

5 . القرار الإحصائي:

وهو القرار الذي يجب اتخاذه حول قبول أو رفض فرضية العدم H_0. وحتى نستطيع اتخاذ ذلك القرار نقوم بالإجراءات التالية:

1 . نحسب قيمة مؤشر الاختبار t المعرّف بالعلاقة (5-2) أو بغيرها، وذلك بالاعتماد على مضمون فرضية العدم وعلى المعلومات المتوفرة من العينة عن المتوسط والانحراف المعياري...إلخ.

2 . نحدد نوع الاختبار، هل هو ثنائي الجانب، أم أحادي يميني، أم أحادي يساري.

3 . نبحث عن القيمة الحرجة المناسبة لنوع الاختبار $Z_{1-\frac{\alpha}{2}}$، أو $Z_{1-\alpha}$، أو $-Z_{1-\alpha}$.

4 . نقارن قيمة مؤشر الاختبار t مع القيمة الحرجة المناسبة ونتخذ القرار المناسب كما يلي:

. إذا كان الاختبار ثنائي الجانب، وكانت $|t| < Z_{1-\frac{\alpha}{2}}$ فإننا نقبل فرضية العدم H_0 والعكس بالعكس.

. إذا كان الاختبار أحادي يميني، وكانت $t < Z_{1-\alpha}$ فإننا نقبل فرضية العدم H_0 والعكس بالعكس.

. إذا كان الاختبار أحادي يساري، وكانت $t > -Z_{1-\alpha}$ فإننا نقبل فرضية العدم H_0 والعكس بالعكس.

بما أن أية فرضية يمكن أن تكون صحيحة أو خاطئة، فإن قرارنا حولها قد يكون بقبولها أو رفضها، لذلك فإننا نواجه عدة حالات متشابكة بالنسبة لصحة القرار الذي سنأخذه حول صحة الفرضية الموضوعة. وهذه الحالات معروضة في الجدول التالي:

نتيجة الاختبار / حقيقة الفرضية H_0	قبول الفرضية H_0	رفض الفرضية H_0
صحيحة	قرار صحيح واحتماله $1-\alpha=\beta$	قرار خاطئ واحتماله $\alpha=1-\beta$
غير صحيحة	قرار خاطئ واحتماله γ	قرار صحيح واحتماله $1-\gamma$

6 . خطوات الاختبار:

يمكننا أن نلخص خطوات الاختبار الإحصائي بما يلي:

1 . تحديد مستوى الدلالة α أو الاحتمال β، حيث $\alpha=1-\beta$.

2 . تحديد الفرضيتين الابتدائية والبديلة بالشكل الذي نراه مناسباً.

3 . سحب عينة بحجم n من المجتمع المدروس.

4 . تحديد مؤشر الاختبار المناسب.

5 . حساب قيمة مؤشر الاختبار اعتماداً على الفرضية الابتدائية H_0 ومعلومات العينة.

6 . إيجاد قيمة $Z_{1-\alpha/2}$ (أو $t_{1-\alpha/2}$) المقابلة للاحتمال $1-\frac{\alpha}{2}$ في حالة الاختبار ثنائي الجانب، أو قيمة $Z_{1-\alpha}$ (أو $t_{1-\alpha}$). المقابلة للاحتمال $1-\alpha$ في حالة الاختبار أحادي الجانب.

7 . إذا كان الاختبار ثنائي الجانب نقارن قيمة مؤشر الاختبار t مع قيمة $Z_{1-\alpha/2}$ (أو $t_{1-\alpha/2}$). فإذا كان $|t| \le Z_{1-\alpha/2}$ فإننا نقبل الفرضية H_0. وإذا كان $|t| \ge Z_{1-\alpha/2}$ فإننا نرفض الفرضية H_0.

أما إذا كان الاختبار أحادي يميني، فإننا نقبل الفرضية H_0 إذا كان كان $t < Z_{1-\alpha}$ والعكس بالعكس.

وأما إذا كان الاختبار أحادي يساري فإننا نقبل الفرضية H_0 ، إذا كانت $t > -Z_{1-\alpha}$ والعكس بالعكس.

5-4- اختبارات حول متوسط المجتمع:

عندما يدور الاختبار حول متوسط المجتمع $\bar{y}$، فإن الفرضية الابتدائية تكون من الشكل:

$$H_0 : \bar{y} = \bar{y}_0 \qquad (3-5)$$

وإن الفرضية البديلة يمكن أن تكون من الشكل:

$$H_1 : \bar{y} \neq \bar{y}_0 \qquad (4-5)$$

وعندها يكون الاختبار ثنائي الجانب.

كما يمكن للفرضية البديلة أن تكون من أحد الشكلين التاليين:

$$H_1 : \bar{y} > \bar{y}_0 \quad , \quad H_1 : \bar{y} < \bar{y}_0$$

وعندها يكون الاختبار أحادي الجانب.

وإن قيمة مؤشر الاختبار تحسب من العلاقة التالية:

$$t = \frac{\bar{x} - \bar{y}_0}{\tilde{\sigma}_{\bar{x}}} = \frac{\bar{x} - \bar{y}_0}{s/\sqrt{n}} \qquad (5-5)$$

فعندما يكون حجم العينة n كبيراً وتكون الفرضية H_1 من الشكل $(H_1 = \bar{y} \neq y_0)$ فإننا نقارن قيمة t مع القيمة $Z_{1-\alpha/2}$ المأخوذة من جدول التوزيع الطبيعي المعياري والمقابلة للاحتمال $\left(1-\frac{\alpha}{2}\right)$ ، فإذا كانت $|t| \leq Z_{1-\alpha/2}$، فإننا نقبل الفرضية H_0. ونرفضها في حالة العكس.

أما عندما تكون الفرضية البديلة H_1 من الشكل $(H_1 = \bar{y} > y_0)$ فإن الاختبار يكون أحادي الجانب يميني، لذلك نقارن قيمة t مع القيمة $Z_{1-\alpha}$ المقابلة للاحتمال $(1-\alpha)$. فإذا كانت $t \leq Z_{1-\alpha}$ فإننا نقبل الفرضية H_0. ونرفضها في حالة العكس.

وعندما تكون الفرضية البديلة من الشكل $(H_1 = \bar{y} < y_0)$ فإننا نقارن قيمة t مع القيمة $(-Z_{1-\alpha})$. فإذا كانت $t > -Z_{1-\alpha}$ فإننا نقبل الفرضية H_0. ونرفضها في حالة العكس.

أما إذا كان حجم العينة n صغيراً $(n < 30)$ فإننا نقارن قيمة مؤشر الاختبار مع قيمة متحول ستودينت. فعندما تكون الفرضية البديلة H_1 من الشكل $(H_1 = \bar{y} \neq \bar{y}_0)$، فإن الاختبار يكون ثنائي الجانب، لذلك نقارن قيمة t مع القيمة $t_{1-\alpha/2}$ المقابلة للاحتمال $\left(1-\frac{\alpha}{2}\right)$، ولدرجة حرية $k = n-1$، فإذا كانت $|t| < t_{(1-\alpha/2,\, n-1)}$ فإننا نقبل الفرضية H_0. ونرفضها في حالة العكس. أما عندما تكون الفرضية البديلة H_1 من الشكل $(H_1 = \bar{y} > \bar{y}_0)$، فإننا نقارن قيمة t مع القيمة $t_{1-\alpha}$ المقابلة للاحتمال $(1-\alpha)$، ولـ $(n-1)$ درجة حرية في توزيع ستودينت، ونقبل الفرضية الابتدائية إذا كانت $t \leq t_{(1-\alpha/2,\, n-1)}$ ونرفضها في حالة العكس.

أما عندما تكون الفرضية البديلة من الشكل $(H_1 = \bar{y} < \bar{y}_0)$ فإننا نقارن قيمة t مع القيمة $(-t_{1-\alpha})$. فإذا كانت $t > -t_{1-\alpha}$ فإننا نقبل الفرضية H_0. ونرفضها في حالة العكس.

مثال 1:

قال أحدهم: إن متوسط دخل الأسرة في مدينته يساوي 5000 ل.س. وللتحقق من ذلك سحبنا عينة عشوائية من 35 أسرة فوجدنا أن متوسط دخولهم كان 4660 ل.س. وأن تباينها كان يساوي $s^2 = 1456000$. اختبر صحة القول السابق باحتمال قدره $\beta = 0,95$ (أو بمستوى دلالة $\alpha = 0,05$).

الحل:

لإجراء الاختبار المطلوب نضع الفرضيتين الابتدائية والبديلة كما يلي:

$$H_0 : \bar{y} = 5000 = \bar{y}_0$$
$$H_1 : \bar{y} \neq 5000$$

نلاحظ أن حجم العينة n كبير نسبياً، وأن $\alpha = 0,05$، ثم نقوم بحساب قيمة مؤشر الاختبار التالي:

$$t = \frac{\bar{x} - \bar{y}_0}{\frac{s}{\sqrt{n}}} = \frac{4660 - 5000}{\frac{1456000}{\sqrt{35}}} = -1,667$$

وبما أن الفرضية البديلة هي من الشكل $H_1 : \bar{y} \neq 5000$، فإن الاختبار ثنائي الجانب، لذلك نبحث عن قيمة $Z_{1-\alpha/2}$ المقابلة للاحتمال $\left(1 - \frac{\alpha}{2}\right)$ في التوزيع الطبيعي المعياري فنجد أن:

$$Z_{1-\alpha/2} = Z_{1-0,05/2} = Z_{0,975} = 1,96$$

وبالمقارنة نجد أن القيمة المطلقة لقيمة المؤشر t أصغر من قيمة $Z_{0,975}$؛ وذلك لأن $|-1,667| < 1,96$؛ أي أن:

$$|t| < Z_{1-\alpha/2}$$

لذلك نقبل الفرضية H_0 والقائلة بأن متوسط الدخل يساوي 5000 ل.س تقريباً.

مثال 2:

يقال أن متوسط سعر مادةٍ ما خلال العام كان 12 ل.س. وعندما سحبنا عينة بحجم $n = 10$ قياساً لسعرها بينت أنه كان كما يلي:

13 , 13 , 13 , 12 , 13 , 14 , 15 , 13 , 12 , 14

اختبر صحة ذلك القول باحتمال قدره $\beta = 0,99$ (أو بمستوى دلالة $\alpha = 0,01$).

الحل:

لإجراء الاختبار المطلوب نضع الفرضيتين كما يلي:

$$H_0 : \bar{y} = 12$$
$$H_1 : \bar{y} > 12$$

نحسب قيمة مؤشر الاختبار من العلاقة التالية:

$$t = \frac{\bar{x} - \bar{y}_0}{\frac{s}{\sqrt{n}}}$$

لذلك نقوم بحساب $\bar{x}$ و s فنجد أن:

$$\bar{x} = \frac{1}{n}\sum^{n} xi = 13,2$$

$$s^2 = \frac{1}{n-1}\sum^{n}\left(xi - \bar{x}\right)^2 = 0,844$$

ومنه نجد أن:

$$s = \sqrt{0,844} = 0,919$$

وبذلك نجد أن قيمة مؤشر الاختبار t تساوي:

$$t = \frac{13,2 - 12}{\frac{0,919}{\sqrt{10}}} = 4$$

وبملاحظة أن حجم العينة صغير وأن شكل الفرضية البديلة H_1، يعني أن الاختبار أحادي الجانب يميني، إذن نبحث عن قيمة $t_{1-\alpha}$ المقابلة للاحتمال $(1-\alpha)$ ولـ $(n-1)$ درجة حرية في توزيع ستودينت فنجد أن:

$$t_{(0,99\ ,\ 9)} = 2,821$$

وبالمقارنة نجد أن القيمة الحقيقية لـ t أكبر من القيمة الجدولية $t_{1-\alpha}$.

أي أن:

$$t > t_{1-\alpha} \Leftarrow 3{,}5 > 2{,}821$$

لذلك نرفض الفرضية H_0 والقائلة بأن متوسط السعر يساوي 12 ل.س ونقبل الفرضية البديلة التي تقول إن متوسط السعر كان أكبر من 12 ل.س.

5-5- اختبارات حول إجمالي المجتمع:

عندما نريد إجراء اختبار حول إجمالي مجتمع ما، فإن الفرضية الابتدائية تكون من الشكل التالي:

$$H_0 : Y = Y_0$$

أما الفرضية البديلة فيمكن أن تكون من أحد الأشكال التالي

اختبار ثنائي الجانب $\quad H_1 : Y \neq Y_0$

$$\left.\begin{array}{l} H_1 : Y > Y_0 \\ H_1 : Y < Y_0 \end{array}\right\} \quad (6\text{-}5)$$

اختبار أحادي الجانب

وإن قيمة مؤشر الاختبار تحسب من العلاقة:

$$t = \frac{\widetilde{Y} - Y_0}{N \cdot \dfrac{s}{\sqrt{n}}} \qquad (7-5)$$

ثم نقوم بمقارنة قيمة t مع القيمة $Z_{1-\alpha/2}$ (أو $t_{1-\alpha/2}$)، إذا كان الاختبار ثنائي الجانب، أو مع القيمة $Z_{1-\alpha}$ (أو $t_{1-\alpha}$)، إذا كان الاختبار أحادي الجانب.

وعندما يكون الاختبار ثنائي وتكون نتيجة الاختبار $|t| \le Z_{1-\alpha/2}$ أو $|t| \le t_{1-\alpha/2}$ فإننا نقبل الفرضية H_0 ونرفضها في حالة العكس.

أما عندما يكون الاختبار أحادي يميني، وتكون القيمة الحقيقية t محققة للمتراجحة $t \le Z_{1-\alpha}$ فإننا نقبل H_0 ونرفضها في حالة العكس.

أما عندما يكون الاختبار أحادي يساري، وتكون القيمة الحقيقية t محققة للمتراجحة $t \geq -Z_{1-\alpha}$ فإننا نقبل H_0 ونرفضها في حالة العكس.

مثال 3:

صرح أحد التجار لمديرية المالية بأن إجمالي ربحه الشهري يساوي 20000 ل.س. وللتأكد من ذلك قام موظفو المالية بتدقيق دفاتر مبيعاته خلال 15 يوماً اختيرت بشكل عشوائي، فوجدوا أن متوسط ربحه اليومي يساوي $\bar{x} = 700$ وأن تباينه يساوي $s^2 = 14650$. اختبر صحة التصريح باحتمال قدره 0,99.

الحل:

قبل كل شيء نضع الفرضيتين التاليتين:

$$H_0 : Y = Y_0 = 20000$$
$$H_1 : Y > 20000$$

لقد وضعنا $H_1 : Y > 20000$ لأن موظفي المالية لا يهتمون فيما إذا كان ربحه اقل من 20000، بل فيما إذا كان ربحه أكبر من 20000 وذلك من أجل جباية الضرائب.

ثم نحسب قيمة مؤشر الاختبار من العلاقة:

$$t = \frac{\widetilde{Y} - Y_0}{N \cdot \frac{s}{\sqrt{n}}} = \frac{N\bar{x} - Y_0}{N \cdot \frac{s}{\sqrt{n}}} = \frac{30 \cdot 700 - 20000}{30 \cdot \sqrt{\frac{14650}{15}}}$$

$$t = 1{,}067$$

ومن جدول قيم توزيع ستودينت نجد أن القيمة الجدولية $t_{1-\alpha}$ تساوي:

$$t_{(0{,}99\,,\,14)} = 1{,}345$$

وبالمقارنة نجد أن القيمة الحقيقية لـ t بدون قيمة مطلقة أصغر من القيمة الجدولية $t_{1-\alpha}$؛ (لأن $1{,}067 < 1{,}345$)، أي أن:

$$t < t_{1-\alpha}$$

لذلك نقبل الفرضية H_0 ونعتبر التاجر صادقاً.

5-6- اختبارات حول النسبة في المجتمع:

إذا كان هدفنا إجراء اختبار حول النسبة R فإننا نضع الفرضية الابتدائية:

$$H_1 : R = R_0 \qquad (8-5)$$

ونضع إحدى الفرضيات البديلة التالية:

$$H_1 : R \neq R_0 \quad , \quad H_1 : R < R_0 \quad , \quad H_1 : R > R_0 \qquad (9-5)$$

ثم نقوم بحساب قيمة مؤشر الاختبار من العلاقة:

$$t = \frac{\widetilde{R} - R_0}{\sqrt{\frac{rq}{n}}} = \frac{r - R_0}{\sqrt{\frac{rq}{n}}} \qquad (10-5)$$

فإذا كان حجم العينة n كبيراً، وكانت الفرضية H_1 من الشكل $H_1 : R \neq R_0$ فإننا نقارن قيمة t مع القيمة $Z_{1-\alpha/2}$ المقابلة للاحتمال $\left(1-\frac{\alpha}{2}\right)$، ولكن إذا كانت H_1 من الشكل $H_1 : R > R_0$، أو من الشكل $H_1 : R < R_0$ فإننا نقارن قيمة t مع القيمة $Z_{1-\alpha}$ المقابلة للاحتمال $(1-\alpha)$ حسب اتجاهه يميني، أو يساري.

أما إذا كان حجم العينة n صغيراً، وكانت الفرضية H_1 من الشكل $H_1 : R \neq R_0$ فإننا نقارن قيمة t مع $t_{1-\alpha/2}$ المقابلة للاحتمال $\left(1-\frac{\alpha}{2}\right)$ ولـ $(n-1)$ درجة حرية في توزيع ستودينت. ولكن إذا كانت الفرضية H_1 من الشكل $H_1 : R > R_0$، أو من الشكل $H_1 : R < R_0$ فإننا نقارن قيمة t مع القيمة $t_{1-\alpha}$ المقابلة للاحتمال $(1-\alpha)$. ولـ$(n-1)$ درجة حرية في توزيع ستودينت وحسب اتجاهه يميني، أو يساري.

مثال 4:

نشرت إحدى الصحف أن نسبة العاطلين عن العمل تساوي 14%. وللتحقق من هذه النسبة سحبت عينة عشوائية بحجم $n=100$ أسرة فكانت نسبة العاطلين عن العمل فيها 20%. اختبر صحة ما نشرته تلك الصحيفة باحتمال قدره 0,95.

الحل:

نضع الفرضيتين التاليتين:

$$H_0 : R = R_0 = 0{,}14$$
$$H_1 : R > R_0$$

ثم نحسب قيمة مؤشر الاختبار من العلاقة:

$$t = \frac{r - R_0}{\sqrt{\frac{rq}{n}}} = \frac{0{,}20 - 0{,}14}{\sqrt{\frac{(0{,}20)(0{,}80)}{100}}} = 15$$

وبما أن حجم العينة كبير، فإننا نقارن قيمة t مع $Z_{1-\alpha}$ (ولكون الاختبار أحادي الجانب) ومن جدول التوزيع الطبيعي رقم III نجد أن:

$$Z_{1-0,5} = Z_{0,95} = 1{,}64$$

وبالمقارنة نجد أن القيمة الحقيقية لـ t (وليست المطلقة) أكبر من القيمة الجدولية لـ $t_{1-\alpha}$؛ (لأن $15 > 1{,}64$)، أي أن:

$$t > z_{1-\alpha}$$

لذلك نرفض الفرضية H_0 والقائلة بأن نسبة العاطلين عن العمل تساوي 0,14.

ونقبل الفرضية البديلة والقائلة بأن نسبة العاطلين عن العمل أكبر من 0,14.

5-7- اختبارات حول الفرق بين متوسطي مجتمعين طبيعيين:

عندما يكون لدينا مجتمعان، متوسطاهما $\bar{y}_1$ و $\bar{y}_2$، فإننا كثيراً ما نحتاج إلى مقارنة هذين المتوسطين عن طريق اختبار الفرق بينهما. لذلك نقوم بسحب عينتين مستقلتين n_1 و n_2 على الترتيب. فإذا كان متوسطا الخاصة المدروسة في هاتين العينتين

يساويان $\bar{x}_1$ و$\bar{x}_2$، وكان تباينيهما s_1^2 وs_2^2 على الترتيب. فإن تقدير الفرق $(\bar{y}_1 - \bar{y}_2)$ يعطى بواسطة الفرق $(\bar{x}_1 - \bar{x}_2)$؛ أي أن:

$$(\bar{y}_1 - \bar{y}_2) = (\bar{x}_1 - \bar{x}_2) \qquad (11-5)$$

ونعلم أن تقدير تباين الفرق $(\bar{x}_1 - \bar{x}_2)$ يعطى بالعلاقة (أنظر العلاقة 4-32) الواردة في الفصل السابق.

$$\sigma^2_{(\bar{x}_1 - \bar{x}_2)} = \frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \qquad (12-5)$$ في حالة السحب مع الإعادة

لذلك فإننا نضع الفرضية الابتدائية كما يلي:

$$H_0 : \bar{y}_1 - \bar{y}_2 = (\bar{y}_1 - \bar{y}_2)_0 \qquad (13-5)$$

أما الفرضية البديلة فيمكن أن تكون من الشكل (اختبار ثنائي الجانب):

$$\bar{y}_1 - \bar{y}_2 \neq (\bar{y}_1 - \bar{y}_2)_0 \qquad (14-5)$$

أو من أحد الشكلين التاليين (اختبار أحادي الجانب):

$$\begin{aligned} &\bar{y}_1 - \bar{y}_2 < (\bar{y}_1 - \bar{y}_2)_0 && \text{يساري} \\ &\bar{y}_1 - \bar{y}_2 > (\bar{y}_1 - \bar{y}_2)_0 && \text{يميني} \end{aligned} \qquad (15-5)$$

أما مؤشر الاختبار فيحسب من العلاقة التالية:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\bar{y}_1 - \bar{y}_2)_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \qquad (16-5)$$

فعندما يكون حجم العينة $n_1 + n_2$ كبيراً، فإننا نقارن قيمة t مع $Z_{1-\alpha/2}$ (إذا كان الاختبار ثنائي الجانب)، فإذا كانت $|t| \leq Z_{1-\alpha/2}$ فإننا نقبل الفرضية H_0، وإلا فإننا نرفضها. أما إذا كان الاختيار إحادي يميني فإننا نقبل H_0 إذا كانت $t \leq Z_{1-\alpha}$. وإذا كان يساري فإننا نقبل H_0 إذا كانت $t > -Z_{1-\alpha}$.

أما عندما يكون حجم العينة $n_1 + n_2$ صغيراً، فإن توزيع مؤشر الاختبار السابق (5-16) لم يعرف حتى الآن. إلا أنه إذا كان تباينا المجتمعين متساويين $\left(\sigma_1^2 = \sigma_2^2\right)$ فإننا نعطي حلاً لهذه القضية من خلال الحالة الخاصة التالية:

حالة خاصة 1:

إذا كان تباينا المجتمعين متساويين؛ أي كان $\sigma_1^2 = \sigma_2^2 = \sigma^2$، فإنه يمكننا كتابة تباين الفرق $\left(\bar{x}_1 - \bar{x}_2\right)$ على الشكل التالي:

$$\sigma^2_{(\bar{x}_1 - \bar{x}_2)} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} = \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2} = \sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right) \qquad (17-5)$$

وإن تقدير القيمة المشتركة للتباينين σ^2 يعطى بواسطة الوسط الحسابي المثقل لتبايني العينتين s_1^2 و s_2^2 ومن خلال العلاقة التالية:

$$\tilde{\sigma}^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = s^2 \qquad (18-5)$$

وبذلك نجد أن تقدير $\sigma^2_{(\bar{x}_1 - \bar{x}_2)}$ في هذه الحالة يعطى بالعلاقة:

$$\tilde{\sigma}^2_{(\bar{x}_1 - \bar{x}_2)} = s^2 \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right) \qquad (19-5)$$

ويأخذ مؤشر الاختبار في هذه الحالة الشكل التالي:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\bar{y}_1 - \bar{y}_2)_0}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_1}}} \qquad (20-5)$$

وهو مقدار خاضع لتوزيع ستودينت ذي $(n_1 + n_2 - 2)$ درجة حرية.

لذلك فإننا نقارن قيمة t مع قيمة متحول (ستودينت) $t_{1-\alpha/2}$ (أو $t_{1-\alpha}$) المقابلة للاحتمال $1 - \frac{\alpha}{2}$ (أو $1 - \alpha$) ولـ $(n_1 + n_2 - 2)$ درجة حرية.

فإذا كانت $|t| \le t_{1-\alpha/2}$ ، وكان الاختيار ثنائي الجانب فإننا نقبل الفرضية H_0 الموضوعة،

أما إذا كانت الاختيار أحادي الجانب ويميني فإننا نقبل H_0 إذا كان $t < Z_{1-\alpha}$.
وعندما يكون أحادي يساري فإننا نقبل H_0 عندما يكون $t > -Z_{1-\alpha}$. وهنا نشير إلى أنه إذا كان $\left(\sigma_1^2 = \sigma_2^2\right)$ وكان الحجم $\left(n_1 + n_2\right)$ كبيراً، فإننا نقارن قيمة t مع $Z_{1-\alpha/2}$ (أو $Z_{1-\alpha}$) وهي قيمة متحول التوزيع الطبيعي المقابلة للاحتمال $1-\frac{\alpha}{2}$، أو لـ $(1-\alpha)$ وذلك حسبما يكون الاختبار ثنائي الجانب، أو أحادي الجانب.

حالة خاصة 2:

إذا سحبنا عينتين بحجمين n_1 و n_2 من مجتمع واحد فإنه يكون لدينا حكماً في هذه الحالة ما يلي:

$$\bar{y}_1 = \bar{y}_2 = \bar{y} \quad , \quad \sigma_1^2 = \sigma_2^2 = \sigma^2 \qquad (21-5)$$

ويأخذ المؤشر المعرف بالعلاقة (5−20) الشكل التالي:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_1}}} \qquad (22-5)$$

ويتحول اختبار الفرضية H_0 إلى اختبار جوهرية الفرق بين متوسطي العينتين $(\bar{x}_1)$ و$(\bar{x}_2)$ (وليس بين $\bar{y}_1$ و $\bar{y}_2$).

ولاتخاذ قرار حول جوهرية الفرق بين $(\bar{x}_1)$ و$(\bar{x}_2)$، فإننا نقارن قيمة t مع $Z_{1-\alpha/2}$، $\left(t_{1-\alpha/2}\right)$، فإذا كانت $|t| \le Z_{1-\alpha/2}$ فإننا نعتبر الفرق بين $(\bar{x}_1 - \bar{x}_2)$ غير جوهري وعائد لأسباب عشوائية.

أما إذا كانت $|t| \ge Z_{1-\alpha/2}$ فإننا نعتبر الفرق بين $(\bar{x}_1 - \bar{x}_2)$ فرقاً جوهرياً وعائد لأسباب غير عشوائية يجب البحث عنها وتحديدها.

مثال 5:

لدراسة كيفية أداء أساتذة الرياضيات في مدرستين A و B، سحبنا منهما عينتين بحجمين $n_1 = 40$ و $n_2 = 50$ طالباً، فوجدنا أن قيم متوسطي علامة الطالب فيهما وتباينيهما كانت تساوي $\bar{x}_1 = 60$ و $\bar{x}_2 = 65$. و $s_1^2 = 160$ و $s_2^2 = 250$.

وبناءً على ذلك اختبر فيما إذا كان الفرق بين متوسطي علامة الرياضيات في المدرستين يقدر بـ 8 علامات، وذلك باحتمال قدره 0,98.

الحل:

نضع الفرضية الابتدائية:

$$H_0 : \bar{y}_1 - \bar{y}_2 = 8$$

ونضع الفرضية البديلة على الشكل التالي:

$$H_1 : \bar{y}_1 - \bar{y}_2 < 8 \qquad \text{(الاختبار أحادي يساري)}$$

ثم نقوم بحساب قيمة مؤشر الاختبار من العلاقة (5-16):

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\bar{y}_1 - \bar{y}_2)_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(65 - 60) - 8}{\sqrt{\frac{160}{40} + \frac{250}{50}}} = -1$$

وبما أن الحجم $n_1 + n_2 = 90$ كبير نسبياً، لذلك فإننا نقارن قيمة t مع $Z_{1-\alpha}$ (الاختبار أحادي الجانب يساري) ومن الجدول رقم III نجد أن:

$$Z_{0,95} = 2,06$$

وبالمقارنة نجد أن القيمة الحقيقية لـ t (وليست المطلقة) أكبر من القيمة $-Z_{1-\alpha}$، أي أن $-1 > -2,06$، أي أن $t > -Z_{1-\alpha}$، لذلك نقبل الفرضية H_0 القائلة بأن متوسط علامة الطالب في المدرسة A يزيد عنه في المدرسة B بمقدار 8 علامات (أي أن الفرق بينهما يساوي 8 علامات وليس أقل من ذلك كما ورد في الفرضية H_0).

مثال 6:

سحبنا عينتين من مجتمع واحد (من العمال مثلاً) بحجمين $n_1 = 40$ و $n_2 = 30$ وكان متوسطا الدخل والتباينين فيهما كما يلي:

$$\bar{x}_1 = 1000 \text{ و } \bar{x}_2 = 1100 \text{ . و } s_1^2 = 50000 \text{ و } s_2^2 = 60000$$

فهل هذه المعطيات تعطينا دليلاً كافياً لاعتبار الفرق بين متوسطي الدخل في العينتين فرقاً جوهريا. اختبر باحتمال قدره 0,95

الحل:

بما أن العينتين سحبتا من مجتمع واحد فهذا يعني أنه لا داعي للفرضيات، وأن مؤشر الاختبار يدل على الفرق بين متوسطي العينتين ويحسب من العلاقة (5−22) حيث يكون لدينا:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_1}}} = \frac{1000 - 1100}{\sqrt{\frac{39 \cdot 50000 + 29 \cdot 60000}{40 + 30 - 2}}\sqrt{\frac{1}{40} + \frac{1}{30}}}$$

$$t = -1{,}78$$

ومن الجدول رقم III نجد أن:

$$Z_{1-\alpha/2} = Z_{0,975} = 1{,}96$$

وبالمقارنة نجد أن القيمة المطلقة لـ t أصغر من القيمة الجدولية $Z_{1-\alpha/2}$ (لأن $|-1{,}78| < 1{,}96$)، أي أن:

$$|t| < z_{1-\alpha/2}$$

ولهذا فإننا نقول بعدم وجود فرق جوهري بين متوسطي العينتين المسحوبتين من مجتمع واحد. وإن قيمة الفرق ناتجة عن أسباب عرضية كالمصادفة والعشوائية.

5−8− اختبارات حول الفرق بين نسبتين في مجتمعين طبيعيين R_1 و R_2 :

عندما نريد اختبار الفرق بين النسبتين $(R_1 - R_2)$ في مجتمعين مختلفين، فإننا نستخدم الفرق بين هاتين النسبتين في العينة وهو $(r_1 - r_2)$.

ونضع الفرضية الابتدائية كما يلي:

$$H_0 : (R_1 - R_2) = (R_1 - R_2)_0 \qquad (23-5)$$

ونضع الفرضية البديلة من أحد الأشكال الثلاثة المذكورة سابقاً:

وعندها يكون الاختبار ثنائي الجانب $H_1 : R_1 - R_2 \neq (R_1 - R_2)_0$

وعندها يكون الاختبار أحادي يميني $H_1 : R_1 - R_2 > (R_1 - R_2)_0$

وعندها يكون الاختبار أحادي يساري $H_1 : R_1 - R_2 < (R_1 - R_2)_0$

أما مؤشر الاختبار فهو على نمط المؤشر (5−16) ويعطى، عندما يكون مجموع حجمي العينتين $(n_1 + n_2)$ كبيراً بالعلاقة:

$$t = \frac{(r_1 - r_2) - (R_1 - R_2)_0}{\sqrt{\frac{r_1 q_1}{n_1} + \frac{r_2 q_2}{n_2}}} \qquad (24-5)$$

ولاتخاذ القرار المناسب نقارن قيمة t مع $Z_{1-\alpha/2}$ أو مع $Z_{1-\alpha}$ حسب نوع الاختبار ووجهته. فإذا كان الاختبار ثنائي الجانب وكان $|t| \leq Z_{1-\alpha/2}$ فإننا نقبل الفرضية H_0، ونرفضها في حالة العكس.

أما إذا كان الاختبار أحادي يميني وكان $t > Z_{1-\alpha}$ فإننا نقبل الفرضية H_0، ونرفضها في حالة العكس.

أما إذا كان الاختبار أحادي يساري وكان $t > -Z_{1-\alpha}$ فإننا نقبل الفرضية H_0، ونرفضها في حالة العكس.

حالة خاصة 1:

إذا كان للمجتمعين المدروسين تباينان متساويان $\sigma_1^2 = \sigma_2^2 = \sigma^2$، فإننا نقدر تباين الفرق $(r_1 - r_2)$ بواسطة العلاقة التالية:

$$t = \frac{(r_1 - r_2) - (R_1 - R_2)_0}{S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \qquad (25-5)$$

حيث أن:

$$S^2 = \frac{n_1 r_1 q_1 + n_2 r_2 q_2}{n_1 + n_2 - 2} \qquad (26-5)$$

ولتحديد نتيجة الاختبار عندما يكون الحجم $(n_1 + n_2)$ صغيراً وكان الاختبار ثنائياً فإننا نقارن قيمة t مع قيمة متحول (ستودينت) $t_{1-\alpha/2}$ المقابلة للاحتمال $1 - \frac{\alpha}{2}$ ولـ $(n_1 + n_2 - 2)$ درجة حرية (وعندما يكون $(n_1 + n_2)$ كبير نسبياً فإننا نقارن قيمة t مع $Z_{1-\alpha/2}$ أو مع $Z_{1-\alpha}$ حسب الحال).

فإذا كانت $|t| \le t_{1-\alpha/2}$ فإننا نقبل الفرضية H_0 الموضوعة.

أما إذا كان الاختبار أحادي يميني فإننا نقارن قيمة t مع أو مع $t_{1-\alpha}$، فإذا كانت $t < t_{1-\alpha}$ فإننا نقبل الفرضية H_0.

أما إذا كان الاختبار أحادي يساري فإننا نقارن قيمة t مع أو مع $(-t_{1-\alpha})$، فإذا كانت $t > (-t_{1-\alpha})$ فإننا نقبل الفرضية H_0.

حالة خاصة 2:

إذا سحبنا العينتين من مجتمع واحد فإنه يكون لدينا $R_1 - R_2$ وبالتالي يكون لدينا حكماً ما يلي:

$$(R_1 - R_2) = 0 \quad , \quad \sigma_1^2 = \sigma_2^2 = \sigma^2$$

وأن مؤشر الاختبار يأخذ الشكل التالي:

$$t = \frac{(r_1 - r_2)}{S \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \qquad (27-5)$$

وتتحول عملية الاختبار إلى اختبار جوهرية الفرق بين النسبتين $(r_1 - r_2)$ في العينتين. فإذا كانت $|t| \le z_{1-\alpha/2}$ فإننا نقول بعدم وجود فرق جوهري بين النسبتين $(r_1 \ , \ r_2)$ في العينتين المسحوبتين.

مثال 7:

لدراسة نسبة المدخنين في مجتمعين فيهما $\sigma_1^2 = \sigma_2^2$ ، سحبنا عينتين بحجمين $n_1 = 10$ و $n_2 = 15$ شخصاً، فوجدنا أن نسبتي المدخنين فيهما تساويان $r_1 = 0,20$ و $r_2 = 0,25$ على الترتيب. فهل هناك دلالة كافية لاعتبار أن نسبتي المدخنين R_1 و R_2 مختلفان إذا كان مستوى الدلالة 0,05؟

الحل:

نضع الفرضيتين الابتدائية والبديلة كما يلي:

$H_0 : R_1 - R_2 = 0$ (عدم وجود فرق جوهري)

$H_1 : R_1 - R_2 \neq 0$ (يوجد فرق جوهري)

ونحسب قيمة مؤشر الاختبار من العلاقة (5-25):

$$t = \frac{(0,2 - 0,25) - 0}{\sqrt{\frac{10(0,2)(0,8) + 15(0,25)(0,75)}{10 + 15 - 2}} \sqrt{\frac{1}{10} + \frac{1}{10}}} = -0,2796$$

وبما أن شكل H_1 يجعل الاختبار ثنائي الجانب، لذلك نبحث عن قيمة $t_{1-\alpha/2}$ المقابلة لـ $\left(1 - \frac{\alpha}{2}\right)$ ولـ $n_1 + n_2 - 2$ درجة حرية في توزيع (ستودينت) فنجد أن:

$$t_{(0,975\ ,\ 10+15-2)} = t(0,975 \cdot 23) = 2,07$$

وبما أن الاختبار ثنائي الجانب وأن t تحقق المتراجحة

$|t| < t_{1-\alpha/2}$، أي أن $|-0,2796| < 2,074$

لذلك نقول بعدم وجود فرق جوهري بين النسبتين المذكورتين، في المجتمعين المذكورين.

مثال 8:

في دراسة لنسبة المصابين بأمراض قلبية في مجتمع ما، سحبنا عينتين بحجمين $n_1 = 12$ و $n_2 = 15$، فوجدنا أن نسبتي المصابين بمرض قلبي فيهما تساويان $r_1 = 0,10$ و $r_2 = 0,15$، فهل يعتبر الفرق بين هاتين النسبتين فرقاً جوهرياً، إذا كان مستوى الدلالة 0,05؟:

الحل:

بما أن العينتين سحبتا من مجتمع واحد فهذا يعني أنه لا حاجة لوضع للفرضيات، وأن مؤشر الاختبار يحسب من العلاقة (5-27) فيكون لدينا:

$$t = \frac{0,10 - 0,15}{\sqrt{\frac{12(0,10)(0,90)+15(0,15)(0,85)}{12+15-2}}\sqrt{\frac{1}{12}+\frac{1}{12}}} = -0,12$$

وبما أن الحجم $n_1 + n_2$ صغير، لذلك فإننا نبحث عن قيمة $t_{1-\alpha/2}$ المقابلة للاحتمال $1-\frac{\alpha}{2}$ ولـ $n_1 + n_2 - 2$ درجة حرية في توزيع (ستودينت) فنجد أن:

$$t_{(0,975\ ,\ 12+15-2)} = t_{(0,975\ ,\ 25)} = 2,06$$

وبالمقارنة نجد أن $|-0,12| < 2,06$، أي أن:

$$|t| < t_{1-\alpha/2}$$

لذلك فإننا نقول بعدم وجود فرق جوهري بين النسبتين المذكورتين في العينتين.

5-9- اختبارات حول القياسات الشاذة لحذفها:

كثيراً ما تتضمن القياسات التي تكون بحوزتنا قيماً متطرفة تختلف كثيراً عن باقي القيم الأخرى، وقد يعود ذلك لأسباب مختلفة، منها:

. خطا القياس الذي يمكن أن ينجم عن سوء استخدام الأجهزة أو تعطلها.

. خطأ النقل أو النسخ أو التسجيل...إلخ.

وحتى لا تشوه هذه القياسات الشاذة النتائج التي سنتوصل إليها، نقوم بدراسة صحة انتمائها إلى مجموعة القياسات المفروضة وذلك باستخدام مبادئ اختبار الفرضيات. فإذا كان الاختبار يؤيد انتمائها نتركها ضمن المجموعة، وإذا كان العكس نقوم بحذفها من المجموعة. ولتوضيح ذلك لنفترض أنه لدينا القياسات التالية:

$$x_0\ ,\ x_1\ ,\ x_2\ ,\ x_3 \qquad x_n$$

حيث إننا نعتبر أن x_0 هو القياس الذي نشك بعدم انتمائه للمجموعة بسبب اختلاف قيمته عن الآخرين.

ويمكننا هنا أن ننطلق من الفرضيتين التاليتين:

أي أن x_0 ينتمي إلى المجموعة لأنه قريب من $\bar{x}$ $\quad H_0 : x_0 = \bar{x}$

أي أن x_0 لا ينتمي إلى المجموعة لأنه أكبر من $\bar{x}$ بكثير $\quad H_0 : x_0 > \bar{x}$

أو ننطلق من الفرضيتين التاليتين:

(ينتمي إلى المجموعة) $\quad H_0 : x_0 = \bar{x}$

(لا ينتمي إلى المجموعة) $\quad H_0 : x_0 > \bar{x}$

وحتى ندرس صحة انتماء x_0 إلى المجموعة نقوم بعزله عن المجموعة، ثم نحسب متوسط العناصر المتبقية $\bar{x}$ وانحرافها المعياري s ونحسب قيمة مؤشر الاختبار التالي:

$$t = \frac{x_0 - \bar{x}}{s} \qquad (28-5)$$

ومؤشر الاختبار t هو متحول خاضع لتوزيع ستودينت ذي $n-1$ درجة حرية. ولاتخاذ قرار حول x_0 نقارن قيمة t مع قيمة متحول (ستودينت) $t_{1-\alpha}$ المقابلة للاحتمال $1-\alpha$ ولـ $n_1 + n_2 - 2$ درجة حرية.

فإذا كان الاختبار أحادياً يمينياً وكانت $t \le t_{1-\alpha}$ فإننا نقبل الفرضية H_0 ونعتبر أن x_0 ينتمي إلى المجموعة. وإذا كان العكس، فإننا نعتبر x_0 غير منتمٍ للمجموعة وهو عبارة عن قيمة شاذة، لذلك نقوم بحذفه من المجموعة بشكل نهائي حتى لا يشوه لنا نتائج التحليل الإحصائي. أما إذا كان الاختبار أحادياً يسارياً فإننا نقارن قيمة t مع $-\left(t_{1-\alpha}\right)$ فإذا كانت $t \ge \left(-t_{1-\alpha}\right)$ فإننا نقبل الفرضية H_0 ونعتبر أن القياس x_0 من المجموعة المفروضة. أما إذا كان العكس، فإننا نرفض H_0 ونعتبر x_0 قياساً شاذاً ونحذفه من المجموعة.

مثال 9:

لتكن لدينا القياسات التالية عن وزن سبعة صناديق من التفاح:

25,5 , 24,1 , 23,5 , 23,4 , 23,3 , 23,2 , 23,8

فهل القياس $x_0 = 25,5$ يعد من قياسات هذه المجموعة. اختبر بمستوى دلالة $0,05$.

الحل:

نضع الفرضيتين كما يلي:

$$H_0 : x_0 = \bar{x}$$
$$H_0 : x_0 > \bar{x}$$

نعزل x_0 ثم نحسب المتوسط $\bar{x}$ والانحراف s للقياسات الأخرى فنجد أن:

$$\bar{x} = \frac{1}{n}\sum x_i = \frac{1}{6}\left[24,1 + 23,5 + 23,4 + 23,3 + 23,2 + 23,8\right] = 23,6$$

$$s^2 = \frac{1}{n-1}\sum \left(x_i - 23,6\right)^2 = 0,12$$
$$s = 0,35$$

وبذلك نجد أن مؤشر الاختبار يساوي:

$$t = \frac{25,5 - 23,6}{0,35} = 5,43$$

ومن جدول ستودينت نجد أن:

$$t_{(0,95\ ,\ n-1)} = t_{(0,95\ ,\ 5)} = 2,02$$

وبالمقارنة نجد أن $t > t_{1-\alpha}$ لأن $5,43 > 2,02$ لذلك نعتبر القياس $x_0 = 25,5$ قياساً شاذاً ونحذفه من المجموعة.

تمرينات

1 . سحبنا عينة بحجم $n=100$ عنصراً من مجتمع طبيعي فوجدنا أن متوسطها $\bar{x}=2,7$، وأن $s^2=23$. اختبر الفرضية $H_0:\bar{y}=3$ بمستوى دلالة $\alpha=0,10$.

2 . ادعى أحد المرشحين أن 80% من السكان سينتخبونه. ولكن عينة من 500 شخصاً بينت أن 300 شخصاً يؤيدون ذلك المرشح. اختبر ادعاء المرشح باحتمال قدره $0,95$.

3 . بينت إحدى الدراسات في العام الماضي أن متوسط دخل العامل كان 1500 ل.س. وفي دراسة حالية شملت 150 عاملاً سحبوا بطريقة عشوائية وجدنا أن متوسط الدخل $\bar{x}=1700$ ل.س. وأن تباينه $s^2=2000$. فهل يعد هذا دليلاً كافياً، وباحتمال قدره $0,95$ على تحسن دخل العمال؟

4 . سحبنا عينين بحجمين $n_1=n_2=11$ عنصراً من مجتمعين طبيعيين متوسطيهما $\bar{y}_1$ و $\bar{y}_2$ ولهما تباين موحد. فوجدنا أن $\bar{x}_1=60$ و $\bar{x}_2=65$ ، وأن $s_1^2=31$ و $s_2^2=45$ فهل تقدم هذه المعلومات دلالة كافية، وباحتمال قدره $0,95$ على وجود فرق بين المتوسطين $\bar{y}_1$ و $\bar{y}_2$.

5 . عند اختبار عمر 16 مصباحاً كهربائياً وجدنا أن تقدير متوسط عمر المصباح H $\bar{x}=3800$ وأن تباينه $s^2=4000$. اختبر ادعاء الشركة بأن متوسط عمر مصابيحها يساوي $4000H$ وذلك بمستوى دلالة $0,05$.

6 . لتقدير نسبة الأميين في مدينتين A وB سحبنا منهما عينتين بحجمين $n_1=30$ و $n_2=40$ عاملاً. فوجدنا أن نسبة الأميين فيهما $R_1=0,10$ و $R_2=0,12$، فهل يعطينا هذه دلالة كافية لاعتبار هاتين النسبتين متساويتين وذلك بمستوى دلالة $0,08$.

7 . لنفترض أنه لدينا القياسات التالية:

100 , 105 , 102 , 110 , 112 , 115 , 120 , 130

اختبر فيما إذا كان القياس 130 ينتمي إلى هذه المجموعة وذلك بمستوى دلالة 0,05.

8 . سحبنا عينين من الخراف بحجمين $n_1 = 10$ و $n_2 = 15$ خروفاً من قطيع واحد. فوجدنا أن متوسط وزن الخروف فيهما كان $\bar{x}_1 = 30$ و $\bar{x}_2 = 33$ ، وأن تباينهما كان $s_1^2 = 300$ و $s_2^2 = 400$ فهل هناك فرق جوهري بين متوسطي هاتين العينتين وذلك بمستوى دلالة 0,05.

الفصل السادس

اختبار كاي مربع χ^2 وتطبيقاته

ويشمل الفقرات التالية:

1 . مقدمة.

2 . اختبار مطابقة التكرارات.

3 . اختبار مطابقة النسب.

4 . اختبار مطابقة توزيع تجريبي مع توزيع نظري.

5 . اختبار استقلالية متحولين.

6 . شروط تطبيق اختبار χ^2 .

7 . اختبار χ^2 في حالة جدول 2×2.

8 . معامل التوافق C.

9 . معامل التوافق T^2 .

10 . اختبار تباين المجتمع بواسطة χ^2 .

11 . اختبار تناظر الجداول المربعة بواسطة χ^2 .

12 . تمرينات.

الفصل السادس

اختبار كاي مربع χ^2 وتطبيقاته

6-1: مقدمة:

يعتبر اختبار χ^2 من الاختبارات الأكثر استخداماً في عمليات التحليل الإحصائي، ويطبق بشكل خاص على المتحولات النوعية أكثر من تطبيقاته على المتحولات الكمية.

ولقد تناول العديد من العلماء هذا الاختبار بالتحليل والتعليل والتعديل، وكان أولهم العالم (Karl Person, 1857- 1935)، ثم تابعه (Snedecor , Brandt)، كما عمل على تعديله وتدقيقه كل من (Cochran, 1954) و(Yéats) و (Ryan, 1960) و (Lewantin, Felsentein, 1965) وغيرهم.

ويطبق اختبار χ^2 على التكرارات النسبية، أو المطلقة للمتحولات النوعية أو الكمية المستوحاة من التجارب الميدانية لمقارنتها بتكرارات متوقعة، أو مفترضة، وذلك بهدف استنتاج بعض الخصائص الإحصائية الملازمة لتلك المتحولات، أو الحكم على توافقها، أو على استقلاليتها عن بعضها البعض.

لذلك سنميز بين نوعين أساسيين لاختبار χ^2 هما:

أ . اختبار x^2 لمتحول واحد:

ويعرف بالعلاقة التالية:

$$\chi^2 = \sum_{i=1}^{k}\left[\frac{(n_i - m_i)^2}{m_i}\right] \qquad (1-6)$$

حيث أن: k هو عدد حالات (أو قيم) المتحول المدروس X، وهي:

$(X : x_1 \ , \ x_2 \ldots\ldots x_k)$.

وأن: n_i هو رقم التكرارات التجريبية المقابلة للحالة (أو القيمة) x_i.

وأن: m_i هو رقم التكرارات المتوقعة، أو المفترضة المقابلة للحالة (أو القيمة) x_i.

وأن فرضية العدم هي: عدم وجود اختلافات جوهرية بين التكرارات التجريبية n_i والتكرارات المتوقعة m_i وذلك لجميع حالات X.

ولقد برهن (بيرسون) على أن المؤشر χ^2 يخضع للتوزيع الاحتمالي $\chi^2_{(x)}$ بـ $(k-1)$ درجة حرية (وذلك لأنه يشترط هنا أن تكون التكرارات النظرية miمرتبطة مع التكرارات الفعلية ni بالعلاقة $\sum n_i = \sum m_i$). لذلك فإنه عندما نقوم بحساب قيمة المؤشر χ^2 فإننا نقارنها مع القيمة الحرجة لمتحول التوزيع $\chi^2_{(x)}$ المقابلة لـ$(k-1)$ درجة حرية ولمستوى الدلالة α. فإذا كانت القيمة المحسوبة χ^2 أكبر من القيمة الحرجة χ^2_α، نرفض فرضية العدم، والعكس بالعكس.

ب . اختبار χ^2 لمتحولين متقاطعين $(x\ ,\ y)$:

ويعرف بالعلاقة التالية:

$$x^2 = \sum_{i=1}^{k} \sum_{j=1}^{r} \left[\frac{(n_{ij} - m_{ij})^2}{m_{ij}} \right] \qquad (2-6)$$

حيث أن: k هو عدد حالات (أو قيم) المتحول X وهي:

$$(X:\ x_1\ ,\ x_2 \ldots . x_i, x_k).$$

وأن: r هو عدد حالات (أو قيم) المتحول Y وهي:

$$(Y:\ y_1\ ,\ y_2 \ldots . . y_i, y_r).$$

وأن: n_{ij} هو رقم التكرارات التجريبية في الخلية $(x_i\ ,\ y_j)$.

وأن: m_{ij} هو رقم التكرارات المتوقعة، أو المفترضة في الخلية $(x_i\ ,\ y_j)$.

وأن فرضية العدم هي: عدم وجود علاقة ارتباط بين المتحولين (x, y)؛ أي أن التكرارات التجريبية n_{ij} موزعة على الخلايا المتقاطعة $(x_i\ ,\ y_j)$ بشكل لا يعطي دليلاً كافياً على ارتباط المتحولين (x, y).

ولقد برهن (بيرسون) على أن المؤشر χ^2 الأخير يخضع للتوزيع الاحتمالي $\chi^2_{(x)}$ ذي $(k-1)\cdot(r-1)$ درجة حرية؛ لذلك نقوم بحساب قيمة المؤشر χ^2 من

العلاقة السابقة ونقارنها مع القيمة الحرجة لمتحول التوزيع $\chi^2_{(x)}$ المقابلة لـ$(r-1)\cdot(k-1)$ درجة حرية ولمستوى الدلالة α. فإذا كانت القيمة المحسوبة χ^2 أكبر من القيمة الحرجة χ^2_α، فإننا نرفض فرضية العدم، والعكس بالعكس.

وسنقوم في هذا الفصل باستخدام هذين المؤشرين لدراسة واختبار الأمور التالية:

. اختبار مطابقة تكرارات توزيعين تجريبيين.

. اختبار مطابقة النسب في عدة عينات.

. اختبار مطابقة توزيع تجريبي مع توزيع نظري.

. اختبارات العلاقات في الجداول من النوع $k\cdot 2$.

. اختبارات الاستقلال في جداول التوافق من النوع $k\cdot 2$. وتقدير درجة الارتباط.

. اختبارات الاتجاه العام.

. اختبارات تناظر الجداول المربعة.

. اختبارات الاستقلال والتجانس.

. اختبارات الانتظام في التوزيع.

. شروط تطبيق اختبار χ^2.

6-2: اختبار مطابقة التكرارات بواسطة χ^2:

لإجراء هذا الاختبار نأخذ المثال الآتي حول التكرارات الخاضعة للتوزيع المنتظم:

مثال 1:

لنفترض أننا قمنا برمي حجر النرد 60 مرة، فكانت نتائج تلك التجارب وتكراراتها المطلقة كما في الجدول الآتي:

نتيجة التجربة	1	2	3	4	5	6		Σ
التكرارات التجريبية	15	7	4	11	6	17		60

فهل تدل هذه النتائج على انتظام التكرارات وسلامة حجر النرد؟

وهنا نلاحظ أن هدف الاختبار في هذا المثال يدور حول سلامة حجر النرد، من خلال انتظام وتساوي التكرارات المطلقة المقابلة لكل وجه من أوجه النرد،

الحل:

نضع الفرضيات كما يلي:

H_0: نفترض إن جميع التكرارات المقابلة لأوجه النرد متساوية وتساوي $\frac{60}{6}=10$.

H_1: يوجد أحد التكرارات على الأقل لا يساوي 10 والنرد غير سليم.

وحتى نستطيع إجراء الاختبار نقوم بحساب قيمة مؤشر الاختبار χ^2 المعرف بالعلاقة (6-1) ونعدّ الجدول الأتي:

نتيجة التجربة	1	2	3	4	5	6	Σ
التكرارات التجريبية n_i	15	7	4	11	6	17	60
التكرارات المتوقعة m_i	10	10	10	10	10	10	60
$(n_i-m_i)^2$	25	9	36	1	16	49	.
$\frac{(n_i-m_i)^2}{m_i}$	2,5	4,9	3,6	0,1	1,6	4,9	$\chi^2=13,6$

وبذلك نجد أن قيمة مؤشر الاختبار χ^2 تساوي:

$$\chi^2=\sum_{i=1}^{k}\frac{(n_i-m_i)^2}{m_i}=13,6$$

ثم نقوم بمقارنة هذه القيمة مع القيمة الحرجة التي نستخرجها من جدول التوزيع $\chi^2_{(x)}$ عند درجة حرية $k-1=6-1=5$ والمقابلة لمستوى دلالة قدره $\alpha=0,05$، فنجد أنها تساوي $\chi^2_\alpha=11,070$.

وبالمقارنة نجد أن القيمة المحسوبة χ^2 أكبر من القيمة الحرجة χ^2_α، لذلك نرفض فرضية العدم H_0 التي تقول إن جميع التكرارات متساوية، ونقبل الفرضية

البديلة H_1 التي تقول بأنه يوجد أحد التكرارات على الأقل لا يساوي 10 وأن النرد غير سليم.

6-3: اختبار مطابقة النسب بواسطة χ^2:

مثال 2:

لنفرض أن دراسةً لسبر آراء الطلاب الذين اتبعوا دورة توعية حول مضار التدخين شملت 500 طالباً وطالبة، أعطتنا النتائج التالية:

حالات الرأي	مؤيد	محايد	معارض	ممتنع	Σ
التكرارات المطلقة	39	336	99	26	500
النسب المئوية	7,8	67,2	19,8	5,2	100

قارن هذه النتائج مع نتائج دراسة سابقة أجريت قبل إجراء الدورة والتي كانت نتائجها كما يلي:

نسبة المؤيدين 7%، ونسبة المحايدين 65%، ونسبة المعارضين 18%، ونسبة الممتنعين 10%.

ثم اختبر فيما إذا كانت هذه النسب متطابقة مع النسب الواردة في الدراسة الحالية، وذلك بمستوى دلالة $\alpha = 0,01$.

الحل:

نضع الفرضيات كما يلي:

H_0: لا توجد فروقات جوهرية بين النسب المئوية المقابلة لحالات الرأي في الدراستين القبلية والبعدية.

H_1: توجد فروقات جوهرية بين النسب المئوية المذكورة.

وحتى نستطيع حساب مؤشر الاختبار χ^2 نحسب التكرارات المطلقة للدراسة السابقة (لعينة بالحجم نفسه 500 طالباً) فنجد أن:

عدد المؤيدين (سابقاً) $m_1 = 500 \times \frac{7}{100} = 35$

عدد المحايدين (سابقاً) $m_2 = 500 \times \frac{65}{100} = 325$

عدد المعارضين (سابقاً) $m_3 = 500 \times \frac{18}{100} = 90$

عدد الممتنعين (سابقاً) $m_4 = 500 \times \frac{10}{100} = 50$

ولحساب χ^2 نعدّ الجدول الأتي:

حالات الرأي	مؤيد	محايد	معارض	ممتنع	Σ
التكرارات الحالية n_i	39	336	99	26	500
التكرارات (السابقة) m_i	35	325	90	50	500
$(n_i - m_i)^2$	16	121	81	576	.
$\frac{(n_i - m_i)^2}{m_i}$	0,458	0,900	0,372	11,2	$\chi^2 = 13,349$

وبذلك نجد أن قيمة مؤشر الاختبار χ^2 تساوي:

$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - m_i)^2}{m_i} = 13,349$$

و بمقارنة هذه القيمة مع القيمة الحرجة لمتحول التوزيع χ^2_α المقابلة لـ $k-1=4-1=3$ درجة حرية ولمستوى دلالة قدره $\alpha = 0,01$ والتي تساوي $\chi^2_x = 11,345$ نجد أن القيمة المحسوبة χ^2 أكبر من القيمة الحرجة χ^2_α، لذلك نرفض فرضية العدم H_0، ونقبل الفرضية البديلة H_1، التي تقول بوجود فروقات جوهرية بين النسب المئوية المقابلة لحالات الرأي حول التدخين بين الدراستين القبلية والبعدية. وهذا يعني أن الدورة قد أثرت على الرأي العام للطلاب الذين اتبعوا الدورة. ولكن هذا الرفض يطرح أمامنا سؤالين هامين وهما:

من أين جاءت هذه الفروقات؟ وفي أي الحالات تظهر؟

للإجابة على هذين السؤالين نقوم بحساب مجالات الثقة لكل من النسب المئوية في الدراسة البعدية، وذلك باستخدام جداء الانحراف المعياري لكل منها بمعامل الاحتمال الطبيعي $Z_{1-\frac{\alpha}{2}}$ للحصول على نصف طول مجال الثقة المطلوب من العلاقة:

$$\varepsilon = Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{P \cdot (1-P)}{n}}$$

فنجد أن نصف طول مجال الثقة للنسبة الأولى $P_1 = 0{,}078$ يساوي عند مستوى دلالة $\alpha = 0{,}05$ ما يلي:

$$\varepsilon_1 = 1{,}96 \cdot \sqrt{\frac{0{,}078 \cdot (1-0{,}078)}{500}} = 0{,}0235$$

وبذلك نجد أن مجال الثقة للنسبة الحقيقية R_1 يساوي:

$$[0{,}078 - 0{,}0235 \le R_1 \le 0{,}078 + 0{,}0235]$$
$$[0{,}0545 \le R_1 \le 0{,}1015]$$

وبما أن النسبة P_1 في الدراسة القبلية والتي تساوي $P_1 = 0{,}07$ تقع ضمن ذلك المجال السابق، فإننا نستنتج أن الفروقات الناجمة عن الدورة لا تعود إلى هذه الحالة؛ أي أن محاضرات الدورة لم تؤثر جوهرياً على نسبة المؤيدين للتدخين.

وكذلك نجد أن نصف طول مجال الثقة للنسبة الثانية والتي تساوي $P_2 = 0{,}672$ يساوي عند مستوى دلالة $\alpha = 0{,}05$ ما يلي:

$$\varepsilon_1 = 1{,}96 \cdot \sqrt{\frac{0{,}672 \cdot (1-0{,}672)}{500}} = 0{,}0412$$

وبذلك نجد أن مجال الثقة للنسبة الحقيقية R_2 يساوي:

$$[0{,}672 - 0{,}0412 \le R_2 \le 0{,}672 + 0{,}0412]$$
$$[0{,}6308 \le R_2 \le 0{,}732]$$

وبما أن نسبة المحايدين في الدراسة القبلية والمساوية $P_2' = 0{,}65$ تقع ضمن

مجال الثقة السابق، لذلك فإننا نستنتج أن الفروقات الناجمة عن الدورة لا تعود إلى هذه الحالة؛ أي أن محاضرات الدورة لم تؤثر جوهرياً على نسبة المحايدين.

وبطريقةٍ مشابهة نجد أن نصف طول مجال الثقة للنسبة الثالثة $P_3 = 0{,}198$ يساوي:

$$\varepsilon_3 = 1{,}96 \cdot \sqrt{\frac{0{,}198 \cdot (1 - 0{,}198)}{500}} = 0{,}0349$$

وبذلك نجد أن مجال الثقة للنسبة الحقيقية R_3 يساوي:

$$[0{,}198 - 0{,}0349 \le R_3 \le 0{,}198 + 0{,}0349]$$
$$[0{,}1631 \le R_3 \le 0{,}2329]$$

وبما أن نسبة المعارضين في الدراسة القبلية والتي تساوي $P_3' = 0{,}18$ تقع ضمن مجال الثقة السابق، لذلك فإننا نستنتج أن الفروقات الناجمة عن الدورة لا تعود إلى هذه الحالة، وأن المحاضرات لم تؤثر جوهرياً على نسبة المعارضين أيضاً.

وأخيراً نجد أن نصف طول مجال الثقة للنسبة الرابعة والتي تساوي $P_4 = 0{,}052$ يساوي:

$$S_4 = 1{,}96 \cdot \sqrt{\frac{0{,}052 \cdot (1 - 0{,}052)}{500}} = 0{,}0195$$

وبذلك نجد أن مجال الثقة للنسبة الحقيقية R_4 يساوي:

$$[0{,}052 - 0{,}0195 \le R_4 \le 0{,}052 + 0{,}0195]$$
$$[0{,}0325 \le R_4 \le 0{,}0715]$$

وبما أن نسبة الممتنعين في الدراسة القبلية والتي تساوي 0,10 لا تقع ضمن ذلك المجال، لذلك نستنتج أن الفروقات الناجمة عن الدورة تعود إلى هذه الحالة؛ أي أن محاضرات الدورة أثرت على نسبة الممتنعين عن إبداء الرأي حول مضار التدخين.

6-4: اختبار مطابقة توزيع تجريبي مع توزيع نظري:

لنفرض أننا أجرينا n تجربة على سلوك متحول عشوائي منقطع X وحصلنا على التوزيع التجريبي الإحصائي له، والذي كان كما يلي:

قيم X	x_1	x_2	x_3	$x_4 \ldots .$	$\ldots x_i \ldots$	$\ldots . x_m$	Σ
التكرارات المطلقة n_i	n_1	n_2	n_3	$n_4 \ldots .$	$\ldots n_i \ldots$	$\ldots . n_m$	n
التكرارات النسبية، أو التوزيع التجريبي	$\frac{n_1}{n}$	$\frac{n_2}{n}$	$\frac{n_3}{n}$	$\frac{n_4}{n} \ldots .$	$\ldots \frac{n_i}{n} \ldots$	$\ldots . \frac{n_m}{n}$	1

ومن خلال تحليل الشكل الذي يأخذه التوزيع التجريبي يمكننا أن نفترض أن توزيع X في المجتمع المدروس يتطابق مع أحد التوزيعات النظرية المعروفة: كالتوزيع المنتظم، أو التوزيع الثنائي، أو توزيع بواسون، أو غيرهما. ولذلك نضع الفرضيات كما يلي:

H_0: إن المتحول العشوائي X يخضع لتوزيع نظري المعين؛ أي أنه لا توجد أية فروقات جوهرية بين قيم التوزيع التجريبي (التكرارات النسبية $\frac{n_i}{n}$) والاحتمالات النظرية المأخوذة من التوزيع المفترض p_i والمقابلة للقيمة x_i.

H_1: إن المتحول العشوائي X لا يخضع للتوزيع النظري المفترض؛ أي أنه توجد فروقات جوهرية بين قيم التوزيع التجريبي، وقيم التوزيع النظري المعين.

وللتحقق من صحة، أو عدم صحة فرضية العدم H_0 نقوم بحساب قيمة مؤشر الاختبار χ^2 من العلاقة:

$$\chi^2 = n\sum_{i=1}^{k} \frac{\left(\frac{n_i}{n} - p_i\right)^2}{p_i} \qquad (3-6)$$

ولقد برهن بيرسون على أن المقدار χ^2 السابق يخضع لتوزيع $\chi^2_{(x)}$ بـ $(k-1)$ درجة حرية لأن التكرارات النسبية مرتبطة بالعلاقة $\sum \frac{n_i}{n} = 1$ (وكذلك بفرض أن

تكون p_i مستقلة عن التكرارات النسبية $\frac{n_i}{n}$ وإلا فإن درجة الحرية تنخفض بمقدار عدد العلاقات بينها).

ويمكننا تحويل العلاقة السابقة إلى الشكل التالي:

$$\chi^2 = n\sum_{i=1}^{k}\frac{(n_i - np_i)^2}{n^2 \cdot p_i} = \sum_{i=1}^{k}\frac{(n_i - np_i)^2}{n \cdot p_i} \qquad (4-6)$$

وبما أن الجداء $n \cdot p_i$ يساوي التكرارات المتوقعة m_i المقابلة للقيمة x_i، لأن جداء الاحتمالات النظرية p_i في حجم العينة n يعطينا التكرارات المتوقعة m_i، وبذلك نجد أن العلاقة (6-4) تأخذ الشكل التالي:

$$x^2 = \sum_{i=1}^{k}\frac{(n_i - m_i)^2}{m_i} \qquad (5-6)$$

وهي العلاقة الأولى نفسها (6-1) التي تخضع للتوزيع $\chi^2_{(x)}$ بـ $(k-1)$ درجة حرية، وهذا يعني أن العلاقة (6-3) هي شكل آخر للعلاقة (6-1).

