بسم الله الرحمن الرحيم

جامعة السودان المفتوحة

برنامج علوم الحاسوب وتقنية المعلومات

الإحصاء التطبيقي

رمز المقرر ورقمه: ريض 2029

المؤلف:

د. أحمد الشيخ محمد أحمد

التحكيم العلمي:

بروفيسور . عبد العزيز محمد أحمد تنقساوي

التصميم التعليمي :

أ . محمد حسن الزبير

التصميم الفني :

أسعد جلال الدين محمد بشير

منشورات جامعة السودان المفتوحة، الطبعة الأولى 2006

جميع الحقوق محفوظة لجامعة السودان المفتوحة، لا يجوز إعادة إنتاج أي جزء من هذا المقرر، وبأي وجه من الوجوه، إلا بعد الموافقة المكتوبة من الجامعة.

مقدمة المقرر

بسم الله الرحمن الرحيم ونصلى ونسلم على رسوله الأمين أشرف الأنبياء والمرسلين وعلى آله وصحبه أجمعين.

مرحباً بك عزيزي الدارس في هذا المقرر، مقرر الإحصاء التطبيقي والذي يتضح لك من اسمه انه يهتم بتطبيقات علم الإحصاء في مجالات الحياة العملية المختلفة، كما ستقف على ذلك بنفسك من خلال دراستك لوحدات المقرر الست المختلفة وبالتأكيد ستلاحظ وتدرك أهمية علم الإحصاء عامة والإحصاء التطبيقي بصفة خاصة ودوره في التقدم والتنمية في وقتنا الحاضر.

تناولنا في الوحدة الأولى من هذا المقرر توزيعات المعاينة للإحصاءات المختلفة مع بعض الأمثلة التوضيحية، أما الوحدة الثانية (مبادئ نظرية التقدير) فقد وضحا من خلالها نظرية تقدير المعالم وخواص المقدر الجيد وشرحنا أيضاً تقدير فترات الثقة لبعض معالم المجتمع وفي الوحدة الثالثة (اختبارات الفروض) عرفنا نوعي الخطأ و أوضحنا الخطوات التي يجب إتباعها لاختبار فرض معين، كما شرحنا أيضاً من خلال هذه الوحدة اختبارات المعنوية مع تقديم أمثلة لبعض الاختبارات. أما الوحدة الرابعة وهي (نظرية العينات) فقد

أوضحنا من خلال هذه الوحدة كيفية تقدير الفترة واختبارات المعنوية وذلك عندما يكون حجم الوحدة صغير، لذلك قدمنا أمثلة لبعض التوزيعات. ومن خلال الوحدة الخامسة (الارتباط والانحدار) فقد قدمنا المقاييس المختلفة لمعرفة مدي ارتباط متغيرين ببعضهما البعض وإيجاد علاقة تربط بين هذين المتغيرين. في الوحدة الأخيرة أوضحنا محاسن ومساوئ الاختبارات اللا معلمية وقدمنا بعض النماذج لهذه الاختبارات.

نأمل عزيزي الدارس عند دراستك لهذا المقرر أن تقوم بحل التدريبات وأسئلة لتقويم الذاتي التي تخللت كل وحدة حتى تتم الفائدة المرجوة.

الأهداف العامة للمقرر



عزيزي الدارس ،

بعد فراغك من دراسة هذا المقرر ينبغي أن تكون قادراً على أن :

(1) تتعرف على بعض المصطلحات الإحصائية .

(2) تحسب بعض المقاييس الإحصائية .

(3) تقدر معالم المجتمع الإحصائي .

(4) تطبق اختبارات الفروض المعنوية لاتخاذ القرار المناسب ، للعينات كبيرة الحجم أو صغيرة الحجم .

(5) تشرح بعض أنواع التوزيعات الإحصائية .

(6) تحلل الانحدار وتقيس درجة الارتباط بين متغيرين .

(7) تطبق طرق الاختبارات حرة التوزيع .

(8) تشرح متى تستخدم الاختبارات حرة التوزيع وتطبقها في حل المسائل .


الوحدة الأولى
مبادئ نظرية العينات
وتوزيعات المعاينة

محتويات الوحدة

الموضوع	الصفحة
المقدمة	3
تمهيد	3
أهداف الوحدة	4
1. مبادئ نظرية العينات وتوزيعات المعاينة	5
1.1 المجتمع والعينة والمعاينة بإرجاع وبدون إرجاع	5
2.1 العينة	6
3.1 المعاينة	8
2 . توزيعات المعاينة	9
1.2 توزيع المعاينة للأوساط	11
2.2 توزيع المعاينة للنسب	18
3.2 توزيع المعاينة للفرق بين الأوساط	21
4.2 توزيع المعاينة للفروق بين النسب	25
الخلاصة	29
لمحة مسبقة عن الوحدة الدراسية التالية	29
إجابات التدريبات	31
إجابات أسئلة التقويم الذاتي	35
مسرد المصطلحات	36
المراجع	37

المقدمة

تمهيد ،،

مرحباً بك عزيزي الدارس في هذه الوحدة والتي تعتبر مدخلاً لمقرر الإحصـــاء التطبيقي حيث سنقدم لك من خلالها تعريفاً للمجتمـــع الإحصـــائي والعينـــة العشـــوائية والمعاينة وطرقها وستجد كذلك تعريفاً وافياً لتوزيعات المعاينة لكل من الوسط الحســـابي والنسب والفرق بين الأوساط والفرق بين النسب مع تقديم بعض الأمثلة وذلـــك لتثبيـــت وتركيز التعاريف والمفاهيم الإحصائية بذهنك أكثر حتـــى يســـهل عليـــك اســـتذكارها ومراجعتها .

لذا ننصحك عزيزي الدارس أن تحل جميع أسئلة التدريبات والتقويم الذاتي التـــي ستجدها في ثنايا هذه الوحدة بنفسك وإذا وجدت أي صعوبة في ذلك يمكنك الرجوع إلـــى مشرفك الأكاديمي. كما نأمل أن تزودنا رأيك ونقدك البنّاء

أهداف الوحدة



عزيزي الدارس ،،

بعد دراستك لهذه الوحدة ينبغي أن تكون قادراً على أن :

(1) توضح شكل توزيع المعاينة النظري للوسط إذا كان توزيع المجتمع طبيعي أو غير طبيعي .

(2) تجد الوسط الحسابي و الانحراف المعياري لتوزيع المعاينة للأوساط .

(3) تجد الوسط الحسابي و الانحراف المعياري لتوزيع المعاينة للنسب.

(4) تجد الوسط الحسابي و الانحراف المعياري لتوزيع المعاينة للفروق بين الأوساط الحسابية .

(5) تجد الوسط الحسابي و الانحراف المعياري لتوزيع المعاينة للفروق بين النسب .

1. مبادئ نظرية العينات وتوزيعات المعاينة

1.1المجتمع والعينة والمعاينة بإرجاع وبدون إرجاع

Population, Sample and Sampling With and Without Replacement

تستخدم لفظة مجتمع في علم الإحصاء لتعنى مجموعة من مفردات تهم الباحث سواء كانت مكونة من بشر أو أشياء يمكن مشاهدتها في وضع معين. ومفردات المجتمع تسمى وحدات أولية (Elementary units) وتمتلك الوحدات الأولية خصائص (Characteristics) وهذه الخصائص قد تكون نوعية (Qualitative) أو كمية (quantitative) فمثلاً في دراسة لميزانية الأسرة في مدينة الخرطوم، يضم المجتمع كل الأسر (households) في مدينة الخرطوم والخصائص التي ينبغي مشاهداتها هي دخل ومنصرفات الأسرة (الوحدة الأولية) وناتج مشاهدة الوحدة الأولية يسمى المشاهدة (Observation) وبما أن المشاهدة تحدث لكل وحدة أولية في عملية التقصي والدراسة، فإن المجتمع يعتبر مجموعة كل المشاهدات التي تمت لدراسة وضع معين. وتصنف المجتمعات إلى نوعين:

(أ) **مجتمع محدود (Finite population)** وهو المجتمع الذي يتكون من عدد محدود من المشاهدات. ومن أمثلة المجتمع المحدود عدد الذكور بين طلاب جامعة الخرطوم.

(ب) **مجتمع غير محدود(Infinite population)** وهو المجتمع الذي يتكون من عدد غير محدود من المشاهدات. ومثال للمجتمع غير المحدود هو رمى قطعة نقود معدنية بصورة مستمرة وغير متناهية.

وفى حالة المجتمع غير المحدود من الصعوبة بمكان تأمين معلومات عنه لاتخاذ قرار معين. وعليه يمكن اتخاذ هذا القرار على أساس عينة مسحوبة(Sample) من هذا المجتمع.

2.1العينة

العينة هي عبارة عن مجموعة جزئية من المجتمع وتتكون من عدد محدود من المفردات أو المشاهدات. ولضمان تمثيل المجتمع بصورة سليمة ينبغي أن تكون العينة عشوائية (random). ويقصد بالعينة العشوائية أن تكون لكل مفردة في المجتمع نفس الفرصة أو الاحتمال في الاختيار ضمن العينة المسحوبة أو بصورة أخرى أن يكون

لكل عينة مسحوبة من المجتمع نفس الفرصة أو الاحتمال في الاختيار. ومن الأسباب التي تجعلنا نلجأ لسحب عينة من المجتمع نستدل من خلالها على خصائص ومعالم المجتمع نذكر ما يلي :

(أ) أن يكون المجتمع غير محدود.

(ب) أن يكون المجتمع محدود ولكن مكون من عدد ضخم من المشاهدات وبالتالي يصعب عملياً دراستها.

(ج) أن تؤدى دراسة المجتمع إلى إفناء مفرداته مثلاً للتأكد من جودة نوع معين من أعواد ثقاب، فان دراسة مجتمع أعواد الثقاب سيؤدى ذلك إلى هلاك كل مفرداته.

(د) عدم توفر ميزانية كافية أو وقت كاف لدراسة المجتمع.

(هـ) لكثير من أنواع البيانات يصعب الوصول للمجتمع. وعمليا ينبغي التعامل مع ما يتوفر من جزء من البيانات.

(و) استحالة الوصول لكل مفردات المجتمع.

3.1المعاينة Sampling

ونقصد بكلمة المعاينة (Sampling) عملية دراسة مجتمع علـــى أســـاس عينـــة مسحوبة من هذا المجتمع. وعند اختيار مفردات عينة من مجتمع، توجد طريقتان لسحب المفردة من المجتمع هما :

(أ) المعاينة بإرجاع (With replacement) أو السحب بإرجاع :

و السحب بإرجاع يعني إرجاع المفردة المسحوبة مرة أخرى للمجتمـــع قبـــل أن تجرى عملية السحب التالية ولهذا يمكن من ناحية نظرية أن يتم سحب مفردة معينـــة أكثر من مرة.

(ب) المعاينة دون إرجاع (Without replacement) أو السحب دون إرجاع :

والسحب دون إرجاع يعني أن المفردة لا تعاد مرة أخرى للمجتمع، وعليه لا تظهـــر المفردة التي تم سحبها غير مرة واحدة فقط. ويترتب على عملية السحب بإرجاع وبدون إرجاع اعتبارات فنية سوف تتم الإشارة إليها في تطوراتنا القادمة.

2. توزيعات المعاينة Sampling Distributions

تتلخص فكرة توزيعات المعاينة في سحب كل العينات الممكنة ذات الحجم المتماثل n من مجتمع وحيد أو من مجتمعين ولكن في هذه الحالة ليس بالضرورة أن تكون أحجام العينات متماثلة بالنسبة للمجتمعين، حيث يمكن اخذ عينات من المجتمع الأول حجمها 20 مثلاً وعينات من المجتمع الثاني حجمها 30.

نفترض أن لدينا مجتمعاً واحداً نرمز لحجمه بالرمز N وأخذنا منه عدداً من العينات بنفس الحجم n ثم حسبنا لكل عينة إحصاءة معينه نرمز لها بالرمز S . ولتبسيط وسهولة العرض، نفترض إننا أخذنا كل العينات الممكنة وحجم كل منها n=25 ثم حسبنا قيمة S لكل عينة، فإننا نحصل علي S_1 ، S_2 ،....., S_k

العينة	الإحصاءة
1	S_1
2	S_2
3	S_3
.....	…
k	S_k

فإننا بصفة عامة نسمى $S_1, S_2, S_3 \ldots\ldots\ldots\ldots S_k$ توزيع معاينة للإحصاءة S.

وينبغي أن نوضح هنا أن لكل توزيع معاينة لإحصاءة معينة وسط حسابي ويرمز لـــه بالرمز μ_s وهو عبارة عن متوسط كل احصاءات العينات التي يمكن اختيارهـــا مـــن المجتمع.

$$\mu_S = (S_1 + S_2 + \ldots\ldots\ldots\ldots\ldots + S_k)/k$$

وأيضا لكل توزيع معاينة لإحصاءة معينة انحراف معياري ويرمز له بالرمز σ_s .

وكما أشرنا يوجد نوعان من المعاينة: المعاينة بإرجاع وبدون إرجاع. فـــإذا ســـحبت مفردة من مجتمع لنا الخيار في إرجاع هذه المفردة أو عدم إرجاعها قبل إجراء الســـحبة التالية. في حالة المعاينة بإرجاع فان المفردة يمكن أن تظهر عدداً من المرات وأمـــا إذا تمت المعاينة دون إرجاع فان المفردة تظهر مره واحدة على الأكثر. ومن ناحية أخـــرى هناك نوعان من المجتمعات يسميان مجتمع محدود ومجتمع غير محدود وهناك ضرورة فنية لتحديد نوع المعاينة ونوع المجتمع لأنه يترتب على ذلك استخدام أو عدم اســـتخدام معامل يسمى بمعامل التصحيح كما سيأتي لاحقاً.

1.2 توزيع المعاينة للأوساط

Sampling Distribution of Means

أفرض أن الاحصاءة S تحت الدراسة هي الوسط الحسابي $\bar{X}$. و عليه ، نظريا، يمكن سحب كل العينات الممكنة ذات نفس الحجم n من مجتمع محدود حجمه N ثم نحسب الوسط الحسابي لكل عينة: $\bar{X}_1,\ \bar{X}_2,\ \bar{X}_3,\ldots..$

في هذا المثال يسمي $\bar{X}_1,\ \bar{X}_2,\ \bar{X}_3,\ldots..$ توزيع المعاينة للأوساط. وكما سبق ذكره فإن لكل توزيع معاينة وسط حسابي وفي هذه الحالة نرمز بالرمز $\mu_{\bar{X}}$ و هو يساوي الوسط الحسابي للمجتمع μ . وكذلك لكل توزيع معاينة انحراف معياري يسمي اصطلاحاً الخطأ المعياري، والخطأ المعياري لتوزيع المعاينة للأوساط يرمز له بالرمز $\sigma_{\bar{X}}$ وهو يساوي، $\frac{\sigma}{\sqrt{n}}$ ، حيث σ الانحراف المعياري للمجتمع و n حجم العينة. وتنطبق هذه القاعدة إذا كان المجتمع غير محدود أو إذا كان محدوداً والمعاينة بالإرجاع. أما إذا كان المجتمع محدوداً والمعاينة بدون إرجاع فينبغي استخدام معامل التصحيح، فيصبح الخطأ المعياري للمتوسط $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$ ، حيث N حجم المجتمع.

ملاحظة:

وننبه هنا إلي نقطتين هامتين:

الأولي: إذا كان المجتمع يتوزع توزيعاً طبيعياً فان توزيع المعاينة للأوساط يتبع التوزيع الطبيعي أيضا.

الثانية: إذا كان المجتمع لا يتوزع طبيعياً، فإن توزيع المعاينة للأوساط يتبع التوزيع الطبيعي بالتقريب شريطة أن يكون حجم العينة كبيراً $(n \geq 30)$ ، وكلما زاد حجم العينة كلما تحسن التقريب.

لنبرز هذه النتائج إليك في المثال التالي:

مثال

أفرض أن مجتمعا يتكون من أربعة ارقام هي 2،4،6،8 ويراد سحب عينة حجمها 2. فإن العينات المختلفة الممكنة كما في الجدول التالي:

جدول كل العينات الممكنة ذات الحجم 2 من المجتمع (8، 6، 4، 2):

رقم العينة	العينة	$\bar{X}$	$p\left(\bar{X}\right)$
1	(2،2)	2	1/16
2	(4،2)	3	1/16
3	(6،2)	4	1/16
4	(8،2)	5	1/16
5	(2،4)	3	1/16
6	(4،4)	4	1/16
7	(2،4)	5	1/16
8	(8،4)	6	1/16
9	(2،6)	4	1/16
10	(4،6)	5	1/16
11	(6،6)	6	1/16
12	(8،6)	7	1/16
13	(2،8)	5	1/16
14	(4،8)	6	1/16
15	(6،8)	7	1/16
16	(8،8)	8	1/16
المجموع	.....	**80**	**1**

و يلخص في الجدول التكراري التالي :

جدول التوزيع التكراري لمتوسط عينه حجمها 2:

$\bar{X}$	f	$p\left(\bar{X}\right)$	$\bar{X}\ p\left(\bar{X}\right)$	$\left(\bar{X}\right)^2\ p\left(\bar{X}\right)$
2	1	1/16	2/16	4/16
3	2	2/16	6/16	12/16
4	3	3/16	12/16	36/16
5	4	4/16	20/16	20/16
6	3	3/16	18/16	108/16
7	2	2/16	14/16	98/16
8	1	1/16	8/16	64/16
المجموع	16	1	80/16=5	440/16

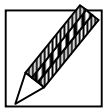
إذاً $\mu_{\bar{X}} = 80/16 = 5\ ;\ \sigma_{\bar{x}}^2 = 440/16 - 5^2 = \frac{5}{2}$

متوسط المجتمع $\mu = 5$

و تباينه $\sigma^2 = \frac{(2-5)^2+(4-5)^2+(6-5)^2+(8-5)^2}{4} = 5$

ونلاحظ أن $\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} = \frac{5}{2}$

تدريب (1)



أحصل على كل العينات الممكنة ذات الحجم 2 والتي يمكن سحبها من مجتمع يتكون من خمسة أرقام: 6و8و10و12و14. وبافتراض أن السحب تم بإرجاع من هذا المجتمع أوجد:

(أ) وسط المجتمع.

(ب) الانحراف المعياري.

(ج) وسط توزيع المعاينة للأوساط.

(د) الانحراف المعياري لتوزيع المعاينة للأوساط.

(هـ) ماذا يسمي المجتمع الذي تم سحب العينات منه.

◄◄ مثال

أفرض أن مجتمعا يتكون من أربعة ارقام هي 2،4،6،8 ويراد سحب عينة حجمها 2. بحيث تتم المعاينة بدون إرجاع فإن العينات المختلفة الممكنة تكون كما في الجدول التالي:

جدول كل العينات الممكنة بدون إرجاع

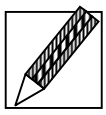
رقم	العينة	$\bar{X}$	$p\left(\bar{X}\right)$	$\bar{X}\ p\left(\bar{X}\right)$	$p\left(\bar{X}\right)$
1	{4,2}	3	1/6	3/6	9/6
2	{6,2}	4	1/6	4/6	16/6
3	{8,2}	5	1/6	5/6	25/6
4	{6,4}	5	1/6	5/6	25/6
5	{8,4}	6	1/6	6/6	36/6
6	{8,6}	7	1/6	7/6	49/6
مجموع		..	1	5	160/6

$$\mu_{\bar{X}} = 5 = \mu$$

$$\sigma_{\bar{x}}^2 = \frac{160}{6} - 5^2 = \frac{5}{3}$$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} = \sqrt{\frac{5}{2}}\sqrt{\frac{4-2}{4-1}} = \sqrt{\frac{5}{2}}.\sqrt{\frac{2}{3}} = \sqrt{\frac{5}{3}}$$

تدريب (2)



أحصل على كل العينات الممكنة ذات الحجم 2 والتي يمكن سحبها من مجتمع يتكون من خمسة أرقام: 6و8و10و12و14. وبافتراض أن السحب تم بدون إرجاع من هذا المجتمع أوجد:

(أ) وسط المجتمع.

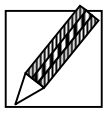
(ب) الانحراف المعياري.

(ج) وسط توزيع المعاينة للأوساط.

(د) الانحراف المعياري لتوزيع المعاينة للأوساط.

(هـ) ماذا يسمي المجتمع الذي تم سحب العينات منه.

تدريب (3)



مجتمع مكون من 15000 مفردة بوسط 120 وانحراف معياري 80. أوجد الوسط والخطأ المعياري لتوزيع المعاينة للوسط عندما يكون حجم العينة :

(أ) 144 (ب) 900.

أسئلة التقويم الذاتي (1)



(1) يتكون مجتمع من المفردات التالية: 11 و 8 و 6 و 3 و 2

(أ) أوجد وسط المجتمع μ وانحرافه المعياري σ

(ب) أحصل على كل العينات الممكنة ذات الحجم 2

(ج) أوجد الوسط الحسابي لكل عينة في الجزء (ب)

(د) أوجد الوسط الحسابي لتوزيع المعاينة للأوساط $\mu_{\bar{x}}$ والانحراف المعياري لتوزيع المعاينة للأوساط $\sigma_{\bar{x}}$

(هـ) تحقق من أن : $\mu_{\bar{x}} = \mu$ و $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$

(2) (أ) ماذا يعني توزيع المعاينة للوسط وكيف يتم الحصول عليه.

(ب) ماذا يعني التوقع والخطأ المعياري لتوزيع المعاينة للوسط.

(3) ما هو شكل توزيع المعاينة النظري للوسط إذا كان المجتمع المسحوبة منه كل العينات الممكنة يتبع نمط التوزيع الطبيعي؟ وما هو شكله إذا كان المجتمع لا يتبع نمط التوزيع الطبيعي.

2.2 توزيع المعاينة للنسب
Sampling Distribution of Proportions

افرض أن لدينا مجتمعا إثنينيا حجمه N ، بمعني أنه من نوعين فقط من الوحدات يرمز لهما مجازاً بنجاح (S) وفشل (F)، حيث احتمال النجاح فى كل مرة يساوي π ، وأن الاحصاءة التي تهمنا هي نسبة وحدات النجاح p في عينة حجمها n فالننظر الى نسب النجاح في كل العينات الممكنة ذات الحجم n: $\{p_1 \ p_2 \ p_3 \ , p_4 .....\}$ تسمي المجموعة توزيع المعاينة للنسبة p. و كما ذكرنا سابقا فإن لهذا التوزيع متوسط وخطأ معياري يرمز لهما بالرمزين μ_p و

$$\sigma_p = \sqrt{\frac{\pi[1-\pi]}{n}}$$

عادة يستخدم الرمز $(1-p)=q$ لنسبة الفشل.

مثال توضيحي

كررت تجربة أربع مرات علي مجتمع إثنيني باحتمال نجاح 0.5 . فإن تكوين العينات المختلفة بالنسبة للناتج كما يلي:

جدول يوضح تكوين العينات المختلفة ونسب النجاح:

رقم العينة	العينة	عدد النجاحات X	نسبة النجاح p
1	SSSS	4	4/4
2	SSSF	3	3/4
3	SSFS	3	3/4
4	SFSS	3	3/4
5	FSSS	3	3/4
6	SSFF	2	2/4
7	SFSF	2	2/4
8	FSSF	2	2/4
9	SFFS	2	2/4
10	FSFS	2	2/4
11	FFSS	2	2/4
12	SFFF	1	1/4
13	FSFF	1	1/4
14	FFSF	1	1/4
15	FFFS	1	1/4
16	FFFF	0	0/4

والجدول التكراري لهذا التوزيع موضح في الجدول ادناه:

الجدول التكراري للتوزيع:

p	f	Pr(p)	p. Pr(p)	P^2 Pr(p)
4/4	1	1/16	4/16	16/256
3/4	4	4/16	12/16	36/256
2/4	6	6/16	12/16	24/256
1/4	4	4/16	4/16	4/256
0/4	1	1/16	0/16	0/256
مجوع	16	1	32/16/=1/2	80/256

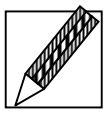
إذاً: $\mu_p = 0.5 = \pi$

$$\sigma_p^{\ 2} = 80/256 - 0.5^2 = \frac{1}{16}$$

لاحظ أن $$\frac{\pi(1-\pi)}{n} = \frac{\frac{1}{2}\left(1-\frac{1}{2}\right)}{4} = \frac{1}{16} = \sigma_p^{\ 2}$$

وينبغي أن نشير إلي أن توزيع المعاينة للنسب يتبع نمط التوزيع الطبيعي فقط في حالة العينات الكبيرة $(n \geq 30)$ ؛ وأما إذا كان حجم العينة صغير فإن النسبة لا تتبع نمط التوزيع الطبيعي لأن المجتمع لا يتبع التوزيع الطبيعي بل توزيع ذات الحدين.

تدريب (4)



فى عينة عشوائية حجمها 144 عامل من مصنع للسكر حيث يبلغ عدد العمال الكلي بالمصنع 1500، وجد أن 60 يصابون بالملاريا كل عام. أوجد وسط توزيع المعاينة لنسب الذين سيصابون بالملاريا كل عام μ_p و الانحراف المعياري σ_p .

3.2 توزيع المعاينة للفرق بين الأوساط

Sampling Distribution of Differences of Means

ونفترض هنا وجود مجتمعيين W_2 , W_1 وقد تم سحب العينات الممكنة من المجتمع الأول ذات الحجم n_1 وسحب العينات الممكنة من المجتمع الثاني ذات الحجم n_2 وليس بالضرورة أن يتساوى n_2 , n_1 كما يوضح الجدول (6) التالي :

جدول العينات المسحوبة من المجتمعين الأول والثاني:

العينات المسحوبة من المجتمع الأول	العينات المسحوبة من المجتمع الثاني
W_{11}	W_{21}
W_{12}	W_{22}
W_{13}	W_{23}
.	.
.	.
.	.
W_{1k}	W_{2r}

حيث يشير الدليل الأول من الشمال رقم المجتمع والدليل الثاني رقم العينة للمجتمع المعنى. ولنفترض أننا حسبنا الوسط الحسابي لكل العينات المسحوبة من المجتمع الأول وكل العينات المسحوبة من المجتمع الثاني كما يلي:

جدول الأوساط الحسابية للعينات المسحوبة من المجتمعين الأول والثاني:

الوسط الحسابي للعينات المسحوبة من المجتمع الأول	السوط الحسابي للعينات المسحوبة من المجتمع الثاني
$\bar{x}_{11}$	$\bar{x}_{21}$
$\bar{x}_{12}$	$\bar{x}_{22}$
.	.
$\bar{x}_{1k}$	$\bar{x}_{2k}$

ولنفترض إننا حصلنا على كل التوافيق (Combinations) الممكنة بين الأوساط الحسابية المحسوبة لعينات المجتمعين ثم حصلنا على الفروق بين الأوساط الحسابية لهذه التوافيق وتسمى مجموعة هذه الفروق بين الأوساط توزيع المعاينة للأوساط الحسابية. ومثلها مثل أية توزيع معاينة لها وسط حسابي وخطأ معياري.

فالوسط الحسابي لتوزيع المعاينة للفروق بين الأوساط يرمز له بالرمز $u_{\bar{x}_1-\bar{x}_2}$

ويساوى الفرق بين وسطي المجتمعين:

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

حيث تمثل μ_1 الوسط الحسابي للمجتمع الأول و μ_2 الوسط الحسابي للمجتمع الثاني.

وأما الخطأ المعياري لتوزيع المعاينة للفروق يرمز له بالرمز $\sigma_{\bar{x}_1-\bar{x}_2}$

ويمكن حسابه باستخدام القاعدة التالية:

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

حيث أن:

σ_1^2 تمثل تباين المجتمع الأول.

σ_2^2 تمثل تباين المجتمع الثاني.

n_1 تمثل حجم العينة المسحوبة من المجتمع الأول.

n_2 تمثل حجم العينة المسحوبة من المجتمع الثاني.

باعتبار أن المجتمعين مستقلان

◄◄ مثال

عينة مكونة من أطوال قامة 6,400 سوداني لها وسط حسابي يساوى 67.85 بوصة وانحراف معياري 2.56 بوصة وعينة أخرى لأطوال قامة 1,600 مصري لها وسط حسابي 68.55 بوصة وانحراف معياري 2.52 بوصة. احصل على الانحراف المعياري لتوزيع المعاينة للفرق بين متوسطي أطوال قامة السودانيين والمصريين.

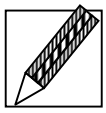
الحل:

$$\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} = \sqrt{\frac{(2.56)^2}{6400}+\left(\frac{2.52}{1600}\right)^2}$$

$$= \sqrt{0.001+0.004} = 0.0707$$

وتم استخدام تباين العينتين كتقدير لتباين المجتمعين.

تدريب (5)



أخذت عينة حجمها 36 من توزيع وسطه 62 وتباينه 16 وأخذت عينة أخري حجمها 64 من توزيع أخر مستقل عن التوزيع الأول ووسطه 80 وتباينه 30. فإذا رمزنا لأوساط مثل هذه العينات بالرمزين $\bar{X}_1$ و $\bar{X}_2$. أوجد $\mu_{\bar{X}_1-\bar{X}_2}$.

4.2 توزيع المعاينة للفروق بين النسب

Sampling Distribution of Differences of Proportions

نفترض وجود مجتمعين W_1 و W_2 وقد تم سحب العينات الممكنة من المجتمع الأول وجميعها ذات حجم متساوي n_1 وسحب العينات الممكنة من المجتمع الثاني وجميعها ذات حجم متساوي n_2 ولكن ليس بالضرورة أن يتساوىn_1 مع n_2 كما يوضح الجدول التالي:

جدول للعينات المسحوبة من المجتمعين الأول والثاني:

العينات المسحوبة من المجتمع الأول	العينات المسحوبة من المجتمع الثاني
w_{11}	w_{21}
w_{12}	w_{22}
w_{13}	w_{23}
.	.
.	.
.	.
w_{1k}	w_{2r}

حيث يشير الدليل الأول لرقم المجتمع والدليل الثاني لرقم العينة للمجتمع المعنى.

ولنفترض أننا حسبنا النسبة لكل العينات المسحوبة من المجتمع الأول وكل العينات المسحوبة من المجتمع الثاني كما **يلي:**

جدول نسب العينات المسحوبة من المجتمعين الأول والثاني:

النسب للعينات المسحوبة من المجتمع الأول	النسب للعينات المسحوبة من المجتمع الثاني
p_{11}	p_{21}
p_{12}	p_{22}
p_{13}	p_{23}
.	.
.	.
.	.
p_{1k}	p_{2r}

ولنفترض أننا حصلنا على كل التوافيق الممكنة بين النسب المحسوبة لعينات المجتمعين، ثم حصلنا على الفروق بين النسب لهذه التوافيق، وتسمى هذه الفروق بين النسب توزيع المعاينة للفروق بين النسب. ووسط توزيع المعاينة للفروق بين النسب يرمز له بالرمز $\mu_{P_1-p_2}$ ويساوى الفرق بين نسبتي المجتمعين أي:

$$\mu_{P_1-P_2} = \mu_{\pi_1} - \mu_{\pi_2} = \pi_1 - \pi_2$$

حيث أن

π_1 : تمثل النسبة في المجتمع الأول

π_2 : تمثل النسبة في المجتمع الثاني

وحتى نستطيع أن نميز بين نسبة العينة ونسبة المجتمع فيرمز لنسبة المجتمع بالحرف التاجي P .

وأما الخطأ المعياري لتوزيع المعاينة للفروق بين النسب فيرمز له بالرمز $\sigma_{P_1-P_2}$ ويساوى

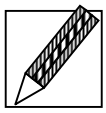
$$\sigma_{P_1-P_2} = \sqrt{\frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}}$$

حيث أن

n_1 و n_2 تمثلان حجمي العينتين المسحوبتين من المجتمع الأول والمجتمع الثاني على التوالي.

وبما أن نسبة النجاح π ونسبة الفشل $(1-\pi)$ مجهولتان للمجتمع يمكن تقديرهما باستخدام نسبة نجاح ونسبة فشل العينة.

تدريب (6)



سحبت عينة عشوائية من سكان مدينة الخرطوم مكونة من 1,000 شخص وقد وجد أن 400 منهم يستهلكون القمح. وسحبت عينة عشوائية أخرى من سكان مدينة مدني مكونة من 800 شخص وقد وجد أن 400 منهم يستهلكون القمح. قدر الخطأ المعياري لتوزيع المعاينة للفرق بين نسبة من يستهلكون القمح في مدينتي الخرطوم ومدني.

أسئلة التقويم الذاتي (2)



1- إذا كانت أجور العمال في المصنع (أ) تتبع نمط التوزيع الطبيعي بوسط 6 دينار يومياً وتباين 0.55 وكانت أجور العمال في المصنع (ب) تتبع نمط التوزيع الطبيعي بوسط 5 دينار يومياً وتباين 0.30 وأخذت عينة حجمها 25 عاملاً من (أ) و 16 عاملاً من (ب). أحصل على وسط توزيع المعاينة للفروق بين الأوساط والانحراف المعياري.

2- إذا كانت نسبة الطالبات بجامعة الخرطوم 0.5 ونسبة الطالبات بجامعة الجزيرة 0.7 وأخذت عينة من طلاب جامعة الخرطوم حجمها 121 وعينة من طلاب جامعة الجزيرة حجمها 196. إذا إفترضنا أن p_1 تمثل نسبة الإناث في جامعة الخرطوم و p_2 تمثل نسبة الإناث في جامعة الجزيرة. أحصل:

[أ] على وسط توزيع المعاينة للفروق بين النسب.

[ب] على الانحراف المعياري لتوزيع المعاينة للفروق بين النسب.

الخلاصة

نأمل عزيزي الدارس بعد انتهائك من دراسة هذه الوحدة أن تكون قد ميــزت بين المعاينة بإرجاع والمعاينة بدون إرجاع وقد أصبحت قادراً علــى أن تحصــل على كل العينات الممكنة لمجتمع محدود عدد مفرداته 5 أو 6 مثلاً

وكذلك أن تكون قد عرفت الفرق بين المعــالم Parameters والإحصــائيات Statistics وأن تكون لديك المقدرة على أن تحصــل علــى الوســط الحســابي والانحراف المعياري لكل من توزيع المعاينة للأوساط و توزيع المعاينــة للنســب توزيع المعاينة للفروق بين الأوساط و توزيع المعاينة للفروق بين النسب .

لمحة مسبقة عن الوحدة التالية

عزيزي الدارس ،،

في الوحدة القادمة سندرس مبادئ نظرية التقدير وتطبيقاتها حيــث ســنتناول مــن خلالها موضوع تقدير المعالم وخواص المقدر الجيد وطرق التقدير لبعض المعالم .

إجابات التدريبات

تدريب (1)

رقم العينة	العينة	$\bar{x}$
1	**(6،6)**	6
2	**(8،6)**	7
3	**(10،6)**	8
4	**(12،6)**	9
5	**(14،6)**	10
6	**(6،8)**	7
7	**(8،8)**	8
8	**(10،8)**	9
9	**(12،8)**	10
10	**(14،8)**	11
11	**(6،10)**	8
12	**(8،10)**	9
13	**(10،10)**	10
14	**(12،10)**	11
15	**(14،10)**	12
16	**(6،12)**	9
17	**(8،12)**	10
18	**(10،12)**	11
19	**(12،12)**	12
20	**(12،14)**	13
21	**(6،14)**	10
22	**(14،8)**	11
23	**(14،10)**	12
24	**(14،12)**	13
25	**(14،14)**	14
المجموع		250

جدول يوضح طريقة الحل الطويلة

$\bar{x}$	$p(\bar{x})$	$\bar{x}.p(\bar{x})$	$(\bar{x}-\mu_{\bar{x}})$	$(\bar{x}-\mu_{\bar{x}})^2$	$(\bar{x}-\mu_{\bar{x}})^2.p(\bar{x})$
6	$1/25$	$6/25$	−4	16	$16/25$
7	$2/25$	$14/25$	−3	9	$18/25$
8	$3/25$	$24/25$	−2	4	$12/25$
9	$4/25$	$36/25$	−1	1	$4/25$
10	$5/25$	$50/25$	0	0	0
11	$4/25$	$44/25$	1	1	$4/25$
12	$3/25$	$36/25$	2	4	$12/25$
13	$2/25$	$26/25$	3	9	$18/25$
14	$1/25$	$14/25$	4	16	$16/25$
1	$\mu_{\bar{x}} = \frac{250}{25} = 10$	—	—	$\sigma_{\bar{x}}^2 = 100/25 = 4$	

نلاحظ أن هذه الطريقة طويلة لحساب $\sigma_{\bar{x}}^2$ ولكن يمكن حلها ببساطة باستخدام القاعدة

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} = \frac{8}{2} = 4$$

(أ) وسط المجتمع

$$M = \frac{6+8+10+12+14}{5} = \frac{50}{5} = 10$$

(ب) الانحراف المعياري

$$\sigma^2 = \frac{(6-10)^2 + (8-10)^2 + (10-10)^2 + (12-10)^2 + (14-10)^2}{5} = \frac{40}{5} = 8$$

(ج) وسط توزيع المعاينة للأوساط

$$\mu_{\bar{x}} = \frac{250}{25} = 10$$

(د) الانحراف المعياري لتوزيع المعاينة للأوساط

$$\sigma_{\bar{x}}{}^2 = \frac{\sigma^2}{n} = \frac{8}{2} = 4$$

$$\sigma_{\bar{x}} = 2$$

(هـــ) يسمى مجتمع محدود

تدريب (2)

رقم العينة	العينة	$\bar{x}$
1	{8،6}	7
2	{10،6}	8
3	{12،6}	9
4	{14،6}	10
5	{10،8}	9
6	{12،8}	10
7	{14،8}	11
8	{12،10}	11
9	{14،10}	12
10	{12،14}	13
المجموع		

ويمكن تكوين جدول التوزيع الاحتمالي كما يلي :

$\bar{x}$	$p(\bar{x})$
7	$1/10$
8	$1/10$
9	$2/10$
10	$2/10$
11	$2/10$
12	$1/10$
13	$1/10$

(أ) وسط المجتمع

$$M = \frac{50}{5} = 10$$

(ب) الانحراف المعياري

$$\sigma^2 = \frac{40}{5} = 8$$

(ج)وسط توزيع المعاينة للأوساط

$$\mu_{\bar{x}} = \frac{100}{10} = 10$$

(د) الانحراف المعياري لتوزيع المعاينة للأوساط

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} = \sqrt{\frac{8}{2}\left(\frac{5-2}{5-1}\right)} = \sqrt{3}$$

(هـ) يسمى مجتمع محدود

تدريب (3)

(أ) وسط توزيع المعاينة للأوساط $\mu_{\bar{x}} = \mu = 120$

الخطأ المعياري لتوزيع المعاينة $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{80}{\sqrt{144}} = \frac{80}{12} = 6.7$

(ب)وسط توزيع المعاينة للأوساط $\mu_{\bar{x}} = \mu = 120$

الخطأ المعياري لتوزيع المعاينة $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{80}{\sqrt{900}} = \frac{80}{30} = 2.7$

تدريب (4)

وسط توزيع المعاينة للنسب $\mu_{\mathrm{p}} = \frac{60}{144} = \frac{5}{12} = 0.42$

مربـــــــــــع الانحـــــــــــراف المعيـــــــــــاري (التبـــــــــــاين)

$$\sigma_{\mathrm{p}}{}^{2} = \frac{\pi(1-\pi)}{n} = \frac{\frac{5}{12}\left(1-\frac{5}{12}\right)}{144} = \frac{35}{144^{2}}$$

الانحراف المعياري $\sigma_{\mathrm{p}} = \frac{\sqrt{35}}{144}$

تدريب (5)

$$\mu_{\bar{\mathrm{x}}_1 - \bar{\mathrm{x}}_2} = \mu_1 - \mu_2 = 62 - 80 - 18$$

تم استخدام وسطي العينتين كتقدير لوسطي المجتمعين

تدريب (6)

$$\hat{\sigma}_{\pi_1 - \pi_2} = \sqrt{\frac{p_1 q_2}{n_1} + \frac{p_2 q_2}{n_2}}$$

$n_1 = 1000$, $p_1 = \frac{400}{1000} = 0.4$, $n_2 = 800$, $p_2 = \frac{400}{800} = 0.5$

$q_1 = 1 - p_1 = 1 - 0.4 = 0.6$, $q_2 = 1 - p_2 = 1 - 0.5 = 0.5$

$$\hat{\sigma}_{\pi_1 - \pi_2} = \sqrt{\frac{(0.4)(0.6)}{1000} + \frac{(0.5)(0.5)}{800}}$$

$$= \sqrt{0.00024 + 0.0003125}$$

$$= \sqrt{0.0005525} = 0.0235$$

إجابات بعض أسئلة التقويم الذاتي

تقويم ذاتي (1)

(أ) وسط المجتمع

$$\frac{42}{6} = 7$$

(ب) الانحراف المعياري

$$\sigma = \sqrt{\frac{84}{6}} = \sqrt{14}$$

(ج) الوسط الحسابي لتوزيع المعاينة للأوساط

$$\mu_{\bar{x}} = \mu = 7$$

(د) الانحراف المعياري لتوزيع المعاينة للأوساط

$$\sigma_{\bar{x}} = \sqrt{\frac{28}{5}}$$

مسرد المصطلحات

مجتمع (Population) : المجموعة الكلية للمفردات أو المشاهدات تحت الدراسة.

مجتمع محدود (Finite Population) : مجتمع يتكون من عدد محدود من المفردات أو المشاهدات.

مجتمع غير محدود (Infinite Population) : مجتمع يتكون من عدد غير محدود من المفردات أو المشاهدات.

عينة (Sample) : مجموعة جزئية من المجتمع وتتكون من عدد محدود من المفردات أو المشاهدات.

عينة عشوائية (Random Sample) : عندما تكون لكل مفردة أو مشاهدة في المجتمع نفس الفرصة أو الاحتمال في الاختيار أو بصورة أخري أن يكون لكل عينة مسحوبة من المجتمع نفس الفرصة أو الاحتمال في الاختيار.

معاينة (Sampling) : عملية دراسة مجتمع على أساس عينة مسحوبة من هذا المجتمع.

معاينة بإرجاع (Sampling with replacement) : يعني ارجاع المفردة المسحوبة مرة أخري للمجتمع قبل أن تجري عملية السحب التالية.

معاينة دون إرجاع (Sampling without replacement) : لا تعاد المفردة المختاره في العينة مرة أخري للمجتمع.

المراجع

- مواري د. شبيجل، "الإحصاء" ، سلسلة ملخصات شوم، الدار الدولية للنشر والتوزيع 1991.

- بول ج.هويل، " المبادئ الأولية في الإحصاء، ترجمة ومراجعة د. بدرية شوقي عبد الوهاب والدكتور محمد كامل الشربيني ، دار جون وايلي وأبنائه.


الوحدة الثانية
مبادئ نظرية التقدير

محتويات الوحدة

الموضوع	الصفحة
المقدمة	41
تمهيد	41
أهداف الوحدة	42
1. نظرية التقدير	43
1.1 تقدير المعالم	43
2.1 خواص المقدر الجيد	44
3.1 التقدير	45
2. تقدير فترة ثقة لمعالم المجتمع	47
1.2 تقدير فترة الثقة للأوساط	48
2.2 تقدير فترة ثقة للنسب	51
3.2 تقدير فترة الثقة للفروق بين المتوسطات	53
4.2 تقدير فترة الثقة للفروق بين النسب	55
الخلاصة	57
لمحة مسبقة عن الوحدة الدراسية التالية	57
إجابات التدريبات	58
إجابات أسئلة التقويم الذاتي	60
مسرد المصطلحات	61
المراجع	62

المقدمة

تمهيد

هذه الوحدة عزيزي الدارس جاءت بعنوان مبادئ نظريـــة التقـــدير حيـــث أن التقدير أحد فرعى الاستدلال الإحصائي، إذ أن الفرع الثاني هـــو إختبـــارات الفـــروض والتي سنتاولها في الوحدة القادمة بإذن الله .

في بداية الوحدة ذكرنا الفـــرق بـــين المقـــاييس الإحصـــائية (Parameters) و الإحصائيات(Statistics) ثم شرحنا كيف يتم تقدير المعالم وما هي خـــواص المقـــدر الجيد وكيفية التقدير بنقطة أو بفترة و وضحنا تقدير فترات الثقة لكـــل مـــن (الوســـط الحسابي ، النسب ، الفروق بين الأوساط الحسابية وأخيراً الفروق بين النسب) . ولـــذا لكي تفهم هذه المواضيع فهماً جيداً ينبغي أن تقوم بحل جميع التدريبات وأســـئلة التقـــويم الذاتي ونحن من جانبنا نرحب بآرائك ومقترحاتك .

أهداف الوحدة



عزيزي الدارس،،

يتوقع منك بعد الانتهاء من دراسة هذه الوحدة أن تكون قادراً على أن:

1- **تشرح** خواص المقدر الجيد.

2- **تعرف** التقدير بنقطة والتقدير بفترة.

3- **تقدر** فترة الثقة للوسط الحسابي.

4- **تقدر** فترة الثقة للنسب.

5- **تقدر** فترة الثقة للفروق بين المتوسطات .

6- **تقدر** فترة الثقة للفروق بين النسب.

1. نظرية التقدير

في بداية هذه الوحدة عزيزي الدارس نود أن نوضح لك معنى كل من :

أ- مصطلح معالم (Parameters)

ب- مصطلح إحصائيات (Statistics)

حيث أن المصلح الأول (معالم) يعني المقاييس الاحصائية المحسوبة على اساس كل مفردات المجتمع .

أما المصطلح الثاني (إحصائيات) فهو يعني المقاييس الاحصائية المحسوبة على أساس مفردات العينة .

1.1 تقدير المعالم (Estimate of Parameters)

تهتم نظرية التقدير باستخدام إحصاءة (Statistic) محسوبة على أساس عينة لتقدير معلم (Parameter) المجتمع المقابل له ويعتبر المعلم ثابت مجهول . ولتباين مفردات العينات المسحوبة من المجتمع قد يكون التقدير المحسوب على أساسها صحيح أو غير صحيح ولا يمكن الجزم بصحة أو عدم صحة التقدير. وبصورة عامة يمكن أن نقول أن التقدير عبارة عن تقريب للمعلم المقابل له. وتتعلق دقة التقدير باعتبارات احتمالية. وهذه الاعتبارات الاحتمالية هي التي تكسب تقويم التقدير بعض من الموضوعية والمصداقية.

2.1 خواص المقدر الجيد Properties of Good Estimators

يمكن تقدير معالم المجتمع بطرق مختلفة ولكن الطريقة الجيدة للتقدير تتمتع بالخواص التالية:

(أولاً) أن يكون المقدر غير متحيز (Unbiased) بمعنى أن متوسط توزيع المعاينة للإحصاءة المعينة ينبغي أن يكون مساو لمعلم المجتمع المقابل له. فمثلاً عند دراساتنا لتوزيع المعاينة للأوساط، وجدنا متوسط توزيع المعاينة للأوساط يساوى وسط المجتمع.

$$u_{\bar{x}} = E\left(\bar{x}\right) = u$$

وفى هذه الحالة فان الوسط الحسابي $\bar{x}$ كمقدر لوسط المجتمع يتمتع بخاصية عدم التحيز. إذا كان متوسط توزيع المعاينة للاحصاءة لا يساوي قيمة المعلم في المجتمع فيسمي الفرق بينهما تحيزاً (Bias)

(ثانياً) خاصية اقل تباين (minimum variance) ونعنى بذلك إذا كان لدينا مقدران مختلفان لمعلم معين وكلاهما يتمتع بخاصية عدم التحيز فإذا كان أحد المقدرين له تباين اقل مقارنة بالآخر. فان المقدر ذو التباين الأقل يتمتع بهذه الخاصية.

