GM – Probabilités et Statistique – Corrigés 10

À noter : les **raisonnements/justifications** des étapes de résolution sont également importantes (pas seulement le résultat final).

En salle

Exercice 1 (a) Variable de réponse : literacy rate (taux d'alphabétisation).
Variables explicatrices : number of daily newspaper copies, number of radios, number of TV sets (all per 1000 people in the population of the country).

(b) $\widehat{y} = 0.51486 + 0.00054\,x_1 - 0.00035\,x_2 + 0.00199\,x_3$, où $\widehat{y}$ = predicted literacy rate (taux d'alphabétisation),

x_1 = the number of daily newspaper copies in the country (per 1000 people),
x_2 = the number of radios in the country (per 1000 people),
x_3 = the number of TV sets in the country (per 1000 people).

(c) Pour pays avec le même nombre de radios et le même nombre de télés par 1000 personnes dans la population, For countries with the same number of radios and same number of TV sets per 1000 people in the population, literacy rate est prédit d'être 0.00054 plus haut pour chaque additionnel journal par 1000 personnes dans la population.

(d) D'abord, à vérifier qu'on n'extrapole pas : toute les valeurs des variables explicatrices se situent dans la plage des données collectées, donc on y va. Deuxièmement, assurez-vous d'avoir les bonnes valeurs pour les bons x.

Alors $\hat{y} = 0.51486 + 0.00054 \times 200 - 0.00035\, times\, 800 + 0.00199\, times\, 250 \approx 0.84$
$\Rightarrow$ on prédit 84% taux d'alphabétisation parmi les résidents du pays .

(e) $\hat{b}_3 \pm \boxed{t_{n-p-1,0.975}}\, SE(\hat{b}_3)$; $n = 10, p = 3$; $\Rightarrow$ $\boxed{0.00199 \pm 2.447 \times 0.00155}$
$(\Rightarrow\ (-0.00180, 0.00578))$

Exercice 2 (a) $\Omega : y = \beta_0 + \beta_1\,x_1 + \beta_2\,x_2 + \epsilon$

(b) $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$

(c) $Var(\hat{\beta}_i) = \sigma^2\ [(\mathbf{X}'\mathbf{X})^{-1}]_{i+1,\,i+1}$, donc $Var(\hat{\boldsymbol{\beta}}) = \underline{\underline{\hat{\sigma}}}^2 \times (z_{11},\ z_{22},\ \underline{\underline{z_{33}}})$
$= 11.06 \times (3.437,\ 0.0014,\ 0.0021) = (38.01, 0.015, 0.023)$.

(d)

source	df	SS	MS	F	p-value
regression	2	184.2	184.2/2 = **92.1**	92.1/11.06 = **8.33**	0.0020
error	22	243.3	243.3/22 = **11.06**		
total	2+22 = **24**	184.2+243.3 = **427.5**			

(e) $R^2 = 1 - SSE/SST = 1 - 243.3/427.5$ $\boxed{\approx 0.43}$

$$R^2_{aj} = 1 - (1 - R^2)\frac{n-1}{n-p-1} = 1 - (1 - 0.43)\left(\frac{24}{22}\right) \quad \boxed{\approx 0.38}$$

(f) $H : \beta_1 = \beta_2 = 0$; A : au moins un $\beta_i \neq 0$
Selon le tableau d'anova, $F_{obs} = 8.33 \sim F_{2,22}$ (en supposant H) ; $p_{obs} = 0.002 < 0.05$, donc on REJETTE H

(g) $\boxed{\hat{\beta}_2 \pm \underline{\underline{\hat{\sigma} \cdot \sqrt{z_{33}}}} \cdot \boxed{t_{n-p-1,0.975}}}$

$\Rightarrow -0.555 \pm \sqrt{11.06 \times 0.0021}\ \boxed{t_{22,0.975}} \Rightarrow \boxed{-0.555 \pm 0.15 \times 2.074}$

(ou -0.555 ± 0.31 ou $[-0.865, -0.245]$)

(h) $SCE(\omega) = 336.5$, $SCE(\Omega) = 243.3$

$$F_{obs} = \frac{[SCE(\omega) - SCE(\Omega)]/(p-q)}{SCE(\Omega)/(n-p-1)} = \frac{(336.5 - 243.3)/(2-1)}{243.3/22} = \frac{93.2}{11.06} =$$

Degrés de liberté numérateur = 1, donc $F_{1,22} = t_{22}^2$, $t_{22,0.975} = 2.074 < |t_{obs} = 2.9|$, donc on REJETTE H

