# Corrigé examen 27 janvier 2021

## Exercice 1

$Y$ : BPchange et deux facteurs : F1 et F2, avec
F1 : "Dose" à 6 niveaux
F2 : "Treatment" à 2 niveaux.

1) On a un modèle d'analyse de variance à 2 facteurs avec une interaction

$$(M1) \quad Y_{ijk} = \mu + (F1)_i + (F2)_j + (F1*F2)_{ij} + \varepsilon_{ijk}$$
$$i = 1, \dots, 6 \quad j = 1,2 \quad k = 1, \dots, n_{ij}$$

$F1_i$ : c'est l'effet du niveau $i$ de F1 sur $y$

$F2_j$ : — " — $j$ F2 — " —

$(F1*F2)_{ij}$ ... de l'interaction ~~des~~ niveaux $i$ de F1 et $k$ de F2 sur $y$

Contraintes : $\sum_{i=1}^{6} F1_i = 0$, $\sum_{j=1}^{2} F2_j = 0$, $\forall i = 1,\dots,6$, $\sum_{j=1}^{2} (F1*F2)_{ij} = 0$

$\forall j = 1,2$, $\sum_{i=1}^{6} (F1*F2)_{ij} = 0$

$\varepsilon_{ijk} \sim \mathcal{N}(0, \sigma^2)$ indép

2) $H_0$ : (M1) non significatif $\Leftrightarrow$ aucun des 2 facteurs et de l'interaction n'influent $y$.

$$\Leftrightarrow \begin{cases} F1_i = 0 \quad \forall i = 1..6, \quad F2_j = 0 \quad \forall j = 1,2 \\ (F1*F2)_{ij} = 0 \quad \forall i = 1..6, \forall j = 1,2 \end{cases}$$

$\Leftrightarrow$ Modèle réduit : $Y_{ijk} = \mu + \varepsilon_{ijk}$

$H_1$ : (M1) significatif $\Leftrightarrow$ $y$ influencée par F1 ou F2 ou F1*F2 $\Leftrightarrow \exists F1_i \neq 0$ ou $\exists F2_j \neq 0$ ou $\exists (F1*F2)_{ij} \neq 0$

Modèle complet : (M1)

Statistique de test :

$$Z = \frac{SM/11}{SR/48} \underset{H_0}{\sim} F(11, 48)$$

(2)

Stat de test: $z = 29{,}45$, pvalue $< 10^{-16}$ $\Rightarrow$ $H_0$ rejetée $\Rightarrow$
$\Rightarrow$ (M1) significatif

3) On teste $F2$ :

$H_0$ : F2 n'influe pas Y si F1 et F1*F2 sont dans le modèle $\Leftrightarrow F2_j = 0 \quad \forall j = 1, 2$
Modèle : $y_{ijk} = \mu + (F1)_i + (F1*F2)_{ij} + \varepsilon_{ijk}$

$H_1$ : F2 influe Y si F1 et F1*F2 sont dans le modèle
$\Leftrightarrow \exists F2_j \neq 0$
Modèle : (M1)

Stat de test : $Z_{F_2} = \dfrac{S_{F_2}/1}{SR/48} \underset{H_0}{\sim} F(1, 48)$

pvalue $= 0{,}0002$ $\qquad z = 15{,}72$
$\Downarrow$
$H_0$ rejetée $\Rightarrow$ F2 influe Y si F1 et F1*F2 dans le modèle.

Pour les autres variables : F1 et F1*F2 influent également Y.
Il ne faut pas enlever des variables.

4) $\hat{\mu} = 11{,}2 \qquad \hat{F1}_1 = -9{,}8 \qquad \hat{F1}_2 = -9{,}2, \ldots \quad \hat{F1}_6 = -(\hat{F1}_1 + \ldots + \hat{F}$

par rapport à la moyenne la Dose 1 diminue la pression artérielle de 9.8

$\widehat{(F1*F2)}_{11} = -2{,}68 \qquad \widehat{(F1*F2)}_{12} = 2{,}68, \quad \boxed{\hat{\sigma} = 4{,}572}$

par rapport à la moyenne la dose 1 et le traitement 1 diminuent la pression de 2.68.

5) $R^2 adj = 0{,}84 \Rightarrow$ (M1) de bonne qualité d'ajustement

Exercice 2. $Y$ = remiss, var. qualitative qui prend 2 valeurs.

Trois variables explicatives numériques :

$X_1$ = cell, $X_2$ = li, $X_3$ = blast.

1) $Y = 0$ pour 18 obs et $= 1$ pour 9 obs $\Rightarrow$ 27 obs.

2) On a un modèle logistique de la forme :

$$(M2) \quad \mathbb{P}[Y_i = 1 | X_{1i}, X_{2i}, X_{3i}) = \frac{\exp(b_0 + b_1 X_{1i} + b_2 X_{2i} + b_3 X_{3i})}{1 + \exp(b_0 + b_1 X_{1i} + b_2 X_{2i} + b_3 X_{3i})}$$

$i = 1, \dots, 27$.

3) $H_0$ : $X_1$ n'influe pas $\mathbb{P}[Y=1]$ si $X_2$ et $X_3$ sont dans le modèle.

$$\Leftrightarrow b_1 = 0 \Leftrightarrow \mathbb{P}[Y_i = 1 | X_{2i}, X_{3i}) = \frac{\exp(b_0 + b_2 X_{2i} + b_3 X_{3i})}{1 + \exp(b_0 + b_2 X_{2i} + b_3 X_{3i})}$$

$H_1$ : $X_1$ influe $\mathbb{P}[Y=1]$ si $X_2$ et $X_3$ sont dans le modèle

$\Leftrightarrow b_1 \neq 0 \Leftrightarrow (M2)$

stat de test $\quad Z_1 = \dfrac{\hat{b}_1}{\hat{\sigma}_{b_1}} \xrightarrow[H_0]{} N(0,1)$

valeur : $z_1 = 1{,}058 \quad$ pvalue $= 0{,}29 \Rightarrow H_0$ acceptée

<u>Pareil</u> : $X_2$ influe, $X_3$ n'influe pas $\mathbb{P}[Y=1]$

4) $\hat{b}_0 = -10{,}6$. $\hat{b}_1 = 7{,}46 > 0 \quad \hat{b}_2 = 3{,}26 > 0 \quad \hat{b}_3 = -0{,}57 < 0$

quand $X_1$ et $X_2$ ~~__~~ augment $\mathbb{P}[Y=1]$ augmente

$\mathbb{P}[Y=1]$ diminue si $X_3$ augmente.

5) $\hat{\mathbb{P}}$ ~~$[Y=0]$~~ : Prévision des $Y=0$ (non rémission

sur les 18 obs, 16 sont bien prévues.

Taux de mauvaise prévision : $\dfrac{2}{18}$

$Y=1$ (rémission) : seulement 4 obs sur 9 sont bien prévues par (M2). Taux d'erreur : $\frac{5}{9}$.

6) $$\mathbb{P}[Y_i = 1 \mid X_{2i}] = \frac{\exp(b_0 + b_1 X_{2i})}{1+\exp(b_0 + b_2 X_{2i})} \quad i = 1 \ldots, 4.$$

modèle de régression logistique avec une seule variable explicative.

7) Les prévisions pour Y sont les mêmes par (M2) et (M3). Ce n'est pas surprenant car (M3) est obtenu à partir de (M2) en enlevant les 2 var. explicatives non significatives