Corrigé examen 30/01/2019

Ex. 1

1) $\mathbb{P}[Y=0 / \text{age} \dots] = \dfrac{\exp(\mu + \alpha_i + \sum_{j=1}^{5} b_j x_j)}{1+\exp(\cdots)}$ Modèle de rég logistique (1) $i = 1, 2$

$X_1 = $ age $X_2 = $ alb, $X_3 = $ creat, $X_4 = $ hgb, $X_5 = $ mspike,

α_1 : l'effet d'être une femme

$\alpha_2 =$ — " — un homme $\boxed{\alpha_2 = 0}$

2) H_0 : (1) non signif $\Leftrightarrow$ aucune var n'influe :

$$\mathbb{P}[Y=0] = \frac{\exp(\mu)}{1+\exp(\mu)}$$

H_1 : (1) signif $\Leftrightarrow$ $\exists$ une var parmi les 7 qui influe $\mathbb{P}[Y=0]$

Modèle (1) :

Stat de test : LR, Score, Wald   p-value $< 0.0001 \Rightarrow H_0$ rejetée

3) H_0 : l'âge n'influe pas | les autres var dans le modèle

$b_1 = 0$ | . . . .

$$\mathbb{P}[Y=0 / \text{sex}, X_2, \dots, X_5] = \frac{\exp(\mu + \alpha_i + \sum_{j=2}^{5} b_j X_j)}{1+\exp(\dots)}$$

H_1 : l'âge influe $\mathbb{P}[Y=0]$ | . . . .

$b_1 \neq 0$

Modèle (1) $1 = \mathbb{P}[Y=0] + \mathbb{P}[Y=1]$

Stat de test : $Z = \left(\dfrac{\hat{b}_1}{\hat{\sigma}_{b_1}}\right)^2 \xrightarrow[H_0]{\mathcal{L}} \chi^2(1)$, $z = 31.1$

p value $< 10^{-4} \Rightarrow$ l'âge influe $\mathbb{P}[Y=0]$ et donc aussi $\mathbb{P}[Y=1]$

H_0 : le genre n'influe pas $\mathbb{P}[Y=0 | X_1, \dots, X_5]$

$\alpha_1 = \alpha_2 = 0$

$$\mathbb{P}[Y=0 | X_1, \dots, X_5] = \frac{\exp(\mu + \sum_{j=1}^{5} b_j X_j)}{1+\exp(\dots)}$$

(2)

H_1 : le genre influe $\mathbb{P}[Y=0|X_1,\dots,X_5]$: $\Rightarrow \alpha_1 \neq \alpha_2$
$\Rightarrow$ modèle (1)

Stat test : $Z = \left(\frac{\hat{\alpha}_1}{\hat{\sigma}_{\alpha_1}}\right)^2 \xrightarrow[H_0]{\mathcal{L}} \chi^2(1)$

valeur de Z : $z = 0,37$ pvalue $= 0,53 \Rightarrow$ le genre n'influe pas $\mathbb{P}[Y=0]$ et donc aussi $\mathbb{P}[Y=1]$

<u>Pour les autres var</u> : seulement "hgb" influe $\mathbb{P}[Y=1]$ en gardant les autres var dans le modèle.

4) $\hat{\mu} = 1.64$ $\hat{\alpha}_1 = 0.14$ $\hat{\alpha}_2 = 0$, $\hat{b}_1 = -0,14$, $\hat{b}_2 = -0,25$
$\hat{b}_3 = 0,20$ $\hat{b}_4 = 0,49$ $\hat{b}_5 = -0,22$

5) $\mathbb{P}[Y=0|X_1,\dots,X_5] = \dfrac{\exp(1,64+0-0,14\cdot 70-0,25\cdot 2+0,20\cdot 1+0,49\cdot 10-0,22\cdot 2)}{1+\exp(\cdot\ \cdot\ )}$

6) $\mathbb{P}[Y=0|X_1,X_4] = \dfrac{\exp(\mu+b_1X_1+b_4X_4)}{1+\exp(\mu+b_1X_1+b_4X_4)}$

7) $\hat{b}_1 = -0,11 < 0 \Rightarrow$ l'âge diminue $\mathbb{P}[Y=0]$, donc augmente la proba de décès

$\hat{b}_4 = 0,32 > 0 \Rightarrow$ si "hgb" augmente, alors la proba de décès diminue.

8) (M1) $\hat{\mu} = 1,64$ $\hat{b}_1 = -0,14$ $\hat{b}_4 = 0,49$

(M2) $\hat{\mu} = 1,23$ $\hat{b}_1 = 0,89$ $\hat{b}_4 = 1,38$

Old Ratio estimate de (M1) $\in$ IC de "Old Ratio est. de (M2)" ; donc on a les mêmes estimations

(3)

9) Sur 57 obs en vie, 18 bien prévues => taux d'erreur pour la survie : $\frac{39}{57} = 68\%$

Sur 182 décès, 170 bien prévus => taux d'erreur du décès : $\frac{18}{182} = 7\%$

Donc le taux d'erreur du décès est faible, pendant que celui de la survie est dix fois plus grand.
Taux général de mauvais classement :

$$\frac{68+7}{200} = 37{,}5\%$$

Exercice 2

Y = rendement

1) $H_0 : Y \sim N$ $\quad H_1 \nsim N$
Stat de test de Shapiro : valeur stat = 0,98 => pvalue = 0,48
=> H_0 acceptée => $Y \sim N$.

