TP 6
Correction de l'exercice 4

Les modifications à apporter sur le corrigé de ce matin est indiquée en bleu

Exercice 4: Test de significativité globale dans le modèle linéaire simple

On considère le modèle linéaire simple:

$$y_i = a_0 + a_1 x_{1,i} + \varepsilon_i, \quad i = 1...n.$$

4.1 Soit F^* la statistique de Fisher pour un test de significativité globale du modèle. Soit $t_{a_1=0}$ la statistique de Student pour un test de l'hypothèse $H_0 : \quad a_1 = 0$. Démontrer que:

$$F^* = t^2_{a_1=0}.$$

Il s'agit simplement de mener le calcul. Dans le cas du test de significativité globale, on a:

$$F^* = \frac{\left(\frac{R^2}{k}\right)}{\left(\frac{1-R^2}{n-(k+1)}\right)} = \frac{R^2}{\left(\frac{1-R^2}{n-2}\right)}.$$

Or souvenons-nous que $R^2 = r^2_{x,y}$ dans le modèle linéaire simple, donc on obtient:

$$F^* = \frac{r^2_{x,y}}{\left(\frac{1-R^2}{n-2}\right)}.$$

Par ailleurs, $1 - R^2 = SCR/SCT$ et donc on obtient:

$$F^* = \frac{r^2_{x,y}}{\left(\frac{SCR}{(n-2)SCT}\right)}.$$

Remplaçons maintenant $r_{x,y}$ par sa définition:

$$r_{x,y} = \frac{Cov(x,y)}{\sqrt{Var(x)Var(y)}}$$

pour obtenir:

$$F^* = \frac{\left(\frac{Cov(x,y)^2}{Var(x)Var(y)}\right)}{\left(\frac{SCR}{(n-2)SCT}\right)}$$

Une simplification vient de la constatation suivante:

$$SCT = \sum_{i=1}^{n}(y_i - \overline{y})^2 = nVar(y)$$

donc on obtient:

$$F^* = \frac{\left(\frac{Cov(x,y)^2}{Var(x)}\right)}{\left(\frac{SCR}{n(n-2)}\right)} = \frac{\left(\frac{Cov(x,y)}{Var(x)}\right)^2}{\left(\frac{SCR}{n(n-2).Var(x)}\right)}.$$

Tout d'abord, notons que:

$$\frac{Cov(x,y)}{Var(x)} = \hat{a}_1$$

et donc:

$$F^* = \frac{(\hat{a}_1)^2}{\left(\frac{SCR}{n(n-2).Var(x)}\right)}.$$

De plus notons que:

$$\frac{SCR}{n-2} = \hat{\sigma}^2$$

et donc:

$$F^* = \frac{(\hat{a}_1)^2}{\left(\frac{\hat{\sigma}^2}{nVar(x)}\right)}.$$

Le dernier point à noter est que:

$$\widehat{Var}(\hat{a}_1) = \frac{\hat{\sigma}^2}{nVar(x)} = (X^tX)^{-1}_{2,2}\hat{\sigma}^2$$

d'après le calcul fait en cours. On obtient donc:

$$F^* = \frac{(\hat{a}_1)^2}{\widehat{Var}(\hat{a}_1)} = \left(\frac{\hat{a}_1}{\sqrt{\widehat{Var}(\hat{a}_1)}}\right)^2.$$

On peut donc conclure:

$$F^* = (t_{a_1=0})^2.$$

4.2 Pourquoi le résultat précédent était-il prévisible? Pourquoi une telle relation ne se généralise-t'elle pas au cas d'un modèle multiple?

Le test de Student qui utilise la statistique $t_{a_1=0}$ *a pour objet de tester si* $a_1 = 0$.

Dans le contexte d'un modèle linéaire simple, le test de significativité globale s'écrit de la même façon: il teste si $a_1 = 0$*!*

Il semble donc "logique" que les deux tests fassent intervenir la même statistique de base.

Pour la généralisation au cas d'un modèle avec k variables explicatives, le test de significativité globale teste si $a_1 = ... = a_k = 0$ *alors qu'un test de Student ne peut tester la nullité que d'un seul coefficient. On ne s'attend donc pas à trouver un lien aussi simple entre les deux tests.*