# Introduction à l'Econométrie - Licence 3

Correction Interro écrite no.1

Rémi Yin

12 mars 2014

## Questions de cours (4 points)

1. **On considère le modèle $y_n = a+bx_n+\varepsilon_n$ pour lequel on dispose de $N$ observations. Qu'appelle-t-on coefficient de détermination ($R^2$) du modèle ? Expliquez pourquoi il est compris entre 0 et 1. (2 points)**
   Le coefficient de détermination est défini par
   $$R^2 = \frac{SCE}{SCT} = \frac{\sum_{n=1}^{N}(\hat{y_n} - \bar{y_N})^2}{\sum_{n=1}^{N}(y_n - \bar{y_N})^2} = 1 - \frac{SCR}{SCT}$$
   Le coefficient de détermination est donc compris entre 0 et 1 car il s'agit d'un rapport entre la variance expliquée par le modèle et la variance totale du phénomène observé.Le pouvoir explicatif des variables utilisées est ainsi d'autant plus fort que le $R^2$ est proche de 1. Bien que cette quantité soit systématiquement utilisée pour évaluer la qualité globale du modèle estimé, son interprétation nécessite beaucoup de précautions. Pour les lecteurs avertis : Le prochain qui me parle du $R^2$...
2. **On considère le modèle $y_n = a + bx_n + \varepsilon_n$ pour lequel on dispose de $N$ observations. Définir les résidus estimés $\hat{\varepsilon_n}$**
   **Quelles sont leurs propriétés ? (2 points)**
   Les résidus estimés sont la différence entre la valeur y observée et la valeur y estimée. Formellement :
   $$\forall n \quad \hat{\varepsilon_n} = y_n - \hat{y_n}$$
   Pour ce qui est des propriétés des résidus, on sait d'après les équations normales (Conditions du premier ordre) du programme de minimisation des erreurs que :
   $$-2\sum_{n=1}^{N}(y_n - \hat{a} - \hat{b}x_n) = 0 \Leftrightarrow \sum_{n=0}^{N}\hat{\varepsilon_n} = 0$$
   $$-2\sum_{n=1}^{N}x_n.(y_n - \hat{a} - \hat{b}x_n) = 0 \Leftrightarrow \sum_{n=0}^{N}\hat{\varepsilon_n}x_n = 0$$
   On peut en déduire que la moyenne empirique des résidus estimés est égale à 0 :
   $$\overline{\hat{\varepsilon_N}} = 0$$

FIGURE 1 – CR7 du Real Madrid

# Exercice 1 (12 points)

On s'intéresse au lien qu'il peut y avoir entre le poids et la taille des buteurs dans les différents championnats Européens de Football en 2014 . On considère le modèle linéaire simple :

$$y_n = a + bx_n + \varepsilon_n \quad \forall n = 1, \dots, N \tag{1}$$

où $y_n$ désigne la poids et $x_n$ la taille des buteurs. On suppose que les résidus sont indépendants et suivent une loi $\mathcal{N}(0; \sigma^2)$

On estime ce modèle par moindres carrés ordinaires et le site de la Fifa nous donne des informations sur $N = 55$ footballeurs. Les observations (en centimètres et en kilogrammes) conduisent aux valeurs suivantes :

$$\overline{x_N} = 183,6 \qquad \overline{y_N} = 79$$

$$Cov_{emp}(x_n; y_n) = 39,2$$

$$s_x = 6,7 \qquad s_y = 7,3$$

où $s_x$ et $s_y$ sont respectivement les **écarts-types** empiriques de la taille et du poids. Les données récoltées pour ce dossier sont disponibles sur le site du jeu Fifa 14 : sofifa.com/fr

1. **Calculez $\hat{a}$ et $\hat{b}$ (2 points)**

$$\hat{b} = \frac{Cov_{emp}(x_n; y_n)}{V_{emp}(x_n} = \frac{39,2}{6,7^2} = 0,87$$

$$\hat{a} = \bar{y_N} - \hat{b}\bar{x_N} = 79 - 0.87 \times 183.6 = -80.7$$

2. **Calculez $\hat{\sigma^2}$ (2 points)**

On souhaite calculer $\hat{\sigma^2} = \frac{SCR}{N-2}$ Or, on sait que :

