8. A study examining differences in life satisfaction between young, middle and older women was conducted. Each woman who participated in the study completed a life satisfaction questionnaire. A high score on the test indicates a higher level of life satisfaction. Test scores are recorded below:

$\bar{y}_{1.} = \frac{23}{5}$

$\bar{y}_{2.} = \frac{49}{6}$

$\bar{y}_{3.} = \frac{52}{6}$

Young	Middle	Older
4	10	7
2	5	7
3	7	9
7	10	8
7	10	9
	7	12
$y_{1.} = 23$	$y_{2.} = 49$	$y_{3.} = 52$

$y_{..} = \sum_{i=1}^{a} \sum_{j=1}^{n_i} y_{ij} = 124$

$\sum_{i=1}^{3} y_{i.} = y_{..} = 23+49+52 = 124$

with $\sum_{i=1}^{3} \sum_{j=1}^{n_i} y_{ij}^2 = 1018$.

(a) Describe the model that you consider more convenient for this situation, indicating the assumptions associated with the chosen model.

(b) Obtain the analysis of variance table.

(c) Test, at a 5% significance level, if the woman age has an effect in life satisfaction. State the hypotheses, test statistic, decision rule and conclusions.

(d) Find a 95% confidence interval estimate for the difference between the mean value of life satisfaction for middle aged women and older women. Is it possible to conclude, at a 5% significance level, that the mean value of life satisfaction is the same for those two groups of women?

(e) Test, at a 5% significance level, if the mean value of life satisfaction of young aged women is less or equal to 5. State the hypotheses, test statistic, decision rule and conclusions. What is the p-value of the test? (Exam 13/01/2017)

a) $Y_{ij} = \mu_i + \varepsilon_{ij} = \mu + \tau_i + \varepsilon_{ij}$ Assumptions:

Anova one-way (Factor - Age) $\varepsilon_{ij} \sim N(0, \sigma^2)$ i.i.d.

$i = 1, 2, 3$ $a = 3$

$j = 1, \ldots, n_i$ $N = \sum_{i=1}^{3} n_i = 5+6+6 = 17$

$n_1 = 5$; $n_2 = 6$; $n_3 = 6$

$E(Y_{ij}) = \mu_i$; $E(Y_{ij}) = \mu_1$ se $i = 1$

b) Anova table

$a = 3$

Source of variations	SS	d.f	$MS = \frac{SS}{d.f}$
treatments	52.163	$(a-1) = 2$	26.082
Errors	61.367	$(N-a) = 14$	$\boxed{4.383}$
total	113,53	$(N-1) = 16$	

$MSE = \hat{\sigma}^2$

SST = SSTR + SSE

$$SST = \sum_{i=1}^{3} \sum_{j=1}^{n_i} y_{ij}^2 - \frac{y_{..}^2}{N} = 1018 - \frac{124^2}{17} \simeq 113.53$$

$$SSTR = \sum_{i=1}^{3} \frac{y_{i.}^2}{n_i} - \frac{y_{..}^2}{N} = \left(\frac{23^2}{5} + \frac{49^2}{6} + \frac{52^2}{6}\right) - \frac{124^2}{17}$$

$$\simeq 52.163$$

$SSE = SST - SSTR$

c) $H_0: \mu_1 = \mu_2 = \mu_3 \iff H_1: \exists (i,j): \mu_i \neq \mu_j$

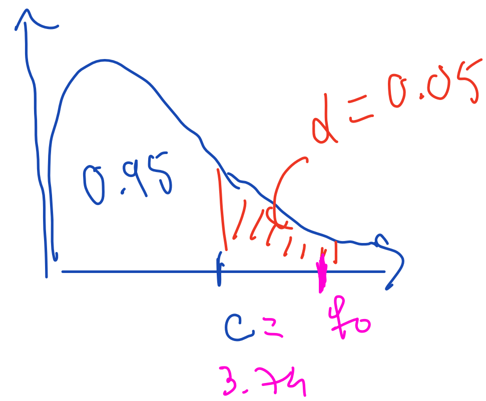
$\tau_1 = \tau_2 = \tau_3 = 0$ $\quad i,j = 1,..3$, $i \neq j$

