ED 10
Correction

Exercice 1 :

Commençons par calculer la moyenne $\overline{x}$ de la réunion des 3 échantillons.

$$\overline{x} = \frac{1}{96}\left(\frac{32\times 81 + 33\times 52 + 31\times 46}{96}\right) = 59.73$$

Dressons le tableau de l'analyse de la variance.

$$SCF = \sum_j n_j(\overline{x}_j - \overline{x})^2 = 32(81-\overline{x})^2 + 33(52-\overline{x})^2 + 31\times(46-\overline{x})^2$$

$$SCR = \sum_j (n_j - 1)S_j^2 = 31\times 6.8^2 + 32\times 5.2^2 + 30\times 6.7^2$$

	Variation	SC	ddl	CM	Fobs
					S_F^2/S_R^2 =
SCF	Factorielle	22293.36	3-1 = 2	11146.48	$\frac{11146.48}{39.20}$ = 284.36
SCR	Résiduelle	3645.6	96-3 = 93	39.20	
SCT	Totale	25938.96	96-1 = 95		

Si l'on admet que les distributions des durées de developpement sont gaussiennes et de même variance, on peut donc appliquer l'analyse de la variance à un facteur et tester :

H_0 : la température n'a pas d'influence sur la durée de developpement du parasites.

H_1 : la température a une influence.

Sous H_0 $F = \dfrac{SCF/_{J-1}}{SCR/_{N-J}} \sim F(J-1, N-J) = F(2,93)$

$F_{théo} = F_{0,95}(2,93) = 3,1$ pour $\alpha = 0,05$, $F_{théo} = F_{0,975}(2,93) = 3,8$ $\alpha = 0,025$

Conclusion : Comme $F_{obs} > F_{théo}$, on rejette H_0 : l'influence de la température est significative.

Exercice 2 :

On dispose de deux échantillons

- Femmes non enceintes : $n_1 = 13$, $\bar{x}_1 = 2.15$, $S_1^2 = 0{,}38$
- Femme enceintes : $n_2 = 11$, $\bar{x}_2 = 4.08$, $S_2^2 = 0{,}46$

Total : $n = n_1 + n_2 = 24$, $\bar{x} = 3.03$

① Première approche : Test de comparaison de deux moyennes
on suppose que les conditions d'application d'un test paramétrique sont réunies (égalité des variances, distributions normales).

$$T = \frac{\bar{X}_1 - \bar{X}_2}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \longrightarrow T(n_1+n_2-2)$$

On estime σ par $\hat{\sigma} = \sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}$

$$= 0{,}416$$

H_0 : Pas d'influence de la grossesse sur la PDE
$(\mu_1 = \mu_2)$

H_1 : Il y a une influence de la grossesse

$$t_{obs} = \frac{2.15 - 4.08}{0{,}416\sqrt{\frac{1}{13}+\frac{1}{11}}} = -7.3$$

$\alpha = 5\%$, $t_{theo} = t_{1-\frac{\alpha}{2}}(n_1+n_2-2) = t_{0{,}975}(22) = 2.074$

Conclusion. Comme $|t_{obs}| > t_{theo}$, on rejette H_0.
Il y a une influence de la grossesse sur la PDE.

A 2ème approche : Analyse de la variance

	Variation	S.C	ddl	CM	Fobs
SCF	Factorielle	22.19	2−1 = 1	22.19	
SCR	Résiduelle	9.16	24−2 = 22	0.416	
Totale					

$\frac{22.19}{0.416} = 53.3$

Ho : la grossesse n'a pas d'influence significative sur la PDE ($\mu_1 = \mu_2$)

H1 = La grosse a de l'influence sur la PDE ($\mu_1 \neq \mu_2$)

$\alpha = 5\%$ $F_{théo} = F_{0,95}(1, 22) = 4.3$

Conclusion :

Comme $F_{obs} > F_{theo}$, on rejette Ho.

2°/ On a 5 échantillons

- à 4 semaines, $n_1 = 12$, $\overline{x_1} = 3.10$, $s_1^2 = 1.411$
- à 5 semaines, $n_2 = 12$, $\overline{x_2} = 5.21$, $s_2^2 = 0.850$
- à 6 semaines, $n_3 = 12$, $\overline{x_3} = 6.93$, $s_3^2 = 2.495$
- à 7 semaines, $n_4 = 12$, $\overline{x_4} = 5.77$, $s_4^2 = 1.742$
- à 8 semaines, $n_5 = 12$, $\overline{x_5} = 7.475$, $s_5^2 = 6.522$

Total $n = 60$, $\overline{x} = 6.10$

$F_{obs} = $

$$F_{obs} = \frac{13.47}{2.60} = 5.17$$

On teste

H_0 : l'âge de la grossesse n'a pas d'influence sur l'activité de PDE (égalité des 5 moyennes)

pour $\alpha = 5\%$

$$F_{théo} = F_{0,95}(4;55) = 2.54$$

Conclusion :

Comme $F_{obs} > F_{théo}$, on rejette H_0 : l'influence de l'âge est significative sur la PDE.