$$SCR = \sum_{n=1}^{N} \hat{\varepsilon_n^2}$$

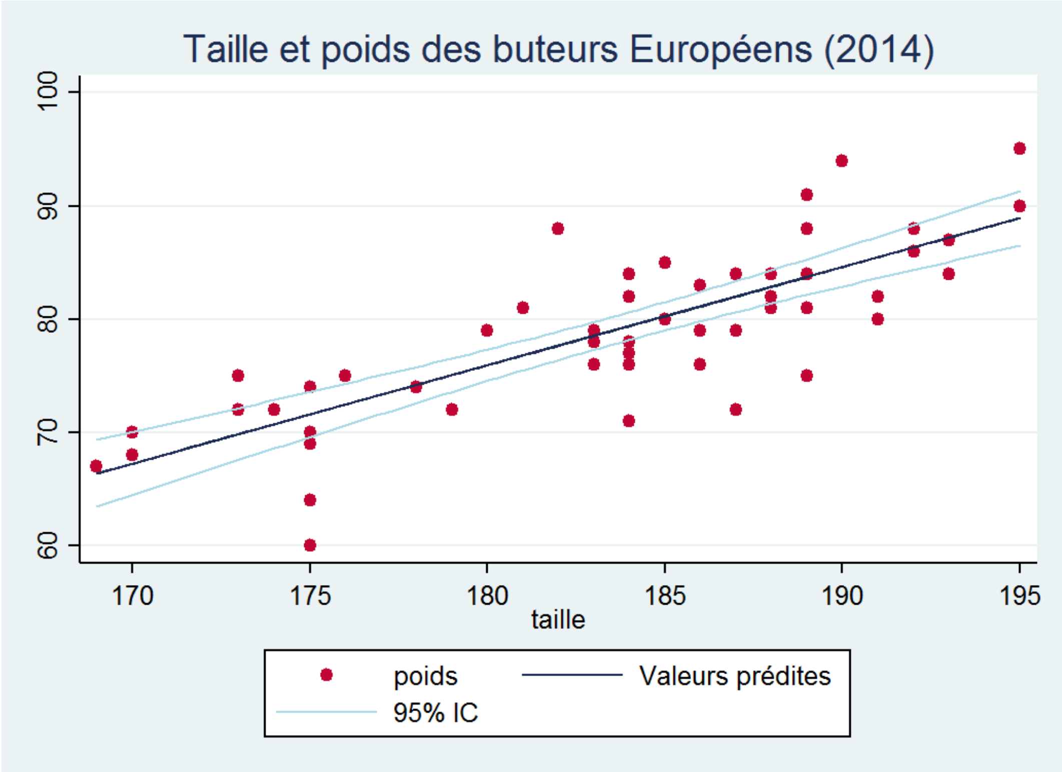


FIGURE 2 – Régression par la méthode des m.c.o

$$
\begin{aligned}
\sum_{n=1}^{N} \hat{\varepsilon}_n^2 &= \sum_{n=1}^{N} (y_n - \hat{y_N})^2 \\
&= \sum_{n=1}^{N} (y_n - \hat{a} - \hat{b}x_n)^2 \\
&= \sum_{n=1}^{N} (y_n - \bar{y_N} + \hat{b}\bar{x_N} - \hat{b}x_n)^2 \\
&= \sum_{n=1}^{N} (y_n - \bar{y_N})^2 + \hat{b}^2(x_n - \bar{x_N})^2 - 2\hat{b}(x_n - \bar{x_N})(y_n - \bar{y_N}) \\
&= N.V_{emp}(y_n) + \hat{b}^2.N.V_{emp}(x_n) - 2\hat{b}.N.Cov_{emp}(x_n; y_n) \\
&= 55 \times 7.3^2 + 0.87^2 \times 55 \times 6.7^2 - 0.87 \times 2 \times 55 \times 39.2 \\
&= 1048.26
\end{aligned}
$$

On en déduit $\hat{\sigma}^2 = \frac{1048.26}{53} = 19.78$

3. **Menez, au seuil de 10%, les tests des significativités du modèle. (2 points)**

On s'intéresse au test $H_0 : b = 0$ contre $H_1 : b \neq 0$ au niveau $\alpha = 10\%$

On sait que :

$$\frac{\hat{b}-b}{\sqrt{\frac{\hat{\sigma}^2}{N.V_{emp}(x_n)}}} \rightarrow \mathcal{T}_{N-2}$$
$$\Rightarrow \frac{\hat{b}}{\sqrt{\frac{\hat{\sigma}^2}{N.V_{emp}(x_n)}}} \overset{H_0}{\rightarrow} \mathcal{T}_{N-2}$$

On rejette donc $H_0$ si

$$\left|\frac{\hat{b}}{\sqrt{\frac{\hat{\sigma}^2}{N.V_{emp}(x_n)}}}\right| > t^*_{0.95;53} = 1.676$$

Or,

$$\left|\frac{\hat{b}}{\sqrt{\frac{\hat{\sigma}^2}{N.V_{emp}(x_n)}}}\right| = 9.7$$

On conclut donc au rejet de $H_0$ au niveau 10%. b est significatif au seuil de 10 %.