فمثلاً، إذا كان لدينا معلم θ ويمكن تقديره بطريقتين مختلفتين $\hat{\theta}_1$ و $\hat{\theta}_2$ فإذا كان $E\left(\hat{\theta}_1\right) = E\left(\hat{\theta}_2\right) = \theta$ وتحقق التالي:

$$\text{var}\left(\hat{\theta}_1\right) < \text{var}\left(\hat{\theta}_2\right)$$

حيث أن $\text{var}\left(\hat{\theta}_1\right)$ يعنى تباين المقدر $\hat{\theta}_1$ و $\text{var}\left(\hat{\theta}_2\right)$ يعنى تباين المقدر $\hat{\theta}_2$ ، فان $\hat{\theta}_1$ يتمتع بخاصية اقل تباين.

(ثالثاً) الاتساق Consistency : أفرض أن $\hat{\theta}$ مقدر غير متحيز لمعلم مجتمع θ حسب من عينة حجمها n. فإذا كانت قيمة $\hat{\theta}$ تقترب من θ كلما زادت قيمة n فيقال أن المقدر $\hat{\theta}$ متسق Consistent

3.1 التقدير Estimation

يتم تقدير المعالم بأحد الطريقتين :

أ- التقدير بنقطة (Point estimation)

يعنى التقدير بنقطة (Point estimation) استخدام قيمة واحدة كتقدير لمعلم المجتمع. فمثلاً إذا كان الوسط الحسابي لعينة من أوزان طلاب فصل دراسي في مدرسة 52 كيلو جراما واستخدم كتقدير لوسط أوزان طلاب المدرسة، فان هذا التقدير يسمى تقديراً بنقطة. التقدير قد يكون قريب جداً لقيمة معلم المجتمع وهو وسط المجتمع في هذه الحالة ولكننا نكاد نجزم بأنه ليس مساو للقيمة الحقيقية. وليست هناك إمكانية

لتقويم درجة عدم الدقة لان الاحتمال المقترن بقيمة وحيدة في التوزيع المتصل (Continuos distributions) يساوى الصفر.

ب- التقدير بفترة (Interval estimation)

وهو أن يتم التقدير بتكوين مدى حول التقدير بنقطة. وكما في المثال السابق يمكن ان نقدر أن وسط وزن التلاميذ يقع بين 50 كجم و 55 كجم حيث نلاحظ أننا في هذا التقدير جعلنا هناك مدى أو فترة طولها 5 وحدات. والتقدير بفترة قد يكون صحيح أو غير صحيح. ولكن تتكون لدينا درجة من " الثقة" بان القيمة الحقيقة للمعلمة يشملها هذا المدى. ويمكن قياس درجة الثقة المرتبطة بطريقة تقدير معينة باستخدام لغة الاحتمالات. ومن محاسن التقدير بفترة هو الإشارة إلى مستوي الثقة (Confidence level) في دقة التقدير الأمر الذي يدعو لتفضيله على التقدير بنقطة.

أسئلة التقويم الذاتي (1)



ماذا يعني كل من :

(أ) التقدير بنقطة

(ب) التقدير بفترة

(ج) مقدر غير متحيز.

2. تقدير فترة ثقة لمعالم المجتمع

أفرض أن لدينا مجتمعاً W يعتمد في توزيعه على معلم θ وأفرض أن $t(\theta, X = t)$ دالة تعتمد علي المعلم وعينة X حجمها n من هذا المجتمع من توزيع معاينة العينة يمكن أيجاد قيمتين a و b $(a < b)$ بحيث

$$p(a < t(\theta_1 \underline{X}) < b) = 1 - \alpha \quad (1)$$

حيث α ثابت صغير محدد سلفاً.

و إذا أمكن إيجاد المتباينة $A(\underline{X}) < \theta\ B(\underline{X})$ من المتباينة $a < t(\theta_1 \underline{X}) < b$ فإننا نسمي الفترة $[A(\underline{X}), B(\underline{X})]$ فترة ثقة $(1-\alpha)$ 100% للمعلم θ ، ويسمي الاحتمال $(1-\alpha)$ معامل أو مستوي الثقة Confidence Coefficient .

◄◄ مثال

أفرض أن عينة عشوائية حجمها n من مجتمع طبيعي بمتوسط μ وتباين معلوم σ^2 أعطت متوسطاً $\bar{X}$. لقد علمنا أن $\bar{X}$ تتوزع طبيعياً بمتوسط μ وتباين $\frac{\sigma^2}{n}$.

إذا $Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$ يتوزع طبيعياً بمتوسط صفر وتباين الواحد. من جداول التوزيع

الطبيعي نجد أن $p\left(-1.96 < \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} < 1.96\right) = 1 - 0.05 = 0.95$

من المتباينه داخل القوس نستنتج:

$$\left(\bar{X} - 1.96\ \sigma / \sqrt{n}\right) < \mu < \left(\bar{X} + 1.96\ \sigma / \sqrt{n}\right)$$

إذاً نسمي الفترة $\bar{X} \pm 1.96\ \sigma / \sqrt{n}$

فترة ثقة $\%100(1 - 0.05) = \%95$ للمتوسط μ

1.2 تقدير فترة الثقة للأوساط

لنفترض الآن أن المعلم المراد تقديره هو متوسط مجتمع μ وأن الإحصاءة تحت الدراسة هى الوسط الحسابي $\bar{X}$ للعينة وبما أن الخطأ المعياري للوسط الحسابي هو $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ ، فبصفة عامة فان تقدير فترة ثقة لوسط المجتمع، يمكن الحصول عليها باستخدام القاعدة **التالية:**

$$\bar{X} \pm Z_C \sigma_{\bar{X}}$$

أو

$$\bar{X} \pm Z_C \frac{\sigma}{\sqrt{n}}$$

حيث Z_C يعتمد على مستوى الثقة المطلوب. فمثلاً لمستوي الثقة 95% لتقدير الوسط الحسابي للمجتمع الطبيعي $\mu_{\bar{X}}$ هي 1.96 و ولـــ99 % هي 2.58. وهذا ينطبق في حالة المعاينة لمتوسط مجتمع غير محدود أو مجتمع محدود ولكن المعاينة بإرجاع. وأما إذا كان المجتمع محدوداً والمعاينة دون إرجاع فإننا نستخدم عامل تصحيح المجتمع المحدود (Finite Population Correction Factor) $(N-n)/(N-1)$

وتصبح القاعدة بصفة عامة **كما يلي:**

$$\bar{X} \pm Z_C \frac{\sigma}{\sqrt{n}} \sqrt{\left(\frac{N-n}{N-1}\right)}$$

حيث أن:

N يمثل حجم المجتمع.

n يمثل حجم العينة المسحوبة من هذا المجتمع.

الانحراف المعياري للمجتمع σ مجهول ولاستخدام القاعدة ينبغي الحصول على تقدير له. وهذا التقدير يكون مرضى إذا كان حجم العينة المستخدمة يساوى 30 أو أكثر $(n \geq 30)$.

◄◄ مثال

والآن لنفترض أن وزارة السياحة في السودان ترغب في تقدير متوسط النفقات للسياح في الزيارة الواحدة. يمكن سحب عينة عشوائية من 100 سائح وتقدير فترة ثقة باستخدام الطريقة التالية:

- الحصول على الوسط الحسابي $\bar{X}$ ونفترض انه 800 دينار.
- الحصول على الخطأ المعياري ويمكن استخدام القاعدة $\sigma/\sqrt{n}$ وبافتراض أن الانحراف المعياري للمجتمع هو 120 دينار فإن $\sigma_{\bar{X}} = 120/\sqrt{100} = 12$.
- اختيار مستوى الثقة: فإذا كانت وزارة السياحة تعتقد أن فترة ثقة بنسبة 95% كافية، فإن $Z_C = 1.96$.
- أحسب حدود الثقة:

حد الثقة الأعلى هو:

$$\bar{X} + Z_C\sigma_{\bar{X}} = 800 + 1.96\ (12) = 823.52 \text{ دينار}$$

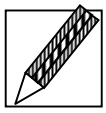
وحد الثقة الأدنى هو:

$$\bar{X} + Z_C\sigma_{\bar{X}} = 800 - 1.96\ (12) = 776.48 \text{ دينار}$$

- بيان فترة الثقة:

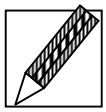
بالنسبة للحالة تحت الدراسة يمكن إعطاء البيان التالي "وسط نفقات السياح بمعدل زيارة واحدة للسودان تقدر ما بين 776.48 دينار و 823.52 دينار بمستوى ثقة 95%".

تدريب (1)



أخذت عينة مكونة من 9 مفردات وحسب الوسط لها فكان $\bar{X} = 20$ فإذا فرضنا أن هذه العينة أخذت من توزيع طبيعي $N(\mu\ ,\ 45)$. أوجد فترة 95% ثقة للوسط μ.

تدريب (2)



أوجد حجم العينة n التي يجب أخذها من توزيع طبيعي $N(\mu,36)$ للحصول على فترة 90% ثقة للوسط وبطول 5.

2.2 تقدير فترة ثقة للنسب

لنفرض أن الإحصاءة تحت الدراسة هي النسبة p (Proportion) . وبما أن الخطأ المعياري للنسبة هو $\sigma_P = \sqrt{\frac{pq}{n}}$ فان تقدير فترة ثقة لنسبة المجتمع هي :

$$P \pm Z_C\ \sigma_P$$

أو

$$P \pm Z_C \sqrt{\frac{Pq}{n}}$$

حيث Z_C يعتمد على فترة الثقة المطلوبة كما سبق ذكره في حالة فترة الثقة لوسط المجتمع.

◄◄ مثال

في استطلاع للرأي العام حول مرشح لدائرة انتخابية سحبت عينة عشوائية حجمها 100 من جميع الناخبين حيث دلت على أن 55 في المائة منهم سيصوتون لصالح هذا المرشح. أوجد حدود ثقة (أ) 95 في المائة و(ب) 99 في المائة لنسبة المساندين لهذا المرشح من بين جميع الناخبين.

الحل:

(أ) حدود ثقة 95% للنسبة P للمجتمع **هي:**

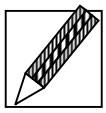
$$P \pm Z_C \sigma_P = P \pm 1.96 \sqrt{\frac{Pq}{n}} = 0.55 \pm 1.96 \sqrt{\frac{(0.55)(0.45)}{100}} = 0.55 \pm 0.1$$

حيث استخدمت نسبة العينة p لتقدير نسبة المجتمع π واستخدام التقريب الطبيعي لتوزيع p

وبالمثل (] حدود ثقة 99% للنسبة π هي:

$$0.55 \pm 2.58 \sqrt{\frac{(0.55)(0.45)}{100}} = 0.55 \pm 0.13$$

تدريب (3)



عينة عشوائية مكونة من 500 مواطن بقرية بولاية الجزيرة، يرغب 240 منهم أن يضاف إلى صهريج مياههم كمية بسيطة من الفلــور. استخدم هذه البيانات لإيجاد 95% فترة ثقة لنسبة المجتمــع الراغــب في إضافة الفلور.

3.2 تقدير فترة الثقة للفروق بين المتوسطات

إذا تم الحصول على الاحصايئتين $\bar{X}_1$, $\bar{X}_2$ من عينتي توزيع معاينتهمــا يتبع نمط التوزيع الطبيعي تقريباً، فإن حدود الثقة للفروق بين معالم المجتمع كما **يلي:**

$$\bar{X}_1 - \bar{X}_2 \pm Z_C \sigma_{X_1 - X_2} = \bar{X}_1 - \bar{X}_2 \pm Z_c \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

◄◄ مثال

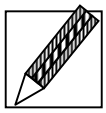
لنفترض أننا اخترنا عشوائياً مجموعتين A و B حيث تتكون المجموعة Aمن 50 تلميذاً والمجموعة B من 100 تلميذاً . أعطيت المجموعة A نوعــاً جديداً من حبوب النوم والمجموعة B نوعاً عادياً من حبوب النوم. فإذا كان متوسط

ساعات النوم للمجموعة A هو 7.82 بانحراف معياري 0.24 ساعة ومتوسط النوم للمجموعة B هو 6.75 بانحراف معيارى 0.3 ساعة. أوجد حدود ثقة 95% للفرق بين متوسطي ساعات النوم الناتجة من استخدام نوعى الحبوب المنومة الجديدة والعادية.

الحل:

$$(\bar{X}_1-\bar{X}_2)\pm Z_C\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma^2{}_2}{n_2}}=(7.82-6.75)\pm 1.96\sqrt{\frac{(0.24)^2}{50}+\frac{(0.3)^2}{100}}=1.07\pm 0.09 \text{ ساعة}$$

تدريب (4)



إذا كان وسط دخل الأسرة السنوي بناء على عينة حجمها 90 أسرة من سكان مدينة امدرمان هو 1540 دينار وكان وسط دخل الأسرة بناء على عينة من 85 من سكان مدينة الخرطوم هو 1100 دينار. وكان الإنحراف المعيارى لمجتمع امدرمان هو 500 دينار ولمجتمع الخرطوم هو 400 دينار. أوجد فترة 99% ثقة للفرق بين وسطي الدخل لمجتمع امدرمان μ_1 ومجتمع الخرطوم μ_2 .

4.2 تقدير فترة الثقة للفروق بين النسب

وبالمثل، فان حدود الثقة للفروق بين النسب في مجتمعين بافتراض أن المجتمعين غير محدودين كما **يلي:**

$$P_1 - P_2 \pm Z_C \sigma_{P_1-P_2} = P_1 - P_2 \pm Z_c \sqrt{\frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}}$$

في كلتا الحالتين تعتمد Z_C على مستوي الثقة المطلوب.

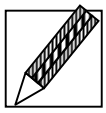
◄◄ مثال

إذا كان لدينا مزرعتان للدواجن A و B. فإذا أخذنا عينة مكونة من 200 بيضة من المزرعة A ووجدنا 15 بيضة منها تالفة، وعينة أخرى حجمها 100 بيضة من المزرعة B ووجدنا بها 12 بيضة تالفة. أوجد حدود ثقة 99% للفروق بين نسب البيض التالفة في المزرعتين.

الحل:

$$P_1 - P_2 \pm Z_C \sqrt{\frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}} = \left(\frac{12}{100} - \frac{15}{100}\right) \pm 2.58 \sqrt{\frac{\left(\frac{12}{100}\right)\left(\frac{88}{100}\right)}{100} + \frac{\left(\frac{15}{200}\right)\left(\frac{185}{200}\right)}{200}} = 0.045 \pm 0.097$$

تدريب (5)



لدراسة المشكلة الصحية المرتبطة بالتدخين فى مـــدينتي واد مـــدني والخرطوم، أخذت عينة من سكان واد مدني حجمها 800 شـــخص ووجد من بينهم 400 مدخناً وأخذت عينة مـــن ســـكان الخرطـــوم حجمها 1200 شخص ووجد من بينهم 800 مدخناً. أوجـــد فتـــرة 95% ثقة للفرق بين نسبتي المدخنين فى واد مدني والخرطوم.

الخلاصة

عزيزي الدارس،،

في نهاية هذه الوحدة نستعرض أهم ما ورد فيها حيث أننا في بداية الوحدة ميزنا بين خواص المعالم وهي المقاييس الاحصائية التي نتحصل عليها من بيانات مفردات المجتمع والاحصائيات وهي المقاييس التي يتحصل عليها من بيانات العينة لنستدل بها علي خصائص ومعالم المجتمع .

كما ذكرنا أن خواص المقدر الجيد هي عجم التحيز وقلة التباين والاتساق ، كذلك وضحنا أن التقدير بنقطة لا يجعلنا نجزم بأن القيمة المقدرة للمعلم هي القيمة الحقيقية له لكن تقدير الفترة يجعلنا أكثر ثقة بأن المعلم المقدر يقع داخل مدى معين .

واختتمنا الوحدة بتقدير فترة الثقة لكل من الوسط الحسابي والنسب والفرق بين الأوساط والفروق بين النسب .

لمحة مسبقة عن الوحدة التالية

الوحدة التالية توضح كيفية اتخاذ القرار المناسب بقبول أو عدم قبول ادعاء معين ولذا سنوضح من خلالها أن هناك نوعين من الخطأ ، كما سنشرح من خلال الوحدة أيضاً خطوات اختبار فرضية أو تصريح ما .

إجابات التدريبات

التدريب (1)

حد الثقة الأعلى هو

$$\overline{x}+z_c\sigma_{\overline{x}} = 20 + 1.96\frac{45}{\sqrt{9}}$$

$$= 20+1.96 \times 15 = 20 + 29.4 = 49.4$$

حد الثقة الأدني هو

$$\overline{x}+z_c\sigma_{\overline{x}} = 20 - 29.4 = -9.4$$

تدريب (2)

$$z_c\sigma_{\overline{x}} = \frac{5}{2}$$

$$1.64\left(\frac{36}{\sqrt{n}}\right) = \frac{5}{2}$$

$$\sqrt{n} = 23.616$$

حجم العينة تقريباً $n = 558$

تدريب (3)

حدود ثقة 95% هي $p \pm z_c\sigma_p$

$$p \pm z_c\sigma_p = \frac{240}{500} \pm 1.96\sqrt{\frac{pq}{n}}$$

$$= 0.48 \pm 1.96\sqrt{\frac{(0.48)(0.52)}{500}}$$

$$= 0.48 \pm 0.044$$

تدريب (4)

حدود فترة الثقة هي $\bar{x}_1 - \bar{x}_2 \pm z_c\ \sigma_{\bar{x}_1 - \bar{x}_2} = \bar{x}_1 - \bar{x}_2 \pm z_\sigma \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

$$\bar{x}_1 - \bar{x}_2 \pm z_c\ \sigma_{\bar{x}_1 - \bar{x}_2} = 1540 - 1100 \pm 2.58 \sqrt{\frac{(500)^2}{90} + \frac{(400)^2}{85}}$$

$$= 440 \pm 17.6$$

تدريب (5)

حدود فترة الثقة هي $p_1 - p_2 \pm z_c\ \sigma_{p_1 - p_2} = p_1 - p_2 \pm z_c \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$

$$p_1 - p_2 \pm z_c\ \sigma_{p_1 - p_2} = \left(\frac{400}{800} - \frac{800}{1200}\right) \pm 1.96 \sqrt{\frac{\frac{1}{2}\cdot\frac{1}{2}}{800} + \frac{\frac{2}{3}\cdot\frac{1}{3}}{1200}}$$

$$= \left(\frac{1}{2} - \frac{2}{3}\right) \pm 1.96 \sqrt{\frac{1}{3200} + \frac{2}{10800}}$$

$$= 0.16667 \pm 0.00098$$

إجابات بعض أسئلة التقويم الذاتي

تقويم ذاتي (2)

$s^2 = 16$, $\bar{x} = 12.4$

حدود فترة الثقة هي

$$\bar{x} \pm z_c \cdot \sigma_{\bar{x}} = 12.4 \pm z_c \sqrt{\frac{4}{25}}$$

$$= 12.4 \pm 1.96 \left(\frac{4}{5}\right)$$

$$= 12.4 \pm 1.57$$

مسرد المصطلحات

- **الاستدلال الاحصائي:** هو عملية استنتاج عن المجتمع مـــن المعلومـــات التـــي توفرها العينة.
- **معلم (Parameter)** : مقياس إحصائي محسوب على أساس كل مفردات المجتمع أو بمعني آخر خاصية وصفية (مثل الوسط والانحراف المعياري) لمجتمع ما.
- **إحصاءة (Statistic)** : مقياس إحصائي محسوب على أساس مفردات العينة أي خاصية وصفية لعينة.
- **عدم التحيز (Unbiased)** : متوسط توزيع المعاينة للاحصاءة المعينة ينبغي أن يكون مساو لمعلم المجتمع المقابل له.
- **التحيز (Bias)** : الفرق بين متوسط توزيع المعاينة للاحصاءة والقيمة الحقيقية لمعلم المجتمع.
- **التقدير بنقطة (Point estimation)** : استخدام قيمة وحيدة كتقدير لمعلم المجتمع.
- **مستوي الثقة (Confidence level) : الإشارة إلى احتمال قبول فرض صحيح.**
- **معامل الثقة (Confidence Coefficient)** : قيمة مقابلة لمستوي الثقة وتستخدم للحصول على التقدير بفترة.

المراجع

- مواري د. شبيجل، "الإحصاء" ، سلسلة ملخصات شوم، الدار الدولية للنشر والتوزيع 1991.

- بول ج.هويل، " المبادئ الأولية في الإحصاء، ترجمة ومراجعة د. بدرية شوقي عبد الوهاب والدكتور محمد كامل الشربيني ، دار جون وايلي وأبنائه.


الوحدة الثالثة
اختبارات الفروض والمعنوية
Hypothesis Testing

محتويات الوحدة

الموضوع	الصفحة
المقدمة	65
تمهيد	65
أهداف الوحدة	66
1. اختبارات الفروض	67
1.1 الفرض الإحصائي	67
2.1 الخطأ من النوع الأول والثاني	68
3.1 مستوى المعنوية	69
2. اختبارات تتضمن التوزيع الطبيعي	70
1.2 خطوات اختبارات الفرض	71
2.2 اختبارات المعنوية	78
الخلاصة	95
لمحة مسبقة عن الوحدة الدراسية التالية	95
إجابات التدريبات	96
إجابات أسئلة التقويم الذاتي	99
مسرد المصطلحات	100
المراجع	101

المقدمة

تمهيد

عزيزي الدارس ،،

في هذه الوحدة تناولنا موضوع اختبارات الفروض والمعنوية ، وكما ذكرنا سابقاً فإن اختبارات الفروض المعنوية هي الشق الثاني للاستدلال الإحصائي إذ أن الشق الأول هو التقدير .

فمن خلال هذه الوحدة عرفنا الفرض الإحصائي وشرحنا الخطأ بنوعيه: الأول والثاني، كما عرفنا أيضاً مستوى المعنوية وذكرنا الخطوات التي ينبغي إتباعها لاختبار فرض معين. كما وضحنا أن هناك اختبار يكون من طرف واحد وآخر من طرفين وكذلك عرفنا كلاً من اختبار المعنوية للوسط ومعنوية النسب واختبار معنوية الفرق بين وسطين ومعنوية الفرق بين نسبتين.

فيجدر بك عزيزي الدارس أن تفهم هذه الوحدة تماماً حتى تدرك أهمية الإحصاء التطبيقي في هذا العصر.

أهداف الوحدة



عزيزي الدارس،،

يتوقع منك بعد الانتهاء من دراسة هذه الوحدة أن تكون قادراً على أن:

1- **تعرف** الفرض الإحصائي .

2- **تشرح** نوعي الخطأ .

3- **تعرف** مستوى المعنوية .

4- **تذكر** الخطوات التي يجب اتباعها عند اختبار فرض ما .

5- **تختبر** معنوية وسط حسابي لمجتمع معلوم تباينه .

6- **تختبر** معنوية النسب .

7- **تختبر** معنوية الفرق بين وسطين .

8- **تختبر** معنوية الفرق بين نسبتين .

1. اختبارات الفروض

كما مر بك سابقاً عزيزي الدارس تعتبر اختبارات الفروض الشق الثاني للاستدلال الاحصائي حيث يتم بناءاً عليها اتخاذ القرار الاحصائي السليم في ضوء بيانات مضمنة في عينة مأخوذة من المجتمع بصورة عشوائية

1.1 الفرض الإحصائي (Statistical Hypothesis)

يعتبر الفرض الإحصائي بيان مؤقت غير حاسم أو تخمين عن معلم واحد أو عدد من معالم مجتمع أو مجموعة من المجتمعات. ونلجأ إلى التخمين في هذه الحالة لأنه لا تتوفر لدينا معلومات أكيدة عن قيم معالم المجتمع. ورغماً عن ذلك ينبغي الوصول إلى قرارات خاصة بقيم هذه المعالم حتى في حالة عنصر عدم التأكد الذي يحيط بها. والهدف الرئيسي للاستدلال الإحصائي هو المساعدة في الوصول إلى قرارات سليمة بصرف النظر عن عنصر عدم التأكد (Uncertainty).

وصنع القرار هو عبارة عن عملية اختيار بين بدائل. وفى إطار نقاشنا لمفردات هذه الوحدة الدراسية، تعنى عملية اتخاذ القرار الوصول إلى استدلال (Inference) عن فرضين نقيضين على أساس معلومات مضمنة في عينة عشـــوائية. وبإعطـــاء مستوى المعنوية المرغوب، نقبل الفرض الأول إذا كانت معلومات العينـــة تعـــززه وتسانده ويتم رفضه إذا كانت المعلومات المستقاة من العينـــة تعطـــى دلـــيلاً ضـــعيفاً لمساندته.

2.1 الخطأ من النوع الأول والثاني
(Type one and Type Two Errors)

في عملية الاستدلال هذه، يمكن الوقوع في احد خطأين:

(أ) – **الخطأ من النوع الأول (Type one Error)** هو رفض فرض كـــان ينبغـــي قبوله.

(ب) **الخطأ من النوع الثاني (Type two Error)** هو قبول فـــرض كـــان ينبغـــي رفضه.

يستخدم كل من الخطأ من النوع الأول والخطأ من النوع الثاني في تقويم قواعد القرار. وللحصول على اختبارات فروض أو قواعد اتخاذ قرارات جيدة ينبغي أن تعمم هذه القواعد بحيث تؤدى إلى التقليل من أخطاء القرار. والصعوبة التي ستواجه الباحث في التطبيق العملي حيث يستخدم عينة واحدة ذات حجم معين فإن سعيه لإنقاص احتمال أحد نوعى الخطأ سيؤدى إلى زيادة في احتمال النوع الأخر. وعادة يكون هناك خطأ اكثر خطورة من الآخر الأمر الذي يتطلب الوصول إلى حل وسط لصالح تحجيم الخطأ الأكثر خطورة. والطريقة المثالية للتقليل من نوعى الخطأ هو زيادة حجم العينة إذا أمكن ذلك. يكون الفرض فى الشكل

" لا يوجد فرق بين....." ويسمي فرض العدم (null hypothesis) ويسمي الثاني الفرض البديل.

3.1 مستوى المعنوية (Level of Significance)

يقصد بمستوى المعنوية أقصى احتمال للوقوع في الخطأ من النوع الأول أي رفض فرض ينبغي قبوله. يرمز لمستوى المعنوية بالرمز α وينبغي تحديده قبل أن يتم سحب العينة من المجتمع حتى لا تؤثر النتائج التي تم الحصول عليها على الاختبار. عادة يستخدم مستوى المعنوية 0.05 و 0.01 في التطبيقات العملية. فمثلاً إذا تم

استخدام 0.05 أو 5% فهذا يعنى أن هناك حوالي 5 فرص في كل 100 فرصة تكـــون عرضة لرفض فرض كان ينبغي قبوله، إي أن الثقة في اتخاذ القرار السليم هي 95%.

إذا وجدت النتيجة معنوية عند مستوى معنوية $\alpha = 0.05$ إي احتمال خطأ من النوع الأول مقداره 5 في المائة توصف النتيجة بأنها "معنوية" Significance . وأمـــا إذا تم استخدام احتمال خطأ من النـــوع الأول مقـــداره 1 فـــي المائـــة إي $\alpha = 0.01$ توصف النتيجة بأنها " عالية المعنوية" Highly Significant. من ناحيـــة إحصـــائية هذان اللفظان لهما معنى دقيق ومحدد.عندما نقول أن النتيجة معنويـــة فإننـــا نقصـــد أن المعلومة المستقاة من العينة لا تؤيد الادعاء في فرض العدم.

2. اختبارات تتضمن التوزيع الطبيعي

ذكرنا من قبل إذا كان حجم العينة "n" يساوى أو اكبر من 30 $(n \geq 30)$ فإن توزيع المعاينة للإحصاءة $\bar{X}$ يتبع نمط التوزيع الطبيعي وأيضا إذا كان المجتمع نفســـه يتبع نمط التوزيع الطبيعي بصرف النظر عن حجم العينة المسحوبة منه، ورمزاً، يكتـــب ذلك كما **يلي:**

$$\sim \text{N } \bar{X} \quad \left(u_S \, , \, \sigma_S{}^2\right)$$

ويمكن التحويل إلى التوزيع الطبيعي المعياري باستخدام القاعدة التالية:

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

وعليه

$$Z \sim N (0,1)$$

أي أن Z تتبع نمط التوزيع الطبيعي المعياري بوسط يساوى الصفر وتبـــاين يساوى الواحد الصحيح.

1.2 خطوات اختبارات الفرض

وفيما يلي سنوضح الخطوات التي ينبغي اتباعها لاختبار فرض معين:

أولاً : صياغة الفرضيات Formulating Hypotheses

الخطوة الأولى هي صياغة فرض العدم Null Hypotheses والتي يرمـــز لها بالرمز H_o والفرض البديل (Alternative Hypothesis) والتي يرمز لها بـــالرمز

H_1 لابد من وجود قيمة افتراضية عن معلم المجتمع تحت الدراسة قد تكون ادعـــاء أو من معلومات سابقة عن المجتمع تم اختبارها بوساطة معلومات مضمنة في عينة.

وعادة يصاغ القرار بعد إتمام الاختبار بالرفض أو عـــدم الـــرفض بالنســـبة للفرض H_0 والسبب هو: كقاعدة عامة من السهل أن ننفى فرض من أقامة دليل قـــاطع عليه. وفى كثير من الأحوال يمكن رفض فرض بدرجة عالية من المعنوية أي بهـــامش خطأ من النوع الأول بسيط للغاية مثلاً 0.0001. ولكن عند عدم رفضه لا يمكن التأكد بثقة كبيرة من دقة هذه النتيجة ولذا لا نقول قبلنا، بل نقول " لا نرفض".

ونقطة أخرى ينبغي الإشارة إليها هي أن الفرض البديل غير محـــدد بصـــورة قاطعة. ويشير عادة إلى عدد من القيم التي يمكن أن يأخذها المعلم تحت الدراســـة إذا كانت تختلف من القيمة الافتراضية المشار إليها في فرض العدم. وهذا يعنى في مقابـــل فرض العدم يوجد عدد من القيم الافتراضية التي يمكن أن ندعيها في الفـــرض البـــديل. وفى الممارسة العملية عادة لا نحدد قيمة في الفرض البديل بصورة دقيقة. وبـــدلاً عـــن ذلك تتم صياغة الفرض البديل بصورة توحي بمدى واسع من القيم البديلة. فمثلاً يمكـــن صياغة الفرض البديل كما يلي: $H_1 : u \neq 70$. وهذا يعنى أنها تغطى كـــل القـــيم فـــي

مجال المعلم ما عدي 70 . في بعض الأحيان يكون كل من فرض العدم والبديل مكونـــا من نقطة واحدة فقط مثلاً $H_0 : \theta = 10$ مقابل $H_1 : \theta = 15$

ثانياً : اختيار مستوى المعنوية Selecting the level of Significance

الخطوة الثانية هي اختيار مستوى المعنوية أي احتمال رفض فرض كان مـــن المفترض قبوله أو احتمال الوقوع في الخطأ من النوع الأول. وفى الدراسات التطبيقيـــة عادة يستخدم مستوى المعنوية $\alpha = 0.05$ أو $\alpha = 0.01$. فإذا كانت $\alpha = 0.05$ وتـــم رفض فرض العدم H_o ، نقول أن نتائج العينـــة تختلـــف معنويـــاً Significantly different من H_o عند مستوى معنوية 5% في المائة.

ثالثاً : تحديد نوع إحصاءة الاختبار Determining the Test Statistic

الخطوة الثالثة تتعلق باختيار إحصاءة الأختبار Test Statistic . وتســـتخدم إحصاءة الاختبار لاختبار فرض العدم عند مستوى معنوية معين. وتعـــرف احصـــاءة الاختبار بأنها دالة فى العينة والمعلم تحت الاختبار بحيث يكـــون توزيعهـــا الاحتمـــالي معروف بالكامل، أي لا يعتمد على المعلم.

رابعاً : اكتشاف توزيع المعاينة لإحصاءة الاختبار
Discovering The Sampling Distribution of the Test Statistics

الخطوة الرابعة هي الحصول على توزيع المعاينة لإحصاءة الاختبار بافتراض صحة فرض العدم H_o .

ويمكن أن يتخذ توزيع المعاينة لإحصاءة الاختبار أشكالاً مختلفة مثل التوزيع الطبيعي المعياري، أو توزيع t أو χ^2 أو F وسندرسها لاحقاً

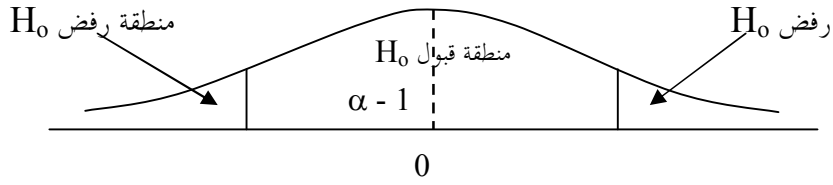
خامساً : تحديد المنطقة الحرجة (أو الرفض)
Defining the Critical (Rejection) Region

بعد تحديد إحصاءة الاختبار وتوزيع معاينتها، تأتى الخطوة الخامسة وهى كيفية تحديد قيم إحصاءة الاختبار التي تتسبب في رفض فرض العدم إذا كانت صحيحة. مجموعة هذه القيم تسمي المنطقة الحرجة Critical Region أو منطقة الرفض Rejection Region . فاحتمال وقوع نتيجة عينة داخل منطقة الرفض إذا كان فرض العدم صحيح يساوى قيمة مستوى المعنوية المحدد سلفاً. قيمة احصاءة الاختبار التي تفصل بين منطقتي الرفض والقبول acceptance region تسمي القيمة الحرجة Critical value

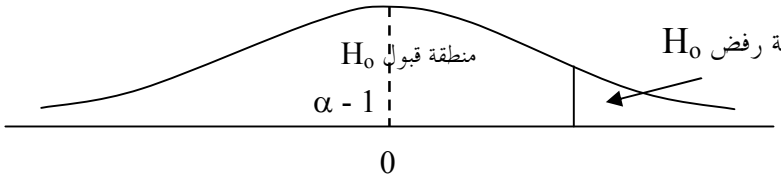
ومن مما يجدر ذكره أن هناك نوعين هناك من الاختبار:

(أ) الاختبار من طرفين (ذيلين) Two-tailed

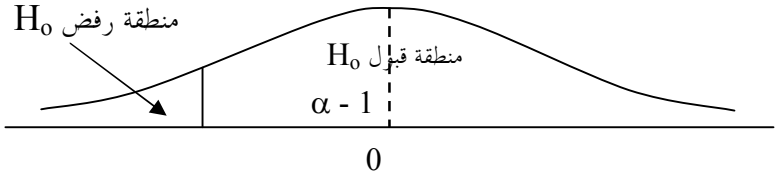
ففي هذه الحالة تتوزع منطقة الرفض على طرفي المنحنى كما موضح بالرسم أدناه(الشكل أ)



شكل (أ)



شكل (ب)



شكل (ج)

(ب) الاختبار من طرف واحد One – tailed test

وفي هذه الحالة تنحصر منطقة الرفض في أحد طرفي التوزيع كما موضح بالرسم (شكل ب ، شكل ج) . ويتحدد ذلك بالصياغة التي يتم بها تحديد الفرض البديل. فإذا

تمت صياغة الفرض البديل بحيث أن قيمة معلم المجتمع أكبر من ، مثلاً $(H_1 : u > 70)$ أو أصغر من $(H_1 : u < 70)$ فإن الاختبار يكون من طرف واحد وعليه منطقة الرفض تنحصر في طرف واحد وهو الذي يشير إليه رمز المتباينة. وإذا كانت الصياغة " لا تساوي" فإن منطقة الرفض تتكون من طرفي التوزيع. والمنطقة التي تقع بين منطقتي الرفض في حالة الاختبار من طرفين تسمى منطقة القبول كما في الشكل (أ) وأما في حالة الاختبار من طرف واحد تكون منطقة القبول أما على الجهة اليسرى كما في الشكل (ب) أو الجهة اليمنى كما في الشكل (ج). ويمكن توضيح ذلك باستخدام الرسومات التالية لحالات معينة يكون فيها توزيع المعاينة للإحصاءة Z يتبع نمط التوزيع الطبيعي المعياري عند مستوى معنوية 5%.

الحالة الأولى:

$$H_0 : u = 70$$
$$H_1 : u \neq 70$$

شكل (أ)

الحالة الثانية

$$H_0 : u = 70$$
$$H_1 : u > 70$$

شكل (ب)

الحالة الثالثة

$$H_0 ; u = 70$$
$$H_1 : u < 70$$

شكل (ج)

القيم 1.96 و 1.64 هي القيم الحرجة Critical Values عند مستوى المعنوية 5% للاختبار من طرفين وطرف واحد على الترتيب. وهى القيم التي تفصل بين منطقة الرفض ومنطقة القبول كما أسلفنا. وبالمثل إذا كان مستوى المعنوية $\alpha = 0.01$ أي 1% فإن منطقة الرفض تفصل بالقيم 2.58 – و 2.58 في حالة الاختبار من طرفين و 2.33 – و 2.33 في حالة الاختيار من طرف واحد علي التوالي.

سادساً : القيام بالعمليات الحسابية

سابعاً : صنع القرار Making The Decisions

نقوم بمقارنة بين القيمة المحسوبة لإحصاءة الاختبار والقيمة الحرجة أو القيمة الجدولية المشار إليها في الخطوة خمسة وتسمى بالجدولية لأنه يمكن الحصول عليها من جدول توزيع الاحصاءة بمعرفة مستوى المعنوية. فإذا كانت القيمة المحسوبة داخل المنطقة الحرجة يرفض فرض العدم أما إذا كانت القيمة المحسوبة داخل منطقة القبول لا يرفض.

2.2 اختبارات المعنوية Test of Hypothesis

لكي تفهم جيداً موضوع اختبارات الفروض نقدم إليك النماذج الآتية للاختبارات :

(أ) اختيار معنوية الوسط الحسابي عندما يكون التباين معلوم

Testing a Hypothesis about u When σ^2 is Known

ليكن لدينا مجتمعاً معلوم التباين ورغبنا في معرفة إذا كان الوسط الحسابي لعينة يختلف معنوياً عن وسط افتراضي. بمعنى آخر نرغب في تحديد فيما إذا كان الوسط الحسابي لعينة معنويا اكبر أو معنوياً أصغر من قيمة افتراضية محددة ففي هذه الحالة سنستخدم الاختبار من طرفين . وعليه إحصائيا يمكن صياغة الفرضين كما يلي:

$H_o : \mu =$ قيمة محددة

$H_1 : \mu \neq$ قيمة محددة

والآن دعنا نقدم اليك هذا المثال ونطبق عليه خطوات اختبارات الفروض

◄◄ مثال

يكسب موظف مبيعات في مصنع سيارات عمولة شهرية متوسطها 400 دولار بانحراف معياري يساوى 5 دولار. تم إنشاء مصنع جديد للسيارات الأمر الذي ترتب عليه انخفاض في عمولة الموظف ودخله. ولكن في نفس الوقت طرأت زيادة على سعر الإنتاج نتيجة للتضخم. وهذا العامل أدى إلى زيادة عمولة موظف المبيعات. وتتشوق إدارة المصنع للكشف عن الاثر الصافي لهذين العاملين (أي إنشاء مصنع جديد وزيادة سعر الإنتاج نتيجة للتضخم) على عمولة الموظف.

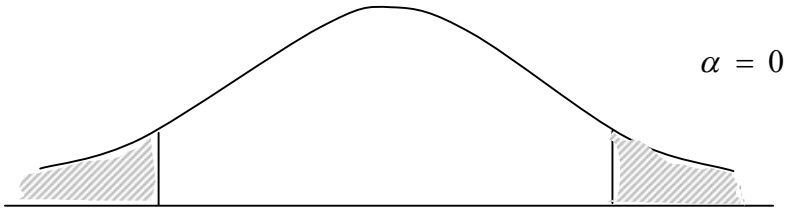
أولا: صياغة الفروض

$H_0 : \mu = 400$ دولار

$H_1 : \mu \neq 400$ دولار

وهذا يعني أن الوسط في الفرض البديل أما اكبر من 400 دولار أو اصغر من 400 دولار.

ثانيا: اختيار مستوى المعنوية



ثالثاً: تحديد إحصاءة الاختبار

وفى هذه الحالة، يختص الاختبار بالوسط الحسابي بافتراض أن الانحراف المعياري للمجتمع معلوم. ويستخدم اختبار الإحصاءة $Z = \left(\bar{X} - u \right) / \sigma_{\bar{X}}$ لاختبار فرض العدم.

رابعاً : اكتشاف توزيع المعاينة لإحصاءة الاختبار:

تتبع $\bar{X}$ نمط التوزيع الطبيعي وخطأها المعياري يساوى $\sigma / \sqrt{n}$. وتوزيع $\left(\bar{X} - u \right)$ طبيعي بوسط حسابي يساوى الصفر ونفس الخطأ المعياري. وعليه توزيع المعاينة للإحصاءة Z والذي يساوى $\left(\bar{X} - u \right) / \frac{\sigma}{\sqrt{n}}$ طبيعي معياري وبافتراض أن فرض العدم صحيح، فإن توزيع Z طبيعي بوسط يساوى الصفر وانحراف معياري يساوى واحد صحيح.

خامساً: تعريف منطقة الرفض (أو المنطقة الحرجة)

الاختبار من طرفين ومستوى المعنوية يساوى 0.05، فإن منطقة الرفض تكون أدنى -1.96 و أعلى 1.96. وتصبح قاعدة القرار كما يلي: ارفض H_0 إذا كانت $Z \le -1.96$ أو إذا كانت $Z \ge 1.96$ وما عدا ذلك لا نرفض فرض العدم H_0.

سادساً: القيام بالعملية الحسابية

- لنفترض أن عينة عشوائية حجمها 100 تم سحبها ووجدت $\bar{X} = 388$ دولار
- فإذا كانت $\sigma = 50$ و n=100 فإن الخطأ المعياري للوسط $\bar{X}$ يساوى:

$$\sigma_{\bar{X}} = \frac{50}{\sqrt{n}} = \frac{50}{\sqrt{100}} = 5$$

ولتكن $H_0 : \mu = 400\$$ وأن $H_1 : \mu \neq 400\$$

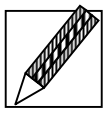
- $Z = \frac{\bar{X} - u}{\sigma_{\bar{X}}} = \frac{388 - 400}{5} = -2.4$

سابعاً: اتخاذ القرار

قيمة Z اقل من 1.96 – وهذا يعنى أنها تقع في منطقة الرفض. وعليه فإن الفرق 12 دولار يعتبر فرق معنوي. وهذا يعنى الصدفة في العينة لا تلعب دوراً في تفسير هذا الفرق.

ويترتب على ذلك اتخاذ القرار التالي: فرض أن وسط العمولة الحقيقي هو 400 دولار غير مقبول عند مستوي دلاله 5%.

تدريب (1)



إدعي قسم الهندسة بمصنع للسيارات بأن معدل استهلاك الوقود لنموذج معين يساوي 40 ميل للجالون. ويرغب قسم الدعاية والاعلان فى اختبار هذا الأدعاء للتأكد من أن الرقم المعلن أقل أو أكبر من 40 ميل للجالون. فإذا كانت $\sigma = 4$ وعينة مكونة من 50 سيارة أعطت $\bar{X} = 36.4$ ميل للجالون. أختبر ادعاء قسم الهندسة باستخدام مستوي معنوية 5%.

(ب) اختبار معنوية النسب

Testing A Hypothesis About A population Proportion

في كثير من الأحيان يتطلب موقف معين حساب نسبية خاصية (Attribute) معينة للمجتمع وليس قيمة الوسط الحسابي لهذا المجتمع كما تم شرحه أعلاه. وتعتبر المشاهدات في هذه الحالة ذات طبيعة نوعية (Qualitative). وعندما تصنف المشاهدات إلى مجموعتين مثلاً ذكور وإناث أو إنتاج مصنع إلى سليم ومعيب، فإن أية مجموعة تعتبر نسبة من المجتمع. المفردة التي تتصف بالخاصية التي يراد إخضاعها للدراسة تسمي مجازاً (نجاح Success) والتي لا تتصف بالخاصية يشار اليها (فشل Failure) ، يرمز لنسبة النوع الاول بالرمز p ونسبة الثانية بالرمز 1-p=q وينصب اهتمامنا في قياس واحدة من النسبتين وعادة نسبة النجاح. وبما أن المجتمع

مكون من نسبتين فقط فأن معرفة نسبة النجاح تؤدى إلى معرفة نسبة الفشل. فإذا كانت

$p = 0.3$ فأن $q = I - P = 0.3 = 0.7$

وإذا كان نسبة مجتمع له خاصية معينة تساوى P وهذه النسبة ليست قريبة من الواحد أو الصفر فإن توزيع المعاينة للنسب والتي لها نفس الخاصية في عينة كبيرة حجمها n يتبع بالتقريب نمط التوزيع الطبيعي بوسط يساوى P وخطا معياري :

$$\sigma_P = \sqrt{\frac{Pq}{n}}$$

ولاختبار معنوية النسب نتبع نفس الخطوات السابقة لاختبار معنوية الوسط الحسابي. ولشرح هذه الخطوات نفترض المثال التالي:

◄◄ مثال

أدعى مدير الهيئة القومية للتلفزيون والإذاعة إن أقل من نصف الشباب الذين يشاهدون التلفزيون لا ينزعجون من العنف في بعض البرامج. اختبر هذا الإدعاء إذا كشف مسح ميداني شمل 1998 من الشباب بأن 48 في المائة منهم ينزعجون من العنف باستخدام مستوى معنوية 5 في المائة.

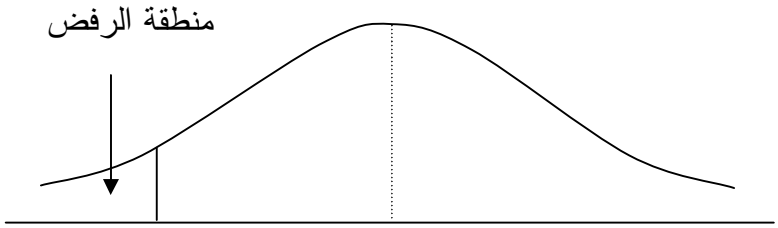
الحل:

بما أن فرض العدم يجب أن يكون بالتساوي فإننا نضع

$H_o : P = 0.5$

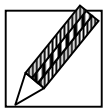
$H_1 : p < 0.5$ (وهو ادعاء المدير)

$$Z = \frac{\hat{P} - P}{\sqrt{\frac{Pq}{n}}} = \frac{0.48 - 0.5}{\sqrt{\frac{(0.5)(0.5)}{1998}}} = -1.79$$



وبما أن القيمة المحسوبة (Z = -1.79) تقع داخل منطقة الرفض المظللـــة، نرفض فرض العدم. وعليه يوجد دليل كاف من العينة يعزز الادعاء بأن أقل من نصف الشباب ينزعجون من برامج العنف بالتلفزيون.

تدريب (2)



إذا كانت نسبة المدخنين في مدينة الخرطوم P1 أخذت عينة حجمها 500 فوجد أن منهم 250 مدخناً ، اختبر الفرض H0: P1= 0.4 مقابل الفرض H1: P1< 0.4 عند مستوى دلالة 1 %

(ج) اختبار معنوية الفرق بين وسطين
Test of A Difference between Two Means

ينصب الاهتمام في مثل هذا النوع من الاختبارات في معرفة إذا تم سحب عينتين من نفس المجتمع أو من مجتمعين مختلفين بنفس المتوسط هل يوجد فرق حقيقي بين وسط المجتمعين. مثلاً، هل هناك فرق حقيقي بين متوسط دخل العمال الذين يعملون في القطاع العام ومتوسط دخل العمال الذين يعملون في القطاع الخاص ؟ وهل هناك فرق حقيقي بين متوسط سرعة سيارة A ومتوسط سرعة سيارة B ؟ وللحصول على إجابة على الأسئلة المطروحة ينبغي اختبار معنوية الفرق بين وسطين. ونتبع نفس الخطوات السابقة لاختبار المعنوية هذا. والمعلومة الإضافية التي نحتاجها هي معرفة طبيعة توزيع المعاينة للفرق بين الأوساط، وقد سبق توضيح ذلك.

