

Notebook

Stat 2

First
Semester
2024/2025



STATISTICS ~ 2

Ch:7 Sampling Distributions and Point Estimation of Parameters.

Statistic $\rightarrow$ Random variable $\rightarrow$ Probability Distributions
"Sampling distributions"

7.1 Point Estimation

Parameter of Population	Point Estimator	Point Estimate
μ	$\hat{\mu}$ or $\bar{x}$	singular numerical value
θ	$\hat{\theta}$	$\hat{\theta}$
σ	δ	5.7 cm
σ^2	δ^2	5.7 ² cm
p	$\hat{p} = \frac{x}{n}$	0.07 cm

μ : average sample single population $\hat{\mu} = \bar{x}$: Sample mean $\bar{x} = \frac{\text{المجموع}}{\text{العدد}}$
 $\hat{\theta}$: single numerical value $M = E(x) = \sum x \cdot f(x)$
 σ^2 : Variance $= \delta^2 = S^2$ $\sigma^2 = V(x) = E(x^2) - (E(x))^2$
 σ : Standard deviation
 p : what we interest to study $= \frac{x}{n}$ x : is the number of items in a random sample of size n .

7.2 Sampling Distributions and the Central Limit Theorem

الاستدلال الإحصائي
 * Statistical inference is concerned with making decisions about a population. $\text{نشاط الاستدلال الإحصائي هو اتخاذ قرارات بشأن المجتمع الإحصائي}$

$x_1, x_2, \dots, x_n$ random variable of size n , every x_i has same Probability distribution

Sampling Distribution: Probability distribution

Cont to Discr $n > 30$

* If we have sample has unknown Prob dist we can suppose the Prob is Normal Dist, just we need large size (n).

These Theorem is called Central Limit Theorem.

* Central Limit Theorem: any unknown Prob Dist "sampling Dist" of the sample mean will still be approximately normal with mean μ and Variance $\frac{\sigma^2}{n}$ تقريباً

إذا احيناً نقرّب أي Dist Normal

* Note $x \sim n(\mu, \sigma^2)$ $Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

* Normal approximately for $\bar{x}$ depends on sample size n.

جداول ال Cumulative Standard Normal Dist
صفحة 764 + 765 بالكتاب

7.3 General Concepts: of Point Estimation.

7.3.1 Unbiased Estimators.

The Point estimator $\hat{\theta}$ is an unbiased estimator for Parameter θ if unbiased $\hat{\theta}$ small estimator call minimum $E(\hat{\theta}) = \theta$ في صحتنا (صالح)

- If the estimator is not unbiased, the the difference $E(\hat{\theta}) - \theta$ is called the bias of the estimator $\hat{\theta}$ if true value equal point estimator, the will be unbiased

7.3.2 Variance of a Point estimator.

If $x_1, x_2, \dots, x_n$ is a random sample of size n from a normal dist with mean μ and variance σ^2 , the sample mean $\bar{x}$ is the MVUE for μ

7.3.3 Standard Error: Reporting a Point Estimate

* The standard error: of an estimator $\hat{\theta}$ is standard deviation, given by $\sigma_{\hat{\theta}} = \sqrt{V(\hat{\theta})}$. if the standard error involves unknown parameter that can be estimated, substitution of those values into $\sigma_{\hat{\theta}}$ produces an estimated standard error, denoted by $\hat{\sigma}_{\hat{\theta}}$

* the standard error of $\bar{x}$ is $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$
if we didn't know σ but substituted the sample standard deviation s into the above equation, the estimated standard error of $\bar{x}$ would be
 $\hat{\sigma}_{\bar{x}} = \frac{s}{\sqrt{n}}$: sample standard deviation

Ch 7 laws

$$\bar{x} = \frac{\sum x_i}{n}, \quad \mu = \bar{x}, \quad \mu = E(x) = \sum x_i \cdot f(x), \quad P = \frac{x}{n}$$

$$P = \frac{x}{n}, \quad \sigma^2 = s^2 = \sigma^2 = V(x) = E(x^2) - (E(x))^2, \quad Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \rightarrow \text{SE Mean}$$

$$\begin{aligned} \text{unbiased} &\rightarrow E(\hat{\theta}) = \theta, & \sigma_{\hat{\theta}} &= \sqrt{V(\hat{\theta})}, & \sigma_{\bar{x}} &= \frac{\sigma}{\sqrt{n}} \\ \text{not unbiased} &\rightarrow E(\hat{\theta}) \neq \theta, & \sigma^2 &= \frac{(x_2 - x_1)^2}{12}, & \hat{\sigma}_{\bar{x}} &= \frac{s}{\sqrt{n}} \end{aligned}$$

$$\mu = \frac{(x_2 + x_1)}{2}$$

$$s^2 = \frac{n \sum x_i^2 - (\sum x_i)^2}{n(n-1)}$$

7.11

7.12

7.22

7.23

sum of squares = $\sigma^2 \times n$

ch8 Statistical Intervals for a Single Sample

* Suppose that $x_1, x_2, \dots, x_n$ is a random sample from a normal dist with unknown mean μ and known variance σ^2 , we know that the sample mean $\bar{x}$ is normally dist with μ and $\frac{\sigma^2}{n}$, we may standardize $\bar{x}$ by subtracting the mean and dividing by the standard deviation, which results in the variable:

$\bar{x} \rightarrow$ normal dist

$\mu_{\bar{x}} \rightarrow \mu$

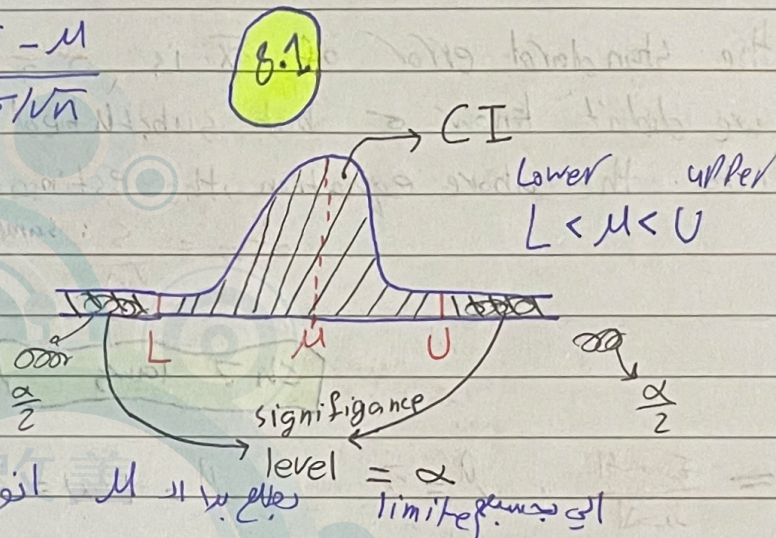
$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

8.1

* Confidence interval:

$$P\{L \leq \mu \leq U\} = 1 - \alpha = CI$$



* Significance level: Prob

* The end points or bounds L and U are called Lower and Upper Confidence limits, respectively,

→ Since Z follows a standard normal distribution, we can write:

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

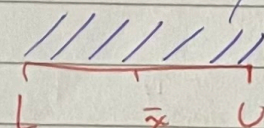
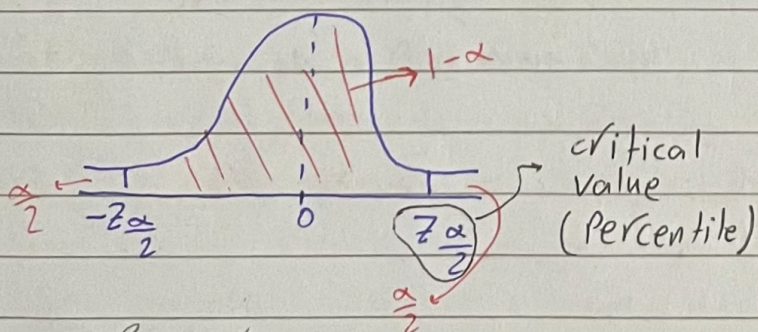
$$P\left\{-Z_{\frac{\alpha}{2}} \leq Z \leq Z_{\frac{\alpha}{2}}\right\} = 1 - \alpha$$

$$P\left(\bar{x} - Z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right) \leq \mu \leq \bar{x} + Z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right)\right) = 1 - \alpha$$

$$\bar{x} - E \leq \mu \leq \bar{x} + E \rightarrow E = Z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right)$$

error

confidence level



σ : standard deviation (given)

n : sample size

$Z_{\frac{\alpha}{2}}$: Percentile (الدرجة)

$$\bar{x} = \frac{\text{مجموع العينة}}{n}$$

* CI length : $= 2E$

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

- 1) as the confidence level increases, CI length increases.
- 2) as the standard deviation increases, length increases.
- 3) as n increases, CI length decreases.

$\beta = S$

Cumulative Standard Normal Distribution جداول جبروت
764 صفحة
765 +
بالكتاب

* one-sided Confidence Bounds

Upper $\rightarrow$

$$\mu \leq U = \bar{x} + Z_{\alpha} \left(\frac{\sigma}{\sqrt{n}} \right)$$

$Z_{\alpha/2}$ بدل

lower $\rightarrow$

$$\bar{x} - Z_{\alpha} \left(\frac{\sigma}{\sqrt{n}} \right) = L \leq \mu$$

Z_{α} بدل

* CI on the Mean of a Normal Dist, σ^2 Unknown

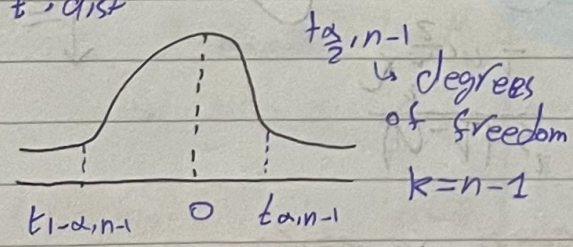
$\bar{x} \rightarrow$ normal Dist
 $S \rightarrow$ sample st. Dev

8.2

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

t Dist جداول جبروت
767 صفحة

$$E = t_{\alpha/2, n-1} \left(\frac{s}{\sqrt{n}} \right)$$



$t_{\alpha, n-1} \rightarrow t_{\alpha/2, n-1}$ حالة one-sided (منسحب)

$$T = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

(Random variable)

