Statistics, lecture 22 : 28 يوتيوب +15 جاكم

تذكّر: لما يحطلك C.I ويطلب $\bar{X}$ ، تجمع الكدين و تقسم كل 2

والفرق بين الكدّين هو ال Error

e.g) A random sample of 400 people with a professional degree taken showed that their mean monthly salary is 450 JD with a standard deviation of 100 JD.

A) Give a 90% C.I. for the mean monthly salary.

sol) $n = 400$, $\bar{X} = 450$, $S = 100$

$C = 0.90$

$Z_c = 1.645 \Rightarrow$ من الجدول

$$E = Z_c \cdot \frac{S}{\sqrt{n}} \simeq 8.23$$

$(450 - 8.23, 450 + 8.23)$

$(441.77, 458.23)$

e.g) It was believed in the Arab world that 50% of persons are smoking. During the year 2000, a sample of 1000 persons showed that the no. of smokers is 620. Establish 95% C.I for the proportion of smokers.

1

sol) $P = 50\%$.

$$P^\wedge = \frac{X}{n} = 0.62.$$

$q^\wedge = 0.38$

$Z_c = 1.96$ "من الجدول"

$$E = Z_c \cdot \sqrt{\frac{P^\wedge q^\wedge}{n}} = 0.03$$

$(0.62 - 0.03, 0.62 + 0.03)$

$(0.59, 0.65)$

e.g) Assume that it's required to estimate the proportion of patients suffering a bad reaction from taking a certain medication P by 95% CI. Determine the sample size needed if the error of estimation is about 0.10 in the following cases:

a) No prior information about $P^\wedge$.

b) Previous study showed that $P^\wedge$ is approximately 0.2

sol) $C = 0.95$ $Z_c = 1.96$ "من الجدول"

a) $P^\wedge = 0.5$

$$n = \left(\frac{Z_c}{E}\right)^2 P^\wedge q^\wedge \simeq 97$$

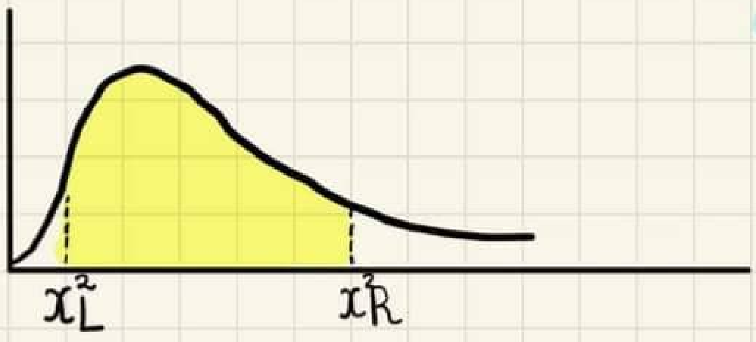
b) $P^\wedge = 0.2$ $q^\wedge = 0.8$

$n \simeq 62$

2

*C.I. for σ^2 and σ *

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$



C.I. for σ^2:

$$\left(\frac{(n-1)S^2}{X^2R}, \frac{(n-1)S^2}{X^2L}\right)$$

C.I. for σ:

$$\left(\sqrt{\frac{(n-1)S^2}{X^2R}}, \sqrt{\frac{(n-1)S^2}{X^2L}}\right)$$

*Given that the population is normally distributed.

*See the Guidelines, book, P 354

Recall: sample variance

$$S^2 = \frac{\Sigma(X-\bar{X})^2}{n-1} \quad \underline{\underline{or}}$$

$$S^2 = \frac{\Sigma X^2}{n-1} - \frac{(\Sigma X)^2}{n(n-1)}$$

see example 2, book, P 355

X^2R: 0.005 بالنسبة من المساحة ع يمينه

X^2L: 0.995 " " " " "

eg) Quality control engineer wishes to weight variation of a new product. A sample of 10 items is taken and provided $\bar{X}=0.6$ kgs and $S=0.4$ kgs. Find a 90% C.I for the variances of all items "Assume that the distribution is normally distributed"

Sol) $n=10$ $\quad \bar{X}=0.6$ $\quad S=0.4$

Area to the right of $X_R^2 = \frac{1-0.90}{2} = 0.05$

Area to the right of $X_L^2 = \frac{1+0.90}{2} = 0.95$

$d.f = 9$

الآن أجد قيمة X_R^2 و X_L^2 من الجدول.

