التقدير الإحصائي واختبارات الفروض

الوسط الحسابي والانحراف المعياري

- ***الوسط الحسابي*** *Arithmetic Mean*

هو أحد مقاييس النزعة المركزية ، ويحسب بقسمة مجموع القيم على عددها ، وتطبق المعادلة التالية .

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

$$= \frac{\sum_{i=1}^{n} x_i}{n}$$

مثـــال (١)

تم زراعة 5 قطع زراعية مختارة عشوائيا بمحصول الذرة ، وكان إنتاج الطن/ هكتار على النحو التالي :

4.7 3.27 4.50 3.23 5.3 **والمطلوب حساب متوسط الإنتاج .**

الحـــل

بتطبيق المعادلة أعلاه ، نجد أن :

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{5.3 + 3.23 + 4.5 + 3.27 + 4.7}{5}$$

$$= \frac{21}{5} = 4.2 \ \ T/H$$

- ***التباين و الانحراف المعياري*** *Variance & Standard Deviation*

١ـ *التباين:* هو أحد مقاييس التشتت ، ويعبر عنه بمتوسط مربعات انحرافات القيم عن

وسطها الحسابي ، ومن ثم يحسب بالمعادلة التالية :

أولا: تباين المجتمع (σ^2)

Population Variance $$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - u)^2}{N} = \frac{\sum_{i=1}^{N} x_i^2}{N} - \mu^2$$

حيث أن μ هي الوسط الحسابي للمجتمع ، N هي حجم المجتمع .

ثانيا: تباين العينة (S^2)

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

حيث أن $\bar{x}$ هو الوسط الحسابي للعينة ، n هي حجم العينة .

وغالبا يكون تباين المجتمع (σ^2) غير معلوم ، ولذا يستخدم تباين العينة (S^2) أعلاه كتقدير غير متحيز لتباين المجتمع (σ^2) .

- ***الانحراف المعياري :*** هو أحد مقاييس التشتت ، ويعبر عنه بالجذر التربيعي الموجب للتباين ، أي أن :

Standard Deviation for sample $$S = \sqrt{\frac{\left(\sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}\right)}{(n-1)}}$$

مثـــال (٢) :

بالتطبيق على بيانات مثال (١)، نجد أن الانحراف المعياري للعينة هو :

x	x^2
5.3	28.09
3.23	10.433
4.5	20.25
3.27	10.693
4.7	22.09
21	91.5558
$\sum x$	$\sum x^2$

$$S=\sqrt{\left(\sum_{i=1}^{n}x_i^2-\frac{(\sum_{i=1}^{n}x_i)^2}{n}\right)\Big/(n-1)}$$

$$=\sqrt{\frac{91.5558-\frac{(21)^2}{5}}{4}}=\sqrt{\frac{3.3558}{4}}=0.915942$$

$$S=\sqrt{\left(\sum_{i=1}^{n}x_i^2-\frac{(\sum_{i=1}^{n}x_i)^2}{n}\right)\Big/(n-1)}$$

$$=\sqrt{\frac{91.5558-\frac{(21)^2}{5}}{4}}=\sqrt{\frac{3.3558}{4}}=0.915942$$

يقصد بتوزيع معاينة الوسط الحسابي للعينة، دراسة خصائص التوزيع الاحتمالي للوسط الحسابي للعينة:

مثـــال

مزرعة (مجتمع) بها عدد 5 عجول رضيعة، أوزانها بالكيلوجرام كالتالي:

55 68 59 64 69

- **متوسط المجتمع:** $\mu=\sum x/N=315/5=63$
- **تباين المجتمع :** $\sigma^2=\frac{\sum_{i=1}^{N}x_i^2}{N}-\mu^2=\frac{19987}{5}-(63)^2=28.4$
- **الانحراف المعياري للمجتمع:** $\sigma=\sqrt{\sigma^2}=\sqrt{28.4}=5.329$

إذا تم سحب عينات مع الإرجاع أحجامها $(n=3)$، نجد أن عدد العينات الممكن سحبها كالتالي:

i	samples			Mean $(\bar{x})$
1	55	68	59	60.667
2	55	68	64	62.333
3	55	68	69	64.000
4	55	59	64	59.333
5	55	59	69	61.000
6	55	64	69	62.667
7	68	59	64	63.667
8	68	59	69	65.333
9	68	64	69	67.000
10	59	64	69	64.000

