جامعة 8 ماي 1945 ـ قالمة

كلية العلوم الاقتصادية والتجارية وعلوم التسيير

قسم علوم التسيير

مطبوعة بيداغوجية في مقياس

الإحصاء 3

المستوى : سنة ثانية

السداسي : الأول

من اعداد

د. دبوش عبد القادر

السنة الجامعية 2017-2018

تقديم

يلعب الإحصاء دور بارز في مختلف الأنشطة، الطبيعية منها والإنسانية. فهو يوفر الأساليب التي تمكن من دراسة مختلف الظواهر، بداية من ملاحظتها، إلى تحليلها وتوقعها. وقد تطورت هذه الأساليب الإحصائية بشكل أصبحت قادرة على التعامل مع مختلف المشكلات التي تواجهنا. لهذا، يأتي هذا المقياس المعنون بـ "الإحصاء **3**". أو كما يطلق عليه "الإحصاء التطبيقي". لتمكين الطلبة من هذه الأساليب الإحصائية بغرض القيام بدراساتهم الميدانية في مذكرات التخرج، وكذا التعامل مع مختلف المشكلات التي تواجههم في الحياة العملية.

إن هذه المطبوعة، عبارة عن محاضرات قدمت لطلبة السنة ثانية ليسانس علوم التسيير بجامعة **8** ماي **1945** قالمة، وطلبة السنة ثالثة مالية بجامعة العربي بن مهيدي أم البواقي. ولقد أثريت بالعديد من التوجيهات والملاحظات من مختلف الزملاء الأساتذة وكذا الطلبة منذ سنة **2007**. كما تعتبر هذه المطبوعة أداة لدراسة عدة مقاييس منها: تسيير الميزانية، الاقتصاد الكلي،...

جاءت هذه المطبوعة في أربعة فصول وفق المقرر الوزاري وهي: المعاينة، التقدير، اختبار الفرضيات ونموذج الانحدار الخطي البسيط. وقد تضمنت العديد من الأمثلة التطبيقية، التمارين المحلولة وكذا مسائل مقترحة، وهذا لكي يتعود الطالب ويتمكن من تطبيق الجانب النظري على الحياة العملية.

في الأخير، لا يسعني إلا أن أتقدم بأسمى عبارات الشكر والتقدير، لكل أساتذتي الذين بذلوا جهدا كبيرا في تكويننا في مرحلة التدرج وما بعد التدرج في هذا الحقل، وأخص بالذكر أ.د شريف غياط وأ.د السعدي رجال.

الفصل الأول: المعاينة Sampling

1 – 1 تمهيد

يهتم علم الإحصاء بجمع وتحليل وعمل استدلالات حول الظاهرة المدروسة. إن هذا التعريف يحمل في دلالته الكثير من الجوانب. فعملية جمع بيانات عن ظاهرة ما لا تعد مسألة بسيطة. بل تستلزم الكثير من الفرضيات والإجراءات لتنفيذها. أولى هذه الإجراءات هي تحديد العينة المراد دراستها، من حيث حجمها وتركيبتها. أما تحليل البيانات فيعتبر أكثر من مجرد معرفة معاني المصطلحات الإحصائية. ويتضمن مزيجا من سلامة التفكير والحدس والخبرة التقنية والفضول. خصوصا أن تحليل البيانات لا يعتبر عملية رتيبة، فلكل بيانات نستخدمها تعتبر وحيدة بطريقة ما. أما الاستدلال حول الظاهرة المدروسة فيعتبر جوهر العمليات الإحصائية لما يوفره لنا من معلومات نعتمد عليها في مختلف الأنشطة الإنسانية.

1 – 2 المعاينة العشوائية البسيطة

غالبا ما نهتم في الحياة العملية باستخلاص نتائج تخص مجموعة كبيرة من المفردات التي تسمى مجتمعا (Population). فبدلا من دراسة كامل المجموعة، وهذا ما يكون مستحيلا في أغلب الأحيان بسبب عاملي الوقت والتكلفة، يمكن أن تتشكل لدينا فكرة عنه بدراسة جزء صغير من هذا المجتمع نسميه العينة (Sample).

1 – 2 – 1 تعريفات

المجتمع:

يعرف المجتمع بأنه: "مجموعة عناصر تشترك في خاصية أو أكثر قد يكون محدود أو غير محدود".

مثال رقم (1–1): نرغب في استخلاص نتائج حول أطوال 10000 طالب من خلال فحص 100 طالب فقط، في هذه الحالة يتكون المجتمع من 10000 طالب، بينما تتكون العينة من 100 طالب مسحوب من هذا المجتمع.

مثال رقم (1–2): نرغب في استخلاص نتائج تخص تجانس قطعة نقدية معينة من خلال رميها بشكل متكرر، يتشكل المجتمع من كل نتائج رمي القطعة النقدية الممكنة، أما العينة فيمكن الحصول عليها بفحص أول 50 رمية مثلا.

نلاحظ أن المجتمع في المثال 1 محدود أما في المثال 2 فهو غير محدود.

العينة:

تعرف العينة بأنها: "جزء مسحوب من مجتمع".

المعلمة (Paramètre):

هي خاصية وصفية لمجتمع ما. مثل الوسط الحسابي μ، الانحراف المعياري δ، النسبة P.

ويعتبر المجتمع معروف عندما نعلم توزيعه الاحتمالي (دالة الاحتمال، دالة التوزيع أو دالة الكثافة)، للمتغير العشوائي المرافق له X.

في المثال 1 إذا كان X متغير عشوائي قيمه هي أطوال الطلاب الـ 10000، بالتالي فإن لـ X التوزيع الاحتمالي وليكن F(x).

الإحصائية (Statistic):

هي خاصية وصفية للعينة، مثل الوسط الحسابي $\bar{x}$، الانحراف المعياري s، النسبة $\bar{p}$.

حيث أن تشكيل العينة ليس الهدف بحد ذاته، بل الهدف هو الوصول إلى استدلالات حول معلمات المجتمع الذي سحبت منه العينة. وبالتالي فإننا ندعو أي مقدار وصفي نحصل علية من عينة لتقدير معالم المجتمع بالإحصائية.

- في الحالة العامة سيكون هناك مقابل كل معلمة في المجتمع إحصائية تحسب من العينة، وسوف نستخدم (μ, δ, P) لنرمز إلى الوسط الحسابي، الانحراف المعياري و النسبة كمعلمات المجتمع، بالمقابل سوف نستخدم $(\bar{x}, s, \bar{p})$ لنرمز إلى الوسط الحسابي، الانحراف المعياري والنسبة كإحصائيات العينة.

1 – 2 – 2 مفهوم المعاينة العشوائية البسيطة (Simple random sampling):

يمكن استخدام العديد من الطرق لاختيار عينة من مجتمع، وتعتبر المعاينة العشوائية البسيطة من أبسط وأهم هذه الطرق. ويختلف تعريف المعاينة العشوائية البسيطة وكيفية استخدامها لسحب عينة من مجتمع فيما إذا كان المجتمع محدود أو غير محدود.

1 – 2 – 2 – 1 المعاينة من مجتمع محدود (Sampling from a Finite Population)

تعد من أبسط أنواع طرق اختيار العينة، وسوف نستخدم مصطلح المعاينة للدلالة عليه، وتعرف المعاينة من مجتمع محدود بأنها: "استخراج n عنصر من المجتمع N بحيث أن كل السحوبات يكون لديها نفس الاحتمال، أي أن كل عنصر يجب أن يسحب بطريقة مستقلة عن العناصر الأخرى".[1]

توجد عدة طرق لاختيار العينة العشوائية البسيطة من مجتمع محدود منها:

[1] - David R. Anderson, Dennis J. Sweeney, Thomas A. Williams (2005). Statistiques pour l'économie et la gestion, De Boeck. France, p 315.

أولا: الطريقة اليدوية

يتم تسجيل وحدات المجتمع في أوراق، ومن ثم يتم خلط هذه الأوراق جيدا، وبعدها يتم سحب عدد مفردات العينة بصفة عشوائية.

نلاحظ أن هذه الطريقة عملية فقط في حالة المجتمعات الصغير الحجم، أما إذا كانت المجتمعات كبيرة الحجم، فهي تصبح مكلفة وتستلزم الكثير من الوقت.

ثانيا: استخدام جداول الأعداد العشوائية

تقوم هذه الطريقة على إعطاء ترتيب محدد لمفردات المجتمع من 1 إلى غاية N، بحيث كل مفردة يكون ترتيبها مكون من عدد أعداد المجتمع. بعد ذلك، نأخذ أحد صفحات الجداول العشوائية، التي يظهر الجدول رقم (1−1) إحداها. نختار عمود* يكون مساويا لعدد أعداد حجم المجتمع، والتي تمثل ترتيب عناصر العينة المأخوذة، مع إلغاء العدد العشوائي الذي يكون خارج العينة.

مثال رقم (1−3): إذا كان لديك مجتمع يتكون من 400 طالب في قسم سنة ثانية جامعي تخصص مالية، ونرغب في أن نأخذ عينة عشوائية بسيطة حجمها 10 من هذا المجتمع.

الحل:

نقوم بترتيب أفرد المجتمع بالصورة التالية: 001، 002، 003، ...،399، 400. بعدها نختار عناصر هذه العينة باستخدام جدول الأعداد، حيث نأخذ العمود الأول والثاني والثالث، والتي هي:

139، 005، 211، 380، 162، 280، 116، 353، 021، 165.

ثالثا: استخدام البرمجيات

توجد العديد من البرمجيات التي تسمح لنا آنيا بالحصول على الأعداد العشوائية، حيث يكفي إدخال بعض المعلومات لها، والنقر للحصول على العينة العشوائية المرغوبة. من هذه البرمجيات نجد: Excel، SPSS، Minitab،...

* يمكن اختيار أي نقطة من جدول الأعداد العشوائية، كنقطة انطلاق. كذلك يمكن أخذ الأسطر لاختيار الأعداد العشوائية،

جدول رقم (1-1): جدول الأعداد العشوائية

Random Number Table

13962	70992	65172	28053	02190	83634	66012	70305	66761	88344
43905	46941	72300	11641	43548	30455	07686	31840	03261	89139
00504	48658	38051	59408	16508	82979	92002	63606	41078	86326
61274	57238	47267	35303	29066	02140	60867	39847	50968	96719
43753	21159	16239	50595	62509	61207	86816	29902	23395	72640
83503	51662	21636	68192	84294	38754	84755	34053	94582	29215
36807	71420	35804	44862	23577	79551	42003	58684	09271	68396
19110	55680	18792	41487	16614	83053	00812	16749	45347	88199
82615	86984	93290	87971	60022	35415	20852	02909	99476	45568
05621	26584	36493	63013	68181	57702	49510	75304	38724	15712
06936	37293	55875	71213	83025	46063	74665	12178	10741	58362
84981	60458	16194	92403	80951	80068	47076	23310	74899	87929
66354	88441	96191	04794	14714	64749	43097	83976	83281	72038
49602	94109	36460	62353	00721	66980	82554	90270	12312	56299
78430	72391	96973	70437	97803	78683	04670	70667	58912	21883
33331	51803	15934	75807	46561	80188	78984	29317	27971	16440
62843	84445	56652	91797	45284	25842	96246	73504	21631	81223
19528	15445	77764	33446	41204	70067	33354	70680	66664	75486
16737	01887	50934	43306	75190	86997	56561	79018	34273	25196
99389	06685	45945	62000	76228	60645	87750	46329	46544	95665
36160	38196	77705	28891	12106	56281	86222	66116	39626	06080
05505	45420	44016	79662	92069	27628	50002	32540	19848	27319
85962	19758	92795	00458	71289	05884	37963	23322	73243	98185
28763	04900	54460	22083	89279	43492	00066	40857	86568	49336
42222	40446	82240	79159	44168	38213	46839	26598	29983	67645
43626	40039	51492	36488	70280	24218	14596	04744	89336	35630
97761	43444	95895	24102	07006	71923	04800	32062	41425	66862
49275	44270	52512	03951	21651	53867	73531	70073	45542	22831
15797	75134	39856	73527	78417	36208	59510	76913	22499	68467
04497	24853	43879	07613	26400	17180	18880	66083	02196	10638
95468	87411	30647	88711	01765	57688	60665	57636	36070	37285
01420	74218	71047	14401	74537	14820	45248	78007	65911	38583
74633	40171	97092	79137	30698	97915	36305	42613	87251	75608
46662	99688	59576	04887	02310	35508	69481	30300	94047	57096
10853	10393	03013	90372	89639	65800	88532	71789	59964	50681

المصدر:

The Rand Corporation (1955). A Million Random Digits with 100,000 Normal Deviates, The Free Press, USA.

1 – 2 – 2 – 2 المعاينة من مجتمع غير محدود (Sampling from a infinite Population)

يمكن أن تجرى عملية سحب عينة من مجتمعات تكون غير محدودة، أي لا يمكن حساب عدد مفرداتها. وتعرف المعاينة من مجتمع غير محدود بأنها: "المعاينة التي يتحقق فيها الشرطان التاليان: (أ) كل عنصر مسحوب يكون من نفس المجتمع. (ب) كل عنصر مسحوب يكون بصورة مستقلة."[1]

مثال رقم (1–4): نرغب في تقدير الوقت المتوسط المستغرق بين إعداد طلبية عشاء وتقديمها في مطعم بين الساعة 07.00 و10.00 مساء.

نلاحظ أنه لا يمكننا تحديد عدد عناصر المجتمع N، بحيث أنه لا نعرف كم من الزبائن سيتقدمون بطلبيات للعشاء خلال الوقت المحدد. لهذا فإن المجتمع في هذه الحالة يعتبر غير محدود. غير أننا نرغب في تشكيل عينة عشوائية بسيطة حجمها n للقيام بتقدير ما هو مطلوب.

للقيام بذلك، يجب أن نتحقق من أن الشرطان السابقان محققان. بالنسبة للشرط (أ) فهو محقق، من خلال كل الزبائن الذين يأتون خلال الوقت المحدد وهو بين الساعة 07.00 و10.00 مساء. الشرط (ب) يعتبر محقق إذا وفقط إذا كان حضور الزبائن إلى المطعم يكون بصورة مستقلة، أي أن حضور زبون ما إلى المطعم لا يؤثر ولا يتأثر بحضور أي زبون أخر إلى المطعم.

بعد التحقق من الشرطان، تبقى طريقة اختيار العينة العشوائية. حيث نلاحظ أن الطرق السابقة لاختيار العينة العشوائية من مجتمع محدود لا يمكن تطبيقها، لأنه لا يمكن تحديد حجم المجتمع. لهذا، يجب إتباع طرق تكفل بقاء تحقق الشرطان. إحدى الطرق المتبعة في هذه الحالة، هي إعطاء خصم معين للزبائن المعتادين لهذا المطعم. بعد أن يتم اختيار اليوم المحدد لسحب العينة، هي أخذ كل زبون يقدم طلبية عشاء بعد الزبون الذي يقدم طلبية عشاء مرفقة بالخصم الممنوح سابق. هذا يكفل أن يكون اختيار العينة بشكل مستقل. حيث يعتبر حضور هذا الزبون يكون بصفة مستقلة عن الآخرين.

1 – 3 توزيع المعاينة (Sampling distribitions)

يعتبر توزيع المعاينة من بين المفاهيم الشائكة. لهذا، ارتأينا أن نقدم شرح تفصيلي لهذا المفهوم وذلك باستخدام المثال رقم 5.

[1] - Ibid, p 318.

مثال رقم (1−5): يملك مدير الموارد البشرية لشركة مجموعة من المعطيات يرغب في استخدامها لتطوير قاعدة بيانات الشركة. سوف نهتم بفئة المسؤولين المقدرين بـ 2500، من حيث أجرهم السنوي ومشاركتهم في تربص تكويني، ولنعتبرهم مجتمعا.

من خلال استخدام البيانات المتاحة يمكن حساب كل من:

- متوسط الأجر السنوي $\mu = 518000$، والانحراف المعياري $\delta = 4000$.
- يوجد 1500 مسؤول تلقوا تكوينا، بالتالي نسبة المسؤولين الذين تلقوا التكوين هي:

$$P = \frac{1500}{2500} = 0.60$$

لنفرض الآن أن البيانات الضرورية لحساب المعلمات السابقة غير متوفرة. السؤال المطروح هنا:

كيف يمكن لمدير الموارد البشرية أن يحصل على بيانات حول معلمات المجتمع التي تهمه؟

- الطريقة التي تستخدم في هذه الحالة هي المعاينة.

لنأخذ مثلا عينة تتكون من 30 مسؤول (جدول رقم 1−2)، نجد هنا أن الوقت والتكلفة اللازمة لإيجاد بيانات 30 مسؤول أقل بكثير من اللازم لإيجاد بيانات المجتمع ككل.

جدول رقم (1−2): الأجر السنوي بالدولار والمشاركة في تربص لعينة عشوائية بسيطة من 30 مسؤول بالشركة.

	الأجر السنوي	التربص التكويني
x_1	49094.30	نعم
x_2	53263.90	نعم
x_3	49643.50	نعم
x_4	49894.90	نعم
x_5	47621.60	لا
x_6	55924.00	نعم
x_7	49092.30	نعم
x_8	51404.40	نعم
x_9	50957.70	نعم
x_{10}	55109.70	نعم
x_{11}	45922.60	نعم

	الأجر السنوي	التربص التكويني
x_{16}	51766.00	نعم
x_{17}	52541.30	لا
x_{18}	44980.00	نعم
x_{19}	51932.60	نعم
x_{20}	52973.00	نعم
x_{21}	45120.90	نعم
x_{22}	51753.00	نعم
x_{23}	54391.80	لا
x_{24}	50164.20	لا
x_{25}	52973.60	لا
x_{26}	50241.30	لا

لا	52793.90	x_{27}	لا	57268.40	x_{12}
نعم	50979.40	x_{28}	نعم	55688.80	x_{13}
نعم	55860.90	x_{29}	لا	51564.70	x_{14}
لا	57309.10	x_{30}	لا	56188.20	x_{15}

المصدر:

David R. Anderson, Dennis J. Sweeney, Thomas A. Williams (2005). Statistiques pour l'économie et la gestion, De Boeck. France, p 322.

يمكن إيجاد الوسط الحسابي والانحراف المعياري للأجر السنوي، وكذا نسبة العمال اللذين شاركوا في التربص من بيانات الجدول رقم (1−1) كالأتي:

$$\bar{X} = \frac{\sum x_i}{n} = \frac{1554420}{30} = 51814$$

$$S = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{325009260}{29}} = 3347.72$$

$$\bar{p} = \frac{19}{30} = 0.63$$

ليكن $\bar{X}$ متغير عشوائي يمثل متوسط الأجر السنوي للعينة. وككل متغير عشوائي لـ $\bar{X}$ أمل رياضي، تباين و توزيع احتمالي.

- التوزيع الاحتمالي لـ $\bar{X}$ يسمى "**توزيع المعاينة**".

في الحالة العامة كل عينة من 30 مسؤول تعطي لنا نتائج تكون مساوية أو مختلفة عن العينات الأخرى، ويوضح الجدول رقم (3−1)، بعض النتائج لـ $(\bar{x}, s, \bar{p})$ 500 عينة مختلفة.

جدول رقم (3−1): قيمة $(\bar{x}, s, \bar{p})$ لـ 500 عينة عشوائية بسيطة تتكون كل منها من 30 عنصر.

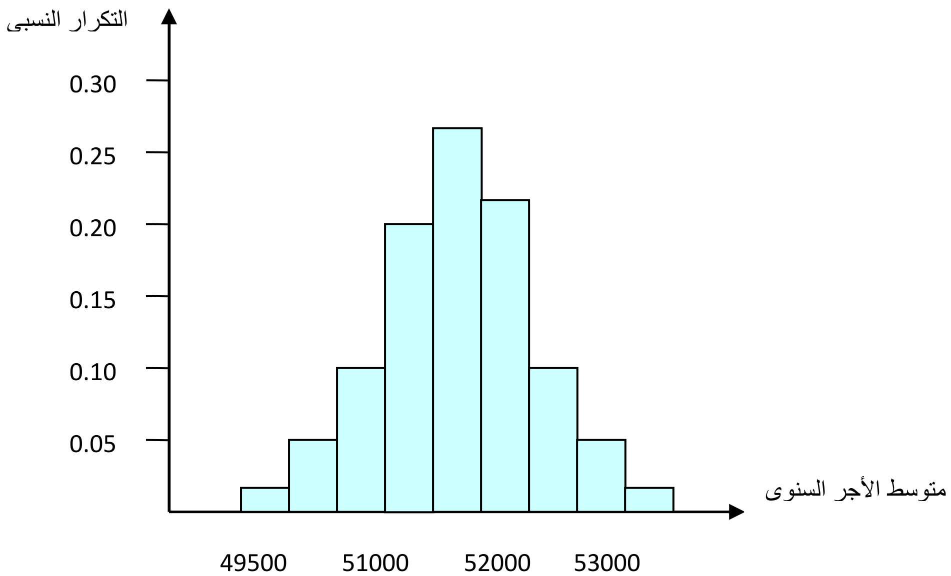
نسبة العينة	الانحراف المعياري للعينة	متوسط العينة	رقم العينة
0.63	3347.00	51814.00	1
0.70	4239.07	52669.70	2
0.67	4433.43	51780.30	3
0.53	3985.32	51587.90	4
0.50	3857.82	51752.00	500

ويوضح الجدول رقم (1−4) التكرار المطلق والنسبي المتعلق بمتوسط الأجر السنوي $\overline{X}$ لـ 500 عينة السابقة. بينما يوضح الشكل رقم (1−1) المدرج التكراري للتكرارات النسبية لقيم $\overline{X}$.

جدول رقم (1−4): التوزيع التكراري لمتوسط الأجر السنوي لـ 500 عينة عشوائية بسيطة يتكون كل منها من 30 عنصر

متوسط الأجر السنوي	التكرار	التكرار النسبي
[49500.00 – 50000.00[	2	0.004
[50000.00 – 50500.00[	16	0.032
[50500.00 – 51000.00[	52	0.104
[51000.00 – 51500.00[	101	0.202
[51500.00 – 52000.00[	133	0.266
[52000.00 – 52500.00[	110	0.220
[52500.00 – 53000.00[	54	0.108
[53000.00 – 53500.00[	26	0.052
[53500.00 – 54000.00[	6	0.012
المجموع	500	1.000

شكل رقم (1−1): مدرج تكراري للتكرارات النسبية لقيم متوسط الأجر السنوي لـ 500 عينة عشوائية بسيطة يتكون كل منها من 30 عنصر



يوضح الشكل رقم $(1-1)$ شكل توزيع المعاينة لـ $\overline{X}$، ونلاحظ أنه جرسي، وكلما اقتربت قيمة $\overline{X}$ من قيمة $\mu = 51800$ كلما زاد تركيز قيمها، ولتحديد طبيعة هذا التوزيع نتبع طرق سوف نتعرض لها في الفصل الثالث.

- في الحياة التطبيقية لا نسحب سوى عينة واحدة من المجتمع.

1 – 3 – 1 توزيع المعاينة للمتوسط Sampling distribution of $\overline{X}$

إذا كانت لدينا مجموعة من العينات مأخوذة من مجتمع، فإن معظم الأوساط الحسابية لهذه العينات تختلف عن بعضها البعض.

- توزيع المعاينة للمتوسط هو التوزيع الاحتمالي لكل القيم الممكنة لمتوسط العينة $\overline{X}$.

يتميز توزيع المعاينة للمتوسط بالخصائص التالية:

أ – الأمل الرياضي Expected Value of $\overline{X}$

يطلق عليه أيضا مصطلح متوسط المتوسطات، نرمز له بالرمز $\mu_{\overline{X}}$ ويكون مساويا لوسط المجتمع $E(\overline{X}) = \mu_{\overline{X}} = \mu$

ب – الانحراف المعياري Standard Deviation of $\overline{X}$

نرمز له بالرمز $\sigma_{\overline{X}}$ ويحسب في حالتين:

المجتمع منته والسحب بدون إعادة	المجتمع غير منته أو السحب مع الإعادة
$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} \times \sqrt{\frac{N-n}{N-1}}$	$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$

بمقارنة العلاقتين السابقتين نلاحظ أن المقدار $\sqrt{\frac{N-n}{N-1}}$ ضروري لحساب الانحراف المعياري لوسط العينة عندما يكون المجتمع منته والسحب بدون إعادة، ويطلق عليه معامل التصحيح.

في أغلب الحالات فإن المجتمع يكون محدود وكبير، بالتالي تكون قيمة معامل التصحيح قريبة من 1. في هذا الحالات عندما يكون المجتمع محدود والسحب بدون إعادة ويكون $n \leq 0.05N$ فلا داعي لاستخدام معامل التصحيح.

بالعودة إلى الإشكالية السابقة، نجد أن:

$$\mu_{\bar{X}} = \mu = 51800$$

$$30 \leq 0.05 \times 2500 \Rightarrow \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{4000}{\sqrt{30}} = 730.30$$

ج – شكل توزيع المعاينة للمتوسط Form of the Sampling Distribution of $\bar{X}$

لتحديد شكل توزيع المعاينة للمتوسط نجد نفسنا في حالتين، الأولى هي أن المجتمع الذي سحبت منه العينة غير معروف توزيعه، والثانية أن المجتمع معروف ويتبع التوزيع الطبيعي. في الحالة الأولى فإننا بحاجة إلى استخدام واحدة من أهم النظريات الإحصائية التي يمكن تطبيقها في توزيع المعاينة للمتوسط وهي:

- **نظرية النهاية المركزية:** عند سحب عينة عشوائية بسيطة ذات حجم n، فإن توزيع المعاينة للمتوسط $\bar{X}$ يمكن أن يقارب إلى التوزيع الطبيعي عندما يكون حجم العينة كبير.

لتوضيح هذه النظرية سوف نتطرق إلى المفاهيم التالية:

المتغير العشوائي: هو متغير ترتبط قيمه باحتمال تحقق تلك القيم.

المتغير العشوائي المتصل: هو المتغير الذي يمكن أن يأخذ عدد لا نهائيا من القيم داخل أي فترة معلومة. احتمال أن يقع X داخل أي فترة يمثله مساحة التوزيع الاحتمالي (دالة الكثافة) داخل هذه الفترة. والمساحة الكلية تحت المنحنى (الاحتمال) تساوي 1.

التوزيع الطبيعي: هو توزيع احتمالي متصل، جرسي الشكل ومتماثل حول الوسط الحسابي. يمتد إلى مالا نهاية في الاتجاهين، ولكن معظم المساحة (الاحتمال) يتركز حول الوسط الحسابي. تكون دالة كثافته بالشكل:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-1}{2}\left(\frac{x-\mu}{\delta}\right)^2} .$$

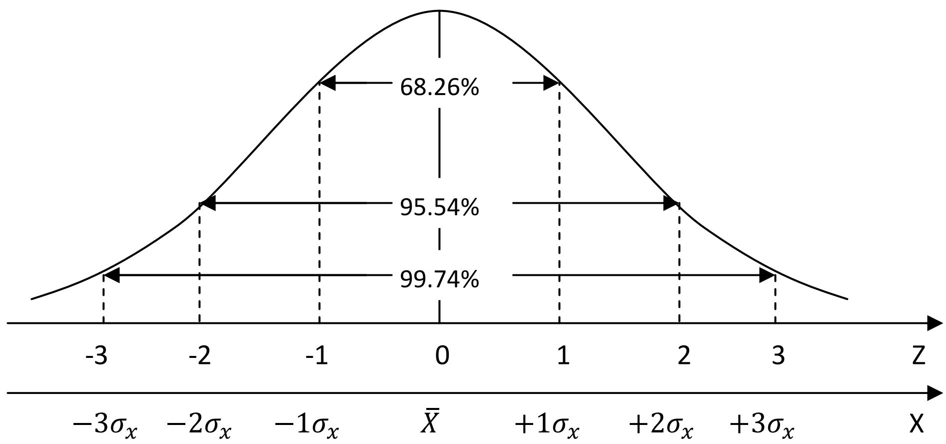
التوزيع الطبيعي القياسي: هو توزيع طبيعي وسطه الحسابي 0، وانحرافه المعياري 1. تكون دالة كثافته بالشكل:

$$f(x) = \frac{1}{\sqrt{2\lambda}} e^{\frac{-1}{2}(\mu)^2} ،$$

يمكن تحويل أي توزيع طبيعي إلى طبيعي قياسي باستخدام العلاقة الشهيرة:

$$Z = \frac{x-\mu}{\sigma}.$$

شكل رقم $(1-2)$: التوزيع الطبيعي



– إن المساحة الكلية المحصورة بين منحنى التوزيع الطبيعي وبين محور السينات تكون مساوية للواحد.

– يعتبر هذا التوزيع متماثل حول وسطه الحسابي. بالتالي، فإن مساحة الجزء الأيسر من التوزيع تكون مساوية لمساحة الجزء الأيمن وتساوي كل منها 0.5. ولحساب المساحة الموجودة في الجهة السالبة فإنه يكفي أن نقوم بتحويلها إلى الجهة الموجبة عن طريق التناظر.

– يمتد هذا التوزيع إلى مالا نهائية في الجهتين.

– يكون التوزيع الطبيعي جرسي الشكل، هذا ما يعني أن معظم المساحة تكون في جوار المتوسط الحسابي، وبالعودة للشكل رقم $(1-2)$، نلاحظ أن 68.26% من مفردات المجتمع (المساحة) تكون قيمها محصورة ضمن المجال $\bar{X} \pm 1\sigma_x$، في حين أن 95.54% تكون قيمها محصورة ضمن المجال $\bar{X} \pm 2\sigma_x$ ، في حين أن 99.74% تكون محصورة قيمها ضمن المجال $\bar{X} \pm 3\delta_x$.

– لحساب أي احتمال مطلوب، وليكن داخل المجال [a–b] مثلا، فعلينا حساب حجم المساحة المحصورة بين الإحداثيات X=a و X=b التي تكتب بالصيغة P(a<X<b) بحيث a<b.

بالعودة إلى نظرية النهاية المركزية و التي تنص على أنه إذا كان x_1، x_2،....x_n متغيرات عشوائية مستقلة ومتماثلة في التوزيع ولها وسط μ و تباين δ^2 فإن:

$$\lim_{n\to\infty} p(a \le \frac{\overline{x}-\mu}{\sigma_{\overline{x}}} \le b) = \frac{1}{\sqrt{2\lambda}} \int_a^b e^{\frac{-1}{2}\mu^2} du$$ ،

بتطبيق هذه النظرية في المعاينة نجد:

أ – إذا كان X يتبع قانون طبيعي بوسط μ وتباين σ^2 $X \mapsto N(\mu,\sigma^2)$ فإن $\overline{X} \mapsto N(\mu_{\overline{x}},\frac{\sigma^2}{n})$ بالتالي:

$$Z = \frac{\overline{x}-\mu_{\overline{x}}}{\sigma_{\overline{x}}} \mapsto N(0,1) .$$

ب – إذا كان X يتبع توزيع غير معروف بوسط μ وتباين σ^2 $X \mapsto N(\mu,\sigma^2)$ فإن $\overline{X} \mapsto N(\mu_{\overline{x}},\frac{\sigma^2}{n})$ بالتالي:

$$Z = \frac{\overline{x}-\mu_{\overline{x}}}{\sigma_{\overline{x}}} \mapsto N(0,1) .$$

مما سبق نستنتج أنه:

- إذا كانت مفردات المجتمع تتبع التوزيع الطبيعي، فإن توزيع المعاينة للمتوسط يكون موزع طبيعيا.
- أما إذا كانت مفردات المجتمع غير معروفة، فإن توزيع المعاينة للمتوسط يكون مقارب للتوزيع الطبيعي عندما $n \ge 30$.

مثال رقم (1–6): بالعودة إلى المثال رقم 5، أوجد شكل توزيع المعاينة للمتوسط مع التمثيل البياني.

شكل توزيع المعاينة لمتوسط الأجر السنوي لمسؤولي الشركة يكون بالصورة التالية:

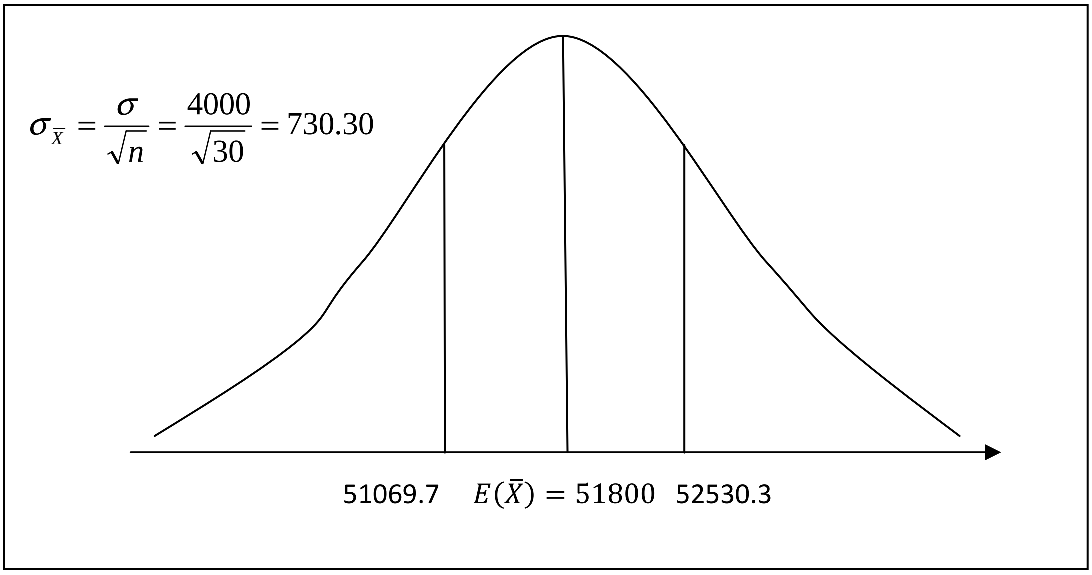
لقد تم إثبات أن

$$\mu_{\bar{X}} = \mu = 51800$$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{4000}{\sqrt{30}} = 730.30$$

استنادا إلى المدرج التكراري الممثل في الشكل رقم (1−1)، يمكن إدراج المنحنى التكراري الذي يمثل شكل توزيع المعاينة للمتوسط، الذي نمثله في الشكل رقم (1−3).

شكل رقم (1−3): شكل توزيع المعاينة لمتوسط الأجر السنوي لمسؤولي الشركة



مثال رقم (1−7): يبلغ متوسط وزن 500 كرية 5.02غ وانحرافها المعياري 0.3غ. أوجد احتمال أن يكون متوسط وزن عينة عشوائية من 100 كرية تم اختيارها من هذه المجموعة:

أ− بين 4.96 − 5.00،

ب− أكبر من 5.10 .

الحل:

لدينا متوسط العينة

$$\mu_{\bar{x}} = \mu = 5.02$$

بما أن المجتمع محدود و $n \geq 0.05N \Rightarrow 100 \geq 0.05(500)$ فإن الانحراف المعياري للتقدير يساوي:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} = \frac{0.3}{\sqrt{100}}\sqrt{\frac{500-100}{500-1}} = 0.027$$

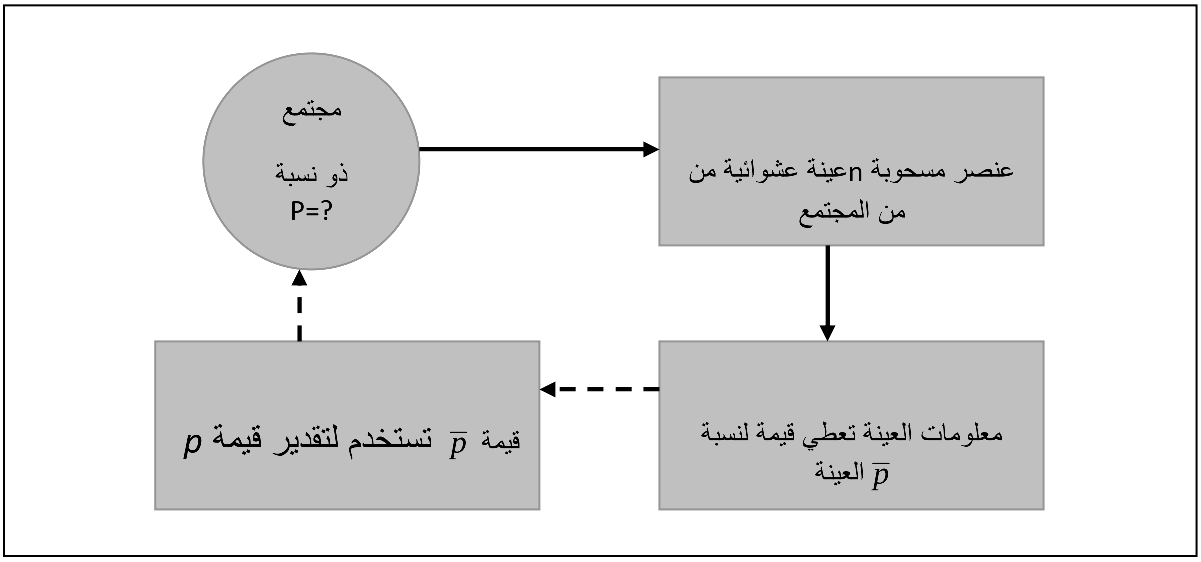
أ –
$$p(4.96 \langle \bar{x} \langle 5.00) \Rightarrow p(\frac{4.96-5.02}{0.027} \langle Z \langle \frac{5.00-5.02}{0.027}) \Rightarrow p(-2.22 \langle Z \langle -0.74)$$
$$= Z(2.22) - Z(0.74) = 0.9868 - 0.7704 = 0.2164$$

ب –
$$p(\bar{x} \rangle 5.1) \Rightarrow p(Z \rangle \frac{5.1-5.02}{0.027}) \Rightarrow p(Z \rangle 2.96) = 1 - Z(2.96) = 1 - 0.9985 = 0.0015$$

1 – 3 – 2 توزيع المعاينة للنسب

في العديد من المجالات الاقتصادية تستخدم نسب العينات $\bar{p}$ لعمل استدلالات حول نسبة المجتمع p، هذه العملية موضحة في الشكل رقم (1–4):

شكل رقم (1–4): عملية إحصائية توضح استخدام نسبة عينة من أجل عمل استدلال حول نسبة المجتمع



المصدر:

David R. Anderson, Dennis J. Sweeney, Thomas A. Williams (2005). Statistiques pour l'économie et la gestion, De Boeck. France, p 343.

لنفرض أنه لدينا مجتمع غير منته ويتبع توزيع ثنائي الحد و أن p و q هما احتمال أن يظهر عنصر ما أو لا يظهر، ولنعتبر كل العينات الممكنة التي حجمها n وسحبت من ذلك المجتمع، ونحدد لكل عينة الإحصائية التي هي نسبة النجاح $\bar{p}$.

- توزيع المعاينة للنسب هو التوزيع الاحتمالي لكل القيم الممكنة لنسبة العينة $\bar{p}$.

يحمل توزيع المعاينة للنسب الخصائص التالية:

أ – الأمل الرياضي:

هو عبارة عن متوسط كل القيم الممكنة لـ $\bar{p}$ ، ونرمز له بالرمز $\mu_{\bar{p}}$ ، ويكون مساويا لنسبة المجتمع $E(\bar{p}) = \mu_{\bar{p}} = p$

ب – الانحراف المعياري:

نرمز له بالرمز $\sigma_{\bar{p}}$ وعلى غرار توزيع المعينة للمتوسط نجد حالتين لحسابه:

المجتمع منته والسحب بدون إعادة	المجتمع غير منته أو السحب مع الإعادة
$\sigma_{\bar{p}} = \sqrt{\frac{pq}{n}} \times \sqrt{\frac{N-n}{N-1}}$	$\sigma_{\bar{p}} = \sqrt{\frac{pq}{n}}$

في أغلب الحالات فإن المجتمع يكون محدود وكبير، بالتالي تكون قيمة معامل التصحيح قريبة من 1. في هذا الحالات عندما يكون المجتمع محدود والسحب بدون إعادة ويكون $n \leq 0.05N$ فلا داعي لاستخدام معامل التصحيح.

بالعودة إلى الإشكالية السابقة نعلم بأن نسبة العمال الذين تلقوا تربصا تكوينيا هو p=0.60 ومع $n \leq 0.05N$ فإننا نلغي معامل التصحيح بالتالي:

$$\mu_{\bar{p}} = p = 0.6$$

$$\sigma_{\bar{p}} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{0.6 \times 0.4}{30}} = 0.0894$$

ج – شكل توزيع المعاينة للنسب:

الآن، وبعد أن عرفنا الوسط والانحراف المعياري لـ $\bar{p}$ ، لنفرض شكل توزيع معاينة للنسب. باستخدام نظرية النهاية المركزية نحصل على النتيجة التالية:

- توزيع المعاينة للنسب يمكن أن يقارب إلى التوزيع الطبيعي، إذا كان حجم العينة كبير $(n \geq 30)$.

مثال رقم (1-8): قدر مدير شركة أن 30% من الطلبيات المقدمة للشركة هي من عملاء جدد. نسحب عينة حجمها 100 من الطلبيات لمعرفة نسبة طلبيات الزبائن الجدد. نتائج هذه العينة سنستخدمها لمعرفة قدرة المدير على التنبؤ.

أ- نفرض أن المدير مصيب ونسبة المجتمع هي 0.3، ما هو توزيع المعاينة لنسبة العينة في هذه الدراسة؟

ب- ما هو احتمال أن يكون توزيع المعاينة للنسبة محصورا بين 0.2 و 0.4؟

ج- ما هو احتمال أن ينحرف توزيع المعاينة للنسبة عن نسبة المجتمع بـ ±0.05؟

الحل:

أ- تحديد توزيع المعاينة لنسبة العينة:

$$\mu_{\bar{p}} = p = 0.3$$

$$\sigma_{\bar{p}} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{0.3\times 0.6}{100}} = 0.0458$$

ب- حساب احتمال أن يكون توزيع المعاينة للنسبة محصورا بين 0.2 و 0.4:

$$p(0.20 < \bar{P} < 0.40)$$

$$= p\left(\frac{\bar{P}-\mu_{\bar{P}}}{\sigma_{\bar{P}}} < Z < \frac{\bar{P}-\mu_{\bar{P}}}{\sigma_{\bar{P}}}\right)$$

$$= p\left(\frac{\bar{P}-\frac{1}{2n}-\mu_{\bar{P}}}{\sigma_{\bar{P}}} < Z < \frac{\bar{P}+\frac{1}{2n}-\mu_{\bar{P}}}{\sigma_{\bar{P}}}\right)$$

$$= p\left(\frac{0.20-\frac{1}{2(100)}-0.30}{0.0458} < Z < \frac{0.40+\frac{1}{2(100)}-0.30}{0.0458}\right)$$

$$= p(-2.29 < Z < 2.29) = 0.978$$

ب- حساب احتمال أن ينحرف توزيع المعاينة للنسبة عن نسبة المجتمع بـ ±0.05:

$$p(0.35 < \bar{P} < 0.35)$$

$$= p\left(\frac{\bar{P}-\mu_{\bar{P}}}{\sigma_{\bar{P}}} < Z < \frac{\bar{P}-\mu_{\bar{P}}}{\sigma_{\bar{P}}}\right)$$

$$= p\left(\frac{\bar{P}-\frac{1}{2n}-\mu_{\bar{P}}}{\sigma_{\bar{P}}} < Z < \frac{\bar{P}+\frac{1}{2n}-\mu_{\bar{P}}}{\sigma_{\bar{P}}}\right)$$

$$= p\left(\frac{0.25-\frac{1}{2(100)}-0.30}{0.0458} < Z < \frac{0.35+\frac{1}{2(100)}-0.30}{0.0458}\right)$$

$$= p(-1.20 < Z < 1.20) = 0.7698$$

1 – 3 – 3 توزيع المعاينة للفروق والمجاميع

لنفترض أنه لدينا مجتمعين، نسحب من المجتمع الأول N_1 عينة حجمها n_1، ونسحب من المجتمع الثاني N_2 عينة حجمها n_2 مستقلة.

يعرف توزيع المعاينة للفرق بين المتوسطين كالتالي:

$$\mu_{\bar{x}_1-\bar{x}_2} = \mu_{\bar{x}_1} - \mu_{\bar{x}_2} = \mu_1 - \mu_2$$

ويكون انحرافه المعياري:

المجتمع منته والسحب بدون إعادة	المجتمع غير منته أو السحب مع الإعادة
$\sigma_{\bar{X}_1-\bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1}(\frac{N_1-n_1}{N_1-1})+\frac{\sigma_2^2}{n_2}(\frac{N_2-n_2}{N_2-1})}$	$\sigma_{\bar{X}_1-\bar{X}_2} = \sqrt{\sigma_{\bar{X}_1}+\sigma_{\bar{X}_2}} = \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$

ملاحظات:

1 – يكون المتغير المعياري $Z = \frac{(\bar{X}_1-\bar{X}_2)-(\mu_{\bar{X}1}-\mu_{\bar{X}2})}{\sigma_{\bar{X}_1-\bar{X}_2}}$ موزع بشكل قريب جدا من التوزيع الطبيعي إذا كان $n_1, n_2 \geq 30$

2 – في حالة كون المجتمعان يتبعان توزيع ثنائي الحد والمجتمع غير منته يكون لدينا:

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

$$\sigma_{\bar{p}_1-\bar{p}_2} = \sqrt{\sigma^2_{\bar{p}_1} + \sigma^2_{\bar{p}_2}}$$

3 – إذا كنا نهتم بتوزيع المعاينة لمجموع وسطين والمجتمع غير منته يكون لدينا:

$$\mu_{\bar{x}_1+\bar{x}_2} = \mu_{\bar{x}_1} + \mu_{\bar{x}_2} = \mu_1 + \mu_2$$

$$\sigma_{\bar{X}_1+\bar{X}_2} = \sqrt{\sigma_{\bar{X}_1} + \sigma_{\bar{X}_2}} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

مثال رقم (1−9):

تنتج شركة A مصابيح متوسط مدة حياتها 1400 سا بانحراف معياري 200 سا. وتنتج شركة B مصابيح متوسط مدة حياتها 1200 سا بانحراف معياري 100 سا.

قمنا باختيار عينة من 125 وحدة من كلتا الشركتين مع الإعادة. أوجد احتمال أن الشركة A تنتج مصابيح كهربائية متوسط مدة حياتها على الأقل أكبر بـ 160 سا من عمر مصابيح الشركة B؟

الحل:

$$\mu_{\bar{x}_1-\bar{x}_2} = \mu_1 - \mu_2 = 1400 - 1200 = 200$$

$$\sigma_{\bar{X}_1-\bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{200^2}{125} + \frac{100^2}{125}} = 20$$

$$p((\bar{x}_1 - \bar{x}_2) \rangle 160) = p(Z \rangle \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sigma_{\bar{X}_1-\bar{X}_2}})$$

$$= p(Z \rangle \frac{160-200}{20}) = p(Z \rangle -2) = 0.9772$$

1 – 4 تمارين محلولة

تمرين رقم 1-1:

نفرض أن X متغير عشوائي يمثل ظهور الصورة عند رمي قطعة نقود متزنة 10 مرات، أوجد:

1- قانون احتمال X؟

2 – احتمال الحصول على صورة بين 3 و 6 باستخدام:

أ- توزيع ثنائي الحد،

ب- التوزيع الطبيعي كتقريب لتوزيع ثنائي الحد مع التمثيل البياني.

الحل:

1- تحديد قانون احتمال X

لدينا X متغير عشوائي كمي منفصل يمثل ظهور الصورة عند رمي قطعة نقود متزنة 10 مرات، بالتالي هذا المتغير يتبع قانون ثنائي الحد بالمعلمات التالية:

n=10, p = 1/2, q =1/2

$$p(x=k)=C_n^k p^k q^{n-k}$$

يتمثل فراغ الحوادث للمتغير العشوائي X في المجموعة التالية:

$$\Omega=\{0,1,2,3,4,5,6,7,8,9,10\}$$

الاحتمالات المناظرة لكل قيم المتغير العشوائي يتم حسابها باستخدام العلاقة:

$$p(x=0)=C_{10}^0(\frac{1}{2})^0(\frac{1}{2})^{10-0}=0.000976=\frac{1}{1024}$$

$$p(x=1)=C_{10}^1(\frac{1}{2})^1(\frac{1}{2})^{10-1}=0.00976=\frac{10}{1024}$$

$$p(x=2)=C_{10}^2(\frac{1}{2})^2(\frac{1}{2})^{10-2}=0.04392=\frac{45}{1024}$$

$$p(x=3)=C_{10}^3(\frac{1}{2})^3(\frac{1}{2})^{10-3}=0.1171=\frac{120}{1024}$$

$$p(x=4)=C_{10}^4(\frac{1}{2})^4(\frac{1}{2})^{10-4}=0.2050=\frac{210}{1024}$$

$$p(x=5)=C_{10}^{5}\left(\frac{1}{2}\right)^{5}\left(\frac{1}{2}\right)^{10-5}=0.2460=\frac{252}{1024}$$

$$p(x=6)=C_{10}^{6}(\frac{1}{2})^{6}(\frac{1}{2})^{10-6}=0.2050=\frac{210}{1024}$$

$$p(x=7)=C_{10}^{7}(\frac{1}{2})^{7}(\frac{1}{2})^{10-7}=0.1171=\frac{120}{1024}$$

$$p(x=8)=C_{10}^{8}(\frac{1}{2})^{8}(\frac{1}{2})^{10-8}=0.04392=\frac{45}{1024}$$

$$p(x=9)=C_{10}^{9}(\frac{1}{2})^{9}(\frac{1}{2})^{10-9}=0.00976=\frac{10}{1024}$$

$$p(x=10)=C_{10}^{10}(\frac{1}{2})^{10}(\frac{1}{2})^{10-10}=0.000976=\frac{1}{1024}$$

يمكن كتابة قانون احتمال X في صورة جدولية كالتالي:

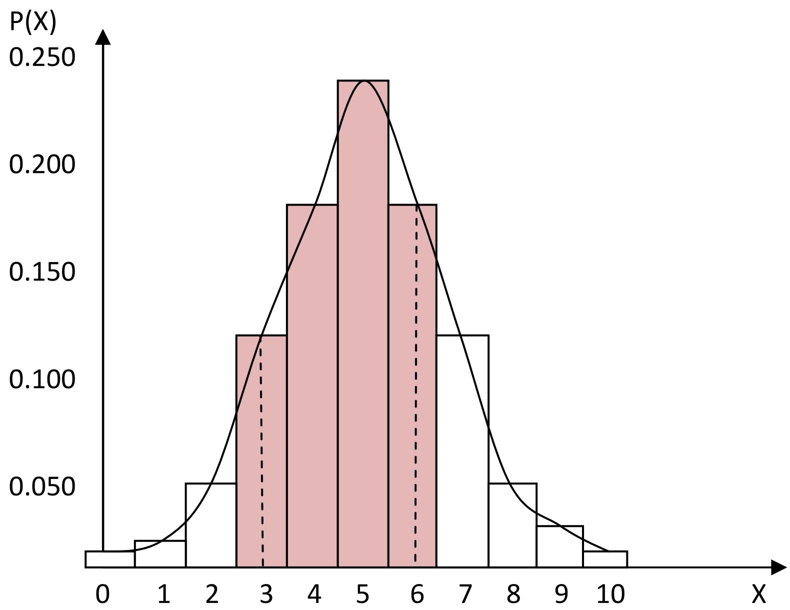
X	0	1	2	3	4	5	6	7	8	9	10
P(X)	$\frac{1}{1024}$	$\frac{10}{1024}$	$\frac{45}{1024}$	$\frac{120}{1024}$	$\frac{210}{1024}$	$\frac{252}{1024}$	$\frac{210}{1024}$	$\frac{120}{1024}$	$\frac{45}{1024}$	$\frac{10}{1024}$	$\frac{1}{1024}$

2- حساب احتمال الحصول على صورة بين 3 و 6 باستخدام:

أ- توزيع ذي الحدين

$$p(3\leq X\leq 6)=p(X=3)+p(X=4)+p(X=5)+p(X=6)$$
$$=\frac{120}{1024}+\frac{210}{1024}+\frac{252}{1024}+\frac{210}{1024}=\frac{792}{1024}=0.7734$$

ب- التوزيع الطبيعي كتقريب لتوزيع ثنائي



إن توزيع الاحتمال للمتغير العشوائي X الذي يمثل عدد مرات ظهور الصورة عند رمي قطعة نقود متزنة 10 مرات ممثل بالشكل السابق. والملاحظ على هذا الشكل أننا عاملنا البيانات وكأنها متصلة (حيث أن الرسم يجب أن يكون بأعمدة لأن المتغير العشوائي منفصل). إن الاحتمال المطلوب ممثل بالمستطيلات المظللة، والتي يمكن أن يتم تقريبها بواسطة المنحنى التكراري (التوزيع الطبيعي) الممثل في نفس الشكل.

