* Normal distribution :

1) is a continuous distribution

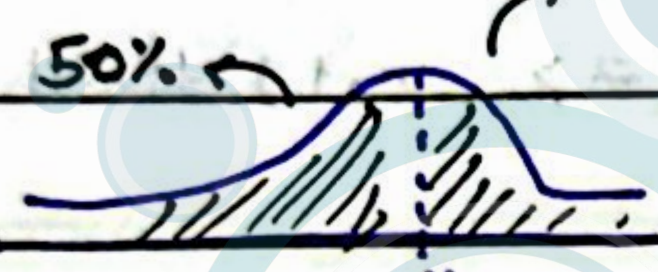
2) $\int_{-\infty}^{\infty} f(x) = 1$ (المساحة كاملة تحت المنحنى = 1)

3) $p(x=k)$ (عند نقطة معينة = صفر)

↳ $\int_{k}^{k} f(x) = 0$ لأنه * area = Probability.

4) لازم نتعامل مع فترة

5) المساواة غير موجودة



symmetric

* الـ Normal بكون (bell shape)

بالمنتصف عنا الـ average

$X \sim N(\mu, \sigma^2)$

the data ← parameters

* إذا الـ X توزيعوا Normal بنحولوا لـ standard normal distribution ← إله جداول

لل Z إلي الـ mean zero والـ variance إله one

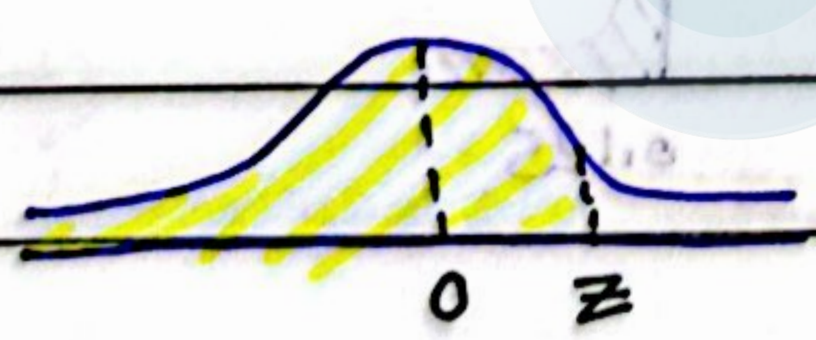
$Z \sim N(0,1)$

* التحويل بكون على هاي العلاقة ← $\boxed{Z = \frac{X-\mu}{\sigma}}$

* جدول الـ Normal تراكمي، بعطي المساحة من كل قيمة الـ Z إلي معي لحد بداية الـ curve.

* الجدول مقسم قسم موجب وقسم سالب (من الصفر)

↲ لحد 3.4 ↲ لحد -3.4



* مثلا لو بدي ألاقي الاحتمال

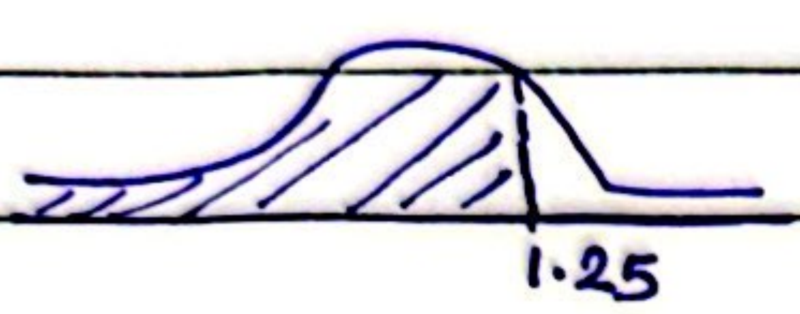
جيب الـ Z = 0.1

0.1 + 0.01

Z	0.01	0.02	...	0.09
0.0				
0.1	# ↳ 0.543795			
⋮				
3.4				

if $Z \sim N(0,1)$

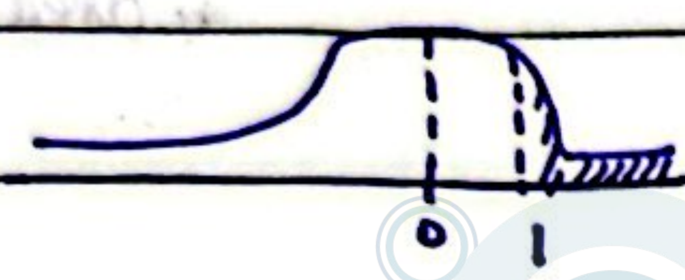
Ex: find: 1) $P(Z<1.25)$



مباشرة من الجدول

$1.2 + 0.05$ المساحة تحت المنحنى $= 0.8944$

$P(Z>1)$

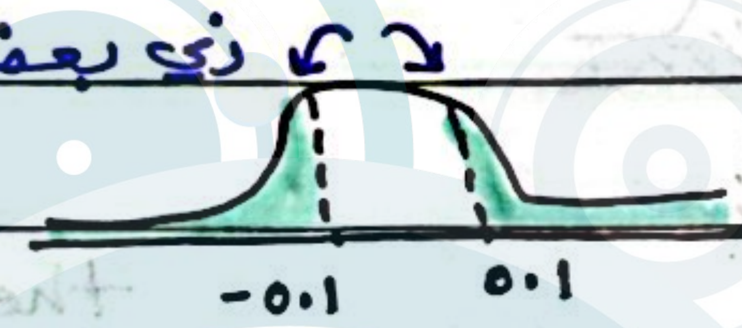


المساحة كلها

$1 - P(Z<1)$ → $1.0 + 0.00$

$1 - 0.8413 = 0.1587$

بما إنه المنحنى متماثل ، احتمال $(Z>0.1)$ مثلاً هو نفسه إحتمال $(Z<-0.1)$



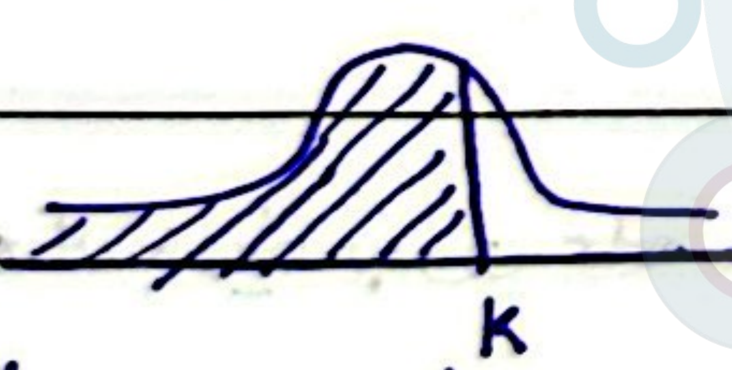
Ex: find the value of k, such that:

1) $P(Z<k) = 0.975$

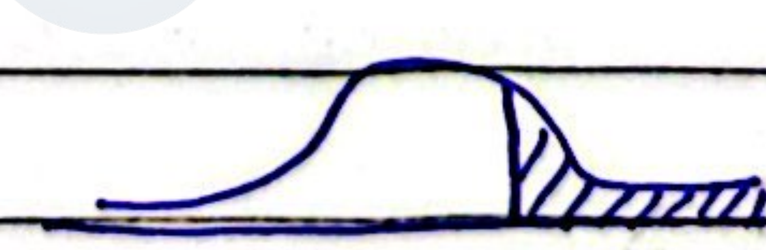
بدور وين 0.975

بالجدول بتطلع $Z = 1.96$

اللي هي K



2) $P(Z>k) = 0.125$



$1 - 0.125 = 0.875$ → بندور عليها بالجدول $\rightarrow K = 1.15$

Ex: $X \sim N(20, 16)$, find:

هاد نوع من الأسئلة بكون بده أحوّل لـ standard .

1) $P(X<20)$

$Z = \dfrac{x - \mu}{\sigma}$ $\quad \mu = 20$, $\sigma = 4$

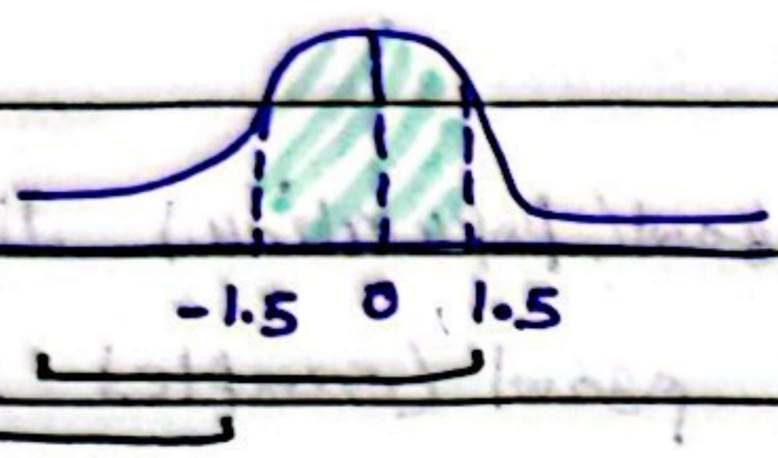
$P(Z<0) = 0.5$ (جدول)

2) $p(14 < X < 26)$

$Z = \frac{X-\mu}{\sigma}$ ⟶ $Z = \frac{14-20}{4} < Z < Z = \frac{26-20}{4}$

p

$\frac{-6}{4} < Z < \frac{6}{4}$ ⟶ $(-1.5 < Z < 1.5)$



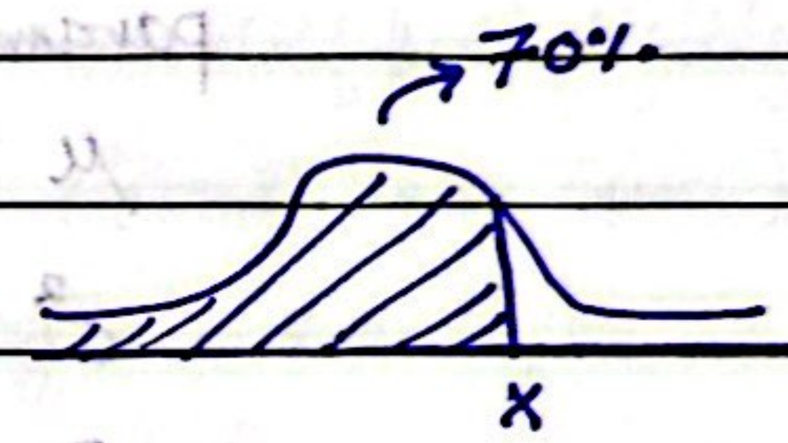
$P(Z<1.5) - p(Z<-1.5)$

$0.9332 - 0.0668 = 0.8664$

* **Ex:** $X \sim N(20, 16)$

the 70^{th} percentile?

↳ يعني شو قيمة X إلي تحتها 70% من القيم ؟



* بنطلع قيمة الـ Z إلي تحتها 70% من القيم و بعدين بحوّلها لـ X .

أول إشي

$P(Z<k) = 0.7$

↳ $k = 0.52$ أقرب إشي

$Z = \frac{X-\mu}{\sigma}$ ⟶ $0.52 = \frac{X-20}{4}$ ⟶ $X = 22.08$

"introduction to stat 2"

(sampling)

* Statistical inference : لما يكون عندي population ما بقدر أعمل عليه statistic معينة فباخذ منه sample يكون (random) .

* الـ sample هاد بعطيني معلومات بتساعدني إني :

1) draw a conclusion } عن الـ population
2) Make a decision

* عندي (3 methods) بقدر أعملها عالـ sample تساعدني لأعمل 1 و 2 للـ population .

1) point estimation ↓ chapter 7

2) statistical intervals ↓ chapter 8

3) test of hypothesis ↓ chapter 9

*** point estimation:** ذكرنا ال population ما بقدر احسب ال parameters فباخذ ال sample
بحسب لل sample ← average ، بسميه $\bar{X}$ ، فال $\bar{X}$ هو ال Mean بال population.
كل parameter بال population إله نظير بال sample بكون point estimator إله كقيمة و اسمه statistic

* يعني بال (population) اسمهم (parameters)
أما بال (sample) اسمهم (statistics).

شو من ال objects بال population →
* احنا مهتمين بال mean ، varience ، proportion بنتمي لـ class معينة.
مثلاً أعمار الناس اللي بال ← (جزء من ال pop بتمثل فئة ما)

parameter	point estimator
μ	$\bar{X}$
σ^2	S^2
P	$\hat{P}$

* ال sample لازم يكون "random".
* لو ناخذ أكثر من sample لنفس ال population ، ال $\bar{X}$ الهم راح تختلف ومش ضروري يكون ال mean واحد من هدول ال $\bar{X}_s$ عادي.
هدول كلهم statistics و point estimators
بس عشان من سامبل لثاني بختلفو ، فال $\bar{X}$ هاي تعتبر "random variable" يعني إله distribution

$\bar{X}_1 = 10 \leftarrow \mu \rightarrow \bar{X}_3 = 9.4$

$\bar{X}_2 = 9.9$

* إحتمال يكون ال pop mean مساوي تمامًا لواحد من ال $\bar{X}$ ← goes to zero
عشان هيك صرنا نستخدم (confedence intervals) بكون ال parameter من ضمنهم بس برضه ما بنعرف بالزبط كم.
* مثلاً: $\underbrace{9}_{\text{lower limit}} < \mu < \underbrace{11}_{\text{upper limit}}$
L ، U بعتمدو عال (sample data) يعني بختلفو من سامبل لثاني ، يعني RVs.

8.1) confidence intervals on the mean : بهاد التشابتر ما بكون بعرف ال mean
بس بعرف ال variance .

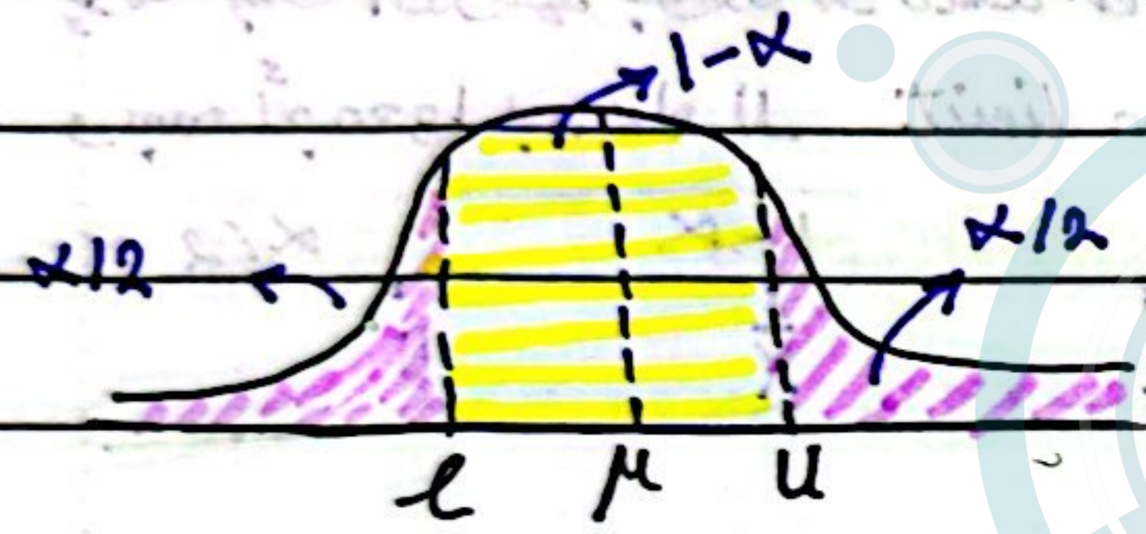
* قلنا كل statistic إله distribution وهاد ال distr إله mean & variance

* إذا ال pop كانت normally distributed معناه $\bar{x}$ بردو هيك . ال mean إله $\mu_{\bar{x}} = \mu_{pop}$

وال $\sigma_{\bar{x}} = \sigma_{pop}/\sqrt{n_{sample}}$ ، n: sample size

* هلأ إحنا بس بتقدر نعمل estimation for the mean ويكون بين ، $l < \mu < u$

ممكن أكبر من ال upper كمان وممكن أصغر من ال lower



* ال probability إنه ال mean يكون ضمن ال u وال l

اسمه (confidence level) ، أما أكبر من u أو أصغر من l

اسمها (significance level)

ال SL اسمها α ، أما ال CL هي ال (1-α)

طبعًا مجموعهم لازم يساوي 1 إلي هي ال (area under the curve) .

* المساحة إلي بعد ال u والي قبل ال l قد بعض ، لإنه ال normal distribution حوالين ال mean بكون symmetric.

تحويل ل Z

* كل normal variable بقدر أحوله standerdization بهاي العلاقة

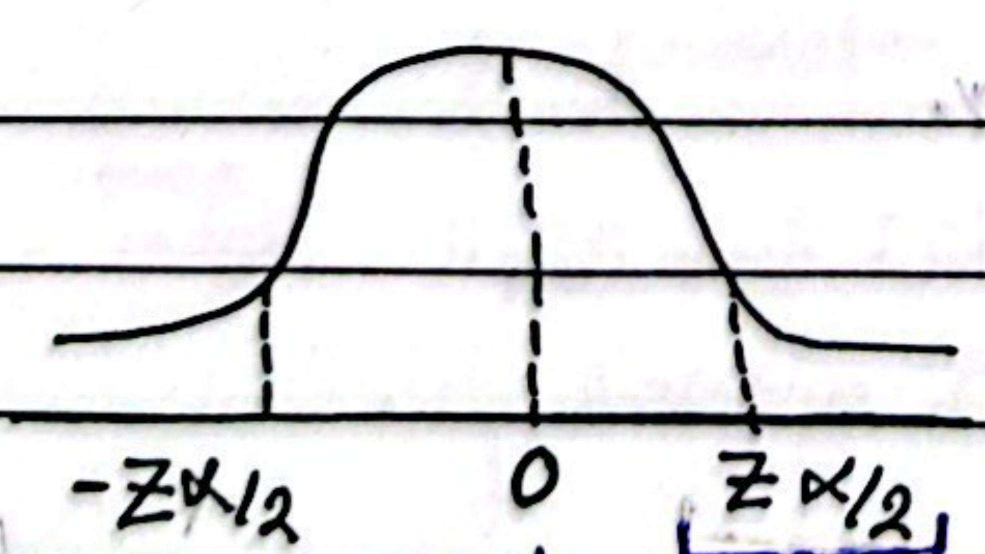
$$\frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}/\sqrt{n}}$$

ال mean إله

ال SD إله

طبعًا أي normal variable بقدر أحوله هيك ، مش بس $\bar{x}$.

ال normal distribution ليتحول ل standard بيبين هيك



هاي الكمية بسميها critical value ، or percentile هي قيمة معينة ل ال standard normal variable (Z) و ال area فوقها هي $\alpha/2$.

هاد μ للdistr مش ل $\bar{x}$.

$\bar{x}$ ال mean إلها ثابت μ .

* بنحول ل standard عشان ال statistical procesures عليه أسهل ، بس احنا بنقدر نشتغل عال Z .
← (جدول)

in normal dist $\rightarrow$ $L < \mu < u$

in standard normal dist $\rightarrow$ $\overset{\text{lower}}{-Z_{\alpha/2}} < Z < \overset{\text{upper}}{Z_{\alpha/2}}$

$$P\left[-Z_{\alpha/2} < Z < Z_{\alpha/2}\right] = 1-\alpha$$

$$-Z_{\alpha/2} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < Z_{\alpha/2} = 1-\alpha \leftarrow$$ μ فكيه

$$\boxed{\bar{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + Z_{\alpha/2}\underbrace{\frac{\sigma}{\sqrt{n}}}_{\text{the SD}}}$$

هاي هي ال formula إلي بنحسب فيها ال confidence interval لل mean ال

$$\bar{X} - E < \mu < \bar{X} + E$$

E: margin of error or range of error

لأنه بس أطرحها بوصل لل lower limit
وبس أجمعها بوصل لل upper limit

$$E: Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$\alpha/2$ — $1-\alpha$ — $\alpha/2$

L — $\bar{X}$ — U

$$\bar{X} - \overset{\text{percentile (table)}}{Z_{\alpha/2}} \underset{\text{population SD (given)}}{\frac{\sigma}{\sqrt{n}}} < \mu < \bar{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$\sqrt{n}$ → sample size (given)

*problem 8-10:

(confidence level) $1-\alpha$ هو ال confidence interval
* دائمًا الرقم إلي قبل ال
بكون على شكل نسبة مئوية.

* السؤال بده تعطيه interval يكون ال mean فيها بنسبة 95%.

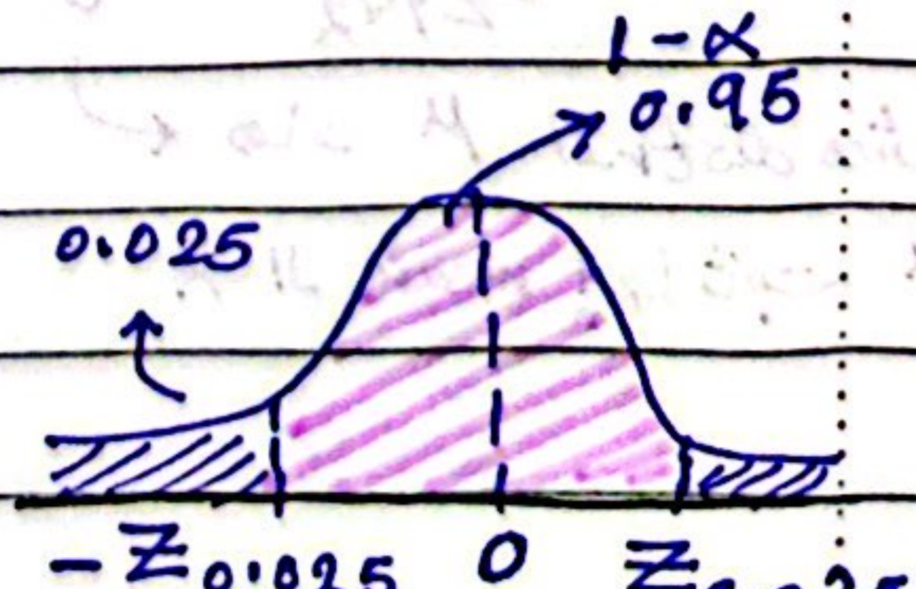
sol: $\bar{X}: 98$, $\sigma = 2$, $n = 9$, و $1-\alpha = 95\%$

so $\alpha = 0.05$, $\alpha/2 = 0.025$

$$\bar{X} - E < \mu < \bar{X} + E$$

$E = Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ و $Z_{0.025}$ * قيمة ال standard normal variable

إلي على يمينها (0.025) أو
على يسارها مساحة (1-0.025)



من الجدول

هي نفسها ال probability ((

* the percentile is the numerical value for the standard normal variable.

مثلاً -3.95 هاي percentile ، التقاطع الـ -3.9 مع 0.05 يعطيني الـ area باليسار ، يعني من $-\infty$ لحد الرقم

↓ The probability

$E = Z_{0.025} \cdot \frac{\sigma}{\sqrt{n}}$ $1.96 * \frac{2}{\sqrt{9}}$

↓ $-1.96 \rightarrow$ باخذها absolute value $= 1.3$

لإنه القانون أصلاً فيه سالب (المتباينة)

$\bar{x} - 1.3 < \mu < \bar{x} + 1.3$

$98 \leftarrow$

$96.7 < \mu < 99.3$

$\overline{96.7 \quad\quad 99.3}$

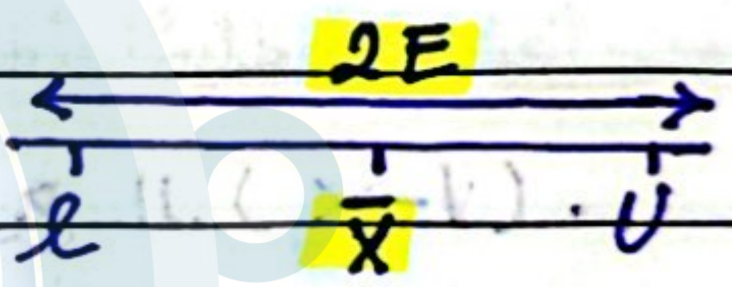
* بقدر بجيب الـ length للـ (confidence interval) بطريقتين طرح الـ lower من الـ upper

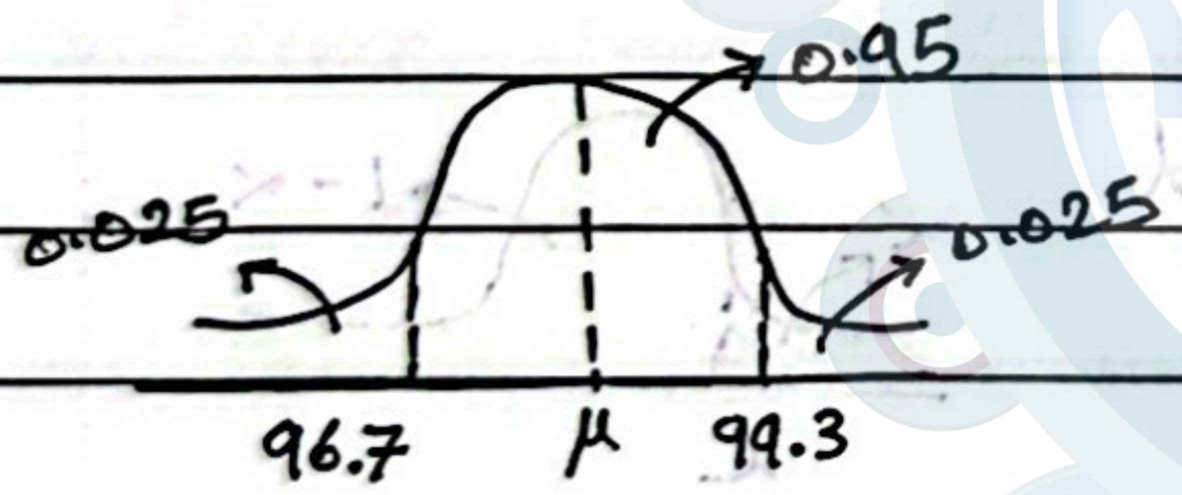
$99.3 - 96.7 = 2.6$ ، أو الـ $2 * \overset{1.3}{\text{error}}$

↓ 2.6   ↓ 2.6

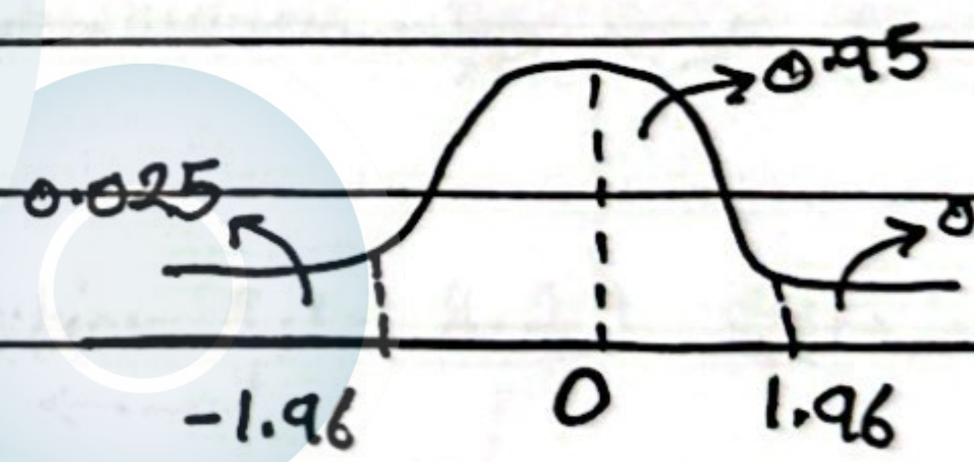
* conclusion: $\boxed{\text{CI Length} = 2E}$ ↳ error

* تعميم





in standard normal distribution →



* الـ CI بقدر أمثلها عالـ normal distri الأصلي أو الـ standard.

* 95% من الـ experaments راح تطلع confidence intervals فيهم الـ population mean or parameter.

معنى 0.95

* يعني بهاد المثال لو أخذنا 100 سامبل راح نحصل على 100 فترة فيهم الـ 95 true mean و 5 will not contain

كلهم إلهم نفس الـ CL ↑

هاد هو الـ relative frequancy interpretation for the CL

* الـ error (E) = $Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ يعتمد على الـ σ والـ n والـ α

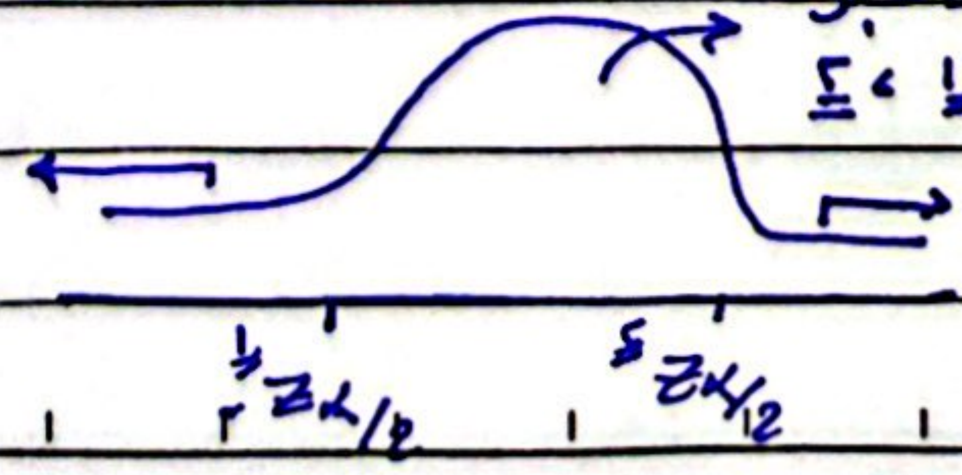
if $\sigma\uparrow \therefore E\uparrow \therefore \text{length}\uparrow$

if $n\uparrow \therefore E\downarrow \therefore \text{length}\downarrow$

if the $CL\uparrow \therefore Z_{\alpha/2}\uparrow \therefore E\uparrow \therefore \text{length}\uparrow$ CI

$\therefore 1-\alpha$

عشان تزيد بدنا نكبّر المساحة ، يعني الـ Z ، بدهم يزيدو لليمين واليسار أكتر



Five Apple

* لو السؤال الماضي نسب 99% بدال 95% :

$\bar{X} = 98$ ، $\sigma = 2$ ، $n = 9$ ، $1-\alpha = 0.99$ ، $\alpha = 0.01$ ، $\alpha/2 = 0.005$

$Z_{\alpha/2} = Z_{0.005} \overset{\text{الجدول}}{=} -2.58$

$E = Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \longrightarrow 2.58 * \frac{2}{\sqrt{9}} = 1.7$ مع ثبات العوامل الأخرى

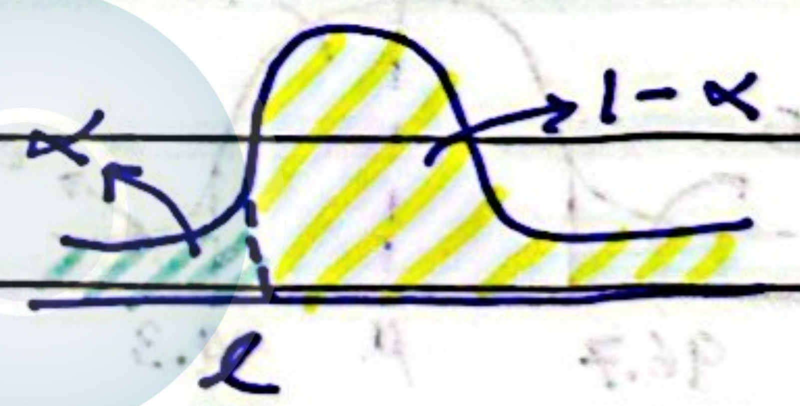
$98 - 1.7 < \mu < 98 + 1.7 \longrightarrow 96.3 < \mu < 99.7$

96.3 — 98 — 99.7 } got wider (more length)

* بس نفحص الـ Parameter بين إشيين بنسميه (2 sided confidence bounds) ، لو بس من جهة وحدة one sided confidence bound . زي كأنه خلينا الـ upper ← ∞ أو الـ lower ← -∞ والجهة التانية خليناها.

* وهون α بتبطل تنقسم ، بتصير الـ area = α خارج الـ (1-α) .

one sided (lower)



$\bar{X} - Z_{\alpha}\frac{\sigma}{\sqrt{n}} < \mu$ (lower)

$\mu < \bar{x} + Z_{\alpha}\frac{\sigma}{\sqrt{n}}$ (upper)

بجيب الحصر هيك

* بدنا نحل نفس المثال إلي فوق بس على lower one sided

$\bar{X} = 98$ ، $\sigma = 2$ ، $n = 9$

$1-\alpha = 0.95$ ، $\alpha = 0.05$

$\bar{X} - Z_{\alpha}\frac{\sigma}{\sqrt{n}} \longrightarrow 98 - 1.65 * \frac{2}{\sqrt{9}} < \mu \longrightarrow 98 - 1.1 < \mu \longrightarrow 96.9 < \mu$

section 8.2

* chose of sample size:

* عشان ما يكون الـ error كبير و الـ CI length كبير، في طريقة لتقليلهم ... منها أنه أتحكم بـ n .

*if we want the 95% CI to be no wider than 2 psi, what sample size is required.

problem 8-10) نفس المسألة لو بدي أضيف هاد الفرع:

the width & length are the same for the interval

means width = 2 psi

$2E = 2$ psi — شو الـ sample size اللازمة

$E = 1$ → ليكون E يساوي واحد

* باختصار، بس خلي n هي موضوع القانون هون ← $E = Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$

$$\boxed{n = \left(Z_{\alpha/2} \frac{\sigma}{E}\right)^2}$$

$Z_{0.025} = 1.96$ → $n = \left(\frac{1.96 \times 2}{1}\right)^2$

$= 15.37$

so $\boxed{16}$ لازم (n) ما تكون أعشار، دايمًا بنعمل rounding up وليس down .

* confidence interval on the mean of a normal distribution, varience unknown:

* الفرق بين (section 8.1 , 8.2)

$$\bar{X} - E < \mu < \bar{X} + E \quad \begin{cases} E = Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \\ E = T_{\alpha/2, n-1} \frac{S}{\sqrt{n}} \end{cases}$$

S: sample SD

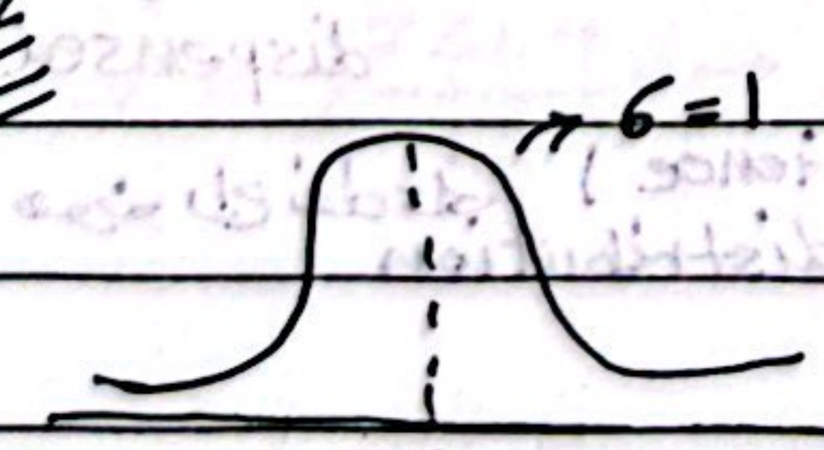
$n-1$: degree of freedom

$t_{\alpha/2, n-1}$ (T percentile)

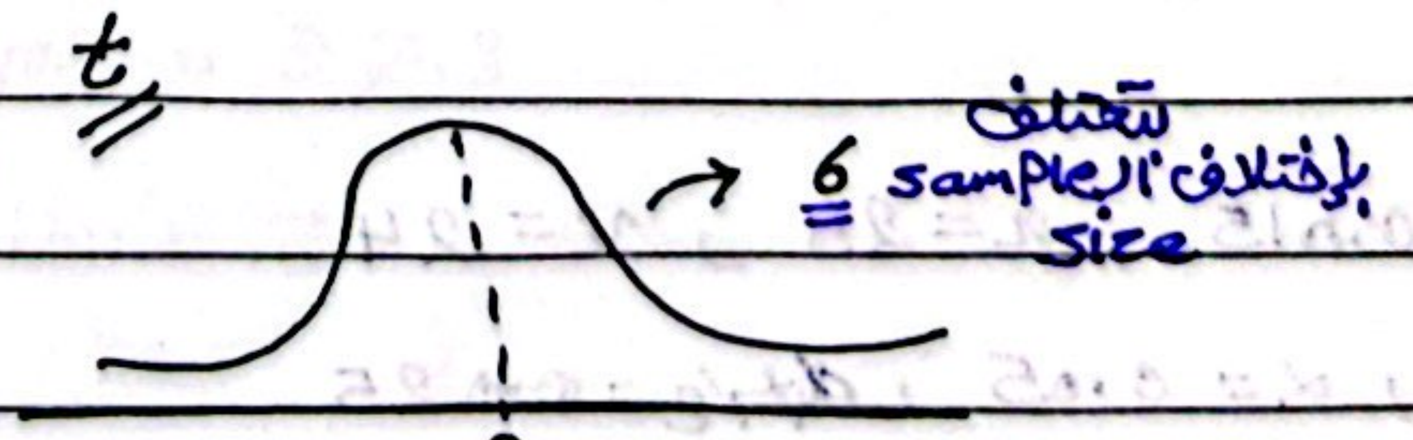
* بنستخدم الـ sample SD لإنه صار الـ varience (unknown) الـ population .