وعند إجراء المطابقة بين أي توزيع تجريبي وآخر نظري، يمكننا استخدام أي من هذه العلاقات، وذلك حسب البيانات المتوفرة لدينا، وسنوضح ذلك من خلال الأمثلة التالية:

مثال 3: حول المطابقة مع توزيع بواسون:

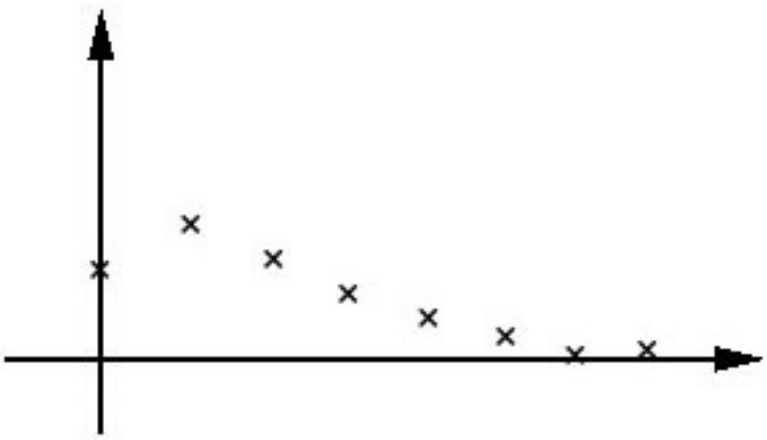
لنفترض أن عدد الأخطاء المطبعية في بحث مؤلف من 60 صفحة كان موزعاً كما في الجدول الآتي:

i التسلسل	1	2	3	4	5	6	7	8	
عدد الأخطاء x_i	0	1	2	3	4	5	6	7	Σ
عدد الصفحات المقابل لعدد الأخطاء (التكرارات n_i)	8	17	16	10	6	2	0	1	60
التكرارات النسبية $\frac{n_i}{n}$	$\frac{8}{60}$	$\frac{17}{60}$	$\frac{16}{60}$	$\frac{10}{60}$	$\frac{6}{60}$	$\frac{2}{60}$	$\frac{0}{60}$	$\frac{1}{60}$	1

والمطلوب إيجاد توزيع نظري منقطع يوافق التوزيع التجريبي الإحصائي السابق، ثم اختيار المطابقة بينهما بمستوى دلالة $\alpha = 0,05$.

الحل:

نقوم برسم الشكل البياني للتوزيع الإحصائي السابق، فنحصل على الشكل التالي:



الشكل رقم (6−1)

وهنا نلاحظ أن هذا الشكل يشبه توزيع بواسون. لذلك نقوم بحساب الاحتمالات النظرية من توزيع بواسون المعرف بالعلاقة:

$$p_x = e^{-\lambda} \cdot \frac{\lambda^x}{xi} \qquad (6-6)$$

ونضع الفرضيتين كما يلي:

H_0: إن عدد الأخطاء المطبعية يخضع لتوزيع بواسون السابق؛ أي نفترض عدم وجود فروقات جوهرية للاحتمالات بين قيم التوزيع الإحصائي وقيم توزيع بواسون.

H_1: إن عدد الأخطاء المطبعية لا يخضع لتوزيع بواسون السابق.

ولكن حتى نستطيع حساب تلك الاحتمالات نحتاج إلى معرفة قيمة λ، أو إلى حساب قيمة تقديرية للمعامل λ، ثم التأكد من أن $E(x)=\lambda$، وأن $\sigma^2(x)=\lambda$؛ أي أن التوقع $E(x)$، والتباين $\sigma^2_{(x)}$ متساويان تقريباً.

لذلك نقوم بحساب التوقع، فنجد أن:

$$E(x)=\sum_{i=1}^{8}\frac{n_i}{n}x_i=\frac{1}{n}\sum n_i x_i$$

$$=\frac{1}{60}\left[8\cdot 0+17\cdot 1+16\cdot 2+10\cdot 3+6\cdot 4+2\cdot 5+0\cdot 6+1\cdot 7\right]$$

$$E(x)=\frac{1}{60}[120]=2$$

ولحساب التباين نجد أن:

$$\sigma^2(x)=\sum_{i=1}^{8}\frac{n_i}{n}(x_i-2)^2=\frac{1}{n}\sum n_i(x_i-2)^2=$$

$$=\frac{1}{60}\left[32+17+0+10+24+18+0+25\right]=$$

$$\sigma^2(x)=\frac{1}{60}[126]=2{,}1$$

وبالمقارنة نجد أن التوقع $E(x)$ والتباين $\sigma^2(x)$ متساويان تقريباً.

وهذا مايبرر الفرضية الابتدائية H_0 التي وضعناها سابقاً، حول خضوع عدد الأخطاء في التقرير البحثي لتوزيع بواسون.

ولكن حتى نختبر صحة تلك الفرضية علينا حساب قيمة χ^2 الفعلية ثم مقارنتها مع القيمة الجدولية لمتحول التوزيع $\chi^2_{(x)}$ المقابلة لـ $k-1=8-1=7$ درجات حرية ولمستوى دلالة $\alpha=0{,}05$.

وحتى نستطيع حساب قيمة χ^2 الفعلية نحتاج إلى حساب الاحتمالات النظرية Pi (أو التكرارات المتوقعة mi) وهذا يقتضي تقدير قيمة المعلم λ اعتماداً على بيانات العينة وعلى خصائص توزيع بواسون، حيث نعلم أن $E(x)=\lambda$ وبناءً على ما تقدم يمكننا وضع التقدير التالي لـ λ.

$$\lambda = 2$$

ثم نعوض ذلك في قانون التوزيع النظري فنحصل على توزيع بواسون الآتي:

$$P_x = e^{-2}\frac{2^x}{x!} \qquad (x: 0\ ,\ 1\ ,\ 2\ ,\ 3.....)$$

ومنه يمكننا حساب الاحتمالات النظرية المقابلة لعدد الأخطاء x_i كما يلي:

$$P_{(x=0)} = P_0 = e^{-2}\frac{2^0}{0!} = e^{-2} = 0{,}1353$$

$$P_{(x=1)} = P_1 = e^{-2}\frac{2^1}{1!} = 2e^{-2} = 0{,}2707$$

$$P_{(x=2)} = P_2 = e^{-2}\frac{2^2}{2!} = 2e^{-2} = 0{,}2707$$

$$P_{(x=3)} = P_3 = e^{-2}\frac{2^3}{3!} = \frac{4}{3}e^{-2} = 0{,}1804$$

$$P_{(x=4)} = P_4 = e^{-2}\frac{2^4}{4!} = \frac{2}{3}e^{-2} = 0{,}0902$$

$$P_{(x=5)} = P_5 = e^{-2}\frac{2^5}{5!} = \frac{4}{15}e^{-2} = 0{,}0361$$

$$P_{(x=6)} = P_6 = e^{-2}\frac{2^6}{6!} = \frac{4}{45}e^{-2} = 0{,}0120$$

$$P_{(x=7)} = P_7 = e^{-2}\frac{2^7}{7!} = \frac{8}{315}e^{-2} = 0{,}0034$$

ولإجراء العمليات اللازمة لحساب χ^2 من العلاقة (6-5) نعد الجدول الآتي:

التسلسل i	1	2	3	4	5	6	7	8	
عدد الأخطاء x_i	0	1	2	3	4	5	6	7	Σ
التكرارات الفعلية $= n_i$	8	17	16	10	6	2	0	1	60
الاحتمالات المتوقعة $= P_i$	0.1353	0.2707	0.2707	0.1804	0.0902	0.0361	0.0120	0.0034	1
$m_i = nP_i$	8.12	16.24	16.24	10.83	5.41	2.17 / 3.1	0.72	0.21	60
$(n_i - m_i)^2$	0.0144	0.5776	0.0576	0.6724	0.3481	0.01	-	-	-
$\frac{(n_i - m_i)^2}{mi}$	0.001777	0.0356	0.00347	0.06214	0.06434	0.003226	-	-	0.17055 $= \chi^2$

ملاحظة:

عند حساب السطرين الأخيرين قمنا بدمج الفئتين الأخيرتين مع الفئة الخامسة لأن تكراراتها المتوقعة كانت أصغر من الواحد.

ومن الجدول السابق نلاحظ أن قيمة χ^2 تساوي:

$$\chi^2 = \sum_{i=1}^{6} \frac{(n_i - m_i)^2}{m_i} = 0{,}17055$$

وبمقارنة هذه القيمة مع القيمة الجدولية لـ χ^2_α المقابلة $k = 8 - 1 - 1 = 6$ درجات حرية، (وذلك لأننا قمنا بتقدير λ من بيانات العينة مما أضاف قيداً جديداً على درجات الحرية وجعلها تنقص بمقدار واحد آخر، فأصبحت $k = 6$ درجة حرية). ومن الجدول نجد أن:

$$x^2_{0,05,6} = 12,6$$

وبالمقارنة نجد أن χ^2 المحسوبة أصغر من χ^2_α الجدولية لذلك نقبل الفرضية الابتدائية H_0 والتي تقول أن عدد الأخطاء χ يخضع لتوزيع بواسون يتوقع قدره $\lambda = 2$ خطيئة في الصفحة.

مثال 4: حول المطابقة مع التوزيع الثنائي:

لنفترض أن لاعباً يقوم برمي 4 قطع من النقود ويسجل في كل مرة عدد الصور التي تظهر من نتيجة التجربة.

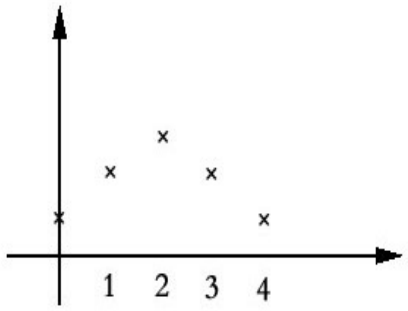
وبعد قيامه بـ 64 تجربة قام بترتيب عدد مرات ظهور الصورة فحصل على التوزيع التكراري المبين في الجدول الآتي:

i	1	2	3	4	5	Σ
xi عدد مرات ظهور الصورة في كل تجربة	0	1	2	3	4	
تكرارات ظهور الصورة ni	1	12	24	18	9	64
$\frac{n_i}{n}$ التكرارات النسبية	$\frac{1}{64}$	$\frac{12}{64}$	$\frac{24}{64}$	$\frac{18}{64}$	$\frac{9}{64}$	1

والمطلوب إيجاد التوزيع الاحتمالي المنقطع الذي يوافق التوزيع التجريبي لعدد مرات ظهور الصورة، ثم اختبار ذلك بمستوى دلالة $\alpha = 0,05$.

الحل:

نقوم برسم الشكل البياني للتوزيع التجريبي فنحصل على الشكل الآتي:



الشكل رقم (6-2)

ومن الشكل نلاحظ أن التوزيع الاحتمالي المناسب هو التوزيع الثنائي (لأنه منقطع وشبه متناظر). وبما أن التوزيع الثنائي يعطى بالعلاقة:

$$Px = C_m^x P^x q^{m-x} \qquad (x: 0\ ,\ 1\ ,\ 2,.....m) \qquad (7-6)$$

حيث إن: P هو احتمال ظهور الصورة، $q = 1 - P$ وm هو عدد الحالات الممكنة للمتحول x، وهنا $m = 4$، ولحساب الاحتمالات النظرية من هذا التوزيع يجب علينا أن نقوم بتقدير الاحتمال P من بيانات العينة، واعتماداً على خواص التوزيع الثنائي نجد أن التوقع الرياضي يساوي:

$$E(x) = m.P \Rightarrow \widetilde{P} = \frac{E(x)}{m}$$

لذلك نحسب التوقع الرياضي من العلاقة:

$$E(x) = \sum \frac{ni}{n} xi = \frac{1}{n}\sum n_i x_i$$

$$= \frac{1}{64}\left[1\cdot 0 + 12\cdot 1 + 24\cdot 2 + 18\cdot 3 + 9\cdot 4\right]$$

$$E(x) = \frac{1}{64}(150) = 2{,}34375$$

وبذلك نجد أن تقدير الاحتمال P يساوي:

$$\widetilde{P} = \frac{2,34375}{4} = 0,5859375$$

وبما أن احتمال ظهور الصورة في أي تجربة يزيد عن 0.5 فإن هذا يعني أن بعض القطع غير متوازنة وإن نتائج التجارب ستكون متحيزة لصالح ظهور الصورة.

وهنا نضع بعض الفرضيات كمايلي:

H_0 : إن عدد ظهور الصورة في كل تجربة يخضع للتوزيع الثنائي الذي فيه $P = 0,5859375$ أي أنه لايوجد فروقات جوهرية بين النتائج التي أعطتها التجارب وبين نتائج ذلك التوزيع .

H_1 : إن عدد ظهور الصورة لا يخضع للتوزيع المذكور.

ولاختبار صحة هذه الفرضية نقوم بحساب الاحتمالات النظرية المقابلة لقيم x، ثم حساب مؤشر الاختبار χ^2 فنجد أن:

$$P_0 = C_4^0 P^0 q^4 = q^4 = (1 - 0,5859375)^4 = 0,029394$$
$$P_1 = C_4^1 P^1 q^3 = 4 \cdot (0,5859375)(1 - 0,5859375)^3 = 0,166383$$
$$P_2 = C_4^2 P^2 q^2 = 6 \cdot (0,5859375)^2 (1 - 0,5859375)^2 = 0,35317$$
$$P_3 = C_4^3 P^3 q^1 = 4 \cdot (0,5859375)^3 (1 - 0,5859375)^1 = 0,33318$$
$$P_4 = C_4^4 P^4 q^0 = P^4 = (0,5859375)^4 = 0,1179$$

ولحساب χ^2 من العلاقة (6-5) نعد الجدول التالي:

i	1	2	3	4	5	$\sum$
xi	0	1	2	3	4	
ni	1	12	24	18	9	64
P_i	0,029394	0.166383	0,35317	0.33318	0.1179	1
$m_i = nP_i$	1,88	10,65	22,60	21,32	7,55	64
$(n_i - m_i)^2$	0,7744	1,8225	1,9519	11,0224	2,1025	
$\frac{(n_i - m_i)^2}{mi}$	0,4119	0,17113	0,08637	0,51699	0,2785	1,4649= χ^2

ومن هذا الجدول نجد أن قيمة χ^2 الفعلية تساوي: $\chi^2 = 1,4649$

ومن أجل المقارنة نبحث في الجدول الملحق (V) عن قيمة χ^2 الجدولية المقابلة لدرجة حرية مساوية $k = 4 - 1 - 1 = 2$ (لقد طرحنا هنا واحداً آخر من درجة الحرية لأننا استفدنا من بيانات العينة في تقدير الاحتمال P). فنجد أنها تساوي $\chi^2_{0,05,2} = 5,991$.

وبالمقارنة نجد أن χ^2 الفعلية أصغر من χ^2_α، لذلك نقبل فرضية العدم H_0 ونقبل بأن عدد مرات ظهور الصورة في كل تجربة يخضع للتوزيع الثنائي الذي فيه $P = 0,5859375$.

فرضية أخرى للحل:

يمكننا أن نحل هذه المسألة بطريقة أخرى وذلك بافتراض أن قيمة الاحتمال P معلومة وتساوي قيمة ثابتة في كل التجارب.

وهنا يمكننا أن نفترض أن جميع القطع متوازنة وإن الاحتمال $P = \frac{1}{2}$. أي أننا سوف لانعتمد على بيانات العينة لتقدير قيمة الاحتمال P، بل نعتبر قيمته معلومة من الافتراض، أو من المجتمع.

ونضع الفرضيات كمايلي:

H_0: إن نتائج التجارب السابقة تتوافق مع التوزيع الثنائي الذي فيه $P = \frac{1}{2}$.

H_1: إن نتائج التجارب السابقة لا تتوافق مع التوزيع الثنائي الذي فيه $P = \frac{1}{2}$ بل تتوافق مع غيره.

وبناء على ذلك يمكننا حساب الاحتمالات النظرية من العلاقة:

$$P_x = C_4^x \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{4-x} = C_4^x \left(\frac{1}{2}\right)^4$$

ولنضع النتائج والحسابات في جدول كالآتي:

i	1	2	3	4	5	$\sum$	
xi	0	1	2	3	4		
ni	1	12	24	18	9	64	
P_i	0,0625	0,25	0,375	0,25	0,0625	1	
$m_i = nP_i$	4	16	24	16	4	64	
$(n_i - m_i)^2$	9	16	0	4	25		
$\frac{(n_i - m_i)^2}{mi}$	2,25	1	0	0,25	6,25	9,75=	χ^2

ومن هذا الجدول نجد أن قيمة χ^2 الفعلية تساوي $\chi^2 = 9,75$

ومن أجل المقارنة نبحث في الجدول الملحق (V) عن قيمة x^2 الجدولية المقابلة لدرجة حرية $k = 4 - 1 = 3$ (لقد بقي عدد درجات الحرية $(k-1)$ لأننا لم نعتمد على بيانات العينة في تقدير الاحتمالP) وبمستوى دلالة $\alpha = 0,05$ فنجد أنها تساوي $\chi^2_{0,053} = 7,81$.

وبالمقارنة نجد أن χ^2 المسحوبة أكبر من قيمة χ^2_α الجدولية، لذلك نرفض فرضية العدم H_0، ونقبل الفرضية H_1 والتي تقول أن توزيع نتائج التجارب السابقة غير خاضع للتوزيع الثنائي الذي فيه $P = \frac{1}{2}$، وهي تخضع لتوزيع احتمالي آخر. ولقد برهنا إن تلك النتائج تتوافق مع التوزيع الثنائي الذي فيه $P = 0,5859$ وهي تميل نحو، أو تتحيز لصالح ظهور الصورة (ابحث عن مكان التحيز من خلال إنشاء مجال ثقة لكل من التكرارات السابقة).

مثال 5: حول المطابقة مع التوزيع الطبيعي:

لنفترض أن دراسة لأعمار المسافرين شملت 100 مسافر، بينت أن التوزيع العمري لهذه العينة كان كما يلي:

i	1	2	3	4	5	0	
الفئة العمرية	[0-10]	[10-20]	[20-30]	[30-40]	[40-50]	[50-60]	Σ
التكرارات المطلقة	5	20	15	45	10	5	100
مركز الفئة	5	15	25	35	45	55	-

والمطلوب بيان فيما إذا كان توزيع أعمار المسافرين يخضع للتوزيع الطبيعي، ثم اختبار ذلك عند مستوى دلالة $\alpha = 0{,}05$.

الحل:

نضع الفرضيات كمايلي:

H_0: إن توزيع أعمار المسافرين يخضع للتوزيع الطبيعي الذي لانعلم توقعه ولا انحرافه المعياري.

H_1: إن توزيع أعمار المسافرين لايخضع للتوزيع الطبيعي.

ولتحديد التوزيع الطبيعي المفترض نقوم بحساب متوسط العينة $\bar{x}$، ونعتبره تقديراً لمتوسط المجتمع μ ، ثم نقوم بحساب تباين العينة s^2 ونعتبره تقديراً لتباين المجتمع σ^2، ونعوضهما في معادلة التوزيع الطبيعي العام فنحصل على الصيغة المحددة للتوزيع الطبيعي المفروض.

وبناء على ذلك نجد أن متوسط العينة يساوي:

$$\bar{x} = \frac{1}{n}\sum n_i x_i = \frac{1}{100}\left[5\cdot 5 + 20\cdot 15 + 15\cdot 25 + 45\cdot 35 + 10\cdot 45 + 5\cdot 55\right]$$

$$\bar{x} = \frac{1}{100}(3000) = 30 \quad \text{عاماً}$$

وأن تباين العينة يساوي:

$$s^2 = \frac{1}{n-1}\sum n_i\left(x_i - \bar{x}\right)^2 = \frac{1}{99}\sum n_i\left(x_i - 30\right)^2\Big)$$

$$s^2 = \frac{1}{99}\begin{bmatrix} 5(5-30)^2 + 20(15-30)^2 + 15(25-30)^2 + \\ + 45(25-30)^2 + 10(25-30)^2 + 5(25-30)^2 \end{bmatrix}$$

$$s^2 = \frac{1}{99}(3125 + 4500 + 375 + 1125 + 1125 + 2250 + 3125)$$

$$s^2 = \frac{1}{99}(14500) = 1465{,}46$$

وبذلك نجد أن الانحراف المعياري للعينة s يساوي:

$$s = \sqrt{146{,}46} = 12{,}1$$

وهكذا يمكننا تقدير متوسط المجتمع وانحرافه المعياري من العلاقتين:

$$\tilde{\mu} = \bar{x} = 30$$

$$\tilde{\sigma} = s = 12{,}1$$

وعندها تأخذ معادلة التوزيع الطبيعي المفترض الشكل الآتي:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{-1}{2}\left(\frac{x-\bar{x}}{\sigma}\right)^2}$$

$$f(x) = \frac{1}{12{,}1\sqrt{2\pi}}e^{-\frac{-1}{2}\left(\frac{x-30}{12{,}1}\right)^2} \qquad (8-6)$$

واعتماداً على هذا التوزيع المفترض نقوم بحساب الاحتمالات النظرية المقابلة للفئات العمرية السابقة فنجد أن:

$$P(0 \le x < 10) = \Phi\left(\frac{10-30}{12{,}1}\right) - \Phi\left(\frac{0-30}{12{,}1}\right)$$
$$= \Phi(-1{,}65) - \Phi(-2{,}48)$$
$$= 1 - \Phi(1{,}65) - [1 - \Phi(2{,}48)]$$
$$= \Phi(2{,}48) - \Phi(1{,}65)$$
$$= 0{,}9934 - 0{,}9505$$
$$= 0{,}0429$$

وكذلك نجد أن:

$$P(10 \le x < 20) = \Phi\left(\frac{20-30}{12{,}1}\right) - \Phi\left(\frac{10-30}{12{,}1}\right)$$
$$= 1 - \Phi(0{,}83) - [1 - \Phi(-1{,}65)]$$
$$= 1 - 0{,}7967 - (1 - 0{,}9505) = 0{,}1538$$
$$P(20 \le x < 30) = \Phi\left(\frac{30-30}{12{,}1}\right) - \Phi\left(\frac{20-30}{12{,}1}\right)$$
$$= 0{,}5 - (1 - 0{,}7967) = 0{,}2967$$
$$P(30 \le x < 40) = \Phi\left(\frac{40-30}{12{,}1}\right) - \Phi\left(\frac{30-30}{12{,}1}\right)$$
$$= 0{,}7967 - 0{,}5 = 0{,}2967$$
$$(40 \le x < 50) = \Phi\left(\frac{50-30}{12{,}1}\right) - \Phi\left(\frac{40-30}{12{,}1}\right)$$
$$= 0{,}9505 - 0{,}7967 = 0{,}1538$$
$$P(50 \le x < 60) = \Phi\left(\frac{60-30}{12{,}1}\right) - \Phi\left(\frac{50-30}{12{,}1}\right)$$
$$= 0{,}9934 - 0{,}9605 = 0{,}0429$$

ولحساب قيمة المؤشر χ^2 نضع تلك الاحتمالات في جدول كالآتي ونضربها بحجم العينة $n = 100$ لنحصل على التكرارات المتوقعة حسب التوزيع المفروض.

i	1	2	3	4	5	6	$\sum$
xi	0-10	10-20	20-30	30-40	40-50	50-60	-
ni	5	20	15	45	10	5	100
P_i	0,0429	0,1538	0,2967	0,2967	0,1538	0,0429	1
$m_i = nP_i$	4,29	15,38	29,67	29,67	15,38	4,29	100
$(n_i - m_i)^2$	0,5041	21,34	215,21	235,0	28,94	0,5041	-
$\frac{(n_i - m_i)^2}{mi}$	0,1175	1,3878	7,253	7,9207	1,8820	0,1175	$18,6785 = \chi^2$

وبذلك نجد أن قيمة χ^2 الفعلية تساوي $\chi^2 = 18,6785$.

ومن الجدول V الملحق نجد أن قيمة χ^2 الجدولية المقابلة لدرجة حرية مساوية لـ $k-2-1=3$ (نقصت درجة بمقدار اثنين لأننا استفدنا من بيانات العينة في تقدير متوسط المجتمع μ وانحرافه المعياري σ) ولمستوى دلالة $\alpha = 0,05$ تساوي $\chi^2_{0,05} = 7,815$ وبالمقارنة نجد أن قيمة χ^2 المسحوبة أكبر من قيمة χ^2_α الجدولية، لذلك نرفض فرضية العدم H_0 التي تقول أن توزيع عمر المسافر لا يخضع للتوزيع الطبيعي الذي قدرنا متوسطه بـ 30 عاماً، وانحرافه المعياري بـ 12,1. ومع ذلك فهو قد يكون خاضعاً لتوزيع طبيعي آخر بمتوسط آخر وانحراف معياري آخر.

6-5: اختبار استقلالية متحولين بواسطة χ^2:

إن دراسة استقلالية متحولين كميين، أو نوعيين بوساطة χ^2 تقتضي أن يأخذ كل منهما عدة حالات مرتبة، أو مبوبة (في جدول مزدوج يسمى جدول التوافق من المرتبة $k \times \ell$)، وهو يتضمن التكرارات المطلقة المقابلة لكل حالة مزدوجة (كلية) لهما. ولتوضيح ذلك نفترض أنه لدينا متحولان X, Y، وأن المتحول X يأخذ K حالة متنافية هي:

$$X : (x_1 \quad x_2 \quad x_3 \quad x_i \ldots\ldots x_k)$$

وأن المتحول Y يأخذ ℓ حالة متنافية هي:

$$Y:\left(y_1 \quad y_2 \quad y_3 \quad y_j.....x_\ell\right)$$

وبما أن المؤشر χ^2 يعرف على التكرارات الفعلية والتكرارات المتوقعة (الفرضية)، فإننا نرمز للتكرارات الفعلية المقابلة للحالة المزدوجة (x_i, y_i) بالرمز n_{ij}، وللتكرارات المتوقعة المقابلة لنفس الحالة (x_i, y_i) بالرمز m_{ij}. ولنفترض أن هذه التكرارات موجودة في الجدول المزدوج كمايلي:

حالات X حالات Y	x_1	x_2	x_3	.....	x_i	.....	x_k	$\sum$
Y_1	n_{11} m_{11}	n_{21} m_{21}	n_{31} m_{31}	.....	n_{i1} m_{i1}	.....	n_{k1} m_{k1}	n'_1 m'_1
Y_1	n_{12} m_{12}	n_{22} m_{22}	n_{32} m_{32}	.....	n_{i2} m_{i2}	.....	n_{k2} m_{k2}	n'_2 m'_2
.....	.....	.....	.....	.....	.....	.....	.....	.....
Y_j	n_{1j} m_{1j}	n_{2j} m_{2j}	n_{3j} m_{3j}	.....	n_{ij} m_{ij}	.....	n_{kj} m_{kj}	n'_j m'_j
.....	.....	.....	.....	.....	.....	.....	.....	.....
Y_ℓ	$n_{1\ell}$ $m_{1\ell}$	$n_{2\ell}$ $m_{3\ell}$	$n_{3\ell}$ $m_{3\ell}$	.....	$n_{i\ell}$ $m_{i\ell}$	.....	$n_{k\ell}$ $m_{k\ell}$	n'_ℓ m'_ℓ
$\sum$	n_1	n_2	n_3	.....	n_i	.....	n_k	n

وهنا نلاحظ أنه قد رمزنا لمجموع التكرارات في العمود i بالرمز n_i، ولمجموع التكرارات في السطر j بالرمز n'_j

أي أن:

$$n_i = \sum_{j=1}^{\ell} n_{ij} \qquad ، \qquad n'_j = \sum_{i=1}^{k} n_{ij}$$

ورمزنا لمجموع التكرارات في جميع الخلايا (الأسطر والأعمدة) بـ n.

أي أن:

$$n = \sum_{j=1}^{\ell} \sum_{i=1}^{k} n_{ij} = \sum_{i=1}^{k} n_i = \sum_{j=1}^{\ell} n'_j$$

ولحساب التكرارات المتوقعة نستخدم طريقة التناسب، التي تقتضي أن يكون عدد التكرارات المتوقعة في الخلية (1، 1) جزءاً من المجموع n'_1 متناسباً مع حصتها من مجموع التكرارات n؛ أي أنه يجب أن يكون متناسباً مع الكسر $\left(\frac{n_1}{n}\right)$، أو بشكل عام مع الكسر $\frac{n_i}{n}$. وبذلك نجد أن التكرار المتوقع m_{11} يحسب من العلاقة:

$$m_{11} = \frac{n_1}{n} \cdot n'_1$$

ويمكن حسابها كجزء من n_1، متناسب مع حصتها $\frac{n'_1}{n}$ فيكون $m_{11} = \frac{n'_1}{n} \cdot n_1$

وفي كلتا الحالتين يكون لدينا:

$$m_{11} = \frac{n_1 \cdot n'_1}{n} = \frac{\text{مجموع السطر الأول} \times \text{مجموع العمود الأول}}{\text{المجموع الكلي n}} \qquad (10-6)$$

وبصورة عامة تحسب التكرارات المتوقعة في جميع خلايا الجدول المزدوج من العلاقة:

$$m_{ij} = \frac{n_i \cdot n'_j}{n} = = \frac{\text{المجموع في السطر j} \times \text{المجموع في العمود i}}{\text{المجموع الكلي للتكرارات}} \qquad (10-6)$$

وأخيراً نجد أن المؤشر χ^2 يعرف بالعلاقة الآتية:

$$\chi^2 = \sum_{j=1}^{\ell} \sum_{i=1}^{k} \left[\frac{(n_{ij} - m_{ij})^2}{m_{ij}} \right] \qquad (11-6)$$

ولقد أشرنا إلى أن المؤشر χ^2 يخضع للتوزيع الاحتمالي $\chi^2(x)$ بدرجة حرية تساوي $v = (k-1)(\ell-1)$.

6-6: شروط تطبيق المؤشر χ^2:

لا يجوز تطبيق المؤشر χ^2 إلا ضمن الشروط التالية:

1 . فرضية العدم، أو الاستقلال تكون كما يلي: إن المتحولين المدروسين $(x \ , \ y)$ مستقلين عن بعضهما البعض، أي عدم وجود علاقة بين المتحولين المدروسين.

2 . الفرضية البديلة تكون كما يلي: إن المتحولين المدروسين غير مستقلين وبالتالي فهما مرتبطان إحصائياً.

3 . أن تكون العينة المسحوبة للدراسة هي عينة عشوائية وأن لايقل حجمها عن 50 وحدة.

4 . أن لا تكون قيمة أي من قيم التكرارات المتوقعة أقل من 5، لأن ذلك يؤثر على عملية التقدير، ويمكن معالجة هذه الحالة عند ظهورها بأن نقوم بدمج عمودين متجاورين، أو سطرين متجاورين حتى نتخلص من تلك التكرارات القليلة. ويمكن أن نتساهل بهذا الشرط واستبداله بشرط آخر، وهو أن لايزيد عدد الخلايا التي يكون تكراراها المتوقع أقل من 3 عن $\frac{1}{5}$ عدد الخلايا.

5 . أن يتم اتخاذ القرار حول استقلال، أو عدم استقلال المتحولين المدروسين بواسطة مقارنة القيمة المحسوبة للمعيار χ^2 مع القيمة الجدولية لمتحول التوزيع χ^2_α المقابلة لمستوى الدلالة α ولدرجة حرية $v = (k-1)(\ell-1)$ ويتم اتخاذ القرار كما يلي:

- إذا كانت القيمة المحسوبة χ^2 أكبر من القيمة الجدولية χ^2_α نرفض فرضية العدم التي تقول باستقلال المتحولين، ونقبل الفرضية البديلة التي تقول بوجود علاقة ارتباطية بينهما وكلما كان الفرق كبيراً كانت متانة الارتباط كبيرة.
- إذا كانت القيمة المحسوبة χ^2 أصغر من القيمة الجدولية χ^2_α نقبل فرضية العدم ونعتبر المتحولين مستقلين عن بعضهما البعض أي أنهما غير مرتبطين.

. ولتسهيل عمليات التطبيق نورد جدولاً مصغراً لقيم χ^2_α الجدولية المقابلة لمستوى دلالة $\alpha = 0{,}05$ والمقابلة لعدة درجات حرية.

جدول: بعض قيم χ^2 عند مستوى دلالة $\alpha = 0{,}05$

درجة الحرية d.f.	قيمة χ^2_α الجدولية المقابلة لدرجة الحرية
1	3,84
2	5,99
3	7,81
4	9,49
5	11,07
6	12,59
7	14,07
8	15,51
9	16,92
10	18,31
11	19,68
12	21,03
13	22,36
14	23,68

مثال 6:

لنفترض أننا قمنا بتبويب بيانات عينة عشوائية عن مبيعات أنواع السيارات ومهن المشترين لها فحصلنا على الجدول التكراري التالي:

جدول: بيانات المثال بـ n_{ij} المقابلة لخلايا المزدوجة:

أنواع السيارات / مهنة المشتري	فخمة	متوسطة	عادية	خفيفة	مجموع التكرارات الهامشية	التكرارات النسبية الهامشية
رجل أعمال	10	8	6	4	28	$\frac{28}{160}$
موظف	15	20	12	5	52	$\frac{52}{160}$
تاجر	20	15	10	2	47	$\frac{47}{160}$
مدرس	5	10	15	3	33	$\frac{33}{160}$
مجموع التكرارات الهامشية (للأعمدة)	50	53	43	14	160	1

والمطلوب : دراسة ارتباط (أو استقلال) مهنة المشتري بنوع السيارة.

الحل:

إن حل هذه المسألة يعتمد على فرضية الاستقلال والتي تعاكس الارتباط والتي تكتب كمايلي:

فرضية العدم: عدم وجود ارتباط بين مهنة المشتري ونوع السيارة التي يشتريها؛ أي أنهما مستقلان .

الفرضية البديلة: يوجد ارتباط بين مهنة المشتري ونوع السيارة التي يشتريها.

فإذا كانت مهنة المشتري مستقلة عن نوع السيارة التي اشتراها، فإن ذلك يعني أن التكرارات الداخلية في كل عمود يجب أن تكون متناسبة مع التكرارات النسبية الهامشية المقابلة لحالات المهن.

أي أنه إذا كانت مبيعات السيارات الفخمة مستقلة عن مهنة المشتري فإن إجمالي حجم مبيعاتها وهو (50) يجب أن يوزع على أصحاب المهن بشكل يتناسب مع نسبة كل مهنة في الدراسة، وهذا يعني أنه إذا كان حجم المبيعات مستقلاً عن المهن فإننا نتوقع أن تكون التكرارات كمايلي:

. أن يكون عدد السيارات الفخمة التي يشتريها العمال:

$$50 \times \frac{28}{160} = 8{,}75 \text{ سيارة فخمة}$$

. أن يكون عدد السيارات الفخمة المتوقعة للموظفين:

$$50 \times \frac{52}{160} = 16{,}25 \text{ سيارة فخمة}$$

. أن يكون عدد السيارات الفخمة المتوقعة للتجار:

$$50 \times \frac{47}{160} = 14{,}69 \text{ سيارة فخمة}$$

. أن يكون عدد السيارات الفخمة المتوقعة للمدرسين:

$$50 \times \frac{33}{160} = 10{,}31 \text{ سيارة فخمة}$$

وبإتباع نفس الطريقة نقوم بحساب التكرارات المتوقعة لأنواع السيارات الأخرى، وذلك بتطبيق العلاقة العامة التالية:

$$m_{ij} = \frac{n_i \cdot n'_j}{n} \qquad (12-6)$$

حيث إن n_i هي المجاميع العمودية للتكرارات، n'_j هي المجاميع السطرية للتكرارات، و n المجموع الكلي لها. فنحصل على الجدول التالي:

جدول: التكرارات المتوقعة لخلايا الجدول m_{ij}

أنواع السيارات / مهنة المشتري	فخمة	متوسطة	عادية	خفيفة	المجموع n'_j
رجل أعمال	8,75	9,28	7,52	2,45	28
موظف	16,25	17,23	13,98	4,55	52
تاجر	14,69	15,57	12,63	4,11	47
مدرس	10,31	10,93	8,87	2,89	33
مجموع التكرارات الهامشية n_i	50	53	43	14	160

وهنا يجب أن نشير إلى أن المجاميع الهامشية للأسطر والأعمدة للتكرارات المتوقعة يجب أن تساوي مقابلاتها في التكرارات الفعلية السابقة.

قبل كل شيء نلاحظ أن جميع التكرارات المتوقعة في عمود السيارات الخفيفة أقل من 5، وهذا خلل في أحد شروط تطبيق المعيار χ^2 (الشرط 4)، لذلك نقوم بدمج عمود السيارات الخفيفة مع العمود المجاور له وهو عمود السيارات العادية، ونجمع التكرارات الفعلية لهما، كما نجمع التكرارات المتوقعة لهما، ونضع النتائج في جدول موحد، فنحصل على الجدول التالي:

جدول التكرارات الفعلية والمتوقعة بعد الدمج

أنواع السيارات / مهنة المشتري	**فخمة**	**متوسطة**	**عادية وخفيفة**	**المجموع n'_j**
رجل أعمال	10	8	10	28
	8,75	9,28	9,79	
موظف	15	20	17	52
	16,25	17,23	18,53	
تاجر	20	15	12	47
	14,69	15,57	16,74	
مدرس	5	10	18	33
	10,31	10,93	11,76	
المجموع n_i	50	53	57	160

وهكذا يصبح الجدول محققاً لجميع شروط تطبيق المعيار χ^2 لذلك نقوم بحساب الحدود المختلفة للتركيب $\frac{(n_{ij}-m_{ij})^2}{m_{ij}}$ ونضعها في الجدول التالي:

جدول: قيم الحدود $\frac{(n_{ij}-m_{ij})^2}{m_{ij}}$

أنواع السيارات / مهنة المشتري	فخمة	متوسطة	عادية وخفيفة	المجموع
رجل أعمال	0,17857	0,17655	0,00009	
موظف	0,09615	0,44532	0,12633	
تاجر	1,91941	0,020867	1,34215	
مدرس	2,73483	0,07913	3,31102	
مجموع التكرارات الهامشية	4,92896	0,72187	4,77959	$10,43043=x^2$

ومن الجدول السابق نجد أن قيمة χ^2 المحسوبة $\chi^2=10,43042$

ومن الجدول (V) نجد أن قيمة χ^2 الجدولية المقابلة لمستوى دلالة $\alpha=0,05$ ودرجة $v=(k-1)(\ell-1)=(4-1)(3-1)=6$ تساوي $\chi^2_{\alpha 6}=12,59$. وبمقارنة هاتين القيمتين نجد أن قيمة χ^2 المحسوبة أصغر من $\chi^2_{\alpha 6}$، لذلك نقبل فرضية العدم، والتي تقول بأن المتحولين مستقلان، ونستنتج أن مهنة المشتري لاترتبط بنوع السيارة التي يشتريها.

6-7 : حالة خاصة اختبار χ^2 عندما يكون لدينا جدول 2 × 2:

عندما يكون جدول التوافق من الدرجة 2 ×2 ويكون حجم العينة n كبيراً فإن جدول التوافق يأخذ الشكل التالي:

جدول : حالة جدول 2 × 2

Yالمتحول / X المتحول	Y1حالة أولى لـ	Y2حالة ثانية لـ	المجموع
X1حالة أولى	a	b	a+b
X2 حالة ثانية	c	d	c+d
المجموع	a+c	b+d	n

فإن صيغة المعيار χ^2 تأخذ شكلاً مبسطاً هو الشكل التالي:

$$\chi^2 = \frac{n(a \cdot d - b \cdot c)^2}{(a+b)(c+d)(a+c)(b+d)} \qquad (13-6)$$

وهو يخضع للتوزيع χ^2 ذي درجة حرية واحدة، لأن عدد درجات الحرية يصبح مساوياً لـ: $v = (k-1)(\ell-1) = (2-1)(2-1) = 1$.

مثال 7:

لدراسة استقلالية نوع الجنس عن نتيجة الامتحان في مقرر الإحصاء سحبنا عينة عشوائية مؤلفة من 145 طالباً فكانت حالاتها وتكراراتها كما في الجدول التالي:

جدول: بيانات العينة

نتيجة الامتحان / نوع الجنس	ناجح	راسب	$\sum$
ذكر	50	40	90
أنثى	30	25	55
$\sum$	80	65	145

وبما أن العينة كبيرة يمكننا مباشرة تطبيق العلاقة السابقة فنجد أن:

$$\chi^2 = \frac{145(50 \cdot 25 - 40 \cdot 30)^2}{90 \cdot 55 \cdot 80 \cdot 65} = 0{,}014$$

وهي قيمة أصغر بكثير من القيمة الجدولية $\chi^2_{\alpha,6} = 3{,}84$ المقابلة لدرجة $v = 1$ ، وبذلك نستنتج مباشرة أن فرضية العدم مقبولة، وأن المتحولين المدروسين مستقلان، أي أن نتيجة الامتحان مستقلة عن نوع الجنس.

*** تصحيح الحالة الخاصة (تصحيح ييتس):**

إذا كان حجم العينة n صغيراً، وكانت الفرضية تقتضي الاستمرار للمتحولات المدروسة، فإن قيمة χ^2 المحسوبة من العلاقة السابقة (6−13) في حالة الجداول من المرتبة 2× 2، تعطينا قيمة متحيزة وأكبر من القيم الحقيقية χ^2 . وللتخلص من ذلك التحيز قدم العالم (ييتس) تصحيحاً لذلك التحيز بإضافة $\frac{1}{2}$ إلى التكرار الذي يقل عن التكرار المتوقع، وبطرح $\frac{1}{2}$ من التكرار الذي يزيد عن التكرار المتوقع . أي بإضافة $\frac{1}{2}$ إلى كل من a ، d وطرح $\frac{1}{2}$ من كل من b ، c، أو بالعكس، وبعد الإصلاح والمعالجة والتعويض في العلاقة السابقة نحصل على العلاقة المصححة الآتية:

$$\chi^2_M = \frac{n\left[\left|a\cdot d - b\cdot c\right| - \frac{n}{2}\right]^2}{(a+b)(c+d)(a+c)(b+d)} \qquad (14-6)$$

وعند الحساب نطرح من القيمة المطلقة للمقدار التالي $\left|a\cdot d - b\cdot c\right|$ المقدار $\frac{n}{2}$ ثم نربع الناتج ونضربه بـ n ونقسمه على الجداءات المذكورة.

مثال 8:

لنحسب قيمة χ^2_M المصححة لمعطيات المثال السابق فنجد أن:

$$x^2_M = \frac{145\left[\left|50\cdot 25 - 40\cdot 30\right| - \frac{145}{2}\right]^2}{90\cdot 55\cdot 80\cdot 65} = 0{,}00285$$

وهي قيمة أصغر من القيمة الأولى لـ χ^2 قبل التعديل. وبما أنها أصغر من القيمة الجدولية $\chi^2_\alpha = 3{,}84$ لذلك نرفض فرضية العدم أيضاً.

6−8 : معامل التوافق C:

لقد قام العالم (كارل بيرسون) بالاستفادة من قيمة χ^2 لتعريف معامل جديد لدراسة الاستقلال، أو الارتباط سماه معامل التوافق، أو الارتباط ويعرف بالعلاقة:

$$C = \sqrt{\frac{\chi^2}{n + \chi^2}}$$

وكلما كانت قيمة C قريبة من الواحد كان التوافق جيداً، وكان الارتباط متيناً، ولكن قيمته تبقى أقل من (1).

مثال 9:

لنقم بحساب معامل التوافق C لبيانات المثال الأسبق (7) فنجد أن:

$$C = \sqrt{\frac{10,43}{160 + 10,43}} = \sqrt{0,061} = 0,247$$

وهي قيمة صغيرة تدل على أن التوافق، أو الارتباط بين مهنة المشتري ونوع السيارة ضعيف جداً.

6-9: معامل التوافق T^2:

ولقد استنبطه العالم (تشبيرو) اعتماداً على المعامل C وتوصل إلى العلاقة التالية:

$$T^2 = \frac{C^2}{(1 - C^2)\sqrt{(k-1)(\ell - 1)}}$$

حيث k عدد الأعمدة و ℓ عدد الأسطر في جدول التوافق لحساب χ^2.

مثال 10:

لنقم بحساب معامل التوافق T^2 للمثال (9) المعتمد على المثال (7) السابق فنجد أن:

$$T^2 = \frac{(0,247)^2}{\left[(1 - (0,247)^2)\right]\sqrt{(4-1)(3-1)}} = \frac{0,061}{(0,567)(2,449)} = 0,0439$$

وهي أيضاً قيمة صغيرة لاتدل على وجود توافق، أو ارتباط بين مهنة المشتري ونوع السيارة.

6-10 : اختبار تباين مجتمع طبيعي بواسطة χ^2:

لقد أشرنا سابقاً إلى أن المقدار:

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

يخضع للتوزيع (x) $\chi^2_{(n-1)}$ ذي الـ $n-1$ درجة حرية.

ولهذا فإنه عند إجراء اختبار أية فرضية حول σ^2 كالفرضية $H_0 : \sigma^2 = \sigma_0^2$، فإننا نستخدم مؤشر الاختبار التالي:

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} \qquad (15-6)$$

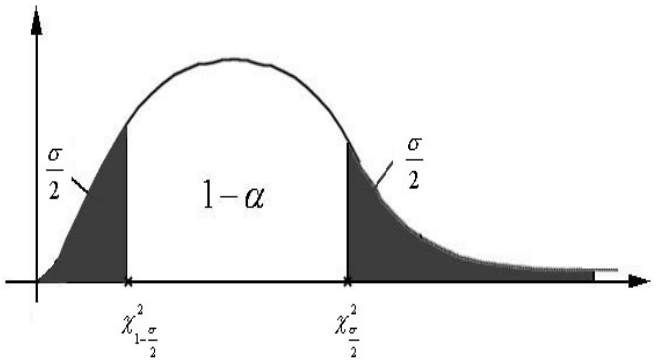
حيث إن s^2 هو تباين العينة ذات الحجم n.

ولتحديد نتيجة الاختبار نميز بين الحالتين التاليتين:

ـ إذا كان الاختبار ثنائي الجانب، فإننا نقارن قيمة المؤشر χ^2 بالقيمتين الجدوليتين $\chi^2_{\left(1-\frac{a}{2},n-1\right)}$ و $\chi^2_{\left(\frac{a}{2},n-1\right)}$. فإذا كانت قيمة χ^2 المحسوبة تحقق المتراجحة المزدوجة:

$$\chi^2_{\left(1+\frac{a}{2},n-1\right)} < \chi^2 < \chi^2_{\left(\frac{a}{2},n-1\right)}$$

فإننا نقبل الفرضية H_0. ونرفضها في حالة العكس، كما في الشكل الآتي:

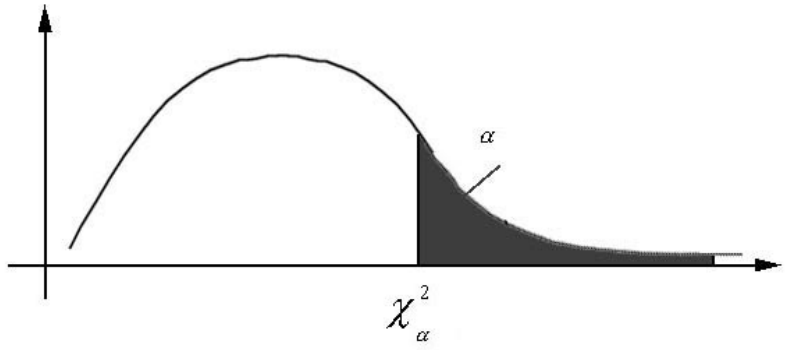


الشكل رقم (6-4)

ـ إذا كان الاختبار أحادي الجانب، فإننا نقارن قيمة المؤشر χ^2 مع القيمة الجدولية $\chi^2_{(a,n-1)}$. فإذا كانت:

$\chi^2 < \chi^2_{(a,n-1)}$

فإننا نقبل الفرضية H_0. ونرفضها في حالة العكس، كما في الشكل الآتي:



الشكل رقم (6-5)

مثال 11:

يدّعي أحدهم أن الانحراف المعياري لمبيعاته اليومية يساوي $\sigma = 10$، ولكن عينة مؤلفة من مبيعات عشرة أيام أعطتنا متوسطاً وتبايناً يساويان:

$\bar{x} = 312 \quad , \quad s^2 = 192$

فهل تقدم هذه المعلومات دلالة كافية على قبول ذلك الادعاء؟

اختبر بمستوى دلالة $\alpha = 0{,}05$.

الحل:

نضع الفرضيتين الابتدائية والبديلة كما يلي:

$H_0 : \sigma^2 = (10)^2 = \sigma_0^2$ (نفترض أن الادعاء صحيح).

$H_1 : \sigma^2 > (10)^2$ (لأننا نهتم بالحالة التي فيها $\sigma^2 > \sigma_0^2$)

ثم نحسب مؤشر الاختبار فنجد أن:

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{(10-1)192}{100} = 17{,}28$$

ومن الفرضية H_1 نلاحظ أن الاختبار أحادي الجانب، لذلك نبحث عن القيمة الجدولية لـ $\chi^2_{(a,n-1)}$ فنجد أن:

$$\chi^2_{(a,n-1)} = \chi^2_{(0,05\ ,\ 9)} = 16,92$$

وبالمقارنة نجد أن $\chi^2 > \chi^2_{\alpha}$ لأن:

$$17,28 > 16,92$$

لذلك نرفض الفرضية H_0، ونعتبر الادعاء باطلاً. ويمكننا متابعة الحل وإيجاد مجال ثقة لـ σ^2 باحتمال $0,90$، وبحيث نترك احتمالاً على كل من اليمين واليسار مساوياً $\frac{a}{2}$ وذلك باستخدام العلاقة التالية:

$$P = \left[\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}} \right] = 1 - \alpha = 0,90$$

ومن الجداول الملحقة نجد أن:

$$\chi^2_{\left(\frac{\alpha}{2},n-1\right)} = \chi^2_{(0,05\ ,\ 2)} = 16,92$$

$$\chi^2_{\left(1-\frac{\alpha}{2},n-1\right)} = \chi^2_{(0,95\ ,\ 2)} = 3,33$$

وبذلك نجد أن مجال الثقة المطلوب هو:

$$P = \left[\frac{9 \cdot 195}{16,9} < \sigma^2 < \frac{9 \cdot 195}{3,33} \right] = 0,90$$

$$P = \left[103,7 < \sigma^2 < 527,82 \right] = 0,90$$

وهو مجال لا يحوي القيمة المدعاة $\sigma^2 = 100$ وهذا يؤكد على رفض فرضية العدم H_0.

مثال 12:

يعتقد عامل أن الانحراف المعياري لقياسات جهازه $\sigma = 2$. وعندما قمنا بثلاث تجارب لقياس شيء واحد حصلنا على القياسات التالية:

$$10,2 \quad , \quad 4,1 \quad , \quad 5,2$$

فهل تتناقض هذه القياسات مع اعتقاد العامل؟

اختبر بمستوى دلالة $\alpha = 0,10$.

الحل:

نضع الفرضيتين الابتدائية والبديلة كما يلي:

$H_0 : \sigma^2 = 4$ (أي نفترض أن اعتقاد العامل صحيح).

$H_1 : \sigma^2 \neq 4$

ولحساب قيمة مؤشر الاختبار نقوم بحساب تباين العينة، فنجد أنه يساوي $S^2 = 10,57$، وبذلك نحصل على أن:

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{2(10,57)}{4} = 5,29$$

ومن شكل H_1 نلاحظ أن الاختبار ثنائي الجانب، لذلك نبحث في جداول $\chi^2_{(x)}$ عن قيمة كل من: $\chi^2_{\left(1-\frac{\alpha}{2}\right)}$ و $\chi^2_{\left(\frac{\alpha}{2}\right)}$ المقابلتين لـ $n-1 = 3-1 = 2$ درجة حرية، فنجد أن:

$$\chi^2_{0,051\cdot 2} = 5,99 \quad , \quad \chi^2_{0,95\cdot 2} = 0,10$$

وبالمقارنة نجد أن قيمة χ^2 المحسوبة تحقق المتراجحة $0,10 \leq 5,29 < 5,99$، لذلك نقبل الفرضية H_0 ونعتبر اعتقاد العامل صحيحاً.

6-11: اختبار تناظر الجداول المربعة بواسطة χ^2:

لنفترض أنه لدينا متحولان XوY، لهما عدد الحالات، أو القيم نفسه، ويساويk، وبذلك فإن جدول التوافق لهذه الحالة يكون من المرتبة $(k \times k)$، ويأخذ الشكل التالي:

حالات X	x_1	x_2	x_3	....	x_i	....	..x_k...	Σ

Y حالات								
y_1	n_{11}	n_{21}	n_{31}	....	x_{i1}	....	n_{k1}	n_1^1
y_2	n_{12}	n_{22}	n_{32}	....	n_{i2}	....	n_{k2}	n_2^1
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
y_i	n_{1j}	n_{2j}	n_3	....	n_{ij}	....	n_{kj}	n_j^1
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
y_k	n_{1k}	n_{2k}	n_{3k}	....	n_{ik}	....	n_{ki}	n_k^1
Σ	n_1	n_2	n_3	....	n_i	....	n_k	n

وهنا نضع الفرضيات كما يلي:

H_0: إن التكرارات n_{ij} = التكرارات n_{ji} (بشكل عام) $i \neq j$

H_1: إن التكرارات $n_{ij} \neq$ التكرارات n_{ji}.

ولاختبار صحة، أو عدم صحة الفرضية H_0، نستخدم المؤشر χ^2 المعرف على نصف الجدول بالعلاقة:

$$\chi_i^2 = \sum_{j=1}^{k-1} \sum_{i>j}^{r} \frac{\left(n_{ij} - n_{ji}\right)^2}{n_{ij} + n_{ji}}$$

وهو مقدار يخضع للتوزيع $\chi^2_{(x)}$ بدرجة حرية تساوي $\frac{k(k-1)}{2}$ درجة حرية.

وهنا نشير إلى أنه يمكننا اعتماد أحد القطرين (الرئيسي، أو الثانوي) لقياس عملية التناظر بالنسبة إليه. ولتحديد أي من القطرين يصلح لقياس عملية التناظر، نقوم بحساب مجموع عناصر التكرارات على كل منهما، ثم اختيار القطر ذي المجموع الأكبر. وسنوضح ذلك من خلال المثال الآتي:

مثال 13:

في دراسة مقارنة لقوة النظر في العين اليمنى مع العين اليسرى لدى 100 طفل، حصلنا على الجدول التالي:

قوة العين اليمنى / قوة العين اليسرى	4	6	8	10	Σ
4	0	10	16	15	41
6	4	2	10	4	20
8	12	4	3	6	25
10	8	4	1	1	14
Σ	24	20	30	26	100

والمطلوب دراسة تناظر هذه المعطيات واختبار تناظرها بمستوى دلالة $\alpha = 0,05$.

الحل:

نلاحظ أن مجموع عناصر القطر الرئيسي = 0+2+3+1=6

وأن مجموع عناصر القطر الثانوي = 8+4+10+15=37.

وبما أن مجموع عناصر القطر الثانوي أكبر من عناصر القطر الرئيسي، لذلك فإننا نعتمد القطر الثانوي مقياساً للتناظر، وبناءً على ذلك نحسب قيمة x^2 فنجد أن:

$$\chi_c^2 = \frac{(12-4)^2}{12+4} + \frac{(4-1)^2}{4+1} + \frac{(0-1)^2}{0+1} + \frac{(2-3)^2}{2+3} + \frac{(10-6)^2}{10+6} + \frac{(16-4)^2}{16+4} =$$

$$\chi_c^2 = 15,2$$

ولمقارنة هذه القيمة مع القيمة الجدولية نلاحظ أن عدد درجات الحرية يساوي:

$$v = \frac{k(k-1)}{2} = \frac{4 \cdot (4-1)}{2} = 6$$ درجة حرية

ومن الجداول نجد أن القيمة الجدولية تساوي:

$$\chi^2_{0,05} = 12,59$$

وبالمقارنة نجد أن قيمة χ^2 المحسوبة أكبر من قيمة χ^2_α الجدولية، لذلك نرفض فرضية العدم، ونقول أن قوة العينين غير متناظرة بالنسبة للقطر الثانوي وذلك باحتمال ثقة قدره 0,95.

تمرينات

1 . إذا كانت علامة الطلاب في مقرر الإحصاء مبوبة كما يلي:

الفئات	[0 – 20[	[20 – 30[	[30 – 40[	[40 – 50[	[50 – 60[	[60 – 70[	[70 – 80[	[80 – 90[	[90 – 100[
التكرارات	5	10	15	20	40	50	30	20	10

والمطلوب: اختبار فيما إذا كانت هذه العلامات خاضعة للتوزيع الطبيعي وذلك بمستوى دلالة $\alpha = 0,05$.

2 . إذا كانت النسب المئوية للزوجات المتعطلات عن العمل حسب العمر موزعة كما يلي:

العمر	[15 – 20[	[20 – 25[	[25 – 30[	[30 – 35[	[35 – 40[	[40 – 45[	[45 – 50[	[50 – 55[	55 +	
النسبة المئوية	2	15	29	25	14	8	2	2	3	100

والمطلوب: اختبار فيما إذا كان هذا التوزيع يخضع للتوزيع الطبيعي وذلك بمستوى دلالة $\alpha = 0,10$.

3 . إذا كانت التكرارات المتقاطعة لدرجتي تعليم الرجل وزوجته كما يلي:

حالة تعليم الزوج / حالة تعليم الزوجة	أمي	ابتدائية	إعدادية	ثانوية	جامعية	عليا
أمية	25	15	10	8	7	5
ابتدائية	10	20	25	15	8	9
إعدادية	8	15	20	35	15	10
ثانوية	7	8	10	30	30	20
جامعية	6	7	8	20	25	30
عليا	5	8	10	15	20	25

والمطلوب: اختبار فيما إذا كانت درجتا تعليم الزوجين مستقلتين عن بعضهما وذلك بمستوى دلالة $\alpha = 0,05$. ثم احسب معامل التوافق C وكذلك معامل التوافق T^2، واستنتج متانة العلاقة بين الدرجتين.

4 . إذا كان عدد المتعطلين عن العمل في الريف والحضر حسب الجنس موزعاً كما يلي (بالآلاف):

المكان / الجنس	الحضر	الريف	المجموع
الذكور	200	150	350
الإناث	120	130	250
المجموع	320	280	600

والمطلوب: دراسة فيما إذا كان هناك اقتران بين المكان والجنس لعدد المتعطلين، أم إنهما مستقلين عن بعضهما. اختبر ذلك بمستوى دلالة $\alpha = 0,05$. ثم احسب معامل التوافق C وكذلك معامل التوافق T^2.

5 . يقول أحدهم أن تباين درجة الحرارة بين الليل والنهار في مدينته لا يتجاوز 25 درجة. والمطلوب اختبار هذا القول بمستوى دلالة $\alpha = 0,10$.

6 . لنفترض أن التكرارات المتقاطعة لألوان عيون الأبناء وأمهاتهم كان كما يلي:

لون عيون الأم / لون عيوب الأبن	أسود	عسلي	زرقاوي	خضراوي
أسود	20	10	8	7
عسلي	10	20	10	8
زرقاوي	8	7	30	20
خضراوي	7	8	20	15

والمطلوب دراسة هذه التكرارات وذلك بمستوى دلالة $\alpha = 0,05$.

الفصل السابع

الاختبارات اللامعلمية من عينة واحدة

يتضمن الفقرات التالية:

1 . مقدمة عامة.

2 . الاختبارات اللامعلمية المعرفة على عينة واحدة.

. اختبار ثنائي الحد.

. اختبار التعاقب.

. اختبار الاقتران.

. اختبار التوافق.

. اختبار التطابق.

. اختبار الارتباط.

3 . الاختبارات اللامعلمية للمتحولات الرتيبة والمجالية.

. اختبارات تطابق التكرارات(χ^2 +كولموغوروف–سميرنوف+ ليليفورز).

. اختبارات التناظر حول الوسيط (اختبار الإشارة + ويلكوكس)

4 . اختبارات الارتباط بين متحولين رتيبين.

. معامل الارتباط الرتبي (معامل سبيرمان).

. معامل كيندال للارتباط الرتبي (من النوع C).

5 . اختبارات الارتباط بين متحول أسمي وآخر رتبي.

. معامل (كوريتون).

. معامل ارتباطية نقطة السلسلة المزدوجة.

الفصل السابع

الاختبارات اللامعلمية من عينة واحدة

7-1 : مقدمة عامة:

لقد لاحظنا في الفصلين السابقين أن اختبارات الفرضيات هي كواشف حساسة لصحة، أو عدم صحة فرضية العدم H_0. ولكن تلك الاختبارات لا تتأثر فقط بوضع الفرضية H_0 نفسها، بل تتأثر بتحقق أو عدم تحقق الشروط المرافقة للاختبار المفروض، وكلما كانت تلك الشروط محققة، كان الاختبار أكثر فاعليةً وقوةً. ولقد رأينا سابقاً أن الاختبارات التي استخدمناها في التحقق من الفرضيات الموضوعة حول متوسط أو إجمالي المجتمع، أو حول النسبة فيه..إلخ، تشترط أن تكون الخاصة المدروسة في المجتمع خاضعة للتوزيع الطبيعي، كما اشترطنا أن تكون العينات عشوائية، أو أن تكون المجتمعات متجانسة، أو أن تكون الخاصة قابلة للقياس..إلخ. ولذلك يجب التأكد من تحقق مثل هذه الشروط (أو مثيلاتها) في البيانات المدروسة قبل تطبيق مؤشر الاختبار عليها واستخلاص النتائج الممكنة منها.

وفي حالة عدم تحقق مثل هذه الشروط أو بعضها، فإن الاختبار يفقد كثيراً من الثقة فيه ويصعب الاعتماد على نتائجه. وبعبارةٍ أخرى، فإن تطبيق الاختبارات الإحصائية على الحالات التي لا تكون فيها الشروط غير محققة يعتبر أمراً خطيراً للغاية، وذلك ليس بفعل البيانات الإحصائية، بل بسبب عدم تحقق أحد، أو بعض الشروط المرافقة للاختبار.

وإضافةً إلى عدم تحقق بعض الشروط السابقة، يواجه الإحصائي في عمله الكثير من المتحولات غير القابلة للقياس الكمّي مثل النوع، والجنس، ودرجة التعليم، ومكان السكن، والولادة، والحالة الزواجية، والحالة التعليمية...إلخ.

ويحتاج التعامل معها إلى أساليب جديدة تتناسب مع تلك المتحولات، ومع نوع البيانات المتوفرة عنها، لذلك تم تصنيف المتحولات إلى صنفين أساسيين هما:

1 . المتحولات النوعية:

وهي المتحولات التي لا يمكن وصفها إلا بواسطة حالات معنية مختلفة بعضها عن بعض بصفة معينة، أو بدرجة محددة، مثل: الجنس، العرق، درجة التعليم، الحالة الزواجية، الحالية العملية، الحالة الصحية ...إلخ. وبتطبيق مبادئ نظرية القياس على هذه المتحولات تم تصنيف المتحولات النوعية حسب سلم القياس فيها إلى مستويين هما:

ـ السلم الاسمي Nominal scale: وهو سلم نوعي ومن أضعف مستويات القياس المستخدمة في الإحصاء، وفيه يتم تصنيف الحالات التي يأخذها المتحول ضمن فئات، أو مجموعات اسمية، وترفق بالتكرارات المطلقة، أو النسبية المقابلة لكل منها، وتوضع في جدول مناسب.

وخير مثال على المتحولات التي تخضع لهذا السلم هي المتحولات التالية:

الجنس: وله حالتان: ذكر، أنثى.

العِرق: وله عدة حالات: الأبيض، الأصفر، الأسود.

الحالة الزواجية: ولها عدة حالات: عازب، متزوج، مطلق، أرمل.

ويشترط هنا أن تكون جميع العناصر ضمن كل حالة متكافئة فيما بينها، وأن لا تكون أفضلية لحالة على أخرى، وهذا يعني أن جميع الذكور ضمن فئة الذكر متساوون، وأن جميع الإناث ضمن فئة الأنثى متكافئات، وأنه لا يمكننا أن نفضل فئة الذكور على فئة الإناث، أو بالعكس...إلخ.

ويمكن الاعتماد على هذا السلم عند تطبيق اختبارات حول النسب المئوية وحول التوافق والتجانس بين الحالات لمتحولين.

. السلم الرتبي Ordinal scale: وهو سلم نوعي شبيه بالسم الاسمي، وفيه يتم تصنيف الحالات التي يأخذها المتحول ضمن فئات اسمية محددة، ولكنه يمكن تفضيل، أو تمييز الحالات المختلفة التي يأخذها المتحول والقيام بترتيبها حسب علاقة ترتيب

معينة، مثل: أكبر، أو أصغر، أو أفضل ...إلخ. ويندرج تحت هذا السلم المتحولات التالية:

درجة التعليم وحالاتها: أمي، ابتدائي، إعدادي، ثانوي، جامعي، عليا.

الرتب الوظيفية وحالاتها: ممتازة، أولى، ثانية، ثالثة، رابعة، خامسة.

وهنا نشير إلى أن العمليتين المعرفتين على هذا السلم هما: عملية التكافؤ ضمن الحالات، والترتيب بين الحالات.

2 . المتحولات الكمّية:

وهي المتحولات التي يمكن وصفها بواسطة القياس الكمّي، والتعبير عن خواصها بواسطة أعداد مرفقة بواحدات قياس معينة. وتقسم هذه المتحولات إلى صنفين أساسيين هما:

ـ متحولات منقطعة: وهي المتحولات التي تأخذ قيم منقطعة صحيحة، أو غير صحيحة، مثل عدد أفراد الأسرة، عدد الطلاب، عدد السيارات...إلخ.

ـ متحولات مستمرة: وهي المتحولات التي تأخذ قيماً مستمرة ومتلاصقة، ويمكنها أن تأخذ أية قيمة مرافقة بين أي قيمتين متجاورتين مهما كانتا قريبتين، مثل درجة الحرارة، العمر، الطول...إلخ.

ويتم قياس المتحولات الكمّية بواسطة سلمين هما:

. السلم المجالي Interval scale: وهو سلم كمّي تكون له نقطة صفر افتراضية، وتكون الأبعاد بين البيانات معروفة ومقاسة بوحدة قياس محددة لكل متحول (كدرجة الحرارة).

ـ السلم النسبي Rational scale: وهو سلم كمّي تكون له نقطة صفر حقيقية تعتبر مبدأً حقيقياً للقياسات وتنسب إليها جميع القياسات، مثل الوزن، والطول، والعمر...إلخ. كما أن النسبة بين أي قيمتين تكون مستقلة عن وحدة القياس.

