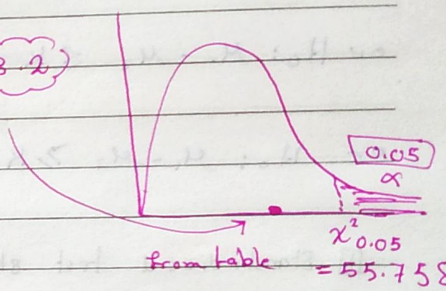
Test for σ^2

eg) A dairy processing company claims that the variance of the amount of fat in the whole milk processed by the company is no more than 0.25. You suspect this is wrong and find that a random sample of 41 milk containers has a variance of 0.27. At $\alpha = 0.05$, is ther enough evidence to reject the company's claim? Assume the population is normally distributed.

$H_0: \sigma^2 \leq 0.25$ (claim) and $H_a: \sigma^2 > 0.25$ (right-tailed test) $n = 41$

$d.f = 41 - 1 = 40$ $s^2 = 0.27$

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{40(0.27)}{(0.25)} = 43.2$$



Accept H_0 → Accept the claim.
Don't reject the claim.

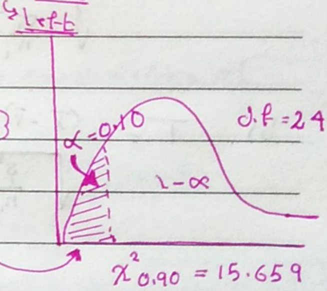
Page 420 Chapter 7. Example (5)

the claim → " the standard deviation is less than 1.4 minutes."

$H_0: \sigma \geq 1.4$ (σ_0) & $H_a: \sigma < 1.4$ (claim) Left

$\boxed{n = 25}$, $\boxed{s = 1.1}$ $\boxed{\alpha = 0.10}$

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{(25-1)(1.1)^2}{(1.4)^2} \approx 14.816$$



[→ reject H_0 → Don't reject the claim] in rejection region ch $\boxed{8}$ [2 Samples]

"The last lecture" :)

If $x_1, x_2, \ldots, x_{n_1} \overset{r.s}{\sim} n(\mu_1, \sigma_1^2)$
& $y_1, y_2, \ldots, y_{n_2} \overset{r.s}{\sim} n(\mu_2, \sigma_2^2)$ } Independent sample

mean estimate for μ_1 : $\bar{x}_1$

" " " μ_2 : $\bar{x}_2$

{1} inDependent Samples

* Test for $\mu_1 - \mu_2$ لمقارنة هل هما متساويات او مين الاكبر و الأصغر منها

$H_0: \mu_1 - \mu_2 = K$ vs $H_a: \mu_1 - \mu_2 \neq K$ → two tailed test

or $H_0: \mu_1 - \mu_2 \leq K$ vs $H_a: \mu_1 - \mu_2 > K$ → right tailed test

or $H_0: \mu_1 - \mu_2 \geq K$ vs $H_a: \mu_1 - \mu_2 < K$ → left tailed test

$\boxed{1}$ Standardized test statistic is :-

a) $Z = \dfrac{(\bar{x} - \bar{y}) - (\mu_1 - \mu_2)_0}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$, if σ_1^2 & σ^2 are known

لن نسأل عن p-value الا بهذه الحالة.

b) $T = \dfrac{(\bar{x} - \bar{y}) - (\mu_1 - \mu_2)_0}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$, if $\sigma_1^2 \neq \sigma^2$ (are unknown)

تذكري: $[d.f = \min\{n_1 - 1, n_2 - 1\}]$

c) $T = \dfrac{(\bar{x} - \bar{y}) - (\mu_1 - \mu_2)_0}{sp \cdot \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$, if $\sigma_1^2 = \sigma^2 = \sigma$ (unknown)

$[d.f = n_1 + n_2 - 2]$

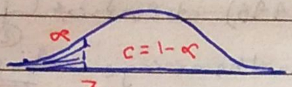
$\hat{\sigma}$

$sp = \hat{\sigma} = \sqrt{\dfrac{(n_1 - 1) S_1^2 + (n_2 - 1) S_2^2}{n_1 + n_2 - 2}}$

COLLEGE

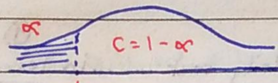
Ch $\boxed{8}$ 2 samples "The last lecture" :)

الطريقة الأولى :



Z_c

Left tailed test



t_c

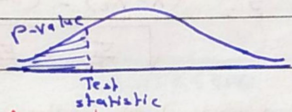
Left tailed test

ثم نختبر test statistic اذا كانت ضمن المنطقة المظللة (reject H_0)

واذا كانت ضمن المنطقة الغير مظللة (Accept H_0) ← Don't reject H_0

↳ failed to reject H_0

الطريقة الأخرى باستخدام P-value .