* مقارنة بين الـ Z والـ t : التنين (normally distributed)

Z



t



$t_{\alpha/2, n-1}$

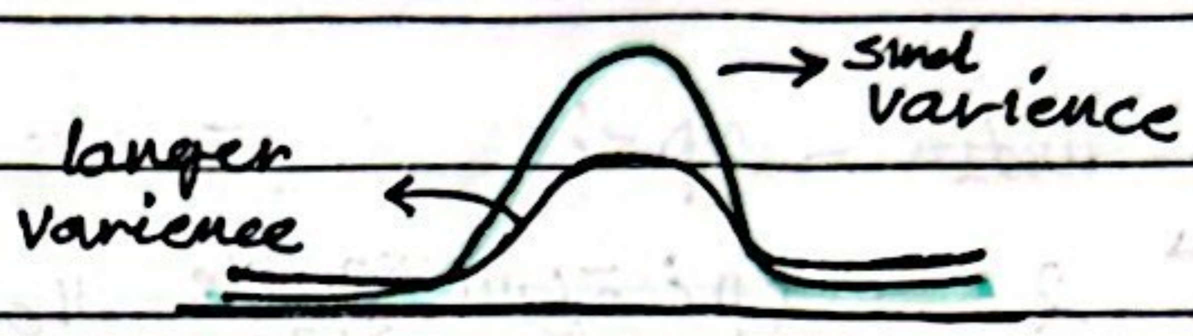
بتأثر بشكل ← الـ distribution وبقيمة الـ Percentile

← نفس إشي بـ Z وبتتحكم بقيمة الـ percentile

* degree of freedom : $\boxed{K = n-1}$
↳ أي الحرية باختيار كل elements السامبل
ما عدا آخر حد.

* إذا قل الـ SD بالـ distri، بزيد ارتفاع الـ curve وبقل عالاطراف عند الـ x axis.
* الـ SD بتعتمد عالـ degree of freedom بالـ t

$$\boxed{\sigma^2 = \frac{K}{K-2}}$$

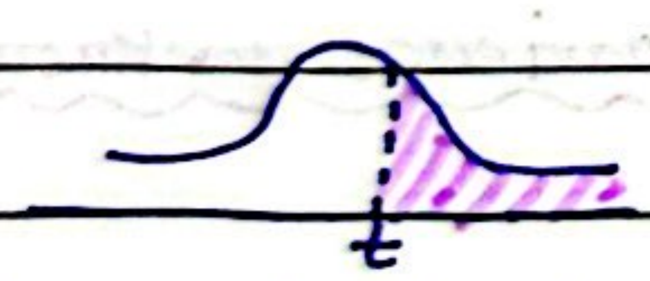


* كل ما زادت الـ degree of freedom كل ما قرب الـ Z من الـ t.
* لنطلع قيمة الـ percentile كاملة لازم نطلع قيمة الـ area والـ dof (V) ← $t_{\alpha/2,\, n-1}$

area ↲ DOF ↲

* أما الـ V تروح لـ ∞ بيصير الـ t = الـ Z.
* مثلاً $t_{0.025,20} \rightarrow 2.086$ ، $1.96 = Z_{0.025}$

أما لو $t_{0.025,\infty} \rightarrow 1.96$



* بالـ t مش مهمة الإشارات، بناخذ قيمة مطلقة، والرقم إلي بطلع بالجدول يكون المساحة على يمين الـ t.

* Ex: a postmix beverage machine is adjusted to release a certain amount of syrup into a chamber where it's mixed with carbonated water. A random sample of 25 beverages was found to have a mean syrup content of $\bar{x} = 1.10$ fluid ounce and a SD of $s = 0.015$ fluid ounce. Find a 95% CI on the mean volume of syrup dispensed.

هون راح نعمل t distribution (unknown varience)

$\bar{x} = 1.1$, $s = 0.015$, $n = 25$, $v = 24$

$1-\alpha = 0.95$, $\alpha = 0.05$, $\alpha/2 = 0.025$

$\bar{x} - E < \mu < \bar{x} + E$

$E = t_{0.025, 24} \frac{s}{\sqrt{n}} \rightarrow 2.064 * \frac{0.015}{\sqrt{25}} = 0.006$

$1.1 - 0.006 < \mu < 1.1 + 0.006 \rightarrow \boxed{1.094 < \mu < 1.106}$

* if one sided CI ?

$\bar{X} - t_{\alpha, n-1} \frac{s}{\sqrt{n}} < \mu$: Lower sided لو طلب السؤال على *

$E = t_{0.05, 24} * \frac{0.015}{\sqrt{25}} \rightarrow 1.711 * \frac{0.015}{\sqrt{25}} = 0.005$

$1.1 - 0.005 < \mu \rightarrow 1.095 < \mu$

* هون ما في choose of sample size

لإنه ال S و ال V بيعتمدو على ال n

Ex: A random sample has been taken from a normal dist, outputs as follow:

variable	N	mean	SE mean	SD	var	sum
X	10	?	0.507	1.605	?	251.848

$\bar{X} = \frac{251.848}{10}$, $var = (1.605)^2$

if

variable	N	(avg) mean	SE mean (standard error)	SD	var	sum
X	?	?	1.58	6.11	?	751.40

مهم

$\boxed{SE_{mean} = \frac{SD}{\sqrt{n}}}$: $1.58 = \frac{6.11}{\sqrt{n}}$ $n = 15$

$\bar{X} = \frac{751.4}{15} \rightarrow 50.1$ $var = 37.3$

$n=6$, sample : 16.8 , 17.2 , 17.4 , 16.9 , 16.5 , 17.1

* ال var غير معروف ، العاد على $\underline{t}$

$\bar{X} - E < \mu < \bar{X} + E$

$E = t_{\alpha/2, n-1} \frac{S}{\sqrt{n}}$

مجموع عددهم

$\bar{X} = 16.98$, $K = 5$, $\boxed{\text{sample SD } (S) = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}} = 0.3504$

$1-\alpha = 99\%$ (من السؤال) [calculate a 99% CI on μ]

$\alpha = 0.01$

$\alpha/2 = 0.005 \rightarrow t_{0.005, 5} = 4.032 \rightarrow E = 4.032 * \frac{0.3504}{\sqrt{6}}$

$= 0.57$

$15.83 < \mu < 16.97$

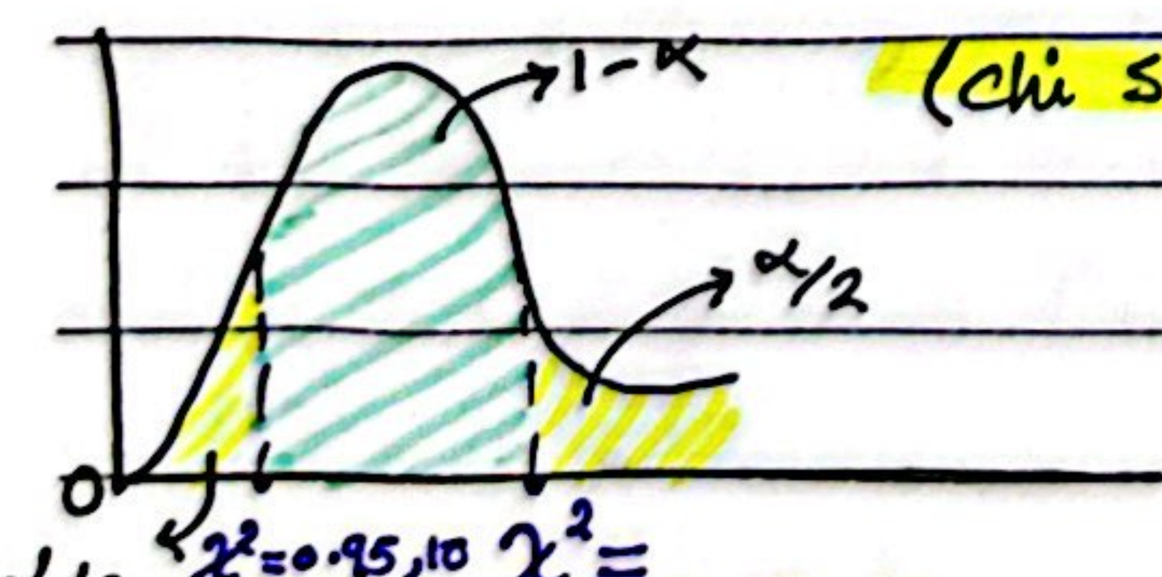
section
* 8.3 : CI on var & SD :

* هنا عنا dist جديد اسمه كاي (χ^2) squared بعتمد عال K ← area

* كل ما زادت K بقترب شكل ال dist من ال normal (Bell shape) .

و كل ما زادت K بتزيد ال symmetry for the distribution

* ما إله أي قيم سالبة لإنه var (σ^2) ، لو μ كاديج . بلش من الصفر (ذيله رايح اليمين) .



(chi square)

← هاد ال Notation معناه المساحة عاليمين

* مثال :

* نفترض ال CI = 90% :

$1-\alpha = 0.9$, $\alpha = 0.1$, $\alpha/2 = 0.05$

* نفترض K = 10

$\chi^2_{0.05, 10} \rightarrow 18.31$ ↳ قيمة كاي على ال dist ال area على يمينها = $\frac{1}{\cdot}$

$\chi^2_{0.95, 10} \rightarrow 3.94$

$\chi^2_{1-\alpha/2, n-1}$ $\chi^2_{\alpha/2, n-1}$

lower upper

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

* عشان الـ percentile موجود بالمقام ، ولازم الـ lower يكون أقل من الـ upper ، بنعوض بالـ lower القيمة الكبيرة وبالـ upper القيمة الصغيرة.
The percentile

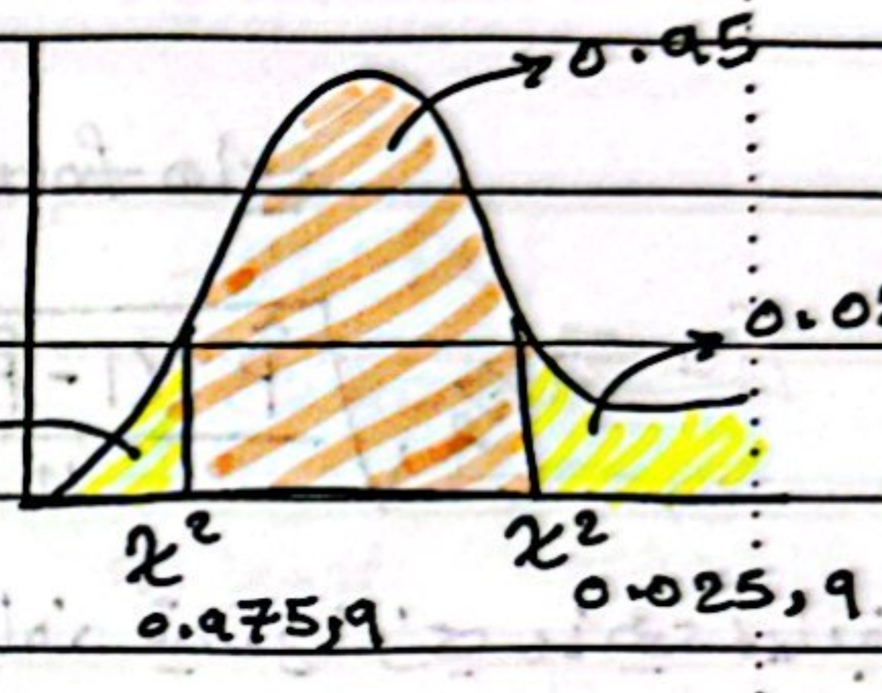
* This is the formula we use to calculate the CI on χ^2 distribution.

* problem 8-50 : $n = 10$, $s = 4.8$, calculate a 95% 2 sided CI for σ.

$$\frac{(10-1)(4.8)^2}{\chi^2_{0.025,9}} \le \sigma^2 \le \frac{(10-1)(4.8)^2}{\chi^2_{1-0.025,9}}$$

$$\frac{207.36}{19.02} \le \sigma^2 \le \frac{207.36}{2.70}$$

$$3.3 \le \sigma \le 8.76$$



* 8.4 : CI on the proportion :

نسبة من الـ data اللي بتنتمي لفئة معينة.

$\bar{X} \rightarrow$ Normal $\begin{cases} Z \\ t \end{cases}$
$S^2 \rightarrow$ chi squared
$\hat{p} \rightarrow$ normal (Z)
} previous sections

* الـ $\hat{p}$ بوصف الـ RV والـ distr والـ μ و σ^2

$\boxed{\mu_{\hat{p}} = P}$ → pop proportion

$\boxed{\sigma_{\hat{p}} = \sqrt{\frac{P(1-P)}{n}}}$

SD or standard error in the proportion ($\hat{p}$)

* أكيد هون كل ما قل n بزيد $\sigma_{\hat{p}}$

$$Z = \frac{\hat{P} - P}{\sigma_{\hat{p}}} \rightarrow \boxed{\frac{\hat{P} - P}{\sqrt{\frac{P(1-P)}{n}}}}$$

For CI →

$$\boxed{\hat{p} - Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < P < \hat{p} + Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \equiv \hat{p} - E < P < \hat{p} + E$$

$$\boxed{E : Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$

* ما بنعرف p فبنستخدم $\hat{p}$ بدالها
(the point estimator)

Five Apple

* problem 8-53 : (بالكتاب)

$$\hat{p} = \frac{13 \rightarrow \text{عدد ال defective}}{300 \rightarrow \text{العدد الكلي}} = 0.043$$

$n = 300$

$1-\alpha = 0.95 \qquad \alpha = 0.05 \qquad \alpha/2 = 0.025 \qquad Z_{0.025} = 1.96$

$$E = Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \longrightarrow 1.96\sqrt{\frac{\frac{13}{300} * \frac{287}{300}}{300}} \longrightarrow 0.023 \quad \text{Approx}$$

* من باب ال checking لل $\hat{p}$ عبارة عن

$\hat{p} - E < p < \hat{p} + E \rightarrow 0.02 < p < 0.066$ شي ، يعني ال lower ممنوع سالب وال upper ممنوع أكبر من واحد .

* عشان أخلي ال E أقل ما يمكن ، بقدر أختار ال n ، فـ عن طريق ال formula هاي (sample size)

$$E = Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \longrightarrow \boxed{n = \left(\frac{Z_{\alpha/2}}{E}\right)^2 \hat{p}(1-\hat{p})}$$

هلأ أنا لو أزيد ال n بقل ال E ولكن بعتمد برضه عال $\hat{p}$ ، متغيرة ، فـ غيا باخذ قيمتها زي ما هي ، يا بفرض أسوء قيمة ممكن $\hat{p}$ تكون ، إلي هي (0.5)

هاي القيمة بتخلي $\hat{p}$ أكبر ما يمكن ، المقدار هاد كامل ل $\hat{p}(1-\hat{p})$ ، يعني أكبر n ممكن تكون minimum

* بستخدم قيمة $\hat{p} = 0.5$ لما يحكيلي بالسؤال (at least) or (regard less) ، إذا ما كان حاكي هيك بستخدم القيمة إلي هو معطيني ياها .

* من وين أجت قيمة 0.5 ؟ من اختبار المشتقة الأولى (القيم القصوى) .

$f(x) = x(1-x)$ $\qquad\qquad 1-2x = 0$

$f(x) = x - x^2$ $\qquad\qquad 2x = 1$

$f'(x) = 1 - 2x$ $\qquad\qquad x = 0.5$

* القيمة القصوى لـ $\hat{p}(1-\hat{p}) = 0.5 * 0.5 = 0.25$

← ما في أي قيمة بتخلي هاد المقدار يزيد عن 0.25

* problem 8-56 : (بالكتاب) $\hat{p} = \frac{823}{1000}$ $\qquad$ b) $n = \left(\frac{1.96}{0.03}\right)^2 * 0.823 * 0.177$

$n = 622$

c) $n = \left(\frac{Z_{\alpha/2}}{E}\right)^2 0.25$

$= \left(\frac{1.96}{0.03}\right)^2 * 0.25$

$= 1068$ (rounding up لفوق)

Review for chapter 8

mean with known variance

$\bar{X} - E < \mu < \bar{X} + E$

$\boxed{E = Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}}$

- No Dof
- choice of n (round up)

$\boxed{\text{Sum of square} = (n-1) * \sigma^2}$

mean with unknown σ^2

$\bar{X} - E < \mu < \bar{X} + E$

$\boxed{E = t_{\alpha/2, n-1} \frac{s}{\sqrt{n}}}$

- Dof ✓
- No choice of n
- $\boxed{SE = \frac{Stdev}{\sqrt{n}}}$

*إذا one sided بس α

CI on σ^2

$\boxed{\frac{(n-1)s^2}{\chi^2_{\alpha/2, n-1}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\alpha/2, n-1}}}$

- Dof ✓
- No choice of n

proportion

$\hat{P} - E < P < \hat{P} + E$

$\boxed{E = Z_{\alpha/2} \sqrt{\frac{\hat{P}(1-\hat{P})}{n}}}$

- No Dof
- choice of n: at least / regardless

افرض $0.5 = \hat{P}$

- $\boxed{SE = \sqrt{\frac{\hat{P}(1-\hat{P})}{n}}}$

chapter 9
"Hypothesis testing"

* Two opposite statements about a pop parameter.

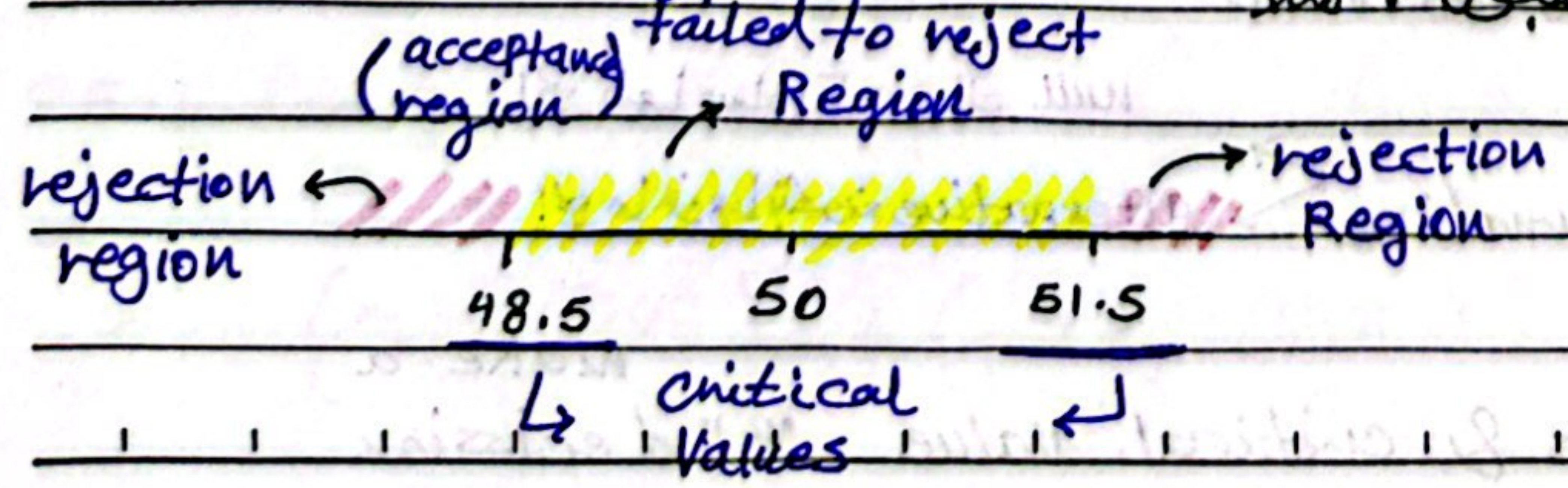
*مثلاً:

$\mu = 50$
$\mu \neq 50$

الـ 2 اسمهم فرضيات / هدول

* إلي فيها مساواة بنسميها null ورمزها H_0
* إلي ما فيها مساواة بنسميها alternative ورمزها H_1
* القيمة إلي بتساويها الـ 50 مثلاً بنسميها [hypothesised value] ورمزها (مسب) الـ parameter وبنعطيها μ_0.

* لو أخدت سامبل وحسبت $\bar{X}$، طلعت القيمة = 42، القيمة كتير بعيدة عن الـ 50، يعني الـ alternative صح، لو $\bar{X} = 50.5$ برفض H_1 هي الصح ولكن في عنا قيمة إسمها [critical value] بتعطيني range بكم بقدر أبعد عن الـ 50 ويكون مقبول. مثلاً:



Five Apple

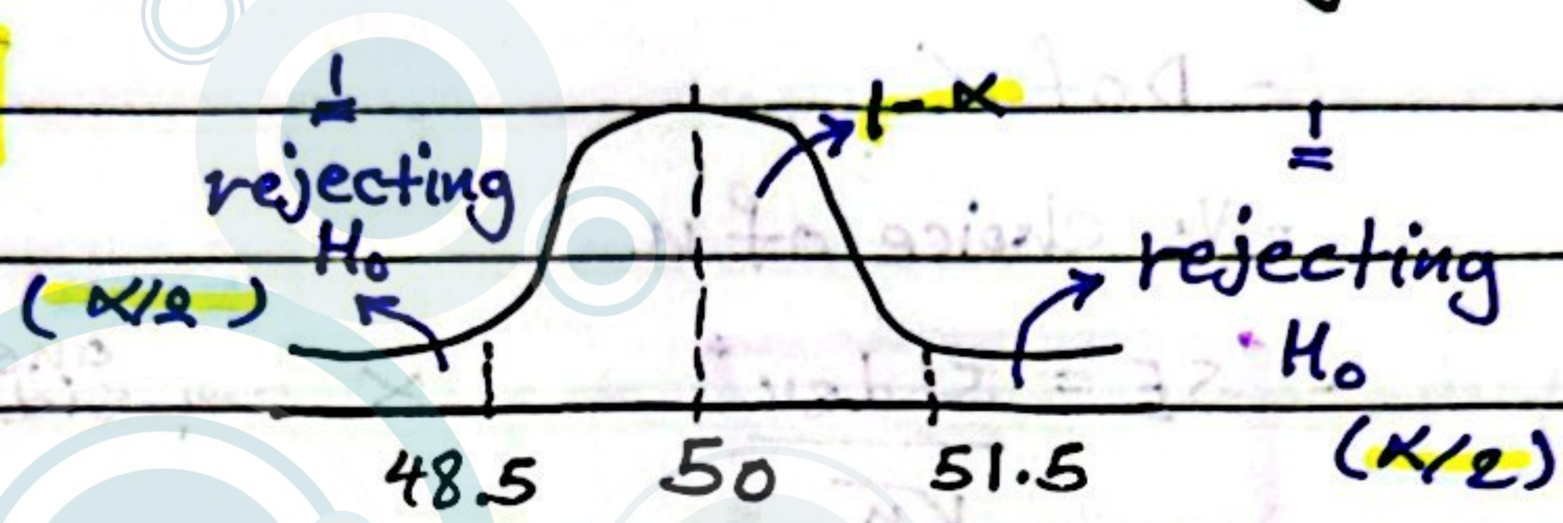
* كلمة accept غلط أحكيها ، لازم (fail to reject)

* بالنسبة للمسألة بتعت $\mu = 50$ / $\mu \neq 50$ ، ف عندي دليل قوي ضد H_0 لأرفضها H_0 false ∴ reject H_0.

* هلأ إحنا عنا pop أخذنا منه sample وحسبنا $\bar{X}$ ، الـ distr للـ $\bar{X}$ هو الـ Normal.

* ما بيجوز بالضرورة H_0 تكون 100% غلط ؟ فعليًا لأ ، لأنّا حسبت RV إلي هو $\bar{X}$ ، غالبًا H_0 غلط بس لو كان في احتمالية حدوث (probability of accurance) إنه أنا رفضت H_0 وهي صح ، بسميها error اسمه type I error

The $\boxed{P(\text{type I error}) = \alpha}$
significance Level



1) مجموعهم = α ∴ the probability of rejecting H_0

أكثر قيمة مكررة لـ α ← $\alpha = 0.05$

* الـ critical values بتحدد قيمتها على أساس المساحة (الإحتمال) على يمينها ويسارها ؛ تكون كل وحدة $\alpha/2$.

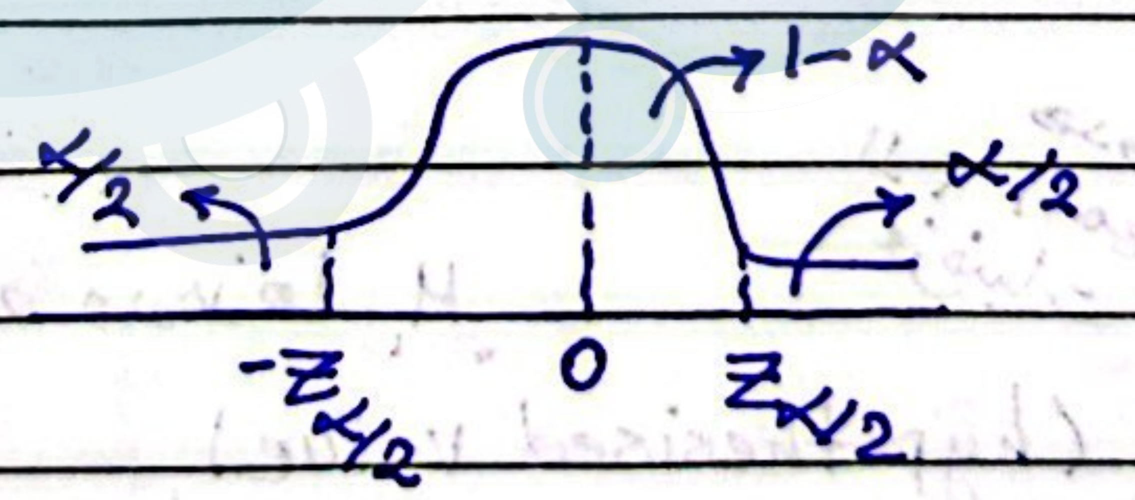
* بعدما أحددهم بحسب $\bar{X}$ وبشوف وين موجود عال dist.
↲ اسمها (test statistics)

* كثير صعب أحدد مين القيمة إلي عيمينها $\alpha/2$ وإلي عيسارها $\alpha/2$ ، الحل لهاي المشكلة هو الـ Standardization.

$$\boxed{Z_0 = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}}$$

نفس نتعامل مع Standard normal distribution →

↳ القيمة إلي حسبها بقرر إذا fail or reject



* بصير هون الـ Z_0 هو الـ test statistic تبعنا.

* كيف نحل أي سؤال الـ hypothesis testing (analysis) ؟

[The procesure]:

mean with known σ^2
mean with unknown σ^2

1) Determine the parameter of interest

variance
proportion

2) Determine the null & alternative
- إلي فيها مساواة هي الـ null
- إلي ما فيها بتكون alternative

3) Calculate test statistic & critical value 4) make a decision

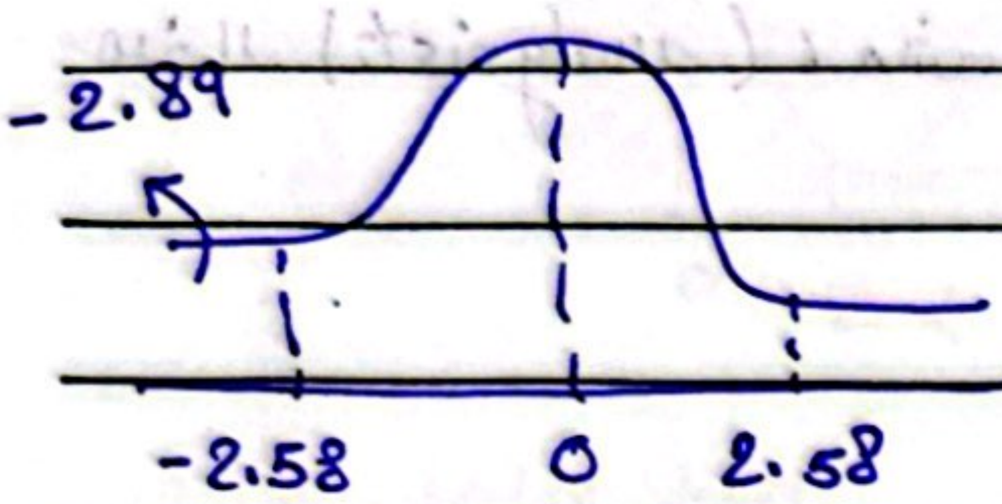
"9.2 : test on the mean, the σ^2 is known:"

*** problem 9-44 :** كتاب

$H_0 : \mu_0 = 3500$ $\alpha = 0.01$ $Z_0 = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \rightarrow \frac{3450-3500}{60/\sqrt{12}}$

$H_1 : \mu_0 \neq 3500$ $\alpha/2 = 0.005$ $= -2.89$

$Z_{\alpha/2} = Z_{0.005} = -2.58$



reject H_0

* لما failed to reject H_0 بس H_0 غلط فبكون error → type II . (H_0 is false)

(عكس type I)

The p (type 2 error) = β

* عشان أحسب β لازم أعرف الـ true value of the mean ، ونحسب إشي اسمه δ (دلتا) ، إلي بتساوي

$\delta = \mu - \mu_0$

الحقيقي ← ← إلي بالفرضية

H_0 is false { → fail to reject (type II error) ; → reject (power of test) = $1-\beta$ }

مقدار قوة الـ test ك rejecting H_0 ← (هاد قرار منطقي وصح)

* as $\alpha \uparrow$, $\beta \downarrow$ as a probability

* لو ازيد n بتقلو هدول ↖

* if $\delta \uparrow$, $\beta \downarrow$

بزيد الفرق بين الـ actual value و hypothesized value 50

* هلأ هاي معناها $Z_{0.05}$ ← هاي الـ area (0.05) .

بدي قيمة Z إلي على يمينها area

أما $\phi(1.5)$ ← الـ area على يسار الـ 1.5 .

← قيمة Z

"* Sol for C in problem 9-44:"

$$\beta = \phi\left(Z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma}\right) - \phi\left(-Z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma}\right)$$

$\delta = \mu - \mu_0$
$= 3470 - 3500$
$= -30$ or 30

$= \phi\left(2.58 - \frac{30\sqrt{12}}{60}\right) - \phi\left(-2.58 - \frac{30\sqrt{12}}{60}\right)$

كالحالتين بتطلع نفس قيمة β

$\beta = \phi(0.85) - \phi(-4.31)$

$0.802 - 0 \qquad \beta = 0.802$

معلومة: أي قيمة $\phi(-4)$ وبعد (لأقل) = صفر
و $\phi(4)$ وبعد = 1

هاد الرقم معناه إنه راح يطلع 80% ايرور II من التجارب إلي راح أعملها. ← لـ طبعًا هاد error كبير، عشان هيك لازم نرفع الـ sample size (n).

معلومة: α is externally determined.
من الـ (analysist)، مش أنا بجطها، بتكون بالسؤال.

"sol for D in problem 9-44"

power = 0.8 , $\beta = 0.2$ $\delta = 3470 - 3500 = -30$ or 30

$$n = \frac{(Z_{\alpha/2} + Z_\beta)^2 \sigma^2}{\delta^2}$$

* إذا أخذت Z (+) باخذ Z (+) والعكس صحيح، ما بعمل آخذ وحدة (+) و وحدة (-)

$$= \frac{(2.58 + 0.84)^2 (60)^2}{30^2} = 47$$

* Excel (lamees): ways to organize data:

1) Pivot , 2) histogram , 3) central theorem

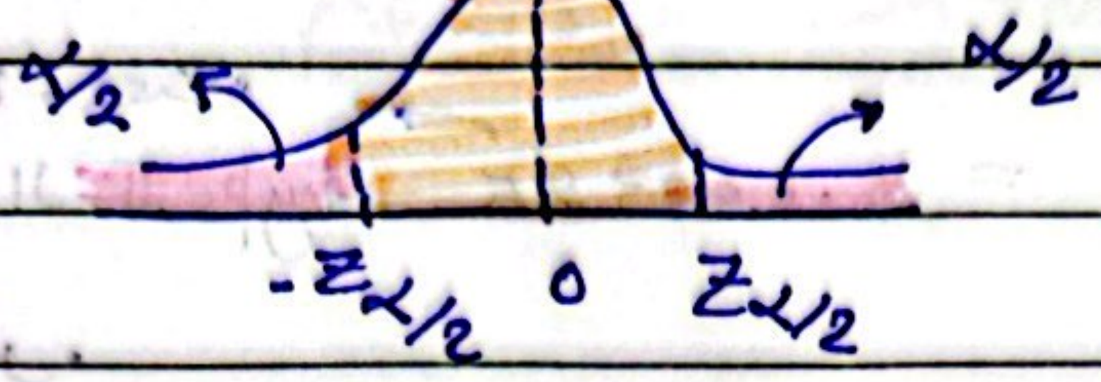
* Back to kaizen:

if $H_0: \mu = 50$
$H_1: \mu \neq 50$ → two sided

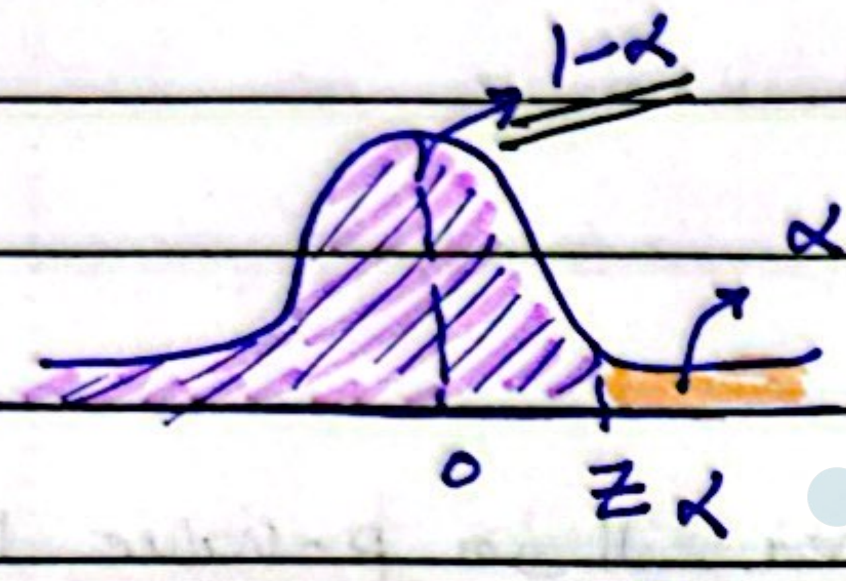
if $\mu > 50$
$\mu < 50$ → one sided

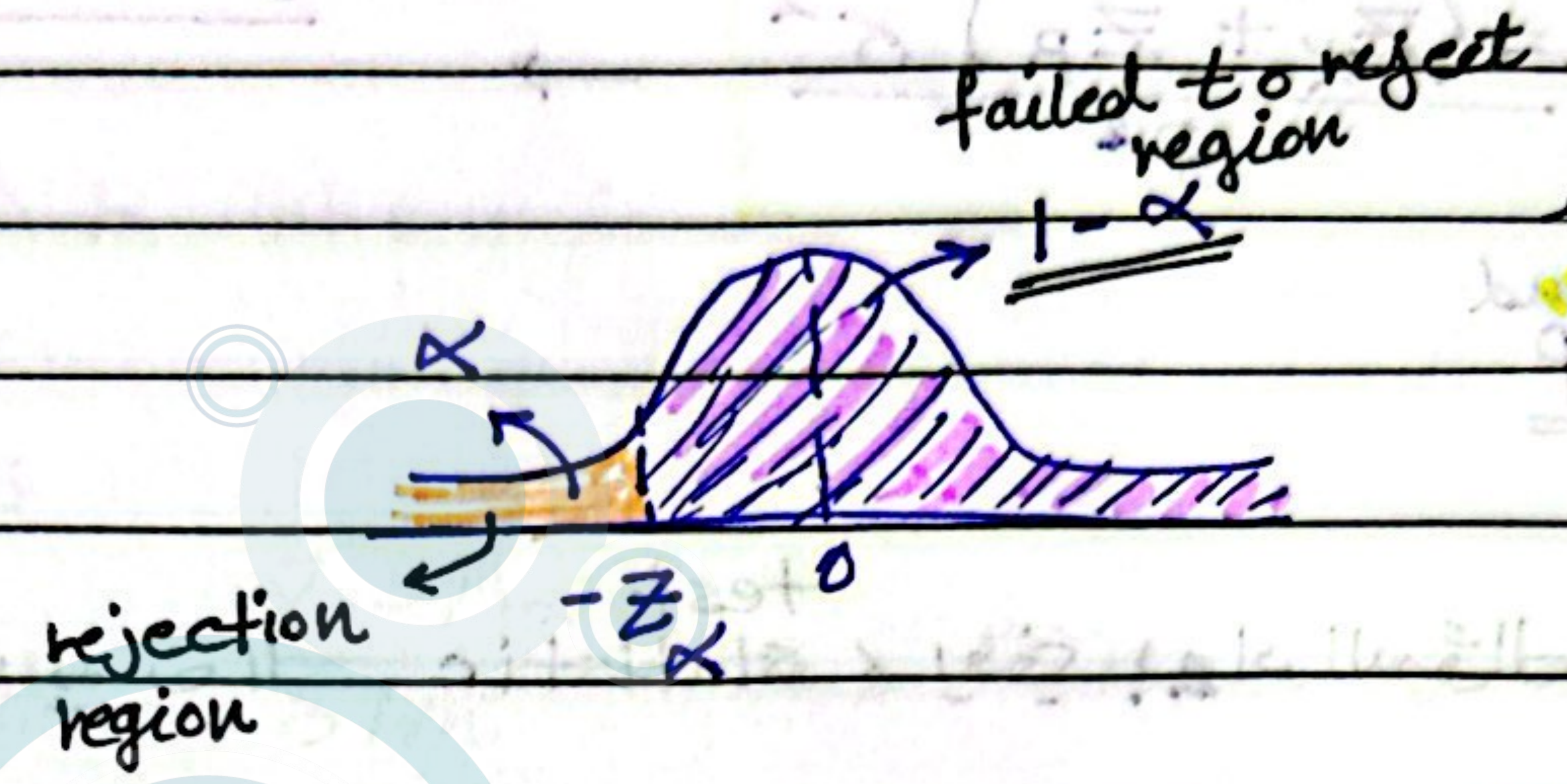
* الـ null هي إلي فيها مساواة، فـ إلي بعملها إعادة كتابة هيك ($\mu = 50$)، ضمنيًا فهي رح (ظاهريًا)
* لما نعطي بدال الـ alternative إنه $\mu \neq 50$، معناه μ يا أكبر يا أصغر من 50، عشان هيك 2-sided
* أمّا لما الـ alternative تكون ∴ $\mu > 50$ one sided - upper -
∴ $\mu < 50$ one sided - lower -

α in 2-sided :



α in one sided :





lower-tailed test

"upper-tailed test"

هاد بال 2-sided كنا ناخد قيمة الـ موجبة والسالبة ، أما بال one-sided إذا lower فبناخد السالب وإذا upper بناخد الموجب.

