

AL ANDALUS UNIVERSITY
FOR MEDICAL SCIENCES
FACULTY OF DENTISTRY

جامعة الأندلس
الخاصة للعلوم الطبية
كلية طب الأسنان

الإحصاء الحيوي (2) BIOSTATISTICS (II)

السنة الثانية

العام الدراسي 2010-2011

إعداد

أ.د. إبراهيم محمد العلي

مفردات المقرر

Random Variables الفصل التمهيدي: المتحولات العشوائية والتوزيعات الاحتمالية
And Probability Distributions

Estimates and Samples الفصل الأول : العينات و التقديرات الاحصائية

Hypothesis Testing with one الفصل الثاني: اختبار الفرضيات من عينة واحدة
Sample

Inference from Two Samples الفصل الثالث: الاستدلال الاحصائي من عينتين

Correlation and Regression الفصل الرابع: الارتباط والانحدار

Analyse of Variance الفصل الخامس: تحليل التباين الأحادي

Nonparametric Statistics الفصل السادس: الإحصاء اللامعلمي

Multinomial Experiments : الفصل السابع: التجارب المتعددة وجداول التقاطع :
Contingency Tables and

المرجع المعتمد:

Biostatistics for the Biological and Health Sciences, Marc M.
Triola, M. D and Mario F. Triola. Pearson Education. 2006.

المراجع الإضافية:

- مبادئ في علم الاحصاء, العلي ابراهيم, جامعة تشرين, 2004
- الاحصاء التطبيقي. العلي ابراهيم و عكروش محمد, جامعة تشرين, 2006
- الموجز في الاحصاء الحيوي وتصميم التجارب, عثمان أحمد الطاهر, محمد عبد العزيز العبد السلام, جعفر محمد الحسن, دار النهضة العربية, بيروت.
- الطرق الاحصائية, محمد صبحي أبو صالح, دار اليازوري العلمية للنشر والتوزيع, عمان

المحاضرة الأولى والثانية

الفصل التمهيدي

المتحولات العشوائية والتوزيعات الاحتمالية

1-2 المتحولات العشوائية :

يمكننا ببساطة أن نعرف المتحولات العشوائية بأنها هي المتحولات التي تأخذ قيمها الممكنة بصورة عشوائية غير معروفة مسبقاً، وبتكرارات معينة. وهي قد تكون مستمرة أو منقطعة، ولتوضيح ذلك نقدم الأمثلة التالية:

2-2 أمثلة بسيطة:

وزن الطفل عند الولادة: وهو متحول عشوائي مستمر ويقاس بالغرام، ويتحول ضمن مجال معين (2....5)، وهو يتغير من طفل لآخر بصورة عشوائية ويأخذ قيمه الممكنة بطريقة غير معروفة مسبقاً وبتكرارات معينة. ويمكن حساب هذه التكرارات وحساب التكرارات النسبية واستخلاص التوزيع التكراري منها.

عدد أفراد الأسرة: وهو متحول عشوائي منقطع ويقاس بالفرد ويتحول ضمن مجال معين (من 1، 2، 3، 15) ويأخذ قيمه الممكنة بتكرارات معينة ويمكننا حساب هذه التكرارات واستخلاص التوزيع التكراري منها؟

عمر الطفل عند ظهور أسنانه اللبنية: وهو متحول عشوائي مستمر ويقاس بالأيام أو الأسابيع أو الأشهر، وهو يختلف من طفل لآخر ويأخذ قيمه الممكنة ضمن مجال معين بصورة عشوائية وغير معروفة مسبقاً، وبتكرارات معينة. ويمكننا حساب هذه التكرارات واستخلاص التوزيع التكراري منها.

عمر الطالب في الجامعة: وهو متحول عشوائي مستمر يختلف من طالب لآخر ويأخذ قيمه الممكنة بصورة عشوائية وغير معروفة مسبقاً وبتكرارات معينة، ويمكننا حساب هذه التكرارات واستخلاص التوزيع التكراري منها.

نتيجة من الأمثلة السابقة: نلاحظ أنه لا بد أن يكون لكل متحول عشوائي عنصران أساسيان:

. مجموعة من القيم الممكنة المنقطعة أو المستمرة ضمن مجال معين (فضاء).

. توزع تكراري يعبر عن الاحتمالات المقابلة لكل قيمة من القيم الممكنة، ويسمى التوزيع الاحتمالي.

ومن خلال معرفتنا لهذين العنصرين يمكن استخلاص العديد من النتائج وحساب جميع المميزات العددية له: كالمتوسط والتباين والانحراف المعياري ... إلخ.

التوزيع الاحتمالي التجريبي: هو عبارة عن التوزيع التكراري (التكرارات النسبية) المقابلة لجميع القيم الممكنة للمتحول العشوائي المدروس والتي تتميز بأنها غير سالبة وإن مجموعها يساوي الواحد.

وهناك نوعان أساسيان من التوزيعات الاحتمالية هما:

1 . التوزيعات الاحتمالية المنقطعة:

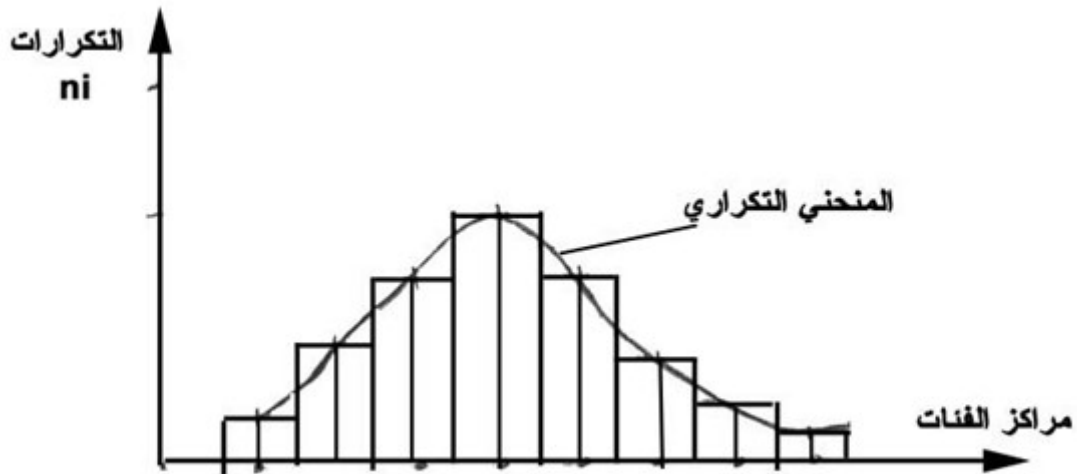
وهي التوزيعات الخاصة بالمتحولات المنقطعة مثل عدد أفراد الأسرة، عدد الطلاب، عدد الضحايا... إلخ، وسوف لن نتعرض لمثل هذه التوزيعات في هذا الفصل.

2 . التوزيعات الاحتمالية المستمرة:

وهي التوزيعات الخاصة بالمتحولات المستمرة أو بالمتحولات التي يمكن اعتبارها مستمرة، مثل درجة الحرارة، دخل الأسرة، درجات الطالب، عمر الطالب... إلخ.

وإذا قمنا بتبويب المعلومات الإحصائية المتوفرة عن أي متحول مستمر X ضمن مجالات محددة فإن التمثيل البياني للترددات النسبية المقابلة لها بواسطة أعمدة تأخذ شكلاً معيناً متناظراً أو مائلاً إلى اليمين أو مائلاً إلى اليسار.

وكمثال على ذلك نأخذ التوزيع التكراري لمستوى الشحوم الثلاثية لدى البالغين والذي يأخذ بعد تبويبه ضمن مجالات محددة شكلاً متناظراً كما في الشكل التالي :

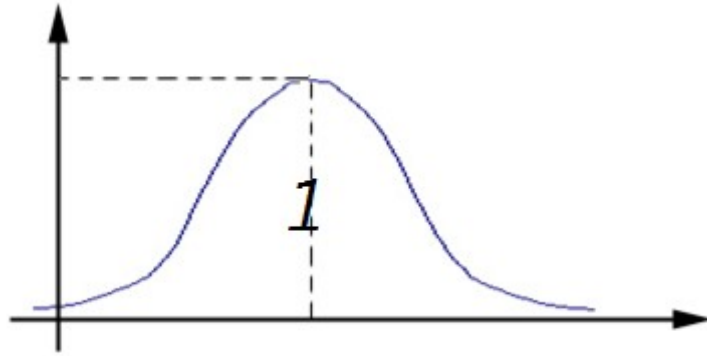


شكل (1-2) نهاية المضلع التكراري

وهنا نلاحظ أن احتمال وقوع المتحول العشوائي X في المجال $[a, b]$ يساوي مساحة العمود الذي يقع فوق ذلك المجال ، وأن مجموع مساحات هذه الأعمدة يساوي الواحد تماماً.

وإذا جعلنا هذه المجالات الجزئية صغيرة جداً وجعلنا حجم العينة n كبيراً فإن أعمدة التوزيع الاحتمالي تصبح على شكل أشرطة عمودية متجاورة .

وإذا قمنا بوصل قمم الأشرطة فإننا سنحصل على منحنٍ انسيابي يسمى منحنى التوزيع الاحتمالي للمتحول العشوائي X ، وهو سيأخذ الشكل الآتي:



شكل (2-2): منحنى التوزيع الاحتمالي

وبما أننا نعتبر المتحول العشوائي X متحولاً مستمراً فإنه يمكننا أن ننتقل إلى مرحلة جديدة ونعتبر عن هذه المنحنيات بواسطة معادلات رياضية معينة وذات متحولات مستمرة.

ولقد توصل الإحصائيون إلى صياغة عدد من المعادلات التي تصلح للتعبير عن الكثير من التوزيعات الاحتمالية. ولكن بما أن هذه المعادلات تتميز ببعض التعقيد قد لا يستوعبه الدارسون في هذه المرحلة، فإننا سنكتفي بإيراد بعض الصيغ الرياضية والأشكال العامة لمنحنيات أهم التوزيعات الاحتمالية.

2-3: بعض قوانين التوزيع الاحتمالية المستمرة :

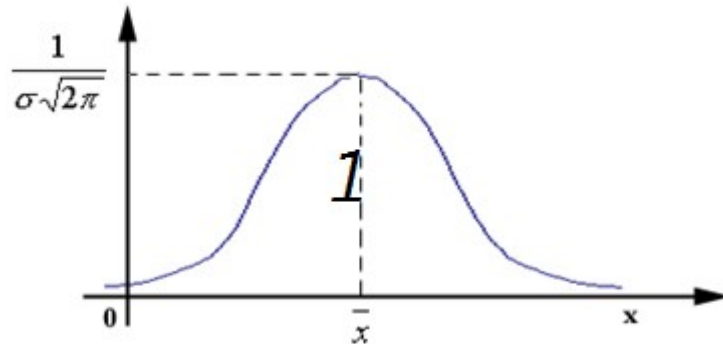
2-3-1: التوزيع الطبيعي العام:

ويعتبر هذا التوزيع أهم التوزيعات الاحتمالية وأكثرها تطبيقاً وانتشاراً، ويصلح لمعظم المتحولات العشوائية، وهو يعرف بالعلاقة التالية:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\bar{x}}{\sigma}\right)^2} \quad -\infty < X < +\infty \quad (8-2)$$

حيث $\bar{x}$ هو المتوسط الحسابي أو التوقع، و σ الانحراف المعياري لـ x .

وهو يرسم المنحنى التالي، والذي يسمى بمنحنى التوزيع الطبيعي العام أو منحنى (غاوص).



الشكل رقم (2-3): الشكل العام للتوزيع الطبيعي

ومن خصائص هذا التوزيع:

. أن مركزه هو المتوسط $\bar{x}$ ، وأن تباينه يساوي σ^2 ، ولذلك يرمز له بالرمز $N(\bar{x}, \sigma^2)$.

. أنه متناظر بالنسبة للمتوسط أو المركز $\bar{x}$.

. أن المساحة تحته تساوي الواحد.

. يقع بكامله فوق محور السينات (المحور الأفقي).

ومن المتحولات الكثيرة التي تخضع لهذا التوزيع نذكر المتحولات التالية:

درجة الطالب في الامتحان: حيث يفترض أنها تتوزع بشكل متناظر حول متوسط معين.

طول المولود أو وزنه عند الولادة: حيث يفترض أنه يتوزع بشكل متناظر حول متوسط معين.

نسبة الذكاء لطلاب مرحلة معينة: حيث يفترض أنها تتوزع بشكل متناظر حول متوسط معين.

مدة حياة الإنسان: حيث يفترض أنها تتوزع بشكل متناظر حول متوسط معين.

حجم المبيعات اليومي في إحدى الصيدليات... إلخ.

ولحساب احتمال أن يأخذ X قيمة أصغر أو تساوي x_1 يجب أن نقوم بإجراء التكامل التراكمي التالي:

$$P(X \leq x_1) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x_1} e^{-\frac{1}{2}\left(\frac{x-\bar{x}}{\sigma}\right)^2} \cdot dx \quad (9-2)$$

ومع أن هذا التوزيع من أكثر التوزيع انتشاراً، إلا أن عملية حساب الاحتمالات فيه اصطدمت بعدة

عوائق رياضية وحسابية. ولقد تم التغلب على هذه العوائق بالاعتماد على حالة خاصة منه تسمى: (قانون

التوزيع الطبيعي المعياري) لأنها تستخدم كمعيار عند حساب الاحتمالات.

2-3-2: التوزيع الطبيعي المعياري:

وهو حالة خاصة من التوزيع الطبيعي العام.

1. أن مركزه ينطبق على مبدأ الإحداثيات، أي أن متوسطه يساوي الصفر $\bar{x} = 0$.

2. أن تباينه $\sigma^2 = 1$ يساوي الواحد.

3. أن المساحة تحته تساوي الواحد.

4. أنه متناظر بالنسبة للمحور العمودي.

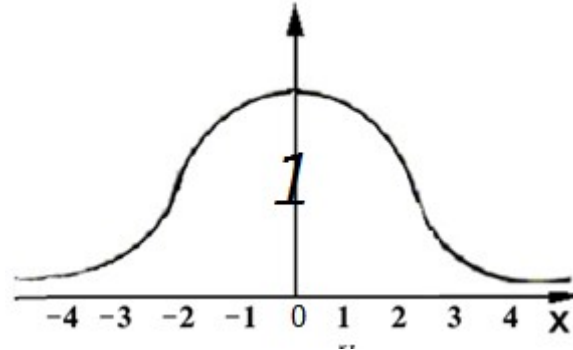
5. لذلك يرمز له بالرمز $N(0,1)$

وتميزاً له عن التوزيع الطبيعي العام نرمز لمتحوله المعياري بالرمز Z وبذلك فإن معادلته تأخذ الشكل

التالي:

$$f(Z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{Z^2}{2}} \quad -\infty < z < \infty \quad (10-2)$$

وهو يرسم على المحاور الإحداثية منحنياً متناظراً بالنسبة للمحور الشاقولي كما هو مبين على الشكل التالي:



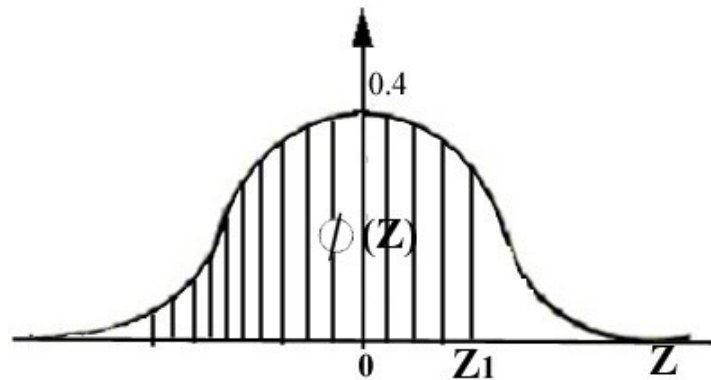
الشكل رقم (2-5): منحنى التوزيع الطبيعي المعياري

وتخضع لهذا التوزيع العديد من المتحولات العشوائية العادية والمتحولات الأخرى الاصطناعية. ويستخدم هذا التوزيع في حساب الاحتمالات الطبيعية الخاصة والعامة. وسنشرح كيفية ذلك بمايلي:

أولاً: كيفية حساب الاحتمالات التراكمية للمتحولات الخاضعة للتوزيع الطبيعي المعياري نفسه:

لحساب هذه الاحتمالات تم إعداد جدول خاص بقيم الاحتمالات التراكمية المختلفة من $(-\infty)$ وحتى قيمة معينة Z_1 للمتحول Z ، وهذه الاحتمالات هي عبارة عن المساحة تحت المنحنى من $(-\infty)$ حتى القيمة المتداولة Z_1 . وهي المساحة المخططة على الشكل (2-6) وسنرمز لها بـ $\Phi(Z)$ ، وبذلك يمكننا أن نحسب الاحتمال التراكمي بواسطة $\Phi(Z)$ ، والذي يسمى بتابع التوزيع الطبيعي المعياري كمايلي:

$$P(Z < z_1) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_1} e^{-\frac{z^2}{2}} dz = \Phi(Z_1) \quad (11-2)$$



الشكل رقم (2-6): تابع التوزيع الطبيعي المعياري $\Phi(Z_1)$

مثال 5: احسب احتمال أن تكون Z أقل من 1.25 أي $P(Z < 1.25)$.

الحل: نلاحظ أن قيمة Z المتداولة تساوي $z_1=1.25$. لذلك نبحث في الجدول (2-1) اللاحق وفي عمود Z عن العدد 1.2 ، ثم نسير في سطره حتى نتقاطع مع العمود الموافق لـ 0.05 فنحصل منه على الرقم 0.8944 المقابل للقيمة المتداولة 1.25. وإن هذا الرقم هو الاحتمال التراكمي المقابل لقيم المتحول التي هي أصغر من 1.25.

3. ونكتب ذلك على الشكل التالي:

$$P(Z < 1.25) = \Phi(1.25) = 0.8944$$

وكذلك نجد على سبيل المثال أن:

$$P(Z < 2.71) = \Phi(2.71) = 0.9966$$

$$P(Z < 0.57) = \Phi(0.57) = 0.7157$$

$$P(Z < 0) = \Phi(0) = 0.500$$

$$P(Z < 1) = \Phi(1) = 0.8413$$

$$P(Z < 2) = \Phi(2) = 0.9772$$

نلاحظ من العلاقة (2-11) أن قيم المتحول Z يمكن أن تكون موجبة أو سالبة، ولكن الجدول (2-1) لا يتضمن غير القيم الموجبة لـ Z . فلماذا ذلك؟. **الجواب:** لأن التوزيع المعياري متناظر، لذلك يمكننا حساب الاحتمالات المقابلة لقيم Z السالبة من جدول قيم الاحتمالات المقابلة لقيم Z الموجبة من الجدول (2-1) ثم طرحها من العدد واحد. وكمثال على ذلك نأخذ القيمة (-1.37) كقيمة متداولة فنجد أن:

$$\begin{aligned} P(Z < -1.37) &= 1 - P(Z < 1.37) = 1 - \Phi(1.37) \\ &= 1 - 0.9147 = 0.0853 \end{aligned}$$

ويمكن التعبير عن ذلك بواسطة العلاقة التالية:

$$P(Z < -z_1) = 1 - P(Z < z_1)$$

$$\Phi(-Z) = 1 - \Phi(Z) \quad (2-12)$$

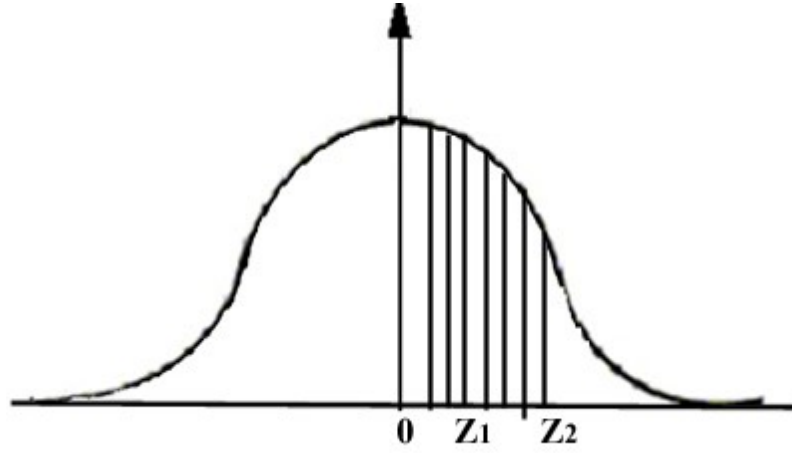
أما لحساب الاحتمال المقابل لمجال جزئي مثل $[z_1, z_2]$ نقوم بطرح الاحتمال التراكمي المقابل لقيمة z_1 الصغيرة من الاحتمال التراكمي المقابل لقيمة z_2 الكبيرة. وكمثال على ذلك نأخذ المجال التالي، [1.28, 2.37] فنجد أن:

$$\begin{aligned} P(1.28 < Z < 2.37) &= P(Z < 2.37) - P(Z < 1.28) \\ &= \Phi(2.37) - \Phi(1.28) = 0.9911 - 0.8997 = 0.0914 \end{aligned}$$

وبصورة عامة يمكننا كتابة ذلك على شكل علاقة عامة كما يلي:

$$P(z_1 < Z < z_2) = \Phi(z_2) - \Phi(z_1) \quad (2-13)$$

وهذا الاحتمال يمثل المساحة تحت المنحني والمحصورة بين Z_1 و Z_2 ، والمبينة على الشكل التالي:

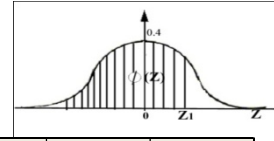


شكل (7.2) الاحتمال على المجال $[z_1, z_2]$

وأخيراً نقدم جدولاً تفصيلياً بقيم تابع التوزيع الطبيعي المعياري $\Phi(z)$ المقابل للقيم الموجبة لـ z والذي يمكن استخدامه لحساب جميع الاحتمالات المختلفة بتطبيق العلاقات والمعادلات السابقة.

جدول (1-2): قيم تابع التوزيع المعياري المعروف على المجال $]-\infty, z_1]$ بالعلاقة:

$$P(Z < z_1) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_1} e^{-\frac{z^2}{2}} dz = \Phi(z_1)$$



Z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91308	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.99900
3.1	0.99903	0.99906	0.99910	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929
3.2	0.99931	0.99934	0.99936	0.99938	0.99940	0.99942	0.99944	0.99946	0.99948	0.99950
3.3	0.99952	0.99953	0.99955	0.99957	0.99958	0.99960	0.99961	0.99962	0.99964	0.99965
3.4	0.99966	0.99968	0.99969	0.99970	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976
3.5	0.99977	0.99978	0.99978	0.99979	0.99980	0.99981	0.99981	0.99982	0.99983	0.99983

ثانياً: حساب الاحتمالات التراكمية للمتحولات الخاضعة للتوزيع الطبيعي العام:

لحساب تلك الاحتمالات ننطلق من نظرية هامة وبسيطة وهي:

نظرية: إذا كان X خاضعاً للتوزيع الطبيعي العام، وكان متوسطه $\bar{x}$ وانحرافه المعياري σ ، فإن المتحول

الاصطناعي Z التالي: $Z = \frac{X - \bar{x}}{\sigma}$ يخضع للتوزيع الطبيعي المعياري $N(0,1)$.

لذلك تُسمى التحويلة $Z = \frac{X - \bar{x}}{\sigma}$ بالدرجة المعيارية أو المتحول المعياري، ويُستفاد من هذه التحويلة في

حساب الاحتمالات العامة، حيث يتم تحويل كل متحول طبيعي عام X إلى المتحول المعياري Z وذلك

بحساب الدرجة المعيارية المقابلة له، ثم استخدام الجدول (2-1) لتابع التوزيع المعياري في حساب

الاحتمالات المطلوبة.

وبشكل عام نجد أنه تكون لدينا عدة حالات لحساب الاحتمالات العامة، هي:

1. حساب الاحتمال العام الذي من الشكل $P(x \leq x_1)$.

