

# Lien entre IPS et goûts d'élèves de

ENZO MERIOUD

RAYAN MOUSTAFID DOS SANTOS

MONSIEUR S. GIBAUD \*

Jeudi 11 Janvier 2024

## Table des matières

<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	Motivations . . . . .	2
1.2	Prérequis Mathématiques, la régression linéaire . . . . .	2
1.3	Explications . . . . .	2
1.4	Organisation de l'étude et résultats obtenus . . . . .	2
1.4.1	Résultats Attendus . . . . .	2
1.4.2	Résultats Obtenus . . . . .	2
1.4.3	Approximations faites . . . . .	3
<b>2</b>	<b>Corps de l'étude</b>	<b>4</b>

---

\*Directeur d'étude

# 1 Introduction

## 1.1 Motivations

### 1.2 Prérequis Mathématiques, la régression linéaire

On a :  $y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_3 x_{3,i} + \dots + \beta_p x_{p,i} + \varepsilon_i$

Avec  $i$  : la  $i$ -ème observation de la variable  $Y$ ,  $\beta$  : les coefficients du modèle,  $x_{1,i}$  à  $x_{p,i}$  les  $i$ -ème observation de la  $j$ -ème variable explicative et  $\varepsilon_i$  l'erreur du modèle.

Cela est équivalent à l'écriture matricielle :  $Y = X\beta + \varepsilon$

Soit :

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, X = \begin{pmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \dots & x_{1K} \\ 1 & x_{21} & \dots & x_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{nK} \end{pmatrix}, \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_n \end{pmatrix}, \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

Donc :

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \dots & x_{1K} \\ 1 & x_{21} & \dots & x_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{nK} \end{pmatrix} \times \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_n \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

### 1.3 Explications

Le modèle linéaire est utilisé dans un grand nombre de champs disciplinaires. Il en résulte une grande variété dans la terminologie. Soit le modèle suivant :

$$Y = X\beta + \varepsilon$$

La variable  $Y$  est appelée variable expliquée ou variable endogène. Les variables  $X$  sont appelées variables explicatives ou variables exogènes.  $\varepsilon$  est appelé terme d'erreur ou perturbation.

On note généralement  $\hat{\beta}$  le vecteur des paramètres estimés. On définit la valeur prédite ou ajustée  $\hat{Y} = X\hat{\beta}$  et le résidu comme la différence entre la valeur observée et la valeur prédite :

$$\hat{\varepsilon} = Y - \hat{Y}$$

On définit aussi la somme des carrés des résidus (SCR) comme la somme sur toutes les observations des carrés des résidus :

$$SCR = \varepsilon' \varepsilon = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

### 1.4 Organisation de l'étude et résultats obtenus

#### 1.4.1 Résultats Attendus

Le But de l'étude était de trouver un lien entre l'IPS des élève et leurs goûts en HGGSP. Effectivement, nous espérions que le coefficients de l'IPS soit le plus important. Malheureusement, d'après la régression linéaire, ce qui influence le plus les goûts des élèves c'est ce que préfère enseigner les professeurs. Cependant, l'IPS est lié au niveau d'étude des professeurs.

#### 1.4.2 Résultats Obtenus

Nous avons donc chercher d'autres résultats avec les données que nous possédons, et c'est ce que l'on a trouvé.

Nous avons d'abord chercher un lien entre **âge** des professeurs et **IPS**. Les coefficients sont cette fois positifs et suffisamment hauts. Plus l'IPS est haut plus l'âge des professeur l'est aussi. Ce résultat est

particulièrement intéressant, effectivement, cela montre que les professeurs plus âgés ont tendance à aller dans des lycées d'élèves plus aisés.

Nous avons ensuite cherché un lien entre **sexe** des professeurs et **IPS**. Les coefficients étaient particulièrement élevés mais négatifs ( $-3.02801885^{14}$  et  $-3.02801885^{14}$ ).

Finalement, nous avons fait une régression linéaire. Mais cette fois-ci, l'IPS est la donnée expliquée. Le coefficient le plus haut parmi les données explicatives est celui des professeurs qui ont fait une Classe préparatoire ECS ou équivalent. Tandis que le plus faible est celui des professeurs qui ont suivi une Faculté de géographie – niveau licence ou équivalent. Ces résultats sont particulièrement intéressants puisque cela montre que lorsque le niveau d'étude des professeurs est haut, l'IPS aussi. Tandis que lorsque le niveau d'étude des professeurs est plutôt bas, l'IPS baisse. Cela peut aller à l'inverse de la pensée commune. Effectivement, on aurait tendance à penser que les professeurs ayant un plus fort niveau d'étude sont placés dans les lycées les moins favorisés.

### 1.4.3 Approximations faites

Par manque d'informations, des approximations ont dû être faites durant l'étude. Effectivement, certains professeurs n'avaient renseigné que le nom de leur académie. Dans ce cas là, nous mettons l'IPS moyen en voie GT de l'académie.

## 2 Corps de l'étude