الـ test statistic ← $\boxed{Z_0 = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}}$

بكتابي

Problem 9-43 :

exceeds 40 hours ∴ $\mu > 40 \longrightarrow \mu > 40 \quad (H_1)$

$\mu \le 40 \longrightarrow \mu = 40 \quad (H_0)$

upper-tailed test
↳ alternative بتحدد من الـ

$\alpha = 0.05$

$Z_{0.05} = -1.65$ باخذ الموجب منها طبعاً

$Z_0 = \frac{40.5 - 40}{1.25 / \sqrt{10}} = 1.26$ ↳ The test statistic came in the "fail to reject region"

c) $\mu = 42$, $\delta = 42 - 40 = 2$

حسب في معادلة الـ β بالـ 2-sided :

$\boxed{\beta = \phi\left[Z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma}\right] - \phi\left[-Z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma}\right]}$

$\boxed{\text{upper} = \beta = \phi\left[Z_{\alpha} - \frac{\delta\sqrt{n}}{\sigma}\right]}$

إذا one side فهي معادلة الـ

$\boxed{\text{lower} = 1 - \phi\left[-Z_{\alpha} - \frac{\delta\sqrt{n}}{\sigma}\right]}$

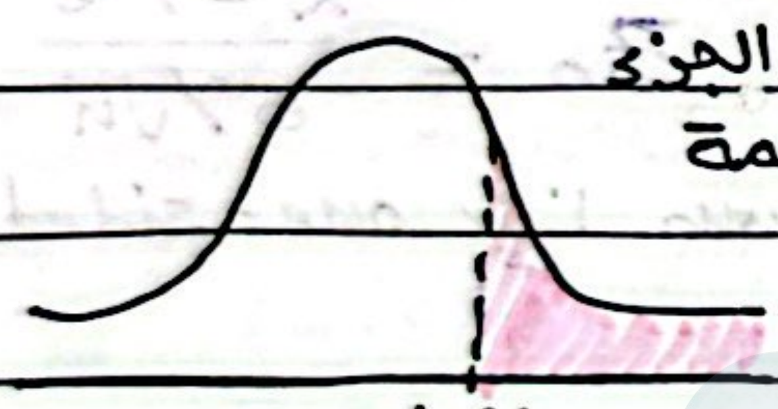
so $\beta = \phi\left[Z_\alpha - \frac{\delta\sqrt{n}}{\sigma}\right] \rightarrow \phi\left(1.65 - \frac{2*\sqrt{10}}{1.25}\right)$

$= \phi(-3.41)$ معنى هاد ال notation إنه بدنا ال area على يسار (-3.41) $= 0.000325$ هاد إشي منيح ، معناه إنه نسبة حدوث ال type II error قليلة كثير.

فرع for D: $\boxed{n = \frac{(Z_\alpha + Z_\beta)^2 \sigma^2}{\delta^2}} \rightarrow \frac{(1.65+1.28)^2 (1.25)^2}{4^2} = 0.8$

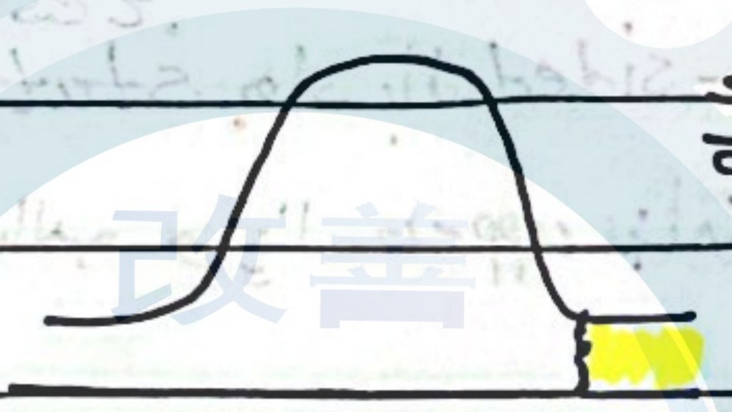
$\approx n = 1$ round up

فرع for B: * ال p-value هو ال area على يمين ال test statistic ، يعني بهاد السؤال عنا الرسمة هيك:



$Z_0 = 1.26$

أما هاي المساحة بتمثل ال p-value



هاد الجزء من الرسمة

$Z_\alpha = 1.65$

هاي المساحة بتمثل α

هاد الجزء من الرسمة

1.26 1.65

وإذا كان lower فهو على يسار ال (test statistic) ، وإذا 2-sided فبنقسم.

* الهدف من ال p-value هو إنه أقارنه ب α ؛ إذا ← "fail to reject region": $(p > \alpha)$

وإذا $(p < \alpha)$ ∴ "reject region"

ومنه بعرف قديه بعيدة عن α.

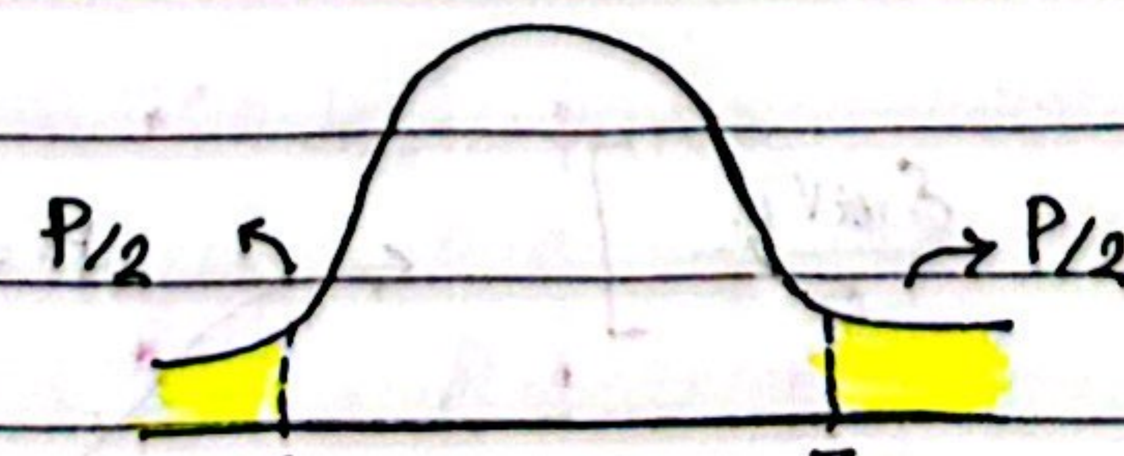
∴ هون ال [p-value] قيمتها = $1 - Z_0 = 0.104$ لإنه الجدول بعطي القيم إلي عاليسار.

$p > \alpha$ ∴ fail to reject

* p-value: the smallest level of significance that would lead to reject H_0.

أقل قيمة ل α إلي بتأدي ال rejection

* إذا 2-sided



Total $p = 2[1-\phi(|Z_0|)]$

$\mu \geq 2.5$ ← the mean at least = 2.5 (مثلاً)

$\mu \leq 2.5$ ← the mean at most = 2.5

* معنى

* معنى → more than ($>$)

less than ($<$)

most of (>0.5) (أغلبية) → أكثر من النص

*9.3

* test on mean when σ^2 is unknown:

→ الفرق عن قبل إنه t بدال Z و s بدال σ

* our test statistic $\boxed{T_0 = \dfrac{\bar{X} - \mu_0}{S/\sqrt{n}}}$, $\nu = n-1$ (dof)

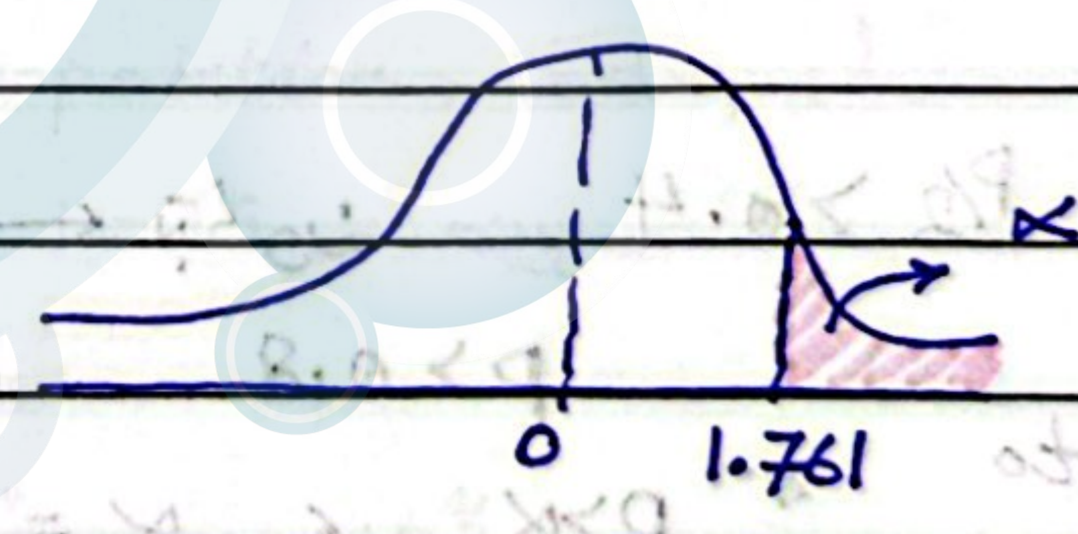
* two tailed test → $t_{\alpha/2,\, n-1}$
* one tailed test → $t_{\alpha,\, n-1}$

* example (9-6): Golf club design

$n = 15$, $\nu = 14$, $\bar{X} = 0.83725$, $S = 0.02456$, $\alpha = 0.05$

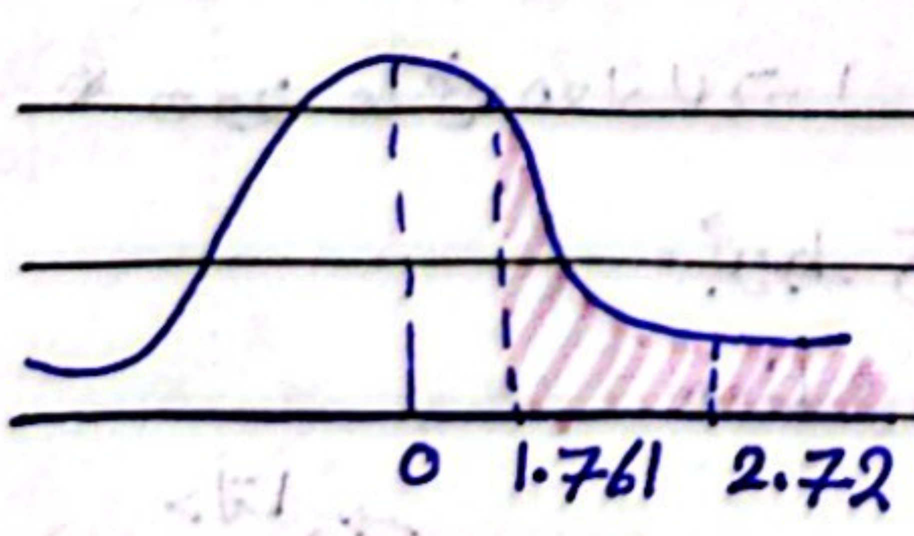
$\mu > 0.82$ → $H_1: \mu > 0.82$] → so upper test

$\mu < 0.82$ $H_0: \mu = 0.82$



$t_{\alpha, n-1} \rightarrow t_{0.05, 14} = 1.761$

$$T_0 = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} = \frac{0.83725 - 0.82}{0.02456/\sqrt{15}} = 2.72$$



so reject H_0

* in p-value?

* بدي أدور على 2.72 بالجدول جوا، بنفس السطر الـ dof = 14

* ما راح نلاقيه نفسه، بنلاقيه بين قيمتين، بنحصر الـ ← test statistic بين قيمتين، يعني بنحصر الـ p-value بين قيمتين (الـ p-value بالجدول هو سطر قيم α)

$0.005 < p < 0.01$

↳ كهاد النشاس ال

$p < \alpha$

So reject

بالكتاب

* Problem 9-58 :

$\bar{X} = 22.496$

$S = 0.378$

أنا مستنتج من ال sample data

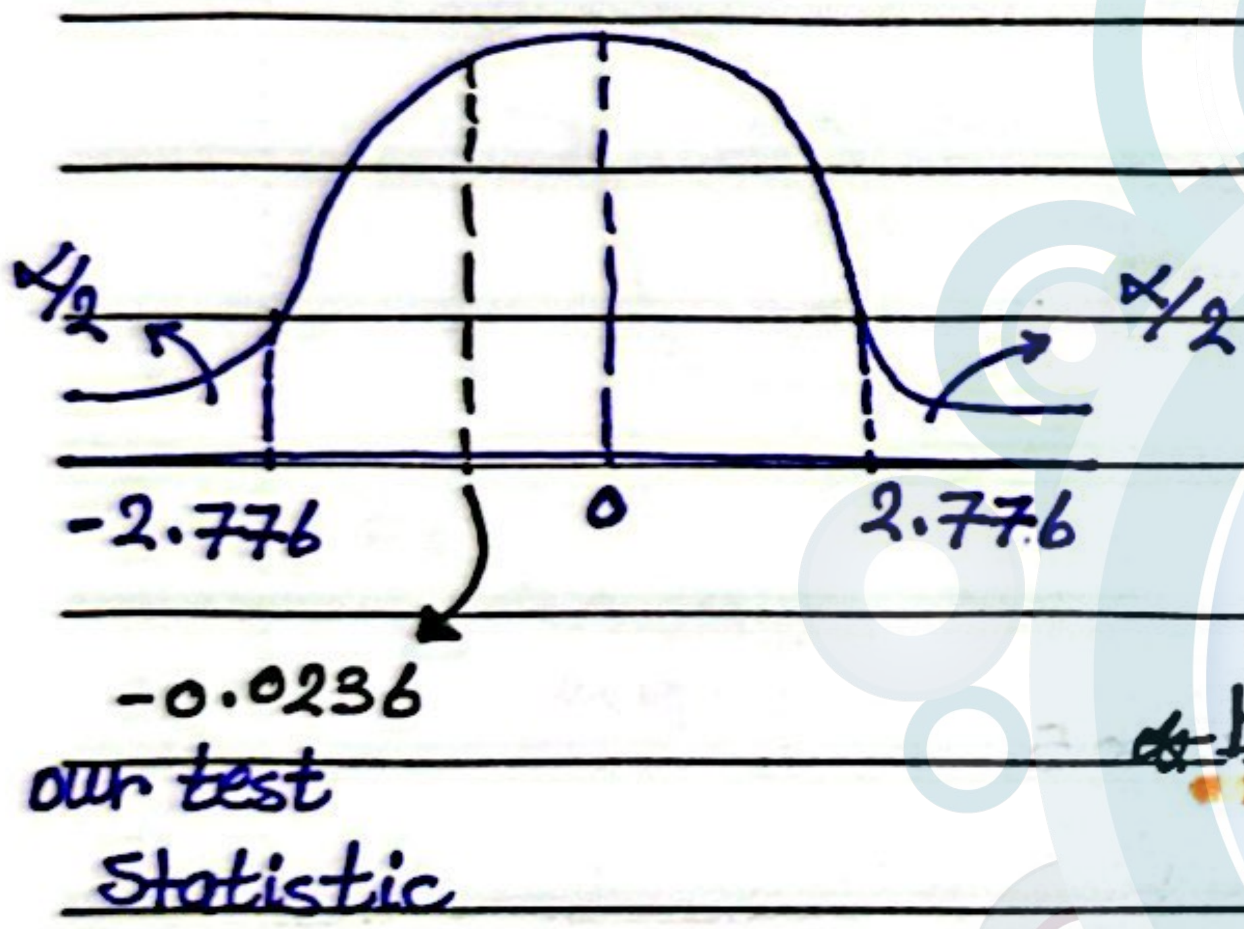
$n = 5$, $v = 4$, $\alpha = 0.05$,

$H_0 : \mu = 22.5$

$H_1 : \mu \neq 22.5$

so two sided

$t_{0.025, 4} ? \; 2.776$

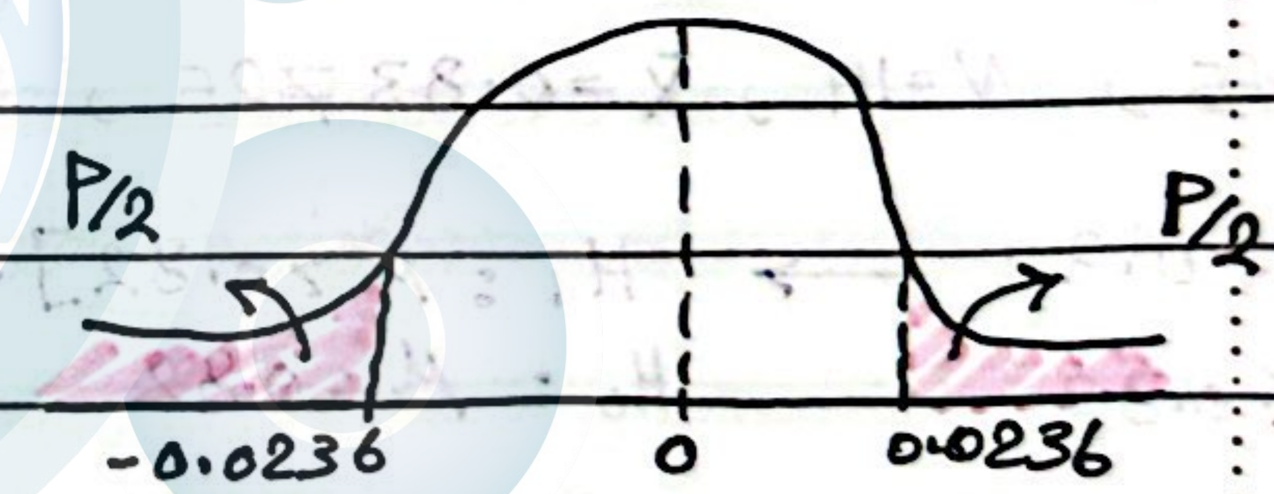


$$T_0 = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

$$= \frac{22.496 - 22.5}{0.378/\sqrt{5}} = -0.0236$$

so fail to reject

* by p-value :



$P/2 > 0.4$ * من جدول t ← بتكون

$p > 0.8$

so fail to reject ← $p > \alpha$ ، $\alpha = 0.05$ بنقارن بـ α ،

* فرع (C) : power??

true mean $\mu = 22.75$, $\delta = 22.75 - 22.5 = 0.25$, $n = 5$

* صون ما في معادلات لـ β ، بنطلعوا من ال OC curves

operating characteristic

↳ ترتبط δ و n و β ببعض ، فلو معنا 2 منهم بعرف منه أطلع الثالث

* what is d?

↳ from OC curve . on the x-axis

$$\boxed{d = \frac{|\mu - \mu_0|}{S}} \text{ or } \boxed{\frac{|\delta|}{S}}$$

دلتا

sample SD

* كيف أتعامل مع الـ oc curves ؟

١) حدد الـ test statistic.

٢) حدد إذا two or one sided.

٣) حدد قيمة α.

* back to the question:

$n=5$, $\delta=0.25$, $d=\frac{|0.25|}{0.378}=0.66$

β ? so $\beta=0.81$

١) حدد d
٢) بقاطعها مع خط الـ n
٣) بشوف كم β (أفقيًا)

& the power $\rightarrow 1-0.81=0.19$

طبعًا في error لأنه تقدير كله

* معي β ومعي δ وبدي n

$\mu=22.75$, $\delta=22.75-22.5=0.25$, power $=0.9$, $\beta=0.1$

$d=\frac{|0.25|}{0.378}=0.66$

* قاطعت d مع β ، طلعت n= 30 تقريبًا

$H_0: \mu=22.5$, $H_1: \mu \neq 22.5$

* فكرة : إذا قاطعت d مع β وما قاطعت أي curve للـ n

أ ← بعمل interpolation شفوي ، باخذ قيمة وسطية بينهم

مثلًا لو بين n=15 و n=10 ، باخذ 13

* section 9.4 : Hypothesis testing on the variance :

$$\boxed{\chi^2_0=\frac{(n-1)S^2}{\sigma_0^2}}$$

test statistic ; S^2 → sample variance ; σ_0^2 → in hypothesis

$\nu = n-1$

* مهم *
* the z table give the area on the left but the t & χ^2 tabels give the area on the right

$\chi^2_{\alpha/2,n-1}$ / $\chi^2_{1-\alpha/2,n-1}$

area ← عاليمين ; area ← عاليمين

↳ upper ; ↳ lower

$\chi^2(1-\alpha)$

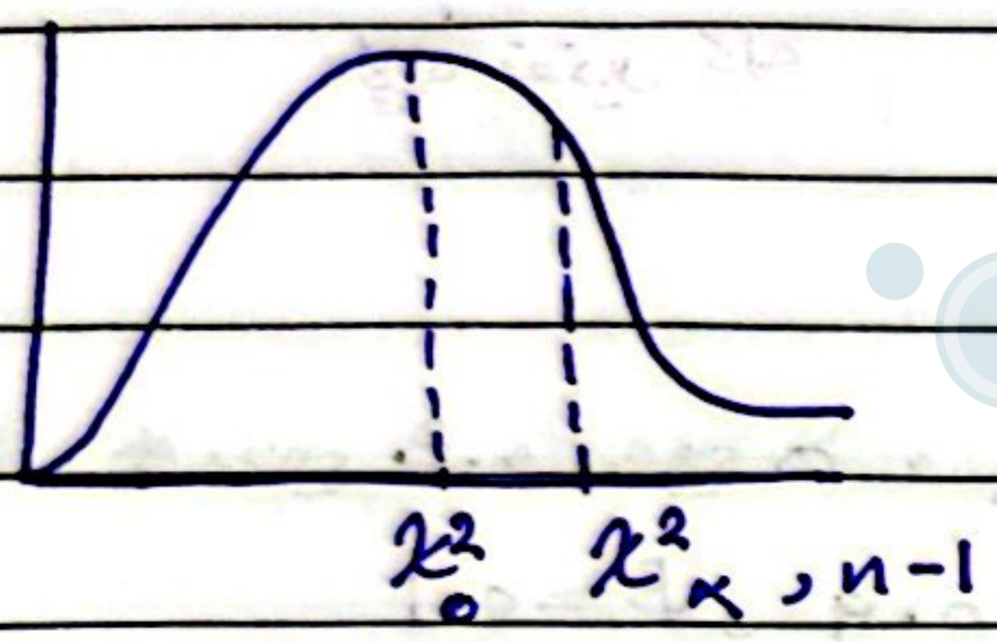
Five Apple

* example 9-8 : automated filling

$\sigma^2 > 0.01$ (exceeds) $\qquad H_1: \sigma^2 > 0.01 \rightarrow$ upper

$\sigma^2 \leq 0.01$ $\qquad H_0: \sigma^2 = 0.01$

$n=20$, $v=19$, $S^2 = 0.0153$, $\alpha = 0.05$

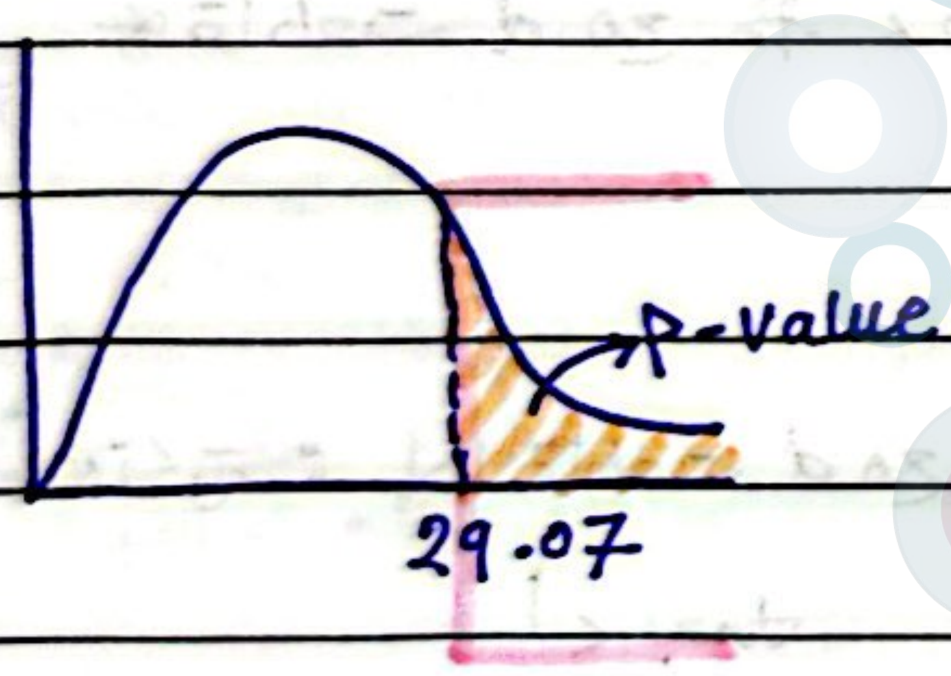


$$\chi^2_{0.05,19} = 30.14$$

$$\chi^2_0 = \frac{(n-1)S^2}{\sigma_0^2} \rightarrow \frac{19 * 0.015}{0.01} = 29.07$$

fail to reject H_0

* p-value :



$p > \alpha$

$0.05 < p < 0.1$

بكتابي

* problem 9-79 :

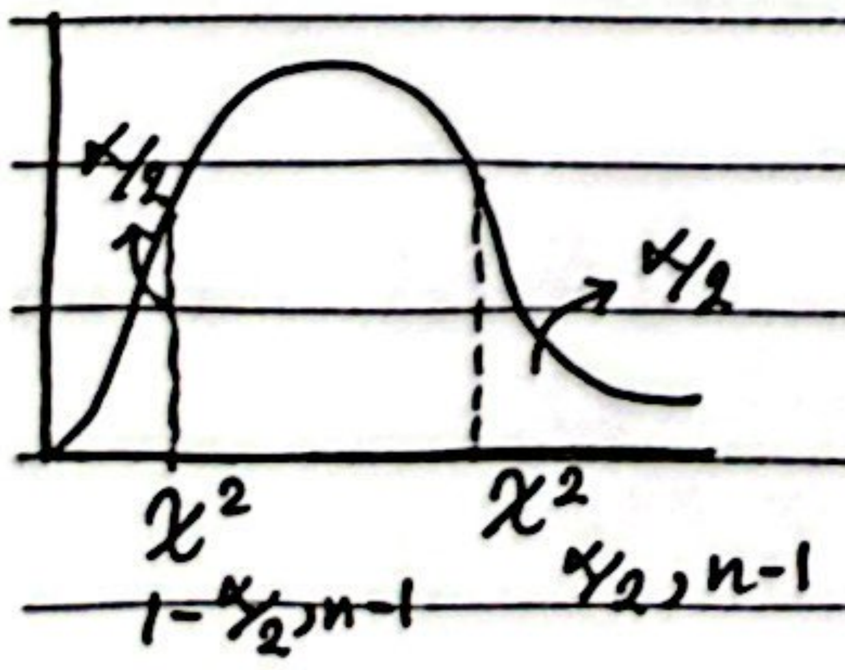
$H_0: \sigma^2 = 0.0625 \leftarrow \sigma = 0.25$

$H_1: \sigma^2 \neq 0.0625 \qquad \sigma \neq 0.25$

$S = 0.37$, $n = 51$

$v = 50$, $\alpha = 0.05$

حولتها عشان أستخدم χ^2



$$\chi^2_{0.025,50} = 71.42$$

$$\chi^2_{0.975,50} = 32.36$$

Five Apple

$$\chi_0^2 = \frac{50 * (0.37)^2}{0.0625} = 109.52$$ reject H_0

p-value :

$\chi^2 >$ القيمة الأكبر $\longrightarrow$ $P/2 < 0.005$ $\quad \alpha > p$

79.49 $\quad p < 0.01$ $\quad$ reject

section 9.5 : test on a population proportion :

$$\boxed{Z_0 = \frac{X - nP_0}{\sqrt{nP_0(1-P_0)}}} \xrightarrow{or} \boxed{Z_0 = \frac{\hat{P} - P_0}{\sqrt{P_0(1-P_0)/n}}}$$

our test statistic

بين قسمة n عليها

P_0 → hypothesised value

$Z_{\alpha/2}$, $Z_{-\alpha/2}$ → 2 sided
Z_α, $-Z_\alpha$ → one sided
} the critical value

for B & n calculations, check page 326 from the book.

problem 9-91 : بشكل عام

$\hat{p} = \frac{16}{200} = 0.08$; $p \geq 0.1$ → at least ; $p < 0.1$

$H_0 : p = 0.1$ $\quad n = 200$

$H_1 : p < 0.1$ $\quad \alpha = 0.01$

↳ so lower

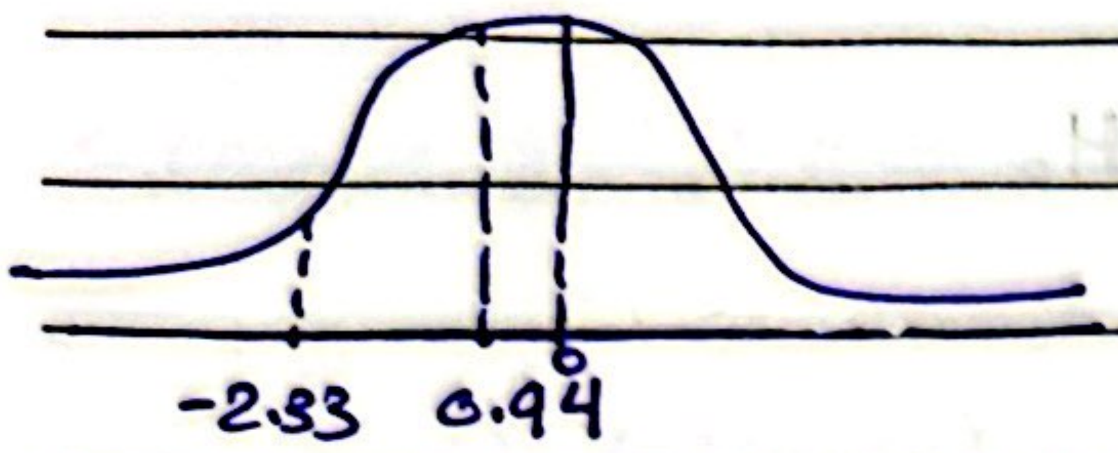
$Z_{0.01} \rightarrow -2.33$

$$Z_0 = \frac{\hat{P} - P_0}{\sqrt{\frac{P_0(1-P_0)}{n}}} = \frac{0.08 - 0.1}{\sqrt{\frac{0.1 * 0.9}{200}}} = -0.94$$ fail to reject H_0

p-value $\longrightarrow 0.174$

$p > \alpha$ → fail to reject

Five Apple



* لو ضفنا فرع ← what is β-error , given that $p = 0.06$

$\beta = 1 - \phi(-0.56) \rightarrow 1 - 0.29 \simeq 0.71$

* What is the required sample size?

$\beta = 0.1$, $p = 0.06$

$n = 629$ (rounding up)

* في عشان أطلع β لسكشن χ^2 هون بتبع الـ λ لازم أطلع

$$\boxed{\lambda = \frac{\sigma}{\sigma_0}}$$

وبكون معنا n ، بنطلع β بالـ OC curve

Chapter (10)

فكرة الشابتر إنه عنا هون (2 populations) ، راح ناخذ من كل واحد (sample) ، والـ parameter of interest

هون عنا الفرق بين هدول الـ 2 ، عنا ٤ حالات ، نفس إلي قبل

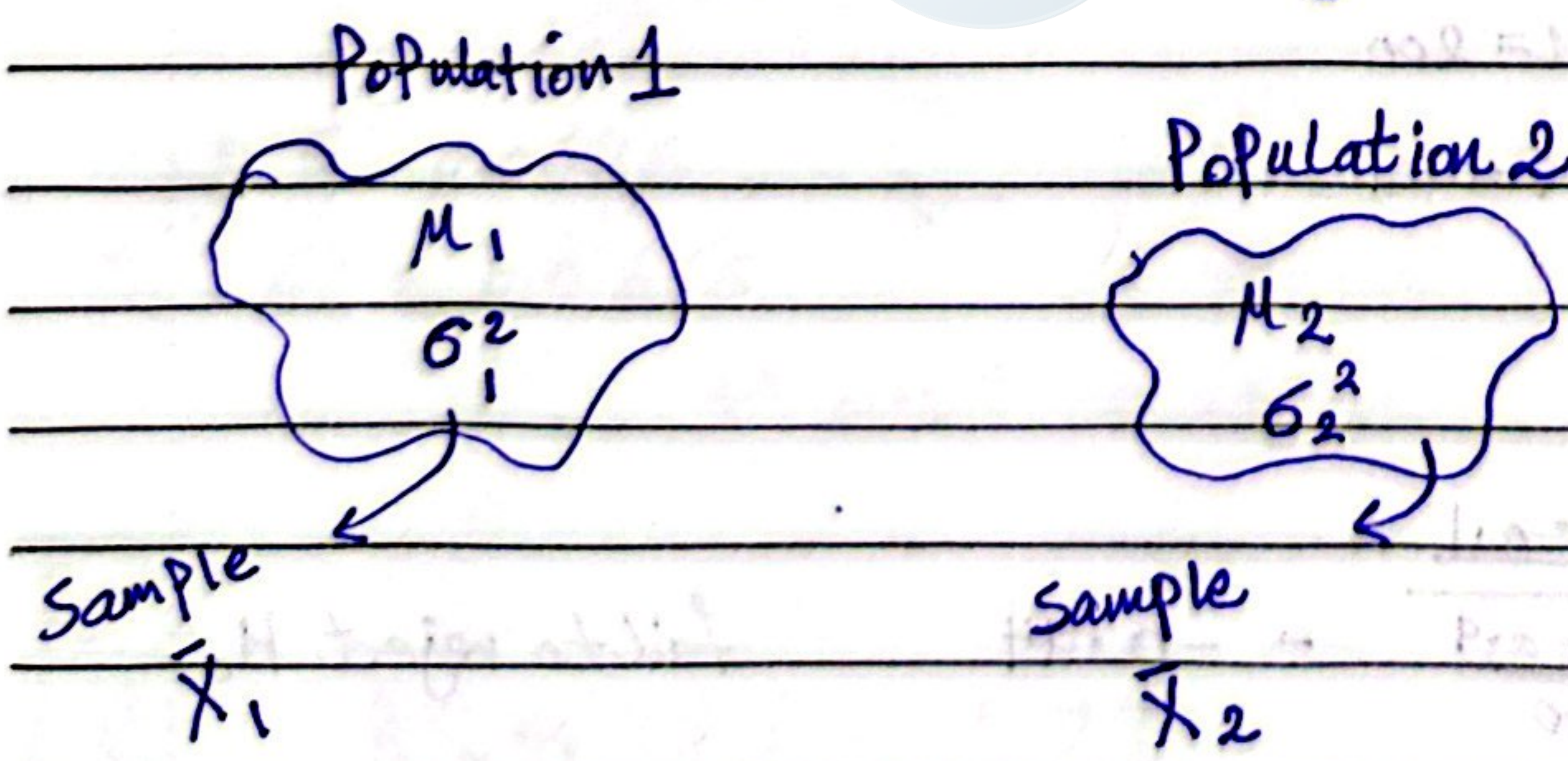
$m \checkmark$, $\sigma^2 \times$

$\mu \checkmark$, $\sigma^2 \checkmark$

$\sigma^2 \checkmark$

$p \checkmark$

* برضو هون بدنا نعمل : 1) Point estimation 2) C.I 3) hypothesis testing



[μ, σ^2 known]

introduction :

هلأ: $\bar{X}$ هو R.v بختلف باختلاف السامبل، وإذا كان normally distributed فإله mean، وال mean
هو ال expected value لل $\bar{X}$، وإلي هو نفسه ال mean لل population، سواء الأول أو الثاني

$$E(\bar{X}) = \mu_{\bar{X}} = \mu_p$$

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

بالحيث $\bar{X}_1$ هو ~ N وال $\bar{X}_2$ ~ N، عادةً
طرحهم برضه بكون ~ N

$\bar{X}_1 - \bar{X}_2$ = Normally distributed variable

- $E(\bar{X}_1 - \bar{X}_2) = E(\bar{X}_1) - E(\bar{X}_2)$
- $\sigma^2(\bar{X}_1 - \bar{X}_2) = \sigma^2(\bar{X}_1) + \sigma^2(\bar{X}_2)$ لأنه σ^2 تراكمي حتى لو انطرح
- $\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

case (1) : $\bar{X}_1 - \bar{X}_2$: "point estimation"

بدنا نعمله standerdization

our test statestic

$$Z_0 = \frac{\overbrace{(\bar{X}_1 - \bar{X}_2)}^{\text{our variable}} - \overbrace{(\mu_1 - \mu_2)}^{\text{it's mean}}}{\underbrace{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}_{\text{it's standard deviation}}}$$

(Z distribution)

case (2) : hypothesis testing :

1) parameter of interesting : $\mu_1 - \mu_2$

2) $H_0 : \mu_1 - \mu_2 = \Delta_0$ هاد بحالة ال 2 sided
$H_1 : \mu_1 - \mu_2 \neq \Delta_0$, Δ_0 : the hypothesised difference between the 2 means

3) $Z_{\alpha/2}$, $-Z_{\alpha/2}$, calculate !