6 10 14

للمتغير اختيار اول 2 صف
الساكن بسب الخيرة لباري
14 حتى تكون $\bar{x} = \mu$

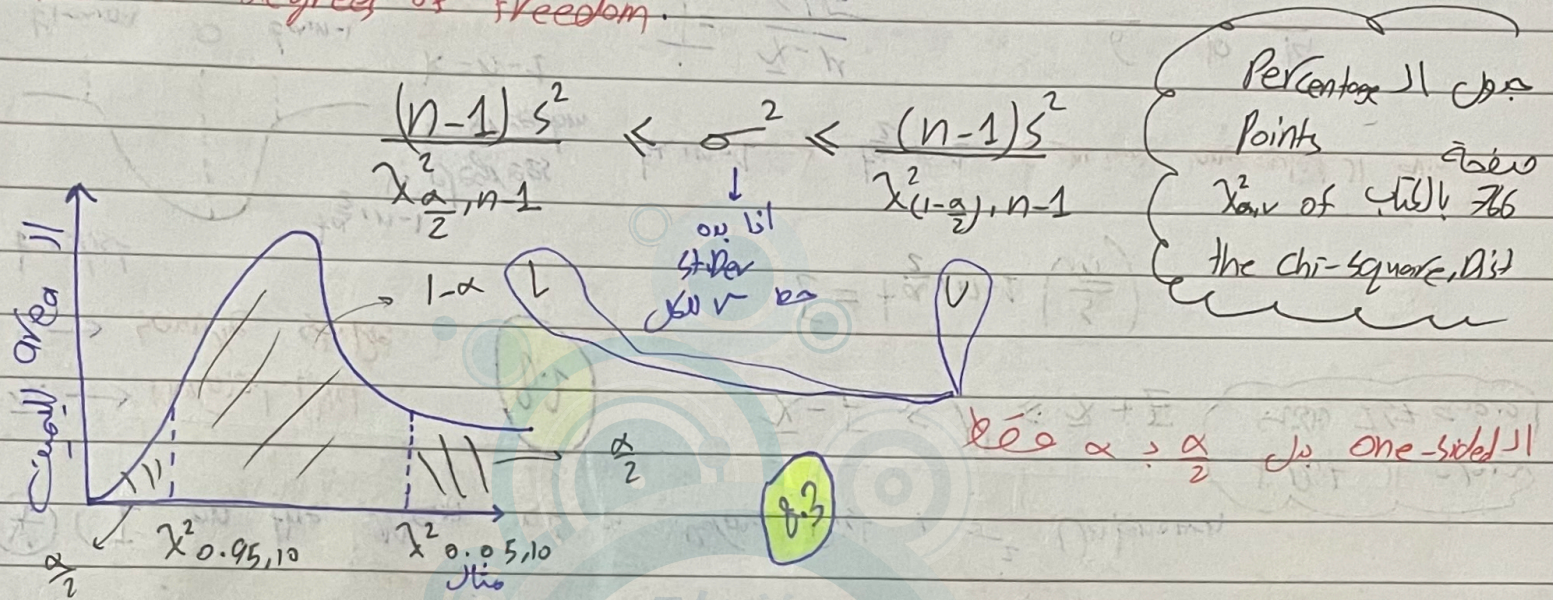
نفس
Sample st. Dev
"S"

$$S = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

$x_i \rightarrow$ values elements

* CI on the Variance and St. Dev of a normal Distribution.

→ Let $X_1, X_2, \dots, X_n$ be a random sample from a normal distribution with mean μ and variance σ^2 , and let s^2 be the sample variance then the random variable $\chi^2 = \frac{(n-1)s^2}{\sigma^2}$ has a chi-square χ^2 dist with $n-1$ degrees of freedom.



* a Large-Sample CI for a Population Proportion:

$P \rightarrow$ Population Proportion

$\hat{P} \rightarrow$ Sample Proportion (Normal Dist)

the standard Error of the Point estimator $\hat{P} = \sqrt{\frac{P(1-P)}{n}}$

$$\mu_{\hat{P}} = P = \frac{x}{n}$$

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

8.9

$$Z = \frac{\hat{P} - P}{\sqrt{\frac{P(1-P)}{n}}} \quad \left| \quad L < P < U \right.$$

$$\downarrow$$

$$-Z_{\alpha/2} < Z < Z_{\alpha/2} \quad \hat{P} = \frac{x}{n}$$

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

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

$$Z = \frac{x - nP}{\sqrt{nP(1-P)}}$$

one-sided

n on list

$$n = \left[\frac{Z_{\alpha/2}}{E} \right]^2 \times \hat{P}(1-\hat{P})$$

at least regardless $P=0.5$

Chapter "8" Laws

8-1 } mean & known σ^2

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

CI length = $2E$: حيز

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

- * No Degrees of freedom

* Choice of $n = \left(\frac{Z_{\alpha} \cdot \sigma}{\frac{2}{E}} \right)^2$

* Confidence level concept

Relationships between CI length & $\frac{E}{Z_{\frac{1}{2}}}$

! one-sided $\mathcal{H} \leftarrow$

Upper: $\mu \leq U = \bar{x} + z_{\alpha} \left(\frac{\sigma}{\sqrt{n}} \right)$

lower: $\bar{x} - z_{\alpha} \left(\frac{\sigma}{\sqrt{n}} \right) = L \leq \mu$

$$\alpha \rightarrow \frac{\alpha}{2} \text{ دہ}$$

* Normal Dist ال جدول *

$\zeta.P \rightarrow 1-2-3-4-5-6$
 $-10-11-14-15-21$

one-sided Lower is always greater than two-sided

Two-sided

ex: Page-7 one " upper " " less "

111

8-2} mean & Unknown σ^2

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

$$E = t_{\alpha, n-1} \left(\frac{s}{\sqrt{n}} \right)$$

* There is Degrees of freedom

• No choice of n

* You must know how to calculate σ_{Der}

$$S = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

elementary xi μ
sample n

$$SE = \frac{\text{st. Dev}}{\sqrt{n}}$$

one-sided test $\rightarrow t_{\alpha, n-1}$ $\sqrt{n}$

* جوں ال E, Dist *

S.P. $\rightarrow$ 24-25-26-28-30
-31-34-38-40

8-3 } variance

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

• There is Degree of freedom

2. No choice of n

$\alpha \rightarrow \frac{\alpha}{2}$ is one-sided $\leftarrow$

* You must know how to calculate chi-sq. * جدول
chi-square

4P → 46-48-49-50-51

8-9} Proportion

$$\hat{p} - E \leq p \leq \hat{p} + E$$

$$E = z_{\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}}$$

2 No Degrees of freedom

* Choice of 'n' $n = \left(\frac{Z_{\alpha/2}}{E} \right)^2 \times p(1-p)$

- Standard Error = $\sqrt{\hat{p}(1-\hat{p})}$

* normal dist $\sqrt{\frac{1}{n}}$ * σ

* at least + regardless $\therefore P=0.5$

 $\alpha \neq \frac{\pi}{2}$ is one side of the line.

LP $\rightarrow$ 59-62-63

Ch:9 Test of Hypotheses for a Single Sample:

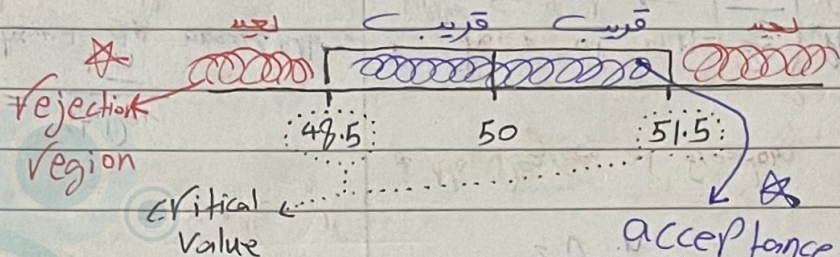
* **Statistical hypothesis**: is a statement about the parameters of one or more populations. (two opposite statements about a population parameter)
 hypothesis value (μ_0)

$H_0: \mu = 50$ null hypothesis

$H_1: \mu \neq 50$ alternative hypothesis Which statement is true?

Statements

☆ : أنا كانت
 قيمة $\bar{x}$ داخل
 المنطقة التي

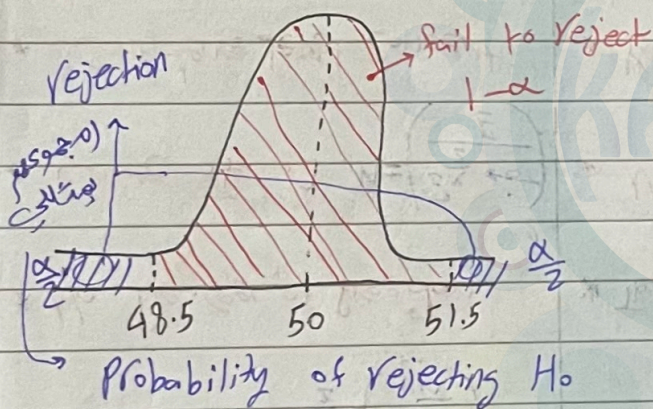


Sample
 بمتوسط $\bar{x} = 42$
 بغير alternative مع

$\bar{x} = 50.5$
 null لا بغير

acceptance
 region
 (fail to reject
 region)

* Population mean $\rightarrow$ Sample $\bar{x}$ $\rightarrow$ Normal dist
 (test statistic)

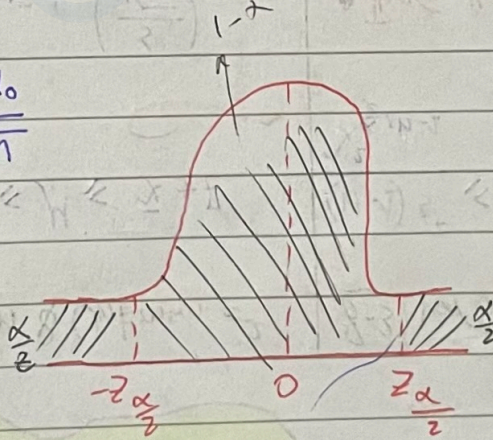


إذا Reject to H_0 بالواقع H_0 is true

$P(\text{type I error}) = \alpha$
 significance level

* standardization: $Z_0 = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$

test
 statistic
 Reject or fail to reject.