ولهذه الأسباب تم تصنيف اختبارات الفرضيات إلى صنفين أساسيين هما:

ـ **اختبارات معلمية** Parametric statistics: وهي الاختبارت التي تتناول المتحولات الكمّية في المجتمع المدروس وتختبر أهم معالمه. مثل اختبارت المتوسط والإجمالي والنسبة والتباين والفرق بين المتوسطين، والفرق بين نسبتين ....إلخ.

وأشهر هذه الاختبارات هي الاختبارات التي تناولناها في الفصل الخامس z و t بجميع أشكالهما.

ـ **اختبارات لا معلمية** Nonparametric statistics: وهي الاختبارات التي تتناول المتحولات النوعية في المجتمع المدروس .

إن أشهر اختبار لهذا الاختبار هو اختبار χ^2 الذي تناولناه في الفصل السادس. ومعامل ارتباط سبيرمان الذي تعرضنا إليه في كتاب مبادئ الإحصاء.

وحتى يكتمل الموضوع سنخصص هذا الفصل لاستعراض بعض الاختبارات اللامعلمية الأخرى، التي لاتتعامل مع الحالات الممكنة للمتحولات، بل مع التكرارات المقابلة لكل حالة من الحالات النوعية أو مع الرتب الموافقة لكل درجة من الدرجات المختلفة.

ومن أهم مميزات الاختبارات اللامعلمية أنها لاتشترط توفر شروط افتراضية كثيرة في المجتمع أو العينة. وإن أغلبها يكتفي بتوفر شروط ذات طبيعة بديهية مثل:

. أن تكون العينة عشوائية، أو أن تكون التجارب مستقلة.

. أن تكون العناصر متكافئة ضمن كل حالة، ولايمكن تفضيل حالة أخرى.

. أن تكون المتغيرات المدروسة مستمرة (في بعض الحالات).

كما أنها تتميز بالبساطة وبسهولة وسرعة الحسابات، ولايحتاج الباحثون عند تطبيقها إلى أساس رياضي كبير، وهي تعطي نتائج سريعة ومضمونه، وخاصة عندما تكون شروط الاختبارات المعلمية غير محققة.

إن مجالات تطبيق الاختبارات اللامعلمية لا تنحصر في التعامل مع المتحولات والبيانات النوعية، بل يمكن تطبيقها أيضاً على البيانات الكمية، حيث يمكننا اعتبار القيم أو المجالات حالات نوعية، إلا أنه لاينصح باستبدال الاختبارات المعلمية

بالاختبارات اللامعلمية وتطبيقها على المتحولات والبيانات الكمية، إلا إذا كانت الشروط المرافقة للاختبار المعلمي غير محققة.

ولقد شهدت الحقبة الأخيرة تطوراً كبيراً للاختبارات اللامعلمية، فظهرت المؤشرات الاختبارية التي تعالج مختلف الحالات العلمية والاقتصادية والاجتماعية... إلخ. لابل تعددت الأشكال والصيغ التي تعالج نفس المشكلة أو تطبق على بيانات نوعية متشابهة.

ولكن تعدد تلك المؤشرات بأشكالها وأنواعها وضع الباحثين أمام مشكلتين جديدتين هما:

. مشكلة تحديد الاختبار اللامعلمي المناسب لتطبيقه على البيانات المتوفرة.

. مشكلة التفضيل بين مؤشرات الاختبار اللامعلمية التي تنطبق على الحالة نفسها.

ولمعالجة هاتين المشكلتين تم وضع بعض المعايير المحددة للاعتماد عليها عند التحديد والتفضيل، وأهم هذه المعايير هي*:

. مقاييس البيانات Measures of data.

. نوع عينة البيانات وحجمها Type and size of sample.

. شكل البيانات والهدف من الاختبار Form of data.

. قوة الاختبار $(1-\gamma)$: Power of test.

. الفعالية النسبية Relative efficiency.

. عدم تحيز الاختبار Unbiased ness of test.

. تماسك الاختبار Consistency of test.

. تحفظ الاختبار Conservation of test.

وتعتبر قوة الاختبار والتي تساوي $(1-\gamma)$ أهم هذه المعايير، ولكن عملية

* انظر: *Segiel P.18-31, Conover P.83-90*

حسابية تحتاج إلى مقدمات وبراهين متعمقة لاتدخل ضمن منهاج هذا الكتاب [**] وبهذا فإننا سنستعرض أهم الاختبارات اللامعلمية وفق الترتيب الآتي:

7-2: الاختبارات اللامعلمية المعرفة على المتحولات الأسمية (من عينة واحدة):

عندما نسحب عينة عشوائية واحدة لإجراء دراسة إحصائية معينة، فإن استمارة البحث تكون متضمنة عدداً محدداً من الأسئلة عن مؤشرات كمية وعدداً آخر عن مؤشرات نوعية، واعتماداً على هذه العينة يمكننا أن نقوم بتصنيف الأجوبة ضمن جداول مختلفة تتضمن البيانات والتكرارات المقابلة لها، ثم ننتقل إلى دراسة خصائص المؤشرات السابقة، وحساب متانة العلاقات الارتباطية بينها، وسنستعرض فيما يلي هذه القضايا حسب أنواع المؤشرات وفق التالي:

وهنا يمكن أن نميز عدة أنواع من المتحولات الأسمية هي:

. متحولات أسمية ثنائية: ويكون لها حالتان فقط مثل: الجنس والحالة الصحية.

. متحولات أسمية متعددة: ويكون لها عدة حالات مثل: الدرجة التعليمية والحالة العملية التي يمكن ندرسها هنا هي:

7-2-1: اختبار ثنائي الحد لعدد مرات الظهور أو النجاح":

ويطبق هذا الاختبار على متحول أسمي واحد، ولهذا يتم تصنيف عناصر العينة العشوائية إلى صنفين متنافيين (صنف يتميز بالخاصة المدروسة، وآخر لايتميز بتلك الخاصية)، وإذا رمزنا لحجم العينة بـ n، ولعدد العناصر التي تتصف بالخاصية المدروسة بـ m، فإن m يمثل عدد مرات ظهور الخاصة المدروسة في العينة العشوائية المسحوبة (أو أن m يمثل عدد مرات النجاح عند إجراء n تجربة مستقلة حول الخاصة المدروسة)، فعندها يمكننا أن نضع نتائج هذه النتائج في جدول صغير كالآتي:

[**] انظر: العلي، ابراهيم، الإحصاء الرياضي، جامعة حلب، 1986 ص.

تحقق الخاصة	نعم	لا	Σ
التكرارات	m	$n-m$	n

وبناءً على ذلك نجد أن نسبة وجود الخاصة المدروسة في العينة تحسب من العلاقة:

$$r=\frac{m}{n} \qquad (1-7)$$

واعتماداً على ذلك نقوم بتقدير النسبة R في المجتمع بواسطة العلاقة:

$$\widetilde{R}=r \qquad (2-7)$$

ولإجراء عملية الاختبار نرمز لعدد مرات ظهور الخاصة المدروسة بالرمز X، وبذلك يكون خاضعاً للتوزيع الثنائي المعروف، والذي يشترط أن يكون احتمال الظهور في كل تجربة ثابتاً ومساوياً لقيمة معينة P.

وبذلك نجد أن المتحول X يأخذ القيم الصحيحة التالية:

$$X: 0\ ,\ 1\ ,\ 2\ ,\ 3\ ,\ 4.....n$$

وإن احتمال أن يأخذ X قيمة معينة مثل m يساوي:

$$P(x=m)=C_n^m\cdot p^m\cdot q^{n-m} \qquad (3-7)$$

حيث $q=1-P$

وإن احتمال أن يأخذ X قيمة أصغر، أو يساوي قيمة ℓ (ℓ قيمة صحيحة) يساوي:

$$P(x\le \ell)=\sum_{m=0}^{\ell} C_n^m\cdot p^m\cdot q^{n-m} \qquad (4-7)$$

وإن احتمال أن يأخذ X قيمة أكبر، أو يساوي قيمة u (u قيمة صحيحة) يساوي:

$$P(x\ge u)=\sum_{m=u}^{n} C_n^m\cdot p^m\cdot q^{n-m} \qquad (5-7)$$

ثم نضع الفرضيتين حول نسبة تواجد الخاصة المدروسة حسب القيود المفروضة علينا، أو حسب المعلومات المتوفرة لدينا سابقاً كما يلي:

فرضية العدم: إن نسبة الخاصة المدروسة في المجتمع تساوي p ؛ أي أن:

$$H_0 : R = p \qquad (6-7)$$

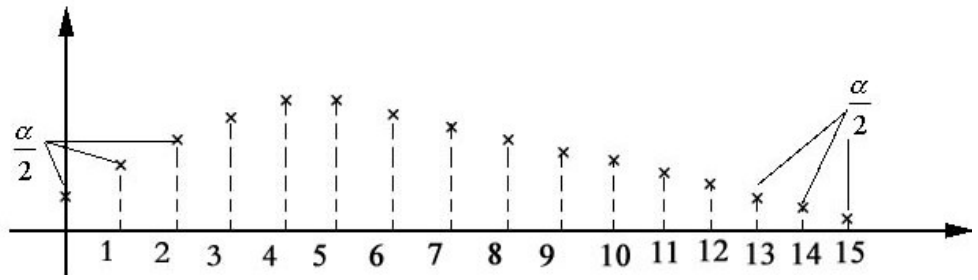
الفرضية البديلة: ويمكن أن تأخذ أحد الأشكال التالية:

$$H_1 : R \neq p \quad , \quad H_1 : R < p \quad , \quad H_1 : R > p$$

أما مؤشر الاختبار في هذه المسألة فهو عبارة عن عدد مرات الظهور الفعلي m مرة خلال إجراء n تجربة مستقلة. وهو عبارة عن قيمة عددية واحدة من قيم المتحول X.

ولاتخاذ قرار حول قبول، أو رفض ذلك العدد m، نقارنه مع منطقتي الرفض والقبول للفرضية H_0. ويتم ذلك حسب الشكل الذي تتخذه الفرضية البديلة H_1.

فإذا كانت الفرضية البديلة من الشكل $H_1 : R \neq p$، فإن الاختبار يكون ثنائي الجانب، وهذا يعني أن منطقة الرفض تقع على الجانبين، ويخصص لكل جانب نصف مستوى الدلالة (أي $\frac{\alpha}{2}$)، كما هو موضح على الشكل الآتي:



الشكل رقم (7-1)

ولتحديد الحد الأدنى لعدد مرات القبول ℓ نقوم بجمع الاحتمالات التراكمية من القيمة 0 حتى القيمة ℓ التي تحقق العلاقة:

$$P(X \leq \ell) = \sum_{m=0}^{\ell} C_n^m \cdot p^m \cdot q^{n-m} = \frac{\alpha}{2} \qquad (7-7)$$

ولتحديد الحد الأعلى لعدد مرات القبول u نقوم بجمع الاحتمالات التراكمية من القيمة u حتى القيمة n التي تحقق العلاقة:

$$P(X \le u) = \sum_{m=u}^{u} C_n^m \cdot p^m \cdot q^{n-m} = \frac{\alpha}{2} \qquad (8-7)$$

ولكن تحقيق المساواة في العلاقتين السابقتين قد لايكون متاحاً في معظم الأحيان، وعندها يكتفى بأقرب قيمة لـ ℓ ولـ u تعطياننا احتمالاً لايتجاوز $\frac{\alpha}{2}$.

أما إذا كانت الفرضية البديلة من الشكل: $H_1 : R < p$

فإننا نقوم بتحديد الحد الأدنى بحيث يكون مقابلاً لكامل مستوى الدلالة α ويحقق العلاقة:

$$P(X \le \ell) = \sum_{m=0}^{\ell} C_n^m \cdot p^m \cdot q^{n-m} = \alpha \qquad (9-7)$$

أما إذا كانت الفرضية البديلة من الشكل: $H_1 : R > p$

فإننا نقوم بتحديد الحد الأعلى بحيث يكون مقابلاً لكامل مستوى الدلالة α ويحقق العلاقة:

$$P(X > u) = 1 - P(x \le u) =$$

$$1 - \sum_{m=0}^{u} C_n^m \cdot p^m \cdot q^{n-m} = 1 - \alpha \qquad (10-7)$$

وتسهيلاً لهذه الحسابات أعدت جداول خاصة لحساب هذه الاحتمالات التراكمية، وهي الجداول (7) الملحقة في آخر هذا الكتاب.

وأخيراً نجد أن اتخاذ القرار المناسب يتم بمقارنة القيمة الفعلية لعدد مرات الظهور m بمجالي منطقة القبول والرفض. فإذا كان واقعاً في منطقة القبول قبلنا الفرضية H_0 واعتبرنا أن النسبة المفترضة P مقبولة.

أما إذا كان واقعاً في منطقة الرفض فإننا نرفض الفرضية H_0 ونقبل الفرضية البديلة H_1 .

حالة العينات الكبيرة:

إذا كان حجم العينة n كبيراً (أكبر من 30) فإنه يمكننا اعتبار التوزيع الطبيعي تقريباً جيداً للتوزيع الثنائي، وعندها يمكن إجراء الاختبار باستخدام مؤشر الاختبار التالي:

$$Z = \frac{m - np}{\sqrt{n \cdot p \cdot q}} \qquad (11-7)$$

ثم نقارن قيمة Z المحسوبة مع القيمة الجدولية $Z_{1-\frac{\alpha}{2}}$ (أو مع $Z_{1-\alpha}$) فإذا كانت $|Z| < Z_{1-\frac{\alpha}{2}}$ نقبل الفرضية H_0، والعكس بالعكس.

أما عندما تكون الخاصة مستمرة (مثل درجة الحرارة) فإن العلاقة (7−11) تصحح وتعدل على الشكل التالي:

$$Z = \frac{(m \pm 0{,}5) - np}{\sqrt{n \cdot P \cdot q}} \qquad (12-7)$$

وذلك بحيث يتم إضافة 0.5 إلى m عندما تكون $m < n \cdot P$ عندما يتم طرح 0.5 من m عندما تكون $m > n \cdot P$.

مثال 1:

تشترط إحدى الشركات على المورد أن لاتزيد نسبة البضاعة المعيبة في البضاعة المستوردة عن 15%، وعند استلام البضاعة سحبت لجان الاستلام عينة صغيرة الحجم n=20 قطعة، فكان عدد القطع المعيبة فيها مساوياً لـ 4 قطع. فهل هذه النتيجة تتفق مع النسبة الموضوعة أم لا؟ اختبر بمستوى دلالة $\alpha = 0.05$.

الحل:

نضع الفرضيتين كمايلي:

$$H_0 : R = 0{,}15 = p$$

$$H_1 : R > 0{,}15 \qquad \text{(والاختبار أحادي يميني)}$$

ولحساب الحد الأعلى لمنطقة القبول u نطبق العلاقة (7−10) كمايلي:

$$P(X>u)=1-P(x\le u)=1-\sum_{m=0}^{u}C_{20}^{m}\cdot p^{m}q^{n-m}=1-0{,}05$$

ومنها نحصل على أن u يجب أن تحقق العلاقة الآتية:

$$\sum_{m=0}^{U}C_{20}^{m}\cdot p^{m}q^{20-m}=0.95$$

ومن الجداول الملحقة نجد أنه عندما يكون $n=20,\quad p=0{,}15$ فإن قيمة الحد الأعلى $u=5$ قطع، ويقابل احتمالاً تراكمياً:

$$\sum_{m=0}^{5}C_{20}^{m}P^{m}q^{20-m}=0.9327$$

وهو أقرب الاحتمالات التي تترك على يمينها مستوى دلالة أكبر من $0{,}05$. وبمقارنة عدد القطع المعيبة $m=4$ مع الحد الأعلى $u=5$، نجد أن $m<u$، لذلك نقبل الفرضية H_0، ونعتبر البضاعة مقبولة، لأن عدد القطع المعيبة يتفق مع النسبة الموضوعة.

مثال 2:

لنفترض أن لجنة أخرى للاستلام قامت بسحب عينة كبيرة بحجم $n=50$ قطعة من البضاعة المذكورة في المثال السابق،، فوجدت أن عدد القطع المعيبة فيها $m=10$ قطع. فهل هذه النتيجة تتفق مع النسبة الموضوعة نفسها؟ أم لا؟

اختبر بمستوى دلالة $\alpha=0{,}05$.

الحل:

نلاحظ أن حجم العينة n أصبح كبيراً، لذلك نطبق اختبار التوزيع الطبيعي، ونضع الفرضيتين كما يلي:

$$H_0: R=0{,}15=p$$

$$H_1: R>0{,}15 \qquad \text{(الاختبار أحادي يميني)}$$

ثم نقوم بحساب مؤشر الاختبار التالي:

$$z = \frac{m - n \cdot P}{\sqrt{n \cdot p \cdot q}} = \frac{10 - 50(0{,}15)}{\sqrt{50(0{,}15)(0{,}85)}} = \frac{2{,}5}{2{,}525} = 0{,}99$$

وبمقارنة هذه القيمة مع قيمة $z_{1-\alpha} = 1{,}64$ نجد أن $z < 1{,}64$، لذلك نقبل الفرضية H_0، ونعتبر البضاعة مقبولة، وتتفق مع النسبة الموضوعة.

مثال 3:

تشير مصادر وزارة التخطيط أن نسبة الأمية بين النساء البالغات في مدينة معينة تبلغ 40%. وللتأكد من ذلك سحبنا عينة عشوائية بحجم $n = 20$ إمرأة بالغة، فوجدنا أن عدد الأميات فيها يساوي $m = 7$ نساء. فهل تعتبر هذه النتيجة متفقة مع مصادر وزارة التخطيط؟ أم لا؟

اختبر بمستوى دلالة $\alpha = 0{,}05$.

الحل:

نضع الفرضيتين كما يلي:

$$H_0 : R = 0{,}40 = p$$

$$H_1 : R \neq 0{,}40$$ (الاختبار ثنائي الجانب)

وبما أن حجم العينة صغير، والاختبار ثنائي الجانب، لذلك نقوم أولاً بحساب الحد الأدنى للظهور ℓ فنجد من الجداول الملحقة عندما تكون $n = 20$، و $P = 0{,}30$ أن قيمة ℓ التي تحقق العلاقة:

$$P(X \leq \ell) = \sum_{m=0}^{\ell} C_n^m \cdot p^m \cdot q^{n-m} = \frac{\alpha}{2} = 0{,}025$$

تساوي $\ell = 3$.

ولحساب الحد الأعلى نبحث عن قيمة u التي تحقق العلاقة التالية:

$$P(X \geq u) = \sum_{m=u}^{n} C_n^m \cdot P^m \cdot q^{n-m} = \frac{\alpha}{2} = 0{,}025$$

أو تحقق العلاقة:

$$P(X \geq u) = 1 - P(X \leq u) = 0{,}025$$

أي التي تحقق العلاقة:

$$P(X \leq u) = 0{,}975$$

ومن الجداول الملحقة نجد أن قيمة u التي تحقق هذه العلاقة هي $u = 11$، وبما أن عدد الأميات في العينة يحقق المتراجحة $3 \leq m \leq 11$، لذلك نقبل الفرضية H_0، ونعتبر أن نتيجة العينة متفقة مع تقديرات الوزارة.

7-2-2: اختبار التعاقب Runs Test:

يدرس اختبار التعاقب المعرف على عينة واحدة مدى تحقق صفة العشوائية بين عناصر العينة المسحوبة، وذلك عن طريق متابعة تعاقب ظهور إحدى الخواص المدروسة فيها. وكمثال على ذلك، نفترض أننا ألقينا قطعة نقود 20 مرة، وكانت نتائج ظهور الصورة H والكتابة T كما يلي:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
H	T	T	H	H	H	T	H	H	T	T	H	T	H	H	T	T	T	H	T
+	−	−	+	+	+	−	+	+	−	−	+	−	+	+	−	−	−	+	−

ومنها نلاحظ أن عدد مرات ظهور الصورة H يساوي 10 مرات، وعدد مرات ظهور الكتابة T يساوي 10 مرات أيضاً.

ولكن عدد مرات تعاقب ظهور الصورة فيساوي 12 مرة فقط (وهو يساوي مجموع عدد المجموعات الموجبة والسالبة في النتائج السابقة).

ومن الناحية النظرية يمكننا أن نتصور أنه عندما تكون الأمور عشوائية تماماً، فيجب أن تكون نتائج الـ 20 رمية على الشكل التالي:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
H	T	H	T	H	T	H	T	H	T	H	T	H	T	H	T	H	T	H	T
+	−	+	−	+	−	+	−	+	−	+	−	+	−	+	−	+	−	+	−

والتي تُظهر أن تعاقب الصورة H يتم بشكل متناوب مع الكتابة T، وأن عدد مرات ظهورها يساوي 10 مرات، وكذلك نجد أن عدد مرات ظهور الكتابة يساوي 10 مرات أيضاً، ولكننا نجد أن عدد مرات التعاقب في هذه الحالة يساوي 20 مرة.

وكذلك يمكننا أن نتصور أنه عندما تكون الأمور غير عشوائية بشكل كامل، فإن عدد مرات ظهور الصورة H يمكن أن يساوي 20 مرة، ويمكن أن يساوي صفراً كأن يكون على الشكل التالي:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
H	H	H	H	H	H	H	H	H	H	T	T	T	T	T	T	T	T	T	T
+	+	+	+	+	+	+	+	+	+	−	−	−	−	−	−	−	−	−	−

ومن الشكل يتبين أن عدد مرات ظهور الصورة H يساوي 10 مرات، وعدد مرات ظهور الكتابة يساوي 10 مرات أيضاً. ولكن عدد مرات التعاقب يساوي مرتين فقط.

ومما تقدم يتضح لنا أنه لقياس صفة العشوائية بين عناصر العينة لا يُكتفى بحساب عدد مرات الظهور، بل لابد من اللجوء إلى حساب عدد مرات التعاقب، والذي يعبر عن عشوائية توارد الخاصة المدروسة مع عناصر العينة المسحوبة، أو التجارب المجراة. ولمعالجة هذا الموضوع نقوم بما يلي:

نرمز لحجم العينة المسحوبة، أو لعدد التجارب المجراة بالرمز n.

ونرمز لعدد مرات ظهور الخاصة المدروسة بالرمز n_1 مرة.

ونرمز لعدد مرات عدم ظهور الخاصة المدروسة بالرمز n_2 مرة.

وعندها يكون لدينا $n_1 + n_2 = n$

كما نرمز لعدد مرات تعاقب الخاصة المدروسة بالرمز r مرة.

ثم نقوم بوضع الفرضيات كما يلي:

الفرضية الابتدائية (فرضية العدم) H_0: إن تعاقب ظهور الخاصة المدروسة يحدث بشكل عشوائي مقبول؛ أي إن عدد مرات التعاقب يساوي $H_0 : r = \frac{n}{2}$.

الفرضية البديلة: إن تعاقب ظهور الخاصة المدروسة لا يحدث بشكل عشوائي؛ أي إن عدد مرات التعاقب يساوي $H_1 : r \neq \frac{n}{2}$

أما مؤشر الاختبار فيتجسد في عدد مرات التعاقب r، ويتم حسابه مباشرةً من عناصر العينة المسحوبة، وهو يمكن أن يأخذ القيم التالية $n.....3\ ,\ 2 : r$. ولاتخاذ القرار المناسب حول قيمة r المحسوبة نعود إلى التوزيع الاحتمالي الخاص به، أو إلى الجدول الملحق في آخر الكتاب، ونستخلص منهما الحدين الأدنى r_1 والأعلى r_2 لقيم r المقابلين للعددين n_1 و n_2، ولمستوى دلالة قدره 0,025 على كل جانب (لأن الاختبار السابق ثنائي الجانب).

فإذا كانت قيمة r تحقق المتراجحة $r_1 \le r \le r_2$، فإننا نقبل الفرضية H_0، والعكس بالعكس. وسيكون رفضنا لـ H_0 وبمستوى دلالة قدره 0,05.

ملاحظة:

إذا كان الاختبار أحادي الجانب $H_1 : r < \frac{n}{2}$، أو $H_1 : r > \frac{n}{2}$)، فإننا نستخدم أحد الجدولين F_1 أو F_2 لاستخراج القيمة الحدية المطلوبة.

*** حالة العينات الكبيرة (n_1 أو $n_2 > 20$):**

إذا كان العدد n_1 أو n_2 أكبر من 20، فإن التوزيع الاحتمالي لعدد مرات التعاقب يقترب من التوزيع الطبيعي الذي توقُّعَه وتباينه يساويان:

$$\mu_r = 1 + \frac{2n_1 \cdot n_2}{n_1 + n_2} \qquad (13-7)$$

$$\sigma_r^2 = \frac{2n_1 \cdot n_2 (n_1 \cdot n_2 - n_1 - n_2)}{(n_1 + n_2)^2 (n_1 + n_2 - 1)} \qquad (14-7)$$

وبذلك نجد أنه لاختبار فرضية العدم يمكننا الاعتماد على مؤشر الاختبار الطبيعي التالي:

$$z = \frac{r - \mu_r}{\sigma_r} \qquad (15-7)$$

$$z = \frac{r - \left(\frac{2n_1 \cdot n_2}{n_1 + n_2} + 1\right)}{\sqrt{\frac{2n_1 \cdot n_2 (2n_1 \cdot n_2 - n_1 - n_2)}{(n_1 + n_2)^2 (n_1 + n_2 - 1)}}} \qquad (16-7)$$

وبعد حساب قيمة z، نقوم بمقارنته مع قيمة $z_{1-\frac{\alpha}{2}}$ الجدولية. فإذا كانت $|z| < z_{1-\frac{\alpha}{2}}$، فإننا نقبل فرضية العدم H_0، والعكس بالعكس.

مثال 4:

يقف الطلبة في صف طوله 30 طالباً لشراء الكتب الجامعية. فإذا كان عدد الإناث $n_1 = 12$، وكان عدد الذكور $n_2 = 18$، وكان عدد مرات التعاقب ضمن الصف $r = 10$، فالمطلوب:

اختبار عشوائية توارد الذكور والإناث في هذا الصف وذلك باحتمال قدره $\alpha = 0{,}05$.

الحل:

نلاحظ أن كلاً من العددين n_1 و n_2، أصغر من 20 مرة، لذلك نطبق أسلوب الجداول فنجد أنه يمكننا وضع الفرضيتين كما يلي:

فرضية العدم: إن توارد الجنسين في الصف يتم بشكل عشوائي، أو أن $r = \frac{n}{2} = 15$.

الفرضية البديلة: إن توارد الجنسين في الصف لا يتم بشكل عشوائي، أو أن $r \neq \frac{n}{2}$.

كما نجد أن الجدول F_1 يعطينا أن الحد الأدنى لعدد مرات التعاقب المقابل للعددين $n_1 = 12$، و $n_2 = 18$، يساوي $r_1 = 9$

ومن الجدول الملحق يعطينا أن الحد الأعلى لعدد مرات التعاقب المقابل للعددين $n_1 = 12$، و $n_2 = 18$، يساوي $r_2 = 21$..

وبمقارنة قيمة r الفعلية مع الحدين الأدنى والأعلى نجد أنه:

$$9 \le r \le 21$$

وهذا يعني أنه يجب علينا قبول فرضية العدم H_0 التي تقول إن توارد الجنسين يتم بشكل عشوائي ضمن الصف المذكور.

مثال 5:

يتوارد الناس على موقف الباص بشكل كيفي. وعند دراسة عشوائية هذا التوارد، أخذنا عينة متواصلة بحجم $n = 50$ فرداً. فوجدنا أن عدد الذكور فيها $n_1 = 30$، وكان عدد الإناث $n_2 = 20$، ثم حسبنا عدد مرات التعاقب فوجدنا أن $r = 22$، والمطلوب:

اختبار فيما إذا كان ذلك التوارد عشوائياً، وذلك بمستوى دلالة قدره $\alpha = 0{,}05$.

الحل:

نلاحظ أن $n_1 > 20$، لذلك نطبق أسلوب الاختبار الطبيعي، ونضع الفرضيتين كما يلي:

$$H_0 : r = \frac{n}{2} = 25$$

$$H_1 : r \neq \frac{n}{2}$$

ثم نقوم بحساب قيمة مؤشر الاختبار من العلاقة:

$$Z=\frac{r-\left(\frac{2n_1\cdot n_2}{n_1+n_2}+1\right)}{\sqrt{\frac{2n_1\cdot n_2\left(2n_1\cdot n_2-n_1-n_2\right)}{\left(n_1+n_2\right)^2\left(n_1+n_2-1\right)}}}=\frac{22-\left(\frac{2\cdot 30\cdot 20}{20+30}+1\right)}{\sqrt{\frac{2\cdot 30\cdot 20\left(2\cdot 3020-30-20\right)}{\left(30+20\right)^2\left(30+20-1\right)}}}$$

$$Z=\frac{22-25}{\sqrt{\frac{1200(1150)}{2500(49)}}}=\frac{-3}{3,35638}=-0,8938$$

وبما أن قيمة Z الجدولية $Z_{1-\frac{\alpha}{2}}$ ، فإننا نجد أن: $|Z|<1,96$، لذلك نقبل فرضية العدم H_0 التي تقول إن توارد الجنسين يتم بشكل عشوائي.

7-2-3: اختبارات الاقتران بين متحولين اسميين ثنائيين Association Tests:

لنفترض أنه لدينا متحولان اسميان x و y، لكل منهما حالتان فقط. ولهما التكرارات المقابلة كما في الجدول الآتي:

y \ x	x_1	x_2
y_1	A	B
y_2	C	D

7-2-3-1: معادل الاقتران الرباعي:

يستخدم هذا المعامل لاختبار وجود اقتران بين هذين المتحولين بطريقة حساب معامل الاقتران Association coefficient، المعرف بالعلاقة التالية:

$$CA=\frac{A\cdot D-B\cdot C}{A\cdot D+B\cdot C} \qquad (17-7)$$

وهو يأخذ قيماً محصورة في المجال $]-1\ ,\ +1[$، وكلما كانت قيمته المطلقة قريبة من الواحد، كان الاقتران بين المتحولين قوياً. أما عندما تكون قيمته المطلقة قريبة من الصفر، أو أصغر من 0,50، فلا نأخذ بالاقتران بين المتحولين.

وحتى لا يحصل أي التباس حول ذلك يفضل عند إجراء الدراسات الدقيقة إجراء اختبار معنوية هذا المعامل من خلال مؤشر الاختبار الطبيعي الآتي:

$$z = \frac{2(CA)}{1-(CA)^2} \cdot \frac{1}{\sqrt{\frac{1}{A}+\frac{1}{B}+\frac{1}{C}+\frac{1}{D}}} \qquad (18-7)$$

وبمقارنة قيمة Z مع قيمة Z الجدولية، نقرر فيما إذا كانت قيمة معامل الاقتران معنوية، أم لا.

مثال 6:

ادرس فيما إذا كان هناك اقتران بين الجنس والنجاح لنتائج الثانوية في إحدى المدارس إذا علمت أن نتيجة الامتحانات أعطتنا الأعداد التالية:

الجنس / النتيجة	ذكر	أنثى
ناجح	215	150
راسب	130	90

الحل:

نقوم بحساب معامل الاقتران بين هذين المتحولين الاسميين، فنجد أن:

$$CA = \frac{A \cdot D - B \cdot C}{A \cdot D + B \cdot C} = \frac{215 \cdot 90 - 150 \cdot 130}{215 \cdot 90 + 150 \cdot 130}$$

$$CA = \frac{-150}{38850} = -0{,}00386$$

وهي قيمة صغيرة جداً تدل على عدم وجود اقتران بين الجنس والنتيجة. وللتأكد من معنوية هذه القيمة نحسب قيمة مؤشر الاختبار Z، فنجد أن:

$$Z=\frac{2(0{,}00386)}{1-(0{,}00386)^2}\cdot\frac{1}{\sqrt{\frac{1}{215}+\frac{1}{150}+\frac{1}{130}+\frac{1}{90}}}=0{,}0772\cdot\frac{1}{\sqrt{0{,}030}}$$

$$Z=0{,}446$$

وبمقارنة هذه القيمة مع القيمة الجدولية لـ Z عند مستوى دلالة $0{,}05$، والتي تساوي $z_{1-\frac{\alpha}{2}}=1{,}96$ نجد أن $Z<Z_{1-\frac{\alpha}{2}}$، لذلك نقبل فرضية العدم H_0 حول CA والتي تقول إن قيمة هذا المعامل معدومة ولا يوجد أي اقتران بيم حالتي الجنس، وحالتي النتيجة.

7-2-4: اختبارات التوافق لمتحولين اسميين Contingency Test:

7-2- 4-1: معامل التوافق الرباعي (2×2):

يطبق هذا المعامل على متحولين اسميين ثنائيين (أو كميين ثنائيين)، ولهما التكرارات المقابلة كما في الجدول الآتي:

x / y	x_1	x_2	المجموع
y_1	A	B	$A+B$
y_2	C	D	$C+D$
المجموع	$A+C$	$B+D$	N

ويعرف معامل التوافق الرباعي بالعلاقة التالية:

$$T=\frac{N(A\cdot D-B\cdot C)^2}{(A+B)(C+D)(A+C)(B+D)} \qquad (19-7)$$

وهو متحول خاضع لتوزيع $\chi^2(x)$ بدرجة حرية واحدة $(\nu=1)$.

ولاتخاذ قرار حول فرضية العدم (عدم التوافق) نقارن قيمة هذا المعامل مع قيمة متحول التوزيع $\chi^2(x)$ المقابلة لمستوى الدلالة ولدرجة حرية واحدة $(\nu = 1)$، ونتخذ القرار المناسب كالعادة.

مثال 7:

لدراسة التوافق بين أجوبة المبحوثين حول سؤالين هما:

السؤال الأول: هو هل تؤيد عمل المرأة؟ نعم لا

السؤال الثاني: هل تؤيد تنظيم الإنجاب؟ نعم لا

أخذنا نتائج الأجوبة وبوبناها كما في الجدول الآتي:

السؤال الأول / السؤال الثاني	نعم	لا	المجموع
نعم	160	50	210
لا	40	130	170
المجموع	200	180	380

والمطلوب دراسة التوافق بين أجوبة المبحوثين على السؤالين السابقين.

الحل:

لدراسة التوافق المطلوب نطبق العلاقة (7−19)، فنجد أن:

$$T = \frac{N(A\cdot D - B\cdot C)^2}{(A+B)(C+D)(A+C)(B+D)} = \frac{380(160\cdot 130 - 50\cdot 40)^2}{210\cdot 170\cdot 200\cdot 180}$$

$$T = \frac{380(353440000)^2}{1285200000} = 104{,}5$$

ومن الجداول الملحقة *(V)* نجد أن قيمة متحول $\chi^2(x)$ المقابلة لمستوى الدلالة ولدرجة حرية واحدة $(\nu = 1)$، تساوي $\chi^2(x) = 3{,}84$.

وبمقارنة قيمة T المحسوبة مع القيمة الحرجة 3,84، نجد أن $T > 3{,}84$ لذلك نرفض فرضية العدم حول عدم التوافق ونقبل الفرضية البديلة التي تقول بوجود توافق بين أجوبة المبحوثين حول السؤالين السابقين. ونستنتج أن معظم الذين يؤيدون عمل المرأة يؤيدون أيضاً تنظيم الإنجاب، والعكس بالعكس.

7-2-4-2: معامل التوافق المتعدد $(k \times l)$ Contingency coefficient:

يعتمد هذا المعامل على اختبار χ^2 الذي درسناه في الفصل السابق ويعرف على متحولين اسميين لهما أكثر من حالتين، ونفترض أن المتحول الأول له k حالة، وأن المتحول الثاني له r حالة. ويأخذ جدول التكرارات المشتركة لهما الشكل التالي:

المتحول الأول x / المتحول الثاني y	x_1	x_2	x_3	...	x_j	...	x_k	المجموع
y_1	n_{11}	n_{12}	n_{13}	...	n_{1j}	...	n_{1k}	n'_1
y_2	n_{21}	n_{22}	n_{23}	...	n_{2j}	...	n_{2k}	n'_2
y_3	n_{31}	n_{32}	n_{33}	...	n_{3j}	...	n_{3k}	n'_3
$\vdots$	$\vdots$	$\vdots$	$\vdots$	...	$\vdots$	...	$\vdots$	$\vdots$
y_i	n_{i1}	n_{i2}	n_{i3}	...	n_{ij}	...	n_{ik}	n'_i
$\vdots$	$\vdots$	$\vdots$	$\vdots$	...	$\vdots$	...	$\vdots$	$\vdots$
y_ℓ	$n_{\ell 1}$	$n_{\ell 2}$	$n_{\ell 3}$	...	$n_{\ell j}$	...	$n_{\ell k}$	n'_ℓ
المجموع	n_1	n_2	n_3	...	n_1	...	n_k	n

وبعد حساب قيمة المتحول χ^2 المعرف بالعلاقة:

$$\chi^2 = \sum_{i=1}^{\ell}\sum_{j=1}^{k}\frac{\left(n_{ij} - E_{ij}\right)^2}{E_{ij}} \qquad (20-7)$$

حيث E_{ij} هي التكرارات المتوقعة، وتحسب من العلاقة $E_{ij} = \dfrac{n_j \cdot n'_i}{n}$، ثم نقوم بحساب معامل التوافق العام من العلاقة (علاقة بيرسون):

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}} \qquad (21-7)$$

ويأخذ هذا المعامل قيمة عددية واقعة في المجال $[1\ ,\ 0]$، وكلما كانت قيمته قريبة من الواحد، كان التوافق بين المتحولين قوياً. ولكننا نشير إلى أن قيمة المعامل C لا يمكنها أن تبلغ العدد (1). ويشترط عند تطبيقه تحقق الشروط المرافقة للاختبار χ^2، (وخاصةً بشرط أن تكون $(E_{ij}\ 25)$).

وبما أننا تعرضنا لهذا المعامل عندما درسنا الاختبارات المعتمدة على المؤشر χ^2 في الفصل السابق، فإننا سوف لن نفصل بشرحه في هذه الفقرة، وننصح الطالب بالعودة إلى الفصل السابق لمتابعة كيفية تطبيقه في جميع الأحوال. كما نشير إلى وجود علاقات أخرى، لمعامل التوافق السابق نذكر منها العلاقات التالي:

. علاقة (كرامر Cramer) وتعرف بالعلاقة:

$$C_1 = \sqrt{\frac{\chi^2}{n(q-1)}} \qquad (22-7)$$

حيث q أصغر العدد ℓ، أو k. و n عدد التكرارات.

. علاقة (ييل Yulle) وتعرف بالعلاقة:

$$C_2 = \frac{\chi^2}{n} \qquad (23-7)$$

. علاقة (تشوبراو Tschuprew) وتعرف بالعلاقة:

$$C_3 = \sqrt{\frac{\chi^2}{n(k-1)(\ell-1)}} \qquad (24-7)$$

. علاقة ( ) وتعرف بالعلاقة:

$$C_4 = \sqrt{\frac{K-n}{K}} \qquad (25-7)$$

حيث أن K الأخيرة تحسب من التكرارات الفعلية والمتوقعة كما يلي:

$$\chi^2 = \sum_{j=1}^{k}\sum_{i=1}^{\ell}\frac{{n_{ij}}^2}{E_{ij}} \qquad (26-7)$$

وتعطينا جميع هذه المعاملات تقديرات لمقدار التوافق بين المتحولين الاسميين إذا كانا متعددي الحالات، وتأخذ قيمتها في المجال $[0\ ,\ 1]$.

وكلما كانت قيمة المعامل المحسوب قريبة من الواحد، كان التوافق جيداً، وبالعكس، كلما كانت قيمته قريبة من الصفر كان التوافق بين حالات المتحولين ضعيفاً.

مثال 8:

احسب معامل التوافق حسب علاقة بيرسون بين متحولي الحالة الاجتماعية والجنس، وذلك من البيانات التالية المأخوذة من عينة الدراسة ذات الحجم $n = 192$.

الحالة الاجتماعية / الجنس	أعزب	متزوج	مطلق	أرمل	المجموع
ذكر	40	50	10	5	105
أنثى	30	35	15	7	87
المجموع	70	85	25	12	192

نقوم أولاً بحساب قيمة χ^2، لذلك نحسب التكرارات المتوقعة (النظرية)، فنجد أن:

$$E_{11} = \frac{105\cdot 70}{192} = 38{,}28 \quad , \quad E_{12} = \frac{105\cdot 85}{192} = 46{,}48$$

$$E_{13} = \frac{105\cdot 25}{192} = 13{,}67 \quad , \quad E_{14} = \frac{105\cdot 12}{192} = 6{,}56$$

$$E_{21} = \frac{87\cdot 70}{192} = 31{,}72 \quad , \quad E_{22} = \frac{87\cdot 85}{192} = 38{,}52$$

$$E_{23} = \frac{87\cdot 25}{192} = 11{,}33 \quad , \quad E_{24} = \frac{87\cdot 12}{192} = 5{,}44$$

وبذلك نجد أن χ^2 تساوي:

$$\chi^2 = \frac{(40-38,28)^2}{38,28} + \frac{(50-46,48)^2}{46,48} + \frac{(10-13,67)^2}{13,67} + \frac{(5-6,56)^2}{6,56}$$
$$+ \frac{(30-31,72)^2}{31,72} + \frac{(35-38,52)^2}{38,52} + \frac{(15-11,33)^2}{11,33} + \frac{(7-5,44)^2}{5,44} =$$
$$\chi^2 = 0,0773 + 0,2666 + 0,9853 + 0,3710 + 0,0933$$
$$+ 0,3217 + 1,1888 + 0,2868 = 3,59$$

وبناءً على ذلك يمكننا حساب قيمة معامل التوافق (علاقة بيرسون) فنجد أن:

$$C_1 = \sqrt{\frac{\chi^2}{\chi^2 + n}} = \sqrt{\frac{3,59}{3,59+192}} = \sqrt{0,0184} = 0,135$$

وهي قيمة صغيرة تدل على عدم وجود توافق بين الحالة الاجتماعية والجنس في العينة المسحوبة.

ملاحظة: كان يمكننا حساب معامل (كرامر) حيث نجد أن:

$$C_1 = \sqrt{\frac{\chi^2}{n(q-1)}} = \sqrt{\frac{3,59}{192(2-1)}} = 0,1367$$

وهي قيمة قريبة من المعامل السابق، وتدل أيضاً على عدم وجود توافق بين الحالة الاجتماعية والجنس.

7-3: الاختبارات المعرفة على المتحولات الرتبية والمجالية (من عينة واحدة):

تتميز المتحولات الرتبية بإمكانية ترتيب حالاتها المختلفة حسب علاقة ترتيب معينة مثل: أكبر، أو أصغر، أو أفضل...إلخ. وبذلك تأخذ هذه الحالات اتجاهاً معيناً (متصاعداً، أو متنازلاً) وتكون مرفقة بالتكرارات المطلقة (أو النسبية) المقابلة لكل حالة فيها.

وأهم الاختبارات المعرفة على المتحولات الرتبية الاختبارات التالية:

7-3-1: اختبارات تطابق التكرارات النسبية مع توزيع احتمالي معين:

لنفترض أن متحولاً رتبياً X له الحالات والتكرارات المطلقة والنسبية التالية:

الحالات المرتبة للمتحول x	x_1	x_2	x_3	…	x_i	…	x_m	المجموع
التكرارات المطلقة n_i	n_1	n_2	n_3	…	n_i	…	n_m	n
التكرارات النسبية p_i	p_1	p_2	p_3	…	p_i	…	p_m	1

حيث إن p_i تحسب من العلاقة التالية: $p_i = \frac{n_i}{n}$

ولنفترض أننا نريد دراسة مدى تطابق، أو توافق هذه التكرارات النسبية مع قيم أحد التوزيعات الاحتمالية النظرية المعطاة بدلالة علاقة رياضية معينة مثل $f(x_i)$، وبذلك يأخذ التوزيع النظري، أو المتوقع الشكل الآتي:

الحالات المرتبة للمتحول X	x_1	x_2	x_3	…	x_i	…	x_m	المجموع
الاحتمالات النظرية المقابلة	$f(x_1)$	$f(x_2)$	$f(x_3)$	…	$f(x_i)$	…	$f(x_m)$	1

ولدراسة مدى التطابق بين هذه التوزيعات نستخدم أحد الاختبارات الآتية:

أ . اختبار χ^2 :

وهو الاختبار الذي درسناه في الفصل السابق، والذي يعرف في حالة متحول واحد بالعلاقة التالية:

$$\chi^2 = \sum_{i=1}^{m} \frac{(n_i - m_i)^2}{m_i} = n \sum_{i=1}^{m} \frac{(p_i - fx_i)^2}{f(x_i)} \qquad (27-7)$$

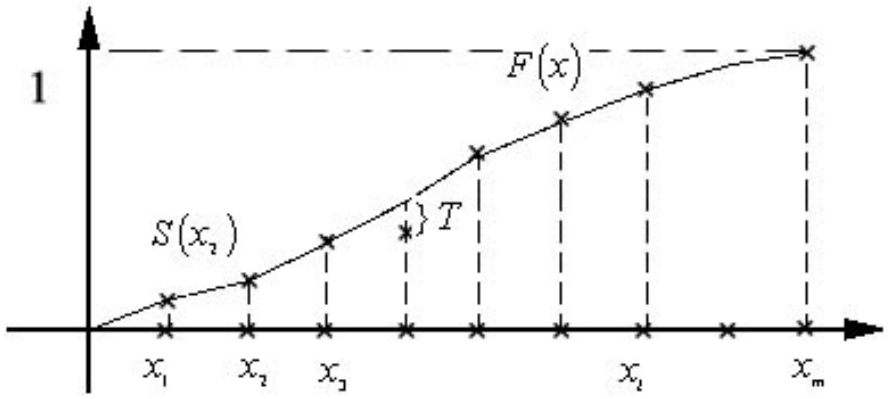
وهناك عدد من تطبيقاته معروضة في الفصل السابق.

ب . اختبار (كولموغوروف . سميرنوف Kolmogorov-smirnov**):**

وهو يعتمد على مقارنة تابع التوزيع الاحتمالي التجريبي $S(x_i)$ (التراكمي) مع تابع التوزيع النظري $F(x_i)$ (التراكمي). وإذا حسبنا هذين التابعين ووضعناهما في جدول نحصل على ما يلي:

الحالات المرتبة للمتحول x	X_1	X_2	X_3	...	X_i	...	X_m	المجموع
تابع التوزيع التجريبي $S(x_i)$	$S(x_1)$	$S(x_2)$	$S(x_3)$	...	$S(x_i)$	...	1	
تابع التوزيع النظري $F(x_i)$	$F(x_1)$	$F(x_2)$	$F(x_3)$		$F(x_i)$		1	

ويمكن للشكل البياني لهذين التابعين أن يأخذ الشكل التالي:



وعندها يمكننا أن نضع الفرضيتين كما يلي:

$$H_0 : S(x_i) = F(x_i) \quad \forall i$$

$$H_1 : S(x_i) \neq F(x_i)$$ (حالة ثنائي الجانب فقط)

وبناءً على ذلك قام (كولموغوروف وسميرنوف) بتعريف مؤشر اختبارهما بالعلاقة التالية:

أعلى قيمة للفرق بالقيمة المطلقة = $T = \sup|S(x_i) - F(x_i)|$

ثم قاما بإعداد جداول خاصة للتوزيع الاحتمالي الذي يخضع له المؤشر T، وهي ملحقة في آخر الكتاب، ومنها يمكننا الحصول على القيم الحرجة T^* المقابلة لمستوى الدلالة α ولحجم العينة n. ولاتخاذ قرار حول الفرضية H_0 نقارن قيمة T

المحسوبة مع القيمة الحرجة T^*. فإذا كانت $T < T^*$ نقبل فرضية العدم H_0، والعكس بالعكس.

مثال 9:

قمنا بتوبيب أفراد عينة من العمال حجمها $n = 25$ فرداً حسب الحالة التعليمية، فحصلنا على الجدول التالي:

الحالات التعليمية لـ X	أمي	ابتدائي	إعدادي	ثانوي	المجموع
التكرارات المطلقة	10	8	5	2	25
التكرارات النسبية	$\frac{10}{25}$	$\frac{8}{25}$	$\frac{5}{25}$	$\frac{2}{25}$	1

فهل هذا التوزيع يخضع للتوزيع المنتظم. اختبر بواسطة اختبار (كولموغوروف وسميرنوف) بمستوى دلالة $\alpha = 0{,}10$.

الحل:

بما أن المطلوب مقارنة هذا التوزيع بالتوزيع المنتظم فإن قيم التوزيع الاحتمالي النظري تأخذ الشكل التالي:

الحالات التعليمية لـ X	أمي	ابتدائي	إعدادي	ثانوي	المجموع
قيم التوزيع النظري	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	1

ولمقارنة هذين التوزيعين نقوم بحساب قيم تابع التوزيعين التجريبي والنظري (تراكمياً) فنحصل على الجدول التالي:

الحالات التعليمية لـ X	أمي	ابتدائي	إعدادي	ثانوي
قيم تابع التوزيع التجريبي $S(x)$	$\frac{10}{25}$	$\frac{18}{25}$	$\frac{23}{25}$	1
قيم تابع التوزيع النظري $F(x)$	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{3}{4}$	1
$S(x_i) - F(x_i)$	0,15	0,22	0,17	0

$$H_0 : S(x_i) = F(x_i) \quad \forall i$$
$$H_1 : S(x_i) \neq F(x_i)$$

ثم نقوم بحساب الفروقات $S(x_i) - F(x_i)$ ونضيفها إلى الجدول السابق، فنحصل على السطر الأخير فيه، وبذلك نجد أن قيمة مؤشر الاختبار تساوي:

$$T = \sup|S(x_i) - F(x_i)| = 0{,}22$$

ومن الجدول (9) الخاص الملحق في آخر الكتاب نجد أن القيمة الحرجة المقابلة لمستوى دلالة $\alpha = 0{,}10$ ولـ $n = 25$ تساوي $T^* = 0{,}238$. وبالمقارنة نجد أن $T < 0{,}238$ لذلك نقبل فرضية العدم H_0، ونعتبر أن التوزيع التجريبي يقارب التوزيع المنتظم.

جـ . اختبار ليليفورز Lilliefors**) الكمي للتطابق مع التوزيع الطبيعي المعياري:**

يطبق هذا الاختبار لدراسة التطابق بين التوزيع التجريبي لمتحول كمي X مع التوزيع الطبيعي المعياري $N(0\ ,\ 1)$. ولنفترض أنه لدينا عينة بحجم n ومبوبة في جدول كالتالي:

قيم X المرتبة	X_1	X_2	X_3	...	X_i	...	X_m	المجموع
التكرارات المطلقة	n_1	n_2	n_3	...	n_i	...	n_m	n
التكرارات النسبية	p_1	p_2	p_3	...	p_i	...	p_m	1
تابع التوزيع التراكمي $S(x_i)$	$S(x_1)$	$S(x_2)$	$S(x_3)$	...	$S(x_i)$	...	1	

ومن هذه البيانات نقوم بتقدير متوسط المجتمع μ وتباينه σ^2 من العلاقتين:

$$\tilde{\mu} = \bar{x} = \frac{1}{n}\sum n_i x_i$$

$$\tilde{\sigma}^2 = s^2 = \frac{1}{n-1}\sum n_i (x_i - \bar{x})^2$$

وبعدها نقوم بحساب القيم المعيارية للمتحول Z من العلاقة:

$$Z_i = \frac{x_i - \bar{x}}{s} \qquad i : 1\ ,\ 2\ ,\ 3 \ldots\ldots m$$

ونضع الفرضيات كما يلي:

H_0: المتحول Z يخضع للتوزيع الطبيعي المعياري $N(0\ ,1)$، وبالتالي يكون المتحول X خاضعاً للتوزيع الطبيعي العام $N\left(\mu\ ,\ \sigma^2\right)$.

H_0: المتحول Z لا يخضع للتوزيع الطبيعي المعياري $N(0\ ,1)$.

ثم نقوم بتعريف مؤشر الاختبار على المتحول Z الذي له تابع التوزيع التجريبي نفسه $S(z_i)$، ومقارنته مع تابع التوزيع الطبيعي المعياري التراكمي $\Phi(x_i)$ (المعرف سابقاً) وذلك باستخدام مؤشر (كولموغوروف وسميرنوف) نفسه الذي يأخذ الشكل التالي:

$$T_z = \sup\left|S(z_i) - \Phi(z_i)\right|$$

ولاتخاذ قرار حول الفرضية H_0 نقارن قيمة T_z مع القيمة الحرجة لها T_z^* المأخوذة من الجدول (10) الخاص بذلك والمقابلة لمستوى الدلالة α ولحجم العينة n. فإذا كانت $T_z < T_z^*$، نقبل الفرضية H_0، والعكس بالعكس.

مثال 10:

لنفترض أن تبويب الطلاب المبحوثين حسب العمر في عينة بحجم $n = 50$ كان كما يلي:

قيم العمر X	20	21	22	23	24	25	26	المجموع
التكرارات المطلقة	5	7	8	10	9	8	3	50
التكرارات النسبية	$\frac{5}{50}$	$\frac{7}{50}$	$\frac{8}{50}$	$\frac{10}{50}$	$\frac{9}{50}$	$\frac{8}{50}$	$\frac{3}{50}$	1
التكرارات النسبية التراكمية $S(x_i)$	$\frac{5}{50}$	$\frac{12}{50}$	$\frac{20}{50}$	$\frac{30}{50}$	$\frac{39}{50}$	$\frac{47}{50}$	1	

والمطلوب دراسة فيما إذا كان العمر X يخضع للتوزيع الطبيعي العام، واختبار ذلك بمستوى دلالة $\alpha = 0{,}05$.

الحل:

نضع الفرضيات كما يلي:

H_0: أعمار المبحوثين تخضع للتوزيع الطبيعي العام $N(\mu,\sigma^2)$.

H_1: أعمار المبحوثين لا تخضع للتوزيع الطبيعي العام.

لذلك نقوم بتقدير كل من μ و σ^2 من العلاقتين:

$$\tilde{\mu} = \bar{x} = \frac{1}{50}[5\cdot 20 + 7\cdot 21 + 8\cdot 22 + 10\cdot 23 + 9\cdot 24 + 8\cdot 25 + 3\cdot 26]$$

$$\tilde{\mu} = \bar{x} = \frac{1147}{50} = 22{,}94$$

$$\tilde{\sigma}^2 = s^2 = \frac{1}{49}[5\cdot(20-22{,}94)^2 + 7\cdot(21-22{,}94)^2 +$$
$$8\cdot(22-22{,}94)^2 + 10\cdot(23-22{,}94)^2 + 9\cdot(24-22{,}94)^2 +$$
$$8\cdot(25-22{,}94)^2 + 3\cdot(26-22{,}94)^2]$$

$$s^2 = \frac{1}{49}[43{,}218 + 26{,}345 + 7{,}069 + 0{,}036 + 10{,}112 + 33{,}95 + 28{,}091]$$

$$s^2 = \frac{1}{49}(148{,}821) = 3{,}037 \qquad s = 1{,}743$$

ومنها نجد أن الانحراف المعياري يقدر بـ $\tilde{\sigma} = s = 1{,}743$.

ثم نحسب قيم المتحول z من العلاقة: $z_i = \dfrac{x_i - \bar{x}}{s}$ ، ونضعها مع التكرارات النسبية التراكمية في جدول كالتالي:

قيم z المعيارية	-2,94	-1,94	-2,94	0,06	1,06	2,06	3,06
التكرارات النسبية التراكمية $S(x_i)$	$\frac{5}{50}$	$\frac{12}{50}$	$\frac{20}{50}$	$\frac{30}{50}$	$\frac{39}{50}$	$\frac{47}{50}$	1
قيم التوزيع التجريبي التراكمية $S(x_i)$	$\frac{5}{50}$	$\frac{12}{50}$	$\frac{20}{50}$	$\frac{30}{50}$	$\frac{39}{50}$	$\frac{47}{50}$	
قيم التوزيع الطبيعي التراكمية $\Phi(x_i)$	0,0164	0,0762	0,1736	0,5239	0,8554	0,9803	1
الفروقات $\lvert\Phi(x_i)-F(x_i)\rvert$	0,0836	0,2138	0,2264	0,0761	0,0754	0,0403	0

ولقد قمنا بحساب الاحتمالات التراكمية النظرية $\Phi(x_i)$ اعتماداً على جدول التوزيع الطبيعي المعياري، فوجدنا مثلاً أن:

$$\Phi(-2,94)=1-\Phi(2,94))=1-0,99835=0,0164$$

ثم حسبنا الفروقات بين كل احتمالين تراكميين، ووضعنا هذه النتائج في الجدول السابق.

ومن هذا الجدول نجد أن قيمة مؤشر (ليليفورز) تساوي:

$$T_1=\sup|S(z_i)-\Phi(z_i)|=0,2264$$

ولاتخاذ قرار حول ذلك التطابق نبحث عن القيمة الحرجة T_1^* من الجدول (15) الخاص والمقابلة لمستوى دلالة $\alpha=0,05$ ولـ $n=50$، فنجد أنها تساوي $T_1^*=\frac{0,886}{\sqrt{50}}=0,1253$. وبالمقارنة نجد أن:

$T>T_1^*$، لذلك نرفض فرضية العدم والتي تقول إن التوزيع الاحتمالي لأعمار المبحوثين يخضع للتوزيع الطبيعي العام، ونقبل الفرضية البديلة H_1 التي تقول إن أعمار المبحوثين لا تخضع للتوزيع الطبيعي العام.

7-3-2: اختبارات التناظر حول الوسيط Me **:**

نعلم أن الوسيط الكمي لمعلومات مرتبة Me يحقق االعلاقة الاحتمالية التالية:

$$P(x > Me) = P(X < Me) = \frac{1}{2}$$

ولدراسة التناظر حول الوسيط نفترض أن المتحول X يأخذ القيم، أو الحالات المرتبة التالية $X_1 \ , \ X_2 \ , \ X_3 ..... X_i ..... X_n$، وعندها يمكننا وضع الفرضيات كما يلي:

قيمة معلومة $H_0 : Me = M_0$

$H_1 : Me < M_0$ أو $H_1 : Me > M_0$ أو $H_1 : Me \neq M_0$.

وبناءً على ذلك يمكننا أن نقيس التناظر حول Me بواسطة أحد الاختبارات الآتية:

أ . اختبار الإشارة Sign Test:

لتطبيق هذا الاختبار نقوم بحساب الفروقات المختلفة لقيم X عن الوسيط المفترض Me؛ أي نحسب الفروقات وإشاراتها:

$$(X_i - M_0) \qquad \forall i : 1 \ , \ 2 \ , \ 3 ..... n$$

ثم نقوم بإحصاء الإشارات الموجبة (+) والإشارات السالبة (−) لهذه الفروقات.

ومن الواضح أنه إذا كان التناظر قائماً حول الوسيط المفترض Me، فإن عدد الإشارات الموجبة K^+ يجب أن يساوي عدد الإشارات السالبة K^-. أما إذا كان عدد الإشارات الموجبة أو السالبة قليلاً فإننا نرفض فرضية العدم حول التناظر، أو حول الوسيط المفترض Me.

ولتحديد العدد الحرج الذي نقبل عنده الفرضية H_0، أو نرفضها، نعتمد على التوزيع الثنائي الذي يكون فيه $P = \frac{1}{2}$، و n . ونحسب العدد k الذي يحقق العلاقة (أو يقارب تحقيقها):

$$P\left(K \leq \overset{+}{K} \right) = \sum_{r=0}^{K^+} C_n \left(\frac{1}{2}\right)^r \left(\frac{1}{2}\right)^{n-r} = \frac{\alpha}{2}$$

واختصاراً لهذه الحسابات نقوم بإعداد العدد الحرج K^* من الجدول (1) الخاص بذلك، وهو العدد الذي يقابل العدد n ومستوى الدلالة $\frac{\alpha}{2}$ (أول α الأحادي الجانب).

فإذا كان عدد الإشارات الموجبة k^+ أكبر أو يساوي من k^* فإننا نرفض الفرضية H_0.

حالة العينات الكبيرة: إذا كان حجم العينة n كبيراً فإننا نستخدم المتحول المعياري التالي:

$$Z = \frac{(K + 0.5) - 0.5n}{0{,}5\sqrt{n}}$$

ونرفض فرضية العدم إذا كانت قيمة Z المحسوبة أكبر أو تساوي $Z_{1-\frac{\alpha}{2}}$.

مثال 11:

لنفترض أن الوقت اللازم لتنفيذ أحد الأعمال بواسطة 12 عاملاً مختلفاً ومنفرداً كان كمايلي:

العامل	1	2	3	4	5	6	7	8	9	10	11	12
الوقت /ساعة/	3	3,1	3,3	3,5	3,8	4	4,2	4	4,5	5	5,5	6

والمطلوب:

دراسة تناظر هذه الأزمنة حول الوسيط، واختبار ذلك بمستوى دلالة $\alpha = 0{,}05$.

الحل:

نضع الفرضيتين كما يلي:

$$H_0 : Me = 4$$
$$H_1 : Me > 4$$

ثم نحسب الفروقات $(X_i - 4)$ وعدد إشاراتها الموجبة فنجد أن:

$(Xi-y):-1,-0.9,-0.7,-0.5,-0.5,-0.2,0,+0.2,+0.5,+1,+1.5,+2$

وهكذا نجد أن عدد الإشارات الموجبة $k^{+}=5$، وعدد الإشارات السالبة $k^{-}=7$. ومن الجدول الخاص بذلك ننجد أن العدد الحرج المقابل الذي يترك على يمينه احتمالاً قدره α هو 8. وبما أن $k^{+}<8$ لذلك نقبل الفرضية H_0 التي تقول أن الوسيط يساوي $Me=4$.

ب . اختبار (ويلكوكسون Wilcoxon **) حول الوسيط:**

ويتناول هذا الاختبار لتحديد الوسيط للمتحولات الكمية المستمرة المأخوذة من عينة عشوائية بحجم n، ومسحوبة من مجتمع متناظر. وبحيث تكون القياسات $x_1\ ,\ x_2\ ,\ x_3.....x_n$ مستقلة.

$$H_0: Me = M_0$$

$$H_1: Me \neq M_0 \text{ أو } H_1: Me < M_0 \text{ أو } H_1: Me > M_0$$

ثم نقوم بالإجراءات التالية:

1 . نحسب الفروقات المختلفة $(x_i - M_o)$ مع إشارتها.

2 . نقوم بوضع رتب للفروقات $|x_i - M_o|$ بقيمها المطلقة، ونرمز لتلك الرموز بـ R_i.

3 . نضع إشارة الفروقات $(x_i - M_o)$ على الرتب R_i.

4 . نجمع الرتب الموجبة، ونرمز لمجموعها بالرمز T^{+}.

5 . نجمع الرتب السالبة، ونرمز لمجموعها بالرمز T^{-}.

6 . نرمز لأصغر العددين (T^{+}، أو T^{-}) بالرمز T ، ونعتبر أن T هو مؤشر الاختبار المطلوب.

7 . نقارن قيمة T مع القيمة الحرجة لها T^{*} المأخوذة من الجدول (12) الخاص بذلك، والتي تقابل مستوى الدلالة α وحجم العينة n.

فإذا كانت $T<T^{*}$، نرفض الفرضية H_0 (انتبه بعكس العادة).

أما إذا كانت $T \geq T^{*}$، نقبل الفرضية H_0 (عكس العادة).

ملاحظة: إذا كان الاختبار أحادي الجانب نقارن T^+ مع T^* المقابلة لـ α و n، أو نقارن T^- مع T^* المقابلة لـ α و n، ونتخذ القرار المناسب (المقارنة عكس العادة).

*** حالة العينات الكبيرة:** إذا كان حجم العينة n كبيراً، فإننا نستخدم المتحول المعياري الآتي:

$$Z = \frac{T - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} \qquad (27-7)$$

وعندما يكون الاختبار أحادي الجانب نستبدل T^+ أو T^- بالعدد T.

ملاحظة:

إذا كانت الرتب تتضمن قيماً مكررة عدد كل منها t نقوم بطرح المقدار $\frac{\Sigma t^3 - \Sigma t}{48}$ من المقدار الموجود تحت الجذر في العلاقة (7−27).

مثال 12:

لنفترض أننا أجرينا اختباراً لذكاء على 15 طالباً فكانت نتائجهم كما يلي:

رقم الطالب	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
IQ	99	100	90	94	135	108	107	111	119	104	127	109	117	105	125

والمطلوب اختبار فيما إذا كان الوسيط مساوياً لـ 107.

الحل:

نضع الفرضيتين كما يلي:

$$H_0 : M = 107$$

$$H_1 : M \neq 107$$

ثم نقوم بإجراء الحسابات اللازمة ووضعها في الجدول التالي:

رقم الطالب	IQ	$D_i = x_i - 107$	رتب الفروقات بالقيمة المطلقة	رتب الفروقات مع الإشارة
1	99	−8	7	−7
2	100	−7	6	−6
3	90	−17	11	−11
4	94	−13	10	−10
5	135	+28	14	+14
6	108	+1	1	+1
7	107	0	−	نعزل الصفر من التحليل
8	111	+4	5	+5
9	119	+12	9	+9
10	104	−3	4	−4
11	127	+20	13	+13
12	109	+2	2,5	+ 2,5
13	117	+10	8	+8
14	105	−2	2,5	− 2,5
15	125	+18	12	+12
				$T^- = 40,5$
				$T^+ = 64,5$

وبما أن مؤشر الاختبار هو العدد T الذي يمثل أصغر العددين T^+ و T^-، نجد أن $T = 40,5$.

وبالعودة إلى الجدول (12) الخاص بقيم T نجد أن القيمة الحرجة للعدد T^* المقابلة لنصف مستوى الدلالة $\alpha = 0,05$ ولـ $n = 14$ (لقد أهملنا إحدى القيم لأن الفرق عندها كان يساوي الصفر) تساوي $T^* = 21$. وبما أن $T > 21$، لذلك نقبل فرضية العدم H_0، ونعتبر أن القيمة 107 هي وسيط رقم الذكاء IQ.