2) (M3) On a un modèle d'analyse de variance avec trois facteurs et interaction d'ordre deux.
Les facteurs : F1 : block, à 6 niveaux $p = 6$
F2 : variété, à 3 niveaux $q = 3$
F3 : niveau engrais à 4 niveaux, $r = 4$.

(M3)
$$Y_{ijk,l} = \mu + (F1)_i + (F2)_j + (F3)_k + (F1*F3)_{ik} +$$
$$+ (F2*F3)_{jk} + (F1*F3)_{ik} + \varepsilon_{ijk,l}$$

$i = 1, \dots, 6 \quad j = 1,2,3 \quad q = 1, \dots, 4 \quad l = 1 \dots, n_{ijk}$

$\sum_{i=1}^{6} F1_i = 0, \quad \sum^{3} F2_j = 0, \quad \sum^{4} F3_k = 0 \quad \sum^{6} (F1*F3)_{ik} = 0 \ \forall k$

4

$F1_i$ l'effet du block i sur le rendement
$F2_j$ —"— variété j —"—
$F3_k$ —"— du niveau d'engrais k - - -
$(F1*F3)_{ik}$ l'effet de l'interaction du block i et des niveau k de l'engrais sur y

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

3) H_0 : (M1) non signif $\Leftrightarrow$ aucun facteur et interaction n'influent le rendement $\Leftrightarrow F1_i = F2_j = F3_k = (F1*F3)_{ik} = (F1*F2)_{ij}$
$= (F2*F3)_{jk} = 0 \quad \forall i,j,k$

Modèle $\quad Y_m = \mu + \varepsilon_m \quad m = 1,\dots,72$

H_1 : (M1) signif $\Leftrightarrow$ au moins un facteur ou interaction infl. y
$\Leftrightarrow \exists i,j,k$ t.q $F1_i \neq 0$ ou $F2_j \neq 0$ - - -
$\Leftrightarrow$ Modèle (M1)

Stat de test: $Z = \dfrac{SM/41}{SR/30} \underset{H_0}{\sim} F(41,30)$

Val stat test $z = 5.423$, pvalue $= 10^{-6} \Rightarrow H_0$ rejetée.

4) H_0 : le block n'influe pas le rendement | les autres dans le modèle
$\Leftrightarrow F1_i = 0 \quad \forall i = 1..6$ | ———
$\Leftrightarrow$ Modèle : $Y_{ijk,l} = \mu + (F2)_j + (F3)_k + (F1*F2)_{ij} + (F2*F3)_{jk} + (F1*F3)_{ik} + \varepsilon_{ijk}$

H_1 : le bloc influe y | - ———
$\Leftrightarrow \exists i$ t.q $F1_i \neq 0$ | - ———

(5)

Modèle : (M3)

Stat test $Z = \frac{SF1/5}{SR/30} \underset{H_0}{\sim} F(5,30)$

la réalisation $z = 15,4$ pvalue$=10^{-7}$ $\Rightarrow$ H_0 rejetée $\Rightarrow$ le block influe | — —

La variété de l'avoine et le niveau d'engrais sont significatifs.

$H_0 : (F1*F2)_{ij} = 0 \mid \forall i, \forall j$ | — —

Modèle $Y_{ijk,e} = \mu + F1_i + F2_j + F3_k + (F1*F3)_{ik} + (F2*F3)_{jk} + \varepsilon_{ijk,e}$

$H_1 : \exists i,j$ t.q $(F1*F2)_{ij} \neq 0$ | — —

Modèle (M.3)

Stat de test $Z = \frac{S_{F1*F2}/10}{SR/30} \underset{H_0}{\sim} F(10,30)$

$z = 2,91 \Rightarrow$ pvalue $= 0,01 \Rightarrow H_0$ rejetée.

les deux autres interactions ne sont pas signif.

5) (M4) $Y_{ijk,e} = \mu + (F1)_i + (F2)_j + F3_k + (F1*F2)_{ij} + \varepsilon_{ijke}$

$\sum F1_i = 0$, $\sum F2_j = 0$, $\sum F3_k = 0$

$\sum_{i=1}^{6} (F1*F2)_{ij} = 0 \quad \forall j$, $\sum_{j=1}^{3} (F1*F2)_{ij} = 0$, $\forall i$

$\varepsilon_{ijk,e} \sim N(0,\sigma^2)$ indep.

6) $\hat{F3}_4 = -(\hat{F3}_1 + \hat{F3}_2 + \hat{F3}_3) = -(-24,5 - 5 + 10) \simeq 19,5$

6

Par rapport à la moyenne, mettre 0,6 d'engrais, augmente le rendement de 19,5

7) $\hat{F3}_1 = -24$

$H_0 : F3_1 = 0 \iff$ par rapport à la moyenne, le pas mettre d'engrais, ne change pas le rendement

$H_1 : F3_1 \neq 0 \iff$ ne pas mettre ~~de bl~~ engrais change le rendement

Stat test. $z = \dfrac{\hat{F3}_1}{\hat{\sigma}_{F3_1}} \sim t(51)$

Stat test $= -9,4$ pvalue $= 10^{-13} \Rightarrow H_1$ acceptée

Ne pas mettre d'engrais, le rendement diminue de 24 % par rapport à la moyenne

8) $\hat{F2}_3 = -(\hat{F2}_1 + \hat{F2}_2) = -(0,5+5,8) = -6,3$

Parce que la variété est signif, alors la variété 3 est moins rentable que la 2, avec un risque de 0,05.

9) $R^2_{adj} = 0,77$ assez bonne qualit