$X_R^2 = 16.919$ $\qquad X_L^2 = 3.325$

C.I $\Rightarrow (0.09, 0.43)$.

* $X_1, X_2, X_3 \ldots X_{n_1} \overset{r.s}{\sim} n(\mu_1, \sigma_1^2)$
* $Y_1, Y_2, Y_3 \ldots Y_{n_2} \overset{r.s}{\sim} n(\mu_2, \sigma_2^2)$

$\bar{X} \sim n(\mu_1, \sigma_1^2/n_1)$
$\bar{Y} \sim n(\mu_2, \sigma_2^2/n_2)$

$$\bar{X}-\bar{Y} \sim n\left(\mu_1-\mu_2, \frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right)$$

The point estimate for $\mu_1-\mu_2$ is $\bar{X}-\bar{Y}$

C.I. for $(\mu_1-\mu_2)$:

$$(\bar{X}-\bar{Y}) \pm Z_c \cdot \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$$

Provided that σ_1, σ_2 are known

*Solve Q29, book, P449

* $$\bar{X}-\bar{Y} \pm t_c \cdot \sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}$$

Provided that $\sigma_1 \neq \sigma_2$ (unknown)
$d.f = \underbrace{\min}\{n_1-1, n_2-1\}$
القيمة الاصغر بينهم

* $$\bar{X}-\bar{Y} \pm t_c \cdot S_p \cdot \sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

Provided that $\sigma_1=\sigma_2=\sigma$ (unknown)
$d.f = n_1+n_2-2$

$$\underset{\hat{\sigma}^2}{S_p} = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$$

*Solve Q23, Book, P457

*Solve Q25, Book, P457

* C.I. for $P_1 - P_2$ *

$$\hat{P}_1 - \hat{P}_2 \pm Z_c \cdot \sqrt{\frac{\hat{P}_1 \hat{q}_1}{n_1} + \frac{\hat{P}_2 \hat{q}_2}{n_2}}$$

*Solve Q23, Book, P475 .

ملاحظة : لنفترض طلعت الفترة (0.1 , 0.3) هنا معناه انه $P_1 - P_2$ محصورة في هاي الفترة وبالتالي قيمتها موجبة مما يعني انه $P_1 > P_2$ فنسبة الطلاب الي بيدرسوا فنون حاليًا اكبر من الي قبل عشر سنوات .

واذا طلعت بالسالب مثلاً ($-0.2, -0.15$) فهذا معناه انه $P_1 < P_2$ يعني نسبة دارسي الفنون حالياً اقل من زمان

اذا طلعت الفترة فيها سالب وفيها موجب مثل $(-0.2, 0.3)$ هون بنحكي انه ما عنّا دليل على وجود فرق بين P_1 و P_2 ، ممكن يكون صفر!

Inferential statistics

- Estimation
  - Point
  - C.I
- Test of hypothesis

Test of hypothesis

Def) Statistical hypothesis: A statement about the population parameter (μ, σ, P)

يعني مثلاً لو بدّي اختبر فرضية انه متوسط دخل الاسرة في الاردن اقل من 800 ، فهاي الفرضية بتوصف μ وبدي اعرف اذا هي صح او خطأ .

* Types of hypothesis:-

1) Null hypothesis, H_0:- Contains equality ($=$ or $\geq$ or $\leq$)

2) Alternative hypothesis, H_a: ($\neq, <, >$)

النوعين من الفرضيات بستخدمهم لحتّى اوصف الـ Parameters ، يعني وحدة فيهم رح تكون الادّعاء الي بدنا ندرسه في السؤال

*See example 1, book, P371

ملاحظة على الفرع 1 :

بيحكيلنا انه $P = 0.61$ فهاي H_0

اذا بدّي عكسها $P \neq 0.61$ وهاي H_a

لكن الفرضيّة الي بتمثّل الادّعاء هي الأولى

الآن بدي اتخذ قرار بالفرض Ho والي هي ممكن تكون صحيحة او خاطئة، وبالنسبة لقراري فأنا رح يكون رفضها او قبولها .

	Ho True	Ho false
Reject Ho	Type 1 error α	✓ $1-\beta$
Accept Ho	✓ $1-\alpha$	Type 2 error β

✓ : رفضت النظرية وهي فعلياً خاطئة بالتالي قرار صائب

✓ : قبلت الفرضية وهي فعلياً صحيحة بالتالي قرار صائب .