لحساب هذا الاحتمال نطرح من طرفي المتراجحة $\bar{x}$ ثم نقسم الناتج على σ فنحصل على ما يلي:

$$P(x \leq x_1) = P\left[\frac{X - \bar{x}}{\sigma} \leq \frac{x_1 - \bar{x}}{\sigma}\right]$$

وباستخدام رمز التحويلة المعيارية السابقة نجد أن:

$$Z = \frac{X - \bar{x}}{\sigma}, \quad Z_1 = \frac{x_1 - \bar{x}}{\sigma}$$

وبناء على ذلك نجد أن الاحتمال السابق:

$$P(x \leq x_1) = P(Z \leq Z_1) = \Phi(Z_1)$$

أي أن حساب الاحتمال السابق يعطى بالعلاقة التالية:

$$P(x \leq x_1) = \Phi\left(\frac{x_1 - \bar{x}}{\sigma}\right) \quad (14 - 2)$$

مثال 6: لنفترض أن X هي درجة الطالب في الامتحان وأن متوسطها $\bar{x} = 65$ وأن انحرافها المعياري σ

$= 14$ ، فأوجد احتمال أن تكون درجة الطالب أقل من 70، وذلك بفرض أن X يخضع للتوزيع الطبيعي

العام.

الحل:

نلاحظ أن X يخضع للتوزيع الطبيعي العام وليس للتوزيع المعياري. ولإيجاد الاحتمال المطلوب

نقوم بتحويل X إلى Z وحساب الدرجة المعيارية المقابلة للقيمة $x_1 = 70$ فنجد أن:

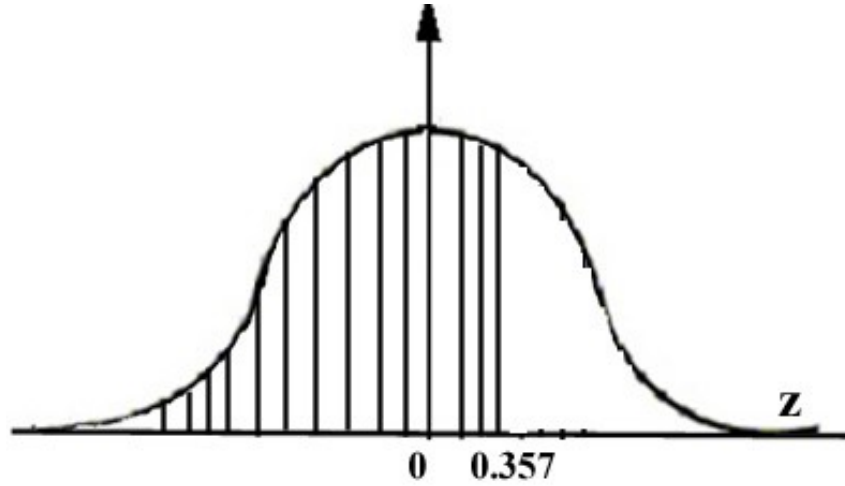
$$Z_1 = \frac{x_1 - \bar{x}}{\sigma} = \frac{70 - 65}{14} = \frac{5}{14} = 0.357$$

ومن التوزيع الطبيعي المعياري نجد أن:

$$P(x < 70) = P(z < 0.357) = \Phi(0.357) = 0.6406$$

وهو الاحتمال المطلوب أي احتمال أن تكون درجة الطالب أقل من 70 يساوي 0.6406 كما هو مبين

على الشكل المخطط التالي:



الشكل (2-8): الاحتمال المطلوب $P(Z < 0.357)$

وهكذا يمكننا أن نحسب جميع الاحتمالات المختلفة الأخرى المشابهة لذلك.

2. حساب الاحتمال العام في مجال محدد $[x_1, x_2]$ من الشكل $P(x_1 \leq X \leq x_2)$:
لحساب هذا الاحتمال نستخدم الأسلوب نفسه فنجد أن:

$$P(x_1 \leq X \leq x_2) = P\left[\frac{x_1 - \bar{X}}{\sigma} \leq \frac{\bar{X} - X}{\sigma} \leq \frac{x_2 - \bar{X}}{\sigma}\right]$$

$$P(x_1 \leq X \leq x_2) = P[z_1 \leq Z \leq z_2] = \Phi(z_2) - \Phi(z_1)$$

$$P(x_1 \leq X \leq x_2) = \Phi\left(\frac{x_2 - \bar{X}}{\sigma}\right) - \Phi\left(\frac{x_1 - \bar{X}}{\sigma}\right) \quad (2-15)$$

مثال 7:

لنأخذ معطيات المثال السابق ونحسب احتمال أن تكون درجة الطالب واقعة في المجال $[55, 80]$ ،
لذلك نقوم في البداية بحساب الدرجة المعيارية المقابلة لكل من طرفي المجال $[x_1, x_2]$ فنجد أن:

$$Z_1 = \frac{x_1 - \bar{X}}{\sigma} = \frac{55 - 65}{14} = -0.714$$

$$Z_2 = \frac{x_2 - \bar{X}}{\sigma} = \frac{80 - 65}{14} = 1.07$$

ونحصل على الاحتمال المطلوب باستخدام الدرجتين المعياريتين لطرفي المجال فنجد أن:

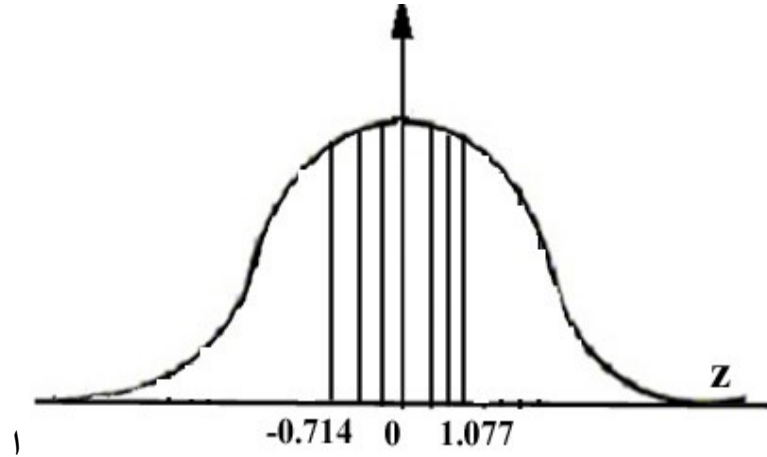
$$P(x_1 \leq X \leq x_2) = \Phi(z_2) - \Phi(z_1)$$

ومن جدول تابع التوزيع المعياري نجد أن:

$$P(55 < X < 80) = P(-0.714 < Z < 1.074) = \Phi(1.074) - \Phi(-0.714)$$

$$= 0.8577 - 0.2389 = 0.6188$$

وهو يمثل المساحة الواقعة فوق المجال $[-0.714, 1.0747]$ للتوزيع الطبيعي المعياري المبين في الشكل (9-2).



الشكل (9-2)

وبصورة عامة يمكننا أن نطبق العلاقة (2-15) السابقة مباشرة لحساب الاحتمالات العامة وذلك باستخدام الصيغة التالية:

$$P(x_1 \leq X \leq x_2) = \Phi\left(\frac{x_2 - \bar{X}}{\sigma}\right) - \Phi\left(\frac{x_1 - \bar{X}}{\sigma}\right)$$

$$P(55 \leq X \leq 80) = \Phi\left(\frac{80-65}{14}\right) - \Phi\left(\frac{55-65}{14}\right)$$

$$= \Phi(1.071) - \Phi(-0.714)$$

$$= \Phi(1.071) - [1 - \Phi(-0.714)]$$

$$= 0.8577 - (1 - 0.7611)$$

$$P(55 \leq X \leq 80) = 0.6188$$

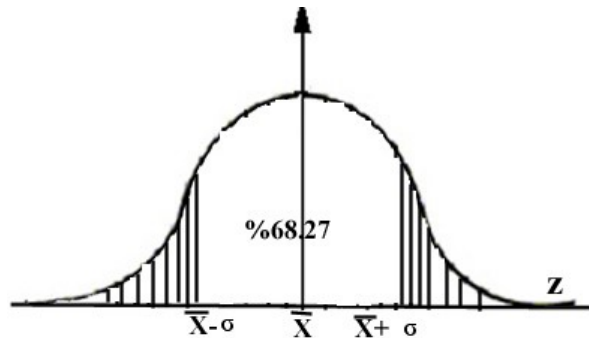
حيث نجد أن:

3 . حساب الاحتمالات المقابلة لمجالات الثقة الأساسية، وهي:

• الاحتمال المقابل لمجال الثقة الأول: وهو الاحتمال المقابل للمجال الذي مركزه المتوسط $\bar{X}$

وطرفاه محددين بالعديدين $\bar{X} + \sigma$ و $\bar{X} - \sigma$ ، أي المحددين بجمع وطرح الانحراف المعياري من

المتوسط $\bar{X}$. كما هو مبين على الشكل (2-10) التالي:



الشكل (2-10) مجال الثقة الأول

إن الاحتمال المطلوب هو:

$$P(\bar{X} - \sigma \leq X \leq \bar{X} + \sigma)$$

لحساب هذا الاحتمال نقوم بإيجاد الدرجة المعيارية المقابلة لكل طرفي المجال كما يلي:

$$z_1 = \frac{(\bar{X} - \sigma) - \bar{X}}{\sigma} = -1$$

$$z_2 = \frac{(\bar{X} + \sigma) - \bar{X}}{\sigma} = +1$$

ومن الجدول المعياري نجد أن الاحتمال المطلوب يساوي:

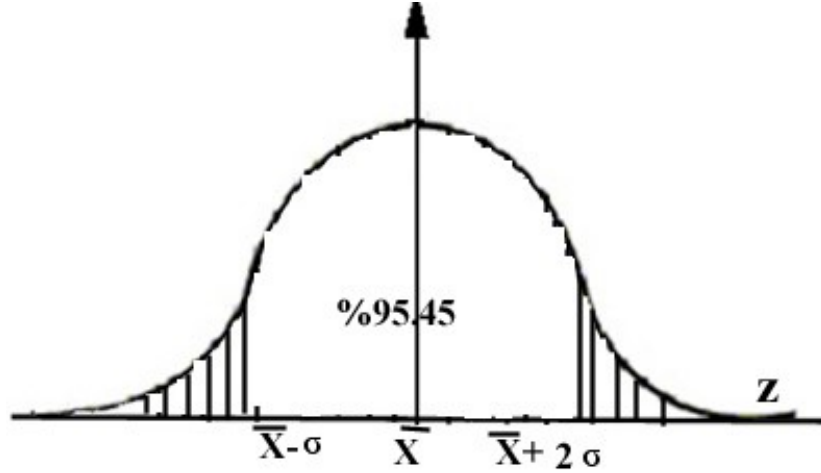
$$\begin{aligned} P(\bar{X} - \sigma \leq X \leq \bar{X} + \sigma) &= \Phi(z_2) - \Phi(z_1) = \Phi(1) - \Phi(-1) \\ &= 0.8413 - (1 - 0.8413) = 0.6826 \end{aligned}$$

وهو يساوي المساحة غير المخططة فوق المجال الأول والمبنية على الشكل (2-10)، وهذا يعني احتمال أن يقع X في ذلك المجال يساوي 0.6826، أي أن 68% فقط من قيمه تقع في ذلك المجال.

• الاحتمال المقابل لمجال الثقة الثاني: وهو الاحتمال المقابل للمجال الذي مركزه المتوسط $\bar{X}$

وطرفاه محددين بالعديدين $\bar{X} + 2\sigma$ و $\bar{X} - 2\sigma$ ، أي المحددين بجمع وطرح ضعفي الانحراف

المعياري إلى المتوسط $\bar{X}$. كما هو مبين على الشكل (2-11) التالي:



الشكل (2-11) مجال الثقة الثاني

إن الاحتمال المطلوب هو:

$$P(\bar{X} - 2\sigma \leq X \leq \bar{X} + 2\sigma)$$

لحساب هذا الاحتمال نقوم بإيجاد الدرجة المعيارية المقابلة لكل من طرفي المجال كما يلي:

$$z_1 = \frac{(\bar{X} - 2\sigma) - \bar{X}}{\sigma} = -2$$

$$z_2 = \frac{(\bar{X} + 2\sigma) - \bar{X}}{\sigma} = +2$$

وبذلك نجد أن الاحتمالات المعيارية تساوي:

$$P(\bar{x} - 2\sigma \leq X \leq \bar{x} + 2\sigma) = \Phi(z_2) - \Phi(z_1) = \Phi(2) - \Phi(-2)$$

ومن جدول تابع التوزيع المعياري نجد أن الاحتمال المطلوب يساوي:

$$P(\bar{x} - 2\sigma \leq X \leq \bar{x} + 2\sigma) = 0.9772 - (1 - 0.9772) = 0.9544$$

وهو يساوي المساحة غير المخططة فوق المجال الثاني والمبنية على الشكل (2-11) أعلاه، وهذا يعني احتمال أن يقع X في المجال الثاني يساوي 0.9544، أي أن حوالي 95% فقط من قيمه الممكنة تقع في ذلك المجال.

• الاحتمال المقابل لمجال الثقة الثالث: وهو $P(\bar{x} - 3\sigma \leq X \leq \bar{x} + 3\sigma)$

بطريقة مشابهة نجد أن احتمال وقوع X في المجال الثالث للثقة يساوي:

$$P(\bar{x} - 3\sigma \leq X \leq \bar{x} + 3\sigma) = 0.9973 \quad (2-14)$$

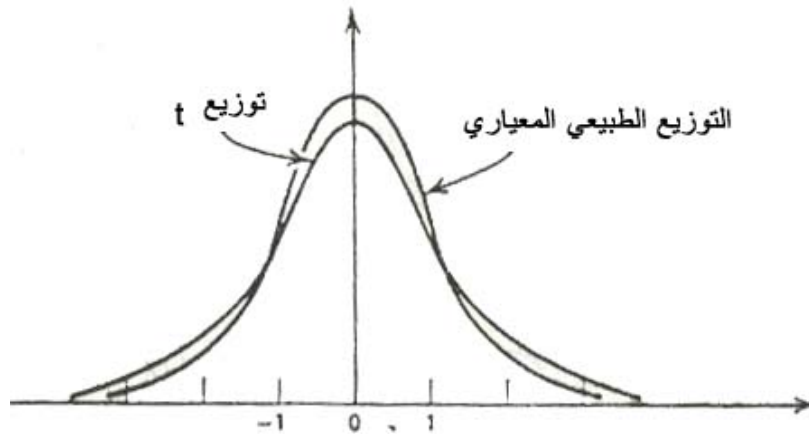
ملاحظة: هناك تطبيقات هامة كثيرة لهذه المجالات ولاحتماالاتها المقابلة لذلك يجب على الطلاب إعطاءها اهتماماً خاصاً.

(2-3-3) توزيع ستودينت (توزيع t):

وهو يستخدم بدلاً عن التوزيع الطبيعي المعياري عندما يكون حجم العينة صغيراً، (أي عندما $n \leq 30$ ، وأن الشكل العام لمنحني هذا التوزيع يشبه منحني التوزيع الطبيعي المعياري فهو متناظر بالنسبة لمبدأ الإحداثيات (أي، مركزه صفر) ولكن انحرافه المعياري يختلف عن الواحد. وأن صيغة معادلته الرياضية هي:

$$f(t) = C \left(1 + \frac{t^2}{k} \right)^{-\frac{k+1}{2}} \quad (2-15)$$

حيث $-\infty < t < +\infty$ ويسمى العدد k درجة الحرية، و C ثابت عددي موجب ومعين ويأخذ شكله البياني الشكل الآتي:



الشكل رقم (2-12) منحني توزيع ستودينت

وغالباً ما يرمز للمتغير العشوائي الخاضع لتوزيع ستوديننت بالرمز t ، ويُستخدم هذا التوزيع في اختبارات الفرضيات المختلفة عندما تكون العينات صغيرة الحجم (حجم العينة n أقل من 30). وعندما يقوم الباحث بحساب قيمة مؤشر الاختبار للفرض الذي يختبره فإنه يحصل على قيمة عددية نرمز لها بـ t . ثم نقوم بمقارنة هذه القيمة مع القيمة الحرجة t_0 المقابلة لمستوى الدلالة ودرجة الحرية، (ونحصل على قيم t_0 من جداول توزيع t الملحق)، فإذا كانت القيمة المطلقة لـ t المحسوبة أصغر من القيمة الحرجة t_0 ، نقبل فرضية العدم الموضوعة، وإذا كان العكس نرفض فرضية العدم الموضوعة ونقول بعكسها لأننا لا نملك دليلاً كافياً على إثباتها.

وسنتطرق إلى هذا الموضوع في فقرة اختبار الفرضيات الواردة لاحقاً.

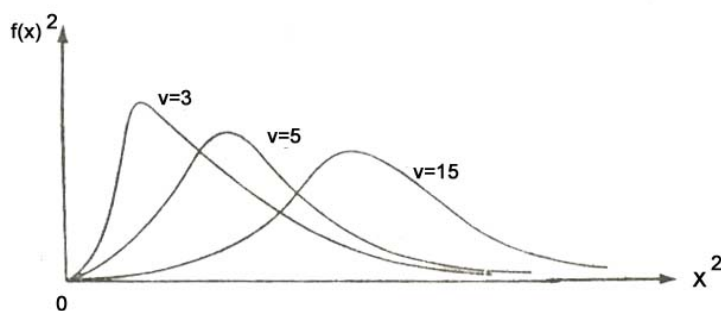
(2-3-4): توزيع كاي مربع χ^2 :

وهو توزيع احتمالي خاص يُستخدم لاختبار مدى تطابق المعطيات الميدانية أو التجريبية مع المعطيات الفرضية أو الضابطة. ولتطبيقه تُستخدم علاقة رياضية معينة لحساب قيمة مؤشر الاختبار χ^2 ثم مقارنتها مع القيمة الحرجة χ_0^2 المقابلة لمستوى الدلالة α ودرجة الحرية المحددة k ، ويتم الحصول على χ_0^2 من جداول خاصة بالتوزيع χ^2 . فإذا كانت قيمة χ^2 المحسوبة أصغر من χ_0^2 الحرجة نقبل فرضية العدم الموضوعة، كما هو موضح على الشكل التالي، وفي حالة العكس نرفض تلك الفرضية لعدم وجود دليل كافٍ على إثباتها.

أما الصيغة الرياضية لتوزيع χ^2 فهي كما يلي:

$$f(x^2) = C \cdot (x^2)^{\frac{k}{2}-1} e^{-\frac{x^2}{2}} : x^2 > 0 \quad (2-16)$$

حيث k درجة الحرية، و C ثابت عددي موجب ومعين ويأخذ شكله البياني الشكل الآتي:



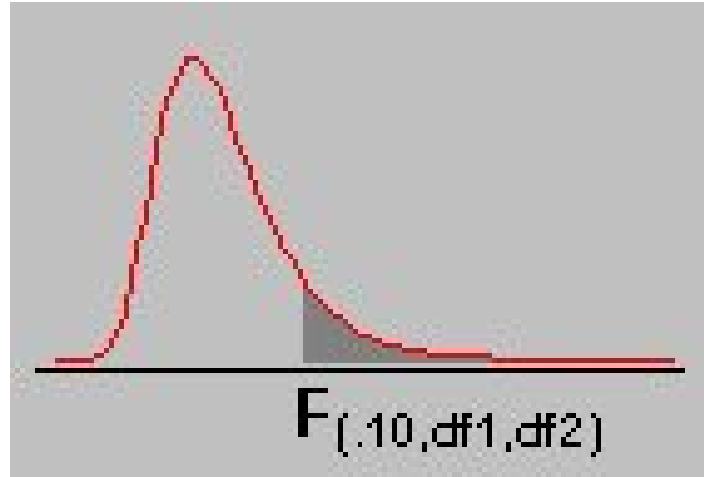
الشكل (2-13) منحنى توزيع χ^2

(2-3-5): توزيع فيشر أو توزيع F:

وهو توزيع احتمالي خاص ومعقد يستخدم لاختبارات التجانس وتحليل التباين ضمن المعطيات وبين العينات، ولتطبيقه تُستخدم علاقة رياضية معينة لحساب قيمة مؤشر الاختبار F ثم مقارنة هذه القيمة مع القيمة الحرجة F_0 المقابلة لمستوى الدلالة α ولدرجتي الحرية في العينتين. فإذا كانت قيمة F المحسوبة أصغر من F_0 الحرجة نقبل فرضية العدم الموضوعة، وفي حالة العكس نرفض تلك الفرضية لعدم وجود دليل كافٍ على إثباتها. أما معادلته الرياضية فهي كما يلي:

$$f(F) = C \cdot \frac{F^{\frac{k}{2}-1}}{k_2 + k_1 F^{\frac{k_2+k_1}{2}}} : F > 0 \quad (17-2)$$

حيث k_1 درجة الحرية للتباين الكبير الأول و k_2 درجة الحرية للتباين الثاني، و C ثابت عددي موجب ومعين وهو من الشكل الآتي:



الشكل (2-14) منحنى توزيع F

تمريعات

1. إذا كان Z خاضعاً للتوزيع الطبيعي المعياري فاحسب الاحتمالات الآتية:
 $P(-1 \leq Z \leq +1)$ ، $P(-2 \leq Z \leq +2)$
 $P(-2.58 \leq Z \leq +2.58)$ ، $P(-3 \leq Z \leq +3)$
2. إذا كانت X قيمة المبيعات اليومية في إحدى الصيدليات تخضع للتوزيع الطبيعي وكان متوسطها $\bar{X} = 3000$ وتباينها $\sigma^2 = 14900$ والمطلوب:
. تحديد الصيغة الرياضية لقانون التوزيع الاحتمالي لـ X .
. حساب احتمال أن تتجاوز قيمة المبيعات اليومية 4000 ل.س، أي $P(X > 4000)$
. حساب الاحتمال التالي: $P(2000 \leq X \leq 4000)$
3. إذا كان وزن الطفل عند الولادة X يخضع للتوزيع الطبيعي العام الذي متوسطه $\bar{x} = 3$ kg وتباينه $\sigma^2 = 0.64$ فالمطلوب:
. تحديد الصيغة الرياضية لقانون التوزيع الاحتمالي لـ X .
. حساب الاحتمالات التالية:
 $P(2 \leq X \leq 4)$ ، $P(X \geq 3)$ ، $P(X \leq 2)$ ، $P(X \geq 4)$
4. إذا كان X هو مقدار الدخل اليومي لعيادة أحد الأطباء وكان خاضعاً للتوزيع الطبيعي الذي متوسطه $\bar{x} = 4000$ وتباينه $\sigma^2 = 490000$ فالمطلوب:
. تحديد الصيغة الرياضية لقانون التوزيع الاحتمالي للمتحول لـ X .
. حساب الاحتمالات التالية:
 $P(X \leq 5000)$ ، $P(X \leq 5000)$ ، $P(3000 \leq X \leq 5000)$
 $P(2000 \leq X \leq 5000)$
5. إذا كان X هو طول الشخص البالغ وكان خاضعاً للتوزيع الطبيعي العام الذي متوسطه $\bar{x} = 170$ وتباينه $\sigma^2 = 25$ فالمطلوب:
. تحديد الصيغة الرياضية للطول لـ X .
. حساب الاحتمالات التالية:
 $P(X \geq 180)$ ، $P(160 \leq X \leq 180)$ ، $P(X \leq 160)$
 $P(\bar{x} - 2\sigma \leq X \leq \bar{x} + 2\sigma)$

المحاضرة الثالثة

الفصل الأول

العينات والتقديرات الاحصائية

Estimates and Samples

مفردات البحث

- 1- تقدير متوسط المجتمع
- 2- تقدير التباين والانحراف المعياري للمجتمع
- 3- تقدير الخطأ المعياري لمتوسط المجتمع
- 4- تقدير نسبة خاصة في المجتمع
- 5- تقدير الفرق بين متوسطي مجتمعين طبيعيين
- 6- التقدير المجالي
- 7- تقدير حجم العينة

المرجع المعتمد: المرجع اسابق من الصفحة 254 إلى الصفحة 313

تمهيد:

تعتبر طريقة العينات من الطرائق الإحصائية المستخدمة في عمليات جمع المعلومات الإحصائية وتحليلها، وذلك لأنها توفر كثيراً من المال والجهد والوقت في الحصول على المعلومات اللازمة. وتعد طريقة العينات في كثير من الأحيان الأسلوب الإحصائي الوحيد الذي يمكن تطبيقه في الحصول على المعلومات، كما في حالة تحليل الدم أو تحليل التربة، أو تحليل المياه، أو مراقبة الإنتاج، أو تقدير صلاحية الأغذية أو الأدوية...إلخ.