* Two-sided Test: $H_0: \mu = \mu_0$
 $H_1: \mu \neq \mu_0$

* one-sided Tests:

$H_0: \mu = \mu_0$ or $H_0: \mu = \mu_0$
 $H_1: \mu > \mu_0$ $H_1: \mu < \mu_0$

* Procedure for analysis:

1. Parameter of interest
2. null & alternative hypothesis
3. calculate test statistic and critical values.
4. Draw a conclusion

9.2

* Tests on the mean, Variance known.

$$H_0: \mu = \mu_0 \text{ null}$$

$$H_1: \mu \neq \mu_0 \text{ alternative}$$

fail to reject H_0 (إذا كان قرار) H_0 is false (بالتالي)

$$P(\text{type II error}) = \beta$$

$$P_{SI} = \frac{\text{Pound}}{m^2}$$

$$P_{AS} = \frac{N}{m^2}$$

* Test statistic: $Z_0 = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$

fake $\left\{ \begin{array}{l} \text{fail to reject type 2 } \beta \\ \text{Reject No error (Power) } 1 - \beta \end{array} \right.$

	H_0 is True	H_0 is False β
fail to reject H_0	no error	type II error
Reject H_0	type I error α	no error $1 - \beta = \text{Power}$

إذا β موجود لا يكون H_0 null أو false وبالتالى β خطأ في μ الـ null

$$\delta = \mu - \mu_0$$

- as α increase, β decrease
- as n increase, β and α decrease
- as δ increase, β decrease

(Cumulative) $\rightarrow$ Standard Normal Distribution $\frac{769}{765}$

$\rightarrow$ to calculate β :

$\Phi(z) \rightarrow$ the Probability to the left of z in standard normal distribution

$$\Phi(x \geq 4) = 0 \quad \Phi(x \leq 4) = 0$$

$$\beta = \Phi\left[\frac{z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma}}{1}\right] - \Phi\left[\frac{-z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma}}{1}\right]$$

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

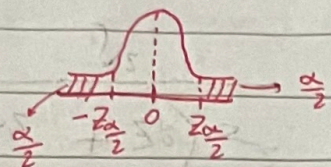
$$d = \frac{|\mu - \mu_0|}{\sigma} = \frac{|\delta|}{\sigma}$$

9-46

* Two-sided

$$H_0: \mu = 50$$

$$H_1: \mu \neq 50$$



* one-sided upper-tailed

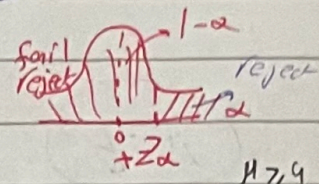
exceeds

$$\mu > 50$$

$$\mu < 50$$

$$H_0: \mu = 50$$

$$H_1: \mu > 50$$



* \beta at upper-tailed

$$\beta = \Phi\left[\frac{z_{\alpha} - \delta\sqrt{n}}{\sigma}\right]$$

* one-sided lower-tailed

at least \alpha \rightarrow

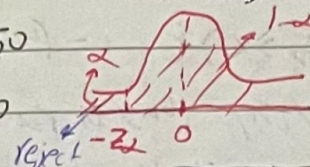
at most \alpha \rightarrow \mu \leq \alpha

$$\mu > 50$$

$$\mu < 50$$

$$H_0: \mu = 50$$

$$H_1: \mu < 50$$



* \beta at lower-tailed

$$\beta = 1 - \Phi\left[-z_{\alpha} - \frac{\delta\sqrt{n}}{\sigma}\right]$$

one-sided \mu \leq \alpha

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

* The P-Value is the smallest level of significance that would lead to rejection of the null hypothesis H_0 with the given data.

[area to the right of test statistic] P-Value

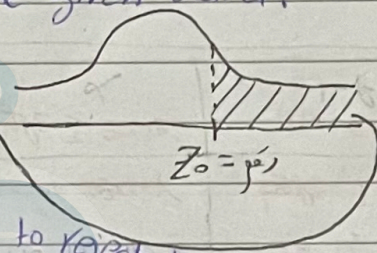
test statistic

major

smaller than \alpha

$$P\text{-value} = 1 - \Phi(z)$$

if $P\text{-value} > \alpha \rightarrow$ fail to reject H_0
 $P\text{-value} < \alpha \rightarrow$ reject H_0



Null Hypothesis: $H_0: \mu = \mu_0$

Test Statistic

$$z_0 = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

Alternative hypothesis

Rejection criteria

$$H_1: \mu \neq \mu_0$$

$$z_0 > z_{\alpha/2, n-1} \text{ or } z_0 < -z_{\alpha/2, n-1}$$

$$H_1: \mu > \mu_0$$

$$z_0 > z_{\alpha, n-1}$$

$$H_1: \mu < \mu_0$$

$$z_0 < -z_{\alpha, n-1}$$

P-values in Hypothesis tests

for a two-sided tailed test

$$H_0: \mu = \mu_0 \quad H_1: \mu \neq \mu_0$$

$$P\text{-value} = 2[1 - \Phi(z_0)]$$

for an upper tailed test

$$H_0: \mu = \mu_0 \quad H_1: \mu > \mu_0$$

$$P\text{-value} = 1 - \Phi(z_0)$$

for a lower tailed test

$$H_0: \mu = \mu_0 \quad H_1: \mu < \mu_0$$

$$P\text{-value} = \Phi(z_0)$$

9.3 Test on the mean of a Normal Dist, variance Unknown

Null Hypothesis : $H_0: \mu = \mu_0$

Test statistic : $T_0 = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$

alternative hypothesis

$H_1: \mu \neq \mu_0$

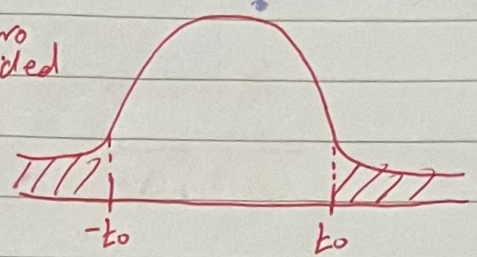
$H_1: \mu > \mu_0$

$H_1: \mu < \mu_0$

Rejection criteria

$T_0 \leftarrow \begin{cases} t_0 > t_{\alpha/2, n-1} \text{ or } t_0 < -t_{\alpha/2, n-1} \\ t_0 > t_{\alpha, n-1} \\ t_0 < -t_{\alpha, n-1} \end{cases}$

two sided

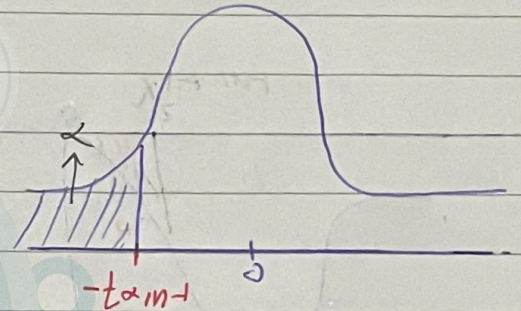
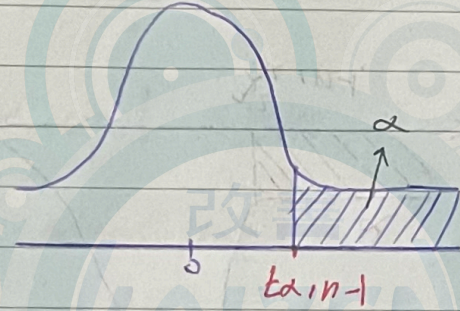
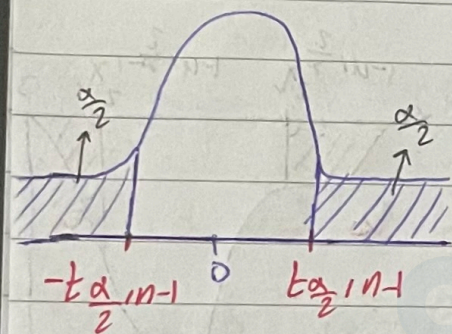


Critical values: $t_{\alpha/2, n-1}$

$\frac{\alpha}{2}$ or α for one-sided test

767

area t-dist

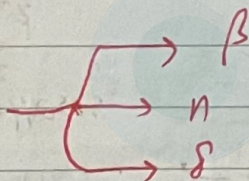


$$* S = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

⊗ Type two error and choice of n :

$$\beta = P\{-t_{\alpha/2, n-1} \leq T_0 \leq t_{\alpha/2, n-1}\}$$

* OC curves



لازم يكون جملتي
تحت نفس مشات
في التوزيع

776

$$J = \frac{|\mu - \mu_0|}{s} = \frac{1.81}{s}$$

2 قوت -3

two or one-sided-2

test ال نوع

$$\text{Power} = 1 - \beta$$

9.4

Hypothesis Tests on the Variance and standard Deviation of a normal Distribution

$$H_0 : \sigma^2 = \sigma_0^2$$

$$H_1 : \sigma^2 \neq \sigma_0^2$$

* Test statistic : $\chi_0^2 = \frac{(n-1)s^2}{\sigma_0^2}$

Chi-square χ^2 degree of freedom = $n-1$

Null Hypothesis : $H_0 : \sigma^2 = \sigma_0^2$

Alternative Hypothesis

$$H_1 : \sigma^2 \neq \sigma_0^2$$

$$H_1 : \sigma^2 > \sigma_0^2$$

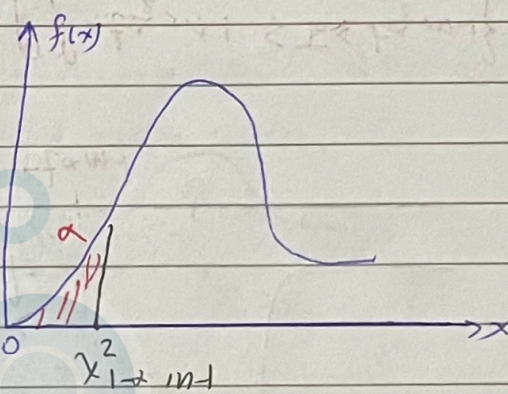
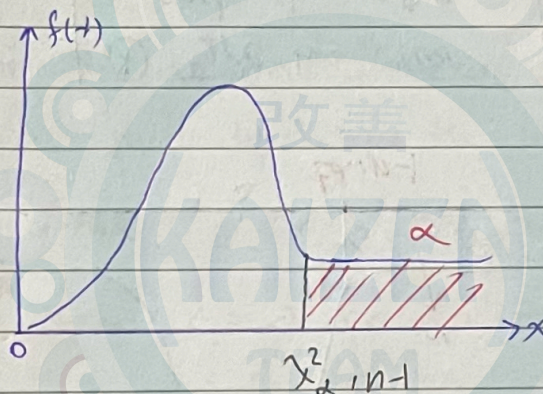
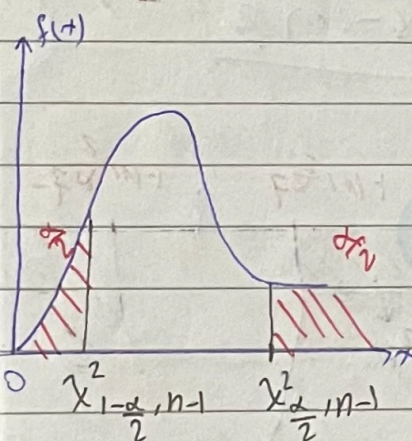
$$H_1 : \sigma^2 < \sigma_0^2$$