والتوزيع الاحتمالي للوسط الحسابي للعينة هو:

$(\bar{x})$	$f(\bar{x})=p(x)$	$\bar{x}\ p(\bar{x})$	$(\bar{x}^2)$	$\bar{x}^2\ p(\bar{x})$
59.333	0.1	5.9333	3520.404889	352.0404889
60.667	0.1	6.0667	3680.484889	368.0484889
61.000	0.1	6.1000	3721.000000	372.1000000
62.333	0.1	6.2333	3885.402889	388.5402889
62.667	0.1	6.2667	3927.152889	392.7152889
63.667	0.1	6.3667	4053.486889	405.3486889
64.000	0.2	12.8000	4096.000000	819.2000000
65.333	0.1	6.5333	4268.400889	426.8400889
67.000	0.1	6.7000	4489.000000	448.9000000
		63.0000		3973.7333334

- **خصـــائص هذا التوزيع:**

***القيمة المتوقعة للوسط الحسابي للعينة:** $E(\bar{x})=\mu_{\bar{x}}=\sum \bar{x}\ p(\bar{x})=63$

أي أن القيمة المتوقعة للوسط الحسابي للعينة يساوي متوسط المجتمع $\mu=63$.

*** تباين الوسط الحسابي للعينة:**

$$\sigma_{\bar{x}}^2=E(\bar{x}^2)-\mu_{\bar{x}}^2=\sum \bar{x}^2\ p(\bar{x})-(63)^2$$
$$=3973.7333334-3969=4.73333$$

أي أن تباين الوسط الحسابي للعينة يساوي:

$$\frac{N-n}{N-1}\times\frac{\sigma^2}{n}=\frac{5-3}{5-1}\times\frac{28.4}{3}=4.73333$$

ونستنتج من ذلك ما يلي:

١- أن إذا كان لدينا مجتمع حجمه N (محدود)، وسحبت منه عينات مع الإرجاع أحجامها n، فإن الوسط الحسابي للعينة $(\bar{x})$ يتبع توزيع احتمالي له الخصائص التالية:

$$E(\bar{x}) = \mu_{\bar{x}} = u$$

$$\sigma_{\bar{x}}^2 = \frac{N-n}{N-1} \times \frac{\sigma^2}{n}$$

٢- إذا كان حجم المجتمع كبير، أو غير محدود نجد أن خصائص توزيع المعاينة للوسط الحسابي للعينة $\bar{x}$ هو:

- **القيمة المتوقعة للوسط الحسابي للعينة:**

$$E(\bar{x}) = \mu_{\bar{x}} = u$$

- **التباين للوسط الحسابي للعينة هو :**

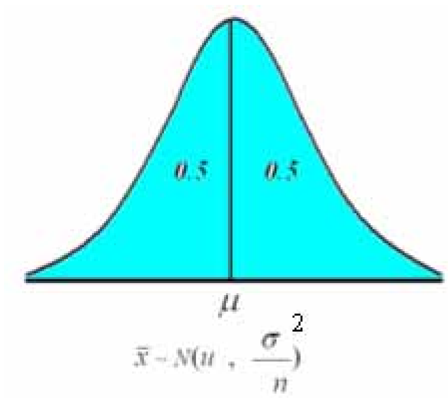
$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$$

وحيث أننا نتعامل مع مجتمعات نظرية ذات توزيعات خاصة مثل التوزيع الطبيعي، وتوزيع ت، وغير ذلك من التوزيعات ، فإن هناك أيضا توزيعات خاصة للوسط الحسابي للعينة وذلك على النحو التالي:

- ***أولا: العينة مسحوبة من مجتمع له توزيع طبيعي ، والتباين σ^2 معلوم.***

إذا كانت العينة مسحوبة من مجتمع له توزيع طبيعي متوسطه u ، وتباينه σ^2 (معلوم)، فإن الوسط الحسابي للعينة له أيضا توزيع طبيعي متوسطه u ، وتباينه $\sigma_{\bar{x}}^2 = \sigma^2/n$ ، أي أن :

