10.1: Inference about difference between two population means, σ_1, σ_2 known.

→ **Assumptions**

1. sample 1 Random
2. sample 2 Random.
3. sample 1 and sample 2 independent.
4. a. population 1 Normal.

   b. population 2 Normal.

OR Both sample are large enough.

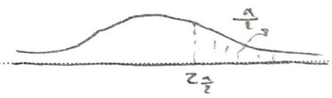
★ large enough samples: $n_1 \geq 30$, $n_2 \geq 30$ } use it if pop 1, 2 not Normal

→ point estimator for $\mu_1 - \mu_2 = \bar{X}_1 - \bar{X}_2$.

→ confidence interval (CI) / interval estimate for $\mu_1 - \mu_2 = (\bar{X}_1 - \bar{X}_2) \mp E$

margin of error $(E) = Z_{\frac{\alpha}{2}} \underbrace{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ → standard error.

→ $(1-\alpha)$ CI for $\mu_1 - \mu_2 = (\bar{X}_1 - \bar{X}_2) \mp Z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$.



→ **Hypothesis test about $\mu_1 - \mu_2$:**

① lower tail test	② upper tail test	③ Two tail test.
$H_0: \mu_1 - \mu_2 \geq D_0$	$H_0: \mu_1 - \mu_2 \leq D_0$	$H_0: \mu_1 - \mu_2 = D_0$
$H_1: \mu_1 - \mu_2 < D_0$	$H_1: \mu_1 - \mu_2 > D_0$	$H_1: \mu_1 - \mu_2 \neq D_0$.

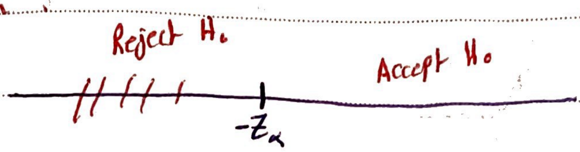
→ **Test statistic:**

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - D_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

By Z-table.

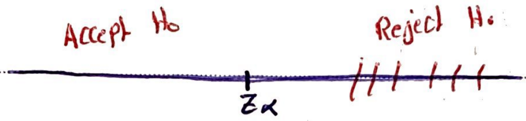
→ Critical value Approach :

① Lower tail test :



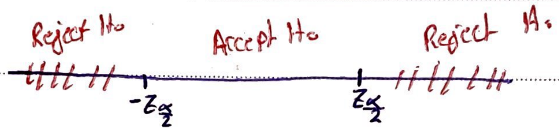
Reject H_0 if $Z \leq -Z_\alpha$.

② upper tail test :



Reject H_0 if $Z \geq Z_\alpha$.

③ Two tail test :

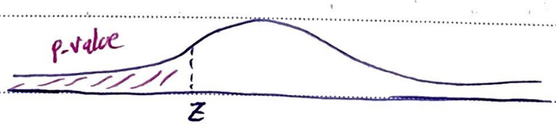


Reject H_0 if $Z \geq Z_{\frac{\alpha}{2}}$ or $Z \leq -Z_{\frac{\alpha}{2}}$.

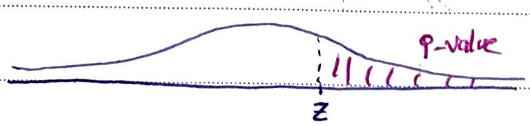
→ P- value Approach :

Reject H_0 if $P\text{-value} \leq \alpha$

① lower tail test :



② upper tail test :



③ Two tail test :

10.2: Inferences about $\mu_1 - \mu_2$, σ_1 and σ_2 unknown.

→ **Assumptions:**

1. Sample 1 Random sample from pop. 1
2. sample 2 Random sample from pop. 2.
3. sample 1 and sample 2 are independent.
4. pop. 1 and pop. 2 have Normal distribution OR sample 1 and sample 2 are large enough.

$\star$ large enough: $n_1 + n_2 \geq 20$ s.t. $n_1 \approx n_2$.

→ point estimater for $\mu_1 - \mu_2 = \bar{x}_1 - \bar{x}_2$.