Rejection Criteria

$$\chi_0^2 > \chi_{\frac{\alpha}{2}, n-1}^2 \text{ or } \chi_0^2 < \chi_{1-\frac{\alpha}{2}, n-1}^2$$

$$\chi_0^2 > \chi_{\alpha, n-1}^2$$

$$\chi_0^2 < \chi_{1-\alpha, n-1}^2$$



* for Type II error and choice of n $\lambda = \frac{s}{\sigma_0}$

9.5 Test on a Population Proportion :

z-dist

$$H_0 : P = P_0$$

$$H_1 : P \neq P_0$$

* Test statistic : $Z_0 = \frac{x - nP_0}{\sqrt{nP_0(1-P_0)}}$
and reject $H_0 : P = P_0$ if
 $Z_0 > Z_{\alpha/2}$ or $Z_0 < -Z_{\alpha/2}$

$$Z_0 = \frac{\hat{P} - P_0}{\sqrt{\frac{P_0(1-P_0)}{n}}}$$

$$Z_0 = \frac{\frac{x}{n} - P_0}{\sqrt{\frac{P_0(1-P_0)}{n}}}$$

* Type II Error and Choice of sample size :

For a two-sided alternative :

$$\beta = \Phi \left[\frac{P_0 - P + Z_{\alpha/2} \cdot \sqrt{P_0(1-P_0)/n}}{\sqrt{P(1-P)/n}} \right] - \Phi \left[\frac{P_0 - P - Z_{\alpha/2} \cdot \sqrt{P_0(1-P_0)/n}}{\sqrt{P(1-P)/n}} \right]$$

If the alternative is $P < P_0$: $\downarrow$

$$\beta = 1 - \Phi \left[\frac{P_0 - P - Z_{\alpha} \cdot \sqrt{P_0(1-P_0)/n}}{\sqrt{P(1-P)/n}} \right]$$

If the alternative is $P > P_0$: $\uparrow$

$$\beta = \Phi \left[\frac{P_0 - P + Z_{\alpha} \cdot \sqrt{P_0(1-P_0)/n}}{\sqrt{P(1-P)/n}} \right]$$

For a two-sided alternative

$$n = \left[\frac{Z_{\alpha/2} \sqrt{p_0(1-p_0)} + Z_{\beta} \sqrt{p(1-p)}}{p - p_0} \right]^2$$

For a one-sided alternative

$$n = \left[\frac{Z_{\alpha} \sqrt{p_0(1-p_0)} + Z_{\beta} \sqrt{p(1-p)}}{p - p_0} \right]^2$$

at least $\geq$ most of 70.5
at most $\leq$ اكبر من النصف

more than $>$
less than $<$

Null hypotheses: $H_0: p = p_0$

Test statistics: $Z_0 = \frac{x - np_0}{\sqrt{np_0(1-p_0)}}$

Alternative Hypothesis

P-value

Rejection criterion for Fixed-Level Tests

$H_1: p \neq p_0$

Probability above $|z_0|$ and below $-|z_0|$, $p = 2[1 - \phi(z_0)]$

$Z_0 > Z_{\alpha/2}$ or $Z_0 < -Z_{\alpha/2}$

$H_1: p > p_0 \cup$

Probability above z_0 , $p = 1 - \phi(z_0)$

$Z_0 > Z_{\alpha}$

$H_1: p < p_0 \rightarrow$

Probability below z_0 , $p = \phi(z_0)$

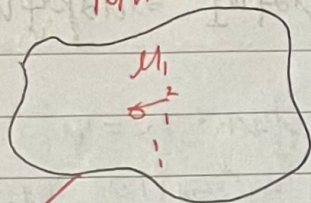
$Z_0 < -Z_{\alpha}$

Ch: 10 Statistical Inference for Two Samples.

10.1

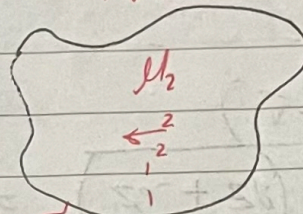
Inference on Difference in mean, variance known:

Population 1



Sample: $\bar{x}_1$

Population 2



Sample: $\bar{x}_2$

single $\left\{ \begin{array}{l} \text{Point Estimation} \\ \text{CI} \\ \text{Hypothesis testing} \end{array} \right.$

The Parameter of interest:

Null $H_0: \mu_1 - \mu_2 = \Delta_0$

alternative $H_1: \mu_1 - \mu_2 \neq \Delta_0$

مثال $-1 \leq \mu_1 - \mu_2 \leq 1.5$

Critical Values $\left\{ \begin{array}{l} Z_{\frac{\alpha}{2}} \\ -Z_{\frac{\alpha}{2}} \end{array} \right.$

$\Delta_0 = \mu_1 - \mu_2$

Δ = true difference of means

* Point Estimator of μ_1 is $\bar{x}_1$
* Point Estimator of μ_2 is $\bar{x}_2$ $\rightarrow$ Point Estimator of $\mu_1 - \mu_2$ is $\bar{x}_1 - \bar{x}_2$

$\rightarrow \bar{x} = \text{Normally Distributed}$

$\rightarrow \mu_{\bar{x}_1} = \mu = E(x)$ expected value of x

$\rightarrow \sigma_{\bar{x}_1} = \frac{\sigma_1}{\sqrt{n}} \rightarrow \sigma_{\bar{x}_1}^2 = \frac{\sigma_1^2}{n}$

$\left\{ \begin{array}{l} \bar{x}_1 - \bar{x}_2 \rightarrow \text{is normally distributed} \\ \text{so normal variable} - \text{normal variable} = \text{normally dist variable} \end{array} \right.$

* Mean and Expected Value: $E(\bar{x}_1, \bar{x}_2) = E(\bar{x}_1) - E(\bar{x}_2) = \mu_1 - \mu_2$

* Variance and σ^2 : $V(\bar{x}_1 - \bar{x}_2) = V(\bar{x}_1) + V(\bar{x}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$

* standard Dev σ : $\sigma(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

* Test Statistic: $Z_0 = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$

جواب (ن.د) 764
765 +
764

Type II Error and choice of sample size

* OC curves: d, β, n } OC curves can be calculated
 test 1) for μ } μ is the mean
 sided-two, one 1) μ } d, n, β
 0.01, 0.05 α 2) μ } one-sided or two-sided μ

$$d = \frac{|\mu_1 - \mu_2 - \Delta_0|}{\sqrt{\sigma_1^2 + \sigma_2^2}} = \frac{|\Delta - \Delta_0|}{\sqrt{\sigma_1^2 + \sigma_2^2}} \quad \left. \begin{array}{l} \text{Two-sided and} \\ \text{one-sided alternative} \end{array} \right\}$$

* If $n_1 \neq n_2$, use $n = \frac{\sigma_1^2 + \sigma_2^2}{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ to calculate n .

$$\beta = \Phi \left[\frac{z_{\alpha/2} - \frac{\Delta - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \right] - \Phi \left[-\frac{z_{\alpha/2} - \frac{\Delta - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \right] \quad \left. \begin{array}{l} \text{Two-sided} \end{array} \right\}$$

$$\beta = \Phi \left[\frac{z_{\alpha} - \frac{\Delta - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \right] \quad \left. \begin{array}{l} \text{one-sided} \\ \text{upper} \end{array} \right\} \quad \beta = \left[1 - \Phi \left[-\frac{z_{\alpha} - \frac{\Delta - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \right] \right] \quad \left. \begin{array}{l} \text{one-sided} \\ \text{lower} \end{array} \right\}$$

Power = $1 - \beta$

* n for a two-sided test on the difference in Means with $n_1 = n_2$, Variance known }
$$n = \frac{(z_{\alpha/2} + z_{\beta})^2 (\sigma_1^2 + \sigma_2^2)}{(\Delta - \Delta_0)^2}$$

* n for a one-sided test on the difference in Means with $n_1 = n_2$, Variance known }
$$n = \frac{(z_{\alpha} + z_{\beta})^2 (\sigma_1^2 + \sigma_2^2)}{(\Delta - \Delta_0)^2}$$

* Confidence Interval on a Difference in Means with Variance known }
$$\bar{x}_1 - \bar{x}_2 - E \leq \mu_1 - \mu_2 \leq \bar{x}_1 - \bar{x}_2 + E$$

 Two sided: $E = z_{\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

* n for a CI on Difference in Means with Variance known }
$$n = \left(\frac{z_{\alpha/2}}{E} \right)^2 (\sigma_1^2 + \sigma_2^2)$$

one-sided Lower-confidence Bound: $\bar{x}_1 - \bar{x}_2 - z_{\alpha} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \leq \mu_1 - \mu_2$ (re)
10.3, 10.4 $\hookrightarrow$ Hypothesis: $\mu_1 - \mu_2 = 0$

$$p\text{-value} = 1 - \Phi(z_0)$$

$H_0: \mu_1 - \mu_2 = 0$	$H_1: \mu_1 - \mu_2 \neq 0$	$Z_0 = Z_{\alpha/2}$	$Z_0 < -Z_{\alpha/2}$
$H_0: \mu_1 - \mu_2 \leq 0$	$H_1: \mu_1 - \mu_2 > 0$	$Z_0 < -Z_{\alpha}$	$Z_0 > Z_{\alpha}$

Normal Distributions, Variances Unknown:

Test statistic, Critical values, Degree of freedom, CI

Diagram illustrating two populations, μ_1 and μ_2 , each with a standard deviation σ^2 . Arrows point from each population to a sample, labeled "sample 1" and "sample" respectively.