Exercice 3 : Modèle linéaire simple :

le modèle de régression simple s'écrit

$$Y_i = a + bx_i + \varepsilon_i$$

Soient les notations suivantes (voir le cours)

$$\bar{x} = \frac{1}{n} \sum x_i \quad , \quad \bar{y} = \frac{1}{n} \sum y_i$$

$$S_X^2 = \frac{1}{n} \sum (x_i - \bar{x})^2 \quad , \quad S_Y^2 = \frac{1}{n} \sum (y_i - \bar{y})^2$$

$$\operatorname{cov}(X,Y) = S_{XY}^2 = \frac{1}{n} \sum (x_i - \bar{x})(y_i - \bar{y})$$

a) Estimation par les moindres carrés des coefficients a et b.

$$b = \frac{\operatorname{Cov}(X,Y)}{V(X)} = -2.0747$$

$$a = \bar{y} - b\bar{x} = 126.7036$$

L'équation de la droite de régression est donc donnée par

$$\underline{y = 126.7036 - 2.0747x}$$

b) On note S_ε^2 la variance résiduelle

$$S_\varepsilon^2 = \frac{\sum (y_i - a - bx_i)^2}{n-2} = 0,6809$$

c) Intervalle de confiance autour de A et B

Rappelons que

$$S_b^2 = \frac{S_\varepsilon^2}{n S_x^2} \quad \text{et} \quad S_a^2 = \frac{1}{n} S_\varepsilon^2 \left(1 + \frac{(\bar{x})^2}{S_x^2}\right)$$

Distribution de b

$$* \ \frac{1}{S_b}(\hat{b} - b) = \sqrt{\frac{n S_x^2}{S_\varepsilon^2}} (\hat{b} - b) \rightsquigarrow T(n-2)$$

$$* \ \frac{1}{S_a}(\hat{a} - a) = \sqrt{\frac{n}{S_\varepsilon^2 \left(1 + \frac{(\bar{x})^2}{S_x^2}\right)}} (\hat{a} - a) \rightsquigarrow T(n-2)$$

Les intervalles de confiance autour de a et b sont obtenus par, pour un risque α.

$$b \in \left[\hat{b} - t_{1-\frac{\alpha}{2}} \sqrt{\frac{n S_x^2}{S_\varepsilon^2}} \, , \, \hat{b} + t_{1-\frac{\alpha}{2}} \sqrt{\frac{n S_x^2}{S_\varepsilon^2}}\right]$$

$$a \in \left[\hat{a} - t_{1-\frac{\alpha}{2}} \sqrt{\frac{n}{S_\varepsilon^2 \left(1 + \frac{\bar{x}^2}{S_x^2}\right)}} \, , \, \hat{a} + t_{1-\frac{\alpha}{2}} \sqrt{\frac{n}{S_\varepsilon^2 \left(1 + \frac{\bar{x}^2}{S_x^2}\right)}}\right]$$

$\alpha = 5\%$ $\quad t_{0,975} = 2.571$

$$a \in [110.91 \, ; \, 142.50]$$

$$b \in [-2.3596 \, ; \, -1,7898]$$

4) Test du coefficient de regression B

$H_0 \quad B = 0 \qquad H_1 \quad B \neq 0$

La statistique du test est :

$$T = \sqrt{\frac{n S_x^2}{S_\varepsilon^2}} \, \hat{b} \quad , \quad t = -18.72$$

$$t_{0,975}(5) = 2.571$$

Conclusion:

Comme $t \notin [-2.571 , 2.571]$, on n'accepte pas Ho.
Compte tenu de la taille de l'échantillon, nous admettons que le coefficient B est significativement $\neq$ de 0.

5) Estimation de la valeur moyenne de Y pour $X = 55$.

Ceci revient à calculer

$$E(Y|X) = A + BX \quad \text{pour } X = 55.$$

$$\text{pour } X = 55 \quad , Y = 126.7036 - 2.0747 \times 55$$

$$= 12.5951 \simeq 12.60 .$$

6) Intervalle de prévision de Y pour $X = 55$

au risque α, l'intervalle de prévision de la valeur prise par Y est

$$\left[\hat{b}x_0 + \hat{a} - t_{1-\frac{\alpha}{2}}(n-2)\, S_\varepsilon \sqrt{1 + \frac{1}{n} + \frac{(\bar{x} - x_0)^2}{n S_X^2}} \right.$$

$$\left. bx_0 + \hat{a} + t_{1-\frac{\alpha}{2}}(n-2)\, S_\varepsilon \sqrt{1 + \frac{1}{n} + \frac{(\bar{x} - x_0)^2}{n S_X^2}} \right]$$

Application:

L'estimation de Y par un intervalle de prédiction au risque $\alpha = 0,05$ est

$$[10,3256 \, ; \, 14.8646] .$$