وزارة التعليم العالي والبحث العلمي

جامعة بابل /كلية التربية للعلوم الصرفة

قسم الرياضيات

اختبار الفرضيات حول

(اختبارات النسب - حسن المطابقة)

مشروع بحث تقدم به الطالب سعد كاظم حسن الى مجلس كلية التربية للعلوم الصرفة قسم الرياضيات

كجزء من متطلبات نيل شهادة البكالوريوس في الرياضيات

اشراف د. كوثر فوزي

1444 هـ

2023م

بِسۡمِ ٱللَّهِ ٱلرَّحۡمَٰنِ ٱلرَّحِيمِ

"وَقُل رَّبِّ زِدۡنِي عِلۡمًا"

صدق اللَّهُ العلي العظيم

«سورة طه: الآية 114»

أهداء

إلى من أُفضِّلها على نفسي، ولِمَ لا؛ فلقد ضحَّت من أجلي ولم تدَّخر جُهدًا في سبيل إسعادي على الدَّوام (أُمِّي الحبيبة). نسير في دروب الحياة، ويبقى من يُسيطر على أذهاننا في كل مسلك نسلكه صاحب الوجه الطيب، والأفعال الحسنة. فلم يبخل عليَّ طيلة حياته (والدي العزيز). الى اختي المرحومة زهراء كامل كزار و اخوتي واخواتي و أصدقائي، وجميع من وقفوا بجواري وساعدوني بكل ما يملكون، وفي أصعدة كثيرة أُقدِّم لكم هذا البحث، وأتمنَّى أن يحوز على رضاكم

شكر وتقدير

لابد لنا ونحن نخطو خطوتنا الاخيره في الحياه الجامعية من وقفه نعود الى اعوام

قضيناها في رحاب الجامعة مع أساتذتنا الكرام الذين قدمو لنا الكثير باذلين بذلك جهود

كبيره في بناء جيل المستقبل لتبعث الامه من جديد .

تحديدا الى دكتورتي التي اشرفت على انجاز هذا البحث الدكتوره "كوثر فوزي " .

وقبل أن نمضي أتقدم بأسمى أيات الشكر والامتنان والتقدير والمحبة الى الذين حملوا

اقدس رسالة في الحياة إلى إلذين مهدوا لنا طريق العلم والمعرفة ...... إلى جميع

أساتذتنا الافاضل

الملخص

في هذا البحث قد درسنا مفهوم الفرضية الإحصائية وانواعها وكيفه_اختيارها

والتعرف على أنواع الأخطاء المترتبة في الاختبار ثم تناول مفهوم النسب وتعين

اختبارات تتعلق بنسبة واحدة ونسبتين مع امثلة الفرضيات ثم تناول موضوع

حسن المطابقة التي تفيد في جودة التوفيق

مقدمة حول البحث

الفرضية الإحصائية هي أي إدعاء أو مقولة تخص وسيط الجمهرة وهذه الفرضية قد تكون صحيحة وقد تكون خاطئة.

كما يطلق عليه تسمية اختبار الفرضيات الإحصائي أو اختبار الفرضيات وهي تعبر عن خوارزمية إحصائية لإتخاذ قرار بشأن فرضية معينة تخص بيانات احصائية (أي تخص أحد الوسطاء مثل المتوسط أو التباين) أو جمهرة احصائية ما. والقرار قد يكون بدعم الفرضية أو رفضها حسب مجال معين من الثقة تحدده طبيعة الدراسة وطبيعة البيانات الإحصائية، ويحدد الاختبار مدى انطباق البيانات المتوفرة مع الفرضية المدروسة مثل وجود علاقة بين خاصيتين لأفراد العينة (الجمهرة) الإحصائية.

الفصل الأول

1-1. أختبار الفرضيات

تقصد في الفرضيات الإحصائية Statistical Hypotheses هي الفرضيات التي تتعلق في المجتمع الإحصائي المسحوبة منه العينة، أو توزيع هذا المجتمع أو معالمه كالوسط الحسابي أو النسبة في المجتمع. والفرضية ما هي إلا تخمين أو استنتاج ذكي مبني على حيثيات معقولة أو منطقية ولكنه ليس مبنياً على حسابات دقيقة خاصة بالمجتمع لأننا نفترض أنه لا يمكن دراسة المجتمع بالكامل عن طريق الحصر الشامل بل نحاول استنتاج أو الاستدلال على مقاييس المجتمع باستخدام بيانات ونتائج العينة. فمثلاً : قد يفترض الباحث أن متوسط الدخل الشهري للفرد في دولة ما هو 200 دولار بناء على ما يراه من مستوى المعيشة في هذا البلد وأوضاعه الاقتصادية)، ويحتاج إلى اختبار علمي (إحصائي) لمعرفة مدى صحة هذه الفرضية أو قد يفترض باحث آخر أن نسبة الناخبين في إحدى الدوائر الذين يؤيدون مرشحاً معيناً لا تقل عن 30 % وهكذا... والمطلوب هو اختيار مدى صحة هذه الفرضيات. أي أن يصل الباحث إلى قرار إما بقبول الفرضية أو عدم قبولها (أي رفضها) وذلك باحتمال معين. وقبل تناول كيفية إجراء الاختبارات الإحصائية نستعرض أولاً بعض المفاهيم والتعريفات الأساسية اللازمة لهذا الموضوع حتى تكون الصورة أكثر وضوحاً..

