# Statistiques appliquées - L3 d'Economie
# Devoir - Groupes 8, 9 & 12

Pierre-Emmanuel PERRIER DE LA BATHIE & Marc SANGNIER

7 décembre 2007

Ce devoir est à rendre pour le 14 décembre 2007.
En annexe, vous trouverez des résultats que vous pouvez utiliser dans vos calculs.

## Lien entre les années d'études et le salaire d'embauche

### Section 1 - Modèle économique

On s'intéresse ici à un modèle (très) simple qui permettrait de d'expliquer le salaire d'embauche par le nombre d'années d'études. Le modèle que nous nous envisageons est :

$$y_i = b + ax_i \tag{1}$$

Dans l'équation (1), $y_i$ représente le salaire d'embauche de l'individu $i$ et $x_i$ le nombre d'années d'études de l'individu $i$. $a$ et $b$ sont des nombres réels.

#### Question 1.1

Quelle est la variable exogène (ou variable explicative) de ce modèle ?

#### Question 1.2

Quelle est la variable endogène (ou variable à expliquer) de ce modèle ?

#### Question 1.3

A priori, quel est le signe du réel $a$ ? Pourquoi ? Comment peut être interprété $b$ ?

#### Question 1.4

Expliquez en quoi la théorie du capital humain peut-être considérée comme théorie sous-jacente de ce modèle ?

### Section 2 - Les données

On dispose de $n = 40$ observations (fictives) du couple $(X; Y)$. $X$ (respectivement $Y$) est la variable aléatoire représentant les années d'études (le salaire de l'individu). Les années sont comptabilisées en années complètées (les redoublements ne sont donc pas pris en compte). Les salaires sont donnés sous la forme d'indices, le plus faible ayant été choisi comme base de calcul, ils sont exprimés dans la même unité monétaire, ils sont donc comparables entre eux. Pour plus de réalité, on peut supposer que ces salaires représentent pour tous les individus la même fraction du salaire mensuel. Ces observations sont rassemblées dans les tableaux présentés en annexe.

#### Question 2.1

Compte tenu de votre connaissance du système éducatif français, à quoi correspond un nombre d'années égal à 12 ? De même, à quoi correspond un nombre d'années d'études égal à 17 ?

#### Question 2.2

Calculez la moyenne empirique de la distribution des salaires. Vous l'appellerez $\bar{Y}$.

**Question 2.3**

Calculez la moyenne empirique de la la distribution des années d'études. Vous l'appellerez $\bar{X}$.

**Question 2.4**

Calculez la covariance du couple $(X;Y)$.

**Question 2.5**

Calculez le coeficient de corrélation linéaire du couple $(X;Y)$.

## Section 3 - Modèle économètrique

On introduit le modèle économètrique suivant ;

$$y_i = b + ax_i + \varepsilon_i \tag{2}$$

Dans l'équation (2), $y_i$ représente le salaire d'embauche de l'individu $i$, $x_i$ le nombre d'années d'études de l'individu $i$ et $\varepsilon_i$ des caractèristiques inobservées (ou non explicitées) de l'individu $i$ ayant un impact sur son salaire d'embauche. $a$ et $b$ sont des nombres réels.

**Question 3.1**

Expliquez l'intérêt d'introduire le terme $\varepsilon_i$.

**Question 3.2**

Quelles sont les hypothèses faites pour estimer ce modèle par la méthode des moindres carrés ordinaires ?

## Section 4 - Méthode des moindres carrés ordinaires

Vous allez maintenant procèder pas à pas à l'estimation de ce modèle. La méthode des moindres carrés ordinaires consiste à minimiser l'écart entre la valeur observée de $y_i$ et sa valeur prédite le modèle. Pour une observation donnée de $x_i$, la valeur prédite est $\hat{y}_i = ax_i + b$. On désire donc déterminer $a$ et $b$ de façon à minimiser pour toutes les observations le terme $y_i - \hat{y}_i$. Pour ce faire, on définit une fonction de perte quadratique, c'est à dire la somme des écarts au carré. Soit $S$ cette fonction que nous allons chercher à minimiser :

$$S = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{3}$$

**Question 4.1**

Montrez que :

$$S = \sum_{i=1}^{n} \varepsilon_i^2 \tag{4}$$

Remarquez que :

$$S = \sum_{i=1}^{n} (y_i - b - ax_i)^2$$

**Question 4.2**

Quelles sont les conditions du premier ordre par rapport à $a$ et $b$ pour minimiser $S$ ?

