الفصل الخامس عشر

التقدير

Estimation

15.1. مقدمة Introduction

التقدير هو استنتاج قيم معلمات المجتمع Population Parameters عن طريق قيم إحصاءات العينة Sample Statistics، فإذا تم التقدير Estimation باستنتاج قيمة وحيدة لمعلمات المجتمع ، فإننا نسمى هذا النوع تقدير النقطة Point Estimate ، أما إذا تم التقدير بمدىً يقع بين نقطتين، فإننا نسمى هذا النوع تقدير الفترة Interval Estimate.

وقبل إجراء عملية التقدير يجب التأكد من أن العينة التي تم سحبها تنتمي إلى المجتمع محل الدراسة أم لا.

15.2. تقدير النقطة Point Estimate

هو تقدير قيمة وحيدة لمعلمات المجتمع Population Parameters أى تقدير بنقطة واحدة، فلو أردنا تقدير متوسط مدة تنفيذ المشروعات التي تقوم شركة ما بتنفيذها، فإحدى هذه الطرق ستكون بأخذ عينات من المشروعات التي تم تنفيذها، ثم نحسب المتوسط الحسابي لهذه المدد، ونعتبرها كتقدير لمتوسط المجتمع Population Mean (μ)، وفى هذه الحالة يجب أن نأخذ بعين الاعتبار أن هذا الاستدلال يشوبه نوع من عدم الدقة إلى حد كبير، ولكننا نلجأ إليه فى الحالات التي لا تتطلب دقة عالية فى الاستنتاج، ولا يترتب عليها قرارات هامة.

15.3. التقدير بفترة Interval Estimation

وفى هذا النوع من التقدير، وحيث إننا نقوم بدراسة جزء من البيانات وليس كل البيانات، فيجب أن يكون لدينا قناعة بأن القيم المحسوبة للإحصاءات Statistics ستختلف قليلاً عن القيم الحقيقية لمعلمات المجتمع Population Parameters، وبعبارة أخرى ليس من المحتمل أن نحصل على قيمة المعلمات تماما، ولذا نلجأ لافتراض مدى أو فترة يكون من الممكن وقوع هذه المعلمات داخلها، أى أننا نعطى قيمتين، و نقول أن قيم المعلمات تقع بين هاتين القيمتين باحتمال معين، وهذه الفترة تسمى فترة الثقة "CI" Confidence Interval.

ولتوضيح هذا المفهوم، نفرض أن إحدى الجرائد اليومية طالعتنا بأن متوسط أعمار المصطافين خلال الصيف المنقضي بلغ 28 عاما للشباب وبلغ 19.5 عاما للفتيات . فهل سألت نفسك كيف تم حساب هذا الرقم ؟ وما هي درجة دقته ؟ وما مدى ثقتنا فى هذه الدقة؟

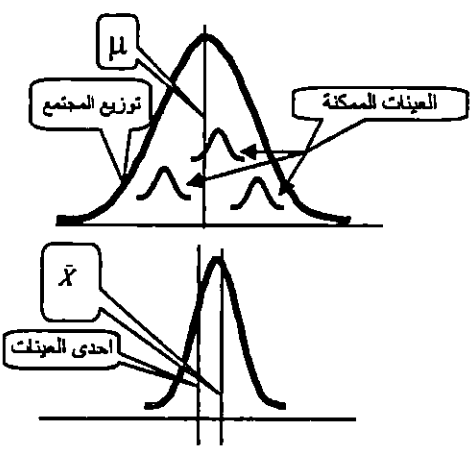
للإجابة على هذه التساؤلات نقول أنه ولصعوبة أن نسأل كل المصطافين عن أعمارهم، فإننا نلجأ إلى أخذ العينات Sampling، ثم نسأل أفراد هذه العينة عن أعمارهم، ثم نحسب المتوسط $\bar{X}$ ، وحيث إن دقة الحسابات تتوقف بدرجة كبيرة على دقة جمع العينات وما يحيط بها من ظروف واحتياطات، منها حجم العينة التي تم انتقاؤها، والفترة الزمنية التي تم أخذ العينات بها، هل هي وقت ذروة أم لا؟ وهل شملت العينة عدد متساو من الجنسين؟ وعوامل أخرى كثيرة تزيد من دقة استنتاجاتنا.