↓
sample 2

↓ ↓
51² 52²
Cura
Cura
Cura

* The pooled estimator of σ^2 is S_p^2

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

Degree of freedom = $n_1 + n_2 - 2$

* Test statistic : $T_0 = \frac{\bar{x}_1 - \bar{x}_2 - \Delta_0}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$

* Critical values: $T_{\frac{\alpha}{2}, n_1+n_2-2}$

في حالة one-sided $b > a$ بن $\frac{a}{2}$

* CI on the Difference in means:

variance unknown, $\sigma_1^2 = \sigma_2^2 = \sigma^2$ (case 2)

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

$$E = t \frac{\alpha}{2}, n_1 + n_2 - 2 \quad \times \quad \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}$$

767 جاب ال t-dist قسم

Null hypothesis $H_0: \mu_1 - \mu_2 = \Delta_0$

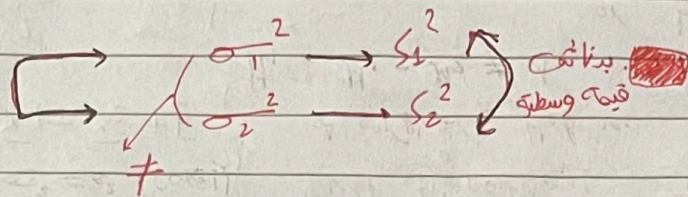
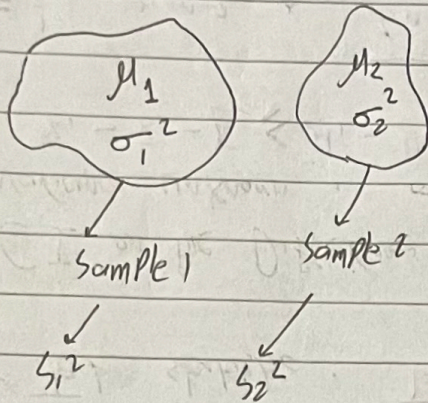
Alternative Hypothesis	Rejection Criterion
$\mu > \mu_0$	$\bar{X} > \bar{x}_c$
$\mu < \mu_0$	$\bar{X} < \bar{x}_c$
$\mu \neq \mu_0$	$\bar{X} > \bar{x}_c$ or $\bar{X} < \bar{x}_c$

$$H_1: \mu_1 - \mu_2 \neq \Delta_0 \quad T_0 \leftarrow t_0 > t_{\frac{\alpha}{2}}, n_1 + n_2 \geq 2$$

or $t_0 < -t \frac{\alpha}{2}, n_1 + n_2 - 2$

$$H_1: \mu_1 - \mu_2 > \Delta_0 \cup t_0 > t_{\alpha, n_1+n_2-2}$$
$$H_1: \mu_1 - \mu_2 < \Delta_0 \quad \vee \quad t_0 < -t_{\alpha, n_1+n_2-2}$$

Case 2: $\sigma_1^2 \neq \sigma_2^2$



$$df = \left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \right)^{-1} \times \text{Degree of freedom}$$

integer

* Test statistic: $T_0 = \frac{\bar{x}_1 - \bar{x}_2 - \Delta_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$

* Critical values: $T_{\frac{\alpha}{2}, df}$
 $\frac{\alpha}{2}$ or α → one-sided or two-sided

* CI on the Difference in means, variance unknown, ($\sigma_1^2 \neq \sigma_2^2$) Case 2

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

$$E = t_{\frac{\alpha}{2}, df} \times \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

767 أو t-dist أو

10.4 Paired t-Test:

Parameter of interest

$$\mu_1 - \mu_2$$

Sample 1	Sample 2	Difference
x_{11}	x_{21}	$x_{11} - x_{21}$
x_{12}	x_{22}	$x_{12} - x_{22}$
$\vdots$	$\vdots$	$\vdots$
x_{1n}	x_{2n}	$x_{1n} - x_{2n}$

* Test statistic: $T_0 = \frac{\bar{D} - \Delta_0}{s_D / \sqrt{n}}$

* degree of freedom is $n-1$

* Critical values $t_{\frac{\alpha}{2}, n-1}$

$\frac{\alpha}{2}$ or α → one-sided or two-sided

Null hypothesis : $H_0: \mu_D = \Delta_0$

Alternative Hypothesis

Rejection Region

$H_1: \mu_D \neq \Delta_0$

$t_0 > t_{\frac{\alpha}{2}, n-1}$ or $t_0 < -t_{\frac{\alpha}{2}, n-1}$

$H_1: \mu_D > \Delta_0$

$t_0 > t_{\alpha, n-1}$

$H_1: \mu_D < \Delta_0$

$t_0 < -t_{\alpha, n-1}$

* CI for μ_D

$\bar{D} - E \leq \mu_D \leq \bar{D} + E$

$S = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$

$E = t_{\frac{\alpha}{2}, n-1} * \frac{SD}{\sqrt{n}}$

one sided $\frac{\alpha}{2}$ do α do

105 Inferences on the Variances of Two Normal Populations

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

F-distribution

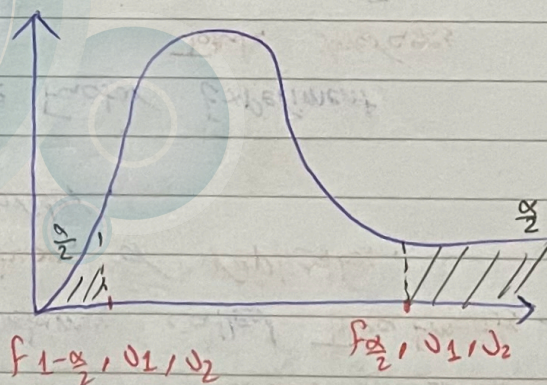
with v_1 and v_2 Degree of freedom

$F = \frac{\frac{\chi_1^2 / v_1}{\chi_2^2 / v_2} = \frac{W/U}{K/V}$

$f_{1-\frac{\alpha}{2}, v_1, v_2}$

$f_{1-\frac{\alpha}{2}, v_1, v_2} = \frac{1}{f_{\frac{\alpha}{2}, v_2, v_1}}$

F-dist 766 area



* test statistic: $F_0 = \frac{S_1^2}{S_2^2}$

* Type two error: $\gamma = \frac{\sigma_1}{\sigma_2}$

* CI on the Ratio of two variances

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

Alternative Hypothesis

Rejection Criterion

$H_1: \sigma_1^2 \neq \sigma_2^2$

$f_0 > f_{\frac{\alpha}{2}, n_1-1, n_2-1}$ or

$f_0 < f_{1-\frac{\alpha}{2}, n_1-1, n_2-1}$

$H_1: \sigma_1^2 > \sigma_2^2$

$f_0 > f_{\alpha, n_1-1, n_2-1}$

$H_1: \sigma_1^2 < \sigma_2^2$

$f_0 < f_{1-\alpha, n_1-1, n_2-1}$

$\frac{S_1^2}{S_2^2} * f_{(1-\frac{\alpha}{2}), v_2, v_1} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{S_1^2}{S_2^2} * f_{\frac{\alpha}{2}, v_2, v_1}$

$v_1 = n_1 - 1$

$v_2 = n_2 - 1$

Design and Analysis of Single-Factor:

Factor		observations or replicates						Totals	Averages
Concentrations		1	2	3	4	5	6		
Treatments	level 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
Balanced means Level		اذا كانت كل ال obs متساوية في كل Level						383	15.96

- * The levels of the factor are sometimes called **Treatments**.
- * Each treatment has six observations or **Replicates**.
- * The runs are run in **random** order.

Typical Data for a Single Factor Experiment

Treatment	observations				Totals	Averages
1	y_{11}	y_{12}	...	y_{1n}	$y_{1.}$	$\bar{y}_{1.}$
2	y_{21}	y_{22}	...	y_{2n}	$y_{2.}$	$\bar{y}_{2.}$
...	...	...	...	...	...	...
a	y_{a1}	y_{a2}	...	y_{an}	$y_{a.}$	$\bar{y}_{a.}$
					$y_{..}$	$\bar{y}_{..}$

- y_{an} : n observation في ال Treatment a
- $y_{a.}$: مجموع ال Treatment a
- $\bar{y}_{a.}$: افرج ال Treatment a
- $y_{..}$: مجموع الجداول
- $\bar{y}_{..}$: افرج الافرج

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

$$Y_{ij} = \mu_{ij} + \epsilon_{ij} \quad \begin{cases} i = 1, 2, \dots, a \\ j = 1, 2, \dots, n \end{cases}$$

* We wish to test the hypotheses:

$$H_0: \tau_1 = \tau_2 = \dots = \tau_a = 0$$

$$H_1: \tau_i \neq 0 \text{ for at least one } i$$

* The sum of squares symbolically

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

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

$$N = a \cdot n$$

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

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

* The sum of squares identity is:

$$\sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{..})^2 = n \sum_{i=1}^a (\bar{y}_{i.} - \bar{y}_{..})^2 + \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{i.})^2$$

* The Expected Value of the treatment sum of squares is:

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

* The Expected Value of the error sum of squares is:

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

* The mean square for treatments:

$$MS_{\text{Treatments}} = \frac{SS_{\text{Treatments}}}{(a-1)}$$

* The mean square for error:

$$MS_E = \frac{SS_E}{a(n-1)}$$

We would reject H_0

$$\text{if } F_0 > F_{\alpha, a-1, a(n-1)}$$

(F_{dist} join me)

* F_0 is called The appropriate test statistic:

$$F_0 = \frac{MS_{\text{Treatments}}}{MS_E}$$

Sum of squares	Degree of freedom	Mean square	The Analysis of Variance for a Single-Factor Experiment, Fixed-Effect model
Treatments $SS_{\text{Treatment}}$	$a-1$	$MS_{\text{Treatment}}$	
Error SS_E	$a(n-1)$	MS_E	
Total SS_T	$an-1$		

في اول جدول ، تعال نفرض $\alpha = 0.01$ كل اشي

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

$a=4$
 $n=6$
 $N=24$

بتحسب الترتيب لكل Element

$$= 7^2 + 8^2 + \dots + 20^2 - \frac{383^2}{24} = 512.96$$

بالجدول كامل واجهوه
وطرح منهم $\frac{(\text{افرج الافرج})^2}{N}$

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

نقسم كل مجموع اصف وبتريجه $\frac{(\text{افرج الافرج})^2}{N}$
ونجمع واقسمه على n وطرح

$$= \frac{(60)^2 + (94)^2 + (102)^2 + (127)^2}{6} - \frac{(383)^2}{24} = 382.79$$