◄◄ مثال

يستخدم جهازان لتعبئة مشروب معين. اختيرت عينة من إنتاج كل جهاز. نرمز للنتائج من الجهاز (أ) بالمجموعة 1 ومن الجهاز (ب) بالمجموعة 2 وهي كما يلي:

الجهاز (أ)	الجهاز (ب)
$n_1 = 50$	$n_2 = 100$
$\bar{X}_1 = 4.53$	$\bar{X}_2 = 4.01$

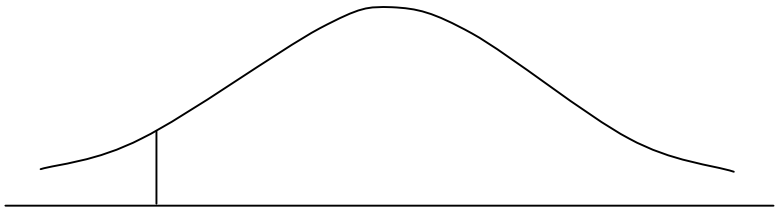
إذا كان الانحراف المعياري للسلعة المعبأة بالجهاز (أ) هو 0.8 كجم وللجهاز (ب) 0.6 كجم أختبر الإدعاء بأن وسط السعة المعبئة بالجهاز (أ) يساوى وسط السعة المعبئة بالجهاز (ب) باستخدام مستوي معنوية 5 في المائة.

الحل:

$$H_o : \mu_1 = \mu_2$$

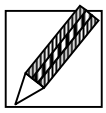
$$H_1 : \mu_1 \neq \mu_2$$

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{4.53 - 4.01}{\sqrt{\frac{(0.8)^2}{50} + \frac{(0.6)^2}{100}}} = 4.06$$



وبما أن القيمة المحسوبة (Z = 4.06) اكبر من القيمة الحرجة (Z* = 1.96) نرفض فرض العدم ونستنتج أن وسطي المجتمعين المقابلين للجهازين غير متساويين.

تدريب (3)



أعطي اختبار ذكاء لمجموعتين من الذكور والإناث النتائج التالية:

النوع	الوسط	الانحراف المعياري	حجم العينة
الإناث	75	15	150
الذكور	70	20	250

هل يوجد فرق معنوي في وسط النتائج التي تم الحصول عليها بين الذكور والاناث. استخدم مستوي دلالة $\alpha = 0.01$

(د) اختبار معنوية الفرق بين نسبتين

Significance of A Difference between Two Proportions

يمكن استخدام الطرق والمفاهيم الخاصة بأختبار معنوية الفرق بين نسبتين للإجابة على مثل الأسئلة **التالية:**

- هل هناك فرق حقيقي بين نسبة تعاطى المخدرات للفئة العمرية 17−25 ونسبة تعاطى المخدرات في الفئة 25− 30؟.
- هل هناك فرق حقيقي بين نسبة الرجال والنساء الذين يشغلون مناصب تنفيذية.

- هل هناك فرق حقيقي بين نسبة الطلاب الذين ينجحون في مادة الإحصـــاء ونسبة الذين ينجحون في مادة الاقتصاد.
- هل هناك فرق حقيقي بين نسبة مبيعات الكمبيوتر من نوع معين في مدينة الخرطوم ونسبته في مدينة مدني.

وللإجابة على مثل هذا النوع من الأسئلة سنفترض أن بيانـــات العينـــة تأتى من مجتمعين مستقلين عن بعضهما. ومن المجتمع الأول نفترض التالي:

p_1 تمثل نسبة المجتمع

n_1 تمثل حجم العينة

X_1 تمثل عدد النجاحات (number of successes)

$$\hat{P}_1 = \frac{x_1}{n_1}$$

وبالمقابل للمجتمع الثاني نفترض **التالي:**

P_2 تمثل نسبة المجتمع

n_2 تمثل حجم العينة

x_2 تمثل عدد النجاحات

$$\hat{P}_2 = \frac{x_2}{n_2}$$

$\hat{P}_1$ تمثل تقدير العينة لنسبة المجتمع π_1

$\hat{P}_2$ تمثل تقدير العينة لنسبة المجتمع الثاني π_2

ونعرف من توزيع ذي الحدين (Binomial Distribution) إذا كانت $n_1 p_1 \geq 5$ و $n_1 q_1 \geq 5$ فإن النسبة $\hat{P}_1$ تتبع نمط التوزيع الطبيعي بالتقريب. والفروق $\hat{P}_1 - \hat{P}_2$ تتبع أيضاً نمط التوزيع الطبيعي بالتقريب. وبما أن وسطى $\hat{P}_1$ و $\hat{P}_2$ هما π_1 و π_2 على التوالي يترتب على ذلك أن وسط الفروق $\hat{P}_1 - \hat{P}_2$ هو $\pi_1 - \pi_2$. وسبق أن ذكرنا أن توزيع النسب $\hat{P}$ خطأه المعياري (Standard error) يساوى $\sqrt{Pq/n}$ وهذا يوحي أن تباين نسبة العينة $\hat{P}_1$ هو $P_1 q_1 / n$. وبنفس المنطق فإن تباين $\hat{P}$ هو $P_2 q_2 / n$ وتباين الفروق بين المتغيرات العشوائية المستقلة هو مجموع تبايناتها (variances) كل على حدة. ويمكن استخدام هذه الخاصية للحصول على :

$$\sigma_{\hat{P}_1 - \hat{P}_2} = \sigma^2_{\hat{P}_1} + \sigma^2_{\hat{P}_2} = \frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}$$

ولكن في كثير من الأحيان لا نستطيع استخدام الصيغة أعلاه لأن قيم π_1 و π_2 مجهولة. ولكن بافتراض أن $\pi_1 = \pi_2$ يمكن تقدير قيمتهما المشتركة بدمج بيانات العينة كما **يلي:**

$$\bar{P} = \frac{x_1 + x_2}{n_1 + n_2}$$

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

وباستخدام التقدير المشترك (Pooled estimate) محل P_1 و P_2 نحصل على الخطأ المعياري للفروق بين نسبتي العينتين كما يلي:

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

وسبق أن ذكرنا أن فروق نسب العينة $\hat{P}_1-\hat{P}_2$ تتبع بالتقريب نمط التوزيع الطبيعي بوسط يساوى $P_1 - P_2$ وخطأ معياري كما هو موضح أعلاه.

ولتثبيت المعلومات أعلاه في ذهنك عزيزي الدارس يمكن تلخيصها في سطور معدودة كما يلي: طالما أن n_1P_1 و n_1q_1 و n_2P_2 و n_2q_2 كلاً على حدة تأخذ قيمة على الأقل مساوية 5، فإن احصاءة الاختبار (test statistic) **التالي:**

$$Z=\frac{\left(\hat{P}_1-\hat{P}_2\right)-\left(\pi_1-\pi_2\right)}{\sqrt{\bar{P}\bar{q}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}=\frac{\hat{p}-\hat{p}-0}{\sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

لان $\pi_1 = \pi_2$ حسب H_0

حيث أن:

$$\hat{P}_1 = \frac{x_1}{n_1} \qquad \text{و} \qquad \hat{P}_2 = \frac{x_2}{n_2}$$

$$\bar{P} = \frac{x_1 + x_2}{n_1 + n_2}$$

$$\bar{q} = 1 - \bar{P}$$

له توزيع معاينة (Sampling Distribution) يتبع بالتقريب نمط التوزيع الطبيعي المعياري. ويطبق اختبار الإحصاءة (Test Statistic) في حالات معينة. وهذه الحالات هي عندما بكون فرض العدم H_0 كما يلي:

$$H : \pi_1 = \pi_{21}$$

$$H_1 : \pi_1 \neq \pi_2 \qquad \text{أو}$$

وإذا رغبنا في اختبار فرض عدم حيث أن الفرق $\pi_1 - \pi_2$ يساوي ثابت خلاف الصفر، يمكن استخدام اختبار إحصاءة مختلفة (*). وعليه نحصر اختبار المعنوية عندما تكون فرضيات العدم كما هو موضح أعلاه.

(*) لاختيار فرضية عدم حيث أن الفرق بين نسبتي مجتمع يساوى ثابت $C \neq 0$ يمكن استخدام الصيغة التالية:

$$Z = \frac{(\hat{P}_1 - \hat{P}_2) - C}{\sqrt{\frac{\hat{P}_1(1-\hat{P}_1)}{n_1} + \frac{\hat{P}_2\,(1-\hat{P}_2)}{n_2}}}$$

مثال

أجري فريق بحث من جامعة الخرطوم في منتصف الستينات من القرن المنصرم مسحاً ميدانياً حول تمويل المستهلك وقد تم اختيار عينة مكونة من 3000 اسرة. و أجرى نفس الفريق دراسة مماثلة في مطلع السبعينات للقرن المنصرم حيث تم اختيار 3120 أسرة. وتوصل المسحان للوقائع التالية:

في منتصف الستينات 7.7 في المائة من الذين تمت معاينتهم يرغبون في شراء أجهزة تلفزيون ولكن 5.9 في المائة من الذين تمت معاينتهم في مطلع السبعينات يخططون لاقتناء أجهزة تلفزيون. هل هناك تغيير معنوي في رغبات المستهلكين لشراء أجهزة تلفزيون بين منتصف الستينات ومطلع السبعينات عند مستوي معنوية 5%.

الحل:

يمكن الإجابة على السؤال في صلب المثال السابق بإتباع الخطوات التالية:

- الفرضيات Hypotheses

$$H_o : \pi_1 = \pi_2$$

$$H_1 \ \pi_1 \neq \pi_2$$

- مستوى المعنوية Level of Significance

$$\alpha = 0.05$$

- إحصاءة الاختبار Test Statistic

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

- توزيع اختبار الإحصاءة Distribution of Z

إذا كانت فرض العدم H_o صحيح، اختبار الإحصاءة Z يتبع نمط التوزيع الطبيعي المعياري.

- منطقة الرفض Rejection Region

قاعدة القرار :

ارفض H_o إذا كانت $Z \leq -1.96$ و $Z \geq +1.96$ ..

وأقبل فرض العدم اذا كان $-1.96 < Z < 1.96$

- **العملية الحسابية Computations Process**

البيانات المعطاة كما يلي:

منتصف الستينات	مطلع السبعينات
$\hat{P}_2 = 0.077$	$\hat{P}_1 = 0.059$
$n_2 = 3000$	$n_1 = 3120$

$$\bar{P} = \frac{n_1 \hat{P}_1 + n_2 \hat{P}_2}{n_1 + n_2} = \frac{3120(0.059) + 3000(0.077)}{3120 + 3000} = 0.0678$$

$$\bar{q} = 1 - 0.0678 = 0.9322$$

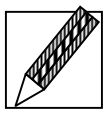
$$Z = \frac{P_1 - P_2}{\sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.059 - 0.077}{\sqrt{(0.0678)(0.9322)\left[\frac{1}{3120} + \frac{1}{3000}\right]}}$$

$$= \frac{0.018}{0.64} = -2.81$$

⑦ القرار Decision

وبما إن القيمة المحسوبة اكبر من القيمة الجدولية (الحرجة) نرفض فرض العدم H_o ونقر بأن هناك تغير معنوي في رغبات الأسرة لشراء أجهزة تلفزيون في الفترة ما بين منتصف الستينات ومطلع السبعينات عند مستوي معنوية 5%.

تدريب (4)



قام أمين مكتبة جامعة الخرطوم بعمل إحصائية للاعارات من كتب الاقتصاد الجزئي وغير كتب الاقتصاد الجزئي فوجد أن 350 كتاب اقتصاد جزئي بين 600 كتاب قد استعيرت فى العام الماضي بينما استعير 450 من غير كتب الاقتصاد الجزئي من بين 750 من هذا النوع. هل توجد فروق جوهرية بين نسب كتب الاقتصاد الجزئي المستعارة ونسب غير كتب الاقتصاد الجزئي. استخدم مستوى دلالة 5%.

الخلاصة

بعد دراستك لهذه الوحدة ينبغي أن تكون قد عرفت أن الفرض الإحصـــائي هـــو تخمين عن معلم واحد كما عرفت أيضاً أن الخطأ نوعين الأول هو رفض فرض يجـــب قبوله ، ويسمى أقصى احتمال للوقوع في هذا النوع من الخطـــأ بمســـتوى المعنويـــة ؛ والثاني هو قبول فرض يجب رفضه .

وأيضاً من خلال هذه الوحدة تعلمنا أن خطوات الاختبار هي كالآتي : صـــياغة الفرض ، اختيار مستوى المعنوية ، تحديد نوع إحصاء الاختبـــار ، اكتشـــاف توزيـــع المعاينة لإحصاءة الاختبار ، تحديد المنطقة الحرجة ، صنع القرار .

كذلك تعلمت كيف تختبر فروض تتضمن التوزيع الطبيعـــي للوســـط الحســـابي والنسب للفروق بين الأوساط وبين النسب .

لمحة مسبقة عن الوحدة التالية

في هذه الوحدة قدمنا بعض التوزيعات الإحصـــائية و وضـــحنا كيفيـــة التقدير بفترة واختبارات المعنوية في حالة العينات الصـــغيرة وكيفيـــة اختيـــار الطريقة الإحصائية السليمة لتطبيق الاختبار .

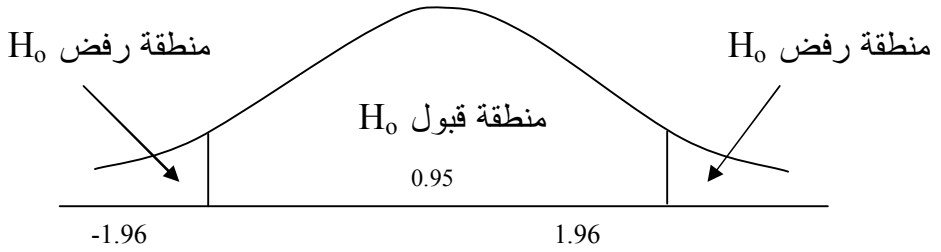
إجابات التدريبات

تدريب (1)

$$H_o : \mu = 40$$

$$H_1 : \mu \neq 40$$

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{36.4 - 40}{4 / \sqrt{50}}$$

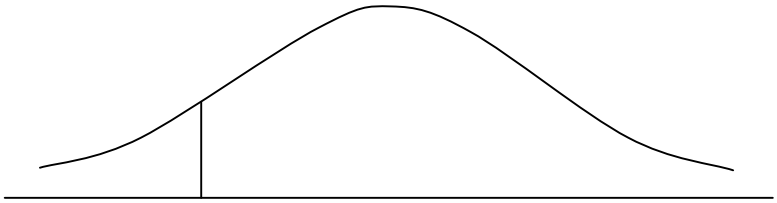


بما أن Z تقع في منطقة الرفض ، إذن نرفض فرضية العدم
بمعنى يرفض إدعاء الشركة بأن متوسط الجالون 40 ميل ونقبل الفرضية البديلة أن المتوسط أكبر أو أقل من 40 ميل .

تدريب (2)

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p\,q}{n}}}$$

$$= \frac{0.5 - 0.4}{\sqrt{\frac{0.4 \times 0.6}{500}}} = -4.5$$



إذن عند مستوى دلالة $\alpha = 0.01$

نرفض $\mu_o : p = 0.4$

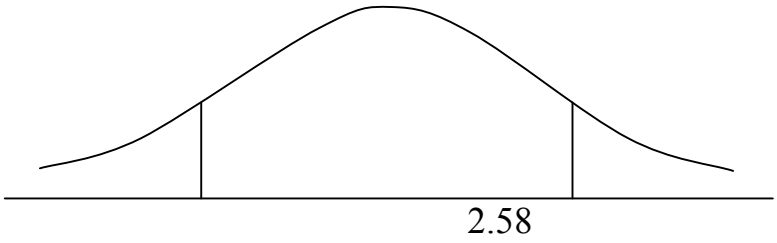
ونقبل الفرض البديل $H_1: p < 0.4$

تدريب (3)

$$H_o : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1^2} + \frac{\sigma_2^2}{n_2^2}}} = \frac{75-70}{\sqrt{\frac{(15)^2}{(150)^2} + \frac{(20)^2}{(250)^2}}} = \frac{5}{\sqrt{\frac{1}{100} + \frac{64}{10000}}} = 39$$



بما أن القيمة المحسوبة $z = 39$ أكبر من Z_c

إذن نرفض فرض العدم $\mu_1 = \mu_2$

ونقبل الفرض البديل ونستنتج أن الوسطين غير متساويين

تدريب (4)

$$H_o : \pi_1 = \pi_2$$

$$H_1 : \pi_1 \neq \pi_2$$

مستوى المعنوية $\alpha = 0.05$

$$z = \frac{p_1 - p_2}{\sqrt{\bar{p}\,\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$\hat{p}_1 = \frac{350}{600} = 0.583 \quad , \quad n_1 = 600$$

$$\hat{p}_2 = \frac{450}{750} = 0.6 \quad , \quad n_2 = 750$$

$$\bar{p} = \frac{350 + 450}{1350} = \frac{800}{1350} = 0.593$$

$$\bar{q} = 1 - 0.593 = 0.407$$

$$z = \frac{p_1 - p_2}{\sqrt{\bar{p}\,\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.583 - 0.6}{\sqrt{(0.593)(0.407)\left(\frac{1}{600} + \frac{1}{750}\right)}}$$

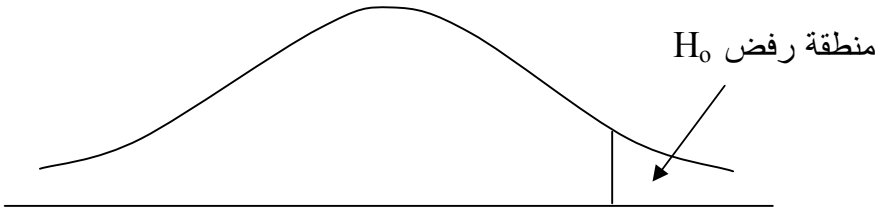
إجابات بعض أسئلة التقويم الذاتي

تقويم ذاتي (1)

$$H_o : \mu_1 = 60$$

$$H_1 : \mu_1 < 60$$

$$z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{70-60}{11/\sqrt{150}} = \frac{10}{11/(12.3)} = 11.18$$



2.33

إذن نرفض فرض العدم $H_o : \mu_1 = 60$

ونقبل الفرض البديل $H_1 : \mu_1 < 60$

مسرد المصطلحات

الفرض الإحصائي (Statistical hypothesis) : بيان مؤقت غير حاسم أو تخمين عن معلم واحد أو عدد من معالم مجتمع أو مجموعة من المجتمعات.

الخطأ من النوع الأول (Type One error) : رفض فرض كان ينبغي قبوله.

الخطأ من النوع الثاني (Type Two error) : قبول فرض كان ينبغي رفضه.

مستوي المعنوية (Level of Significance) : أقصي احتمال للوقوع في الخطأ من النوع الأول أي رفض فرض ينبغي قبوله.

فرض العدم (Null hypothesis) : بيان الفرق الصفرى والذي يختبر مباشرة. وهو يقابل الادعاء الأصلي إذا كان هذا الادعاء يشمل شرط عدم التغير أو عدم وجود فرق.

الفرض البديل (Alternative hypothesis) : بيان يكون صحيح إذا كان فرض العدم غير صحيح.

إحصاءة الاختبار (Test Statistic) : أحصاءة عينة أو قيمة تم الحصول عليها من عينة. وتستخدم لاتخاذ قرار يخص رفض فرض العدم.

المنطقة الحرجة أو منطقة الرفض (Critical or Rejection region) : مجموعة كل قيم أحصاءة الاختبار التي تجعلنا نرفض فرض العدم.

القيمة الحرجة (Critical value) : قيمة تفصل المنطقة الحرجة من قيم أحصاءة الاختبار التي لا تؤدي إلي رفض فرض العدم. والقيمة الحرجة تعتمد على طبيعية فرض العدم وتوزيع المعاينة المعني ومستوي المعنوية α.

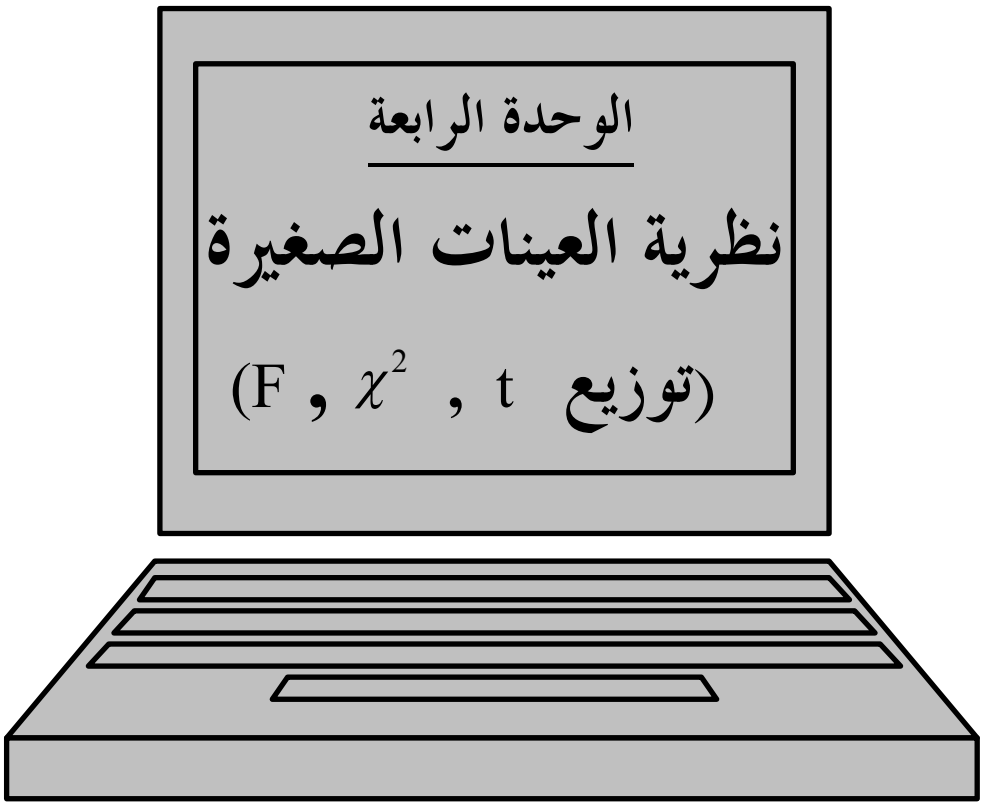
α **(ألفا)** : رمز يستخدم ليمثل إحتمال الخطأ من النوع الأول.

β **(بيتا)** : رمز يستخدم ليمثل إحتمال الخطأ من النوع الثاني.

المراجع

- مواري د. شبيجل، "الإحصاء" ، سلسلة ملخصات شوم، الدار الدولية للنشر والتوزيع 1991.

- بول ج.هويل، " المبادئ الأولية في الإحصاء، ترجمة ومراجعة د. بدرية شوقي عبد الوهاب والدكتور محمد كامل الشربيني ، دار جون وايلي وأبنائه.


الوحدة الرابعة
نظرية العينات الصغيرة
(توزيع t , χ^2 , F)

محتويات الوحدة

الموضوع	الصفحة
المقدمة	105
تمهيد	105
أهداف الوحدة	106
1. العينة الصغيرة Small Sample	107
2. أمثلة لبعض التوزيعات	07
1.2 توزيع ت (t)	109
2.2 توزيع مربع كاي (χ^2 Distribution)	112
3.2 توزيع F (F-distribution)	115
3. فترات الثقة واختبارات الفروض للعينات الصغيرة	117
1.3 تقدير فترة للوسط الحسابي للمجتمع μ عندما يكون الانحراف المعياري σ	118
2.3 اختبار μ عندما يكون حجم العينة صغير	121
3.3 المقارنة بين وسطين للمجتمع في حالة العينات الصغيرة	124
4. اختبارات التباينات (Test of Variances)	128
1.4 اختبار فرضية تباين وحيد	131
2.4 اختبار فرض عن نسبة بين تباينيين	136
3.4 اختبار F وتحليل التباين	146
الخلاصة	165
لمحة مسبقة عن الوحدة الدراسية التالية	166
إجابات التدريبات	167
مسرد المصطلحات	177
المراجع	178

المقدمة

تمهيد

نرحب بك في هذه الوحدة عزيزي الـــدارس وهـــي بعنـــوان نظريـــة العينـــات الصغيرة، إذ أن طريقة التقدير بفترة واختبارات المعنوية تختلف حسب حجم العينة. ففي هذه الوحدة عرفنا لك العينة الصغيرة، وعرفنا لك توزيعـــات ت، ف، مربـــع كـــاي و وضحنا لك خواص هذه التوزيعات.

كذلك شرحنا لك طريقة اختبار فترات الثقـــة واختبـــارات الفـــروض للعينـــات الصغيرة حيث قدمنا بعض الأمثلة كتقدير فترة الثقة واختبار الفرض للوســـط الحســـابي عندما تكون العينة صغيرة الحجم والمجتمع مجهول الانحـــراف المعيـــاري ، وكـــذلك وضحنا كيفية اختبار معنوية الفرق بين وسطين لمجتمعين لهما تباين واحد مجهول وذلك عندما تكون العينات المأخوذة صغيرة .

ثم تناولنا اختبارات التباينات حيث شرحنا الطريقة التي يتم بهـــا اختبـــار تبـــاين وحيد والي يتم بها فرض عن نسبة تباينين وذلك بتقديم بعض الأمثلة . واختتمنا الوحـــدة بتحليل التباين في اتجاه واحد ثم في اتجاهين .

أهداف الوحدة



عزيزي الدارس،،
يتوقع منك بعد الانتهاء من دراسة هذه الوحدة أن تكون قادراً على أن:

1- **تعرف** العينة الصغيرة.

2- **تشرح** توزيع ت وتوضح سماته.

3- **تشرح** توزيع ف وتوضح سماته.

4- **تشرح** توزيع مربع كاي وتوضح سماته.

5- **تقدر** فترة الوسط الحسابي لمجتمع مجهول الانحراف المعياري وذلك من خلال بيانات عينة صغيرة الحجم.

6- **تختبر** معنوية وسط الحسابي لمجتمع مجهول الانحراف المعياري وذلك من خلال بيانات عينة صغيرة الحجم.

7- **تقارن** بين وسطين لمجتمعين في حالة العينات الصغيرة وتباين المجتمعين واحد لكنه مجهول .

8- **تختبر** فرضية تباين وحيد.

9- **تختبر** فرض عن نسبة بين تباينين.

10- **تنشئ** جدول لتحليل التباين.

11- **تختبر** معنوية الفروق بين ثلاثة أو أكثر من المتوسطات .

1. العينة الصغيرة Small Sample

يقصد بالعينة الصغيرة العينة التي يقل حجمها عن 30 مفردة ، ففي الوحدات السابقة كان حجم عينة اكبر من ثلاثين مفردة ($n > 30$) فهي إذاً من العينات الكبيرة (Large Sample).حيث كان توزيع المعاينة للمتوسطات يتبع نمط التوزيع الطبيعي التقريبي ويزداد التقريب جودة كلما زاد حجم العينة n. و لكن عندما نتعامل مع عينة صغيرة ($n<30$) يكون تباين العينة مختلف تماما عن تباين المجتمع. وعليه، سنقع في خطأ جسيم إذا استخدمنا التوزيع الطبيعي التقريبي وحسبنا الخطأ المعياري بإحلال $\hat{s}$ محل σ. وينبغي في حالة العينات الصغيرة استخدام توزيع t لاختبار معنوية وسط المجتمع.

2. أمثلة لبعض التوزيعات

سنحاول التعرف على توزيعات χ^2 ,F, t وإلقاء الضوء على سماتها الرئيسة لأهميتها في اختبارات المعنوية وبناء فترات ثقة لبعض معالم المجتمع.

في دراستنا لاختبارات المعنوية وبناء فترات ثقة في الوحدة الثالثة افترضنا ان تباين المجتمع σ^2 معلوم. وهذا الفرض غير واقعي لان في كثير من التطبيقات و

الممارسات الاحصائية المختلفة تباين المجتمع مجهول وينبغي تقديره باستخدام تباين العينة. وعليه يصبح الخطأ المعياري $\sigma_{\bar{x}}$ لتوزيع المعاينة للاوساط و الذي يساوي $\sigma/\sqrt{n}$ مجهول ولا نستطيع في هذه الحالة استخدام التوزيع الطبيعي المعياري

$$Z = \frac{\bar{X}-u}{\sigma/\sqrt{n}} \quad \text{.....(1)}$$

لبناء فترات ثقة أو اختبار معنوية. وإذا تم إحلال التقدير $\hat{s}$ محل σ في المقام، فان المقدار $\frac{\bar{X}-u}{\hat{s}/\sqrt{n}}$ لا يتبع نمط التوزيع الطبيعي وخاصة عندما يكون حجم العينة صغيراً.

والسؤال عزيزي الدارس الذي يطرح نفسه في هذا الصدد، كيف يمكن اختبار معنوية وسط مجتمع أو الفرق بين وسطي مجتمعين عندما يكون الخطأ المعياري لتوزيع المعاينة للأوساط مجهولاً.

وهنا نسارع لنقول لك إذا كان المجتمع طبيعي فإن المتغير $T = \frac{\bar{X}-u}{\hat{s}/\sqrt{n}}$ بإحلال $\hat{s}$ محل σ يتبع توزيع ت (t).

1.2 توزيع ت (t)

ويعتبر المتغير T متغير عشوائي ويسمى إحصاءة اختبـــار (Test Statistic) حتى لا يحدث خلط بينه وبين الإحصاءات الأخرى التي تستخدم كتقدير لمعالم المجتمع.

وتعطي T بالمعادلة $T = \frac{\bar{x} - \mu}{\hat{s}\sqrt{n}}$

حيث $\bar{x}$ هو وسط العينة ، μ هو وسط المجتمـــع ، n هـــو حجـــم العينـــة ، $\hat{s}$ هـــي الانحراف المعياري للعينة

وينبغي الآن إلقاء الضوء على طبيعة توزيع ت (t) و الفروض الأساسية التـــي يقوم عليها. و أول هذا الفروض هي أن المتغير العشـــوائي x يتبـــع نمـــط التوزيـــع الطبيعي. وتنبع أهمية هذا الفرض من ضرورة الاستقلال الإحصائي لكل مـــن $\hat{s}/\sqrt{n}$ و $\bar{X}$ حتى يمكن توصيف التوزيع ت (t) . ولا يتم الاستقلال الإحصائي بـــين $\hat{s}/\sqrt{n}$ و $\bar{X}$ إلا إذا كان المجتمع يتبع نمط التوزيع الطبيعي. وينبغي أن نشير هنا إلى تـــوفر الاستقلال الإحصائي بين الوسط الحسابي للعينة $\bar{X}$ وتباين العينة s^2 أو $\hat{s}^2$ إذا كان المتغير العشوائي x يتبع نمط التوزيع الطبيعي.

وإذا كان المتغير العشوائي x يتبع نمط التوزيع الطبيعي، فان وسط العينة $\bar{X}$ يتبع أيضا نمط التوزيع الطبيعي. وبدورها $\bar{X} - \mu$، (بسط المقدار T)، يتبع نمط التوزيع الطبيعي حيث أن μ متوسط X وعملية طرحها من $\bar{x}$ لا يؤثر على تشتت (Dispersion) التوزيع ولكن ينقله مقدار μ وحدة جهة اليسار. و النقطة الجوهرية هنا أن $\hat{s}/\sqrt{n}$ والتي تشكل مقام لتوزيع t ليست مثل $\sigma/\sqrt{n}$ ثابتة، فهي تتغير من عينة إلى أخرى. ونتيجة لذلك فإن النسبة T ليست مثل النسبة Z فهي لا تتبع نمط التوزيع الطبيعي بيد أن توزيع ت (t) يقترب كثيرا من توزيع Z خاصة عندما يكون حجم العينة كبيراً. وسط توزيع T يساوي وسط توزيع Z. أي أن E(t) = 0. وتباين t يساوي n/(n-2) وهو اكبر بقليل من الواحد الصحيح.

ودالة كثافة t (Density function) **كما يلي:**

$$f(t) = \frac{k}{\left(1 + t^2/\nu\right)^{(\nu+1)/2}} \quad , \quad -\infty < t < \infty \ldots\ldots\ldots\ldots(2)$$

حيث أن:

v : تمثل درجات الحرية (Degrees of Freedom)

k : ثابت موجب يعتمد على قيمة v

وينبغي الإشارة هنا إلى أن توزيع t يعتمد على معلمة واحدة فقط وهي درجات الحرية v. وبالرغم من ان دالة كثافة وتوزيع t تختلف عن دالة كثافة التوزيع الطبيعي

الا ان بيان (Graph) توزيع ت (t) يشبه الى حد كبير بيان التوزيع الطبيعي فكلاهما متماثل (Symmetric) ولكن توزيع ت (t) منبسط قليلا في منطقة الوسط.

وإذا دققنا النظر في صيغة دالة كثافة توزيع ت (t) نلاحظ ان قيمة ت (t) مربعة (Squared) في المقام الأمر الذي يترتب عليه تساوي دالة الكثافة لقيمتين مختلفتين لـ t إذا كان لهما نفس الحجم المطلق (أي 5 أو 5−). وهذا يفسر تماثل بيان توزيع ت (t) . ونلاحظ أيضا أن درجات الحرية v عبارة عن ثابت موجب.

ونتيجة لذلك فان الأس (v+1)/2 موجب. وهذا يعني ان كل ما زادت قيمة t المطلقة كلما قل احتمال الكثافة (Probability Density) المرتبط بها الامر الذي يوضح ان توزيع ت (t) ذو منوال وحيد (Unimodal). وبما ان التوزيع المتماثل يتساوى فيه الوسط و المنوال، فان منوال توزيع (t) يساوي الصفر (t=0) وعند هذه القيمة فان دالة كثافة توزيع ت(t) تبلغ نهايتها العظمى. ونلاحظ ان القيمة العظمي للدالة تساوي تماما k .

عزيزي الدارس سبق إن أشرنا إلى إن تباين توزيع t اكبر من واحد صحيح ولهذا فان التوزيع يتشتت أكثر من توزيع Z خاصة عندما يكون حجم العينة صغيراً. ويعتمد شكل بيان توزيع t على حجم العينة او درجات الحرية v. ودرجات الحرية بدورها تساوي حجم العينة ناقص واحد (v = n-1) عندما تكون لدينا عينة وحيدة للمشكلة تحت الدراسة. وهذا يعنى ان توزيع t ليس وحيداً مثل التوزيع الطبيعي المعياري، بل يوجد عدد كبير من توزيعات t كل واحد يعتمد على درجة الحرية الخاصة به. وكلما زاد حجم العينة تصبح $\hat{s}^2$ قريبة من σ^2 ونتيجة لذلك يقترب توزيع t من التوزيع الطبيعي المعياري. وعندما يؤول حجم العينة n الى مالا نهاية $(n \to \infty)$فان $\hat{s}^2$ تقترب كثيرا من σ^2 وبالتالي ينتفي الفرق بين التوزيعين.

2.2 توزيع χ^2 (χ^2 Distribution)

توزيع χ^2 مثله مثل توزيع ت (t) يعتبر متغير عشوائي وإحصاءة الاختبار (Test Statistic) وله استخدامات معينة في التطبيقات الإحصائية. وقد أوضحنا أن توزيع t يستخدم في اختبار معنوية وبناء فترات ثقة لمقاييس النزعة المركزية مثل الوسط الحسابي و النسب (Proportions). ولكن في كثير من الأحيان قد يواجه الباحث باختبار معنوية تباين التوزيع. وعليه أحد استخدامات توزيع χ^2 هو اختبار معنوية

التباين (Variance). ويستخدم التوزيع أيضا في اختبار جودة التوفيق (Goodness of Fit). وهو اختبار فرض يتعلق بالفرق بين التكرارات المشاهدة (Observed Frequencies) لعينة و التكرارات المتوقعة (Expected Frequencies) المقابلة لها بناء على توزيع نظري (Theoretical Distribution). واستخدام أخر لتوزيع χ^2 هو اختبار الاستقلال الإحصائي (Statistical Independence) لمتغيرين تصنيفيين وفي بعض الأحيان يسمى اختبار بيرسون للارتباط (Pearson's Test for Association) وله علاقة وثيقة باختبار جودة التوفيق.

وتوزيع χ^2 له علاقة وثيقة بالتوزيع الطبيعي المعياري z مثل توزيع t. وشكل بيان توزيع χ^2 يتغير كلما زاد حجم العينة n. وينبغي توضيح هذه العلاقة بين z , χ^2. فإذا كان المتغير العشوائي x يتبع نمط التوزيع الطبيعي بوسط u وتباين σ^2 أي $Z\sim N\left(u,\sigma^2\right)$ فان المتغير المعياري

$$Z=\frac{x-u}{\sigma}\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(3)$$

يتبع نمط التوزيع الطبيعي بوسط صفر وتباين واحد صحيح أي $Z\sim N$ (0و1) وإذا قمنا بتربيع طرفي المعادلة (3) فهل مربع z (أي z^2) يتبع ايضا التوزيع الطبيعي. الاجابة بالنفي وعليه فان مربع z يتبع توزيع χ^2 بدرجة حرية مساوية للواحد:

$$Z^2 = \frac{(x-u^2)}{\sigma^2} \sim \chi^2_{(1)}$$

و المتغير χ^2 له مجموعة كبيرة من التوزيعات وكل توزيع مقابل لدرجة حرية معينة. ودرجات حرية توزيع χ^2 تساوي حجم العينة n. وبصفة عامة فإن

$$\sum_{i=1}^{n} Z^2 = \sum_{i=1}^{n} \frac{(X_i - u)^2}{\sigma^2} \sim \chi_n^2 \quad \text{.....(4)}$$

تتبع نمط توزيع χ^2 بدرجات حرية تساوي n ودالة كثافة توزيع χ^2 كما **يلي:**

$$f(\chi^2) = k(\chi^2)^{\left(\frac{\vee}{2}-1\right)} e^{-(x^2)/2}, \quad \chi^2 > 0$$

حيث أن:

e = 2.71828

v = درجات الحرية

k = ثابت يعتمد فقط على v

ويعتمد توزيع χ^2 على معلمة واحدة هي درجات الحرية v. وعندما ما تكون درجات الحرية صغيرة فان شكل بيان توزيع χ^2 ذو التواء موجب (Positively Skewed) وكلما زادت درجات الحرية v كلما قل الالتواء نحو اليمين واقترب البيان من التماثل. ويكاد ان يكون متماثل عندما تكون درجة الحرية v تساوي الواحد (v = 1). واما

عندما تكون درجات الحرية كبيرة فان توزيع χ^2 يقترب من التوزيع الطبيعـــي. وفـــي كــل الأحـــوال فـــان التوزيـــع χ^2 متصـــل (Continuous) وذو منـــوال وحيـــد (Unimodal) . ووسط توزيع χ^2 يساوي درجات الحرية v وتباينه يساوي 2v.

3.2 توزيع F (F-distribution)

توزيع F مثله مثل توزيع t وتوزيع χ^2 يعتبـــر متغيـــر عشـــوائي وإحصـــاءة اختبار ويعرف بانه نسبة بين تقديرين مستقلين غير متحيزين لتبايني مجتمعين:

$$F = \frac{\hat{S}_1}{\hat{S}_2} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (5)$$

وتستخدم هذه النسبة لاختبار فرض عن نســـبة تبـــاينين اثنـــين $(\sigma_1^2 = \sigma_2^2)$. وتستخدم أيضا لاختبار فرض تساوي ثلاثة أوساط أو أكثر. وفي بعـــض الأحيـــان قـــد يواجه الباحث بالمقارنة بين تباينين لمجتمعين اثنين لتحديد فيما إذا كان تبـــاين أحـــدهما يختلف عن تباين المجتمع الأخر. ولإلقاء الضوء على طبيعة توزيـــع F، لنفتـــرض ان لدينا متغيرين عشوائيين مستقلين X_1و X_2 وكل واحد منهما يتبع نمط التوزيع الطبيعي أي: $X_1 \sim N(u_1, \sigma_1^2)$. $X_2 \sim N(u_2, \sigma_2^2)$ وإذا أخذنا عينتين من x_1 و x_2 بحجم n_1 و n_2 على التوالي، ثم حصلنا على تقدير غير متحيز لتبايني المجتمعين s_1^2 , s_2^2، فـــان العلاقات التالية تتحقق:

$$\hat{S}_1^{\,2} = \frac{\sigma_1^{\,2}\chi^{2}_{(n_1-1)}}{n_1-1}$$

$$\hat{S}_2^{\,2} = \frac{\sigma_2^{2}\chi^{2}_{(n_2-1)}}{n_2-1}$$

وبما أن $F = \frac{\hat{S}_1^{\,2}}{S_2^{\,2}}$ تحت الفرضية $\sigma_1^{\,2} = \sigma_2^{\,2}$ ، فان :

$$F = \frac{\chi^{2}_{(n_1-1)}/(n_1-1)}{x^{2}_{(n_2-1)}/(n_2-1)} \quad \text{.....(6)}$$

و النسبة F هي عبارة عن نسبة بين توزيعين مستقلين لـــ χ^2 كل مقسوم علـــى درجات الحرية الخاصة به كما توضح المعادلة (6).

ودالة كثافة توزيع F كما **يلي:**

$$f(F) = k\,F^{(V_1/2)-1}\left(1+\frac{\vee_1 F}{\vee_2}\right)^{-(\vee_1+\vee_2)/2} \quad , F > 0$$

حيث أن:

$v_1 = n_1 - 1$ و $v_2 = n_2 - 1$ وهي درجات الحرية في البسط و المقام للنســبة F على التوالي. k ثابت يعتمد فقط على v_1، v_2. وتوزيع F له معلمتـــان فقــط v_2, v_1. ويمكن توصيفه بالكامل بمعرفة المعلمتين.

وبما إن البسط و المقام عبارة عن مقدارين مربعين في النسبة F، فهذا يعنى أن أية قيمة يأخذها المتغير F أما موجبة او تساوي الصفر. ومدى قيم F يمتد من الصفر الى ما لانهاية ($F \geq 0$) وشكل بيان F يعتمد على المعلمين v_2, v_1 وبصفة عامة فهو غير متماثل (Non symmetric) ملتو جهة اليمين. ولكن عندما يزيد مقدار احد المعلمتين او كليهما فان توزيع F يميل نحو التماثل. وسط توزيع F يساوي **التالي:**

$$E(F) = \frac{\vee_2}{\vee_2 - 2}$$

وأما تباين F فهو:

$$E\left\{ \left(F - E\left(F\right) \right)^2 \right\} = \frac{2 \vee^2_2 \left(\vee_1 + \vee_2 - 2 \right)}{\vee_1 \left(\vee_2 - 2 \right)^2 \left(\vee_2 - 4 \right)}$$

و يتضح عندما تكون $v_2 = 2$ فانه لا يعرف له وسط أو تباين وعند $v_2 = 4$ لا يعرف تباين لتوزيع F .

3. فترات الثقة واختبارات الفروض للعينات الصغيرة

سنشرع الآن في توضيح تطبيقات التوزيعات الثلاثة في كيفية بناء فترات ثقة واختبارات الفروض لبعض معالم المجتمعات. وينبغي ان نشير الى نقطتين مهمتين عند

اتخاذ قرارات مبنية على عينات صغيرة (1) تؤدي العينة الصغيرة دائما الى خطأ معياري كبير بسبب التقلبات الكبيرة في عملية المعاينة. وعليه، تعطي العينة الصغيرة معلومات عن المجتمع لا يمكن ان نصفها بانها قاطعة مقارنة بالعينة الكبيرة.

(2) تقدير الخطأ المعياري على اساس عينة صغيرة يختلف بصورة اكبر من القيمة الحقيقية وبالتالي تزداد درجة عدم التأكد (Uncertainty) ومن ثم تتسرب الى حساباتنا وتجعل نتائج العينة الصغيرة اقل دقة.

1.3 تقدير فترة للوسط الحسابي للمجتمع μ عندما يكون الانحراف المعياري σ مجهولاً

Estimation of μ when σ is unknown

للحصول على تقدير الفترة (Interval estimate) لوسط المجتمع عندما يكون الانحراف المعياري للمجتمع مجهولاً وعندما يكون حجم العينة اقل من 30 فان بناء فترة الثقة يتم بنفس الطريقة التي سبق شرحها ماعدا ان التوزيع هو توزيع t بدرجة حرية محددة. وبصورة عامة يمكن كتابة حدود الثقة (Confidence Limits) **كما يلي:**

$$\text{Confidence Limits} = \bar{x} \pm t_{\alpha/2} S_{\bar{x}}$$

◄◄ مثال

عينة عشوائية حجمها 16 لها وسط يساوي 53. ومجموع مربع انحراف القـيم من الوسط الحسابي يساوي 135. أوجد

(أ) حدود ثقة 95%.

(ب) حدود ثقة 99 % لوسط المجتمع.

الحل:

$$\bar{x} = 53$$

$$n = 16$$

$$\sum\left(x - \bar{x}\right)^2 = 135$$

$$S = \sqrt{\frac{\sum\left(x - \bar{x}\right)^2}{n-1}} = \sqrt{\frac{135}{15}} = \sqrt{9} = 3$$

$$df = \vee = n - 1 = 16 - 1 = 15$$

(أ) 95% فترة ثقة لوسط المجتمع

$$\bar{x} \pm t_{\alpha/2}\frac{S}{\sqrt{n}} = 53 \pm (2.13)\frac{3}{\sqrt{16}} = 53 \pm 1.6$$

أذن فترة ثقة 95% للمتوسط هي :

$$\therefore \quad 51.4 < u < 54.6$$

(ب) فترة ثقة 99% لوسط المجتمع

$$\bar{x} \pm t_{\alpha/2} \frac{S}{\sqrt{n}} = 53 \pm (2.95)\frac{3}{\sqrt{16}} = 53 \pm 2.213$$

$$50.787 \quad < \quad u \quad < \quad 55.213$$

وترجمة فترة الثقة باستخدام توزيع t تماثل تلك باستخدام التوزيع الطبيعي. وفي المثال السابق في الجزء (أ) مثلا يمكن ترجمة النتيجة كما يلي: [الفترة 51.4 و 54.6] وتشمل μ ويقيس الاحتمال 0.95 ثقتنا في صحة ما تم بيانه بين قوسين مربعين.

ولكن من الضرورة الإشارة إلى نقطتين فيما يختص بفترات الثقة للعينات الصغيرة وتقدير الأخطاء المعيارية (Standard Error). أولا، مدى فترة الثقة للعينة الصغيرة واسع واقل موثوقية (Precision) من فترات الثقة للعينات الكبيرة. وأحد الأسباب هو أن العينة الصغيرة لها خطأ معياري كبير الأمر الذي يترتب عليه إدخال عنصر عدم التأكد بصورة اكبر في تقدير $S_{\bar{x}}$ ولكن عندما يكون حجم العينة كبيراً، فان s تعطى تقدير افضل للانحراف المعياري للمجتمع σ ومن ثم ينظر الى $S_{\bar{x}}$ بأنها تقدير يعتمد عليه لـ $\sigma_{\bar{x}}$. وهذا هو السبب الرئيسي

لاستخدام التوزيع الطبيعي للاستدلال حتى عندما تكون σ مجهولة طالما أن حجم العينة كبير. وثانيا، اذا كانت σ معلومة فان فترات الثقة لعدد مختلف من العينات سيكون لها نفس الطول بالرغم من اختلاف نقاطها المركزية (mid-point).

وهذا لانه عندما تكون σ معلومة، فان الخطأ المعياري (Standard Error) سيكون ثابت ولكن عندما تكون σ مجهولة فان $S_{\bar{x}}$ لا تكون ثابتة ولكنها تختلف من عينة إلى أخري. ونتيجة لذلك فان فترات الثقة المكونة لأوساط حسابية مختلفة محسوبة من عينات لها نفس الحجم سيكون لها ليس فقط نقاط مركزية مختلفة بل أيضا أطوال غير متساوية. وينبغي أن ننبه هنا عندما تكون σ مجهولة أو معلومة فان ترجمة فترة الثقة تكون بنفس الصورة التى تمت الاشارة اليها من قبل.