7-4 : اختبارات الارتباط بين متحولين رتيبيين (x,y) (من عينة واحدة):

لا تطبق اختبارات الارتباط على المتحولات الاسمية لأنها لا تقبل الترتيب، ولذلك فإن اختيارات الارتباط تشترط أن يكون المتحولان رتبيان على الأقل (يمكن أن يكون مجالي وتحويله إلى رتبي)، ولهما حالات متقابلة مكانياً أو زمانياً وعددها n زوجاً. ولمعالجة هذه الاختبارات نفترض أنه لدينا متحولان رتيبان (x,y)، ولهم الحالات المتقابلة التالية:

$$X : x_1 \ , \ x_2 \ , \ x_3 \ldots\ldots x_n$$
$$Y : y_1 \ , \ y_2 \ , \ y_3 \ldots\ldots y_n$$

ونضع الفرضيتين كما يلي:

H_0: المتحولان X و Y مستقلان أو عدم وجود ارتباط بينهما.

H_1: المتحولان X و Y مرتبطان. (ويكون الاختبار ثنائي الجانب لأن الارتباط يمكن أن يكون موجباً، أو سالباً).

وإن أهم مقاييس الارتباط بين المتحولات الرتيبة هي:

7-4-1: معامل الارتباط الرتبي (معامل سبيرمان):

ويعرف هذا المعامل بالعلاقة التالية:

$$RS = 1 - \frac{6\sum_{i=1}^{n}(r_i - q_i)^2}{n\left(n^2-1\right)} \qquad (28-7)$$

حيث إن R_i هي رتب المتحول الأول X.

وإن q_i هي رتب المتحول الثاني Y.

ولقد برهنا في كتاب مبادئ الإحصاء أن هذا المعامل هو حالة خاصة من معامل الارتباط العادي (معامل بيرسون)، لذلك فإنه يتمتع بجميع خواصه، وأهمها إنه محصور في المجال $[+1 \ , \ -1]$، وكلما كانت قيمته المطلقة $|RS|$ قريبة من الواحد، كان الارتباط قوياً.

ويمكن اختبار معنوية هذا المعامل باستخدام مؤشر التوزيع الطبيعي المعياري التالي:

$$z = \frac{RS - 0}{\frac{1}{\sqrt{n-1}}} = RS \cdot \sqrt{n-1} \qquad (29-7)$$

وبمقارنة قيمة Y المحسوبة مع $Z_{1-\frac{\alpha}{2}}$ الجدولية نتخذ القرار المناسب حول معنوية قيمة RS ولتطبيق هذا المعامل يجب القيام بالإجراءات التالية:

1 . ترتيب حالات المتحول الأول X تصاعدياً (أم تنازلياً) بوضع رتبة لكل حالة من حالاته وإعطائها رقماً صحيحاً من 1 حتى n. ونرمز لهذه الرتب بـ r_i.

2 . ترتيب حالات المتحول الثاني Y تصاعدياً (أم تنازلياً) بوضع رتبة لكل حالة من حالاته وإعطائها رقماً صحيحاً من 1 حتى n. ونرمز لهذه الرتب بـ q_i.

وهنا نشير إلى أن ترتيب حالات كل متحول يتم بشكل مستقل عن الآخر، ولكن بالاتجاه نفسه (تصاعياً، أم تنازلياً).

3 . نقوم بحساب الفروقات بين الرتب المتقابلة $(r_i - q_i)$، ثم نربعها $(r_i - q_i)^2$، ثم نأخذ مجموعها $\sum (r_i - q_i)^2$، ونعوضها في العلاقة (7-29) لحساب قيمة معامل الارتباط الرتبي RS.

مثال 13:

لدراسة العلاقة بين المستوى التعليمي للزوج، والمستوى التعليمي للزوجة، أخذنا عينة عشوائية بحجم $n = 7$ أسر، فحصلنا على البيانات التالية:

التسلسل	1	2	3	4	5	6	7
مستوى تعليم الزوج	أمي	ملم	ابتدائي	إعدادي	ثانوي	متوسط	جامعي
مستوى تعليم الزوجة	ملمة	أمية	إعدادية	ابتدائية	متوسط	ثانوية	ثانوية

والمطلوب دراسة الارتباط بين مستوى تعليم الزوج والزوجة.

الحل:

بما أن مستويي التعليم متحولان رتبيان (قابلان للترتيب)، فإنه يمكننا تطبيق معامل الارتباط الرتبي عليهما لحساب متانة الارتباط بينهما، لذلك وقبل كل شيء نقوم بما يلي:

1 . إعطاء رتب تصاعدية متسلسلة لحالات تعليم الزوج من 1 حتى 7 كما هي موضحة في الجدول.

2 . إعطاء رتب تصاعدية متسلسلة لحالات تعليم الزوجة من 1 حتى 7، علماً بأن كل من الحالتين الأخيرتين (ثانوية وثانوية) يمكن أن تأخذان 5 أو6، ولكن بما أنهما متكافئتان نعطيهما الرقم المتوسط 5,5 كما هو موضح في الجدول.

3 . نحسب الفروقات $(r_i - q_i)$، ثم نربعها، ونأخذ مجموعها ونحسب قيمة معامل الارتباط الرتبي RS.

مستوى تعليم الزوج	أمي	ملم	ابتدائي	إعدادي	ثانوي	متوسط	جامعي
رتب تعليم الزوج r_i	1	2	3	4	5	6	7
مستوى تعليم الزوجة	ملمة	أمية	إعدادية	ابتدائية	متوسط	ثانوية	ثانوية
رتب تعليم الزوجة q_i	2	1	4	3	7	5,5	5,5
$(r_i - q_i)$	−1	+1	−1	+1	−2	+ 0,5	+ 1,5
$(r_i - q_i)^2$	1	1	1	1	4	0,25	2,25

وبذلك نجد أنّ $\sum (r_i - q_i)^2 = 10,5$

وأنّ معامل الارتباط الرتبي بين المستويين يساوي:

$$RS = 1 - \frac{6\sum_{i=1}^{n}(r_i - q_i)^2}{n\left(n^2-1\right)} = 1 - \frac{6\cdot(10{,}5)}{7\cdot(49-1)}$$

$$RS = 1 - \frac{63}{336} = 0{,}8125$$

وأنّ قيمة RS تدل على درجة ارتباط جيد بين مستويي التعليم للزوج والزوجة.

ولدراسة معنوية هذه القيمة نحسب قيمة مؤشر الاختبار Z فنجد أن:

$$Z = RS\cdot\sqrt{n-1} = 0{,}8125\sqrt{7-1} = 1{,}99$$

وبمقارنة هذه القيمة مع قيمة $Z_{1-\frac{\alpha}{2}}$ الجدولية المقابلة لمستوى دلالة $\alpha = 0{,}05$ والتي تساوي $Z_{0{,}975} = 1{,}96$ نجد أن $Z > 1{,}96$، لذلك نرفض فرضية العدم H_0 (التي تقول إن $RS = 0$) ونقبل الفرضية البديلة التي تقول إن قيمة RS هي قيمة معنوية وتدل على وجود ارتباط جيد بين المتحولين المدروسين.

7-4-2: معامل ارتباط (كيندال Kendall) من النوع C:

ويستخدم هذا المعامل لدراسة متانة الارتباط بين متحولين رتيبين على الأقل، وأن حالاتهما المرتبة تتقاطع في جدول التكرارات كما يلي:

حالات x المرتبة / حالات y المرتبة	x_1	x_2	x_3	x_4
y_1	n_{11}	n_{12}	n_{13}	n_{14}
y_2	n_{21}	n_{22}	n_{23}	n_{24}
y_3	n_{31}	n_{32}	n_{33}	n_{34}

ولا يشترط في هذه الحالة أن يكون عدد حالات X مساوياً لعدد حالات Y، ولكن يشترط أن تكون حالات كل منهما مرتبة ترتيباً باتجاه واحد(تصاعدياً، أو تنازلياً). ونرمز لعدد الأسطر بالرمز r، ولعدد الأعمدة بالرمز c، ولأصغرهما بالرمز q. ويعرف معامل كيندال من النوع c بالعلاقة التالية:

$$\tau_C = \frac{q(P-Q)}{N^2(q-1)} \qquad \text{حيث إن } q = \min(r,c)$$

وإن n هو حجم العينة المدروسة.

و إن P تحسب من اليسار إلى اليمين بأخذ جداء كل تكرار في مجموع التكرارات التي دونه (بعد حذف سطره وعموده) ثم أخذ مجموعها؛ أي إن:

$$P = n_{11}(n_{22}+n_{23}+n_{24}+n_{32}+n_{33}+n_{34}) + n_{12}(n_{23}+n_{24}+n_{33}+n_{34}) + n_{13}(n_{24}+n_{34}) + n_{21}(n_{32}+n_{33}+n_{34}) + n_{22}(n_{33}+n_{34}) + n_{23}(n_{34})$$

وإن Q تحسب من اليمين إلى اليسار بأخذ جداء كل تكرار في مجموع التكرارات التي دونه (بعد حذف سطره وعموده) ثم أخذ مجموعها؛ أي إن:

$$Q = n_{14}(n_{23}+n_{22}+n_{21}+n_{33}+n_{32}+n_{31}) + n_{13}(n_{22}+n_{21}+n_{32}+n_{31}) + n_{12}(n_{21}+n_{31}) + n_{24}(n_{33}+n_{32}+n_{31}) + n_{23}(n_{32}+n_{31}) + n_{22}(n_{31})$$

ملاحظة: إن المعامل τ_C يأخذ قيمه في المجال $[+1\ ,\ -1]$، وكلما كانت قيمته المطلقة قريبة من الواحد كان الارتباط متيناً. ويمكن إجراء اختبار لمعنوية المعامل τ_C باستخدام مؤشر التوزيع المعياري التالي:

$$z_C = \frac{\tau_C - 0}{\sqrt{\frac{4(r+1)(c+1)}{9 \cdot n \cdot r \cdot c}}}$$

ثم اتخاذ القرار المناسب.

مثال 14:

لدراسة العلاقة بين مستويي التعليم للزوج، والزوجة، أخذنا عينة عشوائية كبيرة بحجم $n = 300$ أسرة، وبوبنا البيانات التالية:

حالات تعليم الزوج / حالات تعليم الزوجة	أمي	ابتدائي	إعدادي	ثانوي	جامعي
أمية	20	10	15	12	–
ابتدائية	10	15	20	10	5

إعدادية	5	10	25	30	10	
ثانوية	–	3	15	40	45	
المجموع	35	38	75	92	60	300

والمطلوب دراسة الارتباط بين مستوى تعليم الزوج والزوجة.

الحل:

نلاحظ أن المتحولين رتيبان، وأن حالاتهما مرتبة باتجاه تصاعدي. لذلك يمكننا حساب معامل كيندال من النوع c كما يلي:

$$\tau_C = \frac{q(P-Q)}{n^2(q-1)}$$

ولحساب P نأخذ جداء كل تكرار في مجموع التكرارات التي دونه (بعد حذف سطره وعموده) وذلك من اليسار إلى اليمين ثم نأخذ المجموع، فنجد أن:

$$\begin{aligned} P = {} & 20(15+20+10+5+10+25+30+10+3+15+40+45)+ \\ & 10(20+10+5+25+30+10+15+40+45)+ \\ & 15(10+5+30+10+40+45)+12(5+10+45) \\ & 10(10+25+30+10+3+15+40+45)+ \\ & 15(25+30+10+15+40+45)+20(30+10+40+45)+10(10+45) \\ & 5(3+15+40+45)+10(15+40+45)+25(40+45)+30(45) \end{aligned}$$

وأخيراً نجد أن:

$$P = 21675$$

ولحساب Q نأخذ جداء كل تكرار في مجموع التكرارات التي دونه (بعد حذف سطره وعموده) وذلك من اليمين إلى اليسار، فنجد أن:

$$Q = 12(20+15+10+25+10+5+15+3)+15(15+10+10+5)+$$
$$10(10+5)+5(30+25+10+5+40+15+3)+$$
$$10(25+10+5+15+3)+20(10+5+3)+15(5)+10(40+15+3)+$$
$$30(15+3)+25(3)=$$

وأخيراً نجد أن:

$$Q = 4836$$

وبملاحظة أن $q = \min(4,5)$ وأن $n = 300$، نجد أن قيمة معامل كيندال من النوع c تساوي:

$$\tau_C = \frac{4(21675-4836)}{(300)^2(4-1)} = 0,2495$$

وهذا يدل على وجود علاقة موجبة ولكنها ضعيفة بين مستويي تعليم الزوج والزوجة (حسب معطيات المسألة).

ولدراسة معنوية هذه القيمة نحسب مؤشر الاختبار Z_C، فنجد أن:

$$Z_C = \frac{\tau_C - 0}{\sqrt{\frac{4(r+1)(c+1)}{9 \cdot n \cdot r \cdot c}}} = \frac{0,2495}{\sqrt{\frac{4 \cdot 5 \cdot 6}{9 \cdot 300 \cdot 4 \cdot 5}}} = 5,2927$$

وبما أن قيمة $Z_{1-\frac{\alpha}{2}}$ الجدولية المقابلة $\alpha = 0,05$ تساوي 1,96، وأن $Z_C > 1,96$، لذلك نرفض فرضية العدم H_0 (التي تقول إن $\tau_C = 0$) ونعتبر أن القيمة التي حصلنا عليها معنوية، وتدل على وجود ارتباط بين المستويين المدروسين.

7-4-3: معامل الارتباط (غاما G) لمتحولين رتبيين (Goodman-Kruskal):

ويعرف هذا المعامل بالعلاقة التالية:

$$G = \frac{P-Q}{P+Q} \qquad (30-7)$$

حيث P و Q هما المعرفان بمعامل (كيندال) السابق، وهو يأخذ قيمه في المجال $[+1\ ,-1]$.

ويمكن اختبار معنوية القيمة التي نحصل عليها باستخدام مؤشر التوزيع الطبيعي المعياري المعرف بالعلاقة التالية:

$$Z_G = G \cdot \sqrt{\frac{P+Q}{\left[1-G^2\right]}} \qquad (31-7)$$

مثال 15:

لنأخذ المثال السابق ولنحسب المعامل RG لبياناته فنجد أن:

$$RG = \frac{21675-4836}{21675+4836} = \frac{16839}{25851} = 0{,}6352$$

وهو يدل على ارتباط جيد.

7-5: اختبارات الارتباط بين متحول اسمي ومتحول رتبي(من عينة واحدة):

إن أهم هذه الاختبارات هي:

1 . معامل (كوريتون Coreton**):**

وهو يدرس العلاقة بين متحول رتبي مثل الحالة التعليمية، ومتحول اسمي ثنائي مثل الجنس (ذكر، أو أنثى)، أو مكان الإقامة (حضر، أو ريف)، أو الاستجابة (نعم، أم لا)..إلخ. ويعرف معامل (كوريتون) بالعلاقة التالية:

$$RC = \frac{2}{n}(\bar{y}_2 - \bar{y}_1)$$

حيث إن:

. $\bar{y}_1$ هو متوسط رتب حالات المتحول الرتبي المقابلة للخاصة الأولى للمتحول الاسمي (للذكور، أو للحضر...).

. $\bar{y}_2$ هو متوسط رتب حالات المتحول الرتبي المقابلة للخاصة الثانية للمتحول الاسمي (للإناث، أو للريف...).

. n هو حجم العينة العشوائية.

ويأخذ هذا المعامل قيمه في المجال $[+1\ ,\ -1]$.

مثال 16:

ادرس العلاقة الارتباطية بين الحالة التعليمية والجنس من خلال البيانات التالية (عشرة أشخاص).

رقم الشخص	1	2	3	4	5	6	7	8	9	10
الحالة التعليمية	أمي	إعدادي	ابتدائي	ثانوي	إعدادي	جامعي	إعدادي	أمي	ثانوي	ابتدائي
الجنس	ذكر	أنثى	أنثى	أنثى	ذكر	ذكر	أنثى	أنثى	ذكر	أنثى
رتب الحالة التعليمية	1	3	2	4	3	5	3	1	4	2

والمطلوب دراسة الارتباط بين الحالة التعليمية والجنس.

الحل:

نفترض أن حالة الذكر هي الصفة الأولى لمتحول الجنس، وبذلك نجد أن متوسط الرتب التي تقابل حالة الذكر يساوي:

$$\bar{y}_1 = \frac{1+3+5+4}{4} = \frac{13}{4} = 3{,}25$$

ونفترض أن حالة الأنثى هي الصفة الثانية لمتحول الجنس، وبذلك نجد أن متوسط الرتب التي تقابل حالة الأنثى يساوي:

$$\bar{y}_2 = \frac{3+2+4+3+1+2}{6} = \frac{16}{6} = 2{,}5$$

وبذلك نجد أن معامل (كوريتون) بين التعليم والجنس يساوي:

$$RC = \frac{2}{10}(2{,}5 - 3{,}25) = -0{,}15$$

وهذا يدل على ارتباط عكسي ضعيف، وبميل لصالح الذكور ضد الإناث.

7-5-2: معامل ارتباط نقطة السلسلة المزدوجة (Point Biserial coeff):

ويطبق هذا المعامل على متحولين أحدهما كمي مجالي X، والآخر اسمي ثنائي يأخذ القيمة 1 عندما تظهر الصفة المدروسة، والقيمة 0 عند عدم ظهورها. ويعرف هذا المعامل بالعلاقة التالية:

$$R_{PB} = \sqrt{\frac{n_1 \cdot n_0}{n}} \cdot \left(\frac{\bar{x}_1 - \bar{x}_0}{\sqrt{\sum (x - \bar{x})^2}} \right)$$

حيث إن n هو حجم العينة.، و n_1 عدد حالات ظهور الخاصية، و n_0 عدد حالات عدم ظهورها.

وإن $\bar{x}_1$ هو متوسط قيم X المقابلة للقيم 1 التي يأخذها المتحول الاسمي.

وإن $\bar{x}_2$ هو متوسط قيم X المقابلة للقيم 0 التي يأخذها المتحول الاسمي.

وإن $\bar{x}$ هو متوسط قيم X في إجمالي العينة ولجميع قيم X.

ويأخذ هذا المتحول قيمه في المجال $[+1\ ,\ -1]$، وكلما كانت قيمه قريبة من الواحد كان الارتباط متيناً.

مثال 17:

ادرس فيما إذا كانت هناك علاقة ارتباطية بين الدخل الأسبوعي للفرد وحيازته على شهادة عليا، وذلك من البيانات التالية المأخوذة من عينة مؤلفة من 15 موظفاً.

التسلسل	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
الدخل ل.س	2400	1500	3500	3000	1200	2500	1300	2100	2700	2900	1900	2100	1400	1600	1000
حيازة شهادة عليا	1	1	0	1	0	1	0	1	1	1	0	1	0	0	0

لقد رمز بـ 1 لحيازة الموظف على شهادة عليا، وبـ 0 على عدم حيازته عليها.

الحل:

إن هذه العلاقة هي بين متحول كمي X هو مقدار الدخل الأسبوعي للفرد، ومتحول اسمي هو الحيازة على شهادة عليا الذي يأخذ إحدى القيمتين 1 أو 0 حسب ظهور الخاصة، أو عدم ظهورها.

ولدراسة هذه العلاقة نحسب معامل ارتباط نقطة السلسلة المزدوجة، ولذلك نقوم بحساب عناصره حيث نجد أن:

$$n = 15 \quad , \quad n_1 = 8 \quad , \quad n_0 = 7$$

ثم نحسب متوسط قيم X المقابلة للعدد (1) فقط فنجد أن:

$$\overline{x}_1 = \frac{2400 + 1500 + 300 + 2500 + 2100 + 2700 + 2900 + 2100}{8}$$

$$\overline{x}_1 = 2400 L.S.$$

ثم نقوم بحساب متوسط قيم X المقابلة للعدد (0) فقط فنجد أن:

$$\bar{x}_0 = \frac{3500+1200+1300+1900+1400+1600+1000}{7}$$

$$\bar{x}_0 = 1700 L.S.$$

ثم نقوم بحساب المتوسط العام لجميع قيم X حيث نجد أن:

$$\bar{x} = \frac{1}{15}\sum x_i = \frac{1}{15}\cdot 31100 = 2073{,}33\, L.S.$$

وبذلك نجد أن مجوع مربعات الانحرافات عن المتوسط العام يساوي:

$$\sum_{i=1}^{15}(x_i - \bar{x})^2 = \sum_{i=1}^{15}(x_i - 2073{,}33)^2 = 7809299{,}574$$

وأخيراً نقوم بحساب معامل ارتباط السلسلة المزدوجة، فنجد أن:

$$R_{PB} = \sqrt{\frac{8\cdot 7}{15}}\left(\frac{2400-1700}{7809299{,}57}\right) = 0{,}484$$

وهو يدل على درجة ارتباط مقبولة بين الدخل والشهادة العليا.

تمرينات

1 . تشترط وزارة الصناعة على شركة المصابيح ألا تزيد نسبة القطع المعيبة من المصابيح عن 10%، وللتأكد من ذلك سحبنا عينة من 15 قطعة فوجدنا أن فيها 3 قطع معيبة.

فهل هذا يدل على التزام شركة المصابيح بالمواصفات المطلوبة؟ اختبر بمستوى دلالة $\alpha = 0,05$.

2 . لنفترض أن الشركة اعترضت على العينة السابقة وطلبت إجراء اختبار آخر، ولذلك قامت اللجنة بسحب عينة أخرى بحجم n=50 مصباحاً، وكانت تحتوي على 6 قطع معيبة.

فهل تدل هذه النتيجة على التزام شركة المصابيح بالمواصفات المطلوبة؟ اختبر بمستوى دلالة $\alpha = 0,95$.

3 . عند رمي قطعة نقود 100 مرة كان عدد مرات ظهور الصورة 60 مرة، وعدد مرات ظهور الكتابة 40 مرة، وكان عدد مرات التعاقب 40 مرة. فهل تدل هذه النتائج على سلامة القطعة.

4 . في دراسة لقوة العمل حسب الحالة العملية والجنس تبين لنا ما يلي:

الحالة العملية / الجنس	مشتغل	متعطل
ذكور	320	150
إناث	130	80

فهل هناك اقتران بين الجنس والحالة العملية؟ (اختبر بمستوى دلالة $\alpha = 0,05$).

5 . إذا كان توزع السكان لبلدة معينة حسب الحالة التعليمية والجنس كما يلي:

الحالة العملية / الجنس	أمي	ملم	ابتدائي	إعدادي	ثانوي	جامعي	عليا
ذكور	100	130	80	150	200	50	10
إناث	180	150	90	10	70	20	5

والمطلوب:

1 . دراسة وجود ارتباط أو توافق بين الحالة التعليمية والجنس وذلك بمستوى دلالة $\alpha = 0,05$.

2 . اختبار فيما إذا كان التوزيع التجريبي للذكور يخضع للتوزيع المنتظم، وذلك بمستوى دلالة $\alpha = 0,10$.

3 . اختبار فيما إذا كان التوزيع التجريبي للإناث يخضع للتوزيع المنتظم، وذلك بمستوى دلالة $\alpha = 0,10$.

4 . اختبار فيما إذا كان التوزيع التجريبي للذكور يخضع لتوزيع المجموع (ذكور + إناث)، وذلك بمستوى دلالة $\alpha = 0,05$.

5 . اختبار فيما إذا كان التوزيع التجريبي للإناث يخضع لتوزيع المجموع (ذكور + إناث)، وذلك بمستوى دلالة $\alpha = 0,10$.

6 . إذا كان توزع النساء حسب أعمارهن عند الزواج الأول كما يلي:

العمر عند الزواج الأول	15	16	17	18	19	20	21	22	23	24	25	25+
التكرارات	10	15	20	30	50	60	80	100	70	50	30	20

فهل تدل هذه النتائج على أن أعمار النساء عند الزواج تخضع للتوزيع الطبيعي العام؟ وما هو متوسط تباينه؟ اختبر بمستوى دلالة $\alpha = 0,05$.

7 . لنفترض أن تبويب العمل حسب الحالة التعليمية وعدد أفراد الأسرة أعطانا الجدول التالي:

الحالة العملية / عدد أفراد الأسرة	أمي	ابتدائي	إعدادي	ثانوي
3	30	40	50	70
4	40	50	40	80
5	50	60	30	90
6	60	80	20	60
7	70	90	10	40
8	80	70	5	20
9	100	60	8	10

والمطلوب:

حساب معامل ارتباط كيندال من النوع c ثم اختار معنويته بمستوى دلالة $\alpha = 0,05$.

8 . ادرس فيما إذا كان هناك علاقة بين الدخل والجنس من المعطيات التالية:

الدخل	5500	7600	4500	8000	7500	6500	5600	6800
الجنس	ذكر	ذكر	أنثى	ذكر	أنثى	أنثى	ذكر	ذكر

وذلك باستخدام كل من معامل كوريتون، ثم معامل ارتباط نقطة السلسلة المزدوجة.

الفصل الثامن

مراقبة جودة الإنتاج

Quality control of production

1 . مقدمة.

2 . أهمية الرقابة على جودة الإنتاج.

3 . الرقابة الإحصائية على جودة الإنتاج.

4 . الأساليب لإحصائية في مراقبة الجودة.

5 . حالات أخرى في مراقبة جودة الإنتاج.

الفصل الثامن

مراقبة جودة الإنتاج

Quality control of production

8-1- مقدمة:

تعتبر الرقابة على نوعية الإنتاج من الوظائف الهامة في المنشآت الصناعية، التي من خلالها يتم التأكد من أن ما تم إنتاجه مطابق للمواصفات الموضوعة، وذلك من أجل مواظبة التطور الحاصل في إنتاج السلع. وقد برز موضوع الرقابة على نوعية الإنتاج منذ قيام الثورة الصناعية فقد ظهرت شركات كبيرة، وإن معظم هذه الشركات تنتج غالبية المنتجات، مما خلق فيما بينها نوع من التنافس من أجل السيطرة على الحصة الأكبر في السوق.

وحتى تتمكن هذه الشركات على المحافظة على مكانتها في السوق، عليها أن تنتج سلع على مستوى عال من الجودة. ومن أجل ذلك كان لابد من الاهتمام بالرقابة على نوعية المنتجات. وفي بداية العقد الثاني من القرن العشرين، اقترح شيوارت W. Shewhart خرائط الرقابة على نوعية الإنتاج، والتي سميت باسمه "خرائط شيوارت". ثم قام روميج ودوج H.G.Romig and H.F.Dodge بتصميم جداول عينات القبول في عام 1930، ثم تتالى الاهتمام بموضوع الرقابة على جودة المنتجات، فقد اقترح العالم الأمريكي ديمنج E.Deming طريقة إحصائية للرقابة على نوعية الإنتاج إلى الشركات الصناعية اليابانية، والتي كان لها الفضل في كون المنتجات اليابانية تمتلك قوة تنافسية عالمية.

وفي العقد السابع من القرن العشرين ظهر أسلوب آخر من الرقابة على النوعية، ويسمى بحلقات الرقابة على النوعية [1] Quality control Circles، والتي كان لهذا النوع من الرقابة أثر كبير على الجودة التي تتمتع بها المنتجات اليابانية.

---

[1] د. حسين عبد الله حسن التميمي، 1994، إدارة الإنتاج والعمليات، درا الحكمة اليمانية، اليمن ، صنعاء.

8-2: أهمية الرقابة على جودة الإنتاج:

إن مخرجات المشروعات الصناعية قد تكون جيدة ومطابقة للمواصفات الموضوعة، وقد تكون معيبة وغير مطابقة للمواصفات، وللكشف عن مدى سلامة الإنتاج (المخرجات)، ومطابقته للمواصفات الموضوعة، تمَّ تطوير الكثير من الأساليب الإحصائية المستخدمة في الرقابة على نوعية الإنتاج. وللحصول على منتج مطابق للمواصفات لابد وأن يتوفر عدد من الشروط منها:

1 . أن تكون الآلات والأدوات المستخدمة في العملية الإنتاجية سليمة.

2 . أن تتوفر لدى العاملين مهارة عالية في استخدام أدوات الإنتاج.

3 . أن تتوفر لدى العاملين دراية واسعة عن العملية الإنتاجية والمنتج.

4 . مطابقة المواد الأولية المستخدمة في العملية الإنتاجية لشروط المنتج.

5 . أن تتوفر الظروف الملائمة المحيطة بالعملية الإنتاجية.

وعند عدم توفر هذه الشروط في العملية الإنتاجية فإن المنتج يكون غير مطابقاً للمواصفات الموضوعة، وبالتالي تزداد نسبة التالف في الإنتاج، وأيضاً تزداد نسبة الهدر في المواد الأولية مما يؤدي إلى زيادة التكلفة الإنتاجية، وانخفاض الربحية، مما يتطلب من إدارة المشروع البحث والتقصي والكشف عن السبب، وتلافيه لتقليل نسبة التالف ونسبة الهدر في المواد الأولية، ويتم ذلك من خلال المراقبة المستمرة على مراحل العملية الإنتاجية، وفي النتيجة يتم الحصول على منتج مطابق للمواصفات الموضوعة، ويتمتع بجودة عالية ويلبي رغبات المستهلكين. ويستخدم لفظ الجودة للتعبير عن ملائمة السلعة، أو الخدمة المنتجة لأداء وظيفتها الأساسية، والحكم على جودة السلعة، أو الخدمة من وجهة نظر المستهلك تشمل ثلاث نواحي هي [2]:

المظهر والأداء ودرجة الثقة، أو الاعتماد على السلعة، أو الخدمة في أداء وظيفتها.

---

[2] عفيف شريف عبد الله، عطية محمد عطية، 1990، إدارة العملية الإنتاجية، عمان ، دار الفكر.

8-3: الرقابة الإحصائية على جودة الإنتاج:

عندما تكون مواصفات المنتج غير مطابقة للمواصفات الموضوعة مسبقاً، فهذا يعني أنه يوجد تغيّر معين وبنسبة معيّنة في مستوى الجودة للمنتج، وإن مقدار هذه النسبة سوف يؤدي بنا إلى قبول، أو رفض المنتج. فإذا كانت النسبة ضمن الحد المسموح به فإن المنتج يتم قبوله أما إذا كانت النسبة خارج الحد المسموح به فإننا نرفض المنتج. ولكن في الحياة العملية، وباعتبار أننا بصدد دراسة جودة الإنتاج، إننا نلاحظ أن جودة الإنتاج لن تكون أبداً مقداراً ثابتاً ومطلقاً، بل يوجد مقدار معين من التغير في مستوى الجودة عن المستوى النظامي حسب المواصفات الموضوعة، ويمكن أن يعزى سبب هذا التغير إلى أحد السببين التاليين: [3]

1 . أسباب طبيعية (طارئة) Chance causes:

وهذه الأسباب كثيرة جداً وتؤثر بشكل طفيف على جودة الإنتاج وتحصل هذه التغيرات صدفةً كنتيجة طبيعية للعمليات الإنتاجية، ومن الأسباب التي تؤدي إلى تغير جودة الإنتاج ومردها إلى الأسباب الطارئة مثل تدني مهارة العمال، وتغيُّر في نوعية المواد الأولية ... الخ.

2 . أسباب قابلة للتشخيص Assignable causes:

وهذه الأسباب تعود إلى التغيرات في عمليات الإنتاج مثل قدم الآلات المستخدمة في الإنتاج، أو المواد الأولية غير صحيحة... إلخ.

إن وجود أحد هذه الأسباب، أو عدد منها يؤثر على الإنتاجية، وبالتالي على جودة الإنتاج مما يتطلب من مهندسي الإنتاج ومراقبي الجودة العمل على إزالتها.

وعندما تكون الأسباب المؤدية إلى التغير في جودة الإنتاج هي أسباب طارئة فقط، فإننا نقول أن العملية الإنتاجية تحت السيطرة الإحصائية، أو مضبوطة إحصائياً Statistical control، وأن التغيرات الناتجة تعود إلى عامل الصدفة فقط، وبالتالي فإن

---

[3] د. محمد علي الأطرقجي، 1980، الرسائل التطبيقية في الطرق الإحصائية، دار الطليعة، بيروت.

إمكانية التغير في جودة الإنتاج لا يمكن أن تحصل إلا إذا تغيرت عملية الإنتاج نفسها، ومنه يمكن استخلاص نتيجة مفادها أن جميع المشاهدات التي تم اختيارها من أجل اختبار الجودة تكون مسحوبة من مجتمع إحصائي واحد.

أما عندما تكون الأسباب المؤدية إلى التغير في جودة الإنتاج ليست أسباب طارئة فقط، فإننا نقول إن العملية الإنتاجية ليست تحت السيطرة الإحصائية، أو غير مضبوطة إحصائياً، أو خارج المراقبة الإحصائية، وأن التغيرات الناتجة تعود إلى عوامل تؤثر بشكل مباشر في مستوى الجودة للمنتج، وبالتالي فإن وحدات الإنتاج سوف تقع خارج الحدود المسموح بها. وقد تؤدي هذه الأسباب إلى إيقاف العملية الإنتاجية بشكل جزئي، أو كلي مما يجب البحث والتقصي عن هذه الأسباب ومعالجتها.

وبذلك يمكن القول أن الهدف من مراقبة الجودة هو التقصي والبحث عن الأسباب المؤدية إلى التغير في جودة الإنتاج.

8-4: الأساليب الإحصائية في مراقبة الجودة:

لقد ذكرنا سابقاً أنه توجد العديد من الأساليب الإحصائية المستخدمة في مراقبة الجودة، وأن أهمية هذه الأساليب تنحصر في أنها تساهم في كشف الأسباب المؤدية إلى انحرافات مواصفات المنتجات عن المواصفات المقررة للمنتج. بالإضافة إلى أن عملية الرقابة على الجودة لاتنحصر فقط على المنتجات النهائية، وإنما ترافق جميع مراحل العملية الإنتاجية، حتى يتسنى للإدارة الإنتاجية الكشف عن المرحلة من العملية الإنتاجية والتي يكون فيها المنتج غير مطابق للمواصفات والتي تسبب في انحرافات مواصفات المنتج النهائي عن المواصفات المقررة، وبالتالي يمكن استبعاد تلك الأجزاء من العملية الإنتاجية حتى لايلحق الضرر بالمنتج ويصبح دون الحد المسموح به.

ولإجراء عملية الرقابة على جودة الإنتاج يتم سحب عينات عشوائية من السلع المنتجة، ومن ثم يقوم المعنيون بالرقابة على الإنتاج بفحص العينة، والتحقق من مطابقتها للمقاييس والمواصفات الموضوعة. وبناءً على نتائج العينة المأخوذة يمكن

قبـول، أو رفـض الإنتـاج. وتـدعى هـذه العينـات بـ " عينـات القبـول Acceptance Sampling" ويمكن أن تقسم عينات القبول إلى نوعين أساسيين هما[4]:

. عينات الخصائص Sampling by attributes:

وتعتمد على توزيع ثنائي الحد في فحص الإنتاج وتقرير فيما إذا كان مقبولاً أم مرفوضاً. وكمثال على ذلك فإن اللمبات الكهربائية تكون إما جيدة تضيء، أو تكون معطلة فلا تضيء.

. عينات المتغيرات Sampling by Variables:

وتعتمد هذه العينات على التوزيع الطبيعي في فحص الإنتاج، حيث يتم قياس مؤشرات الوزن، الطول، الكثافة، القطر، السماكة...إلخ، ومن ثم تحديد مقدار انحرافات مواصفات المنتج عن المواصفات المحددة مسبقاً، وبعدئذٍ تقرير فيما إذا كانت الانحرافات ضمن المدى المسموح به فيتم قبول المنتج. أما إذا كانت خارج المدى المسموح به فيتم رفض الإنتاج.

ويمكن تقسيم الأساليب الإحصائية في مراقبة الجودة إلى:

. أساليب تستخدم في مراقبة جودة المنتج النهائي.

. أساليب تستخدم في مراقبة الجودة للمنتج في كل مرحلة من مراحل العملية الإنتاجية.

أولاً: الأساليب المستخدمة في مراقبة جودة المنتج النهائي:

تتطلب عملية فحص الإنتاج سحبَ عينةٍ عشوائيةٍ، ومن ثم القيام بفحص وحداتها، وليس بالضرورة فحص الإنتاج بالكامل واستنتاج فيما إذا كان الإنتاج مطابقاً للمواصفات أم لا؟ وذلك من خلال مطابقة العينة للمواصفات الموضوعة. ويتم فحص الإنتاج باستخدام الأساليب الإحصائية التالية:

1 . العينات الفردية Single Sampling [5]:

([4]) عفيف شريف عبد الله، عطية محمد عطية. (1990). إدارة العمليات الإنتاجية، عمان، دار الفكر.

([5]) عفيف شريف عبد الله، عطية محمد عطية. (1990). إدارة العمليات الإنتاجية، عمان، دار الفكر.

وتعني العينة الفردية أن قرار قبول الإنتاج، أو رفضه يتم على أساس نتيجة فحص أول عينة يمكن اختيارها من الإنتاج. ويتم الفحص باستخدام العينات الفردية بالاعتماد على متغيرين اثنين هما: n حجم العينة التي تم اختيارها و X متغير يعبر عن عدد الوحدات المعيبة المسموح بها لكي يتم قبول الإنتاج. فعندما يتم اختيار عينة من الوحدات المنتجة بحجم n وحدة ويتم فحص وحدات هذه العينة فحصاً شاملاً؛ فإذا كان عدد الوحدات المعيبة في العينة أقل من X عندئذٍ يتم قبول الإنتاج. أما إذا كان عدد الوحدات المعيبة في العينة أكبر من X عندئذٍ يتم رفض الإنتاج. ويتم تحديد كلا المتغيرين n و X بالاعتماد على:

. نسبة الوحدات المعيبة من وجهة نظر المنتِج وتدعى مستوى القبول، ونرمز له بـ α.

. نسبة الوحدات المعيبة من وجهة نظر المستهلك وتدعى مستوى التحمل، ونرمز له بـ β.

ثم نقوم باتباع الخطوات التالية:

1 . نقسم مستوى التحمل على مستوى القبول ونبحث عن القيمة الناتجة في جداول إجراءات الفحص باستخدام العينات في العمود الثاني من الجدول رقم (1)، فإذا كانت القيمة الناتجة غير موجودة في الجدول فإننا نستخدم القيمة الأكبر من القيمة الناتجة مباشرةً من الجدول ومن العمود الثاني.

2 . يمكن الحصول على قيمة X من العمود الأول في الجدول رقم (1)، وتتحدد بالقيمة المقابلة للقيمة الناتجة في الخطوة الأولى.

3 . إن أرقام العمود الثالث هي عبارة عن حجم العينة مضروباً في مستوى القبول، وبالتالي يمكن الحصول على حجم العينة n من خلال تقسيم القيمة في العمود الثالث والمقابلة للقيمة الناتجة في الخطوة الأولى على مستوى القبول، وبالتالي نكون قد استطعنا تحديد حجم العينة n وعدد الوحدات المعيبة المسموح بها X ، ومن ثم نقوم بفحص وحدات العينة n فإذا كان عدد الوحدات المعيبة أقل من X

وهو العدد الذي تم تحديده في الخطوة الثانية فإننا نقبل الإنتاج. وإذا كان عدد الوحدات المعيبة أكبر من X فإننا نرفض الإنتاج بكامله.

الجدول رقم (1) جزء من جداول إجراءات الفحص باستخدام العينات

X	$\frac{\beta}{\alpha}$ = مستوى التحمل / مستوى القبول	حجم العينة × مستوى القبول $n \cdot X$
0	44,980	0,052
1	10,946	0,355
2	6,509	0,818
3	4,890	1,366
4	4,057	1,970
5	3,549	2,613
6	3,206	3,286
7	2,957	3,981
8	2,768	4,695
9	2,618	5,426

مثال 1:

يقوم أحد المشروعات الصناعية بإنتاج سلعة ما، وعند وجود 3% من الوحدات المنتجة معيباً يَعتبِر أن الإنتاج مقبول. لنفرض قدوم أحد المستهلكين لشراء كمية من إنتاج هذا المشروع، واعتبر أن وجود 9% وحدات معيبة على الأكثر من الكمية المشتراة أمراً مقبولاً.

المطلوب: استخدام هذه البيانات في الحكم على جودة إنتاج السلعة.

الحل:

لإيجاد حجم العينة وعدد الوحدات المعيبة المسموح بها نتبع ما يلي:

. نقسم مستوى التحمل β على مستوى القبول α فنجد أن:

$$\frac{\beta}{\alpha} = \frac{0,09}{0,03} = 3$$

. نبحث عن القيمة (3) في العمود الثاني فنجد أنها غير موجودة، وبالتالي نأخذ القيمة التي هي أكبر منها مباشرةً وتكون 3,206 وتقابل هذه القيمة قيمة X في العمود الأول وتساوي 6؛ أي $X = 6$ ؛ أي إن عدد الوحدات المعيبة الذي يمكن قبوله هي 6 وحدات على الأكثر.

ـ لتحديد حجم العينة n نقسم الرقم المقابل للقيمة الناتجة في الخطوة الأولى وهو 3,206 على مستوى القبول α فنجد أن حجم العينة يساوي:

$$n = \frac{3,286}{0,03} = 109,53 \approx 110$$ وحدة

حيث يتم تقريب حجم العينة للعدد الصحيح الأكبر.

. يتم القرار بقبول الإنتاج، أو رفضه على أساس اختيار عينة مؤلفة من 110 وحدات منتجة، ويتم فحص هذه الوحدات. فإذا تبين أن عدد الوحدات المعيبة في العينة أكثر من 6 وحدات معيبة عندئذٍ يتم رفض الإنتاج وتلغى الطلبية لعدم مطابقة الإنتاج المواصفات المتفق عليها. وإذا تبين أن عدد الوحدات المعيبة في العينة أقل من 6 وحدات معيبة فإن المستهلك يقر بقبول الإنتاج ومطابقته للمواصفات المتفق عليها.

2 . العينات الثنائية Double Sampling:

ويتم بموجب هذا الأسلوب سحب عينة صغيرة بحجم n_1 من الوحدات المنتجة، ثم يتم تحديد الحد الأدنى المسموح به لعدد الوحدات المعيبة ونرمز له بـ k_1 ونرمز للحد الأعلى المسموح به لعدد الوحدات المعيبة بـ k_2 . وعند إجراء الفحص على وحدات العينة n_1 فإذا كان عدد الوحدات المعيبة في العينة المسحوبة أقل من الحد الأدنى المسموح به k_1 فإنه يتم قبول الإنتاج. أما إذا ازداد عدد الوحدات المعيبة في العينة المسحوبة عن الحد الأقصى المسموح به k_2 فإنه يتم رفض الإنتاج ودون الحاجة إلى أخذ عينة ثانية. أما إذا كان عدد الوحدات المعيبة محصوراً بين الحد الأدنى المسموح به k_1 والحد الأعلى المسموح به k_2، عندئذٍ يتم سحب عينة ثانية n_2. فإذا كان مجموع عدد الوحدات المعيبة في العينتين المسحوبتين أقل من الحد

الأدنى المسموح به k_1 فإنه يتم قبول الإنتاج. أما إذا كان عدد الوحدات المعيبة من العينة الأولى n_1 والعينة الثانية n_2 لا يزيد عن الحد الأقصى المسموح به لعدد الوحدات المعيبة k_2 فإنه يتم قبول الإنتاج، أما إذا زاد عن k_2 فإنه يتم رفض الإنتاج.

مثال 2:

اتفق أحد المستهلكين على شراء 1000 لعبة من ألعاب الأطفال من شركة لإنتاج ألعاب الأطفال واتفقا على اختيار عينة مؤلفة من $n_1 = 30$ لعبة، وفحصها بعد أن تم تحديد الحد الأدنى والحد الأعلى المسموح بهما لعدد اللعب المعيبة في العينة الأولى؛ أي [2–5]، والمطلوب:

1 . ما هو قرار المستهلك عندما يكون عدد اللعب المعيبة في العينة يساوي 1.

2 . ما هو قرار المستهلك عندما يكون عدد اللعب المعيبة في العينة يساوي 6.

3 . ما هو قرار المستهلك عندما يكون عدد اللعب المعيبة في العينة يساوي 3.

الحل:

1 . باعتبار أن عدد اللعب المعيبة في العينة يساوي لعبة واحدة وهي أقل من الحد الأدنى المسموح به $k_1 = 2$ عندئذٍ يقرّ المستهلك بقبول الكمية من الألعاب.

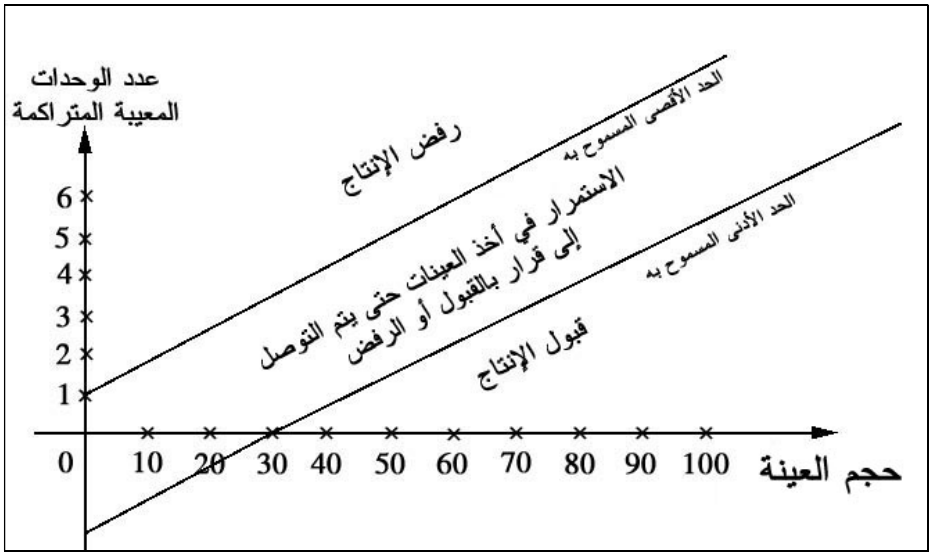
2 . باعتبار أن عدد اللعب المعيبة في العينة يساوي 6 لعب، وهو أكثر من الحد الأعلى المسموح به $k_2 = 5$ عندئذٍ يقرّ المستهلك برفض الكمية من الألعاب.

3 . باعتبار أن عدد اللعب المعيبة في العينة يساوي 3 وهذا العدد محصور بين الحد الأدنى والحد الأعلى المسموح بهما، عندئذٍ نقوم بسحب عينة ثانية n_2 وبحجم $n_2 = 20$ لعبة ويتم فحصها. ولنفرض أنه وجد أن عدد اللعب المعيبة في العينتين بلغ 5 لعب، عندئذٍ يقر بقبول الكمية المشتراة من الألعاب.

3 . العينات المتتالية Sequential Sampling:

وهي عبارة عن امتداد لأسلوب العينات الثنائية حيث يتم بموجب هذا الأسلوب اختيار أكثر من عينتين من كمية محددة من الإنتاج، ثم يتم وضع حدٍ أدنى، وحدٍ أقصى مسموحٌ بهما لعدد الوحدات المعيبة لكل عينة. ومن خلال النتائج المتراكمة

لعملية فحص وحدات العينات المتتالية يتم رفض، أو قبول الإنتاج. فإذا كان عدد الوحدات المعيبة المتراكم أقل من أو يساوي الحد الأدنى المسموح به فإنه يتم قبول الإنتاج. أما إذا كان عدد الوحدات المعيبة المتراكم أكبر من الحد الأقصى المسموح به فإنه يتم رفض الإنتاج. وفي حال وقوع عدد الوحدات المعيبة بين الحدين الأدنى والأقصى المسموح بهما، فإننا نستمر بأخذ عينات جديدة من الكمية المنتجة إلى أن نتحقق من رفض الإنتاج، أو قبوله. ويمكن توضيح ذلك بالشكل الآتي:



الشكل رقم (8–1) يبين منطقتي القبول والرفض باستخدام العينات المتتالية.

يتطلب استخدام العينات المتتالية في اتخاذ قرار حول قبول المنتج، أو رفضه توافر بعض المعلومات عن مخاطر المنتج ونرمز له بالرمز M_1، ومخاطر المستهلك ونرمز له بالرمز M_2، بالإضافة إلى مستوى القبول α ، ومستوى التحمل β.

وبالتالي فإن الحد الأقصى والحد الأدنى المسموح بهما لعدد الوحدات المعيبة هما عبارة عن خطوط مستقيمة، ويمكن رسم كلٍ منهما بتحديد نقطتين فقط والوصل بينهما، حيث يمكن التعبير عن الحد الأقصى بالمعادلة التالية:

$$X_1 = An + c_2 \qquad (1-11)$$

ويمكن التعبير عن الحد الأدنى بالمعادلة التالية:

$$X_2 = An - c_1 \qquad (2-8)$$

حيث:

n حجم العينة.

X_1 و X_2 متغير يعبر عن عدد الوحدات المعيبة المتراكم بالنسبة للحد الأقصى والحد الأدنى.

A ميل خط الانحدار.

c_1 و c_2 ثابتين.

ويمكن إيجاد الثوابت c_1 و c_2 بالعلاقات التالية:

$$c_1 = \frac{\lg \dfrac{(1-M_1)}{M_2}}{\lg \dfrac{\beta(1-\alpha)}{\alpha(1-\beta)}} \qquad (3-8)$$

$$c_2 = \frac{\lg \dfrac{(1-M_2)}{M_1}}{\lg \dfrac{\beta(1-\alpha)}{\alpha(1-\beta)}} \qquad (4-8)$$

ومنه فإن A تساوي:

$$A = \frac{\lg \dfrac{(1-\alpha)}{(1-\beta)}}{\lg \left[\dfrac{\beta(1-\alpha)}{\alpha(1-\beta)}\right]} \qquad (5-8)$$

وبعد إيجاد المعادلات نقوم برسم المستقيمين الممثلين للحد الأدنى والحد الأقصى للوحدات المعيبة المسموح بها. وسنقوم بعرض المثال التالي لكي تتوضح عملية استخدام العينات المتتالية في مراقبة جودة الإنتاج.

مثال 3:

لنأخذ معطيات المثال رقم (1) بالإضافة إلى أن المشروع الصناعي يرغب في أن يحقق قبول إنتاج جيد باحتمال قدره 0,95. ولايرغب المستهلك في قبول إنتاج تصل نسبة المعيب فيه إلى أكثر من 10% من الحالات.

والمطلوب: وضع خطة فحص للإنتاج باستخدام العينات المتتالية.

الحل:

نلاحظ أن مستوى القبول α يساوي: $\alpha = 0,03$ ، ومستوى التحمل β يساوي: $\beta = 0,09$.

مخاطر المنتج M_1 وتساوي $M_1 = 0,05$.

مخاطر المنتج هو عبارة عن احتمال رفض الإنتاج الجيد.

ومخاطر المستهلك M_2 ويساوي $M_2 = 0,10$

مخاطر المستهلك هي عبارة عن احتمال قبول الإنتاج الرديء.

إن المستقيم الذي يعبر عن الحد الأقصى لعدد العيوب المسموح بها يعطى بالعلاقة:

$$X_1 = An + c_2$$

وأن المستقيم الذي يعبر عن الحد الأدنى لعدد العيوب المسموح بها يعطى بالعلاقة:

$$X_2 = An - c_1$$

لنقوم بحساب الثوابت c_1 و c_2 و A من خلال العلاقات السابقة (8-3) و(8-4) و(8-5):

$$C_1 = \frac{\lg \frac{(1-0,05)}{0,10}}{\lg \frac{0,09(1-0,03)}{0,03(1-0,09)}} = 1,937$$

$$C_2 = \frac{\lg \frac{(1-0,10)}{0,05}}{\lg \frac{0,09(1-0,03)}{0,03(1-0,09)}} = 2,486$$

و A تساوي:

$$A = \frac{\lg \frac{(1-0,03)}{(1-0,09)}}{\lg \frac{0,09(1-0,03)}{0,03(1-0,09)}} = 0,055$$

وبالتالي تصبح المعادلات السابقة بالشكل:

$$X_1 = 0,055\,n + 2,486$$
$$X_2 = 0,055\,n - 1,937$$

ولرسم كل من المعادلتين يكفي تعيين نقطتين لكل منهما أي:

عندما يكون حجم العينة صفر فإن: $n = 0 \Rightarrow X_1 = 2,486$

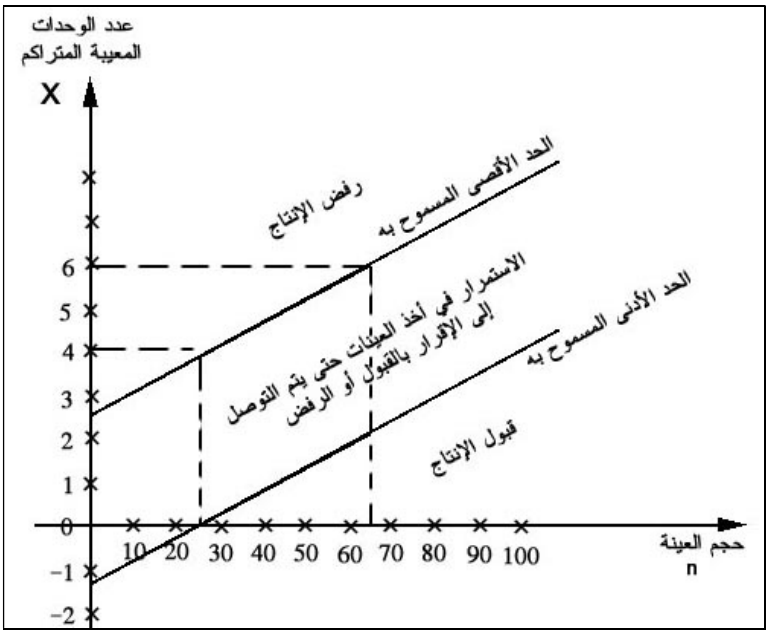
عندما يكون حجم العينة 100 فإن: $n = 100 \Rightarrow X_1 = 7,986$

وبذلك نكون قد حصلنا على نقطتين هما $(0\ ,\ 2,486)$, $(100\ ,\ 7,986)$ نقوم بتحديدهما على جملة محاور إحداثية لتحديد مستقيم الحد الأقصى لعدد الوحدات المعيبة المسموح به.

بالنسبة للمستقيم الثاني نجد:

$$n = 0 \Rightarrow X_2 = -1,937 \Rightarrow \quad (0\ ,\ -1,937)$$
$$n = 100 \Rightarrow X_2 = 3,563 \Rightarrow \quad (100\ ,\ 3,563)$$

وبذلك نكون قد حصلنا على إحداثيات مستقيم الحد الأدنى لعدد الوحدات المعيبة المسموح به. لنحدد المستقيمين على جملة محاور إحداثية فنجد أن:



الشكل رقم (8-2)

من الشكل البياني نلاحظ أن أول قرار بالقبول، أو الرفض يكون عند حجم العينة $n = 25$، فإذا كان عدد الوحدات المعيبة صفراً فإنه يتم قبول الإنتاج، وإذا كان أكبر من 4 وحدات معيبة فإنه يتم رفض الإنتاج فوراً.

أما إذا كان عدد الوحدات المعيبة بين صفر وأربع وحدات فإننا نأخذ عينة ثانية $n_2 = 65$ ونفحص وحداتها فإذا كان عدد الوحدات المعيبة المتراكمة أكبر من 6 فإننا نرفض الإنتاج... وهكذا...

ثانياً: أساليب مراقبة مراحل العملية الإنتاجية:

يهدف هذا الأسلوب إلى مراقبة جودة السلعة في كل مرحلة من مراحل العملية الإنتاجية وذلك بالاعتماد على أساس سحب عينات صغيرة ذات حجوم متساوية ودورية من السلعة في كل مرحلة من مراحل العملية الإنتاجية لها، ومن ثم تجري عليها عملية الاختبار والفحص. فإذا تبين من نتيجة الفحص أنها مطابقة للمعايير والأسس والقياسات الموضوعة، عندئذٍ يُسمح للعملية الإنتاجية بالاستمرار، أما إذا كانت نتيجة فحص السلعة غير مطابقة للمعايير والأسس والقياسات فإنه يتم الإيعاز بإيقاف العملية

الإنتاجية واتخاذ الإجراءات التصحيحية اللازمة إذا توفرت إمكانية لتغيير مسار العملية الإنتاجية. ولذلك تعتبر الرقابة الإحصائية على جودة المنتج أثناء مراحل العملية الإنتاجية وقائية لأنها تهدف إلى كشف انحرافات مواصفات السلعة عن المعايير والأسس الموضوعة لها مسبقاً. كما تعتبر الرقابة الإحصائية علاجية لكونها تبين مواطن الخلل والانحرافات في كل مرحلة وتؤكد على ضرورة اتخاذ الإجراءات العلاجية اللازمة لتصحيح مسار العملية الإنتاجية للسلعة. وبذلك يتوضح الهدف من مراقبة جودة السلعة أثناء مراحل العملية الإنتاجية الذي يتمثل في سرعة اكتشاف الانحرافات عن الحدود المسموح بها فور حدوثها واتخاذ الإجراءات اللازمة بصددها. أما بالنسبة لما يخص حجم العينة فيفضل أن تكون صغيرة جداً؛ أي تتكون من أربع، أو خمس وحدات، ويتم اختيارها على فترات زمنية محدودة وبشكلٍ متتالٍ.

والجدير بالذكر أن انحرافات معايير ومواصفات المنتج عن المعايير والمواصفات المحددة سلفاً ومسبقاً يمكن أن تنشا نتيجة نوعين من المتغيرات هما:

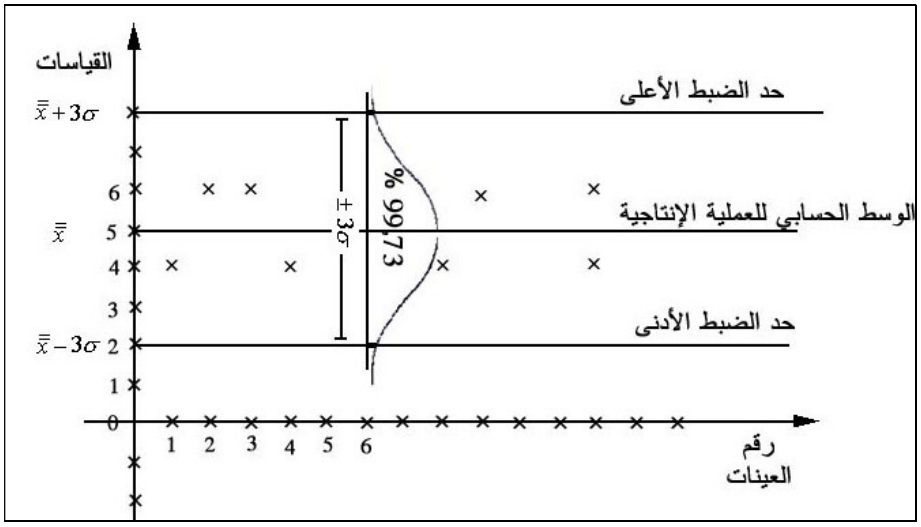
1 . متغيرات، أو انحرافات عادية، أو عشوائية لا يمكن ضبطها، أو التحكم بها، ولا تؤثر على الإنتاج من حيث مطابقته للمواصفات المحددة؛ أي تبقى الانحرافات ضمن الحدود المسموح بها، وبالتالي يمكن القول إن العملية مضبوطة إحصائياً، أو محكّمة، أو تحت السيطرة.

2 . متغيرات، أو انحرافات غير عادية يمكن ضبطها والتحكم فيها، وتؤثر على الإنتاج من حيث مطابقته للمواصفات المحددة، وبالتالي فإن الانحرافات تكون خارج الحدود المسموح بها. وعندئذٍ يمكن القول إن العملية غير مضبوطة، أو غير محكّمة، أو ليست تحت السيطرة.

ومن أهم أساليب الرقابة الإحصائية Statistical Control على مراحل العملية الإنتاجية هي خرائط ضبط الجودة أو خرائط الرقابة Control Charts، حيث يتم إعداد خرائط ضبط الجودة على شكل رسم بياني يتحدد عليه الحد الأدنى والحد الأعلى للضبط، وذلك بالاعتماد على سحب عينات صغيرة وذات حجوم متساوية، وعلى فترات

زمنية متساوية. ومن ثم يتم حساب متوسطات هذه العينات حسب نوع الخريطة (خريطة الوسط الحسابي، خريطة المدى، خريطة الانحراف المعياري، خريطة نسبة وعدد الوحدات المعيبة). وتؤشّر على الخريطة مباشرةً، بالإضافة إلى تحديد المتوسط العام للعينات والذي هو عبارة عن متوسط الأوساط الحسابية للعينات. وعندما تكون متوسطات العينات بين حدي الضبط يمكننا القول إن العملية مضبوطة إحصائياً، أو تحت السيطرة، وهذا يعني أن يعزى جزء كبير من التغيرات في العينات إلى التغيرات العشوائية، أو للصدفة. أما إذا كانت متوسطات العينات خارج حدي الضبط، يمكننا القول إن العملية غير مضبوطة إحصائياً، أو خارج السيطرة.

ويمكن تحديد حدي الضبط العلوي والسفلي من خلال الوسط الحسابي ± 3 انحراف معياري، وبالتالي يُتَوَقْع أن تحصر خريطة الضبط بين حديها 99,73 % من متوسطات العينات المحسوبة، ومنه يخضع التوزيع الاحتمالي لمتوسطات العينات للتوزيع الطبيعي. ويمكن توضيح ما تقدم بالشكل البياني التالي:



الشكل رقم (8-3)

وبالنظر إلى خريطة الضبط الإحصائية يمكن معرفة مستوى جودة الإنتاج ومدى مطابقته للمواصفات والمعايير الموضوعة. وإن الهدف من خريطة الضبط الإحصائية هو التمييز بين الانحرافات الطبيعية ذات المنشأ العشوائي والانحرافات غير الطبيعية

والتي يمكن ضبطها والتحكم فيها، بالإضافة إلى أنها تبين مدى كون العملية الإنتاجية تحت السيطرة الإحصائية أو مضبوطة إحصائياً. ويوجد لدينا نوعان من خرائط ضبط الجودة (خرائط الرقابة على الجودة) وهما:

1 . خرائط الرقابة على المتغيرات (خرائط ضبط المتغيرات) Control Charts for Variables.

2 . خرائط الرقابة على الخصائص (خرائط ضبط المواصفات) Control Charts for Attributes.

1 . خرائط الرقابة على المتغيرات (خرائط ضبط المتغيرات) Control Charts for Variables:

تستخدم هذه الخرائط في الرقابة على متغيرات لها قياسات ومعايير محددة [6] مثل الوزن، الطول، اللون، المتانة، الضغط، درجة الحرارة...إلخ، وذلك لمعرفة انحرافات هذه المتغيرات عن المواصفات المحددة مسبقاً لها. وذلك من خلال الوسط الحسابي والمدى والانحراف المعياري. ومن الخرائط المستخدمة في الرقابة على المتغيرات:

. خريطة ضبط الوسط الحسابي والمدى.

. خريطة ضبط الوسط الحسابي والانحراف المعياري.

أ . خريطة ضبط الوسط الحسابي $\overline{X}$ Charts :

بعد أن تسحب جميع العينات ولتكن m ذات الحجوم المتساوية، خلال مراحل العملية الإنتاجية، فإنه يتم حساب متوسطات العينات $\bar{x}_i$ ، حيث $i = 1 , 2 , ... , m$ ومن ثم يتم حساب المتوسط الحسابي العام والذي نرمز له بـ $\overline{\overline{X}}$ وهو عبارة عن مجموع الأوساط الحسابية للعينات مقسوماً على عدد العينات؛ أي:

([6]) حسين عبد الله التميمي، (1994)، إدارة الإنتاج والعمليات، اليمن، صنعاء، دار الحكمة اليمانية.

$$\bar{\bar{X}} = \frac{\sum_{i=1}^{m} \bar{x}_i}{m} \qquad (6-8)$$

مع العلم أن توزيع الأوساط الحسابية للعينات ياخذ شكل المنحني الطبيعي، وأن الوسط الحسابي للعملية الإنتاجية $\bar{\bar{X}}$ يساوي الوسط الحسابي للمجتمع الإحصائي الذي سحبت منه العينات. وأن تباين الأوساط الحسابية للعينات $\sigma_{\bar{x}}^2$ يساوي تباين المجتمع مقسوماً على حجم العينة n؛ أي:

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} \qquad (7-8)$$

وعندما يكون تباين المجتمع σ^2 مجهولاً فإننا نستخدم تباين العينة s^2 كتقدير غير متحيز لتباين المجتمع، وعندئذٍ تصبح العلاقة (8–7) بالشكل:

$$\sigma_{\bar{x}}^2 = \frac{s^2}{n}$$

حيث:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{m}(x_i - \bar{x})^2$$

وبالتالي فإن حدي الضبط العلوي والسفلي التي يتوقع أن تحصر بينهما 99.73 % من متوسطات العينات المسحوبة تحسب من العلاقات التالية:

$U_{\bar{x}}$ حد الضبط الأعلى للمتوسطات:

$$U_{\bar{x}} = \bar{\bar{X}} + 3\sigma_{\bar{x}} \qquad (8-8)$$

$L_{\bar{x}}$ حد الضبط الأدنى للمتوسطات:

$$L_{\bar{x}} = \bar{\bar{X}} - 3\sigma_{\bar{x}} \qquad (9-8)$$

وبتعويض الانحراف المعياري $\sigma_{\bar{x}}$ في العلاقتين (8–8) و(8–9) نحصل على حدود الضبط:

$$U_{\bar{x}} = \overline{\overline{X}} + 3\left(\frac{\sigma}{\sqrt{n}}\right)$$
$$L_{\bar{x}} = \overline{\overline{X}} - 3\left(\frac{\sigma}{\sqrt{n}}\right) \qquad (10-8)$$

أما عندما نستخدم الانحراف المعياري للعينة كتقدير للانحراف المعياري للمجتمع، فإن حدود الضبط تصبح بالشكل التالي:

$$U_{\bar{x}} = \overline{\overline{X}} + 3\left(\frac{s}{\sqrt{n}}\right)$$
$$L_{\bar{x}} = \overline{\overline{X}} - 3\left(\frac{s}{\sqrt{n}}\right) \qquad (11-8)$$

ويمكن رسم حدي الضبط للمتوسط باستخدام العلاقة التالية:

$$\overline{\overline{X}} \pm A_2 \overline{R} \qquad (12-8)$$

حيث $\overline{R}$ متوسط المدى لجميع العينات ويساوي مدى العملية الإنتاجية ويحسب من العلاقة التالية:

$$\overline{R} = \frac{\sum_{i=1}^{m} R_i}{m} \qquad (13-8)$$

و R_i هي مدى العينات و m ,....., 3 , 2 , $i = 1$

A_2 ثابت يستخرج من جداول إنشاء خرائط الرقابة الإحصائية في الملحق الجدول رقم (3) الملحق في نهاية هذا الفصل. وبعد أن يتم رسم خريطة الرقابة الإحصائية وتوقع عليها متوسطات العينات وحدود الضبط ومتوسط العملية الإنتاجية $\overline{\overline{X}}$ فإنه بالنظر إلى خريطة الرقابة الإحصائية يمكن اتخاذ القرار الإحصائي برفض، أو قبول المنتج. فإذا كانت جميع متوسطات العينات ضمن حدي الضبط فإننا نقر بأن العملية الإنتاجية مضبوطة إحصائياً.