2-1 أنواع الفرضيات

1 - الفرضية العدمية (أو الصفرية) The Null Hypothesis الفرضية العدمية هي الفرضية الأساسية المراد اختبارها. ويرمز لها عادة بالرمز :Ho. هذه الفرضية تأخذ عادة شكل معادلة أو مساواة فمثلاً إذا كانت الفرضية العدمية المراد اختبارها هي أن متوسط دخل الفرد في إحدى المناطق هو 200 دولار شهرياً فإن هذا الفرض يكتب بالرموز كما يلي :

$$Ho: \mu = 200$$

ويقرأ بالشكل التالي :

الفرضية العدمية هي أن متوسط دخل الفرد في المنطقة هو 200 دولاراً شهرياً.

وكمثال آخر : إذا كان الفرض المراد اختباره هو أن نسبة المؤيدين لبرنامج اقتصادي معين بين عمال أحد المصانع هي ، 30 ، فإن هذه الفرضية تكتب بالرموز كما يلي : $Ho: P = 0.30$

وتقرأ بالشكل التالي :

الفرضية العدمية: هي أن نسبة المؤيدين للبرنامج الاقتصادي بين عمال المصنع هي 0.30 وليس شرطاً أن تصاغ الفرض العدمية بالرموز ، فقد يتم التعبير عنه بدون رموز. فقد يريد الباحث أن يختبر ما إذا كانت هناك علاقة بين الأمية والاستعداد للانحراف، أو بين المؤهل العلمي ودرجة الوعي السياسي. فقد يصيغ الباحث الفرضية العدمية بالشكل التالي على سبيل المثال الأمية والاستعداد للانحراف مستقلان (أي) لا توجد علاقة بينهما، أو أن العلاقة بينهما منعدمة.

1.3 - الفرضية البديلة : The Alternative Hypothesis في اختبارات الفرضيات يتحتم وضع فرضية أخرى غير الفرضية العدمية المراد اختبارها تسمى الفرضية البديلة. وهذه الفرضية " هي التي ستقبل في حالة رفض الفرضية العدمية " أي لابد من تحديد فرضية أخرى بديلة في الوقت الذي نحدد فيه الفرضية العدمية، وبالتالي فإن الفرضية البديلة تعرف كما يلي : الفرضية البديلة هي الفرضية الأخرى التي ستقبل في حالة رفض الفرضية العدمية" ويرمز له عادة بالرمز : H_1

والفرضية البديلة لها أهمية كبيرة وبالذات في قياس الظواهر الاجتماعية – كما سوف نرى فهي التي تحدد نوع الاختبار المستخدم لذلك فهو يأخذ أحد أشكال ثلاثة هي :

أ- أن يأخذ شكل " لا يساوي ". وفي هذه الحالة نستخدم ما يسمى اختبار الطرفين فمثلاً : إذا كانت الفرضية العدمية هي أن متوسط الدخل الشهري لفئة معينة في المجتمع هي 200 دولار .

$$Ho: \mu = 200$$

فإن الفرضية البديلة في هذه الحالة تأخذ الشكل التالي :

$$H_1: \mu 200$$

بمعنى أن متوسط دخل هذه الفئة من المجتمع " لا يساوي " 200 دولار شهرياً.

ب- أو أن يأخذ شكل " أكبر من .. وفي هذه الحالة نستخدم ما يسمى" اختبار الطرف الأيمن". فمثلاً: قد تكون الفرضية البديلة كما يلي : $H_1: \mu > 200$ أي أن متوسط الدخل لهذه الفئة من المجتمع أكبر من 200 دولار شهرياً.

ج- وأخيراً قد يأخذ الفرض البديل شكل " أقل من". وفي هذه الحالة نستخدم ما يسمى " اختبار الطرف الأيسر". فمثلاً: قد تكون الفرضية البديلة هي:

$$H_1: \mu < 200$$

أي أن متوسط الدخل لهذه الفئة من المجتمع أقل من 200 دولاراً شهرياً. مهم والخلاصة هي لابد للباحث من تحديد الفرضية البديلة التي لا تخرج عن أحد الأشكال الثلاثة السابقة، وهذا التحديد جداً قبل الدخول فى تفاصيل الاختبار الإحصائي وذلك لأنه هو الذي يحدد نوع الاختبار المستخدم كما سوف نرى.