**Question 4.3**

On suppose les conditions du second ordre vérifiées. Résolvez de façon analytique le système d'équation suivant :

$$\frac{\partial S}{\partial a} = 0 \tag{5}$$

$$\frac{\partial S}{\partial b} = 0 \tag{6}$$

**Question 4.4**

Calculez $\hat{a}$ et $\hat{b}$, les solutions du système précédents :

$$\hat{a} = \frac{\sum_{i=1}^{n}(x_i - \bar{X})(y_i - \bar{Y})}{\sum_{i=1}^{n}(x_i - \bar{X})^2} \tag{7}$$

$$\hat{b} = \bar{Y} - \hat{a}\bar{X} \tag{8}$$

**Question 4.5**

Donnez l'expression du salaire d'embauche théorique $y$ en fonction du nombre d'années d'études $x$.

**Question 4.6**

Montrez que la droite définie à la question précédente passe par le point $(\bar{X}; \bar{Y})$.

**Question 4.7**

Représentez graphiquement votre démarche d'estimation (inutile de représenter les 40 observations).

**Question 4.8**

Ecrivez la dérivée partielle de $y$ par rapport à $x$.

**Question 4.9**

Donnez l'interprétation économique du coefficient $a$ en tenant compte de sa valeur estimée $\hat{a}$.

## Section 5 - Analyse de la variance

L'équation fondamentale d'analyse de la variance est :

$$\sum_{i=1}^{n}(y_i - \bar{Y})^2 = \sum_{i=1}^{n}(\hat{y}_i - \hat{\bar{Y}})^2 + \sum_{i=1}^{n}\varepsilon_i^2 \tag{9}$$

De l'équation (9) on peut déduire l'expression du coefficient de détermination du modèle, $R^2$ :

$$R^2 = \frac{\sum(\hat{y}_i - \bar{Y})^2}{\sum(y_i - \bar{Y})^2} \tag{10}$$

**Question 5.1**

Commentez l'équation fondamentale d'analyse de la variance.

**Question 5.2**

Montrez que le coefficient de détermination peut s'écrire de la façon suivante :

$$R^2 = 1 - \frac{\sum \varepsilon_i^2}{\sum(y_i - \bar{Y})^2} \tag{11}$$

**Question 5.3**

Caculez et interprètez le coefficient de détermination de notre modèle.

## Annexe

Tableaux de données :

| Oservation N°$i$ | 01 | 02 | 03 | 04 | 05 | 06 | 07 | 08 | 09 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Salaire ($y_i$) | 100 | 140 | 115 | 125 | 115 | 110 | 120 | 115 | 145 | 130 |
| Années d'études ($x_i$) | 12 | 17 | 13 | 13 | 12 | 11 | 15 | 15 | 16 | 15 |

| Oservation N°$i$ | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|
| Salaire ($y_i$) | 130 | 150 | 130 | 170 | 150 | 130 | 170 | 175 | 165 | 185 |
| Années d'études ($x_i$) | 16 | 16 | 15 | 17 | 18 | 15 | 18 | 16 | 17 | 19 |

| Observation N°$i$ | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 |
|---|---|---|---|---|---|---|---|---|---|---|
| Salaire ($y_i$) | 140 | 110 | 115 | 130 | 165 | 200 | 150 | 150 | 145 | 125 |
| Année d'études ($x_i$) | 17 | 9 | 13 | 15 | 15 | 20 | 15 | 16 | 13 | 15 |

| Observation N°$i$ | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 |
|---|---|---|---|---|---|---|---|---|---|---|
| Salaire ($y_i$) | 190 | 125 | 120 | 190 | 125 | 140 | 150 | 180 | 140 | 150 |
| Année d'études ($x_i$) | 19 | 13 | 12 | 19 | 17 | 16 | 16 | 17 | 15 | 15 |

Les résultats présentez ci-dessous sont issus des données présentées ci-dessus. Les notations sont celles de l'énoncé.

$$\sum_{i=1}^{n} y_i = 5710$$

$$\sum_{i=1}^{n} x_i = 613$$

$$\sum_{i=1}^{n} (y_i - \bar{Y})^2 = 24947,5$$

$$\sum_{i=1}^{n} (x_i - \bar{X})^2 = 216,775$$

$$\sum_{i=1}^{n} x_i y_i = 89395$$

$$\sum_{i=1}^{n} (x_i - \bar{X})(y_i - \bar{Y}) = 1889,25$$

$$\sum_{i=1}^{n} \varepsilon_i^2 = 8482,199$$