4) Compare P-value with α, or Z_0 and $Z_{\alpha/2}$ (test & critical)

* p-value : the probability above Z_0 and below Z_0

$P = 1 - \phi_{Z_0} \rightarrow$ if above Z_0

$P = \phi_{Z_0} \rightarrow$ if below Z_0

above & below: $P = 2[1 - \phi(|Z_0|)]$

$\Delta_0 \rightarrow$ ممكن تكون أي قيمة بس غالبًا
تكون صفر لإنه ال μ_1 و μ_2 متساويات

* type 2 error & choice of sample size :

$$d = \frac{|\mu_1 - \mu_2 - \Delta_0|}{\sqrt{\sigma_1^2 + \sigma_2^2}} = \frac{|\Delta - \Delta_0|}{\sqrt{\sigma_1^2 + \sigma_2^2}}$$

Δ → true difference, Δ_0 → hypothesised difference

من هاي الصيغة

$$\tilde{n} = \frac{\sigma_1^2 + \sigma_2^2}{\sigma_1^2/n_1 + \sigma_2^2/n_2}$$

→ equivalent sample size

* عندي قيمة ل n_1 و قيمة ل n_2، عشان أستخدم ال OC curves للازم أجيب n مشتركة بين n_1 و n_2، إلا إذا كانوا متساويين بستخدم وحدة منهم.

* عشان نتعامل مع ال OC curves :
1) حدد إذا ال test ال Z & t.
2) حدد إذا one & 2sided.
3) حدد قيمة α.

* B formula for 2 sided :

$$\beta = \phi\left(Z_{\alpha/2} - \frac{\Delta - \Delta_0}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}\right) - \phi\left(-Z_{\alpha/2} - \frac{\Delta - \Delta_0}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}\right)$$

* sample size for a 2 sided test on the difference in means $n_1 = n_2$, σ^2 known

$$n = \frac{(Z_{\alpha/2} + Z_\beta)^2(\sigma_1^2 + \sigma_2^2)}{(\Delta - \Delta_0)^2}$$

Five Apple

* sample size for a one sided test on the difference in means with $n_1 = n_2$, σ^2 known:

$$n = \frac{(Z_\alpha + Z_\beta)^2 (\sigma_1^2 + \sigma_2^2)}{(\Delta - \Delta_0)^2}$$

[for 2 sided]

* Case (3): CI

$$\underbrace{\bar{X}_1 - \bar{X}_2}_{\text{statistical mean difference}} \underbrace{- Z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}_{\text{margin of error}} \le \underbrace{\mu_1 - \mu_2}_{\text{our variable}} \le \bar{X}_1 - \bar{X}_2 + Z_{\alpha/2} \underbrace{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}_{\text{الجذر لحاله هو ال standard deviation for } \mu_1 - \mu_2}$$

* choice of sample size:

$$n = \left(\frac{Z_{\alpha/2}}{E}\right)^2 (\sigma_1^2 + \sigma_2^2)$$

* for one sided CI:

upper: $\mu_1 - \mu_2 \le \bar{X}_1 - \bar{X}_2 + Z_\alpha \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

lower: $\bar{X}_1 - \bar{X}_2 - Z_\alpha \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \le \mu_1 - \mu_2$

* Problem 10.4 :

machine 1		machine 2	
16.03	16.01	16.02	16.03
16.04	15.96	15.97	16.04
16.05	15.98	15.96	16.02
16.05	16.02	16.01	16.01
16.02	15.99	15.99	16.00

Sol: 1) sample 1: $\sigma_1 = 0.02$, $\bar{X}_1 = 16.015$, $n_1 = 10$

sample 2: $\sigma_2 = 0.025$, $\bar{X}_2 = 16.005$, $n_2 = 10$

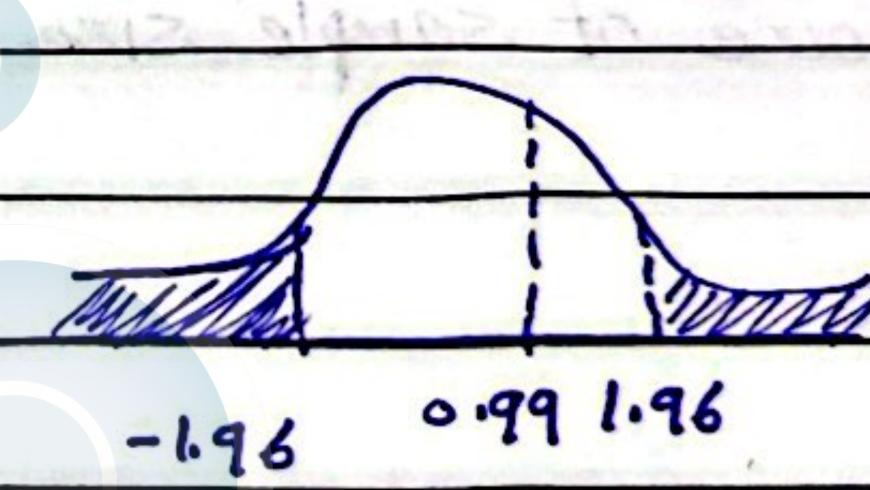
$\alpha = 0.05$

3) $Z_{\alpha/2}$

$\alpha = 0.05$

$\alpha/2 = 0.025$

$Z_{0.025} = 1.96$



2) $H_0 : \mu_1 - \mu_2 = 0$

$H_1 : \mu_1 - \mu_2 \neq 0$ $\}$ ∴ 2-sided

4) $$Z_0 = \frac{\bar{X}_1 - \bar{X}_2 - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \rightarrow \frac{16.015 - 16.005 - 0}{\sqrt{\frac{0.02^2}{10} + \frac{0.025^2}{10}}} = 0.99$$

∴ we fail to reject H_0

p-value : $\frac{\text{p-value}}{2} = 0.161087 \rightarrow 0.32$ ∴ p-value > α ∴ fail to reject

b) $\bar{X}_1 - \bar{X}_2 - E \leq \mu_1 - \mu_2 \leq \bar{X}_1 - \bar{X}_2 + E$

$-0.0098 < \mu_1 - \mu_2 \leq 0.0298$

c) $\Delta = 0.04$, Power ?

$\Delta_0 = 0 \rightarrow$ من السؤال
$\mu_2 = \mu_1$ لانه

Power = 1 - B

$\hookrightarrow \beta = \phi(-1.99) - \phi(-5.91)$

$= 0.023295 - 0 \longrightarrow 0.023295$

power $= 0.976705 \simeq 0.98$

D) $\beta = 0.05$, $\Delta = 0.04$, n ?

$n = 8.35 \simeq 9$

$\hookrightarrow n_1 = 9, n_2 = 9$
من السؤال

B for 1-sided ?

$\longrightarrow$ upper = $\left[\phi \left(Z_\alpha - \frac{(\Delta - \Delta_0)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \right) \right]$

$\longrightarrow$ lower = $\left[1 - \phi \left(-Z_\alpha - \frac{\Delta - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \right) \right]$

Section 10.2 : inference on the difference in means , σ^2 unknown :

هون ال test statistic هو t مش z وفي هنا حالتين :

case 1 :

$\sigma_1^2 = \sigma_2^2 = \sigma^2$

ال σ^2 ال pop متساويات ولكن مش معناها إنه ال s متساويات ، بأخذ قيمة وسطية بينهم وإلي هي

$$\left[S_p^2 = \frac{(n_1 - 1) S_1^2 + (n_2 - 1) S_2^2}{n_1 + n_2 - 2} \right]$$

طبعًا s_p^2 بتتبع الـ t distribution مع v إله كان الـ v الـ 2 pop

$\hookrightarrow n_1 - 1 + n_2 - 1 \rightarrow n_1 + n_2 - 2$

اسمها Pooled estimator

* case 2 :

$\sigma_1^2 \neq \sigma_2^2$

← بنعوض وقتها عن σ_1^2 بـ s_1^2 وعن σ_2^2 بـ s_2^2

وبالنسبة للـ dof هون بصير بدها estimator ، حرف لاتيني اسمه new ← V

$$V = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}} \rightarrow$$

هون بنعمل round down to the nearest integer

مثلًا 23.7 بتصير 23

* test statistic for case 1 :

$$T_0 = \frac{\bar{X}_1 - \bar{X}_2 - \Delta_0}{\sqrt{\frac{sp^2}{n_1} + \frac{sp^2}{n_2}}} \xrightarrow{\text{نفسه}} \frac{\bar{X}_1 - \bar{X}_2 - \Delta_0}{sp\sqrt{1/n_1 + 1/n_2}}$$

* the critical value is:

$T_{\alpha/2},\ n_1 + n_2 - 2$

* test statistic for case 2 :

$$T_0 = \frac{\bar{X}_1 - \bar{X}_2 - \Delta_0}{\sqrt{s_1^2/n_1 + s_2^2/n_2}},$$

* the critical value is

$T_{\alpha/2},\ \underset{\hookrightarrow \text{new}}{V}$

* The CI for case 1 :

$$\bar{X}_1 - \bar{X}_2 - t_{\alpha/2, n_1+n_2-2}\ sp\sqrt{1/n_1 + 1/n_2} \leq \mu_1 - \mu_2 \leq \bar{X}_1 - \bar{X}_2 + t_{\alpha/2, n_1+n_2-2}$$

$$sp\sqrt{1/n_1 + 1/n_2}$$

* CI for case 2 :

$$\bar{X}_1-\bar{X}_2-t_{\alpha/2,v}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}\leq \mu_1-\mu_2\leq \bar{X}_1-\bar{X}_2+t_{\alpha/2,v}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$$

بكتابي

* problem 10-15 :

sample 1	sample 2	
$\bar{X}_1 = 8.73$	$\bar{X}_2 = 8.68$	$H_0: \mu_1-\mu_2=0$
$s_1^2 = 0.35$	$s_2^2 = 0.4$	$H_1: \mu_1-\mu_2\neq 0$
$n_1 = 15$	$n_2 = 15$	two tailed test

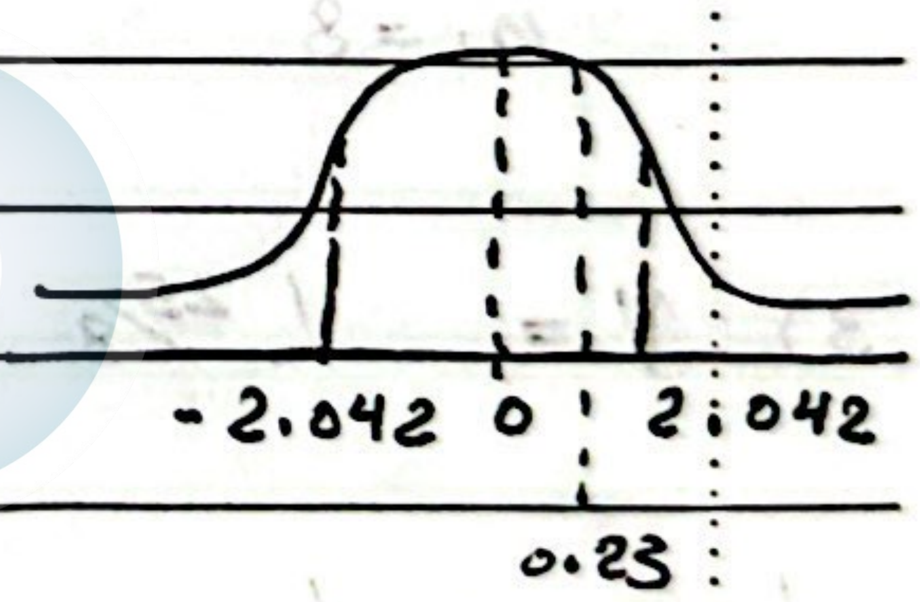
$\alpha = 0.05$

$$sp^2=\frac{14*0.35+16*0.4}{15+17-2}$$

$\bar{X}_1-\bar{X}_2 = 8.73-8.68 = 0.05$

$sp^2 = 0.38 \quad , \quad sp = 0.614$

$t_{\alpha/2,\,n_1+n_2-2} \longrightarrow t_{0.025,30} = 2.042$



$$T_0=\frac{\bar{X}_1-\bar{X}_2-\Delta_0}{sp\sqrt{1/n_1+1/n_2}}=\frac{0.05-0}{0.614*\sqrt{1/15+1/17}}$$

we fail to reject H_0

p-value ?

$\frac{\text{p-value}}{2} > 0.4 \longrightarrow \text{p-value} > 0.8$

CI ?

$1-\alpha = 0.25 \quad , \quad t_{0.025,30} = 2.042$

$$\bar{X}_1-\bar{X}_2-E\leq \mu_1-\mu_2\leq \bar{X}_1-\bar{X}_2+E$$

$E = t_{0.025, 30} * sp\sqrt{1/n_1 + 1/n_2}$

$E = 0.444 \longrightarrow -0.294 \leq \mu_1 - \mu_2 \leq 0.494$

includes the zero

بتاني

*** problem 10-26 :**

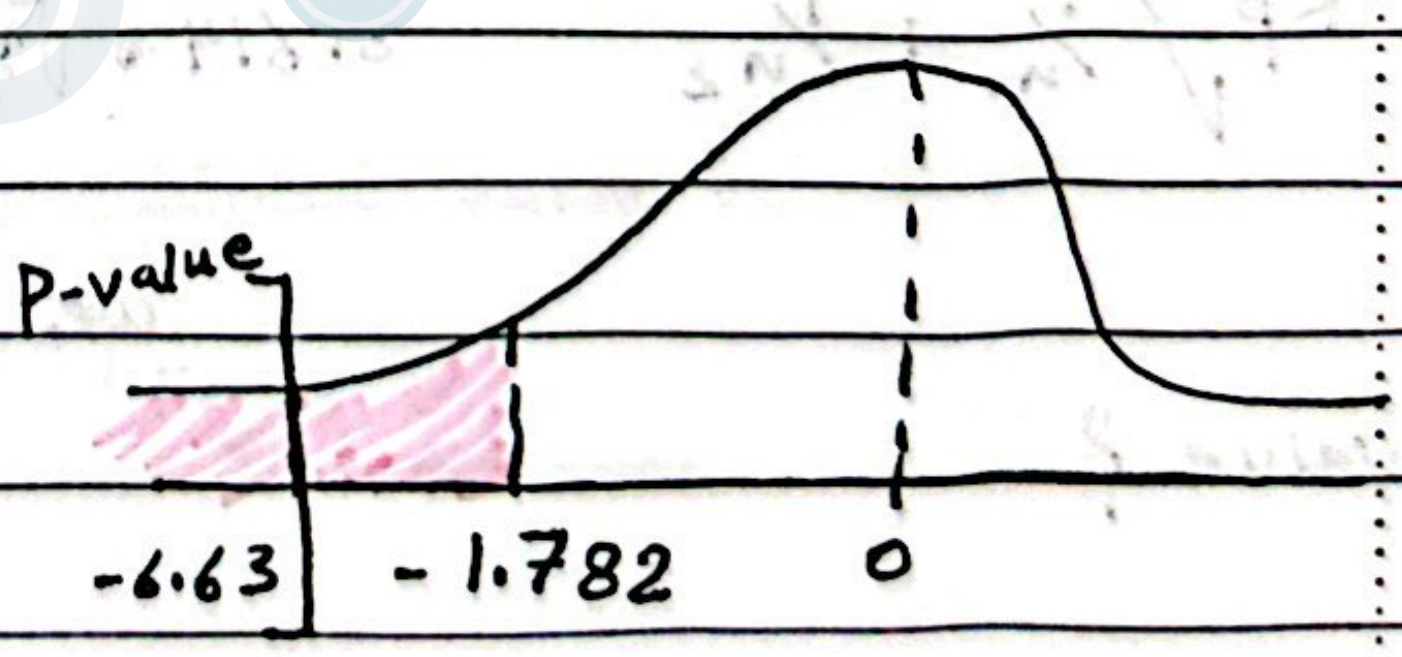
8 rats $\longrightarrow$ sample 1 , 9 rats $\longrightarrow$ sample 2

a) $\mu_2 > \mu_1 \rightarrow$ so $\mu_1 - \mu_2 < 0$, $\mu_1 - \mu_2 \geq 0$, $\sigma_1^2 \neq \sigma_2^2$

sample 1	sample 2	
$\bar{x}_1 = 90$	$\bar{x}_2 = 115$	$H_1: \mu_1 - \mu_2 < 0$
$s_1 = 5$	$s_2 = 10$	$H_0: \mu_1 - \mu_2 = 0$
$s_1^2 = 25$	$s_2^2 = 100$	so it's a lower tailed test
$n_1 = 8$	$n_2 = 9$	

3) $V = \left(25/8 + 100/9\right)^2 \Big/ \frac{(25/8)^2}{7} + \frac{(100/9)^2}{8} = 12.04$ so $V = 12$

$t_{\alpha, V} = t_{0.05, 12} = 17.82$



$t_0 = \dfrac{\bar{x}_1 - \bar{x}_2 - \Delta_0}{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \rightarrow \dfrac{90 - 115 - 0}{\sqrt{\frac{25}{8} + \frac{100}{9}}}$

$= -6.63$, we reject H_0

p-value < 0.0005

b) CI ? $\alpha = 0.05$, $\mu_1 - \mu_2 < \bar{x}_1 - \bar{x}_2 + E$

$E = 1.782 \sqrt{\frac{25}{8}+\frac{100}{9}} = 6.72$ so $\mu_1-\mu_2 \le -18.28$

* section 10.4) paired t-test :

* أي قواعيت متقابلات بين ال samples 2 بينهم علاقة بنسمّيهم Pair

sample 1	sample 2	
x_{11}	x_{21}	→ لو بينهم علاقة ، اسمهم pair
x_{12}	x_{22}	→ 〃 〃 〃
⋮	⋮	
x_{1n}	x_{2n}	

* هون ال Parameter of interest هو كمان $\mu_1-\mu_2$ عن طريق paired t-test .

* the data is collected in pairs.

بنضيف كمان عامود بمثل الفرق بين ال 2 samples هدول ، وهاد العامود هو إلي بمثل ال sample تاعنا ، بنجيبله μ .

	Sample 1	sample 2	D_i
pair 1 ←	x_{11}	x_{21}	$x_{11}-x_{21}$
pair 2 ←	x_{12}	x_{22}	$x_{12}-x_{22}$
	⋮	⋮	⋮

هاد العامود الجديد إله برضه μ, S, n

لازم أعملهم حسابات ، زي chapter-9 إنه سامبل واحد صفت

* problem 10-40) : بلتالي حليه

* problem 10-44) :

* صفحة 376 بالكتاب فيه كلشي عن حسابات paired t test

Test (1)	test (2)	diff
		-0.2
		-0.4
		0
		0.1
		-0.3
		-0.3
		-0.3

-0.3

1) $\bar{D} = -0.2125$

$S_d = 0.1727$

$n = 8$

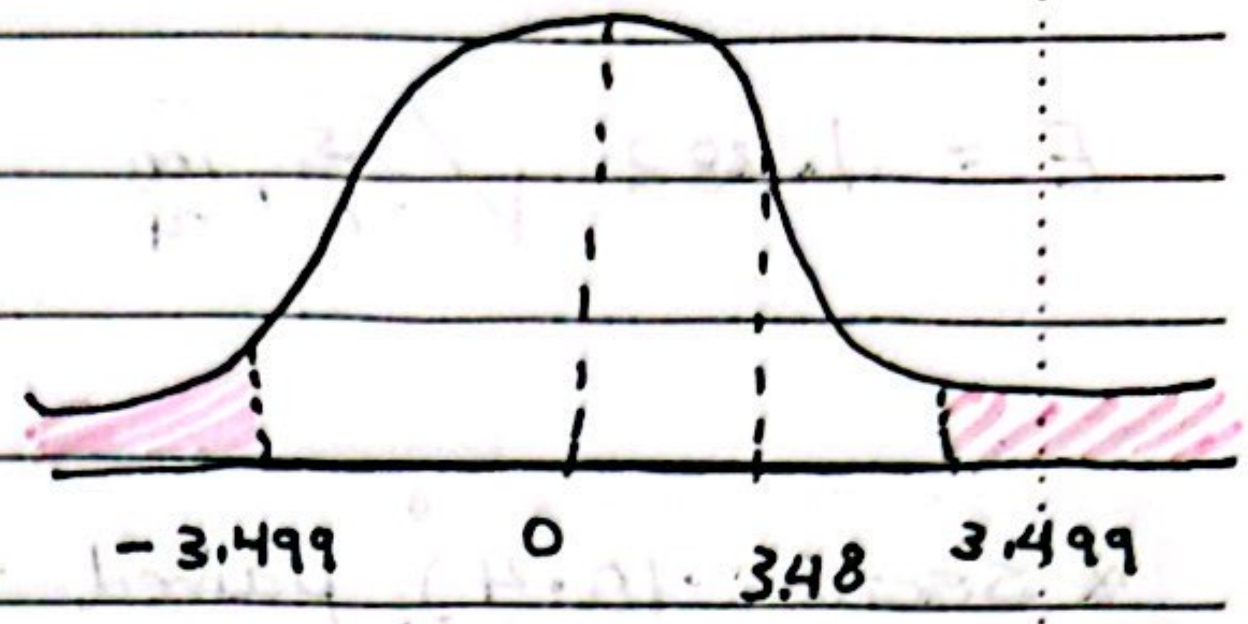
2) $H_0: \mu_D = 0$

$H_1: \mu_D \neq 0$

so, two sided .

3) $t_{0.005,7} = 3.499$

$$t_0 = \frac{\bar{D} - \Delta_0}{S_d / \sqrt{n}} \rightarrow \frac{-0.2125 - 0}{0.1727 / \sqrt{8}} = 3.48$$



so we failed to reject H_0

* The formula for CI :

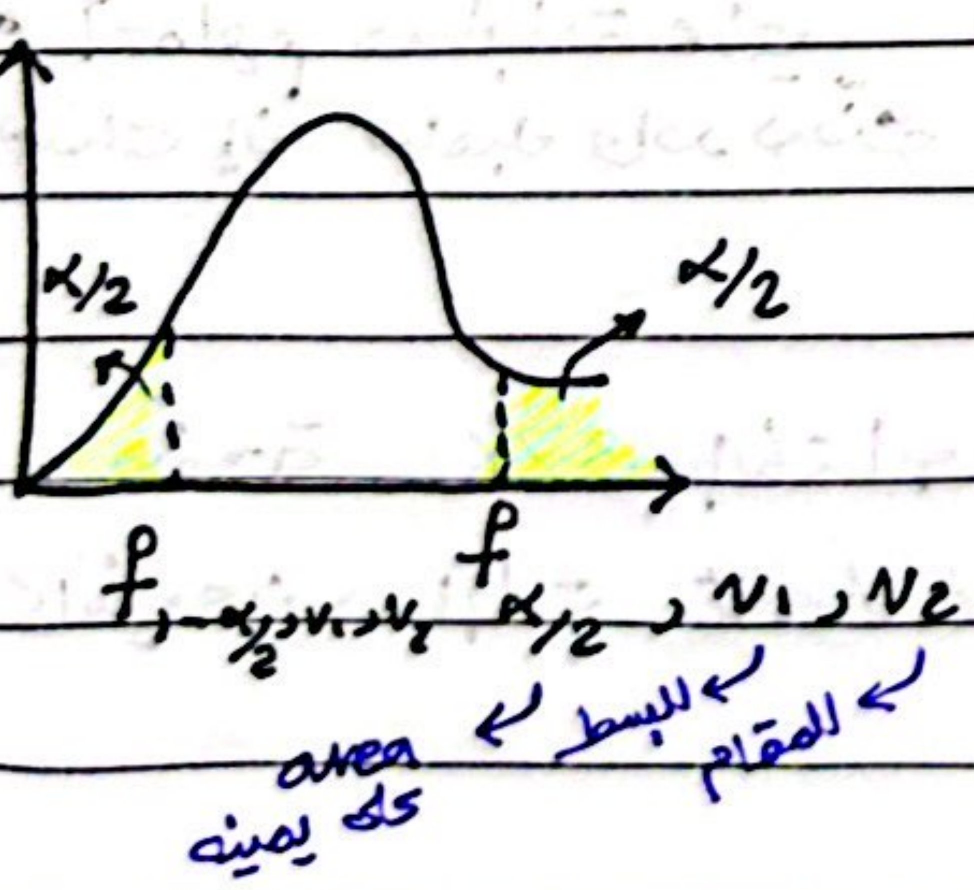
$$\boxed{\bar{d} - t_{\alpha/2, n-1} S_D / \sqrt{n} \leq \mu_D \leq \bar{d} + t_{\alpha/2, n-1} S_D / \sqrt{n}}$$

* section 10.5) : inference on the σ^2s

here we have a new distribution → F distribution

$$\boxed{F = \frac{\chi_1^2 / v_1}{\chi_2^2 / v_2}}$$

هنا 2 samples، كل واحد إله χ^2 خاص فيه و dof خاصة فيه،
لو أقسمهم على بعض بنحصل ال F إلي إلها برضو dof .



* مستحيل يكون سالب زي كاي، بيبلش من الصفر

* for hypothesis : [بنختبر ال ratio بينهم]

1) Parameter of interest : $\frac{\sigma_1^2}{\sigma_2^2}$

2) $H_0 : \sigma_1^2 = \sigma_2^2 \xrightarrow{\text{بتصير}} \frac{\sigma_1^2}{\sigma_2^2} = 1$

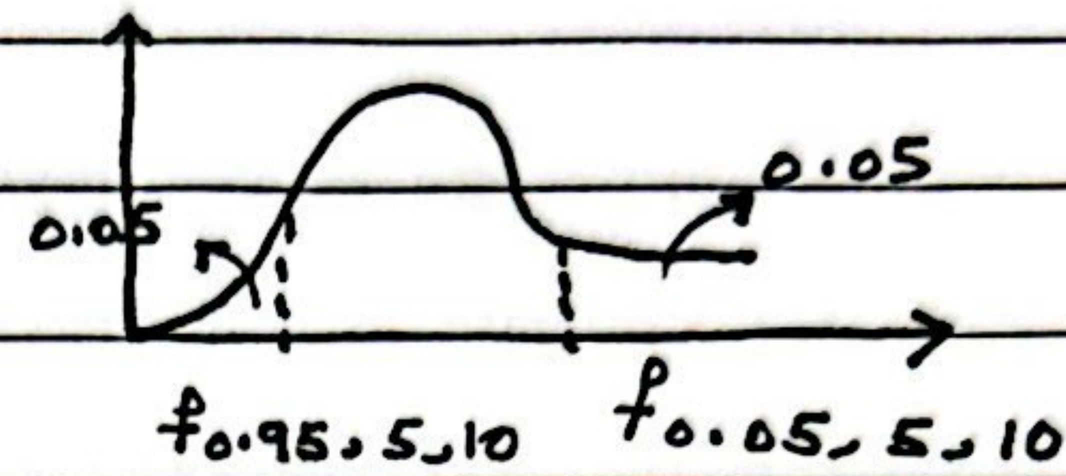
$H_1 : \sigma_1^2 \neq \sigma_2^2 \xrightarrow{\text{بتصير}} \frac{\sigma_1^2}{\sigma_2^2} \neq 1$

الصيغة لـ F بالكتاب هيك : $\boxed{F = \frac{W/u}{Y/v}}$

* جداول الـ F بتعطيني قيم F حسب الـ area إلي هو كاتبها (α) .

* يعني بنختار الـ table حسب الـ α لنجيب قيمة F .

مثلاً



3.33

$\alpha = 0.1$

$\alpha/2 = 0.05$

$u = 5$

$v = 10$

* الجداول الموجودة بس $\alpha/2 =$ 0.1 ، 0.25 ، 0.05 ، 0.025 ، 0.01

* ما في table بحسب $f_{0.95}$ ، ففي صيغة بقدر أحسبها فيها :

$$\boxed{f_{1-\alpha, v_1, v_2} = \frac{1}{f_{\alpha, \underbrace{v_2, v_1}_{\text{معكوسين}}}}}$$

so $\frac{1}{f_{0.05, 10, 5}} = \frac{1}{4.74} = 0.211$

* البوكس بالكتاب صفحة 385 كاتب كلشي عن F .

The CI : $\left[\frac{S_1^2}{S_2^2} f_{1-\alpha/2, n_2-1, n_1-1} \le \frac{\sigma_1^2}{\sigma_2^2} \le \frac{S_1^2}{S_2^2} f_{\alpha/2, n_2-1, n_1-1} \right]$

بكتابي

*Problem 10-54 :

Sample 1: $S_1 = 4.7$, $n_1 = 10$

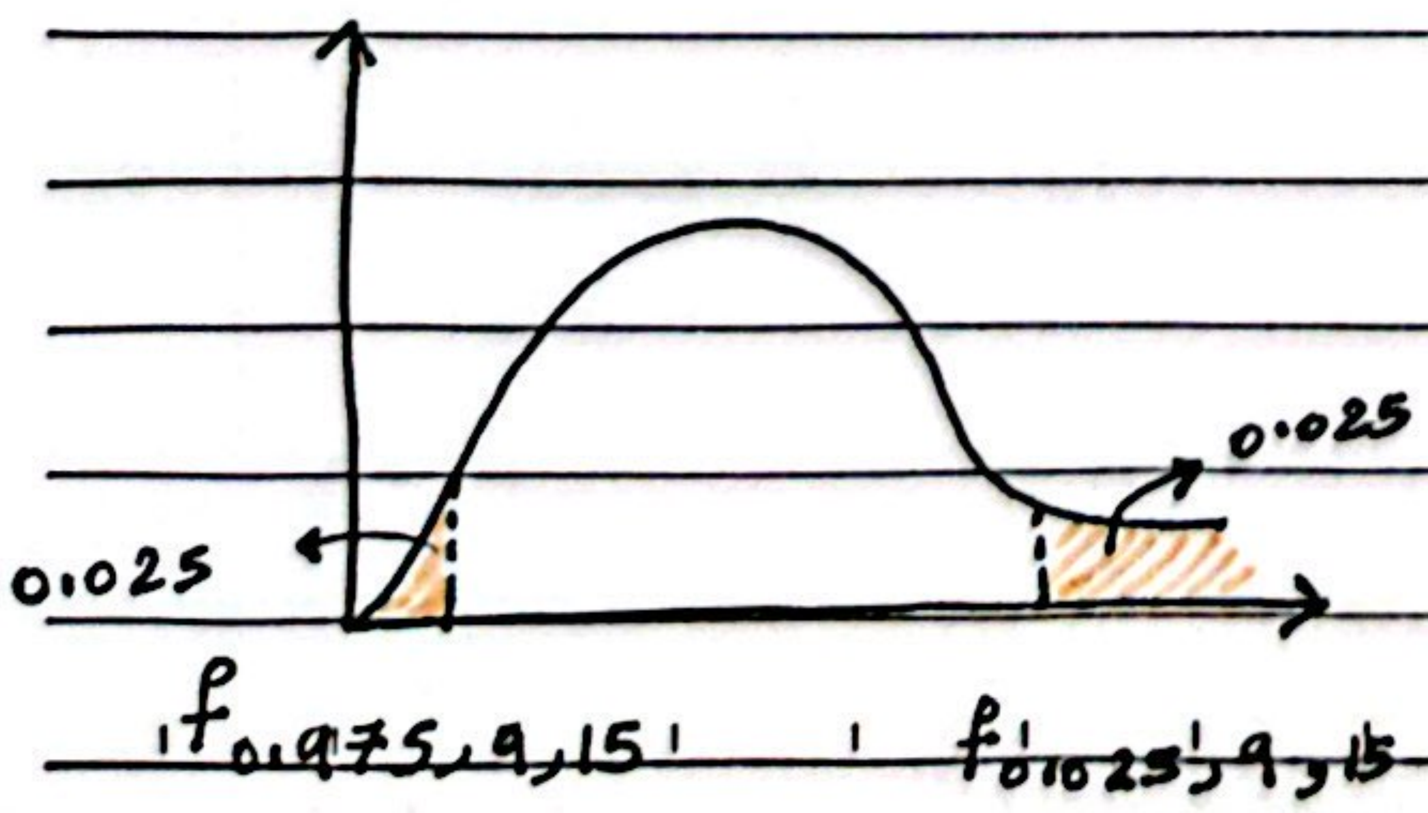
sample 2: $S_2 = 5.8$, $n_2 = 16$

$\alpha = 0.05$

$H_1 : \sigma_1^2 \neq \sigma_2^2 \rightarrow \frac{\sigma_1^2}{\sigma_2^2} = 1$

$H_0 : \sigma_1^2 = \sigma_2^2$

$\hookrightarrow \frac{\sigma_1^2}{\sigma_2^2} \neq 1$



$\hookrightarrow = 3.12$

$$f_{0.975, 9, 15} = \frac{1}{f_{0.025, 15, 9}} = 0.265$$

$$F_0 = \frac{S_1^2}{S_2^2} = \frac{(4.7)^2}{(5.8)^2} = 0.657$$

we fail to reject

Chapter 11

section
11.1) Empirical models:

* the regression is a relationship btw 2 or more vars.

X , y
ind var ↲ ↳ dep var

* relationships in general are 2 types:

1. deterministic (قطعي) → شو ما أعمل تجارب ، بس اثبت قيمة لـ (X) راح يعطيني قيمة ثابتة لـ (y).
2. stochastic (احتمالي) (Probabilistic) → بس أثبت قيمة لـ (X) راح كل مرة يعطي قيمة لـ (y) مختلفة بسبب (random errors) عادي حتى لو بنفس الظروف.

1) زي مساحة الدائرة لو نفس الـ diameter

2) زي مثال التفاعل الكيميائي → var يعني الـ y هاي إله ستانيستك

هنا بدنا نعمل model يكون simple → one ind & one dep var ويكون linear

* احنا مهتمين بالـ non deterministic هون.

* regression analysis: the collection of statistical tools that are used to model and explore relationship btw vars that are related in a non deterministic manner.

X → ind var (predictor)
y → dep var (response)

* هون كنا best fit line ، النقاط إلي عالخط تماماً همة الـ expected values (→ predicted) ، أما إلي بكون فيهم error همة الـ actual (← observed) ، والفرق بينهم هو الـ error الي بالـ scattar diagram.

* expected value → ∴ mean
لكل نقطة عالخط.

slope

* هلأ معادلة الليني هي $y = ax + b$

intercept

intercept → slope

نفس الإشي هون بس in different notations ← $E(y|X) = \beta_0 + \beta_1 X$

بعطيني ال mean ل y بالنسبة لقيم X

* Mean of (y) for a given value of (x)

* Ex: $E(Y/X) = 75 + 15X$ ← هاد الموديل راح يعطينا القيم المتوقعة ل (y) حسب (X)

intercept, slope

يعني لو $X = 1$ ← $y = 90$

هاي مش قيمة حقيقية ، هاي Mean (متوقعة)

حولين ال 90 بكل تجربة والمتوسط تاعهم 90

if $X = 1$ → model → y_1, y_2, y_3 } its mean is $y = 90$

* هون كنا معادلة بتعطي القيمة الحقيقية ل y ، لإنه ضاف ال error

مش تقريبي

الفرق بين ال expected and the actual

هذه (ϵ) إبسيلون

$$y = \beta_0 + \beta_1 X + \epsilon$$

هاي المعادلة هي ال simple linear regression model

* X is the ind var or the (regressor) اسمه

* ϵ is a var also ⟶ * ما بعرف بالزبط كم الفرق.