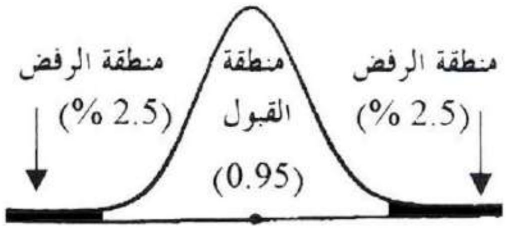
3-1. مستوى المعنوية : Level of Significance

يعتبر مصطلح " مستوى المعنوية " واحداً من أهم المصطلحات المستخدمة في دراسة نظرية اختبارات الفرضيات والمقصود بمستوى المعنوية هو احتمال حدوث الخطأ من النوع الأول ". أو نسبة حدوثه " أي احتمال رفض الفرضية العدمية بينما هي صحيحة ". وعادة ما يرمز إلى مستوى المعنوية بالرمز اللاتيني ألفا α وأشهر قيمتين لمستوى المعنوية هما 5% ، 1% ولكن ليس هناك ما يمنع من أن يأخذ قيما أخرى.

ومن الملاحظات المهمة هنا هو أن مستوى المعنوية " والذي يسمى أحياناً " مستوى الدلالة " هو المكمل لدرجة الثقة " بمعنى أن مجموعهما يساوي 100% أو واحد صحيح. فإذا كانت درجة الثقة% 95 فإن مستوى المعنوية يساوي 5% والعكس صحيح فإذا كان مستوى المعنوية 5% فإن هذا يعني أن درجة الثقة .95% ولعل من أهم الملاحظات هنا هو استخدام تعبير مستوى المعنوية في حالات اختبارات الفرضيات، بينما يستخدم مصطلح درجة أو مستوى الثقة في حالات التقدير.

والفكرة الأساسية في اختبار الفرضية هي تقسيم المساحة تحت المنحني إلى منطقتين: أحداهما تسمى " منطقة القبول " أي منطقة قبول الفرضية العدمية.
والأخرى تسمى " منطقة الرفض"، أي منطقة رفض الفرضية العدمية والتي تسمى أحيانا " بالمنطقة الحرجة Critical region ". والنقطة الجديرة بالملاحظة هنا هي أن منطقة القبول تمثل درجة الثقة، بينما تمثل منطقة الرفض مستوى المعنوية وهناك ثلاث حالات مختلفة لمنطقتي القبول والرفض هي :

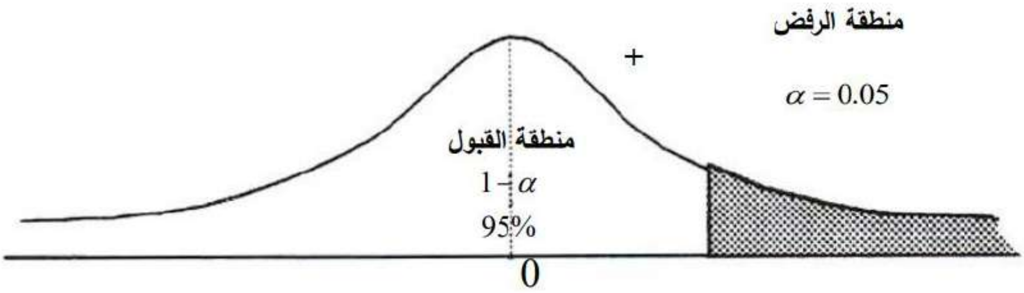
<u>الأولى :</u> إذا كانت الفرضية البديلة تأخذ شكل " لا يساوي " كأن يكون الفرض في هذه الحالة هو أن متوسط دخل الفرد لا يساوي 200 دولاراً فإن منطقة الرفض تكون موزعة على طرفي المنحنى بالتساوي، ويسمى الاختبار في هذه الحالة " اختبار الطرفين "، والذي يأخذ الشكل التالي (بافتراض أن $\alpha = 5\%$) :



أختبار الطرفين

فالفرضية العدمية هنا Ho :μ 200 يعني أن متوسط دخل الفرد يساوي 200 دولار شهريا، والفرضية البديلة في هذه الحالة هو $H_1: \mu \neq 200$ بمعنى أن متوسط دخل الفرد لا يساوي 200 دولار شهرياً. حيث تمثل المنطقة البيضاء غير المظللة منطقة القبول والتي قد تساوي 95% وبالتالي فمنطقة الرفض مقسمة بالتساوي على طرفي المنحني في هذه الحالة تكون قيمة كل منهما % 2.5.
والنتيجة هو أن القرار أيا كان نوعه سيكون بمستوى معنوية 5% بمعنى أن احتمال أو نسبة الخطأ فيه من النوع الأول تساوي 5%.