نلاحظ من خلال الشكل السابق أن المنحنى الطبيعي هو توزيع متصل، لهذا فإننا يجب أن نقوم بتحويل المطلوب من التوزيع الثنائي إلى التوزيع الطبيعي. وهذا عن طريق إضافة أو طرح المقدار 0.5 حسب ما هو مطلوب في السؤال. والهدف من هذه الخطوة، هو ضم المساحة الناقصة من المدرج التكراري إلى الاحتمال المطلوب حسابه على النحو التالي:

ثنائي الحد $p(3 \leq X \leq 6) \backsim B(n,p,q)$

طبيعي $p(2.5 < X \leq 6.5) \backsim N(\mu,\delta^2)$

نقوم بحساب معلمات التوزيع الطبيعي باستخدام العلاقات التالية:

$$\mu = np = 10\left(\frac{1}{2}\right) = 5$$

$$\sigma = npq = 10\left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = 1.58$$

الاحتمال المطلوب يعطى بالعلاقة:

$$\begin{aligned} p(2.5 < X < 6.5) &= p\left(\frac{2.5-5}{1.58} < Z < \frac{6.5-5}{1.58}\right) = p(-1.58 < Z < 0.95) \\ &= [Z_{0.95} - Z_0] + [Z_{1.58} - Z_0] = [0.8289 - 0.5] + [0.9429 - 0.5] \\ &= 0.7718 \end{aligned}$$

نلاحظ أن الاحتمال المحسوب باستخدام التوزيع الطبيعي كتقريب لتوزيع ثنائي الحد أعطى نتائج قريبة للقيمة المحسوبة عن طريق استخدام توزيع ثنائي الحد.

تمرين رقم 1−2:

يتكون مجتمع من خمسة أرقام 1 ، 2 ، 4 ، 6 ، 7. اعتبر كل العينات الممكنة التي يكون حجمها اثنين و التي يمكن سحبها مع الإرجاع من هذا المجتمع. أوجد:

1 – متوسط المجتمع μ؟

2 – الانحراف المعياري للمجتمع σ؟

3 – متوسط توزيع المعاينة للأوساط $\mu_{\bar{X}}$؟

4 – الانحراف المعياري لتوزيع المعينة للأوساط $\sigma_{\bar{X}}$؟

5 – حل المسألة السابقة في حالة المعاينة دون إرجاع؟

الحل:

1 – حساب متوسط المجتمع μ

$$\mu = \frac{1+2+4+6+7}{5} = 4$$

2 – الانحراف المعياري للمجتمع σ

$$\sigma=\sqrt{\frac{\sum(x_i-\mu)^2}{n}}=\sqrt{\frac{\sum x_i^2}{n}-\mu^2}=\sqrt{\frac{106}{5}-4^2}=2.28$$

3 – متوسط توزيع المعاينة للأوساط $\mu_{\bar{x}}$

هناك 5^2 عينة ذات حجم اثنين يمكن سحبها بإرجاع من هذا المجتمع. وهي معروضة في الجدول الموالي:

متوسطاتها الحسابية					العينات				
4	3.5	2.5	1.5	1	(7,1)	(6,1)	(4,1)	(2,1)	(1,1)
4.5	4	3	2	1.5	(7,2)	(6,2)	(4,2)	(2,2)	(1,2)
5.5	5	4	3	2.5	(7,4)	(6,4)	(4,4)	(2,4)	(1,4)
6.5	6	5	4	3.5	(7,6)	(6,6)	(4,6)	(2,6)	(1,6)
7	6.5	5.5	4.5	4	(7,7)	(6,7)	(4,7)	(2,7)	(1,7)

$$\mu_{\bar{x}}=\frac{1+1.5+\cdots.+6.5+7}{25}=4$$

4 – الانحراف المعياري لتوزيع المعينة للأوساط $\delta_{\bar{x}}$

$$\sigma_{\bar{x}}=\sqrt{\frac{\sum(\bar{x}-\mu_{\bar{x}})^2}{n}}=\sqrt{\frac{\sum\bar{x}^2}{n}-\mu_{\bar{x}}{}^2}=\sqrt{\frac{465}{25}-4^2}=1.61$$

هذه النتيجة تؤكد أنه في حالة المجتمعات المحدودة المتضمنة المعاينة مع الإعادة فإن:

$$\sigma_{\bar{x}}=\frac{\delta}{\sqrt{n}}=\frac{2.28}{\sqrt{2}}=1.61$$

5 – في حالة المعاينة بدون إعادة.

أ– متوسط المجتمع وانحرافه المعياري تبقى بدون تغيير.

ب- متوسط توزيع المعاينة للأوساط $\mu_{\bar{x}}$

هناك $C_2^5 = 10$ عينة ذات حجم 2 يمكن سحبها بدون إرجاع. وهي:

متوسطاتها الحسابية					العينات				
3	4	3.5	2.5	1.5	(2,4)	(1,7)	(1,6)	(1,4)	(1,2)
6.5	5.5	5	4.5	4	(6,7)	(4,7)	(4,6)	(2,7)	(2,6)

$$\mu_{\bar{x}} = \frac{1.5 + 2.5 + \cdots + 5.5 + 6.5}{10} = 4$$

4 – الانحراف المعياري لتوزيع المعيانة للأوساط $\sigma_{\bar{x}}$

$$\sigma_{\bar{x}} = \sqrt{\frac{\sum(\bar{x} - \mu_{\bar{x}})^2}{n}} = \sqrt{\frac{\sum \bar{x}^2}{n} - \mu_{\bar{x}}{}^2} = \sqrt{\frac{179.5}{10} - 4^2} = 1.39$$

هذه النتيجة تؤكد أنه في حالة المجتمعات المحدودة المتضمنة المعاينة بدون إعادة فإن:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} = \frac{2.28}{\sqrt{2}}\sqrt{\frac{5-2}{5-1}} = 1.39$$

تمرين رقم 1-3:

إذا كان متوسط الدخل الشهري للأسرة الجزائرية 40000 دج بانحراف معياري 10000 دج ويتبع التوزيع الطبيعي.

ما هو احتمال أن تكون عينة عشوائية حجمها 20 مسحوبة من هذا المجتمع دخلا شهريا وسطيا:

1 – أقل من 35000؟

2 – بين 38000 و 42000؟

3 – أكبر من 38000؟

الحل:

المجتمع غير محدود ويتبع التوزيع الطبيعي. بالتالي:

$$\mu_{\bar{x}} = \mu = 40000$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{10000}{\sqrt{25}} = 2000$$

1 – أقل من 35000.

$$p(\bar{X} < 35000)$$

$$= p\left(Z < \frac{35000 - 40000}{2000}\right)$$

$$= p(Z < -2.5) = 1 - Z_{2.5} = 1 - 0.9938 = 0.0062$$

2 – بين 38000 و42000.

$$p(38000 < X < 42000)$$

$$= p\left(\frac{38000 - 40000}{2000} < Z < \frac{42000 - 40000}{2000}\right)$$

$$= p(-1 < Z < 1) = 2[Z_1 - Z_0] =$$

$$2[0.8413 - 0.5] = 06826$$

3 – أكبر من 38000.

$$p(\bar{X} > 38000)$$

$$= p\left(Z > \frac{38000 - 40000}{2000}\right)$$

$$= p(Z > -1) = Z_1 = 0.8413$$

تمرين رقم 1-4:

إن المدة المتوسطة لكي يصل فيها سكان مدينة شيكاغو إلى العمل هي 31.5 دقيقة، بانحراف معياري قدره 12 دقيقة. نسحب عينة حجمها 50 شخص.

1- أوجد توزيع المعاينة للمتوسط؟

2- ما هو احتمال أن ينحرف توزيع المعاينة للمتوسط بـ 1 دقيقة عن متوسط المجتمع؟

الحل:

1- تحديد توزيع المعاينة للمتوسط:

$$\mu_{\bar{x}} = \mu = 31.5$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{12}{\sqrt{50}} = 1.70$$

2- حساب احتمال أن ينحرف توزيع المعاينة للمتوسط بـ 1 دقيقة عن متوسط المجتمع:

$$p(34.1 < X < 36.1)$$

$$= p\left(\frac{-1}{1.70} < Z < \frac{+1}{1.70}\right)$$

$$= p(-0.59 < Z < 0.59) = 0.4448$$

تمرين رقم (1-5):

تفيد دراسة لمعرفة العامل الأهم الذي يدفع أشخاص إلى اختيار فندق، أجرتها مجلة USA to day بأن 74% يؤكدون أن وجود قاعة لغير المدخنين يعتبر أهم عامل. نسحب عينة حجمها 200 شخص.

1- ما هو توزيع المعاينة لنسبة العينة في هذه الدراسة؟

2- ما هو احتمال أن ينحرف توزيع المعاينة للنسبة عن نسبة المجتمع بـ ± 0.04؟

3- ما هو احتمال أن ينحرف توزيع المعاينة للنسبة عن نسبة المجتمع بـ ± 0.02؟

الحل:

أ- تحديد توزيع المعاينة لنسبة العينة:

$$\mu_{\bar{p}} = p = 74$$

$$\sigma_{\bar{p}} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{0.3 \times 0.6}{100}} = 0.031$$

ب− حساب احتمال أن ينحرف توزيع المعاينة للنسبة عن نسبة المجتمع بـ ± 0.04:

$$p(0.70 < \bar{P} < 0.78)$$

$$= p\left(\frac{\bar{P} - \mu_{\bar{P}}}{\sigma_{\bar{P}}} < Z < \frac{\bar{P} - \mu_{\bar{P}}}{\sigma_{\bar{P}}}\right)$$

$$= p\left(\frac{\bar{P} - \frac{1}{2n} - \mu_{\bar{P}}}{\sigma_{\bar{P}}} < Z < \frac{\bar{P} + \frac{1}{2n} - \mu_{\bar{P}}}{\sigma_{\bar{P}}}\right)$$

$$= p\left(\frac{0.70 - \frac{1}{2(200)} - 0.74}{0.031} < Z < \frac{0.78 + \frac{1}{2(200)} - 0.74}{0.031}\right)$$

$$= p(-1.37 < Z < 1.37) = 0.8249$$

ب− حساب احتمال أن ينحرف توزيع المعاينة للنسبة عن نسبة المجتمع بـ ± 0.02:

$$p(0.72 < \bar{P} < 0.76)$$

$$= p\left(\frac{\bar{P} - \mu_{\bar{P}}}{\sigma_{\bar{P}}} < Z < \frac{\bar{P} - \mu_{\bar{P}}}{\sigma_{\bar{P}}}\right)$$

$$= p\left(\frac{\bar{P} - \frac{1}{2n} - \mu_{\bar{P}}}{\sigma_{\bar{P}}} < Z < \frac{\bar{P} + \frac{1}{2n} - \mu_{\bar{P}}}{\sigma_{\bar{P}}}\right)$$

$$= p\left(\frac{0.72 - \frac{1}{2(200)} - 0.74}{0.031} < Z < \frac{0.76 + \frac{1}{2(200)} - 0.74}{0.031}\right)$$

$$= p(-0.72 < Z < 0.72) = 0.0.5284$$

تمرين رقم (1−6):

يلعب A و B مباراة في الصورة والكتابة، حيث يقوم كل منهم يرمي عملة متوازنة 50 مرة. سوف يكسب A المباراة إذا ظهر في رمياته 5 صور أو أكثر من تلك التي حصل عليها B. بخلاف ذلك يكسب B.

المطلوب: ما هو احتمال فوز A؟

الحل:

$$\mu_{\bar{p}_1-\bar{p}_2} = P_1 - P_2 = 0$$

$$\sigma_{\bar{p}_1-\bar{p}_2} = \sqrt{\frac{p_1q_1}{n_1}+\frac{p_2q_2}{n_2}} = \sqrt{\frac{0.5\times0.5}{50}+\frac{0.5\times0.5}{50}} = 0.1$$

$$p((\bar{p}_1-\bar{p}_2)\rangle 0.1) = p(Z\rangle \frac{(\bar{p}_1-\bar{p}_2)+\frac{1}{2n}}{\sigma_{\bar{p}_1-\bar{p}_2}})$$

$$= p(Z\rangle \frac{0.1-200}{0.1}) = p(Z\rangle 0.9) = 0.1841$$

الفصل الثاني: التقدير Estimation

1-1 تمهيد

يعتبر التقدير الأداة الإحصائيات التي تسمح لنا بإجراء استدلالات حول معالم المجتمع انطلاقا من عينة مسحوبة منه. هذا الاستدلال لا يصح إلا في ظل شروط محددة، تتعلق أساسا بفرضيات تتمحور حول طبيعة بيانات المجتمع والعينة. ويعتبر التقدير من الأهمية بما كان. نظرا لأنه يسمح لنا بان نقوم بتحديد قيم لمعالم مجتمعات يكون عدد أفرادها كبير. ولهذا فإننا سوف نقتصد في الوقت والتكلفة اللازمة لإجراء التجربة حول الظاهرة المراد دراستها. على سبيل المثال، من أجل معرفة مدى شعبية مرشح لانتخابات ما. يكفي أن نأخذ عينة من الناخبين بطرق علمية تكون ممثلة لذلك المجتمع. إنطلاقا من نتائج هذه العينة يمكن للمرشح أن يعرف مدى شعبيته ومدى نجاح حملته الانتخابية، وهذا لتلافي أي قصور ممكن في الأداء.

2-2 مفهوم التقدير

يعرف التقدير بأنه:" عملية استنتاج أو تقدير أحد معالم المجتمع (مثل الوسط الحسابي أو الانحراف المعياري) من الإحصائية المناظرة والخاصة بعينة مسحوبة من المجتمع".[1]

كما يعرف التقدير بأنه: "الحصول على كمية ما أو مقدار عدم التيقن المقترن بها من خلال تعيين قيم عددية للمشاهدة في صيغة تقدير أو مقدر. ويمكن التعبير عن نتائج التقدير على النحو التالي:

- تقدير بنقطة،
- تقدير بمجال ثقة.

من هذه التعريفات نستنتج أن التقدير هو عملية إحصائية تسمح لنا بتحديد قيمة معلمة ما محل اهتمام وذلك انطلاقا من إحصائية عينة مسحوبة عشوائيا من المجتمع. وتسمى القيمة المتحصل عليها بالمقدر Estimator الذي يمكن تعريفه بأنه: "صيغة تحدد كيفية حساب القيم التقديرية لمعلمة مجتمع من خلال بيانات العينة".

3-2 التقدير النقطي Point estimation

كما أسلفنا الذكر، يمكن إعطاء تقديرات لمعالم مجتمع انطلاقا من إحصائيات عينة عشوائية مسحوبة منه بعدد واحد. يسمى حينها هذا المقدر بالمقدر النقطي الذي يعرف بأنه: "عملية تقدير معلمة مجتمع بعدد".

وسنعرض في هذا الجزء تقديرين نقطيين لمتوسط وتباين مجتمع.

2-3-1 تقدير نقطي لمتوسط مجتمع

يعتبر متوسط العينة $\bar{X}$ مقدر نقطي لمتوسط المجتمع μ بحيث:

$$\hat{\mu} = \bar{X}$$

ملاحظة: الرمز $\widehat{\ }$ يستخدم للدلالة على القيمة المقدرة.

لدينا:

$$E(\bar{X}) = E\left(\frac{\sum_{i=1}^{n} x_i}{n}\right) = \left(\frac{\sum_{i=1}^{n} E(x_i)}{n}\right) = \left(\frac{\sum_{i=1}^{n} \mu}{n}\right) = \frac{n\mu}{n} = \mu$$

[1]- دومينيك سالفادور، الإحصاء والاقتصاد القياسي، ديوان المطبوعات الجامعية، الجزائر، 1993، ص 72.

2-3-2 تقدير نقطي لتباين مجتمع

يعرف تباين مجتمع بالعلاقة التالية:

$$\sigma^2 = \frac{\sum_{n=1}^{N}(X_n-\mu)^2}{N} \qquad (1)$$

ويعرف تباين عينة بالعلاقة التالية:

$$S^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{X})^2}{n} \qquad (2)$$

لدينا:

$$E(S^2) = E\left(\frac{\sum_{i=1}^{n}(x_i-\bar{X})^2}{n}\right) \qquad (3)$$

في العلاقة رقم (3) نقوم بوضع:

$$(x_i - \bar{X}) = (x_i - \mu) - (\bar{X} - \mu) \qquad (4)$$

بالتالي:

$$E(S^2) = E\left(\frac{\sum_{i=1}^{n}(x_i-\bar{X})^2}{n}\right) = E\left(\frac{\sum_{i=1}^{n}[(x_i-\mu)-(\bar{X}-\mu)]^2}{n}\right)$$

$$E(S^2) = E\left(\frac{\sum_{i=1}^{n}[(x_i-\mu)^2 - 2(x_i-\mu)(\bar{X}-\mu) + (\bar{X}-\mu)^2]^2}{n}\right)$$

$$E(S^2) = \frac{\sum_{i=1}^{n} E(x_i-\mu)^2}{n} + \frac{\sum_{i=1}^{n} E(\bar{X}-\mu)^2}{n} - 2E\left[\frac{(\bar{X}-\mu)\sum_{i=1}^{n}(x_i-\mu)}{n}\right]$$

مع:

$$\sum_{i=1}^{n}(x_i-\mu) = n(\bar{X}-\mu)$$

بالتالي:

$$E(S^2) = \frac{\sum_{i=1}^{n} E(x_i - \mu)^2}{n} + E(\bar{X} - \mu)^2 - 2E(\bar{X} - \mu)^2$$

$$E(S^2) = E(x_i - \mu)^2 - E(\bar{X} - \mu)^2$$

$$E(S^2) = V(x) - V(\bar{X}) \qquad (5)$$

بالنسبة للمعادلة رقم (5)، يجب الأخذ بعين الاعتبار طبيعة عملية السحب. بالتالي، يجب حساب قيمة $E(S^2)$ في حالة كون العينات مستقلة (السحب مع الإعادة) أو في حالة السحب بدون إعادة.

الحالة الأولى: السحب مع الإعادة

في حالة السحب مع الإعادة فإننا نعلم أن:

$$V(x) = \sigma^2 \; et \; V(\bar{X}) = \frac{\sigma^2}{n} \qquad (6)$$

نقوم بتعويض المعادلة رقم (6) في المعادلة رقم (5):

$$E(S^2) = V(x) - V(\bar{X}) = \sigma^2 - \frac{\sigma^2}{n} = \sigma^2\left(\frac{n-1}{n}\right) \qquad (7)$$

نلاحظ من خلال المعادلة رقم (7) أن المقدار S^2 يعتبر مقدر متحيز[1] للمعلمة σ^2، لأن $E(S^2) \neq \sigma^2$ (إن تحيز المقدار S^2 يساوي $\frac{n-1}{n}$ في حالة السحب مع الإعادة).

لهذا، فإن المقدر غير المتحيز لتباين المجتمع لا يكون S^2، بل S'^2، الذي نحصل عليه على طريق ضرب قيمة S^2 في مقلوب مقدار التحيز $\frac{n}{n-1}$، ويمكننا أن نكتب:

$$\hat{\sigma}^2 = S'^2 = S^2\left(\frac{n}{n-1}\right) = \frac{\sum_{i=1}^{n}(x_i - \bar{X})^2}{n-1}$$

الحالة الثانية: السحب بدون إعادة

في حالة السحب بدون إعادة، فإننا نعلم أن:

$$V(x) = \sigma^2 \; et \; V(\bar{X}) = \frac{\sigma^2}{n}\frac{N-n}{N-1} \qquad (9)$$

[1]- سيتم التعرض لمفهوم التحيز في الجزء 2-6

إذا استبدلنا المعادلة رقم (9) في المعادلة رقم (5)، ينتج لنا:

$$E(S^2) = V(x) - V(\bar{X}) = \sigma^2 - \left(\frac{\sigma^2}{n}\frac{N-n}{N-1}\right) = \sigma^2\left(\frac{N}{n}\frac{n-1}{N-1}\right) \qquad (10)$$

نلاحظ كذلك من خلال المعادلة رقم (10) أن المقدار S^2 يعتبر مقدر متحيز للمعلمة σ^2، لأن $E(S^2) \neq \sigma^2$ (إن تحيز المقدر S^2 يساوي $\frac{N}{n}\frac{n-1}{N-1}$ في حالة السحب بدون الإعادة).

لهذا، فإن المقدر غير المتحيز لتباين المجتمع لا يكون S^2، بل S''^2، الذي نحصل عليه على طريق ضرب قيمة S^2 في مقلوب مقدار التحيز $\frac{n}{N}\frac{N-1}{n-1}$، ويمكننا أن نكتب:

$$\hat{\sigma}^2 = S''^2 = S^2\left(\frac{n}{N}\frac{N-1}{n-1}\right)$$

مثال 2-1: يقوم مصنع بإنتاج قطع ميكانيكية. نقوم بقياس طول كل قطعة في عينة مسحوبة عشوائيا تضم 50 قطعة، فنلاحظ أن المتوسط الحسابي يساوي 64.715 ملم والانحراف المعياري 0.095 ملم. من خلال المعلومات المتوفرة في العينة، أعط تقدير نقطي لمتوسط وتباين المجتمع؟

الحل:

يمكن الحصول على تقدير نقطي لمتوسط وتباين المجتمع باستخدام العبارتين التاليتين:

$$\hat{\mu} = \bar{X} = 64.715\,mm$$

$$\hat{\sigma}^2 = S^2\left(\frac{n}{n-1}\right) = 0.095\left(\frac{50}{50-1}\right) = 0.096$$

2-3-4 تقدير نقطي لنسبة مجتمع

ليكن لدينا مجتمع ما، يتكون من حدثين، الحدث A باحتمال تحقق p، والحدث المتمم للحدث A، باحتمال تحقق q=1-p.

نرمز بـ Xn للمتغير العشوائي الذي يقابله من أجل كل عينة ذات حجم n، عدد العناصر X التي تمثل الحدث A. في هذه الحالة، نحن نبحث عن قيمة النسبة P المجهولة في المجتمع، انطلاقا من إحصائية العينة. لهذا، نقوم بعملية سحب مع الإعادة لعينة حجمها n من هذا المجتمع. ونعرف التكرار المشاهد للحدث A كالتالي:

$$f = \frac{X_n}{n}$$

مع Xn تمثل عدد عناصر الحدث A.

حيث أنه لدينا مجتمع يتكون من حدثين، لكل منهما احتمال نجاح p واحتمال فشل q، والتجربة تتكرر n مرة متماثلة. فإن المتغير العشوائي Xn يتبع توزيع ثنائي الحد مع الخصائص التالية:

$$E(Xn) = np, \quad V(Xn) = npq$$

وحيث أن:

$$X_n \backsim B(n,p) \text{ and } f = \frac{X_n}{n}$$

بالتالي:

$$E(f) = E\left(\frac{X_n}{n}\right) = \frac{E(X_n)}{n} = \frac{np}{n} = p$$

$$V(f) = V\left(\frac{X_n}{n}\right) = \frac{V(X_n)}{n^2} = \frac{npq}{n^2} = \frac{pq}{n}$$

مثال 2−2: تقوم مجموعة طلبة بتحقيق حول مجتمع طلبة علم الاجتماع بجامعة ما. العينة المستخرجة من هذا المجتمع ذات حجم n=135. نريد معرفة نسبة p الطلبة الذين زاولوا دراسات علمية في مستوى الثانوي. من أجل الإسراع في القيام بهذا التحقيق، انقسم الطلبة إلى مجموعتين:

– **المجموعة الأولى:** تحققت من 60 طالب، ووجدوا أن 24 منهم زاولوا دراسات علمية في مستوى الثانوي،

– **المجموعة الثانية:** تحققت من 75 طالب، ووجدوا أن 33 منهم زاولوا دراسات علمية في مستوى الثانوي.

المطلوب: أحسب 3 تقديرات لـ p.

الحل:

$$\hat{P}_1 = f_1 = \frac{24}{60} = 0.4$$

$$\hat{P}_2 = f_2 = \frac{33}{75} = 0.44$$

$$\hat{P}_3 = f_3 = \frac{57}{135} = 0.422$$

2-4 التقدير بمجال ثقة Confidence interval estimation

يعتبر التقدير بمجال ثقة، أحد أهم الأدوات الإحصائية وذات مصداقية كبيرة في الممارسة. وذلك لأنها تعطي قيم تقديرية لمعالم المجتمع من خلال حصرها ضمن مجال محدد ويكون ذلك باحتمال معين. حيث يعرف مجال ثقة بأنه: "المجال الذي تقع ضمنه القيمة الحقيقة لمعلمة ما. ويعبر عن مستوى الاعتقاد بالاحتمال الذي تكون لقيمته علاقة بحجم المجال".

من هذا التعريف، نستنتج أن التقدير بمجال ثقة الذي يرتبط بعينة عشوائية مسحوبة من مجتمع ما، هو مجال من الشكل $[\theta - c, \theta + d]$ الذي يشمل المعلمة الحقيقية θ. إن احتمال أن تقع قيمة المعلمة الحقيقية θ ضمن هذا المجال يقدر بالقيمة $1-\alpha$ التي تسمى درجة الثقة بينما قيمة α تسمى خطأ، ونكتب:

$$p(\theta - c < \theta < \theta + d) = 1 - \alpha$$

2-4-1 التقدير بمجال ثقة لمتوسط

إن تشكيل مجال لمتوسط مجتمع مجهول μ يتبع جملة من المعطيات هي: طبيعة المتغير العشوائي المرافق له X، حجم العينة n، طبيعة السحب (مع الإعادة أو بدون إعادة) ومعلومة تباين X في المجتمع σ^2.

2-4-1-1 التقدير بمجال ثقة لمتوسط باستخدام التوزيع الطبيعي

يمكن استخدام التوزيع الطبيعي لتقدير متوسط مجتمع مجهول في ثلاث حالات هي:

الحالة رقم 1: إذا كان

$\boldsymbol{n \geq 30, \sigma}$ معلوم

وذلك باستخدام العلاقة التالية:

- السحب مع الإعادة

أو المجتمع غير محدود:

- السحب بدون إعادة

و المجتمع محدود:

كما يمكن كتابة العلاقة السابقة بالصورة التالية:

- السحب مع الإعادة

أو المجتمع غير محدود:

- السحب بدون الإعادة

و المجتمع محدود:

ملاحظة: إذا كان $n \leq 0.05N$ فلا داعي لاستخدام معامل التصحيح.

بحيث:

α: الخطأ،

$1-\alpha$: درجة الثقة،

$Z_{1-\frac{\alpha}{2}}$: تستخرج من جدول التوزيع الطبيعي،

$Z_{1-\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right)$: خطأ التقدير.

حيث يمكن للباحث أن يستطيع معرفة تباين المجتمع الأصلي سواء بالرجوع إلى أبحاث سابقة في نفس المجتمع، أو بالرجوع إلى أبحاث سابقة على مجتمعات يعتقد أنها متشابهة مع هذا المجتمع.

مثال رقم 2-3: يبلغ عدد طلبة ميدان الاقتصاد في جامعة ما 1500 طالب. نسحب منهم عينة عشوائية حجمها 100 طالب، فنجد أن متوسط الوزن هو 70.215 كلغ. مع العلم أن الانحراف المعياري للمجتمع هو 10.521 كلغ.

المطلوب: أوجد تقدير لمتوسط وزن جميع الطلبة عند درجة ثقة قدرها 90%؟

الحل:

بما أن حجم العينة أكبر من 30، والانحراف المعياري للمجتمع معلوم، فإننا نستخدم التوزيع الطبيعي للتقدير :

$$p\left(70.215 - 1.65\left(\frac{10.521}{\sqrt{100}}\sqrt{\frac{1500-100}{1500-1}}\right) < \mu < 32 + 1.96\left(\frac{10.521}{\sqrt{100}}\sqrt{\frac{1500-100}{1500-1}}\right)\right) = 0.90$$

$$p(70.215 - 0.517 < \mu < 70.215 + 0.517) = 0.90$$

$$p(69.698 < \mu < 70.732) = 0.90$$

الحالة رقم 2: إذا كان

$\boldsymbol{n} \geq 30, \sigma$, مجهول, $\mathbf{s}$ معلوم

وذلك باستخدام العلاقات التالية.

- السحب مع الإعادة

أو المجتمع غير محدود:

- السحب بدون إعادة

و المجتمع محدود:

ملاحظات:

- إذا كان $n \leq 0.05N$ فلا داعي لاستخدام معامل التصحيح،

- الانحراف المعياري للعينة محسوب بالعلاقة التالية:

$$s = \sqrt{\frac{\sum(x_i - \bar{X})^2}{n-1}}$$

مثال رقم 2−4: دلت قياسات أقطار عينة عشوائية تتألف من 200 كرة تصنعها إحدى الآلات خلال أسبوع أن المتوسط هو 0.824 سم والانحراف المعياري هو 0.042 سم.

المطلوب: أوجد حدود الثقة 90%، 95% و99% لمتوسط قطر كل الكريات؟ قارن بين النتائج؟

الحل:

بما أن حجم العينة أكبر من 30، والانحراف المعياري للمجتمع مجهول، والانحراف المعياري للعينة معلوم، فإننا نستخدم التوزيع الطبيعي للتقدير:

- عند مستوى ثقة قدره 90%

$$p\left(0.824 - 1.65\left(\frac{0.042}{\sqrt{200}}\right) < \mu_1 < 0.824 + 1.65\left(\frac{0.042}{\sqrt{200}}\right)\right) = 0.90$$

$$p(0.824 - 0.0049 < \mu_1 < 0.824 + 0.0049) = 0.90$$

$$p(0.8289 < \mu_1 < 0.8191) = 0.90$$

- عند مستوى ثقة قدره 95%

$$p\left(0.824 - 1.96\left(\frac{0.042}{\sqrt{200}}\right) < \mu_2 < 0.824 + 1.96\left(\frac{0.042}{\sqrt{200}}\right)\right) = 0.95$$

$$p(0.824 - 0.0058 < \mu_2 < 0.824 + 0.0058) = 0.95$$

$$p(0.8298 < \mu_2 < 0.8182) = 0.95$$

- عند مستوى ثقة قدره 99%

$$p\left(0.824 - 2.58\left(\frac{0.042}{\sqrt{200}}\right) < \mu_3 < 0.824 + 2.58\left(\frac{0.042}{\sqrt{200}}\right)\right) = 0.99$$

$$p(0.824 - 0.0076 < \mu_3 < 0.824 + 0.0076) = 0.99$$

$$p(0.8316 < \mu_3 < 0.8164) = 0.99$$

مقارنة النتائج:

عند مقارنة القيم المقدرة لمتوسط المجتمع، نلاحظ أنه مع زيادة درجة الثقة المطلوبة، فإن طول مجال الثقة يزيد أيضا، ويصبح التقدير بمجال أكثر غموضا.

الحالة رقم 3: إذا كان

المجتمع يتبع التوزيع الطبيعي , معلوم $\boldsymbol{n} < 30, \sigma$

وذلك باستخدام العلاقات التالية:

- السحب مع الإعادة

أو المجتمع غير محدود:

- السحب بدون إعادة

والمجتمع محدود:

مثال رقم 2-5: سحبت عينة من 10 عمال من مجتمع كبير انحرافه المعياري 5.1. حيث وجدت أجورهم كما يلي:

الرقم	1	2	3	4	5	6	7	8	9	10
الأجر الشهري	32	40	25	27	42	30	32	34	28	30

المطلوب: تقدير متوسط أجر العمال بدرجة ثقة 95%.

الحل:

- نقوم بحساب متوسط للعينة:

$$\bar{X} = \frac{\sum x_i}{n} = \frac{32 + 40 + \cdots + 30}{10} = 32$$

- نقوم باستخراج قيمة $Z_{\text{الجدولية}}$ من جدول التوزيع الطبيعي:

$$Z_{1-\frac{\alpha}{2}} = Z_{1-\frac{0.05}{2}} = Z_{0.975} = 1.96$$

- حيث أن المجتمع كبير وغير محدود فإننا نقوم بعملية التقدير باستخدام العلاقة:

$$p\left(32-1.96(\frac{5.1}{\sqrt{10}})<\mu<32+1.96(\frac{5.1}{\sqrt{10}})\right)=0.95$$

$$p(32-3.161<\mu<32+3.161)=0.95$$

$$p(28.839<\mu<35.161)=0.95$$

مثال رقم 2−6: لنفرض أن المجتمع الذي سحبت منه العينة محدود ويساوي 100 عامل فقط.

المطلوب: تقدير متوسط أجر العمال بدرجة ثقة 95% .

الحل:

من حل المثال السابق لدينا:

- قيمة المتوسط الحسابي تساوي 32.
- قيمة $Z_{\text{الجدولية}}$ تساوي 1.96.
- حيث أن المجتمع محدود و $n \geq 0.05N$ فإننا نقوم بعملية التقدير باستخدام العلاقة التالية:

$$p\left(\bar{X}-Z_{1-\frac{\alpha}{2}}(\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}})<\mu<\bar{X}+Z_{1-\frac{\alpha}{2}}(\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}})\right)=1-\alpha$$

$$p\left(32-1.96(\frac{5.1}{\sqrt{10}}\sqrt{\frac{100-10}{100-1}})<\mu<32-1.96(\frac{5.1}{\sqrt{10}}\sqrt{\frac{100-10}{100-1}})\right)=1-\alpha$$

$$p(32-3.013<\mu<32+3.013)=0.95$$

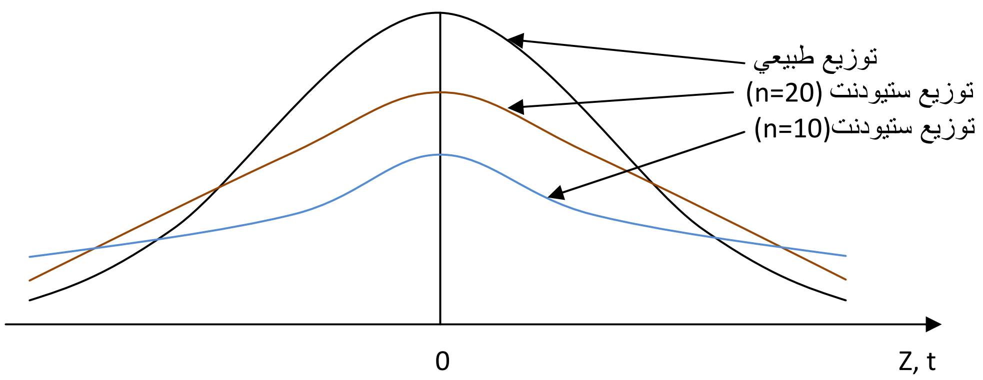
$$p(28.987<\mu<35.013)=0.95$$

2−1−4−2 التقدير بمجال ثقة لمتوسط باستخدام توزيع ستيودنت

في حالة العينات الصغيرة التي حجمها أقل من 30، فإن توزيع المعاينة للمتوسط يتبع توزيع المجتمع؛ إذا كان المجتمع يتبع التوزيع الطبيعي، فإن توزيع المعاينة للمتوسط يتبع التوزيع الطبيعي. ولتقدير متوسط مجتمع مجهول انطلاقا من بيانات عينة في حالة عدم معرفة قيمة تباين المجتمع، فإننا نستخدم توزيع ستيودنت Student.

يعتبر توزيع ستيودنت أو توزيع t أو توزيع العينات الصغيرة المقترح من طرف *William sealy Gosset*، توزيع احتمالي بدرجة حرية نرمز لها بـ *Df*. ويعتبر توزيع ستيودنت على غرار للتوزيع الطبيعي متماثل حول الوسط الحسابي. لكن، هناك توزيع طبيعي واحد، بينما توزيع ستيودنت يتبع حجم *n*. كلما كان حجم *n* صغير كلما كان توزيع ستيودنت مفلطح أكثر؛ أي أن معظم الاحتمال يكون عند الأطراف، ثم مع زيادة حجم *n* يزداد معها تدبب التوزيع، إلى غاية أن تتجاوز *n* قيمة 30، عندها يتساوى تقريبا توزيع ستيودنت مع التوزيع الطبيعي. وهذا ما يظهره الشكل رقم (2−1).

شكل رقم (2−1): مقارنة بين توزيع ستيودنت والتوزيع الطبيعي



ويوجد في الملحق توزيع ستيودنت عند الخطأ α ودرجة الحرية *df*. الذي نستخدمه لاستخراج قيمة ستيودنت الجدولية.

يمكن استخدام توزيع ستيودنت لتقدير متوسط مجتمع مجهول في الحالة التالية:

المجتمع يتبع التوزيع الطبيعي, معلوم s , مجهول σ ,$n < 30$

وذلك باستخدام العلاقات التالية:

− السحب مع الإعادة

أو المجتمع غير محدود:

− السحب بدون إعادة

والمجتمع محدود:

كما يمكن كتابة العلاقة السابقة بالصورة التالية:

– السحب مع الإعادة

أو المجتمع غير محدود:

– السحب بدون الإعادة

و المجتمع محدود:

ملاحظة: إذا كان $n \leq 0.05N$ فلا داعي لاستخدام معامل التصحيح.

بحيث:

α: الخطأ،

$n-1$: درجة الحرية،

$t_{\frac{\alpha}{n-1}}$: تستخرج من جدول ستيودنت،

$t_{\frac{\alpha}{n-1}} (\frac{s}{\sqrt{n}})$: خطأ التقدير.

ملاحظات:

– إذا كان $n \leq 0.05N$ فلا داعي لاستخدام معامل التصحيح،

– الانحراف المعياري للعينة محسوب بالعلاقة التالية:

$$s = \sqrt{\frac{\sum(x_i - \bar{X})^2}{n-1}}$$

مثال رقم 2–7: تظهر البيانات التالية المدة اللازمة بالأيام لإجراء تربص لعينة من عمال شركة.

العامل	1	2	3	4	5	6	7	8
الفترة	52	44	55	44	45	59	50	54
العامل	9	10	11	12	13	14	15	
الفترة	62	46	54	58	60	52	53	

المطلوب: أوجد تقدير بمجال لمتوسط المدة اللازمة لإجراء التربص عند درجة ثقة قدرها 95% مع العلم أنها تتبع التوزيع الطبيعي؟

الحل:

– حساب المتوسط الحسابي والانحراف المعياري للعينة:

$$\bar{X} = \frac{\sum X_i}{n} = \frac{808}{15} = 53.87$$

$$s = \sqrt{\frac{\sum(x_i - \bar{X})^2}{n-1}} = \sqrt{\frac{651.73}{14}} = 6.82$$

– حيث أن حجم العينة أقل من 30، الانحراف المعياري للمجتمع مجهول، الانحراف المعياري للعينة معلوم والمجتمع يتبع التوزيع الطبيعي، فإننا نستخدم توزيع ستيودنت للتقدير مع:

$$\alpha = 0.05, t_{\underset{n-1}{\alpha}} = t_{\underset{15-1}{0.05}} = 2.145$$

وذلك وفقا العلاقة التالية:

$$p\left(\bar{X} - t_{\underset{n-1}{\alpha}}(\frac{s}{\sqrt{n}}) < \mu < \bar{X} + t_{\underset{n-1}{\alpha}}(\frac{s}{\sqrt{n}})\right) = 1 - \alpha$$

$$p\left(53.87 - 2.145(\frac{6.82}{\sqrt{15}}) < \mu < \bar{X} + 2.145(\frac{6.82}{\sqrt{15}})\right) = 0.95$$

$$p(53.87 - 3.78 < \mu < 53.78 + 3.78) = 0.95$$

$$p(53.09 < \mu < \bar{X} + 57.56) = 0.95$$

2-4-1-3 التقدير بمجال ثقة لمتوسط باستخدام نظرية تشيبتشيف

تعتبر نظرية تشيبتشيف chebychev's ذات أهمية، لأنها تعطينا حد أدنى لنسبة البيانات الواقعة في مجال معين عند معرفة متوسطها وتباينها دون الحاجة إلى معرفة طبيعة التوزيع الذي تتبع له هذه البيانات.

وتنص هذه النظرية على أنه: إذا كان لدينا بيانات متوسطها $\bar{X}$ وانحرافها المعياري s، فإن المتباينة التالية تكون محققة:

$$p(|\bar{X}-\mu| \leq ks)) \geq 1-\frac{1}{k^2}$$

بالتالي، وإنه بصرف النظر عن شكل التوزيع، فإن نسبة البيانات (أو المساحة) التي لا تبعد عن المتوسط الحسابي بأكثر من ks فهي لا تقل عن القيمة $1-\frac{1}{k^2}$، مع $k \geq 1$.

تستخدم هذه النظرية في تقدير متوسط مجتمع في الحالة التالية:

المجتمع مجهول , s أو σ ,$\boldsymbol{n} < 30$

نضع $1-\alpha = 1-\frac{1}{k^2}$، بالتالي تصبح العلاقة السابقة بالشكل:

$$p\left(\bar{X}-k\frac{s}{\sqrt{n}} < \mu < \bar{X}-\frac{s}{\sqrt{n}}\right) = 1-\alpha$$

مثال رقم 2-8: حل المثال رقم (2-7) مع العلم أن المجتمع الذي أخذت منه العينة مجهول؟

الحل:

حيث أن حجم العينة أقل من 30، الانحراف المعياري للمجتمع مجهول، الانحراف المعياري للعينة معلوم والمجتمع مجهول، فإننا نستخدم نظرية تشيبتشيف للتقدير، بحيث نضع:

$$\alpha = 0.05,$$

$$1-\frac{1}{k^2} = 1-\alpha \Rightarrow 1-\frac{1}{k^2} = 0.95 \Rightarrow k = 4.47$$

وذلك وفقا للعلاقة التالية:

$$p\left(\bar{X}-k(\frac{s}{\sqrt{n}}) < \mu < \bar{X}+k(\frac{s}{\sqrt{n}})\right) = 1-\alpha$$

$$p\left(53.87 - 4.47(\frac{6.82}{\sqrt{15}}) < \mu < \bar{X} + 4.47(\frac{6.82}{\sqrt{15}})\right) = 0.95$$

$$p(53.87 - 7.87 < \mu < 53.78 + 7.87) = 0.95$$

$$p(46 < \mu < \bar{X} + 61.74) = 0.95$$

2-4-2 التقدير بمجال ثقة لنسبة

إذا كان الهدف هو تقدير نسبة ما في مجتمع يتبع توزيع ثنائي الحد $X \sim B(n,p)$، وذلك باستخدام أحد العينات التي يكون حجمها كبير $n \geq 30$، فإن التقدير يكون وفق العلاقة التالية:

- السحب مع الإعادة

أو المجتمع غير محدود:

- السحب بدون إعادة

والمجتمع محدود:

كما يمكن كتابة العلاقة السابقة بالصورة التالية:

- السحب مع الإعادة

أو المجتمع غير محدود:

- السحب بدون الإعادة

والمجتمع محدود:

ملاحظات:

- إذا كان $n \leq 0.05N$ فلا داعي لاستخدام معامل التصحيح،

- في العلاقات السابقة تعتبر قيمة p في خطأ التقدير $Z_{1-\frac{\alpha}{2}}\sqrt{\frac{pq}{n}}$ مجهولة، حيث أنها القيمة المراد تقديرها. لهذا، فإننا نستخدم الإحصائية $\bar{P}$ كمقدر لـ p من أجل حساب خطأ التقدير الذي يصبح من الشكل $Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{p}\bar{q}}{n}}$.

– تعتبر العلاقة $\sqrt{\frac{\bar{p}\bar{q}}{n-1}}$ تقدير غير متحيز لتباين توزيع المعاينة للنسب $\sigma_{\bar{X}}$، والتي ينبغي اعتمادها عند حساب خطأ التقدير، غير أنه ولأن حجم العينة كبير عادة في الحسابات المتعلقة بالنسب، فإن الفرق في النتائج عند استخدام قيمة n أو عند استخدام قيمة n−1 يتلاشى.

مثال رقم 2−9: في أحد الاستقصاءات المقدمة لقراء جريدة ما، كان احد الأسئلة "هل تقرأ أخبار السينما في الصحف اليومية". وجاءت الأجوبة كما يلي:

الإجابة	عدد المجيبين
نعم	150
لا	50
أحيانا	300

المطلوب: تقدير مجال الثقة لنسبة من يقرأ أخبار السينما بدرجة ثقة 95%؟

الحل:

– نسبة المجيبين بنعم هي:

$$\bar{p} = \frac{150}{500} = 0.3$$

– مجال الثقة يكتب بالشكل:

$$p\left(\bar{P} - Z_{1-\frac{\alpha}{2}}\sqrt{\frac{pq}{n}} < p < \bar{P} - Z_{1-\frac{\alpha}{2}}\sqrt{\frac{pq}{n}}\right) = 1 - \alpha$$

$$p\left(0.3 - 1.96\sqrt{\frac{0.3 \times 0.7}{1000}} < p < 0.3 + 1.96\sqrt{\frac{0.3 \times 0.7}{1000}}\right) = 0.95$$

$$p(0.3 - 0.028 < p < 0.3 + 0.028) = 0.95$$

$$p(0.272 < p < 0.328) = 0.95$$

2-4-3 التقدير بمجال ثقة للفرق أو المجموع بين متوسطين

2-4-3-1 حالة المجتمع يتبع التوزيع الطبيعي والتباين معلوم

لنفرض أن $x_1 = (x_1, x_2, \ldots x_{n1})$ و $x_2 = (x_1, x_2, \ldots x_{n2})$ عينتين مستقلتين بحيث:

$$X_1 \sim N\left(\mu_{\bar{X}}, \sigma^2_{\bar{X}_1}\right), \qquad X_2 \sim N\left(\mu_{\bar{X}}, \sigma^2_{\bar{X}_2}\right)$$

نهتم أولا بتقدير الفرق بين متوسطي المجتمعين $\mu_1 - \mu_2$ الذي يمكن صياغته بالعلاقة التالية:

$$p\left((\bar{X}_1 - \bar{X}_2) - Z_{1-\frac{\alpha}{2}}\left(\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right) < \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + Z_{1-\frac{\alpha}{2}}\left(\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)\right) = 1 - \alpha$$

كما يمكن تقدير المجموع بين متوسطي المجتمعين $\mu_1 + \mu_2$ بالعلاقة التالية:

$$p\left((\bar{X}_1 + \bar{X}_2) - Z_{1-\frac{\alpha}{2}}\left(\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right) < \mu_1 + \mu_2 < (\bar{X}_1 + \bar{X}_2) + Z_{1-\frac{\alpha}{2}}\left(\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)\right) = 1 - \alpha$$

ملاحظة: نستخدم معامل التصحيح إذا دعت الضرورة إلى ذلك وفق القواعد السابقة.

مثال رقم 2-10: يبلغ متوسط الطاقة المحركة الكهربائية لبطاريات تنتجها إحدى الشركات 1.5 فولط، فإذا وصلنا أربعا من هذه البطاريات على التسلسل.

المطلوب: أوجد مجال الثقة 95% للطاقة المحركة الكهربائية الكلية، مع العلم أن الانحراف المعياري للمجتمع يساوي 0.004 فولط؟

الحل:

من المعلوم أن وصل البطاريات على التسلسل يتولد عنه جمع الطاقة المحركة للبطاريات المجموعة. هكذا، نحن لدينا أربع عينات متشابهة، كل عينة حجمها 1 ومتوسطها يساوي 1.5 فولط.

لإيجاد مجال الثقة لمجموع متوسط المجتمع نعتمد على العلاقة:

$$p\left((\bar{X}_1+\bar{X}_2+\bar{X}_3+\bar{X}_4)-Z_{1-\frac{\alpha}{2}}\left(\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}+\frac{\sigma_3^2}{n_3}+\frac{\sigma_4^2}{n_4}}\right)<\mu_1+\mu_2 <(\bar{X}_1+\bar{X}_2+\bar{X}_3+\bar{X}_4)+Z_{1-\frac{\alpha}{2}}\left(\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}+\frac{\sigma_3^2}{n_3}+\frac{\sigma_4^2}{n_4}}\right)\right)=1-\alpha$$

$$p\left((1.5+1.5+1.5+1.5)-1.96\left(\sqrt{\frac{0.004}{1}+\frac{0.004}{1}+\frac{0.004}{1}+\frac{0.004}{1}}\right)<\mu_1+\mu_2 <(1.5+1.5+1.5+1.5)+1.96\left(\sqrt{\frac{0.004}{1}+\frac{0.004}{1}+\frac{0.004}{1}+\frac{0.004}{1}}\right)\right) =0.95$$

$$p\big(6-1.96(0.126)<\mu_1+\mu_2<6-1.96(0.126)\big)=0.95$$

$$p(6-0.247<\mu_1+\mu_2<6-0.247)=0.95$$

$$p(5.753<\mu_1+\mu_2<6.247)=0.95$$

2-3-4-2 حالة المجتمع يتبع التوزيع الطبيعي والتباين مجهول

لنفرض أن $x_1=(x_1,x_2,\ldots x_{n1})$ و $x_2=(x_1,x_2,\ldots x_{n2})$ عينتين مستقلتين بحيث:

$$X_1 \sim N\left(\mu_{\bar{X}},\sigma_{\bar{X}_1}^2\right),\qquad X_2 \sim N\left(\mu_{\bar{X}},\sigma_{\bar{X}_2}^2\right)$$

لتقدير الفرق بين متوسطي المجتمعين $\mu_1-\mu_2$ فإننا نستخدم العلاقة التالية:

$$p\left((\bar{X}_1-\bar{X}_2)-t_{\frac{\alpha}{df}}\left(\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}\right)<\mu_1-\mu_2<(\bar{X}_1-\bar{X}_2)+t_{\frac{\alpha}{df}}\left(\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}\right)\right)=1-\alpha$$

كما يمكن تقدير المجموع بين متوسطي المجتمعين $\mu_1+\mu_2$ بالعلاقة التالية:

$$p\left((\bar{X}_1+\bar{X}_2)-t_{\frac{\alpha}{df}}\left(\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}\right)<\mu_1-\mu_2<(\bar{X}_1+\bar{X}_2)+t_{\frac{\alpha}{df}}\left(\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}\right)\right)=1-\alpha$$

وتعطى درجات حرية لتوزيع ستيودنت بالعلاقة التالية:

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{s_2^2}{n_2}\right)^2}$$

حالة خاصة:

نهتم بتقدير الفرق بين متوسطي المجتمعين $\mu_1 - \mu_2$ بحيث نفرض أن:

$$\sigma_{\bar{X}_1}^2 = \sigma_{\bar{X}_2}^2 = \sigma^2$$

بحيث تعتبر قيمة الانحراف المعياري للمجتمع σ^2 مجهولة أيضا.

نعرف الانحراف المعياري s غير المتحيز الذي يساوي:

$$s = \sqrt{\frac{n_1 - 1}{n_1 + n_2 - 2} s_1^2 + \frac{n_2 - 1}{n_1 + n_2 - 2} s_2^2}$$

ومنه فإن تقدير الفرق بين متوسطي مجتمعين يكتب الصيغة التالية:

$$p\left((\bar{X}_1 - \bar{X}_2) - t_{\frac{\alpha}{n_1+n_2-2}} s\left(\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right) < \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + t_{\frac{\alpha}{n_1+n_2-2}} s\left(\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right)\right) = 1 - \alpha$$

كما يمكن تقدير المجموع بين متوسطي المجتمعين $\mu_1 + \mu_2$ بالعلاقة التالية:

$$p\left((\bar{X}_1 + \bar{X}_2) - t_{\frac{\alpha}{n_1+n_2-2}} s\left(\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right) < \mu_1 + \mu_2 < (\bar{X}_1 + \bar{X}_2) + t_{\frac{\alpha}{n_1+n_2-2}} s\left(\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right)\right) = 1 - \alpha$$

الصعوبة التي تواجه هنا هي فرضية تساوي التباين، والتي هي صعبة التحقق، خصوصا عند اختلاف حجم العينات المستخرجة.

مثال رقم 2–11: بغرض قياس قيمة زاوية A تم إجراء تجربتين مستقلتين.

التجربة الأولى أعطت القياسات التالية: 21.76 ، 20.98 درجة.

