# إحصاء ١١١

# STAT 111

قسم الإحصاء – كلية العلوم

جامعة الملك عبد العزيز

# البـــــــاب الثامن

# التقدير واختبارات الفروض

## مقدمة:

في هذا الباب سوف نتناقش قسمين رئيسيين من الإحصاء الاستدلالي:

(1)تقدير معالم المجتمع.

(2)اختبارات فروض بشأن صحة قيم معالم المجتمعات.

**ونستعرض الآن تعريفين أساسيين في الاستدلال الإحصائي:**

**المعلمة:** هي عبارة عن خاصية أو مقياس يتم حسابها من المجتمع محل الدراسة. مثل الوسط الحسابي للمجتمع $(\mu)$، الإنحراف المعياري للمجتمع $(\sigma)$.

**الاحصاءة:** هي عبارة عن خاصية أو مقياس يتم حسابها من العينة المسحوبة من المجتمع محل الدراسه. مثل الوسط الحسابي للعينة $(\bar{x})$، الانحراف المعياري للعينة $(S)$.

# تقدير متوسط المجتمع:

**التقدير:** هو أسلوب إحصائي مبني على نظريات إحصائية، يستخدم لتقدير معلمة ما محل الاهتمام عن طريق استخدام مقاييس العينة.

هناك أسلوبان لتقدير معلمة المجتمع المجهولة وهما: *التقدير بنقطة* و*التقدير بفترة*.

١- أسلوب التقدير بنقطة:

في أسلوب التقدير بنقطة تستخدم بيانات العينة لتقدير معلمة المجتمع المجهولة بنقطة واحدة فقط، أي بقيمة واحدة فقط.

***ونستخدم الوسط الحسابي للعينة $\bar{x}$ كتقدير نقطة لمتوسط المجتمع $\mu$***

➢ لكل إحصاءة من الإحصاءات التي نحسبها من العينة الخطأ المعياري الخاص بها، فهناك الخطأ المعياري لمتوسط العينة و الخطأ المعياري للانحراف المعياري للعينة و الخطأ المعياري لمعامل الارتباط بين بيانات العينة وغيرها وهذا يدخل تحت موضوع الاستدلال الإحصائي. وسوف نركز فقط على تعريف الخطأ المعياري لمتوسط العينة.

**الخطأ المعياري للمتوسط**: هو الانحراف المعياري لتوزيع مجتمع متوسطات العينات . بمعنى انحراف متوسطات العينات عن متوسط مجتمعها .ويرمز له بالرمز $\sigma_{\bar{X}}$ حيث:

$$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}}$$

أو

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

*ملاحظة:* غالباً ما تكون قيمة الانحراف المعياري للمجتمع $(\sigma)$ غير معلومة ، لذا نقدرها بقيمة الانحراف المعياري للعينة $(S)$.

**مثال:**

**أوجد الخطأ المعياري للوسط الحسابي لعينة حجمها (49) تم سحبها من مجتمع له انحراف معياري يساوي (14).**

الحل: الخطأ المعياري لمتوسط العينة يساوي $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{14}{\sqrt{49}} = 2$ .

**مثال:**

**تم قياس مستوى الأداء على مهارة معينة لأفراد عينة حجمها (36)، فكان الوسط الحسابي والانحراف المعياري للعينة 30 و9 على الترتيب. أوجد الخطأ المعياري لمتوسط العينة.**

الحل:

لأن قيمة الانحراف المعياري للمجتمع ($\sigma$) مجهولة، لذا نقدرها بقيمة الانحراف المعياري للعينة($S$) . فيصبح لدينا تقدير للخطأ المعياري للوسط الحسابي للعينة يساوي:

$$\hat{\sigma}_{\bar{x}} = \frac{S}{\sqrt{n}} = \frac{9}{\sqrt{36}} = \mathbf{1.5}$$

## ٢- أسلوب التقدير بفترة:

وفي أسلوب التقدير بفترة تستخدم بيانات العينة لتقدير معلمة المجتمع المجهولة بفترة من القيم.

وجدنا فيما سبق أن تقدير النقطة نادراً ما يساوي المعلمة التي نرغب في تقديرها، لذلك فإننا نحدد فترة تحتوي على مجموعة من القيم تتضمن فيما بينها قيمة معلمة المجتمع، وتسمى هذه الفترة **بتقدير *(فترة) الثقة.***

***- درجة الثقة:***

هي احتمال وقوع المعلمة في فترة الثقة ويرمز لها بالرمز $(1 - \alpha)$.

***- مستوى المعنوية***

هي احتمال عدم وقوع المعلمة في فترة الثقة ويرمز لها بالرمز $\alpha$ .