Left tailed test

نجد المساحة من (table)

وتكون هي قيمة P-value

فإذا كانت :

if P-value ≤ α , reject H_0

and if P-value > α , Accept H_0

→ ① σ_1 & σ_2 are known ⇒ Z-test

Example 2 ← Page (444) يجب الرجوع الى الكتاب صفحة

Claim: "there is a difference in the mean credit card debts of people in California and Florida."

$H_0: \mu_1 = \mu_2$ → $\mu_1 - \mu_2 = 0$ ↳ $\boxed{K}$ and $H_a: \mu_1 \neq \mu_2$. [Claim] ↳ two tailed test

250 people from each state

$n_1 = n_2 = 250$, the two samples are independent , $\boxed{\alpha = 0.05}$

Assume that $\sigma_1 = 960$, $\sigma_2 = 845$

California : $\bar{x}_1 = 3060$, $n_1 = 250$, $\sigma_1 = 960$

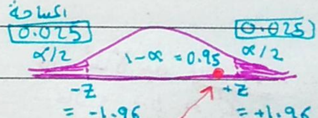
Florida : $\bar{x}_2 = 2910$, $n_2 = 250$, $\sigma_2 = 845$

} are known

Sol)

(Test statistic)

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(3060 - 2910) - 0}{\sqrt{\frac{(960)^2}{250} + \frac{(845)^2}{250}}} = 1.85$$



(You fail to reject the H_0) → (there isn't enough evedince to support the claim) z is not in the rejection rejon.

Example (3) ← (Page 445) بص في الكتاب

Claim :- The averge daily cost ......in Texas (1) is less than the average ....in Virgina (2).

($H_0: \mu_1 \geq \mu_2 \Rightarrow \mu_1 - \mu_2 \geq 0$ $H_a: \mu_1 < \mu_2$ [claim])
→ left tailed test

المعطيات
* Texas :- $\bar{x}_1 = 245$, $n_1 = 25$, $\sigma_1 = 20$
* Virginia :- $\bar{x}_2 = 251$, $n_2 = 20$, $\sigma_2 = 25$

are known

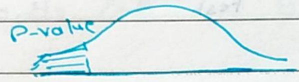
The two samples are independent and that both populations are normally distributed. , $\alpha = 0.01$,

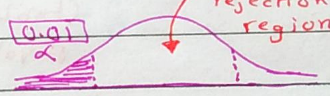
Is there enough evidence to support the claim?

Sol) Test statistic :- $z = \dfrac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} = \dfrac{(245 - 251) - 0}{\sqrt{\dfrac{(20)^2}{25} + \dfrac{(25)^2}{20}}} \approx -0.87$

is not in the rejection region

Accept H_0 → Don't support the claim.





p-value = 0.1922 > α = 0.01 → Accept H_0

(2) σ_1 & σ_2 are unkown ⇒ t-test

Example (1) ← Page 453 يجب الرجوع الى

Claim :- there is a difference in the mean mathematics ...

$H_0: \mu_1 = \mu_2 \rightarrow \mu_1 - \mu_2 = 0$ and $H_a: \mu_1 \neq \mu_2$ [claim]
→ (2 tailed test)

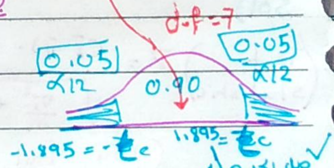
المعطيات
Teacher 1 : $\bar{x}_1 = 473$, $s_1 = 39.7$, $n_1 = 8$ → d.f = 7 ✓
Teacher 2 : $\bar{x}_2 = 459$, $s_2 = 24.5$, $n_2 = 18$ → d.f = 17

σ_1 & σ_2 are unknown ← لاحظوا ان

→ Assume the population variances are not equal ($\sigma_1^2 \neq \sigma_2^2$)