التجربة الثانية أعطت القياسات التالية: 21.64 ، 21.54، 22.32، 20.56، 21.43، 21.07 درجة.

نفرض أن جميع القياسات هي مستقلة وتتبع التوزيع الطبيعي بتباين متساوي.

المطلوب: أوجد تقدير بمجال ثقة للفرق بين قياسي التجربتين عند مستوى خطأ قدره 1%؟

الحل:

نفرض أن قياسات التجربة الأولى من الشكل:

$$X_1 = (X_1, X_2), \qquad X_1 \sim N\left(\mu_{\bar{X}}, \sigma^2_{\bar{X}_1}\right),$$

نفرض أن قياسات التجربة الثانية من الشكل:

$$X_2 = (X_1, \dots, X_6), \qquad X_2 \sim N\left(\mu_{\bar{X}}, \sigma^2_{\bar{X}_1}\right),$$

– حساب المتوسط الحسابي والانحراف المعياري للعينتين:

$$\bar{X}_1 = \frac{\sum X_i}{n} = \frac{42.74}{2} = 21.37$$

$$s_1 = \sqrt{\frac{\sum (x_i - \bar{X})^2}{n-1}} = \sqrt{\frac{0.3121}{1}} = 0.558$$

$$\bar{X}_2 = \frac{\sum X_i}{n} = \frac{128.56}{6} = 21.42$$

$$s_2 = \sqrt{\frac{\sum(x_i - \bar{X})^2}{n-1}} = \sqrt{\frac{1.735}{5}} = 0.589$$

– في هذه الحالة فإن الانحراف المعياري للفرق يكتب بالشكل:

$$s = \sqrt{\frac{n_1 - 1}{n_1 + n_2 - 2} s_1^2 + \frac{n_2 - 1}{n_1 + n_2 - 2} s_2^2}$$

$$s = \sqrt{\frac{2-1}{2+6-2}(0.558)^2 + \frac{6-1}{2+6-2}(0.589)^2} = 0.583$$

– تقدير الفرق بين متوسطي المجتمعين يعطى بالشكل:

$$p\left((\bar{X}_1 - \bar{X}_2) - t_{\frac{\alpha}{n_1+n_2-2}} s\left(\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right) < \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + t_{\frac{\alpha}{n_1+n_2-2}} s\left(\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right)\right) = 1 - \alpha$$

$$p\left((21.37 - 21.42) - 3.707(0.583)\left(\sqrt{\frac{1}{2} + \frac{1}{6}}\right) < \mu_1 - \mu_2 < (21.37 - 21.42) + 3.707(0.583)\left(\sqrt{\frac{1}{2} + \frac{1}{6}}\right)\right) = 0.99$$

$$p(-0.05 - 1.76 < \mu_1 - \mu_2 < -0.05 + 1.76) = 0.99$$

$$p(-1.81 < \mu_1 - \mu_2 < 1.71) = 0.99$$

2–4–4 التقدير بمجال ثقة للفرق أو المجموع بين نسبتين

لتكن لدينا عينتين كبيرتين لمجتمع يتبع توزيع ثنائي الحد.

يمكن إعطاء تقدير للفرق بين نسبتي مجتمعين بالعلاقة التالية:

$$p\left((\bar{P}_1 - \bar{P}_2) - Z_{1-\frac{\alpha}{2}}\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}} < p_1 - p_2 < (\bar{P}_1 - \bar{P}_2) + Z_{1-\frac{\alpha}{2}}\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}\right) = 1 - \alpha$$

كما يمكن إعطاء تقدير للمجموع بين نسبتي مجتمعين بالعلاقة التالية:

$$p\left((\bar{P}_1+\bar{P}_2)-Z_{1-\frac{\alpha}{2}}\sqrt{\frac{p_1q_1}{n_1}+\frac{p_2q_2}{n_2}}<p_1+p_2<(\bar{P}_1+\bar{P}_2)+Z_{1-\frac{\alpha}{2}}\sqrt{\frac{p_1q_1}{n_1}+\frac{p_2q_2}{n_2}}\right)=1-\alpha$$

ملاحظة: نستخدم معامل التصحيح إذا دعت الضرورة إلى ذلك وفق القواعد السابقة.

مثال رقم 2−12: في عينة عشوائية مؤلفة من 400 بالغ و600 مراهق يشاهدون برنامجا تلفزيونيا معينا، أشار 100 بالغ و300 مراهق بأنهم يحبون هذا البرنامج.

المطلوب: أوجد حدود الثقة 95% لفرق تناسبات كل المراهقين والبالغين الذين شاهدوا البرنامج وأحبوه؟

الحل:

نرمز بـ n_1 لعدد عناصر عينة المراهقين، وبـ n_2 لعدد عناصر عينة البالغين.

وبالتالي فإن نسبة المراهقين والبالغين الذين شاهدوا البرنامج وأحبوه في العينتين هي:

$$\bar{p}_1=\frac{300}{600}=0.5,\ \ \bar{p}_1=\frac{100}{400}=0.25$$

تقدير الفرق بين نسبتي كل المراهقين والبالغين الذين شاهدو البرنامج وأحبوه يعطى بالعلاقة التالية:

$$p\left((\bar{P}_1-\bar{P}_2)-Z_{1-\frac{\alpha}{2}}\sqrt{\frac{p_1q_1}{n_1}+\frac{p_2q_2}{n_2}}<p_1-p_2<(\bar{P}_1-\bar{P}_2)+Z_{1-\frac{\alpha}{2}}\sqrt{\frac{p_1q_1}{n_1}+\frac{p_2q_2}{n_2}}\right)=1-\alpha$$

$$p\left((0.5-0.25)-1.96\sqrt{\frac{0.5\times0.5}{600}+\frac{0.25\times0.75}{400}}<p_1-p_2\right.$$

$$\left.<(0.5-0.25)+1.96\sqrt{\frac{0.5\times0.5}{600}+\frac{0.25\times0.75}{400}}\right)=0.95$$

$$p(0.25-0.058<p_1-p_2<0.25-0.058)=0.95$$

$$p(0.192<p_1-p_2<0.308)=0.95$$

2-4-5 التقدير بمجال ثقة لتباين مجتمع

لقد تعرضنا فيما سبق لكيفية، تحديد مجالات ثقة لكل من المتوسط والنسبة، أما في هذا الجزء، نرغب الآن في تقدير مجال ثقة لتباين مجتمع يتبع التوزيع الطبيعي. وكما هو معلوم، فإن للتباين أهمية كبيرة، حيث أنه يحدد لنا مجال انتشار البيانات حول وسطها الحسابي. ولا يخفى أن هذه المعلومة الإحصائية تحدد لنا طبيعة التوزيع الإحصائي. أما في الحياة العملية، فإنه يحدد لنا خصوصا مدى جودة المخرجات من حيث تشابهها ودرجة الاختلافات الموجودة بينها.

ولتحديد فترة ثقة لتباين مجتمع أو الانحراف المعياري للمجتمع، فإننا بحاجة إلى توزيع جديد يسمى توزيع χ^2 chi-square (χ هي حرف لاتيني ينطق كاي). هذا التوزيع على غرار توزيع ستيودنت ينتمي إلى عائلة التوزيعات التي تعتمد على درجات الحرية df. ويعتبر Hershel سنة 1869 أول من اكتشف هذا التوزيع بدرجة حرية قدرها 2 في دراسته حول دقة إطلاق السهام على هدف. وقد ساهم العديد من الرياضيين منذ ذاك الحين في تطويره.

ويمكن كتابة مجال ثقة لتباين مجتمع يتبع التوزيع الطبيعي بالشكل التالي:

$$p\left(\frac{(n-1)s^2}{\chi^2_{\substack{1-\frac{\alpha}{2}\\ n-1}}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{\substack{\frac{\alpha}{2}\\ n-1}}}\right) = 1-\alpha$$

بحيث:

s^2: تمثل قيمة تباين العينة محسوب بالصيغة غير المتحيزة $s^2 = \frac{\sum(x_i-\bar{x})^2}{n-1}$،

χ^2: تستخرج من توزيع كاي مربع بدرجة حرية n−1.

مثال رقم 2-13: تبين من إحدى الدراسات على عينة تحوي 20 سيجارة، أن الانحراف المعياري لكمية النيكوتين فيها هو 1.6 مليغرام، مع العلم أنها تتبع التوزيع الطبيعي.

المطلوب: أوجد تقدير لتباين المجتمع والانحراف المعياري للمجتمع بدرجة ثقة قدرها 95%؟

الحل:

مجال ثقة لتباين مجتمع يعطى بالشكل التالي:

$$p\left(\frac{(n-1)s^2}{\chi^2_{\substack{1-\frac{\alpha}{2}\\ n-1}}} < \sigma^2 < \frac{(n-1)s^2}{{\chi^2}_{\substack{\frac{\alpha}{2}\\ n-1}}}\right) = 1-\alpha$$

$$p\left(\frac{(20-1)1.6^2}{\chi^2_{\substack{1-\frac{0.05}{2}\\ 20-1}}} < \sigma^2 < \frac{(20-1)1.6^2}{{\chi^2}_{\substack{\frac{0.05}{2}\\ 20-1}}}\right) = 1-0.05$$

$$p\left(\frac{(20-1)1.6^2}{32.852} < \sigma^2 < \frac{(20-1)1.6^2}{8.907}\right) = 0.907$$

$$p(1.5 < \sigma^2 < 5.5) = 0.95$$

أما مجال ثقة للانحراف المعياري للمجتمع فهو:

$$p\left(\sqrt{1.5} < \sigma < \sqrt{5.5}\right) = 0.95$$

$$p(1.2 < \sigma^2 < 2.3) = 0.95$$

2-5 خواص المقدر

يعتبر المقدر جيد إذا كان يتصف بالخصائص التالية:

2-5-1 مقدر غير متحيز

يعرف التحيز الذي نرمز له بـ $B(\hat{\theta})$، بأنه متوسط الفرق بين القيمة المقدرة $\hat{\theta}$ والقيمة الحقيقة للمعلمة المناظرة لها في المجتمع θ.

ولكي يكون المقدر غير متحيز يجب أن يكون مقدار التحيز مساويا للصفر، أي:

$$B(\hat{\theta}) = E(\hat{\theta}-\theta) = E(\hat{\theta}) - E(\theta) = E(\hat{\theta}) - \theta = 0 \Rightarrow E(\hat{\theta}) = \theta$$

بالتالي، فالمقدر يكون غير متحيز إذا كان أمله الرياضي مساويا لقيمة معلمة المجتمع.

2-5-2 مقدر ذو أقل تباين

يعطى تباين المقدر $\hat{\theta}$ بالعلاقة التالية:

$$V(\hat{\theta}) = E(\hat{\theta} - E(\hat{\theta}))^2$$
$$= E(\hat{\theta} - \theta)^2$$
$$= E(\hat{\theta} - E(\hat{\theta}) + E(\hat{\theta}) - \theta)^2$$
$$= E[\hat{\theta} - E(\hat{\theta})]^2 + E[(\hat{\theta}) - \theta)]^2$$
$$= \sigma^2(\hat{\theta}) + B^2(\hat{\theta})$$

وذلك لأن:

$$E\{[\hat{\theta} - E(\hat{\theta})][E(\hat{\theta}) - \theta]\} = E[\hat{\theta} - E(\hat{\theta})][E(\hat{\theta}) - \theta] = 0$$

من العلاقة السابقة، يتضح لنا بأن التغيرات في الخطأ الإجمالي للمقدر $\hat{\theta}$ تتبع الخطأ العشوائي (التباين المقاس) $\sigma^2(\hat{\theta})$ المتعلق بهذا المقدر، وإلى مقدار التحيز $B^2(\hat{\theta})$.

على أساس هذه القاعدة، فإن أفضل مقدر هو الذي يكون ذو الخطأ الإجمالي الأقل. عندما يكون لدينا عينة، والتي نحسب على أساسها مقدرين لهما نفس التحيز، فإن الخطأ الإجمالي الأقل هو الذي يسمح لنا باختيار المقدر المرغوب.

2-5-3 مقدر متقارب

لكي يكون المقدر $\hat{\theta}$ متقارب يجب أن تتجه قيمه المقدرة إلى قيمة المعلمة الحقيقة مع زيادة حجم المشاهدات في العينة المسحوبة. ويعتبر المقدر $\hat{\theta}$ متقارب احتماليا إذا كان يحقق الشرط التالي:

$$\forall \varepsilon > 0, p(|\hat{\theta} - \theta| < \epsilon \; if \; n \to \infty$$

على هذا الأساس، لكي يكون المقدر متقارب $\hat{\theta}$ يجب أن يتحقق الشرطان التاليان:

- أن يكون غير متحيز $E(\hat{\theta}) = \theta$

- تباين المقدر يتجه نحو الصفر عندما تزداد المشاهدات إلى المالانهاية $\lim_{n\to\infty}(V(\hat{\theta})) \to 0$

4-5-2 مقدر كفء

ليكن $\hat{\theta}_1$ و $\hat{\theta}_2$ مقدران غير متحيزان ومتقاربان إلى نفس المعلمة θ؛ يعتبر المقدر كفء المقدر ذو التباين الأصغر بينهما، أي:

$$V(\hat{\theta}) = E[\hat{\theta} - E(\hat{\theta})]^2 \text{minimum}$$

متراجحة Cramer-Rao وكفاءة مقدر

نقول عن مقدر غير متحيز أنه كفء، إذا لم يكن هناك مقدر غير متحيز آخر له تباين أقل منه.

ولتحديد كفاءة مقدر نستخدم متراجحة Cramer-Rao التي تنص على:

من بين مجموعة المقدرات غير المتحيزة $\hat{\theta}$ للمعلمة θ، فإن تباين مقدر يجب أن يحقق المتراجحة:

$$\forall \hat{\theta}_n \in \hat{\theta},\ \ V(\hat{\theta}_n) \geq \frac{1}{I_n(\theta)} \qquad (1)$$

بحيث:

$\frac{1}{I_n(\theta)}$: تمثل الحد الأدنى لـ Cramer-Rao

$I_n(\theta)$: كمية المعلومات التي توفرها n عينة للمتغير العشوائي X، وتسمى كذلك كمية معلومات فيشر، وتحدد بالعلاقة التالية:

$$I_n(\theta) = nE[lnf(X,\theta)'_{\theta}]^2$$

$$or \qquad I_n(\theta) = nE\left[\frac{\partial lnf(X,\theta)}{\partial \theta}\right]^2 \qquad (2)$$

بحيث:

$f(X,\theta)$: دالة الكثافة الاحتمالية للمتغير العشوائي X، و θ معلمة مراد تقديرها،

بتعويض (2) في (1) نجد:

$$\forall \hat{\theta}_n \in \hat{\theta},\ \ V(\hat{\theta}_n) \geq \frac{1}{nE\left[lnf(X,\theta)'_{\theta}\right]^2}$$

2-6 تمارين محلولة

تمرين رقم 2-1: عينة تتكون من 532 مشترك في مجلة Business Week تشير إلى أن متوسط تصفحهم الانترنت أسبوعيا هو 6.7 ساعة. إذا كان الانحراف المعياري لهذه العينة هو 5.8 ساعة.

المطلوب: أوجد مجال الثقة 95% لمتوسط تصفح الانترنت من طرف مشتركي هذه المجلة؟

الحل:

المعطيات:

$$n = 532, \qquad \bar{X} = 6.7, \qquad s = 5.8, \qquad \alpha = 0.05$$

في ظل هذه المعطيات فإننا نستخدم التوزيع الطبيعي للتقدير بالصيغة التالية:

$$p\left(\bar{X} - Z_{1-\frac{\alpha}{2}}(\frac{s}{\sqrt{n}}) < \mu < \bar{X} + Z_{1-\frac{\alpha}{2}}(\frac{s}{\sqrt{n}})\right) = 1 - \alpha$$

$$p\left(6.7 - 1.96\left(\frac{5.8}{\sqrt{532}}\right) < \mu < 6.7 + 1.96(\frac{5.8}{\sqrt{532}})\right) = 0.95$$

$$p(6.2 < \mu < 7.2) = 0.95$$

تمرين رقم 2-2: في إحدى نشرات بورصة نيويورك. قدمت عينة تتألف من 10 أسهم، معلومات حول حول ربح كل سهم التي هي: 5، 7، 9، 10، 14، 23، 20، 15، 3، 26.

المطلوب:

1- أوجد مقدر نقطي لمتوسط ربح جميع الأسهم المدرجة في بورصة نيويورك؟

2- أوجد مقدر نقطي للانحراف المعياري لربح جميع الأسهم المدرجة في بورصة نيويورك؟

3- ما هو مجال الثقة 95% لمتوسط ربح جميع الأسهم المدرجة في بورصة نيويورك، مع افتراض أن المجتمع يتبع التوزيع الطبيعي؟

4- علق على نتائج السؤال 3؟

5- نرغب في تقدير بمجال لمتوسط ربح جميع الأسهم المدرجة في بورصة نيويورك؟ كم هو عدد الأسهم التي تسمح لنا بإيجاد خطأ تقدير قدره 2 عند درجة خطأ قدرها 5%.

الحل:

1- حساب المتوسط الحسابي للعينة:

$$\bar{X} = \frac{\sum X_i}{n} = \frac{132}{10} = 13.2$$

وبالتالي فإن المقدر النقطي لمتوسط المجتمع يساوي:

$$\hat{\mu} = \bar{X} = 13.2$$

2- حساب الانحراف المعياري للعينة

$$s = \sqrt{\frac{\sum(x_i - \bar{X})^2}{n}} = \sqrt{\frac{547.6}{10}} = 7.4$$

وبالتالي فإن المقدر النقطي للانحراف المعياري للمجتمع يساوي:

$$\hat{\sigma} = S\sqrt{\frac{10}{10-1}} = 7.8$$

يمكن مباشرة إيجاد النتيجة السابقة على طريق استخدام علاقة الانحراف المعياري

$$\hat{\sigma} = s = \sqrt{\frac{\sum(x_i - \bar{X})^2}{n-1}} = \sqrt{\frac{547.6}{10-1}} = 7.8$$

3- نستخدم توزيع ستيودنت للتقدير نفق العلاقة التالية:

$$p\left(\bar{X} - t_{\substack{\alpha \\ n-1}}\left(\frac{s}{\sqrt{n}}\right) < \mu < \bar{X} + t_{\substack{\alpha \\ n-1}}\left(\frac{s}{\sqrt{n}}\right)\right) = 1 - \alpha$$

$$p\left(13.2 - 2.262\left(\frac{7.8}{\sqrt{10}}\right) < \mu < 13.2 - 2.262\left(\frac{7.8}{\sqrt{10}}\right)\right) = 0.95$$

$$p(7.62 < \mu < 18.78) = 0.95$$

4- التعليق: نلاحظ أن مجال الثقة لمتوسط أرباح الأسهم في بورصة نيويورك يعتبر واسع. لهذا يفضل استخدام عينة أكبر.

5- حساب عدد المشاهدات على طريق العلاقة التالية:

$$Z_{1-\frac{\alpha}{2}}\left(\frac{s}{\sqrt{n}}\right)=2 \Rightarrow n=\left(\frac{Z_{1-\frac{\alpha}{2}}s}{2}\right)^2=\left(\frac{1.96\times 7.8}{2}\right)^2=59$$

تمرين رقم 2-3: يعتبر الوقت والتكلفة من أهم محددات اختيار شركات الطيران من طرف عملاء درجات الأعمال. في دراسة لمجلة USA today، لعينة مؤلفة من 1993 شخص، أجابوا على الاستبيان، أشار 618 منهم إلى أن برنامج الرحلات يعتبر أهم محدد لاختيار شركة الطيران.

المطلوب:

1- ما هو التقدير النقطي لنسبة المجتمع الذين يعتبرون برنامج الرحلات كأهم محدد لاختيار شركة الطيران؟

2- أوجد مجال ثقة 95% لنسبة المجتمع؟

3- كم يجب أن يكون حجم العينة، لكي نحصل على خطأ تقدير قدره 0.01 بدرجة ثقة قدرها 95%؟ هل تنصح مجلة USA today باعتمادها؟ برر إجابتك؟

الحل:

1- يمكن إعطاء تقدير نقطي لنسبة المجتمع كالتالي:

$$p=\hat{p}=\frac{618}{1993}=0.3101$$

2- إيجاد مجال الثقة

$$p\left(\bar{P}-Z_{1-\frac{\alpha}{2}}\sqrt{\frac{pq}{n}}<p<\bar{P}+Z_{1-\frac{\alpha}{2}}\sqrt{\frac{pq}{n}}\right)=1-\alpha$$

$$p\left(0.3101 - 1.96\sqrt{\frac{0.3101 \times 0.6899}{1993}} < p < 0.3101 + 1.96\sqrt{\frac{0.3101 \times 0.6899}{1993}}\right) = 0.95$$

$$p(0.2898 < p < 0.3304) = 0.95$$

3- إيجاد حجم العينة n

$$Z_{1-\frac{\alpha}{2}}\sqrt{\frac{pq}{n}} = 0.01 \Rightarrow n = \frac{\left(Z_{1-\frac{\alpha}{2}}\right)^2 pq}{(0.01)^2} \Rightarrow n = \frac{(1.96)^2 \times 0.3101 \times 0.6899}{0.0001} = 8219$$

حجم العينة المطلوب للحصول على خطأ تقدير يساوي 0.01 هو 8219. لا ننصح الشركة باعتمادها. لأنه كبير جدا، ويتطلب الكثير من الوقت والتكلفة.

تمرين رقم 2-4: بينت عينة عشوائية تتألف من 50 علامة رياضيات من أصل 200 علامة، بأن المتوسط هو 75 و الانحراف المعياري هو 10.

المطلوب:

1 - ما هي حدود الثقة 95% لتقدير متوسط 200 علامة؟

2 - بأي درجة من الثقة نستطيع أن نقول بأن متوسط الـ 200 علامة هو 75 ± 1؟

الحل:

1- المعطيات:

$$n = 50, \qquad N = 200, \qquad \bar{X} = 75, \qquad s = 10, \qquad \alpha = 0.05$$

في ظل هذه المعطيات فإننا نستخدم التوزيع الطبيعي للتقدير، وحيث أن المجتمع محدود فإننا نستخدم معامل التصحيح: الصيغة التالية:

$$p\left(\bar{X} - Z_{1-\frac{\alpha}{2}}\left(\frac{s}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}\right) < \mu < \bar{X} + Z_{1-\frac{\alpha}{2}}\left(\frac{s}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}\right)\right) = 1 - \alpha$$

$$p\left(75-1.96\left(\frac{10}{\sqrt{50}}\sqrt{\frac{200-50}{200-1}}\right)<\mu<75-1.96\left(\frac{10}{\sqrt{50}}\sqrt{\frac{200-50}{200-1}}\right)\right)=0.95$$

$$p(72.6<\mu<77.4)=0.95$$

2- حيث أن خطأ التقدير يساوي 1 فإن:

$$Z_{1-\frac{\alpha}{2}}\left(\frac{s}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}\right)=1\Rightarrow Z_{1-\frac{\alpha}{2}}=\frac{1}{\left(\frac{s}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}\right)}=0.81$$

من جدول التوزيع الطبيعي نستخرج الاحتمال المقابل للقيمة المعيارية 0.81 والموجودة في نصف الجدول والتي هي: 0.7910

إذن لاستخراج درجة الثقة نستخدم العادلة التالية:

$$1-\frac{\alpha}{2}=0.7910\Rightarrow 1-\alpha=0.582$$

وبالتالي فإن درجة الثقة التي تعطي لنا خطا تقدير يساوي 1 هي 58.2%.

تمرين رقم 2-5: ضمن خطتها لإصلاح حركة المرور في المدينة، قامت بلدية قالمة بإجراء مسح ميداني لتحديد حجم الحركة عبر تقاطع رئيسي خلال فترة الصباح. باختيار يوم الأحد لثمانية وثلاثون أسبوعا متتالية، تم عد المركبات التي تمر عبر التقاطع بين الساعة 7:00 و9:00 صباحا، ووجد أن متوسط عدد المركبات للعينة يساوي 1500 سيارة، والانحراف المعياري للعينة يساوي 300 سيارة.

المطلوب: حساب فترة الثقة 99% لمتوسط عدد المركبات في المجتمع؟

الحل:

لأن n أكبر من 30 فإننا نقدر باستخدام التوزيع الطبيعي

$$\mu=\bar{x}\pm Z_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}}=1500\pm 2.58\frac{300}{\sqrt{38}}=1500\pm 125.55$$

تمرين رقم 2-6: في إحدى دراسات بنك Clearview عرض عينتين عشوائيتين مستقلتين كما هو موضح في الجدول التالي:

الوكالة البنكية	عدد الحسابات	متوسط رصيد العينة	الانحراف المعياري لرصيد العينة
Cherry grove	12	1000	150
beechmont	10	920	120

المطلوب:

1- أوجد مجال الثقة 90% للفرق بين متوسط رصيد الحساب بين الوكالتين؟ مع افتراض أن متوسط رصيد حساب العينتين يتبع التوزيع الطبيعي بتباين متساوي.

2- علق على النتيجة؟

الحل:

لإيجاد مجال الثقة نستخدم الانحراف المعياري

$$s = \sqrt{\frac{n_1 - 1}{n_1 + n_2 - 2} s_1^2 + \frac{n_2 - 1}{n_1 + n_2 - 2} s_2^2}$$

$$s = \sqrt{\frac{12 - 1}{12 + 10 - 2}(150)^2 + \frac{10 - 1}{12 + 10 - 2}(120)^2} = 137.31$$

ومنه فإن تقدير الفرق بين متوسطي مجتمعين يكتب الصيغة التالية:

$$p\left((\bar{X}_1 - \bar{X}_2) - t_{\substack{\alpha \\ n_1+n_2-2}} s\left(\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right) < \mu_1 - \mu_2 \right.$$
$$\left. < (\bar{X}_1 - \bar{X}_2) + t_{\substack{\alpha \\ n_1+n_2-2}} s\left(\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right)\right) = 1 - \alpha$$

$$p\left((1000-920)-1.725(137.31)\left(\sqrt{\frac{1}{12}+\frac{1}{10}}\right)<\mu_1-\mu_2\right.$$

$$\left.<(1000-920)+1.725(137.31)\left(\sqrt{\frac{1}{12}+\frac{1}{10}}\right)\right)=0.90$$

$$p(80-101.42<\mu_1-\mu_2<80+101.42)=0.90$$

$$p(80-101.42<\mu_1-\mu_2<80+101.42)=0.90$$

$$p(-21.42<\mu_1-\mu_2<181.42)=0.90$$

2- التعليق: احتمال أن يكون الفرق بين متوسطي رصيدي حسابي الوكالتين محصور بين 21.42- و181.42 هو 0.90. حيث أن المجال يحتوي على قيم سالبة، فهذا يشير إلى إمكانية أن يكون الفرق الحقيقي بين متوسطي المجتمعين سالب، ما يعني أن قيمة μ_2 يمكن أن تكون أكبر من μ_1. أي أن رصيد الحسابات في وكالة beechmont أكبر من رصيد الحسابات في وكالة Cherry grove. كذلك، احتواء مجال الثقة على القيمة 0، يعني أنه يمكن أن لا يكون هناك أي فرق في رصيد الحسابات بين الوكالتين.

تمرين رقم 2-7: إذا كانت $\bar{p}$ هي نسبة النجاح في عينة حجمها n،

1- وضح أن حدود الثقة لتقدير نسبة النجاح في مجتمع p عند درجة ثقة $Z_{1-\frac{\alpha}{2}}$ يعطى بالشكل:

$$p=\frac{\bar{p}+\frac{Z^2_{1-\frac{\alpha}{2}}}{2n}\pm Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}+\frac{Z^2_{1-\frac{\alpha}{2}}}{4n^2}}}{1+\frac{Z^2_{1-\frac{\alpha}{2}}}{n}}$$

2- إلى ماذا تؤول العلاقة السابقة عندما يكون n كبير؟

الحل:

1- لدينا

$$p = \bar{p} \pm Z_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}}$$

$$p - \bar{p} = \pm Z_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}}$$

$$(p-\bar{p})^2 = Z^2_{1-\frac{\alpha}{2}}\frac{p(1-p)}{n}$$

$$n(p^2 - 2p\bar{p} + \bar{p}^2) = Z^2_{1-\frac{\alpha}{2}}p(1-p)$$

$$np^2 - 2np\bar{p} + n\bar{p}^2 = Z^2_{1-\frac{\alpha}{2}}p - Z^2_{1-\frac{\alpha}{2}}p^2$$

$$np^2 - 2np\bar{p} + n\bar{p}^2 - Z^2_{1-\frac{\alpha}{2}}p + Z^2_{1-\frac{\alpha}{2}}p^2 = 0$$

$$(n+Z^2_{1-\frac{\alpha}{2}})p^2 - (2n\bar{p} + Z^2_{1-\frac{\alpha}{2}})p + n\bar{p}^2 = 0$$

نقوم بحل المعادلة السابقة:

$$\Delta = B^2 4AC$$

$$= \left(2n\bar{p} + Z^2_{1-\frac{\alpha}{2}}\right)^2 - 4\left(n + Z^2_{1-\frac{\alpha}{2}}\right)(n\bar{p}^2)$$

$$= 4n^2\bar{p}^2 + 4n\bar{p}Z^2_{1-\frac{\alpha}{2}} + Z^4_{1-\frac{\alpha}{2}} - 4n^2\bar{p}^2 - 4Z^2_{1-\frac{\alpha}{2}}n\bar{p}^2$$

$$= 4n\bar{p}Z^2_{1-\frac{\alpha}{2}} + Z^4_{1-\frac{\alpha}{2}} - 4Z^2_{1-\frac{\alpha}{2}}n\bar{p}^2$$

$$= Z^2_{1-\frac{\alpha}{2}}\left(4n\bar{p} + Z^2_{1-\frac{\alpha}{2}} - 4n\bar{p}^2\right)$$

$$= Z^2_{1-\frac{\alpha}{2}}\left(4n\bar{p}(1-\bar{p}) + Z^2_{1-\frac{\alpha}{2}}\right)$$

$$p_{1,2} = \frac{-B \pm \sqrt{\Delta}}{2A}$$

$$p_{1,2} = \frac{2n\bar{p} + Z^2_{1-\frac{\alpha}{2}} \pm \sqrt{Z^2_{1-\frac{\alpha}{2}}\left(4n\bar{p}(1-\bar{p}) + Z^2_{1-\frac{\alpha}{2}}\right)}}{2(n + Z^2_{1-\frac{\alpha}{2}})}$$

$$p_{1,2}=\frac{\frac{1}{2n}\left(2n\bar{p}+Z^2_{1-\frac{\alpha}{2}}\pm Z_{1-\frac{\alpha}{2}}\sqrt{4n\bar{p}(1-\bar{p})+Z^2_{1-\frac{\alpha}{2}}}\right)}{\frac{1}{2n}\left(2(n+Z^2_{1-\frac{\alpha}{2}})\right)}$$

$$p_{1,2}=\frac{\bar{p}+\frac{Z^2_{1-\frac{\alpha}{2}}}{2n}\pm Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}+\frac{Z^2_{1-\frac{\alpha}{2}}}{4n^2}}}{1+\frac{Z^2_{1-\frac{\alpha}{2}}}{n}}$$

2− عندما تؤول n في العلاقة السابقة إلى ما لانهائية فإنها تصبح بالشكل:

$$p_{1,2}=\bar{p}\pm Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

تمرين رقم 2−8: أعطت عينة عشوائية حجمها 25 متوسطا قدره 80.

المطلوب:

1− أوجد تقدير بمجال ثقة 95% لمتوسط المجتمع في الحالات التالية:

- الانحراف المعياري للمجتمع يساوي 30 والمجتمع يتبع التوزيع الطبيعي،
- الانحراف المعياري للعينة يساوي 30 والمجتمع يتبع التوزيع الطبيعي،
- الانحراف المعياري للعينة يساوي 30 والمجتمع مجهول،

2− قارن بين النتائج؟

الحل:

−1

$n=25, \bar{X}=80, \sigma=30,$ م ي ت ط	$n=25, \bar{X}=80, s=30,$ م ي ت ط	$n=25, \bar{X}=80, s=30,$ م مجهول
نستخدم توزيع طبيعي للتقدير	نستخدم توزيع ستيودنت للتقدير	نستخدم نظرية تشيبتشيف للتقدير
$\mu=\bar{X}\pm Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$ $\mu=80\pm 1.96\frac{30}{\sqrt{n}}$ $\mu=80\pm 11.76$	$\mu=\bar{X}\pm t_{\frac{\alpha}{n-1}}\frac{s}{\sqrt{n}}$ $\mu=80\pm 2.064\frac{30}{\sqrt{n}}$ $\mu=80\pm 11.38$	$\mu=\bar{X}\pm K\frac{s}{\sqrt{n}}$ $\mu=80\pm 4.47\frac{30}{\sqrt{n}}$ $\mu=80\pm 26.82$

2- مقارنة النتائج: إن مجال الثقة باستخدام التوزيع الطبيعي وباستخدام توزيع ستيودنت متقاربان جدا، وهذا بسبب أن حجم العينة يساوي 25، وهي قريبة من 30 التي عندها يتساوى تقريبا كل من التوزيع الطبيعي وتوزيع ستيودنت. على عكس فترة الثقة المحسوبة باستخدام نظرية تشيبتشيف، فهي أوسع بكثير من نظيرتيها. لهذا السبب فهي قليلة الاستخدام، ولكنها تبقى الحل الوحيد إن لم نستطع أن نرفع حجم العينة إلى 30.

تمرين رقم 2-9: لتكن $X=(X_1,X_2,\ldots X_n)$ عينة عشوائية من مجتمع توزيعه $N(\mu,\sigma^2)$. المطلوب:

1 – أثبت أن المقدرات التالية للمعلمة μ غير متحيزة:

$$T_1=\frac{1}{n-1}\sum_{i=1}^{n-1}X_i \qquad ,T_2=\frac{1}{2}\left(\bar{X}+X_1\right) \qquad ,T_3=\frac{1}{3}\left(\bar{X}+X_1+X_2\right)$$

2 – أي المقدرات الثلاثة $i=1,2,3$, T_i لها أصغر تباين؟

الحل:

1- إثبات أن المقدرات غير متحيزة

ليكن يكون المقدر غير متحيز يجب أن يكون $E(T)=\mu$

- المقدر الأول

$$E(T_1)=E\left(\frac{1}{n-1}\sum_{i=1}^{n-1}X_i\right)=\frac{1}{n-1}\sum_{i=1}^{n-1}E(X_i)=\frac{n-1}{n-1}\mu=\mu$$

- المقدر الثاني

$$E(T_2)=E\left(\frac{1}{2}(\bar{X}+X_1)\right)=\frac{1}{2}[E(\bar{X})+E(X_1)]=\frac{1}{2}[\mu+\mu]=\mu$$

- المقدر الثالث

$$E(T_2)=E\left(\frac{1}{3}(\bar{X}+X_1+X_2)\right)=\frac{1}{3}[E(\bar{X})+E(X_1)+E(X_2)]=\frac{1}{3}[\mu+\mu+\mu]$$
$$=\mu$$

المقدرات الثلاث هي مقدرات غير متحيزة لوسط المجتمع.

2- إيجاد تباين المقدرات الثلاث

- المقدر الأول

$$V(T_1)=V\left(\frac{1}{n-1}\sum_{i=1}^{n-1}X_i\right)=\frac{1}{(n-1)^2}\sum_{i=1}^{n-1}V(X_i)=\frac{1}{n-1}\sigma^2$$

- المقدر الثاني

$$V(T_2)=V\left(\frac{1}{2}(\bar{X}+X_1)\right)=\frac{1}{4}[V(\bar{X}+X_1)]=\frac{1}{4}\left[V\left(\frac{\sum_{i=1}^{n}x_i}{n}+X_1\right)\right]$$
$$=\frac{1}{4}\left[V\left(\frac{X_1+\sum_{i=2}^{n}X_i}{n}+X_1\right)\right]=\frac{1}{4}\left[V\left(\frac{\sum_{i=2}^{n}X_i}{n}+\frac{X_1}{n}+X_1\right)\right]$$
$$=\frac{1}{4}\left[V\left(\frac{\sum_{i=2}^{n}X_i}{n}+\frac{(n+1)}{n}X_1\right)\right]$$

$$=\frac{1}{4}\left(\frac{1}{n^2}\sum_{i=2}^{n}V(X_{i)}+\frac{(n+1)^2}{n^2}V(X_1)\right)=\frac{1}{4}\left(\frac{(n-1)}{n^2}\sigma^2+\frac{(n+1)^2}{n^2}\sigma^2\right)$$
$$=\frac{1}{4}\left(\frac{n-1+n^2+2n+1}{n^2}\right)\sigma^2=\frac{n+3}{4n}\sigma^2$$

- المقدر الثالث

$$E(T_3)=V\left(\frac{1}{3}(\bar{X}+X_1+X_2)\right)=\frac{1}{4}[V(\bar{X}+X_1+X_2)]$$
$$=\frac{1}{4}\left[V\left(\frac{\sum_{i=1}^{n}x_i}{n}+X_1+X_2\right)\right]$$
$$=\frac{1}{4}\left[V\left(\frac{X_1+X_2+\sum_{i=3}^{n}X_i}{n}+X_1+X_2\right)\right]$$

$$=\frac{1}{4}\left[V\left(\frac{\sum_{i=3}^{n}X_i}{n}+\frac{X_1}{n}+\frac{X_2}{n}+X_1+X_2\right)\right]$$

$$=\frac{1}{4}\left[V\left(\frac{\sum_{i=3}^{n}X_i}{n}+\frac{(n+1)}{n}X_1+\frac{(n+1)}{n}X_2\right)\right]$$

$$=\frac{1}{9}\left(\frac{1}{n^2}\sum_{i=2}^{n}V(X_{i)}+\frac{(n+1)^2}{n^2}V(X_1)+\frac{(n+1)^2}{n^2}V(X_2)\right)$$

$$=\frac{1}{9}\left(\frac{(n-2)}{n^2}\sigma^2+\frac{(n+1)^2}{n^2}\sigma^2+\frac{(n+1)^2}{n^2}\sigma^2\right)$$

$$=\frac{1}{9}\left(\frac{n-2+2n^2+4n+2}{n^2}\right)\sigma^2=\frac{2n+5}{9n}\sigma^2$$

بمقارنة التباينات الثلاث، نجد أن تباين المقدر الأول هو الأصغر.

تمرين رقم 2−10: أثبت أن المقدر:

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

بأنه يعتبر غير متحيز لتباين المجتمع σ^2؟

الحل:

لأجل أن يكون المقر s^2 مقدر غير متحيز للمعلمة σ^2، يجب أن يكون:

$$E(S^2)=\sigma^2$$

نعلم أنه لكل قيم i فإن:

$$E(x_i)=\mu\,,V(x_i)=\sigma^2$$

لدينا:

$$E(S^2)=E\left(\frac{\sum_{i=1}^{n}(x_i-\bar{X})^2}{n-1}\right)=\frac{\sum_{i=1}^{n}E(x_i-\bar{X})^2}{n-1}=\frac{\sum_{i=1}^{n}\left(x_i^2-2x_i\bar{x}+\bar{X}^2\right)^2}{n-1}=\frac{n}{n-1}\Big(E(x_i^2)-2E(x_i\bar{x})+E(\bar{X}^2)\Big)\quad(1)$$

تحديد قيمة $E(x_i^2)$:

نعلم أن:

$$V(x_i) = E(x_i^2) - \big(E(x_i)\big)^2 \Rightarrow E(x_i^2) = V(x_i) + \big(E(x_i)\big)^2 = \sigma^2 + \mu^2 \quad (2)$$

تحديد قيمة $E\left(\bar{X}^2\right)$:

$$E\left(\bar{X}^2\right) = E(\bar{X}\bar{X}) = E\left(\frac{(\sum_{i=1}^{n} x_i)}{n}\right)\left(\frac{(\sum_{j=1}^{n} x_j)}{n}\right) = \frac{E[(\sum_{i=1}^{n} x_i)(\sum_{j=1}^{n} x_j)]}{n^2} = \frac{\sum_{i\neq j} E(x_i x_j)}{n^2} + \frac{\sum_{i=1}^{n} E(x_i)^2}{n^2} = \frac{1}{n^2} n(n-1)\mu^2 + \frac{n}{n^2}(\sigma^2 + \mu^2) = \mu^2 + \frac{\sigma^2}{n} \quad (3)$$

تحديد قيمة $E(x_i\bar{x})$:

$$E(x_i\bar{x}) = \frac{1}{n}\sum_{i=1}^{n} E\left(x_i x_j\right) = \left(\frac{n-1}{n}\right)\mu^2 + \frac{1}{n}(\sigma^2 + \mu^2) = \mu^2 + \frac{\sigma^2}{n} \quad (4)$$

بتعويض المعادلات (2) ، (3) و (4) في (1) نحصل على:

$$E(S^2) = \frac{n}{n-1}\left(E(x_i^2) - 2E(x_i\bar{x}) + E\left(\bar{X}^2\right)\right) = \frac{n}{n-1}\left(\sigma^2 + \mu^2 - 2\mu^2 - 2\frac{\sigma^2}{n} + \mu^2 + \frac{\sigma^2}{n}\right) = \sigma^2$$

إذن المقدر $S^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{X})^2}{n-1}$ يعتبر مقدر غير متحيز لتباين المجتمع.

الفصل الثالث: اختبار الفرضيات Hypothesis test

3 – 1 تمهيد

لقد رأينا في الفصل الأول والثاني كيف يمكن لنا أن نقوم باستدلالات حول معالم مجتمع انطلاقا من إحصائيات عينة مسحوبة منه. هذه المعلومات المتوفرة تسمح لنا بإجراء المزيد من العمليات الإحصائية حول الظاهرة المدروسة. فقد تتبادر للذهن أحيانا، العديد من الأسئلة حول الظاهرة. مثلا، فهل حقيقة أن متوسط عدد ساعات العمل هو 40 ساعة في الأسبوع. هل نسبة النجاح من مرض ما هي 70%...الخ. هذه الأسئلة تدعى بالفرضيات، يضعها الباحث كتخمين مؤقت لحل تجربته، حيث تقبل أو ترفض. ويوفر لنا الإحصاء من خلال اختبار الفرضيات الوسائل الضرورية اللازمة للقيام بهذا الإجراء.

3-2 المفاهيم الأساسية لاختبار الفرضيات

تعرف الفرضية بأنها: "عبارة (إفادة، تخمين، تصريح، مقولة،...) حول شكل (نمط) توزيع أو خصائص متغير عشوائي أو أكثر". ومثل هذه الفرضيات يمكن صياغتها على أساس التصورات النظرية أو على أساس المعلومات التي توفرها عينة عشوائية من قيم المتغير أو المتغيرات العشوائية الملاحظة أو على أساس أبحاث إحصائية لملاحظات أخرى.[1]

يعتبر اختبار الفروض الإحصائية الوسيلة العلمية للتحقق من النظريات، القوانين أو الملاحظات الموضوعة من طرف الباحث. حيث أنه وفي كثير من الأحيان، يراد الوصول إلى استدلالات عن أعداد كبيرة من المفردات أكبر بكثير من الأعداد المتضمنة فعليا في الدراسة المنجزة.

على سبيل المثال، لدينا المسالة التالية:[2]

يدعي أحد الأطباء المعروفين أن لديه طريقة جديدة أفضل لمعالجة مرض معين. فقد شفي 70% من المرضى العشرة الذين تلقوا هذه المعالجة الجديدة. بينما أشارت إحدى النشرات العلمية المرموقة حول هذا الموضوع إلى شفاء 50% فقط من المرضى في مختلف أرجاء الدولة. اعتمادا على نتائج تجربة الطبيب، هل يمكنك معرفة فيما إذا كان الطبيب قد توصل إلى اكتشاف ما في معالجة هذا المرض؟

لتقييم إدعاء الطبيب، يجب علينا أن نطرح التساؤل التالي: هل النتائج المشاهدة (7 حالات شفاء من أصل 10) بعيدة الاحتمال إذا كانت نسبة الشفاء الحقيقية في المجتمع هي 50%؟

من المؤكد أننا نعلم بأنه إذا كانت نسبة الشفاء الفعلية للمرضى هي النصف تماما، فهذا لا يعني أنه إذا اخترت 10 مرضى في أي وقت، فإن منهم 5 بالضبط سيشفون بهذه المعالجة. إذن، لتقييم إدعاء الطبيب. نأخذ عدد كبير من العينات كل منها مكون من 10 مرضى، وليكن 500 عينة مثلا، ونسجل لكل عينة النسبة المئوية للمتعافين من المرض. إن مبرر القيام بهذه العملية، هو معرفة نوع النتائج الممكنة للعينة وذلك إذا لم تكن المعالجة الجديدة مختلفة عن المعالجة القياسية. يمكننا بعد ذلك تحديد

---

[1]- عبد الحفيظ محمد فوزي مصطفى، الاستدلال الإحصائي (2) نظرية اختبار الفرضيات، مجموعة النيل العربية، مصر، 2002، ص 1.

[2]- ماريجا نوروسيس، تحليل البيانات باستخدام SPSS، شعاع للنشر والعلوم، سورية، 2010، ص ص 262-267.

فيما إذا كان العثور على 70% من حالات الشفاء في عينة مكونة من 10 مرضى هي نتيجة شاذة عندما يكون معدل الشفاء الفعلي هو 50%.

نتائج التجربة نعرضها في الجدول رقم (3-1).

جدول رقم (3-1): نسبة الشفاء لـ 500 عينة من مرض ما

نسبة الشفاء	أقل من 0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	أكبر من 0.9
عدد العينات	3	24	70	98	114	95	66	24	6

من هذا الجدول، يمكننا أن نعرف بشكل تقريبي، ما يجب توقع أن نراه من نتائج مختلفة في العينات المكونة من 10 مرضى. ونلاحظ أنه من أجل معظم العينات فإن نسبة الشفاء قريبة من 50%. وفي الحقيقة، هناك 307 حالة شفاء من أصل 500 عند معدلات الشفاء 40% 50% 60%. وكلما ابتعدت النسبة عن 50% في أي من الاتجاهين، فإنك تحصل على عينات أقل.

يمكن حساب الإحصائيات الوصفية من أجل البيانات التي تم عرضها في الجدول رقم (3-1). حيث نعرضها في الجدول رقم (3-2).

جدول رقم (3-2): إحصائيات وصفية لـ 500 عينة من مرض ما

	N	Minimum	Maximum	Mean	Std. Deviation
X	500	.1000	0.9000	0.5002	0.1622

نلاحظ أن قيمة المتوسط الحسابي هي قريبة جدا من 50%. وهي قيمة المتوسط للمجتمع الذي تم سحب العينات منه. أما الانحراف المعياري للنسب المئوية Std. Deviation فيساوي 16.22%.

على ضوء هذه المعلومات، يمكننا تقدير فيما إذا كانت نتائج الطبيب هي شاذة لو كان معدل الشفاء الحقيقي هو 50%. ونلاحظ أن 66+24+6=96 حالة من أصل 500 تجربة (أي 19.2%) نتجت عند معدلات شفاء 70% أو أكثر. ويدل ذلك على أنه حتى لو كانت المعالجة الجديدة ليست أفضل من المعالجة القياسية، فإنه يمكنك أن تتوقع أن تجد معدلات شفاء مساوية على الأقل لتلك المشاهدة من قبل الطبيب وذلك بمعدل 5/1 تقريبا تعيد فيها التجربة.

وبالطبع، فإنه من الممكن دوما أن تكون المعالجة الجديدة أقل فاعلية عمليا من المعالجة المعتادة. ولذلك، إذا أردنا اختبار الفرضية بأن المعالجة الجديدة لا تختلف عن المعالجة القياسية، فإن عليك تقييم احتمال أن تكون النتائج في أقصى الحدود المشاهدة بأحد الاتجاهين (زيادة معدل الشفاء أو انخفاضه). ومن الجدول رقم (2−1)، يمكننا تقدير بان احتمال وجود 30% فأقل و70% فأكثر من حالات الشفاء هو (96+97)/500 =38.6%.

اعتمادا على هذه المعلومة، هناك أمل ضعيف للاعتقاد بأن الطبيب قد حقق شيئا مهما بالفعل. إذ أن نتائجه ليست على خلاف مع العينات المختارة من المجتمع الذي تبلغ نسبة الشفاء الحقيقية فيه 50%.

تبعا للمثال السابق، فإننا نستنتج أن اختبار الفرضيات يهتم بالمسألة التالية:

ليكن X متغير عشوائي معرف على مجتمع ما ذي قانون هو بدلالة θ ، هل نستطيع افتراض أن $\theta = \theta_0$ معطاة مقدما؟

فرضية العدم والفرضية البديلة Null hypothesis and Alternative hypothesis

لقد قدمنا في المثال السابق، شرح لكيفية اختبار الفرضيات فيما إذا كان لأحد الأدوية الجديدة نفس معدل الشفاء للمعالجة القياسية. ونلاحظ في هذا المثال أن هناك فرضيتان أو إدعاءان يحظيان بالاهتمام. الفرضية الأولى هي أنه لم يحدث شيء: فالدواء الجديد فعال بنفس المستوى كالدواء القياسي. وباستخدام المصطلحات الإحصائية، فإن هذه الفرضية تسمى فرضية العدم. لاحظ أن فرضية العدم محددة بإحكام. إنها تصف حالة افتراضية ولكن دقيقة. أما الفرضية الثانية، وهي الفرضية البديلة فتصف الحالة عندما تكون فرضية العدم خاطئة. وفيما يلي الفرضية البديلة: يغير الدواء الجديد معدل الشفاء.

لهذا فإننا نعرف الاختبار الإحصائي كما يلي: "اختبار إحصائي هو عملية تسمح بحساب قيمة دالة ما لملاحظات عينة واحدة أو أكثر، والتي تؤدي إلى رفض أو قبول بمستوى معنوية فرضية نسميها فرضية عدم H_0 مقابل الفرضية البديلة H_1.

وعند اختبار إحدى الفرضيات إحصائيا، فإننا نفترض بأن فرضية العدم تصف الحالة بشكل صحيح. وفرضية العدم هي الإطار المرجعي الذي تحكم بموجبه على نتائج العينة. نحن نفترض أن معدل الشفاء هي قيمة المجتمع البالغة 50%. وهكذا نجد أن فرضية العدم تصف حالة جيدة التعريف.

إذا كان معدل الشفاء للمجتمع هو 50%، فإنه يمكنك تحديد عدد المرات التي تتوقع أن تجد فيها مختلف نتائج العينات الممكنة، مثل 12 حالة شفاء أو أكثر في عينة مكونة من 20 مريضا. يجب كذلك على فرضية العدم أن تصف الحالة بشكل فريد. مثلا، لا يمكن لفرضية العدم أن تشير إلى أن معدل الشفاء ليس 50%. فمثل هذه العبارة لا تصلح كإطار مرجعي لتقييم نتائج العينة، لأنها تصف عدد من النتائج المحتملة. وفي معظم الأحيان، وعندما تقوم بتنفيذ تجربة أو عملية مسح معينة، فإن فرضية العدم تدعي عكس ما نريد أن نثبت صحته.

يمكن أن تحدد الفرضية البديلة اتجاه الفرق الذي تتوقع مشاهدته. إذا كنت تعلم بأن معدل الشفاء الذي تصفه لا يمكن أن يكون أسوأ من القياسي، فإنه يمكن للفرضية البديلة الإدعاء بأن معدل الشفاء هو أفضل من القياسي. وعلى كل حال، يجب الإشارة إلى اتجاه الفرضية البديلة بشكل مسبق.

منطقة القبول والرفض

منطقة الرفض أو المنطقة الحرجة هي فئة جزئية من فضاء العينة ترفض عند كل نقطة (عينة مشاهدة) من نقاطها فرضية العدم. أما منطقة القبول فهي فئة جزئية من فضاء العينة تقبل عند كل نقطة من نقاطها فرضية العدم.

إحصاء الاختبار Test statistic

هو إحصاء نجري من خلاله الاختبار. أي أن قرار قبول أو رفض فرضية العدم يبنى على أساس قيمة إحصاء الاختبار عند العينة الملاحظة.

رفض فرضية العدم

بما أن فرضية العدم تعمل كإطار مرجعي يتم وفقه تقييم نتائج العينة، فإذا كانت نتائج العينة تبدو غير محتملة عندما تكون فرضية العدم صحيحة، فإنك ترفض فرضية العدم. إذا كان احتمال الحصول على نتائج العينة التي تساوي النتائج المشاهدة (مستوى المعنوية المشاهد) صغيرا، وعادة أصغر من 0.05، فنخول برفض فرضية العدم.