→ standard error $= \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$.

→ margen of error (E) $= t_{\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$.

→ $df = \left\lfloor \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\left(\frac{1}{n_1-1}\right)\left(\frac{s_1^2}{n_1}\right)^2 + \left(\frac{1}{n_2-1}\right)\left(\frac{s_2^2}{n_2}\right)^2} \right\rfloor$ → $\lfloor 9.4 \rfloor = 9$.

→ $1-\alpha$ CI $= (\bar{x}_1 - \bar{x}_2) \mp E$.

→ Hypotheses testing for $\mu_1 - \mu_2$:

① lower tail test

$H_0: \mu_1 - \mu_2 \geq D_0$

$H_1: \mu_1 - \mu_2 < D_0$

② upper tail test

$H_0: \mu_1 - \mu_2 \leq D_0$

$H_1: \mu_1 - \mu_2 > D_0$

③ two tail test

$H_0: \mu_1 - \mu_2 = D_0$

$H_1: \mu_1 - \mu_2 \neq D_0$.

→ Test statistic:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

→ critical value Approach:

① lower tail test:

We reject H_0 if $t_{test} \leq -t_\alpha$.

② upper tail test:

We reject H_0 if $t_{test} \geq t_\alpha$.

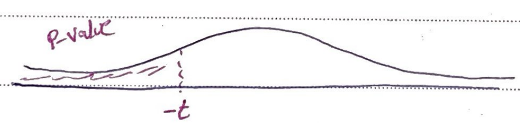
③ two tail test:

We reject H_0 if $t_{test} \geq t_{\frac{\alpha}{2}}$ or $t_{test} \leq -t_{\frac{\alpha}{2}}$.

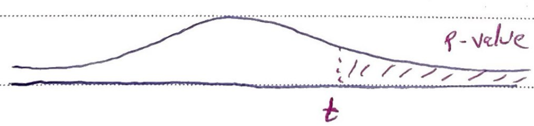
→ p-value Approach:

Reject H_0 if P-value $\leq \alpha$.

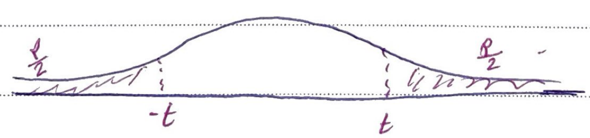
① lower tail test:



② upper tail test:



③ two tail test:



→ If we additionally have $\sigma_1 = \sigma_2$:

✯ Test statistic:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

✯ $df = n_1 + n_2 - 2$

✯ Pooled sample variance:

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

6.3 : Inference about the difference between two population means , matched samples .

Notations :

- n : sample size / # of element
- μ_1 : pop.1 mean
- μ_2 : pop.2 mean
- $\mu_d = \mu_1 - \mu_2$
- $d_i = x_i^{I} - x_i^{II}$
- $\bar{d} = \frac{\sum d_i}{n}$
- $s_d = \sqrt{\frac{\sum (d_i - \bar{d})^2}{n-1}}$

→ $H_0 : \mu_d = \mu_{d,0}$

$H_1 : \mu_d \neq \mu_{d,0}$ } two tailed test

→ Test statistic :

$$t = \frac{\bar{d} - \mu_{d_0}}{\frac{s_d}{\sqrt{n}}} \quad , \; df = n-1 .$$

→ Reject H_0 if p-value $\leq \alpha$

p-value = area in both tails .

Reject H_0 if $|t| \geq t_{\frac{\alpha}{2}}$, $df = n-1$.

→ $(1-\alpha)$ CI for $\mu_d = \bar{d} \mp \underbrace{t_{\frac{\alpha}{2}} \frac{s_d}{\sqrt{n}}}_{\text{margin of error}}$

standard error .

10.4: Inferences about the difference between two population proportions.

→ **Assumptions:**

1. sample 1 and sample 2 Random.
2. samples 1 and 2 are independent.
3. samples 1 and 2 large enough.