(1-1).أنواع العينات:

تصنّف العينات حسب تصميم عمليات المعاينة , والذي يعتمد على طبيعة المجتمع المدروس وعلى هدف الدراسة... وأنواعها هي:

- 1 . **المعاينة العشوائية البسيطة:** وهي التي يتم فيها سحب عينة عشوائية من عناصر المجتمع المدروس وباحتمالات متساوية وتطبق على المجتمعات المتجانسة.
 - 2 . **المعاينة العشوائية الطبقية:** وهي التي يتم فيها تقسيم المجتمع المدروس غير المتجانس إلى طبقات متجانسة تم تطبيق المعاينة البسيطة على كل طبقة على حده، ثم يتم تركيب النتائج وإجراء الحسابات اللازمة.
 - 3 . **المعاينة العنقودية:** وهي التي يتم فيها سحب العينات العنقودية على مرحلتين أو أكثر: تُسحب عينة فروع من الجذع ثم عينات عناصر من الفروع (عينات من المدن، ثم عينات من الأسر في كل مدينة مسحوبة).
 - 4 . **المعاينة المنتظمة :** وهي التي يتم فيها سحب عناصر العينة بطريقة عددية منتظمة, كأن نسحب عنصراً بعد مرور كل k عنصراً.
- اما خطوات البحث المتبعة في طريقة العينات فهي نفسها الخطوات المتبعة في حالة المسح الشامل. غير أن طريقة العينات تحتاج إلى خطوة إضافية وهامة هي خطوة سحب العينة، وهو ما سنعالجه فيما يأتي.

كيفية سحب العينة العشوائية البسيطة:

يُشترك في سحب العينة أن يكون سحباً عشوائياً ولذلك نقوم بما يلي:

- 1 . نحصر عناصر المجتمع المدروس ونحدد عددها N .
 - 2 . نعطي كل عنصر رقماً متسلسلاً من 1 حتى N .
 - 3 . نضع أرقام هذه العناصر في صندوق خاص ونخلطها جيداً.
 - 4 . نقوم بسحب الأرقام من الصندوق بطريقة عشوائية ونخلط الأرقام قبل كل عملية سحب.
- أما السحب بحد ذاته فيمكن أن يكون على شكلين هما:
- . السحب مع الإعادة: وفيه يُعاد الرقم المسحوب إلى الصندوق ليتعرض للسحب مرةً ثانية أو ثالثة...الخ.
 - . السحب بدون إعادة، وفيه لا يُعاد الرقم المسحوب إلى الصندوق وبذلك لا يتعرض للسحب مرةً ثانية.
- ويمكننا سحب الأرقام عشوائياً بواسطة استخدام عجلات الرهان، أو بواسطة جداول الأرقام العشوائية، أو بواسطة علاقات رياضية مبرمجة على الآلات الحاسبة بمفتاح Run. وسنقتصر في بحثنا على العينات الكبيرة المسحوبة مع الإعادة لسهولة التعامل معها.

5. نعتبر الأرقام المسحوبة في الخطوة السابقة هي أرقام العناصر التي ستدخل في العينة، ونبحث عنها في المجتمع لدراستها واستخلاص المعلومات اللازمة منها.

إن عملية السحب العشوائي لعناصر العينة تضمن لنا عدم التحيز في الانتقاء، كما تؤمن تمثيلاً جيداً لجميع عناصر المجتمع المدروس.

أما بالنسبة لحجم العينة اللازم سحبه في دراسة معينة فيتوقف على مقدار تجانس المجتمع بالنسبة للخاصة المدروسة، وعلى مقدار الدقة المطلوبة d وعلى معامل الثقة المفروض Z وعلى مقادير أخرى كالتكلفة وطريقة السحب. وسنعمد في هذا الكتاب على أبسط العلاقات الرياضية التي تعطينا حجم العينة n باحتمال ثقة قدره (0.95) وهي:

$$n = \frac{4\sigma^2}{d^2} \quad (1-3)$$

حيث أن σ^2 هو تباين الخاصة المدروسة في المجتمع ويتم تقديره من خلال تباين عينة اختبارية. وأن d هي الدقة المطلوبة في التقدير وتُحدد مسبقاً من قبل المسؤولين عن البحث.

وقبل أن نبحث في قضايا التقدير نورد بعض التعاريف المتعلقة بالمجتمع والعينة والمؤشرات المستخدمة في تقدير معالم المجتمع.

(2-3) تعريف المجتمع الإحصائي والعينة:

المجتمع الإحصائي : هو جملة العناصر التي تكون مستهدفة بالدراسة وسنرمز لعدد تلك العناصر بـ N ، وللخاصة المستهدفة بالدراسة فيه بـ Y ولقيمتها المختلفة بالرموز.

$Y: y_1, y_2, y_3, y_4, \dots, y_i, \dots, y_n$

ونرمز لمتوسط هذه القيم في المجتمع بالرمز $\bar{y}$ والذي يساوي:

$$\bar{y} = \frac{1}{N} \sum_{i=1}^N y_i$$

ولتباين تلك القيم بالرمز σ^2 ساوي:

$$\sigma^2 = \frac{1}{N} \cdot \sum_{i=1}^N (y_i - \bar{y})^2$$

ولنسبة وجود تلك الخاصة في المجتمع بالرمز R وتساوي:

$$R = \frac{M}{N}$$

حيث M عدد المتميزين بتلك الخاصة.

إن هذه القيم ومتوسطها وتباينها ونسبتها تكون مجهولة في المجتمع، لذلك نسحب عينة من المجتمع بصورة عشوائية، وسنستخدم معلومات هذه العينة لتقدير معالم المجتمع كالمتوسط $\bar{y}$ والتباين σ^2 والنسبة R ... إلخ.

العينة: هي جزء من المجتمع مسحوب عشوائياً منه ومؤلف من عدد العناصر: وليكن n عنصراً. وإذا رمزنا للمعلومات التي تقدمها العينة بالرمز X ولقيمها بالرموز:

$$X: x_1, x_2, x_3, x_4, \dots, x_n$$

ولنفترض أن متوسطها كان يساوي:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (5-3)$$

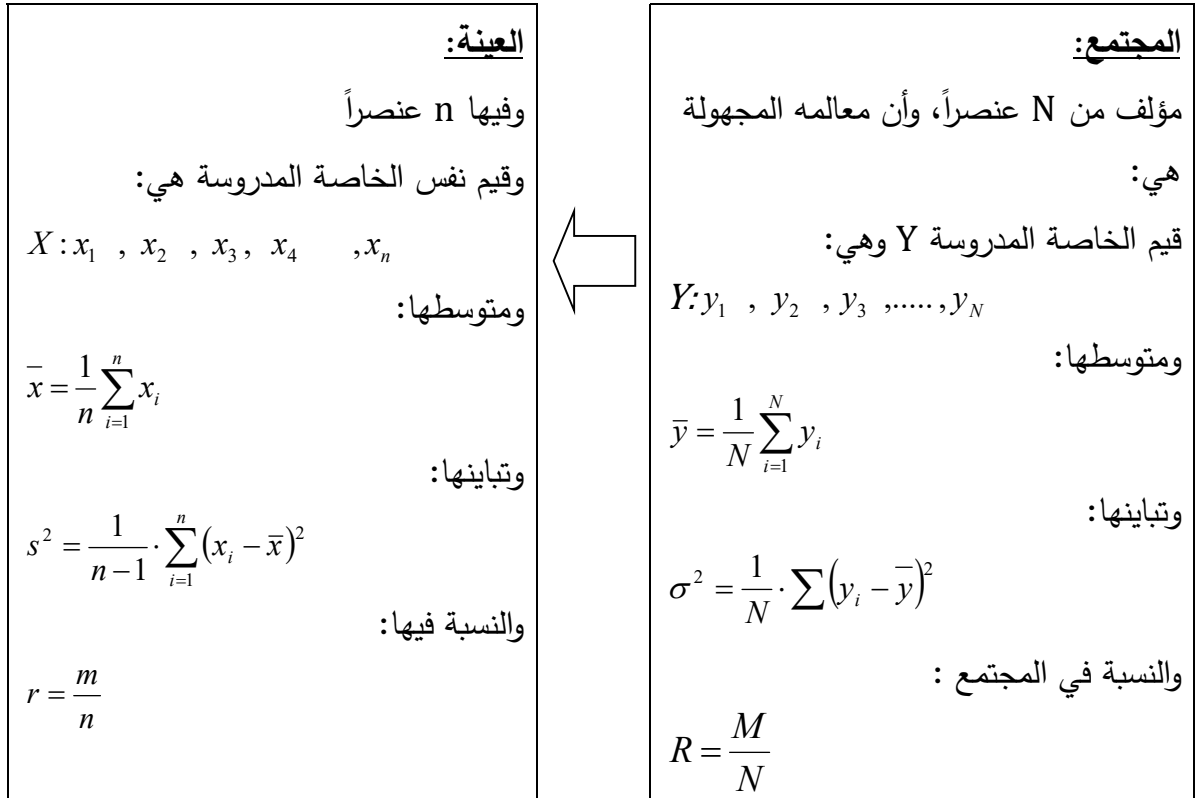
وأن تباينها المصحح كان يساوي:

$$s^2 = \frac{1}{n-1} \cdot \sum (x_i - \bar{x})^2 \quad (6-3)$$

وأن النسبة في العينة كانت تساوي:

$$r = \frac{m}{n} \quad (7-3)$$

ولتوضيح العلاقة بين المجتمع والعينة المسحوبة منه نعرض الشكل البياني التالي:



الشكل (1-11) : العلاقة بين المجتمع والعينة ومعالميهما

وسنرى كيف يمكن استخدام مؤشرات العينة لتقدير معالم المجتمع من خلال الفقرات التالية:

1.1. تقدير متوسط المجتمع

يعرف متوسط المجتمع (ونرمزه $\bar{y}$ أو μ) بأنه حاصل قسمة مجموع عناصره على عددها أي :

$$\bar{y} = \frac{\sum_{i=1}^N y_i}{N}$$

حيث أن N هو عدد عناصر المجتمع و y_i هو العنصر الذي يكون ترتيبه i .

و باعتبار أن متوسط المجتمع غالبا ما يكون مجهول لذلك نقدره من خلال عينة مسحوبة من هذا المجتمع. فلو فرضنا أننا سحبنا عينة من عناصر هذا المجتمع بحجم n عنصر وكانت هذه العناصر هي التالية:

$x_1, x_2, x_3, \dots, x_n$. نستطيع تقدير متوسط المجتمع عن طريق متوسط العينة $\bar{x}$ كالتالي:

$$\tilde{y} = \bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

وهو تقدير غير متحيز وفعال ومتماسك ويعتمد عليه في جميع البحوث والدراسات العلمية.

مثال 1:

لتقدير متوسط سن الزواج عند الشباب السوري في إحدى القرى قمنا بسحب عينة عشوائية من الرجال المتزوجين في هذه القرية مؤلفة من 10 رجال فوجدنا أن أعمارهم عند الزواج الأول كالتالي:

$x_i : 20, 22, 28, 30, 26, 25, 40, 32, 33, 34$ في هذه الحالة نستطيع تقدير متوسط سن الزواج عند الشباب في هذه القرية عن طريق متوسط العينة كما يلي:

$$\tilde{y} = \bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{20 + 22 + \dots + 34}{10} = \frac{290}{10} = 29$$

2.1. تقدير التباين والانحراف المعياري

يعرف تباين المجتمع (ونرمزه σ^2) على الشكل التالي :

$$\sigma^2 = \frac{\sum_{i=1}^N (y_i - \bar{y})^2}{N}$$

وانحرافه المعياري كما يلي:

$$\sigma = \sqrt{\sigma^2}$$

باعتبار أن تباين المجتمع غالبا ما يكون مجهول كما معظم معالم المجتمع نستطيع تقديره من خلال عينة مسحوبة من هذا المجتمع. فلو فرضنا أننا سحبنا عينة من عناصر هذا المجتمع بحجم n عنصر وكانت هذه العناصر هي التالية: $x_1, x_2, x_3, \dots, x_n$. نستطيع تقدير تباين المجمع وانحرافه المعياري عن طريق تباين العينة (S^2) وانحرافها المعياري (S) كما يلي:

$$\tilde{\sigma}^2 = S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

وهو تقدير غير متحيز لتباين المجتمع. أما تقدير الانحراف المعياري للمجتمع فيحسب كالتالي:

$$\tilde{\sigma} = S = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

مثال 2:

بالعودة إلى معطيات المثال الأول , المطلوب تقدير تباين سن الزواج في المجتمع وتقدير انحرافه المعياري.

$$\tilde{\sigma}^2 = S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} = \frac{(20-29)^2 + (22-29)^2 + \dots + (34-29)^2}{10-1} = \frac{328}{9} = 36.44$$

وبذلك نجد أن تباين سن الزواج في هذا المجتمع يقدر بما يلي:

$$\tilde{\sigma}^2 = S^2 = 36.44$$

وأن الانحراف المعياري للعمر في المجتمع يقدر بما يلي:

$$\tilde{\sigma} = S = \sqrt{36.44} = 6.03$$

مثال 3:

إذا اعتبرنا أن X متحول عشوائي يدل على درجات الكولسترول في الدم. وفي عينة مؤلفة من أربعين نزيلا في إحدى المستشفيات تم قياس مستوى الكولسترول فجاءت النتائج كالتالي:

المجموع	70	60	54	50	48	46	42	40	35	32	30	مستوى الكولسترول X_i
40	1	1	2	5	10	8	6	3	2	1	1	عدد المرضى n_i

المطلوب تقدير تباين مستوى الكولستترول وانحرافه المعياري.

نحسب بداية متوسط العينة:

$$\bar{x} = \frac{\sum n_i x_i}{\sum n_i} = \frac{1840}{40} = 46$$

نقوم بعد ذلك بتقدير التباين

$$\tilde{\sigma}^2 = S^2 = \frac{\sum_{i=1}^n n_i (x_i - \bar{x})^2}{n-1} = \frac{1 \times (30-46)^2 + \dots + 1 \times (70-46)^2}{40-1} = \frac{1918}{39} = 49.17$$

ويكون مقدار الانحراف المعياري للمجتمع (الانحراف المعياري للعينة) مساوي :

$$\tilde{\sigma} = S = \sqrt{49.17} = 7.01$$

3.1. تقدير الخطأ المعياري لتقدير متوسط المجتمع

إن تقدير متوسط المجتمع بواسطة متوسط العينة $\bar{x}$ يختلف من عينة لأخرى. وبما أنه يمكننا أن نسحب عدداً كبيراً من العينات، فإن ذلك يعطينا عدداً كبيراً من التقديرات لمتوسط المجتمع $\bar{y}$ وأن هذه التقديرات تختلف عن بعضها البعض، ويكون لها متوسط وتباين وانحراف معياري. فكيف نحسب أو نقدر الانحراف المعياري للتقدير الذي استخدمناه في تقدير متوسط المجتمع $\bar{y}$ ؟

يعطى تباين تقدير متوسط المجتمع بالعلاقة التالية:

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$$

ولكن بما أن تباين المجتمع σ^2 مجهول كما ذكرنا سابقاً فإننا نقدره عن طريق تباين العينة كالتالي:

$$\tilde{\sigma}_{\bar{x}}^2 = \frac{S^2}{n}$$

ونسمي متوسط الخطأ المرتكب من جراء تقدير متوسط المجتمع بواسطة العينة، يسمى بالخطأ المعياري أو بالانحراف المعياري لمتوسط العينة $\bar{x}$ ويرمز له بـ $\tilde{\sigma}_{\bar{x}}$ ، وهو يقدر بواسطة العلاقة التالية:

$$\tilde{\sigma}_{\bar{x}} = \frac{S}{\sqrt{n}}$$

ويعتمد على هذان التباينان بشكل كبير في اختبار الفرضيات.

مثال 4: من بيانات المثال السابق (مثال 3) يمكننا تقدير الخطأ المعياري لمتوسط الكولسترول كالتالي :

$$\tilde{\sigma}_{\bar{x}} = \frac{S}{\sqrt{n}} = \frac{7.01}{\sqrt{40}} = 1.1$$

أي أن متوسط الكولسترول في المجتمع يقدر بـ 46، وبخطأ معياري قدره 1.1. أو بمعنى آخر أن الخطأ المعياري الذي وقع عند تقدير متوسط الكولسترول يقدر بـ 1.1.

4.1. تقدير نسبة خاصة معينة في المجتمع

لنفترض أننا نريد إيجاد نسبة الأفراد الذين يتمتعون بخاصة ما في مجتمع. يمكننا حساب هذه النسبة كالتالي :

$$P = \frac{M}{N}$$

حيث أن M هو عدد الأفراد الذين يتمتعون بتلك الخاصة في المجتمع، N عدد أفراد المجتمع.

في هذه الحالة تكون نسبة الأفراد الذين لا يتمتعون بتلك الخاصة مساوية $Q = 1 - P$.

باعتبار أن عدد الأفراد الذين يتمتعون بهذه الخاصة في المجتمع غالبا ما يكون مجهول كما معظم معالم المجتمع نستطيع تقدير نسبة الافراد الذين يتمتعون بتلك الخاصة من خلال عينة مسحوبة من هذا المجتمع. فلو فرضنا أننا سحبنا عينة من عناصر هذا المجتمع بحجم n عنصر وكانت عدد العناصر الذين يتمتعون بتلك الخاصة مساوي m عنصر، نستطيع تقدير هذه النسبة كالتالي :

$$\tilde{P} = p = \frac{m}{n}$$

في هذه الحالة تكون نسبة الأفراد الذين لا يتمتعون بتلك الخاصة مساوية $\tilde{Q} = q = 1 - p$.

ويعطى تباين العينة بالعلاقة:

$$s^2 = p \cdot q$$

و الخطأ المعياري للمقدر p ، والذي سنرمزه بـ σ_p يقدر بواسطة العلاقة:

$$\sigma_p = \sqrt{\frac{p \cdot q}{n}}$$

مثال 5:

لتقدير نسبة المدخنين من الموظفين، سحبنا عينة عشوائية من الموظفين بحجم 160 موظفاً، فوجدنا أن 35 منهم يدخنون. وبذلك نجد أن نسبة المدخنين في العينة تساوي:

$$p = \frac{m}{n} = \frac{35}{160} = 0.2188 \Leftrightarrow (21.88\%)$$

وبذلك نجد أن نسبة المدخنين في مجتمع الموظفين تقدر بما يلي:

$$\tilde{P} = p = 0.2188$$

وأن الخطأ المعياري في تقدير هذه النسبة يقدر بـ:

$$\sigma_p = \sqrt{\frac{p \cdot q}{n}} = \sqrt{\frac{(0.2188)(1-0.2188)}{160}} = 0.0327$$

أي أن نسبة المدخنين في المجتمع المذكور تقدر بـ 21.88 % وبخطأ معياري قدره 3.27 %.

المحاضرة الرابعة

5.1. تقدير الفرق بين متوسطي مجتمعين طبيعيين:

لنفترض أننا نريد مقارنة متوسطي مجتمعين مختلفين. وكما رأينا سابقا غالبا ما يكون متوسط المجتمع مجهول. وبالتالي لمقارنة متوسطي مجتمعين سوف نعتمد على عينتين مسحوبتين من هذين المجتمعين. فلو فرضنا أننا سحبنا عينة من المجتمع الأول بحجم n_1 وعينة من المجتمع الثاني بحجم n_2 . ولنفترض أن متوسطي هاتين العينتين كانا $\bar{x}_1$ و $\bar{x}_2$ على التوالي. في هذه الحالة وكما رأينا سابقا يمكن أن نوجد تقدير لكل من متوسطي المجتمعين باستخدام متوسطي العينتين:

$$\begin{aligned}\bar{y}_1 &= \bar{x}_1 \\ \bar{y}_2 &= \bar{x}_2\end{aligned}$$

ومن خلال هذين المتوسطين يمكن أن نوجد تقديرا للفرق بين متوسطي المجتمعين كالتالي:

$$\bar{y}_1 - \bar{y}_2 = \bar{x}_1 - \bar{x}_2$$

ويقدر تباين هذا الفرق بمجموع التباينين:

$$\sigma_{\bar{x}_1 - \bar{x}_2}^2 = \sigma_{\bar{x}_1}^2 + \sigma_{\bar{x}_2}^2$$

والخطأ المعياري لهذا التقدير بالعلاقة التالية :

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\sigma_{\bar{x}_1}^2 + \sigma_{\bar{x}_2}^2}$$

وبطريقة مشابهة يمكننا تقدير الفرق بين نسبتي مجتمعين كالتالي:

$$\bar{p}_1 - \bar{p}_2 = p_1 - p_2$$

حيث p_1 نسبة هذه الخاصة في المجتمع الاول و p_1 نسبة تلك الخاصة في العينة المسحوبة منه. و p_2

نسبة هذه الخاصة في المجتمع الثاني و p_2 نسبة تلك الخاصة في العينة المسحوبة منه.

و يكون تباين الفرق بين هاتين النسبتين مساوي:

$$\sigma_{p_1 - p_2}^2 = \sigma_{p_1}^2 + \sigma_{p_2}^2$$

و يعطى تقدير الخطأ المعياري لذلك الفرق من خلال العلاقة التالية:

$$\sigma_{p_1 - p_2} = \sqrt{\sigma_{p_1}^2 + \sigma_{p_2}^2}$$

مثال 6:

نفترض أننا نريد تقدير الفرق بين متوسطي الإنفاق الشهري لطلاب كلية طب الأسنان في جامعتين وتقدير الفرق بين نسبتي الطلاب ممن هم من سكان المدينة. قمنا بسحب عينة من كل مجتمع (جامعة)، العينة الأولى بحجم 30 طالب والثانية بحجم 20 طالب، فكان متوسط الإنفاق في العينة الأولى $\bar{x}_1 = 8000$ وفي العينة الثانية $\bar{x}_2 = 10000$ وكان الانحراف المعياري للعينة الأولى $S_1 = (360)$ والانحراف المعياري للعينة الثانية $S_2 = 150$. كما كانت نسبة الطلاب من سكان المدينة في الجامعة الأولى $p_1 = 0.35$ وفي الثانية $p_2 = 0.27$ على الترتيب. فكيف يمكننا أن نقارن بين هذين المجتمعين ونقدر الفروقات بينهما.

للإجابة على ذلك نقوم بتقدير الفرق بين متوسطي الدخل فنجد أن:

$$\bar{y}_2 - \bar{y}_1 = \bar{x}_2 - \bar{x}_1 = 10000 - 8000 = 2000$$

أما بالنسبة لتقدير الخطأ المعياري لذلك التقدير فيساوي:

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\sigma_{\bar{x}_1}^2 + \sigma_{\bar{x}_2}^2}$$

$$\tilde{\sigma}_{\bar{x}_1 - \bar{x}_2} = \sqrt{\sigma_{\bar{x}_1}^2 + \sigma_{\bar{x}_2}^2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{(360)^2}{30} + \frac{(150)^2}{20}} = 73.79$$

أي أن الفرق بين متوسطي الإنفاق الشهري للطلاب بين هاتين الجامعتين يقدر بـ 2000 ل.س، وبخطأ معياري قدره 73.79 ليرة.

أما بالنسبة لتقدير الفرق بين نسبتي المدخنين فتستخدم العلاقة التالية:

$$\bar{p}_1 - \bar{p}_2 = p_1 - p_2 = 0.35 - 0.27 = 0.08$$

ولتقدير الخطأ المعياري المتعلق بذلك الفرق نطبق العلاقة التالية:

$$\begin{aligned} \tilde{\sigma}_{p_1 - p_2} &= \sqrt{\frac{p_1 \cdot q_1}{n_1} + \frac{p_2 \cdot q_2}{n_2}} = \sqrt{\frac{(0.35)(1-0.35)}{30} + \frac{(0.27)(1-0.27)}{23}} \\ &= \sqrt{0.0076 + 0.0099} = 0.1323 \end{aligned}$$

أي أن الفرق بين نسبتي المدخنين في المدينتين يقدر بـ 0.08 وبخطأ معياري يساوي 0.1323.

6.1 . التقدير المجالي :

بعد أن درسنا عدداً من التقديرات النقطية لبعض مؤشرات المجتمع وتوصلنا إلى علاقات رياضية لحساب كل منها. نشير إلى أن تلك التقديرات غير متحيزة وفعالة ومتماسكة للمؤشر المجهول، إلا أن أياً منها لا يعطينا أية درجة من الثقة فيه. فنحن لانعرف فيما إذا كانت قيمة التقدير النقطي قريبة من القيمة الحقيقية المجهولة أم بعيدة عنها. ومن هنا لابد من إيجاد وسيلة تضمن لنا وباحتمال معين أن تكون تلك القيمة الحقيقية المجهولة واقعة ضمن مجال معين حول القيمة المقدرة.