* بقدر أطلع بعلاقة بين متغيرين ما إلهم دخل ببعض عادي بس وقتها ما راح يكون في السبب منطقي (no regression (causal))

section

11.2) Simple linear regression:

$E(Y|X) = \beta_0 + \beta_1 x$ ← المعادلة هاي

← هدول Parameters الموديل بدنا نعملهم estimation

بال ٣ طرق إلي قبل بردنه ← PE, CI, HT

فوائد ال LR ← prediction, optimization purposes

* حنعطي معادلات لـ β_0 و β_1 ← (اسمهم the least squares estimates) بنحسب فيهم الـ b's
هاد الموديل بأفضل شكل ممكن (best fit line)، والقيم فيهم همة estimators.
معمولين بحيث المربع لمجموع الـ errors يكون أقل ما يمكن.

$$\boxed{\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} y_i x_i - \frac{\left(\sum_{i=1}^{n} y_i\right)\left(\sum_{i=1}^{n} x_i\right)}{n}}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}$$

اسم البسط S_{xy} → sum of squares for the products btw x_s and y_s

اسم المقام S_{xx} → sum of squares for x_s

$$\boxed{s = \sqrt{\frac{\sum (x_i - \bar{y})^2}{n-1}}}$$ ← هدول جايين من قانون ال SD وال variance

الها صيغتين $S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}$

الها صيغتين $S_{xy} = \sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x}) = \sum_{i=1}^{n} x_i y_i - \frac{\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n}$

$$\boxed{\text{so } \hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}}$$

not estimated but → predicted value of y for a given value of X ← $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 X$ بصير المعادلة

الحقيقية بجيبها من هون

$y = \hat{\beta}_0 + \hat{\beta}_1 X + e$

لما استخدم الـ estimators بصير الـ error اسمه (residual)

$(e = \text{actual} - \text{esstimated})$

actual: from sample — estimated: from model

مقترن بالقيم الموجودة بالـ sample

يختلف عن ϵ ← هاد Var

X	Y
3	6

من الـ sample → if $\hat{y} = 5.8$, so $e = 6 - 5.8 = 0.2$

هاد القيمة وحدة معينة (the residual)

هاد ϵ هو RV ← الـ Mean الها = صفر والـ Var (σ^2) تعطى من العلاقة هاي :

لأنه ما بنعرف أي قيمة لـ y راح تطلعلي لما اختار قيمة لـ X

$$\boxed{\sigma^2 = \frac{SS_E}{n-2}}$$ (the estimator of variance)

unknown ← فبعمل estimation

SS_E) the errors sum of squares .

كيف أجيبها ؟ → $$\boxed{SS_E = SS_T - \hat{\beta}_1 S_{xy}}$$

SS_T → Total

هي نفسها S_{yy} ، كيف أجيبها ؟

$$\boxed{SS_T = \sum_{i=1}^{n} (y_i - \bar{y})^2}$$

صيغة ثانية → $$\boxed{\sum_{i=1}^{n} y_i^2 - n\bar{y}^2}$$

يعني كخطوات : بحسب SS_T بعدين SS_E بعدين σ^2 .

Five Apple

ceramic
american society :
Problem

"the least squares" estimators ⟶ همة $\hat{\beta}_1$ و $\hat{\beta}_0$ نفسوم

$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 X$

$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$

$S_{xx} = \sum x_i^2 - \frac{(\sum x_i)^2}{n}$

$S_{xy} = \sum x_i y_i - \frac{\sum x_i \sum y_i}{n}$

$S_{yy} = \sum y^2 - \frac{(\sum y)^2}{n}$

خطوات حل ال linear regression model :

١) بعمل جدول لـ X و y و yX و X^2 و y^2.

٢) بنجمع ، بنطلع ∑ لكل كامود ، و $\bar{X}$ و $\bar{y}$.

٣) احسب S_{xx} و S_{xy} و SS_T.

٤) احسب $\hat{\beta}_0$ و $\hat{\beta}_1$ و اكتب الموديل.

X	y	X^2	y^2	Xy
1100	30.8	1210000	948.64	33880
1200	19.2	1440000	368.64	23040
1300	6	1690000	36	7800
1100	13.5	1210000	182.25	14850
1500	11.4	2250000	129.96	17100
1200	7.7	1440000	59.29	9240
1300	3.6	1690000	12.96	4680

$\bar{X} = 1242.85$ $\bar{y} = 13.1714$ $\sum X^2 = 10930000$ $\sum y^2 = 1737.74$

$\sum X = 8700$ $\sum y = 92.2$ $\sum Xy = 110590$

$S_{xx} = 10930000 - \frac{(8700)^2}{7} = 117142.86$

$S_{xy} = 110590 - \frac{(8700)(92.2)}{7} = -4001.4$ ⟶ عادي يكون سالب

$\hat{\beta}_1 = \frac{-4001.4}{117142.86} = -0.0342$

$\hat{B}_0 = \bar{y} - \hat{B}_1 \bar{X}$

$\hat{\beta}_0 = 13.17 + 0.0342 * (1242.86) = 55.68$

the model ⟶ $\hat{y} = 55.68 - 0.0342X$

a) find an estimate for σ^2?

$SS_T = \sum y_i^2 - n\bar{y}^2$

$1737.74 - (7)(13.17)^2 = 523.6$

$SS_E = SS_T - \hat{\beta}_1 S_{xy}$

$523.6 + 0.0342(-4001.4) = 386.7$

$$\hat{\sigma}^2 = \frac{SS_E}{n-2} = \frac{386.7}{5} = 77.35$$

b) mean prosity? for a 1400° temp

$\hat{y} = 55.68 - 0.0342(1400) \rightarrow 7.8$

• لو طلب نرسم scatter diagram ونوصف العلاقة، عادي يعني بنعين النقاط لـ x و y.
ممكن ما تكون العلاقة لينير أصلاً.

Section 11.4)

• السكشن الماضي كان إنه حسبنا single numerical value يعني Point esstimators لـ β_1 و β_0

• هون بدنا نحسب الـ hypothesis testing لـ β_1 و β_0

↳ هدول statistics بعتمدو ع sample data
وهمه RV_s

$\boxed{\mu_{\hat{\beta}_1} = \beta_1}$, $\boxed{Var\ \hat{\beta}_1 = \frac{\hat{\sigma}^2}{S_{XX}}}$ } properties for the slope

$\boxed{\mu_{\hat{\beta}_0} = \beta_0}$, $\boxed{Var\ \hat{\beta}_0 = \hat{\sigma}^2\left[\frac{1}{n} + \frac{\bar{X}^2}{S_{XX}}\right]}$ } properties for the intercept

E (expected value)

se (standard error)

[how to hypothesis testing]

$H_0: \beta_1 = \beta_{1,0}$ → hypothesised value

$H_1: \beta_1 \neq \beta_{1,0}$

real value

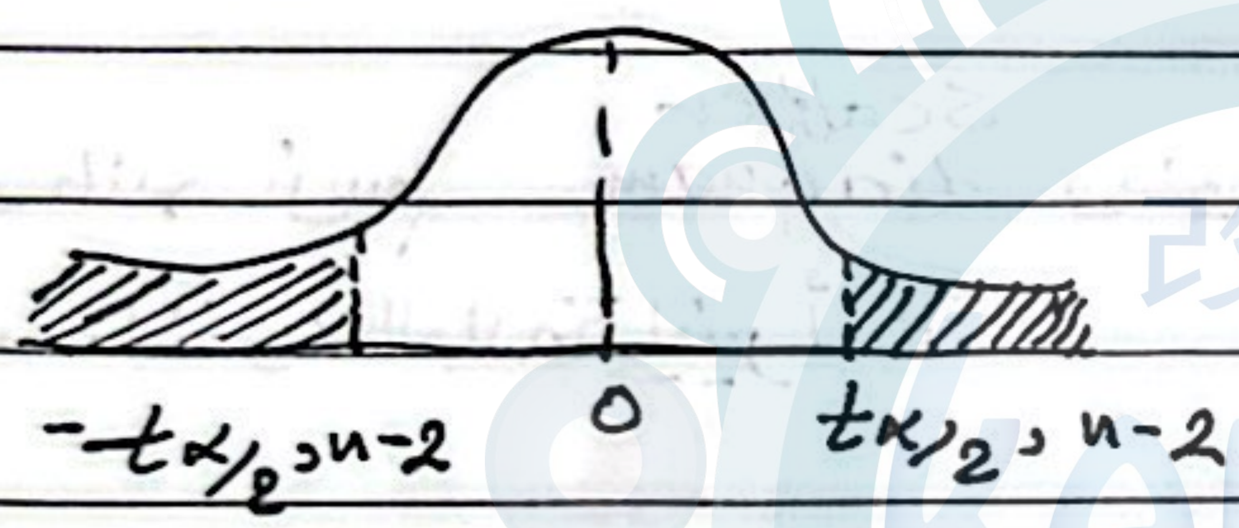
* غالبًا 2 sided

$T_{critical} = T_{\alpha/2, n-2}$ ← من هون بحسب T critical

T dist

$n-2$ ← DoV
دائمًا

$T_0 = \dfrac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\hat{\sigma}^2 / S_{xx}}}$ ← من هون بحسب ال test statistic



for the slope :

* for the intecept :

$H_0: \beta_0 = \beta_{0,0}$ ← hypothesised

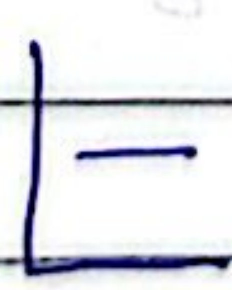
$H_1: \beta_0 \neq \beta_{0,0}$

true
(بالموديل)

$T_{critical} = t_{\alpha/2, n-2}$

$$\boxed{T_0 = \dfrac{\hat{\beta}_0 - \beta_{0,0}}{\sqrt{\hat{\sigma}^2 \left[\frac{1}{n} + \frac{\bar{X}^2}{S_{xx}}\right]}}}$$

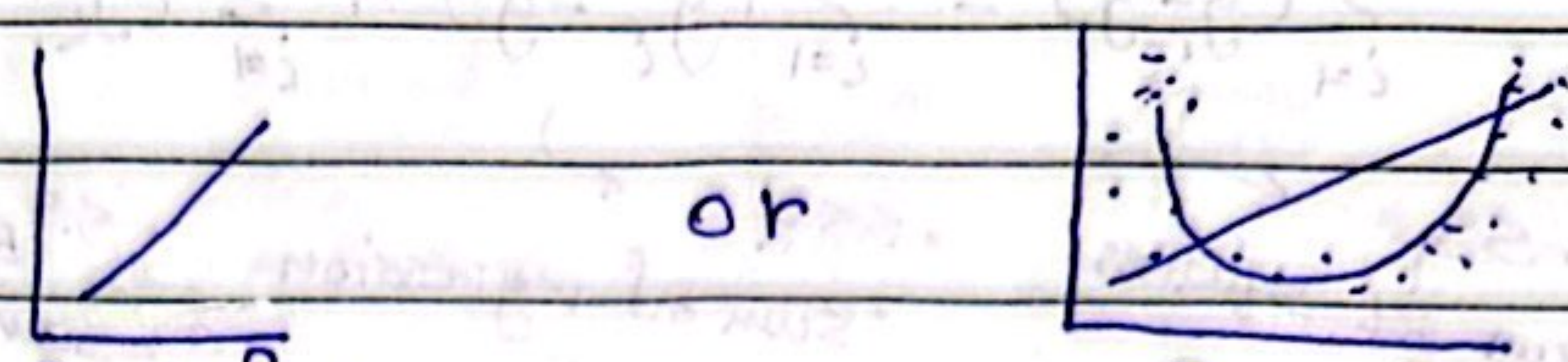
* هنا special case ، هو إنه ال slope β_1 يكون يساوي صفر ، يعني يا العلاقة تكون ثابتة يا أكانت من درجة اللينير ، فالرسمة مثلا تكون

 أو  إلخ الخ

$y = \beta_0 + \beta_2 x^2$ $\quad$ $y = \beta_0$

* لو $\beta_1 \neq 0$ في حالتين ، يا لينير ، يا درجات أعلى من التربيعي

$\beta_1 \neq 0 \longrightarrow$



$y = \beta_0 + \beta_1 x$ or $y = \beta_0 + \beta_1 x + \beta_2 x^2 \ldots\ldots$

$$\left.\begin{array}{l} H_0 : \beta_1 = 0 \\ H_1 : \beta_1 \neq 0 \end{array}\right\}$$

* هاد التست اسمه significance of regression
↳ testing if the slope = or ≠ 0

* عادي بنحسب $T_{critical}$ و بنحسب T_0 للـ β_1 . إذا fail to reject H_0
↳ معناه $\beta_1 = 0$
we are in this case

if rejection of H_0
↳ $\beta_1 \neq 0$
we are in this case

* في كمان طريقة لأحل هاد التست ، إلي هي الـ analysis of var approach
[ANOVA]

* لو بدي أضيف فرع كالسؤال إلي مبناه : ceramic society
test for significance of regression provide p-value for this test.

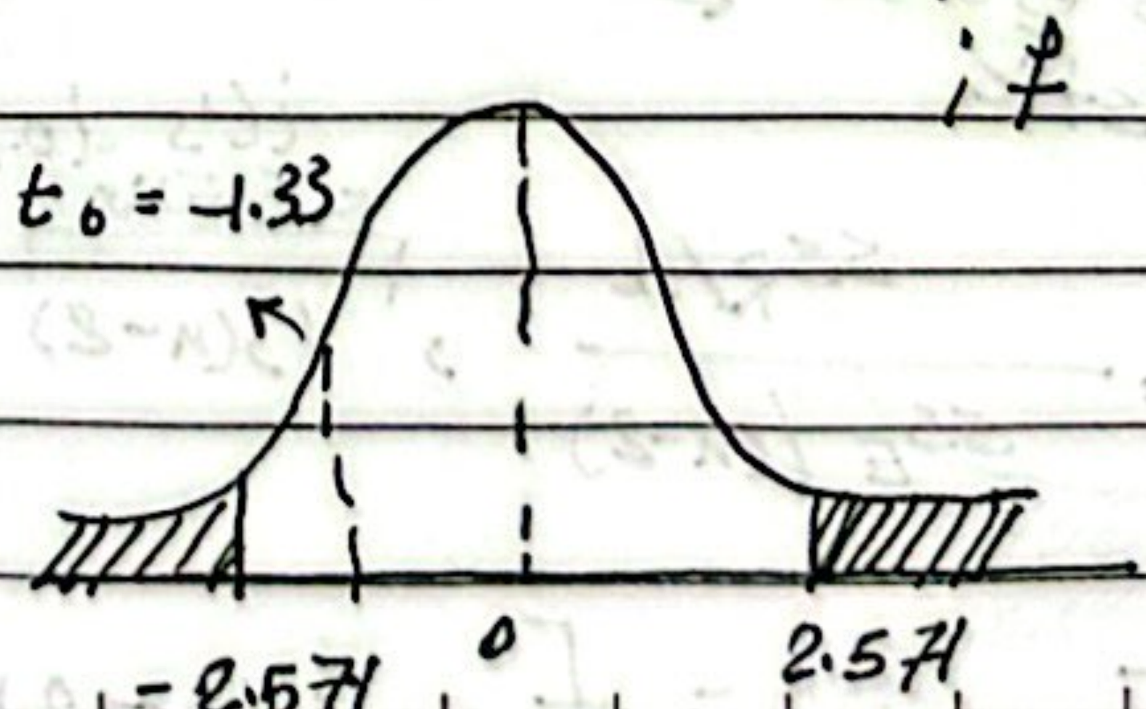
1) $H_0 : \beta_1 = 0$ $H_1 : \beta_1 \neq 0$	2) $t_{\alpha/2, n-2} = t_{0.025, 5} = -2.571$ $T_0 = \dfrac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\hat{\sigma}^2 / S_{xx}}} = \dfrac{-0.0342 - 0}{\sqrt{77.35 / 117142.86}}$	هدول حسبناهم already $\hat{\beta}_1 = -0.0342$ $\hat{\beta}_0 = 55.68$ $\hat{y} = 55.68 - 0.0342x$ $S_{xx} = 117142.86$ $\hat{\sigma}^2 = 77.35$

$t_0 = -1.33$ 3) fail to reject H_0

if $n = 7$, $\alpha = 0.05$

$0.1 < \dfrac{P\text{-value}}{2} < 0.25$

$0.2 < P \text{ value} < 0.5$



So, $\beta_1 = 0$
↳ يا العلاقة اقتران ثابت
↳ يا العلاقة من higher degree بس معامل X = صفر

ANOVA:

بدنا نقسم هون ال variability لقسمين

$$\sum_{i=1}^{n}(y_i-\bar{y})^2 = \sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2 + \sum_{i=1}^{n}(y_i-\hat{y}_i)^2$$

SS_T total sum of squares

SSR sum of regression

SS_E sum of error

يعني

$$y = \underbrace{\beta_0+\beta_1 x}_{\text{the regression component}} + \underbrace{\varepsilon}_{\text{the error component}}$$

the variability in y comes from 2 components (two sources)

identity: هاي اسمها ال

$$SS_T = \underbrace{SS_R}_{\text{بنطلع هاي}} + SS_E$$

بس أكمل ال model تاعي بطلع كل هدول

$$SS_T = \sum_{i=1}^{n}(y_i-\bar{y})^2 = \sum_{i=1}^{n} y_i^2 - n\bar{y}^2$$

بين السكشنز ماشية لأنه تعاملنا معها من قبل

أنا كم بعمل analysis لل variability ، بشوف شو ال sources إلهم ، وبحسبهم ، ففكرة ال ANOVA هي بإختصار partitioning :

المفروض التغيير الأكبر يكون ب SS_R ، لو كان ب SS_E معناه الموديل تبعي مش مناسب

بكون الموديل أحسن لو هاي كانت أكبر ، more adequate

هلا لو أقسم $\boxed{\frac{SS_E}{\sigma^2}}$ بعطيني chi-square var ، وال dof إله $= n-2$

$\boxed{\frac{SS_R}{\sigma^2}}$ // // // // // // // // // $= 1$ دائماً

لإنه بعمل hypothesis testing على one ind var

الصيغ لو نقسم $\frac{SSR/1}{SS_E/(n-2)}$ شو بعطي ؟ F_0

وهاي هي ال test statistic تبعنا

$$F_0 = \frac{SS_R/1}{SS_E/(n-2)} , \overbrace{1,(n-2)}^{\text{it's dof بالترتيب}}$$

$$F_0 = \frac{MSR}{MSE}$$

MSR → mean square regression

MSE → mean square error

$$F_{critical} = F_{\alpha, 1, (n-2)}$$

[ANOVA table]   * كل هاي المعلومات ، نحطها بجدول اسمه

Source of variation	sum of squares (SS)	Dof	mean square	F_0
Regression	$SS_R = \hat{\beta}_1 S_{xy}$	1	$MS_R \to \frac{SS}{dof}$	MS_R / MS_E
Error	$SS_E = SS_T - \hat{\beta}_1 S_{xy}$	$n-2$	$MS_E \to \frac{SS}{dof}$	
Total	$SS_T = \sum y_i^2 - n\bar{y}^2$	$n-1$		

* $\boxed{MS_E = \hat{\sigma}^2}$

F_c (critical) $= F_{\alpha, 1, n-2}$

* نفس مثال الـ ceramic بس بطريقة الـ ANOVA :

* الخطوات : 1) Identity
2) ANOVA table
3) hypothesis testing

محسوبين already:
$\hat{\beta}_1 = -0.0342$
$\hat{\beta}_0 = 55.68$
$S_{xx} = 117142.86$
$S_{xy} = -4001.4$

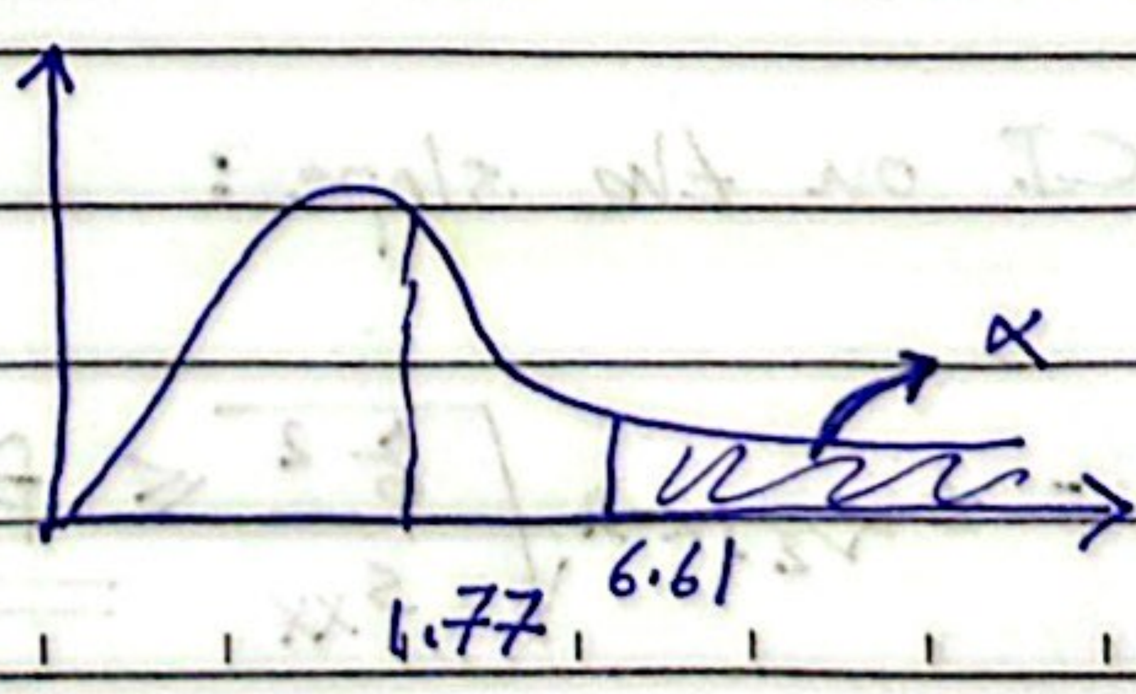
1]
$$SS_T = \sum y_i^2 - n\bar{y}^2 = 1737.74 - (7)(13.17)^2 = 523.6$$
$$SS_R = \hat{\beta}_1 S_{xy} = (-0.0342)(-4001.4) = 136.8$$
$$\to SS_E = 523.6 - 136.8 = 386.8$$

بعد ما أطلع هدول ، بعملهم الـ table .

using $n=7$ & $\alpha = 0.05$

Source	SS	Dof	mean squares	F_0
Reg	136.8	1	136.8	1.77
Error	386.8	5	77.36	
Total	523.6	6		

$F_c \to F_{0.05, 1, 5} = 6.61$



we fail to reject H_0

* Notes :

١) هون كم نفصل على linear regression متغير واحد .
٢) هون لإنه كم نفصل على estimation $\hat{\beta}_0$ و $\hat{\beta}_1$.
يعني 2 parameters .
٣) هاد الـ total بنعوم ، (جمعوم) .

Dof
١   1
٢   n-2
٣   n-1

لش؟؟ SS

* لازم الـ MSR تكون أكبر من الـ MS_E لأعتمد الموديل حقي و زادت قيمة f_0 ، يعني قربنا من منطقة الـ rejection أكثر ، بزيد رفضي لهاي الـ hypothesis ← $H_0 : \beta_1 = 0$

* تكملة من كفيفي :

* SS_R : regression sum of squares sometimes called "model" sum squares.
* SS_E : error sum of squares is called the "residual" " " " .

* $$T_0^2 = \frac{(\hat{\beta}_1)^2 S_{xx}}{\hat{\sigma}^2} = \frac{SSR}{MSE} = \frac{MSR}{\underbrace{MSE}_{}} = f_0$$

لإنه linear ، Dof = 1

هاي نفس صيغة T_0 ، بس ، حسنالها تربيع

$$T_0 = \frac{\hat{\beta}_1}{\sqrt{\hat{\sigma}^2 / S_{xx}}}$$

* الـ T بفحصنا الـ intercept .

$$\boxed{\therefore T_0^2 = f_0}$$

* CI on the slope and intercept :

1] CI on the slope :

$$\hat{\beta}_1 - t_{\alpha/2, n-2} \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}} \leq \beta_1 \leq \hat{\beta}_1 + t_{\alpha/2, n-2} \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}$$

2] CI on the intercept :

$$\hat{\beta}_0 - t_{\alpha/2, n-2}\underbrace{\sqrt{\hat{\sigma}^2\left[\frac{1}{n}+\frac{\bar{x}^2}{S_{xx}}\right]}}_{\text{standard error for the intercept}} \le \beta_0 \le \hat{\beta}_0 + t_{\alpha/2, n-2}\sqrt{\hat{\sigma}^2\left[\frac{1}{n}+\frac{\bar{x}^2}{S_{xx}}\right]}$$

standard error ← for the intercept

* CI on the mean response : $\hat{\mu}_{Y|x_0} = \hat{\beta}_0 + \hat{\beta}_1 x_0$

← من الجدول بحسبها

$$\hat{\mu}_{Y|x_0} - t_{\alpha/2, n-2}\sqrt{\hat{\sigma}^2\left[\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}\right]} \le \underbrace{\mu_{Y|x_0}}_{\text{(response)}} \le \hat{\mu}_{Y|x_0} + t_{\alpha/2, n-2}\sqrt{\hat{\sigma}^2\left[\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}\right]}$$

* section 11.6 :

Y_0 ≡ prediction of new observations → it's estimator is $\hat{Y}_0$

لعادي بس تعويض قيم X إلي معطينا

(Y بطلّع)

* بالرسم بنشوف إنه الأماكن القريبة على $\bar{X}$ تكون أضيّق (narrower) من الأطراف لإنه $\bar{X}$ كم تخلي البسط أكبر بما إنه بطرحها.

CI get narrower !

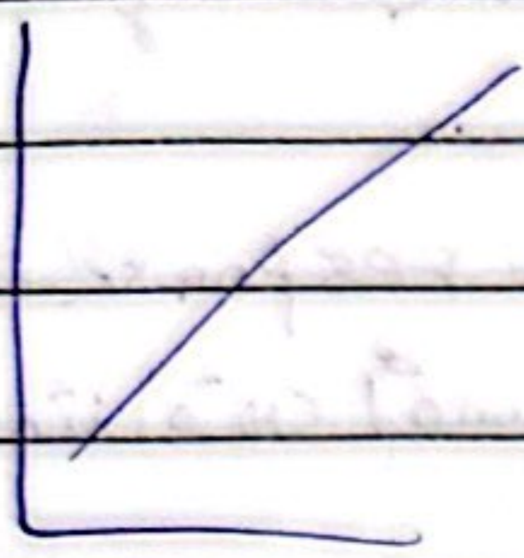
* Prediction interval on future observation Y_0 at the value x_0 is given :

$$\hat{y}_0 - t_{\alpha/2, n-2}\sqrt{\hat{\sigma}^2\left[1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}\right]} \le Y_0 \le \hat{y}_0 + t_{\alpha/2, n-2}\sqrt{\hat{\sigma}^2\left[1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}\right]}$$

CI get wider !

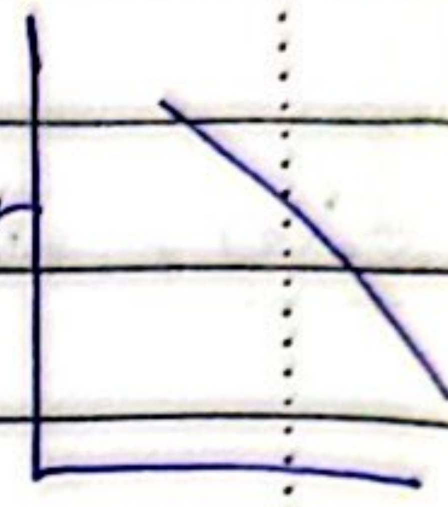
من ال (1) إلي ضفناه هون

* لما ما يكون في علاقة بين متغيرين نحكي uncorrelated



perfect (+) linear correlation

perfect (−) linear correlation



* طبعًا هاد مستحيل ، دائمًا راح يكون في error ، فبنوصف العلاقة بدون كلمة perfect .

* strength of the linear correlation:

1) "r" →
↳ بيعطينا كمية وإشارة

$r = 1$ perfect (+)

$r = -1$ perfect (−)

$r = 0$
↳ no linear relationship

$0 \le r \le 1$

total variation in y = Variation due to regression model + var due to randomness

{ $\epsilon_i \sim N(0, \sigma^2)$ — بتخلي ال y موزعة Normal

Total var in y =

$y = \beta_0 + \beta_1 x + \epsilon$ (real y)

$\hat{y} = \beta_0 + \beta_1 x$ (estimated y)

$y - \hat{y} = \epsilon$ (real − Estimated)

بنا بعمل estimation لقيم β_0 و β_1 ← بصيرو $\hat{\beta}_0$ و $\hat{\beta}_1$

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + e_i$$

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

$$e_i = y_i - \hat{y}_i$$

* Sum of squares $\equiv$ Variability

* $r \longrightarrow$ pearson' correlation coefficient

$$\boxed{r = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}}}$$

هاي إلي بتتحكم بإشارة الـ r

no unit for r

Sign for r is the same for $\hat{\beta}_1$

$$\boxed{R^2 = \frac{\hat{\beta}_1 \, S_{xy}}{S_{yy}}}$$ (coefficient of determination)

$$\boxed{\frac{S.S_{Reg}}{S_{yy}}}$$

القيم إلي بتطلع هون (R^2): هي بكون كم قدرة أفسّر الـvariation.

section 11.7 :

Residual analysis : $e_i = y_i - \hat{y}_i$

actual ←   → من regression بعد ما نعوض قيم x بالموديل ، بكون عنا داتا بجدول

The normality test :

Normal probability plot ⟶ hypothesis testing

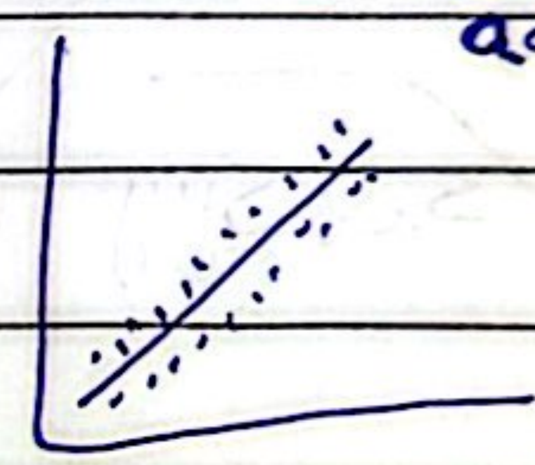
H_0: The residuals are normaly distributed

H_1: " " " " not " " "

بنرسم plot حسب الداتا ، إذا كانو كتير قريبين حال

$N \sim$ lo Lies best fit line

adequate

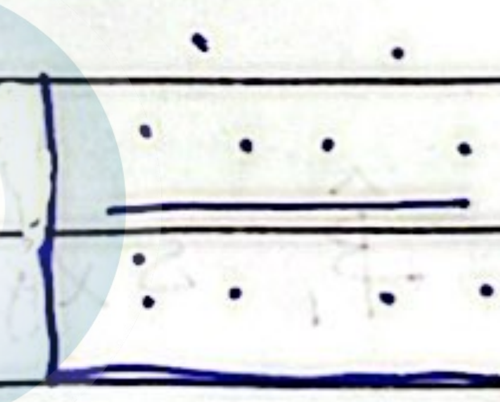


failed to reject H_0

⟷ so it's $\sim N$ & p-value is large

Stdev for the error is $\hat{\sigma}$ & $\hat{\sigma}^2$ is $\frac{SS_E}{n-2}$

not adequate



coefficient of determination (R^2) :

⟶ زي ال residual analysis بس ⟶ أقوى

$$\boxed{R^2 = \frac{SS_R}{SS_T} \text{ or } 1 - \frac{SS_E}{SS_T}}$$

$0 \leq R^2 \leq 1$

← قديش قدر يفسر ال variability بالموديل

يفسّر ≡ accounts

* Section 11.8)

bivariate → متغير بتغير بمتغير تاني

* Correlation coefficient (ص) دوه

$\boxed{ص = \frac{\sigma_{xy}}{\sigma_x \sigma_y}}$ → Covarience $\sigma_{xy} = E\,(x-\mu_x)(y-\mu_y)$

or → + → $x\uparrow \therefore y\uparrow$

→ − → $x\uparrow \therefore y\downarrow$

→ $ص = \sqrt{R^2}$

design and analysis of singe - factor experiment :
the analysis of variance . one ind Var effect يعني → one dep var

* Section 13.2) ex 13-2.1 (سلايدس) :

table 13-1)

---

* chapter 14 : *أول إشي إحنا كم ندرس 2 factors هون مبدئيًا وبعدين راح نخليهم أكثر .

* 2 factors

A : a → levels ← ال index الها i

B : b ← ال index الها j

*نفترض بدنا ندرس العوامل إلي خلّو هاي البلاطة تصير مقعّرة وإلي هي ال Temp و ال speed عن طريق نقيس مسافة ½

		factor A: "speed"	
		L = 10 m/s	H = 20 m/s
factor B "temp"	L 100c	3,4	5,6 → sample2
	H	7,8 (sample 1)	9,10 ↑ المسافات

*هل الفاكتور الأول significant ؟ وهل الثاني ؟

*راح يكون عنا 2 CIs ، لكل عامل واحدة

A: $H_0 : \tau_1 = \tau_2 \ldots\ldots \tau_a = 0$ → if not significant

$H_1 : \tau_i \neq 0$ atleast 1 i not equal , $i = 1, \ldots\ldots a$

B: $H_0 : \beta_1 = \beta_2 \ldots \beta_b = 0$

$H_1 : \beta_j \neq 0$, $j = 1, 2, \ldots b$

for at least one j

* our anova table

y_{ij} ⟶ معناها عندي 2 factors

⟶ بس لازم نغطي ⟶ y_{ijk} (our data point) ال include كمان نعمل samples

* يعني مثلاً، y_{111} معناها first row for the first factor, first column for the second factor, first sample

يعني من الجدول إلي عملناه ، القيمة ٣

Ex: $y_{222} \rightarrow 10$

So $y_{ijk} = \mu + \tau_i + \beta_j + \epsilon_{ijk}$,
$i = 1, 2, \dots a$
$j = 1, 2, \dots b$
$k = 1, 2, \dots n$

μ: the grand mean; τ_i: التأثير من ال factor الأول; β_j: التأثير من ال factor الثاني

* بهاد المثال عنا $2 = n = b = a$

* كل data point راح يكون إلها ijk
وكل تقاطع بالجدول هو treatment compination

Source of variation
factor A
factor B
—
Error
Total

TC ١	TC
TC	TC

* الفاكتور إلي تركناه هولا : factor ~~effect~~

رمزه كابيتال

A = avg of response at high level − avg of resp at low level

B = // − //

Scanned with CamScanner

Source	SS	DoF	MS	F_0	p-value
A					
B					
Error					
Total					

* كيف نحسب ال factor effect ؟

$$A = \frac{6+5+9+10}{4} - \frac{3+4+7+8}{4}$$

$$\frac{30}{4} - \frac{22}{4} \rightarrow 2$$

Factor B \ Factor A	L	H
L	3, 4	5, 6
H	7, 8	9, 10

$$B = \frac{7+8+9+10}{4} - \frac{3+4+5+6}{4}$$

$$\frac{34}{4} - \frac{18}{4} \rightarrow 4$$

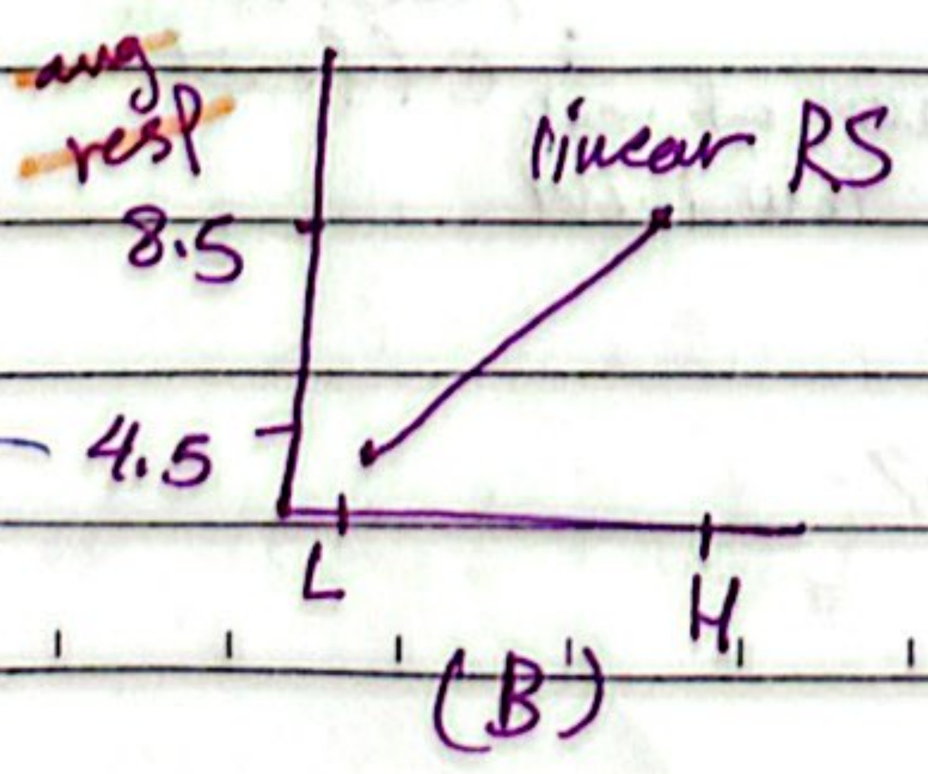
4 data points

هاي القيم هي ال slope

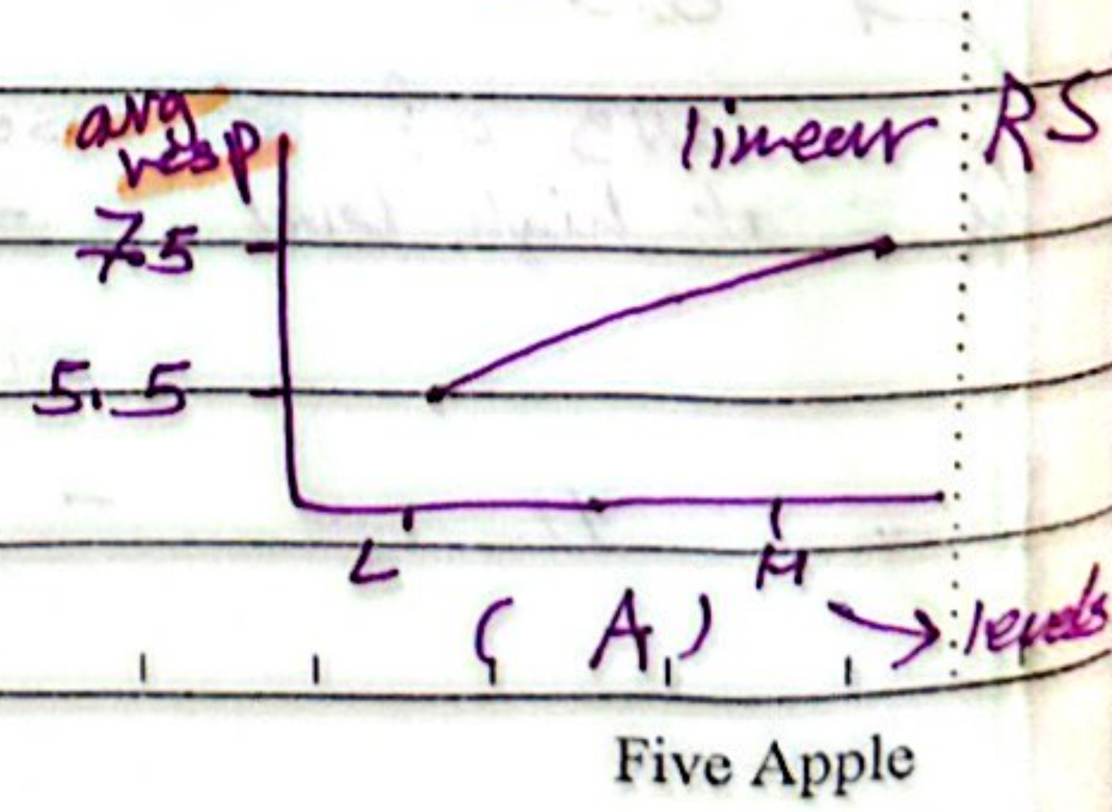
slope B > A

* معنى رقم ٤ إنه التأثير كان ع ال response
ب B أكثر من A (بنقارن حسب كمية وإشارة)
يعني B ← more significant كان لما انتقلنا
من ال High لل low أو العكس.
بعطيني (indication) عن نتيجة الأنوفا.