★ large enough:

pop: $n_1\pi_1 \geq 5$, $n_1(1-\pi_1) \geq 5$

$n_2\pi_2 \geq 5$, $n_2(1-\pi_2) \geq 5$.

sample: $n_1 p_1 \geq 5$, $n_1(1-p_1) \geq 5$.

$n_2 p_2 \geq 5$, $n_2(1-p_2) \geq 5$.

★ Notations:

π_1: proportion in pop. 1

π_2: proportion in pop. 2

p_1: proportion in sample 1

p_2: proportion in sample 2.

n_1: sample 1 size.

n_2: sample 2 size

→ point estimator for $\pi_1 - \pi_2 = p_1 - p_2$.

→ $(1-\alpha)$ CI for $\pi_1 - \pi_2 = (p_1 - p_2) \mp E$.

→ margin of error $(E) = Z_{\frac{\alpha}{2}} \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$

→ standard error of $p_1 - p_2$ $(\sigma_{p_1-p_2}) = \sqrt{\frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}}$.

→ Hypotheses test about $\pi_1 - \pi_2$:

① lower tail test:

$H_0: \pi_1 - \pi_2 \geq 0$

$H_1: \pi_1 - \pi_2 < 0$

② upper tail test:

$H_0: \pi_1 - \pi_2 \leq 0$

$H_1: \pi_1 - \pi_2 > 0$.

③ two tail test:

$H_0: \pi_1 - \pi_2 = 0$

$H_1: \pi_1 - \pi_2 \neq 0$.

→ Remark: under the H_0 when H_0 is true an equality we get $\pi_1 = \pi_2 = \pi$.

→ standard error ($\pi_1 = \pi_2 = \pi$) : $\sigma_{P_1 - P_2} = \sqrt{\pi(1-\pi)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$.

→ Pooled estimate of π when $\pi_1 = \pi_2 = \pi$:

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

→ Test statistic for Hypotheses test about $\pi_1 - \pi_2$:

$$Z = \frac{(P_1 - P_2)}{\sqrt{P(1-P)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}.$$

→ Reject H_0 :

→ If P-value $\leq \alpha$.

→ ① Lower tail test: $Z \leq -Z_\alpha$

② Upper tail test: $Z \geq Z_\alpha$

③ Two tail test: $|Z| \geq Z_{\frac{\alpha}{2}}$.

11.1 Inferences about population variance :

σ^2 : Population variance .
σ : Pop. st. dev
S^2 : sample variance
S : sample st. dev

→ sampling distribution of $\frac{(n-1)S^2}{\sigma^2}$ has chi-squared distribution with $n-1$ degrees freedom.

→ Assuming :

1. The sample is Random .
2. The sample come from a Normal population .

→ $(1-\alpha)$ CI for $\sigma^2 = \left(\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}} , \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}} \right)$, where $df = n-1$, $S^2 = \frac{\Sigma(x_i - \bar{x})^2}{n-1}$.

→ Testing Hypothesis :

① Lower tail test
$H_0 : \sigma^2 \geq \sigma_0^2$
$H_1 : \sigma^2 < \sigma_0^2$

② upper tail test
$H_0 : \sigma^2 \leq \sigma_0^2$
$H_1 : \sigma^2 > \sigma_0^2$

③ Two tail test .
$H_0 : \sigma^2 = \sigma_0^2$
$H_1 : \sigma^2 \neq \sigma_0^2$.

→ Test statistic :

$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$, $df = n-1$, σ_0^2 : hypothesis value .

→ Reject H_0 if :

→ P-value $\leq \alpha$.

→ ① lower tail test
$\chi^2 < \chi^2_{1-\alpha}$

② upper tail test
$\chi^2 > \chi^2_{\alpha}$

③ Two tail test .
$\chi^2 > \chi^2_{\frac{\alpha}{2}}$ or $\chi^2 < \chi^2_{1-\frac{\alpha}{2}}$.

11.2 Inferences about two population variances.

→ sampling distribution of $\frac{S_1^2}{S_2^2}$ when $\sigma_1^2 = \sigma_2^2$ has F distribution with $n_1 - 1$ df for the numerator (البسط) and $n_2 - 1$ df for the denominator (المقام).