2.3 اختبار μ عندما يكون حجم العينة صغير

Testing About μ With Small Sample

لا تختلف طريقة اختبار معنوية وسط المجتمع μ عندما يكون حجم العينة اقل من 30 و σ مجهولة، فإننا نتبع نفس الخطوات التي تم شرحها من قبل ولكن باستخدام توزيع t كأساس لاتخاذ القرار.

◄◄ مثال

لنفترض أن فريق بحث طبي أعلن بان جهوده قد كللت بالنجـــاح فـــي تطـــوير علاج لضغط الدم المرتفع بحيث يمكن تخفيض ضغط المريض بأكثر من 20 نقطة. وبمعرفة هذه المعلومة، طبق طبيب هذا العلاج على 15 من مرضاه الذين يعـــانون من ضغط عالي في الدم. ونتيجة لذلك، وجد الطبيب أن ضغط الدم لمرضاه انخفض بـــ 21.5 نقطة في المتوسط وبانحراف معياري يساوي 1.5 نقطة. ويرغب الطبيب في معرفة إذا كان العلاج فعال كما يدعي فريق البحث الطبي.

الحل:

1- الفرض: (Hypothsis)

$$H_0 : u = 20$$

$$H_1 : u > 20$$

2- مستوى المعنوية (Level of Significance)

$$\alpha = 0.01$$

3- إحصاءة الاختبار (Test Statistic)

$$t = \frac{\left(\bar{x} - u\right)}{S/\sqrt{n}}$$

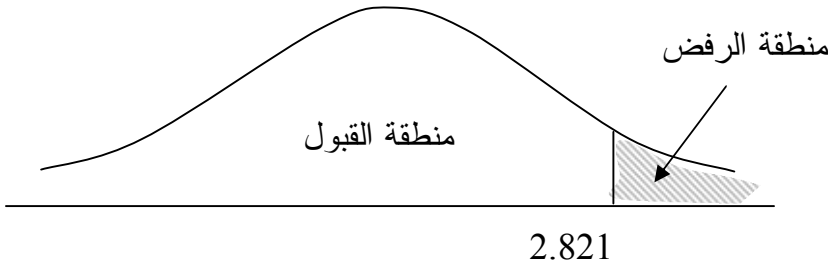
4- توزيع t (Distribution of t)

t تتبع نمط توزيع t بدرجة حرية تساوي 9 اذا كان فرض العدم صحيح.

5- المنطقة الحرجة (Critical Region)

إذا كانت درجات الحرية 9 و $\alpha = 0.01$، لاختبار من طرف واحـــد، فـــان قاعدة القرار كما **يلي:**

ارفض H_o اذا كانت $t > 2.821$، اقبل H_o اذا كانت $t < 2.821$



6- الحسابات (Computations)

$$n = 10 \ , \ \bar{x} = 21.5 \quad , \quad S = 1.5$$

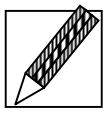
$$S/\sqrt{n-1} = 1.5/\sqrt{10-1} = 0.5$$

$$t = \frac{\bar{x} - u}{S/\sqrt{n-1}} = \frac{21.5 - 20}{0.5} = 3$$

7- القرار (Decision)

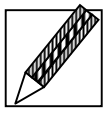
رفض فرض العدم H_o. وهذا يعنى ان الفرق 1.5 نقطة بين وسط العينـــة و القيمـــة الفرضية ذو دلالة إحصائية . وعليه الفرق معنوي احصائياً، وادعاء فريق البحـــث الطبي صحيح.

تدريب (1)



تم اختبار عينة عشوائية حجمها 10 من شحنة كبيرة لمصابيح كهربائية فإذا كانت لدينا البيانات التالية عن زمن حياة المصباح بالآلف الساعات 4.2 ، 4.6 و 3.9 ، 4.1 ، 5.2 ، 3.8 ، 3.9 ، 4.3 ، 4.4 ، 5.9. هل يمكننا قبول فرض أن وسط زمن حياة المصباح تساوي 4000 ساعة.

تدريب (2)



أستخدم نوعان من الدواء على 5 و 7 مرضي. النوع الأول A مستورد والنوع الثاني B محلي. حدث انخفاض في وزن المرضي بعد استخدام الدواء لفترة سنة كما يلي:

الدواء A :	10	12	13	11	14		
الدواء B :	8	9	12	14	15	10	9

هل هنالك فرق معنوي في فاعلية الدوائيين. إذا كان لا يوجد فرق معنوي أي الدوائيين يمكن شرائة.

3.3 المقارنة بين وسطين للمجتمع في حالة العينات الصغيرة

Comparing Two Population Means With Small Samples

بافتراض سحب عينتين مستقلتين من مجتمعين لهما نفس التباين، فانه يمكن استخدام توزيع t للعينة الصغيرة لاختبار معنوية الفرق بين وسطين.

عندما يكون الانحراف المعياري للمجتمع مجهول، تستخدم احصاءة الاختبار

التالية: $t = \dfrac{\bar{x}_1 - \bar{x}_2}{S_d}$

وهذه الإحصاءة تتبع نمط توزيع t بدرجة حرية $n_1 + n_2 - 2$ حيث ان s_d تمثل تقدير الخطأ المعياري للفرق بين وسطي عينتين ويمكن حسابه باستخدام البيانات الخاصة بالعينتين المستقلتين.

ويمكن الحصول على تقدير الانحراف المعياري المشترك عندما تكون لدينا عينتان بافتراض انهما مأخوذتان من مجتمعيين بنفس التباين كالأتي:

$$\hat{\sigma} = \sqrt{\frac{n_1 S_1^{\,2} + n_2 S_2^{\,2}}{n_1 + n_2 - 2}}$$

وعليه يمكن الحصول على S_d كما **يلي:**

$$S_d = \sqrt{\frac{\hat{\sigma}^2}{n_1} + \frac{\hat{\sigma}^2}{n_2}} = \hat{\sigma} \sqrt{\frac{n_1 + n_2}{n_1 n_2}}$$

◄◄ مثال

لنفترض ان مصلحة الضرائب بمدينة الخرطوم ترغب في معرفة اذا كان هناك فرق حقيقي بين متوسط دخل الحدادين ومتوسط دخل النجارين. وقد تم تكليف ادارة البحث بالمصلحة لتجميع البيانات التى ستمكنها من معرفة هذا الفرق. وبالفعل تم الحصول على البيانات **التالية:**

الحدادين	النجارين
$n_1 = 12$	$n_2 = 15$
$\bar{x}_1 = 6000$ دينار	$\bar{x}_2 = 5400$ دينار
$S_1 = 750$ دينار	$S_2 = 600$ دينار

وبناء على هذه المعلومات يمكن إجراء الاختبار كما يلي على أســـاس أن التبـــايين متساويان

1- الفرضيات (Hypotheses)

$$H_o : u = u_2$$

$$H_1 : u_1 \neq u_2$$

2- مستوى المعنوية (Level of Significance)

$$\alpha = 0.05$$

3- احصاءة الاختبار (Test Statistic)

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S_d}$$

4- توزيع t (Distruibution of t)

إذا كان المجتمعان لهما نفس الوسط ونفس الانحراف المعياري، فـــان النســـبة t تتبع نمط توزيع t بدرجات حرية: $25 = 2 - 15 + 12.$

5- منطقة الرفض (Rejection Region)

قاعدة القرار كما يلي:

ارفض H_o إذا كان $t \leq -2.06$ أو $t \geq 2.06$ وأقبل فرضية العدم إذا كان

$-2.06 < t < 2.06$

6- الحسابات (Computations)

$$\hat{\sigma} = \sqrt{\frac{n_1 S_1^{\,2} + n_2 S_2^{\,2}}{n_1 + n_2 - 2}} = \sqrt{\frac{(12)(750)^2 + (15)(600)^2}{12 + 15 - 2}} = 697$$

$$S_d = \hat{\sigma}\sqrt{\frac{n_1 + n_2}{n_1 n_2}} = 697\sqrt{\frac{12 + 15}{(12)(15)}} = 269.74$$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S_d} = \frac{6000 - 5400}{269.74} = 2.224$$

7- القرار (Decision)

تقع قيمة t المحسوبة في منطقة الرفض. وعليه نرفض فرض تساوي متوسط دخل الحدادين ومتوسط دخل النجارين عند مستوى معنوية 5 في المائة. ومن ثم يقبل الفرض البديل باختلاف متوسط دخل الحدادين ومتوسط دخل النجارين.

أسئلة التقويم الذاتي (1)



(1) عينة عشوائية حجمها 16 لها وسط يساوي 53 ومجموع مربع انحراف القيم من الوسط الحسابي هو 135. هل يمكن اعتبار ان هذه العينة مأخوذة ذو من مجتمع له وسط يساوي 56. أحصل على فترة 95% ثقة و فترة 99% لوسط المجتمع.

(2) في عينة عشوائية حجمها 500 شخص من مدينة الخرطوم وجد أن 200 يستهلكون القمح. وفي عينة عشوائية أخري حجمها 400 من مدينة امدرمان وجد أن 220 يستهلكون القمح. هل يوجد فرق معنوي في نسبة من يستهلكون القمح بين مدينتي الخرطوم وامدرمان.

4. اختبارات التباينات (Test of Variances)

ناقشنا فيما سبق اختبارات الفروض الخاصة بالاوساط الحسابية و النسب. ولاختبار هذه الفروض استخدمنا التوزيع الطبيعي المعياري وتوزيع t. وهذان التوزيعان لهما نفس الشكل الجرسي Bell Shapes) وكلتاهما متماثل حول الوسط صفر

وإحصاءات الاختبار تتم حساباتها بصورة متشابهة. والان سنتعرض لتوزيع مختلف تماما لاختبار الفروض الخاصة بتباين المجتمع او انحرافه المعياري. وكما تعلم فان الانحراف المعياري هو الجذر التربيعي للتباين. واذا عرفنا قيمة احدهما نستطيع ان نحصل على قيمة الاخر. ولهذا السبب كلما يقال عن التباين ينطبق ايضا على الانحراف المعياري.

والفرض الأساسي الذي يقوم عليه النقاش في هذا الصدد هو ان توزيع المتغير العشوائي x. طبيعي أي $x \sim N\left(u,\sigma^2\right)$. ويعرف تباين العينة ذات الحجم n كما يلي:

$$S^2 = \frac{\sum_{i-1}^{n}\left(x_1 - \bar{x}\right)^2}{n}$$

وإذا قسمنا طرفي المعادلة على σ^2 نحصل **على:**

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

وإذا ضربنا طرفي المعادلة بحجم العينة n

$$\frac{nS^2}{\sigma^2} = \sum\frac{\left(x_i - \bar{x}\right)^2}{\sigma^2}$$

نلاحظ في الطرف الايمن من المعادلة ان لدينا الوسط الحسابي للعينة $\bar{x}$ وليس وسط المجتمع u و السؤال الذي يطرح نفسه هل المتغير

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

يتبع نمط توزيع χ^2 بدرجات حرية n. والاجابة على السؤال هو ان المقدار يتبع نمط توزيع χ^2 إذا كانت x تتبع نمط التوزيع الطبيعي.و يمكن ببساطة أن نبرهن أن

$$\frac{nS^2}{\sigma^2} \qquad \chi^2_{n-1}$$

أي أن تباين العينة مضروب في n ومقسوم على تباين المجتمع يتبع نمط توزيع χ^2 بدرجات حرية تساوي (n-1) واستخدام $\bar{X}$ بدلا عن μ في المقدار يخفض درجات الحرية من n إلى (n-1). وهذا لان درجة حرية واحدة تفقد عندما تحل إحصاءة (Statistic) محل معلم (Parameter).

وإذا استخدمنا التقدير غير المتحيز (Unbiased) للتباين $\hat{\sigma}^2$ وذلك بقسمة مجموع مربع انحراف القيم من الوسط الحسابي بـ (n – 1) بدلا عن n فان

$$nS^2=\sum\left(x-\bar{x}\right)^2=(n-1)\hat{\sigma}^2$$

ونجد أن

$$\frac{(n-1)\hat{\sigma}^2}{\sigma^2}=\chi^2_{(n-1)}$$

ويمكن استخدام أيا من القاعدتين لاختبار معنوية تباين المجتمع على أساس تباين العينة.

1.4 اختبار فرضية تباين وحيد

(Test of Hypthesis About A Single Variance)

لنفترض إننا نرغب في تحديد إذا كان تباين مجتمع يتغير عبر الزمن أو تبـــاين التوزيع يختلف عن قيمة محددة. ولتحقيق هذه الرغبة نختبر فرض العدم التالي:

$$H_o : \sigma^2 = \sigma_o^2$$

حيث أن σ_o^2 تباين معلوم أو قيمة افتراضية. وتعتمد صياغة الفرض البـــديل على المشكلة موضوع الدراسة.

فإذا كانت x تتبع نمط التوزيع الطبيعي بتباين σ^2 واخترنا عينة حجمها n من مجتمع X أعطت تباين S^2 فانه يمكن استخدام إحصاءة الاختبار التالية:

$$\frac{nS^2}{\sigma_o^2}$$

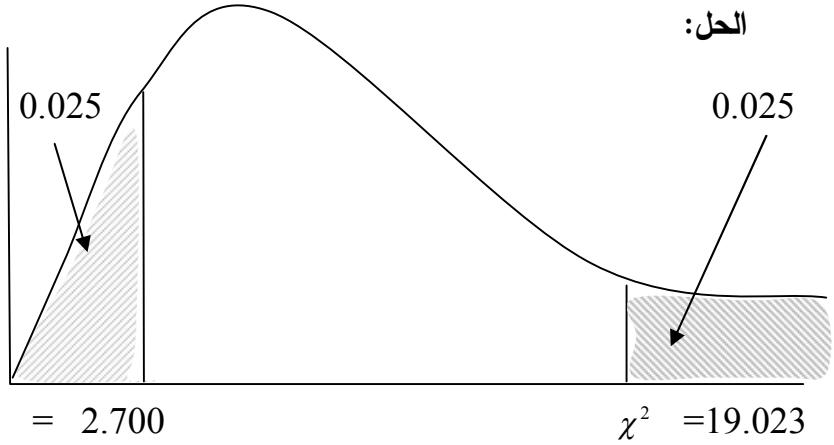
التي تتبع نمط توزيع χ^2 بدرجات حرية n-1 لتحديد إذا كان تباين المتغيـــر العشوائي x يساوي σ_0^2.

وبمجرد تحديد درجات الحرية ومستوى المعنوية α ونوع الاختبار (اختبار من طـــرف واحد، او من طرفين)، يمكن استخدام جـــدول χ^2 للحصـــول علـــى القـــيم الحرجـــة (Critical Values) لـــ χ^2 وأحد السمات الرئيسة لجدول χ^2 هو أن اية قيمة حرجـــة (Critical Value) تفصل المساحة نحو اليمين للقيم المعطـــاة فـــي الصـــف الأعلـــى للجدول.

◄◄ مثال

أوجد القيم الحرجة لتوزيع χ^2 التي تحدد المنطقة الحرجة (Critical Region) ذات المساحة 0.025 عند كل طرف. ولنفترض أن حجم العينة هو n وعليه تكون درجات الحرية هي 10−1 =9

الحل:

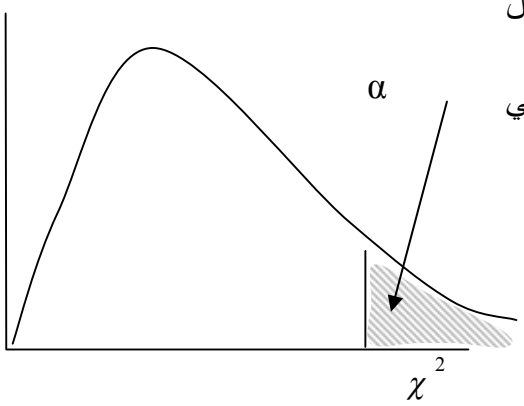


يمكن الحصول على القيمة الحرجة 19.023 يمين الشكل بصورة مباشرة وذلك بتحديد موقع درجات الحرية 9 في العمود الأول يمين جدول χ^2 و 0.025 على الصف الأعلى للجدول. وأما القيمة الحرجة 2.700 شمال الشكل يمكن الحصول عليها أيضا بتحديد موقع درجات الحرية 9 فى العمود الأول يسار جدول χ^2 ولكن ينبغي تحديد (0.975=0.025−1) على الصف الأعلى للجدول حيث أن القيم على الصف

الأعلى هي دائما مساحات (areas) إلى اليمين من القيمة داخل الجدول.عزيزي الدارس بما أن كثيراً من الطلاب يتعثر في استخدام جدول χ^2 لذلك سنوضح لك بالرسم أدنـــاه كيفية الحصول على القيم الحرجة في حالة الاختبارمن طرف واحد أو من طرفين .

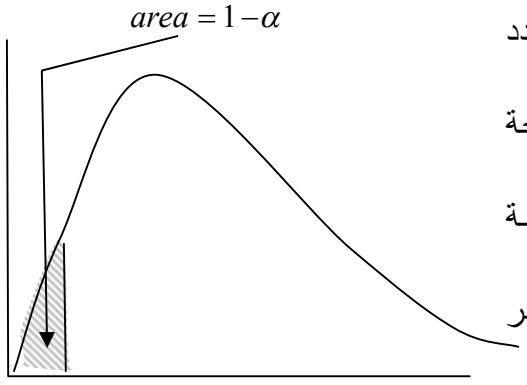
اختبار من طرف واحد **الشكل (أ) الطرف الأيمن**

للحصول على القيمة الحرجة χ^2 في الشـــكل (أ) أستخدم العمود مع المساحة α المعطاة في الصف الأعلى للجدول



اختبار من طرف واحد **الشكل (ب) الطرف الايسر**

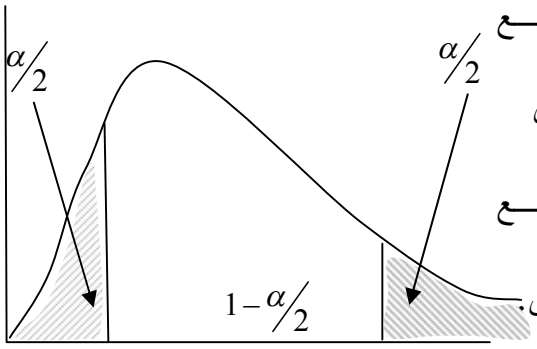
للحصول على القيمة الحرجـــة χ^2 ، حـــدد مساحة المنطقة إلى يمين الشكل أى المســـاحة غير المظللة واستخدام العمود مع هذه القيمـــة في الصف الأعلى. فإذا كان الطرف الأيســـر له قيمة α، استخدم العمود مع القيمة $1-\alpha$ في الصف الأعلى للجدول



اختبار من طرفين **الشكل (ج) الطرفان**

للحصول على χ_R^2 ،استخدام العمود مع المساحة $\alpha/2$ في الصف الأعلى للجدول

للحصول على χ_L^2 ، استخدم العمود مع المساحة $1-\frac{\alpha}{2}$ في الصف الأعلى للجدول.



χ^2_L χ^2_R

◄◄ مثال

ادعت علوية بان الدرجات المعيارية لاختبار الذكاء (IQ) له تباين يساوي 225 فإذا اختيرت عينة عشوائية مكونة من 41 فرداً وتم رصد الدرجات التى حققوها ووجد ان تباينها يساوي 258 . وضح كيف يمكن اختبار ادعاء علوية باستخدام مستوى معنوية 5 في المائة.

الحل:

يمكن اتباع نفس الخطوات التي تم شرحها من قبل

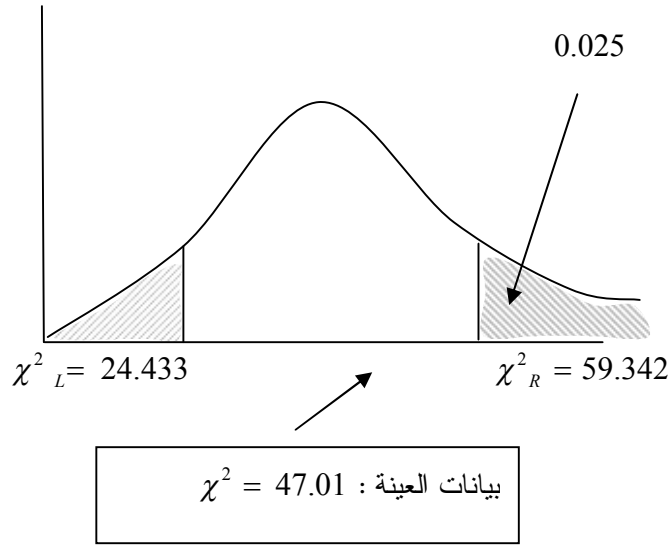
$$H_O : \sigma^2 = 225$$

$$H_1 : \sigma^2 \neq 225$$

ويتضح من صياغة الفرض البديل بان الاختبار من طرفين لان تباين العينـــة إذا كان اكبر أو اقل من 255 يؤدي إلى رفض فرض العدم. وقد تم تحديد مستوى المعنوية : $\alpha = 0.05$ وينبغي هنا استخدام تباين العينة لاختبار الادعاء عن تبـــاين المجتمـــع. ومن المناسب استخدام توزيع χ^2 لان درجات الذكاء (IQ) تتبع نمط التوزيع الطبيعـــي وعليه يمكن حساب إحصاء الاختبار كما يلي:

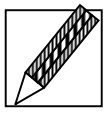
$$\chi^2 = \frac{(n)S^2}{\sigma^2} = \frac{(41)(258)}{225} = 47.01$$

ومن جدول توزيع χ^2 وباستخدام درجات حرية تساوي 40 ومستوى معنويـــة $\alpha = 0.05$ لاختبار من طرفين نرسم الشكل التالي:



وبما أن القيمة المحسوبة لإحصاءة الاختبار ($\chi^2$47.017=) لا تقع داخل منطقة الرفض (أو المنطقة الحرجة) فإننا نقبل فرض العدم حيث لا يوجد دليل كاف يسمح برفض الادعاء أن التباين يساوي 225 عند مستوي دلاله 5%.

تدريب (3)



أعطت عينة عشوائية حجمها 25 انحراف معياري مقداره 8.5. أختبر فرض أن الانحراف المعياري للمجتمع مقداره 10.

2.4 اختبار فرض عن نسبة بين تباينين

(Inference About the ratio of Two Variances)

يستخدم توزيع F – ضمن استخدامات أخرى – لاختبار فرضية نسبة تباين مجتمعين. وسيتم شرح هذا الاستخدام من خلال مثال رقمي.

◄◄ مثال

لنفترض أن شركة الروبي لتصنيع بطاريات السيارات ترغب في مقارنة طريقتين لإنتاج البطاريات من حيث التباين اختارت الشركة عشر بطاريات عشوائياً من كل طريقة فكانت أعمارها كما يلي:

طريقة الإنتاج A	طريقة الإنتاج B
2	3.7
2.1	3.9
2.5	3.9
3	3.9
3.3	4
4.2	4
4.2	4
4.3	4.1
6.8	4.2
7.6	4.3
$S_1^2 = 3.59$	$S_2^2 = 0.03$

وسنفترض أن لدينا مجتمعان مستقلان حيث يتبعا نمط التوزيع الطبيعي. وقد سحبت عينه حجمها n_1 من المجتمع الأول وعينة أخرى حجمها n_2 من المجتمع الثاني. وقد تم حساب تباين العينتين S_1^2 , S_2^2 لاستخدامها لاختبار فرض تساوي تباييني المجتمعين σ_1^2 , σ_2^2.

ويتضح من واقع بيانات الجدول أعلاه ان طريقتي الإنتاج تؤديان الى إنتاج بطاريات حيث بطاريات العينة المنتجة بالطريقة A تباينها اكبر. وعليه يظهر الفرق الرئيسي بين طريقتي الإنتاج من خلال المقارنة بين تباينها. ويعكس التباين الواضح بين

الطريقتين 3.59 و 0.03 الفرق بينهما ولكن ليس في كل الظروف والأحـــوال يتضـــح هذا الفرق من المقارنة بين التباينين. وعليه يصبح من الضرورة البحـــث عـــن طريقـــة اختبار موضوعية. وينبغي ترجيح الفرق بين التباينين بواسطة حجم العينتين لتحديد فيما اذا كان الفرق – حتى في حالة وضوحه التام – معنوياً إحصائيا ام لا.

وسنفترض ان S^2_1 تمثل تباين العينة الاكبر و S^2_2 تمثل تباين العينة الاصـــغر. وتوضح التطبيقات الاحصائية انه اذا كان لدينا مجتمعان يتبعان نمط التوزيـــع الطبيعـــي ولهما

نفس التباين ($\sigma_1^2 = \sigma_2^2$) فان توزيـــع المعاينـــة (Sampling Distribution) للنسبة:

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

يتبع نمط توزيع F .

فإذا كان المجتمعان لهما نفس التباين، فإن النسبة F تقترب من الواحد الصـــحيح حيث أن قيمتى S_1^2 و S_2^2 تقتربان من بعضهما. ولكن إذا كـــان تباينـــا المجتمعـــين يبتعدان عن بعضهما فإن قيمتي S_1^2 و S_2^2 تكونا مختلفتان إلى حد كبير. وباعتبـــار أن

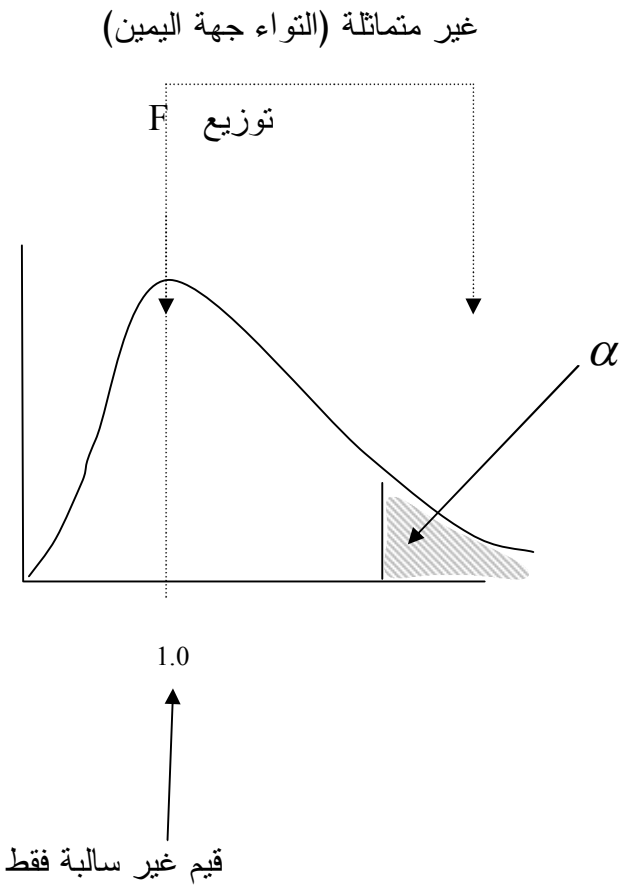
S_1^2 تمثل التباين الأكبر، فان النسبة S_1^2/S_2^2 تصبح كبيرة عندما تكون قيمتي S_1^2 و S_2^2 متباعدتان. ونتيجة لذلك، إذا كانت قيمة F تقترب من الواحد الصحيح فهذا يدل على أن $\sigma_1^2 = \sigma_2^2$. ومن ناحية أخرى، إذا كانت قيمة F تبتعد عن الواحد الصحيح فهذا يدل على أن تباييني المجتمعين غير متساويين. ويمكن الحصول على القيم الحرجة لتوزيع F من الجدول الخاص بها.

ويمكن الحصول على القيم الحرجة من جدول F إذا كان لدينا القيم التالية:

(1) مستوى المعنوية α
(2) درجات الحرية للبسط $(n_1 - 1)$
(3) درجات الحرية للمقام $(n_2 - 1)$

وينبغي التأكد عند استخدام الجدول بأن n_1 تقابل العينة ذات التباين S_1^2 و n_2 تقابل العينة ذات التباين S_2^2 . وينبغي أيضا تحديد مستوى المعنوية α وفيما إذا كان الاختبار من طرف واحد أو من طرفين. وللاختبار من طرف واحد نستخدم مستوى المعنوية α للطرف الأيمن وأما للاختبار من طرفين نقسم مساحة منطقة الرفض بالتساوي بين الطرفين الأيمن والأيسر. ويوضح الشكل التالي الاختبار من طرف واحد.

شكل يوضح اختبار من طرف واحد



يوجد توزيع F مختلف لكل زوجيين من درجات الحرية في البسط والمقام

وينبغي أن نشير إلى أن قراءة القيمة الحرجة من الجدول تتحدد بتقاطع العمـــود الذي يمثل درجات الحرية لتباين العينة S_1^2 (البسط) مع الصف الذي يمثـــل درجـــات

الحرية لتباين العينة S_2^2 (المقام) ، وأن توزيع F غير متماثل وليس له مركز يساوى الصفر كما في التوزيع الطبيعي وتوزيع t. وهذا يعنى أنه لا يمكن الحصول على القيم الحرجة للطرف الأيسر باستخدام سالب القيم الحرجة للطرف الأيمن. ويمكن الحصول على القيم الحرجة للطرف الأيسر بدلاً عن ذلك باستخدام مقلوب القيمة الحرجة للطرف الأيمن بعكس درجات الحرية حيث تصبح درجات حرية البسط هي درجات حرية المقام ودرجات حرية المقام هي درجات حرية البسط، أي:

$$F_{\alpha}(v_1, v_2) = \frac{1}{F_{1-\alpha}(v_2 v_1)}$$

حيث α < 0.5

◄◄ مثال

لاختبار الفرض من طرفين، نحصل فقط على القيمة الحرجة العليا (Upper critical value) . ولنرمز لهذه القيمة بالرمز F_R حيث يشير الدليل R إلى الطرف الأيمن (Right Tail). وللحصول على القيمة الحرجة الدنيا (Lower Critical value) F_L حيث يشير الدليل L إلى الطرف الأيسر (Left Tail) نعكس درجات الحرية ثم نأخذ مقلوب قيمة F المتحصل عليها من الجدول باستخدام درجات الحرية المعكوسة مثلاً أوجد القيم الحرجة F_L و F_R لاختبار فرض من طرفين حيث أن

(أ) $n_1 = 10$	و $n_2 = 10$	و $\alpha = 0.05$
(ب) $n_1 = 10$	و $n_2 = 7$	و $\alpha = 0.05$
(ج) $n_1 = 7$	و $n_2 = 10$	و $\alpha = 0.05$
(د) $n_1 = 25$	و $n_2 = 10$	و $\alpha = 0.02$
(هـ) $n_1 = 10$	و $n_2 = 25$	و $\alpha = 0.02$

الحل:

وبما أن درجات الحرية (n_1-1) و (n_2-1) نحصل على الإجابات التالية:

(أ) $F_R = 4.0260$	و $F_L = 0.2484$
(ب) $F_R = 5.5234$	و $F_L = 0.2315$
(ج) $F_R = 4.3197$	و $F_L = 0.1810$
(د) $F_R = 4.7290$	و $F_L = 0.3071$
(هـ) $F_R = 3.2560$	و $F_L = 0.2115$

◄◄ مثال

تقصى سمسار عقارات عن الفروقات بين أسعار بيع المنـــازل بـــين منطقتـــي الخرطوم وامدرمان وأدعى أن التباين في الأسعار بين المنطقتين متساو وضـــح كيـــف يمكن اختبار هذا الادعاء عند مستوى معنوية $\alpha = 0.05$ علماً بان لـــدينا البيانـــات التالية عن أسعار بيع المنازل بآلاف الدولارات للمنطقتين ولنفترض انه تـــم الحصـــول عليها من مجتمعين يتبعان نمط توزيع الطبيعي.

جدول لأسعار بيع افتراضية للمنازل بمنطقتي الخرطوم وامدرمان بآلاف الدولارات:

منطقة الخرطوم	منطقة امدرمان
270	115
107	136.9
148	121
125	164
127.5	175
125	128
126	147.5
109	147
113.5	105
147	163.75
167	115
	149.165
	120.5
	147
n = 11	n = 14
$\hat{\sigma}^2 = 2124$	$\hat{\sigma}^2 = 456$

ويتضح من واقع بيانات الجدول أن تباين العينة لأسعار بيع منطقة الخرطوم اكبر وعليه سنرمز لهذه المنطقة بالمجموعة (1) . ونرمز لمنطقة امدرمان بالمجموعة (2). ولاختبار فرض العدم.

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

في مواجهة الفرض البديل

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

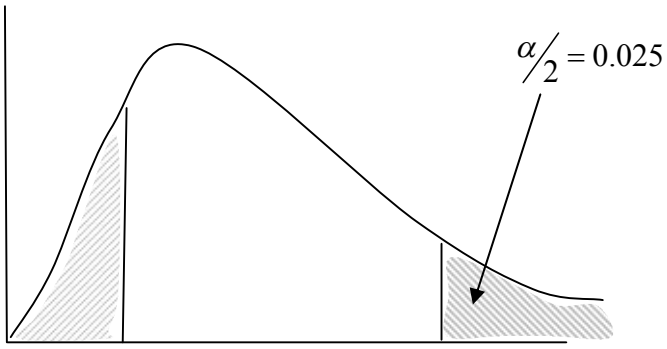
المجموعة (1)
$n_1 = 11$
$S_1^2 = 2122$

المجموعة (2)
$n_2 = 14$
$S_2^2 = 455$

وهذا اختبار من طرفين، وعليه نقسم مستوى المعنوية بين منطقتي الرفض اليمنى واليسرى حيث تكون مساحة كل منطقة 0.025 . وبأتباع العرف أن التباين الأكبر يوضع في البسط، نحصل فقط على القيمة الحرجة في الطرف الأيمن. وبما أن مساحة منطقة الرفض في الطرف الأيمن تساوى 0.025 وأن $n_1 = 11$ و $n_2 =$ 14 فان قيمة F الجدولية (القيمة الحرجة) التي تقابل درجات الحرية 10 في البسط ودرجات الحرية 13 في المقام هي 3.2497. والقيمة المحسوبة هي:

$$F = \frac{\hat{\sigma}^2}{\hat{\sigma}^2} = \frac{2124}{456} = 4.66$$

وبما أن القيمة المحسوبة F = 4.6637 اكبر من القيمة الجدولية F = 3.2497 ، نرفض إدعاء السمسار بأن التباين في أسعار بيع المنازل بمنطقتي الخرطوم وأمدرمان متساو عند مستوي دلالة 0.05



F=3.2497

بيانات العينة F = 4.6637

3.4 اختبار F وتحليل التباين
(F-test and Analysis of Variance)

تحليل التباين طريقة إحصائية صممت خصيصا لاختبار فرضية تساوى ثلاثة أو أكثر من الأوساط لمجتمعات تحت الدراسة. ويتكون تحليل التباين أساسا من تصنيف النتائج الإحصائية من جانب وتصنيفها تقاطعيا من جانب أخر لاختبار إذا كانت أوساط تصنيف معين تختلف معنويا. وبهذه الطريقة يمكن تحديد إذا كان تصنيف معين له أهمية في التأثير على النتائج المتحصل عليها.

وضحنا سابقا كيف يمكن اختبار الفرق بين وسطين باستخدام توزيع t عندما تكون لدينا عينتان. و السؤال الذي يطرح نفسه في هذا المقام كيف يمكن إجراء الاختبار عندما يكون لدينا ثلاثة عينات أو أكثر. وفي هذه الحالة ينبغي البحث عن طريقة بديلة لإجراء مثل هذا الاختبار، وتحليل التباين هو الطريقة المناسبة.

وينبغي أن ننبه هنا أن تحليل التباين الذي ستتم مناقشته ليس القصد منه اختبار معنوية الفرق بين تباين عينتين ولكن القصد منه هو اختبار معنوية الفروق بين أوساط العينات. ويستخدم توزيع F لاختبار معنوية نسبة تباينيين ولكن تم تصميم الاختبار

بحيث إن التباينات التي سيقارن بينها تختلف فقط في حالة إن الأوساط قيد الدراسة غير متجانسة. وبهذه الطريقة فان قيم F المعنوية تشير إلى إن الأوساط تختلف معنويا عن بعضها البعض.

ويقوم وتحليل التباين على الفرضيات التالية:

- التوزيع الطبيعي Normality
- التجانس Homogeneity
- استقلال الأخطاء Independence of Errors

ومن ناحية نظرية فان الإخلال بأي واحد من هذه الفروض يجعل طريقة تحليل التباين غير صالحة للاستخدام لأنها تؤدي إلى نتائج غير مرضية.

ولتسهيل العرض، سيتم فصل ونقاش طريقة تحليل التباين إلى طريقتين :

(أ) تصنيف الاتجاه الواحد (One-way Classification)

(ب) تصنيف الاتجاهين(Two-way Classification).

أ/ تصنيف الاتجاه الواحد (One-way Classification):

افرض أن لدينا مجتمعاً تصنف مفرداته إلي C صنفاً ويراد اختبار عما إذا كانت هذه الأصناف متساوية من حيث المتوسط

$H_0 : u_1 = u_2 = u_3 = \ldots\ldots\ldots\ldots = u_k$

علي الأقل متوسطان غيــر متســاويين

$H_1 :$

والطريقة المتبعة لاختبار هذا الفرض تسمي طريقة تحليل التباين فهي تبني علي الافتراضات الآتية:-

1- كل مشاهدة مكونة من 3 عناصر: الأول خاص بالصنف الذي تتبع له والثاني خطأ المشاهدة الخاص بالمفردة، إضافة لمتوسط عام μ دون اعتبار للصنف:

$$y_{ij} = \mu + e_{ij} \qquad i = 1,2,\ldots\ldots\ldots\ldots C.$$

$$j = 1,2,\ldots\ldots\ldots\ldots n_i$$

2- μ ثابت، α_i أثر الصنف رقم i ومتوسطها صفر.

3- الخطأ e_{ij} متغيرات عشوائية مستقلة بمتوسط صفر وتباين مشترك $\bar{\sigma}^2$.

وتعتمد هذه الطريقة علي تجزئة مجموع المربعات الكلي (SST) $SST = \sum_i \sum_j \left(y_i - \bar{\bar{y}}..... \right)^2$ إلى جزئين: الأول يقيس التباين بين متوسطات الاصناف

ويسمي مجموع مربعات الأصناف $(SSC)=\sum_i\sum_j\left(\bar{y}_i-\bar{y}.....\right)^2$ والثاني يقيس التباين الناتج عن الخطأ ويسمي مجموع مربعات الخطأ (SSE): $(SSE)=\sum_i\sum_j\left(\bar{y}_i-\bar{y}_i\right)^2$ لان

$$\sum_i\sum_j\left(y_{ij}-\bar{y}...\right)=\sum_i\sum_j\left[\left(y_i-\bar{y}...\right)+\left(y_{ij}-\bar{y}_i\right)\right]$$

$$\sum\sum\left(\bar{y}_i-\bar{y}.....\right)+\sum_i\sum_j\left(y_{ij}-\bar{y}_i\right)^2 \quad (1)$$ لأن الحد الأوسط يساوى الصفر، حيث

y_{ij} المشاهدة رقم j في الصنف i

$\bar{y}_i$ متوسط مشاهدات الصنف رقم i

$\bar{y}..$ المتوسط العام لكل المشاهدات.

الجانب الأيسر في (1) يتناسب مع التباين الكلي بين المشاهدات بينما الحد الأول في الجانب الأيمن يتناسب مع التباين للمتوسطات والحد الثاني فى الجانب الأيمن يتناسب مع تباين الخطأ في المشاهدات

من الناحية العملية تحسب مجاميع المربعات هذه كالاتي:

$$SST=\sum_i\sum_j y_{ij}^2-\frac{y^2..}{N} \quad (2)$$

$$SSC=\sum\sum\frac{y_i^2}{n_i}-\frac{y^2..}{N} \quad (3)$$

الفرق $SSC = \Sigma\Sigma y_{ij}^2 - \Sigma\Sigma \frac{y_{ij}^2}{n_i}$

$\left(\frac{y_{....}^2}{N}\right)$ يسمي معامل التصحيح. حيث n_i عدد المشاهدات فى الصنف رقم i ، $y.., \sum_i n_i = N$ المجموع الكلي و y_i مجموع المشاهدات فى الصنف رقم i نلاحظ أنه اذا كانت متوسطات الاصناف متساوية فانها تساوي المتوسط العام وبذلك SST_r = الصفر.

لتحديد معنوية الفروقات بين متوسطات الاصناف فان نقارن SST_r مع SSE بعد قسمة كل منها علي عدد درجات حريته (C-1 , N-C على التوالي) هذه النسبة تتوزع حسب توزيع ف بدرجتي حرية (K-1) و (N-K) :-

$$F = \frac{SSC/(C-1)}{SSE/(N-C)}$$

وتقارن هذه مع القيمة الحرجة من جدول F حسب مستوي المعنوية المحدد، فإذا كانت قيمة F المحسوبة اقل من قيمة F الجدولية فهذا يعنى ان الفرق غير معنوي وصادر اساساً من تقلبات الصدفه في المعاينة.

وعادة يتم تلخيص جدول تحليل التباين في جدول يسمى جدول تحليل التباين ويرمـــز له اختصاراً بالرمز ANOVA كما موضح بالشكل التالي:

جدول تحليل التباين (ANOVA)

Source of Variation	Sum of Squares SS	Degrees of Freedom DF	Mean Squares MS	Variance Ratio F
مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	نسبة التباين F
Between Samples (بين العينات)	SSC	$V_1 = c - 1$	MSC=SSC/(C-1)	MSC/MSE
Within Samples (داخل العينات)	SSE	$V_2 = n - c$	MSE=MSE/(n-C)	
Total الجموع	SST	n -1		

حيث أن:

SST : مجموع المربعات الكلي

SSC : مجموع المربعات بين الأصناف

SSE : مجموع المربعات داخل العينات (الخطأ).

MSC : متوسط مجموع المربعات بين العينات.

MSE: متوسط مجموع المربعات داخل العينات.

وينبغي أن ننبه إلى أن نفس طريقة تحليل التباين تنطبق في حالتي العينات ذات الحجـــم المتساوي أو غير المتساوي.

◄◄ مثال

لنفترض أن باحث يرغب في تقويم وتباين الأداء لمقرر معين يدرس في مدرسة ذات أربعة انهر (1) ,(2), (3), (4) فإذا تم سحب عينة عشوائية مكونة من خمسة درجات من كل نهر لهذا المقرر فكانت **كما يلي:**

(4)	(3)	(2)	(1)
14	13	9	7
8	11	10	6
11	16	13	11
15	7	7	9
12	8	6	7

المطلوب:هل الأداء في هذه الأنهر متساو في المتوسط ؟

الحل:

	العينة (1) X_1	العينة (2) X_2	العينة (3) X_3	العينة (4) X_4
	7	***9***	13	14
	6	10	11	8
	11	13	16	11
	9	7	7	15
	7	6	8	12
المجموع	**40**	**45**	**55**	**60**
$\bar{X}$	**8**	**9**	**11**	**12**

العينة (1)		العينة (2)		العينة (3)		العينة (4)	
x_1^2	x_1	x_2^2	x_2	x_3^2	x_3	x_4^2	x_4
49	7	81	9	169	13	196	14
36	6	100	10	121	11	64	8
121	11	169	13	256	16	121	11
81	9	49	7	49	7	225	15
49	7	36	6	64	8	144	12
$\sum x^2_1$	$\sum x_1$	$\sum x^3_2$	$\sum x_2$	$\sum x_3^2$	$\sum x_3$	$\sum x_4^2$	$\sum x_4$
=336	=40	=435	=45	= 659	=55	=750	=60

مجموع كل المفردات للعينات المختلفة (T) = $\Sigma X_4 + \Sigma X_3 + \Sigma X_2 + \Sigma X_1$

60+55+45+40+=200

معامل التصحيح (Correction Factor) = $\frac{T^2}{N} = \frac{(200)^2}{20} = \frac{40000}{20} = 2000$

مجموع المربعات الكلي (Total Sum of Squares) =

$$\sum x^2_1 + \sum x_2^2 + \sum x_3^2 + \sum x_4^2 - \frac{T^2}{N} = 336 + 435 + 659 + 750 - 2000$$

$$= 2180 - 2000 = 180$$

مجموع المربعات بين العينات (Sum of Squares Between Samples) =

$$\frac{(\sum x_1)^2}{N}+\frac{(\sum x_2)^2}{N}+\frac{(\sum x_3)^3}{N}+\frac{(\sum x_4)^2}{N}-\frac{T^2}{N}=\frac{(40)^2}{5}+\frac{(45)^2}{5}+\frac{(55)^2}{5}+\frac{(60)^2}{5}-2000$$

$$=\frac{1600}{5}+\frac{2025}{5}+\frac{3025}{5}+\frac{3600}{5}-2000$$

=320+405+605+720-200

= 2050-2000 = 50

مجموع المربعات داخل العينـــات (Sum of Squares Within Samples

= مجموع المربعات الكلى – مجموع المربعات بين العينات= 180−50= 130

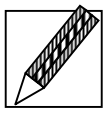
جدول تحليل التباين

مصدر التباين	درجـــــات الحرية	مجمـــــــوع المربعات	متوسط المربعات	ف F
بين الأصناف	3	50	16.667	2.051
الخطأ	16	130	8.125	–
المجموع	19	180	–	–

$$F_{3,16}(0.05)=8.69$$

وبما أن قيمة ف (F) المحسوبة (2.051) أقل من قيمة ف (F) الجدولية (8.69) فإننا نقبل فرض العدم عند مستوي دلالة 5%.

تدريب (5)



تم الحصول علي ثلاثة عينات من مجتمعات تتبع نمط التوزيع الطبيعي ولها تباينات متساوية. أختبر فرض تساوي أوساط الثلاثة عينات.

العينة (1)	العينة (2)	العينة (3)
8	7	12
10	5	9
7	10	13
14	9	12
11	9	14

ب/ تصنيف الاتجاهين (Two-way Classification

يمثل التحليل ذو الاتجاه الواحد والذي تم شرحه أعلاه ابسط نموذج لتحليل التباين. والآن سنتناول بالشرح نموذج آخر وهو نموذج تحليل التباين ذو الاتجاهين. وهذا النموذج معقد بعض الشيء مقارنة بالنموذج الأول.

في كثير من الأحوال قد نرغب في اختبار تأثير نوعان من العوامل (Factors) أو المعالجات (Treatments). فمثلا في بحث لمعرفة اتجاهات تسويق سلعة معينة قد نصنف موظفي البيع طبقا لأعمارهم ومستويات تعليمهم لمعرفة فيما إذا كان هذين العاملين لهما تأثير على حجم المبيعات. وقد ينصب اهتمامنا ليس فقط لمعرفة تأثير كل عامل على حدة على حجم البيع ولكن أيضا التأثير المشترك (Joint effect) لهذين العاملين على حجم البيع. ويطلق على هذا التأثير المشترك لفظة تفاعل (Interaction) حيث أن فئة عمرية معينة قد تعمل أفضل عندما تتمتع بمستوى تعليمي معين.

وسنحصر دراستنا لهذا النموذج على الفروقات فقط الخاصة بالتصنيفات الموصوفة لكل نوع من العوامل تحت الدراسة، بمعنى إننا سوف لا نهتم كثيرا بالتصنيفات الأخرى التي يمكن تضمينها. وسنفترض انه يمكن الحصول على مشاهدتين أو أكثر لكل عامل مشترك بمعنى إذا كانت لدينا مجموعة مكونة من عاملين فانه يمكن الحصول على مشاهدة لكل نوع من العاملين. وهذا قد يمكننا من إجراء اختبار لوجود تأثير التفاعل. وينبغي إن ننبه هنا إذا تم الحصول على مشاهدة واحدة فقط للخلية الواحدة، فانه لا يمكن التعرف على تأثير التفاعل لأنه سيكون ضمن تأثير الخطأ العشوائي. وعليه عندما تكون لدينا مشاهدتين أو أكثر للخلية الواحدة سنتمكن من حساب مجموع مربعات التفاعل الذي يترتب عليه إمكانية فصل تأثير التفاعل من تأثير الخطأ العشوائي.

في هذا الكتاب تقتصر دراستنا للحالات التي ليس فيها تفاعلاً بين العاملين.