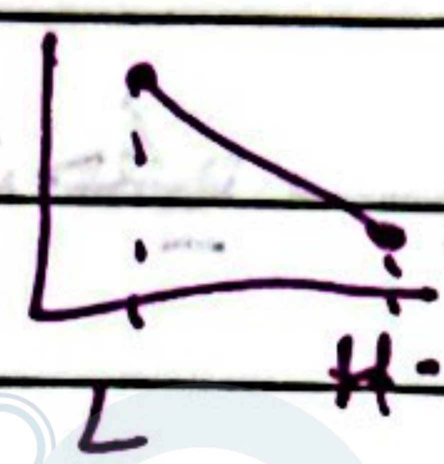
* شو معنى الإشارة؟ كنا رسمة اسمها ~~factor plot~~ main effect ← لإنه كل factor لحاله



(B)



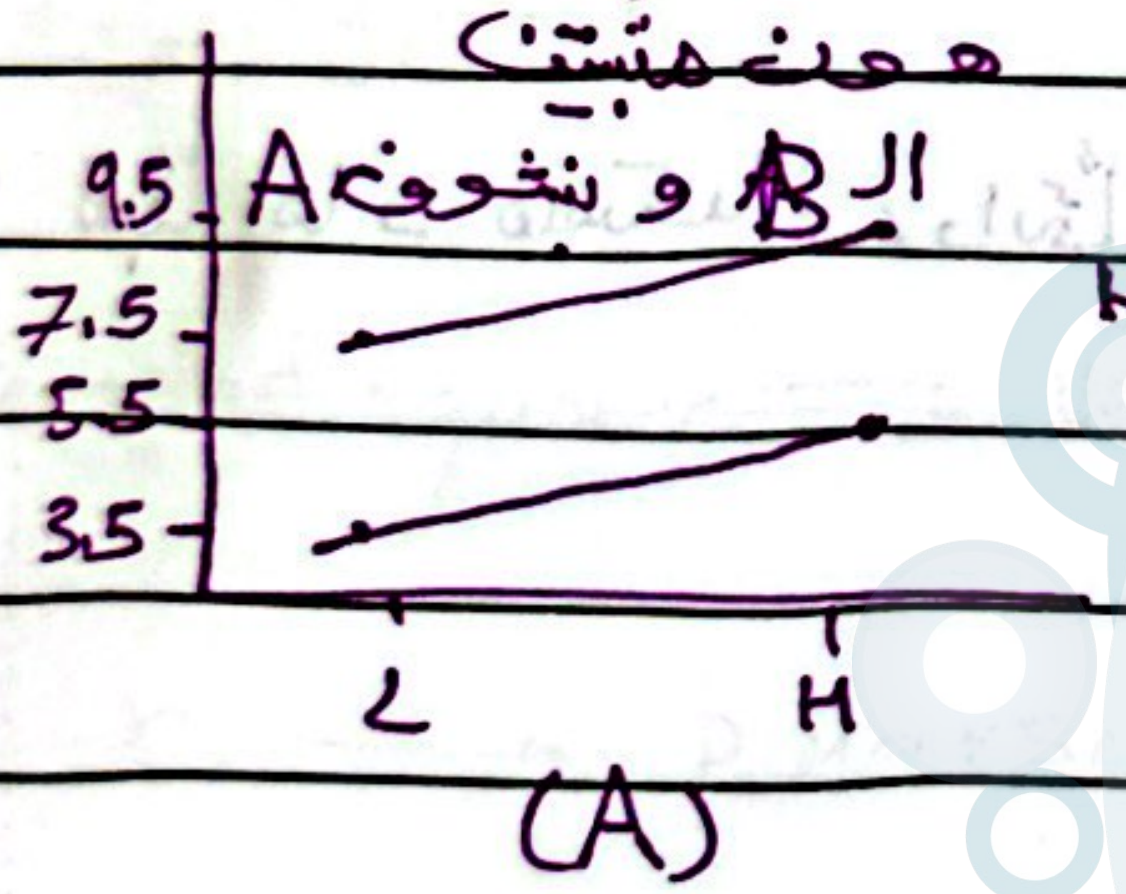
(A)

لو ال sign (−) معناه  معناها لما أختر بال... من low factor

ل high ← ال response بقل.

● ال response هنا كان تجيب التقعر، إلي هو كل ما قل بكون أحسن، يعني المتناقص أحسن .

● في هنا إشي اسمه interaction plot



هون منبتي ال A و بنشوف B

● بنشوف A و B سوا، ال avg ل (L L)

كان ل ($\frac{3+4}{2}$) ← ٣.٥

($\frac{5+6}{2}$) ← 5.5

حيث ال high وال low وكل تقاطعاتهم ل A و B بنفس الرسمة .

Design and Analysis of Single-Factor Experiments: The Analysis of Variance

CHAPTER OUTLINE

LEARNING OBJECTIVES

After careful study of this chapter, you should be able to do the following:

1. Design and conduct engineering experiments involving a single factor with an arbitrary number of levels
2. Understand how the analysis of variance is used to analyze the data from these experiments
3. Assess model adequacy with residual plots
4. Use multiple comparison procedures to identify specific differences between means
5. Make decisions about sample size in single-factor experiments
6. Understand the difference between fixed and random factors
7. Estimate variance components in an experiment involving random factors
8. Understand the blocking principle and how it is used to isolate the effect of nuisance factors
9. Design and conduct experiments involving the randomized complete block design

13-1 Designing Engineering Experiments

Every experiment involves a sequence of activities:

تخمين

1. Conjecture – the original hypothesis that motivates the experiment.
2. Experiment – the test performed to investigate the conjecture.
3. Analysis – the statistical analysis of the data from the experiment.
4. Conclusion – what has been learned about the original conjecture from the experiment. Often the experiment will lead to a revised conjecture, and a new experiment, and so forth.

13-2 The Completely Randomized Single-Factor Experiment

13-2.1 An Example

A manufacturer of paper used for making grocery bags is interested in improving the tensile strength of the product. Product engineering thinks that tensile strength is a function of the hardwood concentration in the pulp and that the range of hardwood concentrations of practical interest is between 5 and 20%. A team of engineers responsible for the study decides to investigate four levels of hardwood concentration: 5%, 10%, 15%, and 20%. They decide to make up six test specimens at each concentration level, using a pilot plant. All 24 specimens are tested on a laboratory tensile tester, in random order. The data from this experiment are shown in Table 13-1.

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$

$H_1: \mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4$] → at least 2

OR anova

13-2 The Completely Randomized Single-Factor Experiment

13-2.1 An Example

as anotation

تقسيمة

$y_{i.} = \sum_j y_{ij}$

↳ the dot means the sum

Table 13-1 Tensile Strength of Paper (psi)

sample size

Hardwood Concentration (%)	Observations 1	2	3	4	5	6	Totals (مجموع ال obser)	Averages ($\bar{y}_{i.}$)
5 (pop₁)	7	8	15	11	9	10	60 $y_{1.}$	10.00
10 (pop₂)	12	17	13	18	19	15	94 $y_{2.}$	15.67
15	14	18	19	17	16	18	102 $y_{3.}$	17.00
20	19	25	22	23	18	20	127 $y_{4.}$	21.17
							$\Sigma =$ 383	15.96

single factor with 4 levels

treatement 1 2 3 4

$\sum_{j=1}^{6} y_{1j}$

$\mu_1 \rightarrow$ تأثير $\tau_1 = 15.96 - 10 = 5.96$

averages for each level

overall mean ↳ يعطي إنطباع عن كل ال data (Mean pop)

ال observation ال row الأول هو sample وإله μ_1 (يعني إله μ خاص فيها)

↳ الإنحرافات هي أساس ال anova table

$SS_T = SS_{treat} + SS_E$

$SS_{total} = (7)^2 + (8)^2 \ldots\ldots (20)^2 - \frac{(383)^2}{4 \times 6} = 612.16$

$SS_{treat} = \frac{60^2 + 94^2 + \ldots\ldots 127^2}{6} - \frac{(383)^2}{4 \times 6} = 382.79$

$SS_{error} = SS_{total} - SS_{treat} = 130.17$

	SS	Dof	Ms	Fo
treatement	382.79	3	127.59	19.59
error	130.17	20	6.51	
Total	512.96	23		

13-2 The Completely Randomized Single-Factor Experiment

13-2.1 An Example

- The levels of the factor are sometimes called **treatments**.

ما غيرنا ال conditions

تكرار

- Each treatment has six observations or **replicates**.

= n

- The runs are run in **random** order.

إلي همه البيانات فوق إنه 24

data points

13-2 The Completely Randomized Single-Factor Experiment

13-2.1 An Example

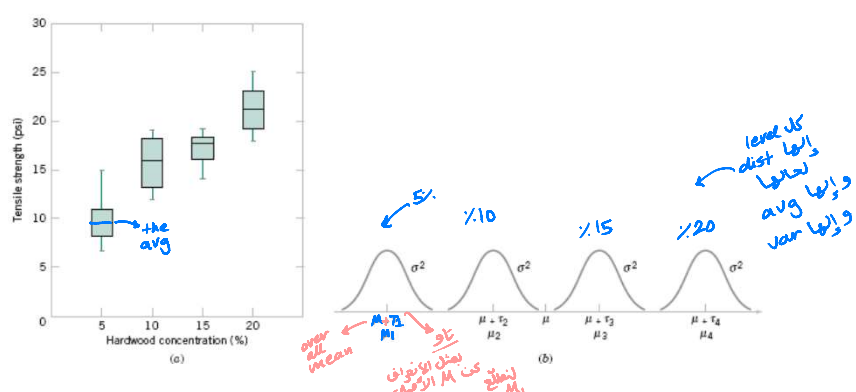


Figure 13-1 (a) Box plots of hardwood concentration data. (b) Display of the model in Equation 13-1 for the completely randomized single-factor experiment

13-2 The Completely Randomized Single-Factor Experiment

الصيغة العامة إله هو الجدول إلي تحت

13-2.2 The Analysis of Variance

Suppose there are a different levels of a single factor that we wish to compare. The levels are sometimes called **treatments**.

الفكرة إنه عملنا تجربة وطلع منها كل الـ observations، هلأ بدنا نعملها تحليل كالـ ANOVA

Table 13-2 Typical Data for a Single-Factor Experiment

Treatment	Observations				Totals	Averages
1	y_{11}	y_{12}	$\cdots$	y_{1n}	$y_{1\cdot}$	$\bar{y}_{1\cdot}$
2	y_{21}	y_{22}	$\cdots$	y_{2n}	$y_{2\cdot}$	$\bar{y}_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
a	y_{a1}	y_{a2}	$\cdots$	y_{an}	$y_{a\cdot}$	$\bar{y}_{a\cdot}$
					$y_{\cdot\cdot}$	$\bar{y}_{\cdot\cdot}$

مج Dot means sum

$\frac{y_{1\cdot}}{n}$

13-2 The Completely Randomized Single-Factor Experiment

13-2.2 The Analysis of Variance

الـ row الواحد في ذي بيجم الـ observations و بختلفوا عن الـ μ بتعتهم

We may describe the observations in Table 13-2 by the linear statistical model:

random error within the same treatement

لكل row

overall mean

treatements

$$Y_{ij} = \mu + \tau_i + \epsilon_{ij} \begin{cases} i = 1, 2, \ldots, a \\ j = 1, 2, \ldots, n \end{cases}$$

أي observations بالجدول

كل row وإله μ ولأنه جاي من pop مختلف

بسبب الـ level

observations

where y_{ij} is a random variable denoting the (ij) th observation, μ is a parameter common to all treatments called the **overall mean**, τ_i is a parameter associated with the i th treatment called the i th **treatment effect**, and ε_{ij} is a random error component.

لأنه ناتج عن إنحراف عن level i الانحراف الـ observation عن الـ overall mean

The model could be written as

$$Y_{ij} = \mu_i + \epsilon_{ij} \begin{cases} i = 1, 2, \ldots, a \\ j = 1, 2, \ldots, n \end{cases}$$

$\mu + \tau_i$

13-2 The Completely Randomized Single-Factor Experiment

13-2.2 The Analysis of Variance

conclusion effects the level only

Fixed-effects Model

The treatment effects are usually defined as deviations from the overall mean so that:

$$\sum_{i=1}^{a} \tau_i = 0$$

الـ dot عملت Replace على الـ j لعاد الـ ∑ بتكون للـ j

Also,

$$y_{i\cdot} = \sum_{j=1}^{n} y_{ij} \qquad \bar{y}_{i\cdot} = y_{i\cdot}/n \qquad i = 1, 2, \ldots, a$$

الصف

$$y_{\cdot\cdot} = \sum_{i=1}^{a} \sum_{j=1}^{n} y_{ij} \qquad \bar{y}_{\cdot\cdot} = y_{\cdot\cdot}/\underline{\underline{N}}$$

$\hookrightarrow a * n$

∑ لكل الـ treatments

13-2 The Completely Randomized Single-Factor Experiment

13-2.2 The Analysis of Variance

the ANOVA

We wish to test the hypotheses:

إنه فش انحرافات، كلهم قد بعض ويساوو صفر

$$H_0: \tau_1 = \tau_2 = \cdots = \tau_a = 0$$
$$H_1: \tau_i \neq 0 \quad \text{for at least one } i$$

لو انحراف واحد لا يساوي صفر، الـ H الصح

The analysis of variance partitions the total variability into two parts.

بقسمو الـ variability بالـ treatment والـ error

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$

$H_1:$ at $\mu_i \neq \mu_j$ for at least

	SS	Dof	MS
treatment	SS_{treat}	$a-1$	$SS_{treat}/(a-1) = MS_{treat}$
error	SS_E	$a(n-1)$	$SS_E / a(n-1) = MS_E$
Total	SS_{Total}	$(an-1)$	

The anova table

MS_E: Point estimator for $\to \sigma^2$

$F_0 = \dfrac{MS\ treat}{MS\ error}$

تقارنهم ببعض $\hookrightarrow f_{\alpha, (a-1), a(n-1)}$

2/27/2024

مهم

13-2 The Completely Randomized Single-Factor Experiment

13-2.2 The Analysis of Variance

Definition → اختصارها ANOVA

The sum of squares identity is

$$\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{..})^2 = n\sum_{i=1}^{a}(\bar{y}_{i.}-\bar{y}_{..})^2 + \underbrace{\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i.})^2}_{\text{error sum of square}} \qquad (13\text{-}5)$$

الفرق التربيع

$\dfrac{y_{..}}{an} = \bar{y}_{..}$

↳ treatement sum of square

or symbolically

$$SS_T = SS_{\text{Treatments}} + SS_E \qquad (13\text{-}6)$$

total sum of square من المشاهدة الأولى لعدد 24

the level

بحسبوا بالطرح

13-2 The Completely Randomized Single-Factor Experiment

13-2.2 The Analysis of Variance

The expected value of the treatment sum of squares is → if null is true

expected value

error ال تحت بتعق

$$E(SS_{\text{Treatments}}) = (a-1)\sigma^2 + n\sum_{i=1}^{a}\tau_i^2$$

→ if H_1 is true → treatment الإنحرافات يلي سببها الإنتقال من تاني مضروبه بـ n

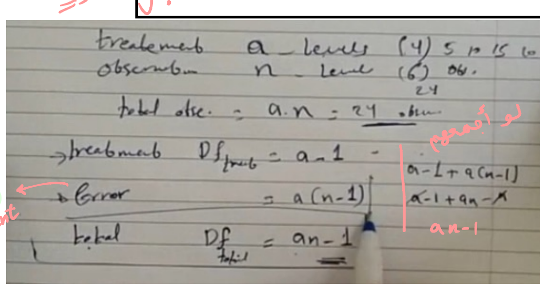
and the expected value of the error sum of squares is

$$E(SS_E) = a(n-1)\sigma^2$$

dof for error ↳ $MS_E = \hat{\sigma}^2 = \dfrac{SS_E}{Dof_E}$ أصلها

The ratio $MS_{\text{Treatments}} = SS_{\text{Treatments}}/(a-1)$ is called the **mean square for treatments**.

1) why?



total - treatment

13-2 The Completely Randomized Single-Factor Experiment

13-2.2 The Analysis of Variance

The appropriate test statistic is

$$F_0 = \frac{SS_{\text{Treatments}}/\overbrace{(a-1)}^{\text{treatment's dof}}}{SS_E/\underbrace{[a(n-1)]}_{\text{error's dof}}} = \frac{MS_{\text{Treatments}}}{MS_E} \qquad (13\text{-}7)$$

الإنحراف إلي ناتج بسبب الإنتقال بين ال treatments

إنحرافات within the treatment

We would reject H_0 if $f_0 > f_{\alpha,a-1,a(n-1)}$

البسط معناها أكبر من المقام بكثير يعني راح نعمل reject إنه الانحرافات بالبسط ≠ صفر

المقارنة

13-2 The Completely Randomized Single-Factor Experiment

13-2.2 The Analysis of Variance

Definition

The sums of squares computing formulas for the ANOVA with equal sample sizes in each treatment are

$$SS_T = \sum_{i=1}^{a} \sum_{j=1}^{n} y_{ij}^2 - \frac{y_{..}^2}{N} \qquad (13\text{-}8)$$

$\rightarrow$ an $\rightarrow$ the sum

and

$$SS_{\text{Treatments}} = \sum_{i=1}^{a} \frac{y_{i.}^2}{n} - \frac{y_{..}^2}{N} \qquad (13\text{-}9)$$

The error sum of squares is obtained by subtraction as

$$SS_E = SS_T - SS_{\text{Treatments}} \qquad (13\text{-}10)$$

13-2 The Completely Randomized Single-Factor Experiment

13-2.2 The Analysis of Variance

Analysis of Variance Table

the ANOVA table

Table 13-3 The Analysis of Variance for a Single-Factor Experiment, Fixed-Effects Model

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Treatments	$SS_{\text{Treatments}}$	$a - 1$	$MS_{\text{Treatments}}$	$\frac{MS_{\text{Treatments}}}{MS_E}$
Error	SS_E	$a(n - 1)$	MS_E	$\frac{SS_{Error}}{dof}$
Total	SS_T	$\underline{an} - 1$		

N

13-2 The Completely Randomized Single-Factor Experiment

Example 13-1

Consider the paper tensile strength experiment described in Section 13-2.1. We can use the analysis of variance to test the hypothesis that different hardwood concentrations do not affect the mean tensile strength of the paper.

The hypotheses are

$$H_0\colon \tau_1 = \tau_2 = \tau_3 = \tau_4 = 0$$
$$H_1\colon \tau_i \neq 0 \text{ for at least one } i \quad .$$

13-2 The Completely Randomized Single-Factor Experiment

Example 13-1

We will use $\alpha = 0.01$. The sums of squares for the analysis of variance are computed from Equations 13-8, 13-9, and 13-10 as follows:

$$SS_T = \sum_{i=1}^{4} \sum_{j=1}^{6} y_{ij}^2 - \frac{y_{..}^2}{N}$$
$$= (7)^2 + (8)^2 + \cdots + (20)^2 - \frac{(383)^2}{24} = 512.96$$
$$SS_{\text{Treatments}} = \sum_{i=1}^{4} \frac{y_{i.}^2}{n} - \frac{y_{..}^2}{N}$$
$$= \frac{(60)^2 + (94)^2 + (102)^2 + (127)^2}{6} - \frac{(383)^2}{24} = 382.79$$
$$SS_E = SS_T - SS_{\text{Treatments}}$$
$$= 512.96 - 382.79 = 130.17$$

13-2 The Completely Randomized Single-Factor Experiment

Example 13-1

The ANOVA is summarized in Table 13-4. Since $f_{0.01,3,20} = 4.94$, we reject H_0 and conclude that hardwood concentration in the pulp significantly affects the mean strength of the paper. We can also find a P-value for this test statistic as follows:

$$P = P(F_{3,20} > 19.60) \simeq 3.59 \times 10^{-6}$$

Since $P \simeq 3.59 \times 10^{-6}$ is considerably smaller than $\alpha = 0.01$, we have strong evidence to conclude that H_0 is not true.

Table 13-4 ANOVA for the Tensile Strength Data

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	f_0	P-value
Hardwood concentration	382.79	3	127.60	19.60	3.59 E-6
Error	130.17	20	6.51		
Total	512.96	23			

$P(F_{3,20} > 19.6)$

بجدول ال F_{dist}

$$SS_T = \sum_{i=1}^{4}\sum_{j=1}^{6} y_{ij}^2 - \frac{y_{..}^2}{N}$$

$$= 7^2 + 8^2 + 15^2 + \dots + 20^2 - \frac{(383)^2}{24}$$

$$SS_T = 512.96$$

$$SS_{treatment} = \sum_{i=1}^{4} \frac{y_{i.}^2}{n} - \frac{y_{..}^2}{N}$$

$$= \frac{60^2}{6} + \frac{94^2}{6} + \frac{102^2}{6} + \frac{127^2}{6} - \frac{383^2}{24}$$

$$SS_{treatment} = \frac{60^2 + 94^2 + 102^2 + 127^2}{6} - \frac{383^2}{24}$$

$SS_{treatment} =$

بنقدر نعمل ال anova table بعد ما نطلع هدول

$$SS_{treatment} = \frac{60^2 + 94^2 + 102^2 + 127^2}{6} - \frac{383^2}{24}$$

$$SS_{treatment} = 382.79$$

$$SS_E = SS_T - SS_{treatment} = 512.96 - 382.79 = 130.17$$

Source of variation	DF	SS	MS	F_0	P-value
treatment (btw)	$a-1$ 3	SS_{tr} 382.79	127.6	19.6	
error (within)	$a(n-1)$ 20	SS_E 130.17	6.51		
total	$N-1$ 23	SS_T 512.96			

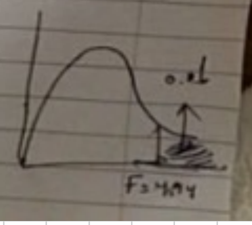
P-value $= P(F_{3,20} > 19.6) = \lll 0.01$

$F_{0.01,3,20} = 4.94$

if $f_0 > f_{\alpha, v_1, v_2}^{a-1,\, a(n-1)}$

$F_{0.01,3,20}$

$19.6 \gg 4.94$



$H_0: \tau_1 = \tau_2 = \tau_3 = \tau_4 = 0$

H_a: At least $\tau_i \neq 0$

fail to reject.

Hardwood conc. is signif.

Table 13-5 Minitab Analysis of Variance Output for Example 13-1

One-Way ANOVA: Strength versus CONC

Analysis of Variance for Strength

Source	DF	SS	MS	F	P
Conc	3	382.79	127.60	19.61	0.000
Error	20	130.17	6.51		
Total	23	512.96			

Individual 95% CIs For Mean Based on Pooled StDev

Level	N	Mean ($\bar{y}_{1.}$)	StDev
5	6	10.000	2.828
10	6	15.667	2.805
15	6	17.000	1.789
20	6	21.167	2.639

Pooled StDev = 2.551

10.0 15.0 20.0 25.0

Fisher's pairwise comparisons

Family error rate = 0.192
Individual error rate = 0.0500

Critical value = 2.086

Intervals for (column level mean) − (row level mean)

	5	10	15
10	−8.739 −2.594		
15	−10.072 −3.928	−4.406 1.739	
20	−14.239 −8.094	−8.572 −2.428	−7.239 −1.094

diff btw 2 means → كل واحد pop

13-2 The Completely Randomized Single-Factor Experiment

Definition

رأحسب ال CI لكل level لحال

A $100(1-\alpha)$ percent confidence interval on the mean of the ith treatment μ_i is

$$\bar{y}_{i\cdot} - t_{\alpha/2,a(n-1)}\sqrt{\frac{MS_E}{n}} \le \mu_i \le \bar{y}_{i\cdot} + t_{\alpha/2,a(n-1)}\sqrt{\frac{MS_E}{n}} \qquad (13\text{-}11)$$

observation N avg ال لكل row (factor levels)

For 20% hardwood, the resulting confidence interval on the mean is

$$19.00 \text{ psi} \le \mu_4 \le 23.34 \text{ psi}$$

$(\bar{y}_{1.} - \bar{y}_{2.}) \rightarrow$ Point estimator for $\mu_1 - \mu_2$

for treatment 1

$$t_0 = \frac{\bar{y}_{1.} - \mu_i}{\sqrt{\frac{MS_E \, \hat{\sigma}^2}{n}}}$$

treatment 1 $\quad t_0 = \dfrac{\bar{y}_{1.} - \mu_i}{\sqrt{\dfrac{MS_E\ \hat{\sigma}^2}{n}}}$ $\quad H_0: \mu_i = \mu$

CI for treatment 1
95 % CI that μ_1 is between

$$\bar{y}_{1.} - t_{\frac{\alpha}{2},20} \cdot \frac{\sigma}{\sqrt{n}} \le \mu_1 \le \bar{y}_{1.} + t_{\frac{\alpha}{2}} \cdot \frac{\sqrt{MS_E}}{\sqrt{n}}$$

$$10 - t_{0.025,20} \cdot \sqrt{\frac{6.51}{6}} \le \mu_1 \le 10 + t_{0.025,20} \sqrt{\frac{6.51}{6}}$$

$$\le \mu_1 \le 10 + 2.086 \times \sqrt{\frac{6.51}{6}}$$

$$\boxed{7.83 \le \mu_1 \le 12.173}$$

for one treatment

Randomized Single-

Definition

only for CI, we use t

A 100(1 − α) percent confidence interval on the difference in two treatment means $\mu_i - \mu_j$ is

formula هاي بقدر أطبقها على أي 2 means
بس لازم ألاقي MS_E

$$\bar{y}_{i\cdot} - \bar{y}_{j\cdot} - t_{\alpha/2,a(n-1)}\sqrt{\frac{2MS_E}{n}} \le \mu_i - \mu_j \le \bar{y}_{i\cdot} - \bar{y}_{j\cdot} + t_{\alpha/2,a(n-1)}\sqrt{\frac{2MS_E}{n}} \qquad (13\text{-}12)$$

For the hardwood concentration example,

$$-1.74 \le \mu_3 - \mu_2 \le 4.40$$

13-2 The Completely Randomized Single-Factor Experiment

An Unbalanced Experiment

عدد ال observation بكل row مختلفين

بكل treatment ما أخذت نفس عدد ال replicates
لـ مثلاً بار 5% أخذ 6 مشاهدات
بار 10% أخذ 5 وهكذا

The sums of squares computing formulas for the ANOVA with unequal sample sizes n_i in each treatment are

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{n_i} y_{ij}^2 - \frac{y_{..}^2}{N} \qquad (13\text{-}13)$$

$N = n_1 + n_2 \dots n_a$

$a\underline{n}$ هون متغيرة

(مجموعهم لحد n_a)

$$SS_{\text{Treatments}} = \sum_{i=1}^{a} \frac{y_{i\cdot}^2}{n_i} - \frac{y_{..}^2}{N} \qquad (13\text{-}14)$$

عدد مشاهدات كل y

and

$$SS_E = SS_T - SS_{\text{Treatments}} \qquad (13\text{-}15)$$

$H_0: \mu_i = \mu_j \quad (\mu_i - \mu_j = 0)$

$H_1: \mu_i \neq \mu_j$

$$t = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_i - \mu_j)}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \qquad \sqrt{MS_E}$$

$H_0: \mu_i = \mu_j \rightarrow CI$

$H_1: \mu_i \neq \mu_j$

$\mu_i \quad \bar{y}_{i.} \quad \bar{y}_{1.}$

$\mu_j \quad \bar{y}_{j.} \quad \bar{y}_{2.}$

$$t_0 = \frac{\bar{y}_{i.} - \bar{y}_{j.} - 0}{\sqrt{\frac{MS_E}{n} + \frac{MS_E}{n'}}} = \frac{\bar{y}_{i.} - \bar{y}_{j.}}{\sqrt{\frac{2MS_E}{n}}}$$

$$\bar{y}_{i.} - \bar{y}_{j.} - t_{\frac{\alpha}{2}, a(n-1)}\sqrt{\frac{2MS_E}{n}} \leq \mu_i - \mu_j \leq \bar{y}_{i.} - \bar{y}_{j.} + t_{\frac{\alpha}{2}, a(n-1)}\sqrt{\frac{2MS_E}{n}}$$

هاي القيمة هي ال LSD

...letely Randomized Single- ...ment

لو عندي 4 مشاهدات ، فأنا بدي أقارن الاولى مع ٢ بعدين ١ مع ٣ ١ مع ٤ ، ٢ مع ١ .... وهكذا

13-2.3 Multiple Comparisons Following the ANOVA

يعني زي 2 pop ، نفس شابتر ١٠

The least significant difference (LSD) is

بتخليني أشوف بالزبط وين صار إنحراف بجيبو.

$$\text{LSD} = t_{\alpha/2, a(n-1)}\sqrt{\frac{2MS_E}{n}} \qquad (13\text{-}16)$$

dof for ε

ال error اي مسموح بال CI

If the sample sizes are different in each treatment:

in the unbalanced case

$$\text{LSD} = t_{\alpha/2, N-a}\sqrt{MS_E\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$$

$n_1 + n_2 \ldots$ بضيفهم كلهم حسب كم tretement كان

13-2 The Completely Randomized Single-Factor Experiment

Example 13-2

We will apply the Fisher LSD method to the hardwood concentration experiment. There are $a = 4$ means, $n = 6$, $MS_E = 6.51$, and $t_{0.025,20} = 2.086$. The treatment means are

$$\bar{y}_{1\cdot} = 10.00 \text{ psi}$$
$$\bar{y}_{2\cdot} = 15.67 \text{ psi}$$
$$\bar{y}_{3\cdot} = 17.00 \text{ psi}$$
$$\bar{y}_{4\cdot} = 21.17 \text{ psi}$$

The value of LSD is $\text{LSD} = t_{0.025,20}\sqrt{2MS_E/n} = 2.086\sqrt{2(6.51)/6} = 3.07$. Therefore, any pair of treatment averages that differs by more than 3.07 implies that the corresponding pair of treatment means are different.

$$\text{LSD} = t_{\frac{\alpha}{2}, a(n-1)}\sqrt{\frac{2MS_E}{n}} \qquad 4(6-1)$$

$$= t_{0.025,20} \times \sqrt{\frac{2 \times 6.51}{6}}$$

$$\text{LSD} = 2.086 \times - = 3.07$$

$$MS_E = \sigma^2$$

13-2 The Completely Randomized Single-Factor Experiment

Example 13-2

The comparisons among the observed treatment averages are as follows:

$$
\begin{aligned}
4 \text{ vs. } 1 &= 21.17 - 10.00 = 11.17 > 3.07 \\
4 \text{ vs. } 2 &= 21.17 - 15.67 = 5.50 > 3.07 \\
4 \text{ vs. } 3 &= 21.17 - 17.00 = 4.17 > 3.07 \\
3 \text{ vs. } 1 &= 17.00 - 10.00 = 7.00 > 3.07 \\
3 \text{ vs. } 2 &= 17.00 - 15.67 = 1.33 < 3.07 \rightarrow \text{no diff (not significant)} \\
2 \text{ vs. } 1 &= 15.67 - 10.00 = 5.67 > 3.07
\end{aligned}
$$

From this analysis, we see that there are significant differences between all pairs of means except 2 and 3. This implies that 10 and 15% hardwood concentration produce approximately the same tensile strength and that all other concentration levels tested produce different tensile strengths. It is often helpful to draw a graph of the treatment means, such as in Fig. 13-2, with the means that are *not* different underlined. This graph clearly reveals the results of the experiment and shows that 20% hardwood produces the maximum tensile strength.

13-2 The Completely Randomized Single-Factor Experiment

Example 13-2

* بالناحية العملية فأنا
بقدر أختار ١٠٪ بدل 15٪
و تعطيني نفس النتيجة ، فهيك بستخدم
resources أقل وتكلفة أقل .

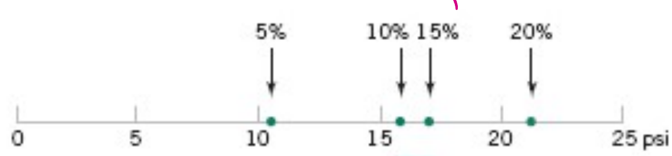
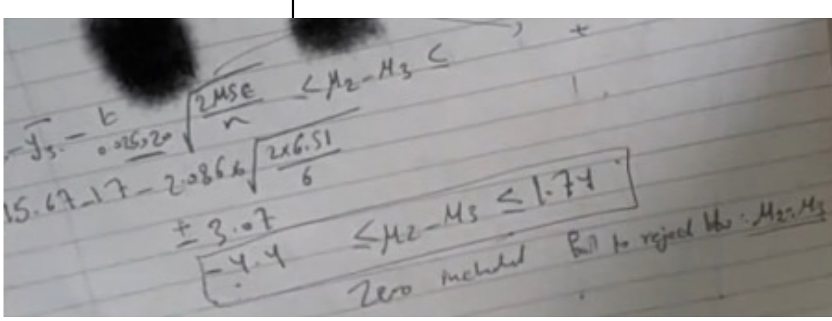


Figure 13-2 Results of Fisher's LSD method in Example 13-2.