Notations:

N_1 : size of pop. 1

N_2 : size of pop. 2

σ_1^2 : variance of pop. 1

σ_2^2 : variance of pop. 2

n_1 : size of sample 1

n_2 : size of sample 2

S_1^2 : variance of sample 1

S_2^2 : variance of sample 2

→ Assuming:

1. sample 1 and sample 2 random
2. sample 1 and sample 2 independent.
3. sample 1 and sample 2 are from Normal population.
4. $\sigma_1^2 = \sigma_2^2$.

→ Testing Hypothesis:

lower ما في

① upper tail test:

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

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

② Two tail test:

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

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

→ Test statistic:

$$F = \frac{S_1^2}{S_2^2} \quad \text{with} \quad df_1 = n_1 - 1, \quad df_2 = n_2 - 1.$$

→ Reject H_0 if:

→ p-value $\leq \alpha$

→ ① upper tail test

$F \geq F_\alpha$

② Two tail test.

$F \geq F_{\frac{\alpha}{2}}$

→ Note:

population 1 is the population with higher sample variance.

12.1 Goodness of fit and Independence :

→ $H_0 : \pi_1 = \pi_{10} , \pi_2 = \pi_{20} , \ldots , \pi_K = \pi_{K0}$

H_1 : The population proportion are not $\pi_1 = \pi_{10}, \ldots, \pi_K = \pi_{K0}$

→ π_i : Population proportion of category i.

π_{i0} : Hypothesied value of the population of category i. $i = 1, \ldots, K$.

→ n : sample size

f_i : observed frequencies.

e_i : expeted frequencies.

→ $e_i = n\,\pi_{i0}$

→ $\sum_{i=1}^{K} f_i = \sum_{i=1}^{K} e_i = n$

→ Test statistic :

$$\chi^2 = \sum_{i=1}^{K} \frac{(f_i - e_i)^2}{e_i}$$

→ Reject H_0 if p-value $\leq \alpha$

Reject H_0 if $\chi^2 \geq \chi^2_\alpha$

df : $K-1$ chi square table.

$e_i \geq 5 \quad \forall i$.

→ Remark

To use Goodness of fit test for multinomial populations we assume :

1. The sample taken is Random.
2. expected frequencies for all categories ; should satisfy the following $e_i \geq 5 \ \forall i$.

12.2: Test of Independence.

→ Null and alternative hypotheses:

H_0: The Row variable and the column variable are independent.

H_1: The Row variable and the column variable are not independent.

→ We need to take a Random sample:

f_{ij} : observed frequency

e_{ij} : expected freq.

n : # of Row.

m : # of columns.

$$e_{ij} \doteq \frac{(\text{Row } i \text{ table})(\text{column } j \text{ table})}{\text{sample size}}.$$

→ Note: $\sum_j \sum_i f_{ij} = \sum_j \sum_i e_{ij} =$ sample size.

→ Test statistic:

$$\chi^2 = \sum_j \sum_i \frac{(f_{ij} - e_{ij})^2}{e_{ij}} \quad \text{with} \quad df = (n-1)(m-1).$$

Assuming: $e_{ij} \geq 5 \quad \forall i \, \forall j$.

→ Rejection Rule:

- Reject H_0 if p-value $\leq \alpha$.
- Reject H_0 if $\chi^2 \geq \chi^2_\alpha$.

→ Note: we use chi-square table.

12.3: Goodness of fit test : Poisson and Normal distribution.

Poisson distribution

→ H_0: The population has a poisson dist.

H_1: The population doesn't have a poisson dist.