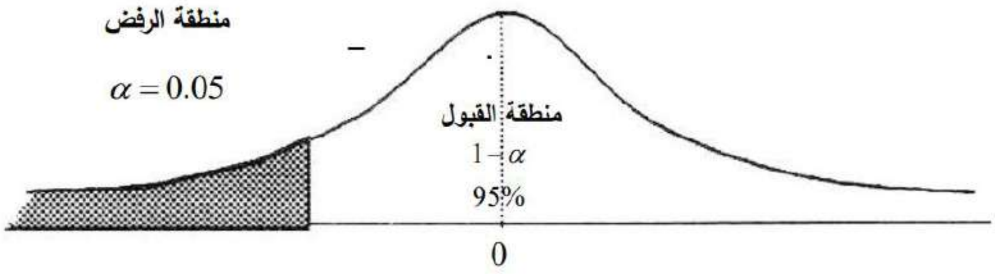
<u>الثانية:</u> إذا كانت الفرضية البديلة تأخذ شكل أكبر" "من" فإن منطقة الرفض تكون مركزة بالكامل في الطرف الأيمن للمنحنى. ويسمى الاختبار في هذه الحالة اختبار الطرف الأيمن، والذي يأخذ الشكل التالي أدناه:



اختبار الطرف الأيمن

فالفرضية العدمية هنا نفس فرض المثال السابق، بينما الفرضية البديلة هي $H1: \mu > 200$بمعنى أن متوسط دخل الفرد أكبر من 200 دولاراً شهرياً. وبالتالي فإن مستوى المعنوية والذي يساوي مثلاً %5% مركز في الطرف الأيمن من المنحنى.

الثالثة : إذا كانت الفرضية البديلة تأخذ شكل " أقل من" فإن منطقة الرفض تكون مركزة بالكامل في الطرف الأيسر للمنحنى. ويسمى الاختبار في هذه الحالة اختبار الطرف الأيسر . والشكل التالي يوضح ذلك:



أختبار الطرف الايسر

مع افتراض ثبات الفرضية العدمية كما في المثال السابق، بينما الفرضية البديلة هي $H_1: \mu < 200$ بمعنى أن متوسط دخل الفرد أقل من 200 دولار شهرياً، وبالتالي فإن مستوى المعنوية والذي يساوي مثلاً 5% مركز في الطرف الأيسر من المنحنى. وسوف نتناول فيما يلي خطوات الاختبار الإحصائي بشيء من التفصيل.

4-1. الخطأ في اتخاذ القرار

ففي حالة قبول الباحث لفرضيته العدمية، فلا مجال للبحث في الفرضية البديلة، أما في حالة حدوث العكس بمعنى رفض الفرضية العدمية فإنه يتحتم في هذه الحالة قبول الفرضية البديلة، على أنه من الجدير بالذكر أن الباحث هنا عرضة للوقوع في الخطأ عند اتخاذ قراره بقبول الفرضية العدمية أو رفضها، فقد يرفض فرضاً هو في الواقع صحيح، وقد يقبل فرضا هو في الواقع غير صحيح. لذلك فقد تم تصنيف هذه الأخطاء إلى نوعين هما:

الخطأ من النوع الأول : Type I error

الخطأ من النوع الأول هو رفض" الفرضية العدمية بينما هي صحيحة ". أي أنه على الرغم من أن الفرضية العدمية في الواقع انها صحيحة وكان من الواجب قبولها فقد تم أخذ قرار خاطئ برفضها. وباختصار شديد فإن الخطأ من النوع الأول هو " رفض الفرضية صحيح".

الخطأ من النوع الثاني : Type II error

وفي المقابل فإن الخطأ من النوع الثاني يعني " قبول الفرضية العدمية بينما هي خاطئة " أي أنه على الرغم من أن الفرضية العدمية خاطئة وكان من الواجب رفضها فقد تم أخذ قرار خاطئ بقبولها وباختصار شديد فإن الخطأ من النوع الثاني هو " قبول الفرضية خاطئ ". وقد يتساءل البعض عند مدى إمكانية تصغير الخطأين معاً ولكن لسوء الحظ لا يمكن تصغيرهما معاً إلى أدنى حد ممكن، ويبدو أن الطريقة الوحيدة المتاحة لذلك هي زيادة (أو تكبير حجم العينة، الأمر الذي قد لا يكون ممكنا في كل الحالات. لذلك فإن الذي يحدث عادة هو تثبيت أحدهما كأن يكون نسبة أو احتمال حدوث الخطأ من النوع الأول ومحاولة تصغير الآخر.

القرار / الفرضية	رفض H_0	قبول H_0
H_0 صحيحة	قرار خاطئ الخطأ من النوع الأول	قرار صحيح
H_0 غير صحيحة	قرار صحيح	قرار خاطئ الخطأ من النوع الثاني

1-5.أمثلة حول الفرضيات

مثال (1) عينة عشوائية حجمها 49 شخصاً اختيرت من أفراد دولة ما، فإذا كان الوسط الحسابي لدخول الأفراد الأسبوعية في العينة هو 75 دولاراً إذا علمت أن الانحراف المعياري للمجتمع لدخول الأفراد يساوي 14 دولاراً. كيف يمكن اختبار الفرض العدمي بأن متوسط الدخل الأسبوعي لمواطني هذه الدولة يساوي 72 دولاراً مقابل الفرض البديل أنه لا يساوي 72 وذلك بمستوى معنوية 5%