تصنف البيانات في حالة تحليل التباين ذو الاتجاهين طبقا لمعيارين أو عـــاملين. وتختلف طريقة تحليل التباين بعض الشيء عن طريقة تحليل التباين ذو الاتجاه الواحـــد. ويأخذ جدول تحليل التباين ذو الاتجاهين الشكل **التالي:**

جدول تحليل تباينين ذو اتجاهين

مصدر التباين Source of variation	مجموع المربعات Sum of Squares	درجات الحرية Degrees of freedom	متوسط مجموع المربعات Mean Sum of Squares	نسبة F ratio of F
بين الأعمدة between columns	SSC	(c-1)	MSC =SSC/(c-1)	F_c = MSC/MSE
بين الصفوف (between Rows)	SSR	(r-1)	MSR=SSR/(r-1)	F_r = MSR/MSE
البـــــاقي أو الخطـــــأ (Residual or Error)	SSE	(c-1)(r-1)	MSE=SSE/(c-1)(r-1)	
المجموع (Total)	SST	N-1		

حيث الصفوف تمثل مستويات أحد العاملين والأعمدة تمثل مستويات العامل الثاني.

حيث أن:

SSC: مجموع المربعات بين الأعمدة بدرجات حرية (c-1)

SSR: مجموع المربعات بين الصفوف بدرجات حرية (r-1)

SSE: مجموع مربعات الأخطاء بدرجات حرية (c-1)(r-1)

SST: مجموع المربعات الكلي بدرجات حرية (N-1)

ويمكن الحصول على مجموع مربعات الأخطاء او الباقي بطرح كل من مجموع المربعات بين الأعمدة وبين الصفوف من مجموع المربعات الكلي كما **يلي:**

$$SSE = SST - (SSC+SSR)$$

و المجموع الكلي لدرجات الحرية = N-1

حيث أن:

c: عدد الأعمدة

r: عدد الصفوف

ويكون حساب مجاميع المربعات كالآتي:-

معامل التصحيح = $CF = \frac{y_{...}^2}{N}$

$$SST = \sum_i \sum_j y_{ij}^2 - CF$$

$$SSC = \sum_i \sum_j \frac{y_{.j}^2}{r} - CF$$

$$SSR = \sum_i \sum_j \frac{y_{i.}^2}{c} - CF$$

الفرق = SSE

حيث $y_{i.}$ مجموع ملاحظات الصنف رقم i

$y_{.j}$ مجموع ملاحظات العمود رقم j

$y_{...}$ مجموع جميع الملاحظات

ومنها تحسب متوسطات المربعات وهي: مجموع المربعات / درجات الحرية

ثم يقارن متوسط مربعات عامل مع متوسط مربعات الخطأ بالقسمة لنجد F (المحسوبة) لتقارن مع القيمة الجدولية.

ومن جهة أخري يمكن الحصول على مجموع المربعات الكلي ومجموع المربعات بين الأعمدة ومجموع المربعات بين الصفوف بنفس الطريقة السابقة. وأما قيم F يمكن الحصول عليها **كالتالي:**

$$F_{(v_1,v_2)} = \frac{MSC}{MSE}$$

حيث أن:

$$(c - 1) = V_1$$
$$(r - 1)(c - 1) = V_2$$

و:

$$F_{(V_1,V_2)} = \frac{MSR}{MSE}$$

حيث أن:

$$(r - 1) = V_1$$
$$(r - 1)(c - 1) = V_2$$

وينبغي أن نشير هنا إلى أن V_1 قد لا تكون متساوية في كلتا الحالتين. في الحالة الأولي تساوي (c−1) وأما في الحالة الأخرى تساوي (r-1).

وتقارن القيم المحسوبة لـF مع القيم الجدولية. وإذا كانت القيمة المحسوبة اكبر من القيمة الجدولية عند مستوى معنوية معين، فترفض فرضية العدم، وبخلاف ذلك تقبل فرضية العدم.

وننبه إلى أن "الباقي" يلعب دور المقياس المعياري لاختبار المعنوية في تحليل التباين ذو الاتجاهين. وهو يمثل حجم التباين الناتج عن عوامل الصدفة (Chance Factors) ويوضح المثال التالي طريقة تحليل التباين ذو الاتجاهين:

◄◄ مثال

عينت شركة لإنتاج الزهور بإحدى ضواحي الخرطوم أربعة بائعين D, C, B, A وتمت مراقبة مبيعاتهم لفصول السنة الثلاثة: الصيف و الشتاء و الخريف. ويوضح الجدول أدناه بيانات المبيعات:

الفصول المناخية	البائعون				المجموع للفصول
	A	B	C	D	
الصيف	38	34	20	36	128
الشتاء	26	29	30	31	116
الخريف	23	27	31	29	110
المجموع للبائعين	87	90	81	96	354

الحل:

تم تصنيف البيانات طبقا لمعيار (أ) البائعون (ب) الفصول المناخيـــة، ولتبســـيط العملية الحسابية يمكن ترميز البيانات (Coding the Data) بطرح عدد ثابـــت مـــن كل مشاهدة وطرح عدد ثابت من كل من المشاهدات لا يغير في نتيجة التحليل .والقصد هو تبسيط العملية الحسابية:.

مجاميع المربعات:

معامل التصحيح $CF = \frac{354^2}{12} = 10443$

مجموع المربعات الكلي ، SST :(درجات الحرية = 12−1=11)

$38^2+ 34^2+20^2 ..... + 96^2 - CF = 10734-10443=291$

مجموع مربعات الصفوف: (درجات الحرية = 3−1=2)

$$\frac{128^2}{4}+\frac{116^2}{4}+\frac{110^2}{4}-CF=10485-10443=42$$

مجموع مربعات الأعمدة (درجات الحرية = 4−1=3)

$$\frac{96^2}{3}+\frac{81^2}{3}+\frac{90^2}{3}+\frac{87^2}{3}-CF=10482-10443=39$$

مجوع مربعات الخطأ = 291−42−49=210

درجات الحرية = 2×3 = 6 (الفرق).

ويمكن إفراغ هذه المعلومات في جدول تحليل التباين التالي:

جدول تحليل التباين

مصادر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	ف
بين الأعمدة (البائعون)	**39**	**3**	**13**	**1>**
بين الصفوف (الفصول)	**42**	**2**	**21**	**1>**
الباقى	**210**	**6**	**35**	
المجموع	**291**	**11**		

ولاختبار فرضية لا يوجد فرق بين مبيعات البائعين والفصول بمعنى أن التباينات الثلاثة المستقلة المقدرة تمثل تقديرات تباين من مجتمع مشترك (Common Population)

وللمقارنة بين تقدير تباين البائعين وتقدير تباين الباقي نحصل **على:**

$$F = \frac{13}{35} \quad \mathbf{1} >$$

وقيمة F الجدولية عند مستوى معنوية 5 في المائة ودرجات حرية $v_1 = 3$ و $v_2 = 6$ هي 4.76. ويتضح أن القيمة المحسوبة اقل من القيمة الجدولية ونستنتج أن مبيعات البائعيين لا تختلف معنوياً

ومن جهة أخري للمقارنة بين تقدير تباين الفصول مع تقدير تباين الباقي، نحصل على:

$$F = \frac{21}{35} \quad \mathbf{1} >$$

وقيمة F الجدولية عند مستوى معنوية 5 في المائة ودرجات حرية $v_1 = 2$, $v_2 = 6$ هي 5.14. ويتضح أن القيمة المحسوبة اقل من القيمة الجدولية وعليه لا يوجد اختلاف معنوي بين الفصول فيما يختص بالمبيعات.

أسئلة التقويم الذاتي (1)

(1) وجد في عينة عشوائية حجمها 8 مفردات أن مجموع مربع انحراف القيم من الوسط الحسابي مقداره 94.5 ووجد في عينة أخري حجمها 10 مفردات أن مجموع مربع الانحرافات يساوي 101.7. أختبر إذا كان الفرق معنوي باستخدام مستوي دلالة 5%.

(2) يعطي الجدول التالي إنتاج عينة مكونة من 15 حيازة لثلاثة أنواع من البذور:

(أ)	(ب)	(ج)
20	18	25
21	20	28
23	17	22
16	25	28
20	15	32

وضح إذا كان هناك فرق معنوي بين أوساط إنتاج الحيازات للثلاثة بذور

(3) يعطي الجدول التالي المبيعات الشهرية (بآلاف الدينارات) لمنشأة معينة في ثلاثة ولايات (الخرطوم والنيل الأزرق والنيل الأبيض) بواسطة أربعة بائعين:

الولاية	البائع (أ)	البائع(ب)	البائع(ج)	البائع(د)
الخرطوم	6	5	3	8
النيل الأزرق	8	9	6	5
النيل الأبيض	10	7	8	7

قم بتكوين جدول تحليل تباين لاختبار إذا كان هناك فرق معنوي:

(أ) بين مبيعات الأربع بائعين. (ب) بين المبيعات فى الثلاثة ولايات.

الخلاصة

عزيزي الدارس

إذا قمنا بتلخيص لأهم ما ورد في هذه الوحدة سنجد أننا قد عرفنا العينة الصغيرة بأنها التي تقل مفرداتها عن 30 والكبيرة عكس ذلك . كما وضحنا أنه في حالة العينات الصغيرة وعدم معرفة تباين المجتمع فإن التوزيع الطبيعي لا يصلح للاختبار وتبين لنا أن توزيع ت يقرب من التوزيع الطبيعي كلما زاد حجم العينة وان وسط توزيع ت هو الصفر وأن تباينه أكبر من الواحد و الشكل البياني له متماثل مثل التوزيع الطبيعي وان توزيع ت ليس وحيداً . ثم وضحنا أن توزيع χ^2 ذو التواء موجب وانه كلما كبرت درجات الحرية اقترب شكل توزيعه من التوزيع الطبيعي .

أما توزيع ف . فقد وضحنا أن شكله يعتمد على درجتي الحرية وملتوٍ جهة اليمين وأن قيمة ف لا تكون سالبة وأن مداه من صفر إلى ما لا نهاية.

وان بناء فترة الثقة للوسط الحسابي لمجتمع مجهول الانحراف المعياري أخذت من عينة صغيرة يتم بنفس طريقة التوزيع الطبيعي إلا أن التوزيع يكون توزيع ت . كما وضحنا انه في حالة العينات الصغيرة وتباين المجتمع مجهول يمكن أن نستخدم توزيع ت لكل من اختبار معنوية الوسط ، المقارنة بين وسطين . وأنه في حالة اختيار فرضية

تباين وحيد نستخدم توزيع χ^2 وفي حالة اختيار فرض عن نسبة تبـــاينين ف أو تحليـــل التباين فإن التوزيع المستخدم هو ف

لمحة مسبقة عن الوحدة التالية

تعرفك الوحدة بكيفية تحديد ما إذا كان هناك علاقة تربط بين متغيـــرين وكيـــف يمكن إيجاد هذه العلاقة وقياس درجة قوتها و نوع العلاقة؛ هـــل العلاقـــة طرديـــة ؟ أم عكسية ؟

إجابات التدريبات

تدريب (1)

$\bar{x} = 4.43$ ألف

$= 4430$

$$s^2 = \frac{(4.2-4.43)^2 + \ldots\ldots + (5.9-4.43)^2}{9}$$

$= \frac{3.921}{9}$ ألف $= 0.4356$ ألف

$= 436$

$S = 20.88$

الفرض $H_o : \mu = 4000$

$H_1 : \mu \neq 4000$

مستوى المعنوية $\alpha = 0.05$

إحصاء الاختبار $t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$

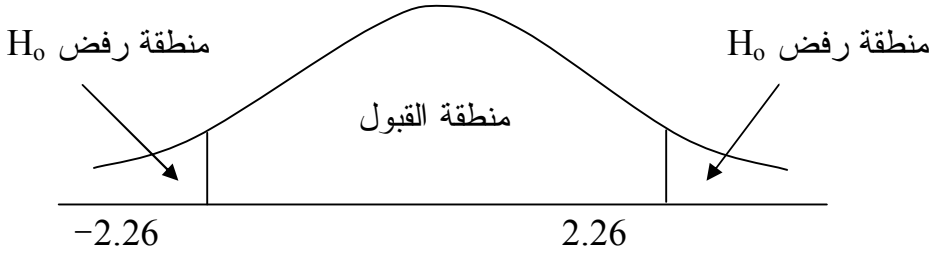
تتبع توزيع t بدرجة حرية q عند صحة فرض العدم

المنطقة الحرجة

الاختبار من طرفين ، $\alpha = 0.05$ ، درجة الحرية q

قاعدة القرار

أرفض H_o إذا كانت $|t| > 2.26$



الحسابات

$$n = 10 \quad , \quad S = 20.88$$

$$S/\sqrt{n-1} = \frac{20.88}{\sqrt{9}} = \frac{20.88}{3} = 6.92$$

$$t = \frac{\bar{x}-\mu}{S/\sqrt{n-1}} = \frac{4430-4000}{6.92} = \frac{430}{6.92} = 62.14$$

القرار يرفض فرض العدم أن وسط حياة المصباح يساوي 4000 ساعة ، وهذا يعني أن فرق 430 ساعة فرق معنوي .

تدريب (2)

$$\bar{x}_1 = 12 \quad , \quad n_1 = 5$$

$$\bar{x}_2 = 11 \quad , \quad n_2 = 7$$

$$s_1^2 = \frac{(10-12)^2 + ... + (14-12)^2}{4} = \frac{10}{4} = 2.5$$

$$s_1 = 1.58$$

$$s_2^2 = \frac{(8-11)^2 + ... + (9-11)^2}{6} = \frac{54}{6} = 9$$

$$s_2 = 3$$

$$\hat{\sigma} = \sqrt{\frac{n_1 s_1^2 + n_1 s_2^2}{n_1 + n_2 - 2}}$$

$$\hat{\sigma} = \sqrt{\frac{5(2.5)+7(9)}{5+7-2}} = \sqrt{\frac{75.5}{10}} = 2.75$$

الفرضيات

$$H_o : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

مستوى المعنوية

$$\alpha = 0.05$$

إحصاء الاختبار

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_d}$$

التوزيع t . باعتبار المجتمعان لهما نفس الانحراف المعياري النسبة t تتبع نمط توزيع t

بدرجة حرية 10= 2 – 7 + 5

منطقة الرفض

قاعدة القرار

أرفض H_o إذا كان $t < -2.23$ أو $t \geq 2.23$

وأقبل H_o إذا كان $-2.23 < t < 2.23$

الحسابات

$$s_d = \hat{\sigma} \sqrt{\frac{n_1+n_1}{n_1 n_2}} =$$

$$= 2.75 \sqrt{\frac{12}{35}}$$

$$= 1.61$$

$$t = \frac{\overline{x}_1 - \overline{x}_2}{s_d}$$

$$t = \frac{12 - 11}{1.61} = \frac{1}{1.61}$$

$$= 0.62$$

نضع قيمة t المحسوبة في منطقة القبول

إذن نقبل تساوي مفعول الدوائين عند مستوى معنوية $\alpha = 0.05$ ، أي لا يوجد فرق معنوي في فاعلية الدوائين عند مستوى معنوية $\alpha = 0.05$.

وعند ذلك نشتري الدواء المحلي .

تدريب (3)

الفرضيات

$$H_o : \sigma^2 = 100$$

$$H_1 : \sigma^2 \neq 100$$

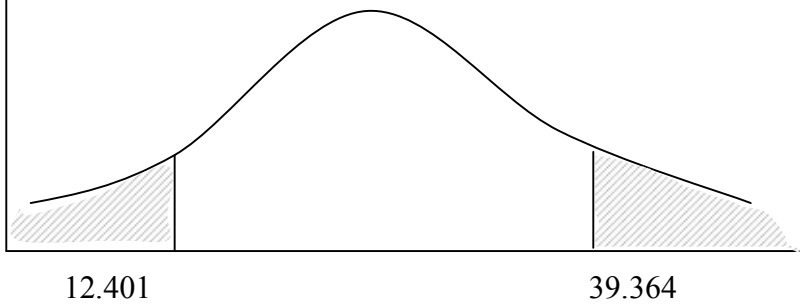
مستوى المعنوية

$$\alpha = 0.05$$

$$\chi^2 = \frac{n\, s^2}{\sigma^2} = \frac{25\ (8.5)^2}{100}$$

$$= 18.06$$

من جدول توزيع χ^2 وبدرجات حرية 24 ومستوي معنوية $\alpha = 0.05$



إذن نقبل فرض العدم بأن التباين يساوي 100 أي الانحراف المعياري يساوي 10، لأنه لا يوجد دليل كافي يسمح برفض فرض العدم H_o عند مستوى معنوية $\alpha = 0.05$.

تدريب (4)

الفرضيات

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

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

المتوسطات

$$\overline{x}_1 = 80 \quad ، \quad \overline{x}_2 = 83$$

التباين

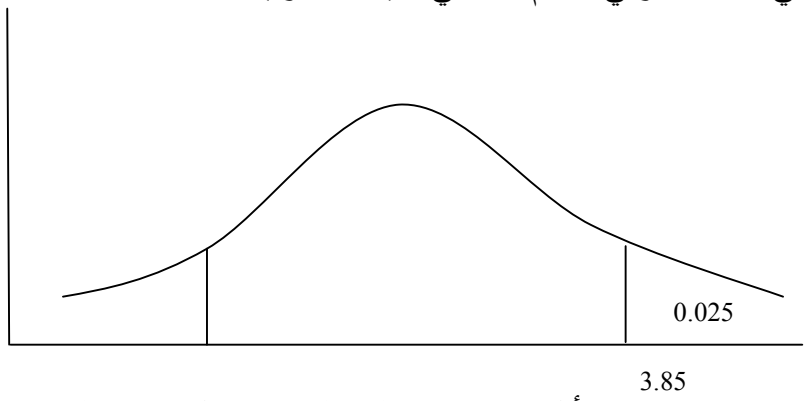
$$\hat{\sigma}_1^2 = 91.65$$

$$\hat{\sigma}_2^2 = 129.8$$

النسبة

$$F = \frac{129.8}{91.65} = 1.416$$

ومساحة منطقة الرفض التي تساوي 0.025 في الطرف الأيمن عندما تكون درجات الحرية في البسط 8 وفي المقام 10 هي القيمة الجدولية $F = 3.85$



وبما أن قيمة F المحسوبه أقل من قيمة F الجدولية إذن نقبل فرض العدم H_o أن المجتمعان لهما نفس التباين عند مستوى دلالة $\alpha = 0.05$.

تدريب (5)

العينة (1)	العينة (2)	العينة (3)
8	7	12
10	5	9
7	10	13
14	9	12
11	9	14
50	40	60
10	8	12

المجموع

المتوسط $\overline{X}$

العينة (1)		العينة (2)		العينة (3)	
X_1^2	X_1	X_2^2	X_2	X_3^2	X_3
64	8	49	7	144	12
100	10	25	5	081	9
049	7	100	10	169	13
196	14	81	9	144	12
121	11	81	9	196	14
530	50	336	40	734	60

المجموع

مجموع القراءات (المفردات)

$$T = 60 + 40 + 50 = 150$$

معامل التصحيح Correction Factor

$$C F = \frac{T^2}{N C} = \frac{150^2}{15} = 1500$$

حيث $N C$ هو عدد المفردات الكلي

مجموع المربعات الكلي

$$SST = \sum x_1^2 + \sum x_2^2 + \sum x_3^2 - C\,F$$

$$STT = 530+336+734 - 1500 = 1600-1500 = 100$$

مجموع المربعات بين العينات SSC = Sum of Squares between Samples

$$SSC = \frac{(\sum x_1)^2 + (\sum x_2)^2 + \sum (x_3)^2 - C\,F}{N}$$

حيث N عدد مفردات كل عينة

$$= \frac{50^2 + 40^2 + 60^2}{5} = \frac{2500+1600+3600}{5} = 1500$$

$$= \frac{7700}{5} - 1500 = 1540 - 1500 = 40$$

مجموع المربعات داخل العينات SSE = Sum of Squares within Samples

$$SSE = SST - SSC$$
$$= 100-40 = 60$$

جدول تحليل التباين ANOVA

مصدر التباين	درجات الحرية	مجموع المربعات	متوسط المربعات	F
بين الأصناف	12	40	20	
الخطأ	12	60	5	4
المجموع	14	100	—	

$$F_{2,\,12\,(0.05)} = 3.89$$

القيمة المحسوبة للنسبة F أكبر من القيمة الجدولية ، إذن يوجد اختلاف معنوي بين متوسطات العينات الثلاث

تدريب (6)

مجاميع المربعات

معامل التصحيح

$$CF = \frac{506^2}{9} = 28448$$

مجموع المربعات الكلي

$$SST = 57^2 + ... + 58^2 - CF$$

$$= 32491 + 3025 + 44891 + 2401 + 2704 + 4624 + 2916 + 2116 + 3364 - 28448$$

$$= 440$$

درجات الحرية

$$9 - 1 = 8$$

مجموع مربعات الصفوف (درجات حرية 2 = 1 − 3)

$$= \frac{179^2 + 169^2 + 158^2}{3} = C\,F = \frac{3204 + 28561 + 24964}{3} - 28448 = 74$$

مجموع مربعات الأعمدة (درجات حرية 2 = 1 − 3)

$$= \frac{160^2 + 153^2 + 193^2}{3} = C\,F = \frac{25600 + 23409 + 37249}{3} - 28448 = 305$$

مجموع مربعات الخطأ $440 - 74 - 305 = 61$

درجات الحرية (الفرق) $4 = 2 \times 2$

جدول تحليل التباين

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	F
بين الأعمدة (المنظفات)	305	2	152.5	F>1
بين الصفوف (الماء)	74	2	37	F>1
الباقي	61	4	15.25	
المجموع	440	8		

للمقارنة بين تقدير تباين المنظفات وتقدير تباين الباقي

$$F = \frac{152.5}{15.25} = 10 > 1$$

قيمة F الجدولية عند مستوى معنوية $\alpha = 0.05$ ودرجات حرية $V_1 = 2$ ، $V_2 = 4$ هي $F = 6.94$ أي أن القيمة المحسوبة أكبر من الجدولية ونستنتج أن أداء المنظفات يختلف معنوياً .

وللمقارنة بين تقدير تباين أنواع الماء وتقدير تباين الباقي

$$F = \frac{37}{15.25} = 2.43$$

وقيمة F الجدولية عند مستوى معنوية $\alpha = 0.05$ ودرجات حرية $V_1 = 2$ ، $V_2 = 4$ هي $F = 6.94$ أكبر من قيمة F المحسوبة ونستنتج أن أداء درجات حرارة الماء لا يختلف معنوياً.

مسرد مصطلحات

❖ **العينة الصغيرة (Small Sample)** : عينة حجمها أقل أو يساوي 30 مفردة، $(n \leq 30)$.

❖ **العينة الكبيرة (Large Sample)** : عينة حجمها اكبر من 30 مفردة $(n > 30)$.

❖ **درجات الحرية (Degrees of freedom)** : بصفة عامة، عندما نستخدم أحصاءة عينة لتقدير معلم المجتمع، تفقد درجة حرية واحدة. وهكذا يمكن القول أن مجموع مربع الانحرافات من الوسط الحسابي $\sum_{i=1}^{n}\left(X_j - \bar{X}\right)^2$ له (n-1) فقط درجات حرية لأن $\bar{X}$ تستخدم لتقدير وسط المجتمع μ.

❖ **المتغير المتصل (Continuous variable)** : متغير يأخذ أي قيمة داخل فترة معينة.

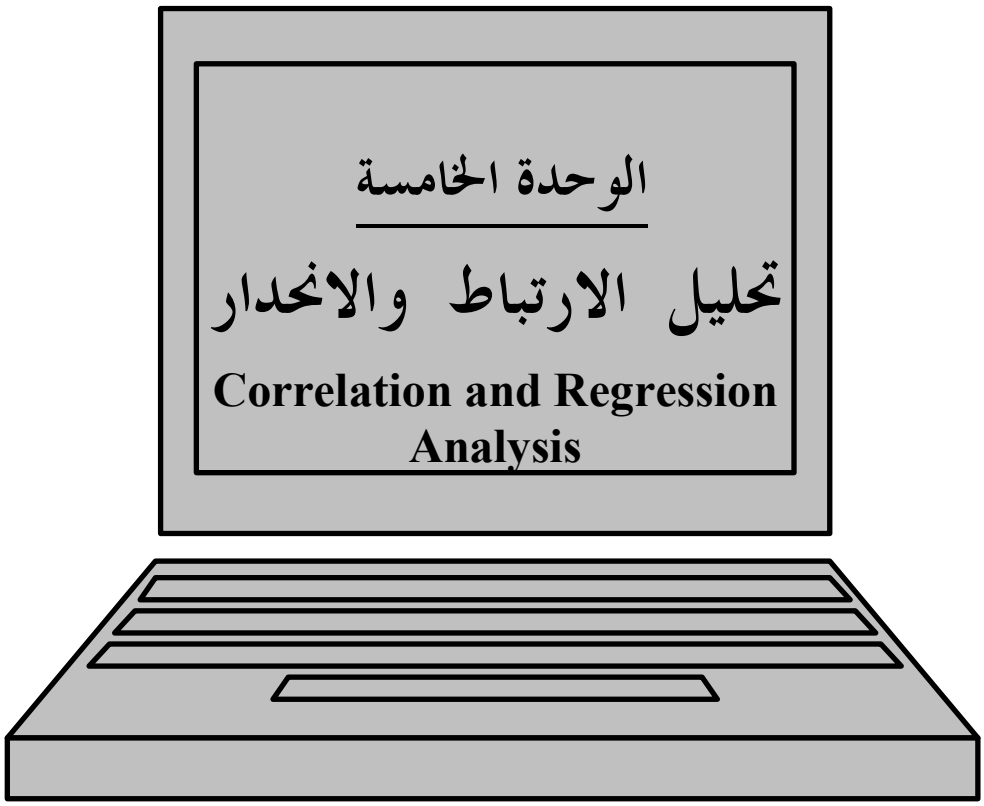
❖ **دالة احتمال الكثافة (Probability density funition)** : التوزيع الأحتمالي لمتغير عشوائي متصل.

❖ **منوال وحيد (Unimodal) : المنوال هو القيمة أو القيم الأكثر تكراراً في التوزيع ولكن عندما تكون هناك قيمة واحدة تتكرر اكثر من غيرها تسمي هذه القيمة بالمنوال الوحيد.**

المراجع

- مواري د. شبيجل، "الإحصاء" ، سلسلة ملخصات شوم، الدار الدولية للنشر والتوزيع 1991.

- بول ج.هويل، " المبادئ الأولية في الإحصاء، ترجمة ومراجعة د. بدرية شوقي عبد الوهاب والدكتور محمد كامل الشربيني ، دار جون وايلي وأبنائه.


الوحدة الخامسة
تحليل الارتباط والانحدار
Correlation and Regression Analysis

محتويات الوحدة

الموضوع	الصفحة
المقدمة	181
تمهيد	181
أهداف الوحدة	182
1. تحليل الارتباط	183
1.1 التغاير	188
2.1 معامل الارتباط الخطي	189
2. الانحدار	199
1.2 الانحدار الخطى البسيط	199
2.2 طريقة المربعات الصغرى العادية	202
3.2 مصدري التباين الكلي للمتغير التابع	215
4.2 معامل التحديد البسيط	217
3. بناء فترة ثقة واختبار المعنوية لمعالم النموذج البسيط	218
الخلاصة	228
لمحة مسبقة عن الوحدة الدراسية التالية	229
إجابات التدريبات	230
مسرد المصطلحات	236
المراجع	237

المقدمة

تمهيد

مرحباً بك في هذه الوحدة من المقرر ، ونذكرك عزيـــزي الـــدارس بأهميـــة الإحصاء التطبيقي في عصرنا الحالي ولكي نفهم هذه الوحدة جيـــداً ينبغـــي أن تكـــون متذكراً لما درسته في مادة الرياضيات في موضوع التفاضل الجزئي وحل المعـــادلات الآنية من الدرجة الأولى في متغيرين باستخدام المصفوفات والمحددات .

بدأت الوحدة بتحليل الارتباط حيث تم شرح الكيفية التي يقدر بها الارتباط بين متغيرين موجب أو سالب أو صفر. كذلك شرحنا معني الارتباط البسيط وكيفية إيجـــاد المعادلة الخطية التي تربط بين المتغيرين ومدى قوة الارتباط بـــين هـــذين المتغيـــرين كذلك تضمنت الوحدة كيفية بناء فترة ثقة لمعلمي المعادلة الخطية واختبار معنوية هذين المعلمين .

كذلك نود أن نشير إلى أن هناك بعض التدريبات وأسئلة التقويم قـــد تخللـــت الوحدة نأمل أنها تحلها جميعاً وإذا واجهتك أي مشـــكلة عليـــك بمراجعـــة المشـــرف الأكاديمي بمركزك كما يمكنك الاستفادة من الوسائط المساندة لهذا المقرر .

أهداف الوحدة



عزيزي الدارس، ،

بعد دراستك لهذه الوحدة ينبغي أن تكون قادراً على أن:

(1) تشرح الارتباط.

(2) تشرح الانحدار.

(3) ترسم لوحة الانتشار.

(4) تحصل على التغاير.

(5) تحسب معامل الارتباط البسيط وتفسره.

(6) تختبر معنوية معامل الارتباط البسيط.

(7) تشرح طريقة المربعات الصغرى.

(8) تجد فترة الثقة لمعلمتي معادلة الانحدار.

(9) تختبر معنوية معلمتي معادل الانحدار.

(10) تختبر معنوية العلاقة بين متغيرين.

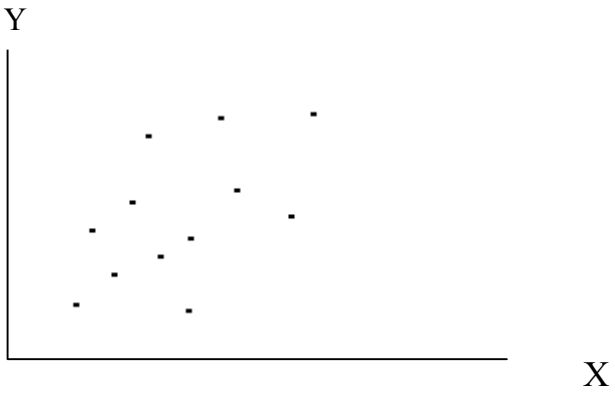
(11) تحصل على معامل التحديد البسيط و تفسره.

1. تحليل الارتباط

تتعلق كل المفاهيم التي تم نقاشها في الوحدات السابقة بمتغير واحد. وتتطلب الممارسة في كثير من الأحوال دراسة أكثر من متغير. فإذا تغيرت قيم متغير وصاحبتها تغيرات في قيم متغير أخر فهذا يعنى أن هناك ارتباط بين المتغيرين. فمثلاً يوجد ارتباط بين سعر سلعة والكميات المطلوبة من هذه السلعة . وتقاس درجة الارتباط بين المتغيرين من خلال تحليل الارتباط. ولنقاش المقاييس المختلفة نفترض أن لدينا بيانات عن المتغير Y وما يقابلها من المتغير X .

X	Y
x_1	y_1
x_2	y_2
x_3	y_3
.	.
.	.
.	.
.	.
x_n	y_n

ويمكن تفريغ هذه البيانات في شكل بياني ذو محورين كما يلي:



لنحسب الوسط الحسابي لكل من المتغيرين:

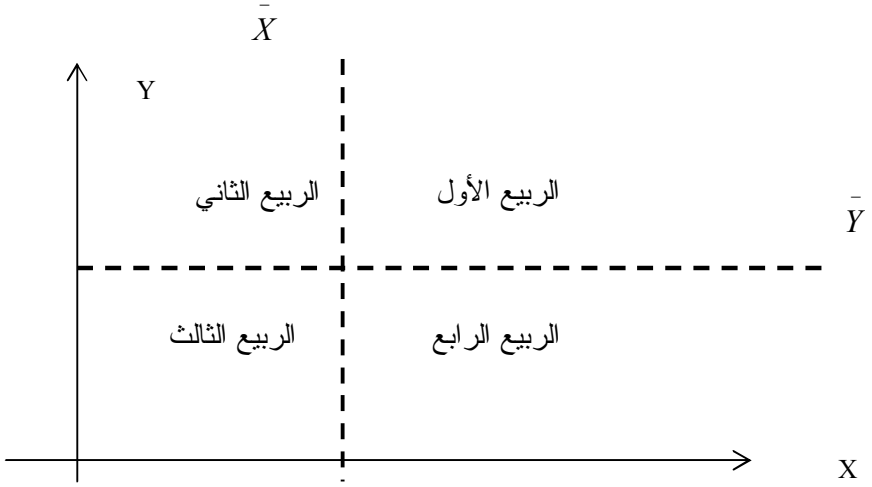
$$\bar{Y} = \frac{\sum Y}{n}$$

$$\bar{X} = \frac{\sum X}{n}$$

ونحدد نقطتي موقع الوسطين على المحورين Y , X وننشئ عمودين من النقطتين:

ويتم تحديد نقطة الوسط الحسابي للمتغيرين على المحورين، الوسط الحسابي للمتغير X على المحور الأفقي والوسط الحسابي للمتغير Y على المحور الرأسي

ومن ثم يمد عمودان من نقطة الوسط الحسابي لكل متغير ليكـــون إحـــداهما موازيـــاً للمحور الرأسي والأخر موازياً للمحور الأفقي كما يوضح الشكل التالي:



ويقسم العمودان $\bar{X}\bar{X}$ و $\bar{Y}\bar{Y}$ الشكل البياني إلى أربعة أقسام ويسمى كـــل قسم ربيع وتسمى عكس اتجاه عقارب الساعة بالربيع الأول والربيع الثـــاني والربيـــع الثالث والربيع الرابع

والآن لنحسب الفرق بين Y و $\bar{Y}$ أي $\left(Y-\bar{Y}\right)$ والفرق بين X و $\bar{X}$ أي $\left(X-\bar{X}\right)$ للنقاط التي تقع في الربيعات الأربعة لمعرفة إذا كان الفرق موجب أم سالب:

الربيع	الفرق	الإشارة	ملحوظات
الأول	$Y - \bar{Y}$ $X - \bar{X}$	$+$ $+$	ينطبق على كل النقاط التي تقع في الربيع الأول
الثاني	$Y - \bar{Y}$ $X - \bar{X}$	$+$ $-$	ينطبق على كل النقاط التي تقع في الربيع الثاني
الثالث	$Y - \bar{Y}$ $X - \bar{X}$	$-$ $-$	ينطبق على كل النقاط التي تقع في الربيع الثالث
الرابع	$Y - \bar{Y}$ $X - \bar{X}$	$-$ $+$	ينطبق على كل النقاط التي تقع في الربيع الرابع

ويتضح من الجدول أعلاه، إذا وقعت كل النقاط أو جلها في الربيع الأول والثالث فان مجموع حواصل ضرب $\left(Y-\bar{Y}\right)$ في $\left(X-\bar{X}\right)$ سيكون موجب، ورمزاً $\sum\left(Y-\bar{Y}\right)\left(X-\bar{X}\right) > 0$

وهذا يعني أن الزيادة في X تقابلها زيادة في Y وبالعكس.

وأما إذا وقعت كل النقاط أو جلها في الربيع الثاني والرابع فان مجموع حواصل ضرب $\left(Y-\bar{Y}\right)$ في $\left(X-\bar{X}\right)$ سيكون سالبا، ورمزاً

$$\sum\left(Y-\bar{Y}\right)\left(X-\bar{X}\right)<0$$

وهذا يعني أن الزيادة في X يقابلها نقصان في Y وبالعكس.

وعليه يمكن استخدام المقدار $\sum\left(Y-\bar{Y}\right)\left(X-\bar{X}\right)$ كمقياس تقريبي للارتباط، فإذا كانت إشارته موجبة فهذا يعنى ارتباط موجب بين Y و X وأما إذا كانت إشارته سالبة فهذا يعنى ارتباط سالب بين X,Y

بيد أن هذا المقدار التقريبي يعانى من عيبين أساسيين:

(أ) يتأثر بعدد مفردات العينة أي حجم العينة n

(ب) يتأثر بوحدات القياس المستخدمة.

وهذا يتطلب تصحيح لهذين العيبين حتى يمكن استخدامه كمقياس للارتباط يعتمد عليه. وتصحيح العيبين بدوره يؤدى إلى مقياسين مفيدين للارتباط:

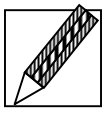
1.1 التغاير Covariance

يستخدم التغاير كمقياس إحصائي لاكتشاف إذا كان هناك ارتباط بين متغيرين مثلاً، X , Y ولكنه لا يوفر معلومة عن درجة الارتباط بين المتغيرين، أي مدى قوة أو ضعف الارتباط ويمكن تعريفه بمتوسط الضربيات $\left(Y-\bar{Y}\right)\left(X-\bar{X}\right)$. وعليه تصبح الصيغة المستخدمة للحصول على التغاير هي

$$Cov\ (X\ ,\ Y) = \frac{\sum \left(Y - \bar{Y}\right)\left(X - \bar{X}\right)}{n}$$

وتكمن فائدة التغاير فقط في معرفة إذا كان الارتباط بين Y و X موجب أم سالب أو صفر. وهذا يعنى أن ما يعتد به في هذه الحالة هو إشارة المقدار وليس حجمه.

تدريب (1)



حصل 8 طلاب على الدرجات التالية في الإحصاء والرياضيات:

درجات الإحصاء:	15	28	20	14	40	60	80	30
درجات الرياضيات:	20	30	50	40	60	30	70	10

أحسب التغاير

2.1 معامل الارتباط الخطي: Linear Correlation Coefficient

يعتبر معامل الارتباط الخطي أهم مقياس للارتباط لأنه عند تعريفه تتم معالجة العيبين الأساسيين للمقياس التقريبي $\sum \left(Y - \bar{Y} \right)\left(X - \bar{X} \right)$. ومن جهة أخرى، يستفاد منه ليس فقط في معرفة اتجاه الارتباط ولكن أيضا مدى درجته ضعفاً وقوة. ويرمز لمعامل الارتباط البسيط المحسوب على أساس عينة بالرمز r ويعرف كما يلى:

$$r = \frac{\sum \left(Y - \bar{Y} \right)\left(X - \bar{X} \right)}{\sqrt{\sum \left(Y - \bar{Y} \right)^2} \sqrt{\sum \left(X - \bar{X} \right)^2}}$$

وتقع r بين سالب واحد وموجب واحد

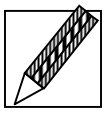
$$-1 \leq r \leq +1$$

وينبغي أن ننبه أن معامل الارتباط البسيط مقياس للارتباط الخطى فقط ولهذا لا يصلح تطبيقه في حالة الارتباط غير الخطى. ونشير إلى أن حجم r مهم للغاية في معرفة درجة الارتباط بين المتغيرين كما يتضح من الجدول التالى:

جدول التعرف على درجة الارتباط بين متغيرين

حجم r	درجة الارتباط	ملحوظات
1	ارتباط موجب تام	هذا يعني كل النقاط في لوحة الانتشار تقع في الربيع الأول والثالث على خط مستقيم ذو ميل موجب
−1	ارتباط سالب تام	هذا يعني كل النقاط في لوحة الانتشار تقع في الربيع الثاني والرابع على خط مستقيم ذو ميل سالب
قريبه من 1	ارتباط موجب قوى	هذا يعني أن كل أو معظم النقاط في لوحة الانتشار تقع في الربيع الأول والثالث ولكن ليس على خط مستقيم
قريبه من -1	ارتباط سالب قوى	وهذا يعنى أن كل النقاط أو معظمها في لوحة الانتشار تقع في الربيع الثاني والرابع ولكن ليس على خط مستقيم.
$0.5 \le r \le 0.7$	ارتباط موجب معتدل	تقع بعض النقاط على الربيع الثاني والرابع
$0.5 \le r \le 0.7$	ارتباط سالب معتدل	تقع بعض النقاط على الربيع الأول والثالث
$r < 0.5$	ارتباط موجب ضعيف	النقاط التي تقع في الربيع الأول والثالث اكثر بقليل من النقاط التي تقع في الربيع الثاني والرابع ولكن ليس في شكل منتظم.
$r < 0.5$	ارتباط سالب ضعيف	النقاط التي تقع في الربيع الثاني والرابع اكثر بقليل من النقاط التي تقع في الربيع الأول والثالث ولكن ليس في شكل منتظم
$r = 0$	ينعدم الارتباط بين المتغيرين	وهذا يعنى أنه لا يوجد ارتباط خطي بين المتغيرين ولكن قد يوجد ارتباط غير خطى بينهما.

تدريب (2)



حصل 8 طلاب على الدرجات التالية في الإحصاء والرياضيات:

درجات الإحصاء:	15	28	20	14	40	60	80	30
درجات الرياضيات:	20	30	50	40	60	30	70	10

أحسب معامل الارتباط بين درجات المقررين

شكل استخدام معامل الارتباط لوصف العلاقة بين متغيرين:

ارتباط قوي (+)	ارتباط معتدل (+)	ارتباط ضعيف (+)	ارتباط ضعيف (−)	ارتباط معتدل (−)	ارتباط قوي (−)
ارتباط تام موجب	ارتباط معتدل موجب	ارتباط متوسط موجب	انعدام الارتباط	ارتباط متوسط سالب	ارتباط معتدل سالب

وتعتبر r إحصاءة عينة (Sample Statistics) لأنها تحسب على أساس مفردات عينة. وعليه يمكن أن ينظر إلى r بمثابة تقدير نقطة (Point estimate) لمعلم المجتمع ρ التي تمثل معامل الارتباط الخطى لكل ازواج المفردات في المجتمع. ويترتب على ذلك كيفية معرفة إذا كانت r معنوية (Significant)

والتقدير r مثله مثل أيه إحصاءة (Statistics) يتغير بتغير العينة المسحوبة من المجتمع. وعندما نحصل على قيمة موجبة أو سالبة للمقدار r لا نعرف على وجه اليقين فيما إذا كانت قيمة ρ المقابله لها أيضا موجبة أو سالبة وذلك لخطأ المعاينة (Sampling error). ولهذا السبب يصبح اختبار المعنوية ضرورة لا غنى عنها.

ونشير هنا أن توزيع المعاينة للإحصاءة r يتغير مع قيمة ρ وحجم العينة n. وبافتراض أن العينتين مسحوبتان من مجتمع مشترك للمتغيرين X و Y يتبع نمط التوزيع الطبيعي، فإن r تميل إلى التوزيع الطبيعي كل ما زاد حجم العينة. وهذا الميل نحو التوزيع الطبيعي يكون اكثر قوة لقيم ρ القريبة من الصفر مقارنة مع قيم ρ القريبة من موجب واحد أو سالب واحد. وبصورة أخري إذا كانت $\rho = 0$ وحجم العينة n كبير فيمكن تقريب توزيع r باستخدام التوزيع الطبيعي الأمر الذي يترتب عليه إمكانية اختبار معنوية فرضية العدم الخاصة بارتباط المجتمع وهى:

$$H_o : \rho = 0$$

في مواجهة الفرض البديل $H_1 : \rho \neq 0$

ويمكن الحصول على القيمة المحسوبة باستخدام القاعدة التالية:

$$t = r \sqrt{\frac{n-2}{1-r^2}}$$

حيث أن

n : حجم العينة

r : معامل ارتباط البسيط

وتتبع النسبة t نمط توزيع t (t-distribution) بدرجات حرية ( )

n =2 إذا كان فرض العدم صحيح.

مثال

يوضح الجدول التالي الكميات المستهلكة (Y) من سلعة ما والـدخل (X) لخمسة مستهلكين لهذه السلعة:

Y	X
3	5
5	9
7	11
9	12
6	13

المطلوب

(أ) رسم لوحة الانتشار والتعليق عليها.

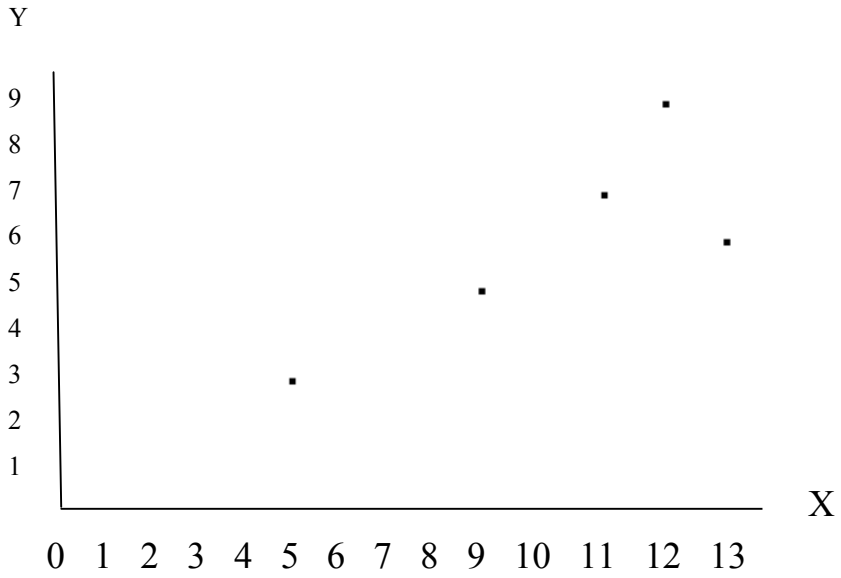
(ب) الحصول على التغاير والتعليق على النتيجة.

(ج) الحصول على معامل الارتباط البسيط والتعليق على النتيجة

(د) اختبار معنوية معامل الارتباط البسيط للمجتمع بأستخدام مستوى معنوية 5 في المائة.

الحل

(أ)



يوضح الشكل علاقة موجبة بين Y و X لأن معظم النقاط تقع في الربيع الأول والثالث

(ب)

Y	X	$Y-\bar{Y}$	$X-\bar{X}$	$\left(Y-\bar{Y}\right)\left(X-\bar{X}\right)$	$\left(Y-\bar{Y}\right)^2$	$\left(X-\bar{X}\right)^2$
3	5	-3	-5	15	9	25
5	9	-1	-1	1	1	1
7	11	1	1	1	1	1
9	12	3	2	6	9	4
6	13	0	3	0	0	9
30	50			23	20	40

$$\bar{Y}=\frac{\sum Y}{n}=\frac{30}{5}=6$$

$$\bar{X}=\frac{\sum X}{n}=\frac{50}{5}=10$$

$$Cov(X,Y)=\frac{\sum\left(Y-\bar{Y}\right)\left(X-\bar{X}\right)}{n}=\frac{23}{5}=4.6$$

وبما أن الإشارة موجبة يوجد ارتباط موجب بين Y , X.

(ج)

$$r = \frac{\sum\left(Y-\bar{Y}\right)\left(X-\bar{X}\right)}{\sqrt{\sum\left(Y-\bar{Y}\right)^2}\sqrt{\sum\left(X-\bar{X}\right)^2}}$$

$$= \frac{23}{\sqrt{20}\sqrt{40}} = \frac{23}{\sqrt{800}} = \frac{23}{28.28} = 0.81$$

يوجد ارتباط موجب قوى بين X و Y

(د)

$$t = r\sqrt{\frac{n-2}{1-r^2}} = 0.81\sqrt{\frac{5-2}{1-(0.81)^2}}$$

$$= 0.81\sqrt{\frac{3}{1-0.66}} = 0.8\sqrt{\frac{3}{0.34}}$$

$$= (0.81)(8.82) = 7.15$$

تساوى القيمة الحرجة لإحصاءة الاختبار t 3.182 عند مستوى معنوية 5 في المائة ودرجات حرية 3. وبما أن القيمة المحسوبة اكبر نرفض فرض

العدم ونقر بوجود ارتباط معنوى بين X و Y. وهذا يعنـــى أن قيمـــة r = 0.81 بدرجات حرية 3 لا يمكن أن تكون ناتجة عن تغيرات الصدفة لوحدها.