→ Take a Random sample of size n

f_i : observed frequencies $\sum f_i = n$.

e_i : expected frequencies $\sum e_i = n$.

$$\boxed{e_i = \frac{\mu^{x_i} e^{-\mu}}{x_i!} \cdot n} \qquad \boxed{\mu = \frac{\sum_{i=1}^{k} x_i f_i}{\sum_{i=1}^{k} f_i}}$$

→ Test statistic :

$$\chi^2 = \sum_{i=1}^{k} \frac{(f_i - e_i)^2}{e_i} \quad \text{with } df = k-2$$

(Assuming $e_i \geq 5 \ \forall i$)

→ Rejection Rule :

- Reject H_0 if $p\text{-value} \leq \alpha$.
- Reject H_0 if $\chi^2 \geq \chi^2_\alpha$.

Normal distribution

→ H_0: The population has a Normal distribution.

H_1: The population doesn't have a Normal distribution.

→ Take a Random sample of size n

f_i: observed frequencies.

e_i: expected frequencies.

→ Notation:

- K: # of categories.

  $K = \frac{n}{5}$

- $e_i = 5 \quad \forall i$

→ Test statistic:

$$\chi^2 = \sum_{i=1}^{K} \frac{(f_i - e_i)^2}{e_i} \qquad \text{with } df = K-3.$$

→ Rejection Rule:

- Reject H_0 if p-value $\leq \alpha$.
- Reject H_0 if $\chi^2 \geq \chi^2_\alpha$.

13.2 : Analysis of variance : testing for the equality of K population means.

→ Testing for the equality of K pop. mean sample mean for treatment j :

$$\bar{x}_j = \frac{\sum_{i=1}^{n_j} x_{ij}}{n_j}$$

→ sample variance for treatment j :

$$s_j^2 = \frac{\sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2}{n_j - 1}$$

→ over sample mean :

$$\bar{\bar{x}} = \frac{\sum_{j=1}^{k}\sum_{i=1}^{n_j} x_{ij}}{n_T}$$

$$\bar{\bar{x}} = \frac{\sum_{j=1}^{k} \bar{x}_j}{k} \quad \text{if } n \text{ are equal.}$$

★ Notations :

- x_{ij} : value of observation i for treatment j.
- n_j : number of observation for treatments j.
- $\bar{x}_j$: sample mean for treatment j.
- s_j^2 : sample variance for treatment j.
- s_j : sample standard deviation for treatment j.

→ Between treatments estimate of population variance :

• Mean square due to treatments.

$$MSTR = \frac{SSTR}{k-1} \quad \text{where} \quad SSTR = \sum_{j=1}^{k} n_j (\bar{x}_j - \bar{\bar{x}})^2 .$$

→ Within treatments estimate of population variance :

• mean square due to error.

$$MSE = \frac{SSE}{n_T - k} \quad \text{where} \quad SSE = \sum_{j=1}^{k} (n_j - 1) s_j^2 .$$

→ Test for the equality of K pop. means :

$H_0 : \mu_1 = \mu_2 = \dots = \mu_k$

H_1 : Not all population means are equal.

→ Test statistic :

$$F = \frac{MSTR}{MSE}$$

→ Rejection Rule :

p-value approach : Reject H_0 if p-value $\leq \alpha$

critical value approach : Reject H_0 if $F \geq F_\alpha$.

→ ANOVA table :

Sorce of variance	df	SS	MS	F
Treatments	$k-1$	SSTR	MSTR	$\frac{MSTR}{MSE}$
Error	n_T-k	SSE	MSE	
Total	n_T-1	SST		

13.3: multiple comparison procedures:

→ FLSD Procedure:

$H_0^{ij}: \mu_i = \mu_j \qquad i \neq j$

$H_1^{ij}: \mu_i \neq \mu_j \qquad i,j \in \{1, \dots, k\}.$

→ Test statistic:

$\bar{X}_i - \bar{X}_j$.

→ Rejection Rule:

Reject H_0 if $|\bar{x}_i - \bar{x}_j| > LSD^{ij}$ where $LSD^{ij} = t_{\frac{\alpha_{cw}}{2}} \sqrt{MSE\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$.

s.t $\alpha_{cw} = \dfrac{\alpha_{EW}}{\binom{K}{2}}$ and $df = n_T - K$.