وللتسهيل فإننا نفترض أنه تم أخذ عينات معبرة عن المجتمع، وبالتالي نفترض أن قيم الإحصاءات Statistics التي تم حسابها من العينة تصلح لتعميمها على معلمات المجتمع Population Parameters .

ولكن ماذا لو طالعتنا صحيفة أخرى بنتائج لنفس الموضوع، فنشرت أن متوسط أعمار المصطافين خلال الصيف المنقضي بلغ 28.5 عاما للشباب وبلغ 20 عاما للفتيات؟

في هذه الحالة، هل نقول أن نتائج كلتا الصحيفتين مختلفة، وبالتالي نشك ونرفض هاتين النتيجتين؟ أم أنها متقاربة بدرجة تجعلنا نقبل النتيجتين معاً؟ فنقول أننا بدرجة ثقة كبيرة متأكدين من أن متوسط أعمار الشباب يتروح من 28 حتى 29 عاماً، وأن متوسط أعمار الشابات يتروح من 19 حتى 20 عاماً.

أي للخروج من هذا المأزق فلابد من إيجاد آلية تسمح بوجود تفاوت مقبول في قيم إحصاءات العينة Sample Statistics، في محاولتنا للتنبؤ بقيم معلمات المجتمع Population Parameters، وهذه الآلية تتحقق من خلال مفهوم مدى الثقة أو فترة الثقة Confidence Interval "CI"، وهي آلية أوجدها الإحصائيون للاستدلال بدرجة ثقة أن القيم التي بين أيدينا هي قيم صحيحة ومقبولة أم لا؟



شكل رقم 15- 1 تعدد العينات للمجتمع الواحد واختلاف متوسطها عن متوسط المجتمع

ويرتكز مفهوم فترة الثقة Confidence Interval "CL"، على أنه إذا كان لدينا المجتمع كما في شكل 15-1، فسيكون لها قيم وحيدة لكل من المتوسط (μ) Mean والانحراف المعياري Standard Deviation (σ) وهي قيم غير معلومة وستظل غير معلومة، وبتعدد العينات المأخوذة

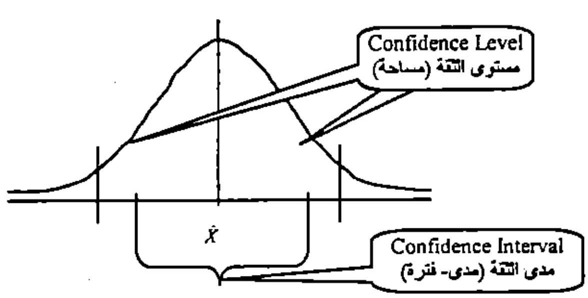
من هذه المجتمعات Populations، فستتعدد قيم المتوسط $\bar{X}$ والانحراف المعيارى (S) لهذه العينات.

وإذا حاولنا استنتاج قيم المعلمات (μ) و(σ) من الـ $\bar{X}$ و(S) الكثيرة والمختلفة، فإننا سنستنتج قيما كثيرة لكل من (μ) و(σ) بالرغم أنها فى واقع الأمر كما قلنا قيما وحيدة،

والحل فى هذه الحالة أن نفترض أن يكون هناك مدى نقبل فيه قيمة متوسط $\bar{X}$ للعينات المختلفة بشرط أن تكون قيمتها قريبة من متوسط المجتمع (μ) وسنطلق على الحدود القصوى و الدنيا لهذا المدى فترة الثقة "CI".