الخطأ من النوع الأول والخطأ من النوع الثاني

يقوم الباحث دائما بإجراء تجربة، يريد من خلالها اختبار فرضية معينة، يحصل من خلالها على نتائج، إما يرفض فرضية العدم وإما يقبلها. نتيجة هذا الاختبار تعتمد على عدة عوامل منها: حجم العينة، طبيعة المجتمع، مستوى الدلالة وعلى تباين الفرق. لهذا فمن الممكن أن تكون نتيجة اختبار الفرضية صحيحة كما قد تكون خاطئة. ونطلق تسمية الخطأ من النوع الأول والثاني كما هو موضح في جدول الأخطاء.

جدول رقم (3-3): جدول الأخطاء

القرار \ الحقيقة	H_0 صحيح	H_0 خاطئ
H_0 صحيح	قرار حسن	خطا من النوع الثاني
H_0 خاطئ	خطأ من النوع الأول	قرار حسن

يظهر لنا الجدول رقم (3-3) أن قرار الباحث الذي توصل إليه من تجربته الخاصة، قد يكون متوافقا مع الحقيقة (المجتمع) أو قد لا يتوافق معه. حيث أنه إذا توافقت نتائج اختبار فرضياته مع الحقيقة فإن القرار المتخذ يكون حسن. أما إن كانت نتيجة اختبار فرضيته خاطئة، بينما في الحقيقة كان يجب أن نقبل فرض العدم، يكون الباحث قد ارتكب خطأ من النوع الأول. وإن كانت نتيجة اختبار فرضيته صحيحة، بينما في الحقيقة كان يجب أن يرفض فرض العدم، يكون الباحث قد ارتكب خطأ من النوع الثاني.

3-3 اختبار الفرضيات المعلمية Parametric Hypothesis test

يتعلق هذا النوع باختبار الفرضيات حول معالم مجتمع ما التي يفترض في هذه البيانات معرفة طبيعة التوزيع الذي يتبعه المتغير العشوائي المرافق لها X.

ويتم إجراء اختبار الفرضيات على طريق إتباع المراحل التالية:

المرحلة الأولى: تحديد صيغة الفرضية المعلمية بشكل عام بإحدى العبارات التالية:

$$H_0: \theta = \theta_0 \qquad H_0: \theta \geq \theta_0 \qquad H_0: \theta \leq \theta_0$$
$$H_1: \theta \neq \theta_0' \qquad H_1: \theta < \theta_0' \qquad H_1: \theta > \theta_0$$

ملاحظة: تضع بعض المراجع فرض العدم دائما في شكل مساواة، والنتيجة واحدة.

المرحلة الثانية: تحديد قاعدة القرار وفق ما يلي:

- تحديد طبيعة التوزيع المستخدم في اختبار الفرضية،
- تحديد مستوى معنوية الاختبار،
- تحديد منطقتي القبول والرفض.

المرحلة الثالثة: يتم حساب قيمة إحصاء الاختبار انطلاقا من عينة مسحوبة من المجتمع. وعلى أساس قاعدة القرار المحددة في المنطقة الثانية، نقبل أو نرفض فرض العدم لصالح الفرض البديل.

3-3-1 اختبار الفرضيات حول وسط مجتمع

3-3-1-1 اختبار الفرضيات حول متوسط مجتمع باستخدام التوزيع الطبيعي

يتم استخدام التوزيع الطبيعي لاختبار فرضيات حول متوسط مجتمع في الحالتين التاليتين:

$n \geq 30, \sigma$ أو s معلومة

$n > 30, \sigma$ معلومة, مجتمع يتبع توزيع طبيعي

مثال رقم 3-1: شركة تنتج مصابيح كهربائية ترغب في معرفة ما إذا كان يمكنها الإدعاء بأن مصابيحها تستمر لمدة 1000 ساعة، لأجل ذلك، قامت بسحب عينة من 100 مصباح، فوجدت أن المتوسط الحسابي هو 980 والانحراف المعياري هو 80.

المطلوب: هل يمكن للشركة أن تقوم بهذا الإدعاء عند مستوى معنوية قدره 5%؟

الحل:

المعطيات:

$$n = 100, \qquad \bar{X} = 80, \quad \mu = 1000 \quad s = 80, \quad \alpha = 0.05,$$

المرحلة الأولى: تحديد الفرض

$$H_0: \mu = 1000$$
$$H_1: \mu \neq 1000$$

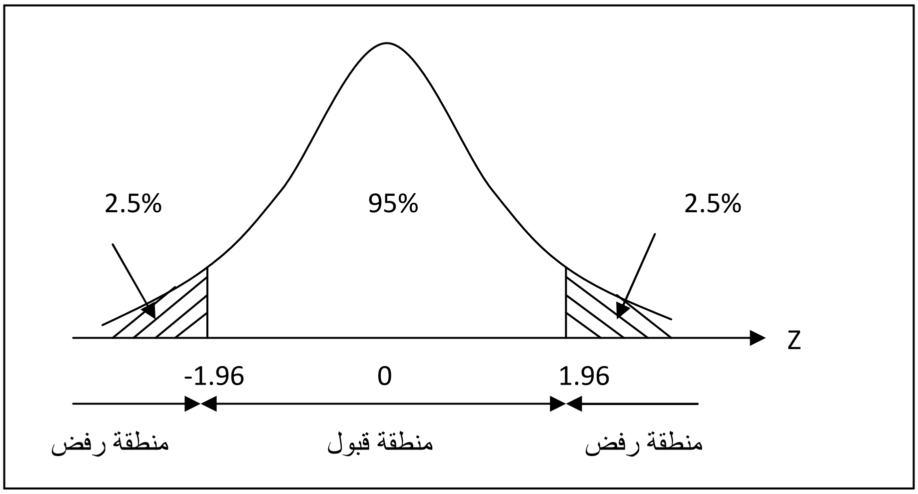
المرحلة الثانية: تحديد قاعدة القرار

على ضوء المعطيات فإننا نستخدم التوزيع الطبيعي لإجراء هذا الاختبار.

بما أن صيغة الفرضية جاءت بالشكل مساواة وعدم مساواة. فهذا يعني أن مستوى المعنوية المحدد لإجراء هذا الاختبار يكون عند ذيلي التوزيع. لأن عدم المساواة يمكن أن تكون من جهة الأكبر من كما يمكن أن تكون من جهة الأصغر من. ويسمى هذا الاختبار باختبار ذو ذيلين.

يتم استخراج قيمة $Z_{\text{الجدولية}}$ التي يتم على أساسها تحديد منطقتي القبول والرفض بالعلاقة:

$$Z_{1-\frac{\alpha}{2}} = Z_{1-\frac{0.05}{2}} = Z_{0.975} = 1.96$$



من خلال الشكل السابق تتضح لنا قاعدة القرار التالية:

- إذا وقعت قيمة $Z_{\text{المحسوبة}}$ في منطقة الرفض فإننا نرفض فرض العدم ونقبل الفرض البديل،

- إذا وقعت قيمة $Z_{\text{المحسوبة}}$ في منطقة القبول فإننا نقبل فرض العدم ونرفض الفرض البديل.

المرحلة الثالثة: اتخاذ القرار

نقوم بإيجاد قيمة إحصاء الاختبار $Z_{\text{المحسوبة}}$ باستخدام العبارة التالية:

$$Z_{\text{المحسوبة}} = \frac{\bar{X} - \mu_0}{\sigma_{\bar{X}}} = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} = \frac{980 - 1000}{80/\sqrt{100}} = -2.5$$

حيث أن قيمة $Z_{\text{المحسوبة}}$ تقع في منطقة الرفض، فإننا نرفض فرض العدم ونقبل الفرض البديل عند مستوى معنوية قدره 5%، ومنه فالشركة لا يمكنها الادعاء بأن مصابيحها تشتغل لمدة 1000 ساعة.

مثال رقم 3−2: تشير ملصقات إحدى علب القهوة، أن كل منها يحتوي على الأقل على 3 رزم. لاختبار مدى صدق هذا الإعلان، قمنا بسحب عينة حجمها 26 علبة، أعطت متوسط حسابي قدره 2.92. إذا علمت من دراسات سابقة أن الانحراف المعياري للمجتمع يساوي 0.18.

المطلوب: اختبر مدى صحة المعلومة الموضوعة في ملصق علب القهوة عند مستوى معنوية قدره 5%.

الحل:

المعطيات:

$$n = 26, \quad \bar{X} = 2.92, \quad \mu = 3 \quad \sigma = 80, \quad \alpha = 0.05,$$

– وضع الفرض

$$H_0: \mu \geq 3$$
$$H_1: \mu < 3$$

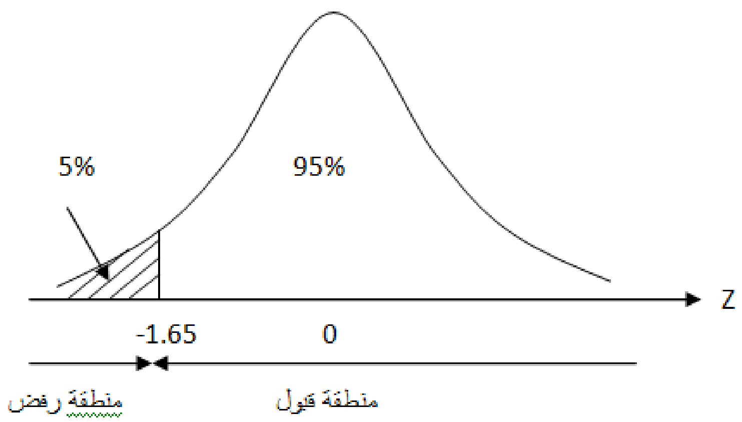
– تحديد قاعدة القرار

على ضوء المعطيات فإننا نستخدم التوزيع الطبيعي للاختبار.

بما أن صيغة الفرضية جاءت بالشكل أعلاه. فهذا يعني أن مستوى المعنوية المحدد لإجراء هذا الاختبار يكون عند ذيل التوزيع الأيسر. ويسمى هذا الاختبار باختبار ذو الذيل الأيسر.

يتم استخراج قيمة $Z_{\text{الجدولية}}$ التي يتم على أساسها تحديد منطقتي القبول والرفض بالعلاقة:

$$Z_{1-\alpha} = Z_{1-0.05} = Z_{0.95} = 1.65$$



- اتخاذ القرار

نقوم بإيجاد قيمة $Z_{\text{المحسوبة}}$ باستخدام العبارة التالية:

$$Z_{\text{المحسوبة}} = \frac{\bar{X}-\mu_0}{\sigma_{\bar{X}}} = \frac{\bar{X}-\mu_0}{s/\sqrt{n}} = \frac{2.92-3}{0.18/\sqrt{26}} = -2.26$$

حيث أن قيمة $Z_{\text{المحسوبة}}$ تقع في منطقة الرفض، فإننا نرفض فرض العدم ونقبل الفرض البديل عند مستوى معنوية قدره 5%، ومنه فالمعلومة الموضوعة على ملصقة علب القهوة سليمة.

3-3-1-2 اختبار الفرضيات حول متوسط مجتمع باستخدام توزيع ستيودنت

يتم استخدام توزيع ستيودنت لاختبار فرضيات حول متوسط مجتمع في الحالة التالية:

مجتمع يتبع توزيع طبيعي, معلومة s, $nn > 30$

مثال رقم 3-3: يعمل نظام إنتاجي في مؤسسة على تعبئة علب بمتوسط وزن قدره 16 أونصة. إذا قامت الآلة بتعبئة كمية أقل، فلن يشير الملصق الموضوع على الكمية، أما إن كانت التعبئة فوق هذا الحد، فالشركة ستتحمل خسارة. من أجل مراقبة العلية الإنتاجية، فإن مديرية الجودة تسحب دوريا عينة عشوائية من 8 علب. حيث أعطت الأوزان التالية:

16.02، 16.22، 15.82، 15.92، 16.22، 16.32، 16.12، 15.92.

المطلوب: إذا افترضنا أن أوزان العلب يتبع التوزيع الطبيعي، فهل يتم توقيف العملية الإنتاجية عند مستوى معنوية قدره 5%؟

الحل:

المعطيات:

$$n = 8, \quad \mu = 16, \quad \alpha = 0.05,$$

المرحلة الأولى: تحديد الفرض

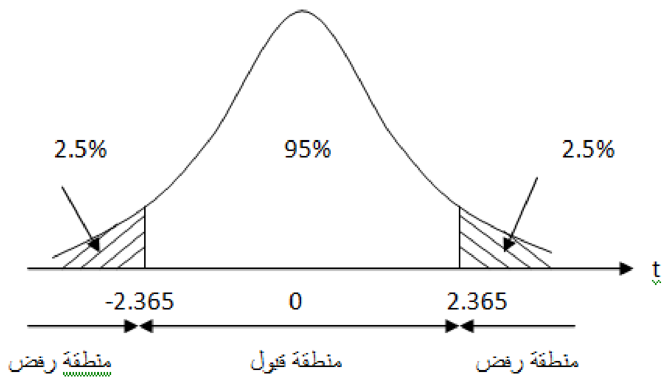
$$H_0: \mu = 16$$
$$H_1: \mu \neq 16$$

المرحلة الثانية: تحديد قاعدة القرار

على ضوء المعطيات فإننا نستخدم توزيع ستيودنت لإجراء هذا الاختبار.

يتم استخراج قيمة $t_{\text{الجدولية}}$ التي يتم على أساسها تحديد منطقتي القبول والرفض بالعلاقة:

$$t_{\substack{\alpha \\ n-1}} = t_{\substack{0.05 \\ 7}} = 2.365$$



من خلال الشكل السابق تتضح لنا قاعدة القرار التالية:

– إذا وقعت قيمة $t_{\text{المحسوبة}}$ في منطقة الرفض فإننا نرفض فرض العدم ونقبل الفرض البديل،

– إذا وقعت قيمة $t_{\text{المحسوبة}}$ في منطقة القبول فإننا نقبل فرض العدم ونرفض الفرض البديل.

المرحلة الثالثة: اتخاذ القرار

– نقوم بحساب قيمة المتوسط الحسابي والانحراف المعياري

$$\bar{X} = \frac{\sum x_i}{n} = \frac{128.56}{8} = 16.07$$

$$s = \sqrt{\frac{\sum(x_i - \bar{X})^2}{n-1}} = \sqrt{\frac{0.22}{7}} = 0.18$$

نقوم بإيجاد قيمة إحصاء الاختبار $t_{\text{المحسوبة}}$ باستخدام العبارة التالية:

$$t_{\text{المحسوبة}} = \frac{\bar{X} - \mu_0}{\sigma_{\bar{X}}} = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} = \frac{16.07 - 16}{0.18/\sqrt{8}} = 1.10$$

حيث أن قيمة $t_{\text{المحسوبة}}$ تقع في منطقة القبول، فإننا نقبل فرض العدم ونرفض الفرض البديل عند مستوى معنوية قدره 5%، ومنه فالشركة لا تقوم بتوقيف العملية الإنتاجية.

مثال رقم 3–4: ترغب شركة أن تعرف بدرجة ثقة 95% ما إذا كان يمكنها الإدعاء بأن صناديق الصابون المسحوق الذي تبيعه تحتوي على أكثر من 500 غ. وتعرف الشركة من الخبرة الماضية، أن أوزان الصابون بالصناديق تتبع التوزيع الطبيعي. لأجل ذلك، أخذت الشركة عينة عشوائية حجمها 25، ووجدت أن المتوسط الحسابي هو 520 والانحراف المعياري هو 75غ.

الحل:

المعطيات:

$$n = 25, \quad \bar{X} = 520, \quad \mu = 500 \quad s = 75, \quad \alpha = 0.05,$$

– وضع الفرض

$$H_0: \mu \leq 500$$
$$H_1: \mu > 500$$

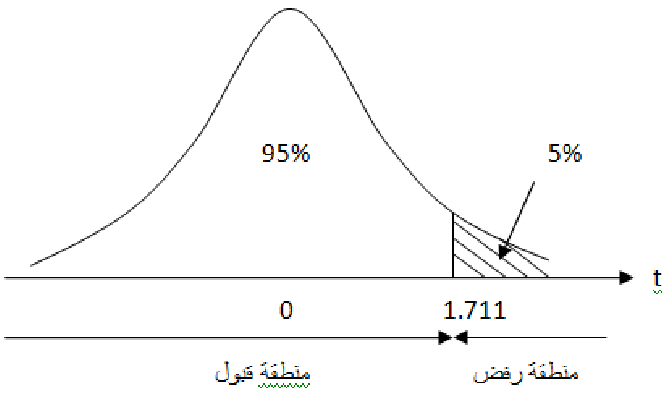
– تحديد قاعدة القرار

على ضوء المعطيات فإننا نستخدم توزيع ستيودنت للاختبار.

بما أن صيغة الفرضية جاءت بالشكل أعلاه. فهذا يعني أن مستوى المعنوية المحدد لإجراء هذا الاختبار يكون عند ذيل التوزيع الأيمن. ويسمى هذا الاختبار باختبار ذو الذيل الأيمن.

يتم استخراج قيمة $t_{\text{الجدولية}}$ التي يتم على أساسها تحديد منطقتي القبول والرفض بالعلاقة:

$$t_{\underset{n-1}{\alpha}} = t_{\underset{24}{0.05}} = 1.711$$



– اتخاذ القرار

نقوم بإيجاد قيمة $t_{\text{المحسوبة}}$ باستخدام العبارة التالية:

$$t_{\text{المحسوبة}} = \frac{\bar{X} - \mu_0}{\sigma_{\bar{X}}} = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} = \frac{520 - 500}{75/\sqrt{25}} = 1.33$$

حيث أن قيمة $t_{\text{المحسوبة}}$ تقع في منطقة القبول، فإننا نقبل فرض العدم ونرفض الفرض البديل عند مستوى معنوية قدره 5%، ومنه فالشركة لا يمكنها الإدعاء بأن صناديق الصابون الذي تبيعه تحتوي على أكثر من 500غ.

3-3-2 اختبار الفرضيات حول نسبة مجتمع

يتم اختبار الفرض حول نسبة مجتمع باستخدام التوزيع الطبيعي.

مثال رقم 3-5: تظهر السجلات أن 60% من الطلبة الذين التحقوا في الماضي بدراسة جامعية معينة، قد حصلوا على الدرجة العلمية خلال ثلاث سنوات. وبالنسبة للملتحقين بالدراسة عام 2014، وعددهم 36، وجد أن 15 طالبا فقط قد حصلوا على الدرجة العلمية سنة 2017.

المطلوب: هل كانت نتائج دفعة طلبة سنة 2014 أسوأ من نتائج الدفعات السابقة عند مستوى معنوية قدره 5%؟

الحل:

المعطيات:

$$p = 0.6, \qquad \alpha = 0.05,$$

- تحديد الفرض

$$H_0: p \geq 0.6$$
$$H_1: p < 0.6$$

- تحديد قاعدة القرار

حيث أن هذا الاختبار هو اختبار الذي الأيسر، فإن منطقة الرفض تقع في المجال $]-\infty, -1.65[$، أما منطقة القبول تقع خارج

- اتخاذ القرار

نقوم بحساب قيمة نسبة وانحراف العينة

$$\bar{p} = \frac{15}{36} = 0.42$$

$$\sigma_{\bar{p}} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{0.6 \times 0.4}{36}} = 0.08$$

نقوم بإيجاد قيمة إحصاء الاختبار $t_{\text{المحسوبة}}$ باستخدام العبارة التالية:

$$Z_{\text{المحسوبة}} = \frac{\bar{p} - p_0}{\sigma_{\bar{p}}} = \frac{0.42 - 0.6}{0.08} = -2.25$$

حيث أن قيمة $Z_{\text{المحسوبة}}$ تقع في منطقة الرفض، فإننا نرفض فرض العدم ونقبل الفرض البديل عند مستوى معنوية قدره 5%، ومنه فإن نتائج طلبة دفعة 2014 كانت أسوا من نتائج الدفعات السابقة.

3-3-3 اختبار الفرضيات للفروق أو المجاميع

3-3-3-1 اختبار الفرضيات للفروق أو المجاميع بين متوسطين

يتم صياغة الفرضيات في هذا الاختبار بإحدى العبارات التالية:

$$H_0: \theta_1 = \theta_2 \qquad H_0: \theta_1 \geq \theta_2 \qquad H_0: \theta_1 \leq \theta_2$$
$$H_1: \theta_1 \neq \theta_2, \qquad H_1: \theta_1 < \theta_2, \qquad H_1: \theta_1 > \theta_2$$

ويختلف التوزيع المستخدم للاختبار حسب الحالتين التاليتين:

أولا: حالة المجتمع يتبع التوزيع الطبيعي والتباين معلوم

إذا كان العينتان كبيرتان*، أو إذا كانت صغيرتان من مجتمع (أو مجتمعان) يتبع التوزيع الطبيعي بتباين معلوم، فإننا نستخدم التوزيع الطبيعي لاختبار الفرضيات حول الفروق أو المجاميع بين المتوسطين، بحيث يحسب إحصاء الاختبار بالعلاقة التالية:

$$Z_{\text{المحسوبة}} = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sigma_{\bar{X}_1 - \bar{X}_2}} = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

مثال رقم 3-6: ترغب مديرة أن تحدد عند مستوى معنوية قدره 5% ما إذا كان الأجر بالساعة للعمال متساوي في مدينتين. لعمل ذلك، تأخذ عينة عشوائية من الأجر بالساعة في كلتا المدينتين. وتحصل على المعلومات التالية (الانحراف المعياري للمجتمع معلوم مسبقا):

$$n_1 = 40, \qquad \bar{X}_1 = 600, \qquad \sigma_1 = 200$$

*نقصد بالعينات الكبيرة هنا هو أن تكون حجم العينتين معا أكبر من 30 (n1 وn2 أكبر من 30)، بينما العينات الصغيرة هي التي يكون حجم العينتين معا أقل من 30 أو حجم عينة واحدة أقل من 30 (n1 و/أوn2 أكبر من 30).

$$n_2 = 54, \qquad \bar{X}_2 = 450, \qquad \sigma_2 = 180$$

الحل:

- تحديد الفرض

$$\begin{matrix} H_0: \mu_1 = \mu_2 \\ H_1: \mu_1 \neq \mu_2 \end{matrix} \; or \; \begin{matrix} H_0: \mu_1 - \mu_2 = 0 \\ H_1: \mu_1 - \mu_2 \neq 0 \end{matrix}$$

- تحديد قاعدة القرار

عند مستوى معنوية قدره5% فإن منطقة القبول تكون ضمن المجال ∓ 1.96 ومنطقة الرفض خارج هذا المجال

- اتخاذ القرار

نقوم بإيجاد قيمة إحصاء الاختبار $Z_{\text{المحسوبة}}$ باستخدام العبارة التالية:

$$Z_{\text{المحسوبة}} = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sigma_{\bar{X}_1 - \bar{X}_2}} = \frac{(600 - 450) - (0)}{\sqrt{\frac{200^2}{40} + \frac{180^2}{54}}} = 1.5$$

حيث أن قيمة $Z_{\text{المحسوبة}}$ تقع في منطقة القبول، فإننا نقبل فرض العدم ونرفض الفرض البديل. ومنه فإن الأجر بالساعة في كلتا المدينتين متساوي عند مستوى معنوية قدره 5%.

ثانيا: حالة المجتمع يتبع التوزيع الطبيعي والتباين مجهول

إذا كانت قيمة التباين مجهولة، فإننا نستخدم التباين المحسوب في العينتين كمقدر لتباين المجتمع. وعلى هذا الأساس فإن إحصاء الاختبار تعطى بالصورة التالية:

$$t_{\text{المحسوبة}} = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sigma_{\bar{X}_1 - \bar{X}_2}} = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

أما درجة حرية الاختبار فتعطى بالعبارة:

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{s_2^2}{n_2}\right)^2}$$

مثال رقم 3-7: قدمت إحدى مؤسسات الإعلام الآلي برمجية جديدة لمعالجة المعلومات تهدف إلى توفير الوقت اللازم لهذه العملية. لأجل اختبار مدى فاعلية البرنامج، تم اختيار عينة عشوائية حجمها 24. حيث تم تطبيق برمجية المعالجة القديمة في 12 عينة، بينما تم تطبيق البرمجية الجديدة على 12 عينة المتبقية. حيث تم التوصل إلى البيانات التالية:

البرمجية الحالية	البرمجية الجديدة
300	274
280	220
344	308
385	336
372	198
360	300
288	315
321	258
376	318
290	310
301	332
283	263

المطلوب: هل تقوم البرمجية الثانية بإنجاز العمل في وقت أقصر من البرمجية الأولى عند مستوى معنوية قدره 5%؟

الحل:

– تحديد الفرض

$$H_0: \mu_1 \le \mu_2 \quad or \quad H_0: \mu_1 - \mu_2 \le 0$$
$$H_1: \mu_1 > \mu_2 \qquad\quad H_1: \mu_1 - \mu_2 > 0$$

- تحديد قاعدة القرار

حساب قيم المتوسط الحسابي والانحراف المعياري

$$\bar{X}_1 = \frac{\sum x_i}{n} = 325, \qquad \bar{X}_2 = \frac{\sum x_i}{n} = 286$$

$$s_1 = \sqrt{\frac{\sum(x_i - \bar{X})^2}{n-1}} \cong 40, \qquad s_2 = \sqrt{\frac{\sum(x_i - \bar{X})^2}{n-1}} \cong 44$$

درجة حرية الاختبار هي:

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{s_2^2}{n_2}\right)^2} = \frac{\left(\frac{40^2}{12} + \frac{44^2}{12}\right)^2}{\frac{1}{12 - 1}\left(\frac{40^2}{12}\right)^2 + \frac{1}{12 - 1}\left(\frac{44^2}{12}\right)^2} = 21.8$$

عند مستوى معنوية قدره5% ودرجة حرية 22 فإن منطقة الرفض تكون ضمن المجال $]1.717, +\infty[$ ومنطقة الرفض خارج هذا المجال.

- اتخاذ القرار

نقوم بإيجاد قيمة إحصاء الاختبار $t_{\text{المحسوبة}}$ باستخدام العبارة التالية:

$$t_{\text{المحسوبة}} = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sigma_{\bar{X}_1 - \bar{X}_2}} = \frac{(325 - 286) - (0)}{\sqrt{\frac{40^2}{12} + \frac{44^2}{12}}} = 2.27$$

حيث أن قيمة $t_{\text{المحسوبة}}$ تقع في منطقة الرفض، فإننا نرفض فرض العدم ونقبل الفرض البديل. ومنه فإن الوقت اللازم لإنجاز العمل باستخدام البرمجية الجديدة أقل من الوقت اللازم لإنجازه باستخدام البرمجية الحالية وهذا عند مستوى معنوية قدره 5%.

حالة خاصة:

في حالة العينات الصغيرة، ومع افتراض أن:

$$\sigma^2_{\bar{X}_1} = \sigma^2_{\bar{X}_2} = \sigma^2$$

بحيث تعتبر قيمة الانحراف المعياري للمجتمع σ^2 مجهولة أيضا.

وعلى هذا الأساس فإن إحصاء الاختبار تعطى بالصورة التالية:

$$t_{\text{المحسوبة}} = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sigma_{\bar{X}_1 - \bar{X}_2}} = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

بحيث:

$$s = \sqrt{\frac{n_1 - 1}{n_1 + n_2 - 2} s_1^2 + \frac{n_2 - 1}{n_1 + n_2 - 2} s_2^2}$$

أما درجة حرية الاختبار فهي: n_1+n_2-2

أما درجة حرية الاختبار فتعطى بالعبارة:

مثال رقم 3-8: حل المثال مع افتراض أن المجتمعان يتبعان التوزيع الطبيعي بتباين متساوي.

الحل:

درجة حرية الاختبار هي: 22=2-12+12

عند مستوى معنوية قدره5% ودرجة حرية 22 فإن منطقة الرفض تكون ضمن المجال $]1.717, +\infty[$ ومنطقة الرفض خارج هذا المجال.

- اتخاذ القرار

نقوم بإيجاد قيمة إحصاء الاختبار $t_{\text{المحسوبة}}$ باستخدام العبارة التالية:

$$s = \sqrt{\frac{n_1 - 1}{n_1 + n_2 - 2} s_1^2 + \frac{n_2 - 1}{n_1 + n_2 - 2} s_2^2} = \sqrt{\frac{12 - 1}{12 + 12 - 2} 40^2 + \frac{12 - 1}{12 + 12 - 2} 44^2} = 42$$

$$t_{\text{المحسوبة}} = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sigma_{\bar{X}_1 - \bar{X}_2}} = \frac{(325 - 288) - (0)}{42\sqrt{\frac{1}{12} + \frac{1}{12}}} = 2.16$$

حيث أن قيمة $t_{\text{المحسوبة}}$ تقع في منطقة الرفض، فإننا نرفض فرض العدم ونقبل الفرض البديل. ومنه فإن الوقت اللازم لإنجاز العمل باستخدام البرمجية الجديدة في المتوسط أقل من الوقت اللازم لإنجازه باستخدام البرمجية الحالية وهذا عند مستوى معنوية قدره 5%.

3-3-3-1 اختبار الفرضيات للفروق بين نسبتين

لاختبار فرضية فروق بين نسبتين مجتمعين كبيرين، فإن صياغة هذه الفرضية تكون بإحدى الصيغ الثلاث التالية:

$$H_0: p_1 = p_2 \qquad H_0: p_1 \geq \theta_2 \qquad H_0: p_1 \leq p_2$$
$$H_1: p_1 \neq p_2{}' \qquad H_1: p_1 < \theta_2{}' \qquad H_1: p_1 > p_2$$

ويحسب إحصاء الاختبار بالصيغة التالية:

$$Z_{\text{المحسوبة}} = \frac{(\bar{p}_1 - \bar{p}_2) - (p_1 - p_2)}{\sigma_{\bar{p}_1 - \bar{p}_2}} = \frac{(\bar{p}_1 - \bar{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}}$$

بحيث:

$$\sigma_{\bar{p}_1 - \bar{p}_2} = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

حيث أن قيمة نسب المجتمع هي مجهولة، وحيث أن فرض العدم يقول بعدم وجود فروق بين نسبتي المجتمع. فإننا توزيع المعاينة للفرق بين النسبتين يعطى بالشكل:

$$\sigma_{\bar{p}_1 - \bar{p}_2} = \sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

مع:

$$\bar{p} = \frac{n_1\bar{p}_1 + n_2\bar{p}_2}{n_1 + n_2}$$

مثال رقم 3−9: نعرض في الجدول الموالي تصريح بالدخل لعينتين مستقلتين يعرضها مكتبان.

	مكتب 1	مكتب 2
حجم العينة	250	300
عدد التصريحات الخاطئة	35	27

المطلوب: هل يوجد فرق في التصريحات الخاطئة بين المكتبين عند مستوى معنوية 10%

الحل:

- تحديد الفرض

$$\begin{matrix} H_0: p_1 = p_2 \\ H_1: p_1 \neq p_2 \end{matrix} \quad or \quad \begin{matrix} H_0: p_1 - p_2 = 0 \\ H_1: p_1 - p_2 \neq 0 \end{matrix}$$

- تحديد قاعدة القرار

عند مستوى معنوية قدره10% فإن منطقة القبول تكون ضمن المجال ∓ 1.65 ومنطقة الرفض خارج هذا المجال

- اتخاذ القرار

حساب قيمة نسب العينة:

$$\bar{p}_1 = \frac{35}{250} = 0.14, \qquad \bar{p}_2 = \frac{27}{300} = 0.09,$$

نقوم بإيجاد قيمة إحصاء الاختبار $Z_{\text{المحسوبة}}$ باستخدام العبارة التالية:

$$\bar{p} = \frac{250(0.14) + 300(0.09)}{250 + 300} = 0.113$$

$$\sigma_{\bar{p}_1 - \bar{p}_2} = \sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = \sqrt{(0.113)(0.887)\left(\frac{1}{250} + \frac{1}{300}\right)} = 0.0271$$

$$Z_{\text{المحسوبة}} = \frac{(\bar{p_1} - \bar{p_2}) - (p_1 - p_2)}{\sigma_{\bar{p}_1 - \bar{p}_2}} = \frac{(0.14 - 0.09) - 0}{0.0271} = 1.85$$

حيث أن قيمة $Z_{\text{المحسوبة}}$ أكبر من $Z_{\text{الجدولية}}$، فإننا نرفض فرض العدم بمستوى معنوية قدره 10%. ومنه، وحسب بيانات العينة. فإن نسبة تصريحات الدخل الخاطئة تعتبر مختلفة.

3-4 اختبار الفرضيات اللامعلمية Non parametric hypothesis test

على عكس اختبار الفرضيات التي تم تناولها في العنصر 3-3، والتي تستوجب معرفة طبيعة التوزيع لإجرائها. فإن الاختبارات اللامعلمية يمكن أن تستخدم في غياب هذا الشرط.

كذلك قد نصادف حالات يصعب إجراء قياسات دقيقة لها مثل ترتيب أصناف من الطعام حسب المذاق، أو ترتيب كفاءة موظفين لأداء مهمة معينة. هذه الحالات وغيرها من الحالات المشابهة التي تكون فيها المقاييس من الشكل الكيفي أو الترتيبي تحتم علينا استخدام الاختبارات اللامعلمية. التي تلخص مزايا استخداماتها في:[1]

- سهولة العمليات الحسابية المستخدمة،
- لا تحتاج إلى شروط كثيرة لذلك فإن إمكانية إساءة استعمالها تعتبر قليلة،
- تستخدم عندما لا تتحقق الشروط اللازمة لتطبيق الاختبارات المعلمية،
- تستخدم في حالة صعوبة الحصول على بيانات دقيقة،
- لا يتطلب استخدامها معرفة دقيقة في مجال الرياضيات والإحصاء،
- لا تشترط كبر حجك العينات.

3-4-1 اختبار كاي-مربع

تعطى قيمة إحصاء الاختبار $\chi^2_{\text{المحسوبة}}$ بالعلاقة التالية:

[1]- محفوظ جودة، التحليل الإحصائي المتقدم باستخدام SPSS، دار وائل، الأردن، 2008، ص 207.

$$\chi^2_{\text{المحسوبة}} = \sum \frac{(f_0 - f_e)^2}{f_e}$$

بحيث:

f_0: تمثل التكرارات المشاهدة،

f_e: تمثل التكرارات المتوقعة (النظرية)،

فإذا كانت $\chi^2_{\text{المحسوبة}}$ أكبر من $\chi^2_{\text{الجدولية}}$ عند مستوى معنوية ودرجة الحرية، فإننا نرفض فرض العدم لصالح الفرض البديل.

يستخدم توزيع كاي-مربع لاختبار:

3-4-1-1 مقارنة التكرارات المشاهدة والمتوقعة

في الحياة العملية، نجد دائما أن القيم المشاهدة لا تكون مساوية للقيم المتوقعة (النظرية). وهذا لأننا نقوم بتجارب عشوائية، لا تكون نتيجتها معروفة مسبقا. وتختلف نتائجها باختلاف التجربة العشوائية المنجزة. مثلا، عند رمي قطعة نقود متزنة 100 مرة. فإننا نتوقع أن نجد 50 صورة بالضبط. غير أنه في النادر ما نحصل على هذه النتيجة. على هذا الأساس نهتم دائما في الإحصاء باختبار الفرضيات ما إن كان الفرق الموجود بين القيم الشاهدة والقيم المتوقعة معنوي. أي أنه راجع لاختلاف في خصائص الظاهرة، وليس اختلاف راجع إلى تعاملنا مع العلوم التجريبية التي تكون نتائجها عموما غير متساوية بالضبط.

ونستخدم اختبار كاي-مربع لمقارنة التكرارات المشاهدة والمتوقعة تحت الفرضية التالية:

H_0 : التكرارات المشاهدة لا تختلف معنويا عن التكرارات المتوقعة

H_1 : التكرارات المشاهدة تختلف معنويا عن التكرارات المتوقعة

وتعطى درجة حرية الاختبار بالعلاقة: $df=c-1$ بحيث c تمثل عدد الفئات.

مثال رقم 3−10: وجد محل تجاري من خبرته الماضية أن 30% من التلفزيونات المباعة من الحجم الصغير، 40% متوسط و30% كبير. لتحديد حجم المخزون الواجب الاحتفاظ به، أخذ المدير عينة عشوائية من 100 من المبيعات الحديثة فوجد أن منها 40 صغير، 40 متوسط و20 كبير.

المطلوب: باستخدام مستوى معنوية 5%، هل يمكن اعتبار أن نمط المبيعات في الماضي لا زال سائدا.

الحل:

− تلخيص المعطيات:

الإجمالي	حجم الشاشة			النمط
	صغير	متوسط	كبير	
100	40	40	20	f_0: التكرارات المشاهدة
100	30	40	30	f_e: التكرارات المتوقعة

− نضع الفرض:

التكرارات المشاهدة لا تختلف معنويا عن التكرارات المتوقعة : H_0

التكرارات المشاهدة تختلف معنويا عن التكرارات المتوقعة : H_1

− تحديد قاعدة القرار :

بمستوى معنوية قدره 5% ودرجة حرية قدرها 3−1=2، فإن منطقة الرفض تكون إلى يمين القيمة الجدولية 5.991.

− اتخاذ القرار

تحديد قيمة إحصاء الاختبار $\chi^2_{\text{المحسوبة}}$ بالعلاقة التالية:

$$\chi^2_{\text{المحسوبة}} = \sum \frac{(f_0 - f_e)^2}{f_e} = \frac{(20-30)^2}{30} + \frac{(40-40)^2}{40} + \frac{(40-30)^2}{30} = 5.83$$

بما أن قيمة $\chi^2_{\text{المحسوبة}}$ أقل من $\chi^2_{\text{الجدولية}}$ فإننا نقبل فرض العدم بأن التكرارات المشاهدة لا تختلف معنويا عن التكرارات المتوقعة عند مستوى معنوية 5%. وبالتالي فإن نمط المبيعات في الماضي لا زال سائدا، ويمكن لصاحب المحل أن يعتمد عليه في تجديد المخزون من هذا المنتج.

ملاحظة: إذا كانت التكرارات المتوقعة في أي فئة أقل من 5، فإنه يجب ضمها للفئة المجاورة.

3-4-1-2 تحديد طبيعة توزيع بيانات

يستخدم توزيع كاي–مربع لمعرفة إن كانت بيانات العينة المسحوبة تتبع توزيع ما (ذي الحدين، طبيعي، بواسوني، هندسي، فوق الهندسي...) تحت الفرضية التالية:

H_0 : البيانات تتبع التوزيع

H_1 : البيانات لا تتبع التوزيع

وهذا عن طريق الخطوات التالية:

- تحليل التوزيع المشاهد،
- اختيار التوزيع المناسب،
- باستخدام بيانات العينة نقوم بتقدير معالم مجتمع نجري الاختبار من خلالها.

حيث تعطى درجة حرية الاختبار بالعلاقة: $df=c-m-1$ بحيث: c تمثل عدد الفئات و m تمثل عدد معالم المجتمع المقدرة من إحصائيات العينة.

مثال رقم 3-11: يمثل لنا الجدول الموالي توزيع القبول لـ 100 طالب في 3 كليات.

عدد الطلاب	25	34	31	10
عدد مرات القبول	0	1	2	3

المطلوب: إذا كان احتمال قبول الطالب في كلية ما هو 0.4. هل يمكن أنيتبع توزيع القول توزيع ذي الحدين عند مستوى معنوية قدره 5%؟

الحل:

– وضع الفرض:

البيانات تتبع توزيع ثنائي الحد : H_0

البيانات لا تتبع التوزيع : H_1

– احتمالات ذي الحدين المناظرة لعدد مرات قبول 0، 1، 2 و3 لأي طالب عند p=0.4 في الجدول الموالي:

التكرار المتوقع	عدد المتقدمين	احتمالات ذي الحدين	التكرار المشاهد	عدد مرات القبول
21.6	100		25	0
43.2	100		34	1
28.8	100	$c_3^2(0.4)^2(0.6)^1 = 0.288$	31	2
6.4	100	$c_3^0(0.4)^0(0.6)^3 = 0.064$	10	3
100		1.000	100	مجموع

– تحديد قاعدة القرار

عند مستوى معنوية قدره 5% ودرجة حرية قدرها df=c−m−1=4−0−1=3، فإن منطقة الرفض تكون إلى يمين قيمة $\chi^2_{\text{الجدولية}}$ التي تساوي 7.815.

– اتخاذ القرار

تحديد قيمة إحصاء الاختبار $\chi^2_{\text{المحسوبة}}$ بالعلاقة التالية:

$$\chi^2_{\text{المحسوبة}} = \sum \frac{(f_0 - f_e)^2}{f_e} = \frac{(25-21.6)^2}{21.6} + \frac{(34-43.2)^2}{43.2} + \frac{(31-28.8)^2}{28.8} + \frac{(10-6.4)^2}{6.4}$$
$$= 4.687$$

بما أن قيمة $\chi^2_{\text{المحسوبة}}$ أقل من $\chi^2_{\text{الجدولية}}$ فإننا نقبل فرض العدم بأن البيانات تتبع التوزيع ثنائي الحد عند مستوى معنوية 5%. وبالتالي توزيع قبول الطلبة إلى الكليات يتبع توزيع ثنائي الحد.

3-4-1-3 اختبارات الاستقلال

تهدف اختبارات الاستقلال إلى اختبار فيما إذا كان متغيرين مستقلان أم لا تحت الفرضية التالية:

H_0 : المتغيرين مستقلين

H_1 : المتغيرين ليس مستقلين

يطلق أيضا على هذا الاختبار باختبار جداول الاقتران، هذا عادة لأن المتغيرين المراد اختبار استقلالهما يدرجان في جدول يسمى بجدول الاقتران الموضح فيما يلي:

	C_1	C_1	…	C_k	sum
r_1	n_{11}	n_{12}	…	n_{1k}	$n_{1.}$
r_2	n_{21}	n_{22}	…	n_{1k}	$n_{1.}$
…	…	…	…	…	…
r_k	n_{m1}	n_{12}	…	n_{mk}	$n_{1.}$
sum	$n_{.1}$	$n_{.2}$	…	$n_{.k}$	$n_{..}$

يتطلب هذا الاختبار إيجاد قيم التكرار المتوقع في كل خلية من خلايا جدول الاقتران وهذا باعتماد العلاقة التالية:

$$f_e = \frac{n_{.1} \times n_{1.}}{n_{..}}$$

درجة حرية هذا الاختبار تعطى بالعلاقة: $df=(C_k-1)(r_k-1)$

لإجراء هذا الاختبار يجب أن تكون كل المشاهدات مستقلة. وهذا يقتضي أن يظهر العنصر مرة واحدة فقط في الجدول. كذلك لا يجب أن تكون فئات المتغير متراكبة. على سبيل المثال، لا يمكن ان تستخدم الفئات العمرية التالية: أقل من 30، ومن 25 إلى 40.

ملاحظة: إذا كان $n_{..}$ أقل من 50، فإننا نستخدم تصحيح ييت وفق الصيغة التالية لحساب إحصاء الاختبار:

$$\chi^2_{\text{المحسوبة}} = \sum \frac{(|f_0 - f_e| - 0.5)^2}{f_e}$$

مثال رقم 3-12: طرح استبيان ABC news عددا من الأسئلة منها: ما هو الأهم بالنسبة للمجتمع: السماح بحرية التعبير للأشخاص حتى إذا كانت تعني قبول العادات السيئة. أم تدعيم العادات الجيدة حتى إذا كانت تعني الحد من حرية التعبير؟

يبين الجدول الموالي إجابات أفراد العينة على هذا السؤال موزعة حسب الجنس

المجموع	أهمية حرية التعبير للمجتمع			
	تدعيم العادات الجيدة/الحد من حرية التعبير	السماح بحرية التعبير/قبول العادات السيئة		
463	195	268	ذكور	الجنس
472	240	232	إناث	
935	435	500	المجموع	

نلاحظ أن 268 (57.9%) من الذكور يفضلون حرية التعبير مقارنة مع حوالي 232 (49.2%) من النساء.

المطلوب: اعتمادا على هذه النتائج، هل تعتقد أنه في مجتمع الولايات المتحدة، للرجال مواقف تختلف عن مواقف النساء حول الأهمية النسبية لحرية التعبير عند مستوى معنوية 5% (هل جنس المشارك وإجابته غير مستقلين)؟

الحل:

- في هذه العينة نلاحظ أن نسبة الرجال أكبر من النساء في دعم حرية التعبير. ولكن كالمعتاد، ليست نتائج العينة هي التي تهمنا، إذ أننا نريد معرفة ما يمكننا استنتاجه حول المجتمع اعتمادا على نتائج العينة المشاهدة. وهذا تحت الفرضية التالية:

H_0 : لا يختلف رأي الرجال والنساء بأن حرية التعبير أكثر أهمية من العادات الجيدة (المتغيرين مستقلان)

H_1 : يختلف رأي الرجال والنساء بأن حرية التعبير أكثر أهمية من العادات الجيدة (المتغيرين غير مستقلين)

- إنشاء جدول التكرار المتوقع:

		أهمية حرية التعبير للمجتمع		
		السماح بحرية التعبير/قبول العادات السيئة	تدعيم العادات الجيدة/الحد من حرية التعبير	المجموع
الجنس	ذكور	$\frac{500 \times 463}{935} = 247.6$	$\frac{435 \times 463}{935} = 215.4$	463
	إناث	$\frac{500 \times 472}{935} = 252.4$	$\frac{435 \times 472}{935} = 219.6$	472
المجموع		500	435	935

– تحديد قاعدة القرار

عند مستوى معنوية 5% ودرجة حرية $df=(C_k-1)(r_k-1)=(2-1)(2-1)=1$ فإن منطقة الرفض تقع إلى يمين قيمة $\chi^2_{\text{الجدولية}}$ التي تساوي 3.841.

– تحديد قيمة إحصاء الاختبار $\chi^2_{\text{المحسوبة}}$ بالعلاقة التالية:

$$\chi^2_{\text{المحسوبة}} = \sum \frac{(f_0 - f_e)^2}{f_e}$$
$$= \frac{(268 - 247.6)^2}{247.6} + \frac{(195 - 215.4)^2}{215.4} + \frac{(232 - 252.4)^2}{252.4} + \frac{(240 - 219.6)^2}{9.6} = 7.16$$

بما أن قيمة $\chi^2_{\text{المحسوبة}}$ أكبر من $\chi^2_{\text{الجدولية}}$ فإننا نرفض فرض العدم بأن الرجال والنساء أجابوا على السؤال بنفس الطريقة. ويبدوا أن النساء يفضلن العادات الجيدة على حرية التعبير أكثر من الرجال.

ويعني الاستقلال أن معرفة قيمة احد المتغيرين من اجل حالة معينة لا يخبرنا أي شيء عن قيمة المتغير الآخر. وحيث أننا قبلنا الفرض البديل بأن جنس المشارك والأهمية التي يعطيها لحرية التعبير هما متغيرين غير مستقلين. بالتالي، إذا عرفت أن المشارك هي أنثى، فإننا نعرف بأن احتمال تفضيلها لحرية التعبير على العادات الجيدة هي أقل مما لو كان المشارك ذكرا.

3-4-2 اختبار تحليل التباين

يستخدم تحليل التباين لاختبار فرض العدم القائل بأن عددا من قيم المتوسط للمجتمعات المستقلة هي متساوية. والتقنية التي نستخدمها هنا هي تحليل التباين والمعروفة اختصارا باسم ANOVA.

وهناك عدد من الفرضيات المطلوبة لتحليل التباين هي:

– سحب عينات عشوائية مستقلة من كل مجتمع، حيث يتم النظر إلى المجموعات التي تتم مقارنتها كمجتمعات متمايزة واضحة المعالم،

– الاستقلال: يعني عدم وجود أي علاقة بين المشاهدات في المجموعات المختلفة، وبين المشاهدات في نفس المجموعة،

– البيانات تكون تتبع التوزيع الطبيعي،

– تساوي قيم التباين.

يجرى اختبار تحليل التباين تحت الفرضية التالية:

$$H_0: \mu_0 = \mu_1 = \mu_1 = \cdots \mu_k = 0$$

$$H_1: \mu_0 \neq \mu_1 \neq \mu_1 \neq \cdots \mu_k \neq 0$$

بينما تعطى إحصاء الاختبار $F_{\text{المحسوبة}}$ (فيشر المحسوبة) من خلال جدول ANOVA التالي:

جدول تحليل التباين ANOVA

مصدر التغير	مجموع المربعات	درجة الحرية	متوسط المربعات	إحصائية F
بين المجموعات	$SSR = r\sum(\bar{X}_j - \bar{\bar{X}})^2$	c−1	$MSR = \dfrac{SSR}{c-1}$	$F = \dfrac{MSR}{MSE}$
داخل المجموعات	$SSE = \sum\sum(x_{ij} - \bar{X}_j)^2 = \sum(n_j - 1)\, S_j^2$	c(r−1)	$MSE = \dfrac{SSR}{c(r-1)}$	
المجموع		cr−1	–	

بحيث:

c: عدد العينات،

r: عدد المشاهدات في كل عينة،

SSR: مجموع مربع الفرق بين المجموعات،

SSE: مجموع مربع الفرق داخل المجموعات،

MSR: متوسط المربعات بين المجموعات،

MSE: متوسط المربعات بين المجموعات،

$c-1$: درجة حرية البسط،

$c(r-1)$: درجة حرية المقام.

إذا كانت $F_{\text{المحسوبة}}$ أكبر من $F_{\text{الجدولية}}$ عند مستوى معنوية ودرجة الحرية، فإننا نرفض فرض العدم لصالح الفرض البديل.

تحليل التغيرية:

في تحليل التباين، يتم تقسيم التغيرية (تجزئة التغيرية) المشاهدة في العينة إلى قسمين:

التغيرية ضمن المجموعات (within groups): هي قيمة تقديرية للتباين اعتمادا على مقدار التغير للمشاهدات ضمن مجموعة معينة. ويساهم تباين العينة لكل مجموعة في القيمة التقديرية للتغيرية ضمن المجموعات،

التغيرية بين المجموعات (Between groups): لدينا متوسط العينة لكل مجموعة من المجموعات في الدراسة. إذا كان لجميع المجموعات نفس العدد من الحالات، فإنه يمكننا أن نجد الانحراف المعياري لقيم المتوسط للعينات، الذي يدل على مقدرا تغير قيم المتوسط للعينات القادمة من نفس المجتمع.

لدينا الآن قيمتان تقديريتان لمدى تغير المشاهدات ضمن مجوعة معينة. وتختلف هاتان القيمتان التقديريتان بشكل كبير، إذ تكون القيمة التقديرية للتباين بين المجموعات صحيحة فقط عندما تكون فرضية العدم صحيحة. إذا كانت فرضية العدم خاطئة، فإن القيمة التقديرية للتباين بين المجموعات ستكون كبيرة جدا. وستكون التغيرية المشاهدة لقيم المتوسط للعينات هي نتيجة عاملين: التغيرية للمشاهدات، ضمن مجموعة ما، والتغيرية لقيم المتوسط للمجموعات.

ملاحظات:

– إذا كانت حجم العينات متساوي في كل المجموعات، فيمكن إجراء تحليل التباين في ظل غياب فرضية التوزيع الطبيعي للبيانات مع تساوي التباين،

– إن رفض فرضية العدم، لا يعني بالضرورة أن متوسطات كل المجموعات مختلفة، بل على الأقل يوجد مجتمعين لهما متوسطين مختلفين.