→ Bonferroni Adjustment:

α_{EW}: experement wise Type I significance

α_{cw}: comparision wise " "

- $\alpha_{EW} = \binom{K}{2} \alpha_{cw}$

- $\alpha_{cw} = \dfrac{\alpha_{EW}}{\binom{K}{2}}$

→ conffedence interval.

$(1-\alpha)$ CI for $\mu_i - \mu_j = (\bar{x}_i - \bar{x}_j) \mp LSD^{ij}$ where $LSD^{ij} = t_{\frac{\alpha}{2}} \sqrt{MSE\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$.

13.6 : Randomized Block design.

↳ contains Treatments and Blocks.

→ ANOVA table : Block Randomized Design :

Source of variance	df	SS	MS	F
Treatments	$k-1$	SSTR	$MSTR = \frac{SSTR}{k-1}$	$F = \frac{MSTR}{MSE}$
Blocks	$b-1$	SSBL	$MSBL = \frac{SSBL}{(b-1)}$	
Error	$(k-1)(b-1)$	SSE	$MSE = \frac{SSE}{(k-1)(b-1)}$	
Total	$n_T - 1$	SST	–	–

* b = # of Blocks.
* BL = Blocks
* $n_T = kb$
* $SST = SSTR + SSBL + SSE$.

→ Hypothesis : $H_0 : \mu_1 = \mu_2 = \dots = \mu_k$

H_1 : Not all μ_j are equal.

→ Rejection Rule :

• critical value : reject H_0 if $F \geq F_\alpha$ with $df_1 = k-1$, $df_2 = (k-1)(b-1)$.

• p-value : reject H_0 if p-value $\leq \alpha$.

→ Notation and def :

- $\bar{X}_{i\circ}$: sample mean of Block i , $i=1,\dots,b$.
- $\bar{X}_{\circ j}$: sample mean of treatments j , $j=1,\dots,K$.
- $\bar{\bar{X}}$: over all mean of all observation .
- $SST = \sum_{j=1}^{K}\sum_{i=1}^{b}(x_{ij}-\bar{\bar{x}})^2$.
- $SSTR = b\sum_{j=1}^{K}(x_{\circ j}-\bar{\bar{x}})^2$.
- $SSBL = K\sum_{i=1}^{b}(x_{i\circ}-\bar{\bar{x}})^2$.
- $SSE = SST - SSTR - SSBL$.

13.7 : Factorial experiments .

→ Notation .

- $a =$ # of levels of factor A .
- $b =$ # of levels of factor B .
- $r =$ # of replications .
- $n_T =$ total number of observaions taken in experement , $n_T = abr$.

→ Hypothesis :

H_0^A : means of factors A are equal

H_1^A : Means of factors A are not equal .

H_0^B : Means of factor B are equal .

H_1^B : Means of factor B are not equal .

H_0^{AB} : Factor A and Factor B have no interaction .

H_1^{AB} : Factor A and Factor B have an interaction .

→ Defenetions

- $SST = SSA + SSB + SSAB + SSE = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r} (x_{ijk} - \bar{\bar{x}})^2$.

- $SSA = br \sum_{i=1}^{a} (\bar{x}_{i\cdot} - \bar{\bar{x}})^2$

- $SSB = ar \sum_{j=1}^{b} (\bar{x}_{\cdot j} - \bar{\bar{x}})^2$

- $SSAB = r \sum_{j=1}^{b} \sum_{i=1}^{a} (\bar{x}_{ij} - \bar{x}_{i\cdot} - \bar{x}_{\cdot j} + \bar{\bar{x}})^2$.

- $SSE = SST - SSA - SSB - SSAB$.

→ ANOVA table : Two Factor Factorial experement.