أما عندما يوجد متوسط إحدى العينات خارج حدي الضبط، فإننا نكون متأكدون باحتمال قدره 99,73 % بأن العملية الإنتاجية قد تغيرت، أو ليست تحت السيطرة الإحصائية، وتحتاج إلى إجراءات تصحيحية. وتستخدم هذه الطريقة عندما تكون العملية الإنتاجية تنتج أعداداً كبيرة من المنتجات، وبالتالي لإجراء عمليات الضبط نسحب عينات صغيرة بانتظام من المجتمع الإحصائي.

مثال 4:

لدراسة الرقابة على إنتاج مصنع تعبئة المياه الغازية ذات الحجم العائلي 2,25 لتر، تم سحب عينات بحجم 4 عبوات بمعدل كل ساعة عينة ولمدة ثماني ساعات ابتداءً من الساعة الثامنة صباحاً.

وكانت البيانات التي تم الحصول عليها مبينة في الجدول التالي:

العينات	سعة العبوات				المجموع
	1	**2**	**3**	**4**	
1	2,25	2,2	2,3	2,4	9,15
2	2,3	2,25	2,2	2,1	8,85
3	2,3	2,25	2,4	2,4	9,35
4	2,4	2,3	2,3	2,1	9,1
5	2,2	2,2	2,25	2,25	8,9
6	2,35	2,35	2,25	2,2	9,15
7	2,25	2,45	2,20	2,4	9,3
8	2,1	2,2	2,4	2,2	8,9

والمطلوب:

1 . ارسم خريطة المتوسط.

2 . تحديد فيما إذا كانت العملية الإنتاجية مضبوطة إحصائياً.

الحل:

لإجراء المطلوب نقوم بحساب متوسطات العينات وتبايناتها، ومن ثم المتوسط الحسابي العام $\overline{\overline{X}}$ والذي يعطى بالعلاقة:

$$\overline{\overline{X}} = \frac{\sum_{i=1}^{m} \overline{x}_i}{m}$$

ونضع النتائج في الجدول التالي:

رقم العينة	1	2	3	4	5	6	7	8	المجموع
متوسطات العينات $\overline{x}_i$	2,29	2,21	2,34	2,28	2,23	2,29	2,33	2,23	18,2
تباين العينات s_i^2	0,0073	0,0073	0,0056	0,0159	0,0009	0,0056	0,0142	0,0159	0,0727

ومنه فإن:

$$\overline{\overline{X}} = \frac{18,2}{8} = 2,275$$

$$s^2 = \frac{\sum_{i=1}^{m} s_i^2}{m} \Rightarrow s^2 = \frac{0,0727}{8} = 0,0091$$

وبالتالي نستطيع حساب تقدير تباين المتوسط بالعلاقة:

$$\tilde{\sigma}_{\overline{x}}^2 = \frac{s^2}{n}$$

$$\tilde{\sigma}_{\overline{x}}^2 = \frac{0,0091}{4} = 0,0023$$

$$\tilde{\sigma}_{\overline{x}}^2 = \sqrt{0,0023} = 0,048$$

وبالتالي نستطيع الآن حساب حدود الضبط لخريطة المتوسط:

الحد الأدنى للضبط:

$$L_{\overline{x}} = \overline{\overline{X}} - 3\left(\frac{s}{n}\right) =$$

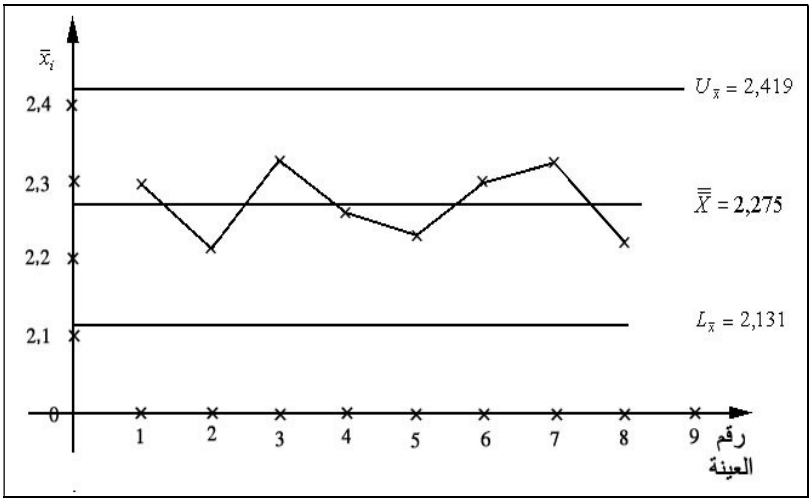
$$L_{\overline{x}} = 2,275 - 3 \cdot 0,048 = 2,131$$

الحد الأعلى للضبط:

$$U_{\bar{x}} = \overline{\overline{X}} + 3\left(\frac{s}{\sqrt{n}}\right)$$

$$U_{\bar{x}} = 2{,}275 + 3 \cdot 0{,}048 = 2{,}419$$

والآن نستطيع رسم خريطة المتوسط بالشكل الآتي:



الشكل رقم (8–4) خريطة المتوسط

نلاحظ أن الأوساط الحسابية للعينات تنحصر ضمن حدي الضبط، وبالتالي فإن العملية الإنتاجية تحت السيطرة الإحصائية (مضبوطة إحصائياً).

ب . خريطة المدى R charts:

يعرف المدى بأنه الفرق بين أكبر قيمة وأصغر قيمة، ويستخدم المدى عندما تكون وحدات العينة قليلة، أما عندما تكون وحدات العينة كبيراً (أي أكبر من 10) فإن استخدام الانحراف المعياري يعطي نتائج أدق من المدى.

إن خريطة المدى تهتم بتشتت العملية الإنتاجية، وإظهار انحرافات مديات العينات عن المتوسط العام للمديات. فقد رأينا عند رسم خريطة المتوسط ووقعنا عليها متوسطات العينات بحيث كانت جميع المتوسطات ما بين حدي الضبط، عند ذلك تأكدنا من أن العملية الإنتاجية تحت السيطرة الإحصائية (مضبوطة إحصائياً). ولكن

ليس من الضروري أن يكون تشتت العملية الإنتاجية ضمن حدي الضبط، بل قد يكون الاختلاف ضمن العينات كبيراً، لذلك من الضروري الاهتمام بتقلب وتغير العملية الإنتاجية ومن أجل ضبط تقلب العملية الإنتاجية يتم إنشاء خريطة لضبط المدى.

إن رسم خريطة المدى R charts لايختلف كثيراً عن خارطة المتوسط إذ أنها أيضاً تتألف من ثلاثة خطوط رئيسية:

. خطان مستقيمان أحدهما علوي ويعبر عن حد الضبط الأعلى، ونرمز له بـ $U_{\overline{R}}$ ، والأخر سفلي و ويعبر عن حد الضبط الأسفل، ونرمز له بـ $L_{\overline{R}}$ ،

ـ خط متوسط يعبر عن المتوسط العام لمديات العينات ونرمز له بـ $\overline{R}$، وإن هذه الخطوط الثلاثة يتم إنشاؤها على محورين إحداثيين حيث يتوضع على المحور الأفقي عدد العينات، ويتوضع على المحور العمودي مديات العينات R_i، حيث $i = 1,2,3.....m$، و m عدد العينات ويتم إيجاد حدود الضبط للمدى $U_{\overline{R}}$ و $L_{\overline{R}}$ وخط المتوسط العام للمديات $\overline{R}$ بالشكل التالي:

. نحسب مدى كل عينة من العينات R_i.

$$R_i = X_{\max} - X_{\min} \qquad (14-8)$$

حيث: $i = 1\ ,\ 2\ ,\ 3....,\ m$

. نحسب المتوسط الحسابي العام للمديات بالعلاقة:

$$\overline{R} = \frac{\sum_{i=1}^{m} R_i}{m} \qquad (15-8)$$

. نحدد حدي الضبط للمدى عند 3 انحرافاً معيارياً، وبالتالي فإن 99,73% من مديات العينات يتوقع أن تنحصر ضمن حدي الضبط، أي أن:

. حد الضبط الأعلى $U_{\overline{R}}$ يساوي:

$$U_{\overline{R}} = \overline{R} + 3\sigma_R \qquad (16-8)$$

. حد الضبط الأدنى $L_{\overline{R}}$ يساوي:

$$L_{\overline{R}} = \overline{R} - 3\sigma_R \qquad (17-8)$$

حيث: σ_R هو عبارة عن الانحراف المعياري للمدى في المجتمع. وعندما تكون σ_R مجهولة نقوم بتقديرها بالانحراف المعياري لمديات العينات عن المتوسط العام للمدى والذي يعطى بالعلاقة:

$$\tilde{\sigma}_R = \sqrt{\frac{\sum_{i=1}^{m}\left(Ri - \overline{R}\right)^2}{m-1}} \qquad (18-8)$$

وبالتالي يصبح حدي الضبط العلوي والسفلي على الترتيب كمايلي:

$$U_{\overline{R}} = \overline{R} + 3\tilde{\sigma}_R$$

$$L_{\overline{R}} = \overline{R} - 3\tilde{\sigma}_R \qquad (19-8)$$

ويمكن إيجاد حدود الضبط العلوية والسفلية لخريطة المدى بالعلاقتين التاليتين:

$$U_{\overline{R}} = D_4\overline{R}$$

$$L_{\overline{R}} = D_3\overline{R} \qquad (20-8)$$

حيث: D_4, D_3 ثوابت تستخرج من جداول إنشاء خرائط الرقابة الإحصائية من الجدول رقم (1) الملحق في نهاية هذا الفصل.

ومن ثم يتم رسم محورين إحداثيين، وينشأ عليهما حدود الضبط للمدى والخط المتوسط الذي يعبر عن المدى العام $\overline{R}$. ونكون قد حصلنا على خريطة المدى ثم يوقّع عليها جميع مديات العينات المحسوبة. فإذا كانت جميع المديات ضمن حدي الضبط عندئذ نقول إن العملية الإنتاجية مضبوطة إحصائياً (تحت السيطرة الإحصائية)، أما إذا كان أحد مديات العينات خارج حدود الضبط فإن العملية الإنتاجية تكون غير مضبوطة إحصائياً أو خارج السيطرة .

مثال 5:

باستخدام بيانات المثال رقم (4). والمطلوب:

ارسم خريطة ضبط المدى، وحدد عليها مديات العينات المأخوذة. ثم استنتج فيما إذا كانت العملية الإنتاجية مضبوطة إحصائياً أم لا؟

الحل:

لإيجاد خريطة المدى نقوم بإجراء الحسابات التالية:

. نحسب مدى العينات الثمانية ثم نحسب الوسط الحسابي العام للمدى $\overline{R}$ ، ومن ثم نحسب $\widetilde{\sigma}_R^2$ ونضع النتائج في الجدول التالي:

رقم العينة	1	2	3	4	5	6	7	8	المجموع
مدى العينة Ri	0,2	0,2	0,15	0,3	0,05	0,15	0,25	0,3	1,6

ومنه الوسط الحسابي العام للمدى يساوي:

$$\overline{R} = \frac{\sum_{i=1}^{m} R_i}{m}$$

$$\overline{R} = \frac{1,6}{8} = 0,2$$

ومن ثم نحسب $\widetilde{\sigma}_R^2$:

$$\widetilde{\sigma}_R^2 = \frac{\sum_{i=1}^{m} \left(Ri - \overline{R}\right)^2}{m-1}$$

$$\widetilde{\sigma}_R^2 = \frac{0,05}{8-1} = 0,007143$$

$$\widetilde{\sigma}_R = 0,085$$

ومنه نحسب حدود الضبط:

. حد الضبط الأعلى:

$$U_{\overline{R}} = \overline{R} + 3\widetilde{\sigma}_R$$
$$U_{\overline{R}} = 0{,}2 + 3 \cdot 0{,}085 = 0{,}455$$

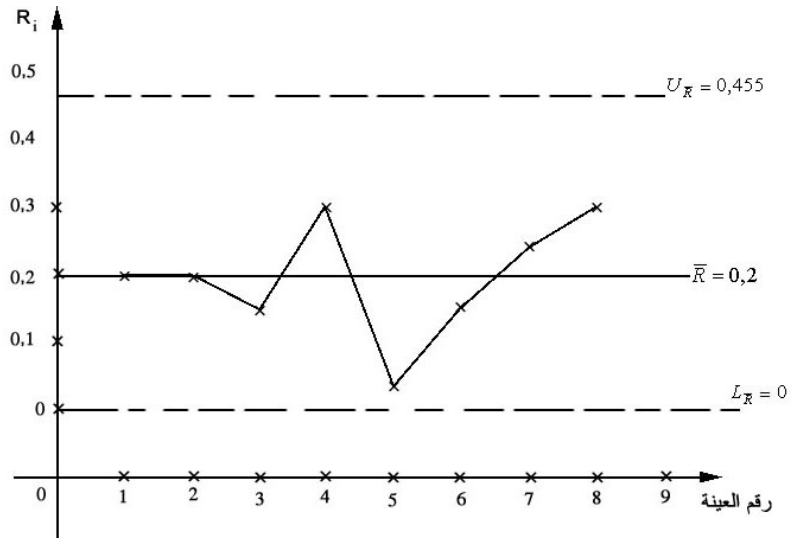
. حد الضبط الأدنى:

$$L_{\overline{R}} = \overline{R} - 3\widetilde{\sigma}_R$$
$$L_{\overline{R}} = 0{,}2 - 3 \cdot 0{,}085 = -0{,}055 = 0$$

ملاحظة:

إذا كان حد الضبط الأدنى ذو قيمة سالبة، عندئذ يُوضع صفرٌ لأن الإنتاج لا يمكن أن يكون ذو قيمة سالبة.

وتتحدد خريطة الضبط للمدى بالشكل التالي:



الشكل رقم (8-5)

وبالنظر إلى خريطة ضبط المدى نلاحظ أن جميع مديات العينات تقع بين حدي الضبط، وبالتالي فإن العملية الإنتاجية تحت السيطرة الإحصائية بالإضافة إلى أنه في العينتين الأولى والثانية كانت العملية مضبوطة تماماً.

جـ . خريطة ضبط الانحراف المعياري S Charts:

وفق الأسلوب السابق نفسه يمكن رسم خريطة الانحراف المعياري. وكما ذكرنا سابقاً فإن الانحراف المعياري للعينات يكون أكثر دقة من المدى في دراسة تشتت الوحدات داخل كل عينة. وبذلك تتم مراقبة انحرافات وحدات العينة عن بعضها البعض. ولإنشاء خريطة ضبط الانحراف المعياري للعينات يتوجب إيجاد حدي الضبط العلوي والسفلي عند 3 انحراف معياري بحيث تحصر 99,73% من المشاهدات، وبالتالي فإن:

. حد الضبط العلوي ونرمز له بـ U_S، ويعطى بالعلاقة:

$$U_S = \overline{S} + 3S \qquad (21-8)$$

. حد الضبط السفلي ونرمز له بـ L_S، ويعطى بالعلاقة:

$$L_S = \overline{S} - 3S \qquad (22-8)$$

حيث: $\overline{S}$ الانحراف المعياري المتوسط للانحرافات المعيارية للعينات، ويساوي:

$$\overline{S} = \frac{\sum_{i=1}^{m} Si}{m} \qquad (23-8)$$

حيث: $i = 1,2,3.....m$ و m عدد العينات.

و S_i الانحرافات المعيارية للعينات والتي حجم كل منها n حيث $j = 1,2,3.....n$.

ويعطى بالعلاقة:

$$S_i = \sqrt{\frac{\sum_{j=1}^{n}\left(x_{ji} - \overline{x}_i\right)^2}{n-1}} \qquad (24-8)$$

حيث: $i = 1\ ,\ 2\ ,\ 3.....m$

وS الانحراف المعياري للانحرافات المعيارية للعينات عن الانحراف المعياري المتوسط $\overline{S}$.

ويعطى بالعلاقة:

$$S=\sqrt{\frac{\sum_{i=1}^{m}\left(S_i-\overline{S}\right)^2}{m-1}} \qquad (25-8)$$

وبتعويض العلاقتين (8–23) و(8–25) في (8–21) و(8–22) نحصل على حدي الضبط العلوي والسفلي.

ويمكن إيجاد حدي الضبط (حدي المراقبة) للانحرافات المعيارية للعينات بالعلاقتين التاليتين:

$$\begin{aligned} U_S &= \beta_4\overline{S} && \text{حد الضبط الأعلى:} \\ L_S &= \beta_3\overline{S} && \text{حد الضبط الأدنى:} \end{aligned} \qquad (26\text{–}8)$$

وعند رسم خريطة الوسط الحسابي والانحراف المعياري فإن حدود الضبط للمتوسط تساوي:

$$\overline{\overline{X}} \pm A_1 \cdot \overline{S}$$

حيث: β_4, β_3, A_1 ثوابت تستخرج من جداول إنشاء خرائط الرقابة الإحصائية. في الملحق الجدول رقم (1)

والآن يمكن رسم حدود الضبط L_S, U_S على محورين إحداثيين حيث يتحدد عدد العينات على المحور الأفقي، ويحدد على المحور العمودي الانحرافات المعيارية للعينات S_i حيث $i = 1,2,3.....m$.

بالإضافة إلى الخط المتوسط $\overline{S}$ والذي يعبر عن متوسط الانحرافات المعيارية للعينات، ونكون قد حصلنا على خريطة ضبط الانحرافات المعيارية S Charrts ، ومن ثم يوقّع عليها الانحرافات المعيارية للعينات S_i، فإذا كانت جميع الانحرافات المعيارية داخل حدي الضبط، فإن العملية مضبوطة إحصائياً أو تحت السيطرة الإحصائية أما إذا وجد أحد الانحرافات المعيارية للعينات خارج حدود الضبط، عندئذ فإن العملية خارج نطاق السيطرة الإحصائية، ويتوجب اتخاذ الإجراءات التصحيحية لتعديل مسار العملية الإنتاجية.

مثال 6:

استخدم بيانات المثال رقم (4)، ومن ثم ارسم خريطة الانحراف المعياري موقعاً عليها الانحرافات المعيارية للعينات، واستنتج فيما إذا كانت العملية الإنتاجية مضبوطة إحصائياً؟

الحل:

لإيجاد خريطة ضبط الانحراف المعياري نحسب الانحرافات المعيارية للعينات S_i، ومن ثم نحسب متوسط الانحراف المعياري للعينات $\overline{S}$، وبعدئذٍ نحسب S ونضع النتائج في الجدول التالي:

رقم العينة	1	2	3	4	5	6	7	8	المجموع
Si الانحراف المعياري للعينات	0,085	0,085	0,075	0,126	0,03	0,075	0,1192	0,1261	0,7212

ومنه فإن:

$$\overline{S} = \frac{\sum_{i=1}^{m} S_i}{m} = \frac{0,7212}{8} = 0,09$$

$$S = \sqrt{\frac{\sum_{i=1}^{m} \left(S_i - \overline{S}\right)^2}{m-1}} = \sqrt{\frac{0,007552}{7}} = 0,0328$$

والآن نقوم بحساب حدي الضبط للانحراف المعياري:

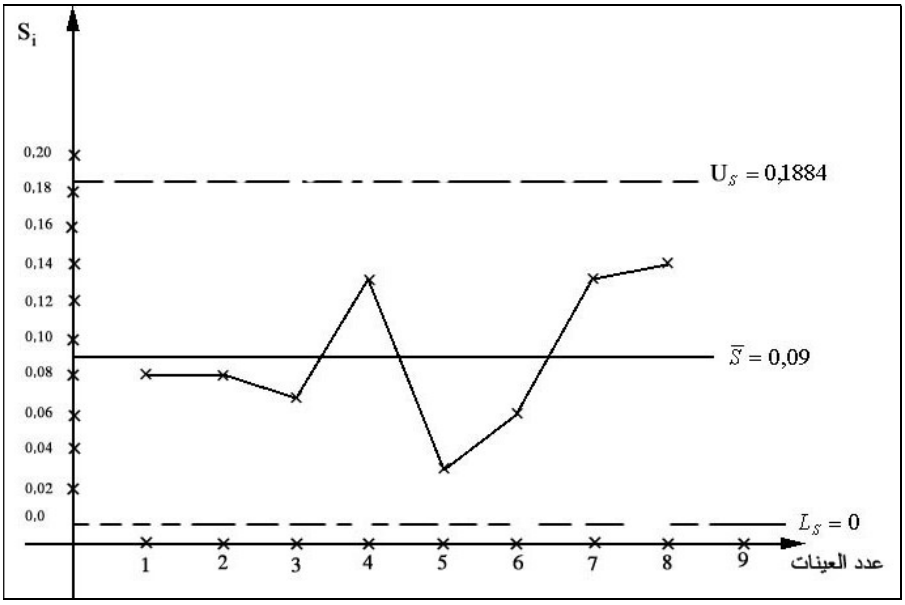
$$U_S = \overline{S} + 3S = 0,09 + 3 \cdot 0,0328 = 0,1884$$

$$L_S = \overline{S} - 3S = 0,09 - 3 \cdot 0,0328 = -0,0084 = 0$$

ملاحظة:

إذا كان حد الضبط الأدنى سالباً فنضعه مساوياً للصفر لأن الإنتاج لايمكن أن يكون سالباً.

والآن نأتي إلى رسم خريطة الضبط للانحراف المعياري:



الشكل رقم (8-6)

وبالنظر إلى خريطة ضبط الانحراف المعياري للعينات نجد أن جميع الانحرافات المعيارية داخل حدي الضبط، وبالتالي فإن العملية تحت السيطرة الإحصائية (مضبوطة إحصائياً).

2 . خرائط الرقابة على الخصائص (خرائط ضبط المواصفات) Control Charts for Attributes:

نستخدم هذه الخرائط عندما يصنف الإنتاج وفقاً لخاصيتين معيبة أو غير معيبة. أو بمعنى آخر عندما تكون محددات المواصفات غير قابلة للقياس بشكل كمي، بل تكون نوعية (جيدة أو غير جيدة)، وبالتالي فإن الإنتاج يمكن أن يكون مطابق للمواصفات أو غير مطابق للمواصفات الموضوعة.

وفي هذه الحالة لايمكن حساب خرائط الوسط الحسابي والمدى في ضبط جودة المنتجات. ويستحسن أن تستخدم خرائط ضبط المواصفات لتقرير فيما إذا كان الإنتاج النهائي مطابقاً للمواصفات أم لا؟ ويوجد نوعان من خرائط الرقابة على المواصفات:

أ. خرائط مراقبة المواصفات النسبية أو خرائط نسبة الوحدات المعيبةP Charrts:

ويتم بموجب هذا الأسلوب حصر جميع الوحدات المنتجة والخاضعة للفحص على أساس خاصية واحدة، أو صفة معينة وتقسيمها إلى وحدات معيبة، والتي تكون غير مطابقة للمواصفات الموضوعة وإلى وحدات غير معيبة (مقبولة)، وهي التي تكون مطابقة للمواصفات الموضوعة.

وتستعمل هذه الخرائط (خرائط نسبة الوحدات المعيبة) عندما نتمكن من حصر جميع الوحدات المنتجة إلى معيبة أو غير معيبة. بالإضافة إلى كون جميع العينات المختارة ذات حجوم ثابتة وعلى فترات زمنية منتظمة.

إن الهدف من إنشاء هذه الخرائط هو معرفة وجود أو عدم وجود تغير أو اختلافات في مواصفات الوحدات المنتجة والذي يعود هذا التغير، إن وجد، إلى خلل في العملية الإنتاجية وليس عائداً إلى أسباب الصدفة.

ولكي يتم التوصل إلى نتيجة مفادها أن العملية الإنتاجية مضبوطة إحصائياً أم لا؟ نأخذ عينات بحجوم ثابتة وعلى فترات زمنية منتظمة، ويفضل أن لايقل حجم العينة عن 50 وحدة منتجة. لأن خرائط المراقبة في حالات النسب تكون أقل حساسية لإظهار التغير في مواصفات الوحدات المنتجة.

ولإنشاء خرائط نسبة الوحدات المعيبة لابد من القيام بسحب عينات ذات حجوم ثابتة وحصر عدد الوحدات المعيبة في كل عينة، ومن ثم حساب نسبة المعيب في كل منها، والذي نرمز له بـ P، ويعطى بالعلاقة:

$$P_i = \frac{\text{عدد الوحدات المعيبة}}{\text{عدد وحدات العينة}}$$

حيث: $i = 1,2,3.....m$، و m عدد العينات.

وبالتالي نحصل على نسبة المعيب في العينات على التوالي:

$$P_1 , P_2 , P_3 , ...P_i , ...P_m$$

ومن ثم نحسب متوسط نسبة المعيب العام ونرمز له بـ $\overline{P}$، ويعطى بالعلاقة:

$$\overline{P} = \frac{\sum_{i=1}^{m} P_i}{m} \qquad (27-8)$$

أو بالعلاقة:

$$\overline{P} = \frac{\text{مجموع الوحدات المعيبة}}{n \times m} \qquad (28-8)$$

والآن بعد أن حسبنا متوسط نسبة المعيب العام $\overline{P}$ نستطيع حساب الانحراف المعياري لتوزيع العينات ونرمز له بـ $S_{\overline{p}}$ من العلاقة التالية:

$$S_{\overline{p}} = \sqrt{\frac{\overline{P}(1-\overline{P})}{n}} \qquad (29-8)$$

حيث n حجم العينة و $S_{\overline{p}}$ تقدير لـ $\sigma_{\overline{P}}$ الانحراف المعياري لنسبة المعيب في المجتمع الإحصائي، وبالتالي تكون حدود الضبط لخريطة نسبة الوحدات المعيبة عند 3 انحراف معياري بحيث تحصر 99,73% من نسب معيب العينات بالشكل التالي:

. حد الضبط الأعلى U_p:

$$Up = \overline{P} + 3S_{\overline{P}} \qquad (30-8)$$

. حد الضبط الأدنى L_p:

$$Lp = \overline{P} - 3S_{\overline{P}} \qquad (31-8)$$

والآن يمكن رسم خريطة ضبط نسبة الوحدات المعيبة بنفس الأسلوب كما سبق حيث نحدد حدود الضبط U_p و L_p والخط المتوسط $\overline{P}$ الذي يعبر عن المتوسط العام لنسب المعيب.

ومن ثم توقع على الخريطة نسب المعيب للعينات. فإذا كانت جميع النسب ضمن حدي الضبط عندئذ فإن العملية الإنتاجية مضبوطة إحصائياً أو محكمة، أما عندما توجد نسبة واحدة على الأقل خارج حدود الضبط فإن العملية الإنتاجية غير

محكمة أو ليست تحت السيطرة بسبب وجود أسباب غير عادية أودت بالعملية إلى التغير عن المواصفات الموضوعة.

مثال 7:

لمراقبة جودة إنتاج مصنع النهضة الصناعية لإنتاج مؤشرات السرعة تم سحب 27 عينة من إنتاج هذا المصنع لفحصها والتأكد من مطابقتها للمواصفات الموضوعة بحيث تتضمن كل عينة 100 مؤشر سرعة بشكل يومي ولمدة سبع وعشرون يوماً، فإذا علمت أن الإنتاج اليومي للمصنع كان 1500 مؤشر سرعة، وقد بينت نتائج الفحص نسبة عدد الوحدات المعيبة في كل عينة، وتم وضعها في الجدول التالي:

رقم العينة	حجم العينة	نسبة عدد الوحدات المعيبة	رقم العينة	حجم العينة	نسبة عدد الوحدات المعيبة	رقم العينة	حجم العينة	نسبة عدد الوحدات المعيبة
1	100	0,06	10	100	0,04	19	100	0,03
2	100	0,05	11	100	0,08	20	100	0,04
3	100	0,07	12	100	0,06	21	100	0,08
4	100	0,06	13	100	0,09	22	100	0,07
5	100	0,04	14	100	0,03	23	100	0,09
6	100	0,02	15	100	0,01	24	100	0,03
7	100	0,08	16	100	0,05	25	100	0,02
8	100	0,03	17	100	0,00	26	100	0,01
9	100	0,08	18	100	0,05	27	100	0,08

والمطلوب: تحديد فيما إذا كانت العملية الإنتاجية تحت السيطرة الإحصائية، أم لا؟ وموضحاً ذلك من خلال خريطة ضبط نسبة الوحدات المعيبة.

الحل:

لتحديد فيما إذا كانت العملية مضبوطة إحصائياً نرسم خريطة نسبة الوحدات المعيبة.

ولرسم الخريطة يتطلب حساب متوسط نسبة المعيب العام $\overline{P}$، ومن ثم الانحراف المعياري لتوزيع العينات $S_{\overline{P}}$. أي:

$$\overline{P} = \frac{\sum_{i=1}^{m} P_i}{m} = \frac{1,35}{27} = 0,05$$

$$S_{\overline{P}} = \sqrt{\frac{0,05(1-0,05)}{100}} = 0,0218$$

وبالتالي فإن حدود الضبط تساوي:

. حد الضبط الأعلى لنسبة الوحدات المعيبة U_P:

$$U_P = \overline{P} + 3S_{\overline{P}}$$

$$U_P = 0,05 + 3 \cdot 0,0218 = 0,1154$$

. حد الضبط الأدنى لنسبة الوحدات المعيبة L_P:

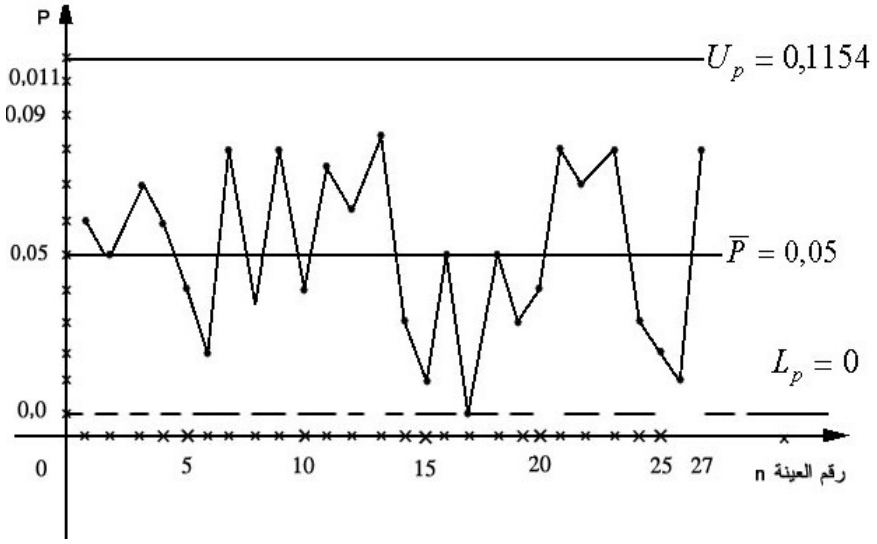
$$L_P = \overline{P} - 3S_{\overline{P}}$$

$$L_P = 0,05 - 3 \cdot 0,0218 = -0,0154 = 0$$

ملاحظة:

عندما يكون حد الضبط الأدنى سالباً فإننا نضع $L_P = 0$، لأن نسبة المعيب لايمكن أن تكون في السالب.

نقوم برسم حدي الضبط الأدنى والأعلى ومتوسط نسبة الوحدات المعيبة على الخريطة ونوقع عليها نسب المعيب للعينات فنحصل على الشكل التالي:



الشكل (8–7) خريطة ضبط نسبة الوحدات المعيبة

وبالنظر إلى خريطة ضبط نسبة الوحدات المعيبة نلاحظ أن جميع الملاحظات كانت ضمن حدي الضبط، وبالتالي فإن العملية الإنتاجية مضبوطة إحصائياً ولاتحتاج إلى تعديل.

ب . خرائط عدد العيوب C Charrts:

تستخدم هذه الخرائط في الحالات التي لايمكن فيها حصر عدد الوحدات غير المعيبة، وإنما يمكن حصر عدد الوحدات المعيبة فقط، مثل عدد الفقاعات في كأس من الزجاج، أو عدد اللطخ السوداء في متر من القماش الأبيض، أو عدد الحروف المغلوطة في الصفحة الواحدة من صفحات الكتاب...إلخ. ويعتبر توزيع بواسون الأساس لخرائط عدد العيوب، حيث يكون الوسط الحسابي $\bar{x}$ مساوياً إلى التباين σ^2، وبالتالي فإن الانحراف المعياري لعدد الوحدات المعيبة يساوي $\sigma = \sqrt{\sigma^2}$، أو يساوي $\sqrt{\bar{x}}$. ولإنشاء خريطة الرقابة لعدد الوحدات المعيبة نحسب متوسط عدد العيوب $\bar{x}$ بالعلاقة التالية:

$$\bar{x} = \frac{\sum_{i=1}^{m} x_i}{m} \qquad (32-8)$$

حيث x_i عدد العيوب في كل عينة و $i = 1 , 2 , 3 , \ldots\ldots, m$ و m عدد العينات.

ثم نحسب الانحراف المعياري:

$$\sigma = \sqrt{\bar{x}} \qquad (33-8)$$

وبالتالي فإن حدي الضبط العلوي والسفلي عند 3 انحراف معياري بحيث تحصر بينها 99,73 % من المشاهدات يعطى بالعلاقتين:

حد الضبط العلوي لعدد العيوب U_C:

$$U_C = \bar{X} + 3\sigma \qquad (34-8)$$

حد الضبط السفلي لعدد العيوب L_C:

$$L_C = \bar{X} - 3\sigma \qquad (35-8)$$

والآن كما سبق يمكن رسم خريطة ضبط عدد الوحدات المعيبة. ويؤشَّر عليها عدد الوحدات المعيبة للعينات ليقرر فيما إذا كانت العملية الإنتاجية مضبوطة إحصائياً أم لا.

مثال 8:

فيما يلي عدد الفقاعات في كل متر مربع من الزجاج، والتي تم اكتشافها في 24 عينة أخذت من إنتاج معمل الزجاج وبطريقة عشوائية، وكانت كما يلي:

رقم العينة	عدد العيوب	رقم العينة	عدد العيوب	رقم العينة	عدد العيوب
1	15	9	8	17	15
2	10	10	6	18	4
3	6	11	7	19	9
4	20	12	10	20	18
5	21	13	16	21	20
6	19	14	19	22	23
7	11	15	18	23	16
8	10	16	8	24	15

والمطلوب: إعداد خريطة مراقبة عدد الوحدات المعيبة. هل العملية الإنتاجية مضبوطة أم لا؟

الحل:

نحسب متوسط عدد العيوب في العينات نجد:

$$\overline{X} = \frac{\sum_{i=1}^{m} X_i}{m} = \frac{324}{24} = 13,5$$

نحسب الانحراف المعياري ويساوي:

$$\sigma = \sqrt{\overline{X}} = \sqrt{13,5} = 3,67$$

نحسب حدود الضبط:

. حد الضبط الأعلى U_C ويساوي:

$$U_C = \overline{X} + 3\sigma$$
$$U_C = 13,5 + 3 \cdot 3,67 = 24,51$$

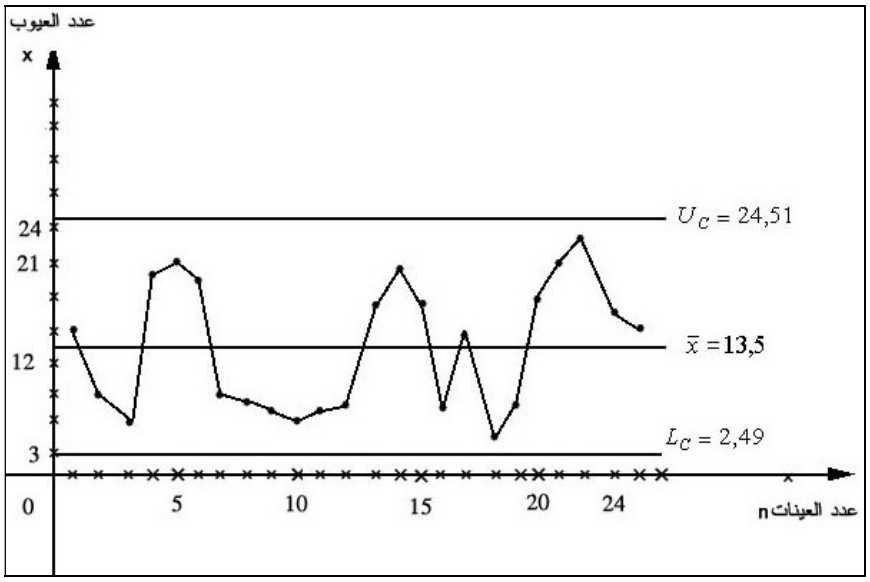
. حد الضبط الأدنى L_C ويساوي:

$$L_C = \overline{X} - 3\sigma$$
$$L_C = 13,5 - 3 \cdot 3,67 = 2,49$$

ملاحظة:

إذا كانت قيمة حد الضبط الأدنى سالبة فإننا نضعها مساوية للصفر لأن عدد الوحدات المعيبة لايمكن أن يكون سالباً.

والآن نقوم برسم خريطة عدد العيوب في العينات:



الشكل (8-8) خريطة عدد العيوب

وبالنظر إلى خريطة ضبط عدد الوحدات المعيبة نلاحظ أن العملية مضبوطة إحصائياً، وأن عملية الإنتاج لاتحتاج إلى تعديل.

8-5: حالات أخرى في مراقبة جودة الإنتاج:

1 . يمكن استخدام التوزيعات التكرارية في مراقبة جودة الإنتاج، وذلك عندما تكون مخرجات أحد خطوط الإنتاج كبيرة. لمعرفة فيما إذا كانت وحدات المنتج مطابقة للمواصفات الموضوعة أم لا؟ نسحب عينات عشوائية وعلى فترات زمنية محددة ومن ثم القيام بأخذ القياسات اللازمة (أوزان، أطوال، مساحات، وحدات معيبة، ... الخ). وتبويبها في فئات ثم رسم المدرج التكراري والمنحني التكراري لها. ومن المفروض أن يكون المنحني التكراري الذي يُمثل هذه القياسات هو المنحني الطبيعي. وبالتالي فإن معظم قياسات المشاهدات سوف تتمركز حول المتوسط $\overline{X}$، إذا كانت العملية الإنتاجية مضبوطة وأن 99.73% من قياسات المشاهدات تكون ضمن $\pm 3\sigma$.

وعندما يطرأ خلل في العملية الإنتاجية فإن قياسات المشاهدات والمعبّر عنها بالمنحني التكراري سوف يكون ملتوياً إلى اليمين أو إلى اليسار وبالتالي سوف يكون

غير مطابقاً للمنحني الطبيعي مما يوحي على أن العملية الإنتاجية غير مضبوطة. ولتوضيح ذلك نأخذ المثال التالي:

مثال 9:

آلة لتعبئة أسطوانات الغاز بوزن 20 كغ بالمتوسط. ولتحديد فيما إذا كانت العملية الإنتاجية تحت السيطرة (مضبوطة).

سحبنا عينات عشوائية بحجم 10 أسطوانات غاز كل ساعتين خلال 24 ساعة عمل، فحصلنا على 120 وحدة معبأة، ومن ثم تم قياس أوزانها وتبويبها في الجدول التالي:

الفئات	التكرارات f
18.25-18.75	7
18.75-19.25	10
19.25-19.75	25
19.75-20.25	30
20.25-20.75	25
20.75-21.25	15
21.25-21.75	8
$\sum$	120

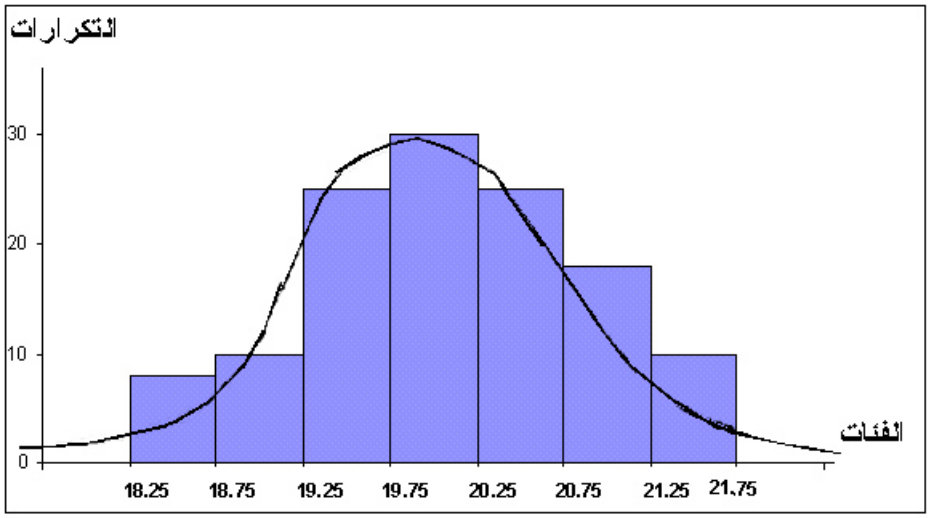
والمطلوب:

1 . ارسم المنحنى التكراري الذي يمثل بيانات الظاهرة المدروسة. هل العملية الإنتاجية مضبوطة (تحت السيطرة)؟

2 . لنفترض أننا في يوم آخر سحبنا نفس العدد من أسطوانات الغاز، وكانت قياساتها موزعة حسب الفئات بالشكل التالي: f :15,20,25,30,10,9,11 . ارسم المنحى التكراري لها ماذا نستنتج؟

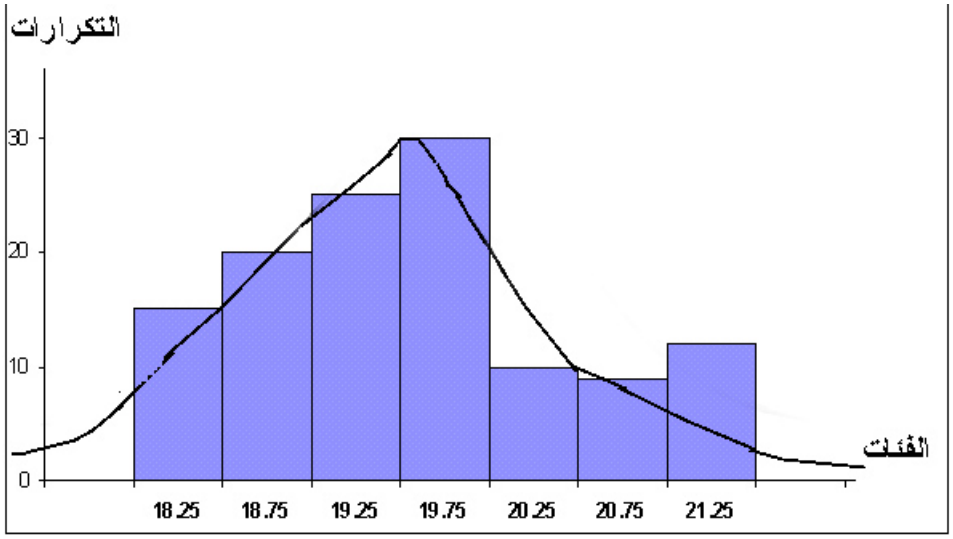
الحل:

نقوم بإعداد المدرج التكراري والمنحني التكراري لبيانات الظاهرة المدروسة.



الشكل رقم (8-9)

نلاحظ من الرسم البياني أن المنحني التكراري قريب جداً من المنحني الطبيعي وأن العملية الإنتاجية مضبوطة (تحت السيطرة)، وأن قياسات معظم المشاهدات تتركز حول المتوسط.



الشكل رقم (8-10)

من خلال رسم المنحني البياني نلاحظ أن المنحني ملتوياً ويوجد عدد من وحدات الإنتاج تتركز في الفئة الأخيرة. وبالتالي نستنتج أنه يوجد خلل في العملية الإنتاجية لأنها أخذت تنحرف عن القياسات المطلوبة.

8-5-2: ويمكن مقارنة احتمال إنتاج الوحدة الجيدة من واقع الاختبار الفعلي مع الاحتمال النظري لإنتاج الوحدة الجيدة، أو مقارنة احتمال إنتاج الوحدة المعيبة من واقع الاختبار الفعلي مع الاحتمال النظري لإنتاج الوحدة المعيبة. وذلك من خلال سحب عينات من المنتجات وتقدير احتمال إنتاج الوحدة الجيدة (أو المعيبة) من واقع الاختبار الفعلي ومقارنته مع الاحتمال النظري لإنتاجها.

فإذا كان هناك فرقٌ كبيرُ بين الاحتمال النظري والاحتمال الفعلي فإن ذلك يعتبر دليلاً كافياً على وجود خلل في العملية الإنتاجية وأنها ليست تحت السيطرة، وبالتالي يجب اتخاذ الإجراءات اللازمة للكشف عن الأسباب وتلافيها، وإذا كان الفرق بين الاحتمالين صغيراً فإن ذلك يُعد دليلاً على أن العملية الإنتاجية تسير وفق المواصفات الموضوعة.

مثال 10:

قام معمل المصابيح الكهربائية باختبار 500 عينة من المصابيح الكهربائية التي ينتجها.

وكانت كل عينة مؤلفة من أربعة مصابيح. وتبين للمعمل أن توزيع هذه العينات وفقاً لعدد المصابيح الجيدة كان كما يلي:

عد المصابيح الجيدة	0	1	2	3	4	المجموع
عدد العينات	15	25	80	180	200	500

والمطلوب:

تحديد فيما إذا كان الإنتاج مطابقاً للمواصفات أم لا؟ إذا علمت أن الاحتمال النظري لإنتاج المصباح الجيد هو $\widetilde{P} = 0,9$.

الحل:

لإيجاد احتمال إنتاج المصباح الجيد من واقع الاختبار الفعلي نتبع ما يلي:

1 . نحسب متوسط عدد المصابيح الجيدة في العينة $\overline{X}$ ، فنجد:

$$\overline{X} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i} = \frac{0 \cdot 15 + 1 \cdot 25 + 2 \cdot 80 + 3 \cdot 180 + 4 \cdot 200}{500} = \frac{1525}{500}$$

$$\overline{X} = 3{,}05$$

2 . نقسم متوسط عدد المصابيح الجيدة في العينة على عدد وحدات العينة فنحصل على احتمال إنتاج المصباح الجيد من واقع الاختبار الفعلي P ؛ أي:

$$P = \frac{\overline{X}}{n} = \frac{3{,}05}{4} = 0{,}763$$

وبمقارنة احتمال إنتاج المصباح الجيد من واقع الاختبار الفعلي $P = 0{,}763$ مع الاحتمال النظري $\widetilde{P} = 0{,}9$ ، نلاحظ أنه يوجد فارق بين الاحتمالين، مما يدل على وجود خلل في العملية الإنتاجية، ويتطلب الأمر البحث والتقصي عن الأسباب وتلافيها.

8-5-3: مقارنة توزيع العينات من واقع الاختبار الفعلي مع التوزيع النظري لهذه العينات:

وذلك من خلال سحب عينات عشوائية من المنتجات وباستخدام الخواص الإحصائية للتوزيع النظري ومقارنتها مع الخواص الإحصائية للتوزيع الفعلي. فإذا كان هناك فرق بين التوزيع الفعلي، لعدد العينات وبين التوزيع النظري (سواءً كانت وحدات العينات جيدة أو معيبة)، فإنه يعتبر دليلاً كافياً على عدم مطابقة الإنتاج للمواصفات الموضوعة.

أما إذا كان غير ذلك، فإن العملية الإنتاجية مضبوطة أو تحت السيطرة الإحصائية.

مثال 11:

باستخدام بيانات المثال السابق رقم (10) يكون المطلوب تحديد فيما إذا كان توزيع العينات من واقع الاختبار الفعلي مطابقاً للتوزيع النظري لها. أو تحديد فيما إذا كانت العملية الإنتاجية مضبوطة أم لا؟

الحل:

نلاحظ أن $\widetilde{q} = 1 - \widetilde{P} = 1 - 0,9 = 0,1 \qquad \widetilde{P} = 0,9$

وبتطبيق قانون التوزيع الثنائي الذي يعطى بالعلاقة:

$$P(X = k) = C_n^k \, P^k q^{n-k}$$

وبالتعويض قيمة $\widetilde{P} = 0,9$ و $\widetilde{q} = 0,1$ نجد:

$$P(X = k) = C_4^k (0,9)^k (0,1)^{4-k}$$

وبفرض أن X متغير عشوائي يعبر عن عدد المصابيح الجيدة في العينة، وبالتالي يأخذ القيم التالية:

$$X : 0 \ , \ 1 \ , \ 2 \ , \ 3 \ , \ 4$$

ومنه نحصل على الاحتمالات المقابلة:

$$P(X = 0) = C_4^0 (0,9)^0 (0,1)^4 = 0,0001$$

$$P(X = 1) = C_4^1 (0,9)^1 (0,1)^3 = 0,0036$$

$$P(X = 2) = C_4^2 (0,9)^2 (0,1)^2 = 0,0486$$

$$P(X = 3) = C_4^3 (0,9)^3 (0,1)^1 = 0,2916$$

$$P(X = 4) = C_4^4 (0,9)^4 (0,1)^0 = 0,6561$$

وللحصول على التوزيع النظري للعينات، نضرب الاحتمالات السابقة بعدد العينات 500؛ أي:

$$0{,}0001 \cdot 500 = 0{,}05 \approx 0$$
$$0{,}0036 \cdot 500 = 1{,}8 \approx 2$$
$$0{,}0486 \cdot 500 = 24{,}3 \approx 24$$
$$0{,}2916 \cdot 500 = 145{,}8 \approx 146$$
$$0{,}6561 \cdot 500 = 328{,}05 \approx \underline{328}$$
$$500$$

والآن نقوم بمقارنة التوزيع النظري للعينات مع التوزيع الفعلي لها من خلال إيجاد الفروق بين التوزيعين؛ أي:

عدد المصابيح الجيدة في العينة	التوزيع الفعلي للعينات	التوزيع النظري للعينات	الفرق
0	15	0	15+
1	25	2	23+
2	80	24	56+
3	180	146	34+
4	200	328	128−
المجموع	500	500	0

وبمقارنة التوزيع الفعلي للعينات مع التوزيع النظري نلاحظ أن التوزيع الفعلي يزيد عن التوزيع النظري بمقدار 128 مصباح في العينات التي يكون فيها عدد الوحدات الجيدة 0 و1 و2 و3 في حين يقل التوزيع الفعلي عن التوزيع النظري في العينة التي يكون فيها عدد الوحدات الجيدة 4 بمقدار 128 مصباحاً. وبالتالي نستنتج أن هذا الفارق يشير إلى أن العملية الإنتاجية غير مضبوطة، مما يتطلب من إدارة المنشأة الإنتاجية الكشف عن الأسباب وتلافيها.

مثال 12:

لبيان مدى سلامة الإنتاج في أحد المصانع لإنتاج الأحذية ضمن قياسات محددة، تم سحب 800 عينة من الوحدات المنتجة (كل وحدة هي عبارة عن زوج من الأحذية) وتتألف كل عينة من 4 أزواج (وحدات منتجة). وتم فحص هذه العينات وتوزيعها حسب عدد الأزواج المعيبة في كلٍ منها، فحصلنا على التوزيع التالي:

عدد العينات	عدد الوحدات المعيبة في العينة
332	0
325	1
120	2
21	3
2	4
800	المجموع

والمطلوب: تحديد مدى سلامة الإنتاج إذا علمت أنه يسمح أن يكون بين كل 100 زوج من الأحذية المنتجة يوجد 20 زوج معيب.

الحل:

نلاحظ أن نسبة الأزواج المعيبة (الاحتمال النظري) يساوي:

$$\widetilde{P} = \frac{20}{100} = 0{,}20$$

1 . مقارنة الاحتمال النظري مع الاحتمال الفعلي لإنتاج زوج الأحذية المعيب؛ أي:

$$\overline{X} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i} = \frac{0 \cdot 332 + 1 \cdot 325 + 2 \cdot 120 + 3 \cdot 21 + 4 \cdot 2}{800} = \frac{636}{800}$$

$$\overline{X} = 0{,}795$$

وهو متوسط عدد الأزواج المعيبة في العينة.

ومن ثم نقسم $\overline{X} = 0,795$ على حجم العينة $n = 4$ فنحصل على احتمال إنتاج زوج الأحذية المعيب من واقع الاختبار الفعلي أي:

$$P = \frac{0,795}{4} = 0,199$$

وبمقارنة الاحتمال النظري مع الاحتمال الفعلي لإنتاج زوج الأحذية المعيب، نجد أن الفارق بين الاحتمالين قريب جداً مما يشير إلى أن العملية الإنتاجية مقبولة ولا تحتاج إلى أي تعديل.

2 . مقارنة توزيع العينات من واقع الاختبار الفعلي مع التوزيع النظري لها وذلك من خلال تطبيق قانون التوزيع الثنائي الذي يعطى بالعلاقة:

$$P(X = k) = C_n^k \, P^k q^{n-k}$$

$$P(X = k) = C_4^k (0,2)^k (0,8)^{4-k}$$

وبفرض أن X متغير عشوائي يعبر عن عدد الأزواج المعيبة في العينة، أي:

$$X : 0 \ , \ 1 \ , \ 2 \ , \ 3 \ , \ 4$$

وبالتالي نحصل على الاحتمالات المقابلة والتوزيع النظري للعينات:

$$P(X = 0) = C_4^0 (0,2)^0 (0,8)^4 = 0,4096 \cdot 800 = 328$$

$$P(X = 1) = C_4^1 (0,2)^1 (0,8)^3 = 0,4096 \cdot 800 = 328$$

$$P(X = 2) = C_4^2 (0,2)^2 (0,8)^2 = 0,1536 \cdot 800 = 123$$

$$P(X = 3) = C_4^3 (0,2)^3 (0,8)^1 = 0,0256 \cdot 800 = 20$$

$$P(x = 4) = C_4^4 (0,2)^4 (0,8)^0 = 0,0016 \cdot 800 = 1$$

$$\overline{800}$$

والآن نقوم بمقارنة التوزيع النظري للعينات مع التوزيع الفعلي لها من خلال إيجاد الفروق بين التوزيعين؛ أي:

عدد الوحدات المعيبة في العينة	التوزيع الفعلي للعينات	توزيع العينات حسب الاحتمال النظري	الفرق
0	332	328	+4
1	325	328	−3
2	120	123	−3
3	21	20	+1
4	2	1	+1
المجموع	800	800	0

وبمقارنة التوزيع الفعلي للعينات مع التوزيع النظري نلاحظ أن التوزيع الفعلي يزيد عن التوزيع النظري بمقدار 6 وحدات في العينات التي يكون فيها عدد الوحدات المعيبة 0 و3 و4، في حين يقل التوزيع الفعلي عن التوزيع النظري في العينات التي يكون فيها عدد الوحدات المعيبة 1 و2 بمقدار 6 وحدات. وبالتالي نستنتج أن هذا الفارق صغير ويشير إلى أن العملية الإنتاجية مضبوطة ولا تحتاج إلى تعديل.

تمرينات

1 . تقوم شركة الزيوت بتعبئة منتجاتها في عبوات سعة الواحدة منها 5 ليتر. ومن أجل الرقابة على العملية الإنتاجية، قرر مدير الإنتاج سحب 8 عينات، حجم كل منها 5 عبوات خلال الثماني ساعات الأولى من يوم العمل، وبعدئذٍ تم فحص هذه العينات وكانت نتيجة الفحص موضحة بالجدول التالي:

العينات / القياسات	1	2	3	4	5	6	7	8
1	4,9	5,5	4,8	4,5	6	4,1	4	5,2
2	5	6	4,7	5,1	6,1	5	6	5,4
3	5	4,5	4,0	5,2	5,9	5,1	5,5	4,9
4	5,2	5	4,5	4,8	3,9	5,4	5,3	4,5
5	5,4	5	5,5	4,9	4	5,5	4,5	6,1

والمطلوب:

تحديد فيما إذا كانت العملية الإنتاجية مضبوطة إحصائياً أم لا؟ وذلك من خلال إعداد خريطة الوسط الحسابي والمدى والانحراف المعياري.

2 . أعد التمرين السابق نفسه ولكن باستخدام جداول Grant في مراقبة الجودة إحصائياً لرسم خرائط المتوسط والمدى والانحراف المعياري.

3 . يقوم معمل الزجاج بإنتاج ألواح زجاجية شفافة قياس 1.5 × 2 ولمعرفة فيما إذا كانت العملية الإنتاجية تحت السيطرة الإحصائية.

سحبنا 20 لوح زجاجي من إنتاج المصنع خلال عشرة أيام، وتم فحص هذه الألواح الزجاجية فوجدت العيوب التالية:

رقم اللوح الزجاجي	عدد العيوب	رقم اللوح الزجاجي	عدد العيوب
1	2	11	9
2	6	12	1
3	8	13	8
4	5	14	5
5	15	15	9
6	3	16	3
7	2	17	1
8	1	18	9
9	0	19	0
10	10	20	12

والمطلوب: تحديد فيما إذا كانت العملية الإنتاجية تحت السيطرة الإحصائية أم لا. من خلال تحديد الحد الأدنى والحد الأقصى لخريطة الرقابة المستخدمة.

4 . وصلت معلومات إلى مدير الإنتاج في مصنع لإنتاج العبوات البلاستيكية سعة واحد ليتر عن وجود بعض العبوات المعيبة (مثقوبة).

ومن أجل معرفة فيما إذا كانت العملية الإنتاجية تحت السيطرة الإحصائية قام بسحب 25 عينة من إنتاج هذا المصنع بمعدل خمس عينات في اليوم وحجم كل عينة 100 عبوة.

وبعد عملية الفحص وجد مدير الإنتاج النتائج التالية:

رقم العينة	عدد العبوات المعيبة	رقم العينة	عدد العبوات المعيبة	رقم العينة	عدد العبوات المعيبة
1	8	10	5	19	5
2	9	11	9	20	6
3	15	12	19	21	2
4	3	13	8	22	19
5	2	14	5	23	11
6	0	15	4	24	0
7	20	16	13	25	15
8	18	17	0		
9	10	18	12		

والمطلوب: هل العملية الإنتاجية تحت السيطرة الإحصائية؟ إعداد خريطة المراقبة الخاصة بالعملية الإنتاجية.

5 . قامت الشركة العامة لصناعة الإلكترونيات باختبار 400 عينة من قطع الغيار وكل عينة مؤلفة من أربع قطع غيار. وقد تبين أن توزيع هذه العينات وفقاً لعدد قطع الغيار الجيدة كما يلي:

التوزيع الفعلي لعدد العينات	عدد قطع الغيار الجيدة
4	0
12	1
68	2
150	3

4	165

فإذا علمت أنه من بين كل 100 قطعة غيار يسمح بأن توجد بينها 5 قطع غيار معيبة.

والمطلوب: مساعدة إدارة الشركة في اتخاذ قرارها حول سلامة الإنتاج ومدى مطابقته للمواصفات الموضوعة. وذلك من خلال الحالتين التاليتين:

1 . مقارنة الاحتمال الفعلي لإنتاج قطعة الغيار الجيدة مع الاحتمال النظري.

2 . مقارنة التوزيع الفعلي للعينات مع التوزيع النظري.

6 . قامت الشركة العامة للسكر بإجراء اختبار حول أوزان أكياس السكر، إذ يبلغ وزن الكيس بالمتوسط 50 كغ. ومن أجل القيام بهذا الاختبار سحبت 20 عينة حجم كلٍ منها 5 أكياس، وبعد إجراء القياسات اللازمة حصلنا على المعلومات التالية:

مدى العينة	متوسط العينة	رقم العينة
2	47	11
1,5	52	12
0,5	51,2	13
0,75	49,5	14
1	50	15
1	50	16
1,5	51	17
2,5	48,2	18
3	47	19
2	50,5	20

مدى العينة	متوسط العينة	رقم العينة
0,5	50	1
1	51	2
1,5	49,5	3
0,5	48,5	4
0,5	49	5
2	51	6
0,75	50,5	7
0,5	51	8
1,5	49,5	9
2,5	48,5	10

والمطلوب:

تحديد فيما إذا كانت العملية الإنتاجية تحت السيطرة الإحصائية وذلك باستخدام خرائط المتوسط والمدى.

ملحق

ثوابت حساب حدود خرائط الرقابة الإحصائية

الجدول رقم (1)

عدد وحدات العينة	A_1	A_2	B_3	B_4	D_3	D_4
2	3,760	1,880	0	3,267	0	3,267
3	2,394	1,023	0	2,568	0	2,575
4	1,880	0,729	0	2,266	0	2,282
5	1,596	0,577	0	2,089	0	2,115
6	1,410	0,483	0,030	1,970	0	2,004
7	1,277	0,419	0,118	1,882	0,076	1,924
8	1,175	0,372	0,185	1,815	0,136	1,864
9	1,094	0,337	0,239	1,761	0,184	1,816
10	1,028	0,308	0,284	1,716	0,223	1,777

الفصل التاسع

تحليل التباين باتجاه واحد

One-way analysis of variance (ANOVA)

1 . مقدمة Introducation.

2 . تصميم تجربة تحليل التباين باتجاه واحد (التصنيف الأحادي).

3 . المقارنات المتعددة للمتوسطات Multiple comparisons for means .

الفصل التاسع

تحليل التباين باتجاه واحد

One-way analysis of variance (ANOVA)

9-1- مقدمة Introducation:

جاء في بحث اختبار الفرضيات، أنه إذا كان لدينا مجتمعين طبيعيين، وأردنا اختبار معنوية الفرق بين متوسطيهما، فإننا نلجأ إلى توزيع ستيودينت أو التوزيع الطبيعي لمعرفة فيما إذا كان الفرق بينهما يعود لأسباب جوهرية أو إذا كان يعود لأسباب عشوائية.

ولكن إذا كان لدينا أربعة مجتمعات وأردنا تقصي الفروق بين متوسطاتها الأربعة فإنه يتوجب علينا إجراء اختبار لكل زوج من المتوسطات على حدة باستخدام اختبار ستودينت (t-test)، وهذا يتطلب منا القيام بإجراء ست مقارنات؛ أي استخدام اختبار ستيودينت ست مرات، فإذا كان مستوى الدلالة Level of significance المستخدم في الاختبار $\alpha = 0,05$ عند كل اختبار على حدة، فإن احتمال الوصول إلى القرار الصحيح للاختبارات الستة يساوي $(0,95)^6 = 0,735$ ، وبالتالي احتمال الوصول إلى قرار واحد أو أكثر غير صحيح يساوي: $1-(0,95)^6 = 0,265$.

لنأخذ حالة أن يكون عدد المجتمعات ستة ولتقصي الفروق بين متوسطاتها الستة فإنه يتوجب علينا إجراء اختبار ستيودينت 15 مرة. فإذا كان مستوى الدلالة المستخدمة $\alpha = 0,05$ عند كل اختبار فإن احتمال الوصول إلى قرار صحيح للاختبارات الخمسة عشرة يساوي $1-(0,95)^{15} = 0,537$. ومنه نستنتج أن احتمال الوصول إلى قرار غير سليم يكبر كلما كبر عدد المجتمعات. لذلك يتطلب منا البحث عن طريقة إحصائية جيدة يمكن استخدامها لمثل هذه الحالات. وإن الاختبار المناسب هو اختبار F (F test) الذي نحصل عليه من تحليل التباينِ Analysis of variance والذي يبنى على تفكيك التباين إلى عدة مركبات جزئية تشكل كل منها قياساً للتغير. فقد تكون التغيرات

داخلية ضمن كل مجتمع من المجتمعات المدروسة أو تكون تغيرات بين المجتمعات المدروسة. ومن أحد طرق التحليل هو تحليل التباين في اتجاه واحد ويدعى أيضاً التصنيف الأحادي One-way classification، لأنه يتضمن عاملاً واحداً فقط. ويستخدم لدراسة الاختلافات (الفروقات) في الأوساط الحسابية لمشاهدات متغير ما تعرض لمؤثر خارجي، ويعتبر من أفضل الأساليب الإحصائية المستخدمة في التحليل عندما تتعلق الدراسة بمقارنة عدة متوسطات لمجتمعات تم سحب منها عينات بحجوم متساوية أو مختلفة.

وكمثال على ذلك، نفرض أنه إذا كان لدينا أربعة طرق للتعليم وأربع شعب للصف الأول الابتدائي وأن أطفال كل شعبة تتلقى تعليمها بإحدى هذه الطرق، هنا من الممكن نستطيع تحديد فيما إذا كان لطرق التدريب تأثير معنوي على تحسين متوسط درجات الطلاب في كل من الطرق الأربعة، أو نستطيع اختبار فيما إذا كانت متوسطات درجات الطلاب في كل من الطرق الأربعة متساوية.

مثال آخر:

إذا كانت لدينا خمسة أصناف من بذور القطن تمت زراعتها في أراض مختلفة، ونريد اختبار فيما إذا كانت الفروق بين متوسطات الإنتاج جوهرية أم لا؟. ويمكن التنوية عند تصميم التجارب الخاصة بتحليل التباين يجب أن تتوفر الصفة العشوائية عند اختيار المجتمعات population أو العينات Sample.

9-2- تصميم تجربة تحليل التباين باتجاه واحد (التصنيف الأحادي):

كما ذكرنا سابقاً أن تحليل التباين في اتجاه واحد هو عبارة عن تصنيف المشاهدات (Observation) التي نحصل عليها عند إجراء التجارب إلى عدد من المجموعات Sets على أساس عامل واحد.

فإذا كان لدينا K مجتمعاً، وأن كل عنصر ينتمي إلى مجتمع واحد فقط من المجتمعات المختلفة. ولنفرض أن هذه المجتمعات مستقلة عن بعضها البعض ويخضع

كل منها للتوزيع الطبيعي بمتوسطات $\mu_1, \mu_2, \mu_3, ...\mu_k$ ، ولكل منها التباين σ^2 نفسه ونريد القيام بمقارنات تتعلق بمقادير هذه المتوسطات.