***فمثلاً*** *إذا كانت درجة الثقة 95% يكون مستوى المعنوية يساوي 0.05.*

## تقدير متوسط المجتمع (μ) باستخدام الوسط الحسابي للعينة $(\bar{X})$ : $(n \geq 30)$

$$\left[\bar{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

- وتحدد قيمة $Z_{\alpha/2}$ حسب درجة الثقة كما يلي:

| درجة الثقة | فإن قيمة $Z_{\alpha/2}$ |
|---|---|
| 90% | 1.65 |
| 95% | 1.96 |
| 99% | 2.58 |

إذا كانت $\sigma$ مجهولة نستخدم S بدلاً منها.

**مثال:**

**مصنع لإنتاج المصابيح الكهربائية، اختير من إنتاجه عينة حجمها (100) مصباح لتقييم جودة الإنتاج. إذا كان الوسط الحسابي لعمر المصباح في العينة المختارة (1200) ساعة وانحرافه المعياري (250) ساعة. قدر بدرجة ثقة 95% متوسط عمر المصباح من إنتاج المصنع كله.**

**الحل:**

حيث أن الانحراف المعياري للمجتمع $\sigma$ مجهول نقدره باستخدام الانحراف المعياري للعينة $S$.

$$n=100,\quad \bar{x}=1200,\quad S=250,\quad 1-\alpha=0.95 \Rightarrow Z_{\alpha/2}=1.96$$

$$\bar{x}-Z_{\alpha/2}\frac{S}{\sqrt{n}} \le \mu \le \bar{x}+Z_{\alpha/2}\frac{S}{\sqrt{n}}$$

$$1200-(1.96)\frac{250}{\sqrt{100}} \le \mu \le 1200+(1.96)\frac{250}{\sqrt{100}}$$

$$1200-49 \le \mu \le 1200+49$$

$$1151 \le \mu \le 1249$$

∴ متوسط عمر المصباح من إنتاج المصنع كله يتراوح بين 1151 و 1249 ساعة بدرجة ثقة 95%. ويدل هذا التقدير أن جودة المصابيح عالية من هذا المصنع.

# اختبار الفروض حول متوسط المجتمع:

هو عبارة عن إجراء إحصائي يعمل على التحقق من صحة أي فرض (إدعاء) حول قيمة متوسط المجتمع μ.

*<u>تتمحور خطوات إجراء أي اختبار للفروض الإحصائية بشكل عام كما يلي:</u>*

- صياغة فرضان يسميان فرض العدم $H_0$ والفرض البديل $H_1$ حول معلمة (أو خاصية) في مجتمع الدراسة.
- حساب بعض الإحصاءات كالمتوسط، والانحراف المعياري...الخ.
- نحسب من قيم الإحصاءات إحصاء الاختبار.
- نتخذ القرار برفض أو عدم رفض فرض العدم.

- فرض العدم (Null Hypothesis) هو ادعاء (أو عبارة) عن معلمة مجتمع يفترض صحته حتى يثبت عكس ذلك.
- الفرض البديل (Alternative Hypothesis) هو ادعاء عن معلمة مجتمع سوف يكون صحيحًا إذا كان فرض العدم غير صحيح.
- إحصاء الاختبار (Test Statistics) هو أسلوب أو طريقة لتحديد قاعدة متى ترفض فرض العدم.

## اختبار الفرض حول متوسط المجتمع (μ) من جانبين في حالة العينات الكبيرة ($n \geq 30$):

1- نضع فرض العدم ($H_0$) والفرض البديل ($H_1$) كما يلي:

$$H_0: \mu = \mu_0 \quad (vs) \quad H_1: \mu \neq \mu_0$$

2- نوجد إحصاء الاختبار ( $Z$ ) كما يلي:

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

فنحصل على ( $Z$ ) المحسوبة

" إذا كانت **σ** مجهولة نستخدم **S** بدلاً منها ".

3- إذا وقعت ( $Z$ ) المحسوبة في منطقة القبول ( $1 - \alpha$ )، ∴ نقبل $H_0$ ونرفض $H_1$.

والعكس إذا وقعت ( $Z$ ) المحسوبة في منطقة الرفض، ∴ نرفض $H_0$ ونقبل $H_1$.

حيث:

إذا كانت $-Z_{\alpha/2} \leq Z \leq Z_{\alpha/2}$ نقبل $H_0$ ونرفض $H_1$، وإذا لم تكن محصورة بين $Z_{\alpha/2}$ و $-Z_{\alpha/2}$ نرفض $H_0$ ونقبل $H_1$.

- قيمة $Z_{\alpha/2}$ هي نفسها المستخدمة في التقدير.