تقدير مجال الثقة لمتوسط المجتمع

عند إنشاء مجال ثقة لمتوسط المجتمع أي مجال يضمن لنا وقوع القيمة الحقيقية لمتوسط المجتمع فيه وفقاً لاحتمالات محددة يمكن أن نميز هنا بين حالتين بحسب إذا كان تباين المجتمع معلوم أو مجهول .

1 - إذا كان تباين المجتمع معلوم (σ^2 معلوم)

عندما نقوم بتقدير متوسط المجتمع من خلال متوسط العينة فإننا نرتكب خطأ وليكن E . في هذه الحالة نعلم أن متوسط المجتمع يقع في المجال :

$$\bar{x} - E < \bar{y} < \bar{x} + E$$

عندئذ إذا كان احتمال الوقوع في الخطأ هو α فإن مقدار الخطأ المرتكب في تقدير متوسط المجتمع E يساوي:

$$E = Z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}$$

حيث أن $Z_{\alpha/2}$ هي القيمة النظرية (الحدية) المقابلة لمساحة $\alpha/2$ عن يمين التوزيع الطبيعي المعياري و $\alpha/2$ عن يساره. في هذه الحالة يكتب مجال الثقة لمتوسط المجتمع على الشكل التالي:

$$\bar{x} - Z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}} < \bar{y} < \bar{x} + Z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}$$

ويشترط لاستخدام هذا التقدير أن تكون العينة عشوائية وأن يتحقق الشرطين التاليين معا أو يتحقق أحدهما على الأقل:

- أن يكون توزيع المجتمع طبيعي

- أن يكون حجم العينة كبيراً ($n > 30$)

2 - إذا كان تباين المجتمع غير معلوم (σ^2 غير معلوم)

في حال كان تباين المجتمع σ^2 غير معلوم نستخدم تباين العينة S^2 كتقدير لتباين المجتمع ويكون مقدار الخطأ المرتكب في تقدير متوسط المجتمع E مساوي:

$$E = t_{\alpha/2} \times \frac{S}{\sqrt{n}}$$

وفي هذه الحالة يمكن إنشاء مجال الثقة لمتوسط المجتمع على الشكل التالي:

$$\bar{x} - t_{\alpha/2} \times \frac{S}{\sqrt{n}} < \bar{y} < \bar{x} + t_{\alpha/2} \times \frac{S}{\sqrt{n}}$$

ويشترط كذلك لاستخدام هذا التقدير أن تكون العينة عشوائية وأن يتحقق الشرطين التاليين معا أو يتحقق أحدهما على الأقل:

- أن يكون توزيع المجتمع طبيعي
- أن يكون حجم العينة كبيرا ($n > 30$)

تقدير مجال الثقة لنسبة خاصة في مجتمع

فيما يتعلق بمجال الثقة لنسبة خاصة في مجتمع وإذا كان احتمال الوقوع في الخطأ هو α فإن مقدار الخطأ المرتكب في تقدير النسبة E يساوي:

$$E = Z_{\alpha/2} \times \sqrt{\frac{p \times q}{n}}$$

في هذه الحالة يكتب مجال الثقة لمتوسط المجتمع على الشكل التالي:

$$p - Z_{\alpha/2} \times \sqrt{\frac{p \times q}{n}} < P < p + Z_{\alpha/2} \times \sqrt{\frac{p \times q}{n}}$$

حيث أن p هي النسبة المحسوبة من العينة.

مثال 7 : أوجد مجال الثقة بمستوى دلالة مقداره 5 % (أو باحتمال مقداره 95 %) لمتوسط الإنفاق الشهري على المكالمات بالهاتف الجوال لطلاب الجامعة، علماً أنه سحبنا عينة عشوائية منهم بحجم $n = 50$ وكان متوسطها $\bar{x} = 1220$ وانحرافها المعياري $S = 75.5$.

نعلم من جدول القيم الحرجة للتوزيع الطبيعي المعياري أن قيمة Z المقابلة لمستوى دلالة مقداره 5% هو $Z_{\alpha/2}$. وبالتالي يكتب مجال الثقة الموافق كالتالي:

$$1220 - 1.96 \times \frac{75.5}{\sqrt{50}} < \bar{y} < 1220 + 1.96 \times \frac{75.5}{\sqrt{50}}$$

وإذا اعتبرنا أن نسبة المدخنين في العينة السابقة هي 0.4 يمكننا إنشاء مجال ثقة لنسبة المدخنين في الجامعة كالتالي:

$$0.4 - 1.96 \times \sqrt{\frac{0.4 \times 0.6}{50}} < P < 0.4 + 1.96 \times \sqrt{\frac{0.4 \times 0.6}{50}}$$

وهو المجال الذي يحوي النسبة الحقيقية المجهولة P باحتمال قدره 95%.

ملاحظة: إذا كان توزيع المجتمع غير طبيعي وكان حجم العينة صغيراً ($n < 30$)

في هذه الحالة يتم استخدام الطرق اللامعلمية في التقدير كما سنرى لاحقاً. ولكن يجب التنويه إلى أنه في بعض الحالات يمكن اعتبار العينة كبيرة إذا كان ($n > 15$) إذا كان توزيع المجتمع طبيعي أو قريب جداً من التوزيع الطبيعي. في هذه الحالة يمكن استخدام مجالات الثقة السابقة الذكر في التقدير المجالي. كما أنه عندما يكون توزيع المجتمع بعيد بشكل كبير عن التوزيع الطبيعي فإننا نحتاج إلى عينة بحجم 50 أو 100 لكي نستطيع اعتبارها عينة كبيرة.

7.1. تقدير حجم العينة

رأينا سابقاً أن الخطأ المرتكب في تقدير متوسط المجتمع والمقابل لمستوى دلالة α يكتب على الشكل التالي:

$$E = Z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}$$

وبالتالي ومن خلال هذه العلاقة يمكننا إيجاد حجم العينة اللازم سحبها لتقدير متوسط المجتمع بحيث لا يتجاوز الخطأ المرتكب المقدار E وذلك بمستوى دلالة α (أو باحتمال قدره $\beta = 1 - \alpha$) كالتالي:

$$n = \left[\frac{Z_{\alpha} \sigma}{E} \right]^2$$

وبالطريقة نفسها يمكننا استنتاج العلاقة التي نحسب منها حجم العينة اللازم لتقدير نسبة خاصة معينة في المجتمع، وهي العلاقة الآتية:

$$n = \frac{(z_{\alpha})^2 p q}{E^2}$$

هذا في حال كان لدينا تقدير للنسبة P . أما إن لم يكن لدينا هكذا تقدير فنستخدم العلاقة التالية:

$$n = \frac{(z_{\alpha})^2 \times 0.25}{E^2}$$

أي أننا نعتبر أن النسبة مساوية لـ 0.5 أي بشكل آخر نأخذ أكبر حجم عينة ممكن لتقليل الأخطاء التي يمكن أن نقع بها وللحصول على تقديرات جيدة.

مثال 8:

أوجد حجم العينة اللازم لتقدير متوسط خاصة معينة في المجتمع، ثم حجمها لتقدير نسبة المدخنين فيه، إذا علمت أننا نريد أن يكون ذلك التقدير باحتمال ثقة قدرة 0.95 (أي بمستوى دلالة أقل من 0.05)، وأن الخطأ في تقدير المتوسط تساوي 5 وفي تقدير نسبة المدخنين 0.10 وأن العينة الاختبارية أعطتنا أن $s^2 = 90$ وأن نسبة المدخنين فيه كانت تساوي $p = 0.6$.

الحل:

من الشروط السابقة نجد أن قيمة z المقابلة لاحتمال الثقة المذكور هي: $Z = 1.96$.
وأن حجم العينة اللازم لتقدير المتوسط يُحسب من العلاقة الآتية:

$$n = \left[\frac{z_{\alpha} \sigma}{E} \right]^2 = \frac{(1.96)^2 \times 90}{(5)^2} = 13.82 \approx 14$$

أي أن الحد الأدنى لحجم العينة الذي يحقق الشروط السابقة لتقدير المتوسط هو 14 فرداً.

أما بالنسبة لحجم العينة اللازم لتقدير النسبة P فنحسبه من العلاقة:

$$n = \frac{(z_{\alpha})^2 p q}{E^2} = \frac{(1.96)^2 \times 0.6 \times 0.4}{0.01} = 92.19 \approx 92$$

وهو الحد الأدنى لحجم العينة اللازم لتحقيق الشروط السابقة في تقدير النسبة P .

تمريعات

- 1 - لنفترض أن مجتمعاً مؤلفاً من العناصر المعلومة التالية:
 $Y_i : 12, 14, 16, 18, 15, 17, 16, 19, 15, 17$
 - اسحب بشكل عشوائي وبدون إعادة عينة بحجم $n = 3$.
 - أوجد تقديراً لمتوسط المجتمع وقارنه مع المتوسط الحقيقي $\bar{y}$.
 - أوجد تقديراً لتباين ذلك المجتمع وقارنه مع التباين الحقيقي.
 - 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 (ألف ل.س).
 $X_i : 12, 10, 14, 18, 17, 11, 19, 20, 13$
 والمطلوب:
 - إيجاد تقدير لمتوسط الدخل في المجتمع $\bar{y}$.
 - إيجاد تقدير لتباين الدخل فيه σ^2 .
 - إيجاد تقدير للتباين σ_x^2 .
 - إيجاد تقدير نسبة الدخل التي هي أقل من 15 وتقدير تباينها.
 - إيجاد مجال الثقة المقابل للاحتمال 0,95 لكل من المتوسط والنسبة.
- 3- احسب حجم العينة اللازم لتحقيق دقة $d=0.01$ ، وباحتمال 0.95 في حالة السحب مع الإعادة عند تقدير نسبة المدخنين علماً بأن التقدير الأولي لنسبة المدخنين كان $p=0.25$.

المحاضرة الخامسة

الفصل الثاني

اختبار الفرضيات من عينة واحدة

Hypothesis Testing with one Sample

مفردات البحث

- 1- الخطوات الأساسية في اختبار الفرضيات
- 2- اختبارات حول متوسط المجتمع: σ معلوم
- 3- اختبارات حول متوسط المجتمع: σ غير معلوم
- 4- اختبارات حول نسبة خاصة في مجتمع
- 5- اختبارات حول الانحراف المعياري والتباين

المرجع المعتمد: المرجع السابق من الصفحة 156 إلى الصفحة 254

تمهيد :

يتعلق اختبار الفرضيات بالمسائل المتعلقة باتخاذ قرار معين يتناول أحد مؤشرات المجتمع كالمتوسط أو النسبة أو الفرق... الخ. ويتضمن الحكم على صحة أو خطأ فرضية ما أو إدعاء حول قيم معينة يمكن تأخذها هذه المؤشرات. فقد يدعي أحدهم على سبيل المثال أن متوسط الدخل الشهري للمحل التجاري في منطقة سكنه يساوي 25000 ليرة، أو أن نسبة البطالة في مدينته تساوي 25%... الخ. هذه الادعاءات هي افتراضات قد تكون صحيحة وقد تكون خاطئة. لذلك لكي نستطيع القول أنها صحيحة أو أنها خاطئة يجب اختبارها على عينة عشوائية من عناصر المجتمع المدروس. ومن خلال هذه العينة نقوم بإجراء الحسابات اللازمة لتقدير قيمة المؤشر الذي نتناوله الفرضية، ثم نقوم بمقارنة هذه القيم المحسوبة مع قيم نظرية أو قيم حرجة لتلك المؤشرات وبناءاً على هذه المقارنة نتخذ القرار المناسب.

1.2. الخطوات الأساسية لاختبار الفرضيات

لإختبار أية فرضية نتبع الخطوات التالية:

- 1- نقوم بداية بتحديد احتمال لرفض الفرضية المختبرة وهو ما نسميه بمستوى الدلالة α (احتمال الوقوع في الخطأ) أو هو احتمال الخطأ من النوع الأول (احتمال رفض فرضية العدم مع أنها صحيحة) وهو احتمال يحدده الباحث إلا أنه غالباً ما يكون مساوياً لـ 0.1 أو 0.05 أو 0.01.
- 2- تحديد الفرضية المختبرة أو ما يسمى فرضية العدم (*Null Hypothesis*) : وهو افتراض يضعه الباحث على شكل مساواة بأن أحد معالم المجتمع وليكن θ (المتوسط أو التباين الخ) يساوي قيمة معينة θ_0 (ادعاء). أو بتعبير آخر تدل على عدم وجود اختلاف جوهري بين الطرفين. ونرمز لفرضية العدم بالرمز H_0 . و نستطيع كتابة هذه الفرضية على الشكل التالي:

$$H_0 : \theta = \theta_0$$

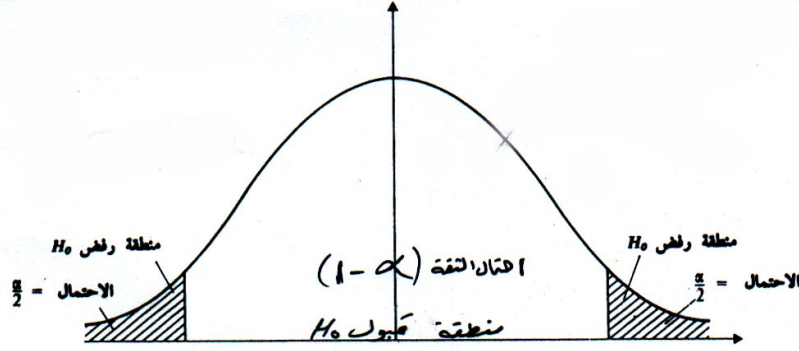
$$H_0 : \theta - \theta_0 = 0$$

أو على الشكل

وهنا نلاحظ أن منطقة قبول هذه الفرضية تشكل مجال الثقة المناسب حول θ_0 وباحتمال ثقة $(1-\alpha)$.

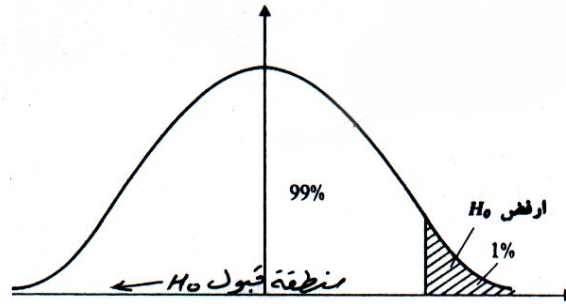
- 3- تحديد الفرضية البديلة (*Alternative Hypothesis*) (وترمز H_1 أو H_a) وتحديد منطقة الرفض: وهي الفرضية التي تقول بأن المعلم θ يختلف عن القيمة θ_0 . وبالتالي تأخذ هذه الفرضية أحد ثلاثة أشكال أساسية هي:

- $H_1 : \theta \neq \theta_0$ وعندها نسمي الاختبار اختبار ثنائي الجانب حيث تقع منطقة رفض فرضية العدم على جانبي منطقة قبولها. ويقابل كل جانب يقابل نصف مستوى الدلالة $\alpha/2$ كما على الشكل التالي:



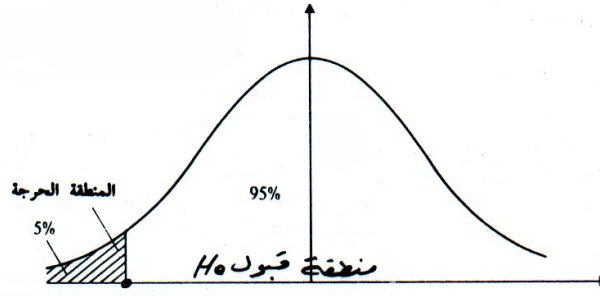
شكل (1-4)

- $H_1 : \theta > \theta_0$ وعندها نسمي الاختبار اختبار أحادي الجانب يميني حيث تقع منطقة رفض فرضية العدم على الجانب الأيمن من منطقة قبولها كما على الشكل التالي:



شكل (2-4)

- $H_1 : \theta < \theta_0$ وعندها يكون الاختبار أحادي الجانب يساري حيث تقع منطقة رفض فرضية العدم على الجانب الأيسر من منطقة قبولها كما على الشكل التالي:



شكل (3-4)

4- تحديد مؤشر الاختبار أو ما يسمى إحصاء الاختبار t (Test Statistic): وهو المؤشر الذي سيستعمل في رفض فرضية العدم و يحسب من معطيات العينة. ويحسب مؤشر الاختبار من خلال تحويل الاحصاء البسيط إلى قيمة معيارية بحيث تكون فرضية العدم صحيحة. وهو عبارة عن متغير يتبع أحد قوانين التوزيع الاحتمالية (التوزيع الطبيعي أو توزيع فيشر.... الخ). وأهم الاحصاءات المستخدمة هي :

- الاحصاء المتعلق بالنسبة:
$$z = \frac{\hat{p} - p}{\sqrt{pq/n}}$$

- الاحصاء المتعلق بالمتوسط
$$t = \frac{\bar{x} - \mu}{S/\sqrt{n}} \quad \text{أو} \quad z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

5- مقارنة القيمة المحسوبة لمؤشر الاختبار بالقيمة الحرجة له t_α (Critical Value) (أي القيمة المحسوبة من قانون التوزيع الاحتمالي النظري) واتخاذ القرار المناسب. وهنا نكون أمام الخيارات التالية :

- الاختبار ثنائي الجانب : نقبل الفرضية H_0 إذا كانت $|t| < t_\alpha$ ونرفضها (وبالتالي نقبل H_1) إذا كانت $|t| > t_\alpha$.

- الاختبار أحادي الجانب يميني $(H_1 : \theta > \theta_0)$: نقبل الفرضية H_0 إذا كانت $t < t_\alpha$ ونرفضها (وبالتالي نقبل H_1) إذا كانت $t \geq t_\alpha$.

- الاختبار أحادي الجانب يساري $(H_1 : \theta < \theta_0)$: نقبل الفرضية H_0 إذا كانت $t > -t_\alpha$ ونرفضها (وبالتالي نقبل H_1) إذا كانت $t \leq -t_\alpha$.

ملاحظة : هناك طرق أخرى للحكم على رفض فرضية العدم وهو مقارنة $P-Value$ مع مستوى الدلالة α (احتمال الوقوع في الخطأ) أو استخدام مجال الثقة للمعلم الاحصائي. و $P-Value$ هو احتمال الحصول على قيمة لمؤشر الاختبار أكبر على الأقل من قيم بيانات العينة عندما تكون فرضية العدم صحيحة. وترفض الفرضية H_0 عندما تكون $P-Value$ أصغر من مستوى الدلالة (0.05 على سبيل المثال)

2.2. اختبارات حول متوسط المجتمع: σ معلوم

يشترط لإجراء هذا الاختبار أن تكون العينة عشوائية وأن يكون تباين المجتمع معلوم وأن يتحقق أحد الشرطين التاليين:

- توزيع المجتمع طبيعي

- حجم العينة أكبر من 30 عنصر (عينة كبيرة)

في هذه الحالة يكون مؤشر الاختبار التالي التوزيع الطبيعي المعياري:

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \sim N(0;1)$$

حيث إن $\bar{x}$ هو متوسط العينة و μ متوسط المجتمع.

مثال 9:

ادعى موظف المالية أن متوسط دخل الطبيب في اليوم يساوي 5000 ليرة سورية. وللتحقق من ذلك سحبنا عينة عشوائية من 35 طبيباً، فوجدنا أن متوسط الدخل فيها كان $\bar{x} = 4660$ ليرة في اليوم، وأن الانحراف المعياري معلوم ويساوي $\sigma = 1207$. **المطلوب :** اختبار صحة هذا الادعاء بمستوى دلالة قدره 0.05 (أي باحتمال ثقة 0.95).

نلاحظ أنه يمكننا كتابة فرضيات الاختبار على الشكل التالي :

$$H_0 : \mu = 5000$$

$$H_1 : \mu \neq 5000$$

وبحساب قيمة مؤشر الاختبار نجد :

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{4660 - 5000}{1207/\sqrt{35}} = -1.66$$

بمقارنة قيمة $Z_{\alpha/2}$ النظرية والتي تساوي 1.96 نجد أن $|Z| = 1.66 < Z_{\alpha/2} = 1.96$ وبالتالي نقبل الفرضية H_0 لأنه ليس لدينا دليلاً كافياً على رفضها.

مثال 10 :

يفترض أحد الأطباء أن متوسط مدة نزيف الدم بعد قلع الضرس يساوي 12 ساعة. وللتأكد من ذلك قمنا بأخذ عينة عشوائية بحجم $n = 11$ شخصاً فوجدنا أن متوسط المدة كان $\bar{x} = 11$ فإذا كان الانحراف المعياري للمجتمع معلوم ويساوي 0.994. فهل يمكن اعتبار مدة النزف أقل من 12 ساعة. اختبر بمستوى دلالة 0.01 (باحتمال ثقة 0.99) إذا علمت أن توزيع المجتمع الأصلي طبيعي.

لاختبار ذلك نضع الفرضيتين العدمية والبديلة كما يلي:

$$H_0 : \mu = 12$$

$$H_1 : \mu < 12$$

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{11 - 12}{0.99/\sqrt{11}} = -3.33$$

بمقارنة قيمة Z_{α} النظرية والتي تساوي -2.33 وبما أن الاختبار أحادي الجانب يساري يكون لدينا $Z = -3.33 < Z_{\alpha} = -2.33$ وبالتالي نرفض الفرضية H_0 ونقبل الفرضية H_1 .

3.2. اختبارات حول متوسط المجتمع: σ غير معلوم

يشترط لإجراء هذا الاختبار أن تكون العينة عشوائية وأن يكون تباين المجتمع معلوم وأن يتحقق أحد الشرطين التاليين:

- توزيع المجتمع طبيعي

- حجم العينة أكبر من 30 عنصر (عينة كبيرة)

في هذه الحالة يتبع مؤشر الاختبار التالي توزيع ستيودنت ب $n - 1$ درجة حرية:

$$z = \frac{\bar{x} - \mu}{S/\sqrt{n}} \sim T_{\alpha}(n-1)$$

حيث إن S هو الانحراف المعياري للعينة .

مثال 11:

من أجل معرفة درجة حرارة الجسم للشخص السليم قمنا بقياس درجة حرارة 106 أشخاص سليمين فكان متوسط درجة الحرارة $99F$ و الانحراف المعياري للعينة يساوي $0.62F$ اختبر بمستوى دلالة مقداره 0.05 الادعاء بأن درجة حرارة الجسم أكبر من $98.6F$.

$$H_0 : \mu = 98.2$$

$$H_1 : \mu > 98.2$$

$$t = \frac{\bar{x} - \mu}{S/\sqrt{n}} = \frac{99 - 98.6}{0.62/\sqrt{106}} = 6.64$$

نلاحظ أن القيمة الحرجة لمؤشر الاختبار بمستوى دلالة 0.05 و 105 درجة حرية تساوي 1.66 وبالتالي نرفض H_0 .

المحاضرة السادسة

4.2. اختبارات حول نسبة خاصة في المجتمع:

إن إجراء أي اختبار حول نسبة خاصة في المجتمع p يجب أن تكون العينة المدروسة عشوائية وأن تكون الظاهرة خاضعة للتوزيع الثنائي (لها وجهان تحقق باحتمال p وعدم تحقق باحتمال $q=1-p$) وأخيرا أن يتحقق الشرطان التاليين: $np \geq 5$ و $nq \geq 5$. لأنه في هذه الحالة يمكن تقريب التوزيع الثنائي للتوزيع الطبيعي بتوقع $\mu = np$ وتباين يساوي $\sigma^2 = npq$. في هذه الحالة يكتب مؤشر الاختبار على النحو التالي:

$$Z = \frac{\hat{p} - p}{\sqrt{pq/n}} \sim N(0;1)$$

حيث أن $\hat{p}$ هي النسبة في العينة.

مثال 12:

في تجارب تهجين نوع من النبات من المتوقع الحصول على نسبة 25 % من هذه البذور من النوع A و الباقي من النوع B . وفي تجربة أجريت على 580 نبتة حصلنا على 26.2 % من البذور من النوع A . فهل نستطيع القول بناء على هذه المعطيات أن نسبة البذور من النوع A مختلفة عن 25 % (اختبر بمستوى دلالة مقداره 5 %).