Source of Variance	df	SS	MS	F	F_α	p-value
Factor A	$a-1$	SSA	$MSA = \frac{SSA}{a-1}$	$F^A = \frac{MSA}{MSE}$	F_α with $a-1$, $ab(r-1)$	
Factor B	$b-1$	SSB	$MSB = \frac{SSB}{b-1}$	$F^B = \frac{MSB}{MSE}$	F_α with $b-1$, $ab(r-1)$	
Interaction AB	$(a-1)(b-1)$	SSAB	$MSAB = \frac{SSAB}{(a-1)(b-1)}$	$F^{AB} = \frac{MSAB}{MSE}$	F_α with $(a-1)(b-1)$, $ab(r-1)$	
Error	$ab(r-1)$	SSE	$MSE = \frac{SSE}{ab(r-1)}$	–	–	–
Total	$n_T - 1$	SST	–	–	–	–

$r \geq 2$.

14.1 : simple linear Regression Model.

→ The simple linear Regression Model :

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

Note :

input
$\underline{X}$ is a Variable.

output
$\underline{y}$ is a Random variable

ε is a Random variable.

→ The simple linear Regression equation.

$$E(y) = \beta_0 + \beta_1 X .$$

→ Estimated simple linear Regression equation.

$$\hat{y} = b_0 + b_1 X .$$

14.2 → Least square method :

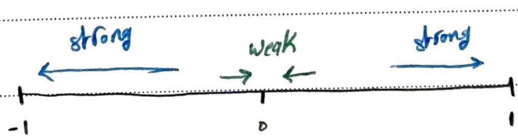
$$b_1 = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2}$$: least square estimate for β_1

$$b_0 = \bar{y} - b_1 \bar{x} .$$: least square estimate for β_0.

14.3 : coeffecient of determination.

→ correlation coeffecient :

- $r_{xy} = \dfrac{S_{xy}}{S_x S_y}$
- $-1 \leq r_{xy} \leq 1$



→ proposition :

- $SST = SSR + SSE$
- $SST = (n-1)\, s_y^2$
- $SSR = b_1^2 (n-1) S_x^2$
- $SSE = SST - SSR$.

→ Coeffeacient of determination :

- $r^2 = \frac{SSR}{SST}$.

- $0 \leq r^2 \leq 1$

- $r^2 = (r_{xy})^2$ → Coeffeacient of determination = (Correlation coeffecient)^2 .

- $r_{xy} = (\text{sign } b_1) \sqrt{r^2}$.

14.5: Testing for signifacance.

→ Model : $Y = \beta_0 + \beta_1 X + \varepsilon$.

→ $H_0: \beta_1 = 0$ means Not significance variable and Model.

$H_1: \beta_1 \neq 0$ means the Model and Variable are significance .

→ Assuming : $E(\varepsilon) = 0$, $var(\varepsilon) = \sigma^2$, ε independent , ε Normal .

طريقة ١

→ test statistic :

- t-test = $t = \dfrac{b_1}{S_{b_1}}$ with $df = n-2$ } Two tail test .

Where $S_{b_1} = \sqrt{\dfrac{MSE}{(n-1)\, S_x^2}}$.

بستخدم قانون

$S_{b_1} = \dfrac{S}{\sqrt{(n-1)S_x^2}}$

→ Rejection Role :

Reject H_0 if $|t| \geq t_{\frac{\alpha}{2}}$.

→ Mean square Error (estimate of σ^2)

$S^2 = MSE = \dfrac{SSE}{n-2}$.

→ standard error of the estimate .

$S = \sqrt{MSE} = \sqrt{\dfrac{SSE}{n-2}}$.

→ Sampling distribution of b_1 :

- $E(b_1) = \beta_1$
- $\sigma_{b_1} = \dfrac{\sigma}{\sqrt{\Sigma(x_i - \bar{x})^2}} = \dfrac{\sigma}{\sqrt{(n-1)\, S_x^2}}$
- Distribution b_1 is Normal .

→ Estimated standared distribution of b_1 :

$$S_{b_1} = \frac{S}{\sqrt{(n-1) S_x^2}}$$

→ $(1-\alpha)$ CI :

$$= b_1 \mp t_{\frac{\alpha}{2}} S_{b_1}$$

2ed → F-test and ANOVA table .