وبما أن قيم $\bar{X}$ و $\bar{Y}$ عادة ما تحتوي علي تقريـــب يستحســـن اســـتخدام القانونين المكافئين التاليين:

$$COV(X_1Y) = \frac{1}{n}\sum XY - \frac{(\sum X)(\sum Y)}{n^2}$$

$$r = \frac{n\sum XY - (\sum X)(\sum Y)}{\sqrt{\left[n\sum X^2 - (\sum X)^2\left[n\sum Y^2 - (\sum Y)^2\right]\right.}}$$

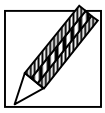
XY	Y^2	X^2	Y	X
15	9	25	3	5
45	25	81	5	9
77	49	121	7	11
108	81	144	9	12
78	36	169	6	13
323	200	540	30	50

$$COV(X,Y) = \left(\frac{1}{5}\right)(323) - \frac{(50)(30)}{25} = 4.6$$

$$r = \frac{(5)(323) - (50)(30)}{\sqrt{(5(540) - (50^2))(5(200) - (30^2))}} = 0.81$$

كما سبق

تدريب (3)

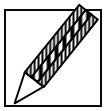


يعطي الجدول التالي أرقام الأسعار والعرض لسلعة معينة في فترة 6 أيام:

اليوم	السبت	الأحد	الاثنين	الثلاثاء	الأربعاء	الخميس
السعر	8	15	20	25	30	20
العرض	28	23	20	15	12	10

أحسب معامل الارتباط الخطي. وما هو الاستنتاج الذي يمكن أن تصل إليه من النتيجة.

تدريب (4)



للحصول على معامل الارتباط الخطي بين X و Y من 12 زوج من المشاهدات، أجريت الحسابات التالية:

$\Sigma X = 30$ ، $\Sigma Y = 5$ ، $\Sigma X^2 = 670$ ، $\Sigma Y^2 = 285$ ، $\Sigma XY = 334$

تمت مراجعة للحسابات ووجد أن الزوج $(X = 11\ ,\ Y = 4)$ نسخ بصورة خاطئة والصحيح هو $(X = 10\ ,\ Y = 14)$ أحصل على معامل الارتباط الخطي الصحيح.

أسئلة التقويم الذاتي (1)

(1) أشرح الصلة بين التغاير للمتغيرين X و Y ومعامل الارتباط البسيط بين المتغيرين X و Y.

(2) هل يوحي معامل الارتباط البسيط السببية (Causation) بين متغيرين أشرح.

(3) أذكر عيوب التغاير كمقياس للارتباط.

2. الانحدار Regression

1.2 الانحدار الخطى البسيط Simple Linear Regression

تحتل دراسة الانحدار أهمية خاصة في مجال الاحصاء التطبيقي وتهدف الي:

(أ) وصف العلاقة بين المتغير التابع والمتغير المستقل وقياس تأثير المتغيرات المستقلة على المتغير التابع.

(ب) التنبؤ بقيم المتغير التابع بمعلومية القيم المناظرة للمتغيرات المستقلة.

(ج) التحكم في الظاهرة تحت الدراسة عن طريق التحكم في المتغيرات المستقلة

وفي هذه الوحدة سنهتم فقط بالانحدار الخطي البسيط. وعليه نفترض أن هناك متغير مستقل واحد في الطرف الأيمن من المعادلة، أي نفترض أن:

$$Y = f(X) \quad \ldots\ldots\ldots\ldots\ldots\ldots (1)$$

حيث يعتبر متغير رياضي (محدد القيم)

وهذا يعنى وجود علاقة دالية بين Y و X . ولتحديد طبيعة العلاقة الدالية نفترض أنها خطية (أي أنها من الدرجة الأولي):

$$Y = a + bX \quad \ldots\ldots\ldots\ldots\ldots\ldots (2)$$

ولفظة خطية تعنى أنه يمكن وصف العلاقة بين Y و X في المتوسط بخط مستقيم. ويمكن تبرير استخدام العلاقة الخطية بأن كثير من العلاقات الكمية بين المتغيرات ذات طبيعة خطية، وأن العلاقات الخطية تعطى تقريب للعلاقات المعقدة والتي يصعب فهمها في كثير من الأحيان. المعادلة (2) أعلاه، تسمى علاقة محددة (Deterministic) وهذا يعنى أن التغيرات التي تحدث في المتغير التابع Y تقابل فقط التغييرات التي تحدث في المتغير المستقبل X . ومثل هذا النوع من النماذج يدرس عادة في الرياضيات والاقتصاد الرياضي. والهدف الرئيسي هو كيفية الحصول على قيمة مقدرة لكل من القاطع a والميل b. ولاستخدام طرق إحصائية للحصول

على قيمة تقديرية لكل من a و b ينبغي تحويل النموذج من نموذج محدد إلى نموذج عشوائي (Stochastic or random) . ويمكن تحويله ببساطة بإضافة مكون عشوائي في الطرف الأيمن للمعادلة (2) ويرمز له بالرمز μ. وعليه يمكن كتابة النموذج العشوائي كما يلي:

$$Y = a + bX + \mu \text{.....(3)}$$

حيث أن:

μ يسمى الحد العشوائي أو الخطأ العشوائي (Error term) وله أيضا بعض المسميات الأخرى وهذا يعني أنه بينما تقاس X قياساً مضبوطاً إلا أن قياس Y يتأثر بعامل خطأ القياس وهو متغير عشوائي.

ويقوم المتغير μ مقام كل المتغيرات المحذوفة التي تؤثر على Y ولكن لسبب أو لآخر غير مضمنة صراحة في الطرف الأيمن من المعادلة. وعليه يمكن وصف المتغير العشوائي بأنه وعاء جامع لكل هذه المتغيرات المحذوفة إضافة للخطأ بالمعني الضيق فهو يحمل أثر هذه المتغيرات على المتغير التابع. وكما ذكرنا فإن الهدف هو كيفية الحصول على قيم تقديرية لكل من a و b . وهناك بعض الطرق المستخدمة للحصول على هذين التقديرين. ولكن سنستخدم هنا طريقة المربعات الصغرى العادية Ordinary Least Squares (OLS)

2.2 طريقة المربعات الصغرى العادية (OLS)

طريقة المربعات الصغرى العادية طريقة إحصائية أصلا عدلت وحورت لتناسب الظاهرة الاجتماعية. وتقوم هذه الطريقة على عدة فروض، فروض خاصة بالمتغير العشوائي μ وفروض خاصة بالمتغير المستقل وفروض خاصة بالعلاقة بين μ و X ونورد هذه الفروض دون الدخول في شرح تفصيلي لها:

(1) μ متغير عشوائي، بمعنى أن له توزيع احتمالي يمكن وصفه بعزوم (Moments) معينة مثل التوقع والتباين.

(2) متوسط μ يساوى الصفر، ورمزاً $E(\mu)=0$.

(3) تباين $\left(\text{var}(\mu)=\sigma_u^2\right)$ μ ثابت عند كل قيم X. ويسمى هذا الفرض بتجانس التباين (Homoscedasticity) .

(4) التغاير بين μ_i μ_j , يساوى صفر، ورمزاً $Cov(\mu,\mu_j)=0$.أي أن القيمة التي يتخذها μ في فترة زمنية معينة مستقله عن القيمة التي يتخذها في فترة زمنية أخري.

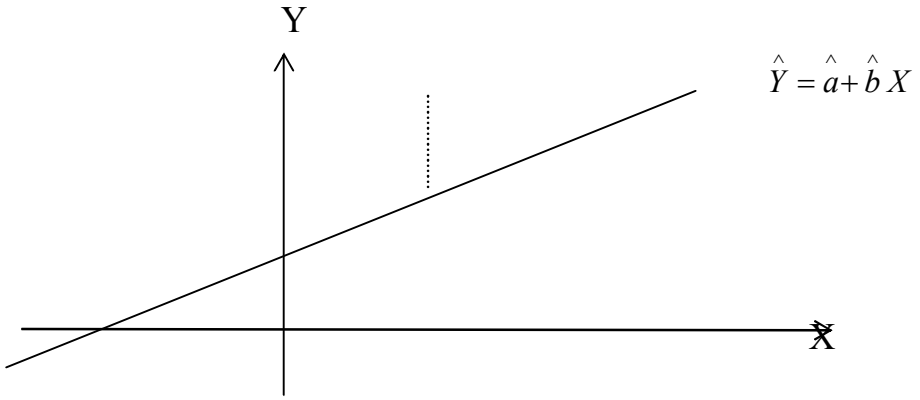
(5) قيم X ثابته لا تتغير من عينة إلى أخرى.أي أن X متغير رياضي.

(6) قيم المتغير المستقل X وقيم u مستقلة عن بعضها البعض بمعنـــى : $Cov(X,\mu)=0$ وهذا يبدو واضحاً من الفرض (5) حيـــث أن قـــيم X ثابتة لا تتغير الأمر الذي يترتب عليه استقلالها عن μ .

وإذا تم الاستيفاء بهذه الفروض يمكن تطبيـــق طريقـــة المربعـــات الصـــغرى العادية. وبما أن الموضوع تحت الدراسة هو انحدار المجتمع (Population Regression) ينبغي سحب عينة عشوائية من مجتمع (X,Y) للحصـــول على قيم تقديرية لمعالم هذا المجتمع. والآن لنفتـــرض أن العينـــة العشـــوائية المسحوبة هي $(X_1,Y_1),(X_2,Y_2),(X_3,Y_3),.....,(X_n,Y_n)$

X	Y
X_1	Y_1
X_2	Y_2
X_3	Y_3
.	.
.	.
.	.
X_n	Y_n

ولنفترض إننا أفرغنا هذه البيانات في الشكل البياني ذو المحـــورين الرأســـي والأفقي باستخدام مقياس الرسم المناسب للحصول على لوحة الانتشار التالية:



وإذا تم تقدير الخط المستقيم الذي يمر بين هذه النقاط بالمعادلة $\hat{Y} = \hat{a} + \hat{b} X$ يتضح انه ينحرف من النقاط المشاهدة حيث توجد نقاط أعلى الخط ونقاط اسفل الخط قلة من النقاط قد يمر عليها الخط. وهذه الانحرافات تسمى اخطاء (errors) . ويمكن حساب الخطأ بالمسافة العمودية بين النقطة المشاهدة والنقطة المقدرة. وننبه هنا إلى أن الخط المقدر $\hat{Y} = \hat{a} + \hat{b} X$ هو الذي يمثل النقاط المقدرة عند قيم X الملاحظة. وبصفة عامة، فان النقطة المشاهدة في الشكل البياني يمثلها الزوج المرتب (X_i , Y_i) والنقطة المقدرة يمثلها الزوج المرتب $\left(\hat{X}_i, \hat{Y}_i\right)$. وعليه لحساب الخطأ e_1 نطرح $\hat{Y}_1$ من Y_1 وهو يمثل المسافة العمودية $Y_1 - \hat{Y}_1$ ، أي أن

$$e_1 = Y_1 - \hat{Y}_1$$

ونكرر ذلك لكل نقاط العينة والتي عددها n للحصول على الأخطاء e_i

$$e_1 = Y_1 - \hat{Y}_1$$

$$e_2 = Y_2 - \hat{Y}_2$$

$$e_3 = Y_3 - \hat{Y}_3$$

$$e_n = Y_n - \hat{Y}_n$$

ونلاحظ هنا عندما تقع النقطة أعلى الخط المقدر، تكون إشارة الخطأ موجبة وعندما تقع النقطة اسفل الخط، تكون إشارة الخطأ سالبة وعندما تقع النقطة على الخط المقدر يساوى الخطأ صفراً. وتقوم طريقة المربعات الصغرى العادية على مبدأ بسيط للغاية وهو أن الخط المقدر الذي يعطى أفضل جودة توفيق (Goodness of fit) هو الخط الذي يكون مجموع مربع أخطائه اقل ما يمكن. وللحصول على مجموع مربع الأخطاء: نربع الأخطاء $e_i(i = 1,2,.....,n)$ ثم نجمعها:

$$e_1^2 = \left(Y_1 - \hat{Y}_1\right)^2$$

$$e_2^2 = \left(Y_2 - \hat{Y}_2\right)^2$$

$$e_3^2 = \left(Y_3 - \hat{Y}_3\right)^2$$

$$e_n^2 = \left(Y_n - \hat{Y}_n\right)^2$$

ويمكن كتابة مجموعها كما يلي:

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i \right)^2 \text{.....(4)}$$

وبما أن موضع الخط المستقيم يتحكم فيه القاطع a والميل b يمكن إدخال هذين الثابتين في الصورة ومن ثم أعاده صياغة مبدأ طريقة المربعات الصغرى العادية. وبالتعويض

في المعادلة (4) نحصل على:

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{\hat{}} \left(Y_i - \hat{a} - \hat{b} X_i \right)^2 \text{.....(5)}$$

والآن يمكن صياغة مبدأ طريقة المربعات الصغرى العادية كما يلي: أختار قيم تقديرية لكل من a و b بحيث أن مجموع مربع الأخطاء $\sum_{i=1}^{n} e_i^2$ يكون أقل ما يمكن. وهذا يعنى أننا نتعامل مع مشكلة الحصول على نهاية صغرى لمجموع مربع الأخطاء وبتطبيق الشرط الضروري للنهاية الصغرى وهو الحصول على التفاضل

الجزئي لمجموع مربعات الأخطاء بالنسبة لكل من $\hat{a}$ و $\hat{b}$ ومساواتها بالصـــفر كما يلي:

$$\frac{\partial \sum e^2}{\partial \hat{a}} = -2\sum\left(Y_i - \hat{a} - \hat{b}X_i\right) = 0$$

وبقسمة طرفي المعادلة على 2− نحصل على

$$\sum\left(Y_i - \hat{a} - \hat{b}X_i\right) = 0$$

وبفك الأقواس وإعادة ترتيب الحدود نحصل على:

$$\sum Y_i = n\hat{a} + \hat{b}\sum X_i \text{..(6)}$$

وبالمثل،

$$\frac{\partial \sum e^2}{\partial \hat{b}} = -2\sum X\left(Y_i - \hat{a} - \hat{b}X_i\right) = 0$$

وبقسمة طرفي المعادلة على 2− نحصل على:

$$\sum X_i\left(Y_i - \hat{a} - \hat{b}X_i\right) = 0$$

وبفك الأقواس وإعادة ترتيب الحدود نحصل على:

$$\sum X_iY_i = \hat{a}\sum X_i + \hat{b}\sum X_i^2 \text{..........................(7)}$$

والمعادلتان (6)(7) تسـميان معادلتـان طبيعيتـان (Normal Equations) وهما معادلتان آنيتان في مجهولين. وبالتالي يمكن حلهما آنيا للحصول على $\hat{a}$ و $\hat{b}$.

ويمكن صياغة المعادلتين الطبيعيتين باستخدام الجبر الخطى كما يلي:

$$X = \begin{bmatrix} X_1 \\ X_2 \\ . \\ . \\ .X_n \end{bmatrix}, Y = \begin{bmatrix} Y_1 \\ Y_2 \\ . \\ . \\ Y_n \end{bmatrix}, \beta \begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix},$$

$$Y = X\beta$$

وبالضرب من اليسار في X'

$$X'Y = X^1 X\beta$$

$$X'X = \begin{bmatrix} n & \Sigma X \\ \Sigma X & \Sigma X^2 \end{bmatrix}$$

$$\hat{B} = \begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix}$$

$$X'Y = \begin{bmatrix} \Sigma Y \\ \Sigma XY \end{bmatrix}$$

وتصبح المعادلتان الطبيعيتان باستخدام الجبر الخطى كما يلي:

$$\left(X'X\right)\hat{B} = X'Y$$

وباستخدام طريقة معكوس المصـــفوفة (Inverse of a matrix method)

نحصل على القيم التقديرية $\hat{a}$ و $\hat{b}$ كما يلي:

$$\hat{B} = \left(X'X\right)^{-1}X'Y$$

ويمكن الحصول على القيم التقديرية $\hat{a}$ و $\hat{b}$ أيضا باستخدام طريقة كرامر (Cramer's method) وهى تقوم على فكرة المحددات

$$\hat{a} = \frac{\begin{vmatrix} \sum Y & \sum X \\ \sum XY & \sum X^2 \end{vmatrix}}{\begin{vmatrix} n & \sum X \\ \sum X & \sum X^2 \end{vmatrix}} = \frac{(\sum Y)\sum\left(X^2\right) - (\sum X)(\sum XY)}{n\sum X^2 - (\sum X)^2}$$

$$\hat{b} = \frac{\begin{vmatrix} n & \sum Y \\ \sum X & \sum XY \end{vmatrix}}{\begin{vmatrix} n & \sum X \\ \sum X & \sum X^2 \end{vmatrix}} = \frac{n\sum XY - (\sum X)(\sum Y)}{n\sum X^2 - (\sum X)^2}$$

وفى كل هذه الطرق للحصول على $\hat{a}$ و $\hat{b}$ استخدمنا القيم المشاهدة للمتغيرين Y و X.

وهناك طريقة أخرى للحصول على $\hat{a}$ و $\hat{b}$ وهى استخدام انحراف قيم كل من Y و X من أوساطها الحسابية. وفى هذه الحالة وكعرف يتم استخدام الحروف الصغيرة (Small letters) لتمثل انحراف القيم من الوسط الحسابي كما يلي:

$$y = Y - \bar{Y}$$

$$x = X - \bar{X}$$

ومن ناحية، أخرى، تستخدم الحروف الكبيرة (Capital Letters) لتمثل القيم الحقيقية. والآن لاستخدام انحراف القيم من أوساطها الحسابية يمكن الحصول على النموذج الجديد بالرجوع إلى المعادلة(3)

$$Y = a + bX + u$$

اجمع وأقسم طرفي هذه المعادلة على n

$$\frac{\Sigma Y}{n} = \frac{na}{n} + b\frac{\Sigma X}{n} + \frac{\Sigma u}{n}$$

$$\bar{Y} = a + b\bar{X} + \bar{u} \ldots\ldots\ldots\ldots\ldots(8)$$

أطرح المعادلة (8) من المعادلة (3)

$$Y - \bar{Y} = b\left(X - \bar{X}\right) + \left(u - \bar{u}\right)$$

أو

$$y = bx + \left(u - \bar{u}\right) \ldots\ldots\ldots\ldots\ldots\ldots(9)$$

لا يوجد سبب يجعل من $\bar{u}$ تساوى الصفر بالرغم من أن $E(u) = 0$. ولكن تستخدم المعادلة (9) فقط في حالة اشتقاق نتائج تخص تحيز أو عد تحيز طريقة التقدير. ولسهولة العرض نستخدم الفرض الضمني أن $\bar{\mu} = 0$ ونكتب النموذج في حالة استخدام انحراف القيم من الوسط الحسابي كما يلي:

$$y = bx + \mu$$

$$y = \hat{b}x + e$$

حيث أن الباقي e (Residual) يعتبر تقدير للمتغير العشوائي μ و $\hat{b}$ تعتبر تقدير للمعلم b .

وعلية:

$$e = y - \hat{b}x \text{.....(10)}$$

ونقوم ألان بتربيع طرفي المعادلة ثم نجمع للحصول على:

$$\sum e^2 = \sum\left(y - \hat{b}x \right)^2$$

ويمكن استخدام مبدأ طريقة المربعات الصغرى العادية للحصول على $\hat{b}$ التي تعطى اقل مجموع مربع أخطاء، وذلك بإيجاد تفاضل $\sum e^2$ بالنسبة لـــ $\hat{b}$ ومساواته بالصفر.

$$\frac{\partial \sum e^2}{\partial \hat{b}} = -2\sum x\left(y - \hat{b}x \right) = 0$$

أقسم طرفي المعادلة على 2 –

$$\sum x\left(y - \hat{b}x \right) = 0$$

$$\sum xy - \hat{b}\sum x^2 = 0$$

$$-\hat{b}\sum x^2 = -\sum xy$$

$$\hat{b} = \frac{\sum xy}{\sum x^2} \text{.....(11)}$$

وهذه الصيغة يمكن استخدامها للحصول علـــى تقـــدير لميـــل المعادلـــة (3).

وللحصول على تقدير للقاطع a نرجع للمعادلـــة (6) وهـــي إحـــدى المعـــادلتين الطبيعيتين :

$$\sum Y = n\hat{a} + \hat{b}\sum X$$

أقسم طرفي هذه المعادلة على n

$$\frac{\sum Y}{n} = \frac{n\hat{a}}{n} + \hat{b}\frac{\sum X}{n}$$

$$\bar{Y} = \hat{a} + \hat{b}\bar{X}$$

$$\hat{a} = \bar{Y} - \hat{b}\bar{X} \quad \text{.....(12)}$$

وللحصول على قيم تقديرية للميل b والقاطع aنستخدم الصـــيغتين (11) و (12) على التوالي في حالة انحراف قيم Yوقيم X من أوساطها الحســـابية كمـــا ذكرنا من قبل.

وبالرجوع إلى المعادلة (11)

$$\hat{b} = \frac{\sum xy}{\sum x^2} \quad \text{.....(11)}$$

وبما أن y متغير عشوائي وبالتالي $\hat{b}$ أيضا متغير عشوائي من الطبيعي تحديد خواص توزيع $\hat{b}$.ودون الدخول في تفاصيل واشتقاقات رياضية فان:

$$E\left(\hat{b}\right)=b$$

أي أن $\hat{b}$ تقدير غير متحيز (Unbiased) للمعلمة b و:

$$\text{var}\left(\hat{b}\right)=\frac{\sigma_u^2}{\sum x^2}$$

ويتضح أن تباين $\hat{b}$ يعتمد فقط على تباين المتغير العشوائي وتباين المتغير المستقل X وعدد مفردات العينة. وبالمثل يمكن الحصول على وسط وتباين القاطع $\hat{a}$

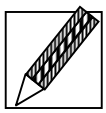
$$E\left(\hat{a}\right)=a$$

$$\text{var}\left(\hat{a}\right)=\frac{\sigma_u^2\sum X^2}{n\sum x^2}$$

حيث أن $\sum x^2=\sum\left(X-\bar{X}\right)^2$. ونحصل على تقدير لتباين المتغير العشوائي كما يلي :

$$\hat{\sigma_u^2}=\left(\sum y^2-\hat{b}\sum xy\right)/(n-2)$$

تدريب (5)



أوجد معادلتا الانحدار للبيانات التالية

Y	91	97	108	121	67	124	51	73	111	57
X	71	75	69	97	70	91	39	61	80	47

3.2 مصدري التباين الكلي للمتغير التابع

عزيزي الدارس سبق أن أوضحنا أن

$$y = \hat{b}x + e$$

وإذا قمنا بتربيع طرفي المعادلة ثم جمعنا نحصل على :

$$\Sigma y^2 = \hat{b}^2 \Sigma x^2 + \Sigma e^2 + 0$$

حيث يمكن أن نبـرهن أن الحـد التقـاطعي (Cross Product) $\hat{b}\Sigma xe$ يساوى صـفر وأن $\hat{b}^2\Sigma x^2 = \hat{b}\Sigma xy$. وعليـه Σy^2 تمثـل التبـاين الكلـى و

$\hat{b}^2\sum x^2 = \hat{b}\sum xy$ تمثل التباين المفسر و $\sum e^2$ تمثل التباين غير المفسر. وبالتـــالي يمكن تقسيم التباين الكلى أو مجموع المربعات الكلى إلى **التالي:**

مجموع المربعات الكلى = مجموع المربعات المفسر+مجمـــوع المربعـــات غيـــر المفسر

Unexplained Sum of Squares + Explained Sum of Squares = Total Sum of Squares

$$USS + ESS = TSS$$

ويمكن كتابة جدول تحليل التباين للانحدار البسيط كما يلي:

جدول تحليل التباين لنموذج الانحدار البسيط

المصدر	مجموع المربعات	درجات الحرية	متوســـــــط المربعات	النسبة F
الانحدار X	$\hat{b}^2\sum x^2 = \hat{b}\sum xy$	1	$\hat{b}\sum xy$	$F = \frac{\hat{b}\sum xy}{\sum e^2/(n-2)}$
الباقي	$\sum e^2 = \sum y^2 - \hat{b}\sum xy$	n-2	$\sum e^2/(n-2)$	
المجموع	$\sum y^2 = \hat{b}\sum xy + \sum e^2$	n - 1		

وقيمة F المحسوبة يمكن استخدامها لاختبار معنوية العلاقة عند مستوى معنوية معين ودرجات حرية 1 و (n - 2)

4.2 معامل التحديد البسيط
(Simple Coefficient of Determination)

معامل التحديد البسيط هو مربع معامل الارتباط البسيط r وعليه يرمز له بالرمز r^2 . وفى حالة انحراف القيم من وسطها الحسابي يمكن الحصول على معامل التحديد البسيط كنسبة بين التباين المفسر ESS والتباين الكلى TSS :

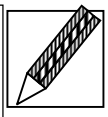
$$r^2 = \left(\frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}} \right)^2 = \frac{\sum xy}{\sum x^2} \cdot \frac{\sum xy}{\sum y^2} = \frac{\hat{b} \sum xy}{\sum y^2} = \frac{ESS}{TSS}$$

وعليه الصيغة التى تستخدم للحصول على معامل التحديد البسيط هي:

$r^2 = \frac{(\sum xy)^2}{\sum x^2 . \sum y^2}$: وعليه قيمة r^2 تقع بين الصفر والواحد الصحيح $(0 \le r^2 \le 1)$

وهى مقياس مهم لجودة توفيق النموذج. فكلما كانت قيمة r^2 قريبة من الواحد فهذا يعنى أن مساهمة المتغير المستقل في تفسير تباين المتغير التابع كبيرة. فمثلاً إذا كانت r^2 = 0.82 . فهذا يعنى أن 82 في المائة من التباين الكلى في المتغير التابع Y يفسرها المتغير المستقل X و 18 في المائة المتبقية تفسرها عوامل أخرى تؤثر على Y ولكن غير مضمنة بصورة صريحة في النموذج.

تدريب (6)



إذا كان لدينا التالي:

التباين الكلي = $\sum\left(Y - \bar{Y} \right)^2 = 38.92$

التباين غير المفسر = $\sum\left(Y - \hat{Y} \right)^2 = 19.7$

جد التباين المفسر ومعامل التحديد البسيط

3. بناء فترة ثقة واختبار المعنوية لمعالم النموذج البسيط

استخدمنا صيغ مختلفة للحصول على تقدير لمعلمتي النموذج a و b . وكل هذا الصيغ تعطى تقديراً بنقطة للمعلمين . ويمكن الحصول على التقدير بفترة للمعلمين بنفس الطريقة التي تحدثنا عنها في الوحدة (الثانية) . فالتقدير بفترة للمعلمين يمكن الحصول عليه باستخدام الصيغتين التاليتين:

$$\hat{a} \pm t_{\alpha/2} \cdot s.e\left(\hat{a} \right)$$

$$\hat{b} \pm t_{\alpha/2} \cdot s.e\left(\hat{b} \right)$$

وتمثل $t_{\alpha/2}$ القيمة الجدولية وتعتمد على درجات الحرية ومستوى المعنوية المعطاة. وترمز s.e إلى الخطأ المعياري للمعلمين $\hat{a}$ و $\hat{b}$ وهما الجذر التربيعي لتباييني المعلمين الذي تمت الإشارة إليه من قبل.

ولاختيار معنوية كل من المعلمين a و b نتبع نفس الخطوات التي تم ذكرها في الوحدة (الثالثة) وهى: صياغة فرض العدم والفرض البديل،وتحديد مستوى المعنوية. وتحديد إحصاءة الاختبار ونستخدم هنا إحصاءة t للحصول على القيمة المحسوبة ثم نحصل على القيمة الحرجة وهى تعتمد على نوع الفرض البديل ومستوى المعنوية ودرجات الحرية وأخيراً مقارنة القيمة المحسوبة مع القيمة الحرجة لاتخاذ القرار السليم بالرفض أو عدم رفض فرض العدم.

ونحصل على القيمة المحسوبة لاختبار المعلمين كما يلي:

(أ) القاطع A

$$t = \frac{\hat{a}}{S.e\left(\hat{a}\right)}$$

حيث أن

$$S.e\left(\hat{a}\right) = \sqrt{\frac{\sigma_u^2 \sum X^2}{n\sum x^2}}$$

(ب) الميل b

$$t = \frac{\hat{b}}{S.e\left(\hat{b}\right)}$$

حيث أن

$$S.e\left(\hat{b}\right) = \sqrt{\frac{\hat{\sigma}_u^2}{\sum x^2}}$$

ونذكر هنا أن تقدير تباين المتغير العشوائي $\hat{\sigma}_u^2$ يمكن الحصول عليه في كلتا الحالتين كما يلي:

$$\hat{\sigma}_u^2 = \frac{\sum y^2 - b\hat{\sum} xy}{n-2}$$

مثال

لنفترض أن لدينا البيانات التالية عن استهلاك فرد (Y) ودخله المتاح (X) بالملاين في الفترة من 1999 حتى 2003 كما يبين الجدول التالي:

السنة	Y	X
1999	2	5
2000	3	4
2001	4	8
2002	7	9
2003	4	9

المطلوب:

(أ) رسم لوحة الانتشار.

(ب) الحصول على تقدير معادلة الانحدار $\hat{Y} = \hat{a} + \hat{b}X$.

(ج) إيجاد فترة ثقة للمعلمة b عند مستوى معنوية 5 في المائة.

(د) اختبار معنوية المعلمة a عند مستوى معنوية 5 في المائة.

(هـ) اختبار معنوية العلاقة عند مستوى معنوية 5 في المائة.

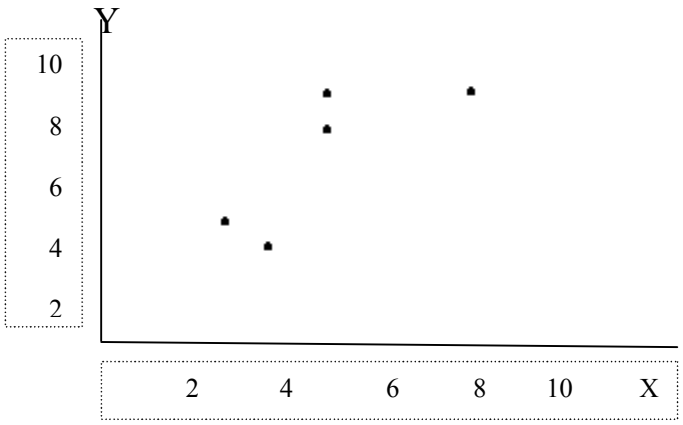
(و) الحصول على معامل التحديد البسيط r^2 وتفسيره.

(ز) التنبؤ بقيمة Y في العام 2004 إذا كانت X في نفس العام هي 10.

الحل

السنة	Y	X	$y = Y - \bar{Y}$	$x = X - \bar{X}$	xy	x^2	y^2	X^2
1999	2	5	-2	-2	4	4	4	25
2000	3	4	-1	-3	3	9	1	16
2001	4	8	0	1	0	1	0	64
2002	7	9	3	2	6	4	9	81
2003	4	9	0	2	0	4	0	81
	20	35			13	22	14	267

(أ)لوحة الانتشار



وبقراءة لوحة الانتشار نتوقع علاقة طردية بين Y و X أي نتوقع أن تكون إشارة الميل b موجبة.

(ب)

$$\bar{y} = \frac{\sum Y}{n} = \frac{20}{5} = 5$$

$$\bar{x} = \frac{\sum X}{n} = \frac{35}{5} = 7$$

$$\hat{b} = \frac{\sum xy}{\sum x^2} = \frac{13}{22} = 0.59$$

$$\hat{a} = 4 - (0.59)(7) = 4 - 4.13 = -0.13$$

وعليه معادلة انحدار Y على X

$$\hat{Y} = -0.13 + 0.59X$$

(ج) لإيجاد فترة الثقة ينبغي الحصول على الخطأ المعياري للمعلمة المقدرة $\hat{b}$

$$S.e\left(\hat{b}\right)=\sqrt{\frac{\hat{\sigma}_u^2}{\sum x^2}}$$

$$\hat{\sigma}_u^2=\frac{\sum y^2-\hat{b}\sum xy}{n-2}=\frac{14-(0.59)(13)}{5-2}=2.11$$

$$S.e\left(\hat{b}\right)=\sqrt{\frac{2.11}{22}}=\sqrt{0.090}=0.31$$

وعليه فترة الثقة للمعلم b عند مستوى ثقة 95 في المائة هي

$$0.59\pm 3.182(0.31)=0.59\pm 0.99=(-0.40\ ,\ 1.58)$$

(د) لاختيار معنوية المعلمة a عند مستوى معنوية 5 في المائة

$$H_0: a=0$$

$$H_1: a\neq 0$$

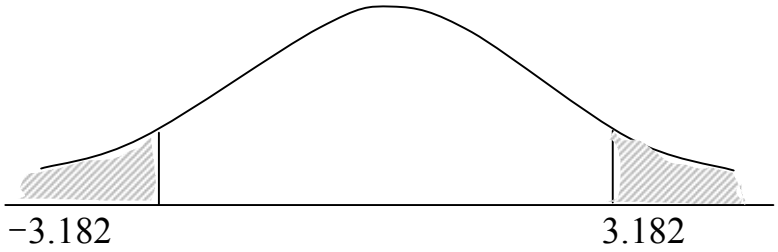
$$t=\frac{\hat{a}}{S.e\left(\hat{a}\right)}$$

$$Var\left(\hat{a}\right)=\frac{\hat{\sigma}_u^2\sum X^2}{n\sum x^2}=\frac{(2.11)(267)}{(5)(22)}=\frac{563.37}{110}=5.12$$

$$S.e\left(\hat{a}\right)=\sqrt{5.12}=2.26$$

$$t = -\frac{0.13}{2.26} = -0.058$$

وقيمة t الحرجة أو الجدولية تساوى 3.182



وبما آن قيمة t المحسوبة اقل من قيمة t الجدولية بإهمال الإشارة نقبل فرضية العدم H_o بأن a غير معنوية إحصائيا.

(هـــ) لاختبار معنوية العلاقة ينبغي انشاء جدول تحليل تباين

المصدر	مجموع المربعات	درجات الحرجة	متوسط المربعات	
الانحـــدار X	$\hat{b}\sum xy = (0.59)(13) = 7.67$	**1**	$\hat{b}\sum xy/1 = 7.67$	$F = \frac{7.67}{2.11} = 3.64$
الباقي	$\sum e^2 == \sum y^2 - \hat{b}\sum xy = 14 - 7.67 = 6.33$	**n-2=5-2=3**	$\sum e^2/(n-2) = \frac{6.33}{3} = 2.11$	–
المجموع	$\sum y^2 = 14$	**4**		–

$$H_0 : b = 0$$
$$H_1 : \ b \neq 0$$

وبما أننا نتعامل مع نموذج بسيط فان فرضية اختبار العلاقة هي نفسها اختبار معنوية معامل المتغير المستقل X. وفى هذه الحالة يمكن الحصول على t بإيجاد الجذر التربيعي لقيمة Fالمحسوبة $t = \sqrt{3.64} = 1.91$. وهذا يعنى عند مستوى معنوية 5 في المائة نقبل فرضية العدم H_0 حيث أن القيمة المحسوبة 1.91 اقل من القيمة الجدولية 3.182. من الخلاصتين عن a و b فانه لا توجد علاقة خطية بين X و Y عند مستوي دلالة 5%.

$$r^2 = \frac{(\sum xy)^2}{\sum x^2 . \sum y^2} = \frac{(13)^2}{(22)(14)} = \frac{169}{286} = 0.59$$

وهذا يعنى أن 59 في المائة من التغيرات التي تحدث في المتغير التابع Y تفسر بالتغيرات التي تحدث في المتغير المستقل X و41 في المائة المتبقية ناتجة من تأثير العوامل الاخرى المحذوفة من الطرف الأيمن.

(ز)

قيمة Y في العام 2004 هي:

$$Y_{2004} = -0.13 + 0.59(10)$$

$$= -0.13+5.9$$

$$= 5.77$$

أسئلة التقويم الذاتي (1)

(1)وضح الفرق بين تحليل الارتباط وتحليل الانحدار.

(2) أشرح لماذا توجد معادلتا انحدار في حالة الانحدار الخطـــي البسيط. وتحت أية شروط توجد معادلة انحدار واحدة.

(3) وضح الفرق بين معامل الارتباط الخطي ومعامل التحديد.

(4) في سجلات مبتورة لتحليل الارتباط، وجدت المعلومات التالية:

تباين $X = 9$

معادلتا الانحدار:

$4X- \quad 5Y+33=0$

$20X-9Y- \quad 107-0$

جد:

(أ) وسط X ووسط Y

(ب) الانحراف المعياري للمتغير Y

(ج) معامل الارتباط الخطي.

الخلاصة

تناولت الوحدة موضوع الارتباط فعرفنا أن $\sum(x-\hat{x})(y-\hat{y})$ هو مقياس تقريبي للارتباط لأنه له عيبين: 1- يتأثر بعدد المفردات 2- يتأثر بوحدات القياس . وأن التغاير هو مقياس من مقاييس الارتباط ويعطي بالمعادلة $\text{cov (x,y)} = \frac{\sum(x-\hat{x})(y-\hat{y})}{n}$ وأن أهم مقاييس الارتباط هو معامل الارتباط الخطي r لأنه تخلص من عيوب المقياس التقريبي حيث :

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

و وضحنا أن $-1 \le r \le 1$ ففي حالة الارتباط الموجب التام $r = 1$ ، وفي الارتباط السالب التام $r = -1$ وأنه بناءً على قيمة r يمكن أن نصف العلاقة بين المتغيرين x , y .

أما في درس الانحدار فقد وضحنا أننا نعرف العلاقة بين المتغيرين في الغالب بأنها خطية حيث $y=a+bx$ ، $\hat{a} = \frac{\sum y \sum x^2 - (\sum x)(\sum xy)}{n\sum x^2 - (\sum x)^2}$

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

أو $\hat{a} = \hat{y} - \hat{b}\,\hat{x}$ ، $\hat{b} = \frac{\sum xy}{\sum x^2}$

كذلك في هذه الوحدة وضحنا أن التباين الكلي TSS ، مجموع المربعات المفسر ESS ، مجموع المربعات غير المفسر USS ، و وجدنا أيضاً أن $r^2=\frac{ESS}{TSS}$ ، وعرفنـــا أن الاحصاءة تستخدم لاختبار المعالم a , b للنموذج البسيط .

لمحة مسبقة عن الوحدة التالية

في هذه الوحدة عالجنا اختبارات تقدير معالم واختبارات فرضـــيات المجتمـــع المجهولة التوزيع مما يعني أنها قد لا تتبع التوزيع الطبيعي أو المتجمعات ذات الإلتواء الواضح أو في حالة العينات الصغيرة أيضاً

إجابات التدريبات

تدريب (1)

درجات الإحصاء	درجات الرياضيات					
x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})(y-\bar{y})$	$(x-\bar{x})^2$	$(y-\bar{y})^2$
15	20	−20.875	−18.75	391.406	435.766	351.563
28	30	−7.875	−8.75	68.906	62.016	76.563
20	50	−15.875	11.25	−178.594	252.016	126.563
14	40	−21.875	1.25	−27.344	478.516	1.563
40	60	4.125	21.25	87.566	17.016	451.563
60	30	24.125	−8.75	−211.094	582.016	76.563
80	70	44.125	31.25	1378.906	1947.016	976.563
30	10	−5.875	−28.75	168.906	43.516	826.563
287	310			1678.568	3808.878	2887.504

$$\bar{x} = \frac{287}{8} = 35.875$$

$$\bar{y} = \frac{310}{8} = 38.75$$

$$\text{cov}(x,y) = \frac{\sum (y-\bar{y})(x-\bar{x})}{n} = \frac{1678.568}{8} = 209.822$$

إذن الارتباط موجب

تدريب (2)

معامل الارتباط

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

$$= \frac{1678.568}{\sqrt{(2887.504) - (3808.878)}}$$

$$= \frac{1678.568}{3316.346} = 0.506$$

يوجد ارتباط موجب معتدل لأن $0.5 \leq r \leq 0.7$

تدريب (3)

x y	y^2	x^2	Y	X
224	784	64	28	8
345	526	225	23	15
400	400	400	20	20
375	225	625	15	25
360	144	900	12	30
200	100	400	10	20
1904	2182	2614	108	118

$$r = \frac{(6)(1904) - (108)(118)}{\sqrt{6(2614) - (118)^2 \quad 6(2182)\text{-}(108)^2}}$$

$$= \frac{11424 - 12744}{\sqrt{[6(2614) - (13924)] - [6(2182)\text{-}(11644)]}}$$

$$= \frac{-1320}{\sqrt{1760 \times 1448}} = -0.83$$

ارتباط سالب قوي .

تدريب (4)

معامل الارتباط قبل تصحيح الخطأ

$$r = \frac{\sum x\,y - (\sum x)(\sum y)}{\sqrt{[n\sum x^2\text{-}(\sum x)^2] \quad [n\sum y^2\text{-}(\sum y)^2]}}$$

$$r = \frac{12(334) - (30)(5)}{\sqrt{[12(670) - (30)^2] - [12(285)\text{-}(5)^2]}}$$

$$r = \frac{3858}{4923.44} = 0.78$$

بعد تصحيح الخطأ

$$\sum x\ y = 334 - 44 + 140 = 430$$

$$\sum x = 30 - 11 + 10 = 29$$

$$\sum y = 5 - 4 + 14 = 15$$

$$\sum x^2 = 670 - 121 + 100 = 649$$

$$\sum y^2 = 285 - 16 + 196 = 465$$

$$= \frac{12(430) - (29)(15)}{\sqrt{[12(649) - (29)^2] - [12(465) - (15)^2]}}$$

$$= \frac{4725}{6099.44} = 0.77$$

تدريب (5)

معادلتا الانحدار

لانحدار y على x نجد أن

$$\sum y_i = n\hat{a} + \hat{b} \sum x_i$$

$$\sum x_i y_i = \hat{a} \sum x_i + \hat{b} \sum x_i^2$$

لانحدار x على y نجد أن

$$\sum x_i = n\hat{a} + \hat{b} \sum y_i$$

$$\sum x_i y_i = \hat{a} \sum y_i + \hat{b} \sum y_i^2$$

$$\sum x = 700$$
$$\sum y = 900$$
$$\sum x^2 = 51868$$
$$\sum x\ y = 66900$$

باستخدام طريقة المحددات

معادلات المربعات الصفري لانحدار y على x .

$$\hat{a} = \frac{\begin{vmatrix} \sum x & \sum y \\ \sum x^2 & \sum x\,y \end{vmatrix}}{\begin{vmatrix} \sum x & n \\ \sum x^2 & \sum x \end{vmatrix}} = \frac{(900)(51868) - (66900)(700)}{(10)(51868) - (700)^2}$$

$$= \frac{-148800}{28680} = -5.2$$

$$\hat{b} = \frac{\begin{vmatrix} \sum y & n \\ \sum x\ y & \sum x \end{vmatrix}}{\begin{vmatrix} \sum x & n \\ \sum x^2 & \sum x \end{vmatrix}} = \frac{(10)(66900) - (700)(900)}{(10)(51868) - (700)^2}$$

$$= \frac{66900 - 630000}{28680} = \frac{39000}{28680} = 1.4$$

$$y = \hat{a} + \hat{b}\ x$$

معادلة انحدار y على x هي :

$$y = 1.4\ x - 5.2$$

وبالمثل يمكن إيجاد انحدار x على y حيث $y = \hat{a} + \hat{b}\ y$

$$\hat{b} = \frac{\begin{vmatrix} \sum \mathrm{x} & n \\ \sum x\ y & \sum \mathrm{y} \end{vmatrix}}{\begin{vmatrix} \sum \mathrm{y} & n \\ \sum y^2 & \sum \mathrm{y} \end{vmatrix}} \qquad \text{و} \qquad \hat{a} = \frac{\begin{vmatrix} \sum y & \sum \mathrm{x} \\ \sum y^2 & \sum \mathrm{x\,y} \end{vmatrix}}{\begin{vmatrix} \sum \mathrm{y} & n \\ \sum y^2 & \sum \mathrm{y} \end{vmatrix}}$$

تدريب (6)

$$\mathrm{USS} + \mathrm{ESS} = \mathrm{TSS}$$
$$\mathrm{ESS} = 38.92 - 19.7$$
$$= 19.22$$

معامل التحديد البسيط

$$r^2 = \frac{ESS}{TSS} \quad = \frac{19.22}{38.92} = 0.49$$

معامل الارتباط

$$r = \sqrt{0.49} = 0.7$$

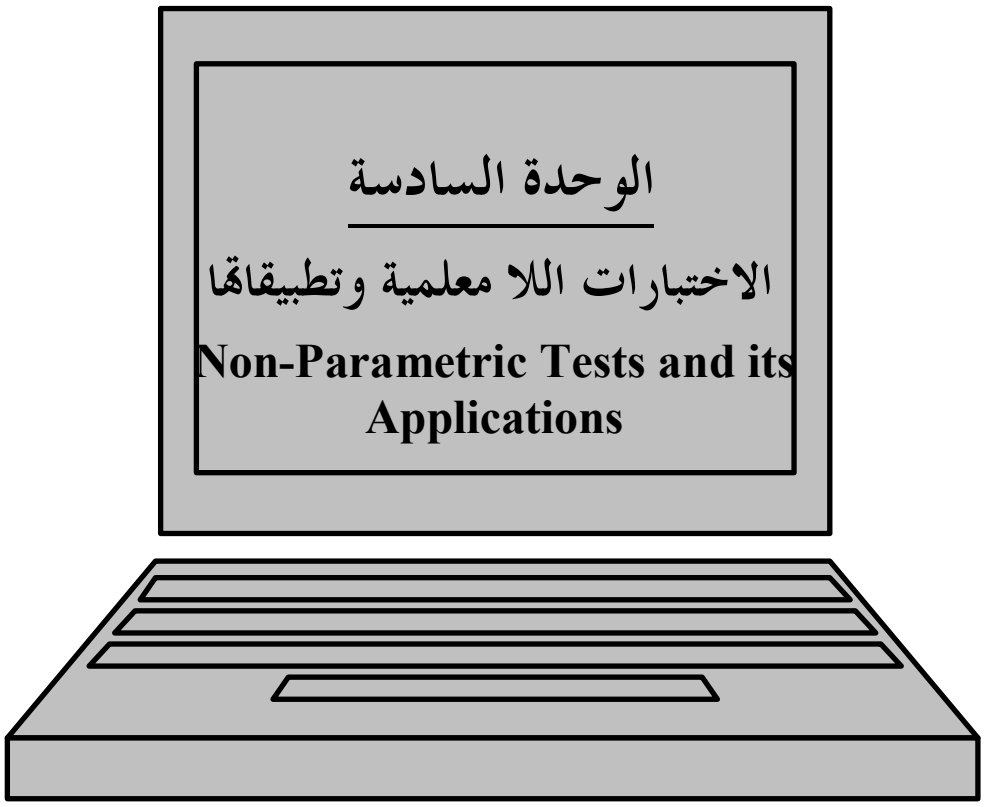
مسرد المصطلحات

- **الارتباط (Correlation)** : هو قياس درجة أو مدي قوة العلاقة بين متغيرين.
- **التغاير (Covariance)** : يعرف التغاير بين المتغيرين X و Y بأنه الوسط الحسابي لحاصل ضرب انحرافات قيم X عن وسطها الحسابي في انحرافات قيم Y عن وسطها الحسابي.
- **معامل الارتباط الخطي (Linear Correlation Coefficient)** : نسبة بين التغاير بين متغيرين X و Y والانحراف المعياري للمتغيرين.
- **الانحدار (Regression)** : مقياس لمتوسط العلاقة بين متغيرين أو أكثر من خلال الوحدات الأصلية للبيانات.
- **البواقي (Residuals)** : انحراف الخط المقدر من جميع القيم المشاهدة ويحسب بطرح القيم المقدرة $\hat{Y}$ من القيم الفعلية Y .
- **جودة التوفيق (Goodness of fit)** : مؤشر لتوضيح مدي اقتراب القيم المقدرة (المحسوبة) باستخدام الخط المقدر من القيم الفعلية.
- **معامل التحديد (Coefficient of Determination)** : هو مربع معامل الارتباط بين المتغير التابع والمتغير المستقل ويقيس مدي نجاح خط الانحدار المقدر في وصف المشاهدات (Observations)
- **التنبؤ (Forecasting)** : إيجاد قيمة المتغير التابع y_0 المناظرة لقيمة معطاة للمتغير المستقل x_0 حيث يفترض أنها صحيحة لا تحتوي على خطأ.