مثال رقم 3-13: تنتج مؤسسة NCP طابعات وأجهزة فاكس في ثلاث وحدات إنتاجية متوطنة في كل من: أطلنطا، دالاس وسياتل. من أجل متابعة معرفة الموظفين لنظام الجودة في التسيير. تم سحب عينة عشوائية من 6 موظفين للإجابة على استبيان في هذا الشأن نعرضها في الجدول الموالي:

المشاهدات	أطلنطا	دالاس	سياتل
1	85	71	59
2	75	75	64
3	82	73	62
4	76	74	69
5	71	69	75
6	85	82	67
متوسط العينة	79	74	66
تباين العينة	34	20	32
الانحراف المعياري للعينة	5.83	4.47	5.66

المطلوب: هل تقييم الموظفين متساوي في المتوسط بين الوحدات الثلاث عند مستوى معنوية 5%؟

الحل:

نضع الفرضية:

$$H_0: \mu_0 = \mu_1 = \mu_1 = 0$$

$$H_1: \mu_0 \neq \mu_1 \neq \mu_1 \neq 0$$

– إعداد جدول تحليل التباين

حساب كل من:

$$\bar{\bar{X}} = \frac{\sum\sum x_{ij}}{n} = \frac{\sum \bar{X}_j}{k} = \frac{79+74+66}{3} = 73$$

$$SSR = r\sum\left(\bar{X}_j - \bar{\bar{X}}\right)^2 = 6[(79-73)^2 + (74-73)^2 + (76-73)^2] = 516$$

$$SSE = \sum\sum\left(x_{ij} - \bar{X}_j\right)^2 = \sum(n_j - 1)\,S_j^2 = (6-1)34 + (6-1)20 + (6-1)32 = 430$$

جدول تحليل التباين ANOVA

مصدر التغير	مجموع المربعات	درجة الحرية	متوسط المربعات	إحصائية F
بين المجموعات	516	3−1=2	$MSR = \frac{516}{2} = 258$	$F = \frac{258}{28.67} = 8.99$
داخل المجموعات	430	c(r−1)=3(6−1)=15	$MSE = \frac{430}{15} = 28.67$	
المجموع		cr−1=3*6−1=17	−	

– تحديد قاعدة القرار

عند مستوى معنوية قدره 5%، ودرجة حرية بسط قدرها 2، ودرجة حرية مقام قدرها 15، فإن منطقة الرفض تكون إلى يمين قيمة $F_{\text{الجدولية}}$ 3.68.

– اتخاذ القرار

بما أن قيمة $F_{\text{المحسوبة}}$ أكبر من $F_{\text{الجدولية}}$ فإننا نرفض فرض العدم القائل بتساوي متوسطات المجتمعات ونقبل الفرض البديل. وبالتالي فإن نتائج الاستبيان المعطى لموظفي الوحدات الإنتاجية الثلاث هي غير متساوية.

3-5 قوة الاختبار ومنحنى توصيف العمليات

3-5-1 قوة الاختبار

القوة (Power) هي مصطلح إحصائي يستخدم لوصف القدرة على رفض فرضية العدم عندما تكون خاطئة. وهي احتمال يتراوح بين 0 و1. وكلما كانت القدرة أكبر، كان من المرجح رفض فرضية العدم عندما تكون خاطئة. وتعتمد القدرة على كبر الفرق الحقيقي، وعلى حجم العينة، وعلى التباين للفرق، وعلى مستوى المعنوية الذي نكون عنده مستعدين لرفض فرضية العدم.

وتدخل جميع العوامل المذكورة في إحصاء الاختبار، التي تحدد فيما إذا كنا سنرفض فرضية العدم. إذا كان البسط كبيرا والمقام صغيرا، فإننا سنحصل على قيمة كبيرة لإحصاء الاختبار. ويكون البسط في الصيغة التي نستخدمها لحساب إحصاء الاختبار من الفرق بين القيمتين. ولذلك، كلما كان الفرق أكبر، كان من الأرجح أن يكون البسط أكبر. أما المقام في الصيغة التي تحسب إحصاء الاختبار فيعتمد على قيم التباين للمجموعات. وسيكون التباين صغيرا إذا كانت قيم التباين للعينات صغيرا، وكان حجم العينة كبيرا.

مثال رقم 3−14: يعرف مركز تجنيد بالجيش من الخبرة الماضية أن وزن المجند يتبع التوزيع الطبيعي بوسط يساوي 80 كلغ وانحراف يساوي 10 كلغ. ويرغب مركز التجنيد في أن يختبر عند مستوى معنوية 1%، ما إذا كان متوسط مجندي هذه السنة أكبر من 80 كلغ. لعمل هذا، تم أخذ عينة عشوائية من 25 مجندا وجد متوسطها 85 كلغ.

المطلوب:

1− كيف يمكن إجراء هذا الاختبار؟

2− أوجد احتمال رفض فرض العدم إذا كانت:

$$\mu = \mu_0 = 80, \mu = 82, \mu = 84, \mu = 85, \mu = 86, \mu = 87$$

3− مثل بيانيا نتائج السؤال 2، مبينا على المحور الرأسي احتمال رفض فرض العدم للقيم المختلفة عندما يكون $\mu > \mu_0$ ؟

4− ماذا يوضح هذا الشكل؟

الحل:

1− اختبار الرفض

− وضع الفرض

$$H_0: \mu \leq 80$$
$$H_1: \mu > 80$$

- قاعدة القرار

عند مستوى معنوية 1% فإن منطقة الرفض تكون إلى يمين القيمة $Z_{\text{الجدولية}}$ التي تساوي 2.33.

- اتخاذ القرار

$$Z_{\text{المحسوبة}} = \frac{85-80}{10/\sqrt{25}} = 2.5$$

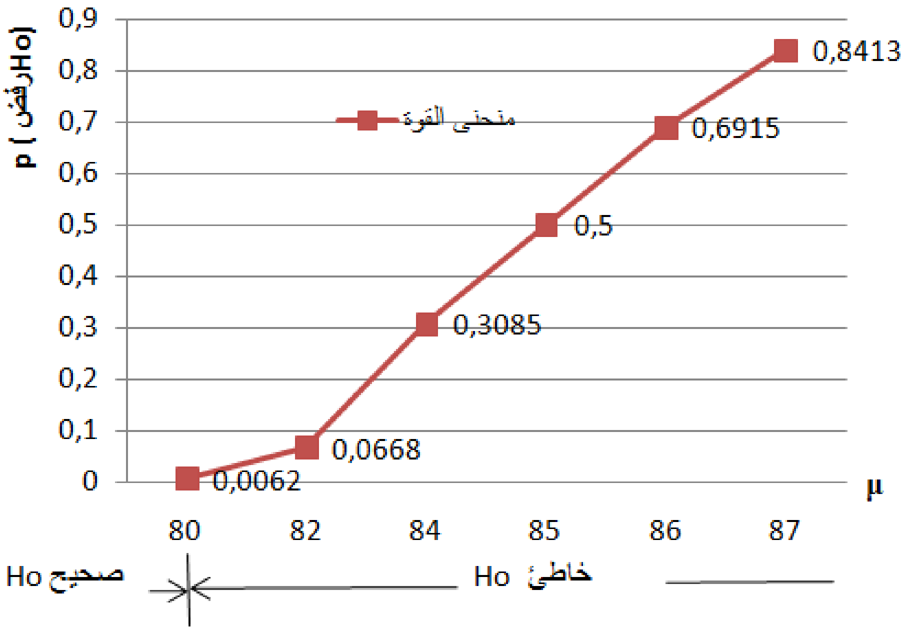
حيث أن قيمة $Z_{\text{المحسوبة}}$ تقع عند منطقة الرفض، فإننا نرفض فرض العدم ونقبل الفرض البديل، وبالتالي فإن متوسط مجندي هذه السنة هو أكبر من 80 كلغ.

بصيغة أخرى، يعني هذا أنه إذا كان متوسط المجتمع هو 80 كلغ. فإن احتمال أن عينة عشوائية مأخوذة من هذا المجتمع تعطي متوسطا قدره 85 كلغ هو أقل من 1%.

2- إيجاد احتمال رفض فرض العدم.

P(رفض فرض العدم)=	Z=	$Z=\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$	μ
0.0062	2.5	$\frac{85-80}{10/\sqrt{25}}$	80
0.0668	1.5	(85−82)/2	82
0.3085	0.5	(85−84)/2	84
0.5000	0	(85−85)/2	85
0.6915	−0.5	(85−86)/2	86
0.8413	−1	(85−87)/2	87

3- التمثيل البياني:



4- يوضح هذا الشكل منحنى القوة، حيث نلاحظ أنه كلما زادت قيمة μ الحقيقية عن μ_0، كلما زادت قوة الاختبار (أي كلما زاد احتمال رفض فرض خاطئ).

3-5-2 منحنى توصيف العليات

يوضح منحنى توصيف العلميات قيم الخطأ من النوع الثاني عند القيم المختلفة عندما يكون $\theta > \theta_0$. ويعتبر هذا المنحى مهم جدا من الناحية العلمية، خصوصا ضمن التجارب التي يؤدي قبول فرض خاطئ فيها إلى نتائج سلبية (على سبيل المثال، قبول فرض دواء على أنه فعال في حين أنه ليس كذلك). في هذه الحالات، فإننا نرغب أن نبقي قيم الخطأ من النوع الثاني صغيرة، حتى لو كان علينا قبول قيم مرتفعة للخطأ من النوع الأول.

مثال رقم 3-15: في المثال رقم 3-14.

1- أوجد احتمال قبول فرض العدم إذا كانت:

$$\mu = \mu_0 = 80, \mu = 82, \mu = 84, \mu = 85, \mu = 86, \mu = 87$$

2- مثل بيانيا نتائج السؤال 2، مبينا على المحور الرأسي احتمال رفض فرض العدم للقيم المختلفة عندما يكون $\mu > \mu_0$ ؟

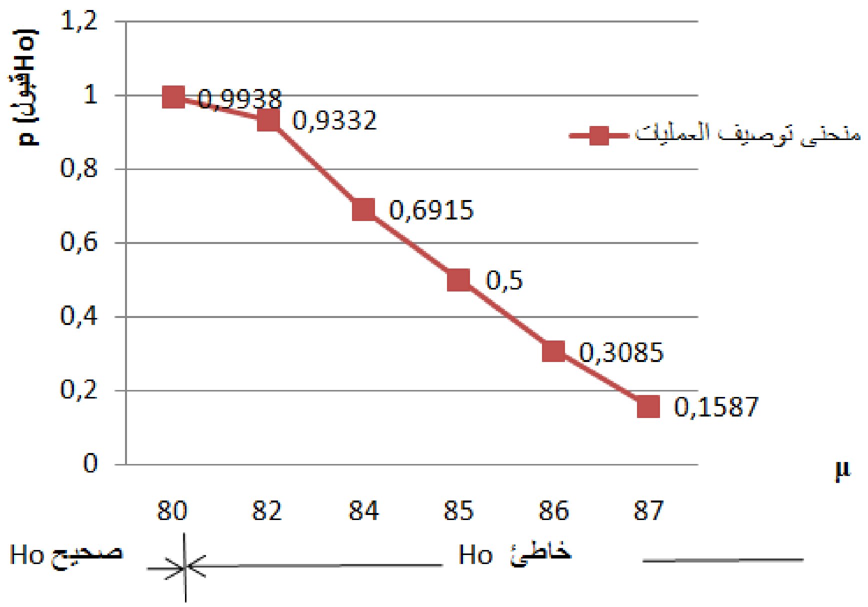
3- علق على الشكل؟

الحل:

1- إيجاد احتمال قبول فرض العدم.

μ	$Z=\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$	Z=	P(قبول فرض العدم)=
80	$\frac{85-80}{10/\sqrt{25}}$	2.5	0.9938
82	(85−82)/2	1.5	0.9332
84	(85−84)/2	0.5	0.6915
85	(85−85)/2	0	0.5
86	(85−86)/2	−0.5	0.3085
87	(85−87)/2	−1	0.1587

2- التمثيل البياني:



3- التعليق: يوضح هذا الشكل منحنى توصيف العلميات، حيث نلاحظ انه كلما زادت قيمة µ الحقيقة عن μ_0، كلما صغرت قيمة الخطأ من النوع الثاني.

3-6 تمارين محلولة

تمرين رقم 3-1:

1 – ما هو احتمال أن نحصل على مابين 40 و 60 مرة صورة ضمنيا في 100 رمية لقطعة نقد متجانسة؟

2 – لاختبار فرضية ما إذا كانت القطعة النقدية متجانسة فعلا فإننا نتبع قاعدة القرار التالية: نقبل الفرضية إذا كان عدد مرات ظهور الصورة هو بين 40 و 60، ونرفض الفرضية فيما عدا ذلك.

أ – ما هو احتمال رفض الفرضية عندما تكون صحيحة فعليا؟

ب – مثل بيانيا وفسر قاعدة القرار ونتيجة الجزء أ؟

ج – ما النتائج التي يمكنك استخلاصها إذا أدت العينة إلى ظهور الصورة 53 مرة؟

الحل:

1- حساب احتمال حصول على مابين 40 و 60 مرة صورة في 100 رمية لقطعة نقد متجانسة.

لدينا:

$$n = 100, p = \frac{1}{2}, q = \frac{1}{2}$$

إذن:

$$p(40 \leq x \leq 60) = \sum_{i=40}^{60} p(x = x_i)$$

نستخدم التوزيع الطبيعي كتقريب لتوزيع ثنائي الحد، ويكون لدينا:

$$\mu = np = 100\frac{1}{2} = 50, \quad \sigma = \sqrt{npq} = \sqrt{100\frac{1}{2}\frac{1}{2}} = 5$$

وبالتالي الاحتمال يساوي:

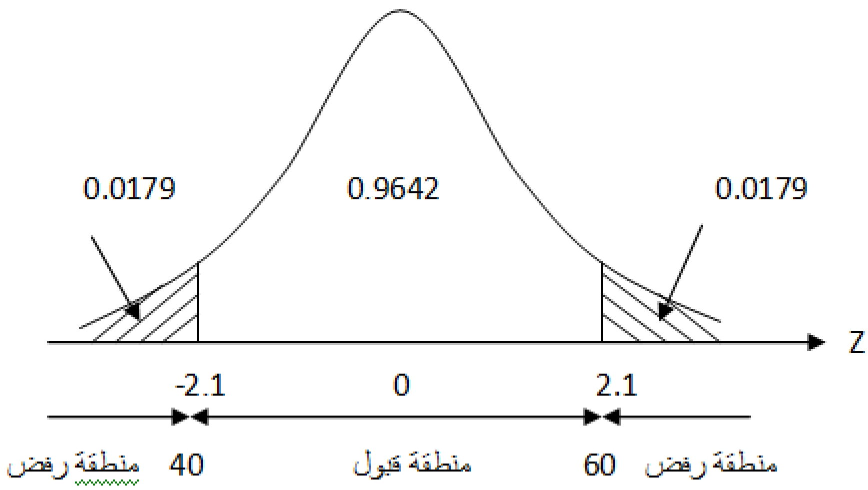
$$p(39.5 < x < 60.5) = p\left(\frac{39.5-50}{5} < Z < \frac{60.5-50}{5}\right)$$

$$p(-2.1 < Z < 2.1) = 0.9642$$

-2

أ- من 1 نجد أن احتمال ظهور الصورة بين 40 و60 مرة هو 0.9642، إذا كانت القطعة النقدية متجانسة، وبالتالي فإن عدم ظهور الصورة بين 40 و60 مرة هو 1-0.9642=0.0358.

ب- التمثيل البياني



التفسير: إذا أنتجت عينة مؤلفة من 100 رمية النقاط Z الواقعة بين -2.1 و2.1، فإننا نقبل الفرضية، بأن القطعة النقدية متجانسة، ونرفضها فيما عدا ذلك، ونقرر بأن القطعة النقدية غير متجانسة.

الخطأ المرتكب عند رفض الفرضية في حين يجب أن تقبل هو الخطأ من النوع الأول، أما احتمال ارتكاب هذا الخطأ هو 0.0358 (المساحة المظللة) وبالتالي فإن مستوى معنوية القرار هو 3.58%.

ج- وفقا لقاعدة القرار فإننا نقبل الفرضية بأن القطعة النقدية متجانسة.

تمرين رقم 3-2: ينتج أحد المصانع نوع من الطوب صلابته تقدر بـ 10ن. ويرغب صاحب المصنع في أن يختبر ما إذا كانت صلابة الطوب الذي ينتجه هي بالضبط 10ن. لأن صلابة أقل لا تكون ملائمة،

وصلابة أكثر تعني تكلفة زائدة للمصنع. لغرض ذلك، أخذ المنتج عينة حجمها 50، وجد أن متوسطها هو 11 ن وانحرافها المعياري هو 2.

المطلوب:

1- قم بإجراء هذا الاختبار عند مستوى معنوية قدره 5%؟

2- أوجد منطقتي القبول والرفض للسؤال 1 مع التمثيل البياني؟

الحل:

1- اختبار الرفض

المعطيات:

$$n = 50, \bar{X} = 11, \mu = 10, \sigma = 2$$

- وضع الفرض

$$H_0: \mu = 10$$
$$H_1: \mu \neq 10$$

- قاعدة القرار

عند مستوى معنوية 5% فإن منطقة الرفض تكون بين $Z_{\text{الجدولية}}$ التي تساوي ± 1.96

- اتخاذ القرار

$$Z_{\text{المحسوبة}} = \frac{11 - 10}{2/\sqrt{50}} = 3.53$$

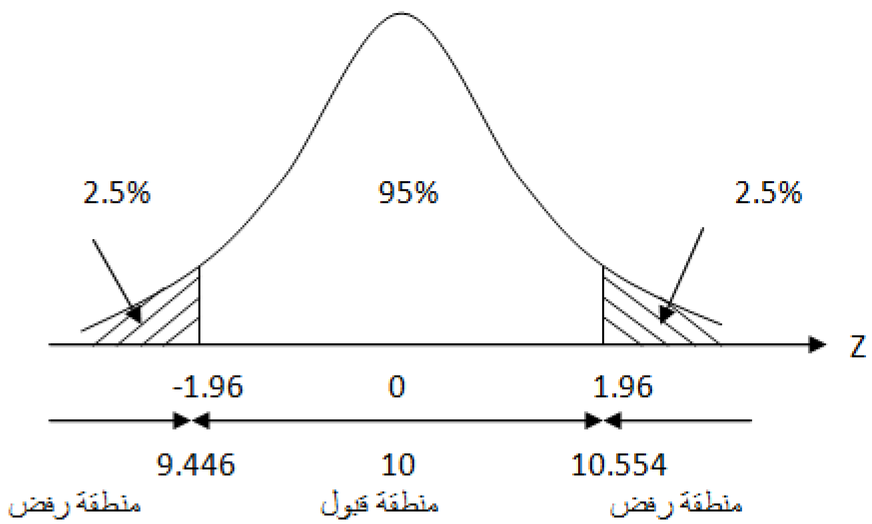
حيث أن قيمة $Z_{\text{المحسوبة}}$ تقع عند منطقة الرفض، فإننا نرفض فرض العدم ونقبل الفرض البديل، وبالتالي فإن متوسط صلابة الطوب المصنوع في هذا المصنع لا يساوي 10ن.

2- تحديد منطقتي القبول والرفض

لإيجاد منطقتي القبول (عند مستوى معنوية 5%) بالنيوتن. فإننا نقوم بعملية تقدير من خلال إيجاد فترة ثقة 95% حول وسط المجتمع المجهول μ_0.

$$\mu_0 \pm Z_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}} = 10 \pm 1.96 \frac{2}{\sqrt{50}} = 10 \pm 0.554$$

أي أنه لقبول Ho عند مستوى معنوية 5%، فإن المتوسط الحسابي للعينة يجب أن يكون أكبر من 10.554ن أو أقل من 9.446ن.



تمرين رقم 3-3:

يدعي متحدث لمكافحة التلوث أن أكثر من 80% من المصانع في المنطقة تستوفي معايير مكافحة التلوث، ولكن واحدة من أنصار مكافحة التلوث لا تصدق هذا الإدعاء، فتأخذ عينة عشوائية من البيانات المنشورة عن مكافحة التلوث في 64 مصنعا في المنطقة، وتجد أن منها 56 مصنعا تستوفي معايير المكافحة.

1 – هل تؤيد بيانات العينة إدعاء المتحدث عند مستوى معنوية قدره 5%؟

2 – هل يتغير القرار إذا كان حجم العينة 124 مع بقاء نسبة المصانع التي تستوفي المعايير كما كانت؟ ماذا تلاحظ؟

الحل:

1- اختبار الرفض

المعطيات:

$$p = 0.80, \bar{p} = \frac{56}{64}$$

- وضع الفرض

$$H_0: p \geq 0.8$$
$$H_1: p > 0.8$$

- قاعدة القرار

عند مستوى معنوية 5% فإن منطقة الرفض تكون إلى يمين $Z_{\text{الجدولية}}$ التي تساوي 1.65.

- اتخاذ القرار

$$Z_{\text{المحسوبة}} = \frac{\bar{p} - p_0}{\sqrt{\frac{pq}{n}}} = \frac{0.875 - 0.80}{\sqrt{\frac{0.8 \times 0.2}{64}}} = 1.5$$

حيث أن قيمة $Z_{\text{المحسوبة}}$ تقع عند منطقة القبول، فإننا نقبل فرض العدم ونرفض الفرض البديل، وبالتالي فإنه لا يوجد أي سند إحصائي لادعاء هذا المتحدث بأن $p > 0.8$.

2- تغيير حجم العينة، مع بقاء نسبة المصانع التي تستوفي شروط على حالها، تصبح قيمة $Z_{\text{المحسوبة}}$:

$$Z_{\text{المحسوبة}} = \frac{\bar{p} - p_0}{\sqrt{\frac{pq}{n}}} = \frac{0.875 - 0.80}{\sqrt{\frac{0.8 \times 0.2}{124}}} = 1.875$$

حيث أن قيمة $Z_{\text{المحسوبة}}$ تقع عند منطقة الرفض، فإننا نرفض فرض العدم ونقبل الفرض البديل، وبالتالي فإنه لا يوجد أي دليل ضد إدعاء المتحدث بأن $p > 0.8$.

الملاحظة: نلاحظ أنه مع زيادة حجم العينة وثبات العوامل الأخرى، قد زاد من احتمال قبول إدعاء المتحدث.

تمرين رقم 4-4:

قام أحد الباحثين باستقصاء أراء العاملين بإحدى الشركات من خلال طرح السؤال التالي ضمن الاستقصاء: إذا أتيح لك العمل في شركة أخرى بنفس الأجر الذي تتقاضاه حاليا وبنفس العلاوات والمكافآت المالية التي تحصل عليها، هل تغير الشركة التي تعمل بها. وكانت الإجابات بنعم هي 4 وبلا هي 96.

إذا علمت أن عدد العمال بهذه الشركة هو 600 عامل.

المطلوب:

هل يستطيع الباحث الدفاع عن فرضه، بأن 10% من العاملين في الشركة لا يرغبون في الاستمرار في العمل فيها بمستوى معنوية 0.05.

الحل:

المعطيات:

$$p = 0.10, \bar{p} = \frac{4}{96+4} = 0.04, N = 600$$

- وضع الفرض

$$H_0: p = 0.1$$
$$H_1: \mu \neq 0.1$$

- قاعدة القرار

عند مستوى معنوية 5% فإن منطقة الرفض تكون ضمن $Z_{\text{الجدولية}}$ التي تساوي± 1.96.

- اتخاذ القرار

$$Z_{\text{المحسوبة}} = \frac{\bar{p} - p_0}{\sqrt{\frac{pq}{n}\frac{N-n}{N-1}}} = \frac{0.04 - 0.10}{\sqrt{\frac{0.04 \times 0.96}{64}\frac{600-100}{600-1}}} = 3.33$$

حيث أن قيمة $Z_{\text{المحسوبة}}$ تقع عند منطقة الرفض، فإننا نرفض فرض العدم ونقبل الفرض البديل، وبالتالي لا يمكن للباحث أن يدافع عن فرضه بأن 10% من العاملين في الشركة لا يرغبون في الاستمرار في العمل بها.

تمرين رقم 4-5:

موزع آلي للمشروبات، يقوم بتقديم القهوة في كؤوس بلاستيكية بمتوسط قدره 15 سل. أجريت دراسة على هذه الآلة، حيث أكدت أن المحتوى الحقيقي يتغير من كأس بلاستيكي لآخر، ويمكن أن يعتبر متغير عشوائي يتبع التوزيع الطبيعي بوسط قدره μ وانحراف معياري قدره 1.5 سل.

تساءل عدد من الزبائن حول القيمة المتوسطة المعلنة من طرف صاحب الآلة. من خلال طرح التساؤل التالي: هل حجم متوسط القهوة المقدمة حقيقية 15 سل أم لا.

للإجابة على هذا التساؤل، قرروا قياس حجم القهوة الموزعة في 100 كوب بلاستيكي مختارة عشوائيا. فوجدوا أن المتوسط الحسابي هو 14.2 سل.

المطلوب:

1- كيف يمكن لنا استخدام هذه المعلومات عند مستوى معنوية قدره 5%؟

2- أوجد قيمة مستوى معنوية الاختبار بافتراض أن مجال قبول فرض العدم هو [14.651 – 15.349]

3- أوجد قيمة درجة ثقة الاختبار إذا كانت القيمة الحقيقة لمتوسط المجتمع تساوي 14.6 سل؟

الحل:

1- نرغب في هذا السؤال أن نقارن بين متوسط المجتمع μ والقيمة المصرح بها له $\mu_0 = 15\text{cl}$. لهذا نقوم بوضع الفرض التالي:

$$H_0: \mu = 15$$
$$H_1: \mu \neq 15$$

- قاعدة القرار

عند مستوى معنوية 5% فإن منطقة الرفض تكون ضمن $Z_{\text{الجدولية}}$ التي تساوي± 1.96.

- اتخاذ القرار

$$Z_{\text{المحسوبة}} = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{14.2 - 15}{1.5/\sqrt{100}} = -5.33$$

بمستوى معنوية قدره 5%، فإن قيمة $Z_{\text{المحسوبة}}$ تقع في منطقة الرفض، وهذا ما يقودنا إلى رفض فرض العدم. وعليه، فإن المحتوى المتوسط للقهوة الموزعة في الكؤوس البلاستيكية لا يطابق القيمة المصرح بها وهي 15 سل، وأن الفرق الموجود بين القيمة المتوسطة للعينة 14.2 والقيمة المتوسطة للمجتمع 15 والذي يساوي 0.8 هو فرق حقيقي.

2- إيجاد قيمة مستوى معنوية الاختبار

استنادا إلى المجال المعطى، فإن قيمة خطأ التقدير تكون مساوية لـ:

$$\mu = \bar{X} \pm Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \Rightarrow 15 - 14.651 = \pm 0.349$$

إذن:

$$Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = 0.349 \Rightarrow Z_{1-\frac{\alpha}{2}} = \frac{0.349 \times \sqrt{100}}{1.5} = 2.32$$

$$1 - \frac{\alpha}{2} = 0.9898 \Rightarrow \alpha \cong 2\%$$

3- إيجاد درجة ثقة الاختبار

نقوم بتحديد منطقتي القبول والرفض

$$\mu = \bar{X} \pm Z_{1-\frac{\alpha}{2}} \Rightarrow 14.6 - 14.2 = \pm Z_{1-\frac{\alpha}{2}} \frac{1.5}{\sqrt{100}}$$

$$14.6 - 14.2 = \pm Z_{1-\frac{\alpha}{2}} \frac{1.5}{\sqrt{100}} \Longrightarrow Z_{1-\frac{\alpha}{2}} = \frac{0.4 \times \sqrt{100}}{1.5} = 2.66$$

$$1 - \frac{\alpha}{2} = 0.9961 \Rightarrow -\alpha = 99.22\%$$

تمرين رقم 4-6:

كانت نسبة الناجح في مجتمع مع في شهادة البكالوريا هي 72%، وبعد إحداث بعض التغييرات في المنظومة التربوية. وجد في عينة من 100 تلميذ أن 80 منهم نالوا الشهادة.

المطلوب: هل حسنت هذه التغييرات من نسبة النجاح في البكالوريا عند مستوى معنوية قدره 5%؟

الحل:

المعطيات:

$$p = 0.72, \bar{p} = \frac{8000}{10000}$$

- وضع الفرض

$$H_0: p \geq 0.72$$
$$H_1: p > 0.72$$

- قاعدة القرار

عند مستوى معنوية 5% فإن منطقة الرفض تكون إلى يمين $Z_{\text{الجدولية}}$ التي تساوي 1.65.

- اتخاذ القرار

$$Z_{\text{المحسوبة}} = \frac{\bar{p} - p_0}{\sqrt{\frac{pq}{n}}} = \frac{0.8 - 0.72}{\sqrt{\frac{0.8 \times 0.2}{100}}} = 2$$

حيث أن قيمة $Z_{\text{المحسوبة}}$ تقع عند منطقة الرفض، فإننا نرفض فرض العدم ونقبل الفرض البديل، وبالتالي فإن التعديلات التي أدخلت على المنظومة التربوية قد حسنت من نسبة النجاح في البكالوريا.

تمرين رقم 4-7: أجريت تجربة لاختبار مدى تلاؤم نوعين من الأعلاف الجديدة للأبقار، حيث خصص لكل عينة 10 بقرات. حيث كان معدل الزيادة في وزن البقرات كما يلي:

علف رقم 1	31	34	29	26	32	35	38	34	30	32
علف رقم 2	28	26	24	28	29	30	29	26	31	29

المطلوب:

هل يوجد فرق معنوي بين نوعي العلف عند مستوى معنوية قدره 5%؟

الحل:

- تحديد الفرض

$$\begin{matrix} H_0: \mu_1 \leq \mu_2 \\ H_1: \mu_1 > \mu_2 \end{matrix} \; or \; \begin{matrix} H_0: \mu_1 - \mu_2 \leq 0 \\ H_1: \mu_1 - \mu_2 > 0 \end{matrix}$$

- تحديد قاعدة القرار

حساب قيم المتوسط الحسابي والانحراف المعياري

$$\bar{X}_1 = \frac{\sum x_i}{n} = 32.1, \qquad \bar{X}_2 = \frac{\sum x_i}{n} = 28$$

$$s_1 = \sqrt{\frac{\sum(x_i - \bar{X})^2}{n-1}} \cong 3.38, \qquad s_2 = \sqrt{\frac{\sum(x_i - \bar{X})^2}{n-1}} \cong 2.11$$

درجة حرية الاختبار هي:

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{s_2^2}{n_2}\right)^2} = \frac{\left(\frac{3.38^2}{10} + \frac{2.11^2}{10}\right)^2}{\frac{1}{10-1}\left(\frac{3.38^2}{10}\right)^2 + \frac{1}{10-1}\left(\frac{2.11^2}{10}\right)^2} = 9.5$$

عند مستوى معنوية قدره5% ودرجة حرية 12 فإن منطقة القبول تكون ضمن المجال $]-2.228, 2.228[$ ومنطقة الرفض خارج هذا المجال.

- اتخاذ القرار

نقوم بإيجاد قيمة إحصاء الاختبار $t_{\text{المحسوبة}}$ باستخدام العبارة التالية:

$$t_{\text{المحسوبة}} = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sigma_{\bar{X}_1 - \bar{X}_2}} = \frac{(31.1 - 28) - (0)}{\sqrt{\frac{3.38^2}{10} + \frac{2.11^2}{10}}} = 2.46$$

حيث أن قيمة $t_{\text{المحسوبة}}$ تقع في منطقة الرفض، فإننا نرفض فرض العدم ونقبل الفرض البديل. ومنه فغنه يوجد اختلاف معنوي بين نوعي العلف وهذا عند مستوى معنوية قدره 5%.

تمرين رقم 4-8:

تنافست ثمانية أحزاب معينة على الفوز بمقاعد البرلمان في اقتراع عام. يبين الجدول الموالي نسبة عدد المقاعد التي توقعها المراقبون قبل الاقتراع وعدد المقاعد التي حصل عليها كل حزب بعد الاقتراع.

الحزب	1	2	3	4	5	6	7	8
النسبة	0.20	0.20	0.15	0.15	0.10	0.10	0.05	0.05
عدد المقاعد	92	105	86	71	33	59	19	35

المطلوب:

اختبر بمستوى معنوية 5% تطابق توقعات المراقبين مع ما أفضت إليه نتائج الاقتراع؟

الحل:

– نضع الفرض:

H_0 : التكرارات المشاهدة لا تختلف معنويا عن التكرارات المتوقعة

H_1 : التكرارات المشاهدة تختلف معنويا عن التكرارات المتوقعة

– بما أن العدد الكلي لعدد مقاعد البرلمان هو 500 (مجموع عدد المقاعد) فإن العدد المتوقع لمقاعد كل حزب E_i بموجب النسبة المتوقعة يساوي:

$$E_i : \ np_1 = 500(0.20) = 100$$

$$np_1 = 500(0.20) = 100$$

$$np_1 = 500(0.15) = 75$$

$$np_1 = 500(0.15) = 75$$

$$np_1 = 500(0.10) = 50$$

$$np_1 = 500(0.10) = 50$$

$$np_1 = 500(0.05) = 25$$

$$np_1 = 500(0.05) = 25$$

- حساب قيمة كاي-مربع المحسوبة

$$\chi^2_{\text{المحسوبة}} = \sum \frac{(f_0 - f_e)^2}{f_e} = \frac{(92-100)^2}{100} + \frac{(105-100)^2}{100} + \cdots + \frac{(35-25)^2}{25} = 15.56$$

- حساب قيمة كاي-مربع الجدولية

عند مستوى معنوية قدره 0.05 ودرجة حرية قدرها 7 فإن قيمة كاي-مربع الجدولية هي 14.07.

حيث أن قيمة كاي-مربع المحسوبة أكبر من قيمة كاي-مربع الجدولية، فإننا نرفض فرض العدم ونقبل الفرض البديل. لهذا فإنه لا تتطابق نسبة توقعات المراقبين لمقاعد كل حزب مع ما أقضت إليه النتائج.

تمرين رقم 4-9:

يعطي الجدول الموالي توزيع درجات اختبار ما لعينة من 100 طالب جامعي.

درجات	350-250	450-350	550-450	650-550	750-650
عدد الطلاب	3	25	50	20	2

المطلوب: باستخدام مستوى معنوية 5% اختبر ما إذا كانت هذه الدرجات تتبع التوزيع الطبيعي.

الحل:

- نضع الفرض:

H_0: المجتمع يتبع الطبيعي

H_1: المجتمع لا يتبع الطبيعي

- حساب قيم المتوسط الحسابي والانحراف المعياري

$$\bar{X} = \frac{\sum x_i}{n} = 493, \qquad s = \sqrt{\frac{\sum (x_i - \bar{X})^2}{n-1}} \cong 80.72$$

- حساب التكرار المتوقع للدرجات باستخدام $\bar{X}$ و s.

حدود الدرجات	$\frac{x-\mu}{\sigma} < Z < \frac{x-\mu}{\sigma}$	P(Z)	f_0	
أقل من 350	$Z < \frac{350-493}{80.72}$	0.0384	3.84	29.81
450−350	$\frac{350-493}{80.72} < Z < \frac{450-493}{80.72}$	0.2597	25.97	
550−450	$\frac{450-493}{80.72} < Z < \frac{550-493}{80.72}$	0.4631	46.31	46.31
650−550	$\frac{550-493}{80.72} < Z < \frac{650-493}{80.72}$	0.2126	21.26	23.88
أكبر من 650	$Z > \frac{750-493}{80.72}$	0.0262	2.62	

- حساب قيمة كاي−مربع المحسوبة

$$\chi^2_{\text{المحسوبة}} = \sum \frac{(f_0 - f_e)^2}{f_e} = \frac{(28-29.81)^2}{29.81} + \frac{(50-46.31)^2}{46.31} + \frac{(22-23.88)^2}{23.88} = 0.54$$

- حساب قيمة كاي−مربع الجدولية

حيث أنه تم تقدير كل من μ و σ من $\bar{X}$ و s، فإن عدد معالم المجتمع المقدرة من إحصائيات العينة m=2.

ddl=c−m−1= 5−2−1=2

$$\chi^2_{\substack{2\\0.05}} = 5.991$$

حيث أن قيمة كاي−مربع المحسوبة تقع في منطقة القبول، فإننا نقبل فرض العدم بأن درجات الاختبار تتبع التوزيع الطبيعي.

الفصل الرابع: نموذج الانحدار الخطي البسيط

4 – 1 تمهيد

كما هو معلوم، فإن الهدف الإجمالي من أي علم هو التغلب على الطبيعة، حيث أنها تفرض صعوبات على الإنسان، ونهدف من خلال تطوير مختلف العلوم إلى محاولة إيجاد سبل للتخفيف والحد من هذه الصعوبات. ويأتي الإحصاء التطبيقي بالعديد من الأدوات والطرق التي تساعد في مجملها على كشف، تحليل وتوقع مختلف الظواهر. ويعتبر نموذج الانحدار الخطي البسيط إحدى فصول الإحصاء التطبيقي، الذي يستخدم لمحاولة دراسة أثر متغير ما على متغير آخر. وهو ما يسمح لنا بفهم الظواهر وتوقع قيمها المقدرة في المستقبل، وهو ما يمكن متخذ القرار من تحديد القرارات المناسبة.

4 – 2 مفهوم نموذج الانحدار الخطي البسيط

4-2-1 تعريف نموذج الانحدار الخطي البسيط

إن الظواهر الاقتصادية بصفة عامة غاية في التشابك والتعقيد، ومن هنا تتضح الغاية من استخدام النماذج وذلك لأجل تسهيل عمليات حل المشاكل الواقعية، سواء كانت بسيطة أم معقدة، إذ ما قام متخذ القرار بالتركيز على الخصائص والأسباب الرئيسية لهذه المشاكل بدلا من دراسة وفحص كل تفاصيل ودقائق المشكلة الواقعية، وهذا التجريد أو التقريب للواقع العملي والذي يمكن إعداده في أشكال متنوعة هو ما يعرف باسم النموذج، ومنه يمكن أن نعرف النموذج بأنه " تمثيل أو تجريد مبسط للواقع العملي في صورة مجموعة من المعادلات والرموز الرياضية، فهو يبين العلاقة المباشرة وغير المباشرة التي تربط العناصر الرئيسية للمشكلة والأفعال وردودها الموجودة في الواقع "[1]. كذلك يمكن تعريف النموذج بمعناه المجرد بأنه "محاكاة علمية لطبيعة الأشياء أو صياغة مفاهيميه". وهذه الصياغة هي صياغة تشكيلية، وذلك باستخدام التحليل. والنموذج اصطلاحا هو "عينة أو مصغر أو قالب ممثل أو شيء مقارب، أو تحويل الظاهرة أو العملية إلى رموز وعلاقات ومعادلات"[2].

وتستخدم النماذج في معظم العلوم لأن الظواهر عادة ما تكون معقدة في الواقع العملي إلى درجة أنه يستحيل دراستها إلا من خلال تمثيلها في نماذج لغرض تبسيطها. و هناك العديد من النماذج و كل منها يتلاءم و طبيعة الظاهرة المدروسة، فنجد النماذج الهندسية، النماذج المادية و النماذج الجبرية[3].

ويعرف النموذج الاقتصادي "بأنه مجموعة من العلاقات بين المتغيرات الاقتصادية لتمثيل ظاهرة معينة بصورة خالية من التفاصيل والتعقيدات ولكنها ممثلة للواقع بهدف تعليلها أو التنبؤ بها و السيطرة عليها"، ويهدف النموذج إلى تقدير قيم عددية لمعلمات علاقة بين متغيرات اقتصادية بغية التنبؤ أو تحليل هيكل اقتصادي أو تقييم سياسة اقتصادية، ويستخدم النموذج الاقتصادي الرموز الرياضية[4].

---

[1] - رجال السعدي، بحوث العمليات، دار رجزو، قسنطينة، 2004، ص ص 13 – 14.
[2] - وليد إسماعيل السيفو، وآخرون، أساسيات الاقتصاد القياسي التحليلي، دار الأهلية للنشر والتوزيع، الأردن، 2006، ص 43.
[3] - محمد صالح تركي القريشي، مقدمة في الاقتصاد القياسي، الوراق للنشر و التوزيع، الأردن، 2004، ص 26.
[4] - حسين علي بخيت، سحر فتح الله، الاقتصاد القياسي، دار اليازوري، الأردن، 2007، ص 22.

كما يعرف النموذج الاقتصادي بأنه مجموعة من العلاقات الاقتصادية التي توضع عادة بصيغ رياضية تسمى معادلة، أو مجموعة معادلات هيكلية، والتي تعكس أو تشرح سلوك أو آلية العلاقات بين المتغيرات الاقتصادية التي تبين عمل منشأة أو قطاع أو اقتصاد بلد معين"[1].

وهناك مجموعة من الخصائص التي يجب أن تتوافر في النموذج الاقتصادي والتي نذكر منها[2]:

– أن يكون مطابقا للنظرية الاقتصادية، بحيث يصف الظاهرة الاقتصادية بشكل صحيح.

– قدرته على توضيح المشاهدات الواقعية، بحيث يكون متناسقا مع السلوك الفعلي للمتغيرات الاقتصادية التي تحدد العلاقة بين هذه المتغيرات.

– دقته في تقدير المعلمات، حيث يجب أن تتطابق هذه التقديرات مع القيم الواقعية للمتغيرات.

– قدرة النموذج على التنبؤ، بحيث يعطي تنبؤات مرضية للقيم المستقبلية للمتغيرات المعتمدة.

– يجب أن يبرز النموذج الاقتصادي العلاقات الاقتصادية بكل بساطة شريطة أن لا يكون على حساب الدقة في التقدير.

وتتكون النماذج من:

- **الثوابت:** وهي قيم ثابتة منذ تاريخ اكتشافها حتى يثبت عكس ذلك علميا، مثل π، e،
- **المعالم:** وهي قيم ثابتة أثناء أداء النموذج،
- **المتغيرات:** وهي مقادير كمية أو وصفية تتغير قيمها من عنصر لآخر.

4-2-2 الصياغة الرياضية لنموذج الانحدار الخطي البسيط

يلجأ الاقتصاديون المستخدمون للأدوات الرياضية في عملية النمذجة إلى عملية التجريد والتبسيط.

* **التجريد؟** يعني اللجوء إلى تبسيط الواقع، وبصورة عقلانية. من خلال القيم بعدة إجراءات ومنها على الأساس:

- التركيز على أهم العناصر التي يبدو تأثيرها فعالا في سير وتطور الظاهرة محل النمذجة،
- الإلقاء جانبا العناصر الثانوية والعارضة ذات الأثر الضئيل،
- تحديد الفروض التي سيبنى النموذج على أساسها.

[1] - وليد إسماعيل السيفو، و آخرون، مرجع سبق ذكره، ص 43.
[2] - حسين علي بخيت، سحر فتح الله، مرجع سبق ذكره، ص 25.

* **التحليل؟** هو وسيلة أساسية تمكن من اختبار مدى صلاحية النموذج، وتتمثل في إجراء عملية فحص من واقع الماضي.

إن دراسة العلاقة بين المتغيرات الاقتصادية يتطلب تحديد المتغيرات المؤثرة في تلك العلاقة. ونموذج الانحدار الخطي البسيط (Simple Linear regression) من أكثر النماذج شيوعا في تمثيل الظواهر الاقتصادية، وهو " نموذج يعبر عن وجود علاقة خطية بين متغير تابع، ومتغير مستقل". من هذا التعريف نجد أن سلوك ظاهرة ما أو متغير معينا يرتبط بواحد من المتغيرات أو العوامل المحددة فقط، يؤثر في وقت واحد و بدرجة متفاوتة من القوة على سلوك الظاهرة المدروسة. ويمكن التعبير عن العلاقة بين المتغير التابع ومجموعة المتغيرات المستقلة وفق نموذج الانحدار الخطي البسيط رياضيا بالصيغة التالية:

$$Y_i = a + bX_i + \mu_i$$

من العلاقة السابقة نجد أن نموذج الانحدار الخطي البسيط يتكون من:

أ – **المتغير التابع** Y_i : المتغير بالمفهوم الرياضي هو ظاهرة يمكن أن تتخذ عدة قيم في الموضوع قيد البحث، وفي المفهوم الاقتصادي من خلال العلاقة السابقة يرمز إلى سلوك الظاهرة التي تستجيب لسلوك المتغيرات المستقلة، وبهذا فهو مخرج النموذج (Output)، ويسمى تابعا لأنه يتبع في سلوكه المتغير المستقل المؤثرة فيه.

ب – **المتغير المستقل** Xi : وهو العنصر المتحكم افتراضا في سلوك الظاهرة المعينة، أي أنه مدخل النموذج(Input) المفسرة للتغيرات في سلوك الظاهرة.

ج – **معلمات النموذج (a,b):** b تمثل نسبة التغير في المتغير التابع عندما يتغير المتغير المستقل بوحدة واحدة. a تسمى بالعنصر الثابت، والذي يمثل قيمة المتغير التابع عندما تكون قيمة المتغير المستقل معدومة.

د – **المتغير العشوائي**μ_i: ويسمى أيضا بحد الخطأ العشوائي، وهو يضم تأثير كل العوامل المعروفة (وغير المدرجة) وغير المعروفة على المتغير التابع.

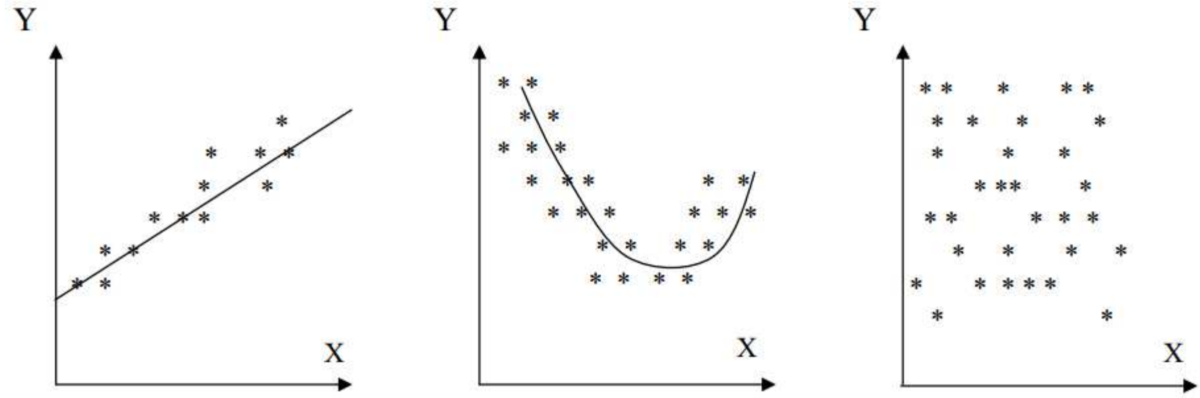
وتعتبر المشاهدات الخاصة بالحد العشوائي مقدرة وغير حقيقية، a,b هي معلمات النموذج التي تحدد طبيعة العلاقة بين المتغيرين. ويسمى هذا النموذج بالخطي البسيط لأن:

- المتغير Yi ينحدر على أساس المتغير Xi،
- بسيط لأنه يمثل العلاقة بين متغيرين فقط،
- خطي لأن العلاقة بين متغيريه تأخذ شكل خط مستقيم.

4-2-3 دراسة الخطية بين المتغيرين التابع والمستقل

لقد أشرنا سابقا إلى أن نموذج الانحدار الخطي البسيط، يفترض وجود علاقة خطية بين المتغيرين، التابع والمستقل. لمعرفة والكشف عن وجد علاقة خطية بين المتغيرين نعتمد على شكل انتشار المتغيرين. وذلك بوضع المتغير التابع على المحور الرأسي والمتغير المستقل على المحور الأفقي، ومن ثم وضع انتشار في شكل نقط لكل الثنائيات (X,Y) للعينة محل الدراسة. ويظهر الشكل رقم (4-1) بعض الأشكال للانتشار.

شكل رقم (4-1): شكل الانتشار



شكل (أ) انحدار خطي بسيط — شكل (ب) انحدار غير خطي بسيط — شكل (ج) لا توجد علاقة انحدار

نلاحظ من خلال الشكل رقم (4-1) ما يلي:

- **شكل (أ):** كلما زادت قيم المتغير المستقل X نلاحظ زيادة في قيم المتغير التابع Y، ونلاحظ أن انتشار هذه القيم هو بشكل خطي تقريبا. أي أن هذه الثنائيات تنتشر حول خط مستقيم بالزيادة والنقصان. بحث أن الزيادة في المتغير المستقل أدت إلى زيادة في المتغير التابع. لهذا فإننا نستنتج وجود علاقة خطية بين المتغيرين،
- **شكل (ب):** نلاحظ أن نقاط انتشار قيم المتغير التابع والمستقل هي بشكل منحني، حيث أن زيادة المتغير المستقل أدت في بداية الأمر إلى انخفاض في المتغير التابع، ومن ثم إلى زيادة فيه. لهذا فإن العلاقة بين المتغيرين تعتبر غير خطية.

- **شكل (ج):** نلاحظ أن نقاط الانتشار في هذا الشكل هي مبعثرة، ولا يمكن تتبع أثر المتغير المستقل على المتغير التابع فيها، وبالتالي لا توجد علاقة بينهما.

على هذا الأساس، وباستخدام شكل الانتشار، يمكن لنا تحديد والكشف علن وجود علاقة خطية بين المتغيرين التابع والمستقل.

3-4 تقدير نموذج الانحدار الخطي البسيط

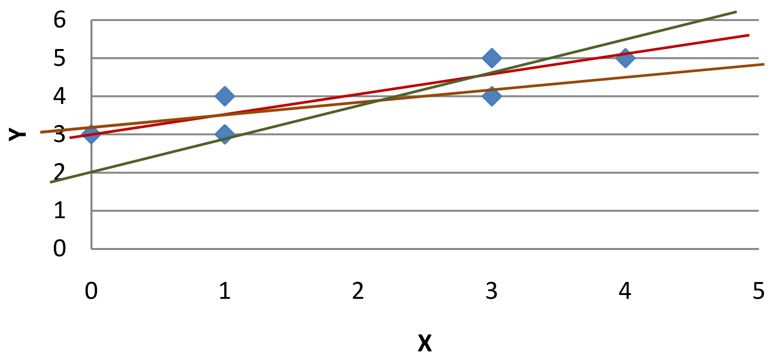
تعتبر عملية تقدير معلمات نموذج الانحدار الخطي البسيط جد مهمة، وذلك من خلال اختيار الخط الذي نوفق من خلاله شكل العلاقة بين المتغيرين التابع والمستقل. فإذا كانت نقاط الانتشار تقع تماما على خط مستقيم، فلا داعي للقلق حول تحديد الخط الملخص للبيانات بالشكل الأفضل. وكل ما علينا عمله هو وصل النقاط المشاهدة للحصول على المستقيم الملائم الأفضل للبيانات. وعندما لا تقع النقاط تماما على الخط المستقيم، فإن اختيار هذا الخط يصبح مسألة أكثر تعقيدا.

وحيث أن في العلوم الإنسانية، من النادر إن لم نقل من المستحيل أن تكون نقاط انتشار الظواهر المدروسة تقع على خط مستقيم، لهذا بحث الرياضيون عن عدة طرق لإيجاد أفضل التوفيقات التي تعمل على تمثيل الظاهرة المدروسة بشكل قريب من الواقع.

على سبيل المثال، لدينا بيانات حول قيم الدخل والاستهلاك لعينة من 6 دول.

4	3	3	1	1	0	X
5	5	4	4	3	3	Y

ونعرض في الشكل الموالي شكل الانتشار مع عدة توفيقات لخط مستقيم لها



يبين الشكل السابق عدة خطوط توفق بيانات العينة، وكل خط يبتعد عن النقاط الحقيقة للانتشار بمسافات معينة. وعلينا البحث عن الخط الوحيد من بين هذه الخطوط الذي يوفق بيانات العينة بشكل أفضل من بقية الخطوط، أي أنه يقترب من النقاط المشاهدة بأكبر قيمة ممكنة. بصيغة أخرى، تكون المسافة الرأسية بين النقاط المشاهدة وبين الخط الذي يمر بينها تكون عند أقل قيمة ممكنة.

لنرمز إلى الفرق الرأسي بين القيم المشاهدة والقيم المقدرة الممثلة بواسطة خط مستقيم بالرمز e_i. نعرف الكمية التالية:

$$e_1^2 + e_2^2 + \cdots + e_i^2$$

من خلال هذا التعريف، نحن نبحث عن المستقيم من جميع المستقيمات الذي يخفض الكمية السابقة إلى أقل قدر ممكن، وهو المستقيم الذي يسمى بخط الانحدار الموفق وفق طريقة المربعات الصغرى (Ordinary least squares). ولاستخدام هذه الطريقة في التقدير يجب أن تتوفر لنا مجموعة من الفرضيات نعرضها في الجزء الموالي.