وللتعميم فإن فترة الثقة "CI" هو مدى لقيم إحصاءات العينة Sample Statistics التي يتم حسابها من العينة و التي من المحتمل بدرجة ثقة عالية أن تحتوى على قيم معلمات المجتمع Population Parameters التي نود حسابها واستنتاجها.

أما مستوى الثقة "CL" Confidence Level فهو (قيمة المساحة تحت المنحنى) احتمال أن تكون القيم التي بين أيدينا صحيحة، وهو يتراوح بين 90% و 99% والشائع منه هو 95%، والشكل 15-2 يعرض مخططا عاما لبيان الفرق بين كل من فترة الثقة "CI" ومستوى الثقة "CL" لمتغير ما.



شكل رقم 15- 2 الفرق بين فترة الثقة ومستوى الثقة

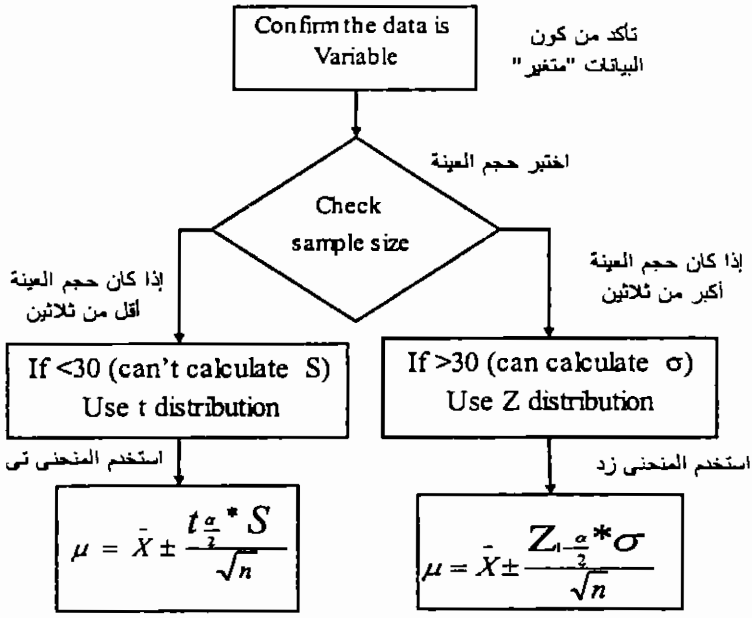
ولمزيد من التوضيح سنتعرض لعدد من معلمات المجتمع الهامة مثل المتوسط والانحراف المعيارى والنسبة، وسندرس فترة الثقة "CI" لهذه المعلمات فى الصفحات القليلة القادمة:

أ- فترة ثقة الوسط الحسابي Confidence Interval For Mean

بالنظر إلى شكل 15-3 ، وكما أسلفنا فإن درجة الدقة عند حساب إحصاءات العينة تعتمد بدرجة عالية على حجم العينة فإذا كان حجم العينة أكبر من 30 عينة (وهذا عدد مناسب إحصائيا لحساب الانحراف المعيارى للمجتمع σ) نستخدم المنحنى Z ، وإذا كان حجم العينة

أقل من 30 عينة فنستخدم المنحنى T (لأن العدد 30 غير مناسب إحصائيا لحساب الانحراف المعيارى للمجتمع σ إذ أن الخطأ يكون كبيرا، ولذا نحسب الانحراف المعيارى للعينة S).

وفى كلتا الحالتين وبمعلومية مستوى الثقة "CL" Confidence Level،وباستخدام الجداول الملحقة فى نهاية هذا الكتاب، يمكننا حساب قيمة كلا من Z و T كل من الجدول الخاص به .