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

$$= 512.96 - 382.79 = 130.17$$

بعد ما تطلع قيمه اعمل قالم جدول :

Sum of squares

DOF

Mean square

$$F_0 = \frac{127.60}{6.51} = 19.60$$

$$SS_{\text{Treatment}} = 382.79$$

$$4 - 1 = 3$$

$$\frac{382.79}{3} = 127.60$$

$$SS_{\text{Error}} = 130.17$$

$$4(6-1) = 20$$

$$\frac{130.17}{20} = 6.51$$

$$SS_T = 512.96$$

$$4 \times 6 - 1 = 23$$

$$F_{\alpha, a-1, a(n-1)}$$

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

$F_0 > F_{\alpha, a-1, a(n-1)}$ we reject H_0

كل اشي تطلع ال P-value :

$$P\text{-Value} = P(F_{3,20} > 19.6) = 3.59 \times 10^{-6}$$

* A $100(1-\alpha)$ Percent confidence interval on the mean of the i^{th} treatment μ_i is :

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

* A $100(1-\alpha)$ Percent confidence interval on the difference in two treatments means $\mu_i - \mu_j$ is :

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

* The least significant difference (LSD) is :

$$LSD = t_{\frac{\alpha}{2}, a(n-1)} \times \sqrt{\frac{2MS_E}{n}}$$

* If the sample sizes are different in each treatment :

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

if $|\bar{y}_{i.} - \bar{y}_{j.}| > LSD \therefore$ Significant

* Residual Analysis and Model checking : Residual = $y_{ij} - \bar{y}_{i.}$

↑
element - average

⊛ The Randomized Block Design : is an extension of the Paired t-test to situations where the factor of interest has more than two levels.

Chemical Type	Fabric Sample					Treatment Totals	Treatment Averages
	1	2	3	4	5	$\sum y_{i.}$	$\bar{y}_{i.}$
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

$\sum y_{.j}$ Block Totals	9.2	10.1	3.5	8.8	7.6	39.2 ($\sum y_{..}$)	
$\bar{y}_{.j}$ Block Averages	2.30	2.53	0.88	2.20	1.96		1.96 ($\bar{y}_{..}$)

A Randomized Complete Block Design with a Treatments and b Blocks

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

y_{ab} : data الموجودة في ال treatment a وال Block b
 $y_{.b}$: مجموع البلوك b
 $y_{a.}$: افرم الترتيبات a
 $\bar{y}_{.b}$: افرم البلوك b
 $\bar{y}_{a.}$: افرم الافرجا -
 $y_{..}$: مجموع المجاميع
 $\bar{y}_{..}$: مجموع الترتيبات a

$$Y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij} \quad \left\{ \begin{array}{l} i = 1, 2, \dots, a \\ j = 1, 2, \dots, b \end{array} \right.$$

* We are interested in testing : $H_0 : \tau_1 = \tau_2 = \dots = \tau_a = 0$

$H_1 : \tau_i \neq 0$ at least one i

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

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

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

$a \rightarrow$ Treatments

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

$b \rightarrow$ Blocks

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

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

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

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

* The expected values of these mean squares are:

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

* The mean square for treatments:

$$MS_{\text{Treatments}} = \frac{SS_{\text{Treatments}}}{a-1}$$

* The mean square for blocks:

$$MS_{\text{Blocks}} = \frac{SS_{\text{Blocks}}}{b-1}$$

* The mean square for error:

$$MS_E = \frac{SS_E}{(a-1)(b-1)}$$

* Residual Analysis and Model checking:

$$e_{ij} = y_{ij} - \hat{y}_{ij}$$

$$\hat{y}_{ij} = \bar{y}_{i.} + \bar{y}_{.j} - \bar{y}_{..}$$

	Sum of Squares	Degrees of freedom	Mean Square	Fo
Treatments	$\sum \sum_{\text{Treatments}}$	$a-1$	$\frac{\sum \sum_{\text{Treatment}}}{a-1}$	$\frac{MS_{\text{Treatment}}}{MS_E}$
Blocks	$\sum \sum_{\text{Blocks}}$	$b-1$	$\frac{\sum \sum_{\text{Blocks}}}{b-1}$	ANOVA for a Randomized Complete Block Design
Error	$\sum \sum_E$	$(a-1)(b-1)$	$\frac{\sum \sum_E}{(a-1)(b-1)}$	
Total	$\sum \sum_T$	$ab-1$		

9 اول جون ، تعال تقریب $\alpha=0.01$ قای کل اشی

$$a=4, b=5$$

$$\sum \sum_T = \sum_{i=1}^4 \sum_{j=1}^5 y_{ij}^2 - \frac{y_{..}^2}{ab} = (1.3)^2 + (1.6)^2 + \dots + (1.4)^2 - \frac{(39.2)^2}{20} = 25.69$$

$$\sum \sum_{\text{Treatments}} = \sum_{i=1}^4 \frac{y_{i.}^2}{b} - \frac{y_{..}^2}{ab} = \frac{(5.7)^2 + (8.8)^2 + (6.9)^2 + (17.8)^2}{5} - \frac{(39.2)^2}{20} = 18.09$$

$$\sum \sum_{\text{Blocks}} = \sum_{j=1}^5 \frac{y_{.j}^2}{a} - \frac{y_{..}^2}{ab} = \frac{(9.2)^2 + (10.1)^2 + (3.5)^2 + (8.8)^2 + (7.6)^2}{4} - \frac{(39.2)^2}{20} = 6.69$$

$$\sum \sum_E = \sum \sum_T - \sum \sum_{\text{Blocks}} - \sum \sum_{\text{Treatments}} = 25.69 - 6.69 - 18.09 = 0.96$$

	Sum of Squares	DOF	Mean Square	Fo	P-value
Treatments	18.04	3	6.01	75.13	$4.79 \cdot 10^{-9}$
Blocks	6.69	4	1.67	$F_{0.01, 3, 12} = 5.95$	
Error	0.96	12	0.08		
Total	25.69	19		$F_0 > F_{\alpha, a-1, (a-1)(b-1)}$	

$$\text{if } F_0 > F_{\alpha, a-1, (a-1)(b-1)}$$

Reject, significant

So we reject H_0

Ch:14 Design of Experiments with Several Factors :

* By a Factorial Experiment we mean that in each complete trial or replicate of the experiment all possible combinations of the levels of the factor are investigated.

(K)

		Factor B	
Factor A		B _{low}	B _{high}
level	A _{low}	10	20
	A _{high}	30	40

(J)

		Factor B	
Factor A		B _{low}	B _{high}
level	A _{low}	10	20
	A _{high}	30	0

" A Factorial Experiment with Two Factors " " A Factor Experiment with Interaction "

- (K)
- Main effect of A : $(\frac{30+40}{2}) - (\frac{10+20}{2}) = 20$
 - Main effect of B : $(\frac{20+40}{2}) - (\frac{10+30}{2}) = 10$
- (J)
- Main effect of A : $(\frac{30+0}{2}) - (\frac{10+20}{2}) = 0$
 - Main effect of B : $(\frac{20+0}{2}) - (\frac{10+30}{2}) = 30$

(K)

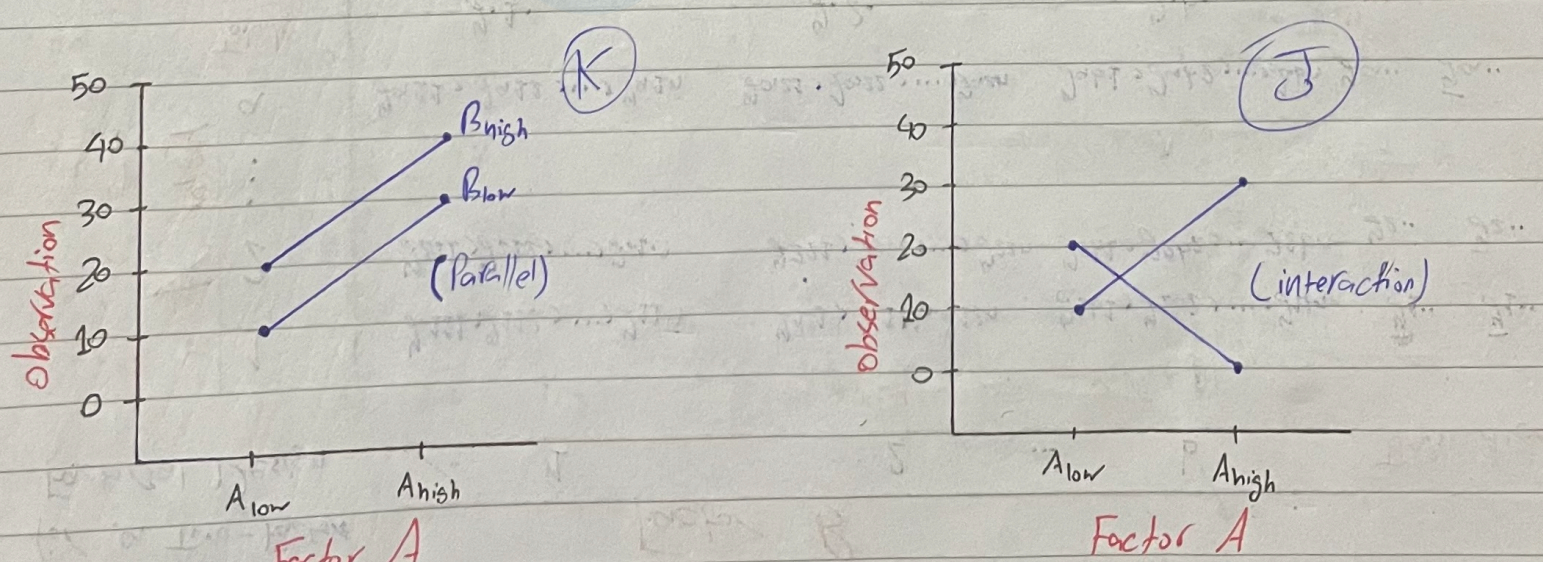
$$AB = (\frac{20+30}{2}) - (\frac{40+10}{2}) = 0$$

Interaction لا يوجد
B و A مع
(B و A مع A و B مع Level اي تأثير لا)