4-3-1 فرضيات طريقة المربعات الصغرى (شروط Gauss-Markov)

إن تقدير نموذج الانحدار الخطي البسيط المكتوب بالعلاقة: $Y_i = a + bX_i + \mu_i$، وفق مفهوم المربعات الصغرى يتطلب أن تتوفر عدة فرضيات هي:

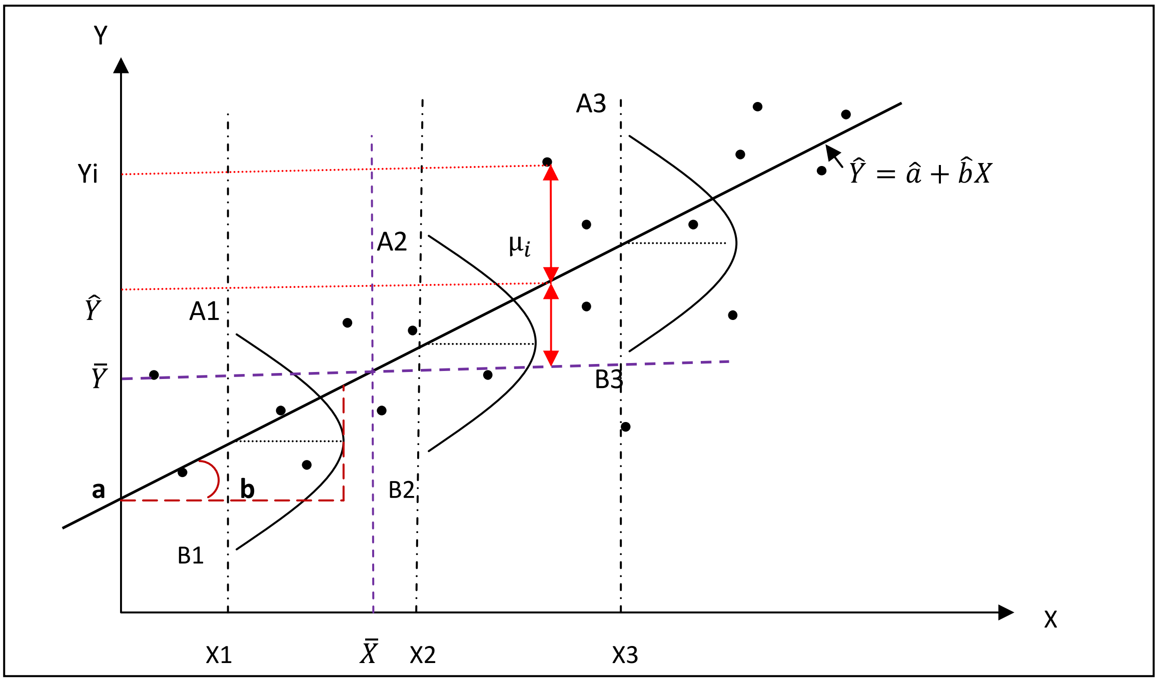
أولا: المتغير μ_i متغير عشوائي

يعني هذا الفرض أن القيمة التي يأخذها هذا المتغير تتوقف على عامل الصدفة، وأن كل قيمة يمكن أن يأخذها لها احتمال معين للتحقق، بحيث تكون هذه الاحتمالات مستقلة.

ثانيا: الأمل الرياضي للأخطاء معدوم

تعني هذه الفرضية أن الأخطاء μ_i لا تدخل في تفسير المتغير التابع، إذ أنها تعبر عن حدود عشوائية لا يمكن قياسها أو تحديد قيمها بدقة. بالنسبة لكل مستوى من المتغير المستقل Xi، هناك عدة قيم للمتغير Yi. وبالتالي هناك قيم مختلفة للأخطاء بعضها يكون موجب وبعضها يكون سالب على النحو الذي يظهره الشكل رقم (4-3).

شكل رقم (4−1): نموذج الانحدار الخطي البسيط



نلاحظ من خلال الشكل رقم (4−1)، أن انتشار قيم المتغير Yi حول كل قيمة Xi، يشكل توزيع احتمالي يمكن تصويره. ويظهر أن هناك قيم موجبة وأخرى سالبة لهذه الأخطاء بحيث يفترض أن أملها الرياضي يكون معدوم. ويمكن التعبير رياضيا على هذا الفرض بالعلاقة التالية:

$$E(e_i) = 0$$

ثالثا: ثبات تباين الأخطاء

يجب أن يكون تباين الأخطاء حول متوسطها الحسابي ثابت كند كل القيم الممكنة للمتغير X.

أي أن انتشار قيم μ_i حول متوسطها الحسابي لا يختلف باختلاف القيم المناظرة للمتغير X. ويمكن تفسير ذلك على وجه التقريب بأن هناك مدى ثابت لتغير قيم $\boldsymbol{\mu_i}$ ، وبالعودة إلى الشكل رقم (4−1) يمكن التعبير عن ثبات التباين عن طريق ثبات المدى بين القيم التالية:

$$A_1{\sim}B_1 = A_2{\sim}B_2 = A_3{\sim}B_3$$

أي أن انتشار قيم μ_i حول القيم المنخفضة للمتغير المستقل X لا يختلف عن انتشار القيم حول القيم المرتفعة له. وهذا الفرض لا يعبر عن حقيقة يمكن التسليم بها، وإنما هو مجرد فرض قابل للمناقشة، قد يصح في بعض الظروف، وقد لا يصح في بعضها الآخر.

ويعبر عن فرض ثبات التباين رياضيا بالعلاقة التالية:

$$Var(\mu_i) = E[\mu_i - E(\mu_i)]^2 = \sigma_i^2$$

رابعا: قيم الخطأ العشوائي مستقلة عن بعضها البعض

أي أن القيمة التي يأخذها الخطأ عند أي مستوى لقيم المتغير المستقل، لا تتأثر بأي قيمة أخرى للخطأ عند أي مستوى آخر للمتغير المستقل. وهذا يعني أن التباين المشترك (التغاير) لأي زوج من قيم الخطأ يساوي الصفر. ويعبر رياضيا عن هذا الفرض بـ:

$$Cov(\mu_i, \mu_j) = 0 \quad \forall_{ij} / \, i \neq j$$

خامسا: الخطأ العشوائي مستقل عن المتغير المستقل

يعني هذا الفرض أن المتغير المستقل لا يكون سبب في حدوث الخطأ العشوائي، أي أنه لا يوجد اتجاه لتغير المتغير المستقل والخطأ العشوائي في نفس الوقت. ويمكن التعبير عن هذا الفرض رياضيا بـ:

$$Cov(\mu_i, x_i) = 0$$

سادسا: توزيع الخطأ العشوائي

لعدة عوامل، واستنادا إلى نظرية النهاية المركزية، نفترض أن الخطأ العشوائي يتبع التوزيع الطبيعي بوسط معدوم وتباين قدره σ_μ^2. ويعبر عن هذا الفرض كالتالي:

$$\mu \sim N(0, \sigma_\mu^2)$$

سابعا: لا توجد هناك أخطاء تجميع

ثامنا: وصف العلاقة سليم

4-3-2 تقدير معالم نموذج الانحدار الخطي البسيط

يمكن تقدير معلمات العلاقة: $Y_i = a + bX_i + \mu_i$، والتي نفترض وجودها في المجتمع باستخدام طريقة المربعات الصغرى Ordinary Least Squares.

نفرض أن العلاقة التي قدرت بهذة الطريقة من بيانات العينة هي $\hat{Y}i = \hat{a} + \hat{b}Xi$ والتي يمكن كتابتها على النحو التالي $Y_i = a + bX_i + e_i$ حيث ei هي البواقي والتي تعتبر تقديرات لعنصر الخطأ أو المتغير العشوائي μ_i.

وطريقة OLS يتم فيها البحث عن قيم $\hat{a}, \hat{b}$ التي تجعل المجموع $\sum ei^2$ أصغر ما يمكن. وهندسيا (أنظر الشكل رقم 4-1) فإن ذلك يعني البحث عن المستقيم الذي له إحداثيات x,y يمر عبر سحابة من النقاط الممثلة للقيم المشاهدة، بشكل يكون فيها مجموع مربع مسافات النقاط إلى المستقيم أصغر ما يمكن، وهذا المسافات مقاسة على المحور الرأسي.

نخلص مما سبق بأن أساس طريقة OLS هو أن نبحث عن الخط المستقيم $\hat{Y}i = \hat{a} + \hat{b}Xi$ الذي يمر بالمشاهدات والذي يجعل قيمة $\sum ei^2$ أقل ما يمكن، و بعبارة أخرى فإننا نختار تلك القيم لـ $\hat{a}, \hat{b}$ التي تؤدي إلى جعل $\sum ei^2$ أصغر ما يمكن، وهكذا فإن تقديرات المربعات الصغرى للمعلمتين a,b هي تلك التقديرات التي تنتج عن تدنية:

$$Min\sum ei^2 = Min\left(Yi - \hat{Y}\right)^2 = Min\left(Yi - \hat{a} - \hat{b}Xi\right)^2$$

الشرط الضروري لكي تكون قيمة $\sum ei^2$ أصغر ما يمكن هو أن تكون المشتقات الجزئية الأولى للعلاقة السابقة بالنسبة لـ $\hat{a}, \hat{b}$ تساوي الصفر.

بأخذ التفاضل الجزئي للعلاقة السابقة بالنسبة للمعامل $\hat{a}$ ومساواته للصفر نحصل على:

$$\frac{\sigma \sum ei^2}{\sigma \hat{a}} = 0$$

$$\Rightarrow \frac{\sigma \sum \left(Yi - \hat{a} - \hat{b}Xi\right)^2}{\sigma \hat{a}} = 0$$

$$\Rightarrow -2\sum \left(Yi - \hat{a} - \hat{b}Xi\right) = 0$$

$$\Rightarrow \sum Yi - n\hat{a} - \hat{b}\sum Xi = 0................(1)$$

تسمى المعادلة (1) بالمعادلة الطبيعية الأولى.

بأخذ التفاضل الجزئي للعلاقة السابقة بالنسبة للمعامل $\hat{b}$ ومساواته للصفر نحصل على:

$$\frac{\sigma\sum ei^2}{\sigma\hat{b}}=0$$

$$\Rightarrow \frac{\sigma\sum\left(Yi-\hat{a}-\hat{b}Xi\right)^2}{\sigma\hat{b}}=0$$

$$\Rightarrow -2\sum\left(Yi-\hat{a}-\hat{b}Xi\right)Xi=0$$

$$\Rightarrow \sum XiYi-\hat{a}\sum Xi-\hat{b}\sum X_i^2=0...............(2)$$

تسمى المعادلة (2) بالمعادلة الطبيعية الثانية.

لدينا المعادلتين 1 و 2 وهما كافيتين لتقدير قيمة المجهولين $\hat{a},\hat{b}$:

$$\begin{cases}\sum Yi=\hat{a}+\hat{b}\sum Xi\\ \sum XiYi=a\sum Xi+\hat{b}\sum X_i^2\end{cases}$$

يمكن حل هذه الجملة آنيا بطريقة المحددات لنحصل على النتيجة التالية:

$$\hat{a}=\frac{\begin{vmatrix}\sum Yi & \sum Xi\\ \sum XiYi & \sum X_i^2\end{vmatrix}}{\begin{vmatrix}n & \sum Xi\\ \sum Xi & \sum X_i^2\end{vmatrix}}=\frac{\sum Yi\sum X_i^2-\sum XiYi\sum Xi}{n\sum X_i^2-\left(\sum Xi\right)^2}$$

$$\hat{b}=\frac{\begin{vmatrix}n & \sum Yi\\ \sum Xi & \sum XiYi\end{vmatrix}}{\begin{vmatrix}n & \sum Xi\\ \sum Xi & \sum X_i^2\end{vmatrix}}=\frac{n\sum XiYi-\sum Xi\sum Yi}{n\sum X_i^2-\left(\sum Xi\right)^2}$$

إعادة صياغة معادلات $\hat{a},\hat{b}$ باستخدام انحرافات قيم المتغيرات عن أوساطها الحسابية:

يمكن التوصل إلى نفس النتيجة باستخدام انحرافات قيم المتغيرات عن أوساطها الحسابية أي باستخدام:

$$x_i=Xi-\overline{X},\ y_i=Yi-\overline{Y}$$

باستخدام تحويلات مناسبة نحصل على:

$$\hat{b} = \frac{\sum x_i y_i}{\sum x_i^2}$$
$$\hat{a} = \bar{Y} - \hat{b}\bar{X}$$

مثال رقم (3−1):

أخذت عينة من 06 مشاهدات حول قيم X,Y حيث X يمثل الدخل، و Y يمثل الاستهلاك. وتقابل العينة النموذج الخطي $Y_i = a + bX_i + U_i$ حيث U مقدار الخطأ العشوائي.

4	3	3	1	1	0	X
5	5	4	4	3	3	Y

المطلوب: قدر العلاقة التابعية الخطية بين الدخل كمتغير مستقل والاستهلاك كمتغير تابع، و فسر معلمات هذه العلاقة؟

الحل:

n	X	Y	$x = X - \bar{X}$	$y = Y - \bar{Y}$	$x.y$	x^2
0	0	3	−2	−1	2	4
1	1	3	−1	−1	1	1
2	1	4	−1	0	0	1
3	3	4	1	0	0	1
4	3	5	1	1	1	1
5	4	5	2	1	1	4
$\sum$	12	24	0	0	6	12

خط الانحدار المقدر يكتب بالشكل: $\hat{Yi} = \hat{a} + \hat{b}Xi$

حساب $\bar{X},\bar{Y}$ $\bar{Y} = \frac{\sum Yi}{n} = \frac{24}{6} = 4$ $\bar{X} = \frac{\sum Xi}{n} = \frac{12}{6} = 2,$

تقدير $\hat{b}$: $\hat{b} = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{6}{12} = \frac{1}{2}$

تقدير $\hat{a}$: $\hat{a} = \overline{Y} - \hat{b}\overline{X} = 4 - 0.5 \times 2 = 3$

ومنه فخط الانحدار المقدر يكتب بالشكل $\hat{Y}i = 3 + 0.5Xi$

تفسير المعلمات:

$\hat{a}$: هي نقطة تقاطع خط الانحدار مع محور العينات، وهي تمثل قيمة الاستهلاك عندما يكون الدخل معدوم (الاستهلاك التلقائي).

$\hat{b}$: هي ميل خط الانحدار، وتمثل الزيادة في قيمة الاستهلاك عندما يزداد الدخل بوحدة نقدية واحدة (الميل الحدي للاستهلاك).

4-4 تشكيل مجالات الثقة للقيم المقدرة

لاحظنا أن تقديرات المربعات الصغرى لنموذج الانحدار الخطي البسيط، قد أعطت لنا تقديرات نقطية، وهذا بالنسبة للمعلمتين المقدرتين $\hat{a}, \hat{b}$. ونرغب في هذا الجزء، في تكوين تقدير بمجال لهاتين المعلمتين المقدرتين، وكذا تقدير بمجال للخطأ العشوائي المقدر.

ليكن x متغير عشوائي يتبع التوزيع الطبيعي حيث $X \mapsto N(\mu, \delta^2)$ وفق قانون الاحتمال:

$$p(X) = \frac{1}{\sqrt{2\lambda\delta^2}} e^{-\frac{1}{2}\left(\frac{X-\mu}{\delta}\right)^2}$$

فإذا كان Z متغير عشوائي آخر حيث $Z = \frac{X-\mu}{\delta}$، فإن Z يتوزع وفق قانون التوزيع الطبيعي المعياري حيث $Z \mapsto N(0,1)$، وبالرجوع إلى توزيع ستيودنت الذي ينص على أنه إذا كان للمتغير العشوائي Z توزيعا طبيعيا معياريا، ولـ V^2 توزيع كيدو $V^2 \mapsto \chi^2(r)$ مستقل بدرجة حرية عددها r، فإن المقدار $t = \frac{Z\sqrt{r}}{v}$ يتوزع وفق توزيع ستيودنت بدرجات حرية r وفق القانون الاحتمالي التالي:

$$p(t) = c(1 + \frac{t^2}{r})^{\frac{-v+1}{2}}$$

4-4-1 الانحراف المعياري للتقدير لكل من $\hat{a}, \hat{b}, \hat{\sigma}_\mu^2$

بالاعتماد على خصائص تقديرات المربعات الصغرى، والتي تعتبر مقدراتها غير متحيزة، فإن الانحراف المعياري للتقدير لكل من $\hat{a}, \hat{b}, \hat{\sigma}_\mu^2$ يعطى بالعلاقات التالية:

$$\hat{\sigma}_\mu^2 = \frac{\sum e_i^2}{n-k}$$

$$\hat{\sigma}_a = \sqrt{\frac{\sum X_i^2}{n\sum x_i^2}\sigma_\mu^2}$$

$$\hat{\sigma}_b = \sqrt{\frac{1}{\sum x_i^2}\sigma_\mu^2}$$

4-4-2 تشكيل مجالات الثقة للمعلمة المقدرة $\hat{a}$

إذا كان $Z = \frac{\hat{a}-a}{\sigma_{\hat{a}}}$ وكان $\hat{\delta}_\mu^2 = \frac{\sum ei^2}{n-2}$ فإن $V^2 = \frac{\sum ei^2}{\sigma_\mu^2} = \frac{(n-2)\hat{\sigma}_\mu^2}{\sigma_\mu^2}$ ذو توزيع كيدو مستقل بـ (n−2) درجة حرية لذلك فإن:

$$t = \frac{Z\sqrt{r}}{v} = \frac{\dfrac{\hat{a}-a}{\sigma_{\hat{a}}}\sqrt{n-2}}{\dfrac{\sqrt{n-2}\sqrt{\hat{\sigma}_\mu^2}}{\sqrt{\sigma_\mu^2}}} = \frac{(\hat{a}-a)\sqrt{n-2}\,\sigma_\mu}{\sigma_{\hat{a}}\sqrt{n-2}\,\hat{\sigma}_\mu} = \frac{(\hat{a}-a)\sigma_\mu}{\dfrac{\sigma_\mu}{\sqrt{\sum x_i^2}}\hat{\sigma}_\mu} = \frac{(\hat{a}-a)\sigma_\mu\sqrt{\sum x_i^2}}{\sigma_\mu\hat{\sigma}_\mu} = \frac{(\hat{a}-a)\sqrt{\sum x_i^2}}{\hat{\sigma}_\mu}$$

ذو توزيع ستيودنت بـ (n−2) درجة حرية.

حيث أن الانحراف المعياري للتقدير للمعلمة المقدرة هي: $\hat{\sigma}_a = \sqrt{\frac{1}{\sum x_i^2}\hat{\sigma}_\mu^2}$، لذا يمكننا تحويل العلاقة بالشكل:

$$\hat{\sigma}_a = \sqrt{\frac{\sum X_i^2}{n\sum x_i^2}\sigma_\mu^2}$$

$$= \sigma_\mu \frac{\sqrt{\sum X_i^2}}{\sqrt{n\sum x_i^2}} \Longrightarrow \hat{\sigma}_{\hat{a}} = \hat{\sigma}_\mu \frac{\sqrt{\sum X_i^2}}{\sqrt{n\sum x_i^2}}$$

إذن: $\frac{\hat{a}-a}{\hat{\sigma}_{\hat{a}}} = t = \frac{\hat{a}-a}{\hat{\sigma}_\mu / \sqrt{\sum x_i^2}}$ حيث $\hat{\sigma}_{\hat{a}}$ هو تقدير للانحراف المعياري لـ $\hat{a}$.

نستنتج من هذا أنه يتم الحصول على مجال الثقة من أجل المعلمة b وفق ما يلي:

$p(-t_{\frac{\alpha}{n-k}} \langle t \langle t_{\frac{\alpha}{n-k}}) = 1-\alpha$ و بتعويض قيمة t نجد:

$$\begin{aligned} & p(-t_{\frac{\alpha}{n-k}} \langle \frac{\hat{a}-a}{\hat{\sigma}_{\hat{a}}} \langle t_{\frac{\alpha}{n-k}}) \\ &= p(-t_{\frac{\alpha}{n-k}} \hat{\sigma}_{\hat{a}} \langle \hat{b}-b \langle t_{\frac{\alpha}{n-k}} \hat{\sigma}_{\hat{a}}) \\ &= p(-\hat{b}-t_{\frac{\alpha}{n-k}} \hat{\sigma}_{\hat{a}} \langle -b \langle -\hat{b}+t_{\frac{\alpha}{n-k}} \hat{\sigma}_{\hat{a}}) \\ &= p(\hat{b}+t_{\frac{\alpha}{n-k}} \hat{\sigma}_{\hat{a}} \rangle b \rangle \hat{b}-t_{\frac{\alpha}{n-k}} \hat{\sigma}_{\hat{a}}) \end{aligned}$$

أي أن مجال الثقة للمعلمة a هو:

$$p\left(\hat{a} - t_{\frac{\alpha}{n-k}} \hat{\sigma}_{\hat{a}} < a < \hat{a} - t_{\frac{\alpha}{n-k}} \hat{\sigma}_{\hat{a}}\right) = 1-\alpha$$

بحيث: $\hat{\delta}_{\hat{a}} = \sqrt{\frac{\sum X^2}{n\sum x_i^2}\hat{\delta}_\mu^2}$

3-4-4 تشكيل مجالات الثقة للمعلمة المقدرة $\hat{b}$

إذا كان $Z = \frac{\hat{b}-b}{\sigma_{\hat{b}}}$ وكان $\hat{\delta}_\mu^2 = \frac{\sum ei^2}{n-2}$ فإن $V^2 = \frac{\sum ei^2}{\sigma_\mu^2} = \frac{(n-2)\hat{\sigma}_\mu^2}{\sigma_\mu^2}$ ذو توزيع كاي-مربع مستقل بـ (n-2) درجة حرية لذلك فإن:

$$t=\frac{Z\sqrt{r}}{v}=\frac{\frac{\hat{b}-b}{\sigma_{\hat{b}}}\sqrt{n-2}}{\frac{\sqrt{n-2}\sqrt{\hat{\sigma}_{\mu}^2}}{\sqrt{\sigma_{\mu}^2}}}=\frac{(\hat{b}-b)\sqrt{n-2}\sigma_{\mu}}{\sigma_{\hat{b}}\sqrt{n-2}\hat{\sigma}_{\mu}}=\frac{(\hat{b}-b)\sigma_{\mu}}{\frac{\sigma_{\mu}}{\sqrt{\sum x_i^2}}\hat{\sigma}_{\mu}}=\frac{(\hat{b}-b)\sigma_{\mu}\sqrt{\sum x_i^2}}{\sigma_{\mu}\hat{\sigma}_{\mu}}=\frac{(\hat{b}-b)\sqrt{\sum x_i^2}}{\hat{\sigma}_{\mu}}$$

ذو توزيع ستيودنت بـ (n−2) درجة حرية.

حيث أن الانحراف المعياري للتقدير للمعلمة المقدرة هي: $\hat{\sigma}_b=\sqrt{\frac{1}{\sum x_i^2}\hat{\sigma}_{\mu}^2}$، لذا يمكننا تحويل العلاقة بالشكل:

$$\sigma_{\hat{b}}=\sqrt{\frac{1}{\sum x_i^2}\hat{\sigma}_{\mu}^2}=\frac{\sigma_{\mu}}{\sqrt{\sum x_i^2}}\Longrightarrow\hat{\sigma}_{\hat{b}}=\frac{\hat{\sigma}_{\mu}}{\sqrt{\sum x_i^2}}$$

إذن: $t=\frac{\hat{b}-b}{\hat{\sigma}_{\mu}/\sqrt{\sum x_i^2}}=\frac{\hat{b}-b}{\hat{\sigma}_{\hat{b}}}$ حيث $\hat{\sigma}_{\hat{b}}$ هو تقدير للانحراف المعياري لـ $\hat{b}$.

نستنتج من هذا أنه يتم الحصول على مجال الثقة من أجل المعلمة b وفق ما يلي:

$p(-t_{\frac{\alpha}{n-k}}\langle t\langle t_{\frac{\alpha}{n-k}})=1-\alpha$ وبتعويض قيمة t نجد:

$$\begin{aligned}&p(-t_{\frac{\alpha}{n-k}}\langle\frac{\hat{b}-b}{\hat{\sigma}_{\hat{b}}}\langle t_{\frac{\alpha}{n-k}})\\&=p(-t_{\frac{\alpha}{n-k}}\hat{\sigma}_{\hat{b}}\langle\hat{b}-b\langle t_{\frac{\alpha}{n-k}}\hat{\sigma}_{\hat{b}})\\&=p(-\hat{b}-t_{\frac{\alpha}{n-k}}\hat{\sigma}_{\hat{b}}\langle -b\langle -\hat{b}+t_{\frac{\alpha}{n-k}}\hat{\sigma}_{\hat{b}})\\&=p(\hat{b}+t_{\frac{\alpha}{n-k}}\hat{\sigma}_{\hat{b}}\rangle b\rangle\hat{b}-t_{\frac{\alpha}{n-k}}\hat{\sigma}_{\hat{b}})\end{aligned}$$

أي أن مجال الثقة للمعلمة b هو:

$$p\left(\hat{b}-t_{\frac{\alpha}{n-2}}\hat{\sigma}_{\hat{b}}<b<\hat{b}-t_{\frac{\alpha}{n-2}}\hat{\sigma}_{\hat{b}}\right)=1-\alpha$$

بحيث: $\hat{\delta}_{\hat{b}} = \sqrt{\frac{1}{\sum x_i^2}\hat{\delta}_\mu^2}$

4-4-4 تشكيل مجال ثقة للخطأ العشوائي $\hat{\sigma}_\mu^2$

يمكن أن نستنتج مجال الثقة لتباين حد الخطأ من خلال التعبير الاحتمالي التالي:

$$p(\chi^2_{\underset{n-k}{1-\alpha}} \langle v \langle \chi^2_{\underset{n-k}{\alpha}}) = 1-\alpha \Rightarrow p(\chi^2_{\underset{n-k}{1-\alpha}} \langle \frac{(n-2)\hat{\sigma}_\mu^2}{\sigma_\mu^2} \langle \chi^2_{\underset{n-k}{\alpha}})$$

$$= p(\frac{\chi^2_{\underset{n-k}{1-\alpha}}}{(n-2)\hat{\sigma}_\mu^2} \langle \frac{1}{\delta_\mu^2} \langle \frac{\chi^2_{\underset{n-k}{\alpha}}}{(n-2)\hat{\sigma}_\mu^2}) = 1-\alpha$$

وهذا التعبير الاحتمالي الأخير يعطي حدي الثقة 1−α لأجل δ_μ^2 وهما:

$$\delta_\mu^2 = \left] \frac{(n-2)\hat{\delta}_\mu^2}{\chi^2_{\underset{n-k}{\alpha}}}, \frac{(n-2)\hat{\delta}_\mu^2}{\chi^2_{\underset{n-k}{1-\alpha}}} \right[$$

مثال رقم (3−2):

بأخذ المثال السابق أوجد مجال ثقة لكل من σ_u^2, a, b بدرجة ثقة 95%؟

__الحل:__

n	X	Y	$x = X - \bar{X}$	$y = Y - \bar{Y}$	$x.y$	x^2	$\hat{Y}$	$ei = Y_i - \hat{Y}_i$	ei^2	$(Y_i - \bar{Y})^2$	X^2
0	0	3	−2	−1	2	4	3	0	0	1	0
1	1	3	−1	−1	1	1	3.5	−0.5	0.25	1	1
2	1	4	−1	0	0	1	3.5	+0.5	0.25	0	1
3	3	4	1	0	0	1	4.5	−0.5	0.25	0	9
4	3	5	1	1	1	1	4.5	+0.5	0.25	1	9
5	4	5	2	1	1	4	5	0	0	1	16
$\sum$	12	24	0	0	6	12		0	1	4	36

1 − حساب $\hat{\sigma}_\mu^2, \sigma_{\hat{b}}, \sigma_{\hat{a}}$

$$\hat{\sigma}_{\mu}^{2}=\frac{\sum ei^{2}}{n-2}=\frac{1}{4},$$

$$\sigma_{\hat{b}}=\sqrt{\frac{1}{\sum x_i^2}\hat{\delta}_{\mu}^{2}}=\sqrt{\frac{1}{12}0.25}=0.144,$$

$$\sigma_{\hat{a}}=\sqrt{\frac{\sum X^2}{n\sum x_i^2}\hat{\sigma}_{\mu}^{2}}=\sqrt{\frac{36}{6\times 12}0.25}=0.353$$

2 – إيجاد مجالات الثقة للمعالم:

$$b\in\left]\hat{b}-t_{\frac{\alpha}{n-k}}\hat{\sigma}_{\hat{b}},\hat{b}+t_{\frac{\alpha}{n-k}}\hat{\sigma}_{\hat{b}}\right[=\left]0.5-2.776\times 0.144,0.5+2.776\times 0.144\right[=\left]0.1,0.9\right[$$

$$a\in\left]\hat{a}-t_{\frac{\alpha}{n-k}}\hat{\sigma}_{\hat{a}},\hat{a}+t_{\frac{\alpha}{n-k}}\hat{\sigma}_{\hat{a}}\right[=\left]3-2.776\times 0.353,3+2.776\times 0.353\right[=\left]2.02,3.98\right[$$

$$\delta_{\mu}^{2}=\left]\frac{(n-2)\hat{\sigma}_{\mu}^{2}}{\chi^{2}_{\frac{\alpha}{n-k}}},\frac{(n-2)\hat{\sigma}_{\mu}^{2}}{\chi^{2}_{\frac{1-\alpha}{n-k}}}\right[=\left]\frac{4\times 0.25}{9.488},\frac{4\times 0.25}{0711}\right[=\left]0.105,1.406\right[$$

4-5 الاختبارات الإحصائية

4-5-1 اختبار ستيودنت (اختبار معنوية المعلمتين $\hat{a},\hat{b}$)

نهتم دائما باختبار معنوية المعلمتين $\hat{a},\hat{b}$. وهذا معناه اختبار فرض أن تكون كل منهما تختلف عن الصفر، وذلك وفق الخطوات التالية:

أولا: بالنسبة للمعلمة $\hat{a}$

نضع الفرض التالي: $\begin{matrix} H_0: a=0 \\ H_1: a\neq 0 \end{matrix}$ ثم نحسب $t_{\text{المحسوبة}}$ $t=\frac{\hat{a}}{\sigma_{\hat{a}}}$ ثم نحدد قيمة $t_{\text{الجدولية}}$.

- إذا كانت قيمة $t_{\text{المحسوبة}}$ تقع ضمن منطقة الرفض، فإننا نرفض H_0 ونقبل H_1، وهذا يعني أن a تختلف معنويا عن الصفر، و بالتالي ليس هناك تناسب بين المتغيرين.

- إذا كانت قيمة t المحسوبة تقع ضمن منطقة القبول، فإننا نقبل H_0 ونرفض H_1، وهذا يعني أن a لا تختلف معنويا عن الصفر، و هذا يدل على وجود تناسب بين المتغيرين X و Y.

ثانيا: بالنسبة للمعلمة $\hat{b}$

نضع الفرض التالي: $\begin{matrix} H_0 : b = 0 \\ H_1 : b \neq 0 \end{matrix}$ ثم نحسب t المحسوبة $t = \dfrac{\hat{b}}{\sigma_{\hat{b}}}$ ثم نحدد قيمة t الجدولية.

- إذا كانت قيمة t المحسوبة تقع ضمن منطقة الرفض، فإننا نرفض H_0 ونقبل H_1، وهذا يعني أن b تختلف معنويا عن الصفر، ما يدل على معنوية العلاقة الخطية بين X و Y، أي أنها غير راجعة للصدفة.
- إذا كانت قيمة t المحسوبة تقع ضمن منطقة القبول، فإننا نقبل H_0 ونرفض H_1، وهذا يعني أن b لا تختلف معنويا عن الصفر، وبالتالي فالعلاقة المقدرة بين X و Y راجعة للصدفة.

مثال رقم (4-3):

بأخذ المثال السابق اختبر معنوية $\hat{a},\hat{b}$؟

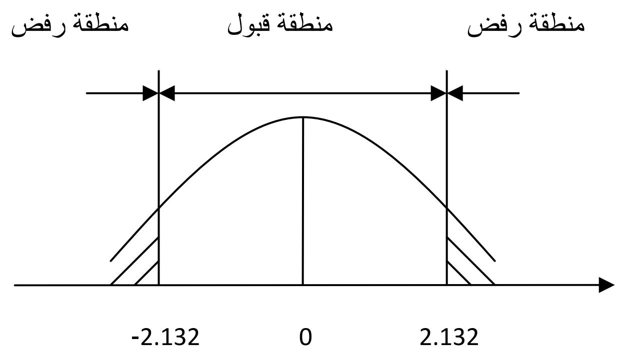
الحل:

أ – اختبار معنوية $\hat{b}$ عند مستوي المعنوية 10%؟

$$H_0 : b = 0$$
$$H_1 : b \neq 0$$

$$t = \frac{\hat{b}}{\sigma_{\hat{b}}} = \frac{0.5}{0.144} = 3.47$$

t الجدولية $=2.132$



بما أن قيمة t المحسوبة تقع ضمن منطقة الرفض، فإننا نرفض H_0 ونقبل H_1، وهذا يعني أن b تختلف معنويا عن الصفر، ما يدل على معنوية العلاقة الخطية بين X و Y، أي أنها غير راجعة للصدفة.

أ – اختبار معنوية $\hat{a}$ عند مستوي المعنوية 10%؟

$$H_0 : a = 0$$
$$H_1 : a \neq 0$$

$$t = \frac{\hat{a}}{\sigma_{\hat{a}}} = \frac{3}{0.353} = 8.49$$

$t_{\text{الجدولية}} = 2.132$

بما أن قيمة t المحسوبة تقع ضمن منطقة الرفض، فإننا نرفض H_0 ونقبل H_1، وهذا يعني أن a تختلف معنويا عن الصفر، و بالتالي ليس هناك تناسب بين المتغيرين.

4-5-2 اختبارات الفروض الخاصة بالمقدرين

أولا: بالنسبة للمعلمة المقدرة $\hat{a}$

إذا ما أردنا اختبار معنوية الفرق بين قيمة a المقدرة و هي $\hat{a}$ وبين قيمة أخرى محددة ولتكن a_0.

أولا: نضع الفرض التالي: $\begin{matrix} H_0 : a = a_0 \\ H_1 : a \neq a_0 \end{matrix}$ ثم نحسب t المحسوبة $t = \frac{\hat{a} - a_0}{\sigma_{\hat{b}}}$ ثم نحدد قيمة t الجدولية.

- إذا كانت قيمة t المحسوبة تقع ضمن منطقة الرفض، فإننا نرض H_0 ونقبل H_1، ومعنى ذلك أن الفرق بين $\hat{a}$ و a_0 معنوي أي غير راجع للصدفة، وبالتالي نقبل فرض أن $\hat{a}$ تختلف معنويا عن a_0.
- إذا كانت قيمة t المحسوبة تقع ضمن منطقة القبول، فإننا نقبل H_0 ونرفض H_1، ومعنى ذلك أن الفرق بين $\hat{a}$ و a_0 ليس معنوي أي راجع للصدفة، وبالتالي نقبل فرض أن $\hat{a}$ لا تختلف معنويا عن a_0.

ثانيا: بالنسبة للمعلمة المقدرة $\hat{b}$

إذا ما أردنا اختبار معنوية الفرق بين قيمة b المقدرة و هي $\hat{b}$ وبين قيمة أخرى محددة ولتكن b_0.

أولا: نضع الفرض التالي: $\begin{matrix} H_0: b = b_0 \\ H_1: b \neq b_0 \end{matrix}$ ثم نحسب t المحسوبة $t = \frac{\hat{b} - b_0}{\sigma_{\hat{b}}}$ ثم نحدد قيمة t الجدولية.

- إذا كانت قيمة t المحسوبة تقع ضمن منطقة الرفض، فإننا نرض H_0 ونقبل H_1، ومعنى ذلك أن الفرق بين $\hat{b}$ و b_0 معنوي أي غير راجع للصدفة، وبالتالي نقبل فرض أن $\hat{b}$ تختلف معنويا عن b_0.
- إذا كانت قيمة t المحسوبة تقع ضمن منطقة القبول، فإننا نقبل H_0 ونرفض H_1، ومعنى ذلك أن الفرق بين $\hat{b}$ و b_0 ليس معنوي أي راجع للصدفة، وبالتالي نقبل فرض أن $\hat{b}$ لا تختلف معنويا عن b_0.

مثال رقم (4-4):

هل يمكن لقيمة b أن تكون أكبر من 0.6 عند مستوي المعنوية 10% ؟

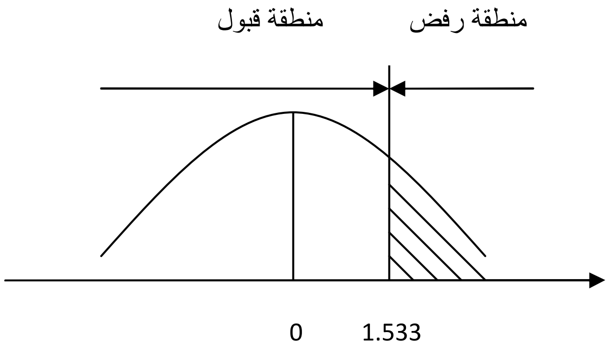
الحل:

اختبار فرض أن b أكبر من 0.6: (اختبار الذيل الأيمن)

$$\begin{matrix} H_0: b = 0.6 \\ H_1: b \rangle 0.6 \end{matrix},$$

$$t = \frac{\hat{b} - b_0}{\sigma_{\hat{b}}} = \frac{0.5 - 0.6}{0.144} = -0.694$$

$t_{\text{الجدولية}}=1.533$



بما أن قيمة t المحسوبة تقع ضمن منطقة القبول فإننا نقبل H0 و نرفض H1، ومنه لا يمكن لقيمة b أن تكون أكبر من 0.6.

3-5-4 اختبار جودة التوفيق والارتباط

في هذا الجزء سوف نبين القدرة التفسيرية للمتغير المستقل في تفسير المتغير التابع.

أولا: معامل التحديد R^2

يعرف معامل التحديد بأنه "نسبة التغير الإجمالي في Y الذي يفسره خط الانحدار أي هو مقدار الدقة في التقدير".

يمكن إيجاد قيمة معامل التحديد كما يلي:

- إثبات أن $\sum e_i = 0$

$$Y_i = \hat{a} + \hat{b}X_i + e_i$$

$$\Rightarrow \sum Y_i = n\hat{a} + \hat{b}\sum X_i + \sum e_i$$

$$\Rightarrow \sum Y_i - n\hat{a} - \hat{b}\sum X_i = \sum e_i$$

$$\Rightarrow \sum Y_i - n(\bar{Y} - \hat{b}\bar{X}) - \hat{b}\sum X_i = \sum e_i$$

$$\Rightarrow \sum Y_i - n(\frac{\sum Y_i}{n} - \hat{b}\frac{\sum X_i}{n}) - \hat{b}\sum X_i = \sum e_i$$

$$\Rightarrow \sum Y_i - \sum Y_i + \hat{b} \sum X_i - \hat{b} \sum X_i = \sum e_i$$

$$\sum e_i = 0$$

من خلال هذه المبرهنة يمكننا إيجاد العلاقة التالية:

التغير الإجمالي بين القيم الحقيقية والمتوسط الحسابي يساوي (أنظر الشكل رقم 4−1):

$$(Y_i - \bar{Y}) = (Y_i - \hat{Y}_i) + (\hat{Y}_i - \bar{Y})$$

$$(Y_i - \bar{Y})^2 = (Y_i - \hat{Y}_i)^2 + (\hat{Y}_i - \bar{Y})^2 + 2[(Y_i - \hat{Y}_i)(\hat{Y}_i - \bar{Y})]$$

$$\sum (Y_i - \bar{Y})^2 = \sum (Y_i - \hat{Y}_i)^2 + \sum (\hat{Y}_i - \bar{Y})^2 + 2 \sum [(Y_i - \hat{Y}_i)(\hat{Y}_i - \bar{Y})]$$

لدينا:

$$\sum [(Y_i - \hat{Y}_i)(\hat{Y}_i - \bar{Y})]$$

$$= \sum (Y_i - \hat{Y}_i) \sum (\hat{Y}_i - \bar{Y})$$

$$= \sum (e_i) \sum (\hat{Y}_i - \bar{Y})$$

$$= 0 \sum (\hat{Y}_i - \bar{Y})$$

إذن:

$$\sum (Y_i - \bar{Y})^2 = \sum (Y_i - \hat{Y}_i)^2 + \sum (\hat{Y}_i - \bar{Y})^2$$

نرمز بـ:

SST: لمجوع المربعات الإجمالي: $\sum (Y_i - \bar{Y})^2$ Sum-of-Squares Total

SSR: لمجوع المربعات الإجمالي: $\sum (\hat{Y}_i - \bar{Y})^2$ Sum-of-Squares Regression

SSE: لمجوع المربعات الإجمالي: $\sum (Y_i - \hat{Y}_i)^2$ Sum-of-Squares Error

بقسمة طرفي المعادلة السابقة على مجموع المربعات الإجمالي نجد:

$$\frac{SST}{SST} = \frac{SSE}{SST} + \frac{SSR}{SST}$$

$$1 = \frac{SSE}{SST} + \frac{SSR}{SST}$$

$$\frac{SSR}{SST} = R^2 = 1 - \frac{SSE}{SST}$$

تتراوح قيمة معامل التحديد بين 0 و1، و كلما اقتربت من 1 كلما زادت دقة التقدير.

ثانيا: معامل الارتباط r

هو درجة التشابك والترابط بين المتغيرين X و Y.

يمكن إيجاد معامل الارتباط من معامل التحديد بالعلاقة التالية:

$r = \pm\sqrt{R^2}$ (تأخذ الإشارة على حسب إشارة المعلمة $\hat{b}$)

تتراوح قيمة معامل الارتباط بين −1 و +1 ، بحيث:

- r = 0 : لا توجد علاقة ارتباط بين X و Y،
- r =−1 : يوجد ارتباط خطي سالب تام،
- r =+1 : يوجد ارتباط خطي موجب تام.

أي أن قوة الارتباط تزداد كلما اقتربت قيمة معامل الارتباط من +1 أو −1 وتنخفض كلما اقتربت من 0.

مثال رقم (4−5):

بأخذ المثال السابق أوجد معامل التحديد والارتباط مع التعليق؟

الحل:

أ – حساب معامل التحديد:

$$R^2 = 1 - \frac{\sum ei^2}{\sum \left(Y_i - \overline{Y}\right)^2} = 1 - \frac{1}{4} = 0.75$$

أي أن خط الانحدار يفسر ما مقداره 75% من التغير في Y و الباقي 25% يعود للخطأ العشوائي.

ب – حساب معامل الارتباط:

$r = \pm\sqrt{R^2} = +\sqrt{0.75} = +0.866$ (يتم أخذ الإشارة الموجبة حسب إشارة $\hat{b}$)

يوجد ارتباط طردي قوي بين المتغيرين.

4-5-4 اختبار فيشر

يهتم اختبار فيشر باختبار المعنوية الكلية للنموذج وسلامة الشكل الرياضي المختار لصياغة العلاقة بين المتغيرين. وهذا تحت الفرض التالي:

$$H_0: a = b = 0$$
$$H_1: a \neq b \neq 0$$

ثم نقوم بإعداد جدول تحليل التباين ANOVA

مصدر التباين	مجموع المربعات	درجة الحرية	متوسط مجموع المربعات	إحصائية F
الانحدار	SSR	1	$MSR = SSR$	$F=\frac{MSR}{MSE}$
البواقي	SSE	n−2	$MSE = \frac{SSE}{n-2}$	
المجموع	$SST=SSR+SSE$	n−1	−	

بعد ذلك نقوم باستخراج قيمة $F_{\text{الجدولية}}$ من جدول فيشر وذلك بدرجة حرية بسط قدرها 1 ودرجة حرية مقام قدرها n−2 ومستوى معنوية مناسب.

- إذا كانت $F_{\text{المحسوبة}}$ أكبر من $F_{\text{الجدولية}}$ فهذا يعني أننا نرفض H_0 ونقبل H_1 ومنه فالنموذج معنوي إجمالا. والشكل الرياضي المختار للمعادلة سليم إحصائيا.

مثال رقم (4-6):

اختبر المعنوية الكلية للنموذج المقدر في المثال السابق عند مستوى معنوية قدره 5%؟

الحل:

نضع الفرض:

$$H_0: a = b = 0$$
$$H_1: a \neq b \neq 0$$

ثم نقوم بإعداد جدول تحليل التباين ANOVA

مصدر التباين	مجموع المربعات	درجة الحرية	متوسط مجموع المربعات	إحصائية F
الانحدار	3	1	$MSR = 3$	$F=\frac{3}{1/4} = 12$
البواقي	1	4	$MSE = \frac{1}{4}$	
المجموع	4	5	–	

عند درجة حرية بسط قدرها 1 ودرجة حرية مقام قدرها 4 ومستوى معنوية قدرها 5%، فإن قيمة فيشر الجدولية تساوي 7.71.

بما أن قيمة $F_{المحسوبة}$ أكبر من $F_{الجدولية}$ فهذا يعني أننا نرفض H_0 ونقبل H_1 ومنه فالنموذج معنوي إجمالا. والشكل الرياضي المختار للمعادلة سليم إحصائيا.

4-6 التوقع باستخدام نموذج الانحدار الخطي البسيط

إن أحد أهم استعمالات خط الانحدار المقدر هو التوقع بالقيم المستقبلية للمتغير التابع Y المقابلة لقيم معينة خاصة بالمتغير المستقلX، و لتكن X_0 على سبيل المثال. يمكن الحصول على التوقع بالصيغة التالية:

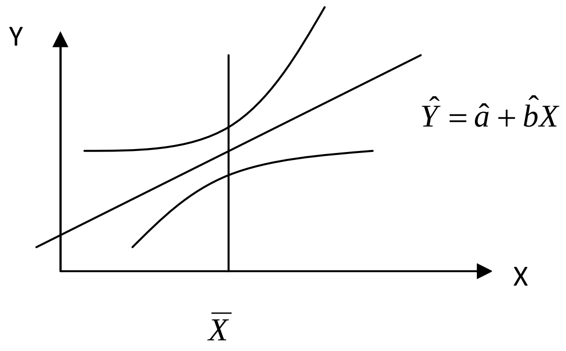
$$\hat{Y}_0 = \hat{a} + \hat{b}X_0$$

ويعتبر هذا التوقع شرطيا حيث $\hat{Y}_0$ مشروطة بقيمة المتغير التفسيري X_0. كما تتوقف صحة التوقع على شرط آخر وهو ثبات العلاقة الهيكلية الرابطة بين X وY بحث تنطبق على فترة التنبؤ مثلما انطبقت على فترة التقدير، وحدد الانحراف المعياري التنبؤ بالعلاقة التالية:

$$\sigma \hat{y}_0 = \sqrt{\hat{\sigma}_\mu^2 \left[1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_i^2}\right]}$$

نلاحظ على هذه العبارة ما يلي:

- تصغر كلما كبر حجم العينة و بالتالي يكون التنبؤ أكثر دقة،
- تصغر كلما اقتربت X_0 من $\bar{X}$ و يكون التوقع أكثر دقة. وهذا ما يمكن توضيحه بالشكل الموالي:



ويمكن حساب مجال ثقة للتوقع بالعلاقة التالية:

$$p(\hat{Y}_0 - t_{\underset{n-2}{\alpha}}\ \sigma\hat{y}_o \langle Y_0 \langle \hat{Y}_0 + t_{\underset{n-2}{\alpha}}\ \sigma\hat{y}_0) = 1-\alpha$$

كما يمكن إجراء اختبارات الفروض حول القيم المتوقع بها حسب الصيغة المعروفة لاختبار ستيودنت كالتالي:

أولا: نضع الفرض التالي: $\begin{matrix} H_0 : Y = Y_0 \\ H_1 : Y \neq Y_0 \end{matrix}$ ثم نحسب t المحسوبة $t = \dfrac{\hat{Y}_0 - Y_0}{\sigma\hat{y}_0}$ ثم نحدد قيمة t الجدولية.

- إذا كانت قيمة t المحسوبة تقع ضمن منطقة الرفض، فإننا نرفض H_0 ونقبل H_1، و بالتالي لا يمكن لقيمة Y أن تساوي Y_0،
- إذا كانت قيمة t المحسوبة تقع ضمن منطقة القبول، فإننا نقبل H_0 ونرفض H_1، و بالتالي يمكن لقيمة Y أن تساوي Y_0.

مثال:

بأخذ المثال السابق تحصل على تنبؤ لقيمة الإنفاق الاستهلاكي لمستوى من الدخل مقداره 2، ثم أحسب فترة الثقة 95% له؟

الحل:

$$\hat{Y}_0 = \hat{a} + \hat{b}X_0 = 3 + 0.5(2) = 4$$

أي أن الإنفاق الاستهلاكي سيبلغ 4 وحدات إذا كان الدخل يساوي 2 وحدة.

حساب:

$$\delta\hat{y}_0 = \sqrt{\hat{\sigma}_\mu^2\left[1+\frac{1}{n}+\frac{\left(X_0-\overline{X}\right)^2}{\sum x_i^2}\right]} = \sqrt{0.25\left[1+\frac{1}{6}+\frac{(2-2)^2}{12}\right]} = 0.54$$

تحديد مجال الثقة للتنبؤ:

$$p(\hat{Y}_0 - t_{\substack{\alpha \\ n-2}}\,\sigma\hat{y}_o \langle Y_0 \langle \hat{Y}_0 + t_{\substack{\alpha \\ n-2}}\,\sigma\hat{y}_0) = 1-\alpha$$

$$\Rightarrow\ p(4-2.776\times0.54\langle Y_0\langle 4+2.776\times0.54) = 0.95$$

$$\Rightarrow\ p(2.5\langle Y_0\langle 5.5) = 0.95$$

4-7 تمارين محلولة

تمرين رقم (4-1):

1 – ما هو الفرق بين البواقي ei و الأخطاء μ_i ؟

2 – ما هو الفرق بين a,b من ناحية و $\hat{a},\hat{b}$ من ناحية أخرى؟

3 – لماذا نأخذ مجموع مربع انحرافات القيم $\sum e_i^2$ عند تقدير معلمات النموذج دون الأخذ مباشرة بمجموع انحرافات القيم ∑ei؟

4 – لماذا نأخذ الانحرافات الرأسية دون الأفقية عند عملية التقدير؟

5– ما هو مفهومك لخاصية انعدام الارتباط الذاتي بين الأخطاء؟

الحل:

1– الفرق بين ei و μ_i:

μ_i: هي حط الخطأ في العلاقة الحقيقية غير المعلومة بين X و Y،

ei: هي البواقي بين كل القيم المشاهدة Y والقيم المقدرة المناظرة لها $\hat{Y}$ في العلاقة المقدرة.

2– الفرق بين a,b من ناحية و $\hat{b}, \hat{a}$.

a,b: هي معالم خط الانحدار الحقيقي غير المعلوم.

$\hat{b}, \hat{a}$: هي معالم خط الانحدار المقدر.

3– لا يمكننا أخذ مجموع انحرافات القيم لكل مشاهدة عن خط الانحدار لأن مجموع هذه الانحرافات مختلفة الإشارة يكون مساويا للصفر.

4– نأخذ الانحرافات الرأسية لأننا نحاول أن نفسر وأن نتوقع بالتغيرات في Y التي تقاس على المحور الرأسي.

5- يشير الارتباط الذاتي بشكل عام إلى وجود ارتباط بين قيم مشاهدات نفس المتغير. وفي نموذج الانحدار، عادة ما يشير الارتباط الذاتي إلى وجود ارتباط بين القيم المتتالية للخطأ العشوائي μ_i . وخاصية انعدام الارتباط الذاتي بين الأخطاء، تفرض أن تكون قيمة معامل التغاير بين قيم الخطأ العشوائي μ_i تساوي 0. ويفسر هذا المفهوم من الناحية الاقتصادية على أن خطأ ما وقع في فترة سابقة لا يؤثر في أخطاء فترات متتالية بطريقة تؤدي إلى تكرار الخطأ.

