**Corrigé type de la série 01 :la régression simple et multiple**

**Master 01 : modèles mathématiques en biologie**

**Exercice 01**

$$Y\_{i}=aX\_{i}+b+ε\_{i}$$

$$Y\_{i}=\hat{a}X\_{i}+\hat{b}+\hat{ε}\_{i}$$

$$\hat{Y}\_{i}=\hat{a}X\_{i}+\hat{b}$$

1.

$$\hat{a}=\frac{Cov (X\_{},Y\_{})}{V(X)}=\frac{\sum\_{i=1}^{1429}\left(x\_{i}-\overbar{x}\right)\left(y\_{i}-\overbar{y}\right)}{\sum\_{1}^{1429}\left(x\_{i}-\overbar{x}\right)^{2}}=\frac{26466}{102924}=0.257$$

$$\hat{b}=\overbar{y}-\hat{a}\overbar{x}=21.1-\left(0.257\right)\left(47.3\right)=8.94$$

$$\hat{Y}\_{i}=0.257X\_{i}+8.94$$

2.

SCT=$\sum\_{1}^{1429}\left(y-\overbar{y}\right)^{2}$=8857

SCE=$\sum\_{1}^{1429}\left(\hat{y}-\overbar{y}\right)^{2}=\hat{a}^{2}\sum\_{1}^{1429}\left(x\_{i}-\overbar{x}\right)^{2}=6805$

SCR=SCT-SCE=2052

|  |  |  |  |
| --- | --- | --- | --- |
| Source de variation | Somme des carrées | D.D.L | Moyenne des carrées |
| X | SCE=$6805$ | k-1=2-1=1 | MCE=SCE/1=6805 |
| $$ε\_{i}$$ | SCR=2052 | n-k=1427 | MCR=SCR/n-k=1.43 |
| Y | SCT=8857 | n-1=1428 |  |

3.

$$r\_{x\_{i},y\_{i}}=\frac{Cov (X\_{},Y\_{})}{\sqrt{V(X)}\sqrt{V(y)}}=0.876$$

La variable dépendante (à expliqué Y) est fortement corrélée positivement (+0.876) avec la variable explicative indépendante X

4.

$$R^{2}=r\_{x\_{i},y\_{i}}^{2}=\frac{SCE}{SCT}=0.76$$

La variable X explique la variable Y de 76% et le reste (24%) due des erreurs (les résidus)

Etant donnée $\sum\_{}^{}e\_{i}^{2}=2052$ et supposant que $ε\_{i}\~\~N(0,σ^{2})$

5.

$$\hat{σ}^{2} \_{ε}=\frac{SCR}{n-2}=\frac{2052}{1427}=1.43$$

6.

$$\hat{σ}^{2} \_{a}=\frac{\hat{σ}^{2} \_{ε}}{\sum\_{1}^{1429}\left(x\_{i}-\overbar{x}\right)^{2}}=1.38.10^{-5}$$

$$\hat{σ}^{2} \_{b}=\hat{σ}^{2} \_{ε}\left[\frac{1}{n}+\frac{\overbar{X}^{2}}{\sum\_{1}^{1429}\left(x\_{i}-\overbar{x}\right)^{2}}\right]=0.032$$

7.

$$\left\{\begin{array}{c}H\_{0}: a=0\\H\_{1}: a\ne 0\end{array}\right.$$

$$T\_{c}=\frac{\hat{a}-a}{σ\_{a}}$$

Sous H0 :
$$T\_{c}=\frac{\hat{a}}{σ\_{a}}=\frac{8.94}{\sqrt{0.032}}=279.37$$

$$T\_{t}=T\_{n-2 }\left(5\%\right)=2.069$$

$T\_{c}>T\_{t}$ : on rejette l’hypothèse nulle $H\_{0}: a=0$ d’ouu le paramètre b est hautement significativement différent de 0.