Type 1 error: Rejecting Ho while it's true

Type 2 error: Accepting Ho while it's false

Accept = Fail to reject "نفس المعنى"

ملاحظة: ($1-\alpha$, α, $1-\beta$, β) تعتبر كلن احتماليات .

* α = Prob.(Type 1 error) = significance level = Prob(reject Ho | Ho true)

* $1-\alpha$ = Prob(Accept Ho while it is true) = C = confidence level

$\alpha + C = 1$ always

* β = Prob(Type 2 error)

* $1-\beta$ = Prob(Reject Ho while it is False) = Power of the test

ملاحظة: لا يمكن السيطرة على α و β معاً يعني رح نخطئ على الاكيد فشو نعمل ؟ بنثبت قيمة α حسب السؤال ونحاول نقلل β .

*See example 2, book, P374

Tests

* Test for μ

$H_0: \mu = \mu_0$ vs $H_a: \mu \neq \mu_0$ or
$H_0: \mu \leq \mu_0$ " $H_a: \mu > \mu_0$ or
$H_0: \mu \geq \mu_0$ " $H_a: \mu < \mu_0$

* Test for P

$H_0: P = P_0$ vs $H_a: P \neq P_0$ or
$H_0: P \leq P_0$ " $H_a: P > P_0$ "
$H_0: P \geq P_0$ " $H_a: P < P_0$

* Tests for σ^2

$H_0: \sigma^2 = \sigma_0^2$ vs $\boxed{H_a: \sigma^2 \neq \sigma_0^2}$ or

$H_0: \sigma^2 \leq \sigma_0^2$ " $\boxed{H_a: \sigma^2 > \sigma_0^2}$ or

$H_0: \sigma^2 \geq \sigma_0^2$ " $\boxed{H_a: \sigma^2 < \sigma_0^2}$

في حالة اللامساواة "زي الاولى" بنسميها

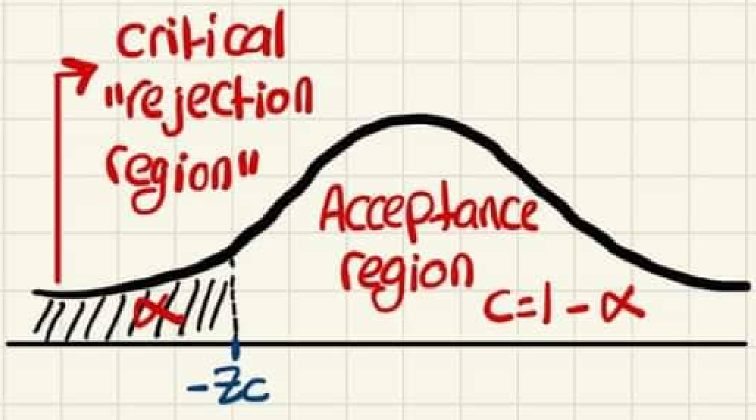
↳ two tailed test

في حالة المساواة "الثانية والثالثة" بنسميها

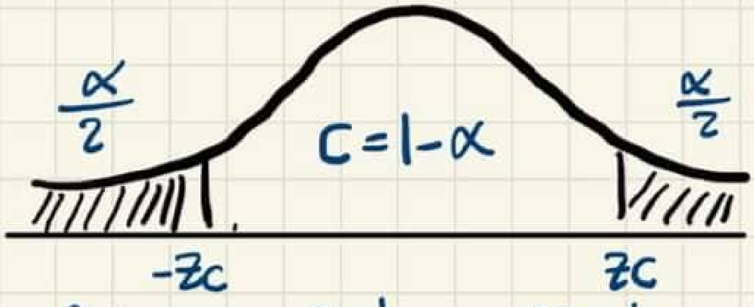
↳ one tailed test

Right tailed test بنسميها "اكبر من" في حالة

left tailed test " "اصغر من" " "