تمرين رقم (4-2):

بفرض توفر البيانات الخاصة بمقدار المبيعات اليومية لإحدى عشر عامل(بالألف دينار) في محل تجاري حسب مدة الخدمة (بالسنين) والتي يلخصها الجدول الموالي:

n	1	2	3	4	5	6	7	8	9	10	11
المبيعات اليومية	7	17	15	14	13	10	8	12	11	9	5
مدة الخدمة	4	13	11	12	12	8	6	11	10	9	3

المطلوب:

1 – ضع هذه البيانات في شكل انتشار؟ ما هي ملاحظاتك هو الشكل المتحصل عليه؟

2- أوجد خط الانحدار المقدر مع التمثيل البياني؟ ثم أحسب كل من: أ- تباين الحد العشوائي، ب- الانحراف المعياري للتقدير لـ $\hat{a}, \hat{b}$؟

3- أوجد مجال الثقة للمعلمة a وهذا بمستوى دلالة قدره 5%؟

4- أوجد مجال الثقة للمعلمة b و هذا بدرجة ثقة قدرها 90%؟

5 – اختبار فرض أن قيمة b = 0.5 وذلك بدرجة ثقة قدرها 80%؟

6 – اختبار فرض أن قيمة a = 1 ، وذلك بدرجة ثقة قدرها 95%؟

7 – نفترض أن قيمة $b > 1.5$ ، هل يمكن قبول هذا الفرض إذا علمت أن $t_{\text{الجدولية}} = 1.833$ ($\alpha =$ 5%) ؟

8 – هل يمكن القول أن خط الانحدار المقدر هو ذو معنوية إحصائية عند مستوى المعنوية 5%؟

9 – هل يمكن القول أن هناك تناسب بين المتغيرين x و y عند مستوى المعنوية 5%؟

10 – أحسب معامل التحديد ومعامل الارتباط مع التعليق على النتيجة؟

11 – أوجد مجال ثقة لتباين حد الخطأ عند مستوى معنوية قدره 5%؟

12 – أوجد توقع لمقدار المبيعات اليومية إذا علمت أن مدة خدمة العامل هي 15 سنة، مع تحديد مجال الثقة 90% له؟

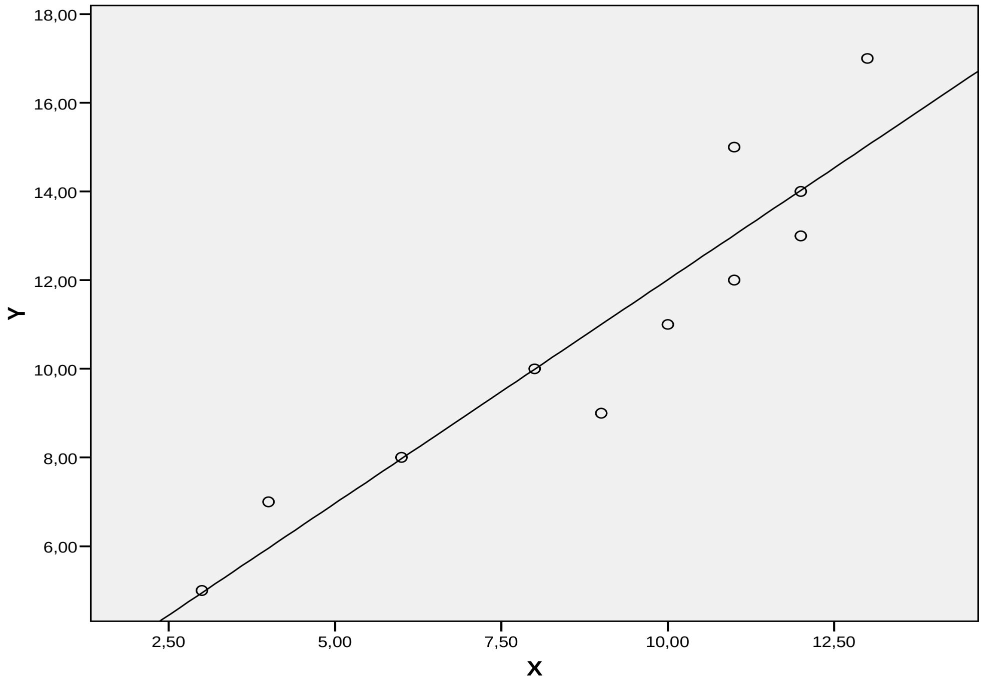
13 – اختبر فرض أن المبيعات اليومية تساوي 14 إذا كانت مدة الخدمة هي 13 سنة عند مستوى معنوية 5%؟

14– اختبر المعنوية الكلية للمعلمات المقدرة عند مستوى معنوية قدره 1%؟

الحل:

1– تمثيل شكل الانتشار مع التعليق:

حيث أن مدة الخدمة هي المتغير المستقل الذي يؤثر في المبيعات اليومية فإننا نعتبرها متغير مستقل X بينما المبيعات اليومية هي المتغير التابع Y.



يظهر لنا شكل الانتشار وجود علاقة خطية بين المتغير المستقل والمتغير التابع.

2- إيجاد خط الانحدار المقدر مع التمثيل البياني:

لإيجاد خط الانحدار المقدر نقوم بإعداد الجدول التالي:

n	X	Y	$x = x - \bar{X}$	$y = y - \bar{Y}$	x*y	x^2	$\hat{Y}$	ei=(yi-$\hat{Y}$)	ei2	X^2	(yi-$\bar{Y}$)^2
1	4	7	-5	-4	20	25	6	1	1	16	16
2	13	17	4	6	24	16	15	2	4	169	36
3	11	15	2	4	8	4	13	2	4	121	16
4	12	14	3	3	9	9	14	0	0	144	9
5	12	13	3	2	6	9	14	-1	1	144	4
6	8	10	-1	-1	1	1	10	0	0	64	1
7	6	8	-3	-3	9	9	8	0	0	36	9
8	11	12	2	1	2	4	13	-1	1	121	1
9	10	11	1	0	0	1	12	-1	1	100	0
10	9	9	0	-2	0	0	11	-2	4	81	4
11	3	5	-6	-6	36	36	5	0	0	9	36
$\sum$	99	121	0	0	115	114		0	16	1005	132

- حساب المتوسط الحسابي:

$$\bar{Y} = \frac{\sum y_i}{n} = \frac{121}{11} = 11$$

$$\bar{X} = \frac{\sum x_i}{n} = \frac{99}{11} = 9$$

– حساب معلمات خط الانحدار:

$$\hat{b} = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{115}{114} \cong 1$$

$$\hat{a} = \bar{Y} - \hat{b}\bar{X} = 11 - 1(9) = 2$$

– خط الانحدار المقدر يكتب بالصورة التالية:

$$\hat{Y} = 2 + X$$

– التمثيل البياني لخط ا لانحدار المقدر موضح في شكل الانتشار السابق.

– حساب:

$$\hat{\sigma}_\mu^2 = \frac{\sum e_i^2}{n-k} = \frac{16}{11-2} = 1.777$$

$$\hat{\sigma}_a = \sqrt{\frac{\sum X_i^2}{n \sum x_i^2} \sigma_\mu^2} = \sqrt{\frac{1005}{11(114)} 1.777} = 0.124$$

$$\hat{\sigma}_b = \sqrt{\frac{1}{\sum x_i^2} \sigma_\mu^2} = \sqrt{\frac{1}{114} 1.777} = 1.193$$

3– إيجاد مجال الثقة للمعلمة a وهذا بمستوى دلالة قدره 5%:

$$a \in \left]\hat{a} - t_{\underset{n-k}{\alpha}} \hat{\sigma}_{\hat{a}}, \hat{a} + t_{\underset{n-k}{\alpha}} \hat{\sigma}_{\hat{a}}\right[=]2 - 2.2.262 \times 1.193, 2 + 2.776 \times 0.353[=]-0.698, 4.698[$$

4– إيجاد مجال الثقة للمعلمة b وهذا بدرجة ثقة قدرها 90%:

$$b \in \left]\hat{b} - t_{\underset{n-k}{\alpha}} \hat{\sigma}_{\hat{b}}, \hat{b} + t_{\underset{n-k}{\alpha}} \hat{\sigma}_{\hat{b}}\right[=]1 - 1.833 \times 0.124 + 1.833 \times 0.124[=]0.772, 1.277[$$

5 – اختبار فرض أن قيمة b = 0.5 وذلك بدرجة ثقة قدرها 80%:

$$H_0 : b = 0.5$$
$$H_1 : b \neq 0.5$$

$$t = \frac{\hat{b} - b_0}{\sigma_{\hat{b}}} = \frac{1 - 0.5}{0.124} = 4.03$$

$t_{\text{الجدولية}} = 1.383$

بما أن قيمة t المحسوبة تقع ضمن منطقة الرفض فإننا نرفض H_0 ونقبل H_1، ومنه لا يمكن لقيمة b أن تأخذ القيمة 0.5.

6 – اختبار فرض أن قيمة a = 1، وذلك بدرجة ثقة قدرها 95%:

$$H_0 : a = 1$$
$$H_1 : a \neq 1$$

$$t = \frac{\hat{a} - a_0}{\sigma_{\hat{a}}} = \frac{2 - 1}{1.193} = 0.84$$

$t_{\text{الجدولية}} = 2.262$

بما أن قيمة t المحسوبة تقع ضمن منطقة القبول فإننا نقبل H_0 ونرفض H_1، ومنه يمكن لقيمة a أن تأخذ القيمة 1.

7 – اختبار فرض أن قيمة b > 1.5:

$$H_0 : b = 1.5$$
$$H_1 : b \succ 1.5$$

$$t = \frac{\hat{b} - b_0}{\sigma_{\hat{b}}} = \frac{1 - 1.5}{0.124} = -4.03$$

$t_{\text{الجدولية}} = 1.383$

بما أن قيمة t المحسوبة تقع ضمن منطقة القبول فإننا نقبل H_0 ونرفض H_1، ومنه لا يمكن لقيمة b أن تكون أكبر من 1.5.

8 – اختبار معنوية المعلمة المقدرة b عند مستوى المعنوية 5%:

$$H_0 : b = 0$$
$$H_1 : b \neq 0$$

$$t = \frac{\hat{b}}{\sigma_{\hat{b}}} = \frac{1}{0.124} = 8.06$$

$t_{\text{الجدولية}}=2.262$

بما أن قيمة t المحسوبة تقع ضمن منطقة الرفض فإننا نرفض H_0 ونقبل H_1، ومنه فخط الانحدار المقدر هو ذو معنوية إحصائية عند مستوى معنوية قدره 5%.

9 – اختبار فرض أن هناك تناسب بين المتغيرين x و y عند مستوى المعنوية 5%:

$$H_0 : a = 0$$
$$H_1 : a \neq 0$$

$$t = \frac{\hat{a}}{\sigma_{\hat{a}}} = \frac{2}{1.193} = 1.67$$

$t_{\text{الجدولية}}=2.262$

بما أن قيمة t المحسوبة تقع ضمن منطقة القبول فإننا نقبل H_0 ونرفض H_1، ومنه يوجد تناسب بين المتغيرين.

10 – حساب معامل التحديد ومعامل الارتباط مع التعليق على النتيجة؟

– حساب معامل التحديد:

$$R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{16}{132} = 0.8787$$

التعليق: المتغير المستقل X يفسر ما مقداره 87.87% من التغير في المتغير التابع Y، والباقي 12.13% يعود للخطأ العشوائي.

– حساب معامل الارتباط:

$$r=\pm\sqrt{R^2}=\sqrt{0.8787}=0.937$$

التعليق: يوجد ارتباط طردي قوي بين المتغيرين.

11 – أوجد مجال ثقة لتباين حد الخطأ عند مستوى معنوية قدره 5%؟

$$\delta_{\mu}^2=\left]\frac{(n-2)\hat{\sigma}_{\mu}^2}{\chi^2_{\underset{n-k}{\alpha}}},\frac{(n-2)\hat{\sigma}_{\mu}^2}{\chi^2_{\underset{n-k}{1-\alpha}}}\right[=\left]\frac{9\times1.777}{16.918},\frac{9\times1.777}{3.325}\right[=\left]0.945,4.809\right[$$

12 – إيجاد توقع لمقدار المبيعات اليومية إذا علمت أن مدة خدمة العامل هي 15 سنة:

$$\hat{Y}_0=\hat{a}+\hat{b}X_0=2+1(15)=17$$

أي أن المبيعات اليومية ستبلغ 17 ألف دينار يوميا إذا كانت مدة الخدمة هي 15 سنة.

التوقع بمجال:

$$\delta y_0=\sqrt{\hat{\sigma}_{\mu}^2\left[1+\frac{1}{n}+\frac{(X_0-\bar{X})^2}{\sum x_i^2}\right]}=\sqrt{1.777\left[1+\frac{1}{11}+\frac{(15-9)^2}{114}\right]}=1.581$$

تحديد مجال الثقة للتوقع:

$$p(\hat{Y}_0-t_{\underset{n-2}{\alpha}}\,\delta y_o\langle Y_0\langle\hat{Y}_0+t_{\underset{n-2}{\alpha}}\,\delta y_0)=1-\alpha$$

$$\Rightarrow p(17-1.833\times1.581\langle Y_0\langle 17+1.833\times1.581)=0.90$$

$$\Rightarrow p(14.1\langle Y_0\langle 19.9)=0.90$$

13 – اختبار فرض أن المبيعات اليومية تساوي 14 إذا كانت مدة الخدمة هي 13 سنة:

$$H_0 : Y_0 = 14$$
$$H_1 : Y_0 \neq 14$$

$$\delta \hat{y}_0 = \sqrt{\hat{\sigma}^2_{\mu}\left[1+\frac{1}{n}+\frac{\left(X_0-\overline{X}\right)^2}{\sum x_i^2}\right]} = \sqrt{1.777\left[1+\frac{1}{11}+\frac{(13-9)^2}{114}\right]} = 1.479$$

$$t = \frac{\hat{Y}_0 - Y_0}{\sigma_{\hat{Y}_0}} = \frac{15-14}{1.479} = 0.67$$

بما أن قيمة t المحسوبة تقع ضمن منطقة القبول فإننا نقبل H_0 ونرفض H_1، ومنه يمكن لقيمة المبيعات أن تساوي 14 إذا كانت مدة الخدمة تساوي 13 سنة.

14- اختبار معنوية المعلمات المقدرة:

نضع الفرض:

$$H_0: a = b = 0$$
$$H_1: a \neq b \neq 0$$

ثم نقوم بإعداد جدول تحليل التباين ANOVA

إحصائية F	متوسط مجموع المربعات	درجة الحرية	مجموع المربعات	مصدر التباين
$F=\frac{116}{1.777} = 65.27$	$MSR = 116$	1	116	الانحدار
	$MSE = 1.777$	9	16	البواقي
	–	10	132	المجموع

عند درجة حرية بسط قدرها 1 ودرجة حرية مقام قدرها 4 ومستوى معنوية قدرها 5%، فإن قيمة فيشر الجدولية تساوي10.56.

بما أن قيمة $F_{\text{المحسوبة}}$ أكبر من $F_{\text{الجدولية}}$ فهذا يعني أننا نرفض H_0 ونقبل H_1 ومنه فالنموذج معنوي إجمالا.

والشكل الرياضي المختار للمعادلة سليم إحصائيا.

تمرين رقم (4-3):

أثبت أن:

$$\hat{b} = \frac{\sum x_i y_i}{\sum x_i^2}$$

لدينا:

$$y_i = Y_i - \bar{Y}, \qquad x_i = X_i - \bar{X}$$

بتعويض العلاقة السابقة في البسط نجد:

$$\sum x_i y_i = \sum (X_i - \bar{X})(Y_i - \bar{Y})$$

$$= \sum (X_i Y_i - Y_i \bar{X} - X_i \bar{Y} + \bar{X}\bar{Y})$$

$$= \sum X_i Y_i - \bar{X} \sum Y_i - \bar{Y} \sum X_i + n\bar{X}\bar{Y}$$

$$= \sum X_i Y_i - \frac{n\bar{X} \sum Y_i}{n} - \frac{n\bar{Y} \sum X_i}{n} + n\bar{X}\bar{Y}$$

$$= \sum X_i Y_i - n\bar{X}\bar{Y} - n\bar{X}\bar{Y} + n\bar{X}\bar{Y}$$

$$= \sum X_i Y_i - n\bar{X}\bar{Y}$$

$$= \sum X_i Y_i - n \frac{\sum X_i}{n} \frac{\sum Y_i}{n}$$

$$= \sum X_i Y_i - \frac{\sum X_i \sum Y_i}{n}$$

كذلك نقوم بتعويض في المقام:

$$\sum x_i^2 = \sum (X_i - \bar{X})^2$$

$$= \sum \left(X_i^2 - 2X_i \bar{X} + \bar{X}^2\right)$$

$$= \sum X_i^2 - 2\bar{X} \sum X_i + n\bar{X}^2$$

$$= \sum X_i^2 - \frac{2n\bar{X}\sum X_i}{n} + n\bar{X}^2$$

$$= \sum X_i^2 - 2n\bar{X}^2 + n\bar{X}^2$$

$$= \sum X_i^2 - n\bar{X}^2$$

$$= \sum X_i^2 - n\left(\frac{\sum x_i}{n}\right)^2$$

$$= \sum X_i^2 - \frac{\sum x_i}{n}$$

لدينا:

$$\hat{b} = \frac{n\sum X_iY_i - \sum X_i \sum Y_i}{n\sum X_i^2 - (\sum X_i)^2}$$

$$= \frac{\sum X_iY_i - \frac{\sum X_i \sum Y_i}{n}}{n\sum X_i^2 - \frac{(\sum X_i)^2}{n}} = \frac{\sum x_iy_i}{\sum x_i^2}$$

مسائل محلولة

مسألة رقم 01:

الأسئلة:

1 – عرف المعلمة والإحصائية؟ ما العلاقة بينهما؟

2 – ما ذا يقصد بتوزيع المعاينة للمتوسط؟

3 – ماذا يعني التقدير بنقطة؟ التقدير بفترة؟ مقدر غير متحيز؟

4 – ماذا يقصد بالخطأ من النوع الأول، الخطأ من النوع الثاني؟

تمرين 01:

تنتج ملبنة نوع من الجبن مسوق تحت اسم (الريادة). الكتلة X معطاة بالغرام، لقطعة من الجبن مسحوبة عشوائيا في الإنتاج، تتبع التوزيع الطبيعي.

نسحب عينة عشوائية بسيطة من 17 وحدة من الجبن، حيث يمثل الجدول التالي أوزانها:

255	251	252	261	252	255	250	255	253	251	257	250	256	253	254	254	250

المطلوب:

1 – أ – أحسب الوسط الحسابي و التباين لهذه العينة؟ ب – أوجد تقدير نقطي لتباين الإنتاج (المجتمع)؟

2 – أوجد تقدير مجالي بدرجة ثقة 95% للكتلة المتوسطة μ للإنتاج (المجتمع)؟

3 – يرغب مسؤول الإنتاج للجبن (الريادة) في معرفة حجم العينة الأدنى لعينة عشوائية بسيطة لحساب مجال ثقة لـ μ ،بدرجة ثقة 95%، وتكون سعته أقل من 1؟

أ – نقول لهذا المسؤول بأننا لا نستطيع الإجابة على هذا السؤال بدون معرفة معلومة حول تباين الإنتاج (المجتمع). لماذا تعتبر معرفة هذا التباين أساسية للإجابة على السؤال؟

ب – يوضح مسؤول الإنتاج بأن $\sigma^2 = 6.25$. عين مجال الثقة المطلوب؟

4 – يدعي مسؤول الإنتاج بأن 15% من الجبن المنتج له كتلة أكبر من 257غ، نسحب عينة عشوائية بسيطة من 200 قطعة. ونجد أن منها 40 وحدة تتجاوز 257غ.

عن طريق نتائج هذه العينة، وعند مستوى معنوية قدره 5%، هل نقبل إدعاء هذا المسؤول؟

تمرين 02:

ليكن $(X_1, X_2, ..., X_n)$ عينة عشوائية بسيطة مسحوبة من مجتمع وسطه μ و تباينه σ^2.

1 – أثبت أن $\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$ هو مقدر غير متحيز للمعلمة μ؟

2 – أثبت أن $W = \frac{1}{n}\sum_{i=1}^{n}(X_i + \overline{X})$ هو مقدر متحيز للمعلمة μ؟ حدد مقدار هذا التحيز؟ كيف نجعل هذا المقدر مقدر غير متحيز؟

الحل:

1 – تعرف المعلمة والإحصائية. والعلاقة بينهما:

ج: المعلمة هي خاصية وصفية للمجتمع، الإحصائية هي خاصية وصفية للعينة

$$\text{المعلمة} \underset{\text{معاينة}}{\overset{\text{تقدير}}{\rightleftarrows}} \text{الإحصائية}$$

2 – القصد بتوزيع المعاينة للمتوسط:

إذا أخذنا عينات عشوائية متكررة، كل منها من حجم n، من مجتمع، وأوجدنا متوسط كل عينة، فإننا نجد أن هذه المتوسطات تختلف عن بعضها البعض، و يسمى التوزيع الاحتمالي لها بتوزيع المعاينة للمتوسط.

<u>3 – مفهوم التقدير بنقطة. التقدير بفترة، مقدر غير متحيز:</u>

– التقدير بنقطة هو التعبير عن معلمة المجتمع المجهولة بعدد واحد.

– التقدير بفترة هو التعبير عن معلمة المجتمع المجهولة بمجال ، مع الاحتمال المناظر.

– يكون المقدر غير متحيز إذا أعطى توزيع المعاينة النظري، الناتج عن المعاينة العشوائية المتكررة من مجتمع، إحصائية مساوية لمعلمة المجتمع.

<u>4 – المقصود بالخطأ من النوع الأول، الخطأ من النوع الثاني:</u>

– يشير الخطأ من النوع الأول إلى رفض فرض صحيح،

– يشير الخطأ من النوع الثاني إلى قبول فرض خاطئ.

حل التمرين الأول:

1 – حساب الوسط الحسابي والتباين:

$$\bar{X} = \frac{\sum X_i}{n} = 253.47 \approx 253.5$$

$$S^2 = \frac{\sum (X_i - \bar{X})^2}{n-1} = 8.515$$

ب – إيجاد تقدير نقطي لتباين الإنتاج

$$\hat{\sigma}^2 = \frac{n}{n-1} S^2 = \frac{17}{16} \times 8.515 = 9.047$$

2 – إيجاد تقدير مجالي للكتلة المتوسطة:

بما أن: المجتمع طبيعي+تباين المجتمع مجهول+n أقل من 30 فإننا نستخدم توزيع ستيودنت.

$$p\left(\bar{X} - t_\alpha \frac{S}{\sqrt{n}} \langle \mu \langle \bar{X} + t_\alpha \frac{S}{\sqrt{n}}\right) = 1 - \alpha$$

$$p\left(253.5 - 2.12 \frac{2.981}{\sqrt{17}} \langle \mu \langle 253.5 + 2.12 \frac{2.981}{\sqrt{17}}\right) = 0.95$$

$$\mu = 253.5 \pm 1.532$$

3 – أ : إيجاد حجم n بحيث يكون المجال سعته =1.

تحت الشروط السابقة في السؤال 2 فإن مجال الثقة يحسب باستخدام توزيع ستيودنت، وبالتالي:

$$p\left(\bar{X} - t_\alpha \frac{S}{\sqrt{n-1}} \langle \mu \langle \bar{X} + t_\alpha \frac{S}{\sqrt{n-1}}\right) = 1 - \alpha$$

سعة المجال هي:

$$\bar{X} + t \frac{S}{\sqrt{n-1}} - \left(\bar{X} - t \frac{S}{\sqrt{n-1}}\right) = 1$$

$$2 \times t \frac{S}{\sqrt{n-1}} = 1$$

نلاحظ من هذه المعادلة أنه في غياب التباين فإننا لا نستطيع حل هذه المعادلة، و بالتالي يجب أن تتوفر لدينا معلومات تكميلية حوله.

ب – لدينا تباين المجتمع = 6.25 و بالتالي تباين المجتمع أصبح معروف، ومنه نستخدم التوزيع الطبيعي للتقدير:

حساب n:

$$2Z_{0.975}\frac{\sigma}{\sqrt{n}}=1\Rightarrow n\approx 97$$

التقدير:

$$p\left(\overline{X}-Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\langle\mu\langle\overline{X}+Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)=0.95$$
$$\mu=253.5\pm 0.50$$

4 – اختبار الفرضية:

$$H_0:P=0.15$$
$$H_1:P\rangle 0.15$$
$$p=\frac{40}{200}=0.2$$
$$\sigma_{\bar{p}}=\sqrt{\frac{pq}{n}}=\sqrt{\frac{0.15\times 0.85}{200}}=0.025$$
$$Z=\frac{0.2-0.15}{0.025}=2$$

بما أن Zالمحسوبة تقع في منطقة الرفض لأن Zالجدولية تساوي1.65 ، فإننا نرفض فرض العدم و نقبل الفرض البديل، و بالتالي متوسط وزن الجبن يتجاوز 257غ في 15% من الإنتاج(نقبل الادعاء).

حل التمرين 05:

1 – إثبات أن المقدر هو غير متحيز:

$$E\left(\overline{X}\right)=E\left(\sum\frac{X_i}{n}\right)=\frac{1}{n}E\left(X_1+.......+X_n\right)$$
$$=\frac{1}{n}\left(EX_1+........EX_n\right)=\frac{1}{n}\sum E\left(X_i\right)=\frac{n\mu}{\mu}=\mu$$

إذن المقدر هو مقدر غير متحيز لمعلمة المجتمع.

2 – إثبات أن W هو مقدر متحيز:

$$\begin{aligned}E(W) &= E\left(\frac{1}{n}\sum\left(X_i+\overline{X}\right)\right)\\&=\frac{1}{n}E\left(\sum\left(X_i+\overline{X}\right)\right)\\&=\frac{1}{n}E\left(\sum X_i+n\overline{X}\right)\\&=\frac{1}{n}\left(\sum EX_i+nE\overline{X}\right)\\&=\frac{1}{n}(n\mu+n\mu)\\&=\mu+\mu\end{aligned}$$

إذن المقدر W هو مقدر متحيز لمعلمة المجتمع، و مقدار هذا التحيز هو μ.

لجعل هذا المقدر غير متحيز يجب أن يكتب بالشكل التالي:

$$W'=\frac{1}{2n}\sum\left(X_i+\overline{X}\right)$$

مسألة رقم 02:

الأسئلة:

1 – ما هو شكل توزيع المعاينة للمتوسط إذا كان أ: المجتمع الأصلي طبيعيا، ب: غير طبيعي؟

2 – ماذا يعني معامل التحديد؟ معامل الارتباط؟

3 – ما هو الهدف من بناء النماذج؟ ما هي مكوناتها؟ قدم مثالا توضح فيه إجابتك؟

تمرين 01:

أجريت دراسة لتحديد ما إذا كان متوسط الإنفاق الشهري للأسر، يختلف عن متوسط الإنفاق الشهري للمجتمع ككل الذي يساوي 32000 دج. أخذت عينة عشوائية من 36 أسرة من هذا المجتمع، ووجد متوسطها 34000 دج وانحرافها المعياري 8000 دج.

المطلوب: ما النتائج التي يمكنك أن تتحصل عليها من هذه الدراسة عند مستوى المعنوية 5%؟.

تمرين 02:

ضمن خطتها لإصلاح حركة المرور في المدينة، قامت بلدية قالمة بإجراء مسح ميداني لتحديد حجم الحركة عبر تقاطع رئيسي خلال فترة الصباح. باختيار يوم الأحد لثمانية أسابيع متتالية، تم عد

المركبات التي تمر عبر التقاطع بين الساعة 7:00 و 9:00 صباحا، ووجد أن متوسط عدد المركبات للعينة يساوي 1500 سيارة، و الانحراف المعياري للعينة يساوي 300 سيارة.

المطلوب: حساب فترة الثقة 99% لمتوسط عدد المركبات في المجتمع باعتبار أن توزيع المجتمع طبيعي؟

تمرين 03:

لمقارنة ثلاث أنواع من الأدوية لمعالجة الصداع، أخذت مجموعة من 12 شخصا يعانون من الصداع، وقسموا عشوائيا إلى ثلاث مجموعات وأعطيت كل مجموعة نوعا من الأدوية وتم رصد زمن الشفاء بالدقائق. زمن الشفاء موزع طبيعيا بتباين متساوي. وكانت النتائج كما يلي:

T_3	T_2	T_1
58	56	80
52	22	53
41	44	55
53	46	56

المطلوب: هل هناك فروق بين الأدوية الثلاثة عند مستوي دلالة $\alpha = 0.05$ ؟

الحل:

1 – شكل توزيع المعاينة للمتوسط إذا كان أ: المجتمع الأصلي طبيعيا، ب: غير طبيعي:

أ – يكون توزيع المعاينة للمتوسط أيضا طبيعيا،

ب– طبقا لنظرية النهاية المركزية فإن توزيع المعاينة لمتوسط العينة يقترب من التوزيع الطبيعي مع تزايد حجم العينة، و يكون التقريب جيدا كلما كان n أكبر من 30.

2 – معنى معامل التحديد، معامل الارتباط:

معامل التحديد: هو نسبة التغير الإجمالي في Y الذي يفسره خط الانحدار

معامل الارتباط: هو درجة الاقتران بين متغيرين أو أكثر، وهذا لا يعني وجود علاقة سببية بينهما.

3 – الهدف من بناء النماذج، ومكوناتها، مع مثال:

ج: عموما الهدف من بناء النماذج هو التنبؤ بسلوك الظاهرة محل الدراسة، وهي تتكون من الثوابت، المعالم و المتغيرات.

مثال: توزيع بواسون $p(x=k)=\lambda^{k}\frac{e^{-\lambda}}{k!}$ ، e ثابت، λمعلم، k متغير.

حل تمرين 01:

نستخدم توزيع ستيودنت لأن n أقل من 30، تباين المجتمع مجهول و المجتمع يتبع التوزيع الطبيعي.

$$H_0 : u=32000$$

$$H_1 :\neq u\ 32000$$

$$t_{0.05} = 2.060 \qquad t=\frac{\bar{x}-\mu}{s/\sqrt{n}}=\frac{32000-34000}{8000/\sqrt{26}}=-1.27$$

بما أن t المحسوبة تقع في منطقة القبول فإننا نقبل فرض العدم و بالتالي فإن متوسط الإنفاق الشهري للأسر لا يختلف عن متوسط الإنفاق الشهري للمجتمع ككل

تمرين 02:

لأن n أكبر من 30 فإننا نقدر باستخدام التوزيع الطبيعي

$$\mu=\bar{x}\pm Z_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}}=1500\pm 2.58\frac{300}{\sqrt{38}}=1500\pm 125.55$$

تمرين 03:

إعداد جدول تحليل التباين:

$$H_0:\mu_1=\mu_2=\mu_3$$
$$H_1:\mu_1\neq\mu_2\neq\mu_3$$

$$\bar{x}_1=51, \bar{x}_2 45, \bar{x}_3=60, \bar{\bar{x}}=52$$

$$SSR=r\sum\left(\bar{x}_j-\bar{\bar{x}}\right)^2=4(1+49+64)=456$$

$$SSE=\sum\sum\left(x_{ij}-\bar{x}_j\right)^2=49+1+100+4+121+256+1+16+400+81+49+16=1094$$

مصدر التغير	متوسط المربعات	درجات الحرية	مجموع المربعات	إحصاء فيشر
تفسره الأدوية	SSR=456	c−1=2	MSR=228	F=1.87
الخطأ	SSE=1094	c(r−1)=9	MSE=121.55	
المجموع	SST=1550	rc−1=11		

$F_{0.05}$ عند درجة حرية للبسط 2 و المقام 9 هي 4.26،

بما أن قيمة فيشر المحسوبة أقل من قيمة فيشر الجدولية فإننا نقبل فرض العدم و بالتالي نقبل فرض بأن أنواع الأدوية الثلاث متشابهة.

مسألة رقم 03

تمرين رقم 01 :

يضمن أحد الصانعين أن المنتجات المصنوعة في آلته يكون متوسط طولها 20 سم وتباينها 4 سم. من أجل مراقبة مدى ضبط الآلة، نقوم بسحب عينات عشوائية لأجل مقارنتها مع المتوسط النظري.

1 – تتكون العينة المسحوبة من 100 قطعة. أوجد مجال الثقة لمتوسط المجتمع بمستوى معنوية قدره 5%؟

2 – تتكون العينة المسحوبة من 25 قطعة. أوجد مجال الثقة الجديد بمستوى معنوية قدره 5% مع افتراض أن طول هذه المنتجات يتبع التوزيع الطبيعي؟

3 – تتكون العينة المسحوبة من 10 قطع، نقوم بقياس طول كل قطعة، النتائج كالتالي:

22، 22، 18، 24، 18، 15.5، 18، 16، 24.5، 18

أوجد مجال الثقة الجديد بمستوى معنوية قدره 5% مع افتراض أن طول هذه المنتجات يتبع التوزيع الطبيعي بتباين مجهول؟

4 – قارن وعلق على:

أ – الافتراضات التي بنيت عليها التجارب الثلاث؟

ب – تحديد ما إن كان هناك تأثير لحجم العينة على النتائج؟

تمرين رقم 02:

نقوم بعملية انحدار لمحاولة بناء نموذج يربط بين حجم المحصول من الحبوب (مليون طن) وكميات الأمطار المتساقطة (100ملم) خلال 5 سنوات. يظهر الجدول الموالي البيانات.

كمية الأمطار	4	6	6	3	1
حجم المحصول	36	47	43	34	20

1 – قدر نموذج الانحدار المعطى؟ مع التمثيل البياني؟

2 – أوجد تقدير بمجال لمعلمة ميل خط الانحدار وذلك عند درجة ثقة قدرها 95%؟

3 – هل يمكن القول أن خط الانحدار المقدر ذو دلالة إحصائية عند مستوى المعنوية 5%؟

4 – أوجد معامل التحديد؟ علق على النتيجة؟

5 – قدر بمجال قيمة المحصول إن كانت الكمية المتساقطة تبلغ 1000 ملم وهذا عند درجة ثقة قدرها 90%؟

الحل:

نفرض أن X متغير عشوائي يمثل طول المنتج.

حسب تصريح الصناع، فإن لـ X الخصائص التالية:

$$E(X) = \mu = 20\ cm, \quad \sigma_x^2 = 4\ cm$$

1- في الحالة الأولى:

يعطى مجال الثقة لطول المجتمع بمستوى معنوية 5% لعينة تتكون من 100 منتج بالعلاقة التالية:

$$p\left(\bar{X} - Z_{1-\frac{\alpha}{2}}(\frac{\sigma}{\sqrt{n}}) < \mu < \bar{X} + Z_{1-\frac{\alpha}{2}}(\frac{\sigma}{\sqrt{n}})\right) = 1 - \alpha$$

ولذلك لأن حجم العينة أكبر من 30.

بالتالي:

$$p\left(20 - 1.96\left(\frac{2}{\sqrt{100}}\right) < \mu < 20 + 1.96(\frac{2}{\sqrt{100}})\right) = 0.95$$

$$p(19.608 < \mu < 20.392) = 0.95$$

2- في الحالة الثانية:

حجم العينة n=25، تباين المجتمع معلوم والمجتمع يتبع التوزيع الطبيعي. بالتالي فإن $\bar{X}$ سيكون موزع طبيعيا ويكون مجال الثقة المطلوب هو:

$$p\left(\bar{X} - Z_{1-\frac{\alpha}{2}}(\frac{\sigma}{\sqrt{n}}) < \mu < \bar{X} + Z_{1-\frac{\alpha}{2}}(\frac{\sigma}{\sqrt{n}})\right) = 1 - \alpha$$

$$p\left(20 - 1.96\left(\frac{2}{\sqrt{25}}\right) < \mu < 20 + 1.96(\frac{2}{\sqrt{25}})\right) = 0.95$$

$$p(19.216 < \mu < 20.784) = 0.95$$

3- في الحالة الثالثة:

حجم العينة n=10، تباين المجتمع مجهول والمجتمع يتبع التوزيع الطبيعي، في هذه الحالة يجب حساب قيمة المتوسط الحسابي والانحراف المعياري للعينة:

$$\bar{X} = \frac{\sum X_i}{n} = 19.6\ cm$$

$$S=\sqrt{\frac{\sum(x_i-\bar{X})^2}{n-1}}=3.29\ cm$$

على ضوء هذه المعطيات، فإننا نستخدم توزيع ستيودنت لتشكيل مجال ثقة وفق الصيغة التالية:

$$p\left(\bar{X}-t_{\frac{\alpha}{n-1}}\left(\frac{s}{\sqrt{n}}\right)<\mu<\bar{X}+t_{\frac{\alpha}{n-1}}\left(\frac{s}{\sqrt{n}}\right)\right)=1-\alpha$$

$$p\left(19.6-2.262(\frac{3.29}{\sqrt{10}})<\mu<19.6+2.262(\frac{3.29}{\sqrt{10}})\right)=0.95$$

$$p(17.250<\mu<21.950)=0.95$$

4- المقارنة والتعليق:

أ- في الحالة رقم 01، لم يتم وضع فرضية التوزيع الطبيعي للطول، وذلك لأن حجم العينة أكبر من 30. بالتالي في هذه الحالة فإن نظرية النهاية المركزية يمكن تطبيقها. في الحالة الثانية، حجم العينة أقل من 30. ولهذا تم وضع فرضية التوزيع الطبيعي لطول المنتج. أما في الحالة الثالثة، وحيث أن حجم العينة صغير ويساوي 10. لهذا أيضا تم افتراض التوزيع الطبيعي لطول المنتج.

ب- نلاحظ أن مجالات الثقة لطول المنتجات في المجتمع المتحصل عليها عند مستوى معنوية 5% تتبع حجم العينة، حيث كانت النتائج كما يلي:

$$n=100\quad p(19.608<\mu<20.392)=0.95$$

$$n=25\quad p(19.216<\mu<20.784)=0.95$$

$$n=10\quad p(17.250<\mu<21.950)=0.95$$

نلاحظ أن الفرق بين مجال الثقة لما حجم العينة 10 و25 يعتبر كبير مقارنة بالفرق بين مجال الثقة لما حجم العينة يساوي 25 و100. بالتالي يمكن أن نقرر بأن حجم العينة المساوي لـ 10 يعتبر صغير جدا مقارنة بحجم العينة المساوي لـ 25 الذي يعتبر نوعا ما كبير. وحجم العينة المساوي لـ 100 الذي يعتبر كبير جدا. يمكننا أن نقرر على ضوء ذلك أن أفضل حجم عينة يمكن استخدامه لتشكيل مجالات الثقة لمراقبة مدى ضبط هذه الآلة يتراوح بين 25 و100.

حل التمرين رقم 02:

1- لإيجاد خط الانحدار المقدر نقوم بإعداد الجدول التالي:

n	X	Y	$x = x - \bar{X}$	$y = y - \bar{Y}$	x*y	x^2	$\hat{Y}$	ei=(yi-$\hat{Y}$)	ei2	(yi-$\bar{Y}$)^2
1	4	36	0	0	0	0	36	0	0	0
2	6	47	2	11	22	4	45.556	1.444	2.0851	121
3	6	43	2	7	14	4	45.556	-2.556	6.5331	49
4	3	34	-1	-2	2	1	31.222	2.778	7.7173	4
5	1	20	-3	-16	48	9	21.666	-1.666	2.7755	256
$\sum$	99	121	0	0	86	18		0	19.111	430

- حساب المتوسط الحسابي:

$$\bar{Y} = \frac{\sum y_i}{n} = \frac{180}{5} = 36$$

$$\bar{X} = \frac{\sum x_i}{n} = \frac{20}{5} = 4$$

- حساب معلمات خط الانحدار:

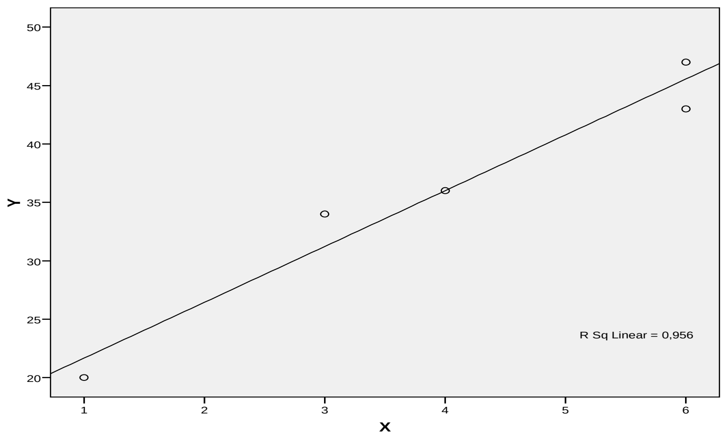
$$\hat{b} = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{86}{18} \cong 4.778$$

$$\hat{a} = \bar{Y} - \hat{b}\bar{X} = 36 - 4.778(4) = 16.888$$

- خط الانحدار المقدر يكتب بالصورة التالية:

$$\hat{Y} = 16.888 + 4.778X$$

- التمثيل البياني لخط الانحدار المقدر:



2- تقدير بمجال لمعلمة ميل خط الانحدار وذلك عند درجة ثقة قدرها 95%:

- حساب:

$$\hat{\sigma}_\mu^2 = \frac{\sum e_i^2}{n-k} = \frac{19.111}{5-2} = 6.37$$

$$\hat{\sigma}_b = \sqrt{\frac{1}{\sum x_i^2}\sigma_\mu^2} = \sqrt{\frac{1}{18}6.37} = 0.595$$

$$b \in \left]\hat{b} - t_{\frac{\alpha}{n-k}}\hat{\sigma}_{\hat{b}}, \hat{b} + t_{\frac{\alpha}{n-k}}\hat{\sigma}_{\hat{b}}\right[$$

$$= \left]4.778 - 3.812 \times 0.595, 4.778 + 3.812 \times 0.595\right[= \left]2.885, 6.671\right[$$

3 – اختبار معنوية المعلمة b عند 5%:

$$H_0 : b = 0$$

$$H_1 : b \neq 0$$

$$t = \frac{\hat{b}}{\sigma_{\hat{b}}} = \frac{4.778}{0.595} = 8.030$$

$t_{\text{الجدولية}} = 3.182$

بما أن قيمة t المحسوبة تقع ضمن منطقة الرفض فإننا نرفض H_0 ونقبل H_1، ومنه فخط الانحدار المقدر هو ذو معنوية إحصائية عند مستوى معنوية قدره 5%.

4 – أوجد معامل التحديد؟ علق على النتيجة؟

$$R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{19.111}{430} = 0.9555$$

التعليق: المتغير المستقل X (كمية الأمطار المتساقطة) يفسر ما مقداره 95.55% من التغير في المتغير التابع Y (حجم المحصول من الحبوب) ، والباقي يعود للخطأ العشوائي.

5 – قدر بمجال قيمة المحصول إن كانت الكمية المتساقطة تبلغ 1000 ملم وهذا عند درجة ثقة قدرها 90%؟

$$\hat{Y}_0 = \hat{a} + \hat{b}X_0 = 16.888 + 4.778(10) = 64.668$$

التوقع بمجال:

$$\delta\hat{y}_0 = \sqrt{\hat{\sigma}_\mu^2\left[1+\frac{1}{n}+\frac{\left(X_0-\overline{X}\right)^2}{\sum x_i^2}\right]} = \sqrt{6.37\left[1+\frac{1}{5}+\frac{(10-4)^2}{18}\right]} = 4.515$$

تحديد مجال الثقة للتوقع:

$$p(\hat{Y}_0 - t_{\underset{n-2}{\alpha}}\delta\hat{y}_o \langle Y_0 \langle \hat{Y}_0 + t_{\underset{n-2}{\alpha}}\delta\hat{y}_0) = 1-\alpha$$

$$\Rightarrow p(64.668 - 2.353\times 4.515 \langle Y_0 \langle 4.668 + 2.353\times 4.515) = 0.90$$

$$\Rightarrow p(54.044 \langle Y_0 \langle 75.292) = 0.90$$

Examen de Statistique Appliquée aux Problèmes Décisionnels

Durée : deux heures. Les calculatrices sont autorisées ; tout autre matériel électronique est interdit.

Exercice 1. (5 pts)

Chacun des dix items suivants comporte trois réponses possibles dont une seule est exacte ; entourez sans justification la réponse qui vous semble correcte et rayez les autres. Une réponse juste apporte 0,5 point ; une réponse fausse enlève 0,25 points ; une absence de réponse n'apporte ni n'enlève de points.

Le nombre d'appels reçus par un standard en une minute est distribué selon la loi de Poisson $\mathscr{P}(2)$.

La variable aléatoire X comptant le nombre d'appels reçus en une minute est distribuée selon $\mathscr{P}(2)$; ainsi pour tout $k \in \mathbb{N}$: $\mathbb{P}(X = k) = e^{-2} \times 2^k/k!$.

1. La probabilité de recevoir exactement un appel en une minute est :

   (a) $\boxed{0,27}$ (b) $\cancel{0,31}$ (c) $\cancel{0,35}$.

   $\mathbb{P}(X = 1) = e^{-2} \times 2^1/1! \approx 0,27$

2. La probabilité de recevoir au plus un appel en une minute est :

   (a) $\cancel{0,31}$ (b) $\cancel{0,37}$ (c) $\boxed{0,41}$.

   $\mathbb{P}(X \leq 1) = \sum_{k=0}^{1} \mathbb{P}(X = k) = \sum_{k=0}^{1} e^{-2} \times 2^k/k! \approx 0,41$

3. La probabilité de recevoir trois appels au plus dans la minute si l'on a déjà reçu deux appels est :

   (a) $\cancel{0,66}$ (b) $\boxed{0,76}$ (c) $\cancel{0,86}$.

   $\mathbb{P}(X \leq 3|X \geq 2) = \mathbb{P}(\{X \leq 3\} \cap \{X \geq 2\})/\mathbb{P}(X \geq 2) = \mathbb{P}(2 \leq X \leq 3)/(1 - \mathbb{P}(X < 2)) = \mathbb{P}(2 \leq X \leq 3)/(1 - \mathbb{P}(X \leq 1)) = (\sum_{k=2}^{3} e^{-2} \times 2^k/k!)/(1 - \sum_{k=0}^{1} e^{-2} \times 2^k/k!) \approx 0,76$

40% des clients de South Face sont des hommes.

En supposant que le genre d'un client de South Face est indépendant de celui d'un autre client, le nombre Y d'hommes dans un échantillon de n clients est distribué selon $\mathscr{B}(n; 0,4)$; ainsi pour tout $k \in \mathbb{N} \cap [0, n]$: $\mathbb{P}(Y = k) = \binom{n}{k} \times 0,4^k \times 0,6^{1-k}$ et $E(Y) = n \times 0,4$.

4. Parmi deux-cents clients choisis au hasard, le nombre de femmes observé est, en moyenne :

   (a) $\cancel{100}$ (b) $\cancel{110}$ (c) $\boxed{120}$.

   $n = 200$; le nombre moyen de femmes dans l'échantillon est : $E(200 - Y) = 200 - E(Y) = 200 - 200 \times 0,4 = 120$.

5. La probabilité d'observer deux hommes exactement parmi dix clients est :

   (a) $\cancel{0,09}$ (b) $\boxed{0,12}$ (c) $\cancel{0,15}$.

   $n = 10$; $\mathbb{P}(Y = 2) = \binom{10}{2} \times 0,4^2 \times 0,6^8 \approx 0,12$.

6. La probabilité d'observer une majorité d'hommes parmi cinq clients est :

   (a) $\boxed{0,32}$ (b) $\cancel{0,42}$ (c) $\cancel{0,52}$.

   $n = 5$; $\mathbb{P}(Y \geq 3) = \sum_{k=3}^{5} \binom{5}{k} \times 0,4^k \times 0,6^{5-k} \approx 0,32$.

L'âge des clients du Lounge est distribué selon une loi normale de moyenne 40 et de variance 100.

A désigne l'âge d'un client choisi au hasard : $A \sim \mathscr{N}(40, 100)$. On note ϕ la fonction de répartition de $\mathscr{N}(0, 1)$.

7. La proportion de clients du Lounge âgés de trente ans au moins est :

   (a) $\cancel{0,64}$ (b) $\cancel{0,74}$ (c) $\boxed{0,84}$.

   $\mathbb{P}(A \geq 30) = 1 - \phi((30 - 40)/\sqrt{100}) = 1 - \phi(-1) = \phi(1) \approx 0,84$.

8. Le nombre moyen de jeunes[a] observé parmi deux-cents clients du Lounge est :

   (a) $\cancel{158}$ (b) $\boxed{168}$ (c) $\cancel{178}$.

   $200 \times \mathbb{P}(A < 50) = 200 \times \phi((50 - 40)/\sqrt{100}) = 200 \times \phi(1) \approx 200 \times 0,84 = 168$.

[a]moins de cinquante ans

9. La probabilité qu'un client du Lounge choisi au hasard ait de trente à cinquante ans vaut :

   (a) ~~0,48~~ (b) ~~0,58~~ (c) $\boxed{0,68}$.

   $\mathbb{P}(30 \leq A \leq 30) = \phi((50-40)/\sqrt{100}) - \phi((30-40)/\sqrt{100}) = \phi(1) - \phi(-1) == 2\phi(1) - 1 \approx 2 \times 0,84 - 1 = 0,68.$

10. L'âge au delà duquel se situent 25% des clients du Lounge[b] est :

    (a) $\boxed{47}$ (b) ~~51~~ (c) ~~54~~.

    On cherche a de sorte que : $0,25 = \mathbb{P}(A \geq a) = 1 - \phi((a-40)/\sqrt{100})$. Ainsi : $\phi((a-40)/\sqrt{100}) = 0,75 \approx \phi(0,67)$ et $a \approx 0,67 \times \sqrt{100} + 40 \approx 47$.

Exercice 2. (6pts)

Le salaire[c] d'un médecin vaudois choisi au hasard est distribué selon une loi normale de moyenne μ et de variance σ^2. Trente-et-un médecins vaudois sont tirés au sort : douze de ces médecins sont des hommes, le salaire moyen de l'échantillon est : $\bar{X} = \frac{1}{31}\sum_{i=1}^{31} X_i = 4,9$ et la variance des salaires est : $S'^2 = \frac{1}{30}\sum_{i=1}^{31}(X_i^2 - \bar{X})^2 = 1,1$ (X_i est le salaire du médecin i).

1. On s'intéresse au salaire moyen μ de l'ensemble des médécins vaudois.

   (a) Déterminez un intervalle de confiance à 95% de μ en supposant :

   i. $\sigma^2 = 0,9$
   Puisque σ^2 est connue : $IC_{0,95}(\mu) = \bar{X} \pm z_{0,975} \times \sigma/\sqrt{n} = 4,9 \pm 1,96 \times \sqrt{0,9}/\sqrt{31} \approx [4,6; 5,2]$.

   ii. σ^2 inconnu.
   Puisque σ^2 est inconnue : $IC_{0,95}(\mu) = \bar{X} \pm t_{n-1;0,975} \times S'/\sqrt{n} = 4,9 \pm 2,042 \times \sqrt{1,1}/\sqrt{31} \approx [4,5; 5,3]$.

   (b) Dans le cas i, quelle devrait être la taille de l'échantillon pour que la longueur de l'intervalle de confiance à 90% de μ soit $0,5$ au plus ?
   Comme σ^2 est connue, la longueur de $IC_{0,90}(\mu)$ est : $2 \times z_{0,95} \times \sigma/\sqrt{n}$. On cherche n de sorte que : $2 \times z_{0,95} \times \sigma/\sqrt{n} \leq 0,5$; ainsi : $n \geq (2 \times z_{0,95} \times \sigma/0,5)^2 = (2 \times 1,64 \times \sqrt{0,9}/0,5)^2 \approx 39$. L'échantillon devrait comporter 39 médecins au moins pour que la longueur de l'intervalle de confiance à 90% de μ ne dépasse pas $0,5$.

   (c) Dans le cas ii, que devrait valoir α $(0 < \alpha < 1)$ pour que la longueur de l'intervalle de confiance $1-\alpha$ de μ soit $0,64$ au plus ?
   Comme σ^2 est inconnue, la longueur de $IC_{1-\alpha}(\mu)$ est : $2 \times t_{n-1;1-\alpha/2} \times S'/\sqrt{n}$. On cherche α de sorte que : $2 \times t_{n-1;1-\alpha/2} \times S'/\sqrt{n} \leq 0,64$; ainsi : $t_{30,1-\alpha/2} \leq 0,64 \times \sqrt{31}/(2 \times \sqrt{1,1}) \approx 1,698 \approx t_{30;0,95}$ dont on déduit : $1-\alpha/2 \leq 0,95$ et $\alpha \geq 0,10$. Le niveau de confiance doit valoir 90% au plus pour que $IC_{1-\alpha}(\mu)$ ait une longueur inférieure ou égale à $0,64$.

2. On s'intéresse à la proportion p de femmes parmi les médecins vaudois.

   (a) Donnez un intervalle de confiance $1-\alpha$ $(0 < \alpha < 1)$ de p.
   $IC_{1-\alpha}(p) = F \pm z_{1-\alpha/2} \times \sqrt{F(1-F)/n}$ où F désigne la proportion de femmes dans un échantillon aléatoire de n médecins vaudois. Puisque l'échantillon de 31 médecins observé comporte 19 femmes : $IC_{1-\alpha}(p) = 19/31 \pm z_{1-\alpha/2} \times \sqrt{19/31(1-19/31)/31}$.