شكل رقم 15- 3 طريقة حساب فترة الثقة للمتوسط بمعلومية حجم العينة

حيث:

$Z_{1-\frac{\alpha}{2}}$ Is The Area Corresponding to the Specified Level of Significance? المساحة المناظرة لمستوى ثقة معين فى المنحنى "زد"

$\bar{X}$ Is the Sample Average متوسط العينة

$t_{\alpha/2}$ Is the Area Corresponding to the Specified Level of Significance? المساحة المناظرة لمستوى ثقة معين فى المنحنى "تى"

σ_x Is the Population Standard Deviation الانحراف المعيارى للمجتمع

S_x Is the Sample Standard Deviation الانحراف المعيارى للعينة

n Is the Sample Size حجم العينة

α Is the Level of Significance مستوى المعنوية

- قيمة "زد" للعينات الكبيرة والانحراف المعيارى غير معروف **Z Values (Large Sample & Unknown Standard Deviation)**

فى هذا النوع من المسائل ومن خلال الأرقام المعطاة وباستخدام برنامج المينيتاب Minitab نقوم بما يلي:

1- نحسب σ من القوائم Minitab>Stat> Basic Statistics> Display Descriptive Statistics، ومن خلال نافذة النتائج نحدد قيمة الانحراف المعيارى.

2. نختبر طبيعية البيانات Normality.

3. نحسب فترة الثقة Confidence Interval من القوائم Minitab>Stat> Basic Statistics> One Sample Z، ثم ندخل قيمة الانحراف المعيارى فتظهر قيمة فترة الثقة.

جدول رقم 15- 1 بيانات المثال 15-1

64.5332	65.5729	64.9075
64.2781	63.8751	65.2609
65.1536	64.0903	63.8409
65.4734	64.2034	64.1978
65.1076	65.1233	65.6074
65.8507	64.2369	64.4147
64.9157	65.0779	64.8279
63.4503	65.2493	64.0263
65.6395	63.0102	64.9753
63.64	64.4785	63.0809

مثال رقم 15- 1 :

لدينا مجموعة الأرقام كما بالجدول 15-1 لعدد 30عينة، والمطلوب حساب فترة ثقة المتوسط μ بدرجة ثقة 95%.

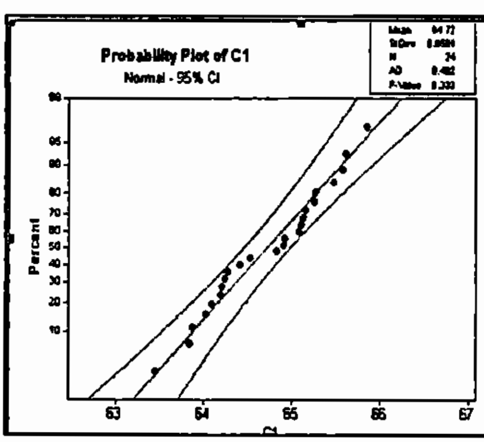
الحل:

1-ندخل الأرقام إلى المينيتاب Minitab

2-نختبر طبيعية البيانات Normality Test من القوائم:-

Minitab>Graph > Probability Plot

فيظهر الشكل 15-4.



شكل رقم 15- 4 نتيجة الخطوة رقم 2

3- Minitab>Stat> Basic Statistics> Display Descriptive Statistics

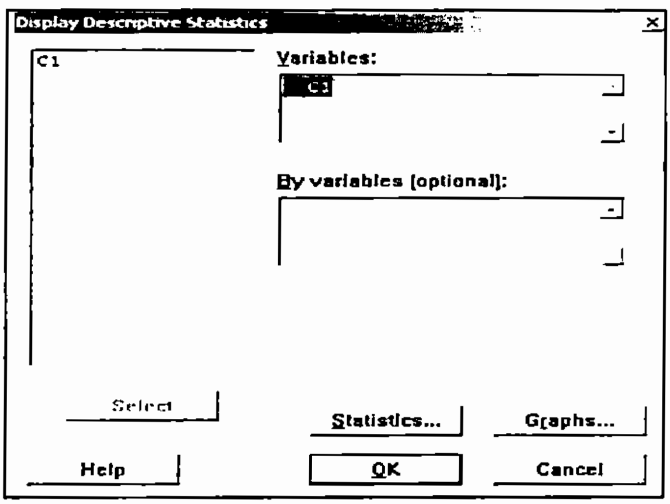
4- من شكل 15-5 فى نافذة النتائج Session Window نحدد قيمة الانحراف المعيارى S