باعتبار أن شروط الاختبار محققة نستطيع كتابة فروض الاختبار على الشكل التالي:

$$H_0 : p = 0.25$$

$$H_1 : p \neq 0.25$$

$$Z = \frac{\hat{p} - p}{\sqrt{pq/n}} = \frac{0.262 - 0.25}{\sqrt{0.25 \times 0.75 / 580}} = 0.76$$

نلاحظ أن القيمة الحرجة للتوزيع الطبيعي المعياري عند مستوى دلالة هو 1.96 و بمقارنة هذه القيمة مع القيمة الحسوبة السابقة نجد أن $Z_{\alpha/2} = 1.96 < Z = 0.76$ وبالتالي نقبل H_0 بأن نسبة البذور من النوع A لا تختلف جوهريا عن 25 % .

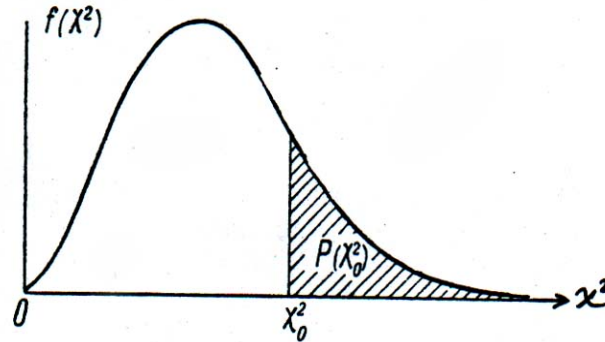
5.2. اختبارات حول تباين مجتمع طبيعي

في كثير من المؤسسات الانتاجية والخدمية يكون الهدف تحسين نوعية منتجاتها من خلال خفض التباين في جودة هذه المنتجات. وبالتالي لا يكون الهدف فقط انتاج منتج بمتوسط مقبول بل يتعدى ذلك إلى انتاج منتج قليل التباين بين وحداته. لذلك نرى أن اختبارات التباين تستخدم بكثرة في مراقبة الجودة.

لإجراء اختبار حول التباين نستخدم مؤشر اختبار χ^2 التالي:

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

حيث أن S^2 هو تباين العينة و σ^2 تباين المجتمع (إدعاء حول تباين المجتمع). حيث أن توزيع كاي مربع له الشكل التالي:



مثال (13):

يقول أحدهم أن تباين درجات الطلاب في مقرر الإحصاء يساوي 256 وللتحقق من ذلك سحبنا عينة من هؤلاء الطلاب بحجم $n = 28$ ، فوجدنا أن تباينها يساوي $s^2 = 236$. اختبر صحة القول السابق بمستوى دلالة قدره 0.10.

نضع الفرضيتين العدمية والبديلة كما يلي:

$$H_0 : \sigma^2 = 256$$

$$H_1 : \sigma^2 \neq 256$$

ثم نحسب قيمة مؤشر الاختبار فنجد أن:

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} = \frac{(28-1) \cdot 236}{256} = 24.89$$

ومن جداول χ^2 نجد أن قيمة $\chi^2_{\alpha/2}(27) = 16.15$ المقابلة لمستوى الدلالة 0.10 ولـ (28-1) درجة حرية.

وبمقارنة قيمة χ^2 المحسوبة مع χ^2 الحرجة، نجد أن $\chi^2 = 24.89 < \chi^2_{\alpha/2}(27) = 16.15$. لذلك نقبل فرضية عدم القائلة بأن تباين المجتمع المفروض يساوي 256.

المحاضرة السابعة

الفصل الثالث

الاستدلال الاحصائي من عينتين

Inference from Two Samples

مفردات البحث

1- اختبار الفرق بين متوسطين في عينتين مستقلتين

2- اختبارات الفرق بين نسبتي في مجتمعين طبيعيين

3- اختبار الأزواج المتقابلة

4- مقارنة الاختلاف (التباين) بين عينتين

المرجع المعتمد: المرجع السابق من الصفحة 254 إلى الصفحة 313

1.3. اختبار الفرق بين متوسطين في عينتين مستقلتين

عندما نكون بصدد مقارنة متوسطي عينتين يجب تحقق الشروط التالية:

- العينتان عشوائيتان
- العينتان مستقلتان (غير مرتبطتان)
- العينتان كبيرتان ($n_1 > 30$ و $n_2 > 30$) أو توزيع المجتمعين طبيعي.

في هذه الحالة يمكن مقارنة المتوسطان باستخدام مؤشر الاختبار التالي:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim T(df)$$

حيث $\bar{x}_1, \bar{x}_2$ متوسطي العينتين. μ_1, μ_2 متوسطي المجتمعين (الادعاء). S_1^2, S_2^2 تبايني العينتين. (df) درجات الحرية وتساوي أصغر القيمتين $n_1 - 1$ و $n_2 - 1$.

مثال (14):

نفترض أننا نريد أن ندرس أداء الطالب في مقرر الرياضيات في مدرستين مختلفتين، فسحبنا عينة عشوائية من طلاب المدرسة الأولى بحجم $n_1 = 40$ وعينة عشوائية من طلاب المدرسة الثانية بحجم $n_2 = 50$ فكان متوسطا الدرجات فيهما $\bar{x}_1 = 65$, $\bar{x}_2 = 60$ وتبايناهما $s_1^2 = 160$, $s_2^2 = 250$. والمطلوب اختبار الفرضية التي تقول أن الفرق بين متوسطي الدرجات في المدرستين غير جوهري وذلك بمستوى دلالة قدره 0.05.

الحل:

نضع الفرضيتين العدمية والبديلة كما يلي:

$$H_0 : \mu_1 - \mu_2 = 0$$

$$H_1 : \mu_1 - \mu_2 \neq 0$$

ثم نقوم بحساب مؤشر الاختبار فنجد أن:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \frac{(65 - 60) - 0}{\sqrt{\frac{160}{40} + \frac{250}{50}}} = +1.67$$

وبمقارنة ذلك مع القيمة الحرجة المعيارية $t_{\alpha/2}(39) = 2.02$ المقابل مستوى الدلالة 0.05 باتجاهين نجد

أن $t_{\alpha/2}(39) = 2.02 > t = 1.67$ لذلك نرفض الفرضية H_0 وبالتالي نقر بوجود فرق جوهري بين

متوسطي درجات الطلاب في مقرر الرياضيات في هاتين المدرستين.

حالة خاصة 1

إذا كان **تباينا المجتمعين معلومين** فإن اختبار الفرق بين المتوسطين يتم باستخدام التوزيع الطبيعي مع

استبدال تبايني العينتين في المؤشر السابق بتباين المجتمعين. أي :

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0;1)$$

حيث $\sigma_1^2; \sigma_2^2$ تبايني المجتمعين .

حالة خاصة 2

إذا كان **تباينا المجتمعين غير معلومين لكنهما متساويان** فإن اختبار الفرق بين المتوسطين يتم

باستخدام التوزيع الطبيعي باستخدام المؤشر التالي :

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}} \sim T(n_1 + n_2 - 2)$$

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \quad \text{حيث}$$

مثال (15):

لدراسة متوسط النفقات الشهرية لطلاب السنة الاولى في كلية طب الاسنان على المواد المخبرية سحبنا عينتين بحجمين $n_1=40$, $n_2=30$ من طلاب جامعتين ، فكان متوسطا النفقات في العينتين : $\bar{x}_1 = 1000$, $\bar{x}_2 = 1100$ وكان تباينا المجتمعين $\sigma_1^2 = 50000$, $\sigma_2^2 = 60000$ على الترتيب. فهل تدلنا هذه المعطيات على وجود فرق جوهري بين متوسطي هاتين العينتين؟. اختبر ذلك بمستوى دلالة 0.05.

الحل:

$$H_0: \mu_1 - \mu_2 = 0$$

$$H_1: \mu_1 - \mu_2 \neq 0$$

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(1000 - 1100) - 0}{\sqrt{\frac{50000}{40} + \frac{60000}{30}}} = -1.75$$

وبمقارنة القيمة المطلقة لـ Z مع قيمة $Z_{\alpha/2}$ المقابل لنصف مستوى الدلالة 0.025 فنجد أن $|Z|$ أصغر من $Z_{0.025} = 1.96$ لذلك لا نستطيع أن نقول بوجود فرق جوهري بين متوسطي العينتين. وهذا يعني أن الفرق الحاصل بينهما ناتج عن الجوانب العشوائية.

2.3. اختبار الفرق بين نسبتي

لاختبار الفرق بين نسبتي يجب أن تكون العينتان عشوائيتان وغير مرتبطتان كما أن عدد الأفراد الذين يحققون الخاصة والذين لا يحققون الخاصة في العينة يجب أن يتجاوز كل منهما الخمسة. عند تحقق هذه الشروط نستطيع اختبار الفرق بين النسبتين باستخدام مؤشر الاختبار التالي :

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\bar{p}\bar{q}}{n_1} + \frac{\bar{p}\bar{q}}{n_2}}} \sim N(0,1)$$

حيث $\hat{p}_1, \hat{p}_2$ النسب الحسوبة من العينتين. p_1, p_2 النسب في المجتمع (الادعاء المفترض).
 $\bar{p} = \frac{x_1 + x_2}{n_1 + n_2}$ و x_1 عدد الأفراد الذين يحققون الخاصة في العينة الأولى (عدد حالات تحقق الخاصة)
و x_2 عدد الأفراد الذين يحققون الخاصة في العينة الثانية. و $\bar{q} = 1 - \bar{p}$.

مثال (16):

لدراسة نسبي المدخنين في مجتمعين سحبنا منهما عينتين بحجم $n_1 = 300$ و $n_2 = 200$ فوجدنا
أن عدد المدخنين فيهما يساوي 60 و 50 على الترتيب. اختبر الفرضية القائلة بعدم وجود فرق جوهري
بين النسبتين في المجتمعين ، وذلك بمستوى دلالة قدره 0.05

الحل:

نضع فرضية العدم القائلة بعدم وجود فرق بين النسبتين في المجتمعين كما يلي:

$$H_0: p_1 - p_2 = 0$$

$$H_1: p_1 - p_2 \neq 0$$

ثم نحسب مؤشر الاختبار:

$$\hat{p}_1 = \frac{x_1}{n_1} = \frac{60}{300} = 0.2 \quad \hat{p}_2 = \frac{x_2}{n_2} = \frac{50}{200} = 0.25$$

$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{60 + 50}{300 + 200} = 0.22$$

$$\bar{q} = 1 - \bar{p} = 1 - 0.22 = 0.78$$

$$Z = \frac{(0.20 - 0.25) - 0}{\sqrt{\frac{(0.22)(0.78)}{300} + \frac{(0.22)(0.78)}{200}}} = -1.32$$

وبمقارنة القيمة المطلقة لـ Z مع القيمة الحرجة المعيارية $Z_{\alpha/2}$ المقابل لنصف مستوى الدلالة
0.05 والذي يساوي $Z_{0.025} = 1.96$ ، نجد أن القيمة المطلقة لـ Z أصغر من $Z_{\alpha/2}$. إذن نقبل الفرضية
العدمية والقائلة بعدم وجود فرق جوهري بين نسبي المدخنين في المجتمعين.

حالة خاصة:

كما رأينا سابقا كنا نختبر فرضية أن الفرق بين نسبتي خاصة في مجتمعين معدوم ($p_1 - p_2 = 0$) وذلك باعتبار أن هاتين النسبتين في المجتمعين مجهولتان وكنا نفترض أن توزيع الفرق بين النسبتين ($p_1 - p_2$) هو توزيع يقترب من الطبيعي ولذلك كنا نقدر تباين الفرق بين النسبتين كالتالي:

$$\sigma_{\hat{p}_1 - \hat{p}_2}^2 = \frac{\overline{pq}}{n_1} + \frac{\overline{pq}}{n_2}$$

وبالتالي كنا نستخدم مؤشر الاختبار السابق.

عندما نفترض أن الفرق بين النسبتين غير معدوم أي أن ($p_1 - p_2 \neq 0$) أو بتعبير آخر ($p_1 - p_2 = c$) فإننا نقدر تباين الفرق بين النسبتين كالتالي:

$$\sigma_{\hat{p}_1 - \hat{p}_2}^2 = \sigma_{\hat{p}_1}^2 + \sigma_{\hat{p}_2}^2 = \frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}$$

وبالتالي يمكن في هذه الحالة أن نستخدم مؤشر الاختبار التالي:

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}} \sim N(0;1)$$

مثال 17:

في دراسة لتأثير المصانع على الإصابة بمرض الربو إدعى أحدهم أن نسبة المصابين بهذا المرض في المناطق القريبة من المناطق الصناعية تبلغ 10 % بينما تبلغ هذه النسبة 4 % في المناطق البعيدة عنها. وللتأكد من ذلك تم أخذ عينتين من سكان مدينتين بحجم 40 شخص من إحدى المدن القريبة من إحدى المدن الصناعية و العينة الأخرى بحجم 50 من مدينة بعيدة عن المناطق الصناعية فتبين أن نسبة المصابين بالربو في المدينة القريبة بلغت 9 % وفي المدينة البعيدة بلغت 5 %. فهل تقدم هذه النتائج دليلا على صحة الادعاء السابق. اختبر بمستوى دلالة 0.05 .

$$H_0 : p_1 - p_2 = 0.06$$

$$H_1 : p_1 - p_2 \neq 0.06$$

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}} = \frac{(0.09 - 0.05) - 0.06}{\sqrt{\frac{0.09 \times 0.91}{40} + \frac{0.05 \times 0.95}{50}}} = -0.36$$

وبمقارنة القيمة المطلقة لـ Z مع القيمة الحرجة المعيارية $Z_{\alpha/2}$ المقابل لنصف مستوى الدلالة 0.05 والذي يساوي $Z_{0.025} = 1.96$ ، نجد أن القيمة المطلقة لـ Z أصغر من $Z_{\alpha/2}$. إذن نقبل الفرضية العدمية والقائلة بأن الفرق بين النسبتين يساوي 0.06 .

المحاضرة الثامنة

. اختبارات حول الأزواج المتقابلة (Matched Paires)

رأينا فيما سبق كيف كنا نختبر الفرق بين عينتين مستقلتين. في هذه الفقرة سوف نختبر الفرق بين عينتين مرتبطتين. أي أن كل قيمتين متقابلتين من المشاهدات في العينتين مرتبطتين فيما بينهما بعلاقة الأزواج المتقابلة (Matched Pairs). إذا لإجراء مثل هذا الاختبار يجب أن تكون العينة عشوائية و مكونة من مشاهدات على شكل أزواج متقابلة وأن يكون عدد هذه الأزواج ($n > 30$) أو أن يكون توزيع الفرق بين قيم أزواج المشاهدات توزيعاً طبيعياً أو شبه طبيعي.

لنرمز ب d : للفرق بين كل زوج من القيم

μ_d : القيمة المتوسطة للفروقات d في المجتمع ولجميع الأزواج

$\bar{d}$: : القيمة المتوسطة للفروقات d في العينة

S_d : الانحراف المعياري للفروقات d في العينة

يمكننا الآن صياغة مؤشر الاختبار على النحو التالي:

$$t = \frac{\bar{d} - \mu_d}{S_d / \sqrt{n}} \sim T(n-1)$$

مثال 18:

في سبيل معرفة فيما إذا كان نوع بذور الذرة تأثير على إنتاجية الهكتار قام أحد المزارعين باستخدام نوعين من البذار و قام بزراعتهم في قطع متقابلة من الأراضي مع استخدام نفس كمية السماد فحصل على النتائج التالية:

1511	1316	1612	1444	2060	1961	2108	2496	1910	1935	1903	محصول النوع الأول
1535	1443	1542	1482	2122	1925	2180	2463	2011	1915	2009	محصول النوع الثاني
-24	-127	70	-38	-62	36	-72	33	-101	20	-106	الفرق

الحل:

$$H_0: \mu_d = 0$$

$$H_1: \mu_d \neq 0$$

بحساب متوسط الفروقات نجد :

$$\bar{d} = \frac{\sum d}{n} = \frac{-371}{11} = -33.727$$

$$S_d^2 = \frac{\sum (d - \bar{d})^2}{n - 1} = \frac{(-106 - (-33.727))^2 + (20 - (-33.727))^2 + \dots}{10} = 4378.618$$

$$\Rightarrow S_d = \sqrt{4378.618} = 66.171$$

$$t = \frac{\bar{d} - \mu_d}{S_d / \sqrt{n}} = \frac{-33.727 - 0}{66.171 / \sqrt{11}} = -1.69$$

وبمقارنة القيمة المطلقة لـ t مع القيمة الحرجة المعيارية $t_{\alpha/2}$ المقابل لنصف مستوى الدلالة 0.05 و 10 درجات حرية و الذي يساوي $t_{0.025}(10) = 2.22$ نجد أن القيمة المطلقة لـ t أصغر من $t_{\alpha/2}$. إذن نقبل الفرضية العدمية والقائلة بعدم وجود فرق جوهري بين نوعي البذار أو بتعبير آخر لا يؤثر نوع البذار على كمية المحصول.

4.3. مقارنة الاختلاف (التباين) بين عينتين

عندما يراد مقارنة تشتت عينتين نستخدم تبايني هاتين العينتين. ويشترط لإجراء هذا الاختبار أن يكون المجتمعان الذان سحبت منهما العينتان مستقلين وأن يكونا ذو توزيع طبيعي. إذا تحققت هذه الشروط يمكن مقارنة هذين المجتمعين باستخدام اختبار فيشر التالي:

$$F = \frac{S_1^2}{S_2^2} \sim F(n_1 - 1; n_2 - 1)$$

حيث أن S_1^2 هو التباين الأكبر.

ولإتخاذ القرار المناسب تتم مقارنة قيمة F المحسوبة سابقا مع قيمة F النظرية المقابلة ل $n_1 - 1$ درجة حرية في البسط و $n_2 - 1$ درجة حرية في المقام $F_{\alpha/2}(n_1 - 1; n_2 - 1)$. فإذا كانت $F > F_{\alpha/2}$ فإننا نرفض H_0 ونقبل H_1 . أما إذا كانت أصغر منها فإننا نقبل H_0 .

مثال 19 :

لاختبار تجانس تباين الوزن لدى طلاب الجامعة رأينا أن نقسم المجتمع إلى مجتمعين :
(1) مجتمع الذكور (2) مجتمع الإناث . وأن نضع الفرضيتين كما يلي :

$$H_0 : \sigma_1^2 = \sigma_2^2$$

$$H_1 : \sigma_1^2 > \sigma_2^2$$

ولنفترض أننا سحبنا عينتين عشوائيتين : من الذكور بحجم $n_1 = 25$ ومن الإناث بحجم $n_2 = 20$ فوجدنا أن تباين العينتين هو $S_1^2 = 64$ و $S_2^2 = 25$.
ولإجراء الاختبار نحسب قيمة مؤشر الاختبار F فنجد أن :

$$F = \frac{S_1^2}{S_2^2} = \frac{64}{25} = 2.56$$

ولإجراء المقارنة نبحث في جدول التوزيع F عن قيمة F الحرجة المقابلة لدرجتي الحرية $n_1 - 1 = 24$ و $n_2 - 1 = 19$ ولمستوى الدلالة $\alpha = 0.05$ فنجد أن :

$$F_\alpha = 2.11$$

وبذلك نجد أن قيمة F المحسوبة (2.56) أكبر من قيمة F_α الحرجة (2.11) . ولهذا فإننا نرفض فرضية العدم H_0 ونقبل الفرضية البديلة التي تقول أن تباين الأوزان لدى الطلاب الذكور أكبر من تباين الأوزان لدى الطالبات الإناث .

تمريعات

- 1- سحبا عينتين بحجمين $n_1 = n_2 = 11$ عنصراً من مجتمعين طبيعيين متوسطيهما $\bar{Y}_1$, $\bar{Y}_2$ ولها تباين موحد. فوجدنا أن $\bar{x}_1 = 60$, $\bar{x}_2 = 65$ وأن $s_1^2 = 31$ وأن $s_2^2 = 45$ فهل تقدم هذه المعلومات دلالة كافية ، وباحتمال قدره 0.90 ، على وجود فرق بين المتوسطين $\bar{Y}_1$, $\bar{Y}_2$
- 2- لتقدير نسبة الأمية في مدينتين A , B سحبا منهما عينتين بحجمين $n_1 = 30$ و $n_2 = 40$ عاملاً. فوجدنا أن نسبة الأميين فيهما $r_1 = 0.10$, $r_2 = 0.12$ فهل هذا يعطينا دلالة كافية لاعتبار هاتين النسبتين متساويتين وذلك بمستوى دلالة 0.08.

المحاضرة التاسعة

الفصل الرابع

الإرتباط والإنحدار

Correlation and Regression

مفردات البحث

- 1- مفهوم الارتباط .
- 2- معامل الارتباط الخطي
- 3- مفهوم الانحدار الخطي
- 4- تفسير معادلة الانحدار و استخدامها في التنبؤ
- 5- مجالات التغير والتنبؤ

المرجع المعتمد: المرجع السابق من الصفحة 424 إلى الصفحة 487

1.4. مفهوم الارتباط (Correlation)

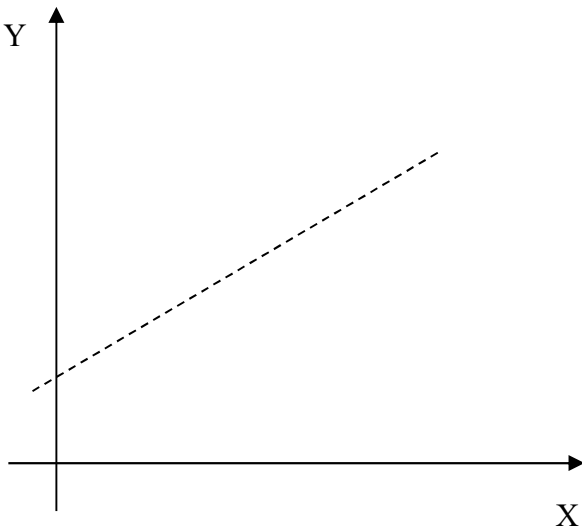
يتناول هذا الفصل دراسة العلاقة الارتباطية بين متحولين. فنقول أن متغيرين مرتبطان عندما يترافق تغير أحدهما بتغير الآخر. ولا يعني الارتباط بالضرورة أن أحد المتغيرين يؤثر أو يتأثر بالآخر بل يكفي أن يكون تغير أحدهما مترافق مع تغير الآخر.

لنفرض أنه لدينا متغيرين X و Y وأننا استطعنا الحصول على سلسلة من القياسات لهذين المتغيرين بحيث كانت كالتالي:

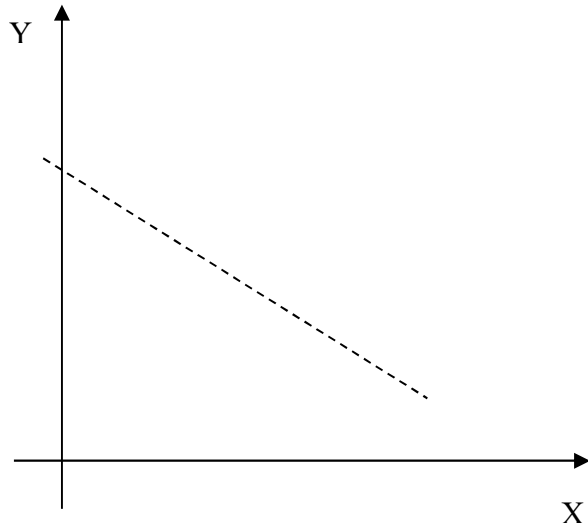
$$X : x_1, x_2, \dots, x_n$$

$$Y : y_1, y_2, \dots, y_n$$

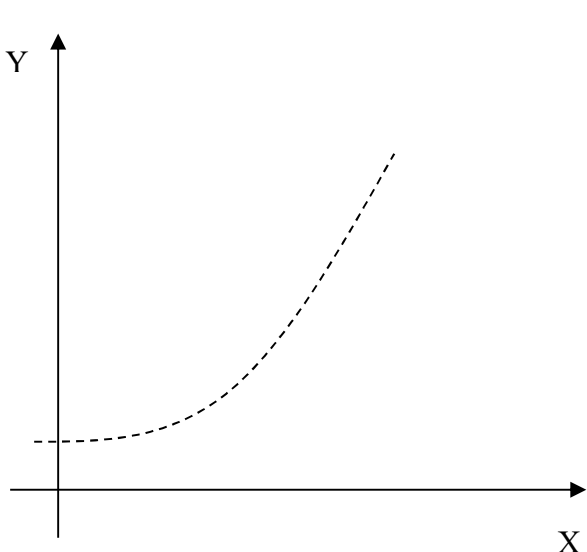
بتمثيل هذه النقاط على المحاور الاحداثية نحصل على ما يسمى بشكل الانتشار. ويعطينا شكل الانتشار فكرة أولية عن قوة الارتباط واتجاهه. ويمكن أن نحصل على أحد الأشكال التالية:



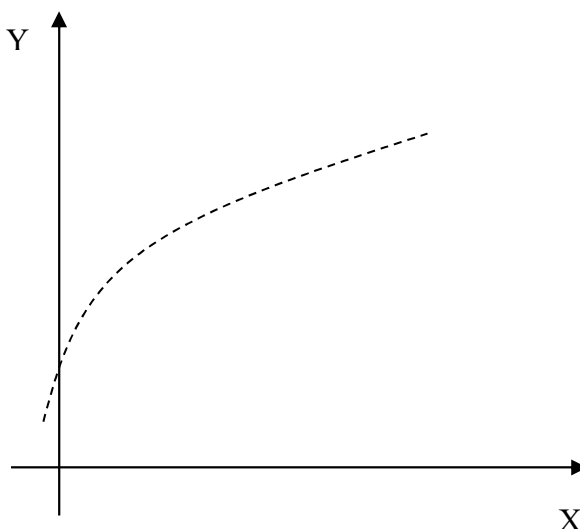
العلاقة طردية وخطية بين المتغيرين



العلاقة عكسية وخطية بين المتغيرين



العلاقة طردية و غير خطية بين المتغيرين



العلاقة طردية و غير خطية بين المتغيرين

بالطبع هناك أشكال أخرى كثيرة للعلاقات بين المتغيرات ولكن نكتفي بذكر هذه الأنواع للتوضيح وسوف نقتصر في دراستنا هذه على العلاقات الخطية.