Figure 13-2 Results of Fisher's LSD method in Example 13-2

← لو أعمل CI للفرق بين M_2 و M_3 :



ال error ناتج withen each treatment، كل وحدة إلها error لحال

13-2 The Completely Randomized Single-Factor Experiment

الفرق بين ال real value و ال fitted value

13-2.5 Residual Analysis and Model Checking

actual – fitted

* there is 3 assumptions for for the residual :
1) fixed variance.
2) Normally distributed.
3) Mean for residuals equal zero.
4) مجموع قيم الإنحرافات عن ال avg بكل level بال residuals بطلع صفر
وال over all mean برضه مجموعه = صفر

Table 13-6 Residuals for the Tensile Strength Experiment

for errors

Hardwood Concentration (%)	Residuals					
5	7-10 −3.00	−2.00	5.00	1.00	−1.00	0.00
10	8-10 −3.67	1.33	−2.67	2.33	3.33	−0.67
15	−3.00	1.00	2.00	0.00	−1.00	1.00
20	−2.17	3.83	0.83	1.83	−3.17	−1.17

من هاد الجدول

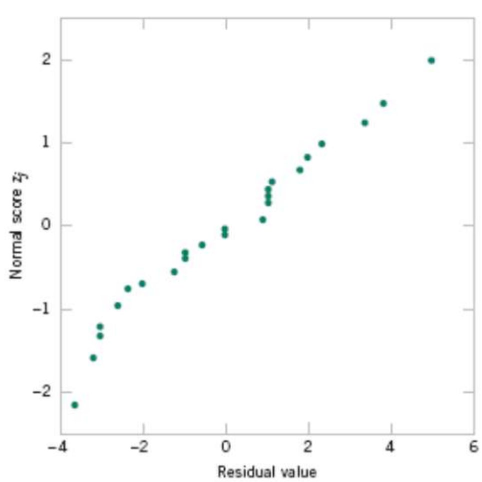
Hardwood Concentration (%)	Observations 1	2	3	4	5	6	Totals	Averages
5	7	8	15	11	9	10	60	10.00
10	12	17	13	18	19	15	94	15.67
15	14	18	19	17	16	18	102	17.00
20	19	25	22	23	18	20	127	21.17
							383	15.96

13-2 The Completely Randomized Single-Factor Experiment

13-2.5 Residual Analysis and Model Checking

to check what is the distribution shape

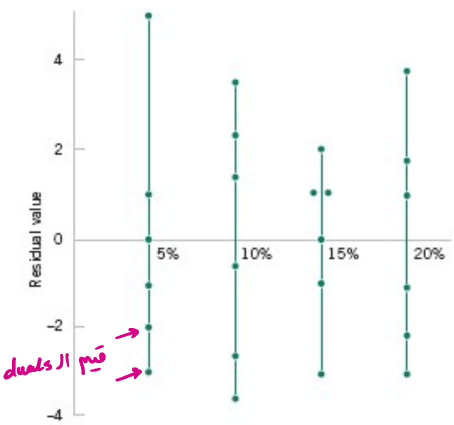
Figure 13-4 Normal probability plot of residuals from the hardwood concentration experiment.



13-2 The Completely Randomized Single-Factor Experiment

13-2.5 Residual Analysis and Model Checking

بنقارن فوق
وتحت ال zero

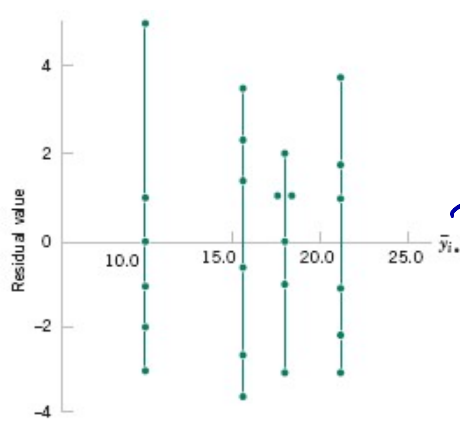


قيم ال residuals

Figure 13-5 Plot of residuals versus factor levels (hardwood concentration).

13-2 The Completely Randomized Single-Factor Experiment

13-2.5 Residual Analysis and Model Checking



factor level or $\bar{y}_{i.}$

Figure 13-6 Plot of residuals versus $\bar{y}_i$

13-3 The Random-Effects Model

13-3.1 Fixed versus Random Factors

In many situations, the factor of interest has a large number of possible levels. The analyst is interested in drawing conclusions about the entire population of factor levels. If the experimenter randomly selects a of these levels from the population of factor levels, we say that the factor is a **random factor.** Because the levels of the factor actually used in the experiment were chosen randomly, the conclusions reached will be valid for the entire population of factor levels. We will assume that the population of factor levels is either of infinite size or is large enough to be considered infinite. Notice that this is a very different situation than we encountered in the fixed effects case, where the conclusions apply only for the factor levels used in the experiment.

13-3 The Random-Effects Model

13-3.2 ANOVA and Variance Components

The linear statistical model is

$$Y_{ij} = \mu + \tau_i + \epsilon_{ij} \begin{cases} i = 1, 2, \ldots, a \\ j = 1, 2, \ldots, n \end{cases}$$

The variance of the response is $V(Y_{ij}) = \sigma_\tau^2 + \sigma^2$

Where each term on the right hand side is called a **variance component.**

13-3 The Random-Effects Model

13-3.2 ANOVA and Variance Components

For a random-effects model, the appropriate hypotheses to test are:

$$H_0: \sigma_\tau^2 = 0$$
$$H_1: \sigma_\tau^2 > 0$$

The ANOVA decomposition of total variability is still valid:

$$SS_T = SS_{\text{Treatments}} + SS_E$$

13-3 The Random-Effects Model

13-3.2 ANOVA and Variance Components

The expected values of the mean squares are

In the random-effects model for a single-factor, completely randomized experiment, the expected mean square for treatments is

$$E(MS_{\text{Treatments}}) = E\left(\frac{SS_{\text{Treatments}}}{a-1}\right) = \sigma^2 + n\sigma_\tau^2 \qquad (13\text{-}21)$$

and the expected mean square for error is

$$E(MS_E) = E\left[\frac{SS_E}{a(n-1)}\right] = \sigma^2 \qquad (13\text{-}22)$$

13-3 The Random-Effects Model

13-3.2 ANOVA and Variance Components

The estimators of the variance components are

$$\hat{\sigma}^2 = MS_E \qquad (13\text{-}24)$$

and

$$\hat{\sigma}_\tau^2 = \frac{MS_{\text{Treatments}} - MS_E}{n} \qquad (13\text{-}25)$$

13-3 The Random-Effects Model

Example 13-4

In *Design and Analysis of Experiments*, 5th edition (John Wiley, 2001), D. C. Montgomery describes a single-factor experiment involving the random-effects model in which a textile manufacturing company weaves a fabric on a large number of looms. The company is interested in loom-to-loom variability in tensile strength. To investigate this variability, a manufacturing engineer selects four looms at random and makes four strength determinations on fabric samples chosen at random from each loom. The data are shown in Table 13-7 and the ANOVA is summarized in Table 13-8.

Table 13-7 Strength Data for Example 13-4

	Observations					
Loom	1	2	3	4	Total	Average
1	98	97	99	96	390	97.5
2	91	90	93	92	366	91.5
3	96	95	97	95	383	95.8
4	95	96	99	98	388	97.0
					1527	95.45

Table 13-8 Analysis of Variance for the Strength Data

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	f_0	P-value
Looms	89.19	3	29.73	15.68	1.88 E-4
Error	22.75	12	1.90		
Total	111.94	15			

13-3 The Random-Effects Model

Example 13-4

From the analysis of variance, we conclude that the looms in the plant differ significantly in their ability to produce fabric of uniform strength. The variance components are estimated by $\hat{\sigma}^2 = 1.90$ and

$$\hat{\sigma}_\tau^2 = \frac{29.73 - 1.90}{4} = 6.96$$

Therefore, the variance of strength in the manufacturing process is estimated by

$$\widehat{V(Y_{ij})} = \hat{\sigma}_\tau^2 + \hat{\sigma}^2 = 6.96 + 1.90 = 8.86$$

Most of this variability is attributable to differences between looms.

13-3 The Random-Effects Model

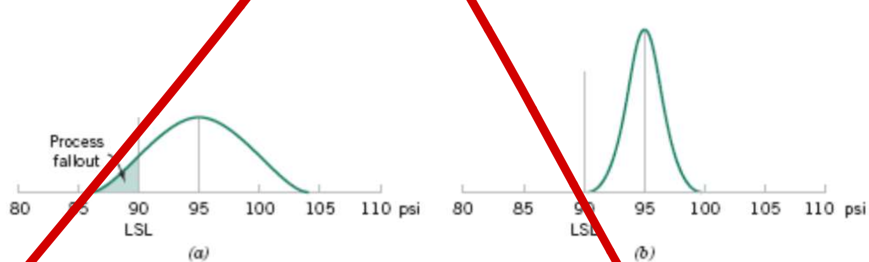


Figure 13-8 The distribution of fabric strength. (a) Current process, (b) improved process.

مطلوب

إكستنشن لل Paired T test
بس بدل pair صارو أكتر من 2
بسميه Block

13-4 Randomized Complete Block Designs

13-4.1 Design and Statistical Analyses

The **randomized block design** is an extension of the paired t-test to situations where the factor of interest has more than two levels.

type1
type 2
type 3

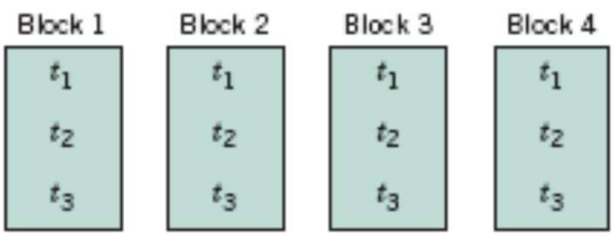


Figure 13-9 A randomized complete block design.

13-4 Randomized Complete Block Designs

13-4.1 Design and Statistical Analyses

For example, consider the situation of Example 10-9, where two different methods were used to predict the shear strength of steel plate girders. Say we use four girders as the experimental units.

Table 13-9 A Randomized Complete Block Design

Treatments (Method)	Block (Girder) 1	2	3	4
1	y_{11}	y_{12}	y_{13}	y_{14}
2	y_{21}	y_{22}	y_{23}	y_{24}
3	y_{31}	y_{32}	y_{33}	y_{34}

هاد كامل هو Block
ال Sample هو ال Block

13-4 Randomized Complete Block Designs

13-4.1 Design and Statistical Analyses

General procedure for a randomized complete block design:

Table 13-10 A Randomized Complete Block Design with a Treatments and b Blocks

Treatments	Blocks 1	2	…	b	Totals	Averages
1	y_{11}	y_{12}	…	y_{1b}	$y_{1\cdot}$	$\bar{y}_{1\cdot}$
2	y_{21}	y_{22}	…	y_{2b}	$y_{2\cdot}$	$\bar{y}_{2\cdot}$
⋮	⋮	⋮		⋮	⋮	⋮
a	y_{a1}	y_{a2}	…	y_{ab}	$y_{a\cdot}$	$\bar{y}_{a\cdot}$
Totals	$y_{\cdot 1}$	$y_{\cdot 2}$	…	$y_{\cdot b}$	$y_{\cdot\cdot}$	
Averages	$\bar{y}_{\cdot 1}$	$\bar{y}_{\cdot 2}$	…	$\bar{y}_{\cdot b}$		$\bar{y}_{\cdot\cdot}$

عدد البلوكس

total treatments

نزوع j ~~b~~

كنا ناخذ difference avg هلأ بناخذ

avg treatments

overall avg

عدد levels ال factor

نزوع i

نزوع i ~~a~~

given data

Calculated data

13-4 Randomized Complete Block Designs

13-4.1 Design and Statistical Analyses

The appropriate linear statistical model:

$$Y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij} \begin{cases} i = 1, 2, \ldots, a \\ j = 1, 2, \ldots, b \end{cases}$$

We assume

بتقارن Blocks مع بعض

- treatments and blocks are initially fixed effects
- blocks do not interact
- $\sum_{i=1}^{a} \tau_i = 0$ and $\sum_{j=1}^{b} \beta_j = 0$

→ anova table:

Source	Dof	MS	F
treat	$a-1$	$\frac{SS_{treat}}{a-1}$	$\frac{MS_{treat}}{MSE}$
Block	$b-1$	$\frac{SS_{block}}{b-1}$	MS_{block}/MS_E
error	$(a-1)(b-1)$	SS_E/dof	
total	$ab-1$		

total − (treat + Block)

13-4 Randomized Complete Block Designs

13-4.1 Design and Statistical Analyses

We are interested in testing:

$$H_0: \tau_1 = \tau_2 = \cdots = \tau_a = 0$$
$$H_1: \tau_i \neq 0 \text{ at least one } i$$

The **sum of squares identity for the randomized complete block design** is

$$\sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij} - \bar{y}_{..})^2 = b\sum_{i=1}^{a}(\bar{y}_{i\cdot} - \bar{y}_{..})^2 + a\sum_{j=1}^{b}(\bar{y}_{\cdot j} - \bar{y}_{..})^2$$
$$+ \sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij} - \bar{y}_{\cdot j} - \bar{y}_{i\cdot} + \bar{y}_{..})^2 \qquad (13\text{-}27)$$

treat

block

error

or symbolically

$$SS_T = SS_{\text{Treatments}} + SS_{\text{Blocks}} + SS_E$$

13-4 Randomized Complete Block Designs

13-4.1 Design and Statistical Analyses

The mean squares are:

$$MS_{\text{Treatments}} = \frac{SS_{\text{Treatments}}}{a - 1}$$
$$MS_{\text{Blocks}} = \frac{SS_{\text{Blocks}}}{b - 1}$$
$$MS_E = \frac{SS_E}{(a - 1)(b - 1)}$$

13-4 Randomized Complete Block Designs

13-4.1 Design and Statistical Analyses

The expected values of these mean squares are:

$$E(MS_{\text{Treatments}}) = \sigma^2 + \frac{b\sum_{i=1}^{a}\tau_i^2}{a-1}$$

$$E(MS_{\text{Blocks}}) = \sigma^2 + \frac{a\sum_{j=1}^{b}\beta_j^2}{b-1}$$

$$E(MS_E) = \sigma^2$$

13-4 Randomized Complete Block Designs

13-4.1 Design and Statistical Analyses

Definition

هاي المعادلات
إلي بستخدمها
بال calculations

The computing formulas for the sums of squares in the analysis of variance for a randomized complete block design are

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij}^2 - \frac{y_{..}^2}{ab} \quad (13\text{-}29)$$

$$SS_{\text{Treatments}} = \frac{1}{b}\sum_{i=1}^{a} y_{i.}^2 - \frac{y_{..}^2}{ab} \quad (13\text{-}30)$$

هدول
زي إلي
قبل عادي

هاد
الاختلاف
الوحيد

$$SS_{\text{Blocks}} = \frac{1}{a}\sum_{j=1}^{b} y_{.j}^2 - \frac{y_{..}^2}{ab} \quad (13\text{-}31)$$

and

أسباب ال total
variability

$$SS_E = SS_T - SS_{\text{Treatments}} - SS_{\text{Blocks}} \quad (13\text{-}32)$$

13-4 Randomized Complete Block Designs

13-4.1 Design and Statistical Analyses

Table 13-11 ANOVA for a Randomized Complete Block Design

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Treatments	$SS_{\text{Treatments}}$	$a - 1$	$\frac{SS_{\text{Treatments}}}{a-1}$	$\frac{MS_{\text{Treatments}}}{MS_E}$
Blocks	SS_{Blocks}	$b - 1$	$\frac{SS_{\text{Blocks}}}{b-1}$	
Error	SS_E (by subtraction)	$(a-1)(b-1)$	$\frac{SS_E}{(a-1)(b-1)}$	
Total	SS_T	$ab - 1$		

أثر اختلاف
كذا

13-4 Randomized Complete Block Designs

Example 13-5

An experiment was performed to determine the effect of four different chemicals on the strength of a fabric. These chemicals are used as part of the permanent press finishing process. Five fabric samples were selected, and a randomized complete block design was run by testing each chemical type once in random order on each fabric sample. The data are shown in Table 13-12. We will test for differences in means using an ANOVA with $\alpha = 0.01$.

Table 13-12 Fabric Strength Data—Randomized Complete Block Design

	Fabric Sample					Treatment Totals	Treatment Averages
Chemical Type	1	2	3	4	5	$y_{i\cdot}$	$\bar{y}_{i\cdot}$
1	1.3	1.6	0.5	1.2	1.1	5.7	1.14
2	2.2	2.4	0.4	2.0	1.8	8.8	1.76
3	1.8	1.7	0.6	1.5	1.3	6.9	1.38
4	3.9	4.4	2.0	4.1	3.4	17.8	3.56
Block totals $y_{\cdot j}$	9.2	10.1	3.5	8.8	7.6	39.2($y_{\cdot\cdot}$)	
Block averages $\bar{y}_{\cdot j}$	2.30	2.53	0.88	2.20	1.90		1.96($\bar{y}_{\cdot\cdot}$)

13-4 Randomized Complete Block Designs

Example 13-5

The sums of squares for the analysis of variance are computed as follows:

$$SS_T = \sum_{i=1}^{4}\sum_{j=1}^{5} y_{ij}^2 - \frac{y_{..}^2}{ab}$$

$$= (1.3)^2 + (1.6)^2 + \cdots + (3.4)^2 - \frac{(39.2)^2}{20} = 25.69$$

$$SS_{\text{Treatments}} = \sum_{i=1}^{4} \frac{y_{i.}^2}{b} - \frac{y_{..}^2}{ab}$$

$$= \frac{(5.7)^2 + (8.8)^2 + (6.9)^2 + (17.8)^2}{5} - \frac{(39.2)^2}{20} = 18.04$$

13-4 Randomized Complete Block Designs

Example 13-5

$$SS_{\text{Blocks}} = \sum_{j=1}^{5} \frac{y_{.j}^2}{a} - \frac{y_{..}^2}{ab}$$

$$= \frac{(9.2)^2 + (10.1)^2 + (3.5)^2 + (8.8)^2 + (7.6)^2}{4} - \frac{(39.2)^2}{20} = 6.69$$

$$SS_E = SS_T - SS_{\text{Blocks}} - SS_{\text{Treatments}}$$
$$= 25.69 - 6.69 - 18.04 = 0.96$$

The ANOVA is summarized in Table 13-13. Since $f_0 = 75.13 > f_{0.01,3,12} = 5.95$ (the P-value is 4.79×10^{-8}), we conclude that there is a significant difference in the chemical types so far as their effect on strength is concerned.

13-4 Randomized Complete Block Designs

Example 13-5

Table 13-13 Analysis of Variance for the Randomized Complete Block Experiment

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	f_0	P-value
Chemical types (treatments)	18.04	3	6.01	75.13	4.79 E-8
Fabric samples (blocks)	6.69	4	1.67		
Error	0.96	12	0.08		
Total	25.69	19			

$\frac{1.67}{0.08}$

كثير صغيرة يعني significant عال tensile strength

13-4 Randomized Complete Block Designs

Minitab Output for Example 13-5

Table 13-14 Minitab Analysis of Variance for the Randomized Complete Block Design in Example 13-5

Analysis of Variance (Balanced Designs)

Factor	Type	Levels	Values				
Chemical	fixed	4	1	2	3	4	
Fabric S	fixed	5	1	2	3	4	5

Analysis of Variance for strength

Source	DF	SS	MS	F	P
Chemical	3	18.0440	6.0147	75.89	0.000
Fabric S	4	6.6930	1.6733	21.11	0.000
Error	12	0.9510	0.0792		
Total	19	25.6880			

treatement

Block

F-test with denominator: Error
Denominator MS = 0.079250 with 12 degrees of freedom

Numerator	DF	MS	F	P
Chemical	3	6.015	75.89	0.000
Fabric S	4	1.673	21.11	0.000

13-4 Randomized Complete Block Designs

13-4.2 Multiple Comparisons

Fisher's Least Significant Difference for Example 13-5

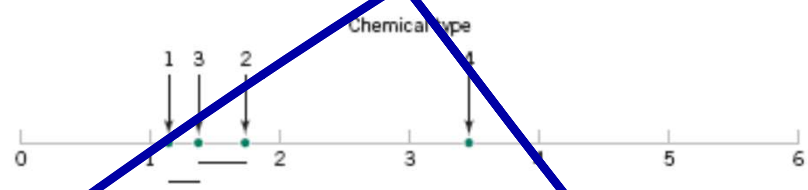
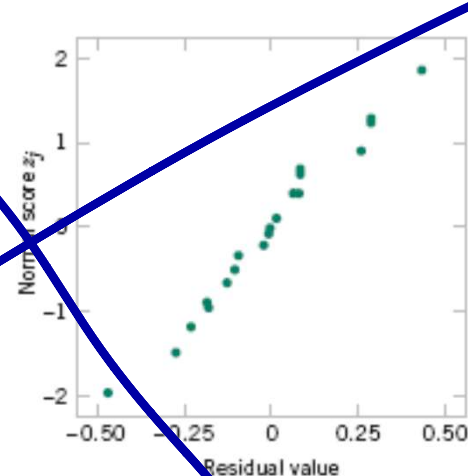


Figure 13-10 Results of Fisher's LSD method.

13-4 Randomized Complete Block Designs

13-4.3 Residual Analysis and Model Checking

Figure 13-11 Normal probability plot of residuals from the randomized complete block design.



13-4 Randomized Complete Block Designs

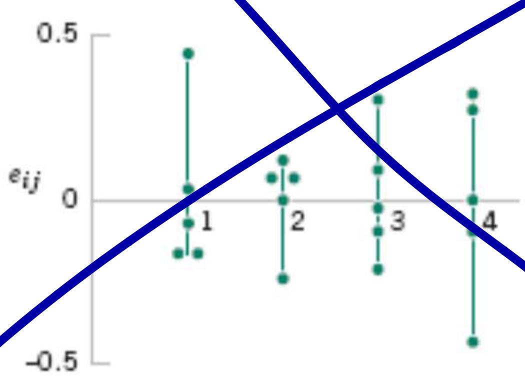


Figure 13-12 Residuals by treatment.

13-4 Randomized Complete Block Designs

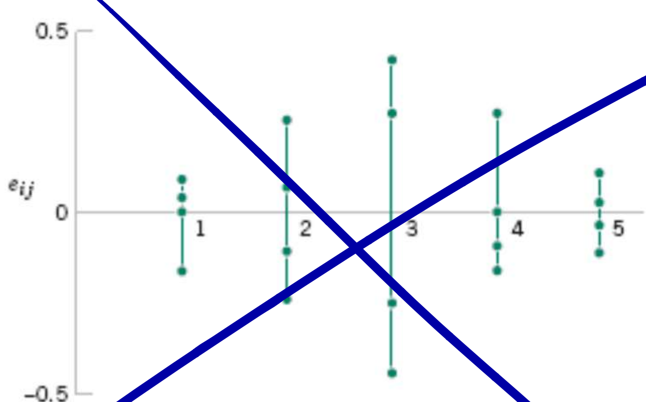


Figure 13-13 Residuals by block.

13-4 Randomized Complete Block Designs

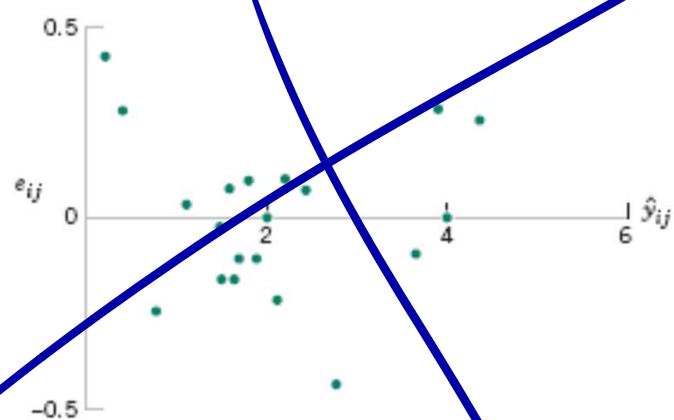


Figure 13-14 Residuals versus $\hat{y}_{ij}$.

IMPORTANT TERMS AND CONCEPTS

Analysis of variance (ANOVA)
Blocking
Complete by randomized experiment
Expected mean squares
Fisher's least significant difference (LSD) method
Fixed factor
Graphical comparison of means
Levels of a factor
Mean square
Multiple comparisons
Nuisance factors
Random factor
Randomization
Randomized complete block design
Residual analysis and model adequacy checking
Sample size and replication in an experiment
Treatment effect
Variance component

Design of Experiments with Several Factors

كان one factor بالشابتر الماضي .

Example

Suppose that we wish to study the factors that affect the surface roughness produced by a polishing operation.

Three factors are considered relevant:

Speed of the grinding wheel
Feed or speed of lateral movement and
Roughness of the grinding wheel

Each factor has several levels

Speed: 50 – 100 – 200 RPM

Feed: 0.1 – 0.2 – 0.4 - 0.5 mm/sec

Roughness: soft – medium – hard

Speed	Feed	Roughness	Speed	Feed	Roughness	Speed	Feed	Roughness
50	0.1	soft	100	0.1	soft	200	0.1	soft
50	0.1	medium	100	0.1	medium	200	0.1	medium
50	0.1	hard	100	0.1	hard	200	0.1	hard
50	0.2	soft	100	0.2	soft	200	0.2	soft
50	0.2	medium	100	0.2	medium	200	0.2	medium
50	0.2	hard	100	0.2	hard	200	0.2	hard
50	0.4	soft	100	0.4	soft	200	0.4	soft
50	0.4	medium	100	0.4	medium	200	0.4	medium
50	0.4	hard	100	0.4	hard	200	0.4	hard
50	0.5	soft	100	0.5	soft	200	0.5	soft
50	0.5	medium	100	0.5	medium	200	0.5	medium
50	0.5	hard	100	0.5	hard	200	0.5	hard

Factorial Experiments

By a **factorial experiment** we mean that in each complete trial or replicate of the experiment all possible combinations of the levels of the factors are investigated.

& How to calc the num of runs in an experiment?

(factor) → pow

(Level)

Case of two levels

Speed	Feed	Roughness
50	0.1	soft
50	0.1	hard
50	0.5	soft
50	0.5	hard
400	0.1	soft
400	0.1	hard
400	0.5	soft
400	0.5	hard

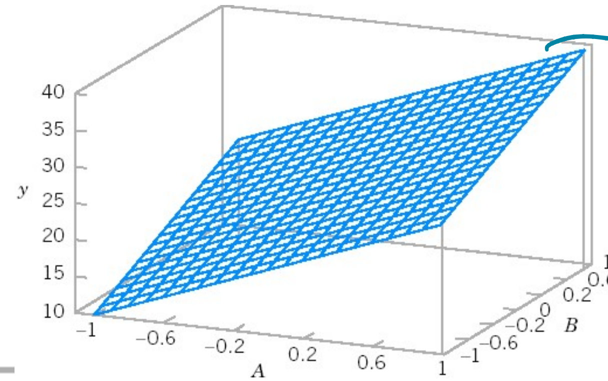
Case of two factors

		Speed		
		50	100	200
Roughness	soft	10, 8,11	12, 8,16	20,22,14
	medium	14,19,18	20,17,16	20,16,15
	hard	12,16,20	19,20,15	22,23,18

Factorial Experiment with Two Factors

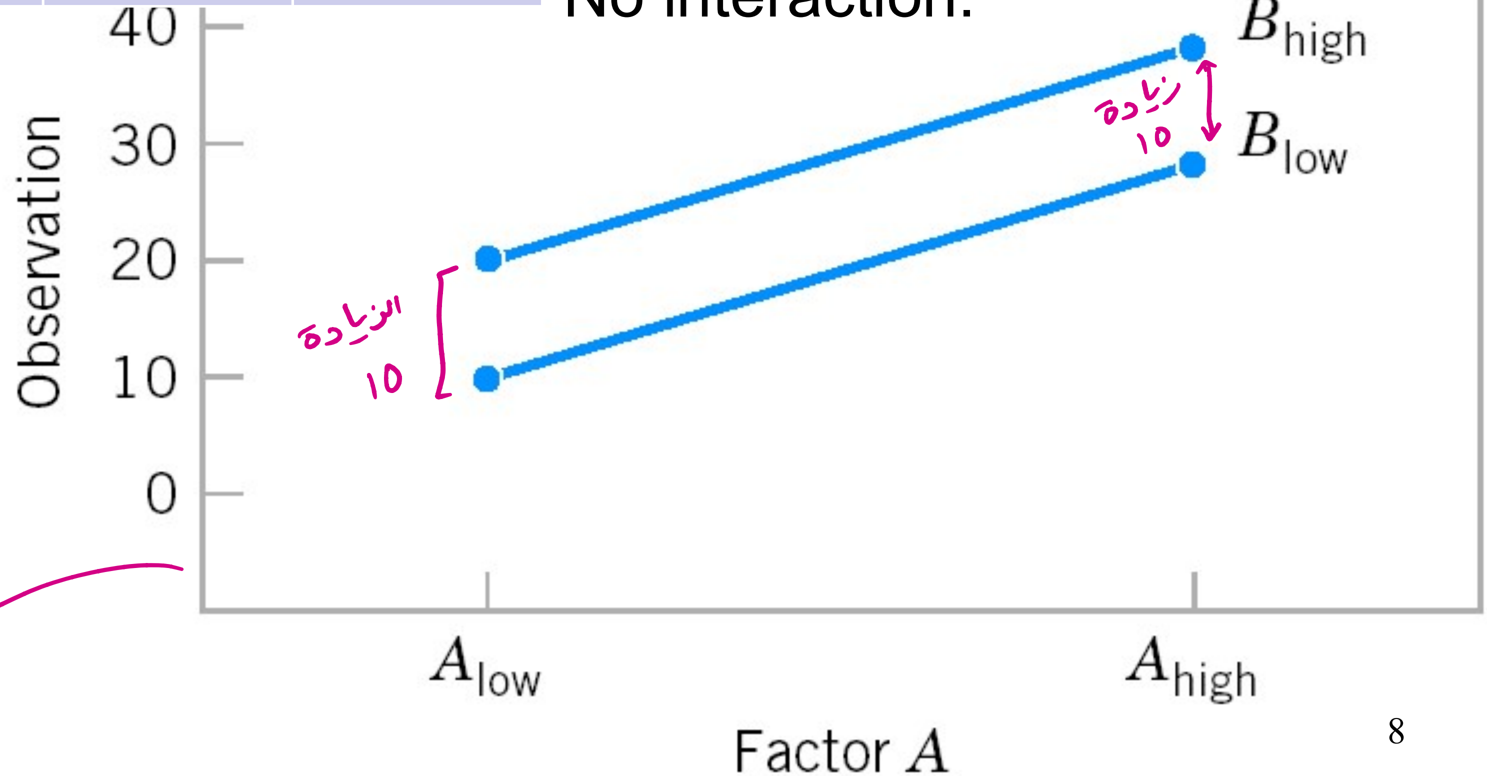
	Factor B	
Factor A	B_{low}	B_{high}
A_{low}	10	20
A_{high}	30	40

resp زاد



linear
+ كم تزيد

No interaction.



الزيادة ١٠

زيادة ١٠

no interaction
(ما اختلطو ببعض)

* كيف نحسب ال main effect ؟

ال avg change بال response عند الإنتقال من ال low level لل high level.

{ قديش تغيّر ال response لما تغيّر ال factor

Main effect of A:

$$\frac{30+40}{2} - \frac{10+20}{2}$$

$35 - 15 \rightarrow 20$

معناها لما انتقل بـ A من low لـ high بزيد ال response بـ نفس إشي هم 20

more significant (more important)

of B:

$$\frac{20+40}{2} - \frac{10+30}{2} = 10$$

بزيد 10 بـ B

لنحسب ال interaction :

بناخد القيم diagonal

$$\frac{10+40}{2} - \frac{20+30}{2}$$

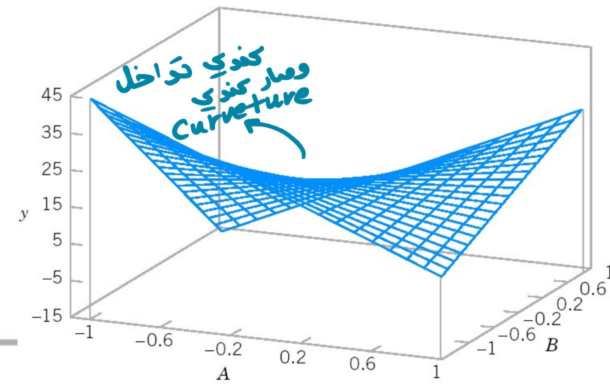
$= 0$

معناه no interaction btw A & B

متوازيين

Factorial Experiment with Two Factors

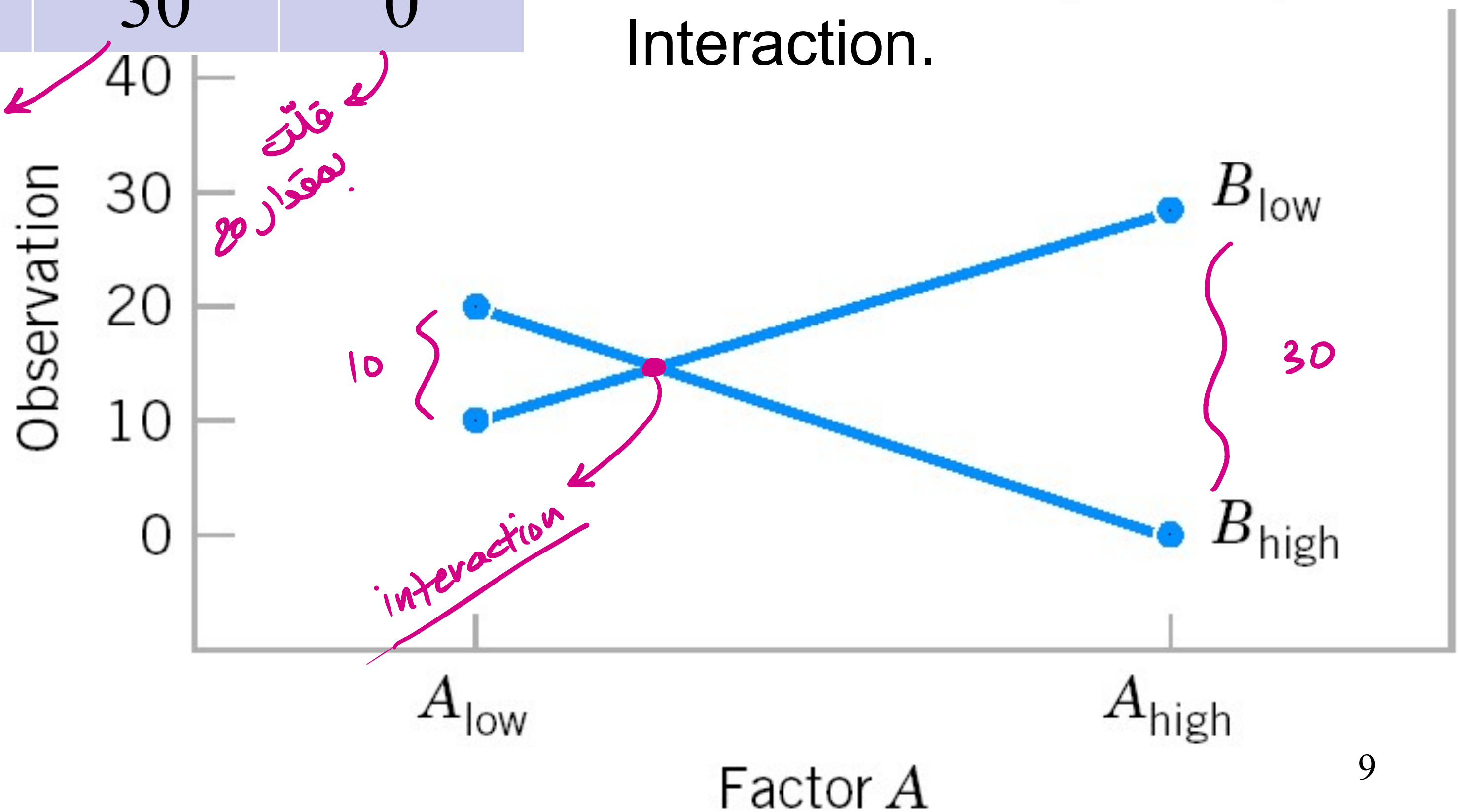
	Factor B	
Factor A	B_{low}	B_{high}
A_{low}	10	20
A_{high}	30	0



زادت بمقدار 20

قلت بمقدار 20

Interaction.