8.

$$\left\{\begin{array}{c}H\_{0}: a=b=0\\H\_{1}: a\ne b\ne 0\end{array}\right.$$

$$F\_{c}=\frac{SCE}{{SCR}/{n-2}}=\frac{R^{2}}{{1-R^{2}}/{n-2}}=4.5710^{3}$$

$$F\_{t}=F\_{\left(1 , n-2\right)}\left(5\%\right)=3.84$$

$$F\_{c}>F\_{t}:le modèle est lobalement significatif$$

**Exercice 02 :**

1. Modèle 01 : On pose le modèle linéaire suivant :

$$Y\_{i}=aX\_{i}+b+ε\_{i}$$

$$Y\_{i}=\hat{a}X\_{i}+\hat{b}+\hat{ε}\_{i}$$

$$\hat{Y}\_{i}=\hat{a}X\_{i}+\hat{b}$$

$$\hat{a}=\frac{Cov (X\_{},Y\_{})}{V(X)}=0.479$$

$$\hat{b}=\overbar{y}-\hat{a}\overbar{x}=36$$

$$\hat{Y}\_{i}=0.479X\_{i}+36$$

2. Modèle 02 : On pose le modèle linéaire suivant :

$$X\_{i}=cY\_{i}+d+ε\_{i}$$

$$X\_{i}=\hat{c}Y\_{i}+\hat{d}+\hat{ε}\_{i}$$

$$\hat{X}\_{i}=\hat{c}Y\_{i}+\hat{d}$$

$$\hat{c}=\frac{Cov (X\_{},Y\_{})}{V(Y)}=1.042$$

$$\hat{d}=\overbar{x}-\hat{c}\overbar{y}=-3.83$$

$$\hat{X}\_{i}=1.042Y\_{i}-3.83$$

3.

$$r\_{x\_{i},y\_{i}}=\frac{Cov (X\_{},Y\_{})}{\sqrt{V(X)}\sqrt{V(y)}}=0.707$$

$$R^{2}=r\_{x\_{i},y\_{i}}^{2}=0.707^{2}=0.5$$

4.

$$\hat{a}.\hat{c}=0.479\*1.042=0.5=R^{2}$$

**Exercice 03**

Soit le modèle de régression multiple $Y\_{i}=a\_{0}+a\_{1}X\_{i1}+a\_{2}X\_{i2}+ε\_{i}$ et étant donner :

1.

$$X^{'}X=\left(\begin{matrix}25&0&0\\0&9.3&5.4\\0&5.4&12.7\end{matrix}\right)$$

2.

$$r\_{x\_{1},x\_{2}}=\frac{Cov (X\_{1},X\_{2})}{\sqrt{V(X\_{1})}\sqrt{V(X\_{2})}}$$

$$\overbar{X}\_{1}=\frac{1}{25}\sum\_{}^{}X\_{1i}=\frac{1}{25}\left(0\right)=0$$

$$\overbar{X}\_{2}=\frac{1}{25}\sum\_{}^{}X\_{2i}=\frac{1}{25}\left(0\right)=0$$

$$V\left(X\_{1}\right)=\frac{1}{n}\sum\_{}^{}x\_{1i}^{2}-\overbar{X}\_{1}^{2}=\frac{1}{25}\left(9.3\right)-0=0.372$$

$$V\left(X\_{2}\right)=\frac{1}{n}\sum\_{}^{}x\_{2i}^{2}-\overbar{X}\_{2}^{2}=\frac{1}{25}\left(12.7\right)-0=0.508$$

$$Cov \left(X\_{1},X\_{2}\right)=\frac{1}{n}\sum\_{}^{}X\_{1}X\_{2}-\overbar{X}\_{1}\overbar{X}\_{2}=\frac{1}{25}\left(5.4\right)-0=0.216$$

$$r\_{x\_{1},x\_{2}}=\frac{Cov (X\_{1},X\_{2})}{\sqrt{V(X\_{1})}\sqrt{V(X\_{2})}}=0.498$$

**Exercice 04**



1.