**مثال:**

**إذا كان متوسط الزيادة في أجور العاملين في إحدى المؤسسات عام 1998 هو (36) عشرات الريالات، وفي عام 2001م أخذت عينة من (64) فرداً من العاملين في هذه المؤسسة، فوجد أن الوسط الحسابي للزيادة في أجورهم (40) عشرات الريالات والانحراف المعياري (8) عشرات الريالات. هل يدل ذلك على أن متوسط الزيادة في أجور العاملين في المؤسسة عام 2001م قد اختلف عن متوسط الزيادة في الأجور عام 1998م؟ وذلك عند مستوى معنوية $\alpha = 0.05$.**

**الحل:**

فرض العدم والفرض البديل:

$$H_0 : \mu = 36 \quad (\text{vs}) \quad H_1 : \mu \neq 36$$

معطيات:

$$n = 64, \quad \bar{x} = 40, \quad S = 8, \quad \alpha = 0.05 \Leftrightarrow Z_{\alpha/2} = 1.96$$

إحصاء الاختبار $(Z)$:

$$Z = \frac{\bar{x} - \mu_0}{S/\sqrt{n}} = \frac{40 - 36}{8/\sqrt{64}} = 4$$

القرار:

**نرفض** فرض العدم $H_0$ ونقبل الفرض البديل $H_1$؛ أي أن متوسط الزيادة في الأجور عام 2001م قد اختلف عن متوسط الزيادة في الأجور عام 1998م. وذلك عند مستوى معنوية $\alpha = 0.05$.

استخدام قيمة ( P ) لإختبار الفروض الاحصائية:

- إذا كانت قيمة P ( P- value ) أقل من مستوى المعنوية ( $\alpha$ ) ∴ نرفض $H_0$ ونقبل $H_1$.
- إذا كانت قيمة P ( P- value ) أكبر من مستوى المعنوية ( $\alpha$ ) ∴ نقبل $H_0$ و نرفض $H_1$.

**مثال:**

**أدعى أحد الباحثين في دراسة له حول متوسط الاستهلاك لأحد المواد الغذائية الأساسية أنه يساوي (1000) كيلوجرام في إحدى الأحياء بمدينة جدة. وبعد جمع وإدخال البيانات في الحاسب الآلي ومعالجتها إحصائياً وباستخدام برنامج Excel، حصل على النتيجة التالية:**

| | |
|---|---|
| **P-Value =** | **0.60** |

**ما هو القرار حول متوسط الاستهلاك، هل يختلف عن (1000) كيلوجرام؟ اختبر هذا الفرض عند مستوى معنوية 0.05.**

**الحل:**

من المعطيات نجد أن فرض العدم والفرض البديل هي على النحو الآتي:

$$H_0 : \mu = 1000 \quad \text{(vs)} \quad H_1 : \mu \neq 1000$$

حيث أن قيمة P-Value $< \alpha$ فإننا لا نرفض فرض العدم، أي أن متوسط الاستهلاك للمادة الغذائية الأساسية يساوي (1000) كيلوجرام في إحدى الأحياء بمدينة جدة، وذلك عند مستوى معنوية 0.05.

**مثال:**

**أدعى أحد الباحثين أن متوسط سرعة القراءة لدى طلاب إحدى الجامعات تساوي (200) كلمة في الدقيقة. وبعد جمع وإدخال البيانات في الحاسب الآلي ومعالجتها إحصائياً وباستخدام برنامج Excel، حصل على النتيجة التالية:**

| | |
|---|---|
| P-Value = | 0.0001 |

**هل متوسط سرعة القراءة لدى هؤلاء الطلاب يختلف عن (200) كلمة في الدقيقة، عند مستوى معنوية 0.10.**

**الحل:**

من المعطيات نجد أن فرض العدم والفرض البديل هي على النحو الآتي:

$$H_0 : \mu = 200 \quad (\text{vs}) \quad H_1 : \mu \neq 200$$

حيث أن قيمة $\text{P-Value} < \alpha$ فإننا نرفض فرض العدم ونقبل الفرض البديل، أي أن متوسط سرعة القراءة لا يساوي (200) كلمة في الدقيقة لدى هؤلاء الطلاب، وذلك عند مستوى معنوية 0.10.