(J)

$$AB = (\frac{20+30}{2}) - (\frac{10+0}{2}) = 20$$

Interaction يوجد
Main effect و Interaction
Main effect



Data Arrangement for a Two-Factor Factorial Design

Factor B

Factor (3 level) A

1

2

...

b

Totals Averages

y_{ij} (مجموع البيانات)

Replicates (التكرار)

$y_{1..}$ $\bar{y}_{1..}$

$y_{2..}$ $\bar{y}_{2..}$

$y_{a..}$ $\bar{y}_{a..}$

$y_{.1.}$ $\bar{y}_{.1.}$

$y_{.2.}$ $\bar{y}_{.2.}$

$y_{.b.}$ $\bar{y}_{.b.}$

$y_{...}$ $\bar{y}_{...}$

$y_{...}$ $\bar{y}_{...}$

$y_{...}$ $\bar{y}_{...}$

$y_{...}$ $\bar{y}_{...}$

$y_{...}$ $\bar{y}_{...}$

$y_{...}$ $\bar{y}_{...}$

$y_{...}$ $\bar{y}_{...}$

$y_{...}$ $\bar{y}_{...}$

$y_{...}$ $\bar{y}_{...}$

$y_{...}$ $\bar{y}_{...}$

$y_{...}$ $\bar{y}_{...}$

$y_{...}$ $\bar{y}_{...}$

$y_{...}$ $\bar{y}_{...}$

$y_{...}$ $\bar{y}_{...}$

$y_{...}$ $\bar{y}_{...}$

$y_{...}$ $\bar{y}_{...}$

$y_{...}$ $\bar{y}_{...}$

$y_{...}$ $\bar{y}_{...}$

$y_{...}$ $\bar{y}_{...}$

$y_{...}$ $\bar{y}_{...}$

$y_{...}$ $\bar{y}_{...}$

$y_{...}$ $\bar{y}_{...}$

$y_{...}$ $\bar{y}_{...}$

$y_{...}$ $\bar{y}_{...}$

$y_{...}$ $\bar{y}_{...}$

$y_{...}$ $\bar{y}_{...}$

$y_{...}$ $\bar{y}_{...}$

$y_{...}$ $\bar{y}_{...}$

$y_{...}$ $\bar{y}_{...}$

y_{abn} : data الموجودة في a level من فكتور A و b level من فكتور B و n Replicates
 $y_{a..}$: مجموع a level من فكتور A و b level من فكتور B و n Replicates
 $y_{.b.}$: مجموع b level من فكتور B و a level من فكتور A و n Replicates
 $\bar{y}_{a..}$: افرج من a level من فكتور A و b level من فكتور B و n Replicates
 $\bar{y}_{.b.}$: افرج من b level من فكتور B و a level من فكتور A و n Replicates
 $y_{...}$: مجموع الجميع
 $\bar{y}_{...}$: افرج الجميع

$$y_{i..} = \sum_{j=1}^b \sum_{k=1}^n y_{ijk}$$

$$\bar{y}_{i..} = \frac{y_{i..}}{bn}$$

$i = 1, 2, \dots, a$

$$y_{.j.} = \sum_{i=1}^a \sum_{k=1}^n y_{ijk}$$

$$\bar{y}_{.j.} = \frac{y_{.j.}}{an}$$

$j = 1, 2, \dots, b$

$$y_{ij.} = \sum_{k=1}^n y_{ijk}$$

$$\bar{y}_{ij.} = \frac{y_{ij.}}{n}$$

$$y_{...} = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n y_{ijk}$$

$$\bar{y}_{...} = \frac{y_{...}}{abn}$$

* The observations may be described by the linear statistical model:

$$Y_{ijk} = \mu + \gamma_i + \beta_j + (\gamma\beta)_{ij} + \epsilon_{ijk} \quad \begin{cases} i = 1, 2, \dots, a \\ j = 1, 2, \dots, b \\ k = 1, 2, \dots, n \end{cases}$$

where ϵ_{ijk} are normal random variables.

* The Hypotheses that will be tested:

- $H_0: \gamma_1 = \gamma_2 = \dots = \gamma_a = 0$
 $H_1: \text{at least one } \gamma_i \neq 0$] No main effect of factor A
- $H_0: \beta_1 = \beta_2 = \dots = \beta_b = 0$
 $H_1: \text{at least one } \beta_j \neq 0$] No main effect of factor B
- $H_0: (\gamma\beta)_{11} = (\gamma\beta)_{12} = \dots = (\gamma\beta)_{ab} = 0$
 $H_1: \text{at least } (\gamma\beta)_{ij} \neq 0$] No interaction

* Computing formulas for the sum of squares:

$$SS_T = SS_A + SS_B + SS_{AB} + SS_E$$

$$SS_T = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n Y_{ijk}^2 - \frac{Y_{...}^2}{abn}$$

* The sum of squares identity:

$$SS_A = \frac{\sum_{i=1}^a Y_{i..}^2}{bn} - \frac{Y_{...}^2}{abn}$$

$$\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (Y_{ijk} - \bar{y}_{i..})^2 = bn \sum_{i=1}^a (\bar{y}_{i..} - \bar{y}_{...})^2$$

$$SS_B = \frac{\sum_{j=1}^b Y_{.j.}^2}{an} - \frac{Y_{...}^2}{abn}$$

$$+ an \sum_{j=1}^b (\bar{y}_{.j.} - \bar{y}_{...})^2$$

$$SS_{AB} = \frac{1}{n} \sum_{i=1}^a \sum_{j=1}^b Y_{ij.}^2 - \frac{Y_{...}^2}{abn} - SS_A - SS_B$$

$$+ n \sum_{i=1}^a \sum_{j=1}^b (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2$$

$$SS_E = SS_T - SS_A - SS_B - SS_{AB}$$

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F ₀
A treatments	$\sum S_A$	$a-1$	$MS_A = \frac{\sum S_A}{a-1}$	To test $H_0: \tau_i = 0$ $F_0 = \frac{MS_A}{MSE}$
B treatments	$\sum S_B$	$b-1$	$MS_B = \frac{\sum S_B}{b-1}$	To test $H_0: \beta_j = 0$ $F_0 = \frac{MS_B}{MSE}$
Interaction	$\sum S_{AB}$	$(a-1)(b-1)$	$MS_{AB} = \frac{\sum S_{AB}}{(a-1)(b-1)}$	To test $H_0: (\tau\beta)_{ij} = 0$ $F_0 = \frac{MS_{AB}}{MSE}$
Error	$\sum S_E$	$ab(n-1)$	$MSE = \frac{\sum S_E}{ab(n-1)}$	
Total	$\sum S_T$	$abn-1$	The ANOVA Table $H_0: \tau_0$	

Reject if $F_0 > F_{\alpha, ab(n-1)}$

Example ~
FACTOR B

$a=3$
 $b=2$ $\alpha=0.05$
 $n=3$

Primer Type	Dipping	Spraying	Totals	Averages
FACTOR A	1 12.8 4.0, 4.5, 4.3 5.4, 4.9, 5.6 15.9 28.7 4.78			
	2 15.9 5.6, 4.9, 5.4 5.8, 6.1, 6.3 18.2 34.1 5.68			
	3 11.5 3.8, 3.7, 4.0 5.5, 5.0, 5.0 15.5 27.0 4.50			
Totals		40.2 + 49.6 = 89.8		4.98
Averages		4.47 5.52		4.98

$$\sum S_T = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n y_{ijk}^2 - \frac{y_{...}^2}{abn} = (4.0)^2 + (4.5)^2 + \dots + (5.0)^2 - \frac{(89.8)^2}{18} = 10.72$$

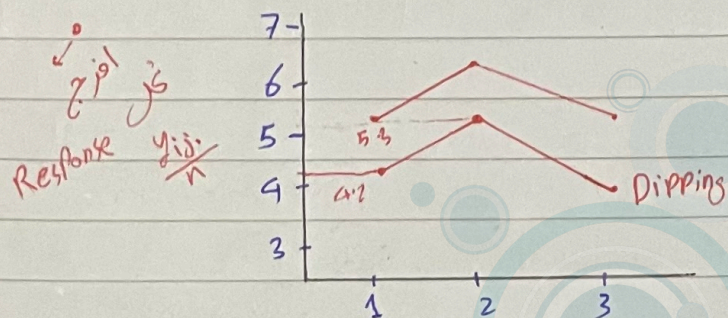
$$\sum S_E = 10.72 - 4.58 - 4.91 - 0.24 = 0.99$$

$$\sum S_A = \sum_{i=1}^a y_{i..}^2 - \frac{y_{...}^2}{bn} = \frac{28.7^2 + 34.1^2 + 27.0^2}{2 \times 3} - \frac{89.8^2}{2 \times 3 \times 3} = 4.58$$

$$\sum S_B = \sum_{j=1}^b y_{.j.}^2 - \frac{y_{...}^2}{an} = \frac{40.2^2 + 49.6^2}{3 \times 3} - \frac{89.8^2}{2 \times 3 \times 3} = 4.91$$

$$\sum S_{AB} = \sum_{i=1}^a \sum_{j=1}^b \frac{y_{ij.}^2}{n} - \frac{y_{...}^2}{abn} = \sum S_A - \sum S_B = 12.8^2 + 15.9^2 + 11.5^2 + 15.9^2 + 18.2^2 + 15.5^2 - \frac{89.8^2}{2 \times 3 \times 3}$$

Source of Variation	Sum of Squares	DOF	MS	F ₀	P-Value
Primer types (A)	4.58	2	$2.29 = 2.29 \div \frac{1}{2} = 2.63$		2.7×10^{-5}
Application methods (B)	4.91	1	4.91	61.38	5.0×10^{-7}
Interaction (AB)	0.24	2	0.12	1.5	0.2621
Error	0.99	12	0.08		
Total	10.72	17			



$$f_{\alpha, \alpha-1, ab(n-1)} = f_{0.05, 2, 12} = 3.8$$

$$f_{\alpha, b-1, ab(n-1)}$$

$$= f_{0.05, 1, 12} = 4.75$$

* Expected Mean Square:

$$E(MS_A) = \sigma^2 + \frac{bn \sum_{i=1}^a \gamma_i^2}{a-1}$$

$$E(MS_B) = \sigma^2 + \frac{an \sum_{j=1}^b \beta_j^2}{b-1}$$

$$E(MS_{AB}) = \sigma^2 + \frac{n \sum_{i=1}^a \sum_{j=1}^b (\gamma\beta)_{ij}^2}{(a-1)(b-1)}$$