ومن أجل ذلك نسحب عينة عشوائية بسيطة من كل مجتمع من هذه المجتمعات بحجم n . وقد تكون حجوم العينات المسحوبة متساوية وقد تكون غير متساوية ولنأخذ حالة كون العينات متساوية ونرمز بـ x_{ij} إلى القيمة العددية المشاهدة التي ترتيبها i في العينة والمسحوبة من المجتمع j :

حيث: $i = 1,2,3.....n$, $j = 1,2,3.....k$ وبالتالي فإن:

x_{11} تمثل القيمة الأولى المشاهدة في العينة المسحوبة من المجتمع الأول.

x_{23} تمثل القيمة الثانية المشاهدة في العينة المسحوبة من المجتمع الثالث.

.....

x_{n4} تمثل القيمة الأخيرة المشاهدة في العينة المسحوبة من المجتمع الرابع.

وهكذا.....

لنرمز بـ T_1 إلى مجموع القياسات (القيم العددية) في العينة الأولى، ولمتوسطها بـ $\overline{x}_1$

لنرمز بـ T_2 إلى مجموع القياسات (القيم العددية) في العينة الثانية ، ولمتوسطها بـ $\overline{x}_2$

لنرمز بـ T_3 إلى مجموع القياسات (القيم العددية) في العينة الثالثة ، ولمتوسطها بـ $\overline{x}_3$

.....

ولنرمز بـ T_k إلى مجموع القياسات (القيم العددية) في العينة k ، ولمتوسطها بـ $\overline{x}_k$

ونرمز لمجموع القياسات (القيم العددية) في جميع العينات بـ T ، ولمتوسطها العام بـ $\overline{X}$ أي:

$$T = \sum_{j=1}^{k} T_j \qquad (1-9)$$

$$\overline{X} = \frac{T}{N}$$

حيث: $N = n \cdot k$ في حالة تساوي حجوم العينات المسحوبة. وأن:

$$T_j = \sum_{i=1}^{n} T_{i\,j} \qquad (2-9)$$

$$\overline{x_j} = \frac{\sum_{i=1}^{n} x_{i\,j}}{n} = \frac{T_j}{n}$$

ويمكن تلخيص ماتقدم في الجدول رقم (1) التالي:

الجدول رقم (1) يوضح قيم المشاهدات للعينات المسحوبة والمتساوية

المجتمعات / المشاهدات	1 2 3 .....j.....k	Σ
1 2 3 : i : : n	x_{11} x_{12} x_{13}.....$x_{i\,j}$.....x_{1k} x_{21} x_{22} x_{23}.....$x_{2\,j}$.....x_{2k} x_{31} x_{32} x_{33}.....$x_{3\,j}$.....x_{3k} ⋮ ⋮ ⋮ ⋮ ⋮ x_{i1} x_{i2} x_{i3}.....x_{ij}.....x_{ik} ⋮ ⋮ ⋮ ⋮ ⋮ x_{n1} x_{n2} x_{n3}....$x_{n\,j}$....x_{nk}	
	n n n.....n.....n	$n \cdot k = N$
$\sum_{i=1}^{n} x_{ij} = T_j$	T_1 T_2 T_3.....T_J.....T_K	$\sum_{j=1}^{k} T_j = T$
$\frac{T_j}{n} = \overline{x}_j$	$\overline{x}_1$ $\overline{x}_2$ $\overline{x}_3$.....$\overline{x}_J$.....$\overline{x}_K$	$\frac{T}{N} = \overline{X}$

وبما أننا نرغب في معرفة فيما إذا كان هناك فروق بين متوسطات المجتمعات التي سحبت منها العينات العشوائية، لابد من إيجاد طريقة لاختبار فرضية العدم التالية:

متوسطات المجتمعات متساوية:

$$H_0 : \mu_1 = \mu_2 = \mu_3 = ... \mu_k \qquad (3-9)$$

أي أن متوسطات المجتمعات متساوية مقابل الفرضية البديلة أنه يوجد متوسطان على الأقل غير متساويين : H_1

ومن أجل ذلك لابد من القيام بما يلي:

9-2-1: صياغة النموذج العام:

من الجدول رقم (1) نلاحظ أن كل قيمة مشاهدة مثل x_{ij} في كل مجموعة (معالجة) حيث: $i = 1,2,3.....n$ قد تزيد أو تنقص عن متوسطها $\bar{x}_{ij}$ حيث $j = 1,2,3.....k$ ، وبالتالي فإن x_{ij} هي عبارة عن متحولات عشوائية تتوزع حول المتوسط μ_j وفق التوزيع الطبيعي ومستقلة عن بعضها البعض ولها نفس التباين σ^2. وبالتالي نستطيع كتابة كل مشاهدة من x_{ij} على الشكل التالي:

$$x_{ij} = \mu_j + \varepsilon_{ij} \Leftrightarrow \varepsilon_{ij} = x_{ij} - \mu_j \qquad (4-9)$$

حيث: ε_{ij} تشير إلى مقدار انحراف المشاهدة i في المعالجة j عن متوسط المجتمع μ_j وهي متحولات عشوائية مستقلة عن بعضها البعض، تتوزع وفق التوزيع الطبيعي بمتوسط يساوي الصفر وتباين σ^2.

إن المتوسط العام ونرمز له بـ μ لـ k مجتمعاً متوسطاتها $\mu_1, \mu_2, \mu_3, ... \mu_k$ هو عبارة عن متوسط المتوسطات والذي يعطى بالعلاقة:

$$\mu = \frac{\sum_{j=1}^{n} \mu_j}{k} \qquad (5-9)$$

وبذلك نجد أن متوسط كل مجتمع من المجتمعات μ_j قد يزيد أو ينقص عن المتوسط العام μ ، نرمز لهذا الفرق بـ α_j أي:

$$\alpha_j = \mu_j - \mu \qquad (6-9)$$

أو :

$$\mu_j = \mu + \alpha_j$$

حيث: α_j أعداد ثابتة تشير إلى تأثير المجتمع j.

وبتعويض العلاقة (9–6) في العلاقة (9–4) نتمكن من صياغة النموذج التالي:

$$x_{ij} = \mu + \alpha_j + \varepsilon_{ij} \qquad (7-9)$$

حيث: $i = 1,2,3.....n$

$j = 1,2,3.....k$

μ و α_j أعداد ثابتة وأن: $\sum_{j=1}^{k} \alpha_j = 0$

9–2–2: فرضيات الاختبار:

من خلال النموذج رقم (9–7) نجد أنه إذا كانت متوسطات جميع المجتمعات متساوية، فإن $\alpha_1 = \alpha_2 = ... = \alpha_j = ...\alpha_k = 0$ وبالتالي فإن تأثير كل مجتمع يصبح مساوياً الصفر، وإذا كان يوجد متوسطان على الأقل غير متساويين فإن واحدة على الأقل من α_j لاتساوي الصفر، وبالتالي فإن الفرضية رقم (9–3) القائلة بأن جميع متوسطات المجتمعات متساوية مقابلة الفرضية البديلة H_1 والقائلة بأنه يوجد متوسطان على الأقل غير متساويين تصبح مكافئة للفرضية التالية:

الفرضية الابتدائية:

$$H_0 = \alpha_1 = \alpha_2 = ... = \alpha_j = ...\alpha_k = 0 \qquad (8-9)$$

أي أن جميع متوسطات المجتمعات متساوية، وحيث أن: $\alpha_j = \mu_j - \mu$

مقابل الفرضية البديلة (توجد واحدة على الأقل من α_j لاتساوي الصفر): H_1 :

9-2-3: حساب التباينات:

نعلم أن مفهوم التباين الكلي يعتمد على مقدار مجموع المربعات الكلي the total sum of squares والذي هو عبارة مجموع مربعات انحرافات القيم التي تضمها التجربة عن متوسطها الحسابي العام $\overline{X}$ ونرمز له بـ SST والذي يعطى بالعلاقة:

$$SST = \sum_{j=1}^{k}\sum_{i=1}^{n} (x_{ij} - \overline{X})^2 \qquad (9-9)$$

وسنقوم الآن بتقسيم مجموع المربعات الكلي STT إلى مركباته الأساسية على الشكل التالي:

$$SST = \sum_{j=1}^{k}\sum_{i=1}^{n}(x_{ij} - \bar{x}_j + \bar{x}_j - \overline{X})^2$$

$$= \sum_{j=1}^{k}\sum_{i=1}^{n}\left[\left(x_{ij} - \bar{x}_j\right) + \left(\bar{x}_j - \overline{X}\right)\right]^2$$

$$SST = \sum_{j=1}^{k}\sum_{i=1}^{n}\left[\left(x_{ij} - \bar{x}_j\right)^2 + 2\left(\bar{x}_{ij} - \bar{x}_j\right)\left(\bar{x}_j - \overline{X}\right) + \left(\bar{x}_j - \overline{X}\right)^2\right]$$

$$SST = \sum_{j=1}^{k}\sum_{i=1}^{n}\left(x_{ij} - \bar{x}_j\right)^2 + 2\sum_{j=1}^{k}\sum_{i=1}^{n}\left(x_{ij} - \bar{x}_j\right)\left(\bar{x}_j - \overline{X}\right) + \sum_{j=1}^{k}\sum_{i=1}^{n}\left(\bar{x}_j - \overline{X}\right)^2 \quad (10-9)$$

إن الحد الثاني على يمين المساواة يساوي الصفر لأنه عبارة عن المجموع $\left(\bar{x}_i - \bar{x}\right)\sum_{i=1}^{n}\left(x_{ij} - \bar{x}_j\right)$ مضروباً بـ $\left(\bar{x}_j - \bar{x}\right)$ وهو يساوي الصفر لأنه يمثل انحرافات القيم عن متوسطها الحسابي. إذاً العلاقة (9-10) تصبح بالشكل التالي:

$$SST = \sum_{j=1}^{k}\sum_{i=1}^{n}\left(x_{ij} - \bar{x}_j\right)^2 + \sum_{j=1}^{k}\sum_{i=1}^{n}\left(\bar{x}_j - \overline{X}\right)^2$$

$$SST = \sum_{j=1}^{k}\sum_{i=1}^{n}\left(x_{ij} - \bar{x}_j\right)^2 + n\sum_{j=1}^{k}\left(\bar{x}_j - \overline{X}\right)^2 \qquad (11-9)$$

إن الحد الأول $\sum_{j=1}^{k}\sum_{i=1}^{n}\left(x_{ij} - \bar{x}_j\right)^2$ هو عبارة عن مجموع مربعات انحرافات قيم مشاهدات العينات عن متوسطاتها الموافقة لها $\bar{x}_i$ ويرمز له بـ SSE ويدعى بتباين الأخطاء أو مجموع مربعات البواقي، وتعتمد قيمته على التغيرات التصادفية التي ترافق

التجربة. ويطلق عليه أيضاً مجموع المربعات داخل المجموعات the within Groups sum of squares

والحد الثاني $n\sum_{j=1}^{k}\sum_{i=1}^{n}\left(\bar{x}_j - \bar{X}\right)^2$ هو عبارة عن مجموع مربعات انحرافات متوسطات العينات عن المتوسط العام، ويرمز له بـ SSA ويدعى بتباين المعاملات أو مجموع المربعات بين المجموعات The among groups sum of squares. وكلما كانت قيم متوسطات العينات قريبة من بعضها فإن قيمة SSA تكون منخفضة.

وبالتالي يمكن أن نعبر عن العلاقة (9−11) بالشكل التالي:

$$SST = SSE + SSA$$

حيث:

$$SSE = \sum_{j=1}^{k}\sum_{i=1}^{n}\left(x_{ij} - \bar{x}_j\right)^2 \qquad (12-9)$$

$$SSA = n\sum_{j=1}^{k}\left(\bar{x}_j - \bar{\bar{X}}\right)^2 \qquad (13-9)$$

ومن الصعوبة إجراء الحسابات الخاصة بالعلاقات (9−9) و(9−12) و(9−13)، لذلك نستبدلها بعلاقات أخرى أسهل بكثير تمكننا من استخدام الآلات الحاسبة, وذلك بالاعتماد على الخاصة التي مفادها أن مجموع مربعات انحرافات مجموعة من القيم x_i، وليكن عددها N عن متوسطها الحسابي $\bar{x}$ يساوي مجموع مربعات القيم $\sum_{i=1}^{N} x_i^2$ مطروحاً منها حاصل قسمة مربع مجموع هذه القيم على عددها N والتي تأخذ الشكل العام التالي:

$$\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2 = \sum_{i=1}^{N} x_i^2 - \frac{\left(\Sigma x_i\right)^2}{N}$$

وبناءً على هذه الخاصة تصبح العلاقات (9−9) و (9−12) و(9−13) على الشكل التالي:

$$SST = \sum_{j=1}^{k}\sum_{i=1}^{n} x_{ij}^2 - \frac{T^2}{N} \qquad (14-9)$$

حيث: $N = n \cdot k$ في حالة العينات متساوية الحجم.

$$SSE = \sum_{j=1}^{k}\sum_{i=1}^{n} x_{ij}^2 - \frac{\sum_{j=1}^{k} T_j^2}{n} \qquad (15-9)$$

$$SSA = \frac{\sum_{j=1}^{k} T_j^2}{n} - \frac{T^2}{N} \qquad (16-9)$$

وبما أننا نريد اختبار فرضية $\mu_1 = \mu_2 = ...\mu_k$ بأن هذه العينات جاءت فرضاً من مجتمع طبيعي واحد تباينه σ^2، فمن مجموع المربعات نستطيع الحصول على تقديرين مستقلين لـ σ^2 هما:

. التقدير الأول لـ σ^2 : إن SSE هو عبارة عن مجموع مربعات انحرافات قيم مشاهدات العينات عن متوسطاتها الموافقة لها.

فإذا قمنا بتقسيم هذا المجموع على $nk - k = N - k$ باعتبار أن حجوم العينات متساوية فإننا نحصل على التباين داخل المجموعات within groups variance ونرمز له بـ MSE ويساوي:

$$MSE = \frac{\sum_{i=1}^{n}\sum_{j=1}^{k}\left(x_{ij} - \bar{x}_j\right)^2}{N-k}$$

أو:

$$MSE = \frac{SSE}{N-k} \qquad (17-9)$$

وهو تقدير غير متحيز لـ σ^2، ومبني على N-k من درجات الحرية.

. التقدير الثاني لـ σ^2 : إن التقدير الثاني لـ σ^2 يمكن حسابه من تباين متوسط العينة المسحوبة من المجتمع والذي يعطى بالعلاقة :

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} \Rightarrow \sigma^2 = n.\sigma_{\bar{x}}^2 \qquad (18-9)$$

وعند حساب متوسطات العينات $\bar{x}_1, \bar{x}_2, \bar{x}_3, ... \bar{x}_k$ فإنها تشكل عينة بحجم k من المجتمع متوسطها الحسابي $\overline{X}$ وبالتالي فإن تباين هذه المتوسطات يعطى بالعلاقة:

$$\sigma_{\bar{x}}^2 = \frac{\sum_{j=1}^{k} \left(\bar{x}_j - \overline{X}\right)^2}{k-1} \qquad (19-9)$$

وهو تقدير غير متحيز لـ σ_n^2 وبالتعويض في العلاقة في (9-18) نحصل على التقدير الثاني لـ σ^2 ونرمز له بـ MSA ويعبر عن التباين بين المجموعات Among groups variance ويساوي:

$$MSA = \frac{n.\sum_{j=1}^{k} \left(\bar{x}_j - \overline{X}\right)^2}{k-1} \qquad (20-9)$$

نلاحظ أن البسط ماهو إلا عبارة عن SSA وبالتالي فإن MSA يصبح بالشكل التالي:

$$MSA = \frac{SSA}{k-1} \qquad (21-9)$$

وهو تقدير لـ σ^2 مبني على k-1 من درجات الحرية. ويعتبر تقديراً غير متحيزاً لـ σ^2 عندما تكون الفرضية الابتدائية صحيحة.ويعطي تقديراً أكبر لـ σ^2 عندما تكون الفرضية البديلة H_1 صحيحة.

أما بالنسبة لدرجات الحرية الخاصة بـ SST والموافقة للعدد الكلي للقياسات التي تتضمنها التجربة هي: $N-k+k-1=N-1$ درجة حرية.

9-2-4: حساب نسبة التباين The variance Ratio:

ونحصل عليه من خلال نسبة MSA (التباين بين المجموعات) إلى MSE (التباين داخل المجموعات) ونرمز لهذه النسبة بـ $\widetilde{F}$ أي:

$$\widetilde{F} = \frac{MSA}{MSE} \qquad (22-9)$$

وهـذه النسـبة تخضـع لتوزيـع F بـدرجات حريـة Degrees of freedom $V_1 = K - 1,\ V_2 = N - K$، وتصغر قيمة F كلما كان تقدير تباين المجتمع الناتج عن الفروق بين أوساط المجموعات أصغر والعكس صحيح.

9-2-5: تنظيم جدول تحليل التباين(The ANOVA table):

يمكن تشكيل جدول يتضمن ماتقدم من خطوات يدعى بجدول تحليل التباين الأحادي (جدول تحليل التباين باتجاه واحد) وفق الشكل التالي:

الجدول رقم (2) لتحليل التباين في اتجاه واحد The ANOVA table

مصدر التشتت (التغير)	مجموع المربعات	درجات الحرية	متوسط المربعات	نسبة التباين $\widetilde{F}$	قيمة F الجدولية
مابين المجموعات أو العينات (المعالجات)	SSA	$V_1 = k - 1$	$MSA = \frac{SSA}{k-1}$	$\widetilde{F} = \frac{MSA}{MSE}$	نبحث عنها حسب مستوى الدلالة α
داخل العينات (المعالجات)	SSE	$V_2 = N - k$	$MSE = \frac{SSE}{N-k}$		
المجموع	SST	$N - 1$			

9-2-6: اختبار الفرضية واتخاذ القرار الإحصائي Tests of hypothesis:

إذا كانـت H_0 غيـر صـحيحة أي أن العينـات مـأخوذة مـن مجتمعـات تختلـف متوسـطاتها عـن بعضـها، وبالتـالي فـإن MSA تعطـي تقدير لـ σ^2 أعلـى مـن الواقـع بالمقارنـة مـع MSE وبالتـالي فـإن النسبة F تتجاوز F_α (باعتبار أننـا نستخدم الطـرف الأيمـن لتوزيـع F كمنطقـة حرجـة Critical region لاختبـار الفرضـية H_0). وبالتـالي نرفض H_0 إذا كان:

$$\widetilde{F} > F_{(\alpha, V_1, V_2)}$$

وبالتـالي نسـتنتج أنـه يوجـد متوسـطين علـى الأقـل مـن المتوسـطات $\mu_1, \mu_2, \mu_3, \ldots \mu_k$ غير متساويين.

وعندما يكون MSA أصغر من MSE فإن النسبة لاتتجاوز F_α وبالتالي نقبل الفرضية H_0 عندما:

$$\widetilde{F} < F_{(\alpha,V_1,V_2)}$$

ونقول أن جميع متوسطات المجتمعات $\mu_1, \mu_2, \mu_3, ...\mu_k$ متساوية.

مثال 1:

قام أحد الباحثين بإجراء تجارب حول استخدام أربعة أنواع من الأسمدة وتأثيرها على مردود إنتاج القطن خلال ست تجارب وكانت المعلومات التي حصل عليها كما يلي:

i \	1	2	3	4
1	50	48	42	64
2	52	56	48	56
3	53	60	50	55
4	60	62	50	54
5	49	50	47	48
6	54	54	51	47

والمطلوب: اختبار فيما إذا كان هناك فرق بين متوسطات الإنتاج (أو تأثيرات الأسمدة على مردودية إنتاج القطن) بمستوى دلالة 0.05.

الحل:

لإجراء اختبار فيما إذا كان هناك فرق بين متوسطات الإنتاج نتبع الخطوات التالية:

1 . تحديد فرضيات الاختبار:

الفرضية الابتدائية:

$$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$$

لايوجد فرق بين متوسطات الإنتاج وأن العينات الأربعة مأخوذة من مجتمع واحد.

الفرضية البديلة: توجد واحدة من α_j لا تساوي الصفر: H_1 أي يوجد فرق بين متوسطات الإنتاج وأن العينات الأربعة ليست مأخوذة من مجتمع واحد.

2 . إعداد الجدول المساعد التالي:

الجدول المساعد

j / i	1	2	3	4	Σ
1	50	48	42	64	
2	52	56	48	56	
3	53	60	50	55	
4	60	62	50	54	
5	49	50	47	48	
6	54	54	51	47	
n	6	6	6	6	$N=24$
$\sum_{i=1}^{n} x_{ij} = T_j$	318	330	288	324	$T=1260$
$\frac{T_j}{n} = \bar{x}_j$	53	55	48	54	$\bar{x}=52,5$

3 . حساب التباينات:

$$SST = \sum_{i=1}^{n} \sum_{j=1}^{k} x_{ij}^2 - \frac{T^2}{2}$$

$$SST = \left[(50)^2 + (52)^2 + (53)^2 + (60)^2 + \ldots\ldots\ldots + (48)^2 + (47)^2\right] - \frac{(1260)^2}{24}$$

$$SST = 66794 - 66150 = 644$$

$$SSA = \frac{\sum_{j=1}^{k} T_j^2}{n} - \frac{T^2}{N}$$

$$SSA = \frac{1}{6}\left[(318)^2 + (330)^2 + (288)^2 + (324)^2\right] - \frac{(1260)^2}{24}$$

$$SSA = 66324 - 66150 = 174$$

$$SSE = SST - SSA$$

$$SSE = 644 - 174 = 470$$

. التباين داخل المجموعات (المعالجات):

$$MSE = \frac{SSE}{N-k} = \frac{470}{24-4} = 23,5$$

. التباين ما بين المجموعات (المعالجات):

$$MSA = \frac{SSA}{k-1} = \frac{174}{4-1} = 58$$

4 . نسبة التباين:

$$\widetilde{F} = \frac{MSA}{MSE} = \frac{58}{23,5} = 2,47$$

وهذا يعني أن التباين بين المجموعات يعادل ضعفين ونصف تقريباً التباين داخل المجموعات.

5 . جدول تحليل التباين:

جدول تحليل التباين في اتجاه واحد

مصدر التشتت (التغير)	مجموع المربعات	درجات الحرية	متوسط المربعات	نسبة التباين	قيمة F الجدولية
مابين المجموعات	$SSA = 174$	$V_1 = k-1 = 3$	$MSA = 58$	$\widetilde{F} = \frac{MSA}{MSE} = \frac{58}{23,5} = 2,47$	3,10
داخل المجموعات	$SSE = 470$	$V_2 = N-K = 20$	$MSE = 23,5$		
المجموع	$SST = 644$	$N-1 = 23$			

6 . اختبار الفرضية واتخاذ القرار الإحصائي:

لاختبار الفرضية نقارن قيمة $\widetilde{F}$ المحسوبة مع القيمة الجدولية لـ $F_{0,95}$ عند درجات حرية $V_1 = 3$, $V_2 = 20$ والتي تساوي:

$$F_{0,95(3,20)} = 3,10$$

وبالمقارنة نجد أن: $\widetilde{F} < F_{0,95(3,20)}$

وبالتالي نقب ل الفرضية القائلة بعدم وجود فروق بين متوسطات الإنتاج. أو لاتوجد فروق بين تأثيرات الأسمدة المتنوعة على إنتاج القطن.

*** حالة كون حجوم العينات غير متساوية:**

عندما تكون حجوم العينات المسحوبة من المجتمعات غير متساوية، أي أن:

حجم العينة الأولى n_1

حجم العينة الثانية n_2

حجم العينة الثالثة n_3

.....

حجم العينة j يساوي n_j

.....

حجم العينة k يساوي n_k

وبالتالي فإن:

$$N = n_1 + n_2 + n_3 + ... + n_j + ... + n_k$$

أو:

$$N = \sum_{j=1}^{k} n_j$$

ويوضح الجدول قيم المشاهدات للعينات المسحوبة غير المتساوية بالشكل التالي:

الجدول رقم (3) يوضح قيم مشاهدات العينات غير المتساوي

المجتمعات / المشاهدات	1 2 3 .....j.....k							Σ
1	x_{11}	x_{12}	x_{13}	.....	x_{ij}	.....	x_{1k}	
2	x_{21}	x_{22}	x_{23}	.....	x_{2j}	.....	x_{2k}	
3	x_{31}	x_{32}	x_{33}	.....	x_{3j}	.....	x_{3k}	
⋮	⋮...	⋮...	⋮...	⋮...	⋮...	⋮...	⋮...	
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	
i	x_{i1}	x_{i2}	x_{i3}	.....	x_{ij}	.....	x_{ik}	
⋮	⋮...	⋮...	⋮...	⋮...	⋮...	⋮...	⋮...	
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	
n	x_{n1}	x_{n2}	x_{n3}	.....	x_{nj}	.....	x_{nk}	
حجم العينات	n_1	n_2	n_3	…	n_j	⋯	n_k	$N=\sum_{j=1}^{k} n_j$
$\sum_{i=1}^{n_i} x_{ij} = T_j$	T_1	T_2	T_3	…	T_j	⋯	T_k	$\sum_{j=1}^{k} T_j = T$
$\bar{x}_j$	$\bar{x}_1$	$\bar{x}_2$	$\bar{x}_3$	…	$\bar{x}_j$	⋯	$\bar{x}_k$	

وبالتالي هنا يطرأ تغيير على مجموع المربعات بحيث تأخذ العلاقات (9-14) و(9-15) و(9-16) الأشكال التالية:

$$SST = \sum_{j=1}^{k}\sum_{i=1}^{nj} x_{ij}^2 - \frac{T^2}{\sum_{j=1}^{k} n_j} \qquad (23-9)$$

$$SSE = \sum_{j=1}^{k}\sum_{i=1}^{nj} x_{ij}^2 - \frac{\sum_{j=1}^{k} T_j^2}{n_j} \qquad (24-9)$$

$$SSA = \frac{\sum_{j=1}^{k} T_j^2}{n_j} - \frac{T^2}{\sum_{j=1}^{k} n_j} \qquad (25-9)$$

كما أن التباين داخل المجموعات الذي نرمز له بـ MSE مبني على $\sum_{j=1}^{k}(n_j - 1)$ من درجات الحرية. أو: $\sum_{j=1}^{k} n_{j-k}$

وبالتالي فإن جدول تحليل التباين في اتجاه واحد عندما تكون حجوم العينات غير متساوية يأخذ الشكل التالي:

مصدر التشتت (التغير)	مجموع المربعات	درجات الحرية	متوسط المربعات	نسبة التباين $\widetilde{F}$
مابين المجموعات (المعالجات)	SSA	$V_1 = k - 1$	$MSA = \frac{SSA}{K-1}$	$\widetilde{F} = \frac{MSA}{MSE}$
داخل العينات (داخل المجموعات)	SSE	$V_2 = \sum_{j=1}^{k} n_j - k$	$MSE = \frac{SSE}{\sum_{j=1}^{k} n_j - k}$	
المجموع	SST	$\sum_{j=1}^{k} n_j - 1$		

وما تبقى من خطوات كما هي في حالة حجوم العينات المتساوية.

مثال 2:

قام أحد منتجي المنظفات بعرض إنتاجه بثلاث طرق وهي:

الطريقة الأولى: عبوات بلاستيكية . وقد تم عرض المنتج في (6) مخازن.

الطريقة الثانية: أكياس نايلون . وقد تم عرض المنتج في (4) مخازن.

الطريقة الثالثة: غير معبأة . وقد تم عرض المنتج في (8) مخازن.

إذا علمت أن هذه المخازن تم اختيارها بشكل عشوائي.

وقد كانت المبيعات الشهرية في كل مخزن وحسب كل طريقة كما في الجدول التالي:

الطريقة j / المخزنة i	1	2	3
1	50	45	55
2	55	35	30
3	40	40	60
4	35	40	45
5	60	-	60
6	66	-	55
7	-	-	65
8	-	-	70

والمطلوب:

اختبار تأثير الطرق الثلاث على مبيعات المنظفات إذا علمت أن المبيعات مقدرة بالكغ وذلك بمستوى دلالة 0,05 .

الحل:

1 . تحديد الفرضيات:

إن الفرضية الابتدائية : H_0 والمراد اختبارها هي عدم وجود فرق بين متوسطات مبيعات المنظفات حسب الطرق الثلاث مقابل الفرضية البديلة : H_1 والقائلة يوجد متوسطين على الأقل غير متساويين ويمكن التعبير عن هذه الفرضيات بالشكل:

الفرضية الابتدائية:

لا يوجد فرق بين متوسطات المبيعات حسب الطرق الثلاث

$$H_0 : \mu_1 = \mu_2 = \mu_3$$

مقابل الفرضية البديلة:

يوجد متوسطين على الأقل غير متساويين : H_1

فإذا كانت متوسطات المجتمعات متساوية فإن العينات قد سحبت من مجتمع واحد.

وبالتالي فإن تأثير كل طريقة يساوي الصفر. ومنه فإن الفرضية السابقة تصبح بالشكل:

الفرضية الابتدائية:

لا يوجد تأثير للطرق الثلاث على مبيعات المنظفات $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = 0$

مقابل الفرضية البديلة:

H_1 : توجد واحدة على الأقل من α_j لا تساوي الصفر

يوجد تأثير لطرق البيع على مبيعات المنظفات.

2 . إعداد الجدول المساعد التالي:

j / i	1	2	3	Σ
1	50	45	55	
2	55	35	30	
3	40	40	60	
4	35	40	45	
5	60	-	60	
6	66	-	55	
7	-	-	65	
8	-	-	70	
n_j	6	4	8	$N = \sum_{j=1}^{k} n_j = 18$
$\sum_{i=1}^{nj} x_{ij} = T_j$	306	160	440	$T = \sum_{j=1}^{k} T_j = 906$
$\frac{T_j}{N} = \bar{x}_j$	51	40	5	

3 . حساب التباينات:

نقوم بحسابها من العلاقات (9−23) و(9−24) و(9−25) فنجد:

$$SST = \sum_{j=1}^{k} \sum_{i=1}^{nj} x_{ij}^2 - \frac{T^2}{\sum_{j=1}^{k} n_j}$$

$$SST = \begin{bmatrix} (50)^2 + (55)^2 + ... + (66)^2 + ... + (45)^2 + \\ ..... + (40)^2 + ... + (55)^2 + ... + (70)^2 \end{bmatrix} - \frac{820836}{18}$$

$$SST = 48056 - \frac{820836}{18} = 2454$$

$$SSA = \frac{\sum_{j=1}^{k} T_j^2}{n_j} - \frac{T^2}{\sum_{j=1}^{k} n_j}$$

$$SSA = \left[\frac{(306)^2}{6} + \frac{(160)^2}{4} + \frac{(440)^2}{8} \right] - \frac{(906)^2}{18}$$

$$SSA = 46206 - 45602 = 604$$

ويمكن حساب SSE بالشكل:

$$SSE = SST - SSA$$

$$SSE = 2454 - 604 = 1850$$

التباين داخل المجموعات (المعالجات):

$$MSE = \frac{SSE}{\sum_{j=1}^{k} n_j - k}$$

$$MSE = \frac{1850}{18 - 3} = 123{,}33$$

التباين ما بين المجموعات (المعالجات):

$$MSA = \frac{SSA}{k - 1}$$

$$MSA = \frac{604}{3 - 1} = 302$$

4 . نسبة التباين:

$$\widetilde{F} = \frac{MSA}{MSE} = \frac{302}{123,33} = 2,45$$

وهذا يعني أن التباين بين المجموعات يعادل ضعفين ونصف تقريباً التباين داخل المجموعات.

5 . جدول تحليل التباين:

جدول تحليل التباين

مصدر التغير	مجموع المربعات	درجات الحرية	متوسط المربعات	نسبة التباين $\widetilde{F}$	قيمة F الجدولية
مابين المجموعات (المعالجات)	SSA	$V_1 = k - 1$ $V_1 = 3 - 1 = 2$	$MSA = 302$	$\widetilde{F} = \frac{MSA}{MSE}$ $\widetilde{F} = 2,45$	3.69
داخل المجموعات (داخل العينات)	SSE	$V_2 = \sum_{j=1}^{k} n_j - k$ $V_2 = 18 - 3 = 15$	$MSE = 123,33$		
المجموع	SST	$\sum_{j=1}^{k} n_j - 1$ $18 - 1 = 17$			

6 . اختبار الفرضية واتخاذ القرار الإحصائي:

لاختبار الفرضية نقارن $\widetilde{F}$ المحسوبة مع القيمة الجدولية لـ $F_{0,95}$ عند درجات حرية $V_1 = 2$ $\quad V_2 = 15,$ أي:

$$F_{0,95(2,15)} = 3,68$$

وبالمقارنة نجد أن $\widetilde{F} < F_{0,95(2,15)}$

وبالتالي نقبل الفرضية القائلة بعدم وجود فروق بين متوسطات المبيعات في الطرق الثلاث، وبالتالي لا يوجد تأثير لكل طريقة على كمية المبيعات.

3-9: المقارنات المتعددة للمتوسطات Multiple Comparisons For Means:

عندما نقوم بإجراء تحليل التباين ونختبر الفرضية H_0 والقائلة بأن متوسطات المجتمعات متساوية مقابل الفرضية H_1 القائلة بأنه يوجد متوسطين على الأقل من المتوسطات غير متساوية. فعندما تكون H_0 صحيحة (مقبولة) فإننا نقر بأن متوسطات المجتمعات المأخوذة منها العينات متساوية.

أما إذا كانت H_1 صحيحة (أي: H_0 مرفوضة) فإننا نقول أنه يوجد متوسطين على الأقل من المتوسطات غير متساوية، ولكننا لا نعلم أي من المتوسطات غير المتساوية. وللكشف عن المتوسطات غير المتساوية توجد عدة اختبارات في هذا المجال منها: اختبار بونفيروني Bonferroni test واختبار تيوكي Tukey test .

أما هنا فسوف نستخدم اختبار شفييه Scheffes test وهذا الاختبار يعتمد على المقارنات المتعددة للمتوسطات مثنى مثنى.

فإذا كان لدينا K مجتمعاً وأن متوسطاتها $\mu_1, \mu_2, ...\mu_j, ...\mu_k$ ونريد اختبار الفرضية:

الفرضية الابتدائية: $H_0 : \mu_1 = \mu_2 = ...\mu_k$ (جميع المتوسطات متساوية)

مقابل الفرضية البديلة: ليست جميع المتوسطات متساوية H_1.

لنفترض أنه تم رفض الفرضية H_0 وبالتالي نقبل الفرضية H_1 والقائلة بأنه يوجد متوسطين على الأقل غير متساويين. ولمعرفة أي المتوسطات غير متساوية نتبع الخطوات التالية:

1 . نجري جميع المقارنات لمتوسطات المجتمعات مثنى مثنى أي:

$\mu_1 = \mu_2$, $\mu_1 = \mu_3$ $\mu_1 = \mu_4$..... $\mu_1 = \mu_k$

$\mu_2 = \mu_3$, $\mu_2 = \mu_4$ $\mu_2 = \mu_5$..... $\mu_2 = \mu_k$

$\mu_3 = \mu_4$, $\mu_3 = \mu_5$ $\mu_3 = \mu_6$..... $\mu_3 = \mu_k$

.....

هكذا.....

أو بشكل عام : $\mu_i = \mu_j$

حيث:

$i = 1,2,3.....k-1$

$j = 1,2,3.....k$

2 . نقوم بكتابة فرضيات جميع المقارنات السابقة بالشكل التالي:

a) الفرضية الابتدائية: $H_0 : \mu_1 - \mu_2 = 0$

الفرضية البديلة: $H_1 : \mu_1 - \mu_2 \neq 0$

b) الفرضية الابتدائية: $H_0 : \mu_1 - \mu_3 = 0$

الفرضية البديلة: $H_1 : \mu_1 - \mu_3 \neq 0$

وهكذا .....

f) الفرضية الابتدائية: $H_0 : \mu_i - \mu_j = 0$

الفرضية البديلة: $H_1 : \mu_i - \mu_j \neq 0$

حيث:

$i = 1,2,3.....k-1$

$j = 1,2,3.....k$

3 . نقوم باختبار هذه الفرضيات مع التمييز بين حالتين:

في حالة حجوم العينات غير متساوية:

إن :

$$N = \sum_{j=1}^{k} n_j = n_1 + n_2 + n_3..... + n_k$$

وبالتالي فإن درجات الحرية تكون:

$$\sum_{j=1}^{k} n_j - k = N - K$$

ثم نحسب المقدار $\widetilde{T}$ والذي يعطى بالعلاقة:

$$\widetilde{T} = \frac{(\bar{x}_i - \bar{x}_j)^2}{s^2\left(\frac{1}{n_i} + \frac{1}{n_j}\right)} \qquad (26-9)$$

في حالة حجوم العينات المتساوية فإن:

$$n_1 = n_2 = n_3 = ...n_k$$

وبالتالي فإن درجات الحرية تكون:

$$nk - k = N - K$$

و s^2 تساوي متوسطات مربعات الخطأ (داخل المعالجات) وتؤخذ من جدول تحليل التباين مباشرة أي:

$$s^2 = MSE = \frac{SSE}{N-K}$$

وبالتالي تصبح العلاقة (9–26) بالشكل التالي:

$$\widetilde{T} = \frac{(\bar{x}_i - \bar{x}_j)^2}{s^2\frac{2}{n}} \qquad (27-9)$$

4 . نحسب قيمة المقدار $F_{(1-\alpha,V_1,V_2)}$ ثم نضربه بـ $(K-1)$ حيث K عدد المجتمعات ونرمز له بـ T^2 أي:

$$T^2 = (k-1)\cdot F_{(1-\alpha,V_1,V_2)} \qquad (28-9)$$

5 . نقارن $\widetilde{T}$ و T^2 فإذا كانت $\widetilde{T} < T^2$: فإننا نقبل H_0 والقائلة بأن $\mu_i - \mu_j = 0$ ، وبالتالي فإن المتوسطين متساويين.

وإذا كانت $\widetilde{T} > T^2$: فإننا نرفض H_0 ونقبل H_1 والقائلة بأن $\mu_i - \mu_j \neq 0$ وبالتالي فإن المتوسطين غير متساويين.

ولمعرفة أي المتوسطين أكبر فإننا نقارن $\sqrt{\widetilde{T}}$ و $T = \sqrt{T^2}$ فإذا كان:

$$\frac{\bar{x}_i - \bar{x}_j}{s\sqrt{\frac{2}{n}}} > T \qquad (29-9)$$

عندئذ نقول أن μ_i أكبر من μ_j $(\mu_i > \mu_j)$.

وإذا كان:

$$\frac{\bar{x}_i - \bar{x}_j}{s\sqrt{\frac{2}{n}}} < -T \qquad (30-9)$$

عندئذ نقول أن μ_i أصغر من μ_j $(\mu_i < \mu_j)$.

وفي حال كون حجوم العينات غير متساوية فإننا نستبدل : $\left(s\sqrt{\frac{2}{n}}\right)$ بـ $s\sqrt{\frac{1}{n_i}+\frac{1}{n_j}}$ في العلاقتين (9–29) و(9–30).

وبإجراء جميع الاختبارات بالنسبة لكل من i و j نستطيع معرفة أي متوسطات المجتمعات متساوية وأياً منها مختلفة.

مثال 3:

تم استخدام خمسة أنواع مختلفة من التكنولوجيا A_1, A_2, A_3, A_4, A_5 في زراعة البطاطا ولمدة خمس سنوات وكانت نتائج التجربة (الإنتاج) مقدراً بالطن مبينة في الجدول التالي:

التكنولوجيا / المشاهدات	A_1	A_2	A_3	A_4	A_5
1	11	15	10	8	19
2	10	6	15	4	15
3	9	4	18	6	17
4	8	6	12	7	15
5	12	8	18	10	15
6	10	3	11	13	15

والمطلوب:

1 . اختبار تأثير أنواع التكنولوجيا المستخدمة على إنتاج البطاطا.

2 . إذا كان هناك فرق بين متوسطات الإنتاج فأي المتوسطات مختلفة مستخدماً مستوى دلالة 0,05 .

الحل:

1 . تحديد فرضيات الاختبار.

الفرضية الابتدائية:

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$$

لايوجد فرق بين متوسطات الإنتاج حسب التكنولوجيا المستخدمة.

الفرضية البديلة:

H_1 ليست كل المتوسطات متساوية

فإذا كانت متوسطات المجتمعات متساوية فإن العينات قد سحبت من مجتمع واحد، وبالتالي فإن تأثير كل تكنولوجيا على الإنتاج يساوي الصفر ومنه فإن الفرضية السابقة تصبح بالشكل التالي:

الفرضية الابتدائية:

$$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = 0$$

لايوجد تأثير للتكنولوجيا المستخدمة على إنتاج البطاطا.

الفرضية البديلة:

H_1 توجد واحدة من α_j لاتساوي الصفر:

يوجد تأثير للتكنولوجيا المستخدمة على إنتاج البطاطا.

2 . إعداد الجدول المساعد:

التكنولوجيا / المشاهدات	A_1	A_2	A_3	A_4	A_5	Σ
1	11	15	10	8	19	
2	10	6	15	4	15	
3	9	4	18	6	17	
4	8	6	12	7	15	
5	12	8	18	10	15	
6	10	3	11	13	15	
n	6	6	6	6	6	$N = nK = 30$
$\sum_{i=1}^{n} x_{ij} = T_j$	60	42	84	48	96	$T = \sum_{j=1}^{k} T_j = 330$
$\overline{x}_j$	10	7	14	8	16	$\overline{X} = 11$

3 . حساب التباينات:

$$SST = \sum_{i=1}^{n} \sum_{j=1}^{k} x_{ij}^2 - \frac{T^2}{N}$$

$$SST = \begin{bmatrix} (11)^2 + (10)^2 + ... + (10)^2 + ... + (15)^2 + \\ ..... + (3)^2 + ... + (19)^2 + ... + (15)^2 \end{bmatrix} - \frac{(330)^2}{30}$$

$$SST = 4218 - 3630 = 588$$

$$SSA = \frac{\sum_{j=1}^{k} T_j^2}{n} - \frac{T^2}{N}$$

$$SSA = \frac{1}{6}\left[(60)^2 + (42)^2 + (84)^2 + (48)^2 + (96)^2\right] - \frac{(330)^2}{30}$$

$$SSA = 3990 - 3630 = 360$$

$$SSE = SST - SSA$$

$$SSE = 588 - 360 = 228$$

التباين داخل المجموعات (المعالجات):

$$MSE = \frac{SSE}{N-k} = \frac{228}{30-5} = 9{,}12$$

التباين ما بين المجموعات (المعالجات):

$$MSE = \frac{SSA}{k-1} = \frac{360}{5-1} = 90$$

4 . نسبة التباينين:

$$\widetilde{F} = \frac{MSA}{MSE} = \frac{90}{9{,}12} = 9{,}87$$

وهذا يعني أن التباين بين المجموعات يعادل تقريباً عشرة أضعاف التباين داخل المجموعات.

5 . جدول تحليل التباين:

جدول تحليل التباين

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	نسبة التباين	قيمة F الجدولية
مابين المجموعات	$SSA = 360$	$V_1 = k-1$ $V_1 = 5-1 = 4$	$MSA = 90$	$\widetilde{F} = \frac{MSA}{MSE}$ $\widetilde{F} = \frac{90}{9{,}12} = 9{,}87$	2.76
داخل المجموعات	$SSE = 228$	$V_2 = N-k$ $V_2 = 30-5 = 25$	$MSE = 9{,}12$		
المجموع	$SST = 588$	$N-1=$ $30-1 = 29$			

6 . اختبار الفرضية واتخاذ القرار الإحصائي:

لاختبار الفرضية نقارن قيمه $\widetilde{F}$ المحسوبة مع القيمة الجدولية لـ $F_{1-\alpha}$ أي $F_{0,95}$ عند درجات حرية $V_1 = 4$ $V_2 = 25,$ التي تساوي:

$$F_{0,95(4,25)} = 2{,}76$$

وبالمقارنة نجد أن $\widetilde{F} > F_{0,95(4,25)}$

وبالتالي نرفض الفرضية الابتدائية H_0 والقائلة بعدم وجود فروق بين متوسطات الإنتاج ونقبل الفرضية البديلة H_1 والقائلة بأنه يوجد فروق بين متوسطات الإنتاج ومنه يوجد تأثير للتكنولوجيا المستخدمة على إنتاج البطاطا.

2 . مقارنة المتوسطات:

باعتبار رفضنا الفرضية الابتدائية H_0 وقبلنا الفرضية البديلة H_1 ، وبالتالي يتوجب علينا الكشف عن المتوسطات غير المتساوية. ومن أجل القيام بذلك لابد من إجراء ما يلي:

1 . نجري المقارنات التالية:

$$\mu_1 = \mu_2,\ \ \mu_1 = \mu_3,\ \ \mu_1 = \mu_4,\ \ \mu_1 = \mu_5$$
$$\mu_2 = \mu_3,\ \ \mu_2 = \mu_4,\ \ \mu_2 = \mu_5$$
$$\mu_3 = \mu_4,\ \ \mu_3 = \mu_5$$
$$\mu_4 = \mu_5$$

2 . نقوم بكتابة فرضيات جميع المقارنات السابقة:

a) الفرضية الابتدائية: $H_0 : \mu_1 - \mu_2 = 0$

الفرضية البديلة: $H_1 : \mu_1 - \mu_2 \neq 0$

b) الفرضية الابتدائية: $H_0 : \mu_1 - \mu_3 = 0$

الفرضية البديلة: $H_1 : \mu_1 - \mu_3 \neq 0$

.....

وبشكل عام:

f) الفرضية الابتدائية: $H_0 : \mu_i - \mu_j = 0$

الفرضية البديلة: $H_1 : \mu_i - \mu_j \neq 0$

حيث:

$$i = 1,2,3.....k-1$$

$j = 1, 2, 3.....k$

3 . تقوم باختبار الفرضيات السابقة مع الأخذ بعين الاعتبار أن حجوم العينات متساوية:

أي أن درجات الحرية تساوي:

$$nK - K = N - K = 30 - 5 = 25$$

نحسب المقدار:

$$\widetilde{T} = \frac{\left(\overline{x}_i - \overline{x}_j\right)^2}{s^2 . \frac{2}{n}}$$

بالنسبة لـ $\mu_1 - \mu_2 = 0$

$$\widetilde{T}_1 = \frac{(10-7)^2}{9{,}12 . \frac{2}{6}} = 2{,}96$$

بالنسبة لـ $\mu_1 - \mu_3 = 0$

$$\widetilde{T}_2 = \frac{(10-14)^2}{9{,}12 . \frac{2}{6}} = 5{,}26$$

بالنسبة لـ $\mu_1 - \mu_4 = 0$ نجد: $\widetilde{T}_3 = 1{,}32$

بالنسبة لـ $\mu_1 - \mu_5 = 0$ نجد: $\widetilde{T}_4 = 11{,}84$

بالنسبة لـ $\mu_2 - \mu_3 = 0$ نجد: $\widetilde{T}_5 = 16{,}12$

بالنسبة لـ $\mu_2 - \mu_4 = 0$ نجد: $\widetilde{T}_6 = 0{,}33$

بالنسبة لـ $\mu_2 - \mu_5 = 0$ نجد: $\widetilde{T}_7 = 26{,}65$

بالنسبة لـ $\mu_3 - \mu_4 = 0$ نجد: $\widetilde{T}_8 = 11{,}84$

بالنسبة لـ $\mu_3 - \mu_5 = 0$ نجد: $\widetilde{T}_9 = 1{,}32$

بالنسبة لـ $\mu_4 - \mu_5 = 0$ نجد: $\widetilde{T}_{10} = 21{,}05$

4 . نحسب المقدار:

$$T^2 = (k-1).F_{(1-\alpha,v_1,v_2)}$$
$$T^2 = (5-1).F_{(0,95,4,25)}$$
$$T^2 = 4 \cdot 2,76 = 11,04$$

5 . نقارن $\widetilde{T}$ مع T^2 فنجد أن $\widetilde{T} < T^2$ عند الفرضيات التالية:

$$H_0 = \mu_1 - \mu_2 = 0$$
$$H_0 = \mu_1 - \mu_3 = 0$$
$$H_0 = \mu_1 - \mu_4 = 0$$
$$H_0 = \mu_2 - \mu_4 = 0$$
$$H_0 = \mu_2 - \mu_5 = 0$$

وبالتالي لانرفض أياً منها بالإضافة إلى أن المتوسطات المقارنة مثنى مثنى متساوية.

أما بالنسبة للفرضيات الباقية نلاحظ أن $\widetilde{T} > T^2$ وبالتالي نرفض H_0 عندها وأن المتوسطات الباقية مثنى مثنى غير متساوية وبتطبيق العلاقة (9−29) نجد:

بالنسبة لـ $\mu_1 - \mu_5 = 0$:

$$\frac{\bar{x}_1 - \bar{x}_5}{s^2.\sqrt{\frac{2}{n}}} = \frac{10-16}{3,02 \cdot \sqrt{\frac{2}{6}}} = -3,44 < -T$$

أي: $-3,44 < -3,32$

وبالتالي $\mu_1 < \mu_5$

بالنسبة لـ $\mu_2 - \mu_3 = 0$:

$$\frac{\bar{x}_2 - \bar{x}_3}{1,74} = \frac{7-14}{1,74} = -4,02 < -T$$

وبالتالي $\mu_2 < \mu_3$

بالنسبة لـ $\mu_2 - \mu_5 = 0$:

$$\frac{\bar{x}_2 - \bar{x}_5}{1,74} = \frac{7-16}{1,74} = -5,17 < -T$$

وبالتالي $\mu_2 < \mu_5$

بالنسبة لـ $\mu_3 - \mu_4 = 0$:

$$\frac{\bar{x}_3 - \bar{x}_4}{1,74} = \frac{14-8}{1,74} = 3,45 > 3,32$$

وبالتالي فإن:

$$\mu_3 > \mu_4$$

بالنسبة لـ $\mu_4 - \mu_5 = 0$:

$$\frac{\bar{x}_4 - \bar{x}_5}{1,74} = \frac{8-16}{1,74} = -4,6 < -T$$

وبالتالي فإن:

$$\mu_4 < \mu_5$$

9-4: شروط تطبيق تحليل التباين:

افترضنا في دراستنا لتحليل التباين تحقق بعض الفرضيات التي تتلخص يكون المجتمعات الإحصائية موزعة توزعاً طبيعياً ولها التباين نفسه، وأن العينات التي سحبناها من هذه المجتمعات هي عينات عشوائية بسيطة ومستقلة عن بعضها بعضاً. فإذا كانت هذه الشروط كافية، فإنها ليست ضرورية حكماً، على الرغم من أهميتها التطبيقية.

تجدر الإشارة مبدئياً إلى أن معادلة تحليل التباين التي تُظهر توزيع التباين يمكن استخدامها بهدف تحليلي دون أي قيد أو شرط ولاضرورة للشروط إلا عندما ننتقل إلى مجال التقدير والاستدلال والاختبار الإحصائي بالاعتماد على هذه المعادلة.

ثانياً، لاضرورة لتحقق الشروط جميعها معاً في آن واحد، فمثلاً، لاضرورة لكون المجتمعات الإحصائية موزعة توزعاً طبيعياً لتقدير الأوساط والتباينات، على الأقل إذا لم نلجأ إلى تحديد مجالات ثقة.

أما فيما يتعلق باختبار الفرضية، فإن عدن تحقق بعض الشروط قد يؤثر في النتيجة ولكن هذا التأثير يبقى ضعيفاً ودون أهمية قصوى، إلا في حالة كون المجتمعات الإحصائية متباينة جداً وبعيدة عن الطبيعة وخاصة إذا كانت العينات صغيرة الحجم جداً. وفي هذه الحالة، يفضل ألا نستخدم الاختبار التي تعتمد على تحليل التباين.

بالنسبة إلى شرط كون تباينات المجتمعات الإحصائية متساوية، فهذا الشرط ثانوي عندما تكون العينات ذات حجوم متساوية، ويعكس تطبيق اختبار F بكل ثقة. أما عندما تكون العينات ذات حجوم مختلفة، فإن عدم تحقق هذا الشرط له تأثير بالغ، خاصة إذا كانت العينات ذات الحجوم الصغيرة تتوافق والمجتمعات الإحصائية ذات التباينات الأعظمية.

في هذه الحالة الأخيرة، يوجد عدة طرق لمعالجة الوضع، منها إجراء تحويلات في المتحولات بغاية جعل التوزيعات أقرب إلى الطبيعية وتثبيت تبايناتها [7]. وإذا لم تجد هذه التحويلات نفعاً فإننا نلجأ إلى الاختبارات الحرة التي لاتخضع لأي شرط.

9-5: علاقة تطبيق التباين باختبار ستيودنت وبمعامل الارتباط:

في الحالة الخاصة لوجود مجتمعين فقط يمكننا أن نكتب:

$$SSB = n_1\left(\bar{x}_1 - \bar{x}\right)^2 + n_2\left(\bar{x}_2 - \bar{x}\right)^2$$

و

$$\bar{x} = \left(n_1 \cdot \bar{x}_1 + n_2 \cdot \bar{x}_2\right)^2 / \left(n_1 + n_2\right)$$

ومنه:

[7] من بعض هذه التحويلات: تحويل الجذر التربيعي . التحويل اللوغاريتمي . التحويل المثلث.

$$SS=\frac{n_1\left[(n_1+n_2)\bar{x}_1-\left(n_1\cdot\bar{x}_1+n_2\cdot\bar{x}_2\right)\right]^2+n_2\left[(n_1+n_2)\bar{x}_2-\left(n_1\cdot\bar{x}_1+n_2\cdot\bar{x}_2\right)\right]^2}{(n_1+n_2)^2}$$

$$SS=\frac{n_1n_2^2\left(\bar{x}_1-\bar{x}_2\right)^2+n_2n_1^2\left(\bar{x}_1-\bar{x}_2\right)^2}{(n_1+n_2)^2}=\frac{n_1n_2^2\left(\bar{x}_1-\bar{x}_2\right)^2}{n_1+n_2}$$

$$\frac{\frac{SSB}{SSE}}{n_1+n_2-2}=\frac{\frac{n_1n_2}{n_1n_2}\left(\bar{x}_1-\bar{x}_2\right)^2}{\frac{SSE_1+SSE_2}{n_1+n_2-2}}=\frac{\left(\bar{x}_1-\bar{x}_2\right)^2}{\frac{SSE_1+SSE_2}{n_1+n_2-2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

وبما أن SSE_1, SEE_2 يمثلان مجموع مربع الانحرافات:

$$\widetilde{F}=\frac{\left(\bar{x}_1-\bar{x}_2\right)^2}{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

وبمقارنة هذه القيمة مع قيمة t في اختبارات ستودينت نجد أن $\widetilde{F}=\widetilde{t}^{\,2}$ كما أنه لدينا $F_{1-a}=t^2_{1-\frac{a}{2}}$.

ينتج مما سبق أن تحليل التباين بالنسبة إلى مجتمعين يكافئ ستيودنت t، وهو من هذا المنطلق تعميم لاختبار ستيودنت t*.

تمرينات

1 . اختار أحد الباحثين في كلية الزراعة ثلاث عينات من الخراف من إحدى المزارع لتربية الخراف وبشكل عشوائي. ثم طبق على كل عينة نوع خاص من التغذية وبعد شهر وجد أن الزيادة الحاصلة في الوزن (بالكلغ) كانت كمايلي:

ترتيب الخروف	1	2	3	4	5	6	7	8	9	10
العينة A	5	7	4	3	4	5	6			
العينة B	3	4	4	3	5	4	6	5		
العينة C	6	7	4	5	8	7	4	5	8	6

والمطلوب:

معرفة فيما إذا كان هناك تأثير لنوع التغذية على الزيادة في وزن الخراف مستخدماً مستوى دلالة 0,01 ثم 0,05 .

2 . لاختبار تأثير أربعة أنواع من التكنولوجيا A,B,C,D على إنتاج القمح، فقد تم اختيار أربع قطع من الأرض متجانسة تماماً. ثم تم تطبيق كل نوع من التكنولوجيا على قطعة أرض واحدة ولمدة أربع سنوات فكانت النتائج مقدرة بالكيلوغرام/ الدونم كما يلي:

نوع التكنولوجيا / العام	A	B	C	D
1	200	315	270	250
2	250	320	290	260
3	225	290	310	270
4	240	280	330	240

والمطلوب: اختبار مدى تأثير التكنولوجيا المستخدمة على إنتاج القمح مستخدماً مستوى دلالة 0,05 .

3 . تم تطبيق ثلاثة نماذج من طرق تعليم مادة الرياضيات على ثلاث شعب من طلاب الصف السادس الابتدائي، حيث كل شعبة تضم مجموعة من الطلاب وبعد الامتحان كانت نتائجهم مبينة في الجدول التالي:

رقم الطالب	1	2	3	4	5	6	7	8	9
النموذج A	90	43	80	65	60	70	–	–	–
النموذج B	90	68	40	24	65	35	40	49	–
النموذج C	65	48	40	60	68	72	52	50	55

والمطلوب:

اختبار فيما إذا كان لنماذج طرق التعليم تأثير على معدلات الطلبة مستخدماً مستوى دلالة 0,05.

4 . تم تطبيق أربعة أنواع من أنظمة التغذية على أربعة مجموعات من المواليد في بداية الشهر الرابع، وعند بلوغهم الخمسة أشهر من العمر كانت الزيادة الحاصلة في أوزانهم كما يلي:

رقم المولود	1	2	3	4	5	6	7	8	9	10	11
نظام (1)	2,2	2,4	3	3,5	1,5	3,2	2,8	2,7	2,5	2,5	3
نظام (2)	2,3	2,7	2,5	3,4	3,9	3,8	3,7	2,9	2,9	2,4	2,4
نظام (3)	4,2	4,3	4,5	2,5	2,6	2,7	3	3	4,1	4,2	2,5
نظام (4)	2,5	2,7	2,8	2,9	3,5	3,7	3,8	3,2	3,4	3,5	2,4

والمطلوب:

اختبار فيما إذا كان هناك تأثير لنظام التغذية على الزيادة من الوزن مستخدماً مستوى الدلالة 0,05.

5 . قامت شعبة الامتحانات بإعداد جدول تتعلق بعلامات بعض الطلبة لعدة مقررات، وكانت النتائج كما يلي (العلامة من مئة).

	العينة الأولى	العينة الثانية	العينة الثالثة	العينة الرابعة
المقرر الأول	80	65	79	92
المقرر الثاني	60	74	53	71
المقرر الثالث	45	52	60	48
المقرر الرابع	52	72	82	42
المقرر الخامس	90	98	76	87

هل يمكننا أن نعد هذه العينات الأربع قد سحبت من مجتمعات ذات أوساط متماثلة إذا كان مستوى الدلالة 5%، و 1%.

2 . أراد أحد المزارعين أن يجرب نوعاً جديداً من الأسمدة وذلك للحكم إذا كان هذا السماد يزيد من إنتاجية الأرض أو لا. من أجل هذا قام هذا المزارع بزراعة 12 قطعة من الأرض الزراعية، مساحاتها متساوية، استخدم في 7 منها السماد الجديد، وفي 5 منها السماد القديم، وقد سجل إنتاجية هذه القطع معبراً عنها بنسب مئوية وفق التالي:

سماد قديم	سماد جديد
85	89
88	90
78	82
70	75
75	70
	92
	86

فهل هنالك فرق بين متوسط نسبة الإنتاجية لكل واحدة من هاتين

مستوى دلالة قدره 5%، لخمس فرق كرة قدم اعتماداً على ولنفترض أن النتائج التي حصلنا عليها في الاختبار هي (العلامة من 20).

الفريق الأول	الفريق الثاني	الفريق الثالث	الفريق الرابع
15	14	17	19
12	15	18	18
18	16	16	12
14	17	15	17
18	15	14	18
17	16	18	14

4 . رغب مدير الإعلان في إحدى دور النشر أن يحدد إذا ماكان فرق بين متوسط عدد المجلات المباعة لدى مجموعة من مراكز وذلك خلال عدة أسابيع، بالاعتماد على المعلومات التالية:

المركز الأول	المركز الثاني	المركز الثالث	المركز الرابع
20	52	36	16
16	44	40	20
32	56	36	32
16	36	40	20

1 . هيء جدول توزيع التباين.

2 . اختبر فرضية عدم وجود فرق بين متوسط عدد المجلات المباعة مدجموعة مراكز البيع ضمن مستوى دالة قدره 0.01.

5 . ادعى أحد مقتنيي منظفات الغسيل أن منتجه يزيل جميع بقع الشحوم والسيارات مهما كان نوع الماء. وللتأكد من صحة مركز حماية المستهلك إلى اختيار هذا المنظف في خمسة .

الفصل العاشر

تحليل التباين باتجاهين (التصنيف الثنائي)

Two - Way Analysis of Variance

ويشمل الفقرات التالية:

1 . مقدمة.

2 . تصميم تجربة تحليل التباين باتجاهين وبمشاهدة واحدة:

. صياغة النموذج العام.

. فرضية الاختبار.

. حساب التباينات.

. حساب نسبة التباين.

. تنظيم جدول تحليل التباين باتجاهين.

. اختبار الفرضية واتخاذ القرار.

3 . تمرينات.

الفصل العاشر

تحليل التباين باتجاهين (التصنيف الثنائي)

Two - Way Analysis of Variance

10-1: مقدمة Introduction:

يستخدم تحليل التباين باتجاهين لتحليل التجارب التي تكون مصممة وفقاً لعاملين اثنين ودراسة تأثيرها على متغير ما. ويتضمن كل عامل عدداً من الصفات أو الخواص. ويدعى أحد العاملين بمتحول الأعمدة وتصنف فيه المشاهدات في مجموعات. ويدعى العامل الآخر بمتحول الصفوف، ويتضمن عدداً من المستويات المختلفة. وتصنف الملاحظات المشاهدة حسب هذه المستويات، وهذا ما يسمى بقطاع تام العشوائية Completely Randomized Blocks .

فإذا كان لدينا عدة أصناف من بذور القطن وعدة أنواع من السماد، ونريد دراسة تأثير صنف البذور ونوعية السماد على إنتاج القطن. فعند تصميم مثل هذه التجربة يكون لدينا عدة تراكيب مؤلفة من صنف البذور ونوع السماد. وما يهمنا هو اختبار فيما إذا كان يوجد تركيب معين من صنف البذور ونوعية السماد بحيث يؤدي إلى متوسط إنتاج أفضل من بين جميع التراكيب الأخرى المحتملة.

مثال آخر، عندما نستخدم عدة طرق لعرض منتج ما في عدة مستويات من المناطق، ونرغب في دراسة تأثير طرق عرض المنتج ومستويات المناطق على بيع المنتج.

10-2: تصميم تجربة تحليل التباين باتجاهين وبمشاهدة واحدة (تصميم قطاع تام العشوائية Randomized Complete Block Design):

لقد ذكرنا أن تحليل التباين باتجاهين يستخدم لتحليل التجارب التي تدرس تغيرات أحد المؤشرات من جراء تأثير عاملين معينين (الأعمدة والصفوف). ونفترض في هذا الفصل أن تقاطع الصف والعمود يشكل خلية تتضمن مشاهدة واحدة. وأنه لا يوجد

تأثير مشترك بين الأعمدة والصفوف، أو بمعنى آخر، إن تأثير الأعمدة يكون نفسه لكل صف من الصفوف. فإذا كان متحول الأعمدة يتألف من j عموداً، ويتألف متحول الصفوف من i صفاً، فإننا نرمز بـ x_{ij} إلى قيمة المشاهدة في الصف i، والموافقة للعمود j، حيث:

$$i = 1\ ,\ 2\ ,\ 3\ ,\ .....\ ,\ n$$

$$j = 1\ ,\ 2\ ,\ 3\ ,\ ....\ ,\ k$$

كما نفترض أن المشاهدات x_{ij} اختيرت بشكل عشوائي من مجتمعات طبيعية تبايناتها متساوية وأوساطها الحسابية μ_{ij}.

وبالتالي فإن:

x_{11} هي المشاهدة الموافقة للصف الأول والعمود الأول.

x_{24} هي المشاهدة الموافقة للصف الثاني والعمود الرابع.

.

.

.

x_{n5} هي المشاهدة الموافقة للصف n والعمود الخامس.

.

.

.

x_{nk} هي المشاهدة الموافقة للصف n والعمود k.

نرمز بـ $T_{i.}$ إلى مجموع قياسات (المشاهدات) الصف i أي:

$$T_{i.} = \sum_{j=1}^{k} x_{ij} \qquad (1-10)$$

وتعبر النقطة عن جميع المشاهدات الموافقة للدليل الذي وضعت النقطة محله.

وبالتالي فإن:

$T_{1\bullet}$ ترمز إلى مجموع قياسات الصف الأول المأخوذة على الدليل j حيث $j = 1 , 2 , 3 , ... , k$.

$T_{2\bullet}$ ترمز إلى مجموع قياسات الصف الثاني المأخوذة على الدليل j حيث $j = 1 , 2 , 3 , ... , k$.

.

.

$T_{n\bullet}$ ترمز إلى مجموع قياسات الصف n المأخوذة على الدليل j حيث $j = 1 , 2 , 3 , ... , k$.

ونرمز بـ $\bar{x}_{i\bullet}$ إلى متوسط قياسات الصف i ويعطى هذا المتوسط بالعلاقة:

$$\bar{x}_{i.} = \frac{\sum_{j=1}^{k} x_{ij}}{k} = \frac{T_{i\bullet}}{k} \qquad (2-10)$$

وتشير النقطة هنا عن أن المتوسط تم أخذه بالنسبة لجميع المشاهدات الموافقة للدليل الذي وضعت النقطة محله، وبالتالي فإن:

$\bar{x}_{1\bullet}$ ترمز إلى متوسط قياسات الصف الأول المأخوذة على الدليل j حيث:

$j = 1 , 2 , 3 , ... , k$

$\bar{x}_{2\bullet}$ ترمز إلى متوسط قياسات الصف الثاني المأخوذة على الدليل j حيث:

$j = 1 , 2 , 3 , ... , k$

.....