الحل: أن المعلومات المتوافرة $n = 49, \sigma = 14, \bar{X} = 75, \mu = 72$

1- الفرضية العدمية : أن متوسط المجتمع للدخل الاسبوعي يساوي 72 وبالرموز :

$Ho: \mu=72$

2- الفرضية البديلة : أن المتوسط المجتمع للدخل الاسبوعي لا يساوي 72

وبالرموز : $H_1: \mu \neq 72$

3- الإحصائية : بما أن العينة التباين أو الانحراف المعياري معلوم فإن الإحصائية في حالة اختبار الوسط تأخذ الشكل التالي :

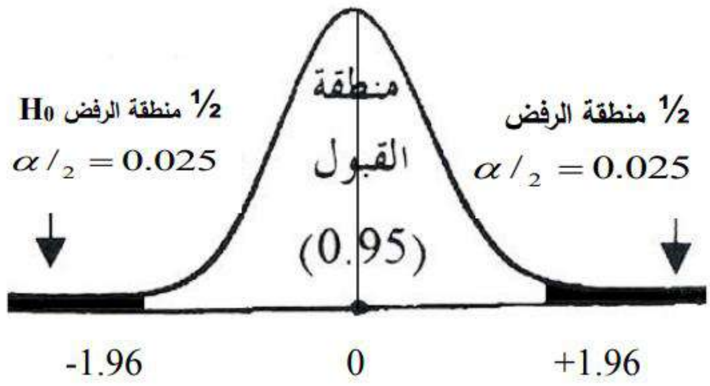
$$Z_{\text{cal}} = Z_{\bar{X}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{75 - 72}{\frac{14}{\sqrt{49}}} = \frac{3}{\frac{14}{7}} = \frac{3}{2} = 1.5$$

أي أن القيمة الاحصائية المحسوبة تساوي 1.5

$$Z_{tab} = Z_{\frac{\alpha}{2}} = Z_{\frac{0.05}{2}} = Z_{0.025} = \pm 1.96$$

و أن القيمة الاحصائية الجدولية ذات اختبار طرفين تساوي ± 1.96

4- حدود منطقتي القبول والرفض : نحصل عليها من التوزيع الطبيعي المعياري حيث مستوى المعنوية 5% وبما أن الفرضية البديلة هو : " لا يساوي " فإن ما يستعمل في هذه الحالة هو اختبار الطرفين كما في الشكل التالي:



وقد حصلنا على حدود منطقتي القبول والرفض وذلك بقسمة درجة الثقة (المكملة لمستوى المعنوية والتي تساوي 0.95 وبالكشف في جدول التوزيع الطبيعي المعياري عن Z نجد أنها تساوي 1.96 وحيث أنها موزعة على طرفي المنحنى بالتساوي فنضع إشارة موجبة في النصف الأيمن، وإشارة سالبة في النصف الأيسر ، أي أن منطقة القبول تبدأ من القيمة 1.96- وتستمر حتى القيمة 1.96+ (أي أن أي قيمة محصورة بين هاتين القيمتين تكون في منطقة القبول، وأي قيمة خارج هذه الحدود تكون في منطقة الرفض).

5- المقارنة والقرار : وبمقارنة قيمة الإحصائية المحسوبة من الخطوة رقم 3 (والتي تساوي 1.5) بحدود منطقتي القبول والرفض (من الخطوة رقم 4) نجد أنها تقع في منطقة القبول لذلك فإن القرار هو : قبول الفرضية العدمية التي تنص بأن متوسط دخول الأفراد الأسبوعية في هذه الدولة يساوي 72 دولاراً وذلك بمستوى معنوية % 5.

2-1. اختبارات النسب :

ان موضوع النسب كثيرا ما يستخدم في البحوث والدراسات لتفسير الظواهر قيد البحث سواء اكانت أدارية ، طبية، نفسية ...... الخ . والنسبة هي ناتج قسمة الجزء على الكل ومن الامثلة التي توضح فائدة النسبة في الوصف والتحليل الاحصائي ، نسبة النجاح في امتحان اجري لمجموعة من الطلبة أو نسبة الاجابات الصحيحة في احدى الاختبارات الى المجموع الكلي للإجابات وذلك لتحديد مستوى الاسئلة او مستوى الطلبة .... الخ. فاذا تم تسمية الاجابات الصحيحة بنسبة النجاح فان النسبة المتبقية تسمى الفشل ، وتتضمن اختبارات النسب موضوعين هما :

2-2. أنواع اختبارات النسب

أ- اختبار يتعلق بنسبة واحدة

ب- اختبار يتعلق بنسبتين

(أ) الاختبار الذي يتعلق بنسبة واحدة : Single proportion test لتكن (p) تمثل نسبة النجاح وترغب بأجراء اختبار عليها بمقارنتها بقيمة معلومة نسبة معلومة ولتكن (p_o) فان نسبة الفشل تمثل ب (q_0) وان قيمتها تساوي $(q= 1-p)$:test .