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3
C1	30	0	65.282	0.187	1.025	63.257	64.425	65.206	66.156

شكل رقم 15- 5 نتيجة الخطوة رقم 4

5- Minitab>Basic Statistics>1- Sample Z

6- فيظهر الشكل 15-6



شكل رقم 15- 6 نتيجة الخطوة رقم 5

7- نعوض عن قيمة الانحراف المعيارى σ ، ثم من نافذة النتائج Session Window تظهر فترة الثقة CI

Variable	N	Mean	StDev	SE Mean	95% CI
C1	30	65.2820	1.0253	0.1871	(64.9152; 65.6488)

ومعنى ذلك أننا واثقون بنسبة 95% من أن قيمة المتوسط (μ) Mean تقع بين القيمتين 64.9152 & 65.6488.

- قيمة "تى" للعينات الصغيرة والانحراف المعيارى غير معروف T Values (Small Sample & Unknown Standard Deviation)

فى هذا النوع ونظراً لأن حجم العينة يكون صغيراً أى أقل من 30 عينة، فإن حساب الانحراف المعيارى للمجتمع σ يكون ذا نسبة خطأ عالية، ولذا لا نحسبه وانما نحسب الانحراف المعيارى للعينة S، ونطبق نفس الخطوات كما فى النوع السابق ولكن على المنحنى T.

مثال رقم 15- 2:

لدينا مجموعة الأرقام كما بالجدول 15-2 لعدد 24عينة، والمطلوب حسابالمتوسط μ بدرجة ثقة 95%.

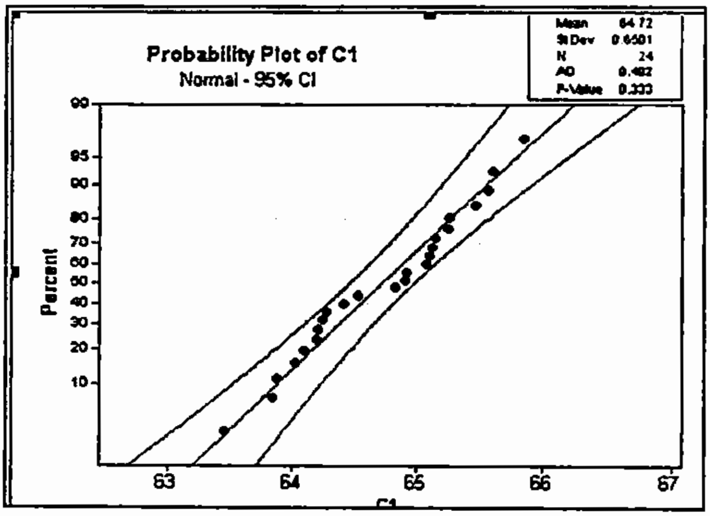
جدول رقم 15- 2 بيانات المثال 15-2

64.5332	65.5729	64.9075
64.2781	63.8751	65.2609
65.1536	64.0903	63.8409
65.4734	64.2034	64.1978
65.1076	65.1233	65.6074
65.8507	64.2359	64.4147
64.9157	65.0779	64.8279
63.4503	65.2493	64.0263

الحل:

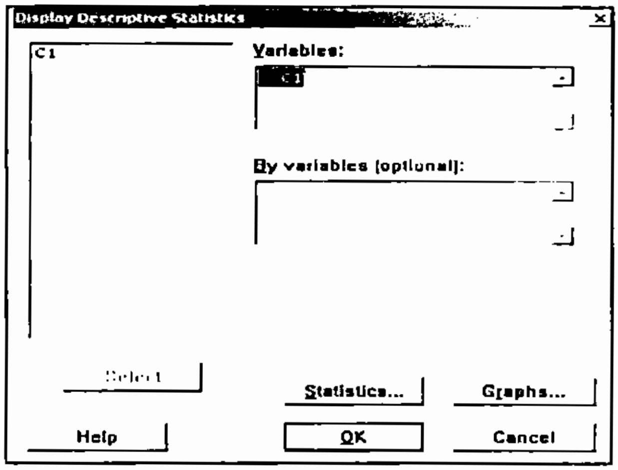
1- ندخل الأرقام إلى المينيتاب Minitab، ونختبر طبيعيةُ البيانات Normalityكما فى شكل 15-7

Minitab>Graph > Probability Plot



شكل رقم 15- 7 نتيجة حل المثال رقم 15-2

2- Minitab>Basic Statistics>1- Sample t



شكل رقم 15- 8 نتيجة الخطوة رقم 5

3- نعوض عن قيمة الانحراف المعيارى للعينة S ، فيظهر شكل 15-8.

4- من نافذة النتائج Session Window يظهر فترة الثقة CI.

One-Sample T: C1

Variable	N	Mean	StDev	SE Mean	95% CI
C1	24	64.7198	0.6501	0.1327	(64.4452; 64.9943)

ومعنى ذلك أننا بدرجة ثقة 95% متأكدون من أن المتوسط (μ) يقع فى المدى من 64.4452 إلى 64.9943.

مثال رقم 15- 3:

صنعت إحدى الشركات المتخصصة سبيكة لاستعمالها فى أحد أنواع الصواريخ، وسجلت قوة السبيكة على 20 قطعة، فوجدت أن الوسط الحسابي 37.8، والانحراف المعياري 2.8، والمطلوب معرفة معدل قوة السبيكة بفترة ثقة 90%، وهل تحتوى هذه الفترة على المتوسط المصرح به أم لا؟

الحل

1. بفرض أن قوة السبيكة تخضع لتوزيع طبيعى Normal

$$\mu = \bar{X} \pm \frac{t_{\frac{\alpha}{2}} S_X}{\sqrt{n}}$$

من جدول T وبدرجة حرية =19 فإن $t_{\frac{\alpha}{2}} = t_{\frac{0.1}{2}} = t_{0.05} = 1.729$

بالتعويض $\mu = 37.8 \pm \frac{1.729*2.8}{\sqrt{20}} = 37.8 \pm 1.08$

ومعنى ذلك أن فترة الثقة Confidence Interval هي من 38.88 إلى 36.72.

2. وحيث إن المعدل μ قيمة مجهولة ، فإننا لا نستطيع أن نجزم بأن الفترة التي حسبناها تحوى μ أم لا ، ولكن إذا كررنا العملية السابقة في الخطوة 1 لعدد كبير من العينات، لأصبح عندنا ثقة أن 90% من هذه الفترات تحوى μ وفقط 10% منها لا تحوى μ.
(محمد صبحي 2001)

ب- فترة ثقة التباين Confidence Interval For Variance

$$\frac{S^2(n-1)}{X^2_{1-\alpha/2,n-1}} \le (\sigma^2) \le \frac{S^2(n-1)}{X^2_{\alpha/2,n-1}}$$

بنفس المفهوم الذى تم تطبيقه سابقا يمكن إجمال خطوات الحل فيما يلي:

1. ندخل الأرقام إلى الميني‌تاب Minitab
2. نختبر طبيعية البيانات Normality من القوائم :-

Minitab>Graph > Probability Plot

3. نحسب S من القوائم :-

Minitab>Stat> Basic Statistics> Display Descriptive Statistics

4. نعوض في المعادلة للحصول على فترة ثقة الانحراف المعياري Confidence Interval For Standard Deviation

مثال رقم 15- 4:

عينة عشوائية حجمها 20 وكان التوزيع مجتمع طبيعي Normal Population وكانت $S^2 = 15$ أوجد فترة ثقة 95% للتباين σ^2.