Source	df	SC	CM	F	p
(h) Groups	(a) # groups − 1 $\Longrightarrow$ $4-1=\mathbf{3}$	(d) MS group × df group 116×3 (= **348**)	116	(g) $\frac{\text{MS group}}{\text{MS error}}$ $\Longrightarrow$ $116/10 = \mathbf{11.6}$	1.8×10^{-5}
Error	(b) df total - df group $\Longrightarrow$ $39-3=\mathbf{36}$	360	(f) $\frac{\text{SS error}}{\text{df error}}$ $\Longrightarrow$ $360/36 = \mathbf{10}$		
Total	(c) # obs − 1 $\Longrightarrow$ $4 \times 10 - 1 = \mathbf{39}$	(e) SS group + SS error $\Longrightarrow$ $116 \times 3 + 360$ (=**708**)			

Exercice 3 (a)-(h) (voir le tableau ci-dessus)

(i) 1. Les paramètres sont $\mu_1, \mu_2, \mu_3, \mu_4$, les moyennes des 4 populations (les 4 groupes)
2. $H : \mu_1 = \mu_2 = \mu_3 = \mu_4$ vs. $A : \exists \mu_i \neq \mu_j, i \neq j$
3. $F_{obs} = 11.6$
4. $p_{obs} = 1.8 \times 10^{-5}$
5. $p_{obs} < \alpha = 0.05 \implies$ REJECT H.

(j) df numerateur = 3, df denominateur = 36

À domicile

Exercice 1 (a) (voir la fiche des sorties annotées)

(b) $n - 1 = 17 + 3 = 18 + 2 = 20$, donc $n = 20 + 1 = 21$

(c) $SCE = 178.8$, $CME = 178.8/17 = 10.5$; $\hat{\sigma} = \sqrt{10.5} = 3.24$

(d) $\hat{\beta}_2 \pm ES(\hat{\beta}_2) \cdot t_{n-p-1,0.975} \implies 1.3 \pm 0.4\, t_{17,0.975} \implies 1.3 \pm 0.4 \times 2.11$ (ou 1.3 ± 0.84 ou $[0.46, 2.14]$)

(e) $H : \omega$ est le vrai modèle ($\beta_3 = 0$)
$A : \Omega$ est le vrai modèle ($\beta_3 \neq 0$)

$$F_{obs} = \frac{[SCE(\omega) - SCE(\Omega)]/(p-q)}{SCE(\Omega)/(n-p-1)} = \frac{(188.8 - 178.8)/(3-2)}{178.8/17} = \frac{10}{10.5} = 0.95$$

Degrés de liberté numérateur = 1, donc $F_{1,17} = t_{17}^2$, alors $|t| = \sqrt{0.95} = 0.97$;
$t_{17,0.975} = 2.11 > |t_{obs} = 0.97|$, donc on NE REJETTE PAS H.

Exercice 2 (a) Tableau d'ANOVA :

Source	df	SC	CM	F	p
Pizza variety	$3 - 1 = \mathbf{2}$	23.0	$\mathbf{23.0/2 = 11.5}$	$\mathbf{11.5/17.6 = 0.65}$	0.53
Error	$20 - 2 = \mathbf{18}$	$339.9 - 23.0 = \mathbf{316.9}$	$\mathbf{316.9/18 = 17.6}$		
Total	20	339.9			

(b) NON : $p = 0.53 > \alpha = 0.05 \Rightarrow$ NE REJETTE PAS H que les moyennes de graisse des trois variétés sont égales.

(c) NON : les tests post-hoc sont pour évaluer LESQUELLES des moyennes sont differentes – qui n'a pas de sens lorsque le test global n'a pas detecté une différence (significative).

Exercice 3 (a) Échantillons indépendants (attribution aléatoire aux groupes), des écarts-types égaux pour chaque groupe, et soit les mesures sont normalement distribuées dans chaque groupe, soit les tailles des échantillons sont suffisamment grandes que les moyens peuvent être supposés être Normalement distribué (TCL).

(b) Tableau d'ANOVA :

Source	df	SC	CM	F	p
Treatments	$5 - 1 = \mathbf{4}$	$11681.4 - 4800 = \mathbf{6881.4}$	$6881.4/4 = \mathbf{1720.4}$	$1720.4/80 = \mathbf{21.5}$	4.8×10^{-11}
Error	60	$60 \times 80 = \mathbf{4800}$	80		
Total	$60 + 4 = \mathbf{64}$	11681.4			

(c) $H : \mu_1 = \mu_2 = \mu_3$
$A : \exists\, \mu_i \neq \mu_j$ (au moins une moyenne diffèrent des autres)
$p_{obs} = 0.04359$, donc on REJETTE l'hypothesèse NULLE. Au moins un groupe a une moyenne significativement different des autres.

(d) Plus on effectue de tests, plus la probabilité globale d'une erreur de Type I (faussement rejetant une hypothèse nulle qui est vraie).