فرضية الاختبار

فرضية العدم

$Ho: P= Po$ ضد الفرضية البديلة او $H1: P\# Po$ $H1: P > Po$ احصاءة الاختبار : ان احصاءة الاختبار التي يجري اختبارها ناتجة من تقريب توزيع ثنائي الحدين الى التوزيع الطبيعي وصيغتها كالاتي : $H1: P< Po$ $P\text{-}Po$ $Z= n$ حيث ان 2 احصاءة الاختبار تتوزع توزيع طبيعي قياسي . P : نسبة النجاح في العينة Po: قيمة معلومة (نسبة معينة). do نسبة الفشل (po-1) .

3-2. اختبار النسبة لمجتمع ذي توزيع ثنائي الحدين :

في هذا المبحث سوف نتطرق إلى حالتين لحجوم العينات وهي عندما يكون حجم العينة كبير وعندما يكون حجم العينة صغير وكمـــا يلي:-

أولاً: عندما يكون حجم العينة كبير:-

ليكن x متغير عشوائي وصفي يشير إلى عدد حالات النجاح في (n) من المحاولات المستقلة (حيث أن كبيرة باحتمال نجاح المحاولة قدره (P) وحيث أن هكذا متغير يتوزع على وفق توزيع ثنائي الحدين بوسط حسابي هو (n) وتباين (n) ، حيث أن [q=1-p والتـ تمثل عدد حالات الفشل مقسومة على عدد الحالات الكلية، واستنادا إلى ما تم ذكره سابقا فإن الدرجة المعيارية في توزيع ثنائي الحدين يؤول توزيعها إلى التوزيع الطبيعي القياسي عندما تكون (n) كبيرة وبالتالي فإن:-

$$X \sim N(np, npq)$$

وان الدرجة المعيارية تصبح :

$$Z = \frac{x - np}{\sqrt{npq}} \sim N(0.1)$$

ولغرض الوصول إلى صيغة نهائية لمعيار الاختبار لاختبار فرضية العدم $H_0: P = P_0$: حيث أن P_0 تمثل قيمة نسبة معطاة) ضد أي فرضية بديلة أخرى، فسوف نفترض أن:

$$\hat{P} = \frac{X}{n}$$

تمثل تقدير نسبة عدد حالات النجاح من عدد المحاولات الكلية محسوبة من إجمالي البيانات التي تم الحصول عليها من العينة، وعليه فأن معيار الاختبار يصبح

$$Z = \frac{n\hat{p} - np}{\sqrt{npq}}$$

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

ويمكن الاستعاضة عن قيمة p بالقيمة المعطاة تحت فرضية العدم وأن (1) تمثل مكملة هذه القيمة.

ثانياً: عندما يكون حجم العينة صغير:- لغرض اختبار النسبة عندما يكون حجم العينة صغير سوف نأخذ حالة وحسب نوع الفرضية البديلة وكما يلي:-

إذا كانت فرضيتا الاختبار:

$H_0: P = P_0$
vs $H_1: P < P_o$

قيمة أكبر عدد صحيح $(K'_{\propto/2})$ يحقق الشرط المذكور هو : $X_t = 5, K'_{\frac{\alpha}{2}} = 5$

4-2. اختبار الفرق بين نسبتين :

لنفترض أن X_1 متغير عشوائي يمثل عدد حالات النجاح في العينة الأولي والتي يكون عدد المحاولات المستقلة فيها n_1 وباحتمال نجاح المحاولة قدره P_1 ، وأن X_2متغير عشوائي يمثل عدد حالات النجاح في من المحاولات المستقلة وباحتمال نجاح المحاولة قدره P_2، علما أن الصفة المدروسة في العينتين هي نفس الصفة، وحيث أن كل منهم لــه توزيع ثنائي الحدين سوف نلاحظ ما يلي:- إن الوسط الحسابي والتباين لـ X_1 هماP_1n_1 و $P_1n_1\ q_1$على التوالي.

وأن الوسط الحسابي والتباين لـ X_2 هما n_2p_2 و $n_2p_2q_2$ على التوالي.

وعلى فرض أن العينتين مستقلتين عن بعضهما، وبناءً على ما تم ذكره سابقاً فأنه عندما يكون حجم العينة كبيرا يكون:-

$$\hat{p}_1 = \frac{x_1}{n_1} \sim N\left(p_1, \frac{p_1 q_1}{n_1}\right)$$
$$\hat{p}_2 = \frac{k_2}{n_2} \sim N\left(p_2, \frac{p_2 q_2}{n_2}\right)$$

لذالك فأن :

$$(\hat{p}_1 - \hat{p}_2) \sim N\left(p_1 - p_2, \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}\right)$$

وبالتالي فأن معيار الاختبار المناسب هنا هو :

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}} \sim (0,1)$$

وفي أغلب الحالات فأن الفرضية التي سنكون في صدد اختبارها هي:

$$H_o: p_1 = p_2 = p \Rightarrow p_1 - p_2 = o$$

وهذا يعني أن صيغة معيار الاختبار تحت فرضية العدم تكون:

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0,1)$$

وبما أن قيم كل من q , p مجهولة، إذن يجب تقديرهما وكما يلي:

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} \hat{q} = 1 - \hat{p}$$

وبالتالي تكون صيغة معيار الاختبار كما يلي:

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

حيث أن :

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

2-5. اختبار الفرق بين النسب

أفرض أن لدينا (k) من العينات بـ (k) من المتغيرات العشوائية وأن

$(i = 1,2, \dots, k)x_i$ متغير عشوائي يمثل عدد حالات النجاح في n_i من المحاولات المستقلة وباحتمال نجاح المحاولة قدره (P_i) علماً أن الصفة المدروسة في العينتين هي بنفس الصفة وأن كل منهم له توزيع ثنائي الحدين بوسط حسابي : $n_i p_i$ وتباين $n_i p_i q_i$، وأن حجوم العينات كبيرة، فأن معيار الاختبار الملائم لاختيار الفرضية

$$H_o: p_1 = p_2 = \cdots = p_k = p$$

ان هنالك على الأقل نسبة واحدة لاتساوي : vs $H_1: p$

$$x^2 = \sum_{i=1}^{k} \frac{\left(x_j - n_i\hat{p}\right)^2}{n_i\hat{p}\hat{q}}$$

حيث ان :

$$\hat{p} = \frac{\sum_{i=1}^{k} xi}{\sum_{i=1}^{k} ni}, q = 1 - \hat{p}$$

الفصل الثالث

3-1. حسن المطابقة

اختبار حسن المطابقة (جودة التوفيق) إن فكرة هذا الاختبار تتمحور حول هدف واحد وهو بيان مدى مطابقة التكرار المشاهد مع التكرار النظري (المتوقع) لظاهرة معينة وعلى أساس قياسات عينة معينة حيث يعتبر هذا الاختبار من أهم استخدامات توزيع مربع كاي (X^2)

ولغرض الوصول إلى الهدف المذكور في يجب تحديد مقدار انحرافات التكرارات المشاهدة عن التكرارات النظرية (المتوقعة) فإذا كان المقدار مقدار الانحراف قليل فذلك يعني أن بيانات العجينة المختارة تمثل المجتمع المسحوبة منه العينة، أما إذا كان مقدار الانحراف كبير فهذا يعني أن البيانات تم سحبها من مجتمع بتوزيع احتمالي غير محدد (غير معروف) أو أن هذه البيانات تم سحبها من مجتمع بتوزيع احتمالي يختلف عن التوزيع الاحتمالي المفترض وبناء على ما تقدم فأن فرضيتا الاختبار الملائمة هنا هي:

H_0 : ان بيانات العينة تتوزع على وفق توزيع (معين)

vs H_1 : ان بيانات العينة لا تتوزع على وفق التوزيع المذكور

او (إن) بيانات العينة تتوزع على وفق توزيع آخر أما معيار الاختبار الملائم فهو :

$$x^2 = \sum_{i=1}^{k} \frac{(o_i - E_i)^2}{E_i} \sim x^2_{(k-1)}$$

حيث أن:

K: تمثل عدد فئات (أو مستويات) العامل.

o_i التكرار المشاهدات للفئة .

E_i : التكرار المتوقع للفئة .

وإن قيمة التكرار المتوقع تحسب على وفق

$$E_i = P_2 \sum_{i=1}^{k} O_i$$

وأن : P_i تمثل احتمال تحقق الفئة أو المستوى i .

مع الأخذ بنظر الاعتبار أن مجموع التكرارات المتوقعة يجب أن يساوي مجموع التكرارات المشاهدة .

أي أن:-

$$\sum^{k} E_i = \sum^{k} O_i$$

علماً أن معيار الاختبار المذكور لا يمكن استخدامه إلا بعد التحقق مما . يلي:-

(1) إن مستويات المتغير المدروس (x) [أو فئاته] يجب أن تكون مستقلة فيما بينها، أي لا يجوز أن تنتمي مفردة معينة إلى أكثر من مستوى أو فئة من مستويات x

(2) أن يكون حجم العينة كبير، بمعنى آخر أن مجموع التكرارات المشاهدة يجب أن يكون أكبر من 50 مفردة، أما في حالة عدم تحقق الشرط فتدمج المستويات أو الفئات مع بعضها السابقة مع اللاحقة أو العكس] .

(3) من المعروف مما سبق أن درجة حرية الاختبار هي (k-1) وهذا القيد المطروح من عدد المستويات ناتج من أن

$$\left[\sum_{i=1}^{k} E_i = \sum_{i=1}^{k} O_i\right]$$

في حالة استخدام تقديرات لمعلمات في توزيع معين فيتم طرح عدد المعلمات المقدرة من درجة الحرية المذكورة فعلى سبيل المثال في التوزيع الطبيعي يجب تقدير معلمتين هما σ^2, μ، فبذلك تصبح درجة الحرية (k-1-2).