Le nombre d’observation est n=33

2.

$$X^{'}X^{-1}=\left(\begin{matrix}0.03&0&0\\0&0.03&-0.01\\0&-0.01&0.02\end{matrix}\right)$$

$$\hat{β}=\left(X^{'}X\right)^{-1}X^{'}Y=\left(\begin{matrix}0.03&0&0\\0&0.03&-0.01\\0&-0.01&0.02\end{matrix}\right)\left(\begin{matrix}132\\44\\92\end{matrix}\right)=\left(\begin{matrix}3.96\\0.4\\1.4\end{matrix}\right)$$

3.

SCT=Y’Y=$\sum\_{}^{}y\_{i}^{2}=\sum\_{1}^{33}\left(y\_{i}-\overbar{y}\right)^{2}+n\overbar{y}^{2}=150+33\left(\frac{132}{33}\right)=678$

SCE=$\hat{Y'}\hat{Y}=\hat{β}^{'}\left(X^{'}X\right)\hat{β}=\hat{β}^{'}X^{'}Y=\left(\begin{matrix}3.96&0.4&1.4\end{matrix}\right)\left(\begin{matrix}132\\44\\92\end{matrix}\right)=670$

SCR=SCT-SCE=678-670=8

|  |  |  |  |
| --- | --- | --- | --- |
| Source de variation | Somme des carrées | D.D.L | Moyenne des carrées |
| X | SCE=$670$ | k-1=3-1=2 | MCE=SCE/2=335 |
| $$ε\_{i}$$ | SCR=8 | n-k-1=30 | MCR=SCR/n-k=0.26 |
| Y | SCT=678 | n-1=32 |  |

4.

$$\hat{σ}\_{β}^{2}=\hat{σ}\_{ε}^{2}(X^{'}X)^{-1}$$

$$\hat{σ}\_{ε}^{2}=\frac{SCR}{n-k-1}=0.26$$

$$\hat{σ}\_{β}^{2}=0.26\left(\begin{matrix}0.03&0&0\\0&0.03&-0.01\\0&-0.01&0.02\end{matrix}\right)=\left(\begin{matrix}0.0078&0&0\\0&0.0078&-0.0026\\0&-0.0026&0.0052\end{matrix}\right)$$

5.

$$R^{2}=r\_{x\_{i},y\_{i}}^{2}=\frac{SCE}{SCT}=0.98$$

6.

$$\left\{\begin{array}{c}H\_{0}: β\_{1}=0\\H\_{1}: β\_{1}\ne 0\end{array}\right.$$

$$T\_{c}=\frac{\hat{β\_{1}}-β\_{1}}{σ\_{β\_{1}}}$$

Sous H0 :
$$T\_{c}=\frac{\hat{β\_{1}}}{σ\_{β\_{1}}}=\frac{0.4}{\sqrt{0.0078}}=4.52$$

$$T\_{t}=T\_{n-3 }\left(5\%\right)=2.042$$

$T\_{c}>T\_{t}$ : on rejette l’hypothèse nulle $H\_{0}: β\_{1}=0$ d’ouu le paramètre $β\_{1}$ est hautement significativement différent de 0.

8.

$$\left\{\begin{array}{c}H\_{0}: β\_{0}=β\_{1}=β\_{2}=0\\H\_{1}: β\_{0}\ne β\_{1}\ne β\_{2}\ne 0\end{array}\right.$$

$$F\_{c}=\frac{\frac{SCE}{k}}{{SCR}/{n-3}}=\frac{R^{2}}{{1-R^{2}}/{n-2}}=1288.46$$

$$F\_{t}=F\_{\left(k , n-3\right)}\left(5\%\right)=3.32$$

$$F\_{c}>F\_{t}:le modèle est lobalement significatif$$