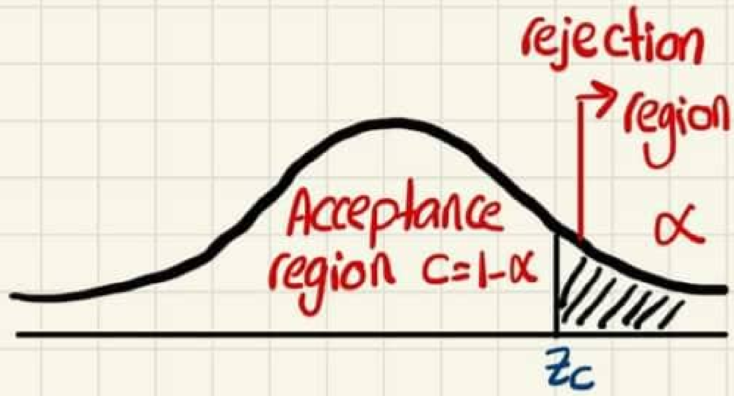


Left tailed



لـ الكتاب بسميها Z_0 ... نفس الاشي

two tailed



طيب شو فائدة ال Z_c و ال regions هاي في الجواب: بتساعدنا نتخذ قرار القبول او الرفض للنظرية ، وح نشوف كيف بس بالبداية في اشياء مهم نعرفها.

* Test statistic "Point estimate"		Standardized test statistics
$\mu \longrightarrow$	$\bar{X} \longrightarrow$	$Z=\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} \sim n(0,1)$ or $t=\frac{\bar{X}-\mu_0}{S/\sqrt{n}} \sim t(n-1)$
		or $Z=\frac{\bar{X}-\mu_0}{S/\sqrt{n}} \sim n(0,1)$ σ is unknown, $n \geq 30$
$P \longrightarrow$	$\hat{P} \longrightarrow$	$Z=\frac{\hat{P}-P_0}{\sqrt{\frac{P_0(1-P_0)}{n}}} \sim n(0,1)$
$\sigma^2 \longrightarrow$	$S^2 \longrightarrow$	$Z=\frac{S^2(n-1)}{\sigma_0^2} \sim \chi^2(n-1)$

الآن كيف s2 نقرّر القبول او الرفض لـ بنحسب standardized statistics

و بنقارن القيمة بـ Zc ، اذا اقل منها فالقرار هو Rejection واذا اكبر منها

فالقرار هو Fail to reject = Accept .

* في كمان طريقة من خلال حساب P-value ومقارنتها بـ α ، فإذا

كانت اقل من α او تساويها فالقرار "reject Ho" ، اما اذا كانت اكبر من α

فالقرار Accept Ho

*See example 4, book, P 388 : s2 اوضحه والله بتفاصيله

Ha: $\mu < 13$ "Claim"

Ho: $\mu \geq 13$ ↲ عكسها

$n=32$ $\bar{X}=12.9$ $\sigma=0.19$ بما انه σ معروفة s2 نستخدم Z

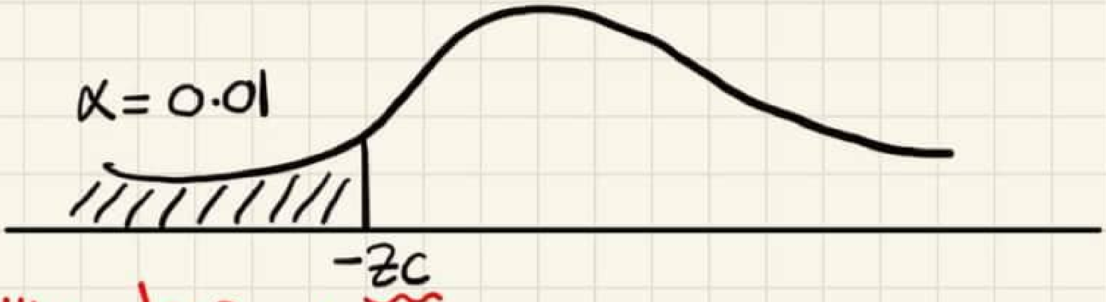
$\alpha=0.01$

الطريقة الاولى: من خلال Standardized statistics

انتبه انه μ_0 بنجيبها من Ho

$$z = \frac{\bar{x} - \mu^{\circ}}{\sigma/\sqrt{n}} = \frac{12.9 - 13}{0.19/\sqrt{32}} = -2.98$$

بما انه $Ha: \mu < 13$ فهنا "left tailed test"