الحل :

حيث إن $1-\alpha = 0.95$ إذن $\frac{\alpha}{2} = 0.025$ وأيضا $1-\frac{\sigma}{2} = 0.975$

وكذلك درجة الحرية Degree of Freedom = 20-1=19

بالتعويض فى المعادلة $\frac{S^2(n-1)}{X^2_{1-\alpha/2,n-1}} \le (\sigma^2) \le \frac{S^2(n-1)}{X^2_{\alpha/2,n-1}}$

$$\sqrt{\frac{15*19}{8.907}} \le (\sigma^2) \le \sqrt{\frac{15*19}{32.852}}$$

$$\sqrt{31.997} \le (\sigma^2) \le \sqrt{8.675}$$

ج- فترة ثقة النسبة بين تباينين Confidence Interval For The Ratio of Two Variances

فى هذا النوع نستخدم جداول توزيع F (جداول F مرفقة بملاحق الكتاب) وبشرط أن تكون العينـات مسـحوبة مـن مجتمعـين طبيعيـين مختلفـين Two Different Normal Populations

$$\frac{S_2^2}{S_1^2} * F\left[\frac{\alpha}{2}; n_1-1, n_2-1\right], \frac{S_2^2}{S_1^2} * F\left[1-\frac{\alpha}{2}; n_1-1, n_2-1\right]$$

مثال رقم 15- 5:

أخذت عينة عشوائية حجمها 9 من مجتمع يخضع للتوزيع الطبيعىNormal Population وكانت قيمة التباين هى 65.4، وأخذت عينة عشوائية أخرى حجمها 12 من مجتمع طبيعى Normal Population أخر مستقل عن الأول، وكانت قيمة التباين 127.3 والمطلوب ايجاد فترة 90% ثقة للنسبة $\frac{\sigma_2^2}{\sigma_1^2}$.

الحل: حيث إن شروط تطبيق المعادلة متحققة وهى وجود مجتمعين مستقلين Two Independent Populations،وحيث إن $1-\alpha = 0.9$ إذن $\frac{\alpha}{2} = 0.05$

ومن جداول توزيع F نجد:

$$F[0.05;8,11] = 2.95$$

بالتعويض $\left[\frac{127.3}{65.4} * 0.3, \frac{127.3}{65.4} * 2.95\right]$

أى أن النسبة المطلوبة هى $[0.588, 5.742]$ (محمد صبحى،2001)

د- فترة ثقة للنسبة Confidence Interval For Proportion

$$\bar{P}-Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{P}(1-\bar{P})}{n}}<P<\bar{P}+Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{P}(1-\bar{P})}{n}}$$

مثال رقم 15- 6:

احسب فترة ثقة 95% لنسبة عدد الخدمات الغير سليمة، أخذت عينة عشوائية حجمها 400 خدمة فوجد أن عدد الخدمات الغير سليمة هي 100 خدمة، والمطلوب حساب فترة الثقة؟

الحل: حيث أن نسبة الخدمات الغير سليمة هي $\bar{P}=\frac{100}{400}=\frac{1}{4}$ وحيث إن حجم العينة كبير فيمكن استخدام المعادلة

$$\bar{P}-Z_{0975}\sqrt{\frac{\frac{1}{4}*\frac{3}{4}}{400}}<P<\bar{P}+Z_{0.975}\sqrt{\frac{\frac{1}{4}*\frac{3}{4}}{400}}$$

$$\frac{1}{4}-1.96\sqrt{\frac{\frac{1}{4}*\frac{3}{4}}{400}}<P<\frac{1}{4}+1.96\sqrt{\frac{\frac{1}{4}*\frac{3}{4}}{400}}$$

$$0.21<P<0.29$$

(محمد صبحى،2001)