$\exists i\ \tau_i \neq 0$

under H_0

$$F_0 = \frac{MSTR}{MSE} \underset{H_0}{\sim} F(2,14)$$

rej H_0 if $F_0 > c$



$\alpha = 0.05 = P(\text{rej } H_0 \mid H_0 \text{ is true})$

$$c = F^{-1}_{F_{(2,14)}}(0.95) = 3.74$$

$C.R = [3.74 ; +\infty[$

observed value of F_0:

$$f_0 = \frac{26.082}{4.383} = 5{,}950$$

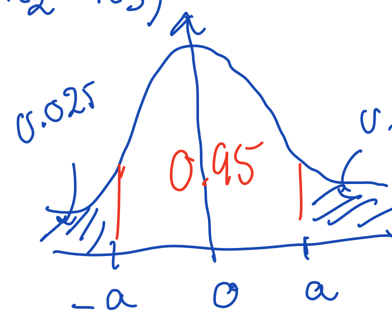
Decision: As $f_0 \in C.R$, for $\alpha = 0.05$,
we rej H_0 => Life satisfaction is
Affected by age

d) $CI_{95\%}(\mu_2-\mu_3) = ?$

Pivotal Quantity : $T = \dfrac{(\bar{Y}_{2.} - \bar{Y}_{3.}) - (\mu_2 - \mu_3)}{\sqrt{MSE\left(\frac{1}{n_2} + \frac{1}{n_3}\right)}} \sim t_{(14)}$

$N-a$

choosing the symmetrical C.I



$$P(-a \le T \le a) = 0.95$$

$$a = F^{-1}_{t(14)}(0.975) = 2.145$$

$$P\left(-a \le \frac{(\bar{Y}_{2.} - \bar{Y}_{3.}) - (\mu_2 - \mu_3)}{\sqrt{MSE\left(\frac{1}{n_2} + \frac{1}{n_3}\right)}} \le a\right) = 0.95 \iff$$

$$P\left(-a\sqrt{MSE\left(\tfrac{1}{n_2}+\tfrac{1}{n_3}\right)} \le (\bar{Y}_{2.} - \bar{Y}_{3.}) - (\mu_2 - \mu_2) \le a\sqrt{\quad}\right) = 0.95$$

$$P\left(-(\bar{Y}_{2.} - \bar{Y}_{3.}) - a\sqrt{\quad} \le -(\mu_2 - \mu_3) \le -(\bar{Y}_{2.} - \bar{Y}_{3.}) + a\sqrt{\quad}\right) = 0.95$$

$$\iff P\left((\bar{Y}_{2.} - \bar{Y}_{3.}) - 2.145\sqrt{MSE\left(\tfrac{1}{n_2}+\tfrac{1}{n_3}\right)} \le (\mu_2 - \mu_3) \le\right.$$

$$\left. \le (\bar{Y}_{2.} - \bar{Y}_{3.}) + 2.145\sqrt{MSE\left(\tfrac{1}{n_2}+\tfrac{1}{n_3}\right)}\right) = 0.95$$

Concretization:

$$CI_{95\%}(\mu_2-\mu_3) = \left[(\bar{y}_{2.}-\bar{y}_{3.}) \pm 2.145\sqrt{MSE\left(\frac{1}{6}+\frac{1}{6}\right)}\right]$$

$$= \left[(\bar{y}_{2.}-\bar{y}_{3.}) \pm 2.145\sqrt{4.383 \times 2/6}\right] = [-3.093; 2.093]$$

$$\bar{y}_{2.}-\bar{y}_{3.} = \left(\frac{49}{6}-\frac{52}{6}\right) = -0.5$$

$\bar{Y}_{2.}$ R.V. $\bar{Y}_{2.} \sim N\left(\mu_2; \frac{\sigma^2}{n_2}\right)$

$H_0: \mu_2 \stackrel{?}{=} \mu_3$ vs $H_1: \mu_2 \neq \mu_3$ $\quad \alpha = 0.05$

$\mu_2-\mu_3 = 0$ vs $\mu_2-\mu_3 \neq 0$

As $0 \in CI_{95\%}(\mu_2-\mu_3)$ we do not rej H_0

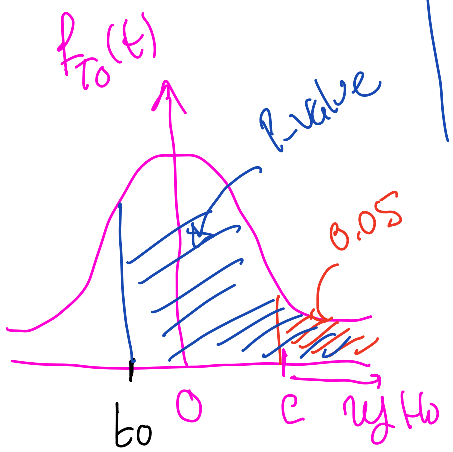
d) $\mu_1 = E[Y_{1j}]$ $\quad i=1$

$H_0: \mu_1 \leq 5$ vs $H_1: \mu_1 > 5$ (one-sided test)

Pivotal Quantity $T = \dfrac{\bar{Y}_{1.}-\mu_1}{\sqrt{\frac{MSE}{5}}} \sim t_{(N-a)}$ $\quad 14$