On s'intéresse au test $H_0 : a = 0$ contre $H_1 : a \neq 0$ au niveau $\alpha = 10\%$ On sait que :

$$\frac{\hat{a}-a}{\sqrt{\frac{\hat{\sigma}^2}{N.V_{emp}(x_n)}(s_x^2+\bar{x_N}^2)}} \rightarrow \mathcal{T}_{N-2}$$
$$\Rightarrow \frac{\hat{a}}{\sqrt{\frac{\hat{\sigma}^2}{N.V_{emp}(x_n)}(s_x^2+\bar{x_N}^2)}} \overset{H_0}{\rightarrow} \mathcal{T}_{N-2}$$

On rejette $H_0$ si $\left|\frac{\hat{a}}{\sqrt{\frac{\hat{\sigma}^2}{N.V_{emp}(x_n)}(s_x^2+\bar{x_N}^2)}}\right| > t^*_{0.95;53} = 1.676$

Or,

$$\left|\frac{\hat{a}}{\sqrt{\frac{\hat{\sigma}^2}{N.V_{emp}(x_n)}(s_x^2+\bar{x_N}^2)}}\right| = 4.90$$

On conclut donc au rejet de $H_0$ au niveau 10%. a est significatif au seuil de 10%.

4. **Menez, au seuil de 10% le test d'hypothèse** $H_0$ $b = 0,8$ **contre l'hypothèse** $H_1$ $b \neq 0,8$ **(1,5 point)**

On s'intéresse au test $H_0 : b = 0.8$ contre $H_1 : b \neq 0.8$ au niveau $\alpha = 10\%$.

$$\frac{\hat{b}-0.8}{\sqrt{\frac{\hat{\sigma}^2}{N.V_{emp}(x_n)}}} \overset{H_0}{\rightarrow} \mathcal{T}_{N-2}$$

Il s'agit d'un test **bilatéral** :

On rejette donc $H_0$ si $\left|\frac{\hat{b}-0.8}{\sqrt{\frac{\hat{\sigma}^2}{N.V_{emp}(x_n)}}}\right| > t^*_{0.95;53} = 1.676$

Or,

$$\left|\frac{\hat{b}-0.8}{\sqrt{\frac{\hat{\sigma}^2}{N.V_{emp}(x_n)}}}\right| = 0.78$$

On ne rejette pas $H_0$ au seuil 10%. b est significativement égal à 0.8 au seuil de 10%

5. **Menez, au seuil de 10% le test d'hypothèse** $H_0$ $b \leq 0,8$ **contre l'hypothèse** $H_1 :$ $b > 0,8$ **(1,5 point)**

   On s'intéresse au test $H_0 : b = 0.8$ contre $H_1 : b > 0.8$ au niveau $\alpha = 10\%$.

$$\frac{\hat{b}-0.8}{\sqrt{\frac{\hat{\sigma}^2}{N.V_{emp}(x_n)}}} \xrightarrow{H_0} \mathcal{T}_{N-2}$$

   Il s'agit d'un test **unilatéral** :

   On rejette donc $H_0$ si $\frac{\hat{b}-0.8}{\sqrt{\frac{\hat{\sigma}^2}{N.V_{emp}(x_n)}}} > t^*_{0.90;53} = 1.299$

   Or,

$$\frac{\hat{b}-0.8}{\sqrt{\frac{\hat{\sigma}^2}{N.V_{emp}(x_n)}}} = 0.78$$

   On ne rejette pas $H_0$ au seuil 5%. b est significativement inférieur à 0.8.