→ $\alpha = 0.10$ → $\alpha/2 = 0.05$

Sol) Test statistic :- $t = \dfrac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} = \dfrac{(473 - 459) - 0}{\sqrt{\dfrac{(39.7)^2}{8} + \dfrac{(24.5)^2}{18}}} \approx 0.922$



لن نسأل عن P-value في حالة χ^2 و t .. فقط في حالة Z test

→ نحصل على critical value فقط → Accept H_0 → reject the claim

خطيرة: P-value عكس مهمة بالسؤال وليس علينا تقرر

P-value ≤ α → reject H_0
P-value > α → Accept H_0

Example 2 Page 454

Claim:- " the mean driving cost per mile of a manufactur's sedans is less than that of its leading competitor."

$H_0: \mu_1 \geq \mu_2 \rightarrow \mu_1 - \mu_2 \geq 0$ $H_a: \mu_1 < \mu_2$ [claim]

↳ left tailed test

Manufacturer: $\bar{x}_1 = 0.48$, $s_1 = 0.05$, $n_1 = 30$

Competitor: $\bar{x}_2 = 0.51$, $s_2 = 0.07$, $n_2 = 32$

> σ_1^2 & σ_2^2 are unknown

$\boxed{t\text{-test}}$

Assume the population variances are equal. ($\sigma_1^2 = \sigma_2^2$)

$\alpha = 0.05$

$d.f = n_1 + n_2 - 2$
$= 30 + 32 - 2 = 60$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_0}{\underbrace{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}_{\text{standard error}}} = \frac{(0.48 - 0.51) - 0}{0.0155416} \approx -1.930$$

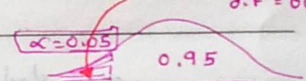
in rejection region

standard error $S_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$

$$\rightarrow = \sqrt{\frac{(30-1)(0.05)^2 + (32-1)(0.07)^2}{60}} \sqrt{\frac{1}{30} + \frac{1}{32}} \approx 0.0155416$$

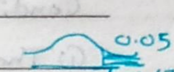
reject H_0 ⇒ there is enough evidence to support the claim.

$d.f = 60$



$-1.671 = t$

لأنه متماثل اخذ السالب



symmetric $t = 1.671$

8.4 Test for $P_1 - P_2$:

$H_0: P_1 - P_2 = K$ vs $H_a: P_1 - P_2 \neq K$

$H_0: P_1 - P_2 \leq K$ vs $H_a: P_1 - P_2 > K$

$H_0: P_1 - P_2 \geq K$ vs $H_a: P_1 - P_2 < K$

① the samples are
a) randomly selected
b) independent
c) large ...

standardized test statistic is :-

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (P_1 - P_2)_0}{\sqrt{\bar{p}\,\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

نستخدمها في p-value

larg samples:
$n_1\bar{p} \geq 5$, $n_2\bar{p} \geq 5$
$n_1\bar{q} \geq 5$, $n_2\bar{q} \geq 5$

n_1, x
$\hat{p}_1 = \frac{x}{n_1}$
n_2, y
$\hat{p}_2 = \frac{y}{n_2}$
$\bar{p} = \frac{x+y}{n_1+n_2}$, $\bar{q} = 1 - \bar{p}$

A Two-Sample z-test for the Difference Between Proportions

Example [1] Page 472 [1]

$H_0: p_1 = p_2 \rightarrow \boxed{k=0}$ [claim] and $H_a: p_1 \neq p_2 \rightarrow$ 2 tailed test

Passenger cars $n_1 = 200$, $\hat{p}_1 = 0.910$, $x_1 = 182$ $\rightarrow \hat{p} = \frac{x_1}{n_1} \rightarrow x_1 = \hat{p} n_1$

Pickup trucks $n_2 = 250$, $\hat{p}_2 = 0.832$, $x_2 = 208 \rightarrow x_2 = \hat{p} n_2$

large samples

$\alpha = 0.10$

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)_0}{\sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{(0.910 - 0.832) - (0)}{\sqrt{(0.8667)(0.1333)\left(\frac{1}{200} + \frac{1}{250}\right)}} = 2.42$$

in rejection region

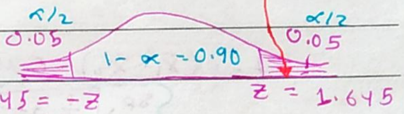
$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{182 + 208}{200 + 250} = 0.8667, \quad \bar{q} = 1 - 0.8667 = 0.1333$$

P-vale = 0.0156

z

$< \alpha \rightarrow$ reject H_0

← reject H_0
reject the claim



8.3 (2) (1) Dependent Samples

- Testing the differenca between means [Dependent samples]

اذا عملنا التجربة على نفس الاشخاص قبل وبعد شيء معين ..

