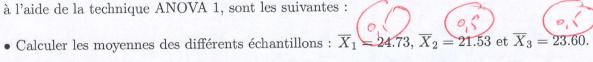
Corrigé type Examen

Solution de l'Exercice 1 Les étapes qu'on doit suivre pour réaliser le test

$$H_0: \ '' \ \mu_1 = \mu_2 = \mu_3 = \mu \ '' \ \text{contre} \ H_1: \ '' \ \exists \ i,j \in \{1,2,3\} \ \text{tel que} \ \mu_i \neq \mu_j \ '',$$

à l'aide de la technique ANOVA 1, sont les suivantes :



• Calculer la moyenne globale de toutes les observations : $\overline{X} = \frac{1}{n}(n_1\overline{X}_1 + n_2\overline{X}_2 + n_3\overline{X}_3) = 23.2889.$



• Compléter le tableau de l'ANOVA à un seul facteur :

	Somme des carrés	Degrés de libertés	Carré moyen	ratio	Ficher
source de variation	SC	ddl	CM	F_{obs}	C
Inter-groupe (31.5911	0,5 2	15.7956	12.02	3.6823
Intra-groupe	19.7067	0, 15	·, 1.3138	0,1	
Total	51.2978	0, 17			

• Décision : on constate que $f_{obs}=12.02~>~f_{\alpha}=3.6823~(pour~un~risque~de~\alpha=5\%),~donc~les$ espaces moyens occupés par les informations sont significativement différents d'une bases de données à une autre. Cela signifié que le facteur bases de données influe sur l'espace mémoire occupé par les informations stockées.



Solution de l'Exercice 2

							Somme
$X \mu g/\mu l$	0	20	40	60	80	100	300
Y	0	0.205	0.331	0.515	0.584	0.671	2.3060
X^2	0	400	1600	3600	6400	10000	22000
Y^2	0	0.0420	0.1096	0.2652	0.3411	0.4502	1.2081
X * Y	0	4.10	13.24	30.90	46.72	67.10	162.06

Afin de modéliser ces données, nous avons proposé le modèle linéaire suivant :

$$Y = a x + b.$$

1. Calcul des estimateurs des paramètres a et b. On a :

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} x_i = \frac{1}{6} 300 = 50.$$

$$\overline{Y} = \frac{1}{n} \sum_{i=1}^{n} x_i = \frac{1}{6} 3002.3060 = 0.3843.$$

$$Cov(x,y) = \frac{1}{n} \sum_{i=n}^{n} x_i y_i - \overline{X} \overline{Y} = \frac{1}{6} (162.06) - (50) (0.3843) = 7.7950$$

$$Var(x) = \frac{1}{n} \sum_{i=1}^{n} x_i^2 - \overline{X}^2 = \frac{1}{6} (22000) - (50)^2 = 1166.6667$$

alors,

$$\hat{a} = \frac{Cov(x,y)}{Var(x)} = \frac{\frac{1}{n} \sum_{i=n}^{n} x_i y_i - \overline{X} \overline{Y}}{\frac{1}{n} \sum_{i=1}^{n} x_i^2 - \overline{X}^2} = 0.0067.$$

$$\hat{b} = \overline{Y} - \hat{a}\overline{X} = 0.0503,$$

de ce fait la droite de régression de l'absorbance (Y) en fonction de la concentration (x) est donnée par:

 $\hat{Y} = 0.0067 \ x + 0.0503.$

2. Quelle absorbance prévoyez-vous à une concentration 40 $\mu g/\mu l$? Que peut-on conclure?

$$\hat{Y} = 0.0067 (40) + 0.0503 = 0.3183.$$

On constate que la valeur de régression est très proche de la vraie valeur (0.331), donc à priori le modèle retenu est adéquate pour la représentation des données du tableau.

3. Calcul du coefficient de corrélation linéaire.

$$r = r(x, y) = \frac{Cov(x, y)}{\sigma_x \sigma_y} = 0.9851,$$

 $r=r(x,y)=\frac{Cov(x,y)}{\sigma_x\sigma_y}=0.9851,$ avec $\sigma_y=\sqrt{var(Y)}=\sqrt{\frac{1}{n}\sum\limits_{i=1}^ny_i^2-\overline{Y}}=\sqrt{0.0536}=0.2316;$ La valeur du coefficient de corrélation est très proche de 1, i.e. X et Y sont fortement linéairement liés donc le modèle est efficace ce qui confirme les résultats de la question 3). les résultats de la question 3).

4. Pour un seuil de risque $\alpha=5\%,$ le modèle proposé est-il pertinent? Pour répondre à cette question on utilise le test de validation du modèle (Fisher). On d'une part

$$f_c = \frac{\sum_{i=1}^{n} (\hat{y} - \overline{Y})^2 / 1}{\sum_{i=1}^{n} (y_i - \hat{y})^2 / (n-2)} = \frac{0.3124 / 1}{0.0095 / (6-2)} = 131.5368,$$

et d'autre par

$$f_{\alpha} = f(1, n-2, 1-\alpha) = f(1, 4, 0.95) = 7.71.$$

On constate que $f_c > f_\alpha$, alors on accepte le modèle proposé, c'est-à-dire le modèle est valide (pertinent)