   (b) Pour quelles valeurs de α, l'intervalle de confiance réalisé contient la valeur $p = 0,5$?
   On cherche α de sorte que : $19/31 - z_{1-\alpha/2} \times \sqrt{19/31(1-19/31)/31} \leq 0,5$. Donc : $z_{1-\alpha/2} \geq (19/31 - 0,5)/\sqrt{19/31(1-19/31)/31} \approx 1,29$ dont on déduit : $1-\alpha/2 \geq 0,901$ et $\alpha \leq 0,198$.

Exercice 3. (5pts)

Cent-deux péruviens sont choisis au hasard : quarante sont des femmes ; la moyenne et la variance de leurs tailles conditionnellement au genre sont consignées dans Table 1. On suppose que la taille d'une péruvienne/d'un péruvien choisi(e)

		moyenne	*variance*
sexe	H	166	144
	F	162	121

Table 1: *Moyenne et variance des tailles (mesurées en cm) de quarante femmes et de soixante-deux hommes choisis au hasard dans la population péruvienne*

au hasard est une variable gaussienne.

1. On souhaite tester $\mathscr{H}_0 : \mu_{\mathrm{H}} = \mu_{\mathrm{F}}$ contre $\mathscr{H}_0 : \mu_{\mathrm{H}} \neq \mu_{\mathrm{F}}$ où μ_{H} et μ_{F} désignent respectivement la taille moyenne de l'ensemble des hommes et celle de l'ensemble des femmes, au Pérou.

   (a) On suppose que la variance des tailles est égale à 130 cm^2 chez les hommes comme chez les femmes.

[b] ou de façon équivalente le troisième quartile des âges
[c] dans une unité monétaire non précisée

i. Quelle est la décision du test au seuil de 5% ?
La variance des tailles des péruviens hommes σ_H^2 et la variance des tailles des péruviennes σ_F^2 sont connues ; la décision du test repose sur la valeur observée de : $Z = (\bar{X}_H - \bar{X}_F)/\sqrt{\sigma_H^2/n + \sigma_F^2/m}$ où $\bar{X}_H$ désigne la taille moyenne dans l'échantillon de $n = 62$ hommes ($\bar{X}_H = 166$) et $\bar{X}_F$ la taille moyenne de l'échantillon de $m = 40$ femmes ($\bar{X}_F = 162$). Puisque le test est bilatéral, on rejette $\mathscr{H}_0$ au seuil de 5% si : $|Z| > z_{0,975} \approx 1,96$. Ici, $|Z| = |(166 - 162)/\sqrt{130/62 + 130/40}| \approx 1,73 \leq 1,96$ donc on ne rejette pas $\mathscr{H}_0$ au seuil de 5% : il se peut que la taille moyenne des péruviennes soit égale à celle des péruviens.

ii. Pour quelles valeurs du seuil rejette-t-on $\mathscr{H}_0$?
$\mathscr{H}_0$ est rejetée au seuil de α dès lors que : $1,73 > z_{1-\alpha/2}$. Or $1,73$ est le quantile d'ordre $0,958$ de $\mathscr{N}(0,1)$; on en déduit : $0,958 > 1 - \alpha/2$ et $\alpha > 0,084$. Ainsi, l'hypothèse $\mathscr{H}_0$ est rejetée pour toute valeur du seuil supérieure à $8,4\%$.

(b) On suppose que la variance de la taille des hommes est inconnue mais égale à celle des femmes.

i. Quelle est la décision du test au seuil de 5% ?
σ_H^2 et σ_F^2 sont inconnues mais égales ; la décision du test repose sur la valeur observée de : $T = (\bar{X}_H - \bar{X}_F)/\tilde{S}$ où : $\tilde{S} = \sqrt{(1/n + 1/m) \times ((n-1)S_H'^2 + (m-1)S_F'^2)/(n+m-2)}$, les statistiques $S_H'^2$ et $S_F'^2$ désignant la variance des tailles dans l'échantillon d'hommes et de femmes. Puisque le test est bilatéral, on rejette $\mathscr{H}_0$ au seuil de 5% si : $|T| > t_{n+m-2;0,975} = t_{100;0,975} \approx 1,984$. Ici, $\tilde{S} = \sqrt{(1/62 + 1/40) \times (61 \times 144 + 39 \times 121)/(62 + 40 - 2)} \approx 2,36$ et $|T| = |(166 - 162)/2,36| \approx 1,697 \leq 1,984$; donc on ne rejette pas $\mathscr{H}_0$ au seuil de 5% : on juge à nouveau que la taille moyenne des péruviennes peut être la même que celle des péruviens.

ii. La p-valeur du test peut-elle être inférieure à 4% ?
Si la p-valeur du test était inférieure à 4%, l'hypothèse $\mathscr{H}_0$ serait rejetée pour toute valeur du seuil supérieure à 4%. Or, au seuil de 5% l'hypothèse $\mathscr{H}_0$ n'est pas rejetée. Donc la p-valeur ne peut pas être inférieure à 4%.

2. Au seuil de 5%, doit-on rejeter l'hypothèse : la population péruvienne est majoritairement féminine[d] ?

L'hypothèse à tester est $\mathscr{H}_0 : p \leq 0,5$ contre $\mathscr{H}_1 : p > 0,5$ où p désigne la proportion d'hommes dans la population péruvienne. La décision du test repose sur la valeur observée de $Z = \sqrt{102}(F - 0,5)/\sqrt{0,5 \times (1 - 0,5)}$ où F désigne la proportion d'hommes parmi 102 habitants du Pérou. Puisque le test est unilatéral à droite, on rejette $\mathscr{H}_0$ au seuil de 5% si $Z > z_{0,95} \approx 1,64$. Ici, $F = 62/102$ et $Z = \sqrt{102} \times (62/102 - 0,5)/0,5 = 2,18 > 1,64$. $\mathscr{H}_0$ est donc rejetée au seuil de 5% : la population péruvienne ne peut pas être composée majoritairement de femmes.

Exercice 4. (4pts)

Table 2 donne la répartition de soixante sujets par rhésus et par groupe sanguin.

		groupe			
		A	B	AB	O
rhésus	+	5	15	5	5
	−	5	10	5	10

Table 2: *Répartition observée de soixante sujets par rhésus et par groupe sanguin*

On souhaite tester au seuil de 5% l'indépendance du groupe et du rhésus d'un sujet choisi au hasard.

1. Enoncez les hypothèses $\mathscr{H}_0$ et $\mathscr{H}_1$.

L'hypothèse à tester est $\mathscr{H}_0$: *groupe et rhésus sont indépendantes* contre $\mathscr{H}_1$: *groupe et rhésus ne sont pas indépendantes*.

2. Sur quelle statistique la décision du test repose-t-elle ?

La décision du test repose sur la valeur de $D = \sum_{i=1}^{2}\sum_{j=1}^{4}(N_{i,j} - N_{i,\bullet}N_{\bullet,j}/60)^2/(N_{i,\bullet}N_{\bullet,j}/60)$ où : $N_{i,j}$ désigne le nombre de sujets parmi soixante, observés dans la classe i du rhésus et dans la classe j du groupe, où $N_{i,\bullet} = \sum_{j=1}^{4} N_{i,j}$ $(i = 1,2)$ sont les effectifs marginaux du rhésus et $N_{\bullet,j} = \sum_{i=1}^{2} N_{i,j}$ $(j = 1,\ldots,4)$ les effectifs marginaux du groupe.

3. Quelle est la distribution de la statistique de test sous $\mathscr{H}_0$?

Sous $\mathscr{H}_0$, la statistique D est distribuée selon une loi de χ^2 à $(4-1) \times (2-1) = 3$ degrés de liberté.

4. Dressez le tableau des effectifs théoriques sous hypothèse d'indépendance.

Les effectifs théoriques sous $\mathscr{H}_0$ sont les coefficients $N_{i,\bullet}N_{\bullet,j}/60$ $(i = 1,2,\ j = 1,\ldots,4)$; leur valeur est consignée dans Table 3 avec celle des effectifs marginaux.

5. Quelle est la valeur observée de la statistique de test ?

La valeur observée de la statistique de test est : $D = (5-5)^2/5 + (15-12,5)^2/12,5 + \cdots + (10-7,5)^2/7,5 = 8/3 \approx 2,67$.

[d] on testera au seuil de 5% : $\mathscr{H}_0 : p \leq 0,5$ *vs* $\mathscr{H}_1 : p > 0,5$ où p désigne la proportion d'hommes dans la population péruvienne

		groupe				
		A	B	AB	O	*total*
rhésus	+	5	12,5	5	7,5	30
	−	5	12,5	5	7,5	30
total		10	25	10	15	60

Table 3: *Effectifs théoriques sous hypothèse d'indépendance*

6. Quelle est la valeur critique du test ?

   Puisque le seuil est de 5%, la valeur critique du test est le quantile d'ordre $0,95$ de χ^2_3 soit $7,815$.

7. Déterminez la p-valeur du test[e].

   La p-valeur du test est la probabilité d'obtenir, sous $\mathscr{H}_0$, une valeur de la statistique de test plus atypique encore que celle observée. Ainsi, $p\text{-val} = \mathbb{P}(\chi^2_3 > 8/3) = 1 - \mathbb{P}(\chi^2_3 \leq 8/3) = 1 - 0,554 = 0,446$.

8. Enoncez la décision du test de deux façons :

   i en comparant la valeur observée de la statistique de test à la valeur critique.
   La valeur observée de la statistique de test ($D = 8/3$) est inférieure à la valeur critique ($7,815$) correspondant au seuil de 5% ; ainsi, au seuil de 5%, on ne rejette pas $\mathscr{H}_0$: le groupe et le rhésus sont indépendants.

   ii en comparant la p-valeur au seuil de risque.
   La p-valeur ($44,6\%$) est supérieure au seuil de risque (5%) ; ainsi, au seuil de 5%, on ne rejette pas $\mathscr{H}_0$: le groupe et le rhésus sont indépendants.

---

[e] la loi de χ^2 à trois degrés de liberté est inférieure à $8/3$ avec probabilité $0,554$

مسائل واختبارات مقترحة

جـــامعة 8 مـاي 1945 قالمـــــــــة
كلية العلوم الاقتصادية والتجارية وعلوم التسيير
قسم علوم التسيير: سنة ثانية مالية

التاريخ: 2016/01/25
المدة: 01 ساعة و 30 دقيقة

اختبار في مقياس: الإحصاء 3

تمرين رقم 01 (05ن)

نرمي قطعة نقود متزنة 12 مرة، أحسب احتمال الحصول على ما بين 7و8 صور باستخدام

1 – توزيع ذي الحدين؟

2 – التوزيع الطبيعي كتقريب لتوزيع ذي الحدين؟

3 – مثل بيانيا النتائج السابقة؟

تمرين رقم 02 (05ن)

يحتوي صندوق على 8 كرات أوزانها معطاة بالغرام كالتالي:

10.5، 19.5، 11.5، 7.5، 3.5، 2.5، 2.5، 8.5.

1 - قدر نسبة الكرات التي أوزانها أكبر من 10 غرام ضمن مجموع الكرات والتي عددها 200 وهذا باستخدام مستوى معنوية 5%؟

2 - هل يمكن القول بأن نسبة الكرات التي أوزانها أقل من 10غ ضمن مجموع الكرات تفوق 35% عند $\alpha=0.05$؟

3 – مثل بيانيا نتائج السؤال 2؟

تمرين رقم 03 (05ن)

1 – نرمي قطعة نقود متزنة 30 مرة. ما هو احتمال أن نحصل على أقل من 18 مرة صورة؟

2 – لديك الجدول الممثل لعدد الطلبة الناجحين:

	اقتصاد	تسيير	تجارة	الإجمالي
ذكور	25	20	30	75
إناث	50	45	30	125
الإجمالي	75	65	60	200

المطلوب: إعداد جدول التكرارات المتوقعة (جدول الاقتران)

تمرين رقم 04 (05ن)

وجد باحث أن متوسط دخل الأسر في مدينة (أ) يبلغ 40.000 دج بانحراف قدره 10.000 دج لعينة قدرها 30 أسرة. بينما يبلغ في مدينة (ب) 42.000 بانحراف قدره 8.000 دج لعينة قدرها 40 أسرة.

1 – أوجد تقدير لمتوسط دخل الأسر في المجتمع باستخدام نتائج المدينة (أ) عند مستوى معنوية قدره 5%؟

2 – هل دخل الأسر في المدينة (أ) يساوي دخل الأسر في المدينة (ب) عند مستوى المعنوية 5%؟

جـــــامعة 8 مـاي 1945 قالمـــــــــــة
كلية العلوم الاقتصادية والتجارية وعلوم التسيير
قسم علوم التسيير: سنة ثانية مالية

التاريخ: 2011/02/13
المدة: 01 ساعة و 30 دقيقة

تمرين رقم 01: (05ن)

لدينا مجتمع يتبع التوزيع الطبيعي بمتوسط غير معلوم وتباين قدره 25. سحبنا منه عينة من حجم n=20 ، ووجدنا أن المتوسط الحسابي هو 11.8.

المطلوب:حدد مجال الثقة للمتوسط عند مستوى الخطأ قدره $\alpha = 5\%$ ؟

تمرين رقم 02: (05ن)

لدى بنك محلي صغير 1450 حساب ادخار شخصي، برصيد متوسط قدره 3000دج وانحراف معياري 1200دج،إذا أخذ البنك عينة عشوائية من 100 حساب.

المطلوب:

1 – ما احتمال أن متوسط المدخرات لهذه الحسابات المائة سيكون أقل من 2800دج؟

2 – وضح إجابتك بتمثيل بياني؟

تمرين رقم 03: (10ن)

في دراسة إحصائية لتحديد العوامل المؤثرة على جذب الاستثمار الأجنبي المباشر (IDE)إلى الجزائر، قام الباحث – أ – بتعيين نموذج انحدار خطي بسيط لدراسة أثر الناتج المحلي الإجمالي (PIB) على استقطاب (IDE) إلى الجزائر، ثم سحب عينة من 10 سنوات، وكانت بياناتها كالتالي:

السنوات	2002	2003	2004	2005	2006	2007
PIB مليار دج	18	20	25	28	30	35
IDE مليون دج	16	16	15	22	24	27

نفرض أن العينة تقابل نموذج الانحدار الخطي $Y_i = a + bX_i + u_i$

المطلوب:

1 – قدر خط انحدار العينة السابقة باستخدام طريقة المربعات الصغرى، مع التمثيل البياني؟

2 – أوجد العدد الحقيقي d بحيث $p(|b-0.713|\langle d) = 0.95$ ؟ ثم استنتج مجال الثقة للمعلمة b؟

3 – هل يمكن استخدام هذا النموذج للتنبؤ بحجم (IDE) عند مستوى معنوية 5%؟. إذا كان الجواب نعم فكم يبلغ حجم (IDE) المتدفق إلى الجزائر في سنة 2008 مع العلم أن PIB=40؟ ما هو مقدار الدقة في التنبؤ (التقدير)؟

4 – قام الباحث – ب – بدراسة أخرى لنموذج مماثل ووجد أن خط الانحدار المقدر هو $\widehat{Y} = 1 + 0.8X$، هل يمكن اعتماد نموذج الباحث – ب -، على ضوء النتائج التي توصلت إليها سابقا، وذلك بمستوى معنوية 10%؟

5 – هل يمكن أن يبلغ الاستثمار الأجنبي المباشر المتدفق إلى الجزائر قيمة أكبر من 28 مليون دج، إذا كان الناتج المحلي الإجمالي يساوي 38 مليار دج عند مستوى دلالة قدره 5%؟

ملاحظة: استخدم ثلاث أرقام بعد الفاصلة بالتقريب لكل الحسابات.

جـــامعة 8 مـاي 1945 قالمـــــة
كلية العلوم الاقتصادية والتجارية وعلوم التسيير
قسم علوم التسيير: سنة ثانية مالية

التاريخ: 2015/06/09
المدة: 01 ساعة و 30 دقيقة

اختبار استدراكي في مقياس: إحصاء 3

تمرين رقم 01 (06ن)

يبلغ متوسط وزن كريات حديدية 22.4كلغ، بانحراف معياري قدره 0.048 كلغ. نسحب عينة عشوائية بسيطة حجمها 36 من هذا المجتمع.

1 – ما هو احتمال أن يكون متوسط العينة أقل من 22.39؟

2 – ما هو احتمال أن يكون متوسط العينة محصورا بين 22.38 و22.41؟

2 – ما هو احتمال أن يكون متوسط العينة أكبر من 22.42؟

تمرين رقم 02 (6ن)

شركة بها 500 سلك ناقل للتيار. تم اختيار 40 سلك عشوائيا، فأظهرت أن متوسط قوة المقاومة لها هي 2400 وانحرافها المعياري هو 150.

1 – ما هي 95% حدود الثقة لتقدير متوسط المقاومة بالنسبة للأسلاك ككل ؟

2 – ما هي درجة الثقة التي يمكن أن نقول بها أن متوسط المقاومة للأسلاك ككل هو 2400 ∓ 35؟

3 – مثل بيانيا نتائج السؤال 1؟

تمرين رقم 03 (04ن)

يوضح لنا الجدول الموالي عدد الطلبة في فوجين A وB الذين نجحوا، والذين رسبوا في امتحان أعطى للفوجين.

	ناجح	راسب
الفوج A	72	17
الفوج B	64	23

- اختبر فرض أنه لا توجد فروق بين الفوجين عند مستوى المعنوية 5%؟

تمرين رقم 04 (04ن)

يعزى دائما تقدم العلوم للتجريب. في هذا الصدد، تعتبر نماذج الانحدار الخطي، من بين أهم الأساليب الإحصائية، التي توفر لنا تعميمات عن الظواهر المراد دراستها. وضح ذلك؟

ملاحظة:

تستخدم ثلاث أرقاب بعد الفاصلة بالتقريب.

جـــامعة 8 مـاي 1945 قالمــــــــة
كلية العلوم الاقتصادية والتجارية وعلوم التسيير
قسم علوم التسيير: سنة ثانية مالية

التاريخ: 2016/06/21
المدة: 01 ساعة و 30 دقيقة

اختبار استدراكي في مقياس: إحصاء 3

الأسئلة (4.5)

1- ما المقصود باختبار الفرضيات، وكيف نقع في خطأ من النوع الأول؟

2- عرف المعلمة والإحصائية، وما هي العلاقة بينهما؟

3- ما هي الحالات التي نستخدم فيها اختبار كاي-مربع لاختبار فرض؟

تمرين رقم 01 (5.5ن)

نفرض أن نقاط مقياس الإحصاء لطلبة قسم معين تتبع التوزيع الطبيعي بوسط يساوي 72 وانحراف معياري قدره 10.

1- نختار طالب واحد عشوائيا. ما هو احتمال أن يكون له نقطة أكبر من 80؟

2- نسحب عينة عشوائية تتكون من 10 طلبة. ما هو احتمال أن يكون متوسط نقاطها أكبر من 80؟

3- أجب عن السؤال 2، في حالة كون المجتمع غير معروف توزيعه؟

تمرين رقم 02 (10ن)

يرغب باحث في معرفة أثر سعر المنتج الذي يبيعه على الكمية المباعة من نفس السلعة. لغرض ذلك، قام بجمع البيانات التالية

المشاهدة	1	2	3	4	5	6
سعر المنتج	3	5	6	7	7	9
الكمية المباعة	12	10	9	9	8	5

المطلوب:

1- مثل بيانيا شكل الانتشار؟ علق عليه؟

2- أوجد تقدير لخط الانحدار؟ مع التمثيل البياني على نفس الشكل؟

3- هل يعتبر النموذج المقدر مقبول إحصائيا عند مستوى معنوية قدره 10%؟

4- أوجد معامل التحديد ومعامل الارتباط؟ علق على النتائج؟

5- هل يمكن القول أن ميل خط الانحدار المقدر يكون أكبر من 0.8 عند مستوى معنوية قدره 10%؟

6- هل يمكن القول أن الكمية المباعة تكون أكبر من أقل من 7 إذا كان سعر المنتج 8 عند مستوى معنوية قدره 5%؟

ملاحظة:

تستخدم ثلاث أرقاب بعد الفاصلة بالتقريب.

جـــامعة 8 مـاي 1945 قالمـــــــة
كلية العلوم الاقتصادية والتجارية وعلوم التسيير
قسم علوم التسيير: سنة ثانية مالية

التاريخ: 2016/01/25
المدة: 01 ساعة و 30 دقيقة

اختبار في مقياس: الإحصاء 3

تمرين رقم 01 (05ن)

نرمي قطعة نقود متزنة 12 مرة، أحسب احتمال الحصول على ما بين 7و8 صور باستخدام

1 – توزيع ذي الحدين؟

2 – التوزيع الطبيعي كتقريب لتوزيع ذي الحدين؟

3 – مثل بيانيا النتائج السابقة؟

تمرين رقم 02 (05ن)

يحتوي صندوق على 8 كرات أوزانها معطاة بالغرام كالتالي:

10.5، 19.5، 11.5، 7.5، 3.5، 2.5، 2.5، 8.5.

1 - قدر نسبة الكرات التي أوزانها أكبر من 10 غرام ضمن مجموع الكرات والتي عددها 200 وهذا باستخدام مستوى معنوية 5%؟

2 - هل يمكن القول بأن نسبة الكرات التي أوزانها أقل من 10غ ضمن مجموع الكرات تفوق 35% عند $\alpha=0.05$؟

3 – مثل بيانيا نتائج السؤال 2؟

تمرين رقم 03 (05ن)

1 – نرمي قطعة نقود متزنة 30 مرة. ما هو احتمال أن نحصل على أقل من 18 مرة صورة؟

2 – لديك الجدول الممثل لعدد الطلبة الناجحين:

	اقتصاد	تسيير	تجارة	الإجمالي
ذكور	25	20	30	75
إناث	50	45	30	125
الإجمالي	75	65	60	200

المطلوب: إعداد جدول التكرارات المتوقعة (جدول الاقتران)

تمرين رقم 04 (05ن)

وجد باحث أن متوسط دخل الأسر في مدينة (أ) يبلغ 40.000 دج بانحراف قدره 10.000 دج لعينة قدرها 30 أسرة. بينما يبلغ في مدينة (ب) 42.000 بانحراف قدره 8.000 دج لعينة قدرها 40 أسرة.

1 – أوجد تقدير لمتوسط دخل الأسر في المجتمع باستخدام نتائج المدينة (أ) عند مستوى معنوية قدره 5%؟

2 – هل دخل الأسر في المدينة (أ) يساوي دخل الأسر في المدينة (ب) عند مستوى المعنوية 5%؟

جـــامعة 08 مـــاي 1945 قالمــــة
كلية العلوم الاقتصادية والتجارية وعلوم التسيير
قسم علــوم التسييــر: سنــــة ثانية ماليــة

التاريخ: ../2013/06
المدة: ساعة و نصف

*** اختبار استدراكي في مقياس: احتمالات و إحصاء تطبيقي ***

التمرين رقم 01: 07ن

أخذ مدير مصنع عينة عشوائية من 100 يوم من الاجازات المرضية، ووجد أن 30% من القوة العاملة في المصنع في فئة العمر 20-29 قد أخذوا اجازة مرضية 26 يوما من الاجمالي 100 يوم، وأن 40% من القوة العاملة في فئة العمر 30-39 قد أخذوا 37 يوما، وأن 20% في فئة العمر 40-49 قد أخذوا 24 يوما، وأن 10% في فئة العمر 50 فأكثر قد اخذوا 13 يوما اجازة مرضية.

كيف يمكن للمدير عند مستوى معنوية 5% أن يختبر الفرض أن العمر ليس عاملا في أخذ اجازة مرضية؟

تمرين رقم 02: 05ن

يرغب باحث في تقدير متوسط الأجر الأسبوعي لعدة آلاف من العاملين بأحد المصانع في حدود زائد ناقص 20 دولار، وبدرجة ثقة 99%. ويعرف الباحث من خبرته الماضية أن توزيع الأجر الأسبوعي للعاملين يتبع التوزيع الطبيعي بانحراف معياري قدره 40 دولار.

ما هو الحد الأدنى لحجم العينة المطلوب؟

تمرين رقم 03: 08ن

لدراسة مدى تأثير سعر سلعة على كميتها المباعة، تم أخذ عينة من 5 مبيعات في فترات مختلفة يلخصها الجدول الموالي:

سعر السلعة X	13	13	10	9	5
الكمية المباعة Y	9	10	12	13	16

المطلوب:

1 – مثل البيانات السابقة في شكل انتشار؟

2 – قدر معادلة خط الانحدار؟ ومثلها بيانيا على نفس الشكل السابق؟

3 – هل يمكن أن يبلغ ميل خط الانحدار قيمة أكبر من -3 عند مستوى معنوية 5%؟

4 – اختبر معنوية خط الانحدار المقدر عند مستوى معنوية 5%؟

5 – احسب معامل التحديد ومعامل الارتباط؟ علق على النتائج؟

6 – اوجد تنبؤ بمجال لكمية السلعة المباعة عند سعر 12 وبمستوى معنوية 5%؟

وزارة التعليم العالي و البحث العلمي
جامعة 8 ماي 1945 قالمة
كلية العلوم الاقتصادية والتجارية وعلوم التسيير
قسم علوم التسيير

السنة الثانية تخصص علوم التسيير
مقياس الإحصاء 3
التاريخ : 25 / 01 / 2016
المدة : ساعتان

امتحان السداسي الثالث

أسئلة نظرية :

1) عرف المصطلحات التالية : المعاينة ، التقدير ، النموذج ، طريقة المربعات الصغرى .

2) إذا قيل لك أن التقدير بنقطة أفضل من التقدير بمجال لأنه محدد و دقيق ، هل هذا صحيح ؟

3) في التقدير واختبار الفرضيات المعلمية علمت أننا قد نستخدم التوزيع الطبيعي أو توزيع ستيودنت حسب الحالة ، بين ما هي الحالة التي يستخدم فيها كل منهما ؟ ثم أرسم في رسمين بيانيين توضيحيين الفرق بينهما من ناحية توزيع المساحة حول المتوسط ؟

4) لعلك سمعت بالقصة الطريفة التي حدثت بين دونالد فيشر وذواقة الشاي في لندن منذ ما يقارب القرن والتي كانت المنطلق في ميلاد فكرة اختبار الفرضيات ، بين من خلال هذه القصة ما تفهمه من اختبار الفرضيات بشكل عام وفيما تستخدم ؟

5) علمت أن اختبار كيدو (كاي مربع) يستخدم في حالتين ، ما هما ؟

6) اشرح لماذا سمي **الانحدار الخطي البسيط** بهذه التسمية بالتحديد ؟

التمرين (1):

أخذت عينة عشوائية حجمها 144 بمتوسط 300 وانحراف معياري 100 ، من مجتمع به 5000 مفردة، ويتبع التوزيع الطبيعي.

1) أوجد التقدير بمجال لمتوسط المجتمع بحيث تكون ثقتنا 90% من أن ذلك المجال يتضمن متوسط المجتمع ؟

2) كيف يكون مجال الثقة لمتوسط المجتمع إذا أردنا أن نرفع مستوى الثقة إلى 95% ثم إلى 99% ؟

3) أوجد مجال الثقة لمتوسط المجتمع عند 95% إذا كان حجم العينة 30 و المجتمع غير محدود و التوزيع مجهول ؟

4) أوجد مجال ثقة متوسط المجتمع إذا كان حجم العينة 16 والمجتمع كبير جدا عند 95 % ؟

5) ماذا تلاحظ : (أ) من مقارنة السؤالين 1و2 ؟ . (ب) من مقارنة نتائج مجال الثقة بمستوى الثقة 95% للأسئلة 2، 3 و4 ؟

التمرين (2) :

من المعلوم أن واحد من عشرة (10%) من الشباب ـ الذين لهم اهتمام خاص بتسريحة شعورهم ـ يفضلون الجال من نوع (A) ، قامت إحدى شركات انتاج الجال بدعاية عن هذا النوع من الجال ، وللحكم على مدى نجاح هذه الحملة الدعائية أخذت عينة حجمها 200 شاب ، ووجد من بينهم 26 يفضلون الجال من النوع (A) .

1) هل هذه البيانات دليلا كافيا على ازدياد نسبة الشباب الذين يفضلون الجال من النوع (A) عند مستوى معنوية 5% ؟

2) هل تتغير النتيجة عند مستوى المعنوية 1% ؟

التمرين (3) :

يحتاج صاحب مصنع طائرات أن يشتري صفائح ألمنيوم بسمك (0.05 in) ، الصفائح الأقل سمكا غير ملائمة والأكثر سمكا أثقل من اللازم . يأخذ المنتج عينة عشوائية من 100 صحيفة من مورد الصفائح ويجد أن متوسط سمكها (0.048 in) وانحراف معياري (0.01 in) .

1) هل يجب على المنتج شراء صفائح الألمنيوم من هذا المورد إذا كان يرغب في أن يتخذ قرارا عند مستوى معنوية 5% ؟

2) هل يتغير القرار إذا كان مستوى الثقة 1% ؟

3) اختبر الفرض إذا كان أخذت عينة من 25 صفيحة عند 5% ؟

التمرين (4):

يوضح الجدول التالي بيانات عن علامات 12 طالب في امتحان إحصاء أثناء دراستهم في الجامعة (X) وفي امتحان شامل لمسابقة توظيف (Y) بعد التخرج :

علامات مادة الإحصاء (X)	18	14	10	15	7	12	13	8	9	17	15	12
علامات مسابقة التوظيف (Y)	20	11	14	16	10	10	17	11	12	20	18	12

ويبين الجدول التالي مختلف الحسابات التي أجريت بهدف صياغة نموذج انحدار بسيط:

$\sum(Y_i-\bar{Y})^2$	$\sum ei^2$	$\sum x^2$	$\sum xy$	$\sum X^2$	$\sum XY$	$\sum x$	$\sum y$	$\sum Y$	$\sum X$
158.25	55.982	135	118	2010	2255	0	0	171	150

حيث أن : $ei = Y_i - \hat{Y}$ ، $x = X - \bar{X}$ ، $y = Y - \bar{Y}$

المطلوب :

1) بناءا على الجدول السابق أوجد خط الانحدار المقدّر ؟ ثم أحسب كلاً من: أ- تباين الحد العشوائي $\hat{\sigma}_u^2$، ب- الانحراف المعياري للتقدير لـ $\hat{a},\hat{b}$ ؟ $\sigma_{\hat{b}}^2,\sigma_{\hat{a}}^2$ ؟

2) أوجد مجال الثقة للمعلمتين a و b وهذا بمستوى دلالة قدره 5% .

3) اختبر فرض أن قيمة b=1 بدرجة ثقة قدرها 95% ؟

4) هل يمكن القول أن خط الانحدار المقدر هو ذو دلالة إحصائية عند مستوى المعنوية 5 % ؟

5) هل يمكن القول أن هناك تناسب بين المتغيرين x و y عند مستوى المعنوية 5% ؟

6) أحسب معامل التحديد و معامل الارتباط مع التعليق على النتيجة ؟

7) إذا علمت أن الطالب "عصام" حصل على العلامة 16 في الإحصاء أثناء الدراسة لكنه تغيب عن هذه المسابقة، فما هي العلامة التقديرية التي تتوقع أن يحصل عليها ؟

... ولِكُلِّ مُجتهِدٍ نَصيب

Partiel Statistiques Appliquées

Mardi 16 janvier 2007 : 8h30 - 11h30
Cours de F. GARDES

Sont autorisées les calculatrices. Les six parties sont indépendantes les unes des autres.

Partie I : Question de cours (1,5 points)

Traitez au choix l'une des questions A ou B.

A. Pourquoi dit-on que les estimateurs par les MCO d'un modèle économique sont des variables aléatoires ?

B. Enoncez le théorème Central Limite. Donnez-en une application.

Partie II : Question de cours (2,5 points)

Traitez au choix l'une des questions A ou B.

A. Expliquez :

1. Le principe de la méthode des MCO pour l'estimation d'un modèle linéaire : $y = X\beta + u$
2. Le principe de la méthode du Maximum de Vraisemblance.

B.

1. Justifiez la distance utilisée dans le tests du χ^2 :

$$D_n = \sum_{i=1}^{r}\sum_{j=1}^{s}\frac{(n_{ij} - n_{i.}.n_{.j}/n)^2}{n_{i.}.n_{.j}/n} = n\left(\sum_{i=1}^{r}\sum_{j=1}^{s}\frac{{n_{ij}}^2}{n_{i.}n_{.j}} - 1\right) = n\left(\sum_{i=1}^{r}\sum_{j=1}^{s}\frac{{p_{ij}}^2}{p_{i.}p_{.j}} - 1\right)$$

   pour n_{ij} le nombre d'individus correspondant aux items de réponse i et j de deux caractères qualitatifs (à respectivement r et s modalités), $n_{i.}$ et $n_{.j}$ les sommes marginales, (avec $p_{ij} = \frac{n_{ij}}{n}$, $p_{i.} = \frac{n_{i.}}{n}$ et $p_{.j} = \frac{n_{.j}}{n}$) et n la taille de la population enquêtée.
2. Donnez un exemple économique où ce test peut être utilisé.

Partie III : Exercice (5 points)

1. Définissez la convergence en Probabilité et la convergence en Moyenne Quadratique.
2. Quelle relation existe-t-il entre elles ?
3. On considère un estimateur T_n sans biais de l'élasticité de la consommation permanente C^p par rapport au revenu permanent Y^p, obtenue par les MCO sur une enquête de consommation auprès de n ménages. Expliquez ce que signifie cette hypothèse d'absence de biais.
4. Supposons que la valeur vraie de cette élasticité soit 1. Démontrez que l'estimateur T_n est convergent en faisant l'hypothèse que la variance de T_n tend vers 0 quand la taille de l'échantillon tend vers l'infini.

5. Rappelez la formule matricielle qui permet d'obtenir cet estimateur des MCO par l'estimation de l'équation :
$$lnC_i^p = \beta_1 lnY_i^p + \beta_2 D_i + \beta_3 + u_i = X_i\beta + u_i$$
du logarithme de la consommation permanente lnC_i^p (de la famille i) par rapport au logarithme du revenu permanent de cette même famille lnY_i^p, à la taille de la famille D_i et à une constante β_3, ces trois variables explicatives étant rassemblées dans le vecteur X_i, avec un résidu u_i (on donnera la formule pour le vecteur de paramètres β, vecteur colonne contenant β_1 dans sa première ligne).
6. Quelles sont les propriétés de cet estimateur sous les hypothèses habituelles ?

Partie IV : Question de cours (2 points)

Traitez les deux questions ci-dessous :

1. Définissez les deux risques d'erreurs dans la théorie des tests. Pourquoi privilégie-t-on le risque de première espèce ?
2. On veut tester l'hypothèse que l'élasticité de la consommation permanente par rapport au revenu est égale à 1. En reprenant les notations de la partie précédente, énoncez l'hypothèse nulle et l'hypothèse alternative et caractérisez l'erreur de première espèce.

Partie V : Exercice (6 points)

Nous disposons d'un échantillon de 25 étudiants tiré de manière aléatoire dans la population étudiante ayant souscrit un emprunt. La dette moyenne dans cet échantillon est de 10290 euros. On considère que la dette d'un étudiant ayant souscrit un emprunt suit une loi normale $N(m, \sigma^2)$. L'écart-type théorique de cette dette, sur l'ensemble de la population étudiante ayant souscrit un emprunt, est supposé connu : $\sigma = 2500$ euros.

1. Construisez un intervalle bilatéral de confiance à 90% pour estimer la dette moyenne m de l'ensemble de la population étudiante ayant souscrit un emprunt.
2. Même question pour un intervalle bilatéral de confiance à 99%.
3. Expliquez l'effet de l'augmentation du niveau de confiance sur la longueur de l'intervalle.
4. On suppose dans cette question (et uniquement dans cette question) qu'on ne connaît pas l'écart-type théorique σ de la dette des étudiants ayant souscrit un emprunt. Par contre, on dispose de l'écart-type empirique s obtenu à partir du même échantillon : $s = 2000$ euros. Déterminez un intervalle bilatéral de confiance à 90% et un intervalle bilatéral de confiance à 95% pour la dette moyenne m.
5. On veut tester l'hypothèse que la dette moyenne de la population étudiante qui souscrit un emprunt est inférieure ou égale à 9300 euros. Ecrivez les hypothèses nulle et alternative et, connaissant l'écart-type théorique de 2500 euros, effectuez le test pour les deux seuils $\alpha = 5\%$ et $\alpha = 1\%$.
6. Pour un seuil $\alpha = 1\%$, quelle taille d'échantillon est nécessaire pour obtenir une réponse négative à un test d'une dette moyenne (de la population étudiante ayant souscrit un emprunt) inférieure ou égale à 9300 euros ?

Partie VI : Exercice (3 points)

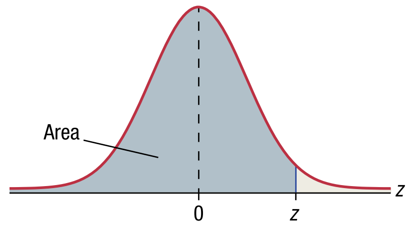
1. Soit (X_i) un échantillon tiré de façon i.i.d dans une loi normale $N(m, \sigma^2)$. Calculez l'estimateur du Maximum de Vraisemblance du paramètre m. Comment s'appelle l'estimateur que vous trouvez ?
Indication : Fonction de densité de la loi normale $\mathcal{N}\left(m, \sigma^2\right)$:
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} \qquad x \in R$$
2. On dispose d'un échantillon i.i.d tiré dans la loi de densité :
$$f(y; a) = ay^{a-1} \qquad\qquad avec\ \ 0 < y < 1$$
Donnez un estimateur de a suivant la méthode des moments.

المصادر

- الجبوري شلال حبيب عبد الله، الإحصاء التطبيقي، دار الحكمة للطباعة والنشر، العراق، 1991.
- السعدي رجال، بحوث العمليات، دار رجزو، الجزائر، 2004.
- إسماعيل سعيد السيد علي، مبادئ الإحصاء الوصفي والتطبيقي، مؤسسة حورس الدولية، الأردن، 2011.
- حسين علي بخيت، سحر فتح الله، الاقتصاد القياسي، دار اليازوري، الأردن، 2007.
- دومينيك سالفادور، الإحصاء والاقتصاد القياسي، ديوان المطبوعات الجامعية، الجزائر، 1993.
- شبيجل، مواري ر.، الإحصاء، الدار الدولية للاستثمارات الثقافية، مصر، 2008.
- عبد الحفيظ محمد فوزي مصطفى، الاستدلال الإحصائي (2) نظرية اختبار الفرضيات، مجموعة النيل العربية، مصر، 2002.
- علاء الدين محمد، فرحان نور الدين حسن، مبادئ الأسلوب الإحصائي، هيئة المعاهد الفنية، العراق، 1988.
- ماريجا نوروسيس، تحليل البيانات باستخدام SPSS، شعاع للنشر والعلوم، سورية، 2010.
- محفوظ جودة، التحليل الإحصائي المتقدم باستخدام SPSS، دار وائل، الأردن، 2008.
- محمد صالح تركي القريشي، مقدمة في الاقتصاد القياسي، الوراق للنشر و التوزيع، الأردن، 2004.
- نوروسيس ماريجا، تحليل البيانات باستخدام SPSS، شعاع للنشر والعلوم، سورية، 2010.
- وليد إسماعيل السيفو، وآخرون، أساسيات الاقتصاد القياسي التحليلي، دار الأهلية للنشر والتوزيع، الأردن، 2006
- حسين علي بخيت، سحر فتح الله، الاقتصاد القياسي، دار اليازوري، الأردن، 2007.
- Abdelkader Elkhider, El mustapha Kchirid, Bachir Lakhdar et Abdelhamid El Bouhadi, Statistique Decisionnelle : Echantillonnage, estimation, comparaison et test, Imprimerie Papeterie el Watanya, Maroc, 2004.
- Allan G. Bluman, Elementary Statistics: A step by step approach, -Hill Irwin, USA, 2005. 2004.
- David R. Anderson, Dennis J. Sweeney, Thomas A. Williams . Statistiques pour l’économie et la gestion, De Boeck, 2005, France.
- Douglas A. Lind, William G. Marchal and Samuel A. Wathen, Statistical techniques in business and economics, Mcgraw-Hill Irwin, USA, 2005.
- The Rand Corporation . A Million Random Digits with 100,000 Normal Deviates, The Free Press, 1955, USA

B Standard Normal Distribution

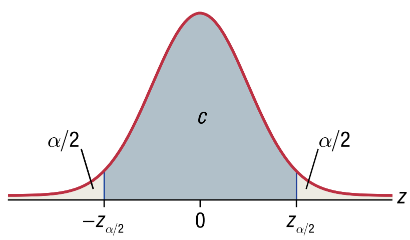
Numerical entries represent the probability that a standard normal random variable is between $-\infty$ and z where $z = \dfrac{x-\mu}{\sigma}$.



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

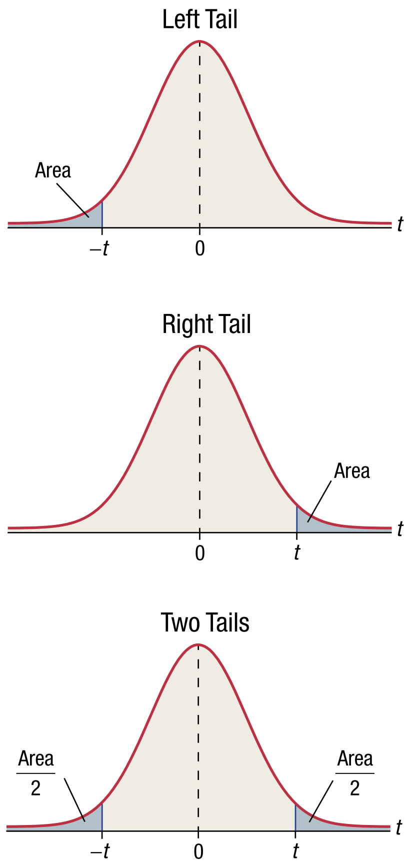
Critical Values of z

Level of Confidence, c	$\alpha = 1 - c$	$z_{\alpha/2}$
0.80	0.20	1.28
0.85	0.15	1.44
0.90	0.10	1.645
0.95	0.05	1.96
0.98	0.02	2.33
0.99	0.01	2.575

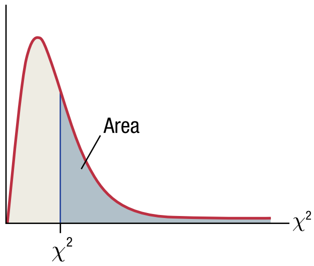


D Critical Values of *t*

	Area in One Tail				
	0.100	0.050	0.025	0.010	0.005
	Area in Two Tails				
df	**0.200**	**0.100**	**0.050**	**0.020**	**0.010**
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
31	1.309	1.696	2.040	2.453	2.744
32	1.309	1.694	2.037	2.449	2.738
34	1.307	1.691	2.032	2.441	2.728
36	1.306	1.688	2.028	2.434	2.719
38	1.304	1.686	2.024	2.429	2.712
40	1.303	1.684	2.021	2.423	2.704
45	1.301	1.679	2.014	2.412	2.690
50	1.299	1.676	2.009	2.403	2.678
55	1.297	1.673	2.004	2.396	2.668
60	1.296	1.671	2.000	2.390	2.660
70	1.294	1.667	1.994	2.381	2.648
80	1.292	1.664	1.990	2.374	2.639
90	1.291	1.662	1.987	2.368	2.632
100	1.290	1.660	1.984	2.364	2.626
120	1.289	1.658	1.980	2.358	2.617
200	1.286	1.653	1.972	2.345	2.601
300	1.284	1.650	1.968	2.339	2.592
400	1.284	1.649	1.966	2.336	2.588
500	1.283	1.648	1.965	2.334	2.586
750	1.283	1.647	1.963	2.331	2.582
1000	1.282	1.646	1.962	2.330	2.581
∞	1.282	1.645	1.960	2.326	2.576



H Critical Values of χ^2



Area to the Right of the Critical Value of χ^2

df	0.995	0.990	0.975	0.950	0.900	0.100	0.050	0.025	0.010	0.005
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.195	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169

I Critical Values of *F* (Area = 0.050)

Numerator Degrees of Freedom

Denominator Degrees of Freedom	1	2	3	4	5	6	7	8	9
1	161.4476	199.5000	215.7073	224.5832	230.1619	233.9860	236.7684	238.8827	240.5433
2	18.5128	19.0000	19.1643	19.2468	19.2964	19.3295	19.3532	19.3710	19.3848
3	10.1280	9.5521	9.2766	9.1172	9.0135	8.9406	8.8867	8.8452	8.8123
4	7.7086	6.9443	6.5914	6.3882	6.2561	6.1631	6.0942	6.0410	5.9988
5	6.6079	5.7861	5.4095	5.1922	5.0503	4.9503	4.8759	4.8183	4.7725
6	5.9874	5.1433	4.7571	4.5337	4.3874	4.2839	4.2067	4.1468	4.0990
7	5.5914	4.7374	4.3468	4.1203	3.9715	3.8660	3.7870	3.7257	3.6767
8	5.3177	4.4590	4.0662	3.8379	3.6875	3.5806	3.5005	3.4381	3.3881
9	5.1174	4.2565	3.8625	3.6331	3.4817	3.3738	3.2927	3.2296	3.1789
10	4.9646	4.1028	3.7083	3.4780	3.3258	3.2172	3.1355	3.0717	3.0204
11	4.8443	3.9823	3.5874	3.3567	3.2039	3.0946	3.0123	2.9480	2.8962
12	4.7472	3.8853	3.4903	3.2592	3.1059	2.9961	2.9134	2.8486	2.7964
13	4.6672	3.8056	3.4105	3.1791	3.0254	2.9153	2.8321	2.7669	2.7144
14	4.6001	3.7389	3.3439	3.1122	2.9582	2.8477	2.7642	2.6987	2.6458
15	4.5431	3.6823	3.2874	3.0556	2.9013	2.7905	2.7066	2.6408	2.5876
16	4.4940	3.6337	3.2389	3.0069	2.8524	2.7413	2.6572	2.5911	2.5377
17	4.4513	3.5915	3.1968	2.9647	2.8100	2.6987	2.6143	2.5480	2.4943
18	4.4139	3.5546	3.1599	2.9277	2.7729	2.6613	2.5767	2.5102	2.4563
19	4.3807	3.5219	3.1274	2.8951	2.7401	2.6283	2.5435	2.4768	2.4227
20	4.3512	3.4928	3.0984	2.8661	2.7109	2.5990	2.5140	2.4471	2.3928
21	4.3248	3.4668	3.0725	2.8401	2.6848	2.5727	2.4876	2.4205	2.3660
22	4.3009	3.4434	3.0491	2.8167	2.6613	2.5491	2.4638	2.3965	2.3419
23	4.2793	3.4221	3.0280	2.7955	2.6400	2.5277	2.4422	2.3748	2.3201
24	4.2597	3.4028	3.0088	2.7763	2.6207	2.5082	2.4226	2.3551	2.3002
25	4.2417	3.3852	2.9912	2.7587	2.6030	2.4904	2.4047	2.3371	2.2821
26	4.2252	3.3690	2.9752	2.7426	2.5868	2.4741	2.3883	2.3205	2.2655
27	4.2100	3.3541	2.9604	2.7278	2.5719	2.4591	2.3732	2.3053	2.2501
28	4.1960	3.3404	2.9467	2.7141	2.5581	2.4453	2.3593	2.2913	2.2360
29	4.1830	3.3277	2.9340	2.7014	2.5454	2.4324	2.3463	2.2783	2.2229
30	4.1709	3.3158	2.9223	2.6896	2.5336	2.4205	2.3343	2.2662	2.2107
40	4.0847	3.2317	2.8387	2.6060	2.4495	2.3359	2.2490	2.1802	2.1240
60	4.0012	3.1504	2.7581	2.5252	2.3683	2.2541	2.1665	2.0970	2.0401
120	3.9201	3.0718	2.6802	2.4472	2.2899	2.1750	2.0868	2.0164	1.9588
∞	3.8415	2.9957	2.6049	2.3719	2.2141	2.0986	2.0096	1.9384	1.8799

I Critical Values of *F* (Area = 0.010)

Numerator Degrees of Freedom

Denominator Degrees of Freedom	1	2	3	4	5	6	7	8	9
1	4052.1807	4999.5000	5403.3520	5624.5833	5763.6496	5858.9861	5928.3557	5981.0703	6022.4732
2	98.5025	99.0000	99.1662	99.2494	99.2993	99.3326	99.3564	99.3742	99.3881
3	34.1162	30.8165	29.4567	28.7099	28.2371	27.9107	27.6717	27.4892	27.3452
4	21.1977	18.0000	16.6944	15.9770	15.5219	15.2069	14.9758	14.7989	14.6591
5	16.2582	13.2739	12.0600	11.3919	10.9670	10.6723	10.4555	10.2893	10.1578
6	13.7450	10.9248	9.7795	9.1483	8.7459	8.4661	8.2600	8.1017	7.9761
7	12.2464	9.5466	8.4513	7.8466	7.4604	7.1914	6.9928	6.8400	6.7188
8	11.2586	8.6491	7.5910	7.0061	6.6318	6.3707	6.1776	6.0289	5.9106
9	10.5614	8.0215	6.9919	6.4221	6.0569	5.8018	5.6129	5.4671	5.3511
10	10.0443	7.5594	6.5523	5.9943	5.6363	5.3858	5.2001	5.0567	4.9424
11	9.6460	7.2057	6.2167	5.6683	5.3160	5.0692	4.8861	4.7445	4.6315
12	9.3302	6.9266	5.9525	5.4120	5.0643	4.8206	4.6395	4.4994	4.3875
13	9.0738	6.7010	5.7394	5.2053	4.8616	4.6204	4.4410	4.3021	4.1911
14	8.8616	6.5149	5.5639	5.0354	4.6950	4.4558	4.2779	4.1399	4.0297
15	8.6831	6.3589	5.4170	4.8932	4.5556	4.3183	4.1415	4.0045	3.8948
16	8.5310	6.2262	5.2922	4.7726	4.4374	4.2016	4.0259	3.8896	3.7804
17	8.3997	6.1121	5.1850	4.6690	4.3359	4.1015	3.9267	3.7910	3.6822
18	8.2854	6.0129	5.0919	4.5790	4.2479	4.0146	3.8406	3.7054	3.5971
19	8.1849	5.9259	5.0103	4.5003	4.1708	3.9386	3.7653	3.6305	3.5225
20	8.0960	5.8489	4.9382	4.4307	4.1027	3.8714	3.6987	3.5644	3.4567
21	8.0166	5.7804	4.8740	4.3688	4.0421	3.8117	3.6396	3.5056	3.3981
22	7.9454	5.7190	4.8166	4.3134	3.9880	3.7583	3.5867	3.4530	3.3458
23	7.8811	5.6637	4.7649	4.2636	3.9392	3.7102	3.5390	3.4057	3.2986
24	7.8229	5.6136	4.7181	4.2184	3.8951	3.6667	3.4959	3.3629	3.2560
25	7.7698	5.5680	4.6755	4.1774	3.8550	3.6272	3.4568	3.3239	3.2172
26	7.7213	5.5263	4.6366	4.1400	3.8183	3.5911	3.4210	3.2884	3.1818
27	7.6767	5.4881	4.6009	4.1056	3.7848	3.5580	3.3882	3.2558	3.1494
28	7.6356	5.4529	4.5681	4.0740	3.7539	3.5276	3.3581	3.2259	3.1195
29	7.5977	5.4204	4.5378	4.0449	3.7254	3.4995	3.3303	3.1982	3.0920
30	7.5625	5.3903	4.5097	4.0179	3.6990	3.4735	3.3045	3.1726	3.0665
40	7.3141	5.1785	4.3126	3.8283	3.5138	3.2910	3.1238	2.9930	2.8876
60	7.0771	4.9774	4.1259	3.6490	3.3389	3.1187	2.9530	2.8233	2.7185
120	6.8509	4.7865	3.9491	3.4795	3.1735	2.9559	2.7918	2.6629	2.5586
∞	6.6349	4.6052	3.7816	3.3192	3.0173	2.8020	2.6393	2.5113	2.4074