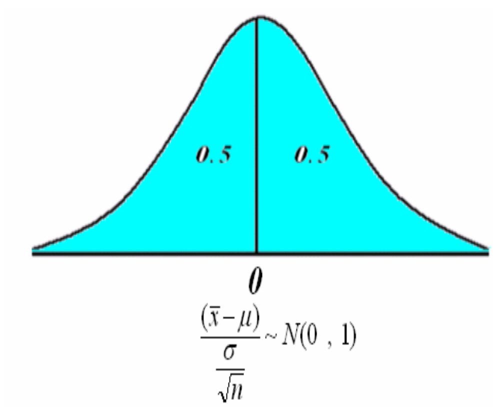
$$\bar{x} \sim N\left(u \ , \ \frac{\sigma^2}{n}\right)$$



ومن ثم يصبح للمقدار:

$$z = \frac{(\bar{x} - \mu)}{\frac{\sigma}{\sqrt{n}}}$$

توزيع طبيعي معياري، متوسطه صفرا، وتباينه واحد، أي أن:

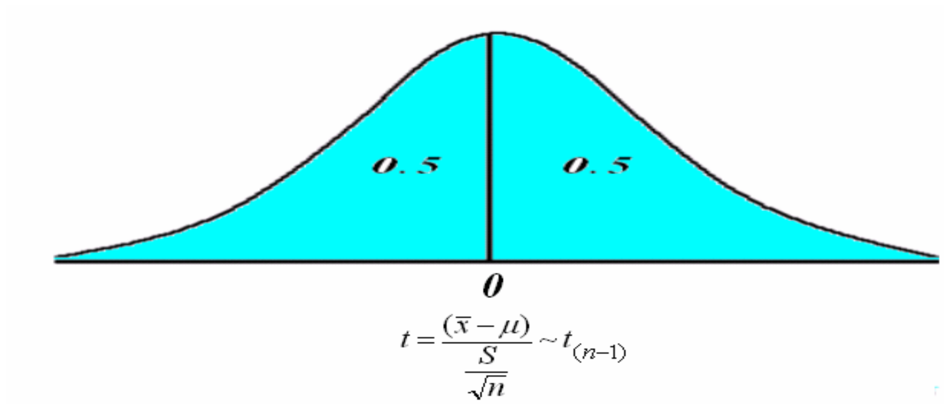


- ***ثانيا: العينة مسحوبة من مجتمع له توزيع طبيعي، والتباين σ^2 غير معلومة***

إذا كانت العينة مسحوبة من مجتمع له توزيع طبيعي ، متوسطه u ، وتباينه σ^2 (غير معلوم) ، فإن توزيع معاينة المقدار

$$t = \frac{(\bar{x} - \mu)}{\frac{S}{\sqrt{n}}} \sim t_{(n-1)}$$

يتبع التوزيع t بدرجات حرية $(n-1)$ ، أى أن



حيث أن S^2 هو تباين العينة، ويستخدم كتقدير غير متحيز لتباين المجتمع .

ثالثا: إذا كانت العينة مسحوبة من مجتمع له توزيع طبيعي، وتباين المجتمع σ^2 غير معلوم ، وكان حجم العينة كبير **$(n>30)$ ، فإن:**

ـ التوزيع الحقيقي للمقدار:

$$t=\frac{(\overline{x}-\mu)}{\frac{S}{\sqrt{n}}}\sim t_{(n-1)}$$

هو التوزيع t بدرجات حرية $(n-1)$.

ـ التوزيع التقريبي هو التوزيع الطبيعي المعياري.

$$\lim_{n\to\infty} t_{(n-1)}\sim z=\frac{(\overline{x}-\mu)}{\frac{S}{\sqrt{n}}}$$

هناك بعض المفاهيم التي يجب التعرف عليها، ومنها ما يلي:

- ***المعلمة*** *Parameter*: **هي قيم حقيقية لمؤشرات المجتمع تحسب باستخدام بيانات المجتمع ككل.**
- ***الإحصاء*** *Statistics*: **هي مؤشرات العينة ، وتحسب باستخدام بيانات العينة ، وتستخدم كتقدير لقيمة المعلمة في المجتمع .**
- ***المقدر*** *Estimator* **:** **هو إحصاء أو معادلة تستخدم للحصول على تقدير لمعلمة المجتمع.**