$\bar{x}_{n\bullet}$ ترمز إلى متوسط قياسات الصف n المأخوذة على الدليل j حيث:

$j = 1 , 2 , 3 , ... , k$

ونرمز بـ $T.j$ إلى مجموع قياسات مشاهدات العمود j أي:

$$T_{\bullet j} = \sum_{i=1}^{n} x_{ij} \qquad (3-10)$$

وبالتالي فإن:

$T_{.1}$ ترمز إلى مجموع قياسات العمود الأول المأخوذة على الدليل i حيث:

$i = 1 \ , \ 2 \ , \ 3 \ , \ldots \ , \ n$

$T_{.2}$ ترمز إلى مجموع قياسات العمود الثاني المأخوذة على الدليل i حيث:

$i = 1 \ , \ 2 \ , \ 3 \ , \ldots \ , \ n$

.....

$T_{.k}$ ترمز إلى مجموع قياسات العمود k المأخوذة على الدليل i حيث:

$i = 1 \ , \ 2 \ , \ 3 \ , \ldots \ , \ n$

ونرمز بـ $\bar{x}_{\bullet j}$ إلى متوسط قياسات العمود j ويعطى هذا المتوسط بالعلاقة:

$$\bar{x}_{.j} = \frac{\sum_{i=1}^{n} x_{ij}}{n} = \frac{T_{\bullet j}}{n} \qquad (4-10)$$

وبالتالي فإن:

$\bar{x}_{\bullet 1}$ ترمز إلى متوسط قياسات العمود الأول المأخوذة على الدليل i .

$\bar{x}_{\bullet 2}$ ترمز إلى متوسط قياسات العمود الثاني المأخوذة على الدليل i .

.....

$\bar{x}_{\bullet k}$ ترمز إلى متوسط قياسات العمود k المأخوذة على الدليل i .

ونرمز بـ T إلى المجموع الكلي للقياسات ويساوي:

$$T = \sum_{j=1}^{k} \sum_{i=1}^{n} x_{ij} \qquad (5-10)$$

ولحساب متوسط القياسات لجميع قيم المشاهدات ونرمز له بـ $\overline{X}$ نجد أنه يساوي:

$$\overline{X} = \frac{\sum_{j=1}^{k} \sum_{i=1}^{n} x_{ij}}{n.k} \qquad (6-10)$$

أو:

$$\overline{X} = \frac{T}{n.k}$$

حيث:

$i = 1,2,3.....n$

$j = 1,2,3.....k$

وبالتالي يمكن توضيح ما تقدم من خلال الجدول التالي:

الجدول رقم (10– 1) تصميم تجربة تحليل التباين باتجاهين وبمشاهدة واحدة.

الأعمدة / الصفوف	1	2	3	……j……			k	$\sum_{j=1}^{k} x_{ij} = T_{i\bullet}$	$\bar{x}_{i\bullet} = \frac{T_{i\bullet}}{K}$
1	x_{11}	x_{12}	x_{13}	.....	x_{ij}	.....	x_{1k}	$T_{1\bullet}$	$\bar{x}_{1\bullet}$
2	x_{21}	x_{22}	x_{23}	.....	x_{2j}	.....	x_{2k}	$T_{2\bullet}$	$\bar{x}_{2\bullet}$
3	x_{31}	x_{32}	x_{33}	.....	x_{3j}	.....	x_{3k}	$T_{3\bullet}$	$\bar{x}_{3\bullet}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
i	x_{i1}	x_{i2}	x_{i3}	.....	x_{ij}	.....	x_{ik}	$T_{i\bullet}$	$\bar{x}_{i\bullet}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
n	x_{n1}	x_{n2}	x_{n3}	.....	x_{nj}	.....	x_{nk}	$T_{n\bullet}$	$\bar{x}_{n\bullet}$
$\sum_{i=1}^{n} x_{ij} = T_{\bullet j}$	$T_{\bullet 1}$	$T_{\bullet 2}$	$T_{\bullet 3}$	…	$T_{\bullet j}$	⋯	$T_{\bullet k}$	$T = \sum_{j=1}^{k} \sum_{i=1}^{n} x_{ij}$	–
$\frac{T_{\bullet j}}{n} = \bar{X}_j$	$\bar{x}_{\bullet 1}$	$\bar{x}_{\bullet 2}$	$\bar{x}_{\bullet 3}$	…	$\bar{x}_{\bullet j}$	⋯	$\bar{x}_{\bullet k}$		$\bar{X} = \frac{T}{n \cdot K}$

وبما أننا نرغب في معرفة فيما إذا كانت تأثيرات جميع الصفوف متساوية وكذلك تأثيرات جميع الأعمدة متساوية، إذاً لابد من القيام بما يلي:

10–2–1: صياغة النموذج العام:

لنفترض أن x_{ij} هي قيمة المشاهدة الموافقة للصف i والعمود j، وأن تأثيرات كل من الصفوف والأعمدة تجميعية (أي أنه لا يوجد تفاعل بين الصفوف والأعمدة). وبالتالي فإن متوسط الخلية الموافقة للصف i والعمود j ونرمز لها بـ μ_{ij} يمكن كتابتها بالشكل التالي:

$$\mu_{ij} = \mu + \beta_i + \alpha_j \qquad (7-10)$$

حيث:

α_j تشير إلى تأثير العمود j و $j = 1,2,3.....k$

β_i تشير إلى تأثير العمود i و $i = 1,2,3.....n$

وأن كل مشاهدة مثل x_{ij} يمكن كتابتها بالشكل:

$$x_{ij} = \mu_{ij} + \varepsilon_{ij} \qquad (8-10)$$

وبتعويض العلاقة (10- 7) في (10- 8) تصبح كل مشاهدة مكتوبة على الشكل التالي:

$$x_{ij} = \mu + \beta_i + \alpha_j + \varepsilon_{ij} \qquad (9-10)$$

حيث:

$i = 1,2,3.....n$

$j = 1,2,3.....k$

μ متوسط ثابت بالنسبة لجميع المشاهدات (الخلايا).

β_i – تشير إلى تأثير الصف i (القطاع i)، وتعني مقدار تأثير القطاع i في تعيين مقدار μ_{ij} وتختلف من صف إلى آخر. وأن:

$$\sum_{i=1}^{n} \beta_i = 0$$

α_j – تشير إلى تأثير العمود j (المعالجة j)، وتعني مقدار تأثير المعالجة j في تعيين مقدار μ_{ij} ، وأن:

$$\sum_{j=1}^{k} \alpha_j = 0$$

ε_{ij} – متحولات عشوائية مستقلة عن بعضها البعض وتتوزع وفق التوزيع الطبيعي الذي متوسطه صفر وبتباين موحد σ^2 . وتمثل جميع مصادر التباين التي لم تحسب.

10-2-2: فرضية الاختبار:

من خلال النموذج الذي تم عرضه فإننا نسطيع اختبار فرضية أن تأثيرات جميع الصفوف (القطاعات) متساوية أي أن: $\beta_i = 0$ وذلك بغض النظر فيما إذا كان يوجد تأثيرات للأعمدة (المعالجات). وكذلك نستطيع اختبار فرضية أن تأثيرات جميع الأعمدة متساوية أي أن $\alpha_j = 0$ وذلك بغض النظر فيما إذا كان يوجد تأثيرات للصفوف وذلك مع الافتراض بعدم وجود تفاعل بين الصفوف والأعمدة. وبالتالي فإن β_i, α_j تصبح ثوابت ولذلك يدعى تحليل التباين باتجاهين بـ (نموذج التأثيرات الثابتة Fixed Effects model) ويمكن صياغة فرضيات الاختبار بالشكل التالي:

فرضية الحالة الأولى:

الفرضية الابتدائية: $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = ... \alpha_k = 0$

أو باختصار على الشكل التالي: $H_0 : \alpha_j = 0$

(تأثيرات المعالجات تساوي الصفر) $j = 1, 2, 3..... k$

الفرضية البديلة:

(توجد واحدة على الأقل من α_j لاتساوي الصفر): $H_1 : \alpha_j \neq 0$

$j = 1, 2, 3..... k$

ويتم اختبار هذه الفرضية بشكل مستقل عن تأثيرات الصفوف.

فرضية الحالة الثانية:

الفرضية الابتدائية: $H_0 : \beta_1 = \beta_2 = \beta_3 = ... \beta_k = 0$

(تأثيرات القطاعات تساوي الصفر) $H_0 : \beta_i = 0$

$i = 1, 2, 3..... n$

الفرضية البديلة:

(توجد واحدة على الأقل من β_i لاتساوي الصفر): $H_1 : \beta_i \neq 0$

حيث: $i = 1, 2, 3..... n$

ويتم اختبار هذه الفرضية بشكل مستقل عن تأثيرات الأعمدة (المعالجات).

10-2-3: حساب التباينات:

لحساب تباينات تحليل التباين باتجاهينوالتي تعزى لمصادرها المختلفة ننطلق من التباين الكلي total Variance وهو عبارة عن مجموع مربعات انحرافات القيم المشاهدة التي تتضمنها التجربة Experiment عن وسطها الحسابي العام $\overline{X}$ ونرمز للتباين الكلي بـ SST ويعرف بالعلاقة:

$$SST = \sum_{j=1}^{k}\sum_{i=1}^{n} (x_{ij} - \overline{X})^2 \qquad (10-10)$$

يمكن كتابة الطرف الأيمن من العلاقة (10- 10) بالشكل التالي:

$$SST = \sum_{j=1}^{k}\sum_{i=1}^{n} \left(\overline{x}_{i\bullet} - \overline{X}\right)^2 + \sum_{j=1}^{k}\sum_{i=1}^{n} \left(\overline{x}_{\bullet j} - \overline{X}\right)^2 +$$

$$+ \sum_{j=1}^{k}\sum_{i=1}^{n} \left[\left(x_{ij} - \left(\overline{x}_{i\bullet} - \overline{X}\right) - \left(\overline{x}_{\bullet j} - \overline{X}\right) - \overline{X}\right)\right]^2$$

وإن $\overline{x}, \overline{x}_{\bullet j}, \overline{x}_{i\bullet}$ معرفة بالعلاقات (10- 2) و(10- 4) و(10- 6) على الترتيب.

$$SST = K\sum_{i=1}^{n} \left(\overline{x}_{i\bullet} - \overline{X}\right)^2 + n\sum_{j=1}^{k} \left(\overline{x}_{\bullet j} - \overline{X}\right)^2 +$$

$$+ \sum_{j=1}^{k}\sum_{i=1}^{n} \left[x_{ij} - \overline{x}_{i\bullet} - \overline{x}_{\bullet j} + \overline{X}\right]^2 \qquad (11-10)$$

ومن خلال العلاقة (10- 11) فإن التباين الكلي يتألف من ثلاث مركبات هي:

المركبة الأولى: وتعزى للصفوف (القطاعات) Blocks ونرمز لها بـ SSB وتقيس التباين مابين القطاعات ويعبر عنه بمجموع المربعات للصفوف Sum of squares for rows ويعطى بالعلاقة:

$$SSB = K\sum_{i=1}^{n} \left(\overline{x}_{i\bullet} - \overline{X}\right)^2 \qquad (12-10)$$

المركبة الثانية: وتعزى للأعمدة (المعالجات) Treatments ونرمز لها بـ SSC وتقيس التباين ما بين المجموعات (المعالجات) ويعبر عنه بمجموع المربعات للأعمدة Sum of squares for columns ويعطى بالعلاقة:

$$SSC = n\sum_{j=1}^{k}\left(\bar{x}_{.j} - \bar{\bar{X}}\right)^2 \qquad (13-10)$$

المركبة الثالثة: وتعزى للخطأ Error المرتكب ونرمز لها بـ SSE وتمثل التباين الناجم عن الخطأ ويعبر عنها بمجموع المربعات للخطأ Sum of squares for Error ويعطى بالعلاقة التالية:

$$SSE = \sum_{j=1}^{k}\sum_{i=1}^{n}\left[x_{ij} - \bar{x}_{i\bullet} - \bar{x}_{\bullet j} + \bar{\bar{X}}\right]^2 \qquad (14-10)$$

وبالتالي فإن العلاقة (10−11) يمكن التعبير عنها بالشكل التالي:

$$SST = SSB + SSC + SSE \qquad (15-10)$$

ويمكن كتابة العلاقات (10−10) و(10−12) و(10−13) بأشكال تسهل عملية حسابها باستخدام الآلات الحاسبة وتصبح على الشكل التالي:

$$SST = \sum_{j=1}^{k}\sum_{i=1}^{n} x_{ij}^2 - \frac{T}{n.k}^2 \qquad (16-10)$$

ودرجات الحرية له $nk-1$ درجة حرية.

$$SSB = \frac{1}{k}\sum_{i=1}^{n} T_{i\bullet}^2 - \frac{T}{n.k}^2 \qquad (17-10)$$

ودرجات الحرية له $n-1$ درجة حرية.

$$SSC = \frac{1}{n}\sum_{j=1}^{k} T_{\bullet j}^2 - \frac{T}{n.k}^2 \qquad (18-10)$$

ودرجات الحرية له $(k-1)$ درجة حرية.

ويدعى المقدار $\frac{T}{n.k}^2$ بحد التصحيح ويرمز له بـ C أي:

$$C = \frac{T}{n.k}^{2} \qquad (19-10)$$

وباستخدام العلاقة (10−15) نحصل على SSE بعد أن نقوم بحساب SST و SSB و SSC أي:

$$SSE = SST - SSB - SSC \qquad (20-10)$$

ودرجات الحرية له $(n-1)(k-1)$ درجة حرية.

ومن المفروض أن تتساوى درجات الحرية لطرفي العلاقة (10−15) ويمكن التأكد من ذلك بالمطابقة بين درجات الحرية أي:

$$SST = SSB + SSC + SSE$$
$$nk - 1 = (n-1) + (k-1) + (n-1)(k-1) \qquad (21-10)$$
$$nk - 1 = nk - 1$$

10−2−4: حساب نسبة التباين:

لحساب نسبة التباين لابد في البداية من حساب:

ـ متوسط مربعات الأعمدة (المعالجات) Mean of squares for columns ونرمز له بـ MSC ويعطى بالعلاقة:

$$MSC = \frac{SSC}{k-1}$$

ـ متوسط مربعات الصفوف (القطاعات) Mean of squares for blocks ونرمز له بـ MSB ويعطى بالعلاقة:

$$MSB = \frac{SSB}{n-1}$$

ـ متوسط مربعات الخطأ Mean of squares for Error ونرمز له بـ MSE ويعطى بالعلاقة:

$$MSE = \frac{SSE}{(n-1)(k-1)}$$

ومن أجل اختبار فرضية الحالة الأولى نحسب نسبة متوسط مربعات الأعمدة (المعالجات) إلى متوسط مربعات الخطأ أي:

$$\widetilde{F}_1 = \frac{SSC}{MSE}$$

وهذه النسبة تخضع لتوزيع F بدرجتي حرية $V_1 = (k-1)$ و $V_2 = (k-1)(n-1)$

ومن أجل اختبار فرضية الحالة الثانية نحسب نسبة متوسط مربعات الصفوف (القطاعات) إلى متوسط مربعات الخطأ أي:

$$\widetilde{F}_2 = \frac{MSB}{MSE}$$

وهذه النسبة تخضع لتوزيع F بدرجتي حرية $V_1 = (n-1)$ و $V_2 = (k-1)(n-1)$

10-2-5: تنظيم جدول تحليل التباين باتجاهينThe ANOVOA table:

يمكن تشكيل جدول يتضمن الحسابات السابقة التي تم إجراؤها وفق الشكل التالي:

الجدول رقم (10- 2) تحليل التباين باتجاهين بمشاهدة واحدة

مصدر التباين (التغير)	مجموع المربعات	درجات الحرية	متوسط المربعات	F نسبة التباين (التغير)	قيمة F الجدولية
مابين المجموعات (المعالجات)	SSC	$k-1$	$MSC = \frac{SSC}{k-1}$	$\widetilde{F}_1 = \frac{MSC}{MSE}$	$F_{1-\alpha}, V_1, V_2$
ما بين الصفوف (القطاعات)	SSB	$n-1$	$MSB = \frac{SSB}{n-1}$	$\widetilde{F}_2 = \frac{MSB}{MSE}$	$F_{1-\alpha}, n-1$
الخطأ	SSE	$(n-1)(k-1)$	$MSE = \frac{SSE}{(n-1)(k-1)}$		
المجموع	SST	$nk-1$			

10-2-6: اختبار الفرضية واتخاذ القرار الإحصائي:

لاختبار الفرضية الأولى والقائلة بأن تأثير جميع الأعمدة تساوي الصفر؛ أي:

$$H_0 : \alpha_j = 0 \qquad j = 1,2,3.....k$$

لابد من تحديد النقطة الحرجة عند مستوى دلالة α؛ أي: $F_{(1-\alpha,v_1,v_2)}$

فإذا كانت: $\widetilde{F}_1 = \dfrac{MSC}{MSE} < F_{1-\alpha(v_1,v_2)}$

فإننا نقبل H_0 ونقول إن تأثير جميع الأعمدة (المعالجات) تساوي الصفر.

وإذا كانت: $\widetilde{F}_1 = \dfrac{MSC}{MSE} > F_{(1-\alpha,v_1,v_2)}$

فإننا نرفض H_0 ونقبل الفرضية البديلة H_1 ونقول إنه يوجد واحدة من α_j على الأقل لاتساوي الصفر.

لاختبار الفرضية الثانية والقائلة بأن تأثير جميع الصفوف (القطاعات) يساوي الصفر أي:

$$H_0 = \beta_i = 0 \qquad i = 1,2,3.....n$$

لابد من تحديد النقطة الحرجة عند مستوى دلالة α أي: $F_{(1-\alpha,v_1,v_2)}$

فإذا كانت: $\widetilde{F}_2 = \dfrac{MSB}{MSE} < F_{(1-\alpha,v_1,v_2)}$

فإننا نقبل H_0 ونقول إن تأثير جميع الصفوف (القطاعات) تساوي الصفر.

وإذا كانت: $\widetilde{F}_2 = \dfrac{MSB}{MSE} > F_{(1-\alpha,v_1,v_2)}$

فإننا نرفض H_0 ونقبل الفرضية البديلة H_1 ونقول أنه يوجد واحدة من β_j على الأقل لا تساوي الصفر.

مثال 1:

تم تطبيق أربعة أنواع من طرق التعليم على طلاب الصف الثاني الابتدائي في ثلاثة أنواع من المدارس وفي اختبار نهاية العام كانت درجات الطلاب كما يلي:

طرق التعليم / المدارس	1	2	3	4
A	8	7	5	9
B	4	9	6	7
C	8	4	7	6

والمطلوب:

1 . اختبار فيما إذا كانت تأثيرات جميع طرق التعليم متساوية.

2 . اختبار فيما إذا كانت تأثيرات أنواع المدارس متساوية، وذلك عند مستوى دلالة 0,05 .

الحل:

1 . يمكن صياغة النموذج العام بالشكل التالي:

$$x_{ij} = \mu + \beta_i + \alpha_j + \varepsilon_{ij}$$

حيث:

$$i = 1,2,3$$

$$j = 1,2,3,4$$

2 . فرضيات الاختبار: يمكن صياغة الفرضيات حسب الحالتين السابقتين كما يلي:

أ . الفرضية الابتدائية: $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$

أي أن تأثيرات جميع طرق التعليم تساوي الصفر.

. الفرضية البديلة: $H_1 : \alpha_j \neq 0$

يوجد تأثير لطريقة واحدة على الأقل لاتساوي الصفر

ب . الفرضية الابتدائية: $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$

أي أن تأثيرات جميع أنواع المدارس تساوي الصفر.

. الفرضية البديلة: $H_1 : \beta_i \neq 0$

يوجد تأثير نوع واحد على الأقل من المدارس لاتساوي الصفر

3 . حساب التباينات:

في البداية نقوم بتشكيل الجدول المساعد التالي:

طرق التعلم / أنواع المدارس	1	2	3	4	$\sum_{j=1}^{k} x_{ij} = T_{i\bullet}$	$\bar{x}_{i\bullet} = \frac{T_{i\bullet}}{k}$
A	8	7	5	9	29	7,25
B	4	9	6	7	26	6,5
C	8	4	7	6	25	6,25
$\sum_{i=1}^{n} x_{ij} = T_{\bullet j}$	20	20	18	22	$T = \sum_{j=1}^{4}\sum_{i=1}^{3} x_{ij} = 80$	
$\bar{x}_{\bullet j} = \frac{T_{\bullet j}}{n}$	6,67	6,67	6	7,33		$\overline{X} = \frac{\mathrm{T}}{n.k} = 6,67$

ثم نقوم بحساب مجاميع المربعات:

. مجموع المربعات الكلي (التباين الكلي):

$$SST = \sum_{j=1}^{4}\sum_{i=1}^{3} x_{ij}^2 - \frac{T^2}{n.k}$$

$$SST = \left[(8)^2 + (7)^2 + (5)^2 + \ldots\ldots\ldots + (7)^2 + (6)^2\right] - \frac{(80)^2}{3 \cdot 4}$$

$$SST = 566 - 533,33 = 32,67$$

. مجموع مربعات الأعمدة (المعالجات):

$$SSC = \frac{1}{n}\sum_{j=1}^{k} T_{\bullet j}^2 - \frac{T^2}{n.k}$$

$$SSC = \frac{1}{3}\left[(20)^2 + (20)^2 + (18)^2 + (22)^2\right] - \frac{(80)^2}{3 \cdot 4}$$

$$SSC = 536 - 533,33 = 2,67$$

. مجموع مربعات الصفوف (القطاعات):

$$SSB = \frac{1}{k}\sum_{i=1}^{n} T_{i\bullet}^2 - \frac{T^2}{n.k}$$

$$SSB = \frac{1}{4}\left[(29)^2 + (26)^2 + (25)^2\right] - \frac{(80)^2}{3 \cdot 4}$$

$$SSB = 535{,}5 - 533{,}33 = 2{,}17$$

. مجموع مربعات الخطأ:

$$SSE = SST - SSC - SSB$$

$$SSE = 32{,}67 - 2{,}67 - 2{,}17$$

$$SSE = 27{,}83$$

وللتأكد من أن درجات الحرية متطابقة بين طرفي العلاقة نكتب:

$$SST = SSC + SSB + SSE$$

$$nk - 1 = (k-1) + (n-1) + (n-1)(k-1)$$

$$3 \cdot 4 - 1 = (4-1) + (3-1) + (3-1)(4-1)$$

$$11 = 3 + 2 + 6$$

نلاحظ أن درجات الحرية متساوية بين طرفي العلاقة.

4 . حساب نسبة التباين:

في البداية نقوم بحساب متوسطات المربعات:

. متوسطات مربعات الأعمدة (المعالجات) MSC:

$$MSC = \frac{SSC}{k-1} = \frac{2{,}67}{4-1} = 0{,}89$$

. متوسطات مربعات الصفوف (القطاعات) MSB:

$$MSB = \frac{SSB}{n-1} = \frac{2{,}17}{3-1} = 1{,}085$$

. متوسطات مربعات الخطأ MSE:

$$MSE = \frac{SSE}{(n-1)(k-1)} = \frac{27,83}{(3-1)(4-1)} = 4,64$$

ثم نقوم بحساب F_1 نسبة متوسط مربعات الأعمدة (المعالجات) إلى متوسط مربعات الخطأ أي:

$$\widetilde{F}_1 = \frac{MSC}{MSE} = \frac{0,89}{4,64} = 0,192$$

نحسب النقطة الحرجة عند مستوى دلالة $\alpha = 0,05$ ودرجات حرية $V_1 = (K-1) = 3$ و $V_2 = (k-1)(n-1) = 6$

فنجد أن:

$$F_{(1-\alpha, v_1, v_2)} = F_{(0,95\ ,3,\ 6)} = 4,76$$

ونحسب F_2 نسبة متوسط مربعات الصفوف (القطاعات) إلى متوسط مربعات الخطأ أي:

$$\widetilde{F}_2 = \frac{MSB}{MSE} = \frac{1,085}{4,64} = 0,234$$

نحسب النقطة الحرجة عندما $\alpha = 0,05$ ودرجات حرية $V_1 = (n-1) = 2$ و $V_2 = (k-1)(n-1) = 6$

فنجد أن:

$$F_{(1-\alpha, v_1, v_2)} = F_{(0,95\ ,2,\ 6)} = 5,14$$

5 . تنظيم جدول تحليل التباين:

نشكل جدول تحليل التباين ونضع فيه الحسابات السابقة على الشكل التالي:

مصدر التباين (التغير)	مجموع المربعات	درجات الحرية	متوسط المربعات	نسبة التباين $\widetilde{F}$
مابين المجموعات (الأعمدة) (المعالجات)	SSC	$k-1=3$	$MSC=\frac{SSC}{k-1}$ $=0,89$	$\widetilde{F_1}=\frac{MSC}{MSE}$ $=0,192$
مابين الصفوف (القطاعات)	SSB	$n-1=2$	$MSB=\frac{SSB}{n-1}$ $=1,085$	$\widetilde{F_2}=\frac{MSB}{MSE}$ $=0,234$
الخطأ	SSE	$(n-1)(k-1)=6$	$MSE=\frac{SSE}{(n-1)(k-1)}$ $=4,64$	
المجموع	SST	$nk-1=11$		

6 . اختبار الفرضية واتخاذ القرار الإحصائي:

بالنسبة للفرضية الأولى نلاحظ أنه بمقارنة $\widetilde{F_1}=0,192$ مع $F_{(1-\alpha,v_1,v_2)}=4,76$ نجد أن:

$$\widetilde{F_1}=F_{(1-\alpha,v_1,v_2)}$$

وبالتالي فإننا نقبل الفرضية الابتدائية H_0 ونقول إن تأثيرات جميع الأعمدة تساوي الصفر، أي:

$$\alpha_1=\alpha_2=\alpha_3=\alpha_4=0$$

بالنسبة للفرضية الثانية نلاحظ أنه بمقارنة: $\widetilde{F_2}=0,234$ مع $F_{(1-\alpha,v_1,v_2)}=5,14$ نجد أن:

$$\widetilde{F_2}<F_{(1-\alpha,v_1,v_2)}$$

وبالتالي فإننا نقبل الفرضية الابتدائية H_0 ونقول إن تأثيرات جميع الأعمدة تساوي الصفر، أي:

$$\beta_1=\beta_2=\beta_3=0$$

مثال 2:

تم استخدام أربعة أنواع من طرق العرض لمنتج معين في خمسة أنواع من المخازن التجارية فكانت قيم إجمالي المبيعات الشهرية (بالآلاف) كما في الجدول التالي:

طرق العرض / أنواع المخازن	A_1	A_2	A_3	A_4
B_1	50	51	55	59
B_2	59	61	60	65
B_3	55	56	60	67
B_4	69	70	66	77
B_5	70	72	70	78

والمطلوب:

1 . اختبار فيما إذا كان يوجد فروق بين تأثير طرق العرض على إجمالي المبيعات.

2 . اختبار فيما إذا كان يوجد فروق بين تأثير الأنواع المختلفة من المخازن على إجمالي المبيعات.

وذلك ضمن مستوى دلالة 0,05 ، ثم بمستوى دلالة 0,01.

الحل:

1 . يمكن صياغة النموذج لهذه التجربة بالشكل التالي:

$$x_{ij} = \mu + \beta_i + \alpha_j + \varepsilon_{ij}$$

حيث:

$$i = 1,2,3,4,5$$

$$j = 1,2,3,4$$

2 . فرضيات الاختبار:

أ . الفرضية الابتدائية: $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$

أو $H_0 : \alpha_j = 0$

لا توجد فروق بين تأثير الأنواع المختلفة من طرق العرض.

. الفرضية البديلة:

توجد طريقة واحدة على الأقل لا تساوي الصفر $H_1 : \alpha_j \neq 0$

ب . الفرضية الابتدائية: $H_0 : \beta_i = 0$

$i = 1,2,3,4,5$

لا توجد فروق بين تأثيرات الأنواع المختلفة من المخازن.

. الفرضية البديلة: $H_1 : \beta_j \neq 0$

$i = 1,2,3,4,5$

يوجد تأثير لنوع واحد على الأقل من أنواع المخازن لايساوي الصفر.

3 . حساب التباينات:

نشكل الجدول المساعد التالي:

طرق العرض / أنواع المخازن	A_1	A_2	A_3	A_4	$= T_{i\bullet}$	$\bar{x}_{i\bullet} = \frac{T_{i\bullet}}{k}$
B_1	50	51	55	59	215	53,75
B_2	59	61	60	65	245	61,25
B_3	55	56	60	67	238	59,5
B_4	69	70	66	77	282	70,5
B_5	70	72	70	78	290	72,5
$T_{.j}$	303	310	311	346	$T = 1270$	
$\bar{x}_{\bullet j} = \frac{T_{\bullet j}}{n}$	60,6	62	62,2	69,2		$\bar{X} = 63,5$

. التباين الكلي:

$$SST = \sum_{j=1}^{4}\sum_{i=1}^{5} x_{ij} - \frac{T^2}{n.k}$$

$$SST = \left[(50)^2 + (59)^2 + (55)^2 + ... + (77)^2 + (78)^2\right] - \frac{(1270)^2}{5 \cdot 4}$$

$$SST = 81898 - 80645 = 1253$$

مجموع مربعات الأعمدة (المعالجات):

$$SSC = \frac{1}{n}\sum_{j=1}^{k} T_{\bullet j}^2 - \frac{T^2}{n.k}$$

$$SSC = \frac{1}{5}\left[(303)^2 + (310)^2 + (311)^2 + (346)^2\right] - \frac{(1270)^2}{5 \cdot 4}$$

$$SSC = 80869{,}2 - 80645 = 224{,}2$$

مجموع مربعات الصفوف (القطاعات):

$$SSB = \frac{1}{k}\sum_{i=1}^{n} T_{i\bullet}^2 - \frac{T^2}{n.k}$$

$$SSB = \frac{1}{4}\left[(215)^2 + (245)^2 + (238)^2 + (282)^2 + (290)^2\right] - \frac{(1270)^2}{5 \cdot 4}$$

$$SSB = 81629{,}5 - 80645 = 984{,}5$$

مجموع مربعات الخطأ:

$$SSE = SST - SSC - SSB$$

$$SSE = 1253 - 224{,}2 - 984{,}5$$

$$SSE = 44{,}3$$

وللتأكد من تطابق درجات الحرية بين طرفي العلاقة نكتب:

$$SST = SSC + SSB + SSE$$

$$nk - 1 = (k-1) + (n-1) + (n-1)(k-1)$$

$$5 \cdot 4 - 1 = (4-1) + (5-1) + (5-1)(4-1)$$

$$19 = 3 + 4 + 12$$

4 . نسبة التباين:

. متوسطات مربعات الأعمدة (المعالجات) MSC:

$$MSC = \frac{SSC}{k-1} = \frac{224,2}{3} = 74,73$$

. متوسطات مربعات الصفوف (القطاعات) MSB:

$$MSB = \frac{SSB}{n-1} = \frac{984,5}{4} = 246,125$$

. متوسطات مربعات الخطأ MSE:

$$MSE = \frac{SSE}{(n-1)(k-1)} = \frac{44,3}{4 \cdot 3} = 3,69$$

نسبة متوسط مربعات الأعمدة (المعالجات) إلى متوسط مربعات الخطأ أي:

$$\widetilde{F}_1 = \frac{MSC}{MSE} = \frac{74,73}{3,69} = 20,25$$

نحسب النقطة الحرجة عند مستوى دلالة $\alpha = 0,05$ ودرجات حرية $V_1 = (k-1) = 3$ و $V_2 = (n-1)(k-1) = 12$

ومنه:

$$F_{(1-\alpha, v_1, v_2)} = F_{(0,95\ ,3,\ 12)} = 3,49$$

نسبة متوسط مربعات الصفوف (القطاعات) إلى متوسط مربعات الخطأ أي:

$$\widetilde{F}_2 = \frac{MSB}{MSE} = \frac{246,125}{3,69} = 66,7$$

النقطة الحرجة عند $\alpha = 0,05$ ودرجات حرية $V_1 = (n-1) = 4$ و $V_2 = (n-1)(k-1) = 12$، ومنه نجد أن:

$$F_{(1-\alpha, v_1, v_2)} = F_{(0,95\ ,4,\ 12)} = 3,26$$

5 . جدول تحليل التباين:

نضع النتائج السابقة في جدول تحليل التباين التالي:

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	نسبة التباين $\widetilde{F}$	قيمة F الجدولية
مابين المجموعات (الأعمدة) (المعالجات)	SSC	$V_1 = k-1 = 3$	$MSC = \frac{SSC}{k-1} = 74,73$	$\widetilde{F}_1 = \frac{MSC}{MSE} = 20,25$	3,49
مابين الصفوف (القطاعات)	SSB	$V_2 = n-1 = 4$	$MSB = \frac{SSB}{n-1} = 246,125$	$\widetilde{F}_2 = \frac{MSB}{MSE} = 66,7$	3,26
الخطأ	SSE	$(n-1)(k-1) = 12$	$MSE = \frac{SSE}{(n-1)(k-1)} = 3,69$		
المجموع	SST	$nk-1 = 19$			

6 . اختبار الفرضية واتخاذ القرار الإحصائي:

بالنسبة للفرضية الأولى نلاحظ أنه بمقارنة: $\widetilde{F}_1 = 20,25$ مع $F_{(0,95\ ,3\ ,\ 12)} = 3,49$، نجد أن:

$$\widetilde{F}_1 > F_{(0,95\ ,3\ ,\ 12)}$$

وبالتالي فإننا نرفض الفرضية الابتدائية H_0 ونقبل الفرضية البديلة H_1 ونقول أنه توجد فروقات بين الأنواع المختلفة من طرق عرض المنتج على مستوى دلالة $0,05$.

بالنسبة للفرضية الثانية نلاحظ أنه بمقارنة: $\widetilde{F}_2 = 66,7$ مع $F_{(0,95\ ,4\ ,\ 12)} = 3,26$، نجد أن:

$$\widetilde{F}_2 > F_{(0,95\ ,4\ ,\ 12)}$$

وبالتالي فإننا نرفض الفرضية الابتدائية H_0 ونقبل الفرضية البديلة H_1 ونقول أنه توجد فروقات بين الأنواع المختلفة من المخازن على مستوى دلالة $0,05$.

* اختبار الفرضيات السابقة على مستوى دلالة $\alpha = 0,01$

بالنسبة للفرضية الأولى:

في البداية نحسب النقطة الحرجة F عند مستوى دلالة $\alpha = 0{,}01$ وبدرجات حرية $V_1 = (k-1) = 3$ و $V_2 = (n-1)(k-1) = 12$

ومنه فإن:

$$F_{(\alpha,v_1,v_2)} = F_{(0{,}01\ ,3,\ 12)} = 5{,}95$$

وبمقارنة $\widetilde{F}_1 = 20{,}25$ مع $F(0{,}01{,}3{,}12)$ نجد أن:

$$\widetilde{F}_1 > F_{(0{,}01\ ,3\ ,\ 12)}$$

وبالتالي فإننا نرفض الفرضية الابتدائية H_0 ونقبل الفرضية البديلة H_1 .

بالنسبة للفرضية الثانية :

نحسب النقطة الحرجة عند مستوى دلالة $\alpha = 0{,}01$ وبدرجات حرية $V_1 = (n-1) = 4$ و $V_2 = (n-1)(k-1) = 12$

ومنه نجد أن:

$$F_{(\alpha,v_1,v_2)} = F_{(0{,}01\ ,4,\ 12)} = 5{,}41$$

وبمقارنة $\widetilde{F}_1 = 20{,}25$ مع $F(0{,}01{,}3{,}12)$ نجد أن:

$$\widetilde{F}_2 > F_{(0{,}01\ ,4\ ,\ 12)}$$

وبالتالي فإننا نرفض الفرضية الابتدائية H_0 ونقبل الفرضية البديلة H_1 .

نلاحظ أننا رفضنا الفرضية الابتدائية لكلتا الفرضيتين الأولى والثانية عند مستوى دلالة 0,05 و 0,01 وقبلنا الفرضية البديلة H_1 لهما.

تمرينات

1 . تم استخدام ثلاثة أنواع من الأسمدة على خمسة أنواع من الذرة الصفراء تمت زراعتها في 15 قطعة أرض ذات خصوبة متساوية واختيرت بشكل عشوائي لمعرفة تأثير نوع السماد المستخدم وتأثير نوع الذرة الصفراء ، فكان الإنتاج مقدراً بالكيلوغرام وفق الجدول التالي:

نوع الذرة الصفراء / نوع السماد	A_1	A_2	A_3	A_4	A_5
B_1	60	55	50	45	70
B_2	65	50	65	75	65
B_3	50	48	70	60	60

المطلوب:

1 . اختبار فيما إذا كان يوجد فروق بين تأثير الأنواع المختلفة من الذرة الصفراء.

2 . اختبار فيما إذا كان يوجد فروق بين تأثير الأنواع المختلفة من السماد . وذلك على مستوى دلالة 0,05.

2 . لنفرض أنه تم إجراء اختبار لمعرفة تأثير جنس الشخص (ذكور . إناث) وتأثير الشريحة العمرية على متوسط عدد الحوادث في العام.

ومن أجل ذلك حصلنا على البيانات اللازمة والموافقة لثمانية شرائح عمرية والمبينة في الجدول التالي:

الجنس / الشرائح العمرية	ذكور	إناث
20-25	7	6
25-30	6	5
30-35	6	5
35-40	5	4
40-45	6	5
45-50	6	6
50-55	7	8
55-60	8	9

والمطلوب:

1 . اختبار فيما إذا كان يوجد أي تأثير لجنس الشخص على متوسط عدد الحوادث في العام.

2 . اختبار فيما إذا كان يوجد تأثير للشريحة العمرية على متوسط عدد الحوادث في العام.

استعمل مستوى دلالة 0,05 ومن ثم مستوى دلالة 0,01.

3 . من أجل دراسة اختبار تأثير أنواع العلف وتأثير أنواع الأبقار على متوسط الزيادة في الوزن تم استخدام أربعة أنواع من الأعلاف وثلاثة أنواع من الأبقار، حيث تم اختيار بقرة واحدة من كل نوع وبشكل عشوائي لنوع واحد من العلف.

وبعد مرور شهر حصلنا على متوسط الزيادة في الوزن بالكيلوغرام. لكل متوافقة من المتوافقات والمبينة في الجدول التالي:

أنواع الأبقار / أنواع الأعلاف	K	A	R
C	20	22	28
O	30	29	26
R	25	23	30
M	28	30	22

والمطلوب:

اختبار فيما إذا كان يوجد تأثير أي لنوع العلف ونوع الأبقار على متوسط الزيادة في الوزن بمستوى دلالة 0,05.

4. قامت مجموعة من الطلاب بإجراء تجربة لمعرفة فيما إذا كان يوجد تأثير للمستوى التعليمي للأسرة وتأثير مكان الإقامة على متوسط عدد الأطفال ومن أجل ذلك تم جمع المعلومات اللازمة ووضعت في الجدول التالي:

المستوى التعليمي / مكان الإقامة	إعدادية	ثانوية	جامعية	عليا
ريف	6	5	4	2
أحياء شعبية	5	5	4	1
مدينة	4	4	3	1

والمطلوب:

اختبار فيما إذا كان يوجد أي تأثير للمستوى التعليمي أو لمكان الإقامة على متوسط عدد أفراد الأسرة مستخدماً مستوى دلالة 0,05.

مثال 5:

استخدم بيانات المثال رقم (3) ثم احسب معامل التحديد ومعامل عدم التحديد.

الحل:

معامل التحديد يعطى بالعلاقة :

$$R^2 = \frac{SSR}{SST} = \frac{11,89}{13,68} = 0,87$$

أي أن معامل التحديد يفسر *87%* من التباين الكلي، وأن خط الانحدار يمثل توفيقاً جيداً لشكل الانتشار.

أما معامل عدم التحديد فيساوي:

$$1 - R^2 = 1 - 0,0,13$$

أي أن معامل عدم التحديد *13%* يشير إلى الجزء العشوائي من التباين الكلي الذي لم يتم تفسيره.

مثال 6:

إذا كان لدينا بيانات عن أرباح ومبيعات إحدى شركات القطاع الخاص لإنتاج الغسالات الأوتوماتيكية معطاة بالجدول التالي خلال عشرة أشهر.

الأشهر	1	2	3	4	5	6	7	8	9	10
الأرباح Y (مائة ألف ليرة سورية)	4	5	7	5	6	9	8	10	9	12
المبيعات X(مائة ألف ليرة سورية)	13	11	15	12	12	15	17	20	18	23

والمطلوب:

1 . أوجد معادلة خط الانحدار Y على X.

2 . إذا كانت $X = 25$ و $X = 30$ ، فما هي الأرباح المقدرة $\hat{Y}$.

3 . اختبر الفرضية $H_0 : \beta = 0$ عند مستوى دلالة 0,05 ثم عند 0,01.

4 . اختبر الفرضية $H_0 : \beta = 0$ عند مستوى دلالة 0,05 باستخدام إحصاء t.

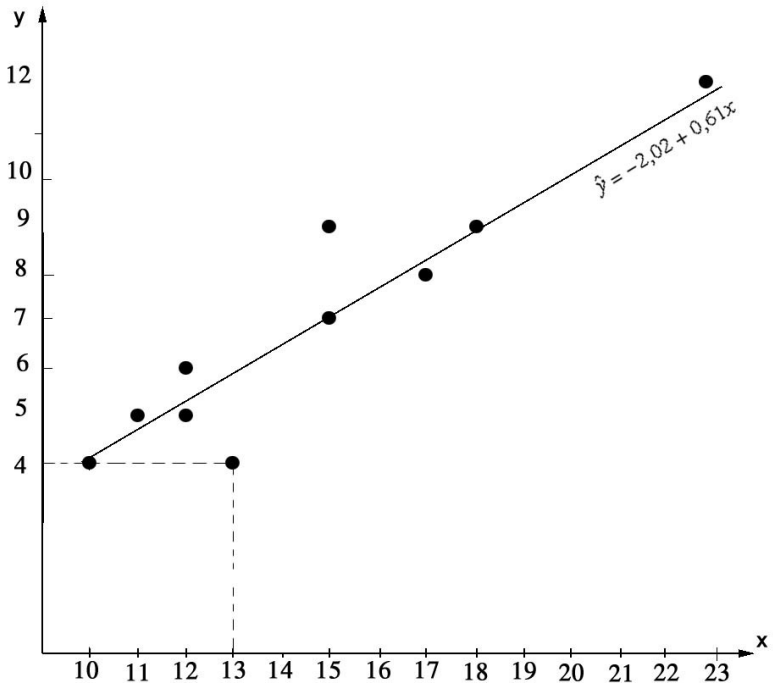
5 . اختبر الفرضية $H_0 : \beta = 0,8$ عند مستوى دلالة 0,05 .

6 . إيجاد مجال الثقة لمعامل الانحدار الحقيقي β ، عندما $\alpha = 0{,}05$.

7 . إيجاد معامل التحديد R^2 .

الحل:

1 . لإيجاد المعادلة نرسم شكل الانتشار فنجد أنه يأخذ الشكل التالي:



شكل رقم (11–11) شكل الانتشار وخط الانحدار

نلاحظ أن شكل الانتشار يوحي بأنه لمعادلة من الدرجة الأولى من الشكل:

$$\widetilde{Y} = a + bX$$

لإيجاد ثوابت هذه المعادلة نستخدم المعادلات الطبيعية التالية:

$$\sum_{i=1}^{n} y_i = na + b\sum_{i=1}^{n} x_i$$

$$\sum_{i=1}^{n} y_i x_i = a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2$$

لإيجاد حدود هاتين المعادلتين نشكل الجدول المساعد التالي:

i	y_i	x_i	$y_i x_i$	x_i^2	y_i^2	$\hat{y}_i$
1	4	13	52	169	16	5,91
2	5	11	55	121	25	4,69
3	7	15	105	225	49	7,13
4	5	12	60	144	25	5,3
5	6	12	72	144	36	5,3
6	9	15	135	225	81	7,13
7	8	17	136	289	64	8,35
8	10	20	200	400	100	10,18
9	9	18	162	324	81	8,96
10	12	23	276	529	144	12,01
	75	156	1253	2570	621	75

نعوض في المعادلات الطبيعية:

$$75 = 10a + 156b$$

$$1253 = 156a + 2570b$$

بحل هاتين المعادلتين حلاً مشتركاً نجد قيمة الثابتين a, b. أو نحسبهما مباشرة من العلاقتين:

$$a = \bar{y} - b\bar{x}$$

$$b = \frac{n\sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n} y_i \cdot \sum_{i=1}^{n} x_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

حيث:

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{75}{10} = 7,5$$

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{156}{10} = 15,6$$

ومنها نجد أن:

$$b = \frac{10 \cdot 1253 - 75 \cdot 156}{10 \cdot 2570 - (156)^2} \approx 0,61$$

$$a = 7,5 - 0,61 \cdot 15,6 \approx -2,02$$

وبالتالي فإن معادلة خط الانحدار تأخذ الشكل التالي:

$$\hat{y} = -2,02 + 0,61x$$

ويمكن رسم هذه المعادلة على نفس شكل الانتشار السابق.

للحصول على قيمة $\hat{y}$ المقدرة عند كل قيمة من قيم X نعوض في معادلة خط الانحدار، أي:

$$\hat{y}_1 = -2,02 + 0,61(13) = 5,91$$

$$\hat{y}_2 = -2,02 + 0,61(11) = 4,69$$

.....

$$\hat{y}_{10} = -2,02 + 0,61(23) = 12,01$$

ويمكن وضع النتائج في العمود الأخير من الجدول المساعد.

2 . وللحصول على قيمة $\hat{y}$ المقدرة عندما $x = 25$,$x = 30$ نعوض في معادلة خط الانحدار المقدرة فنجد:

$$x = 25 \Rightarrow \hat{y} = -2,02 + 0,61(25) = 13,23$$

$$x = 30 \Rightarrow \hat{y} = -2,02 + 0,61(30) = 16,28$$

3 . لاختبار الفرضية $H_0 : \beta = 0$ مستوى دلالة $\alpha = 0,05$ ثم عند $\alpha = 0,01$ مقابل الفرضية $H_1 : \beta \neq 0$.

ونقوم بإجراء الخطوات التالية:

*** حساب مجاميع المربعات للانحرافات الآتية:**

مجموع المربعات الكلي SST:

$$SST = \sum_{i=1}^{n} y_i^{\,2} - \frac{\left(\sum y_i\right)^2}{n}$$

$$SST = 621 - \frac{(75)^2}{10} = 58,5$$

. مجموع المربعات للانحدار SSR ويساوي:

$$SSR = b^2\left[\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right]$$

$$SSR = (0{,}61)^2\left[2570 - 10\cdot(15{,}6)^2\right] = 50{,}75$$

مجموع المربعات للأخطاء SSE ويساوي:

$$SSE = SST - SSR$$

$$SSE = 58{,}5 - 50{,}75 = 7{,}75$$

*** حساب متوسط المربعات:**

متوسط المربعات للانحدار MS_x :

$$MSx = \frac{SSR}{1} = \frac{50{,}75}{1} = 50{,}75$$

متوسط المربعات للخطأ $S_{y\cdot x}^2$:

$$S_{y\cdot x}^2 = \frac{SSE}{n-2} = \frac{7{,}75}{8} = 0{,}97$$

ومنه الخطأ المعياري للتقدير:

$$S_{y\cdot x} = \sqrt{0{,}97} \approx 0{,}99$$

نلاحظ أن الخطأ المعياري للتقدير صغير، وبالتالي فإن القيم الفعلية قريبة من خط الانحدار.

***حساب النسبة $\widetilde{F}$:**

$$\widetilde{F} = \frac{MS_x}{S_{y\cdot x}^2} = \frac{50{,}75}{0{,}97} = 52{,}32$$

ونضع النتائج السابقة في جدول تحليل التباين التالي:

جدول تحليل التباين

مصدر التغير	مجموع المربعات	درجات الحرية	متوسط المربعات	النسبة $\widetilde{F}$	قيمة $F_{1-\alpha}$ الجدولية
بسبب الانحدار	SSR=50,75	1	$MSx = 50,75$	$\widetilde{F} = 52,32$	5,32
بسبب الأخطاء	SSE=7,75	n - 2 = 8	$S^2_{yx} = 0,97$		
المجموع	SST=58,5	n - 1 = 9			

*** إحصاء الاختبار** F:

نقارن قيمة $\widetilde{F}$ المسحوبة مع قيمة F الجدولية عند مستوى دلالة $\alpha = 0,05$ و $V_1 = 1$ و $V_2 = n - 2 = 8$ فنجد:

$$F_{[0,05,1,8]} = 5,32$$

نلاحظ أن: $\widetilde{F} > F_{[0,05,1,8]}$ ؛ لأن: $52,32 > 5,32$

لذلك نرفض الفرضية الابتدائية H_0 ونقبل الفرضية البديلة H_1 ، ونستنتج بأنه توجد علاقة جيدة للانحدار بين X و Y.

اختبار الفرضيتين: $H_0 : \beta = 0$ و $H_1 : \beta \neq 0$ عند مستوى دلالة 0,01.

نبحث عن قيمة F الجدولية عندما $\alpha = 0,01$ و بدرجات حرية $V_1 = 1, V_2 = 8$ فنجد أن: $F_{[0,01,1,8]} = 11,26$

وهنا نلاحظ أن: $\widetilde{F} > F_{[0,01,1,8]}$

لذلك نرفض الفرضية الابتدائية H_0 ونقبل الفرضية البديلة H_1

4 . اختبار الفرضية $H_0 : \beta = 0$ باستخدام إحصاء t.

مقابل الفرضية $H_1 : \beta \neq 0$

نستخدم التوزيع التالي:

$$\widetilde{t} = \frac{b - \beta}{S_b}$$

حيث:

$$S_b = \sqrt{\frac{S_{y \cdot x}^2}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}}$$

$$S_b = \sqrt{\frac{0,97}{2570 - 10 \cdot (15,6)^2}} = 0,084$$

بالتعويض نجد:

$$\widetilde{t} = \frac{b - \beta}{S_b} = \frac{0,61 - 0}{0,084} = 7,26$$

نبحث عن قيمة t الجدولية عند مستوى دلالة 0,05 وبدرجة حرية $n - 2 = 10 - 2 = 8$.

ونلاحظ أولاً أن: $\alpha = 0,05 \Rightarrow \frac{\alpha}{2} = 0,025 \Rightarrow 1 - \frac{\alpha}{2} = 0,975$

ومن الجدول IV لتوزيع ستودينت نجد أن:

$$t_{(0,975,8)} = 2,306$$

وبمقارنة $\widetilde{t}$ المحسوبة مع t الجدولية فنجد أن:

$$\widetilde{t} > t_{(0,975,8)}$$

لذلك نرفض الفرضية الابتدائية H_0 ونقبل الفرضية البديلة H_1، ونستنتج أن معامل الانحدار الحقيقي للمجتمع الإحصائي يختلف عن الصفر. ويمكن الاعتماد على قيم المتغير X في التنبؤ بقيم المتغير Y.

5 . اختبار الفرضية: $H_0 : \beta = 0,8$ عند مستوى دلالة 0,05.

باعتبار أن معامل الانحدار الحقيقي يختلف عن الصفر إذاً لنختبر الفرضية التالية:

الفرضية الابتدائية: $H_0 : \beta = 0,8$

الفرضية البديلة: $H_1 : \beta \neq 0,8$

نستخدم إحصاء t:

$$\tilde{t} = \frac{b - \beta_0}{S_b}$$

$$\tilde{t} = \frac{0,61 - 0,8}{0,084} = -2,26$$

وبمقارنة t الجدولية مع $\tilde{t}$ نجد أن:

$$|\tilde{t}| < t_{(0,975,8)}$$

$$|2,26| < 2,306$$

وبالتالي نقبل الفرضية H_0 ونقول إن معامل الانحدار الحقيقي للمجتمع الإحصائي يمكن أن يأخذ القيمة 0,8.

6 . إيجاد مجال الثقة لمعامل الانحدار:

يعطى مجال الثقة بالعلاقة:

$$I_\beta = P\left[b - t_{\left(1-\frac{\alpha}{2},n-2\right)} \cdot S_b \leq \beta \leq b + t_{\left(1-\frac{\alpha}{2},n-2\right)} \cdot S_b \right] = 1 - \alpha$$

$$I_\beta = P[0,61 - 2,306 \cdot 0,084 \leq \beta \leq 0,61 + 2,306 \cdot 0,084] = 0,95$$

$$[0,416 \leq \beta \leq 0,804] = 0,95$$

نلاحظ أن مجال الثقة يتضمن القيمة السابقة $\beta_0 = 0,8$ ، وأنها قد تكون قيمة لمعامل انحدار للعلاقة بين X و Y. وإن مجال الثقة لايتضمن القيمة صفر، وبالتالي فإن الصفر لايمكن أن يكون قيمة مقدرة لـ β.

7 . معامل التحديد:

يعطى معامل التحديد بالعلاقة:

$$R^2 = \frac{SSR}{SST} = \frac{50,75}{58,5} \approx 0,87$$

أي أن معامل التحديد يفسر 87% من التباين الكلي، ويمثل خط الانحدار توفيقاً جيداً لشكل الانتشار. أما معامل عدم التحديد فيساوي:

$$1 - R^2 = 1 - 0{,}87 = 0{,}13$$

ويشير معامل عدم التحديد الذي يساوي 13% إلى الجزء العشوائي من التباين الكلي الذي لم يتم تفسيره.

11-4: معامل الارتباط الخطي Coefficient Of Linear Correlation:

لقد بحثنا في موضوع إيجاد العلاقة الرياضية التي تربط بين المتغيرين X و Y (معادلة خط الانحدار)، ومن خلال هذه العلاقة نستطيع أن نتنبأ بقيم المتغير Y عند معرفتنا بقيم المتغير X.

وسوف نقوم في هذه الفقرة بدراسة الارتباط بين المتغيرين. ويقصد بالارتباط دراسة العلاقة بين قيم المتغير X وقيم المتغير Y من خلال مقاييس عددية لقياس قوة هذا الارتباط، وتدعى هذه المقاييس بمعامل الارتباط. ومن أشهر مقاييس الارتباط الخطي، معامل ارتباط بيرسون Pearson. ويستخدم هذا المعامل لقياس قوة العلاقة الخطية مابين المتغيرين X و Y، وذلك عندما تكون تراتيب قيم X وتراتيب قيم Y عبارة عن مشاهدات رقمية أو كمية. بالإضافة إلى أنه يقيس قوة واتجاه العلاقة بين المتغيرين. فعندما يزداد المتغير Y نتيجة زيادة ما في المتغير X فنقول إن العلاقة بين المتغيرين هي علاقة طردية، وعندها فإن إشارة معامل الارتباط هي إشارة موجبة. وكمثال على ذلك نأخذ المتغير X الذي يعبر عن ظاهرة المبيعات و المتغير Y الذي يعبر عن ظاهرة الأرباح، فنلاحظ أنه كلما زادت المبيعات كلما زادت الأرباح، وبالتالي فإن العلاقة بين ظاهرتي المبيعات والأرباح هي علاقة طردية.

ولكن عندما يزداد المتغير X ويقابل هذه الزيادة انخفاضاً في المتغير Y أو بالعكس، فنقول إن العلاقة بين المتغيرين هي علاقة عكسية، وإن إشارة معامل الارتباط

هي إشارة سالبة. ومثل ذلك، مثل العلاقة بين ظاهرتي الأرباح والمصاريف. فكلما ارتفعت المصاريف، كلما انخفضت الأرباح.

ويجب الإشارة إلى توخي الحذر عند دراسة الارتباط بين المتغيرات الاقتصادية، وذلك عندما تكون العلاقة بينهما غير سببية. فمثلاً قد نحصل على قيمة لمعامل الارتباط عند دراسة العلاقة بين أعداد الخريجين من الجامعة وأعداد الوفيات. هنا قد نلاحظ علاقة بين الظاهرتين، ولكن هذه العلاقة غير صحيحة لأنها حصلت نتيجة تأثرها بمتغير خارجي. وعندما يستبعد أثر المتغير الخارجي فإن العلاقة بين المتغيرين تصبح صفراً أي أن معامل الارتباط يساوي الصفر.

ويجب الإشارة إلى أن قيمة معامل الارتباط والذي نرمز له بـ r تتراوح ما بين: $-1 \le r \le +1$.

عندما:

$r = 1$ هذا يعني أن الارتباط بين المتغيرين X و Y تاماً، وفي اتجاه طردي.

$r = -1$ هذا يعني أن الارتباط بين المتغيرين X و Y تاماً، وفي اتجاه عكسي.

$r = 0$ هذا يعني أن الارتباط بين المتغيرين X و Y معدوم (لاتوجد علاقة بين X و Y)

وتزداد قوة الارتباط كلما اقتربت قيمة معامل الارتباط من الواحد سواءً كانت العلاقة طردية أو عكسية، وتضعف هذه العلاقة كلما اقتربت قيمة معامل الارتباط من الصفر.

وأخيراً، يجب أن ننوه إلى أن الجذر التربيعي لمعامل التحديد للنموذج الخطي يعطي معامل الارتباط أي: $\sqrt{r^2} = \pm r$ ، بالإضافة إلى أن الإشارة الجبرية لمعامل الارتباط تكون مساوية للإشارة الجبرية لمعامل الانحدار b.

وعند دراسة قوة الارتباط بين متغيرين X و Y يجب أن نفرق بين نوعين من البيانات:

. بيانات غير مبوبة.

. بيانات مبوبة.

11-4-1- معامل الارتباط الخطي في حالة البيانات غير المبوبة:

إذا كان لدينا n مشاهدة من الثنائيات لقيم Y, X والتي يمكن كتابتها بالشكل:

$$X: x_1 \quad x_2 \quad x_3 \quad ..... \quad x_i \quad ..... \quad x_n$$
$$Y: y_1 \quad y_2 \quad y_3 \quad ..... \quad y_i \quad ..... \quad y_n \qquad (41-11)$$

لقد قام العالم بيرسون Pearson بتعريف معامل للارتباط الخطي، وهو الذي يقيس قوة الارتباط بين المتغيرين X, Y ، ويحدد طبيعة التغير سواءً بالزيادة أو بالنقصان، يعرف بأنه مجموع حاصل ضرب القيم المعيارية للمتغيرين المراد دراسة العلاقة بينهما مقسوماً على n-1 ، ويعطى بالعلاقة التالية:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(n-1)Sx \cdot Sy} \qquad (42-11)$$

حيث:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$S_x = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

المتوسط الحسابي لـ y :

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$$

الانحراف المعياري لـ y:

$$S_y = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}}$$

ولصعوبة إجراء الحسابات في العلاقة (11–42) يمكن أن نستخدم صيغة أخرى لمعامل الارتباط تسهل إجراء الحسابات باستخدام الآلات الحاسبة وتعطى هذه الصيغة بالعلاقة التالية:

$$r = \frac{n\sum_{i=1}^{n} x_i \cdot y_i - \left(\sum_{i=1}^{n} x_i\right) \cdot \left(\sum_{i=1}^{n} y_i\right)}{\sqrt{n\sum_{i=1}^{n} x_i^2 - \left(\sum x_i\right)^2} \cdot \sqrt{n\sum_{i=1}^{n} y_i^2 - \left(\sum y_i\right)^2}} \qquad (43-11)$$

أو بواسطة العلاقة الآتية:

$$r = \frac{\sum_{i=1}^{n} x_i \cdot y_i - \bar{x} \cdot \bar{y} \cdot n}{\sqrt{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2} \cdot \sqrt{n\sum_{i=1}^{n} y_i^2 - n\bar{y}^2}} \qquad (44-11)$$

مثال 7:

استخدم بيانات المثال رقم (6)، ثم احسب معامل الارتباط، وفسر النتيجة.

الحل:

لحساب معامل الارتباط نستخدم العلاقة (11–43)، ولأن قيم المقادير فيها مأخوذة من الجدول المساعد في حل المثال رقم (6) أي:

$$\sum_{i=1}^{n} x_i \cdot y_i = 1253 \qquad \sum_{i=1}^{n} x_i = 156$$

$$\sum_{i=1}^{n} x_i^2 = 2570 \qquad \sum_{i=1}^{n} y_i = 75$$

$$\sum_{i=1}^{n} y_i^2 = 621$$

وبالتعويض في العلاقة (11–43) نجد:

$$r = \frac{10 \cdot 1253 - 156 \cdot 75}{\sqrt{10 \cdot 2570 - (156)^2} \cdot \sqrt{10 \cdot 621 - (75)^2}} \approx 0{,}93$$

أي أن الارتباط قوي جداً بين الأرباح والمبيعات، وإن العلاقة بينهما هي علاقة طردية، فكلما زادت المبيعات، زادت الأرباح. وإن إشارة معامل الارتباط هي من إشارة معامل الانحدار.

11-4-2: معامل الارتباط الخطي في حالة البيانات المبوبة:

إذا كان لدينا متغيران X, Y وكانت قيم كل منهما تقابل عدداً من التكرارات، فإن معامل الارتباط الخطي لبيرسون Pearson هو معامل الارتباط الخطي نفسه في حالة البيانات غير المبوبة ولكن بعد أن نقوم بتثقيل قيم كل من X و Y بتكراراتها الموافقة لها. وبالتالي فإن معامل الارتباط الخطي في حالة البيانات المبوبة يعطى بالعلاقة التالية:

$$r = \frac{n\sum x' \cdot y' \cdot n_{x \cdot y} - \left(\sum x' \cdot n_x\right)\left(\sum y' \cdot n_y\right)}{\sqrt{n\sum x'^2 \cdot n_x - \left(\sum x' \cdot n_x\right)^2} \cdot \sqrt{n\sum y'^2 \cdot n_y - \left(\sum y' \cdot n_y\right)^2}} \quad (45-11)$$

حيث:

n مجموع التكرارات الكلية.

n_{xy} التكرارات المشتركة والمقابلة لقيم للثنائية (x , y).

n_y التكرارات المقابلة لقيم Y.

n_x التكرارات المقابلة لقيم X.

y'_i مراكز فئات المتغير Y.

x'_i مراكز فئات المتغير X.

مثال 8:

أوجد معامل الارتباط الخطي بين الأجور وكمية الإنتاجلمجموعة مكونة من 50 عاملاً إذا علمت أن فئات الأجور وفئات كمية الإنتاج لمجموعة العمال كانت كما في الجدول التالي:

فئات الأجور (ألف ليرة سورية) / كمية الإنتاج	[2–4[	[4–6[	[6–8[	[8–10[	[10–12]	مجموع التكرارات
[2–6[	2	3	4	–	–	9
[6–10[	3	4	5	4	–	16
[10–14[	–	2	3	6	1	12
[14–18 [	–	–	1	3	5	9
[18–22[	–	–	–	1	3	4
مجموع التكرارات	5	9	13	14	9	50

الحل:

لإيجاد معامل الارتباط في هذه الحالة نستخدم العلاقة (11–45):

$$r=\frac{n\sum x'\cdot y'\cdot n_{xy}-\left(\sum x'\cdot n_x\right)\left(\sum y'\cdot n_y\right)}{\sqrt{n\sum x_i'^2\cdot n_x-\left(\sum x'\cdot n_x\right)^2}\cdot\sqrt{n\sum y_i'^2\cdot n_y-\left(\sum y'\cdot n_y\right)^2}}$$

ولإيجاد مقادير حدود العلاقة نشكل الجدول المساعد التالي:

y' / x'	3	5	7	9	11	n_x	$x_i'n_x$	$x'^2{}_in_x$	$x'\cdot y'\cdot n_{yx}$
4	2 / 24	3 / 60	4 / 112			9	36	144	196
8	3 / 72	4 / 160	5 / 280	4 / 288		16	128	1024	800
12		2 / 120	3 / 252	6 / 648	1 / 132	12	144	1728	1152
16			1 / 112	3 / 432	5 / 880	9	144	2304	1424
20				1 / 180	3 / 660	4	80	1600	840
n_y	5	9	13	14	9	50	532	6800	4412
$y_i'n_y$	15	45	91	126	99	376			
$y'^2\cdot n_y$	45	225	637	1134	1089	3130			
$x'\cdot y'\cdot n_{xy}$	96	340	756	1548	1672	4412			

وبالتعويض في العلاقة السابقة نجد أن:

$$r = \frac{50 \cdot 4412 - 532 \cdot 376}{\sqrt{50 \cdot 6800 - (532)^2} \cdot \sqrt{50 \cdot 3130 - (376)^2}} = 0{,}29$$

وهذا يعني أن الارتباط ضعيف جداً بين الأجور وعدد الوحدات المنتجة.

5-11: اختبار دلالة معامل الارتباط ρ :

لقد تم حساب معامل الارتباط r الذي يقيس قوة واتجاه العلاقة بين المتغيرين X و Y وذلك بالاعتماد على بيانات عينة بحجم n من المشاهدات أخذت من المجتمع الإحصائي الذي يخضع للتوزيع الطبيعي.

لنفترض أن معامل الارتباط r الذي تم حسابه بناءً على بيانات العينة هو تقدير لمعامل الارتباط ρ في المجتمع الإحصائي الذي يدل على وجود علاقة خطية بين المتغيرين X و Y. فإذا أخذنا جميع العينات الممكنة ذات الحجم n من المجتمع الإحصائي، فإن كل عينة تعطينا قيمة لمعامل الارتباط r وبالتالي فإننا نحصل على عدد كبير من القيم لـ r بعدد العينات التي تم تشكيلها ويمكن أن يعتبر كل منها تقديراً لمعامل ارتباط المجتمع ρ. وإذا افترضنا أن: $\rho = 0$ فإن التوزيع:

$$\tilde{t} = \frac{r}{\sqrt{(1 - r^2)/(n-2)}} \qquad (46-11)$$

يخضع لتوزيع ستودينت بـ $n-2$ درجة حرية.

وحتى نتبين أن r يشير إلى وجود علاقة خطية بين المتغيرين X و Y يجب أن نختبر الفرضية الابتدائية: $H_0 : \rho = 0$ مقابل الفرضية البديلة: $H_1 : \rho \neq 0$ عند مستوى دلالة α.

ولاختبار هذه الفرضية نقارن $\tilde{t}$ مع t الجدولية فإذا كانت $\tilde{t} > t_{\left(1-\frac{\alpha}{2}, n-2\right)}$ فإننا نرفض الفرضية H_0 ونقبل الفرضية H_1 ونستنتج أنه توجد علاقة بين المتغيرين X و Y في المجتمع عند مستوى دلالة α.