$\alpha = 0.05$

$\downarrow$



test statistic (Ho is true)

$$T_0 = \frac{\bar{Y}_{1.} - 5}{\sqrt{\frac{MSE}{5}}} \sim t_{(14)}$$

rej Ho if $T_0 > c = F^{-1}_{t_{(14)}}(0.95)$

$= 1.761$

$\downarrow$

table

$CR = [1.761; +\infty[$ $\qquad MSE = \hat{\sigma}^2 =$

$$t_0 = \frac{\overset{4.6}{\bar{y}_{1.}} - 5}{\sqrt{\frac{4.383}{5}}} \simeq -0.427 \qquad \bar{y}_{1.} = \frac{23}{5} = 4.6$$

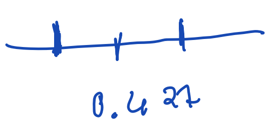
Decision: As $t_0 \notin CR$, for $\alpha = 0.05$, not rej Ho.

P-value $= P(T_0 \geq t_0) = P(T_0 \geq -0.427) =$

$1 - F_{T_0}(-0.427) = F_{T_0}(\underline{0.427}) =$ $\begin{cases} \underline{0.6} < \text{P-value} < \underline{0.7} \\ \Rightarrow \text{rej Ho } \forall \alpha \geq 0.7 \\ \bar{\text{n}} \text{ rej Ho } \forall \alpha \leq 0.6 \end{cases}$

$\downarrow$

table



$\forall \alpha \in]0.6; 0.7[$ we ~~don~~ do not know a

6. We analyzed the prices (in tens of euros) of three equivalent models of televisions of three well-known brands. We obtained the following results:

Brand	Sample size	$\bar{y}_{i.}$	s_i
A	18	50.5	2.1
B	25	53.8	1.9
C	20	64.3	2.7

$$\sum_{i=1}^{3} (n_i-1) s_i^2 + n_i \bar{y}_i^2 = \sum_i \sum_j y_{ij}^2$$

(a) Show that the observations suggest the existence of differences among the actual average price of each brand.

(b) Compare the brands in pairs and give your conclusions.

$y_{..} = 18 \times 50.5 + 25 \times 53.8 + 20 \times 64.3$

a) $y_{ij} = \mu_i + \varepsilon_{ij}$

y_{ij} - price (in tens of euros) of the i-th brand from the j-th model

$i = 1, 2, 3$

$N = \sum_{i=1}^{a} n_i = 63$

$H_0: \mu_1 = \mu_2 = \mu_3$ vs $H_1: \mu_i \neq \mu_j$ for some (i,j)

ANOVA table

Source of variation	SS	d f	MS	f_0
treatments				
Error				
total				

$$SSTR = \sum_i^a n_i(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2 = \sum_i^a \underbrace{\frac{y_{i\cdot}^2}{n_i}} - \boxed{\frac{y_{\cdot\cdot}^2}{N}} \qquad a-1 \qquad MSTR = \frac{SSTR}{a-1}$$

$$SSE = \sum_i^a \sum_j^{n_i} (y_{ij} - \bar{y}_{i\cdot})^2 \qquad \sum_i n_i \bar{y}_i^2 \qquad N-a \qquad MSE = \frac{SSE}{N-a}$$

$$SST = \sum_i^a \sum_j^{n_i} (y_{ij} - \bar{y}_{\cdot\cdot})^2 = \sum_i^a \sum_j^{n_i} y_{ij}^2 - \frac{y_{\cdot\cdot}^2}{N} \quad N-1$$

$$SSTR = \sum_i^a n_i(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2 = \sum_i^a \frac{y_{i\cdot}^2}{n_i} - \frac{y_{\cdot\cdot}^2}{N} \qquad a-1 \qquad MSTR = \frac{SSTR}{a-1}$$

$$SSE = \sum_i^a \sum_j^{n_i} (y_{ij} - \bar{y}_{i\cdot})^2 \qquad N-a \qquad MSE = \frac{SSE}{N-a}$$

$$SST = \sum_i^a \sum_j^{n_i} (y_{ij} - \bar{y}_{\cdot\cdot})^2 = \sum_i^a \sum_j^{n_i} y_{ij}^2 - \frac{y_{\cdot\cdot}^2}{N} \quad N-1$$

$y_{\cdot\cdot}^2$

$$SST = \sum_i \sum_j (y_{ij} - \bar{y}_{\cdot\cdot})^2 = \sum_i \sum_j y_{ij}^2 - \frac{y_{\cdot\cdot}^2}{N}$$

$$s_i^2 = \frac{\sum_i (y_{ij} - \bar{y}_i)^2}{n_i - 1} = \frac{\sum_i y_{ij}^2 - n_i \bar{y}_i^2}{n_i - 1}$$

$$s_i^2 = \frac{\sum_i (y_{ij} - \bar{y}_{i\cdot})^2}{n_i - 1} = \frac{\sum_i y_{ij}^2 - n_i \bar{y}_{i\cdot}^2}{n_i - 1}$$