إن شكل الانتشار يعطينا فكرة أولية عن العلاقة ولكنها غير واضحة وغير دقيقة. لكي نتأكد من شكل

العلاقة واتجاهها يمكن أن نحسب معامل الارتباط البسيط أو ما يسمى معامل بيرسون (Pearson's Coefficient).

2.4. معامل الارتباط الخطي (معامل بيرسون Pearson's Coefficient) .

ويستخدم هذا المعامل للكشف عن قوة العلاقة الارتباطية بين متحولين واتجاه هذه العلاقة. نرمز لمعامل الارتباط بالرمز r و يعرف بالعلاقة التالية:

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{n \sigma_x \sigma_y}$$

كما يمكن كتابته بالشكل:

$$r = \frac{n \sum xy - (\sum x)(\sum y)}{\sqrt{n(\sum x^2) - (\sum x)^2} \sqrt{n(\sum y^2) - (\sum y)^2}}$$

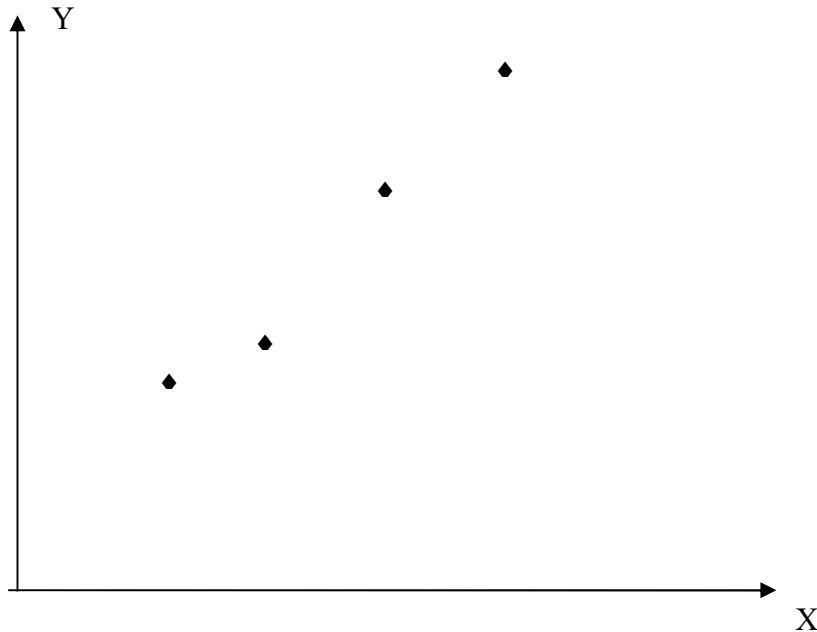
بعض خواص معامل الارتباط الخطي :

- 1- إن قيمة معامل الارتباط r محصورة في المجال $[-1, +1]$.
- 2- إذا كانت قيمة r قريبة من $(+1)$ فإنها تدل على ارتباط متين طردي (موجب) .
وإذا كانت قيمة r قريبة من (-1) فإنها تدل على ارتباط متين عكسي (سالب) .
وإذا كانت قيمة r قريبة من (0) فإنها تدل على عدم وجود ارتباط بين المتحولين .
- 3- إن قيمة معامل الارتباط لا تتأثر بواحدات القياس لأي من المتحولين y أو x . لذلك يمكننا استخدام أية وحدات قياس لهما عند حسابه .
- 4- إن قيمة معامل الارتباط تعبر عن متانة الارتباط الخطي فقط (أي الذي يكون على شكل مستقيم) ولا يصح تطبيقه على العلاقات غير الخطية.

مثال (20) : ارسم الشكل البياني (شكل الانتشار) ثم احسب معامل الارتباط الخطي r لعلاقة دخل الطبيب y مع عدد الزوار x من القيم المتقابلة لكل منهما خلال خمسة أيام والتي كانت كما يلي :

رقم اليوم	1	2	3	4	5
عدد الزوار x_i	2	4	6	8	10
قيمة الدخل (ألف) y_i	4	5	8	13	15

إن شكل الانتشار يشير إلى علاقة خطية بين المتغيرين.



ولحساب معامل الارتباط الخطي نعد الجدول المساعد التالي :

I	x_i	y_i	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
1	2	4	-4	-5	20	16	25
2	4	5	-2	-4	8	4	16
3	6	8	0	-1	0	0	1
4	8	13	2	4	8	4	16
5	10	15	4	6	24	16	36
المجموع Σ	30	45	0	0	60	40	94
النتائج	$\bar{x} = 6$	$\bar{y} = 9$				$\sigma_x^2 = 8$	$\sigma_y^2 = 18.8$

وبتعويض البيانات النهائية لهذا الجدول في العلاقة (4-5) نجد أن :

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{n \sigma_x \sigma_y} = \frac{60}{5 \times \sqrt{8} \times \sqrt{18.8}} = 0.978$$

وهي قيمة كبيرة وتدل على وجود ارتباط متين بين x و y وهو ارتباط طردي (موجب) .

اختبار معنوية معامل الارتباط الخطي

يمكن اختبار معنوية معامل الارتباط لمعرفة فيما إذا كان الارتباط بين متحولين معنوي أم لا . ويمكن إجراء هذا الاختبار بطريقتين:

الطريقة الأولى:

تعتمد الطريقة الأولى على اعتبار معامل الارتباط متحولاً عشوائياً، يخضع لتوزيع احتمالي

معين، وإن متوسطه أو مركزه يساوي قيمة ما ρ وتباينه يساوي $\sigma^2 = \frac{1-r^2}{n-2}$. لذلك نستخدم الاختبار

التالي :

يوجد ارتباط خطي معنوي $H_0 : \rho = 0$

لا يوجد ارتباط خطي معنوي $H_1 : \rho \neq 0$

ونستخدم مؤشر الاختبار التالي:

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} \sim T(n-2)$$

وبمقارنة قيمة هذا المؤشر مع القيمة الحرجة $t_{\alpha/2}$ المقابلة ل $(n-2)$ درجة حرية نقبل أو نرفض H_0 . فإذا كانت القيمة المطلقة لـ t المحسوبة أصغر من $t_{\alpha/2}$ الجدولية نقبل الفرضية العدمية H_0 ونقبل بعدم وجود ارتباط خطي معنوي بين المتحولين المدروسين. أما إذا كانت $|t| > t_{\alpha/2}$ فإننا نرفض H_0 ونقر بمعنوية قيمة r وبوجود ارتباط بين المتحولين x, y .

الطريقة الثانية :

تعتمد هذه الطريقة على مقارنة قيمة r مع قيمة حرجة معطاة في جدول خاص. فإذا كان $|r| > r_\alpha$ فإننا نرفض H_0 ونستنتج بأن الارتباط الخطي معنوي. وإذا كان $|r| \leq r_\alpha$ فإننا نقبل H_0 ونستنتج بأن الارتباط الخطي غير معنوي.

مثال 21 :

لنفترض أن عينة من $n=10$ أزواج من الذكور والإناث أظهرت أن معامل الارتباط بين أطوالهم كان $r = 0.7$ و المطلوب اختبار معنوية هذا المعامل بمستوى دلالة $\alpha = 0.05$.

يوجد ارتباط خطي معنوي $H_0 : \rho = 0$

لا يوجد ارتباط خطي معنوي $H_1 : \rho \neq 0$

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} = \frac{0.7}{\sqrt{\frac{1-(0.7)^2}{10-2}}} = 2.72$$

ومن جداول توزيع ستودينت نجد أن قيمة t المقابلة لـ 8 درجات حرية ولنصف مستوى الدلالة $\alpha/2 = 0.025$ ، تساوي 2.306 . وبمقارنة القيمة المحسوبة لـ t مع القيمة الحرجة لها نجد أن $t > t_{\alpha/2}$ ، لذلك نرفض الفرضية H_0 ونعتبر أن قيمة معامل الارتباط هي قيمة معنوية أي هناك ارتباط معنوي بين أطوال الذكور والإناث.

المحاضرة العاشرة

4. مفهوم الانحدار الخطي (Linear Regression)

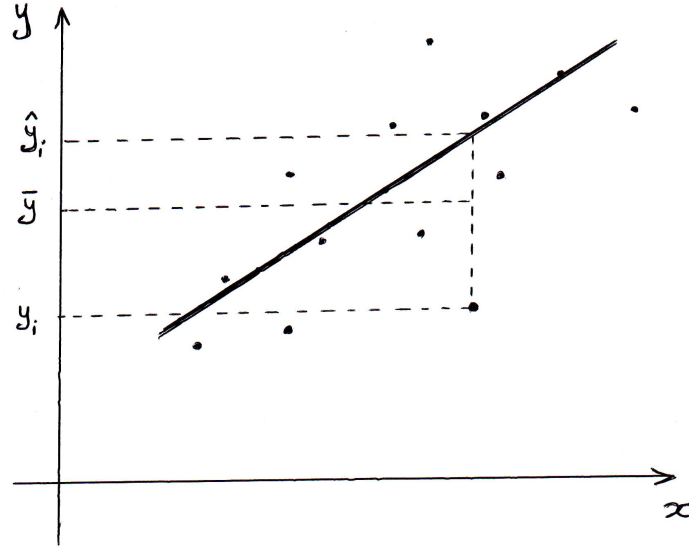
رأينا فيما سبق أننا نستطيع التحقق من وجود ارتباط خطي بين متغيرين باستخدام معامل الارتباط الخطي. يمكننا هذا المعامل من معرفة فيما إذا كان هذان المتحولان يتغيران بنفس الاتجاه دون معرفة فيما إذا كان أحدهما يتأثر بالآخر أو أي منهما المتغير التابع وأي منهما المؤثر. في سبيل معرفة ذلك يمكن استخدام طريقة أخرى لدراسة العلاقة الارتباطية بين المتغيرات وهي طريقة الانحدار.

في طريقة الانحدار نقوم بتمثيل العلاقات الارتباطية بين متحولين x و y بواسطة معادلة المستقيم إذا كان شكل الانتشار للنقاط الفعلية المتقابلة (x_i, y_i) شبيهاً بخط مستقيم أو تنتشر حول خط مستقيم. ونفترض في هذه العلاقة أن المتغير X هو المتغير الذي يؤثر على المتغير الآخر ونطلق عليه اسم المتغير المستقل (Independent Variable) أو المتغير التفسيري (Explicative Variable) أو المتغير الخارجي (Exogenous Variable). بينما نسمي المتغير Y بالمتغير التابع (Dependent Variable) أو المتغير المفسر (Explaid Variable) أو المتغير الداخلي (Endogenous Variable). وعندها نفترض أن معادلة التمثيل (بدلالة القيم المتقابلة) لها الشكل التالي :

$$\hat{y}_i = \hat{\alpha} + \hat{\beta}x_i$$

حيث أن $\hat{\alpha}$ و $\hat{\beta}$ هي القيم المقدرة للمعاملات α و β المجهولتان.

نلاحظ من الرسم البياني أن المستقيم الممثل للعلاقة لا يمر بجميع النقاط الفعلية (القيم الحقيقية). ويكون هذا المستقيم ممثلاً جيداً للعلاقة بين هذين المتغيرين بقدر ما يقترب من النقاط. تحدد قيمة كل من $\hat{\alpha}$ و $\hat{\beta}$ شكل المستقيم ودرجة ميلانه لذلك عند اختيار هذه القيم (تقديرها) نحرص على أن يكون الفرق بين القيم المقدرة بحسب العلاقة والقيم الفعلية $(y_i - \hat{y}_i)$ أقل ما يمكن. أي بعبارق أخرى أن تكون الأخطاء المرتكبة أقل ما يمكن.



لتقدير معاملات العلاقة الخطية السابقة هناك طرق متعددة أهمها طريقة المربعات الصغرى (Ordinary leqst Squares). وتعتمد هذه الطريقة على تصغير مجموع مربعات الفروق بين القيم الفعلية والمقدرة الذي يؤول إلى حل جملة المعادلتين التاليتين حلا مشتركا باحدى الطرق المعروفة:

$$\begin{aligned}\sum y_i &= n\alpha + \beta \sum x_i \\ \sum y_i x_i &= \alpha \sum x_i + \beta \sum x_i^2\end{aligned}$$

وتسمى هاتان المعادلات بالمعادلات الطبيعية.

مثال (22) : أوجد معادلة المستقيم لتمثيل علاقة دخل الطبيب y مع عدد الزوار x خلال عشرة أيام والتي كانت كما يلي :

اليوم	1	2	3	4	5	6	7	8	9	10
عدد الزوار x	5	9	6	8	10	11	12	7	4	3
الدخل y	4	5	8	13	15	13	12	9	6	8

الحل : لنفترض أن معادلة التمثيل من الشكل $\hat{y}_i = \hat{\alpha} + \hat{\beta}x_i$

ولحساب الثابتين $\hat{\alpha}$ و $\hat{\beta}$ نعد الجدول التالي .

رقم اليوم	1	2	3	4	5	6	7	8	9	10	Σ
x_i	5	9	6	8	10	11	12	7	4	3	75
x_i^2	25	81	36	64	100	121	144	49	16	9	645
y_i	4	5	8	13	15	13	12	9	6	8	93
$x_i y_i$	20	45	48	104	150	143	144	63	24	24	765

وبالتالي يكون لدينا :

$$93 = 10\alpha + 75\beta$$

$$765 = 75\alpha + 645\beta$$

وبالحل المشترك نجد: $\beta = 0.8$ و $\alpha = 3.1$ وبالتالي تكتب معادلة التمثيل على الشكل التالي:

$$\hat{y}_i = 3.1 + 0.8x_i$$

طريقة أخرى لإيجاد معاملات العلاقة الخطية باستخدام معامل الارتباط

يمكن إيجاد معاملات العلاقة الخطية باستخدام معامل الارتباط كالتالي:

$$\beta = r_{xy} \frac{\sigma_y}{\sigma_x}$$

$$\alpha = \bar{y} - \beta \bar{x} \quad \text{و}$$

مثال (23) :

لدينا البيانات التالية عن حجم الاستثمار في القطاع الصحي وحجم العائد من هذا الاستثمار :

حجم الاستثمار	حجم العائد
5	20
7	40
8	55
10	70
15	85

والمطلوب إيجاد معامل الارتباط الخطي ثم أوجد العلاقة التي تبين تأثير الاستثمار على العائد (أوجد معاملات هذه العلاقة) إذا علمنا أن العلاقة بين حجم الاستثمار وحجم الناتج المحلي هي علاقة خطية.

الحل : باستخدام الجدول المساعد التالي:

i	X	Y	$(X - \bar{X})$	$(Y - \bar{Y})$	$(X - \bar{X})(Y - \bar{Y})$	$(X - \bar{X})^2$	$(Y - \bar{Y})^2$
1	5	20	4-	34-	136	16	1156
2	7	40	2-	14-	28	4	196
3	8	55	1-	1	1-	1	1
4	10	70	1	16	16	1	256
5	15	85	6	31	186	36	961
Σ	45	270	0	0	365	58	2570
المتوسط	9	54				$\sigma_x^2 = 11.6$	$\sigma_y^2 = 514$

يمكن أن نكتب

$$R_{xy} = \frac{\sum (x - \bar{x})(y - \bar{y})}{n \sigma_x \sigma_y} = \frac{365}{5 \sqrt{11.6} \sqrt{514}} = 0.945$$

إن قيمة معامل الارتباط قريبة من الواحد مما يدل على أن الارتباط قوي بين المتحولين. كما أن القيمة موجبة مما يعني أن الارتباط طردي أو موجب .

المعادلة من الشكل $y = \alpha + \beta x$ و يمكن إيجاد المعاملات كالتالي :

$$\beta = R_{xy} \frac{\sigma_y}{\sigma_x} = 0.945 \frac{\sqrt{514}}{\sqrt{11.6}} = 6.3$$

$$\alpha = \bar{y} - \beta \bar{x} = 54 - 6.3 \times 9 = -2.7$$

$$y = 6.3x - 2.7$$

إذا العلاقة من الشكل

4.4. تفسير معادلة الانحدار و استخدامها في التنبؤ

رأينا سابقا أن معادلة الانحدار الخطي تكتب على الشكل:

$$y = \alpha + \beta x$$

تمثل β قيمة التغير الحدي أو الهامشي (Marginal Change) في المتغير التابع Y . أي مقدار التغير في قيمة Y عندما تتغير X بمقدار وحدة واحدة. بينما تمثل α قيمة المتغير Y عندما تكون قيمة X معدومة.

قبل التنبؤ بقيم Y يجب معرفة قدرة المعادلة السابقة على عكس حقيقة العلاقة بين المتغيرين. لأي بمعنى آخر ما نسبة ما يفسره المتغير X من التغيرات الكلية للمتغير Y . إن معرفة هذه النسبة يركز على مفهوم معامل التحديد.

نسمي الفرق بين قيم y الفعلية ومتوسطها $\bar{y}$ بالانحراف الكلي للمتغير Y (Total Deviation). هذا الانحراف عن المتوسط يمكن أن يقسم إلى قسمين: انحراف مفسر (Explained Deviation) وهو عبارة عن الفرق بين القيم المقدرة للمتغير التابع $\hat{y}$ ومتوسطها $\bar{y}$ أي $(\hat{y} - \bar{y})$. ويسمى بالانحراف المفسر لأن ما يفسره هو خط الانحدار. فالقيم المقدرة هي القيم الناتجة عن معادلة الانحدار. القسم الثاني هو الانحراف غير المفسر (Unexplained Deviation) وهو الفرق بين القيم الفعلية والقيم المقدرة $(y - \hat{y})$ ولا يعرف مصدر هذا الانحراف لذلك سمي بانحراف غير مفسر.

بناءً عليه يمكن أن نقسم التباين الكلي للمتغير التابع (Total Variation) إلى تباين مفسر (Explained Variation) وتباين غير مفسر (Unexplained Variation) . أي :

$$(Total Variation) = (Explained Variation) + (Unexplained Variation)$$

ونعبر عن ذلك رياضياً بالعلاقة:

$$\sum (y - \bar{y})^2 = \sum (\hat{y} - \bar{y})^2 + \sum (y - \hat{y})^2$$

وبالتالي كلما كان التباين المفسر كبيراً مقارنة بالتباين غير المفسر كلما كانت معادلة الانحدار الخطي ممثلة جيداً للعلاقة بين المتغيرين. لذلك نعرف معامل التحديد (Coefficient of Determination) بالعلاقة التالية:

$$r^2 = \frac{Explained\ Variation}{Total\ Variation} = \frac{\sum (\hat{y} - \bar{y})^2}{\sum (y - \bar{y})^2}$$

يبين هذا المعامل مقدار التغير في المتحول Y المفسر بمعادلة الانحدار الخطي. فمثلا إذا كان معامل الارتباط الخطي بين متغيرين يساوي $r = 0.95$ فإن معامل التحديد يساوي $r^2 = (0.95)^2 = 0.9$ أي أنه بالمتوسط فإن 90 % من التغير الكلي في المتحول Y يمكن أن يفسر بالعلاقة $y = \alpha + \beta x$. وبالتالي هناك 10 % من تغيرات Y تبقى دون تفسير.

بعد معرفة مدى جودة العلاقة الخطية لتمثيل العلاقة بين المتغيرين X و Y يمكن استخدام هذه العلاقة في التنبؤ. أي في التنبؤ بقيم المتغير التابع وفقا لقيم محددة للمتغير المستقل. والتنبؤ له نوعان تنبؤ نقطي وتنبؤ مجالي:

التنبؤ النقطي بقيم Y (Point Estimate)

من خلال معادلة الانحدار الخطي وبعد إيجاد المقدرات $\hat{\alpha}$ و $\hat{\beta}$ يمكن تقدير قيم المتغير Y المقابلة لقيم محددة ل X . فمثلا بالرجوع إلى المثال السابق وجدنا أن العلاقة بين حجم الاستثمار في القطاع الصحي وحجم العائد من هذا الاستثمار كانت ممثلة بالعلاقة $y = 6.3x - 2.7$. لو أردنا معرفة حجم العائد من استثمار 20 وحدة نقدية لوجدنا أنها تساوي:

$$\hat{y} = 6.3x - 2.7 = 6.3 \times 20 - 2.7 = 123.3$$

التنبؤ المجالي بقيم Y (Prediction Interval)

ذكرنا سابقا عند بحث مجالات الثقة أن التقدير النقطي غير دقيق أحيانا وفي هذه الحالة يمكن إيجاد مجال للتنبؤ بقيم المتغير Y .

بداية نعرف الخطأ المعياري للتقدير (Standard Error of Estimate) بالعلاقة :

$$S_e = \sqrt{\frac{\sum (y - \hat{y})^2}{n - 2}} = \sqrt{\frac{\sum y^2 - \alpha \sum y - \beta \sum xy}{n - 2}}$$

وبالتالي يمكن إيجاد مجال ثقة للمتغير y كالتالي:

$$\hat{y} - E < y < \hat{y} + E$$

حيث أن E تمثل الخطأ في تقدير y عندما تكون x مساوية لقيمة محددة ولتكن x_0 وهي تساوي:

$$E = t_{\alpha/2} S_e \sqrt{1 + \frac{1}{n} + \frac{n(x_0 - \bar{x})^2}{n(\sum x^2) - (\sum x)^2}}$$

و بالتالي نستطيع كتابة مجال الثقة للقيمة y كالتالي:

$$\hat{y} - E = t_{\alpha/2} S_e \sqrt{1 + \frac{1}{n} + \frac{n(x_0 - \bar{x})^2}{n(\sum x^2) - (\sum x)^2}} < y < \hat{y} + E = t_{\alpha/2} S_e \sqrt{1 + \frac{1}{n} + \frac{n(x_0 - \bar{x})^2}{n(\sum x^2) - (\sum x)^2}}$$

أي أنه عندما تكون x مساوية لقيمة محددة x_0 تكون قيمة y واقعة في هذا المجال باحتمال قدره $(1 - \alpha)$.

مثال 24 :

لدراسة جودة تمثيل المعادلة التي حصلنا عليها في المثال السابق نقوم بإيجاد القيم النظرية لـ $\hat{y}_i$ من معادلة التمثيل المحسوبة ثم نقوم بحساب مجموع مربعات انحرافات القيم النظرية $\hat{y}_i$ عن القيم الفعلية المقابلة y_i لذلك نعد الجدول المساعد التالي :

I	1	2	3	4	5	6	7	Σ
Y_i القيم الفعلية لـ	5	7	9	10	11	13	15	70
القيم المقطرة من المعادلة $\tilde{y}_i = 0.786x_i + 3.716$	5.29	6.86	8.43	10	11.57	13.14	14.71	70
$y_i - \tilde{y}_i$	-0.29	0.14	0.57	0	0.57	-0.14	0.29	0
$(y_i - \tilde{y}_i)^2$	0.0841	0.0196	0.325	0	0.325	0.0196	0.0841	0.8574

ومن نتائج حسابات هذا الجدول نجد أن تباين التمثيل (التباين غير المفسر) يساوي :

$$S^2 = \frac{\sum (y_i - \tilde{y}_i)^2}{n-2} = \frac{0.8574}{5} = 0.171$$

ولقد وجدنا سابقاً أن التباين الكلي لـ Y يساوي : $\sigma_y^2 = 10$.

