

Devoir 2

Partie théorique

T-1

#1

On veut montrer que $\bar{e} = \sum_{i=1}^n \frac{e_i}{n} = 0$. On peut donc développer la formule de \bar{e} :

$$\begin{aligned}\bar{e} &= \sum_{i=1}^n \frac{e_i}{n} = \sum_{i=1}^n \frac{Y_i - \hat{Y}_i}{n} \\ &= \bar{Y} - \frac{\sum_{i=1}^n x'_i \hat{\beta}}{n} \\ \sum_{i=1}^n x'_i \hat{\beta} &= \hat{\beta}_0 \sum_{i=1}^n 1 + \hat{\beta}_1 \sum_{i=1}^n x_{i,1} + \dots + \hat{\beta}_{p'} \sum_{i=1}^n x_{i,p'} \\ \frac{\sum_{i=1}^n x'_i \hat{\beta}}{n} &= \hat{\beta}_0 + \hat{\beta}_1 \bar{x}_1 + \dots + \hat{\beta}_{p'} \bar{x}_{p'}\end{aligned}$$

Tel que donné dans l'énoncé, on prend pour acquis que:

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{x}_1 - \dots - \hat{\beta}_{p'} \bar{x}_{p'}$$

Ainsi, on obtient:

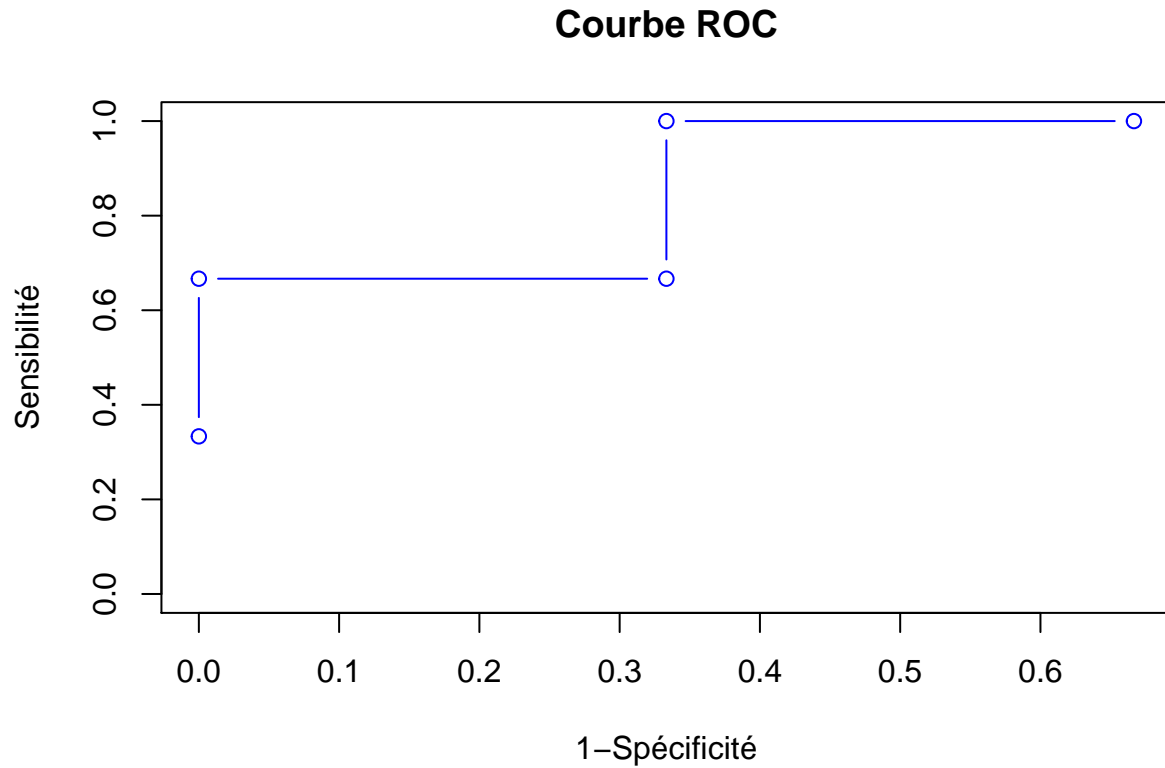
$$\begin{aligned}\bar{e} &= \bar{Y} - \frac{\sum_{i=1}^n x'_i \hat{\beta}}{n} \\ &= \bar{Y} - \left(\hat{\beta}_0 + \hat{\beta}_1 \bar{x}_1 + \dots + \hat{\beta}_{p'} \bar{x}_{p'} \right) \\ &= \bar{Y} - \left(\left[\bar{Y} - \hat{\beta}_1 \bar{x}_1 - \dots - \hat{\beta}_{p'} \bar{x}_{p'} \right] + \hat{\beta}_1 \bar{x}_1 + \dots + \hat{\beta}_{p'} \bar{x}_{p'} \right) \\ &= \bar{Y} - \bar{Y} \\ \bar{e} &= 0\end{aligned}$$