مثلاً يتم تجربة دواء من قبل الاطباء على نفس المرضى ويتم فحصهم قبل اخذ الدواء وبعده.

Conditions :-

1. The sample are randomly selected.
2. the samples are dependent (paired).
3. The populations are normally distributed or the number n of pairs of data is at least 30.

x_i	y_i	$d_i = x_i - y_i$	d_i^2
		Σd_i	Σd_i^2

$$\boxed{\bar{d} = \frac{\Sigma d_i}{n}} \qquad s_d = \sqrt{\frac{\Sigma d_i^2}{n-1} - \frac{(\Sigma d_i)^2}{n(n-1)}}$$

C.I for μ_d : $\bar{d} \pm t_c \cdot \dfrac{s_d}{\sqrt{n}}$

$$\bar{d} - t_c \frac{s_d}{\sqrt{n}} < \mu_d < \bar{d} + t_c \cdot \frac{s_d}{\sqrt{n}}$$

standardized test statistic is : $t = \dfrac{\bar{x} - \mu_d}{s_d/\sqrt{n}}$, d.f = n-1

Example $\boxed{1}$ Page 462

Athlete	1	2	3	4	5	6	7	8	
vertical jump..(x_i) before..	24	22	25	28	35	32	30	27	$\alpha = 0.10$ claim :
vertical jump..(y_i) After using..	26	25	25	29	33	34	35	30	(After > Before) y_1 x_1
Defference (d_i)	-2	-3	0	-1	+2	-3	-5	-3	$\rightarrow \Sigma d_i = -14$
$(d_i)^2$	4	9	0	1	4	9	25	9	$\rightarrow \Sigma d_i^2 = 56$

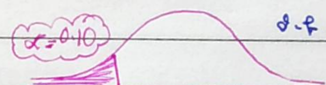
$$s_d = \sqrt{\frac{\Sigma d_i^2}{n-1} - \frac{(\Sigma d_i)^2}{n(n-1)}} = \sqrt{\frac{65}{8} - \frac{(-14)^2}{8(7)}} = 2.1213$$

claim: $(x_1 - y_1) < 0 \rightarrow H_a: \mu_d < 0$ [claim] and $H_0: \mu_d \geq 0$

↳ Left tailed test

$$\bar{d} = \frac{\Sigma d}{n} = \frac{-14}{8} = -1.75$$

d.f = 8-1 = $\boxed{7}$



$t = -1.415$ ← مصطلح الموجب من الجدول

The standardized test statistic is $t = \dfrac{\bar{d} - \mu_d}{s_d/\sqrt{n}} = \dfrac{-1.75 - 0}{2.1213/\sqrt{8}} \approx -2.333$

in rejection region

23. Drug Testing. Page 469

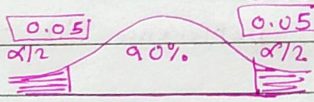
Patient	1	2	3	4	5	6	7	8	$(\sum d)$
without drug x_i	1.8	2.0	3.4	3.5	3.7	3.8	3.9	3.9	
using the drug y_i	3.0	3.6	4.0	4.4	4.5	5.2	5.5	5.7	
$d_i\ (x_i - y_i)$	-1.2	-1.6	-0.6	-0.9	-0.8	-1.4	-1.6	-1.8	$\sum d_i = -9.9$
$(d_i)^2$	1.44	2.56	0.36	0.81	0.64	1.96	2.56	2.24	$\sum d_i^2 = 13.57$

$$s_d = \sqrt{\frac{\sum d_i^2}{n-1} - \frac{(\sum d_i)^2}{n(n-1)}} = \sqrt{\frac{(13.57)}{7} - \frac{(-9.9)^2}{8(7)}} = \sqrt{0.18839}$$

$$s_d = 0.434, \quad \bar{d} = \frac{\sum d_i}{n} = \frac{-9.9}{8} = -1.238$$

d.f = 7

$$\bar{d} - t_c \frac{s_d}{\sqrt{n}} < \mu_d < \bar{d} + t_c \frac{s_d}{\sqrt{n}}$$