وبذلك نجد أن قيمة معامل التحديد R^2 تساوي :

$$R^2 = 1 - \frac{S^2}{\sigma_y^2} = 1 - \frac{0.171}{10} = 0.982$$

وهي قيمة قريبة جداً من الواحد وهي تدل على أن جودة تمثيل المعادلة للعلاقة بين Y و X هو تمثيل جيد جداً . وهو يعني أن تغيرات X تفسر 98.29% من تغيرات Y وإن الباقي يتم بسبب تغيرات المتحولات الأخرى غير الداخلة في معادلة التمثيل .

تمرينات:

- 1- لنفترض أنه لدينا المعلومات التالية عن عدد السكان من القطر وقيمة الناتج الإجمالي (مليار ل.س) خلال الفترة 1980-1986 :

العام	1980	1981	1982	1983	1984	1985	1986
عدد السكان x مليون نسمة	8	9	9.3	9.6	10	10.3	10.6
الناتج الاجمالي y مليار ل.س	51	56	58	59	57	58.5	58.7

المصدر : المجموعة الإحصائية 1988 ص 5-502

- والمطلوب : 1- رسم شكل الانتشار لـ x و y واقتراح نوع معادلة التمثيل .
- 2- حساب معامل الارتباط بين x و y .
- 3- ايجاد معادلة التمثيل الخطية بين x و y .
- 4- دراسة جودة التمثيل للمعادلة السابقة .
- 5- التنبؤ بقيمة الناتج عندما يصبح عدد السكان 12 مليونا

المحاضرة الحادية عشر

الفصل الخامس

الاحصاء الالاعلمي

Nonparametric Statistics

مفردات البحث

1- اختبار الإشارة

2- اختبار ويلكوكسون للأزواج المتقابلة

3- اختبار ويلكوكسون للرتب في عينتين مستقلتين

4- اختبار كروسكال والس

5- اختبارات الارتباط الرتبي

المرجع المعتمد: المرجع السابق من الصفحة 566 إلى الصفحة 609

1- اختبار الإشارة Sign Test

يعتبر اختبار الإشارة من الاختبارات الهامة في الدراسات المتعلقة بالمعطيات النوعية. ويعتمد على مقارنة إشارة الفروقات بين أزواج العناصر المتقابلة. ويستخدم بشكل رئيسي في اختبار الفرضيات المتعلقة بالأزواج المتقابلة في العينة، اختبار الفرضيات المتعلقة بالمعطيات الاسمية، الفرضيات المتعلقة بوسيط مجتمع. واستخدام هذا الاختبار يتطلب أن تكون المعطيات قد سحبت بشكل عشوائي ولكنه لا يتطلب أن يكون للمجتمع توزيع معين كالتوزيع الطبيعي مثلا.

يتحقق هذا الاختبار من فرضية العدم أن مجموع الإشارات السالبة مساوية لعدد الإشارات الموجبة أي لا يوجد فروق جوهرية وإنما هذه الفروق المشاهدة يمكن أن توجد صدفة.

يميز اختبار الإشارة بين العينات الصغيرة والعينات الكبيرة:

حالة العينات الصغيرة ($n \leq 25$)

إذا رمزنا x^+ لعدد الإشارات الموجبة و x^- لعدد الإشارات السالبة، يكتب مؤشر الاختبار على الشكل التالي:

$$x = \min(x^+, x^-)$$

وباستخدام جدول التوزيع الثنائي (أو جداول أخرى مماثلة) يتم إيجاد الاحتمال المقابل للقيمة x ولحجم العينة n . ويتم مقارنة هذا الاحتمال مع مستوى الدلالة المفروض (0.01, 0.05, 0.1). فإذا كان هذا الاحتمال أصغر من 0,05 مثلا فإننا نرفض فرضية العدم التي تنص على عدم وجود فروق جوهرية.

حالة العينات الكبيرة ($n > 25$)

في حال كانت العينة كبيرة نستخدم مؤشر الاختبار التالي:

$$z = \frac{(x + 0.5) - \frac{n}{2}}{\frac{\sqrt{n}}{2}}$$

وهذا المؤشر يتبع التوزيع الطبيعي المعياري لذلك تتم مقارنة قيم المؤشر مع القيم الحرجة للتوزيع الطبيعي المعياري.

ملاحظة: عندما يكون الفرق بين المشاهدات معدوم تستبعد هذه المشاهدات من العينة.

مثال 30 :

بهدف معرفة تأثير نوع البذور على إنتاج الذرة قام وليام غوسيه (Student) باستخدام المعطيات التالية التي تبين إنتاج نوعين من البذور في وحدة المساحة (الهكتار).

الإنتاج (كيلو غرام/هكتار)											
1511	1316	1612	1444	2060	1961	2108	2496	1910	1935	1903	النوع الأول
1535	1443	1542	1482	2122	1925	2180	2463	2011	1915	2009	النوع الثاني

بناءً على المعطيات السابقة هل نستطيع القول أن لنوع البذور تأثير على حجم إنتاج الذرة.

الحل:

H_0 : لا يوجد فروق. (وسيط الفروقات يساوي الصفر)

H_1 : يوجد فروق. (وسيط الفروقات مختلف عن الصفر)

مؤشر الاختبار هو عدد الإشارات الأقل تكراراً وهو في مثالنا يساوي 4 .

الإنتاج (كيلو غرام/هكتار)											
1511	1316	1612	1444	2060	1961	2108	2496	1910	1935	1903	النوع الأول
1535	1443	1542	1482	2122	1925	2180	2463	2011	1915	2009	النوع الثاني
-	+	-	+	-	+	-	-	+	-	-	الفرق

ومن جدول القيم الحرجة لإختبار الإشارة نلاحظ أن القيمة الحرجة المقابلة لحجم عينة 11 ومستوى دلالة 0,05 هي القيمة 1 . لذلك نقبل فرضية العدم بعدم وجود فروق جوهرية بين النوعين.

2- اختبار ويلكوكسون للأزواج المتماثلة Wilcoxon Signed-Ranks Test for Matched Pairs

يستخدم اختبار ويلكوكسون في الأزواج المتماثلة في معطيات العينات المؤلفة من أزواج متماثلة أو متقابلة. وبشكل أو باخر تكون مرتبطة. ويعتمد على اختبار فرضية أن مجتمع الفروقات لهذه الأزواج له وسيط يساوي الصفر. أي:

الأزواج المتماثلة ذات فروقات مسحوبة من مجتمع له وسيط يساوي الصفر : H_0

الأزواج المتماثلة ذات فروقات مسحوبة من مجتمع له وسيط مختلف عن الصفر : H_1

كما يستخدم هذا الإختبار لمعرفة فيما إذا كانت عينة قد سحبت من مجتمع له وسيط محدد.

طريقة إجراء الإختبار:

1- نحسب الفرق بين كل زوج من المشاهدات ولنرمزه d . لكن يجب الإنتباه إلى استبعاد الأزواج التي تكون فروقاتها معدومة.

2- نعطي هذا الفروقات الترتيب المناسب مع اهمال اشارات هذه الفروق. وإذا تساوت الرتب نأخذ متوسطها.

3- نعطي هذه الرتب الإشارات المقابلة التي أهملناها سابقا.

4- نقوم بحساب مجموع الرتب لكل اشارة (السالبة والموجبة)

5- نسمي T مجموع الرتب الأقل. ويكون هو مؤشر الإختبار

6- نرفض H_0 في حال كان $T \geq T_\alpha$. حيث يتم إيجاد T_α من الجدول الموافق لهذا الإختبار.

شروط إجراء الإختبار:

1- المشاهدات على شكل أزواج متماثلة

2- مجتمع الفروقات له توزيع شبه متناظر أو متناظر حول الوسيط

يميز اختبار ويلكوسون للرتب للأزواج المتقابلة بين العينات الصغيرة والكبيرة. ففي حال كانت العينة صغيرة نطبق الإختبار السابق. أما في حال كانت العينة كبيرة (أكبر من 30) نستخدم مؤشر الإختبار التالي:

$$z = \frac{T - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$$

الذي يتبع التوزيع الطبيعي المعياري $N(0,1)$

مثال 31:

لنستخدم معطيات المثال السابق ولنختبر فيما إذا كان هناك فروق جوهرية محصول البذار الأول والثاني باستخدام اختبار ويلكوسون للأزواج المتماثلة.

الحل:

الإنتاج (كيلو غرام/هكتار)											
1511	1316	1612	1444	2060	1961	2108	2496	1910	1935	1903	النوع الأول
1535	1443	1542	1482	2122	1925	2180	2463	2011	1915	2009	النوع الثاني
-106	+ 20	- 101	+ 33	- 72	+ 36	- 62	- 38	+ 70	- 127	- 24	الفرق d
10	1	9	3	8	4	6	5	7	11	2	رتب الفرق
-10	1	-9	3	-8	4	-6	-5	7	-11	-2	رتب مع إشارة

لا يوجد فروق بين المحصولين (وسيط مجتمع الفروقات يساوي الصفر) : H_0

يوجد فروق بين المحصولين (وسيط مجتمع الفروقات مختلف عن الصفر) : H_1

نلاحظ أن مجموع رتب الإشارة الأقل يساوي 15 . والقيمة الحرجة المقابلة لحجم عينة 11 ومستوى دلالة 0,05 هو 11. وبما أن قيمة المؤشر أكبر من القيمة الحرجة فإننا نقبل فرضية العدم . أي لا يوجد فروق ذات دلالة احصائية بين نوعي البذور.

المحاضرة الثانية عشر

3- اختبار ويلكوكسون للرتب في عينتين مستقلتين Wilcoxon Ranks-Sum Test for Two Independent Samples

يستخدم هذا الاختبار للتحقق من أن عينتين قد سحبت من مجتمعين لهما نفس الوسيط. أي :

العينتان سحبت من مجتمعين لهما نفس الوسيط : H_0

العينتان سحبت من مجتمعين لهما وسيطان مختلفان: H_1

طريقة إجراء الاختبار:

1- نعتبر أن العينتان هما عينة واحدة ونقوم بترتيب المشاهدات من الأدنى إلى الأعلى (إعطائها رتب من 1, 2, ... وهكذا) . فإذا كانت الرتب متساوية نأخذ متوسطها.

2- نقوم بحساب مجموع الرتب لكل عينة.

3- نقوم بحساب قيمة Z كما سنرى لاحقا.

شروط إجراء الاختبار:

1- يشترط أن تكون العينتان مستقلتان

2- عدد المشاهدات في كل عينة لا يقل عن 10

3- لا يوجد فرضيات خاصة حول توزيع المجتمع

فإذا افترضنا أن :

n_1 حجم العينة الأولى

n_2 حجم العينة الثانية

R_1 مجموع رتب العينة الأولى

R_2 مجموع رتب العينة الثانية

R مجموع رتب أحد العينتين

μ_R متوسط قيم العينة المعتمدة (الأولى أو الثانية)

σ_R الانحراف المعياري للعينة المعتمدة

في هذه الحالة يكتب مؤشر الإختبار على الشكل التالي:

$$z = \frac{R - \mu_R}{\sigma_R}$$

$$\sigma_R = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}} \quad \text{و} \quad \mu_R = \frac{n_1 (n_1 + n_2 + 1)}{2} \quad \text{حيث}$$

وتحت فرضية العدم يتبع هذا المؤشر التوزيع الطبيعي المعياري $N(0,1)$

ملاحظة : يعتبر هذا الإختبار مكافئ لإختبار Mann-Whitney التالي:

$$z = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}} \quad \text{حيث} \quad U = n_1 n_2 + \frac{n_1 (n_1 + 1)}{2} - R$$

مثال 32:

ليكن لدينا عينتين مؤلفة من 13 رجل و 12 امرأة ونريد اختبار وجود فروق جوهرية في مقاييس والنساء:

الرجال	النساء
23.8	19.6
23.2	23.8
24.6	19.6
26.2	29.1
23.5	25.2
24.5	21.4
21.5	22
31.4	27.5
26.4	33.5
22.7	20.6
27.8	29.9
28.1	17.7
25.2	

الحل:

H_0 : العینتان لهما نفس الوسيط (الرجال والنساء لهما مقایيس متساوية) :

H_1 : العینتان لهما وسيطين مختلفين (الرجال والنساء لهما مقایيس مختلفة):

نقوم بحساب رتب العینتين على الشكل التالي:

الرجال		النساء	
القياس	الرتبة	القياس	الرتبة
23.8	11.5	19.6	2.5
23.2	9	23.8	11.5
24.6	14	19.6	2.5
26.2	17	29.1	22
23.5	10	25.2	15.5
24.5	13	21.4	5
21.5	6	22	7
31.4	24	27.5	19
26.4	18	33.5	25
22.7	8	20.6	4
27.8	20	29.9	23
28.1	21	17.7	1
25.2	15.5		
$n_1 = 13$	$R_1 = 187$	$n_2 = 12$	$R_2 = 138$

نقوم بحساب مؤشر الاختبار:

$$R = R_1 = 187$$

$$\mu_R = \frac{n_1(n_1 + n_2 + 1)}{2} = \frac{13(13 + 12 + 1)}{2} = 169$$

$$\sigma_R = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}} = \sqrt{\frac{(13)(12)(13+12+1)}{12}} = 18.38$$

$$z = \frac{R - \mu_R}{\sigma_R} = \frac{187 - 169}{18.38} = 0.98$$

نلاحظ أن قيمة مؤشر الاختبار أصغر من القيمة الحرجة بمستوى دلالة 0.05 (1.96). وبالتالي نقبل فرضية العدم بعدم وجود فروق ذات دلالة احصائية بين قيس الرجال والنساء.

4- اختبار كروسكال والس Kruscal-Wallis Test

يستخدم هذا الاختبار في حال كان لدينا عدة عينات مستقلة ونريد أن نختبر فيما إذا كانت هذه العينات مسحوبة من مجتمعات لها نفس الوسيط. أي:

H_0 : العينات سحبت من مجتمعات لها نفس الوسيط :

H_1 : العينات سحبت من مجتمعات لهما وسيطاء مختلفة:

ويعتبر هذا الاختبار بديل لاختبار تحليل التباين ANOVA الذي يشترط أن يكون توزيع المجتمعات طبيعي. بينما لا يشترط اختبار كروسكال والس مثل هذا الشرط.

طريقة إجراء الاختبار:

1- نعتبر أن العينات المختلفة تشكل عينة واحدة ونقوم بترتيب قيمها

2- نقوم بحساب مجموع الرتب لكل عينة

3- نقوم بحساب مؤشر الاختبار التالي:

$$H = \frac{12}{N(N+1)} \left(\frac{R_1^2}{n_1} + \frac{R_2^2}{n_2} + \frac{R_3^2}{n_3} + \dots + \frac{R_k^2}{n_k} \right) - 3(N+1)$$

حيث أن:

N مجموع أحجام العينات

R_i مجموع رتب العينة i

تحت فرضية العدم يتبع المؤشر السابق توزيع $\chi^2(k-1)$

مثال 33 :

قام أحد الباحثين بدراسة أثر كل من الماء والسماذ على وزن نوع من الفاكهة فحصل على النتائج التالية:

اضافة الماء والسماذ	اضافة السماذ	اضافة الماء	عدم اضافة شئى
2.03	0.23	1.34	0.15
0.27	0.04	0.14	0.02
0.92	0.34	0.02	0.16
1.07	0.16	0.08	0.37
2.38	0.05	0.08	0.22

فهل نستطيع القول أن هناك فروق جوهرية في تأثير الماء والسماذ على حجم الثمار.

الحل:

نقوم بحساب الرتب كالتالي:

اضافة الماء والسماذ		اضافة السماذ		اضافة الماء		عدم اضافة شئى	
الرتبة	الوزن	الرتبة	الوزن	الرتبة	الوزن	الرتبة	الوزن
19	2.03	12	0.23	18	1.34	8	0.15
13	0.27	3	0.04	7	0.14	1.5	0.02
16	0.92	14	0.34	1.5	0.02	9.5	0.16
17	1.07	9.5	0.16	5.5	0.08	15	0.37
20	2.38	4	0.05	5.5	0.08	11	0.22
$R_4 = 85$	$n_4 = 5$	$R_3 = 42.5$	$n_3 = 5$	$R_2 = 37.5$	$n_2 = 5$	$R_1 = 45$	$n_1 = 5$

$$\begin{aligned}
H &= \frac{12}{N(N+1)} \left(\frac{R_1^2}{n_1} + \dots + \frac{R_k^2}{n_k} \right) - 3(N+1) \\
&= \frac{12}{20(20+1)} \left(\frac{(45)^2}{5} + \frac{(37.5)^2}{5} + \frac{(42.5)^2}{5} + \frac{(85)^2}{5} \right) - 3(20+1) \\
&= 8.21
\end{aligned}$$

نلاحظ أن قيمة مؤشر الاختبار 8,21 أكبر من قيمة كاي مربع الجدولية بدرجة حرية 0,05 والتي تساوي 7,8 وبالتالي نرفض فرضية العدم بعدم وجود فروق جوهرية ونقبل الفرضية البديلة بأن هناك فروق جوهرية في وسطاء المجموعات الأربع.

ملاحظة: رفض فرضية العدم يعني أن هناك معالجة واحدة على الأقل تعطي نتائج مختلفة . لمعرفة أي منها هو سبب الاختلاف يجب مقارنة كل مجموعتين بشكل منفصل.

5- اختبارات الارتباط الرتبي Rank Correlation

هناك العديد من الاختبارات التي تستخدم لدراسة الارتباط بين المتحولات ولكن من أهمها :

Spearman's Rank Correlation Test

يستخدم هذا الاختبار لدراسة الارتباط بين متغيرين وهو يساوي :

$$R_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

حيث أن d_i هي الفرق بين الرتب لكل زوج من المتحولين .

يأخذ هذا المعامل قيمه بين -1 و +1 . وكلما كانت قيمته قريبة من طرفي المجال كلما كان الارتباط بين المتحولين أقوى. ولمعرفة معنوية الارتباط أو الدلالة الإحصائية لهذا الارتباط نميز بين العينات الكبيرة والعينات الصغيرة:

العينات الصغيرة ($n \leq 30$):

نقارن قيمة المعامل السابق مع القيم الجدولية لهذا الإختبار

العينات الكبيرة ($n > 30$):

نستخدم مؤشر الإختبار التالي:

$$z = \frac{R_s}{1} \frac{1}{\sqrt{n-1}}$$

الذي يتبع التوزيع الطبيعي المعياري.

مثال 34 :

في دراسة هدفت إلى معرفة درجة ارتباط مستوى الأجور بالجهد المبذول من قبل العامل أو الموظف قمنا بترتيب الأجور والجهد لعشر وظائف فحصلنا على النتائج التالية:

الوظيفة	ترتيب الأجور	ترتيب الجهد	الفرق d	d^2
1	2	2	0	0
2	6	7	1	1
3	3	6	3	9
4	5	4	1	1
5	7	5	2	4
6	10	8	2	4
7	9	9	0	0
8	8	10	2	4
9	4	3	1	1
10	1	1	0	0
المجموع				24

لنحسب معامل الارتباط:

$$R_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} = 1 - \frac{6(24)}{10(10^2 - 1)} = 0.855$$

نلاحظ أن الارتباط جيد . ومن جدول القيم الحرجة لسبيرمان نلاحظ أن القيمة الحرجة للاختبار هي ± 0.64 وهي قيمة أقل من قيمة معامل الارتباط أي نرفض فرضية العدم ونقر بوجود ارتباط بين كل من المتغيرين أي ($R_s \neq 0$).

المحاضرة الثالثة عشر

الفصل السادس

تحليل التباين

Analyse of Variance

مفردات البحث

- 1- تحليل التباين أحادي الإتجاه
- 2- اختبار بونفيروني

المرجع المعتمد: المرجع السابق من الصفحة 530 إلى الصفحة 565

6 - 1 : تحليل التباين أحادي الاتجاه ANOVA one way

يعتبر تحليل التباين تعميم لفكرة اختبار الفرق بين متوسطي مجتمعين طبيعيين حيث يتناول هذا التحليل دراسة واختبار الفرضيات حول وجود أو عدم وجود فروقات بين متوسطات أحد المتحولات في عدة مجتمعات (ثلاثة مجتمعات فأكثر) . فإذا كان لدينا عدد من المجتمعات (k) مجتمع و ($k \geq 3$) يعتمد تحليل التباين على اختبار الفرضية التالية :

فرضية العدم (جميع المتوسطات متساوية) ونكتبها كما يلي :

$$H_0 : \bar{y}_1 = \bar{y}_2 = \bar{y}_3 = \dots = \bar{y}_k$$

الفرضية البديلة (يوجد فرق بين متوسطي مجتمعين على الأقل) ونكتبها كما يلي :

$$H_1 : \bar{y}_i \neq \bar{y}_j \quad \text{من أجل } i \neq j$$

وحيث أن : $i, j : 1, 2, 3, \dots, k$

ولإجراء هذا الاختبار نقوم بالإجراءات التالية :

1- نسحب عينة عشوائية من كل مجتمع من المجتمعات المذكورة بحجوم مختلفة أو متساوية ونرمز

لها على الترتيب بالرموز $n_1, n_2, \dots, n_k$ ونرمز لمجموعها بالرمز n (أي أن $\sum n_j = n$) .

2- نحسب متوسطات المتحول المدروس في هذه العينات ونرمز لها على الترتيب بالرموز

$$\bar{x}_1, \bar{x}_2, \dots, \bar{x}_k$$

3- نحسب المتوسط العام لقيم المتحول المدروس في جميع هذه المجتمعات وذلك من خلال حساب متوسط المتوسطات في العينات وذلك بتطبيق العلاقة :

$$\bar{x} = \frac{\sum_{j=1}^k n_j \bar{x}_j}{n}$$

4- نحسب التباينات ضمن أو داخل كل عينة على حدة ولنفترض أنها كانت تساوي على الترتيب ما يلي :

$$S_1^2, S_2^2, \dots, S_k^2$$

6- نحسب متوسط التباينات الداخلية فنحصل على ما يسمى بتباين الأخطاء ويحسب من العلاقة :

$$SSE = \sum_{j=1}^k (n_j - 1) S_j^2$$

6- ثم نحسب التباين بين العينات وذلك بأخذ متوسط مربعات انحرافات المتوسطات العينية $\bar{x}_j$ عن المتوسط العام $\bar{x}$ ونرمز له بالرمز SSB ويحسب من العلاقة :

$$SSB = \sum_{j=1}^k n_j (\bar{x}_j - \bar{x})^2$$

7- نحسب التباين العام لجميع القياسات في جميع العينات والذي يساوي متوسط مربعات انحرافات تلك القياسات عن المتوسط العام $\bar{x}$ ونرمز له بالرمز SST ونحسبه من العلاقة :

$$SST = \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x})^2$$

8- ومن خواص هذه التباينات أنها مرتبطة ببعضها البعض وأن التباين العام SST يساوي مجموع التباينين SSE و SSB أي أن :

$$SST = SSE + SSB$$

9- نقوم بحساب متوسطات التباينات السابقة كالتالي:

$$MSE = \frac{SSE}{n - k} = \frac{\sum_{j=1}^k (n_j - 1) S_j^2}{n - k}$$

$$MSB = \frac{SSB}{k - 1} = \frac{\sum_{j=1}^k n_j (\bar{x}_j - \bar{x})^2}{k - 1}$$

$$MST = \frac{SST}{n-1} = \frac{\sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x})^2}{n-1}$$

حساب قيمة مؤشر الاختبار F المعروف بالعلاقة :

$$F = \frac{MSB}{MSE}$$

وإن المؤشر F يمثل متحولاً عشوائياً ويخضع لتوزيع فيشر ذي $(k-1)$ درجة حرية للبسط و $(n-k)$ درجة حرية للمقام .