6. **Construisez un intervalle de confiance à 5 % pour le paramètre** $b$ **(1 point)** cf. Cours magistraux et TD de statistiques du premier semestre pour la procédure de l'intervalle de confiance. On veut :

$$P\left( - t_{(0.975,53)} < \frac{\hat{b}-0.8}{\sqrt{\frac{\hat{\sigma}^2}{N.V_{emp}(x_n)}}} < t_{(0.975,53)}\right) = 95\%$$

   Les calculs permettent d'écrire l'intervalle de confiance sous la forme suivante :

$$\begin{aligned} IC(b) &= \left[\hat{b} \pm 2.009\sqrt{\frac{19.78}{55 \times 6.7^2}}\right] \\ &= \left[0.69; 1.049\right] \end{aligned}$$

7. **Construisez un intervalle de prévision au niveau de confiance 95% lorsque** $x_{N+1} = 195$ **(2 point)**

   On réalise l'estimation ponctuelle : $y^*_{N+1} = -80.7 + 0.87 \times 195 = 88.95$

   On sait que :

$$\frac{Y^*_{N+1} - Y_{N+1}}{\sqrt{\hat{\sigma}^2 \lambda_{N+1}}} \rightsquigarrow \mathcal{T}_{N-2}$$

Avec

$$\lambda_{N+1} = 1 + \frac{1}{N} + \frac{(x_{N+1} - \bar{x_N})^2}{N.V_{emp}(x_n)}$$

On veut :

$$P\left( - t_{(0.975,53)} < \frac{Y^*_{N+1} - Y_{N+1}}{\sqrt{\hat{\sigma}^2 \lambda_{N+1}}} < t_{(0.975,53)} \right) = 95\%$$

Les calculs permettent d'écrire l'intervalle de prévision sous la forme suivante :

$$\begin{aligned} IC(b) &= \left[ Y_{N+1} \pm 2.009 \sqrt{1 + \frac{1}{55} + \frac{(195 - 1836.6)^2}{55 \times 7.3^2}} \right] \\ &= \left[ 79.73; 98.16 \right] \end{aligned}$$

Ainsi, pour un buteur mesurant 1,95m, on peut affirmer avec une certitude de 95% qu'il pèse entre 79,73 et 98,16 kilos.

## Exercice 2 (3 points)

On considère le modèle linéaire simple :

$$y_n = a + bx_n + \varepsilon_n \quad \forall n = 1, \dots, N \tag{2}$$

pour lequel on dispose de $N$ observations.

On note $R^2$ le coefficient de détermination obtenu lorsqu'on estime ce modèle par m.c.o et $\rho^2_{xy}$ le coefficient de corrélation empirique entre $x$ et $y$.

1. **Montrez que $R^2 = \rho^2_{xy}$ (3 points)**
   Cf, exercice de TD pour la correction.

Table 1 – Extrait du tableau Taille-Poids des buteurs en 2014 (Données http ://sofifa.com/fr)

| Nom | Taille | Poids |
|---|---|---|
| Ibrahimovic | 195 | 95 |
| falcao | 176 | 75 |
| Van Persie | 187 | 72 |
| Suarez | 181 | 81 |
| Cavani | 184 | 71 |
| Lewandowski | 184 | 78 |
| Tevez | 173 | 75 |
| Benzema | 187 | 79 |
| Gomez | 189 | 88 |
| Neymar | 175 | 64 |
| Balotelli | 189 | 88 |
| Higuain | 184 | 82 |
| Huntelaar | 186 | 83 |
| Negredo | 186 | 79 |
| Drogba | 189 | 91 |
| Jovetic | 183 | 79 |
| Mandzukic | 187 | 84 |
| Dzeko | 193 | 84 |
| Diego Costa | 188 | 81 |
| Soldado | 179 | 72 |
| David Villa | 175 | 69 |
| Eto'o | 180 | 79 |
| Martinez | 185 | 85 |
| Hernandez | 175 | 60 |
| Sturridge | 183 | 76 |
| Rossi | 173 | 72 |
| Llorente | 195 | 90 |
| Palacio | 175 | 70 |
| Fernando Torres | 183 | 78 |
| Berbatov | 189 | 75 |
| Klose | 184 | 84 |
| Cardosa | 188 | 84 |
| Bony | 182 | 88 |
| Benteke | 192 | 86 |
| Cardozo | 193 | 87 |
| Ba | 189 | 84 |
| Milito | 183 | 78 |
| Kiessling | 191 | 80 |
| Di Natale | 170 | 68 |
| Defoe | 170 | 70 |
| Barrogios | 189 | 81 |
| Lukaku | 190 | 94 |
| Doumbia | 178 | 74 |
| Remy | 184 | 76 |
| Giroud | 192 | 88 |
| ... | ... | ... |