المراجع

- مواري د. شبيجل، "الإحصاء" ، سلسلة ملخصات شوم، الدار الدولية للنشر والتوزيع 1991.

- بول ج.هويل، " المبادئ الأولية في الإحصاء، ترجمة ومراجعة د. بدرية شوقي عبد الوهاب والدكتور محمد كامل الشربيني ، دار جون وايلي وأبنائه.


الوحدة السادسة
الاختبارات اللا معلمية وتطبيقاتها
Non-Parametric Tests and its Applications

محتويات الوحدة

الموضوع	الصفحة
المقدمة	241
تمهيد	241
أهداف الوحدة	242
1.الاختبارات اللا معلمية وتطبيقاتها	243
1.1 متطلبات الاختبارات	243
2.1 محاسن الاختبارات اللامعلمية	245
3.1 مساوئ الاختبارات اللامعلمية	246
2. اختبارات تتضمن إشارات الفروق	247
1.2 اختبار الإشارات	248
2.2 اختباراشارة الرتبة (اختبار ولكوكسن)	257
3. اختبارات تتضمن الرتب	262
1.3 اختبار ولكوكسن لمجموع الرتب	262
2.3 اختبار كروسكال - واليس	268
3.3 اختبار مان–وتنى	273
الخلاصة	281
إجابات التدريبات	282
مسرد المصطلحات	288
المراجع	289

المقدمة

تمهيد

هذه هي الوحدة السادسة في هذا المقرر و قد أوضحت الوحدة الحالات التي تطبق فيها الاختبارات اللا معلمية كما قارنت هذه الوحدة بين الطرق المعلمية و الطرق اللا معلمية للاختبارات .

تناولنا أيضاً في هذه الوحدة الاختبارات التي تبنى علي أساس إشارات الفروق فقط والاختبارات التي تراعي إشارات الفروق بالإضافة إلي ترتيب هذه الفروق وتناولنا أيضاً من خلال الوحدة الاختبارات التي تبني على أساس مجموع الرتب .

كما ناقشنا الحالات التي تستخدم فيها كل من هذه الاختبارات وأوضحنا مزايا وعيوب كل منها واختتمنا الوحدة باختبار كروسكال – أليس . كذلك تضمنت الوحدة بعض التدريبات وأسئلة التقويم الذاتي. نأمل أن يكون مدي استيعابك و تحصيلك لهذه الوحدة ممتازاً.

أهداف الوحدة



ينبغي عزيزي الدارس بعد فراغك من هذه الوحدة أ ن تكون قادراً على أن:

(1) تعرف الاختبار اللا معلمي

(2) تذكر نماذج لاختبارات لا معلمية

(3) تشرح الحالات التي تستخدم فيها الاختبارات اللا معلمية

(4) تعدد محاسن الاختبارات اللا معلمية

(5) تعدد مساوئ الاختبارات اللا معلمية

(6) تطبق علي اختبار الاشارة للعينة الواحدة

(7) تطبق علي اختبار الاشارة لعينة ذات زوج

(8) تطبق علي اختبار وليكلسون لاشارات الفروق وحجمها

(9) تطبق علي اختبارات وليكلسون لمجموع الرتب

(10) تحلل التباين لإختبار فرضية تساوي متوسطات لمجتمعات لا تتبع التوزيع الطبيعي

1.الاختبارات اللا معلمية وتطبيقاتها

1.1 متطلبات الاختبارات

عزيزي الدارس تفترض الطرق المعلمية (Parametric Methods) عند التقدير واختبارات المعنوية أن توزيع المجتمع معلوم، عادة الطبيعي، ولكن هناك أوضاع قد لا يستوفى فيها هذا الفرض الأساسي. ويضطر الباحث في هذه الحالة إلى استخدام الطرق غير المعلمية (Nonparametric Methods) . ولقد رأينا أن معظم الاختبارات الإحصائية التي تعرضنا لها في الوحدات السابقة تتطلب فروض أساسية ينبغي استيفائها لتطبيقها بصورة سليمة. واحد من هذه الفروض هو أن يتبع المجتمع نمط التوزيع الطبيعي. وتسمح هذه الاختبارات إلى حد كبير بالتجاوز عن فرض التوزيع الطبيعي إذا كانت العينات المسحوبة من هذه المجتمعات كبيرة كما أوضحنا سابقاً. ولكن إذا كانت العينة مسحوبة من مجتمع واضح الالتواء أو غير طبيعي وحجم هذه العينة صغير فان الاختبارات الإحصائية القائمة عليها سوف لا تعطى نتائج يمكن الاعتماد عليها.

وتقوم هذه الاختبارات أيضا على فرض آخر هو أن معظمها صيغ بفهم أن القيم المجهولة لمعالم المجتمع يمكن تقديرها من عينات اختيرت بصورة عشوائية من هذا المجتمع. وعليه في هذه الحالة يمكن الحصول على تقديرات جيدة لمعالم المجتمع من إحصاءات العينة المقابلة لها. ولا نستطيع الحصول على تقديرات جيدة في حالة البيانات الوصفية (nominal data) للطبيعة المختلفة لمثل هذا النوع من البيانات.

وبما أن هناك كثير من الأوضاع التي لا تسمح بتوفير مثل هذه الفروض، فقد أبتدع علماء الإحصاء طرقاً بديلة يمكن استخدامها لاختبار فروض (Hypotheses) باستخدام بيانات لا تتبع نمط التوزيع الطبيعي أو لا يمكن الحصول على إحصاءات عينة يمكن الركون إليها . وتسمى هذه الطرق بالاختبارات حرة التوزيع (Distribution – Free Tests) أو الاختبارات اللامعلمية لأنها لا تعتمد على معالم المجتمع. وعليه ينبغي أن ننبه بان المصطلحان غير متطابقين تماماً. فلفظة توزيع حر تصف طريقة لاختبار فرض أو الحصول على تقدير فترة وهذه الطريقة لا تعتمد على طبيعة توزيع المجتمع تحت الدراسة. ومن جهة أخرى، تعنى لفظة لا معلمي أن الفرض قد تمت صياغته ليس بدلالة قيمة تخمينية معينة لمعالم المجتمع. و لكن قد جرت العادة على استخدام مصطلح لا معلمي للإشارة لكلتا الاختبارين.

➢ **ملاحظة**

وتجدر الإشارة أيضا إلى انه لا يجوز استخدام الاختبارات اللامعلمية إذا توافرت متطلبات الاختبارات المعلمية. ويرجع سبب ذلك إلى أن قوة (Power) الاختبارات اللامعلمية ضعيفة مقارنة بالاختبارات المعلمية. ولكن في حالة استخدام الاختبارات المعلمية يستحسن سحب عينة كبيرة لأنه كل ما زاد حجم العينة زادت قوة الاختبار عند مستوى معنوية معين.

2.1 محاسن الاختبارات اللامعلمية

عزيزي الدارس يمكن أن نجمل مزايا الاختبارات اللامعلمية في النقاط الآتية:

(أ) لا يتطلب تطبيق الاختبارات معرفة توزيع المجتمع الذي تم سحب العينة منه فيما إذا كان يتبع نمط التوزيع الطبيعي أم خلافه.

(ب) سهلة الفهم والتطبيق خاصة إذا كان حجم العينة صغير.

(ج) لا تتطلب معظم الاختبارات اللامعلمية حسابات مطولة ومعقدة ولهذا السبب فهي لا تستغرق كثيراً من الوقت.

(د) تطبيق الاختبارات اللامعلمية على كل أنواع البيانات سواء كانت خاضعة لقياس كمي أو نوعي مثل البيانات الترتيبية والبيانات الوصفية.

(هـ) يمكن استخدام الطرق اللامعلمية في حالة العينات بالغة الصغر خاصة في مجالات الأبحاث التي لا يمكن الحصول فيها على عدد كبير من البيانات مثل المجالات الطبية والصحية.

(و) تقوم الطرق اللامعلمية على فرضيات محدودة وغير صارمة مقارنة مع الطرق المعلمية.

3.1 مساوئ الاختبارات اللامعلمية

كما علمنا مزايا الاختبارات المعلمية ينبغي أن نعلم ايضاً عيوب هذه الاختبارات والتي يمكن ان نلخصها كما يلي :

(أ) تميل الطرق اللامعلمية نحو إهدار المعلومات بسبب تحويل البيانات الكمية الحقيقية إلى بيانات نوعية.

(ب) تتمتع الطرق اللامعلمية بحساسية اقل مقارنة بالطرق المعلمية المقابلة لها الأمر الذي يترتب عليه الحاجة الي الحصول على دليل قوى قبل رفض فرضية العدم. وتجدر الإشارة هنا إلى أن كفاءة (Efficiency) الطريقة اللامعلمية تعتبر مقياس جيد لدرجة حساسيتها. وتؤدى زيادة حجم العينة إلى التغلب على تدنى حساسية الطرق اللامعلمية. ونلاحظ من الجدول أدناه أن تدنى الحساسية لا يشكل عائق كبير على استخدامها:

جدول يبين كفاءة الطرق اللامعلمية

التطبيق	الاختبار المعلمى	الاختبار اللامعلمى	كفـــاءة الاختبـــار اللامعلمـــى مـــع مجتمـــع طبيعـــي (normal)
عينتان مستقلتان	اختبار (t) أو (z)	اختبار الإشارة ولكوكسن لإشارات الرتب	0.63 0.95
عينتان غير مستقلتين	اختبار (t) أو (z)	ولكوكسن لمجموع الرتب	0.95
عدد من العينات المستقلة	تحليـــل التبـــاين (اختبار F)	اختبـــار كروســـكل – واليس	0.95
الارتباط	الارتباط الخطى	ارتباط الرتب	0.91

2. اختبارات تتضمن إشارات الفروق
(Test Involving Sign of Differences)

إذا تمت مقارنة عينتين لتحديد فيما إذا كان هناك فرق معنـــوي بـــين وســـطين لمجتمعين، فانه يمكن استخدام أشارت الفروق بين أزواج مفردات العينتين لاختبار هـــذا الفرض. وسنتناول نوعين من الاختبارات التي تشتمل على الإشارات:

1) **اختبار الإشارات:** (Sign Test)
2) **اختبار إشارة الرتبة** (Signed - rank test)

1.2 اختبار الإشارات (Sign Test)

سبق أن ناقشنا اختبار ت (t - Test) وذكرنا أن المجتمع المسحوبة منه العينة العشوائية ينبغي أن يتبع نمط التوزيع الطبيعي. وعند اختبارنا، أيضا، للفرق بين وسطي مجتمعين افترضنا أن تباييني المجتمعين متساويان. وقد لا يتحقق أحد الفرضين أو كلاهما في كثير من الأحوال الأمر الذي يترتب عليه صعوبة استخدام اختبار ت . إذا كان المتغيران مرتبطان (بمعني كل قيمة ملاحظة للمتغير تقابلها قيمة للمتغير الأخر) فيمكن استخدام ما يسمي باختبار الإشارة.

ويستخدم اختبار الإشارة فقط إشارات الفروق بين أزواج المشاهدات. ويقدم هذا الاختبار بديل مريح لاختبار الفروض الخاص بالفرق بين وسطين عندما تكون العينتان تحت الدراسة غير مستقلتين أي إحداهما يعتمد على الأخر. وكما يوحي اسم الاختبار فانه يعتمد على اتجاه الإشارة (موجبة أم سالبة) كما أسلفنا للفرق بين زوجين من المفردات وليس مقدار حجم هذا الفرق. وعليه لاستخدام اختبار الإشارة نحسب:

- عدد الإشارات الموجبة (+)
- عدد الإشارات السالبة (−)
- عدد الأصفار (حيث لا تعتبر موجبة أو سالبة)

وتصاغ فرضية العدم كما **يلي:**

$$H_O : \quad p = 0.5$$

أي احتمال إشارة موجبه يساوي احتمال إشارة سالبه.

وإذا كان الفرق بين مفردتي الزوج ناتج عن طريق الصدفة فان احتمال الإشارة الموجبة (+) لأي زوج معين يساوى 0.5 كما احتمال الإشارة السالبة (−). وإذا إفترضنا أن S هي العدد الأقل من الإشارتين ، فان S تتبع نمط توزيع ذي الحدين (Binomial distribution) باحتمال مقداره 0.5 أي $P = 0.5$ و دليل n .

ويمكن الحصول على القيمة الحرجة (Critical Value) لاختبار من طرفين عند مستوى معنوية 5 في المائة أي $\alpha = 0.05$

$$K = \frac{(n-1)}{2} - (0.98)\sqrt{n}$$

حيث أن n تمثل حجم العينة.

وتأسيسا على ذلك، يمكن رفض فرضية العدم (H_O) إذا كانت S اقل من أو تساوى $(S \le K)$ k لاختبار الإشارة.

ويوجد نوعان لاختبار الإشارة:

(أ) **اختبار الإشارة للعينة الواحدة** (One – Sample sign test).

(ب) **اختبار الإشارة لعينة ذات زوج** (Paired – Sample sign test).

(أ) اختبار الإشارة للعينة الواحدة

وتتم عملية الاختبار للعينة الواحدة كما يلي: نفترض أن فرض العدم هي $u = u_O$ ثم يصاغ الفرض البديل حسب ما هو مطلوب للاختبار باستخدام عينة عشوائية حجمها n وتستبدل أية قيمة مفردة في العينة اكبر من u_O بإشارة موجبة وأية قيمة مفردة في العينة اقل من u_O بإشارة سالبة ونتجاهل قيم مفردات العينة التي تساوى u_O ثم نختبر فرضية العدم بأن الإشارات الموجبة والسالبة قيم لمتغير عشوائي يتبع نمط توزيع ذي الحدين باحتمال P=0 .5.

مثال

ادعى مركز لتعليم الطباعة في ثمانينات القرن المنصرم بأنه يمكن تدريب الشخص في فترة شهرين على الطباعة بمتوسط لا يقل عن 60 كلمة في الدقيقة. وقد تم اختيار عينة عشوائية مكونة من 15 متدرباً وأجلسوهم لاختبار طباعة عند نهاية

فترة التدريب. يعطى الجدول التالي عدد الكلمات التي تمت طباعتها في الدقيقة لكل واحد من المتدربين. اختبر الفرض أن وسط سرعة الطباعة للمتدربين على الأقل 60 كلمة في الدقيقة.

المتدرب	الكلمات في الدقيقة	الفرق
(أ)	78	+18
(ب)	66	+6
(ت)	51	−9
(ث)	73	+13
(ج)	64	+4
(ح)	58	−2
(خ)	80	+20
(د)	75	+15
(ز)	84	+24
(س)	61	+1
(ش)	56	−4
(ص)	77	+17
(ض)	54	−6
(ط)	57	−3
(ز)	60	0
		S = 5

الحل:

الخطوة (1):

صياغة فرض العدم والفرض البديل:

$$H_o : u = 60$$

$$H_1 : u \rangle 60$$

الخطوة (2):

تحديد مستوى المعنوية :

$$\alpha = 0.05$$

الخطوة (3):

عدد الإشارات الموجبة (+)	= 9
عدد الإشارات السالبة (−)	= 5
عدد الأصفار	= 1
المجموع	15

وبما أن عدد الإشارات الموجبة لا يتساوى مع عـــدد الإشـــارات الســـالبة،فنريد اختبار أن ذلك ناتج عن طريق الصدفة. وعليه عدد الإشارات الأقل S (يســـاوى 5) يتبع نمط توزيع ذي الحدين باحتمال 0.5. وعليه يمكن قراءة القيمة الحرجة من جـــدول القيم الحرجة أو حسابه بالقاعدة أعلاه، ثم مقارنتة بقيمة S .

$$K = \frac{(n-1)}{2} - (0.98)\sqrt{n}$$

حيث تمثل n مجموع عدد الإشارات الموجبة والسالبة

$$K = \frac{(14-1)}{2} - (0.98)\sqrt{14} = 2.83$$

وبما أن S اكبر من K نقبل فرض العدم.

(لاحظ أن القبول يتم إذا كانت القيمة المشاهدة اكبر من K)

(ب) اختبار الإشارة لعينة ذات زوج

اختبار الإشارة له تطبيقات هامة، أيضا، عندما تشكل بيانات زوجية الأساس لاختبار فرضية يحتوى على عينتين غير مستقلتين (Two dependent Samples).

» مثال

لنفترض أن شركة دال ترغب في معرفة إذا كانت الإجازات السنوية تساعد علي ارتفاع إنتاجية العامل ولتحقيق هذه الرغبة بدأت الشركة في تجميع بيانات عن الإنتاج الأسبوعي لمجموعة مكونة من 22 عاملاً للأسبوع قبل بداية إجازة العامل والأسبوع بعد انتهاء إجازته. ولنفترض أن X تمثل الإنتاج الأسبوعي قبل بداية الإجازة Y تمثل الإنتاج الأسبوعي بعد انتهاء الإجازة كما هو موضح في الجدول أدناه. والمطلوب هو اختبار فرضية أن فترة الإجازة ليست لها تأثير فعال على إنتاجية العامل؟

أرقام العمل	قبل الإجازة (x)	بعد الإجازة (y)	إشارة الفرق (y-x)
(1)	78	74	–
(2)	81	85	+
(3)	74	72	–
(4)	85	87	+
(5)	80	82	+
(6)	73	73	0
(7)	96	92	–
(8)	73	76	+
(9)	79	81	+
(10)	80	84	+
(11)	81	77	–
(12)	83	80	–
(13)	92	90	–
(14)	86	88	+
(15)	70	78	+
(16)	90	85	–
(17)	78	74	–
(18)	82	89	+
(19)	81	81	0
(20)	89	92	+
(21)	84	87	+
(22)	72	79	+

الخطوة (1) صياغة فرض العدم والفرض البديل:

H_O : الإجازة السنوية ليس لها تأثير فعال على إنتاجية العامل.

H_1 : الإجازة السنوية لها تأثير فعال على إنتاجية العامل .

الخطوة (2) :

عدد الإشارات الموجبة (+) = 12

عدد الإشارات السالبة (−) = 8

عدد الأصفار = 2

وعليه $n = 20 \ , \ S = 8$

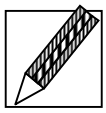
$$K = \frac{(n-1)}{2} - (0.98)\sqrt{n}$$

$$= \frac{20-1}{2} - (0.98)\sqrt{20}$$

$$= 9.5 - (0.98)(4.47) = 4.38$$

وبما آن S > K تقبل فرض العدم.

تدريب (1)



استخدم اختبار الإشارة لتحديد فيما إذا كان هناك فـــرق في برنامج لتحصيل مبالغ مالية قبل وبعد أدخال تعـــديل عليه، عليه استخدم مستوي دلالة 5%.

قبل التعديل	بعد التعديل
30	32
28	29
34	33
35	32
40	37
42	43
33	40
38	41
34	37
45	44
28	27
27	33
25	30
41	38
36	36

2.2 اختبار إشارة الرتبة (اختبار ولكوكسن) (WilcoxonTest)

يتميز اختبار ولكوكسن بأنه يستخدم إشارات الفروق إلى جانب حجمها. ومن هنا يظهر الاختلاف بينه وبين اختبار الإشارة حيث تستخدم فقط إشارات الفروق لتنفيذ الاختبار. وفى هذا الجزء سنناقش اختبار ولكوكسن لإشارات الرتب (Wilcoxon Signed – rank) والذى يأخذ حجم الفرق في الحسبان. وهذا الاختبار يعطى نتائج افضل من اختبار الإشارات لأنه يستخدم مزيداً من المعلومات مقارنة باختبار الإشارة. بيد أن اختبار ولكوكسن لإشارات الرتب يتطلب أن تكون العينات تحت الدراسة مسحوبتان من مجتمعين لهما توزيع مشترك. وهو يختلف عن اختبار ت (t) للبيانات المزاوجة (Paired data)لأنه لا يتطلب أن يكون المجتمع الذي ينبغي دراسته هو مجتمع يخضع للتوزيع الطبيعي أو مجتمع موصوف رياضيا بنموذج هو النموذج الطبيعي.

مثال

يعطى الجدول أدناه بيانات عن 13 شخصاً تم اختبار تفكيرهم المنطقي. ثم أعطوا حبوب مهدئة واختبروا مرة أخرى. استخدم اختبار ولكوكسن لإشارات الرتب للاستيثاق من الادعاء بان الحبوب المهدئة ليست لها تأثير على التفكير المنطقي، أي لا يوجد فرق معنوي بين تناول وعدم تناول الحبوب المهدئة. افترض أن مستوى المعنوية $\alpha = 0.05$.

أرقام الأشخاص	قبل تناول المهدئ	بعد تناول المهدئ	الفرق	رتب الفرق	تخصيص الرتب
.1	65	66	−1	1	−1
.2	70	73	−3	2	−2
.3	84	88	−4	3	−3
.4	74	62	+12	10	+10
.5	76	76	0	–	–
.6	90	79	+11	8.5	+8.5
.7	86	75	+11	8.5	+8.5
.8	84	79	+5	4	+4
.9	76	83	−7	5	−5
.10	64	55	+9	6	+6
.11	70	60	+10	7	+7
.12	80	65	+15	11	+11
.13	81	64	+17	12	+12

والآن يمكن تلخيص طريقة الاختبار في الخطوات التالية:

1- لكل زوج من البيانات، أوجد الفرق (d) بطرح المفردة الأولى من المفردة الثانية. احتفظ بالإشارات وتجاهل الأزواج ذات الفرق (d) الذي يساوى صفراً $(d = 0)$

2- تجاهل إشارات الفروق ثم قم بترتيب الفروق من الأصغر إلى الأكبر. وعندما يحدث تعادل في الرتب نعتمد وسط تراتيبها الافتراضية ويخصص وسط الرتب المتعادلة لكل فرق له نفس المقدار المطلق. فمثلاً الشخصان رقم (6) ورقم (7) من المفترض أن تكون رتبتيهما الثامن والتاسع وعليه نأخذ وسطهما $[(8+9)/2 = 8.5]$ وتخصص الرتبة (8.5) لكل من الشخصين رقم (6) ورقم (7).

3- أعطى كل رتبة إشارة الفرق الذي جاءت منه .

4- أوجد مجموع القيم المطلقة للرتب السالبة وأيضاً مجموع القيم المطلقة للرتب الموجبة.

5- أرمز للمجموع الأصغر في الخطوة (4) بالرمز (T) .

6- أعتبر n هو عدد الأزواج التي لا يساوى الفرق بين مفرديتها (d) الصفر.

7- إذا كانت $n \leq 30$ ، أستخدم الجدول للحصول على القيمة الحرجة لـ T. أرفض فرضية العدم إذا كانت قيمة (T) المحسوبة من بيانات العينة اقل من أو تساوى القيمة الحرجة من الجدول. وخلاف ذلك اقبل فرضية العدم. وأما إذا كانت $n > 30$ أحسب اختبار الإحصاء Z باستخدام التقريب التالي:

$$Z = \frac{T - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} \quad \sim \quad N\ (0,\ 1)$$

وعند استخدام هذه القاعدة، نحصل على القيم الحرجة لاختبار الإحصاءة Z من جدول توزيع Z . لا نرفض فرض العدم إذا كانت القيمة المحسوبة أقل من القيمة الحرجة لـ Z. وخلاف ذلك نرفض فرض العدم . والآن سنتبع هذه الخطوات لاختبار فرض العدم بأن المهدئ ليس له تأثير على التفكير المنطقي.

الخطوة (1):

نحصل على عمود الفروقات (d) بطرح مفردة بعد التناول من مفردة قبل التناول المقابلة لها. ونتجاهل الفرق ذو القيمة صفر.

الخطوة (2):

بتجاهل الإشارات، نرتب الفرو قات من الأدنى إلى الأعلى. ونعالج الفروق المتساوية بالطريقة التي تم شرحها أعلاه.

الخطوة (3):

نحصل على عمود رتب الإشارات بتخصيص نفس إشارة الفرق للرتبة التي تقابلها كما في العمود الأخير للجدول أعلاه. وإذا كان المهدئ ليس له تأثير، نتوقع أن يكون عدد الإشارات الموجبة مساوي لعدد الإشارات السالبة.

الخطوة (4):

أحسب مجموع الرتب للفروق السالبة وأيضا مجموع الرتب للفروق الموجبـــة .

ومن بيانات الجدول أعلاه نحصل على :

مجموع الرتب للفروق السالبه = 1+2+3+5 = 11

مجموع القيم المطلقة للرتب الموجبة 10+8.5+8.5+4+6+7+11+12+ = 67

الخطوة (5):

ويقوم الاختبار على المجموع الأصغر للمجموعين، ويرمز له بالرمز T . وعليه T = 11

الخطوة (6):

وبما أن عدد الفروق (d) الذي لا يساوى الصفر 12 فان n = 12

الخطوة (7):

عندما تكون $n = \leq 30$ نستخدم الجدول المخصص لقيم T الحرجة. وأما إذا كانت $n > 30$ نستخدم التوزيع الطبيعي التقريبي وقد سبقت الإشـــارة للصـــيغة التـــي ينبغي استخدامها في هذه الحالة للحصول على القيمة المحسوبة للمتغيـــر Z ومقارنتهـــا بالقيمة الجدولية. وبما أن T = 11 اقل من القيمة الحرجة 14، نرفض فرض العدم. ويتضح أن المهدئ له تأثير معنوي على التفكير المنطقي للشخص.

3. اختبارات تتضمن الرتب

1.3 اختبار ولكوكسن لمجموع الرتب
(Wilcoxon Rank – Sum Test)

والآن سنلقى الضوء على الاختبار الثاني لولكوكسن ويسمى أختبـــار مجمـــوع الرتب ويطبق في حالة عينتين مستقلتين أي لا يوجد بينهما ترابط.

ونختبر فرض العدم: أن عينتين مستقلتين مسحوبتان من مجتمعين لهما نفس التوزيع. وأما الفرض البديل هو اختلاف التوزيعين بصورة ما. ولا تتطلب هـــذه الطريقـــة أن يتبع التوزيع النمط الطبيعي. وعليه تصبح هذه الطريقة بديل مفيـــد لاختبـــار الفرضـــية الخاصة بوسطى مجتمعين والتي تتطلب توفر فرض التوزيع الطبيعي. ويستخدم اختبـــار ولكوكسن لمجموع الرتب أيضا في حالة البيانات ذات الطبيعة الترتيبية – مثل البيانـــات التي تكون في شكل رتب.

وكما أشرنا من قبل، أن هذا الاختبار يتمتع بدرجـــة كفـــاءة مقـــدارها 0.95 مقارنة مع اختبار Z و t . ويفضل هذا الاختبـــار فـــي بعـــض الأحيـــان علـــى الاختبارات المعلمية المقابلة لها لكفاءته العالية وسهولة حسابه.

مثال توضيحي:

يعطى الجدول التالي عينتين عشوائيتين عن مرتبات المدرسين بولايتي الخرطوم والجزيرة. اختبر فرض أن مرتبات المدرسين متساوية في الولايتين عند مستوى معنوية 5 في المائة.

ولاية الخرطوم	الرتبة	ولاية الجريرة	الرتبة
36,200	28	22.200	8
32,500	25	22.200	10
31,000	23	21,500	5.5
31,100	24	22,700	9
33,700	26	24,000	11
28,400	20	20,500	1
29,100	21	21,000	4
28,000	18	24,100	12
27,900	17	21,500	5.5
33,900	27	22,000	7
39,900	29	25,700	13
		28,100	19
		20,900	3
		20,800	2
		26,400	14
		29,900	22
		26,600	15
		26,700	16
$n_1 = 11$ $R_1 = 258$ $\frac{R_1}{n_1} = \frac{258}{11} = 23.45$		$n_2 = 18$ $R_2 = 177$ $\frac{R_2}{n_2} = \frac{177}{18} = 9.83$	

الحل:

مجتمعا المرتبات متطابق : H_O

مجتمعا المرتبات غير متطابق : H_1

قد يقود التفكير إلى حل هذه المسألة إلى استخدام اختبار ت t لمقارنة وسطين لعينتين مستقلتين. ولكن اختبار ت t كما نعلم يتطلب أن تكون العينتان مأخوذتان من مجتمعين يتبعان نمط التوزيع الطبيعي. ولكن مجتمعا المرتبات في الولايتين لا يتبع نمط التوزيع الطبيعي الأمر الذي يتطلب استخدام اختبار غير معلمي.

ونقوم الآن بترتيب مفردات العينتين وعددها يساوى 29. ونبدأ بأقل مرتب وهو 20.500 ويعطى الرتبة (1) وتعطى الرتبة (2) للمرتب 20.800 وهكذا حتى نصل إلى أعلي مرتب وهو 39.900 ويعطى الرتبة (29). والرتب موضحة في عمودين منفصلين كلتاهما مقابل لمفردات العينتين. ونلاحظ أن هناك تعادل بين الرتبتين (5) و (6) وعليه تتم معالجتهما كما سبق توضيحه من قبل وذلك بأخذ وسط الرتبتين . وتمثل R مجموع رتب احدي العينتين. وإذا أخذنا مرتبات ولاية الخرطوم نحصل على:

$$R = 28+25+23+24+26+20+21+18+17+27+29 = 258$$

وبافتراض صحة فرضية العدم بأن المجتمعين متطابقان وبما أن حجم كل عينة على حدة اكبر من 10، فان توزيع المعاينة لـ R يتبع نمط التوزيع الطبيعي بالتقريب:

$$Z = \frac{R - u_R}{\sigma_R}$$

حيث آن:

$$u_R = \frac{n_1(n_1 + n_2 + 1)}{2}$$

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

وحيث أن

n_1 : يمثل حجم العينة التي حصلنا منها على R

n_2 : يمثل حجم العينة الاخرى

R : يمثل مجموع الرتب للعينة التي حجمها n_1

وكما هو موضح أعلاه، مجموع رتب ولاية الخرطوم R = 258

وحجم عينة مرتبات ولاية الخرطوم هو 11 وحجم عينة ولاية الجزيرة هو 18. وعليه يمكن الحصول على u_R و σ_R و Z .

$$u_R = \frac{n_1(n_1 + n_2 + 1)}{2} = \frac{11(11 + 18 + 1)}{2} = 165$$

$$\sigma_R = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}} = \sqrt{\frac{(11)(18)(11 + 18 + 1)}{12}} = 22.25$$

$$Z = \frac{R - u_R}{\sigma_R} = \frac{258 - 165}{22.25} = 4.18$$

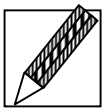
وتشير قيمة Z الكبيرة الموجبة إلى أن الرتب العليا توجد بصورة غير تناسبية في ولاية الخرطوم وبينما تشير قيمة Z الكبيرة السالبة إلى أن ولاية الخرطوم لها نصيب غير تناسبي في الرتب الدنيا. و يوجد دليل قوى في كلتا الحالتين ضد الدعوى بتساوي المرتبات بين ولايتي الخرطوم والجزيرة. ويتضح من ذلك أن الاختبار من طرفين (Two-tailed) . وبما أن القيم الحرجة (الجدولية) لـ Z هي 1.96 و 1.96 - ، فان القيمة المحسوبة لـ Z تقع في منطقة الرفض. و عليه نرفض فرضية العدم بتساوي المرتبات في كل من ولايتي الخرطوم والجزيرة.

ويمكن الحصول على نفس النتيجة بتبادل العينتين حيث تصبح في هذه الحالة

R= 177 و u_R = 270 ، σ_R = 22.25 Z = - 4.18

و يتضح من هذا المثال أن المبدأ الرئيس الذي تقوم عليه طريقة ولكوكسن يتمثل في انه إذا تم سحب عينتين من مجتمعين متطابقين ورتبت مفردات العينتين كمجموعة واحدة من القيم فان انتشار الرتب العليا والرتب الدنيا ينبغي أن يكون متماثلاً بين العينتين. وعليه إذا وجدنا الرتب العليا (أو الرتب الدنيا) غالبة في إحدى العينتين فهذا يقود إلى الشك في أن المجتمعين غير متطابقين. وننبه هنا بالرغم من أن الاختبار يسمى اختبار مجموع الرتب ولكن الحقيقة هي أن وسط مجموع الرتب هو الذي يؤثر على النتيجة. ويضع هذا الاختبار أيضا الحجم النسبي لمفردات العينة في الحسبان لذا يفضل علي اختبار الإشارة كما ذكرنا من قبل.

تدريب (3)



قام طلبة إحصاء ذكور وإناث بتقويم أداء رئيس جمعية باستخدام معايير مختلفة وحصلوا علي النتائج التالية، استخدم اختبار ولكوكسن لمجموع الرتب عند مستوي دلالة 5% لتحديد فيما إذا كان تقويم الذكور والإناث متماثل:

الذكور	27	36	42	57	88	92	60	60	43	29	76	79
الاناث	21	43	38	40	40	60	87	72	73	10	12	–

2.3 اختبار كروسكال – واليس (Kruskal – Wallis Test)

استخدمنا طريقة تحليل التباين في الوحدة الخامسة لاختبار فرض أن اختلاف متوسطات عدد من العينات ناتج عن طريق الصدفة. وتم اختبار هذا الفرض باستخدام إحصاءة F. وتتطلب هذه الإحصاءة أن تكون توزيعات المجتمعات المسحوبة منها العينات العشوائية طبيعية وان تكون تبايناتها (Variances) متساوية تقريباً. ونطرح الآن اختبار كروسكل – واليس اللامعلمى كبديل لاختبار F في الحالات التي تكون فيها المجتمعات تحت الدراسة لا تتبع نمط التوزيع الطبيعي. يختبر اختبار كروسكال – واليس فرض العدم أن عدد من العينات العشوائية تأتى من مجتمعات متطابقة . ويرمز لاحصاءة اختبار كروسكال–والبس بالرمز H. ولتنفيذ الاختبار، نحسب إحصاءة الاختبار H وله توزيع يمكن تقريبه بتوزيع χ^2 طالما كان حجم كل عينة على حدة لا يقل عن 5 مفردات وعند استخدام χ^2 في هذا الإطار، فان درجات الحرية تساوى (k-1) حيث k تمثل عدد العينات تحت الدراسة. وتستخدم الصيغة التالية للحصول على قيمة H المحسوبة.

$$H = \frac{12}{N(N+1)}\left(\frac{R_1^2}{n_1}+\frac{R_2^2}{n_2}+\ldots\ldots\ldots\ldots+\frac{R_k^2}{n_k}\right)-3(N+1) \qquad \sim \sim \chi^2_{k-1}$$

حيث أن :

N : العدد الكلى لمفردات كل العينات موضع الدراسة

R_1 : مجموع الرتب للعينة الأولى

R_2 : مجموع الرتب للعينة الثانية

R_k : مجموع الرتب للعينة k

k : عدد العينات موضع الدراسة.

وفكرة اختبار كروسكال – واليس هي استبدال المفردات المشاهدة بمجاميع الرتب المقابلة لها. ومن ثم نشرع في حساب إحصاءة الاختبار H وهي أساسا مقياس لتباين مجاميع الرتب: R_1 , R_2 ..... و R_k . وإذا توزعت الرتب بصورة متماثلة بين العينات، فان قيمة H تكون نسبياً صغيرة. ولكن إذا كانت العينات مختلفة فان الرتب تصبح صغيرة لبعض العينات وكبيرة للعينات الاخرى الأمر الذي يترتب عليه كبر قيمة H في المحصلة الأخيرة. ونتيجة لذلك، فان القيم الكبيرة لـ H هي التي تقود لرفض فرض العدم بان العينات جاءت من مجتمعات متطابقة. ومن ثم، فان اختبار كروسكال–واليس اختبار من طرف واحد وهو الطرف الأيمن.

ونلخص ما فات في التالي: إذا كان فرض العدم صحيح وكل عينة على حـــدة تتكـــون على الأقل من 5 مفردات، فان توزيع المعاينة لـــــ H يمكـــن تقريبـــه بتوزيـــع χ^2 بـــــدرجات حريـــــة تســـــاوى (k-1) . وعليـــــه نـــــرفض فـــــرض العـــــدم $H_o : u_1 = u_2 = \ldots\ldots\ldots = u_K$ ونقبل الفرض البديل بان الأوساط μ_s غير متســـاوية عند مستوى معنوية α ودرجات حرية (k-1) عندما يكون $H > \chi^2_\alpha$ ولا تســـتخدم χ^2 التقريبية إذا كان حجم العينة اقل من 5 مفردات لأي من العينات.

» مثال

ألحق عدد من المتدربين بصورة عشوائية إلى مجموعات يتم تدريبها على ممارسة الفحص الصناعي بطرق ثلاثة مختلفة. وتختبر نوعية أداء الفحـــص عنـــد نهايـــة مـــدة التدريب. وتوضح البينات التالية نتائج الطرق الثلاثة:

الطريقة الأولى : 80 و 83 و 79 و 85 و 90 و 68

الطريقة الثانية : 82 و 84 و 60 و 72 و 86 و 67 و 91

الطريقة الثالثة : 93 و 65 و 77 و 78 و 88 .

استخدم اختبار H لتحديد إذا كانت الطرق الثلاثة لها نفـــس الفعاليـــة عنـــد مستوى معنوية 5 في المائة.

الحل:

اعتبر مفردات الطرق الثلاثة مجموعة واحدة تتكون من 18 ، ثم رتب أحجام هذه المفردات تصاعدياً وخصص الرتبة الصحيحة للمفردة المقابلة لها كما يوضح الجدول أدناه:

المفردات	الرتب	رتب الطريقة الأولى	رتب الطريقة الثانية	رتب الطريقة الثالثة
60	1	9	10	18
65	2	11	12	2
67	3	8	1	6
68	4	13	5	7
72	5	16	14	15
77	6	4	13	
78	7		17	
79	8			
80	9			
82	10			
83	11			
84	12			
85	13			
86	14			
88	15			
90	16			
91	17			
93	18			
	المجموع	61	62	48

$$R_1 = 61 \quad n_1 = 6$$
$$R_2 = 62 \quad n_2 = 7$$
$$R_3 = 48 \quad n_3 = 5$$
$$N = 18$$

وباستخدام الصيغة H

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

$$= \frac{12}{(18)(19)}\left[\frac{(61)^2}{6} + \frac{(62)^2}{7} + \frac{(48)^2}{5}\right] - 3(19)$$

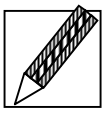
$$= \frac{12}{342}[620.17 + 549.14 + 460.8] - 57$$

$$= 57.197 - 57 = 0.197$$

$$v = k - 1 = 3 - 1 = 2$$

وأما القيمة الجدولية لـــ χ^2 عند مستوى معنوية 5 في المائة ودرجات حريـــة تساوى 2 هي 5.991. وبما أن القيمة المحسوبة أقل من القيمة الجدولية نقبـــل فـــرض العدم ونستنتج أن الطرق الثلاثة لها نفس الفعالية.

تدريب (4)



اختار عالم اجتماع أفراد من ثلاثة خلفيات أسرية مختلفة : أسر مستقرة وأسر تتميز بنسبة طلاق عالية وأسر تمر بمرحلة انتقالية. تمت معاينة لمجموعات الأفراد هذه ورتبوا على أساس تكيفهم الاجتماعي وأعطيت النتائج كما يلي:

الأسر المستقرة	الأسرة ذات نسبة الطلاق العالية	الأسر التي تمر بمرحلة انتقالية
108 (8.5)	113(10)	92(1)
104 (4)	106(7)	123(13)
103 (3)	108(8.5)	126(14)
97(2)	127(15)	105(5.5)
118 (12)	114(11)	105(5.5)

الأرقام داخل الأقواس تمثل الرتب. استخدم اختبار كروسكال–واليس عند مستوي دلالة 5% لاختبار ادعاء أن العينات الثلاث سحبت من مجتمعات متماثلة.

3.3 اختبار مان–وتنى (Mann-Whitney Test)

تم وصف هذا الاختبار في العام 1947 بواسطة عالمين هما مان ووتنى. ولا يتطلب هذا الاختبار أي فروض مقيدة عن طبيعة المجتمع الذى نأخذ منه العينة العشوائية. والفرض الوحيد الذى يتطلبه هذا الاختبار هو أن قيم المتغير العشوائى الذى على أساسه تتم المقارنة بين مجموعتين أن يكون ذو توزيع متصل (Continuous)

ولكن حتى على مستوى الممارسة العملية ليست هناك صعوبة تواجه هذا الاختبار إذا لم يتحقق هذا الفرض.

وكما وضحنا من قبل، يتجاهل اختبار الإشارة للمقارنة بين توزيعين لمجتمعين احجام المفردات المتزاوجه وبالتالي يصرف النظر عن معلومات مفيدة يمكن استخدامها لاختبار فرض العدم ويسعى اختبار مان ووتنى للاستفادة من هذه المعلومة باستخدام الحجم النسبي للمفردات. وفرض العدم الذي نرغب في اختباره هو أن العينتين مأخوذتين بصورة مستقلة من مجتمع واحد أو من مجتمعين لهما نفس الوسط $(u_1 = u_2)$ دون أن نفترض أن هذين المجتمعين يتبعان نمط التوزيع الطبيعي. وقد يكون الاختبار من طرف واحد أو من طرفين. ومن ثم يمثل اختبار مان – وتنى بديلاً مفيداً لاختبار ت (t) المعلمى عند عدم استيفاء الفروض التى يقوم عليها هذا الاختيار. ويسمى اختبار مان–وتنى باختيار U .

ويستخدم اختبار مان – وتنى عادة بدلاً عن اختبار الإشارة أو اختبار ولكوكسن عندما تكون العينتان تحت الدراسة مستقلتان. ولنفترض أن عينتين (حجم الأولى n_1 وحجم الثانية n_2) أختيرتا بشكل مستقل عن بعضهما وأن المفردات في مجموعها (n_1+n_2) تم ترتيبها أما تصاعدياً أو تنازلياً.

ويقوم اختبار فرض العدم – بان العينتين أخذتا من مجتمعين متطـــابقين – علـــى R_1 وهو مجموع رتب العينة الاولى أو R_2 وهو مجموع رتب العينـــة الثانيـــة. ولا نهـــتم كثيراً عند تطبيق هذا الاختبار أي من العينتين نسميها العينة الاولى او العينة الثانية.

وإذا كان حجم العينتين هو n_1 و n_2 فان مجموع R_1 , R_2 هـــو ببســـاطة مجموع الأعداد الصحيحة الموجبة الأولى من واحد الي $(n_1 + n_2)$ أي

$$R_1 + R_2 = \frac{(n_1 + n_2)(n_1 + n_2 + 1)}{2}$$

وذلك لآن حاصل جمع متوالية من الأعداد الطبيعية من 1 إلى n يعطي بالقانون

$$1+2+3+\ldots+n = n\left(\frac{n+1}{2}\right)$$

وتمكنا القاعدة أعلاه من الحصول على R_2 إذا كانت R_1 معروفة والعكـــس صـــحيح. وإحصاءة الاختبار U تعرف كما يلي:

$$U_1 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1 \ldots\ldots\ldots\ldots\ldots\ldots(1)$$

أو

$$U_2 = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - R_2 \ldots\ldots\ldots\ldots\ldots\ldots(2)$$

واذا كان حجم العينتين كلاً على حدة اقل من 10 (ولبعض الإحصائيين أقل من 8) ينبغي استخدام جداول خاصة لتنفيذ الاختبار. وقد تؤدى الصيغتان (1) و (2) إلى مقدارين مختلفين لاحصاءة الاختبار U ويستخدم أصغر المقدارين لأجراء الاختبار. ويرمز للمقدار الكبير بالرمز U' . وقبل استخدام الاختبار، ينبغي التأكد إذا تم الحصول على U أو U' بالمقارنة مع المقدار $(n_1 n_2)/2$. فإذا كانت قيمة U المحسوبة اكبر من المقدار $(n_1 n_2)/2$ فهي تمثل U' وقيمة U يمكن الحصول عليها باستخدام المعادلة:

$$U = n_1 n_2 - U'$$

» مثال

لنفترض أن لدينا رغبة في معرفة فيما إذا كان حجم المبيعات السنوية الذي حققته مجموعة من البائعين من حملة الشهادات الجامعية يختلف عن حجم المبيعات السنوية الذي حققته مجموعة أخرى لا تحمل شاهدات جامعية. ولنفترض أن (أ) يمثل المجموعة الأولى (ب) يمثل المجموعة الثانية من البائعين على التوالي. فإذا أخذت عينتان مستقلتان حجم الأولى $(n_1 = 10)$ من البائعين الجامعيين وحجم الثانية $(n_2 = 21)$ من البائعين غير الجامعين، وضح كيفية استخدام اختبار مان - وتني في ضوء البينات المعطاة في الجدول التالي:

المجموعة (أ)			المجموعة (ب)		
رقم البائع	حجم المبيعات السنوية	الرتبة	رقم البائع	حجم المبيعات السنوية	الرتبة
1	82	24	1	92	31
2	75	19	2	90	29.5
3	70	15	3	90	29.5
4	65	11	4	89	28
5	60	8	5	86	27
6	58	7	6	85	26
7	50	4.5	7	83	25
8	50	4.5	8	81	22.5
9	46	3	9	81	22.5
10	42	2	10	78	21
			11	76	20
			12	73	18
			13	72	17
			14	71	16
			15	68	14
			16	67	13
			17	66	12
			18	64	10
			19	63	9
			20	52	6
			21	40	1
$R_1 = 98$				$R_2 = 398$	

الحل:

يوضح الجدول حجم المبيعـــات ورتبهـــا ومجمـــوع الرتـــب R_1 ، R_2 .

وبالتعويض المباشر في القاعدة نحصل على قيمة U .

$$U = (10)(21) + \frac{10(10+1)}{2} - 98 = 167$$

أن U=167 اكبر من $n_1 n_2 / 2 = 105$ يرمز لها بالرمز U' وعليـــه U التي تستخدم في الاختبار هي:

$$U = n_1 n_2 - U' = (10)(21) - 167 = 43$$

وعندما يكون حجم العينتين n_2 و n_1 كلاً على حدة اقل من 20 ، فـــان قيمة U الصغرى تقارن بالقيم الحرجة في الجدول U لتحديد رفض أو قبول فرضية العدم (عدم وجود فرق بين الوسطين) . والآن في هذا المثـــال n_2 اكبـــر مـــن 20 يستخدم التوزيع الطبيعي التقريبي بدلاً عن جدول U .

وكلما زادت n_1 و n_2 يقترب توزيع U بسرعة من التوزيع الطبيعي. وفـــى الحقيقة عندما تكون n_1 أو n_2 مساوية 20، فان U تتبع نمـــط التوزيـــع الطبيعـــي التقريبي وتوقعها هو:

$$E(U) = \frac{n_1 n_2}{2}$$

وخطأها المعياري هو:

$$\sigma_U = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$$

وعليه يمكن حساب إحصاءة الاختبار Z كما يلي:

$$Z = \frac{U - E(U)}{\sigma_U}$$

وعند استخدام توزيع المعاينة الطبيعي التقريبي للمتغير U، ليس هناك فرق إذا استخدمنا قيمة U الصغيرة أو الكبيرة لان القيمة المطلقة للمتغير Z تظل كما هي. وللاختبار من طرف واحد تقع المنطقة الحرجة تحت القيمة السالبة للمتغير Z إذا استخدمت القيمة الصغيرة للمتغير Z. وتقع أعلى القيمة الموجبة للمتغير Z إذا استخدمت القيمة الكبيرة للمتغير U. ولاختبار الطرفين، كما في هذا المثال، نحصل على :

$$E(U) = \frac{(10)(21)}{2} = 105$$

$$\sigma_U = \sqrt{\frac{(10)(21)(10+21+1)}{12}} = \sqrt{\frac{6720}{12}} = \sqrt{560} = 23.66$$

وهكذا فان:

$$Z = \frac{43 - 105}{23.66} = 2.62$$

وإذا كان مستوى المعنوية هو $\alpha = 0.01$ فان القيمتين الحرجتين للمتغير Z هما ± 2.575 . وعليه، نرفض فرض العدم ونستنتج أن حجم المبيعات السنوية الذي تحققه مجموعة الباعة غير الجامعيين لا يساوى حجم المبيعات السنوية الذي يحققه الباعة الجامعيون.