مثال 9:

لنعود إلى المثال رقم (7) ونختبر فيما إذا كانت قيمة r تدل على وجود علاقة ارتباط بين الأرباح والمبيعات عند مستوى دلالة $0,05$.

الحل:

. نكتب الفرضيات:

الفرضية الابتدائية: $H_0 : \rho = 0$

لا توجد علاقة بين X و Y.

الفرضية البديلة: $H_1 : \rho \neq 0$

توجد علاقة بين X و Y.

. نحسب قيمة $\tilde{t}$ فنجد أن:

$$\tilde{t} = \frac{r}{\sqrt{\left(1 - r^2\right) / (n-2)}}$$

$$\tilde{t} = \frac{0,93}{\sqrt{\left[1 - (0,93)^2\right] / (10-2)}} = 7,157$$

. نبحث عن قيمة t الجدولية فنجد أن:

$$t_{\left(1-\frac{\alpha}{2}, n-2\right)} = t_{(0,975\ ,\ 8)} = 2,306$$

ـ بالمقارنة نجد أن $\tilde{t} > t_{\left(1-\frac{\alpha}{2}, n-2\right)}$ ولذلك نرفض الفرضية H_0 ونستنتج أنه توجد علاقة بين الأرباح والمبيعات في المجتمع عند مستوى دلالة $\alpha = 0,05$.

11-6: حالات خاصة لمعامل الارتباط:

11-6-1: معامل الارتباط للرتب سبيرمان: Spearman's coefficient of Rank Correlation:

يستخدم هذا المعامل عندما تكون لدينا البيانات وصفية ومرتبة بين المتغيرين X و Y مثل تقدير معدلات الطلاب ومراتب الموظفين حسب السلم الوظيفي ...وهكذا.

ويتم حساب معامل الارتباط بين تراتيب المشاهدات لـ X و Y تصاعدياً أو تنازلياً حسب قيمها. ويرمز لمعامل الارتباط للرتب بـ r_s ويعطى بالعلاقة التالية:

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d^2}{n\left(n^2 - 1\right)} \qquad (47-11)$$

حيث أن:

n عدد المشاهدات.

d الفرق بين رتبتي المتغيرين المتقابلين.

ويستخدم هذا المعامل عندما تكون البيانات أقل من 30 مشاهدة. ويمكن إيجاد قيمة معامل الارتباط للرتب باتباع الخطوات التالية:

1 . نعطي رتباً لقيم كل من المتغيرين X و Y تصاعدياً أو تنازلياً، ففي الحالتين تأخذ القيمة الأولى الرتبة (1) وتأخذ القيمة الثانية الرتبة (2)... وهكذا.

2 . إذا تساوت رتبتان فإننا نأخذ متوسط الرتبتين ونضعه رتبةً لكلٍ منهما.

3 . نوجد الفرق بين كل رتبتين متقابلتين لـ X و Y ثم نربّع هذا الفرق.

4 . نطبق العلاقة (11−47) فنحصل على معامل الارتباط للرتب r_s.

مثال 10:

أوجد معامل الارتباط للرتب بين الدخل والإنفاق لعينة مكونة من 10 أسر، إذا توافرت عن هذه العينة المعلومات التالية: (القيم بمئات الليرات السورية).

الدخل X	7	9	11	11	12	14	19	22	22	25
الإنفاق Y	7	8	11	9	9	12	18	19	18	22

الحل:

لإيجاد معامل الارتباط للرتب نشكل الجدول التالي حسب الخطوات السابقة:

X	Y	رتب X	رتب Y	d	d^2
7	7	1	1	0	0
9	8	2	2	0	0
11	11	3,5	5	-1,5	2,25
11	9	3,5	3,5	0	0
12	9	5	3,5	1,5	2,25
14	12	6	6	0	0
19	18	7	7,5	-0,5	0,205
22	19	8,5	9	-0,5	0,205
22	18	8,5	7,5	1	1
25	22	10	10	0	0
Σ	–	–	–	–	9

نطبق العلاقة:

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d^2}{n\left(n^2 - 1\right)}$$

$$r_s = 1 - \frac{6 \cdot 6}{10\left[(10)^2 - 1\right]} \approx 0{,}96$$

وهو ارتباط طردي وقوي جداً بين الدخل والإنفاق.

11-6-2: معامل الاقتران Coefficient of Association:

يستخدم هذا المعامل لإيجاد الارتباط بين المتغيرين X و Y وذلك عندما تكون لدينا البيانات وصفية عن المتغيرين ولا يمكن وضع رتب لهما، بحيث أن كل متغير مقسم إلى مجموعتين، أو أن لكل متغير صفتان. مثلاً، بالنسبة للتدخين فإن الشخص (يدخن، لا يدخن)، وبالنسبة للزواج (متزوج، غير متزوج)، وبالنسبة للتعليم (متعلم، غير متعلم)... وهكذا. ويمكن التعبير عن ذلك بالجدول التالي:

الجدول رقم (2) يبين التكرارات المشتركة بين الصفات:

X / Y	الصفة الأولى	الصفة الثانية
الصفة الأولى	a	b

الصفة الثانية	c	d

حيث: d,c,b,a هي التكرارات المشتركة بين الصفات.

ويرمز لمعامل الاقتران بالرمز r_c ويعطى بالعلاقة التالية:

$$r_c = \frac{ad - bc}{ad + bc} \qquad (48-11)$$

علماً بأن $-1 \le r_c \le +1$.

مثال 11:

إذا كان لدينا المتغير X الذي يعبر عن التعليم وفق الصفتين (متعلم، غير متعلم)، والمتغير Y الذي يعبر عن الصفتين (يعمل، لا يعمل)، وعند جمع البيانات عن 150 عاملاً في إحدى المؤسسات حصلنا على الجدول التالي:

Y \ X	متعلم	غير متعلم
يعمل	60	30
لا يعمل	25	35

والمطلوب: إيجاد معامل الاقتران (معامل الارتباط) بين المتغيرين X و Y.

الحل:

نطبق العلاقة التالية:

$$r_t = \frac{ad - bc}{ad + bc}$$

$$r_t = \frac{60 \cdot 35 - 30 \cdot 25}{60 \cdot 35 + 30 \cdot 25} = 0{,}47$$

أي يوجد ارتباط طردي متوسط بين التعليم والعمل.

تمرينات

1 . تم سحب عينة عشوائية مؤلفة من 12 ثنائية من القيم لـ X و Y ومنه حصلنا على النتائج التالية:

$$\sum x_i^2 = 8040 \quad , \quad \sum x_i = 300 \quad , \quad \sum y_i = 342$$
$$\sum y_i^2 = 11380 \quad , \quad \sum y_i x_i = 9020$$

والمطلوب:

1 . أوجد خط انحدار y على x إذا علمت أن العلاقة بينهما خطية من الدرجة الأولى ثم ارسمه.

2 . إذا كانت $X = 30$ فما هي قيمة y المقدرة؟

3 . أوجد مجال الثقة لمعامل خط الانحدار β باحتمال قدره $1 - \alpha$.

4 . اختبر الفرضية $H_0 : \beta = 2$ عند مستوى دلالة 0,01.

2 . إذا كانت لدينا بيانات عن الإنفاق والدخل لمجموعة من الأسر مبينة في الجدول التالي:

رقم الأسرة	1	2	3	4	5	6	7	8	9	10
الدخل	8	10	12	12	13	15	20	22	23	25
الإنفاق	6	8	10	11	12	13	15	17	20	22

والمطلوب:

. أوجد خط انحدار Y على X ثم ارسمه على شكل الانتشار نفسه.

. أوجد قيمة y المقدرة عندما يصبح الدخل $x = 30$ و $x = 35$.

ـ اختبر الفرضية $H_0 : \beta = 0$ عند مستوى دلالة 0,05 باستخدام إحصاء الاختبار F.

. اختبر الفرضية $H_0 : \beta = 0$ عند مستوى دلالة 0,05 باستخدام الإحصاء t.

. أوجد مجال الثقة لمعامل خط الانحدار β باحتمال قدره $1 - \alpha = 0,95$.

. اختبر الفرضية $H_0 : \beta = 3$ عند مستوى دلالة $\alpha = 0,05$ وفسّر النتيجة.

. إيجاد معامل التحديد وفسّر النتيجة.

3 . استخدم معلومات التمرين رقم (2) وأوجد معامل الارتباط وفسر النتيجة، ثم قارن بين معامل الارتباط ومعامل التحديد، واختبر الفرضية التالية: $H_0 : \rho = 0$ عند مستوى دلالة $\alpha = 0,05$.

4 . من أجل دراسة العلاقة الارتباطية بين عمر الأب وعدد الأطفال في أحد المصارف، قامت مجموعة من الطلاب بجمع البيانات عن عينة من الموظفين فكانت نتائج الاستبيان كما في الجدول التالي:

عدد الأطفال / عمر الأب	[0−2]	[2−4]	[4−6]	[6−8]	[8−10]	[10−12]
]20−24[	10	6	−	−	−	−
]24−28[	8	5	2	−	−	−
]28−32[	4	2	4	1	−	−
]32−36[	1	2	8	3	−	−
]36−40[	−	1	3	7	5	−
]40−44[	−	−	2	10	3	1
]44−48[	−	−	1	4	2	2
]48−52[	−	−	−	2	1	3
]52−56[	−	−	−	1	1	3
]56−60[	−	−	−	−	−	2

والمطلوب:

1 . أوجد معامل الارتباط بين عمر الأب وعدد الأطفال وفسر النتيجة.

2 . اختبر فرضية فيما إذا كانت r تدل على وجود علاقة بين عمر الأب وعدد الأطفال عند مستوى دلالة 0,05 و 0,01.

5 . في دراسة اجتماعية مكونة من 5 عائلات عن الوضع المالي لكل فرد من أسرتيّ الزوج والزوجة وذلك لمعرفة تأثير الحالة المادية في الزواج بين الأسر. وكانت المعلومات التي تم الحصول عليها مبينة في الجدول التالي:

الحالة المادية لأسرة الزوج X	جيدة	منخفضة	متوسطة	ممتازة	متوسطة
الحالة المادية لأسرة الزوجة Y	ممتازة	جيدة	جيدة	ممتازة	متوسطة

والمطلوب: حساب معامل ارتباط الرتب.

6 . أوجد معامل الاقتران بين العمل والزواج لمجموعة من الأفراد وفق الصفتين (يعمل، لا يعمل) و (متزوج، غير متزوج) وكانت البيانات التي تم الحصول عليها مبينة في الجدول التالي:

العمل / الزواج	يعمل	لا يعمل
متزوج	60	30
غير متزوج	25	35

الفصل الثاني عشر

الانحدار الخطي المتعدد

Multiple linear regression and correlation

ويتضمن الفقرات التالية:

1 . مقدمة.

2 . إيجاد معادلة الانحدار الخطي المتعدد:

. حساب قيم المعاملات.

. حساب معامل التحديد ومعامل الارتباط المتعدد.

. حساب معاملات الارتباط الجزئية.

3 . الاختبارات الإحصائية المتعلقة بالانحدار المتعدد.

4 . إيجاد معادلة الانحدار الخطي المتعدد باستخدام المصفوفات.

5 . تمرينات.

الفصل الثاني عشر

الانحدار الخطي المتعدد

Multiple linear regression and correlation

12-1 : مقدمة Introduction:

لقد درسنا في الفصل السابق مسألة إيجاد العلاقة الرياضية بين متغيرين أحدهما مستقل X والآخر تابع Y، ولكن واقع الحياة الاقتصادية غالباً لايتوافق مع هذا النوع من العلاقات. بل نلاحظ أن أية ظاهرة من الظواهر سواء كانت اقتصادية أو اجتماعية تتأثر بالعديد من المتغيرات المؤثرة فيها، سواء كانت متغيرات رئيسية أو متغيرات ثانوية. فعلى سبيل المثال نلاحظ أن كمية استهلاك الطاقة الكهربائية للأسرة يتأثر بعدد أفراد الأسرة والأدوات الكهربائية المتوفرة لديها وبمساحة المنزل ... الخ. أي أننا نكون أمام متغير تابع واحد Y وعدداً من المتغيرات المستقلة $X_1, X_2, \ldots\ldots\ldots, X_n$ التي تؤثر في التغير التابع، ونريد إيجاد العلاقة الرياضية التي تربط بين هذين النوعين من المتغيرات، وهذا ما يدعى بالانحدار المتعدد. وقد تمثل هذه العلاقة بتابع جداء، أو تابع أسّي، أو خطي ..إلخ. وسنكتفي في هذا الفصل بدراسة العلاقة التي تُمَثل بتابعٍ خطي يتضمن عدداً من المتغيرات، وعندئذٍ تدعى هذه العلاقة بالانحدار الخطي المتعدد وذلك باستخدام المعادلات الطبيعية وباستخدام المصفوفات.

12-2:إيجاد معادلة الانحدار الخطي المتعدد باستخدام المعادلات الطبيعية:

إذا كان لدينا متغير Y يمثل ظاهرة ما وكانت هذه الظاهرة تتأثر بعدة متغيرات مستقلة، ولتكن $X_1, X_2, X_3, \ldots\ldots, X_k, \ldots\ldots, X_n$. ولدراسة تأثير هذه المتغيرات على المتغير التابع Y نسحب عينة من المشاهدات والتي تخص الظاهرة المدروسة وبحجم n مشاهدة والتي يمكن ترتيبها بالشكل التالي:

$$
\begin{array}{lccccccc}
Y: & y_1 & y_2 & y_3 & .... & y_i & .... & y_n \\
X_1: & x_{11} & x_{12} & x_{13} & ..... & x_{1i} & .... & x_{1n} \\
X_2: & x_{21} & x_{22} & x_{23} & ..... & x_{2i} & .... & x_{2n} \\
X_3: & x_{31} & x_{32} & x_{33} & ..... & x_{3i} & ..... & x_{3n} \\
. & & & & & & & \\
. & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
. & & & & & & & \\
X_k & x_{k1} & x_{k2} & x_{k3} & ..... & x_{ki} & ..... & x_{kn} \\
. & & & & & & & \\
. & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
. & & & & & & & \\
X_m & x_{m1} & x_{m2} & x_{m3} & ..... & x_{mi} & ..... & x_{mn}
\end{array}
\qquad (1-12)
$$

ولنفترض أن علاقة الانحدار بين Y والمتغيرات $X_1, X_2, X_3 .... X_k .... X_m$ هي علاقة خطية من الشكل التالي:

$$\widetilde{Y} = b_0 + b_1 X_1 + b_2 X_2 + ... + b_k X_k + ... + b_m X_m + e \quad (2-12)$$

حيث: $b_0, b_1, b_2, b_3, ..., b_m$ ثوابت عددية مجهولة.

ويمكن كتابة العلاقة (12-22) بدلالة قيم متحولاتها، أي:

$$\widetilde{Y}_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + b_3 x_{3i} + ... + b_k x_{ki} + ... + b_m x_{mi} + e_i \quad (3-12)$$

وبعد تحديد علاقة انحدار Y على المتغيرات المستقلة $X_1, X_2 .... X_m$ تظهر لنا مسألة إيجاد قيم ثوابت العلاقة.

12-2-1: إيجاد قيم ثوابت معادلة الانحدار الخطي المتعدد:

وللتبسيط في دراسة هذا النوع من العلاقات سوف نقوم بدراسة الانحدار الخطي المتعدد من خلال انحدار Y على متغيرين مستقلين X_1 و X_2 وبالتالي فإن علاقة الانحدار الخطي المتعدد تأخذ الشكل التالي:

$$\widetilde{Y} = b_0 + b_1 X_1 + b_2 X_2 + e \quad (4-12)$$

وتكتب العلاقة (12-4) بدلالة قيم متحولاتها كما يلي:

$$\widetilde{Y}_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + e_0 \qquad (5-12)$$

حيث b_2 , b_1 , b_0 ثوابت عددية مجهولة. e – تمثل الخطأ العشوائي.

وتدعى b_0, b_1, b_2 بمعاملات الانحدار الجزئية Partial regression coefficients.

ولتقدير العلاقة بين X_2, X_1, Y يجب تقدير الثوابت b_0, b_1, b_2، ونأخذ علاقة الانحدار الخطي المتعدد ذات الشكل التالي:

$$\widetilde{Y} = b_0 + b_1 X_1 + b_2 X_2 \qquad (6-12)$$

ولإيجاد ثوابت علاقة الانحدار الخطي المتعدد نشكل تابع المربعات الصغرى لقيم المتحولات التالي:

$$F = \sum_{i=1}^{n} (y_i - \widetilde{y}_i)^2 = \sum_{i=1}^{n} (y_i - b_0 - b_1 x_{1i} - b_2 x_{2i})^2 \to \min \quad (7-12)$$

نقوم بإيجاد المشتقات الجزئية للتابع F بالنسبة للثوابت b_0, b_1, b_2 ، ونجعلها تساوي الصفر ومن ثم نقوم بالإصلاحات الرياضية فنحصل على المعادلات الطبيعية التالية:

$$\begin{aligned}
&\sum_{i=1}^{n} y_i = n b_0 + b_1 \sum_{i=1}^{n} x_{1i} + b_2 \sum_{i=1}^{n} x_{2i} \\
&\sum_{i=1}^{n} y_i x_{1i} = b_0 \sum_{i=1}^{n} x_{1i} + b_1 \sum_{i=1}^{n} x_{1i}^2 + b_2 \sum_{i=1}^{n} x_{2i} x_{1i} \qquad (8-12) \\
&\sum_{i=1}^{n} y_i x_{2i} = b_0 \sum_{i=1}^{n} x_{2i} + b_1 \sum_{i=1}^{n} x_{i1} x_{2i} + b_2 \sum_{i=1}^{n} x_{2i}^2
\end{aligned}$$

وبحل جملة هذه المعادلات بإحدى الطرق المعروفة نحصل على قيم الثوابت b_0, b_1, b_2 (معاملات الانحدار الجزئية) ثم نعوض في العلاقة (6–12) فنحصل على العلاقة المقدرة لانحدار Y على X_1 و X_2.

ولكن لتبسيط العمليات الحسابية. نقسم المعادلة الأولى من جملة العلاقات (8–12) على n فنجد أن:

$$\overline{y} = b_0 + b_1 \overline{x}_1 + b_2 \overline{x}_2 \qquad (9-12)$$

ومن هذه العلاقة نحصل على قيمة الثابت b_0 من العلاقة:

$$b_0 = \bar{y} - b_1\bar{y}_1 - b_2\bar{y}_2 \qquad (10-12)$$

وبـإجراء الحسـابات بدلالـة انحرافـات القـيم عـن متوسـطها الحسـابي بالنسـبة للمتغيرات X_1, X_2, Y أي:

$$d_{yi} = y_i - \bar{y}$$
$$d_{1i} = x_{i1} - \bar{x}_1$$
$$d_{2i} = x_{i2} - \bar{x}_2$$

وبتعويض هذه الانحرافات في جملة المعادلات الطبيعية (12-8) فإننا نجد أن المعادلة الأولى سوف تختفي لأن مجموع انحرافات القيم عن وسطها الحسابي يساوي الصفر، وبالتالي فإن قيمة b_0 تساوي الصفر وتختصر جملة المعادلات السابقة إلى الشكل التالي:

$$\Sigma d_y \cdot d_1 = b_1\Sigma d_1^2 + b_2\Sigma d_1 \cdot d_2 \qquad (11-12)$$
$$\Sigma d_y \cdot d_2 = b_1\Sigma d_1 \cdot d_2 + b_2\Sigma d_2^2$$

ويمكن إيجاد الحدود الواردة في هاتين المعادلتين من خلال العلاقات التالية:

$$\Sigma d_y \cdot d_1 = \sum_{i=1}^{n} x_{1i} \cdot y_i - \frac{(\sum_{i=1}^{n} x_{1i})(\sum_{i=1}^{n} y_i)}{n}$$

$$\Sigma d_1^2 = \sum_{i=1}^{n} x_{1i} - \frac{(\sum_{i=1}^{n} x_{1i})^2}{n}$$

$$\Sigma d_2^2 = \sum_{i=1}^{n} x_{2i}^2 - \frac{(\sum_{i=1}^{n} x_{2i})^2}{n} \qquad (12-12)$$

$$\Sigma d_y \cdot d_2 = \sum_{i=1}^{n} x_{2i} \cdot y_i - \frac{(\sum_{i=1}^{n} x_{2i})(\sum_{i=1}^{n} y_i)}{n}$$

$$\Sigma d_1 \cdot d_2 = \sum_{i=1}^{n} x_{1i} \cdot x_{2i} - \frac{(\sum_{i=1}^{n} x_{1i})(\sum_{i=1}^{n} x_{2i})}{n}$$

وبحل جملة المعادلتين (12- 11) نحصل على قيم معاملات الانحدار الجزئي b_2, b_1 ثم نعوضها في (12- 10) فنحصل على b_0.

ويمكن تفسير معامل الانحدار الجزئي b_k بأنه يقيس معدل التغير المتوقع في المتغير التابع Yنتيجة التغير الذي يطرأ على المتغير المستقل X_k بمقدار وحدة واحدة مع بقاء أثر المتغيرات الأخرى المستقلة ثابتاً.

وبالتالي يمكن التنبؤ بقيم المتغير التابع Y إذا عُلمت قيم المتغيرات المستقلة الأخرى. إذا كانت إشارة أحد معاملات الانحدار الجزئية سالبة فإن ذلك المتغير يؤثر عكسياً على قيمة المتغير التابع Y. أي إذا كانت معادلة الانحدار الخطي المتعدد وعلى سبيل المثال، من الشكل:

$$\widetilde{Y} = b_0 + b_1 X_1 - b_2 X_2$$

فإننا نقول أن المتغير X_2 يؤثر عكسياً على المتغير التابع $\widetilde{Y}$.

مثال 1:

أراد أحد الباحثين دراسة العلاقة بين الإنفاق السنوي على الغذاء ومقدار الدخل وعدد أفراد الأسرة. ومن أجل ذلك جمع معلومات عن 15 أسرة فكانت البيانات التي حصل عليها مبينة في الجدول التالي:

رقم الأسرة	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Σ
الإنفاق السنوي Y (ألف ليرة)	70	62	65	70	75	80	85	75	68	80	88	90	94	98	100	1200
مقدار X_1 دخل (ألف ليرة)	100	90	95	110	115	120	135	140	115	115	120	130	125	140	150	1800
عدد X_2 أفراد الأسرة	3	4	4	5	6	7	8	5	6	8	8	7	7	8	9	95

البيانات فرضية.

والمطلوب:

إيجاد علاقة الانحدار بين Y الإنفاق السنوي، ومقدار الدخل X_1، وعدد أفراد الأسرة X_2، وذلك بفرض أن الإنفاق السنوي هو عبارة عن علاقة خطية لمقدار الدخل وعدد أفراد الأسرة.

ثم إيجاد القيمة المقدرة للإنفاق السنوي لأسرة مقدار دخلها 170 ألف ليرة سورية وعدد أفرادها 8 أفراد.

الحل:

باعتبار أن العلاقة خطية، فإن معادلة الانحدار المتعدد هي من الشكل التالي:

$$Y = b_0 + b_1 X_1 + b_2 X_2 \qquad (1)$$

نكتب هذه العلاقة بدلالة قيم متحولاتها فتأخذ الشكل التالي:

$$y = b_0 + b_1 x_{1i} + b_2 x_{2i} \qquad (2)$$

ولإيجاد ثوابت هذه العلاقة نستخدم المعادلات الطبيعية التالية:

$$\begin{aligned}
&\sum_{i=1}^{n} y_i = nb_0 + b_1 \sum_{i=1}^{n} x_{1i} + b_2 \sum_{i=1}^{n} x_{2i} \\
&\sum_{i=1}^{n} y_i x_{1i} = b_0 \sum_{i=1}^{n} x_{1i} + b_1 \sum_{i=1}^{n} x_{1i}^2 + b_2 \sum_{i=1}^{n} x_{1i} x_{2i} \qquad (3) \\
&\sum_{i=1}^{n} y_i x_{2i} = b_0 \sum_{i=1}^{n} x_{2i} + b_1 \sum_{i=1}^{n} x_{1i} x_{2i} + b_2 \sum_{i=1}^{n} x_{2i}^2
\end{aligned}$$

ولإيجاد حدود هذه المعادلات نشكل الجدول المساعد التالي:

رقم الأسرة	Y	X_1	X_2	$y_i \cdot x_{1i}$	$y_i \cdot x_{2i}$	$x_{1i} \cdot x_{2i}$	x_{1i}^2	x_{2i}^2	$\hat{y}_i$	y_i^2
1	70	100	3	7000	210	300	10000	9	62,39	4900
2	62	90	4	5580	248	360	8100	16	62,74	3844
3	65	95	4	6175	260	380	9025	16	64,29	4225
4	70	110	5	7700	350	550	12100	25	72,39	4900
5	75	115	6	8625	450	690	13225	36	77,39	5625
6	80	120	7	9600	560	840	14400	49	82,39	6400
7	85	135	8	11475	680	1080	18225	64	90,49	7225
8	75	140	5	10500	375	700	19600	25	81,69	5625
9	68	115	6	7820	408	690	13225	36	77,39	4624
10	80	115	8	9200	640	920	13225	64	84,29	6400
11	88	120	8	10560	704	960	14400	64	85,84	7744
12	90	130	7	11700	630	910	16900	49	85,49	8100
13	94	125	7	11750	658	875	15625	49	83,94	8836
14	98	140	8	13720	784	1120	19600	64	92,04	9604
15	100	150	9	15000	900	1350	22500	81	98,59	10000
Σ	1200	1800	95	146405	7857	11725	220150	647		98052

وبالتعويض في المعادلات الطبيعية نجد:

$$1200 = 15 \cdot b_0 + 1800 \cdot b_1 + 95 \cdot b_2$$
$$146405 = 1800 b_0 + 220150 b_1 + 11725 b_2$$
$$7857 = 95 b_0 + 11725 b_1 + 647 b_2$$

وبحل هذه المعادلات بإحدى الطرق المعروفة نجد أن معاملات الانحدار الجزئي تساوي:

$$b_0 \approx 21$$
$$b_1 \approx 0{,}31$$
$$b_2 \approx 3{,}45$$

وبالإمكان الحصول على نفس قيم معاملات الانحدار الجزئي باستخدام العلاقات (11–12) وحساب حدودها بالعلاقات التالية:

$$\Sigma d_y \cdot d_1 = \sum_{i=1}^{n} x_{1i} \cdot y_i - \frac{(\sum_{i=1}^{n} x_{1i})(\sum_{i=1}^{n} y_i)}{n}$$

$$\Sigma d_y \cdot d_1 = 146405 - \frac{1800 \cdot 1200}{15} = 2405$$

$$\Sigma d_y \cdot d_2 = \sum_{i=1}^{n} x_{2i} \cdot y_i - \frac{(\sum_{i=1}^{n} x_{2i})(\sum_{i=1}^{n} y_i)}{n}$$

$$\Sigma d_y \cdot d_2 = 7857 - \frac{95 \cdot 1200}{15} = 257$$

$$\Sigma d_1^2 = \sum_{i=1}^{n} x_{1i}^2 - \frac{\left(\sum_{i=1}^{n} x_{1i}\right)^2}{n} = 220150 - \frac{(1800)^2}{15} = 4150$$

$$\Sigma d_1^2 = \sum_{i=1}^{n} x_{2i}^2 - \frac{\left(\sum_{i=1}^{n} x_{2i}\right)^2}{n} = 647 - \frac{(95)^2}{15} = 45{,}33$$

$$\Sigma d_1 \cdot d_2 = \sum_{i=1}^{n} x_{1i} \cdot x_{2i} - \frac{\left(\sum_{i=1}^{n} x_{1i}\right)\left(\sum_{i=1}^{n} x_{2i}\right)}{n}$$

$$\Sigma d_1 \cdot d_2 = 11725 - \frac{1800 \cdot 95}{15} = 325$$

وبالتالي تأخذ العلاقة (12– 11) الشكل التالي:

$$2405 = 4150b_1 + 325\ b_2$$
$$257 = 325b_1 + 45{,}33\ b_2$$

بحل المعادلتين معاً نجد أن معاملات الانحدار الجزئي تساوي:

$$b_1 \approx 0{,}31$$
$$b_2 \approx 3{,}45$$

وبالتعويض في (12– 10) نحصل على b_0 أي:

$$b_0 = 21$$

نلاحظ أن معاملات الانحدار الجزئي هي نفسها التي حصلنا عليها باستخدام المعادلات الطبيعية. وبالتعويض في العلاقة (1) نجد أن معادلة الانحدار الخطي المتعدد المقدرة تصبح بالشكل التالي:

$$\widetilde{Y} = 21 + 0{,}31X_1 + 3{,}45X_2 \qquad (4)$$

وهذا يعني أن الإنفاق السنوي للأسرة يساوي مبلغاً 21000 ليرة سورية بالإضافة إلى مقدارين آخرين يتعلقان بـ:

أولاً: الدخل السنوي للأسرة مضروباً بمعامل الانحدار $b_1 = 0{,}31$.

ثانياً: عدد أفراد الأسرة مضروباً بمعامل الانحدار $b_2 = 3{,}45$.

ولتقدير مقدار الإنفاق السنوي للأسرة التي دخلها 170 ألف ليرة سورية، وعدد أفرادها 8 أفراد، نعوض في معادلة الانحدار الخطي المتعدد رقم (4) فنجد:

$$\widetilde{Y} = 21 + 0{,}31(170) + 3{,}45(8) = 101{,}3\ S.P$$

أي أن الإنفاق السنوي لهذه الأسرة يقدر بـ 101300 ليرة سورية.

12–2–2: معامل التحديد ومعامل الارتباط الخطي المتعدد:

يعرف معامل التحديد المتعدد بأنه نسبة التباين المفسر من X_1 , X_2 إلى التباين الكلي، ونرمز له بـ R^2 ويعطى بالعلاقة التالية:

$$R^2 = \frac{\sum_{i=1}^{n} (\tilde{y}_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} = \frac{SSR}{SST} \qquad (13-12)$$

أو:

$$R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \qquad (14-12)$$

وبكتابة البسط في العلاقة (12−14) على الشكل التالي:

$$\Sigma(y_i - \tilde{y})^2 = \Sigma(y_i - \hat{y})(y_i - \hat{y})$$

وباستخدام أسلوب انحرافات القيم عن أوساطها الحسابية، وإجراء الإصلاحات الرياضية نجد أن معامل التحديد المتعدد يعطي بالعلاقة التالية:

$$R^2 = \frac{b_1 \sum d_1 \cdot d_y + b_2 \sum d_2 \cdot d_y}{\sum d_y^2} = \frac{SSR}{SST} \qquad (15-12)$$

حيث:

$$\sum d_y^2 = \sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}$$

ونشير أخيراً إلى أن معامل الارتباط المتعدد هو الجذر التربيعي لمعامل التحديد المتعدد.

مثال 2:

استخدم بيانات المثال رقم (1) ثم احسب معامل التحديد ومعامل الارتباط الخطي المتعدد.

الحل:

$$\sum d_y^2 = 98052 - \frac{(1200)^2}{15} = 2052$$

وأن:

$$\sum d_1 d_y = 2405 \ , \quad \sum d_2 d_y = 257$$

نعوض في العلاقة (12−15) فنجد أن:

$$R^2 = \frac{0{,}31 \cdot 2405 + 3{,}45 \cdot 257}{2052} = 0{,}795$$

ويمكن إيجاد معامل التحديد من العلاقة (12−14) مباشرة كما يلي:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} = 1 - \frac{421{,}29}{2052} = 0{,}795$$

ويشير معامل التحديد إلى أنه يمكننا تفسير 795% من التغير في الإنفاق على الغذاء عن طريق معرفتنا بالتغيرات الكلية في X_1 و X_2.

ومنه فإن معامل الارتباط الخطي المتعدد يساوي: $R = 0{,}892$ وهو ارتباط قوي بين المتغيرات.

12−2−3: حساب معاملات الارتباط الجزئية Partial Correlation:

إن معاملات الارتباط الجزئية تقيس العلاقة الحقيقية بين متغيرين بعد استبعاد أثر المتغيرات الأخرى على المتغيرين معاً. فعندما ندرس علاقة انحدار Y على X_1 و X_2 فإن معامل الارتباط الجزئي بين Y و X_1 يقيس لنا العلاقة بين Y و X_1 فقط بعد استبعاد أثر X_2 من المتغيرين Y و X_1 معاً. وبمعنى آخر يستثني حقيقة وجود المتغير X_2 في نموذج الانحدار.

كما أن معامل الارتباط الجزئي بين Y و X_2 يقيس لنا العلاقة بين Y و X_2 فقط بعد استبعاد أثر X_1 من المتغيرين Y و X_2 معاً وبمعنى آخر يستثني حقيقة وجود المتغير X_1 في نموذج الانحدار. لنرمز بـ $r_{y1\cdot 2}$ لمعامل الارتباط الجزئي بين Y و X_1 وتثبيت المتغير X_2 (استبعاد أثر X_2) ويعطى هذا المعامل بالعلاقة التالية:

$$r_{y1\cdot 2} = \frac{r_{y1} - r_{y2} \cdot r_{12}}{\sqrt{\left(1 - r_{12}^2\right)\left(1 - r_{y2}^2\right)}} \qquad (16-12)$$

ولنرمز بـ $r_{y2\cdot 1}$ لمعامل الارتباط الجزئي بين Y و X_2 وتثبيت المتغير X_1 (استبعاد أثر X_1) ويعطى هذا المعامل بالعلاقة التالية:

$$r_{y2\cdot 1} = \frac{r_{y2} - r_{12} \cdot r_{y1}}{\sqrt{\left(1 - r_{12}^2\right)\left(1 - r_{y1}^2\right)}} \qquad (17-12)$$

ويمكن أن نذكر أن الإشارة الجبرية لمعامل الارتباط الجزئي هي نفس الإشارة لمعامل الانحدار الجزئي. فإذا كانت $r_{y1\cdot 2}$ سالبة فإن b_1 سالبة. ولإيجاد معاملات الارتباط الجزئية في العلاقات (12- 16) و(12- 17) من المفروض أن نقوم بإيجاد معاملات الارتباط الجزيئي (نصف الجزء) Part or semi partial correlation بين كل متحولين على حدة، وبالتالي نحصل على معاملات للارتباط الجزيئي، وتدعى أيضاً بمعاملات الارتباط البسيط، التي تستخدم في تقييم الأهمية النسبية للمتغيرات المستقلة في التنبؤ بقيم المتغير التابع Y، ويعطى معامل الارتباط بين المتغيرين X_1 و X_2 ونرمز له بـ r_{12} بالعلاقة التالية:

$$r_{12} = \frac{\sum d_1 \cdot d_2}{\sqrt{\sum d_1^2 \cdot \sum d_2^2}} \qquad (18-12)$$

ونرمز لمعامل الارتباط بين Y و X_1 بـ r_{y1}، والذي يعطى بالعلاقة التالية:

$$r_{y1} = \frac{\sum d_1 \cdot d_y}{\sqrt{\sum d_1^2 \cdot \sum d_y^2}} \qquad (19-12)$$

ونرمز لمعامل الارتباط بين Y و X_2 بـ r_{y2}، والذي يعطى بالعلاقة التالية:

$$r_{y2} = \frac{\sum d_2 \cdot d_y}{\sqrt{\sum d_2^2 \cdot \sum d_y^2}} \qquad (20-12)$$

وتقيس معاملات الارتباط الجزيئي (نصف الجزء) العلاقة بين كل متغيرين وإهمال المتغير الآخر (افتراض عدم وجوده).

مثال 3:

استخدم بيانات المثال رقم (1) ثم احسب معاملات الارتباط الجزئية.

الحل:

لحساب معامل الارتباط الجزئي بين Y و X_1 واستبعاد أثر المتغير X_2 . وأيضاً حساب معامل الارتباط الجزئي بين Y و X_2 واستبعاد أثر المتغير X_1 ، يتوجب علينا حساب معامل الارتباط الجزيئي (وهو نفسه معامل الارتباط البسيط)، وتحسب هذه المعاملات من العلاقات (12− 18) و(12− 19) و(12− 20) كمايلي:

. معامل الارتباط الجزئيئ بين X_1 و X_2 يساوي:

$$r_{12} = \frac{\sum d_1 \cdot d_2}{\sqrt{\sum d_1^2 \cdot \sum d_2^2}}$$

$$r_{12} = \frac{325}{\sqrt{4150 \cdot 45{,}33}} \approx 0{,}75$$

ومنه:

$$r_{12}^2 = 0{,}5625$$

. معامل الارتباط الجزيئي بين Y و X_1 ويساوي:

$$r_{y1} = \frac{\sum d_1 \cdot d_y}{\sqrt{\sum d_1^2 \cdot \sum d_y^2}}$$

$$r_{y1} = \frac{2405}{\sqrt{4150 \cdot 2052}} \approx 0{,}82$$

ومنه: $r_{y1}^2 = 0{,}6724$

. معامل الارتباط الجزيئي بين Y و X_2 ويساوي:

$$r_{y2} = \frac{\sum d_2 \cdot d_y}{\sqrt{\sum d_2^2 \cdot \sum d_y^2}}$$

$$r_{y2} = \frac{257}{\sqrt{45{,}33 \cdot 2052}} = 0{,}84$$

ومنه:

$$r_{y2}^2 = 0{,}7056$$

والآن نقوم بحساب معاملات الارتباط الجزئية.

ـ معامل الارتباط الجزئي بين Y و X_1 واستبعاد أثر المتغير X_2 ويحسب من العلاقة:

$$r_{y1\cdot 2} = \frac{r_{y1} - r_{y2} \cdot r_{12}}{\sqrt{\left(1 - r_{12}^2\right)\left(1 - r_{y2}^2\right)}}$$

$$r_{y1\cdot 2} = \frac{0{,}82 - 0{,}84 \cdot 0{,}75}{\sqrt{(1 - 0{,}5625)(1 - 0{,}7056)}} = 0{,}53$$

ومنه:

$$r_{y1\cdot 2}^2 = 0{,}281$$

وهذا يعني أن الدخل X_1 يفسر 28,1% من التغير في الإنفاق بعد أن كان عدد الأطفال X_2، قد فسر كل ما يستطيع من التغير في الإنفاق.

. معامل الارتباط الجزئي بين Y و X_2 واستبعاد أثر المتغير X_1 ويحسب من العلاقة:

$$r_{y2\cdot 1} = \frac{r_{y2} - r_{12} \cdot r_{y1}}{\sqrt{\left(1 - r_{12}^2\right)\left(1 - r_{y1}^2\right)}}$$

$$r_{y2\cdot 1} = \frac{0{,}84 - 0{,}75 \cdot 0{,}82}{\sqrt{(1 - 0{,}5625)(1 - 0{,}6724)}} \approx 0{,}60$$

ومنه فإن:

$$r_{y2\cdot 1}^2 = 0{,}36$$

وهذا يعني أن عدد الأطفال X_2 يفسر 36% من التغير في الإنفاق بعد أن كان الدخل قد فسر كل ما يستطيع من التغير في الإنفاق. وبالمقارنة نجد أن المتغير X_2 يعتبر عاملاً مهماً أكثر من X_1 في التنبؤ بالإنفاق.

12-3: الاختبارات الإحصائية المتعلقة بالانحدار الخطي المتعدد:

12-3-3: اختبار معنوية معاملات الانحدار الجزئية :

لاختبار معنوية معاملات الانحدار الجزئية فيما إذا كانت تختلف عن الصفر، نقوم بتحديد الفرضية التي مفادها أن بيانات العينة مسحوبة من مجتمع إحصائي لايوجد فيه انحدار جزئي لـ Y على المتحولات X_1 و X_2 ..... X_m. أي أن $\beta_1, \beta_2 ..... \beta_m$ تساوي الصفر، لذلك نضع الفرضيات التالية:

. الفرضية الابتدائية:

$$H_0 : \beta_1 = \beta_2 = ... = \beta_k = ... = \beta_m = 0$$

أي جميع معاملات الانحدار الجزئية في المجتمع الإحصائي تساوي الصفر.

. الفرضية البديلة:

$$H_1 : \beta_k \neq 0$$

أي أن واحدة على الأقل من معاملات الانحدار الجزئية في المجتمع الإحصائي لا تساوي الصفر.

ومن أجل اختبار هذه الفرضية نستخدم اختبار t الإحصائي الذي يعطى بالعلاقة:

$$\widetilde{t}_k = \frac{b_k - \beta_k}{S_{bk}} \qquad (21-12)$$

والذي يخضع لتوزيع t بدرجات حرية $n-m-1$ وبمستوى دلالة α.

وعندما $k=1$ فإن:

$$S_{b1}^2 = \frac{SSE}{n-m-1} \cdot \frac{\sum d_1^2}{\sum d_1^2 \cdot \sum d_2^2 - \left(\sum d_1 \cdot d_2\right)^2} \qquad (22-12)$$

و يمكن أن نحصل على SSE من العلاقة (12- 14) بالشكل التالي:

$$R^2 = \frac{SST - SSE}{SST}$$

$$SST \cdot R^2 = SST - SSE$$

وبالإصلاح نجد:

$$SSE = SST(1 - R^2)$$

$$SSE = \sum d_y^2 (1 - R^2) \qquad (23-12)$$

وعندما $k = 2$ فإن:

$$S_{b2}^2 = \frac{SSE}{n-m-1} \cdot \frac{\sum d_2^2}{\sum d_1^2 \cdot \sum d_2^2 - \left(\sum d_1 \cdot d_2\right)^2} \qquad (24-12)$$

ثم نقارن قيمة $\tilde{t}$ مع t النظرية، وبالتالي إما أن نرفض الفرضية الابتدائية أو نقبلها، ويمكن إيجاد مجال الثقة لمعاملات الانحدار الجزئية بالعلاقة التالية:

$$I_\beta = P\left[b_k - t_{(1-\frac{\alpha}{2},n-m-1)} \cdot S_{bk} < \beta_k < b_k + t_{(1-\frac{\alpha}{2},n-m-1)} \cdot S_{bk} \right] = 1 - \alpha \quad (25-12)$$

ويدعى المقدار $\sqrt{\frac{SSE}{n-m-1}}$ بالخطأ المعياري للتقدير ويرمز له بـ S_{yx}، ويساوي

$$S_{yx} = \sqrt{\frac{SSE}{n-m-1}} = \sqrt{\frac{\sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2}{n-m-1}}$$

وهو عبارة عن مقياس لقابلية التغير حول الانحدار.

مثال 4:

استخدم بيانات المثال رقم (1) ثم اختبر معنوية معاملات الانحدار الجزئية، وأوجد مجالات الثقة عند مستوى دلالة $\alpha = 0{,}05$.

الحل:

نضع الفرضيات كما يلي:

. الفرضية الابتدائية:

$$H_0 : \beta_1 = \beta_2 = 0$$

جميع معاملات الانحدار الجزئية تساوي الصفر في المجتمع الإحصائي.

. الفرضية البديلة:

$$H_1 : \beta_k \neq 0$$

من أجل واحدة على الأقل لا تساوي الصفر.

نحسب قيمة $\widetilde{t_1}$ المقدرة بالنسبة لـ b_1:

$$\widetilde{t_1} = \frac{b_1 - \beta_1}{S_{b1}}$$

و:

$$SSE = \sum d_y^2 \left(1 - R^2\right) = 2052(1 - 0{,}795) = 420{,}66$$

وبالتعويض في العلاقة (12- 22) نجد:

$$S_{b1}^2 = \frac{420{,}66}{15 - 2 - 1} \cdot \frac{45{,}33}{4150 \cdot 45{,}33 - (325)^2} \approx 0{,}0193$$

ومنه:

$$S_{b1} = 0{,}139$$

نعوض في العلاقة (12- 21) نجد:

$$\widetilde{t_1} = \frac{0{,}31 - 0}{0{,}139} = 2{,}23$$

إن قيمة t النظرية عند درجات حرية $n - m - 1 = 15 - 2 - 1 = 12$، وعند مستوى دلالة $\alpha = 0{,}05$ تساوي:

$$t_{(0{,}975,12)} = 2{,}179$$

وبالمقارنة نجد أن: $\widetilde{t_1} > t_{1-\frac{\alpha}{2}}$، وبالتالي نرفض الفرضية الابتدائية ونقبل الفرضية البديلة. وأن b_1 لا تساوي الصفر، وبالتالي فإن X_1 تساهم في تحديد الاختلافات الكلية في المتغير التابع.

مجال الثقة فيساوي بالنسبة β_1 :

$$I_{\beta} = P[0,31 - 2,179 \cdot 0,139 < \beta_1 < 0,31 + 2,179 \cdot 0,139] = 0,95$$
$$[0,0071 < \beta_1 < 0,61]$$

وبالتالي فإن معامل انحدار المجتمع الإحصائي β_1 يقع ضمن حدي المجال المذكور.

. نحسب قيمة $\widetilde{t}_2$ بالنسبة b_2:

$$S_{b2}^2 = \frac{420,66}{12} \cdot \frac{4150}{4150 \cdot 45,33 - (325)^2} = 1,764$$

ومنه:

$$S_{b2} = 1,33$$

ثم نعوض في العلاقة (12- 21) فنجد أن:

$$\widetilde{t}_2 = \frac{3,45 - 0}{1,33} \approx 2,59$$

وبالمقارنة نجد أن: $\widetilde{t}_2 > t_{1-\frac{\alpha}{2}}$.

ولذلك نرفض الفرضية الابتدائية ونقبل الفرضية البديلة. أي أن b_2 لاتساوي الصفر، وبالتالي فإن X_2 تساهم في تحديد الاختلافات الكلية في المتغير التابع، وبشكل أكبر من X_1 .

مجال الثقة بالنسبة β_2 يساوي:

$$I_{\beta} = P[3,45 - 2,179 \cdot 1,33 < \beta_2 < 3,45 + 2,179 \cdot 1,33] = 0,95$$
$$[0,552 < \beta_2 < 6,348]$$

وبالتالي فإن معامل انحدار المجتمع الإحصائي β_2 يقع ضمن حدي المجال المذكور.

12-3-2: اختبار معنوية معامل الارتباط الخطي المتعدد:

وهنا نريد اختبار فرضية عدم وجود ارتباط بين المتغير Y والمتغيرين المستقلين X_1 و X_2 ومن أجل ذلك يجب تحديد الفرضيات التالية:

. الفرضية الابتدائية:

$$H_0 : \rho_{y\cdot 12} = 0$$

أي عدم وجود ارتباط بين التابع Y والمتغيرين X_1 و X_2 ، وبالتالي فإن:

$$b_0 = b_1 = b_2 = 0$$

. الفرضية البديلة:

$$H_1 : \rho_{y\cdot 12} \neq 0$$

أي يوجد ارتباط بين التابع Y والمتغيرين X_1 و X_2 ، ومن أجل واحدة على الأقل من b_0, b_1, b_2 لاتساوي الصفر.

ومن أجل اختبار هذه الفرضية نحسب النسبة $\widetilde{F}$ التي تعطى بالعلاقة التالية:

$$\widetilde{F} = \frac{R^2 / m}{\left(1 - R^2\right) / n - m - 1} \qquad (26-12)$$

ثم نقوم بمقارنة $\widetilde{F}$ مع قيمة F النظرية عند مستوى دلالة α، وبدرجات حرية $V_1 = m$ و $V_2 = n - m - 1$، فإذا كانت $\widetilde{F} < F_{\alpha(m,n-m-1)}$ فإننا نقبل H_0 والتي تنص على عدم وجود ارتباط بين المتغير التابع Y والمتغيرات المستقلة X_1 و X_2 . أما إذا كانت $\widetilde{F} > F_{\alpha(m,n-m-1)}$ فإننا نرفض H_0 ونقبل H_0 والتي تنص على وجود ارتباط بين المتغير التابع Y والمتغيرات المستقلة X_1 و X_2.

21-3-3- اختبار معنوية معامل الارتباط الجزئي:

ولإجراء هذا الاختبار يتوجب علينا اختبار معنوية معامل الارتباط الجزئي بين Y وكل متغير من المتغيرات المستقلة، وأن هذا الاختبار يعتمد على توزيع t بدرجات حرية $n - m - 1$، وبمستوى دلالة α.

أ . اختبار معنوية معامل الارتباط الجزئي بين Y و X_1 بعد استبعاد أثر المتغير X_2 :

ولإجراء هذا الاختبار لابد من تحديد الفرضيات الإحصائية:

الفرضية الابتدائية:

$$H_0 : r_{y1\cdot 2} = 0$$

أي عدم وجود ارتباط بين المتغيرين Y و X_1 .

الفرضية البديلة:

$$H_1 : r_{y1\cdot 2} \neq 0$$

أي يوجد ارتباط بين المتغيرين Y و X_1 .

ثم نحسب قيمة $\tilde{t}_1$ من العلاقة:

$$\tilde{t}_1 = \frac{\left|r_{y1\cdot 2}\right| \cdot \sqrt{n-k-1}}{\sqrt{1-r_{y1\cdot 2}^2}} \qquad (27-12)$$

وبعدئذ نقارن قيمة $\tilde{t}_1$ المقدرة مع t النظرية ضمن مستوى دلالة α، وبدرجات حرية $n-m-1$ فإذا كانت: $\tilde{t}_1 < t_{\left(1-\frac{\alpha}{2},n-m-1\right)}$، نقبل الفرضية الابتدائية، وبالتالي لا يوجد ارتباط بين Y و X_1 .

أما إذا كانت: $\tilde{t}_1 > t_{\left(1-\frac{\alpha}{2},n-m-1\right)}$، نرفض الفرضية الابتدائية، ونقبل الفرضية البديلة، ونقول أنه يوجد ارتباط بين Y و X_1 .

ب . اختبار معنوية معامل الارتباط الجزئي بين Y و X_2 بعد استبعاد أثر المتغير X_1 .

ولإجراء هذا الاختبار لابد من تحديد الفرضيات.

الفرضية الابتدائية:

$$H_0 : r_{y2\cdot 1} = 0$$

أي عدم وجود ارتباط بين المتغيرين Y و X_2 .

الفرضية البديلة:

$$H_1 : r_{y2\cdot 1} \neq 0$$

يوجد ارتباط بين المتغيرين Y و X_2 .

ثم نحسب قيمة $\widetilde{t}_2$ المقدرة، والتي تعطي بالعلاقة التالية:

$$\widetilde{t}_2 = \frac{\left|r_{y2\cdot1}\right| \cdot \sqrt{n-m-1}}{\sqrt{1-r_{y2\cdot1}^2}} \qquad (28-12)$$

وبعدئذ نقارن قيمة $\widetilde{t}_2$ المقدرة مع t النظرية ضمن مستوى دلالة α، وبدرجات حرية $n-m-1$ فإذا كانت:

$\widetilde{t}_2 < t_{\left(1-\frac{\alpha}{2},n-m-1\right)}$، نقبل الفرضية الابتدائية، وبالتالي لا يوجد ارتباط بين Y و X_2 .

أما إذا كانت:

$\widetilde{t}_2 > t_{\left(1-\frac{\alpha}{2},n-m-1\right)}$، نرفض الفرضية الابتدائية، ونقبل الفرضية البديلة، ونقول أنه يوجد ارتباط بين Y و X_2 .

مثال 5:

استخدم بيانات المثال رقم (1) ، ثم اختبر معنوية معامل الارتباط الجزئي المتعدد ومعنوية معامل الارتباط عند مستوى دلالة $\alpha = 0{,}05$.

الحل:

1 . اختبار معنوية معامل الارتباط الخطي المتعدد:

. تحديد الفرضيات:

الفرضية الابتدائية:

$$H_0 : \rho_{y\cdot12} = 0$$

$$b_0 = b_1 = b_2 = 0$$

الفرضية البديلة:

$$H_1 : \rho_{y\cdot12} \neq 0$$

من أجل واحدة من b_2 , b_1 , b_0 على الأقل لا تساوي الصفر.

ثم نحسب قيمة $\hat{F}$ من العلاقة (12- 26) فنجد:

$$\widetilde{F} = \frac{0,795/2}{(1-0,795)/(15-2-1)} = 23,27$$

وإن قيمة F النظرية عند مستوى دلالة 0,05، وبدرجات حرية $V_1 = 2$ و $V_2 = 12$، تساوي

$F_{0,95(2,12)} = 3,89$ وبمقارنة $\hat{F}$ المقدرة مع F النظرية نجد أن:

$$\hat{F} > F_{0,95(2,12)}$$

وبالتالي نرفض الفرضية الابتدائية H_0 والقائلة بعدم وجود ارتباط بين المتغير Y والمتغيرات المستقلة X_1 و X_2 ، ونقبل الفرضية البديلة التي تنص على وجود ارتباط بين المتغير Y والمتغيرات المستقلة .

2 . اختبار معنوية معامل الارتباط الجزئي:

أ . اختبار معنوية $r_{y1\cdot 2}$:

الفرضية الابتدائية:

$$H_0 : r_{y1\cdot 2} = 0$$

عدم وجود ارتباط بين المتغيرين Y و X_1.

الفرضية البديلة:

$$H_1 : r_{y1\cdot 2} \neq 0$$

وجود ارتباط بين المتغيرين Y و X_1.

. حساب قيمة $\widetilde{t}_1$ المقدرة:

$$\widetilde{t}_1 = \frac{\left|r_{y1\cdot 2}\right| \cdot \sqrt{n-m-1}}{\sqrt{1-r_{y1\cdot 2}^2}} = \frac{0,53 \cdot \sqrt{15-2-1}}{\sqrt{1-0,281}} = 2,17$$

نحسب قيمة t النظرية ضمن مستوى دلالة $\alpha = 0,05$، وبدرجات حرية $n-k-1=12$ فنجد:

$$t_{\left(1-\frac{\alpha}{2}/n-k-1\right)} = t_{(0,975\ ,\ 12)} = 2,179$$

وبمقارنة $\widetilde{t}_1$ المقدرة مع t النظرية نجد أن:

$$\hat{t}_1 > t_{(0,975,12)}$$

وبالتالي نرفض H_0، ونقبل H_1 والقائلة بوجود ارتباط بين Y و X_1 .

ب . اختبار معنوية $r_{y2\cdot1}$:

. تحديد الفرضيات:

الفرضية الابتدائية:

$$H_0 : r_{y2\cdot1} = 0$$

عدم وجود ارتباط بين المتغيرين Y و X_2.

الفرضية البديلة:

$$H_1 : r_{y2\cdot1} \neq 0$$

يوجد ارتباط بين المتغيرين Y و X_2.

حساب قيمة $\widetilde{t}_2$ المقدرة:

$$\widetilde{t}_2 = \frac{\left|r_{y2\cdot1}\right| \cdot \sqrt{n-m-1}}{\sqrt{1-r_{y2\cdot1}^2}} = \frac{0,6 \cdot \sqrt{15-2-1}}{\sqrt{1-0,36}} = 2,598$$

نحسب قيمة t النظرية ضمن مستوى دلالة $\alpha = 0,05$، وبدرجات حرية $n-k-1=12$ فنجد:

$$t_{(0,975\ ,\ 12)} = 2,179$$

وبمقارنة $\widetilde{t}_2$ المقدرة مع t النظرية نجد أن $\widetilde{t}_2 > t_{(0,975,12)}$ لذلك نرفض الفرضية الابتدائية، ونقبل الفرضية البديلة، ونقول أنه يوجد ارتباط بين Y و X_2 .

12-4: إيجاد معادلة الانحدار الخطي المتعدد باستخدام المصفوفات:

إن البيانات التي حصلنا عليها من جراء سحب عينة وبحجم n مشاهدة، والتي تخص الظاهرة المدروسة والموجودة تحت الرقم (1) يمكن التعبير عنها على شكل جملة معادلات، وكما نعلم أننا نستطيع كتابة جملة المعادلات بشكل مصفوفي كما يلي:

$$Y = X \cdot b + \varepsilon$$

حيث: ε متغير الخطأ العشوائي.

وإن معادلة الانحدار الخطي المتعدد المقدرة يمكن التعبير عنها بالشكل التالي:

$$\widetilde{Y} = X \cdot b \qquad (30-12)$$

وللحصول على تقديرات لمعاملات الانحدار الجزئية b_0 , b_1 , b_2 والتي تجعل F أصغر ما يمكن نكتب:

$$F = \sum_{i=1}^{n} \left(Yi - \hat{Y}i\right)^2 \Rightarrow Min \qquad (31-12)$$

ويمكن كتابة العلاقة (12- 31) بالشكل التالي:

$$F = \left[Y - X.b\right]' \cdot \left[Y - X.b\right] \qquad (32-12)$$

نلاحظ أن $\left[Y - X.b\right]$ هو عبارة عن مصفوفة عمود ومنقولها $\left[Y - X.b\right]'$ هو عبارة عن مصفوفة سطر. وباشتقاق F بالنسبة لـ b وجعلها تساوي الصفر وبإجراء الإصلاحات الرياضية فإننا نحصل على العلاقة التالية:

$$b = \left[X' \cdot X\right]^{-1} \cdot X'Y \qquad (33)$$

حيث:

b مصفوفة عمود تحوي معاملات الانحدار الجزئية b_0 , b_1 , b_2 . و X' منقول المصفوفة X والتي هي عبارة عن القيم الموافقة للمتغيرات المستقلة مضافاً إليها متغير وهمي كل قيمة من قيمه تساوي الواحد. وY مصفوفة عمود وتضم قيم المتغير التابع.

وبإيجاد نواتج العلاقة (12- 33) نحصل على معاملات الانحدار الجزئية b_0 , b_1 , b_2، وبتعويضها في العلاقة (12- 6) نحصل على معادلة الانحدار الخطي المتعدد المقدرة.

أما بالنسبة لمصفوفة التباينات لمعاملات الانحدار الجزئية فتعطى بالعلاقات التالية:

$$S_{bk}^2 = S_{yx}^2 \cdot [X'X]^{-1} \qquad (34-12)$$

حيث:

$$S_{yx}^2 = \frac{SSE}{n-m-1}$$
$$SSE = \sum d_y^2 \left(1 - R^2\right)$$

ومنه فإن:

$$S_{b0}^2 = S_{yx}^2 \cdot C_{00}$$
$$S_{b1}^2 = S_{yx}^2 \cdot C_{11} \qquad (35-12)$$
$$S_{b2}^2 = S_{yx}^2 \cdot C_{22}$$

حيث: C_{00} , C_{11} , C_{22} هي قيم القطر الرئيسي في المصفوفة $[X' \cdot X]^{-1}$.

ويعطى معامل التحديد المتعدد بالعلاقة التالية:

$$R^2 = \frac{b'[X' \cdot y] - n\overline{Y}^2}{[Y' \cdot Y] - n\overline{Y}^2} \qquad (36-12)$$

مثال 6:

استخدم بيانات المثال رقم (1)، والمطلوب:

1. إيجاد معاملات الانحدار الجزئية.

2 . اكتب معادلة الانحدار الخطي المتعدد المقدرة.

3 . أوجد معامل التحديد ومعامل الارتباط.

4 . أوجد تباينات معاملات الانحدار الجزئية.

الحل:

1 . لإيجاد معاملات الانحدار الجزئية نطبق العلاقة التالية:

$$b=\left[X'\cdot X\right]^{-1}\cdot X'Y \qquad (1)$$

ونحسب $\left[X'\cdot X\right]$ فنجد:

$$\left[X'\cdot X\right]=\begin{bmatrix}1 & 1 & 1 & 1 & ..... & 1 & 1 & 1\\ 100 & 90 & 95 & 110 & ..... & 125 & 140 & 150\\ 3 & 4 & 4 & 5 & ..... & 7 & 8 & 9\end{bmatrix}\cdot\begin{bmatrix}1 & 100 & 3\\ 1 & 90 & 4\\ 1 & 95 & 4\\ 1 & 110 & 5\\ : & : & :\\ 1 & 125 & 7\\ 1 & 140 & 8\\ 1 & 150 & 9\end{bmatrix}=$$

$$=\begin{bmatrix}15 & 1800 & 95\\ 1800 & 220150 & 11725\\ 95 & 11725 & 647\end{bmatrix}$$

لحساب $\left[X'\cdot X\right]^{-1}$ نقوم بحساب قيمة المعين $\left|X'\cdot X\right|$ فنجد أن:

$$\left|X'\cdot X\right|=1237625$$

ثم نقوم بحساب المصفوفة المساعدة $\left[X'\cdot X\right]$ فنجد أن:

$$\left[\widetilde{X}'\cdot X\right]=\begin{bmatrix}4961425 & -50725 & 190750\\ -50725 & 680 & -4875\\ 190750 & -4875 & 62250\end{bmatrix}$$

وبذلك نجد أن:

$$\left[X'\cdot X\right]^{-1}=\frac{1}{\left|X'\cdot X\right|}\cdot\left[X'\cdot X\right]=\frac{1}{1237625}\cdot\left[X'\cdot X\right]$$

وأن:

$$[X' \cdot Y] = \begin{bmatrix} 1200 \\ 146405 \\ 7857 \end{bmatrix}$$

ومنه:

$$b = [x^{\nu} \cdot x]^{-1} \cdot x' \cdot y$$

$$b = \frac{1}{1237625} \begin{bmatrix} 4961425 & -50725 & 190750 \\ -50725 & 680 & -4875 \\ 190750 & -4875 & 62250 \end{bmatrix} \cdot \begin{bmatrix} 1200 \\ 146405 \\ 7857 \end{bmatrix} =$$

$$b = \begin{bmatrix} 21 \\ 0,31 \\ 3,45 \end{bmatrix} = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix}$$

2 . تصبح معادلة الانحدار الخطي المقدرة المتعدد بالشكل التالي:

$$\widetilde{Y} = 21 + 0,31X_1 + 3,45X_2$$

3 . إيجاد معامل التحديد:

$$\overline{Y} = \frac{\sum_{i=1}^{n} Yi}{n} = \frac{1200}{15} = 80$$

نطبق العلاقة (12− 36) فنجد:

$$R^2 = \frac{b'[X' \cdot Y] - n\overline{y}^2}{[Y' \cdot Y] - n\overline{y}^2} = 0,82$$

ومنه فإن معامل الارتباط يساوي:

$$R = 0,91$$

ويعود الاختلاف في الأرقام بسب تقريب قيم معاملات الانحدار الجزئية لذلك سوف نستخدم $R^2 = 0,795$.

4 . تباينات معاملات الانحدار الجزئية وتعطى بالعلاقة:

$$S_{bk}^2 = S_{yx}^2 \cdot [X'X]^{-1} \qquad (2)$$

وأن:

$$S_{yx}^2 = \frac{SSE}{n-m-1}$$
$$SSE = \sum d_y^2 (1 - R^2)$$
$$SSE = 2052(1 - 0{,}795) = 420{,}66$$

ومنه فإن:

$$S_{yx}^2 = \frac{420{,}66}{15-2-1} = 35{,}055$$

وبتطبيق العلاقة (12- 35) نجد أن:

$$S_{b0}^2 = 35{,}055 \cdot \frac{4961425}{1237625} = 140{,}53$$

ومنه:

$$S_{b0} = 11{,}85$$

$$S_{b1}^2 = 35{,}055 \cdot \frac{680}{1237625} = 0{,}0193$$

ومنه:

$$S_{b1} = 0{,}139$$
$$S_{b2}^2 = 35{,}055 \cdot \frac{62250}{1237625} = 1{,}763$$

ومنه:

$$S_{b2} = 1{,}33$$

وهي التباينات نفسها التي حصلنا عليها سابقاً.

تمرينات

1 . إذا كانت معادلة الانحدار الخطي المتعدد من الشكل:

$$\widetilde{Y} = 15 + 12{,}8X_1 + 0{,}08X_2$$

والمطلوب:

1 . تفسير معاملات معادلة انحدار Y على X_2, X_1 .

2 . ما هي قيمة Y المقدرة إذا علمت أن $X_2 = 150, X_1 = 35$.

2 . لإيجاد معادلة الانحدار الخطي المتعدد للتنبؤ عن الأرباح من قيمة المبيعات X_1 والمصاريف X_2 . ثم جمع بيانات عن عشرة محال تجارية حول الأرباح Y والمبيعات X_1 والمصاريف X_2 الشهرية، والموضحة في الجدول التالي:

الأرباح (ألف ليرة)	10	15	17	18	19	21	23	24	28	30
المبيعات(ألف ليرة)	25	27	33	35	39	37	42	42	45	50
المصاريف الإدارية (ألف ليرة)	9	8	6	7	5	4	6	7	5	5

والمطلوب:

1 . تحديد معاملات الانحدار الجزئية.

2 . كتابة معادلة التقدير الخطي المتعدد وتفسير المعادلة الناتجة.

3 . تقدير حجم الأرباح إذا علمت أن $X_1 = 60$ و $X_2 = 9$.

4 . إيجاد معامل التحديد ومعامل الارتباط وتفسير كل منهما.

5 . إيجاد معاملات الارتباط الجزئية وتفسير النتائج.

6 . اختبر معنوية معاملات الانحدار الجزئية عند مستوى دلالة $\alpha = 0{,}05$ ، ثم $\alpha = 0{,}01$ وإيجاد مجال الثقة لكل منها.

7 . اختبر فرضية: $H_0 : \rho_{y\cdot 12} = 0$ عند مستوى دلالة $\alpha = 0{,}05$.

8 . اختبر الفرضيات:

$$H_0 : \gamma_{y1\cdot 2} = 0$$

$$H_0 : \gamma_{y2\cdot 1} = 0$$

عند مستوى دلالة $\alpha = 0{,}05$.

3 . إذا كانت العلاقة بين سعر المنزل Y وبين مساحته X_1 بالمتر المربع و X_2 عمره من الشكل:

$$Y = b_0 + b_1 x_1 + b_2 x_2$$

وحصلنا على البيانات الخاصة بعشرة منازل والمبينة بالجدول التالي:

السعر Y (ألف ليرة)	800	750	900	950	1000	850	900	975	1000	1100
مساحة المنزل X_1 (متر مربع)	100	90	100	110	115	100	100	95	105	110
عمر المنزل X_2	8	9	10	7	8	7	6	5	4	3

والمطلوب:

1 . أوجد معادلة الانحدار الخطي المتعدد والمقدرة بطريقة المصفوفات.

2 . إيجاد معامل التحديد ومعامل الارتباط.

3 . إيجاد تباينات معاملات الانحدار الجزئية.

4 . إجراء كافة الاختبارات الإحصائية المتعلقة بالانحدار المتعدد.