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

* Residual : $e_{ijk} = y_{ijk} - \bar{y}_{ij}$

* Model for a Three-Factor Factorial Experiment:

$$Y_{ijkl} = \mu + \gamma_i + \beta_j + \tau_k + (\gamma\beta)_{ij} + (\gamma\tau)_{ik} + (\beta\tau)_{jk} + (\gamma\beta\tau)_{ijk} + \epsilon_{ijkl}$$

$$i = 1, 2, \dots, a$$

$$j = 1, 2, \dots, b$$

$$k = 1, 2, \dots, c$$

$$l = 1, 2, \dots, n$$

* Analysis of Variance Table for the Three-Factor Fixed Effect Model:

Source of Variation	Sum of Squares	DOF	Mean Square	F ₀
A	$\sum \sum \gamma_i^2$	$a-1$	MS_A	Mean Square MS _E
B	$\sum \sum \beta_j^2$	$b-1$	MS_B	
C	$\sum \sum \tau_k^2$	$c-1$	MS_C	
AB	$\sum \sum (\gamma\beta)_{ij}^2$	$(a-1)(b-1)$	MS_{AB}	
AC	$\sum \sum (\gamma\tau)_{ik}^2$	$(a-1)(c-1)$	MS_{AC}	
BC	$\sum \sum (\beta\tau)_{jk}^2$	$(b-1)(c-1)$	MS_{BC}	
ABC	$\sum \sum \sum (\gamma\beta\tau)_{ijk}^2$	$(a-1)(b-1)(c-1)$	MS_{ABC}	
Error	$\sum \sum \sum \epsilon_{ijkl}^2$	$abc(n-1)$	MS_E	
Total	$\sum \sum \sum y_{ijkl}^2$	$abcn-1$		

Simple Linear Regression and Correlation:

* Expected Value of y for a given value of x :

$$E(y|x) = \mu_{y|x} = \beta_0 + \beta_1 x$$

slope

11.1

$$V(y|x) = \sigma^2$$

intercept

* The Simple linear regression model is given by:

True Value of y $\leftarrow$ $y = \beta_0 + \beta_1 x + \epsilon$ $\rightarrow$ Random error term

* The n observations in the sample can be expressed as:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i = 1, 2, \dots, n$$

* The fitted or estimated regression line is therefore:

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

Note that each pair of observations satisfies the relationship:

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

Residual $\epsilon_i = y_i - \hat{y}_i$

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

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

sum of squares for x

sum of squares for products

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

The least squares estimates

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

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n}$$

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

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

$* E(\epsilon) = 0$ $* \hat{\sigma}^2 = \frac{SS_E}{n-2}$ $* SS_E = SS_T - \hat{\beta}_1 SS_{xy}$
 $* V(\epsilon) = \sigma^2$ $* E(SS_E) = (n-2) \hat{\sigma}^2$

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

: $\sum$ $\sum$ $\sum$ $\sum$ $\sum$

x	y	x^2	y^2	xy

(1) $\sum xy$ $\sum x^2$ $\sum y^2$ $\sum xy$ $\sum x^2$ $\sum y^2$
 (2) $\sum xy$ $\sum x^2$ $\sum y^2$ $\sum xy$ $\sum x^2$ $\sum y^2$
 (3) $\sum xy$ $\sum x^2$ $\sum y^2$ $\sum xy$ $\sum x^2$ $\sum y^2$

*** Properties of the Least Squares Estimators:**
 • Slope properties: • Intercept properties:

$E(\hat{\beta}_1) = \beta_1$ $SE = \text{standard error}$ $E(\hat{\beta}_0) = \beta_0$
 $V(\hat{\beta}_1) = \frac{\hat{\sigma}^2}{SS_{xx}}$ $V(\hat{\beta}_0) = \left[\frac{1}{n} + \frac{\bar{x}^2}{SS_{xx}} \right] \cdot \hat{\sigma}^2$

*** Hypothesis Tests in Simple Linear Regression:**
 (Significance of regression) fail to reject

"Slope"

$H_0: \beta_1 = \beta_{1,0}$
 $H_1: \beta_1 \neq \beta_{1,0}$

Test statistic: $T_0 = \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\frac{\hat{\sigma}^2}{SS_{xx}}}}$

Critical Value: $t_{\alpha/2, n-2}$

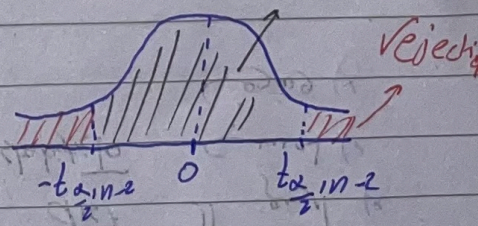
Page-33

Joint T-dist

"Intercept"

$H_0: \beta_0 = \beta_{0,0}$
 $H_1: \beta_0 \neq \beta_{0,0}$

Test statistic: $T_0 = \frac{\hat{\beta}_0 - \beta_{0,0}}{\sqrt{\hat{\sigma}^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{SS_{xx}} \right]}}$



t-dist $\frac{t_{\alpha/2, n-2}}{2} < P\text{-Value} < \frac{t_{\alpha/2, n-2}}{2}$

11.4

خطوات الحل
 (x) حساب كل شيء
 Test all
 (x) حساب
 (x) dist
 (x) test statistic
 (x) conclusion

* Analysis of Variance Approach to test Significance of Regression

The analysis of variance Identity is: $\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2$

The analysis of Variance Symbolically: $SS_T = SS_R + SS_E$

$SS_R = \hat{\beta}_1 \cdot S_{xy}$ $SS_E = SS_T - \hat{\beta}_1 \cdot S_{xy}$ $SS_T = \sum y_i^2 - n\bar{y}^2$

$MSR = SS_R$ $MSE = \frac{SS_E}{n-2} = \sigma^2$

Source of Variation	Sum of Squares	Degree of Freedom	Mean Square
Regression	SS_R	1	MSR
Error	SS_E	$n-2$	MSE
Total	SS_T	$n-1$	$F_0 = \frac{MSR}{MSE}$

critical value

if $F_0 > F_{\alpha, 1, n-2}$: Reject

on the slope β_1 * Confidence Intervals * on the intercept β_0

$$\hat{\beta}_1 - t_{\alpha/2, n-2} \cdot \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}} < \beta_1 < \hat{\beta}_1 + t_{\alpha/2, n-2} \cdot \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}} \quad \hat{\beta}_0 - t_{\alpha/2, n-2} \cdot \sqrt{\hat{\sigma}^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right]} < \beta_0 < \hat{\beta}_0 + t_{\alpha/2, n-2} \cdot \sqrt{\hat{\sigma}^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right]}$$

on the Mean Response $\hat{\mu}_{y|x_0}$

11.5

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

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

11.7

* The Coefficient of Determination: $R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$

10.6

Inference on Two Population Proportions:

$$H_0: P_1 = P_2$$

$$H_1: P_1 \neq P_2$$

$$\hat{P}_1 = \frac{x_1}{n_1}$$

$$\hat{P}_2 = \frac{x_2}{n_2}$$

$$V(\hat{P}) = \frac{P(1-P)}{n}, \sigma_{\hat{P}} = \sqrt{\frac{P(1-P)}{n}}$$

$$\text{Test Statistic: } Z_0 = \frac{\hat{P}_1 - \hat{P}_2 - (P_1 - P_2)}{\sqrt{\frac{P_1(1-P_1)}{n_1} + \frac{P_2(1-P_2)}{n_2}}}$$

$$E(\hat{P}_1 - \hat{P}_2) = E(\hat{P}_1) - E(\hat{P}_2) = P_1 - P_2$$

$$\text{Critical Values: } Z_{\frac{\alpha}{2}}, Z_{\frac{\alpha}{2}}$$

$$V(\hat{P}_1 - \hat{P}_2) = V(\hat{P}_1) + V(\hat{P}_2)$$

$$\bar{P} = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2}$$

$$\hat{P}_1 = \hat{P}_2 = P \text{ (Pooled estimator)} \quad \hat{P} = \frac{x_1 + x_2}{n_1 + n_2} \rightarrow Z_0 = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{\hat{P}(1-\hat{P}) \left[\frac{1}{n_1} + \frac{1}{n_2} \right]}}$$

Alternative Hypothesis

P-Value

Rejection Criterion

$$H_1: P_1 \neq P_2$$

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

$$Z_0 > Z_{\frac{\alpha}{2}}, Z_0 < -Z_{\frac{\alpha}{2}}$$

$$H_1: P_1 > P_2$$

$$P = 1 - \phi(Z_0)$$

$$Z_0 > Z_{\alpha}$$

$$H_1: P_1 < P_2$$

$$P = \phi(Z_0)$$

$$Z_0 < -Z_{\alpha}$$

$$\beta = \phi \left[\frac{Z_{\frac{\alpha}{2}} \sqrt{\bar{P}(1-\bar{P}) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} - (P_1 - P_2)}{\sigma_{\hat{P}_1 - \hat{P}_2}} \right] - \phi \left[\frac{-Z_{\frac{\alpha}{2}} \sqrt{\bar{P}(1-\bar{P}) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} - (P_1 - P_2)}{\sigma_{\hat{P}_1 - \hat{P}_2}} \right]$$

Two-sided

$\sigma_{\hat{P}_1 - \hat{P}_2}$

page-35

$\sigma_{\hat{P}_1 - \hat{P}_2}$

$$H_0: p_1 > p_2$$

$$\beta = \phi \left[z \frac{\sqrt{p(1-p) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} - (p_1 - p_2)}{\sigma_{p_1 - p_2}} \right]$$

$$H_2: p_1 < p_2$$

$$\beta = 1 - \phi \left[\frac{-z \sqrt{p(1-p) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} - (p_1 - p_2)}{\sigma_{p_1 - p_2}} \right]$$

$$N = \left[z_{\frac{\alpha}{2}} \cdot \frac{(p_1 + p_2)(q_1 + q_2) + z_{\beta} \sqrt{p_1 q_1 + p_2 q_2}}{2} \right]^2$$

$$q_1 = 1 - p_1 \quad q_2 = 1 - p_2$$

CI:

$$\hat{p}_1 - \hat{p}_2 - E \leq p_1 - p_2 \leq \hat{p}_1 - \hat{p}_2 + E, \quad E = z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$