11- حتى نتخذ القرار المناسب حول الفرضية H_0 نقوم بمقارنة قيمة F المحسوبة من العلاقة السابقة مع قيمة F_α الحرجة المقابلة لدرجتي الحرية $(k-1)$ للبسط و $(n-k)$ للمقام وعند مستوى دلالة (α) ونتخذ القرار حسب الحال :

إذا كانت $F < F_\alpha$ فإننا نقبل فرضية العدم H_0 ونقول أن جميع متوسطات المجتمعات متساوية أما إذا كانت $F > F_\alpha$ فإننا نرفض فرضية العدم H_0 ونقبل الفرضية البديلة H_1 التي تقول أن متوسطات أو بعض متوسطات هذه المجتمعات غير متساوية (اثنتين منها على الأقل).
يمكن تلخيص الخطوات السابقة ووضعها في جدول خاص يسمى بجدول تحليل التباين ANOVA ويأخذ الشكل التالي :

جدول تحليل التباين ANOVA

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	F	P-Value
بين العينات أو المجتمعات	SSB	$k - 1$	$MSB = \frac{SSB}{k - 1}$	$F = \frac{MSB}{MSE}$	
داخل العينات	SSE	$n - k$	$MSE = \frac{SSE}{n - k}$		
التباين الكلي	SST	$n - 1$			

ملاحظة 1 :

إن قيمة احتمال الدلالة P هي قيمة الاحتمال الذي يتبقى على يسار القيمة المحسوبة F ويتم حسابه آلياً اعتماداً على التكامل . ويمكننا الاعتماد عليه في اتخاذ القرار حول قبول أو رفض فرضية العدم H_0 .
فإذا كانت قيمة هذا الاحتمال أصغر من مستوى الدلالة α فإن هذا يعني أن قيمة F المحسوبة المقابلة لـ P أكبر من قيمة F_α الحرجة المقابلة لـ α وعندها نرفض الفرضية H_0 .
أما إذا كانت قيمة P أكبر من مستوى الدلالة α فإن هذا يعني أن قيمة F المحسوبة والمقابلة لـ P تكون أصغر من القيمة الحرجة F_α المقابلة لـ α وعندها نقبل فرضية العدم H_0 .

ملاحظة 2 :

إذا كانت نتيجة الاختبار رفض فرضية العدم H_0 لعدم تساوي بعض المتوسطات فإنه لا بد لنا من أن نتعرف أو نحدد مصادر عدم التساوي , لذلك نقوم بإجراء مقارنات ثنائية بين كل متوسطين من مجتمعين مختلفين باستخدام اختبار بونفيروني التالي :

اختبار المقارنة لبونفيروني Bonferroni Comparaison Test

إذا كنا بصدد معرفة إذا كان الاختلاف بين المتوسطات ناتج عن الاختلاف بين العينة i والعينة j فإننا نستخدم الاختبار التالي :

$$t = \frac{\bar{x}_i - \bar{x}_j}{\sqrt{MSE \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}}$$

وهذا المؤشر يتبع توزيع ستيودنت ب $(n - k)$ درجة حرية.

مثال 29 :

لدراسة متوسط عدد أفراد الأسرة في 3 ثلاث قرى (مجتمعات) سحبنا عينة عشوائية من الأسرة من كل قرية بحجم $n_1 = 4$, $n_2 = 6$, $n_3 = 3$ فكانت النتائج كما يلي :

متوسط العينة	عدد أفراد الأسرة حسب نتائج الدراسة	حجم العينة	المجتمع
8	10 7 6 9	4	القرية الأولى j = 1
6	7 8 6 6 54	6	القرية الثانية j = 2
5	5 3 4 7 6	5	القرية الثالثة j = 3

والمطلوب اختبار فرضية تساوي متوسط عدد أفراد الأسرة في هذه القرى :

الحل :

$$H_0 : \bar{y}_1 = \bar{y}_2 = \bar{y}_3$$

$$H_1 : \bar{y}_i \neq \bar{y}_j$$

إن المتوسط العام $\bar{x}$ يساوي :

$$\bar{x} = \frac{\sum n_j \bar{x}_j}{n} = \frac{4 \times 8 + 6 \times 6 + 5 \times 5}{4 + 6 + 5} = 6.2$$

وبعد ذلك نقوم بحساب المجاميع التالية :

$$SST = \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x})^2 = (10 - 6.2)^2 + (7 - 6.2)^2 + (6 - 6.2)^2 + \dots + (6 - 6.2)^2 = 50.4$$

$$SSE = \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2 = (10 - 8)^2 + (7 - 8)^2 + (6 - 8)^2 + (9 - 8)^2 + (7 - 6)^2 + \dots + (6 - 5)^2 = 30$$

$$SSB = SST - SSE = 50.4 - 30 = 20.4$$

وبناءً على الحسابات السابقة نعد جدول تحليل التباين التالي :

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	قيمة F المحسوبة	F_α
SSB	SSB = 20.4	k - 1 = 2	$MSB = \frac{SSB}{k - 1} = \frac{20.4}{2} = 10.2$	$F = \frac{MSB}{MSE} = \frac{10.2}{2.5} = 4.08$	3.89
SSE	SSE = 30	N - k = 12	$MSE = \frac{SSE}{n - k} = \frac{30}{12} = 2.5$		
SST	SST = 50.4	N - 1 = 14			

وبمقارنة قيمة F المحسوبة (4.08) مع قيمة F الحرجة المقابلة لدرجتي حرية (2 , 12) ولمستوى الدلالة ($\alpha = 0.05$) والتي تساوي 3.89 نجد أن $F > F_\alpha$ لذلك نرفض فرضية العدم H_0 التي تقول أن متوسطات عدد أفراد الأسرة في هذه القرية متساوية .
ولتحديد مصدر عدم تساوي هذه المتوسطات نقوم بمقارنة كل زوج فيها وذلك بتطبيق اختبار بونفيروني فنجد أن :

$$t_{1,2} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{MSE \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} = \frac{8 - 6}{\sqrt{2.5 \left(\frac{1}{4} + \frac{1}{6} \right)}} = 1.95$$

$$t_{1,3} = \frac{\bar{x}_1 - \bar{x}_3}{\sqrt{MSE \left(\frac{1}{n_1} + \frac{1}{n_3} \right)}} = \frac{8 - 5}{\sqrt{2.5 \left(\frac{1}{4} + \frac{1}{5} \right)}} = 2.82$$

$$t_{2,3} = \frac{\bar{x}_2 - \bar{x}_3}{\sqrt{MSE \left(\frac{1}{n_2} + \frac{1}{n_3} \right)}} = \frac{6 - 5}{\sqrt{2.5 \left(\frac{1}{6} + \frac{1}{5} \right)}} = 1.04$$

بمقارنة القيم السابقة مع القيمة الحرجة ل t المقابلة ل 12 درجة حرية والتي تساوي 2.17 نجد أن الفرق ناتج عن الفرق بين العينة الأولى والثالثة.

المحاضرة الرابعة عشر

الفصل السابع

التجارب المتعددة وجداول الاحتمال

Multinomial Experiments and Contingency Tables

مفردات البحث

- 1- التجارب المتعددة : جودة التوفيق
- 2- الجداول الاحتمالية : الاستقلال والتجانس
- 3- اختبار ماكنمار للأزواج المتقابلة

المرجع المعتمد: المرجع السابق من الصفحة 488 إلى الصفحة 529

1.5. التجارب المتعددة : جودة التوفيق (Goodness-of-Fit)

رأينا سابقا كيف يمكن إجراء اختبارات متعددة جميعها تتعلق ببيانات كمية. فيما يلي سوف نقوم باستعراض بعض الاختبارات المتعلقة بالبيانات النوعية وهي البيانات التي تتألف من عدة مجموعات منفصلة تسمى وحدات أو خلايا (Cells). وسنعمد بشكل رئيسي في هذه الاختبارات على توزيع χ^2 . عندما يكون لدينا هذا النوع من البيانات قد يكون الهدف هو معرفة فيما إذا كان توزيع هذه البيانات يتوافق (Agrees or Fits) مع توزيع معين.

نقول عن التجارب أنها تجارب متعددة أو ذات حالات متعددة (Multinomial Experiments) إذا حققت عدة شروط:

1- عدد التجارب محددة

2- التجارب مستقلة

3- كل نتائج التجربة يمكن تصنيفها في مجموعة من الحالات أو الفئات

4- احتمالات هذه الفئات تبقى ثابتة في كل تجربة

عندما يكون لدينا مثل هذا النوع من التجارب يكون هدفنا اختبار فيما إذا كانت التكرارات المشاهدة في هذه التجربة تتوافق مع توزيع معين. وبالتالي نختبر فيما إذا كان التوزيع التكراري المشاهد يتوافق مع توزيع تكراري نظري محدد. ولهذا سمي هذا الاختبار اختبار جودة التوفيق (Goodness-of-Fit Test).

إذا رمزنا :

- O : التكرارات المشاهدة (Observed)

- E : التكرارات المتوقعة (Expected)

- k : عدد الفئات أو الحالات

- n : عدد التجارب

نستطيع كتابة م{شر الاختبار على الشكل التالي:

$$\chi^2 = \sum \frac{(O - E)^2}{E} \quad \chi^2 (k - 1)$$

وهو اختبار أحادي الجانب يميني. أما بالنسبة للتكرار المتوقع فيتم حسابه كمايلي:

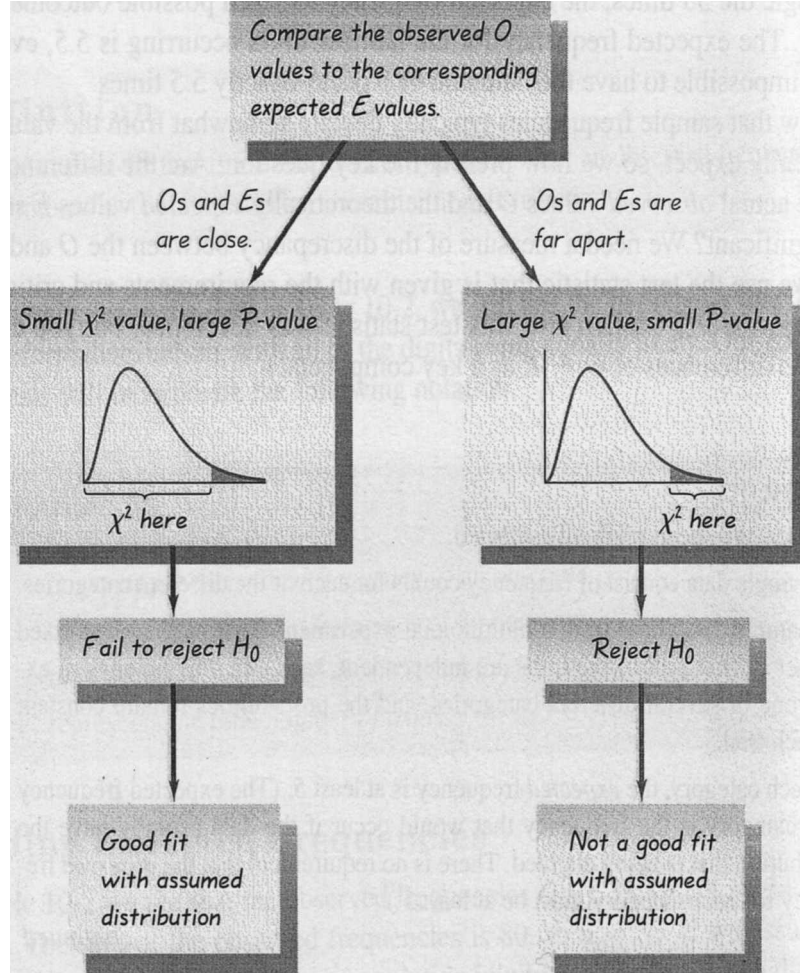
- إذا كانت التكرارات المتوقعة متساوية في هذه الحالة كل تكرار متوقع يساوي مجموع

التكرارات المشاهدة مقسوما على عدد المجموعات أي : $E = n/k$.

- إذا كانت التكرارات المتوقعة غير متساوية في هذه الحالة كل تكرار متوقع يساوي

مجموع التكرارات المشاهدة مضروباً باحتمال كل مجموعة أي : $E = np$.

بشكل عام يمكن تلخيص الاختبار كالتالي:



مثال 25 :

اختبر بمستوى دلالة 0.05 فيما إذا كانت هناك فروقات جوهرية بين التكرارات المطلقة لإجابات

36 طالباً حول أحد البرامج التعليمية والتي كانت كما يلي:

الإجابة	موافق	لا أدري	معارض	المجموع
التكرارات	13	7	16	36

لإجراء الاختبار المطلوب نضع فرضية العدم على الشكل التالي:

H_0 : لا يوجد فروقات جوهرية بين التكرارات النسبية للإجابات (احتمالات الإجابة متساوية)

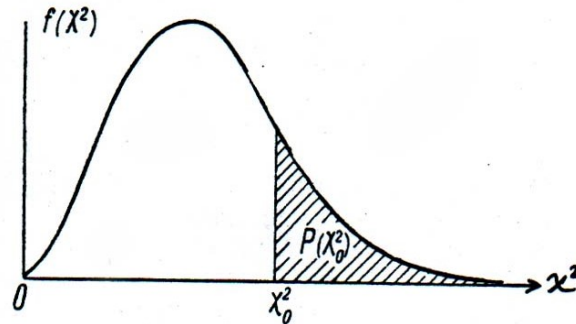
H_1 : هناك على الأقل واحد من التكرارات النسبية للإجابات مختلف عن الآخر

باعتبار أن التكرارات المتوقعة متساوية يكون كل تكرار متوقع مساوي : $E = n/k = 36/3 = 12$.

$$\chi^2 = \sum \frac{(O - E)^2}{E} = \frac{(13-12)^2}{12} + \frac{(7-12)^2}{12} + \frac{(16-12)^2}{12} = 4.33$$

من جدول التوزيع χ^2 نجد أن القيمة الحرجة $\chi_{0.05}^2$ المقابلة ل (3-1) درجة حرية تساوي $\chi_0^2 = 5.99$.

وبمقارنة قيمة χ^2 المحسوبة مع $\chi_{0.05}^2$ الحرجة، نجد أن χ^2 أصغر من $\chi_{0.05}^2$ الحرجة. لذلك نقبل الفرضية H_0 القائلة بعدم وجود فروقات جوهرية بين تكرارات إجابات المبحوثين، ونرفض الفرضية البديلة القائلة بوجود فروقات بين تلك التكرارات. والشكل التالي يوضح ذلك.



2.5. الجداول الاحتمالية (المزدوجة): الاستقلال والتجانس (Contingency Tables)

الجداول المزدوجة هي عبارة عن جداول تتضمن تكرارات متعلقة بمتغيرين نسمي الأول متغير السطر والآخر متغير العمود. فمثلا إذا كان لدينا متغيرين أحدهما يدل على منطقة السكن والآخر يدل على نتيجة الطالب نستطيع تمثيل ذلك في جدول كالتالي:

منطقة السكن			
		مدينة	ريف
نتيجة الطالب	ناجح	50	75
	راسب	35	40

في مثل هذا النوع من الجداول يكون الهدف اختبار فرضية العدم بأن المتغيرين في السطر والعمود مستقلان (غير مرتبطين). لاختبار مثل هذه الفرضية نستخدم اختبار χ^2 ولكن يتوجب تحقق الشروط التالية :

- بيانات العينة عبارة عن تكرارات لملاحظات مستقلة وكل مشاهدة يمكن أن بمتغيرين.
- بيانات العينة عشوائية
- فرضية العدم هي أن متغيرات السطر والعمود مستقلة والفرضية البديلة هي أن هذه المتغيرات مرتبطة
- كل تكرار يجب أن لا يقل عن 5 . وليس من الضروري أن يكون توزيع المجتمع طبيعياً.

عند تحقق هذه الشروط يمكن إجراء اختبار الاستقلالية التالي:

$$\chi^2 = \sum \frac{(O - E)^2}{E} \quad \chi^2 (r-1)(c-1)$$

حيث أن r هو عدد الأسطر و c هو عدد الأعمدة. وهذا الاختبار هو اختبار أحادي الجانب يميني. أما حساب التكرارات المتوقعة فيتم بحسب القاعدة التالية: التكرار المتوقع للخلية يساوي مجموع سطرها مضروباً بمجموع عمودها ومقسوماً على المجموع الكلي.

مثال 26:

يبين الجدول التالي تكرارات إجابات 70 طالباً حسب نوع الإجابة والجنس وذلك حول أحد البرامج أو الأسئلة المطروحة للبحث والتي كانت كما يلي :

الجنس نوع الإجابة	الذكور	الإناث	المجموع
موافق	16	9	25
معارض	14	6	20
المجموع	30	15	45

لإجراء اختبار حول وجود علاقة بين نوع الإجابة والجنس نضع الفرضيات كما يلي:

H_0 : يوجد استقلال بين نوع الإجابة والجنس

H_1 : يوجد ارتباط بين نوع الإجابة والجنس

نوجد أولاً التكرارات المتوقعة لكل خلية :

حساب التكرارات المتوقعة

الجنس الإجابة	الذكور	الإناث	المجموع
موافق	$\frac{25 \times 30}{45} = 16.66$	$\frac{25 \times 15}{45} = 8.33$	25
معارض	$\frac{20 \times 30}{45} = 13.33$	$\frac{20 \times 15}{45} = 6.66$	20
المجموع	30	15	45

$$\chi^2 = \sum \frac{(O - E)^2}{E} = \frac{(16 - 16.66)^2}{16.66} + \frac{(9 - 8.33)^2}{8.33} + \frac{(14 - 13.33)^2}{13.33} + \frac{(6 - 6.66)^2}{6.66} = 0.18$$

من جدول التوزيع χ^2 نجد أن القيمة الحرجة $\chi^2_{0.05}$ المقابلة ل (2-1)(2-1) أي لدرجة حرية واحدة

تساوي $\chi^2_{0.05} = 3.84$. وبمقارنة قيمة χ^2 المحسوبة مع $\chi^2_{0.05}$ الحرجة، نجد أن χ^2 أصغر من $\chi^2_{0.05}$

الحرجة. لذلك نقبل الفرضية H_0 القائلة باستقلالية المتغيرين . أي لا تتعلق الموافقة أو عدم الموافقة على البرنامج بجنس الشخص.

اختبار التجانس (Test of Homogeneity)

في اختبار التجانس نقوم بالتحقق من أن لمجتمعات مختلفة نفس النسب من خواص معينة. فمثلا إذا كنا بصدد اختبار فيما إذا كان نسبة مدخني السيكارة إلى مدخني النرجيلة هي واحدة بين الذكور والاناث. في مثل هذه الحالة نستخدم اختبار χ^2 السابق.

مثال 27 :

يدعي أحد مراكز الأبحاث أن المستجوبون الذكور في قضية ما يميلون لإعطاء إجابة بالقبول أكثر إذا كان المستجوب أنثى ونريد التحقق من أن الاجابات لا تتعلق بجنس المستجوب. لذلك قام أحد الباحثين بإجراء بحث حول هذا الموضوع من خلال استجواب مجموعة من الرجال حول احدى القضايا فحصل على النتائج التالية:

المجموع	أنثى	ذكر	جنس القائم على الاستبيان نوع الاجابة
868	308	560	موافقة
332	92	240	عدم موافقة
1200	400	800	المجموع

المطلوب التحقق من أن نسبة الموافقين على غير الموافقين هي نفسها في المجتمعين أي سواء كان المستجوب ذكرا أم أنثى وذلك بمستوى دلالة 5 %.

إن نسبة الأجوبة الموافقة على غير الموافقة هي نفسها عند المستجوبين الذكور والاناث : H_0

إن نسبة الأجوبة الموافقة على غير الموافقة مختلفة عند المستجوبين الذكور والاناث : H_1

بحساب التكرارات المتوقعة نجد:

حساب التكرارات المتوقعة

جنس القائم على الاستبيان نوع الاجابة	الذكور	الإناث
موافق	$\frac{868 \times 800}{1200} = 578.66$	$\frac{868 \times 400}{1200} = 289.33$
معارض	$\frac{332 \times 800}{1200} = 221.33$	$\frac{332 \times 400}{1200} = 110.66$

$$\chi^2 = \sum \frac{(O - E)^2}{E} = \frac{(560 - 578.66)^2}{578.66} + \frac{(308 - 289.33)^2}{289.33} + \frac{(240 - 221.33)^2}{221.33} + \frac{(92 - 110.66)^2}{110.66} = 6.52$$

من جدول التوزيع χ^2 نجد أن القيمة الحرجة $\chi^2_{0.05}$ المقابلة لدرجة حرية واحدة تساوي $\chi^2_{0.05} = 3.84$. وبمقارنة قيمة χ^2 المحسوبة مع $\chi^2_{0.05}$ الحرجة، نجد أن χ^2 أكبر من $\chi^2_{0.05}$ الحرجة. لذلك نرفض الفرضية H_0 القائلة بأن النسب متساوية (تجانس المجتمعين). أي أن الإستجابة تختلف إذا كان المستجوب ذكر عنها إذا كان المستجوب أنثى.

ملاحظة: إن الاختبار السابق يتطلب بأن يكون التكرار المتوقع في كل خلية أكبر من 5. هذا ضروري لكي يتبع المؤشر السابق توزيع χ^2 بشكل تقاربي. إذا لم تكن الحالة كذلك يمكن استخدام توزيع فيشر F . فتوزيع فيشر في حال كان لدينا جدول 2×2 هو توزيع دقيق ولا يحتاج لتقريب.

3.5 اختبار ماكنمار للأزواج المتقابلة (McNemar's Test for Matched Paires)

في الجداول السابقة من الشكل 2×2 رأينا من شروط الاختبارات السابقة أن تكون البيانات مستقلة. كيف إذا يمكن إجراء الإختبار إذا كانت البيانات مرتبطة (غير مستقلة). فلو فرضنا مثلاً أن مجموعة من المرضى قد خضعت لنوعين من العلاج أو الأدوية وأن نتيجة العلاج كانت كالتالي:

العلاج بالنوع الأول			
		شفى	لم يشف
العلاج بالنوع الثاني	شفى	a	b
	لم يشف	c	d

إن a تدل على عدد الأشخاص الذين شفوا سواء باستخدام المعالجة الأولى أو الثانية. و b تدل على عدد الأشخاص الذين شفوا باستخدام امعالجة الثانية لكنهم لم يشفوا عند استخدام النوع الأول من العلاج. في مثل هذه الحالة نكون بصدد اختبار الفرضية الإبتدائية التالية: إن التكرارات b و c تحدث بنفس النسبة. في هذه الحالة يمكن استخدام اختبار ماكنمار التالي:

$$\chi^2 = \frac{(|b-c|-1)^2}{b+c} \square \chi^2(1)$$

إذا يستخدم هذا الاختبار عندما يكون لدينا متغيرين اسميين لكل منهما حالتين (قبول أو رفض. خطأ أو صح.... الخ) باستخدام تكرارات الأزواج المتقابلة. ويشترط لإجراء هذا الاختبار أن تكون العينة عشوائية وأن تكون البيانات إسمية على شكل أزواج متقابلة تصنف في نوعين. وأن تحقق التكرارات الشرط التالي: $b+c \geq 10$. كما يجدر الذكر بأن الاختبار باتجاه واحد (يميني).

مثال 28 :

في المثال السابق لنفرض أن توزيع المرضى الذين شفوا والذين لم يشفوا باستخدام نوعين من المعالجة كان كالتالي:

العلاج بالنوع الأول			
		شفى	لم يشفى
العلاج بالنوع الثاني	شفى	12	8
	لم يشفى	40	20

في هذه الحالة نكون بصدد اختبار فرضية العدم التالية:

H_0 : إن نسبة الذين لم يشفوا باستخدام العلاج الأول أو العلاج الثاني هي واحدة

H_1 : إن هاتين النسبتين مختلفتين

$$\chi^2 = \frac{(|b-c|-1)^2}{b+c} = \frac{(|8-40|-1)^2}{8+40} = 20.02$$

من جدول التوزيع χ^2 نجد أن القيمة الحرجة $\chi^2_{0.05}$ المقابلة لدرجة حرية واحدة تساوي $\chi^2_{0.05} = 3.84$. وبمقارنة قيمة χ^2 المحسوبة مع $\chi^2_{0.05}$ الحرجة، نجد أن χ^2 أكبر من $\chi^2_{0.05}$ الحرجة. لذلك نرفض الفرضية H_0 القائلة بأن النسب متساوية. أي أن الإستجابة لنوعية العلاج مختلفة وفي مثالنا هذا نلاحظ أن العلاج بالطريقة الأولى أكثر فاعلية.