$-t_c = \boxed{-1.895}$

$$\left(-1.238 - \left(-1.895 \cdot \frac{0.434}{\sqrt{8}}\right) < \mu_d < -1.238 + \left(-1.895 \cdot \frac{0.434}{\sqrt{8}}\right)\right)$$

~~-1.5288~~ ~~$< \mu_d <$~~ $= -0.947 < \mu_d < -1.529$

For $\mu \rightarrow$ point estimate $(\bar{x})$

C.I $\Rightarrow$ $\bar{x} \pm z_c \cdot \frac{\sigma}{\sqrt{n}}$ (σ known)

$\bar{x} \pm t_c \cdot \frac{s}{\sqrt{n}}$ (σ unknow)

$\bar{x} \pm z_c \cdot \frac{s}{\sqrt{n}}$ ($n \geq 30$)

test statistic $\Rightarrow$

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \quad (\sigma \text{ known})$$

$$T = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \quad (\sigma \text{ unknown})$$

For p → point estimate $\hat{p} = \frac{x}{n}$

C.I ⇒ $\hat{p} \pm z_c \sqrt{\frac{\hat{p}\hat{q}}{n}}$

test statistic is ⇒ $\dfrac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}}$

For σ^2 → point estimate is S^2

C.I ⇒ $\left(\dfrac{(n-1)s^2}{\chi^2_R}, \dfrac{(n-1)s^2}{\chi^2_L}\right)$

test statistic is: $\dfrac{(n-1)s^2}{\sigma_0^2}$, d.f $= n-1$

For $\mu_1 - \mu_2$ → point estimate is $\cancel{\bar{X}_1 - \bar{X}_2}$ $\bar{x} - \bar{y}$

C.I → $\cancel{\hat{p}_1 - \hat{p}_2}$

- $(\bar{x}-\bar{y}) \pm z_c \cdot \sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$ (σ_1, σ_2 known) d.f = min $\{n_1 - 1, n_2 - 1\}$
- $(\bar{x}-\bar{y}) \pm t_c \sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}$ ($\sigma_1 \neq \sigma_2$ unknown) d.f $= n_1 + n_2 - 2$
- $(\bar{x}-\bar{y}) \pm t_c \, \hat{\sigma} \cdot \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}$ ($\sigma_1 = \sigma_2$ unknown)

* $\hat{\sigma} = \sqrt{\dfrac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}}$

Test statistic is :- $Z = \dfrac{(\bar{x}-\bar{y}) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$, ($\sigma_1, \sigma_2$ known)

$T = \dfrac{(\bar{x}-\bar{y}) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ ($\sigma_1 \neq \sigma_2$) unknown, d.f = min $\{n_1 - 1, n_2 - 1\}$

$T = \dfrac{(\bar{x}-\bar{y}) - (\mu_1 - \mu_2)_0}{\hat{\sigma}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$, $\sigma_1 = \sigma_2 = \sigma$ unknown, d.f $= n_1 + n_2 - 2$

For $P_1 - P_2$ → Point estimate is $\hat{P}_1 - \hat{P}_2$

C.I → $(\hat{P}_1 - \hat{P}_2) \pm Z_c \cdot \sqrt{\frac{\hat{P}_1 \hat{q}_1}{n_1} + \frac{\hat{P}_2 \hat{q}_2}{n_2}}$

test statistic is :-

$$Z = \frac{(\hat{P}_1 - \hat{P}_2) - (P_1 - P_2)_0}{\sqrt{\bar{p}\,\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

حيث :- $\bar{p} = \frac{x_1 + x_2}{n_1 + n_2}$, $\bar{q} = 1 - \bar{p}$

* Paired data:-

C.I $\bar{d} \pm t_c \frac{S_d}{\sqrt{n}}$, $df = n-1$

Test statistic :- $\frac{\bar{d} - (\mu_d)_0}{S_d / \sqrt{n}}$

- $\mu_d < 0$ → Before < after
  $\mu_d > 0$ → Before > after
  $\mu_d = 0$ → Before = after.