# اختبار مربع كاي:

يستخدم هذا الاختبار في الحالات الآتية :جودة التوفيق، الاستقلال، والتجانس. وسنكتفي بدراسة هذا الاختبار في حالة الاستقلال. وفكرة تطبيق اختبار مربع كاي للاستقلال هي أنه في كثير من المسائل التطبيقية نحتاج إلى التعرف عما إذا كان هناك علاقة بين صفتين من صفات مجتمع ما أم لا .**فمثلاً** قد نحتاج إلى معرفة هل هناك علاقة بين مستوى الدخل ومستوى التعليم؟ أو هل توجد علاقة بين التدخين والإصابة بسرطان الرئة؟ وهكذا... وللإجابة على هذا السؤال يتم تصنيف بيانات كل صفة من الصفتين بحيث يكون مقياس البيانات اسمي أو ترتيبي وبعد ذلك يتم عد التكرارات في كل فئة (خلية) من هذا التصنيف ثم نستخدم اختبار مربع كاي للاستقلال الذي يبحث هل هناك علاقة بين الصفتين أم لا دون التعرض لمستوى هذه العلاقة من كونها قوية أو متوسطة أو ضعيفة.

## خطوات الاختبار:

أ- نضع فرض العدم $H_0$: لا توجد علاقة بين الصفتين.
والفرض البديل $H_1$: توجد علاقة بين الصفتين.

ب- نكون جول التوافق، وهو يحتوي على التكرارات المشاهدة ( $o_i$ ) ( ويعطى في المسألة ).

جـ- نحسب التكرار المتوقع ( $E_i$ ) لكل خلية ( لكل تكرار مشاهد ) من العلاقة:

$$E_i = \frac{\text{مجموع الصف الذي به الخلية} \times \text{مجموع العمود الذي به الخلية}}{\text{مجموع التكرارات}\left(\text{حجم العينة}\right)}$$

د- نوجد $\chi^2$ المحسوبة ( الفعلية ) من العلاقة:

$$\chi^2 = \sum \frac{(o_i - E_i)^2}{E_i}$$

هـ- نوجد $\chi^2$ النظرية ( الجدولية ) بدرجات حرية من العلاقة:

$$\chi^2_{[(r-1)(c-1),\alpha]}$$

حيث:

| | |
|---|---|
| **r** | **عدد الصفوف** |
| **c** | عدد الأعمدة |
| **α** | مستوى المعنوية |

و- إذا وقعت ($\chi^2$) المحسوبة في منطقة القبول، ∴ نقبل $H_0$ ونرفض $H_1$، أي لا توجد علاقة بين الصفتين. والعكس إذا وقعت ($\chi^2$) المحسوبة في منطقة الرفض، ∴ نرفض $H_0$ ونقبل $H_1$، أي توجد علاقة بين الصفتين.

**مثال:**

**أدعى أحد الباحثين الاجتماعين أن عدد ساعات النوم له علاقة بمستوى أداء الطلاب بالمرحلة الثانوية في الاختبار النهائي، وبسؤال عينة من الطلاب بمدارس مختلفة عن عدد ساعات النوم خلال أيام الاختبارات النهائية تم تقسيمهم حسب إجاباتهم إلى فئتين: فئة عدد ساعات النوم مثالي والأخرى غير مثالي وتم تصنيفهم حسب مستوى الأداء في الاختبارات فحصلنا على النتائج التالية:**

| مستوى الأداء | عدد ساعات النوم | | $\Sigma$ |
|---|---|---|---|
| | غير مثالي | مثالي | |
| ضعيف | 60 | 700 | 760 |
| متوسط | 40 | 430 | 470 |
| عالي | 70 | 200 | 270 |
| $\Sigma$ | 170 | 1330 | 1500 |

**1. فرض العدم والفرض البديل هو:**

$H_0$: لا توجد علاقة بين عدد ساعات النوم ومستوى أداء الطلاب بالمرحلة الثانوية في الاختبار النهائي.

$H_1$: توجد علاقة بين عدد ساعات النوم ومستوى أداء الطلاب بالمرحلة الثانوية في الاختبار النهائي

2. التكرار المتوقع لعدد الطلاب الذين عدد ساعات نومهم مثالي ومستوى أدائهم عالي يساوي

$$E_i = \frac{\text{مجموع الصف الذي به الخلية} \times \text{مجموع العمود الذي به الخلية}}{\text{مجموع التكرارات (حجم العينة)}}$$

$$E_{32} = \frac{1330 \times 270}{1500} = 239.4$$

3. قيمة $\chi^2$ النظرية (الجدولية) عند مستوى المعنوية $\alpha = 0.01$

$$\chi^2{}_{[(3-1)(2-1),0.01]} = \chi^2{}_{[(2)(1),0.01]} = 9.210$$

**4. إذا كانت قيمة إحصاء الاختبار تساوي $\chi^2 = 69.88$ ، فإن القرار هو**

$$\chi^2_{\text{المحسوبة}} > \chi^2_{\text{الجدولية}}$$

∴ نرفض فرض العدم ونقبل الفرض البديل

<u>*أي توجد علاقة.*</u>