- ***التقدير*** *Estimate* **: هو القيمة العددية للمقدر.**
- ***التقدير لنقطة*** *Point Estimate*: **هو حساب قيمة باستخدام بيانات العينة كتقدير لمعلمة المجتمع، أي أن التقدير بنقطة هو قيمة تقدير واحدة.**
- ***التقدير لمدي*** *Interval Estimate* **: هو تقدير مدي يقع داخله معلمة المجتمع ، وهذا المدي يتراوح بين حدين أدنى وأعلي يسميان بحدي الثقة.**
- ***معايير اختيار المقدر الجيد:***

١- **عدم التحيز** Unbiased ness

٢- **الاتســـاق** Consistency.

٣- **الكفاءة النسبية** Relative Efficiency.

٤- **الكفاية** Sufficiency.

تقدير فترة ثقة لمتوسط المجتمع μ

يقصد بتقدير فترة ثقة للمتوسط في المجتمع (μ) هو إيجاد الحدين الأدنى، والأعلى اللذان يقع بينهما قيمة متوسط المجتمع(μ) باحتمال قدرة $(1-\alpha)$، ويسمى المقدار $(1-\alpha)\%$ مستوى الثقة، ويطلق على القيمة α بمستوى المعنوية وسوف يأتي تعريفه فيما بعد.

ولذا يحسب حدي الثقة لمتوسط المجتمع (μ) عند مستوى ثقة $(1-\alpha)\%$، باستخدام الصيغ التالية:

١- **إذا كان الانحراف المعياري في المجتمع معروفا.**

$$\bar{x}-\sigma_{\bar{x}}\, z_{(1-\alpha/2)} < \mu < \bar{x}+\sigma_{\bar{x}}\, z_{(1-\alpha/2)}$$

حيث أن: $\sigma_{\bar{x}} = \sigma/\sqrt{n}$

٢- **إذا كان الانحراف المعياري في المجتمع غير معروف،** $n \le 30$

$$\bar{x}-S_{\bar{x}}\, t_{(n-1,\, 1-\alpha/2)} < \mu < \bar{x}+S_{\bar{x}}\, t_{(n-1,\, 1-\alpha/2)}$$

حيث أن : $S_{\overline{x}} = S/\sqrt{n}$ ، $n \le 30$

٣- إذا كان الانحراف المعياري في المجتمع غير معروف ، $n \succ 30$.

$$\overline{x} - S_{\overline{x}}\ z_{(1-\alpha/2)} < \mu < \overline{x} + S_{\overline{x}}\ z_{(1-\alpha/2)}$$

تطبيقـــات

تطبيق (١)

إذا كان وزن الدجاج بالجرام في أحد المزارع بعد *45* يوم يتبع توزيع طبيعي، سحبت عينة عشوائية من أحد المزارع المنتجة للدجاج حجمها *25* دجاجة، ووجد أن متوسط وزن الدجاج في العينة *890* جرام، والانحراف المعياري لها *200* جرام، والمطلوب :

١- قدر فترة ثقة لمتوسط وزن الدجاج في المزرعة باستخدام مستوي ثقة *95%*.

٢- إذا علم من الخبرات السابقة أن تباين وزن الدجاج في المزرعة هو *62500* ، قدر فترة ثقة لمتوسط وزن الدجاج في المزرعة باستخدام مستوي ثقة *99%* .

٣- فسر النتائج السابقة.

الحـل (١)

١- تقدير فترة ثقة لمتوسط وزن الدجاج في المزرعة باستخدام مستوي ثقة *95%* .

بما أن :

ـ الوزن يتبع توزيع طبيعي. ـ التباين σ^2 مجهول.

ـ حجم العينة صغير $n = 25 \le 30$

إذا توزيع المعاينة المستخدم هو توزيع *t* بدرجات حرية *(n-1)*.