#4

Pour faire la courbe ROC, il faut calculer la valeur de $p_i = P(Y_i = 1, x_i) = \frac{1}{1 + \exp(-(-4.2 + 1.2x_i))}$. Par la suite on calcule une prédiction de \hat{Y}_i selon un certain seuil u_k : si $p_i > u_k$, alors $\hat{Y}_i = 1$, sinon $\hat{Y}_i = 0$. On calcule alors la sensibilité et la spécificité pour différents seuils. Voici un tableau des résultats (\hat{Y}_i est représenté comme étant Y^k selon la valeur de u_k), ainsi que le graphique de la courbe ROC:

Observation	x_i	p_i	Y_i	Y^1	Y^2	Y^3	Y^4	Y^5
u_k	NA	NA	NA	0.1	0.2	0.5	0.8	0.9
obs 1	1	0.0474258731775668	0	0.0	0.0	0.0	0.0	0.0
obs 2	2	0.141851064900488	0	1.0	0.0	0.0	0.0	0.0
obs 3	3	0.354343693774204	1	1.0	1.0	0.0	0.0	0.0
obs 4	4	0.645656306225795	0	1.0	1.0	1.0	0.0	0.0
obs 5	5	0.858148935099512	1	1.0	1.0	1.0	1.0	0.0
obs 6	6	0.952574126822433	1	1.0	1.0	1.0	1.0	1.0

Métriques	Y^1	Y^2	Y^3	Y^4	Y^5
VP	3.0000000	3.0000000	2.0000000	2.0000000	1.0000000
FN	0.0000000	0.0000000	1.0000000	1.0000000	2.0000000
VN	1.0000000	2.0000000	2.0000000	3.0000000	3.0000000
FP	2.0000000	1.0000000	1.0000000	0.0000000	0.0000000
Sensibilité	1.0000000	1.0000000	0.6666667	0.6666667	0.3333333
Spécificité	0.3333333	0.6666667	0.6666667	1.0000000	1.0000000



Partie pratique

P1

```
#Modèle complet
modele_complet=lm(SOMA ~ WT2+HT2+WT9+HT9+LG9+ST9, data = data_tp1)

#Modèle final
#modele_final=lm(SOMA ~ WT2+WT9+HT9+ST9, data = data_tp1)

Y<-data_tp1$SOMA
si <- studres(modele_complet) # residus studentises
hatYi <- modele_complet$fitted.values # valeurs ajustées
i <- 1:length(Y)

ols_plot_resid_fit(modele_complet)
```

```

# Résidus pour chaque observation
plot(i,si,xlab="i",ylab="si",main="Résidus de chaque observation")
abline(h=0,lty=2)

# QQ-plot
ols_plot_resid_qq(modele_complet)

# Tests de normalité
ols_test_normality(modele_complet)

# Transformation de Box-Cox
boxcox(modele_complet)

#####
#Influence
#####
# Valeurs des  $h_{ii}$ 
ols_leverage(modele_complet)

# DFBETAS
ols_plot_dfbetas(modele_complet)

# DFFITS
ols_plot_dffits(modele_complet)

# Distances de Cook
ols_plot_cooksd_chart(modele_complet)

# Residus vs  $h_{ii}$ 
ols_plot_resid_lev(modele_complet)

# covratios
covratio(modele_complet)

# tableau résumé
influence.measures(modele_complet)

```