أسئلة التقويم الذاتي (1)

أعطيت مجموعتين من الرجال والنساء بعد أن تم أختيارهما عشوائياً استبانة صممت لقياس إحساسهم نحو عقوبة الأعدام، ورصدت النتائج كما يلي

الرجال	67	72	48	30	92	51	15	5	87	91	54	66	72	98	97	75	74
النساء	20	40	37	42	51	15	68	35	12	31	85	–	–	–	–	–	–

اختبر ادعاء انه لا يوجد فرق في إحساس كل من الرجال والنساء نحو عقوبة الاعدام عند مستوي دلالة 5%

الخلاصة

والآن يمكننا أن نلخص هذه الوحدة بالشكل التالي:

تطبق الاختبارات اللا معلمية عندما يكون توزيع المجتمع مجهول أو غير طبيعي أو لا يمكن الحصول على إحصاءات عينة يمكن الرجوع إليها مع ملاحظة أنه لا تطبق الاختبارات اللا معلمية إلا عندما يتعذر تطبيق الاختبارات المعلمية وذلك لضعف قوتها ولكن لهذه الاختبارات محاسن حيث أنها لا تتطلب نوع توزيع المجتمع بالإضافة لأنها سهلة التطبيق ولا تتطلب حسابات معقدة وتجري في وقت وجيز كما أنها تطبق لجميع أنواع البيانات وتطبق في حالات العينات الصغيرة ، إضافة لكل ذلك بنيت على فروض محددة وغير صارمة ولكن لها عيوب إذ أنها تضيع المعلومات لتحويل البيانات الكمية إلى بيانات نوعية وأيضاً نجد أن درجة حساسيتها أقل مقارنة بحساسية المعلمية لكن زيادة حجم العينة تزيد من حساسية الاختبار .

تناولنا خلال الوحدة نوعين من الاختبارات التي تتضمن الإشارة؛ الأول اختبار الإشارة بنوعيه (اختبار الإشارة للعينة الواحدة واختبار الإشارة لعينة ذات زوج) والثاني اختبار إشارة الرتبة. وكذلك شرحنا في هذه الوحدة اختبار ولكلسون لمجموع الرتب و وضحنا أنه يطبق عندما تكون هناك عينتان مستقلتان ولذلك يمكن أن يطبق لاختبار تطابق توزيع مجتمعين. وطرحنا من خلال هذه الوحدة اختبار كروسكال – أليس كبديل لاختبار ف لتحليل تباين عدد من المتوسطات واختبار معنوية هذه المتوسطات عندما لا تتبع المجتمعات التوزيع الطبيعي . وفي نهاية الوحدة قدمنا اختبار مان وتني و وضحنا أنه لا يضع أي قيود على طبيعة توزيع المجتمع وأوضحنا أن هذا الاختبار بديل مناسب لاختبار الإشارة واختبار ويلكلسون عندما تكون العينتان مستقلتان .

إجابات التدريبات

تدريب (1)

قبل التعديل x	بعد التعديل y	إشارة الفرق y – x
30	32	+
28	29	+
34	33	–
35	32	–
40	37	–
42	43	+
33	40	+
38	41	+
34	37	+
45	44	–
28	27	–
27	33	+
25	30	+
41	38	–
36	36	0

H_o : لا يوجد فرق في البرنامج قبل وبعد التعديل

H_1 : يوجد فرق في برنامج التحصيل قبل وبعد التعديل

عدد الاشارات الموجبة 8

عدد الاشارات السالبة 6

عدد الأصفار 1

$n = 15$, $s = 6$

$$K=\frac{n-1}{2}-(0.98)\sqrt{n}$$

$$=\frac{14}{2}-(0.98)\sqrt{15}$$

$$= 7-3.7955 = 3.2045$$

بما أن $S > K$

إذن نقبل فرض العدم

تدريب (2)

قبل التدخين	بعد التدخين	الفرق	رتب الفرق	تخصيص الرتب
68	69	−1	1.5	−1.5
72	76	−4	6	−6
69	68	+1	1.5	+1.5
70	73	−3	5	−5
70	72	−2	3.5	−3.5
74	76	−2	3.5	−3.5
66	66	0	–	–
71	71	0	–	–

مجموع الرتب للفروق الموجبة 1.5

مجموع الرتب للفروق الموجبة 19.5

H_o : التدخين لا يؤثر على معدل النبض

H_1 : التدخين يؤثر على معدل النبض

إذن $T = 1.5$

$n = 6$ عدد الفروق التي لا تساوي الصفر

نقارن $T = 1.5$ بالقيمة الحرجة الجدولية t .

فإذا كانت $T < t$ فإننا نرفض فرض العدم و إلا نقبل فرض العدم .

تدريب (3)

الذكور	الرتبة	الاناث	الرتبة
27	4	21	3
36	6	43	11.5
42	10	38	7
57	13	40	8.5
88	22	40	8.5
92	23	60	15
60	15	87	21
60	15	72	17
43	11.5	73	18
29	5	10	1
76	19	12	2
79	20	0	
$n_1 = 12$ $R_1 = 163.5$ $\frac{R_1}{n_1} = 13.625$		$n_2 = 11$ $R_2 = 112.5$ $\frac{R_2}{n_2} = 10.227$	

صياغة الفرض

H_o : التقويم (تقويم الذكور و الاناث) متماثل

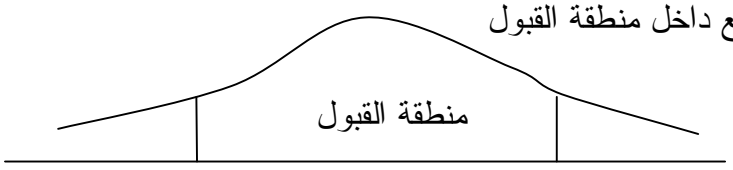
H_1 : التقويم (تقويم الذكور و الاناث) غير متماثل

$$\mu_r = \frac{1}{2}\, n_1(n_1 + n_2 + 1)$$
$$= \frac{1}{2}\,(12)\,(24)$$
$$= 144$$
$$\sigma_R = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$$
$$= \sqrt{\frac{12(11)(24)}{12}}$$
$$= 16.248$$

$$R = R_1 = 163.5$$
$$Z = \frac{R - \mu_R}{\sigma_R}$$
$$= \frac{163.5 - 144}{16.248}$$
$$= 1.2$$

بما أن القيمة الحرجة الجدولية عند الاختبار من طرفين هي $Z = 1.96$ أو $Z = -1.96$

ونجد أن 1.2 تقع داخل منطقة القبول



-1.96 1.96

إذن نقبل فرض العدم وهو أن تقويم الذكور والاناث متماثل .

تدريب (4)

المقررات	الرتب	رتب الأسر المستقرة	رتب الأسر ذات نسبة الطلاق العالية	رتب الأسر التي تمر بمرحلة انتقالية
92	1	8.5	10	1
97	2	4	7	13
103	3	3	8.5	14
104	4	2	15	5.5
105	5.5	12	11	5.5
105	5.5			
106	7			
108	8.5			
108	8.5			
113	10			
114	11			
118	12			
123	13			
126	14			
127	15			

H_o : العينات الثلاث أخذت من مجتمعات متماثلة

H_1 : العينات الثلاث أخذت من مجتمعات غير متماثلة

$n_1 = 5$ $R_1 = 29.5$

$n_2 = 5$ $R_2 = 51.5$

$n_3 = 5$ $R_3 = 39$

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

$$= \frac{12}{15 \times 16}\left[\frac{29.5^2}{5}+\frac{51.5^2}{5}+\frac{39^2}{5}\right]-3(16)$$

$$= 50.435 - 48 = 2.435$$

درجة الحرية

$$V = k - 1 = 3 - 1 = 2$$

القيم الجدولية لـــ χ^2 عند مستوى $\alpha = 0.05$ ودرجة حرية 2 هي 5.991

إذن نقبل فرض العدم أي العينات الثلاث جاءت من مجتمعات متماثلة .

لأن $2.435 < 5.991$

مسرد المصطلحات

- **الطرق المعلمية (Parametric Methods)** : طرق تقوم على فرض أن توزيع المجتمع معلوم وعادة الطبيعي.
- **الطرق غير المعلمية (Nonparametric Methods)** : طرق لا تتطلب معرفة توزيع المجتمع الذي تم سحب العينة منه فيما إذا كان يتبع نمط التوزيع الطبيعي أم خلافه.

المراجع

- مواري د. شبيجل، "الإحصاء" ، سلسلة ملخصات شوم، الدار الدولية للنشر والتوزيع 1991.

- بول ج.هويل، " المبادئ الأولية في الإحصاء، ترجمة ومراجعة د. بدرية شوقي عبد الوهاب والدكتور محمد كامل الشربيني ، دار جون وايلي وأبنائه.

Z- table

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

t - Table

df\p	0.40	0.25	0.10	0.05	0.025	0.01	0.005	0.0005
1	0.324920	1.000000	3.077684	6.313752	12.70620	31.82052	63.65674	636.6192
2	0.288675	0.816497	1.885618	2.919986	4.30265	6.96456	9.92484	31.5991
3	0.276671	0.764892	1.637744	2.353363	3.18245	4.54070	5.84091	12.9240
4	0.270722	0.740697	1.533206	2.131847	2.77645	3.74695	4.60409	8.6103
5	0.267181	0.726687	1.475884	2.015048	2.57058	3.36493	4.03214	6.8688
6	0.264835	0.717558	1.439756	1.943180	2.44691	3.14267	3.70743	5.9588
7	0.263167	0.711142	1.414924	1.894579	2.36462	2.99795	3.49948	5.4079
8	0.261921	0.706387	1.396815	1.859548	2.30600	2.89646	3.35539	5.0413
9	0.260955	0.702722	1.383029	1.833113	2.26216	2.82144	3.24984	4.7809
10	0.260185	0.699812	1.372184	1.812461	2.22814	2.76377	3.16927	4.5869
11	0.259556	0.697445	1.363430	1.795885	2.20099	2.71808	3.10581	4.4370
12	0.259033	0.695483	1.356217	1.782288	2.17881	2.68100	3.05454	4.3178
13	0.258591	0.693829	1.350171	1.770933	2.16037	2.65031	3.01228	4.2208
14	0.258213	0.692417	1.345030	1.761310	2.14479	2.62449	2.97684	4.1405
15	0.257885	0.691197	1.340606	1.753050	2.13145	2.60248	2.94671	4.0728
16	0.257599	0.690132	1.336757	1.745884	2.11991	2.58349	2.92078	4.0150
17	0.257347	0.689195	1.333379	1.739607	2.10982	2.56693	2.89823	3.9651
18	0.257123	0.688364	1.330391	1.734064	2.10092	2.55238	2.87844	3.9216
19	0.256923	0.687621	1.327728	1.729133	2.09302	2.53948	2.86093	3.8834
20	0.256743	0.686954	1.325341	1.724718	2.08596	2.52798	2.84534	3.8495
21	0.256580	0.686352	1.323188	1.720743	2.07961	2.51765	2.83136	3.8193
22	0.256432	0.685805	1.321237	1.717144	2.07387	2.50832	2.81876	3.7921
23	0.256297	0.685306	1.319460	1.713872	2.06866	2.49987	2.80734	3.7676
24	0.256173	0.684850	1.317836	1.710882	2.06390	2.49216	2.79694	3.7454
25	0.256060	0.684430	1.316345	1.708141	2.05954	2.48511	2.78744	3.7251
26	0.255955	0.684043	1.314972	1.705618	2.05553	2.47863	2.77871	3.7066
27	0.255858	0.683685	1.313703	1.703288	2.05183	2.47266	2.77068	3.6896
28	0.255768	0.683353	1.312527	1.701131	2.04841	2.46714	2.76326	3.6739
29	0.255684	0.683044	1.311434	1.699127	2.04523	2.46202	2.75639	3.6594
30	0.255605	0.682756	1.310415	1.697261	2.04227	2.45726	2.75000	3.6460
inf	0.253347	0.674490	1.281552	1.644854	1.95996	2.32635	2.57583	3.2905

Chi-Squire table

\area	.995	.990	.975	.950	.900	.750	.500	.250	.100	.050	.025	.010	.005
1	0.00004	0.00016	0.00098	0.00393	0.01579	0.10153	0.45494	1.32330	2.70554	3.84146	5.02389	6.63490	7.87944
2	0.01003	0.02010	0.05064	0.10259	0.21072	0.57536	1.38629	2.77259	4.60517	5.99146	7.37776	9.21034	10.59663
3	0.07172	0.11483	0.21580	0.35185	0.58437	1.21253	2.36597	4.10834	6.25139	7.81473	9.34840	11.34487	12.83816
4	0.20699	0.29711	0.48442	0.71072	1.06362	1.92256	3.35669	5.38527	7.77944	9.48773	11.14329	13.27670	14.86026
5	0.41174	0.55430	0.83121	1.14548	1.61031	2.67460	4.35146	6.62568	9.23636	11.07050	12.83250	15.08627	16.74960
6	0.67573	0.87209	1.23734	1.63538	2.20413	3.45460	5.34812	7.84080	10.64464	12.59159	14.44938	16.81189	18.54758
7	0.98926	1.23904	1.68987	2.16735	2.83311	4.25485	6.34581	9.03715	12.01704	14.06714	16.01276	18.47531	20.27774
8	1.34441	1.64650	2.17973	2.73264	3.48954	5.07064	7.34412	10.21885	13.36157	15.50731	17.53455	20.09024	21.95495
9	1.73493	2.08790	2.70039	3.32511	4.16816	5.89883	8.34283	11.38875	14.68366	16.91898	19.02277	21.66599	23.58935
10	2.15586	2.55821	3.24697	3.94030	4.86518	6.73720	9.34182	12.54886	15.98718	18.30704	20.48318	23.20925	25.18818
11	2.60322	3.05348	3.81575	4.57481	5.57778	7.58414	10.34100	13.70069	17.27501	19.67514	21.92005	24.72497	26.75685
12	3.07382	3.57057	4.40379	5.22603	6.30380	8.43842	11.34032	14.84540	18.54935	21.02607	23.33666	26.21697	28.29952
13	3.56503	4.10692	5.00875	5.89186	7.04150	9.29907	12.33976	15.98391	19.81193	22.36203	24.73560	27.68825	29.81947
14	4.07467	4.66043	5.62873	6.57063	7.78953	10.16531	13.33927	17.11693	21.06414	23.68479	26.11895	29.14124	31.31935
15	4.60092	5.22935	6.26214	7.26094	8.54676	11.03654	14.33886	18.24509	22.30713	24.99579	27.48839	30.57791	32.80132

Continue ----- Chi-Squire table

16	**5.14221**	**5.81221**	**6.90766**	**7.96165**	**9.31224**	**11.91222**	**15.33850**	**19.36886**	**23.54183**	**26.29623**	**28.84535**	**31.99993**	**34.26719**
17	**5.69722**	**6.40776**	**7.56419**	**8.67176**	**10.08519**	**12.79193**	**16.33818**	**20.48868**	**24.76904**	**27.58711**	**30.19101**	**33.40866**	**35.71847**
18	**6.26480**	**7.01491**	**8.23075**	**9.39046**	**10.86494**	**13.67529**	**17.33790**	**21.60489**	**25.98942**	**28.86930**	**31.52638**	**34.80531**	**37.15645**
19	**6.84397**	**7.63273**	**8.90652**	**10.11701**	**11.65091**	**14.56200**	**18.33765**	**22.71781**	**27.20357**	**30.14353**	**32.85233**	**36.19087**	**38.58226**
20	**7.43384**	**8.26040**	**9.59078**	**10.85081**	**12.44261**	**15.45177**	**19.33743**	**23.82769**	**28.41198**	**31.41043**	**34.16961**	**37.56623**	**39.99685**
21	**8.03365**	**8.89720**	**10.28290**	**11.59131**	**13.23960**	**16.34438**	**20.33723**	**24.93478**	**29.61509**	**32.67057**	**35.47888**	**38.93217**	**41.40106**
22	**8.64272**	**9.54249**	**10.98232**	**12.33801**	**14.04149**	**17.23962**	**21.33704**	**26.03927**	**30.81328**	**33.92444**	**36.78071**	**40.28936**	**42.79565**
23	**9.26042**	**10.19572**	**11.68855**	**13.09051**	**14.84796**	**18.13730**	**22.33688**	**27.14134**	**32.00690**	**35.17246**	**38.07563**	**41.63840**	**44.18128**
24	**9.88623**	**10.85636**	**12.40115**	**13.84843**	**15.65868**	**19.03725**	**23.33673**	**28.24115**	**33.19624**	**36.41503**	**39.36408**	**42.97982**	**45.55851**
25	**10.51965**	**11.52398**	**13.11972**	**14.61141**	**16.47341**	**19.93934**	**24.33659**	**29.33885**	**34.38159**	**37.65248**	**40.64647**	**44.31410**	**46.92789**
26	**11.16024**	**12.19815**	**13.84390**	**15.37916**	**17.29188**	**20.84343**	**25.33646**	**30.43457**	**35.56317**	**38.88514**	**41.92317**	**45.64168**	**48.28988**
27	**11.80759**	**12.87850**	**14.57338**	**16.15140**	**18.11390**	**21.74940**	**26.33634**	**31.52841**	**36.74122**	**40.11327**	**43.19451**	**46.96294**	**49.64492**
28	**12.46134**	**13.56471**	**15.30786**	**16.92788**	**18.93924**	**22.65716**	**27.33623**	**32.62049**	**37.91592**	**41.33714**	**44.46079**	**48.27824**	**50.99338**
29	**13.12115**	**14.25645**	**16.04707**	**17.70837**	**19.76774**	**23.56659**	**28.33613**	**33.71091**	**39.08747**	**42.55697**	**45.72229**	**49.58788**	**52.33562**
30	**13.78672**	**14.95346**	**16.79077**	**18.49266**	**20.59923**	**24.47761**	**29.33603**	**34.79974**	**40.25602**	**43.77297**	**46.97924**	**50.89218**	**53.67196**

F- table (α = 0.1)

df2/df1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	INF
1	39.86346	49.50000	53.59324	55.83296	57.24008	58.20442	58.90595	59.43898	59.85759	60.19498	60.70521	61.22034	61.74029	62.00205	62.26497	62.52905	62.79428	63.06064	63.32812
2	8.52632	9.00000	9.16179	9.24342	9.29263	9.32553	9.34908	9.36677	9.38054	9.39157	9.40813	9.42471	9.44131	9.44962	9.45793	9.46624	9.47456	9.48289	9.49122
3	5.53832	5.46238	5.39077	5.34264	5.30916	5.28473	5.26619	5.25167	5.24000	5.23041	5.21562	5.20031	5.18448	5.17636	5.16811	5.15972	5.15119	5.14251	5.13370
4	4.54477	4.32456	4.19086	4.10725	4.05058	4.00975	3.97897	3.95494	3.93567	3.91988	3.89553	3.87036	3.84434	3.83099	3.81742	3.80361	3.78957	3.77527	3.76073
5	4.06042	3.77972	3.61948	3.52020	3.45298	3.40451	3.36790	3.33928	3.31628	3.29740	3.26824	3.23801	3.20665	3.19052	3.17408	3.15732	3.14023	3.12279	3.10500
6	3.77595	3.46330	3.28876	3.18076	3.10751	3.05455	3.01446	2.98304	2.95774	2.93693	2.90472	2.87122	2.83634	2.81834	2.79996	2.78117	2.76195	2.74229	2.72216
7	3.58943	3.25744	3.07407	2.96053	2.88334	2.82739	2.78493	2.75158	2.72468	2.70251	2.66811	2.63223	2.59473	2.57533	2.55546	2.53510	2.51422	2.49279	2.47079
8	3.45792	3.11312	2.92380	2.80643	2.72645	2.66833	2.62413	2.58935	2.56124	2.53804	2.50196	2.46422	2.42464	2.40410	2.38302	2.36136	2.33910	2.31618	2.29257
9	3.36030	3.00645	2.81286	2.69268	2.61061	2.55086	2.50531	2.46941	2.44034	2.41632	2.37888	2.33962	2.29832	2.27683	2.25472	2.23196	2.20849	2.18427	2.15923
10	3.28502	2.92447	2.72767	2.60534	2.52164	2.46058	2.41397	2.37715	2.34731	2.32260	2.28405	2.24351	2.20074	2.17843	2.15543	2.13169	2.10716	2.08176	2.05542
11	3.22520	2.85951	2.66023	2.53619	2.45118	2.38907	2.34157	2.30400	2.27350	2.24823	2.20873	2.16709	2.12305	2.10001	2.07621	2.05161	2.02612	1.99965	1.97211
12	3.17655	2.80680	2.60552	2.48010	2.39402	2.33102	2.28278	2.24457	2.21352	2.18776	2.14744	2.10485	2.05968	2.03599	2.01149	1.98610	1.95973	1.93228	1.90361
13	3.13621	2.76317	2.56027	2.43371	2.34672	2.28298	2.23410	2.19535	2.16382	2.13763	2.09659	2.05316	2.00698	1.98272	1.95757	1.93147	1.90429	1.87591	1.84620
14	3.10221	2.72647	2.52222	2.39469	2.30694	2.24256	2.19313	2.15390	2.12195	2.09540	2.05371	2.00953	1.96245	1.93766	1.91193	1.88516	1.85723	1.82800	1.79728
15	3.07319	2.69517	2.48979	2.36143	2.27302	2.20808	2.15818	2.11853	2.08621	2.05932	2.01707	1.97222	1.92431	1.89904	1.87277	1.84539	1.81676	1.78672	1.75505
16	3.04811	2.66817	2.46181	2.33274	2.24376	2.17833	2.12800	2.08798	2.05533	2.02815	1.98539	1.93992	1.89127	1.86556	1.83879	1.81084	1.78156	1.75075	1.71817
17	3.02623	2.64464	2.43743	2.30775	2.21825	2.15239	2.10169	2.06134	2.02839	2.00094	1.95772	1.91169	1.86236	1.83624	1.80901	1.78053	1.75063	1.71909	1.68564
18	3.00698	2.62395	2.41601	2.28577	2.19583	2.12958	2.07854	2.03789	2.00467	1.97698	1.93334	1.88681	1.83685	1.81035	1.78269	1.75371	1.72322	1.69099	1.65671
19	2.98990	2.60561	2.39702	2.26630	2.17596	2.10936	2.05802	2.01710	1.98364	1.95573	1.91170	1.86471	1.81416	1.78731	1.75924	1.72979	1.69876	1.66587	1.63077
20	2.97465	2.58925	2.38009	2.24893	2.15823	2.09132	2.03970	1.99853	1.96485	1.93674	1.89236	1.84494	1.79384	1.76667	1.73822	1.70833	1.67678	1.64326	1.60738
21	2.96096	2.57457	2.36489	2.23334	2.14231	2.07512	2.02325	1.98186	1.94797	1.91967	1.87497	1.82715	1.77555	1.74807	1.71927	1.68896	1.65691	1.62278	1.58615
22	2.94858	2.56131	2.35117	2.21927	2.12794	2.06050	2.00840	1.96680	1.93273	1.90425	1.85925	1.81106	1.75899	1.73122	1.70208	1.67138	1.63885	1.60415	1.56678
23	2.93736	2.54929	2.33873	2.20651	2.11491	2.04723	1.99492	1.95312	1.91888	1.89025	1.84497	1.79643	1.74392	1.71588	1.68643	1.65535	1.62237	1.58711	1.54903
24	2.92712	2.53833	2.32739	2.19488	2.10303	2.03513	1.98263	1.94066	1.90625	1.87748	1.83194	1.78308	1.73015	1.70185	1.67210	1.64067	1.60726	1.57146	1.53270
25	2.91774	2.52831	2.31702	2.18424	2.09216	2.02406	1.97138	1.92925	1.89469	1.86578	1.82000	1.77083	1.71752	1.68898	1.65895	1.62718	1.59335	1.55703	1.51760
26	2.90913	2.51910	2.30749	2.17447	2.08218	2.01389	1.96104	1.91876	1.88407	1.85503	1.80902	1.75957	1.70589	1.67712	1.64682	1.61472	1.58050	1.54368	1.50360
27	2.90119	2.51061	2.29871	2.16546	2.07298	2.00452	1.95151	1.90909	1.87427	1.84511	1.79889	1.74917	1.69514	1.66616	1.63560	1.60320	1.56859	1.53129	1.49057
28	2.89385	2.50276	2.29060	2.15714	2.06447	1.99585	1.94270	1.90014	1.86520	1.83593	1.78951	1.73954	1.68519	1.65600	1.62519	1.59250	1.55753	1.51976	1.47841
29	2.88703	2.49548	2.28307	2.14941	2.05658	1.98781	1.93452	1.89184	1.85679	1.82741	1.78081	1.73060	1.67593	1.64655	1.61551	1.58253	1.54721	1.50899	1.46704
30	2.88069	2.48872	2.27607	2.14223	2.04925	1.98033	1.92692	1.88412	1.84896	1.81949	1.77270	1.72227	1.66731	1.63774	1.60648	1.57323	1.53757	1.49891	1.45636
40	2.83535	2.44037	2.22609	2.09095	1.99682	1.92688	1.87252	1.82886	1.79290	1.76269	1.71456	1.66241	1.60515	1.57411	1.54108	1.50562	1.46716	1.42476	1.37691
60	2.79107	2.39325	2.17741	2.04099	1.94571	1.87472	1.81939	1.77483	1.73802	1.70701	1.65743	1.60337	1.54349	1.51072	1.47554	1.43734	1.39520	1.34757	1.29146
120	2.74781	2.34734	2.12999	1.99230	1.89587	1.82381	1.76748	1.72196	1.68425	1.65238	1.60120	1.54500	1.48207	1.44723	1.40938	1.36760	1.32034	1.26457	1.19256
inf	2.70554	2.30259	2.08380	1.94486	1.84727	1.77411	1.71672	1.67020	1.63152	1.59872	1.54578	1.48714	1.42060	1.38318	1.34187	1.29513	1.23995	1.16860	1.00000

F- table (α = 0.05)

df2/df1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	INF
1	161.4476	199.5000	215.7073	224.5832	230.1619	233.9860	236.7684	238.8827	240.5433	241.8817	243.9060	245.9499	248.0131	249.0518	250.0951	251.1432	252.1957	253.2529	254.3144
2	18.5128	19.0000	19.1643	19.2468	19.2964	19.3295	19.3532	19.3710	19.3848	19.3959	19.4125	19.4291	19.4458	19.4541	19.4624	19.4707	19.4791	19.4874	19.4957
3	10.1280	9.5521	9.2766	9.1172	9.0135	8.9406	8.8867	8.8452	8.8123	8.7855	8.7446	8.7029	8.6602	8.6385	8.6166	8.5944	8.5720	8.5494	8.5264
4	7.7086	6.9443	6.5914	6.3882	6.2561	6.1631	6.0942	6.0410	5.9988	5.9644	5.9117	5.8578	5.8025	5.7744	5.7459	5.7170	5.6877	5.6581	5.6281
5	6.6079	5.7861	5.4095	5.1922	5.0503	4.9503	4.8759	4.8183	4.7725	4.7351	4.6777	4.6188	4.5581	4.5272	4.4957	4.4638	4.4314	4.3985	4.3650
6	5.9874	5.1433	4.7571	4.5337	4.3874	4.2839	4.2067	4.1468	4.0990	4.0600	3.9999	3.9381	3.8742	3.8415	3.8082	3.7743	3.7398	3.7047	3.6689
7	5.5914	4.7374	4.3468	4.1203	3.9715	3.8660	3.7870	3.7257	3.6767	3.6365	3.5747	3.5107	3.4445	3.4105	3.3758	3.3404	3.3043	3.2674	3.2298
8	5.3177	4.4590	4.0662	3.8379	3.6875	3.5806	3.5005	3.4381	3.3881	3.3472	3.2839	3.2184	3.1503	3.1152	3.0794	3.0428	3.0053	2.9669	2.9276
9	5.1174	4.2565	3.8625	3.6331	3.4817	3.3738	3.2927	3.2296	3.1789	3.1373	3.0729	3.0061	2.9365	2.9005	2.8637	2.8259	2.7872	2.7475	2.7067
10	4.9646	4.1028	3.7083	3.4780	3.3258	3.2172	3.1355	3.0717	3.0204	2.9782	2.9130	2.8450	2.7740	2.7372	2.6996	2.6609	2.6211	2.5801	2.5379
11	4.8443	3.9823	3.5874	3.3567	3.2039	3.0946	3.0123	2.9480	2.8962	2.8536	2.7876	2.7186	2.6464	2.6090	2.5705	2.5309	2.4901	2.4480	2.4045
12	4.7472	3.8853	3.4903	3.2592	3.1059	2.9961	2.9134	2.8486	2.7964	2.7534	2.6866	2.6169	2.5436	2.5055	2.4663	2.4259	2.3842	2.3410	2.2962
13	4.6672	3.8056	3.4105	3.1791	3.0254	2.9153	2.8321	2.7669	2.7144	2.6710	2.6037	2.5331	2.4589	2.4202	2.3803	2.3392	2.2966	2.2524	2.2064
14	4.6001	3.7389	3.3439	3.1122	2.9582	2.8477	2.7642	2.6987	2.6458	2.6022	2.5342	2.4630	2.3879	2.3487	2.3082	2.2664	2.2229	2.1778	2.1307
15	4.5431	3.6823	3.2874	3.0556	2.9013	2.7905	2.7066	2.6408	2.5876	2.5437	2.4753	2.4034	2.3275	2.2878	2.2468	2.2043	2.1601	2.1141	2.0658
16	4.4940	3.6337	3.2389	3.0069	2.8524	2.7413	2.6572	2.5911	2.5377	2.4935	2.4247	2.3522	2.2756	2.2354	2.1938	2.1507	2.1058	2.0589	2.0096
17	4.4513	3.5915	3.1968	2.9647	2.8100	2.6987	2.6143	2.5480	2.4943	2.4499	2.3807	2.3077	2.2304	2.1898	2.1477	2.1040	2.0584	2.0107	1.9604
18	4.4139	3.5546	3.1599	2.9277	2.7729	2.6613	2.5767	2.5102	2.4563	2.4117	2.3421	2.2686	2.1906	2.1497	2.1071	2.0629	2.0166	1.9681	1.9168
19	4.3807	3.5219	3.1274	2.8951	2.7401	2.6283	2.5435	2.4768	2.4227	2.3779	2.3080	2.2341	2.1555	2.1141	2.0712	2.0264	1.9795	1.9302	1.8780
20	4.3512	3.4928	3.0984	2.8661	2.7109	2.5990	2.5140	2.4471	2.3928	2.3479	2.2776	2.2033	2.1242	2.0825	2.0391	1.9938	1.9464	1.8963	1.8432
21	4.3248	3.4668	3.0725	2.8401	2.6848	2.5727	2.4876	2.4205	2.3660	2.3210	2.2504	2.1757	2.0960	2.0540	2.0102	1.9645	1.9165	1.8657	1.8117
22	4.3009	3.4434	3.0491	2.8167	2.6613	2.5491	2.4638	2.3965	2.3419	2.2967	2.2258	2.1508	2.0707	2.0283	1.9842	1.9380	1.8894	1.8380	1.7831
23	4.2793	3.4221	3.0280	2.7955	2.6400	2.5277	2.4422	2.3748	2.3201	2.2747	2.2036	2.1282	2.0476	2.0050	1.9605	1.9139	1.8648	1.8128	1.7570
24	4.2597	3.4028	3.0088	2.7763	2.6207	2.5082	2.4226	2.3551	2.3002	2.2547	2.1834	2.1077	2.0267	1.9838	1.9390	1.8920	1.8424	1.7896	1.7330
25	4.2417	3.3852	2.9912	2.7587	2.6030	2.4904	2.4047	2.3371	2.2821	2.2365	2.1649	2.0889	2.0075	1.9643	1.9192	1.8718	1.8217	1.7684	1.7110
26	4.2252	3.3690	2.9752	2.7426	2.5868	2.4741	2.3883	2.3205	2.2655	2.2197	2.1479	2.0716	1.9898	1.9464	1.9010	1.8533	1.8027	1.7488	1.6906
27	4.2100	3.3541	2.9604	2.7278	2.5719	2.4591	2.3732	2.3053	2.2501	2.2043	2.1323	2.0558	1.9736	1.9299	1.8842	1.8361	1.7851	1.7306	1.6717
28	4.1960	3.3404	2.9467	2.7141	2.5581	2.4453	2.3593	2.2913	2.2360	2.1900	2.1179	2.0411	1.9586	1.9147	1.8687	1.8203	1.7689	1.7138	1.6541
29	4.1830	3.3277	2.9340	2.7014	2.5454	2.4324	2.3463	2.2783	2.2229	2.1768	2.1045	2.0275	1.9446	1.9005	1.8543	1.8055	1.7537	1.6981	1.6376
30	4.1709	3.3158	2.9223	2.6896	2.5336	2.4205	2.3343	2.2662	2.2107	2.1646	2.0921	2.0148	1.9317	1.8874	1.8409	1.7918	1.7396	1.6835	1.6223
40	4.0847	3.2317	2.8387	2.6060	2.4495	2.3359	2.2490	2.1802	2.1240	2.0772	2.0035	1.9245	1.8389	1.7929	1.7444	1.6928	1.6373	1.5766	1.5089
60	4.0012	3.1504	2.7581	2.5252	2.3683	2.2541	2.1665	2.0970	2.0401	1.9926	1.9174	1.8364	1.7480	1.7001	1.6491	1.5943	1.5343	1.4673	1.3893
120	3.9201	3.0718	2.6802	2.4472	2.2899	2.1750	2.0868	2.0164	1.9588	1.9105	1.8337	1.7505	1.6587	1.6084	1.5543	1.4952	1.4290	1.3519	1.2539
inf	3.8415	2.9957	2.6049	2.3719	2.2141	2.0986	2.0096	1.9384	1.8799	1.8307	1.7522	1.6664	1.5705	1.5173	1.4591	1.3940	1.3180	1.2214	1.0000

F- table (α = 0.025)

df2/df1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	INF
1	647.7890	799.5000	864.1630	899.5833	921.8479	937.1111	948.2169	956.6562	963.2846	968.6274	976.7079	984.8668	993.1028	997.2492	1001.414	1005.598	1009.800	1014.020	1018.258
2	38.5063	39.0000	39.1655	39.2484	39.2982	39.3315	39.3552	39.3730	39.3869	39.3980	39.4146	39.4313	39.4479	39.4562	39.465	39.473	39.481	39.490	39.498
3	17.4434	16.0441	15.4392	15.1010	14.8848	14.7347	14.6244	14.5399	14.4731	14.4189	14.3366	14.2527	14.1674	14.1241	14.081	14.037	13.992	13.947	13.902
4	12.2179	10.6491	9.9792	9.6045	9.3645	9.1973	9.0741	8.9796	8.9047	8.8439	8.7512	8.6565	8.5599	8.5109	8.461	8.411	8.360	8.309	8.257
5	10.0070	8.4336	7.7636	7.3879	7.1464	6.9777	6.8531	6.7572	6.6811	6.6192	6.5245	6.4277	6.3286	6.2780	6.227	6.175	6.123	6.069	6.015
6	8.8131	7.2599	6.5988	6.2272	5.9876	5.8198	5.6955	5.5996	5.5234	5.4613	5.3662	5.2687	5.1684	5.1172	5.065	5.012	4.959	4.904	4.849
7	8.0727	6.5415	5.8898	5.5226	5.2852	5.1186	4.9949	4.8993	4.8232	4.7611	4.6658	4.5678	4.4667	4.4150	4.362	4.309	4.254	4.199	4.142
8	7.5709	6.0595	5.4160	5.0526	4.8173	4.6517	4.5286	4.4333	4.3572	4.2951	4.1997	4.1012	3.9995	3.9472	3.894	3.840	3.784	3.728	3.670
9	7.2093	5.7147	5.0781	4.7181	4.4844	4.3197	4.1970	4.1020	4.0260	3.9639	3.8682	3.7694	3.6669	3.6142	3.560	3.505	3.449	3.392	3.333
10	6.9367	5.4564	4.8256	4.4683	4.2361	4.0721	3.9498	3.8549	3.7790	3.7168	3.6209	3.5217	3.4185	3.3654	3.311	3.255	3.198	3.140	3.080
11	6.7241	5.2559	4.6300	4.2751	4.0440	3.8807	3.7586	3.6638	3.5879	3.5257	3.4296	3.3299	3.2261	3.1725	3.118	3.061	3.004	2.944	2.883
12	6.5538	5.0959	4.4742	4.1212	3.8911	3.7283	3.6065	3.5118	3.4358	3.3736	3.2773	3.1772	3.0728	3.0187	2.963	2.906	2.848	2.787	2.725
13	6.4143	4.9653	4.3472	3.9959	3.7667	3.6043	3.4827	3.3880	3.3120	3.2497	3.1532	3.0527	2.9477	2.8932	2.837	2.780	2.720	2.659	2.595
14	6.2979	4.8567	4.2417	3.8919	3.6634	3.5014	3.3799	3.2853	3.2093	3.1469	3.0502	2.9493	2.8437	2.7888	2.732	2.674	2.614	2.552	2.487
15	6.1995	4.7650	4.1528	3.8043	3.5764	3.4147	3.2934	3.1987	3.1227	3.0602	2.9633	2.8621	2.7559	2.7006	2.644	2.585	2.524	2.461	2.395
16	6.1151	4.6867	4.0768	3.7294	3.5021	3.3406	3.2194	3.1248	3.0488	2.9862	2.8890	2.7875	2.6808	2.6252	2.568	2.509	2.447	2.383	2.316
17	6.0420	4.6189	4.0112	3.6648	3.4379	3.2767	3.1556	3.0610	2.9849	2.9222	2.8249	2.7230	2.6158	2.5598	2.502	2.442	2.380	2.315	2.247
18	5.9781	4.5597	3.9539	3.6083	3.3820	3.2209	3.0999	3.0053	2.9291	2.8664	2.7689	2.6667	2.5590	2.5027	2.445	2.384	2.321	2.256	2.187
19	5.9216	4.5075	3.9034	3.5587	3.3327	3.1718	3.0509	2.9563	2.8801	2.8172	2.7196	2.6171	2.5089	2.4523	2.394	2.333	2.270	2.203	2.133
20	5.8715	4.4613	3.8587	3.5147	3.2891	3.1283	3.0074	2.9128	2.8365	2.7737	2.6758	2.5731	2.4645	2.4076	2.349	2.287	2.223	2.156	2.085
21	5.8266	4.4199	3.8188	3.4754	3.2501	3.0895	2.9686	2.8740	2.7977	2.7348	2.6368	2.5338	2.4247	2.3675	2.308	2.246	2.182	2.114	2.042
22	5.7863	4.3828	3.7829	3.4401	3.2151	3.0546	2.9338	2.8392	2.7628	2.6998	2.6017	2.4984	2.3890	2.3315	2.272	2.210	2.145	2.076	2.003
23	5.7498	4.3492	3.7505	3.4083	3.1835	3.0232	2.9023	2.8077	2.7313	2.6682	2.5699	2.4665	2.3567	2.2989	2.239	2.176	2.111	2.041	1.968
24	5.7166	4.3187	3.7211	3.3794	3.1548	2.9946	2.8738	2.7791	2.7027	2.6396	2.5411	2.4374	2.3273	2.2693	2.209	2.146	2.080	2.010	1.935
25	5.6864	4.2909	3.6943	3.3530	3.1287	2.9685	2.8478	2.7531	2.6766	2.6135	2.5149	2.4110	2.3005	2.2422	2.182	2.118	2.052	1.981	1.906
26	5.6586	4.2655	3.6697	3.3289	3.1048	2.9447	2.8240	2.7293	2.6528	2.5896	2.4908	2.3867	2.2759	2.2174	2.157	2.093	2.026	1.954	1.878
27	5.6331	4.2421	3.6472	3.3067	3.0828	2.9228	2.8021	2.7074	2.6309	2.5676	2.4688	2.3644	2.2533	2.1946	2.133	2.069	2.002	1.930	1.853
28	5.6096	4.2205	3.6264	3.2863	3.0626	2.9027	2.7820	2.6872	2.6106	2.5473	2.4484	2.3438	2.2324	2.1735	2.112	2.048	1.980	1.907	1.829
29	5.5878	4.2006	3.6072	3.2674	3.0438	2.8840	2.7633	2.6686	2.5919	2.5286	2.4295	2.3248	2.2131	2.1540	2.092	2.028	1.959	1.886	1.807
30	5.5675	4.1821	3.5894	3.2499	3.0265	2.8667	2.7460	2.6513	2.5746	2.5112	2.4120	2.3072	2.1952	2.1359	2.074	2.009	1.940	1.866	1.787
40	5.4239	4.0510	3.4633	3.1261	2.9037	2.7444	2.6238	2.5289	2.4519	2.3882	2.2882	2.1819	2.0677	2.0069	1.943	1.875	1.803	1.724	1.637
60	5.2856	3.9253	3.3425	3.0077	2.7863	2.6274	2.5068	2.4117	2.3344	2.2702	2.1692	2.0613	1.9445	1.8817	1.815	1.744	1.667	1.581	1.482
120	5.1523	3.8046	3.2269	2.8943	2.6740	2.5154	2.3948	2.2994	2.2217	2.1570	2.0548	1.9450	1.8249	1.7597	1.690	1.614	1.530	1.433	1.310
inf	5.0239	3.6889	3.1161	2.7858	2.5665	2.4082	2.2875	2.1918	2.1136	2.0483	1.9447	1.8326	1.7085	1.6402	1.566	1.484	1.388	1.268	1.000

F- table (α = 0.01)

df2/df1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	INF
1	4052.181	4999.500	5403.352	5624.583	5763.650	5858.986	5928.356	5981.070	6022.473	6055.847	6106.321	6157.285	6208.730	6234.631	6260.649	6286.782	6313.030	6339.391	6365.864
2	98.503	99.000	99.166	99.249	99.299	99.333	99.356	99.374	99.388	99.399	99.416	99.433	99.449	99.458	99.466	99.474	99.482	99.491	99.499
3	34.116	30.817	29.457	28.710	28.237	27.911	27.672	27.489	27.345	27.229	27.052	26.872	26.690	26.598	26.505	26.411	26.316	26.221	26.125
4	21.198	18.000	16.694	15.977	15.522	15.207	14.976	14.799	14.659	14.546	14.374	14.198	14.020	13.929	13.838	13.745	13.652	13.558	13.463
5	16.258	13.274	12.060	11.392	10.967	10.672	10.456	10.289	10.158	10.051	9.888	9.722	9.553	9.466	9.379	9.291	9.202	9.112	9.020
6	13.745	10.925	9.780	9.148	8.746	8.466	8.260	8.102	7.976	7.874	7.718	7.559	7.396	7.313	7.229	7.143	7.057	6.969	6.880
7	12.246	9.547	8.451	7.847	7.460	7.191	6.993	6.840	6.719	6.620	6.469	6.314	6.155	6.074	5.992	5.908	5.824	5.737	5.650
8	11.259	8.649	7.591	7.006	6.632	6.371	6.178	6.029	5.911	5.814	5.667	5.515	5.359	5.279	5.198	5.116	5.032	4.946	4.859
9	10.561	8.022	6.992	6.422	6.057	5.802	5.613	5.467	5.351	5.257	5.111	4.962	4.808	4.729	4.649	4.567	4.483	4.398	4.311
10	10.044	7.559	6.552	5.994	5.636	5.386	5.200	5.057	4.942	4.849	4.706	4.558	4.405	4.327	4.247	4.165	4.082	3.996	3.909
11	9.646	7.206	6.217	5.668	5.316	5.069	4.886	4.744	4.632	4.539	4.397	4.251	4.099	4.021	3.941	3.860	3.776	3.690	3.602
12	9.330	6.927	5.953	5.412	5.064	4.821	4.640	4.499	4.388	4.296	4.155	4.010	3.858	3.780	3.701	3.619	3.535	3.449	3.361
13	9.074	6.701	5.739	5.205	4.862	4.620	4.441	4.302	4.191	4.100	3.960	3.815	3.665	3.587	3.507	3.425	3.341	3.255	3.165
14	8.862	6.515	5.564	5.035	4.695	4.456	4.278	4.140	4.030	3.939	3.800	3.656	3.505	3.427	3.348	3.266	3.181	3.094	3.004
15	8.683	6.359	5.417	4.893	4.556	4.318	4.142	4.004	3.895	3.805	3.666	3.522	3.372	3.294	3.214	3.132	3.047	2.959	2.868
16	8.531	6.226	5.292	4.773	4.437	4.202	4.026	3.890	3.780	3.691	3.553	3.409	3.259	3.181	3.101	3.018	2.933	2.845	2.753
17	8.400	6.112	5.185	4.669	4.336	4.102	3.927	3.791	3.682	3.593	3.455	3.312	3.162	3.084	3.003	2.920	2.835	2.746	2.653
18	8.285	6.013	5.092	4.579	4.248	4.015	3.841	3.705	3.597	3.508	3.371	3.227	3.077	2.999	2.919	2.835	2.749	2.660	2.566
19	8.185	5.926	5.010	4.500	4.171	3.939	3.765	3.631	3.523	3.434	3.297	3.153	3.003	2.925	2.844	2.761	2.674	2.584	2.489
20	8.096	5.849	4.938	4.431	4.103	3.871	3.699	3.564	3.457	3.368	3.231	3.088	2.938	2.859	2.778	2.695	2.608	2.517	2.421
21	8.017	5.780	4.874	4.369	4.042	3.812	3.640	3.506	3.398	3.310	3.173	3.030	2.880	2.801	2.720	2.636	2.548	2.457	2.360
22	7.945	5.719	4.817	4.313	3.988	3.758	3.587	3.453	3.346	3.258	3.121	2.978	2.827	2.749	2.667	2.583	2.495	2.403	2.305
23	7.881	5.664	4.765	4.264	3.939	3.710	3.539	3.406	3.299	3.211	3.074	2.931	2.781	2.702	2.620	2.535	2.447	2.354	2.256
24	7.823	5.614	4.718	4.218	3.895	3.667	3.496	3.363	3.256	3.168	3.032	2.889	2.738	2.659	2.577	2.492	2.403	2.310	2.211
25	7.770	5.568	4.675	4.177	3.855	3.627	3.457	3.324	3.217	3.129	2.993	2.850	2.699	2.620	2.538	2.453	2.364	2.270	2.169
26	7.721	5.526	4.637	4.140	3.818	3.591	3.421	3.288	3.182	3.094	2.958	2.815	2.664	2.585	2.503	2.417	2.327	2.233	2.131
27	7.677	5.488	4.601	4.106	3.785	3.558	3.388	3.256	3.149	3.062	2.926	2.783	2.632	2.552	2.470	2.384	2.294	2.198	2.097
28	7.636	5.453	4.568	4.074	3.754	3.528	3.358	3.226	3.120	3.032	2.896	2.753	2.602	2.522	2.440	2.354	2.263	2.167	2.064
29	7.598	5.420	4.538	4.045	3.725	3.499	3.330	3.198	3.092	3.005	2.868	2.726	2.574	2.495	2.412	2.325	2.234	2.138	2.034
30	7.562	5.390	4.510	4.018	3.699	3.473	3.304	3.173	3.067	2.979	2.843	2.700	2.549	2.469	2.386	2.299	2.208	2.111	2.006
40	7.314	5.179	4.313	3.828	3.514	3.291	3.124	2.993	2.888	2.801	2.665	2.522	2.369	2.288	2.203	2.114	2.019	1.917	1.805
60	7.077	4.977	4.126	3.649	3.339	3.119	2.953	2.823	2.718	2.632	2.496	2.352	2.198	2.115	2.028	1.936	1.836	1.726	1.601
120	6.851	4.787	3.949	3.480	3.174	2.956	2.792	2.663	2.559	2.472	2.336	2.192	2.035	1.950	1.860	1.763	1.656	1.533	1.381
inf	6.635	4.605	3.782	3.319	3.017	2.802	2.639	2.511	2.407	2.321	2.185	2.039	1.878	1.791	1.696	1.592	1.473	1.325	1.000