3-2. امثلة

مثال / كانت درجات الحرارة العظمى والمسجلة على مدى ى أيام أحد الأسابيع في أحد مناطق بغداد في الشهر الثاني كما يلي:-

الجمعة الخميس الأربعاء الثلاثاء الاثنين الأحد السبت

: اليوم 15 15 16 12 13 19 15 : درجة الحرارة العظمى.

هل أن درجات الحرارة تتوزع على شكل منتظم على ايام الأسبوع عند مستوى معنوية 0.05؟

الحل: أن فرضيتا الاختبار هنا إن درجات الحرارة تتوزع على شكل منتظم (متساوي) على أيام الأسبوع : H_0

أن درجات الحرارة لا تتوزع على شكل منتظم علـــى أيـــام الأسبوع : vs H_1

وحيث أننا نتكلم عن توزيع منتظم فإن احتمال ظهور أي يوم بدرجـــة حرارة تساوي ليوم آخر هي (1/7) للكل ولغرض تسهيل العمل الحسابي لحساب قيمة لمعيار الاختبار نكون الجدول الآتي:

يبين حساب متطلبات الأعمدة فيه لتسهيل العمل الحسابي

اليوم	O_i	P_i	$E_i = P_i \sum_{i=1}^{k} O_i$	$O_i - E_i$	$\frac{(O_i - E_i)^2}{E_i}$
السبت	15	1/7	15	0	0
الأحد	19	1/7	15	4	1.0667
الاثنين	13	1/7	15	-2	0.2667
الثلاثاء	12	1/7	15	-3	0.6
الأربعاء	16	1/7	15	1	0.0667
الخميس	15	1/7	15	0	0
الجمعة	15	1/7	15	0	0
المجموع	105	1		0	1.999

$$x^2 = \sum_{i=1}^{k} \frac{(o_i - E_i)^2}{E_i} = 1.999$$

ومن جداول توزيع مربع – كاي (X^2) نجد ان القيمة النظرية لـــ X^2 عند مستوى معنوية 0.05 ودرجة حرية K-1=7-1=6 هي :- $x^2_{(0.05,6)} = 12,59$

وحيث أن قيمة معيار الاختبار النظرية أكبر من قيمة كعيار الاختبار المحسوبة إذن نقبل فرضية العدم Ho أي أن درجات الحرارة تتوزع بشكل منتظم على أيام الأسبوع في إحدى مناطق بغداد عند مستوى معنوية 0.05 وأن متوسط درجة الحرارة اليومي خلال هذا الأسبوع هي (15) درجة مئوية.

مثال/ من خلال الخلفية النظرية لقوانين الوراثة تم معرفة أن مجموعة من العجول يمكن أن تصنف إلى ثلاثة أصناف وفق سلالاتها بالنسب التالية :3 :5 10 ففي عينة قوامها (1000) عجل وبعد أجراء الفحوص على مجموعة العجول هذه تبين أن عدد هذه العجول وفق أصنافها الثلاث هي:- الصنف: 1 2 3

عدد العجول: 149 348 503

هل تعتقد أن هذه النتائج تتطابق مع القانون الوراثي عند مستوى معنوية 0.01 ؟

الحل : أن فرضيتا الاختبار الملائمة:

إن مجموعة العجول تتوزع حسب سلالاتها بنسب 3: 5: 10 H_0 إن مجموعة العجول تتوزع حسب سلالاتها بنسب تختلف عن نسب قوانين الوراثة لغرض تسهيل الحسابات نكون الجدول أدناه :

يبين حساب المتطلبات (الأعمدة لتسهيل حساب x^2

الصنف	عدد العجول oi	P_i	E_i	$(o_i - E_i)^2 / E_i$
1	149	3/18	166.67	1.859
2	348	5/18	277.78	17.751
3	503	10/18	555.55	4.971
المجموع	1000	18/18	1000	24.58

حيث أن P للصنف الأول هو عبارة عن $\frac{3}{3+5+10}$ وهكذا لباقي الأصناف إذن قيمة معيار الاختبار هي:

$$x^2 = \sum_{i=1}^{k} \frac{(o_i - E_i)^2}{E_i} = 24.58$$

ومن جداول توزيع مربع كاي (x^2) وعند مستوى معنوية (0.01). ودرجة حرية (2 = 3-1) تكون قيمة معيار الاختبار النظرية هي:

$$x^2_{(0.01,2)} = 9.2103$$

وحيث أن قيمة معيار الاختبار المحسوبة أكبر من قيمة معيار الاختبار النظرية نرفض فرضية العدم H_o ونقبل الفرضية البديلة H_1 عند مستوى معنوية 0.01 ، أي أن عدد العجول لا يتوزع حسب سلالاتها وعلى وفق نسب قوانين الوراثة والتي هي 3 5 10

المصادر

1- كتاب الفرضيات الإحصائية – كمال علوان خلف المشهداني – الدكتور عماد حازم عبود

2- الإحصاء الرياضي – امير حنة هرمز

3- الإحصاء الوصفي و التطبيقي والحيوي – محمد حسين محمد رشيد

4- كتاب الإحصاء التطبيقي – عزام عبد الرحمن صبري