Source of variation	df	SS	MS	F	
Regression	1	SSR	MSR	$\frac{MSR}{MSE}$.	upper
Error	n-2	SSE	MSE		
Total	n-1	$\underbrace{SST}_{14.2}$	-		

→ MSR = SSR

$$MSE = \frac{SSE}{n-2}$$

→ Rejection Rule :

- Reject H_0 if $F \geq F_\alpha$ with $df_1 = 1$ and $df_2 = n-2$.

- Reject H_0 if p-value $\leq \alpha$.

14.6: using the estimated regression equation for estimation and predection

→ Point estimation : $\hat{y} = b_0 + b_1 X$.

→ Interval Estimation :

$\hat{y}(p) = xp$

1. confidence interval for the mean value of y .

$$(1-\alpha)\,CI \text{ for } \underset{E(y_p)}{} = \hat{y}_p \mp S\, t_{\frac{\alpha}{2}} \sqrt{\frac{1}{n} + \frac{(x_p - \bar{x})^2}{\Sigma(x_i - \bar{x})^2}}$$

where $\hat{y}_p = b_0 + b_1 x_p$ and $df = n - 2$.

2. predection interval for y :

$$(1-\alpha)\,PI \text{ for } y_p = \hat{y}_p \mp S\, t_{\frac{\alpha}{2}} \sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{\Sigma(x_i - \bar{x})^2}}$$

where $\hat{y}_p = b_0 + b_1 x_p$ and $df = n - 2$.

$$S = \sqrt{MSE}$$

14.7: computer solution:

Excel & [illegible]

14.8:

15.1 : Multiple Regression Model.

→ Model (Multiple linear Regression Model): $y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon$.

→ Multiple linear Regression Equation : $E(Y) = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \dots + \hat{\beta}_p X_p$

→ Estimated multiple linear Regression : $\hat{y} = b_0 + b_1 X_1 + \dots + b_p X_p$.

15.2 : least square Method :

$\sum_{i=1}^{n} (y_i - \hat{y}_i)^2$. } By Excel.

15.3 : Multiple coeffecient of determination :

→ $SST = SSR + SSE$.

→ Multiple coeffecient of determinations : $R^2 = \frac{SSR}{SST}$.

→ Adjusted Multiple coeffecient of determinations : $adj\,R^2 = 1 - (1 - R^2)\left(\frac{n-1}{n-p-1}\right)$.

15.5 : Testing for significance .

→ Model : $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \varepsilon$.

→ $H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$

H_1 : Not all β_j are zero .

→ ANOVA Table :

Source of variation	df	SS	MS	F
Regression	p	SSR	MSR	$\frac{MSR}{MSE}$
Error	$n-p-1$	SSE	MSE	
Total	$n-1$	SST	–	

• $MSR = \frac{SSR}{p}$ • $MSE = \frac{MSE}{n-p-1}$ • $F = \frac{MSR}{MSE}$ with $df_1 = p$ and $df_2 = n-p-1$

• Reject H_0 if $F \geq F_\alpha$ or p-value $\leq \alpha$

• F-test only one time .

→ $H_0 : \beta_j = 0$

$H_1 : \beta_j \neq 0$

→ Test statistic : $t = \frac{b_j}{s_{b_j}}$

→ Reject H_0 if $|t| \geq t_{\frac{\alpha}{2}}$ or p-value $\leq \alpha$ with $df = n-p-1$.

• t-test p-times .

→ $S = \sqrt{S^2} = \sqrt{MSE}$: standared error of the estimate .

→ $\sigma_{b_j} = \sqrt{Var(b_j)}$: standard deviation of b_j .

→ S_{b_j} = estimated standard deviation of b_j .

→ Molticolinearity :

input Variable $x_1, x_2, \dots x_p$

Some times some x_i is dependent on the other x_js , this case is Known as multicolinearity .

→ Variance inflation factor (ViF)

$$ViF = \frac{1}{1 - R_j^2}$$

If $ViF(x_i) \geq 10$ then x_i should be elimeneted .

- R_j^2 : Multiple coeffecient of determination for x_i as a function of the other x_j , s .
- function means multiple regression .

Done ..

الباقي من Notes الدكتور