حيث أن:

$$S_{\overline{x}} = \frac{S}{\sqrt{n}} = \frac{200}{\sqrt{25}} = \frac{200}{5} = 40 \quad ، \quad \overline{x} = 890$$

$$df = (n-1) = 25 - 1 = 24 \quad ، \quad 1 - \alpha/2 = 0.975\ ,\ \alpha/2 = 0.025\ ,\ \alpha = 0.05$$

إذا $t_{(n-1,\ 1-\alpha/2)} = t_{(24,\ 0.975)} = 2.064$

ويكون حدي الثقة هما:

$$\bar{x}-S_{\bar{x}}\, t_{(n-1,\,1-\alpha/2)}<\mu<\bar{x}+S_{\bar{x}}\, t_{(n-1,\,1-\alpha/2)}$$

$$890-40(2.064)<\mu<890+40(2.064)$$
$$807.44<\mu<972.56$$

إذا : $$P(807.44<\mu<972.56)=0.95$$

أي أننا علي ثقة *95%* بأن متوسط وزن الدجاج في المزرعة يتراوح بين *807.44* ، *972.56* جرام.

لو تم سحب *100* عينة عشوائية من هذه المزرعة حجم كل منها *25* دجاجة ، فإنه من المحتمل أن يكون هناك *95%* من فترات الثقة تشمل الفترة من *807* إلى ، *973* جرام.

٢- فترة ثقة لمتوسط وزن الدجاج في المزرعة إذا علم تباين وزن الدجاج في المزرعة باستخدام مستوي ثقة *99%*.

في هذه الحالة التباين معلوم، لذا يكون توزيع المعاينة للوسط الحسابي المستخدم هو التوزيع الطبيعي القياسي، ويكون حدي الثقة هما:

$$\bar{x}-\sigma_{\bar{x}}\, z_{(1-\alpha/2)}<\mu<\bar{x}+\sigma_{\bar{x}}\, z_{(1-\alpha/2)}$$

حيث أن :

$$\sigma_{\bar{x}}=\frac{\sigma}{\sqrt{n}}=\frac{\sqrt{62500}}{\sqrt{25}}=\frac{250}{5}=50 \qquad \bar{x}=0.89$$

$$Z_{(1-\alpha/2)}=Z_{(0.995)}=2.58 \qquad 1-\alpha/2=0.995\ ,\ \alpha/2=0.005\ ,\ \alpha=0.01$$

إذا حدي الثقة هما :

$$890-50(2.58)\prec\mu\prec 890+50(2.58)$$
$$761\prec\mu\prec 1019$$

أي أن : $$P(761\prec\mu\prec 1019)=0.99$$

تطبيق (٢)

تنتج إحدى الشركات الغذائية نوع من العصير المشكل زنة العبوة *125* جرام ، قام مدير مراقبة الإنتاج بسحب عينة عشوائية حجمها *36* عبوة وقياس كمية الكربوهيدرات بالجرام ، ووجد أن متوسط كمية الكربوهيدرات *12* جرام ، والانحراف المعياري *2.4* جرام

، فإذا أراد قسم مراقبة الإنتاج تقدير فترة ثقة *95%* لمتوسط كمية الكربوهيدرات في العبوات إذا كان وزن الكربوهيدرات يتبع التوزيع الطبيعي.

الحل

للوصول إلى فترة الثقة المطلوبة، يتبع الآتي:

- **بما أن كمية الكربوهيدرات تتبع توزيع طبيعي ، تباينه σ^2 غير معلوم**
- **إذا توزيع المعاينة الدقيق هو توزيع *t* بدرجات حرية *{(n-1)=(36-1)=35}* ،**
- **وحيث أن حجم العينة كبير *n=36>30* ، فإنه يمكن تقريب توزيع *t* إلى التوزيع الطبيعي القياسي.**
- **ويكون حدي الثقة هما:**

$$\bar{x} - S_{\bar{x}}\ z_{(1-\alpha/2)} \le \mu \le \bar{x} + S_{\bar{x}}\ z_{(1-\alpha/2)}$$

$$12 - \frac{0.18}{\sqrt{36}}(1.96) \le \mu \le 12 - \frac{0.18}{\sqrt{36}}(1.96)$$

$$12 - 0.0588 \le \mu \le 12 + 0.0588$$

$$11.94 \le \mu \le 12.06$$

إذا أننا نثق بنسبة *95%* بأن كمية الكربوهيدرات يتراوح وزنها ما بين *(11.94* إلى *12.06)* جرام.