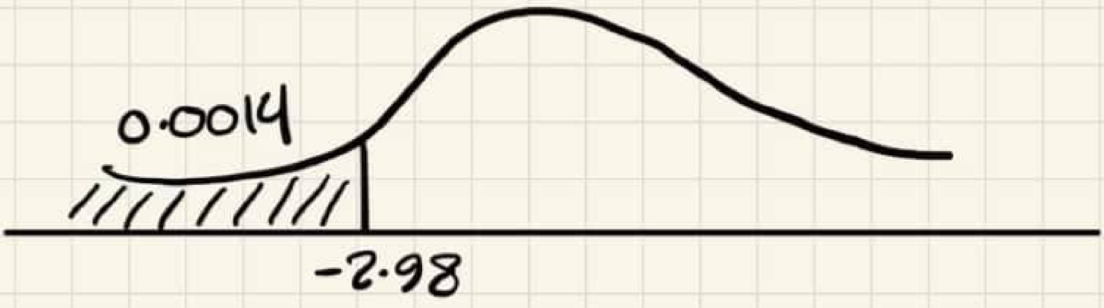


← بجيب قيمتها من الجدول

$-Zc = -2.33$ "اخذت المساحة الاقرب ل 0.01"

$-2.98 < -2.33$ وعليه القرار هو rejection of Ho واحتمالية الخطأ بالقرار هي 0.01 .

* الطريقة الثانية: عن طريق حساب P-value ... كيف احسبها في هذا المثال بحسب standardized statistic والمساحة تحته هي P-value



rejection of Ho اذاً $0.01 > 0.0014$

*see example 5 , book, P 389:   دعنو رح احلّه بالتفصيل ^^

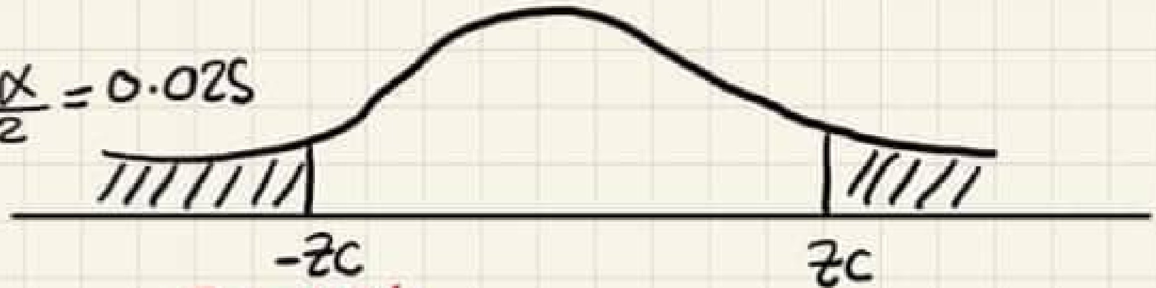
$H_0: \mu = 68.3$ $H_a: \mu \neq 68.3$ "Claim" $n = 25$

$\bar{X} = 67.2$ $\sigma = 3.5$ $\alpha = 0.05$

Sol) $z = \dfrac{\bar{X} - \mu^{\circ}}{\sigma/\sqrt{n}} = -1.57$

بما انه "Ha ≠ " فهذا : two tailed test

$\frac{\alpha}{2} = 0.025$

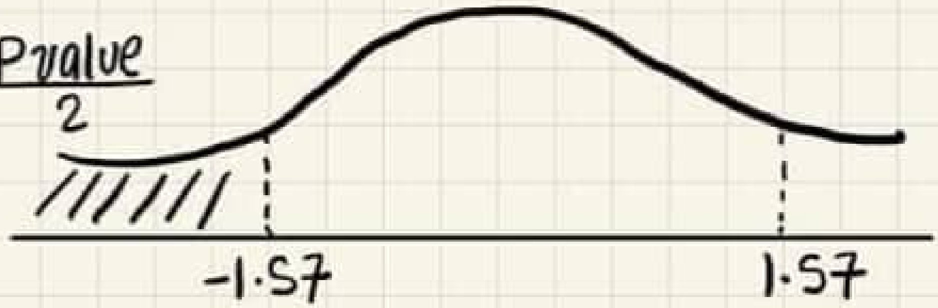


$Z_c = -1.96$

Fail to reject Ho اذاً $-1.96 < -1.57$

* الحل عن طريق حساب P-value :

$\dfrac{P\text{-value}}{2} = 0.0582 \Leftarrow \dfrac{P\text{value}}{2}$



$P\text{-value} = 0.1164 > \alpha$

Fail to reject Ho

لاحظ: الاشي الي بيسأل عنه السؤال هو Ha وليس Ho واحنا
دائمًا بناخذ Ho عكس Ha

14