A

$$\sum_i y_{ij}^2 = \boxed{(n_i - 1)\, s_i^2 + n_i \bar{y}_{i\cdot}^2}$$

s_1

s_2

s_3

$$s_1^2 = \frac{\sum_{j=1}^{n_1} (y_{1j} - \bar{y}_{1\cdot})^2}{n_1 - 1} = \frac{\sum_j y_{1j}^2 - n_1 \bar{y}_{1\cdot}^2}{n_1 - 1}$$

s_2^2

s_3^2

$(2.1)^2$

$$\sum_i \sum_j y_{ij}^2 = 17 \times s_1^2 + 18 \times 50.5^2 + 24\,(1.9)^2 + 25 \times (53.8)^2$$

$$+\, 19 \times (2.7)^2 + 20 \times (64.3)^2 = 201255.4$$

$$SST = 201255.4 - \frac{(18 \times 50.5 + 25 \times 53.8 + 20 \times 64.3)^2}{63}$$

$$= 2341.134$$

$$SSTR = 18 \times 50.3^2 + 25\,(53.8)^2 + 20\,(64.3)^2 -$$

$$- \frac{(18 \times 50.5 + 25 \times 53.8 + 20 \times 64.3)^2}{63}$$

$$= 2041.014$$

$$SSE = SST - SSTR = 2341.134 - 2041.014$$

$$= 300.12$$

Anova table

Source	SS	df	MS	f_0
treat.	2041.014	2 ✓	1020.506	204.019
Error	300.12	$N-a=60$ ✓	5.002	
total	2341.134	$N-1=62$		

$$f_0 = \frac{MSTR}{MSE}$$

Under H_0 $F_0 \sim F_{(2,60)}$ ✓

rej H_0 if $F_0 > c$

$$\text{P-value} = P(F_0 \geq 204.019) =$$

$$= 1 - F_{F_0}(204.016) = 1 - \underline{\underline{0.99916}}$$

$$\simeq 8.32 \times 10^{-4}$$

rej H_0 $\forall \alpha \geq 8.32 \times 10^{-4}$ rej H_0 $\forall \alpha$

b) $\mu_1, \mu_2 ; \mu_3$ $\binom{3}{2}$

$\mu_1 - \mu_2$ ✓

$\mu_1 - \mu_3$ ✓

$\mu_2 - \mu_3$ ✓

$CI_{\underline{\underline{95\%}}}(\mu_1 - \mu_2)$ $CI_{\underline{\underline{95\%}}}(\mu_1 - \mu_3)$

$(1-\alpha)$ $\alpha = 5\%$ $CI_{\underline{\underline{95\%}}}(\mu_2 - \mu_3)$

Bonferroni correction

$1 - \alpha/2 =$ $g = 3$ $1 - \alpha/_{2\times g} = 1 - \frac{0.05}{2\times 3}$

0.975 $\underline{\underline{= 0.99}}$ ✓

Pivotal Quantity:

$$T = \frac{\bar{Y}_{i.} - \bar{Y}_{j.} - \boxed{(\mu_i - \mu_j)}}{\sqrt{MSE\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}} \sim t_{(N-a)}$$

60

$F^{-1}_{t_{(60)}}(0.99) = 2.39$

$$P(-2.39 \leq T \leq 2.39) = 0.95$$

$\vdots$

$$C.I_{95\%}(\mu_1 - \mu_2) = \left[\bar{Y}_{1.} - \bar{Y}_{2.} \pm 2.39 \times \sqrt{MSE\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}\right]$$

$$C.I_{95\%}(\mu_2 - \mu_3) = \left[\bar{Y}_{2.} - \bar{Y}_{3.} \pm 2.39 \sqrt{MSE\left(\frac{1}{n_2} + \frac{1}{n_3}\right)}\right]$$

$$C.I_{95\%}(\mu_1 - \mu_3) = \left[\bar{Y}_{1.} - \bar{Y}_{3.} \pm 2.39 \sqrt{MSE\left(\frac{1}{n_1} + \frac{1}{n_3}\right)}\right]$$

$$CI_{95\%}(\mu_1 - \mu_2) = (-4.952 \, ; -1.648)$$

$$CI_{95\%}(\mu_1 - \mu_3) = (-15.537 \, ; -12.063)$$

$$CI_{95\%}(\mu_2 - \mu_3) = (-12.104 \, ; -8.896)$$

$$\left.\begin{array}{l} \mu_2 > \mu_1 \\ \mu_3 > \mu_1 \\ \mu_3 > \mu_2 \end{array}\right\} \text{Brand A is cheaper!!}$$