ae for A:

$$\frac{30+0}{2} - \frac{10+20}{2} = 0$$

و

هش معناها إنه not significant

الموضوع بعتمد على ال B levels

ال A بتأثر بس لما أحسبه مع B

لنحسب

ال interaction

effect بنعمل ال

for AB

B :

$$\frac{20+0}{2} - \frac{10+30}{2} = -1$$

$$\frac{20+30}{2} - \frac{10-0}{2}$$

$$= 20$$

في interaction كبير ، وأكبر من A لحالها أو B لحالها

يعني أثر تغييرهم مع بعض هو إلي بأثر

Data Arrangement for a Two-Factor Factorial Design

k : for the replicates

sample size

		Factor B					
		1	*2*	…	*b*	*Total*	*Ave*
Factor A	1	$y_{111}, y_{112}, \ldots y_{11n}$	$y_{121}, y_{122}, \ldots y_{12n}$		$y_{1b1}, y_{1b2}, \ldots y_{1bn}$	$y_{1..}$	$y_{1..}/bn$
	2	$y_{211}, y_{212}, \ldots y_{21n}$	$y_{221}, y_{222}, \ldots y_{22n}$		$y_{2b1}, y_{2b2}, \ldots y_{2bn}$	$y_{2..}$	$y_{2..}/bn$
	a	$y_{a11}, y_{a12}, \ldots y_{a1n}$	$y_{a21}, y_{a22}, \ldots y_{a2n}$		$y_{ab1}, y_{ab2}, \ldots y_{abn}$	$y_{3..}$	$y_{3..}/bn$
Total		$y_{.1.}$	$y_{.2.}$		$y_{.b.}$		
Ave		$y_{.1.}/an$	$y_{.2.}/an$		$y_{.b.}/an$		

a levels for factor A

Two-factor factorial experiments

The observations may be described by the linear statistical model:

$$Y_{ijk} = \overbrace{\mu + \tau_i + \beta_j + \underbrace{(\tau\beta)_{ij}}_{\text{the interactions}} + \varepsilon_{ijk}}^{\text{all sources of variability}} \quad \begin{cases} i = 1,2,\ldots,a \\ j = 1,2,\ldots,b \\ k = 1,2,\ldots,n \end{cases}$$

راح يكون عندي ٣ فرضيات ، وحدة لكل إنحراف

for τ , for β , for $(\tau\beta)$

where ε_{ijk} are normal random variables

Statistical Analysis of the Fixed-Effects Model

لكل صف

$$y_{i..} = \sum_{j=1}^{b} \sum_{k=1}^{n} y_{ijk} \qquad \bar{y}_{i..} = \frac{y_{i..}}{bn} \qquad i = 1, 2, \ldots, a$$

مقام ال a

$$y_{.j.} = \sum_{i=1}^{a} \sum_{k=1}^{n} y_{ijk} \qquad \bar{y}_{.j.} = \frac{y_{.j.}}{an} \qquad j = 1, 2, \ldots, b$$

$$y_{ij.} = \sum_{k=1}^{n} y_{ijk} \qquad \bar{y}_{ij.} = \frac{y_{ij.}}{n} \qquad \begin{array}{l} i = 1, 2, \ldots, a \\ j = 1, 2, \ldots, b \end{array}$$

$$y_{...} = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} y_{ijk} \qquad \bar{y}_{...} = \frac{y_{...}}{abn}$$

لكل البيانات

The hypotheses that will be tested

1. H_0: $\tau_1 = \tau_2 = \cdots = \tau_a = 0$ (no main effect of factor A)
   H_1: at least one $\tau_i \neq 0$

2. H_0: $\beta_1 = \beta_2 = \cdots = \beta_b = 0$ (no main effect of factor B)
   H_1: at least one $\beta_j \neq 0$

3. H_0: $(\tau\beta)_{11} = (\tau\beta)_{12} = \cdots = (\tau\beta)_{ab} = 0$ (no interaction)
   H_1: at least one $(\tau\beta)_{ij} \neq 0$

The Sum of Squares Identity

The SS for 2-factor ANOVA is

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(y_{ijk}-\bar{y}_{\cdots})^2 = bn\sum_{i=1}^{a}(\bar{y}_{i\cdot\cdot}-\bar{y}_{\cdots})^2$$

القيمة الحقيقية

ال overall avg

$$+\, an\sum_{j=1}^{b}(\bar{y}_{\cdot j\cdot}-\bar{y}_{\cdots})^2$$

$$+\, n\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{y}_{ij\cdot}-\bar{y}_{i\cdot\cdot}-\bar{y}_{\cdot j\cdot}+\bar{y}_{\cdots})^2$$

هاي الطريقة صعبة

$$+\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(y_{ijk}-\bar{y}_{ij\cdot})^2$$

$$SS_T = SS_A + SS_B + SS_{AB} + SS_E$$

ناتج عن A

Computing formulas for the sum of squares

هاي الطريقة أحسن

مجموعها الكل

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk}^2 - \frac{y_{...}^2}{abn}$$

كل قيمة بالجدول تربيع

عدد المشاهدات الكلية

سطر سطر

$$SS_A = \frac{\sum_{i=1}^{a} y_{i..}^2}{bn} - \frac{y_{...}^2}{abn}$$

$$SS_B = \frac{\sum_{j=1}^{b} y_{.j.}^2}{an} - \frac{y_{...}^2}{abn}$$

$$SS_{AB} = \frac{1}{n}\sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij.}^2 - \frac{y_{...}^2}{abn} - SS_A - SS_B$$

$$SS_E = SS_T - SS_A - SS_B - SS_{AB}$$

الصيغة العامة ل
2 factor factorial experiments

The ANOVA Table

بقارنها مع
f_{cr}

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
A treatments	SS_A	$a-1$	$MS_A = \dfrac{SS_A}{a-1}$	$\dfrac{MS_A}{MS_E}$
B treatments	SS_B	$b-1$	$MS_B = \dfrac{SS_B}{b-1}$	$\dfrac{MS_B}{MS_E}$
Interaction	SS_{AB}	$(a-1)(b-1)$	$MS_{AB} = \dfrac{SS_{AB}}{(a-1)(b-1)}$	$\dfrac{MS_{AB}}{MS_E}$
Error	SS_E	$ab(n-1)$	$MS_E = \dfrac{SS_E}{ab(n-1)}$	
Total	SS_T	$\underline{abn} - 1$		

كل المشاهدات

The tests of hypotheses

To test H_0: $\tau_i = 0$ use the ratio

$$F_0 = \frac{MS_A}{MS_E}$$

To test H_0: $\beta_j = 0$ use the ratio

$$F_0 = \frac{MS_B}{MS_E}$$

To test H_0: $(\tau\beta)_{ij} = 0$ use the ratio

$$F_0 = \frac{MS_{AB}}{MS_E}$$

Aircraft primer paints are applied to aluminum surfaces by two methods: dipping and spraying. The purpose of the primer is to improve paint adhesion, and some parts can be primed using either application method. The process engineering group responsible for this operation is interested in learning whether three different primers differ in their adhesion properties. A factorial experiment was performed to investigate the effect of paint primer type and application method on paint adhesion. For each combination of primer type and application method, three specimens were painted, then a finish paint was applied, and the adhesion force was measured. The data from the experiment are shown in Table 14-5. The circled numbers in the cells are the cell totals $y_{ij\cdot}$. The sums of squares required to perform the ANOVA are computed as follows:

Example

- Aluminum surfaces of planes are coated with primer then paint.
- We are interested in the adhesion strength of the paint.
- It is suspected that the type of primer used and the method of applying it affect the adhesion strength.
- There are three types of primers and two methods of application; dipping and spraying.
- The factorial experiment: try all combinations of primer type and application method.
- Three replicates of each combination are made.

What is the total number of trials or specimens?

Example

Primer Type	Dipping		Spraying		$y_{i\cdot\cdot}$
1	4.0, 4.5, 4.3	12.8	5.4, 4.9, 5.6	15.9	28.7
2	5.6, 4.9, 5.4	15.9	5.8, 6.1, 6.3	18.2	34.1
3	3.8, 3.7, 4.0	11.5	5.5, 5.0, 5.0	15.5	27.0
$y_{\cdot j\cdot}$	40.2		49.6		$89.8 = y_{\cdots}$

مجموعهم

٣ عينات

مجموع الصف كامل

مجموع ال٣

مجموع الثلاث

مجموعهم ال٣

factor A

→ 3 types → $a = 3$

2 methods → $b = 2$

3 replicates → $n = 3$

$abn = 3 * 2 * 3 = 18$

Example

$$SS_T = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} y_{ijk}^2 - \frac{y_{\cdots}^2}{abn}$$

$$= (4.0)^2 + (4.5)^2 + \cdots + (5.0)^2 - \frac{(89.8)^2}{18} = 10.72$$

كل سطر

a

$$SS_{\text{types}} = \sum_{i=1}^{a} \frac{y_{i\cdot\cdot}^2}{bn} - \frac{y_{\cdots}^2}{abn}$$

$$= \frac{(28.7)^2 + (34.1)^2 + (27.0)^2}{\underline{\underline{6}}} - \frac{(89.8)^2}{18} = 4.58$$

$2*3 = bn$

سطر

$$SS_{\text{methods}} = \sum_{j=1}^{b} \frac{y_{\cdot j \cdot}^2}{an} - \frac{y_{\cdots}^2}{abn}$$

b

$$= \frac{(40.2)^2 + (49.6)^2}{9} - \frac{(89.8)^2}{18} = 4.91$$

كل عامود

$an = 3*3$

Example

$$SS_{\text{interaction}} = \sum_{i=1}^{a}\sum_{j=1}^{b}\frac{y_{ij\cdot}^2}{n} - \frac{y_{\cdots}^2}{abn} - SS_{\text{types}} - SS_{\text{methods}}$$

$$= \frac{(12.8)^2 + (15.9)^2 + (11.5)^2 + (15.9)^2 + (18.2)^2 + (15.5)^2}{3 \rightarrow n}$$

$$- \frac{(89.8)^2}{18} - \underbrace{4.58}_{SS\,A} - \underbrace{4.91}_{SS\,B} = 0.24$$

and

$$SS_E = SS_T - SS_{\text{types}} - SS_{\text{methods}} - SS_{\text{interaction}}$$
$$= 10.72 - 4.58 - 4.91 - 0.24 = 0.99$$

Example

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	f_0	P-Value
Primer types	4.58	2	2.29	28.63	2.7 × E-5
Application methods	4.91	1	4.91	61.38	5.0 × E-7
Interaction	0.24	2	0.12	1.50	0.2621
Error	0.99	12	0.08		
Total	10.72	17			

factor A

factor B

AB

10^{-5} صغيرة يعني significant

أصغر من f_α معناها fail to reject

0.05 أكبر يعني no effect

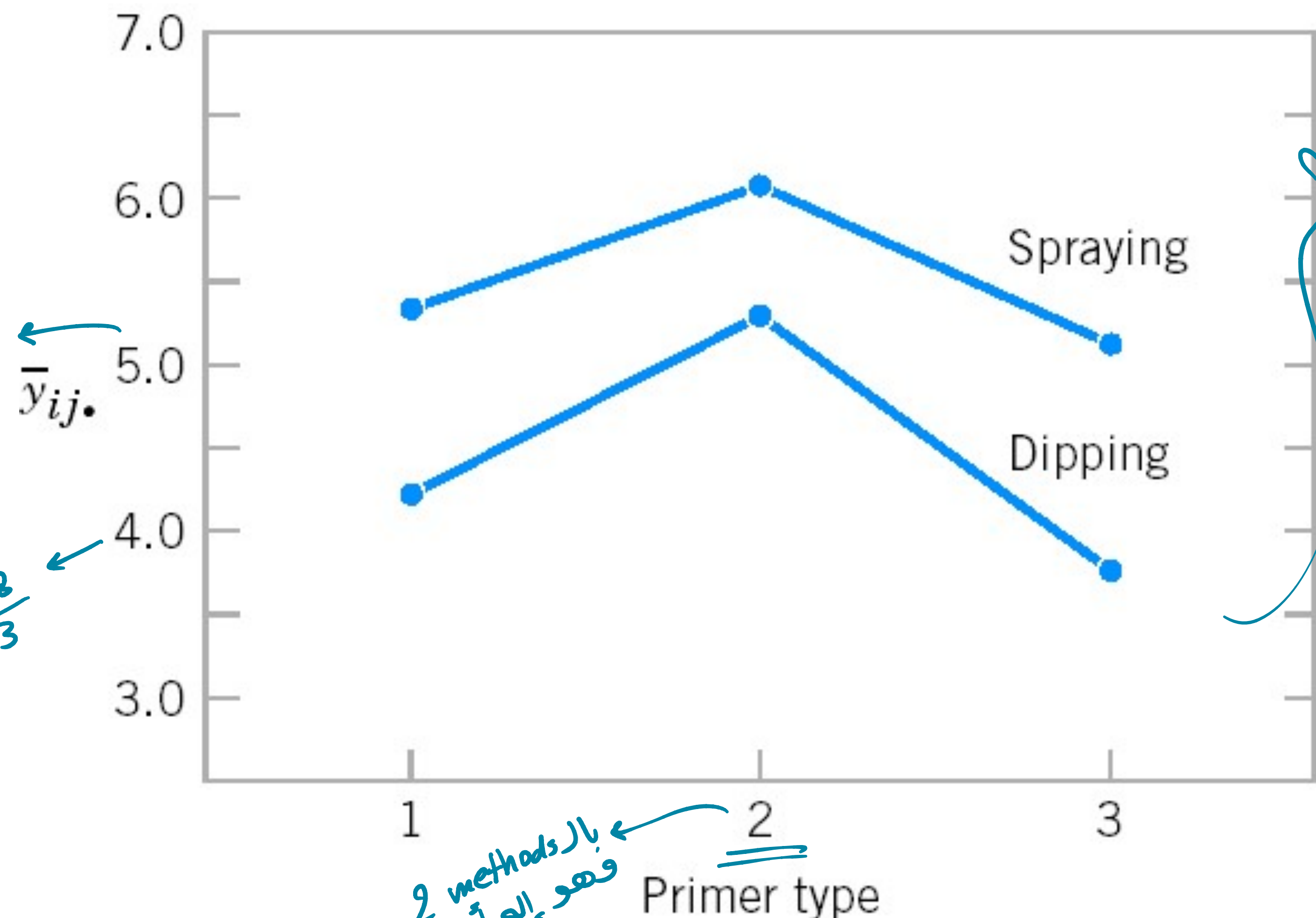
fail to reject H_0

هون بالعكس

$$Suppose\ \alpha = 0.05 \text{ then : } f_{0.05,2,12} = 3.89 \text{ and } f_{0.05,1,12} = 4.75$$

Conclusions?

إذا p-value أقل من 0.05 معناته significant



$\frac{15.9}{3}$

$\frac{12.8}{3}$

no interaction تقريبًا Parallel

بال 2 methods فهو إله أعلى قيمة التصاق

Graph of average adhesion force versus primer types for both application methods.

Conclusions ?

Model Adequacy Checking

real　　expected

$$\text{Residual}: e_{ijk} = y_{ijk} - \bar{y}_{ij.}$$

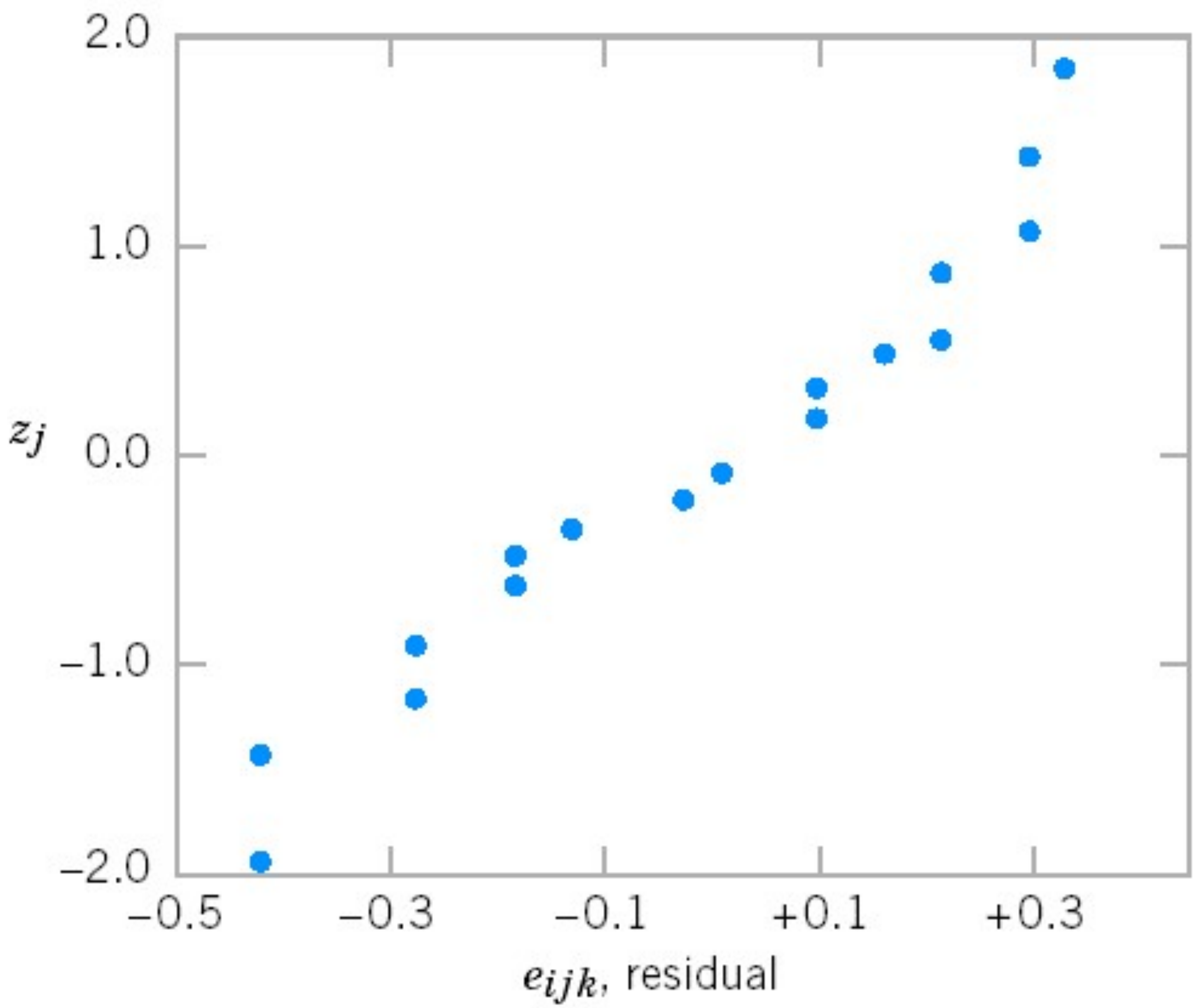
expected value

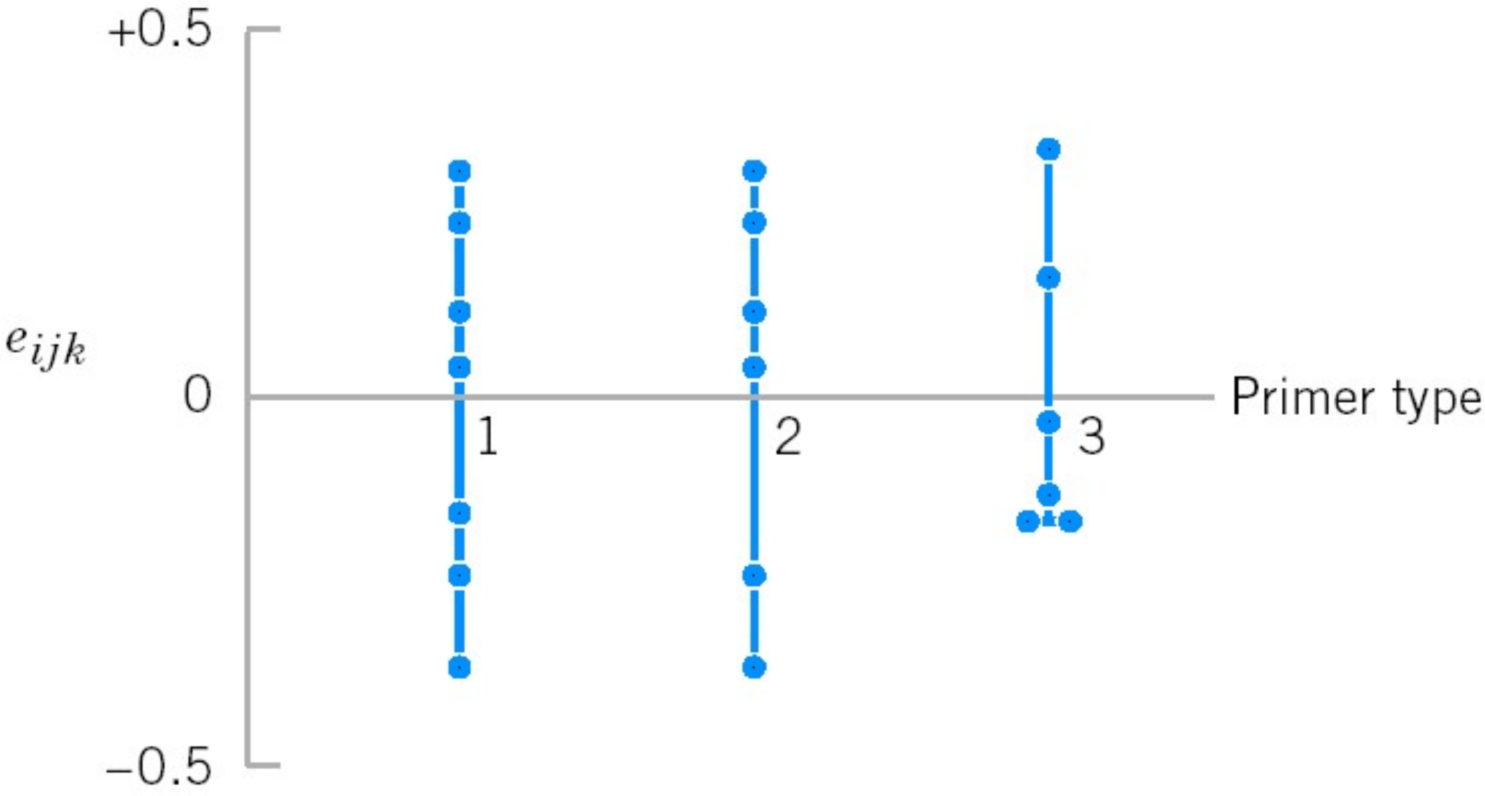
$\frac{12.8}{3} = 4.26$

5.04 − 4.26

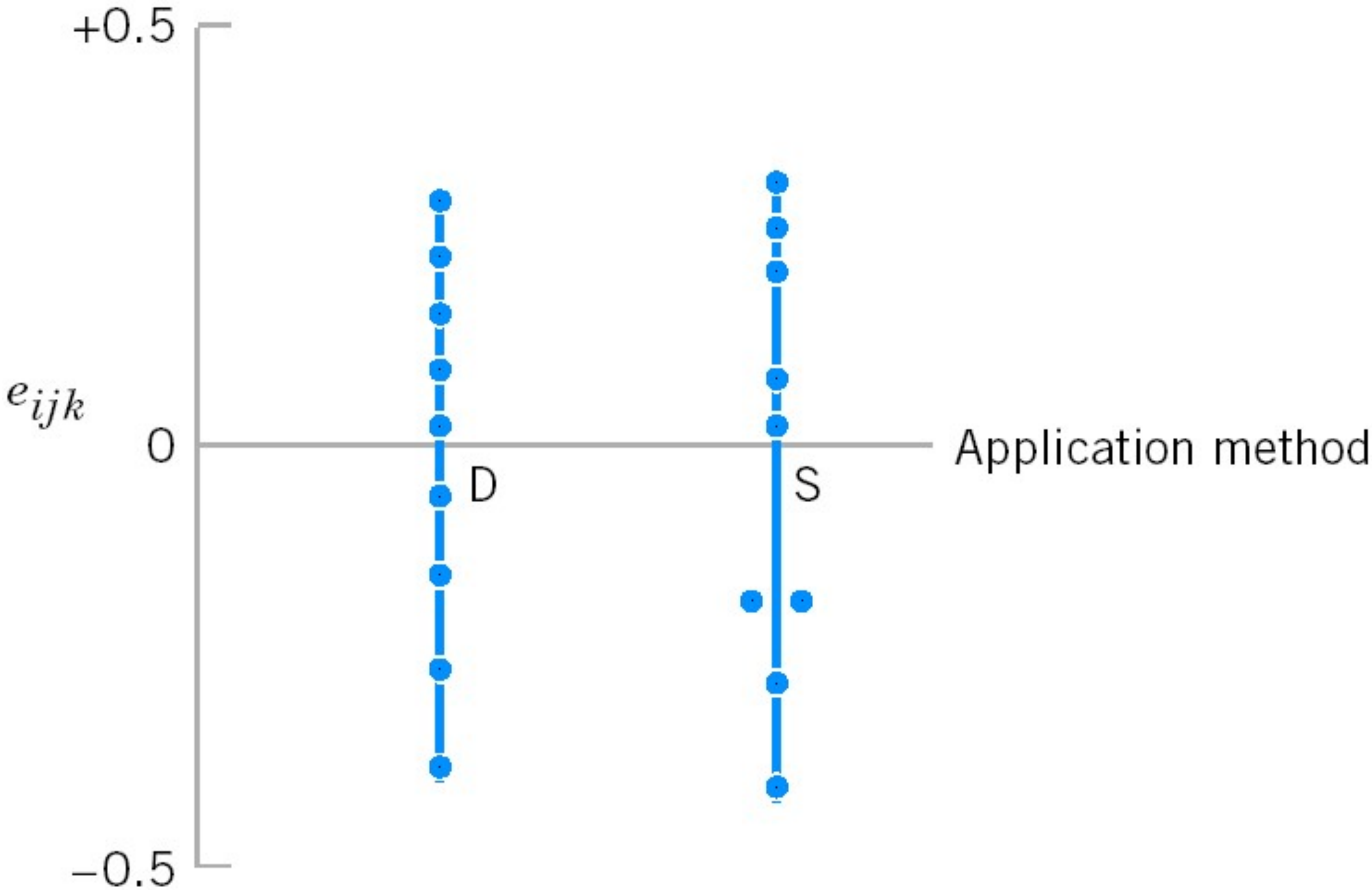
Primer Type	Application Method					
	Dipping			Spraying		
1	−0.27,	0.23,	0.03	0.10,	−0.40,	0.30
2	0.30,	−0.40,	0.10	−0.27,	0.03,	0.23
3	−0.03,	−0.13,	0.17	0.33,	−0.17,	−0.17

Normal probability plot of the residual

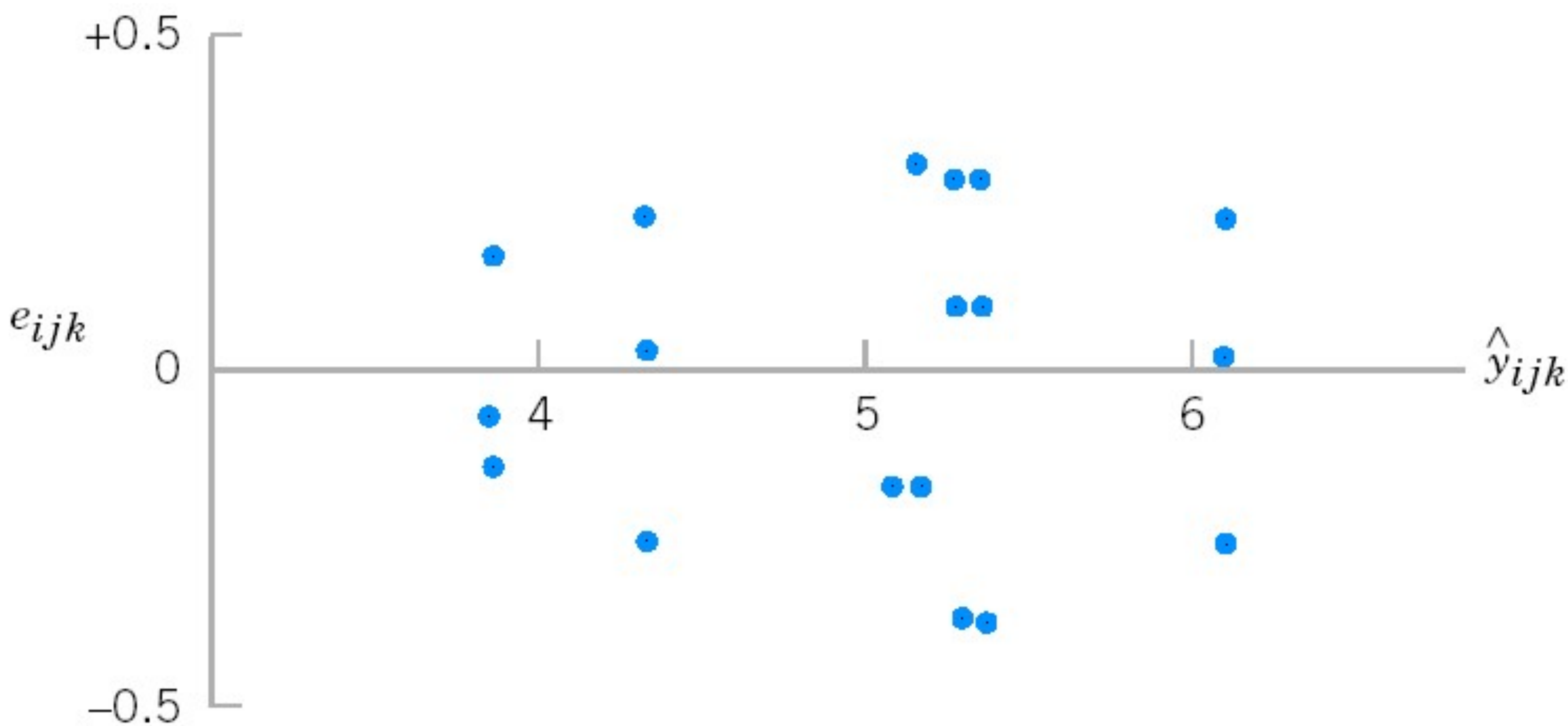




Plot of residuals versus primer type



Plot of residuals versus application method.



Plot of residuals versus predicted values.

Model for a three-factor factorial experiment

$$Y_{ijk\underline{l}} = \mu + \overbrace{\tau_i + \beta_j + \gamma_j + (\tau\beta)_{ij} + (\tau\gamma)_{ik} + (\beta\gamma)_{jk} + (\tau\beta\gamma)_{ijk}}^{\text{7 sources of variations}} + \varepsilon_{ijkl} \begin{cases} i = 1,2,\ldots,a \\ j = 1,2,\ldots,b \\ k = 1,2,\ldots,c \\ l = 1,2,\ldots,n \end{cases}$$

replicates

ال interaction إلهم كلهم سوا

Analysis of Variance Table for the Three-Factor Fixed Effects Model

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Expected Mean Squares	F_0
A	SS_A	$a-1$	MS_A	$\sigma^2 + \dfrac{bcn\sum\tau_i^2}{a-1}$	$\dfrac{MS_A}{MS_E}$
B	SS_B	$b-1$	MS_B	$\sigma^2 + \dfrac{acn\sum\beta_j^2}{b-1}$	$\dfrac{MS_B}{MS_E}$
C	SS_C	$c-1$	MS_C	$\sigma^2 + \dfrac{abn\sum\gamma_k^2}{c-1}$	$\dfrac{MS_C}{MS_E}$
AB	SS_{AB}	$(a-1)(b-1)$	MS_{AB}	$\sigma^2 + \dfrac{cn\sum\sum(\tau\beta)_{ij}^2}{(a-1)(b-1)}$	$\dfrac{MS_{AB}}{MS_E}$
AC	SS_{AC}	$(a-1)(c-1)$	MS_{AC}	$\sigma^2 + \dfrac{bn\sum\sum(\tau\gamma)_{ik}^2}{(a-1)(c-1)}$	$\dfrac{MS_{AC}}{MS_E}$
BC	SS_{BC}	$(b-1)(c-1)$	MS_{BC}	$\sigma^2 + \dfrac{an\sum\sum(\beta\gamma)_{jk}^2}{(b-1)(c-1)}$	$\dfrac{MS_{BC}}{MS_E}$
ABC	SS_{ABC}	$(a-1)(b-1)(c-1)$	MS_{ABC}	$\sigma^2 + \dfrac{n\sum\sum\sum(\tau\beta\gamma)_{ijk}^2}{(a-1)(b-1)(c-1)}$	$\dfrac{MS_{ABC}}{MS_E}$
Error	SS_E	$abc(n-1)$	MS_E	σ^2	
Total	SS_T	$abcn-1$			

EXAMPLE 14-2 Surface Roughness

A mechanical engineer is studying the surface roughness of a part produced in a metal-cutting operation. Three factors, feed rate (A), depth of cut (B), and tool angle (C), are of interest. All three factors have been assigned two levels, and two replicates of a factorial design are run. The coded data are shown in Table 14-10.

Example 14-2

We think the surface roughness produced in metal cutting operations depends on three factors of interest;

feed rate, A, depth of cut, B and tool angle, C

All factors are replicated 2 times for the full design.

8 for each

$2^3 \rightarrow$ 8 runs

Feed Rate (A)	Depth of cut (B)			
	0.025”		0.040”	
	Tool Angle (C)		Tool Angle (C)	
	15°	25°	15°	25°
20” /min	9 7	11 10	9 11	10 8
30” /min	10 12	10 13	12 15	16 14

Solution

Example 14-2

Feed Rate (A)	Depth of cut (B)			
	0.025”		0.040”	
	Tool Angle (C)		Tool Angle (C)	
	15^{o}	25^{o}	15^{o}	25^{o}
20” /min	9 7	11 10	9 11	10 8
30” /min	10 12	10 13	12 15	16 14

For computing SS_{Feed}

Guess SS_{Depth}

Example 14-2

Feed Rate (A)	Depth of cut (B)			
	0.025"		0.040"	
	Tool Angle (C)		Tool Angle (C)	
	15°	25°	15°	25°
20" /min	9 7	11 10	9 11	10 8
30" /min	10 12	10 13	12 15	16 14

For computing SS_{Depth}

Guess SS_{Angle}

Example 14-2

Feed Rate (A)	Depth of cut (B)			
	0.025”		0.040”	
	Tool Angle (C)		Tool Angle (C)	
	15°	25°	15°	25°
20” /min	9 7	11 10	9 11	10 8
30” /min	10 12	10 13	12 15	16 14

For computing SS_{Angle}

Guess $SS_{Feed\ x\ Depth}$

Example 14-2

Feed Rate (A)	Depth of cut (B)			
	0.025”		0.040”	
	Tool Angle (C)		Tool Angle (C)	
	15°	25°	15°	25°
20” /min	9 7	11 10	9 11	10 8
30” /min	10 12	10 13	12 15	16 14

For computing $SS_{\text{Feed x Depth}}$

Guess $SS_{\text{Feed x Angle}}$

Example 14-2

Feed Rate (A)	Depth of cut (B)			
	0.025”		0.040”	
	Tool Angle (C)		Tool Angle (C)	
	15º	25º	15º	25º
20” /min	9 7	11 10	9 11	10 8
30” /min	10 12	10 13	12 15	16 14

For computing $SS_{\text{Feed x Angle}}$

Guess $SS_{\text{Depth x Angle}}$

Example 14-2

Feed Rate (A)	Depth of cut (B)			
	0.025”		0.040”	
	Tool Angle (C)		Tool Angle (C)	
	15°	25°	15°	25°
20” /min	9 7	11 10	9 11	10 8
30” /min	10 12	10 13	12 15	16 14

For computing $SS_{\text{Depth x Angle}}$

Guess $SS_{\text{Feed x Depth x Angle}}$

Example 14-2

Feed Rate (A)	Depth of cut (B)			
	0.025”		0.040”	
	Tool Angle (C)		Tool Angle (C)	
	15°	25°	15°	25°
20” /min	9 7	11 10	9 11	10 8
30” /min	10 12	10 13	12 15	16 14

For computing $SS_{\text{Feed x Depth x Angle}}$

Guess SS_T

Example 14-2

Feed Rate (A)	Depth of cut (B)			
	0.025”		0.040”	
	Tool Angle (C)		Tool Angle (C)	
	15°	25°	15°	25°
20” /min	9 7	11 10	9 11	10 8
30” /min	10 12	10 13	12 15	16 14

For computing SS_T

Minitab ANOVA

ANOVA (Balanced Designs)

Factor	Type	Levels	Values	
Feed	fixed	2	20	30
Depth	fixed	2	0.025	0.040
Angle	fixed	2	15	25

Analysis of Variance for Roughness

Source	DF	SS	MS	F	P
Feed	1	45.563	45.563	18.69	0.003
Depth	1	10.563	10.563	4.33	0.071
Angle	1	3.063	3.063	1.26	0.295
Feed*Depth	1	7.563	7.563	3.10	0.116
Feed*Angle	1	0.062	0.062	0.03	0.877
Depth*Angle	1	1.563	1.563	0.64	0.446
Feed*Depth*Angle	1	5.062	5.062	2.08	0.188
Error	8	19.500	2.437		
Total	15	92.938			

Conclusions ?

lem. The F-ratios for all three main effects and the interactions are formed by dividing the mean square for the effect of interest by the error mean square. Since the experimenter has selected $\alpha = 0.05$, the critical value for each of these F-ratios is $f_{0.05,1,8} = 5.32$. Alternately, we could use the P-value approach. The P-values for all the test statistics are shown in the last column of Table 14-11. Inspection of these

Most likely, both feed rate and depth of cut are important process variables.

Practical Interpretation: Further experiments might study the important factors in more detail to improve the surface roughness.

P-values is revealing. There is a strong main effect of feed rate, since the F-ratio is well into the critical region. However, there is some indication of an effect due to the depth of cut, since $P = 0.0710$ is not much greater than $\alpha = 0.05$. The next largest effect is the AB or feed rate $\times$ depth of cut interaction.