STID - IUT Université Paris Descartes **Modèle linéaire**
2AFA
2020-2021

**Feuille de TD 1 : Régression linéaire simple**

## Exercice 1 *Le ressort*

Une approche tant expérimentale que théorique tend à montrer que l'allongement d'un ressort est directement proportionnel à la force appliquée à ce ressort. Ceci peut s'exprimer par la relation $L = kF$ où $L$ représente l'allongement du ressort, $F$ la force appliquée et $k$ le coefficient de raideur du ressort. Pour déterminer la raideur d'un ressort particulier, 10 expériences (considérées comme indépendantes) sont pratiquées en TP :

| $N^o$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $x_i$ | 1 | 5 | 10 | 15 | 20 | 25 | 30 | 35 | 40 | 50 |
| $y_i$ | 2 | 9,9 | 20 | 30,2 | 39,9 | 50 | 60,2 | 68,9 | 80,3 | 99,6 |

où $x_i$ représente la force appliquée lors de la $i$ème expérience et $y_i$ l'allongement obtenu. On donne :

$$\sum_{i=1}^{10} x_i = 231, \quad \sum_{i=1}^{10} y_i = 461, \quad \sum_{i=1}^{10} (x_i - \bar{x})^2 = 2\,264,9$$

$$\sum x_i y_i = 15\,162, \quad \sum_{i=1}^{10} (y_i - \bar{y})^2 = 8\,993,46, \quad \sum_{i=1}^{10} (x_i - \bar{x})(y_i - \bar{y}) = 4\,512,9.$$

Après avoir représenté sur un graphique les différentes mesures et ajusté la droite des moindres carrés au nuage de points, l'expérimentateur affirme que le coefficient de raideur du ressort est égal à 2. On cherche à confirmer ou infirmer cette valeur au risque 5%.

1. Reproduire l'expérience de l'expérimentateur : représenter sur un graphique les différentes mesures, tracer approximativement la droite des moindres carrés et calculer graphiquement sa pente.
2. Calculer les estimateurs des moindres carrés $a$ et $b$ de $\alpha$ et $\beta$ et l'estimateur $s^2$ de la variance $\sigma^2$ dans le modèle de régression linéaire où l'allongement du ressort est la variable d'intérêt et la force appliquée la variable explicative.
3. Donner, au niveau de confiance 95%, des intervalles de confiance pour $\alpha$ et $\beta$.
4. Tester le caractère significatif de la relation linéaire entre l'allongement du ressort et la force appliquée au ressort. Que peut-on en conclure ?

**Exercice 2** ***Rendement du blé***

Pour étudier la relation qui peut exister entre le rendement de blé et la quantité d'engrais utilisée dans les parcelles d'une région donnée, des agronomes recueillent sur $n = 7$ parcelles, la quantité d'engrais $x_i$ (en quintaux) utilisée sur la $i$ème parcelle, et le rendement de blé $y_i$ (en kg) mesuré correspondant sur la $i$ème parcelle. Les données sont les $n = 7$ couples $(x_i, y_i)$ et sont reportées dans le tableau ci-dessous

| $x_i$ | 100 | 200 | 300 | 400 | 500 | 600 | 700 |
|---|---|---|---|---|---|---|---|
| $y_i$ | 40 | 50 | 50 | 70 | 65 | 65 | 80 |

On donne :

$$\sum_{i=1}^{n} x_i = 2\,800, \quad \sum_{i=1}^{n} y_i = 420, \quad \sum x_i^2 = 1\,400\,000,$$
$$\sum x_i y_i = 184\,500, \quad \sum y_i^2 = 26\,350.$$

1. Ecrire le modèle linéaire permettant d'étudier le lien qui existe entre le rendement de blé et la quantité d'engrais utilisée. Préciser les hypothèses de modélisation.
2. Calculer les estimateurs des moindres carrés $a$ et $b$ de $\alpha$ et $\beta$ et l'estimateur $s^2$ de la variance $\sigma^2$.
3. Donner, au niveau de confiance 95%, des intervalles de confiance pour $\alpha$ et $\beta$.
4. Tester le caractère significatif de la relation linéaire entre le rendement de blé et la quantité d'engrais utilisée. Que peut-on en conclure ?
5. Construire une prévision du rendement de blé que l'on obtiendrait sur une parcelle où la quantité d'engrais utilisée serait de 450 kg. Donner l'intervalle de prévision associé au niveau de confiance 95%.
6. Vérifier les calculs en utilisant `R`.

**Exercice 3** ***Fréquence cardiaque***

La "fréquence seuil" d'un sportif amateur est sa fréquence cardiaque mesurée après 3/4 d'heure d'un effort soutenu de course à pied. Dans cette étude, on cherche à savoir si l'âge d'un sportif amateur a une influence sur sa fréquence seuil. On dispose pour cela de $n = 20$ couples de mesures $(x_i, y_i)$ où $x_i$ est l'âge et $y_i$ la fréquence seuil. On obtient les résultats suivants :

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = 1991, \quad \sum_{i=1}^{n}(y_i - \bar{y})^2 = 189,2, \quad \sum(x_i - \bar{x})(y_i - \bar{y}) = -195,4$$
$$\sum(y_i - \hat{y}_i)^2 = 170, \quad \bar{x} = 35,6, \quad \bar{y} = 170,2.$$

1. Indiquer la variable explicative et la variable à expliquer.
2. Calculer les estimateurs des moindres carrés $a$ et $b$ de $\alpha$ et $\beta$ et l'estimateur $s^2$ de la variance $\sigma^2$.
3. Calculer une estimation de la variance de $a$.
4. Calculer le coefficient de détermination $R^2$. Commenter la qualité de l'ajustement des données au modèle.
5. Tester l'hypothèse $(H_0) \; : \; \alpha = 0$ contre l'hypothèse $(H_1) \; : \; \alpha \neq 0$ pour un risque de 5%. Conclure sur la question de l'influence de l'âge sur la fréquence seuil.
6. Donner, au niveau de confiance 99%, des intervalles de confiance pour $\alpha$ et $\beta$.

**Exercice 4** ***Plus théorique I***

On suppose que l'on est dans le cas d'un modèle de régression linéaire simple de la forme

$$Y_i = \alpha x_i + \beta + \varepsilon_i \quad \text{où} \quad \forall i = 1 \dots n \quad \varepsilon_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2). \tag{1}$$

On suppose connaître $\beta$, mais pas $\alpha$. Le but de cet exercice est donc de trouver un estimateur de $\alpha$ sachant que $\underline{\beta \text{ est connu.}}$

1. Calculer $\mathbb{E}(Y_i)$ et en déduire la dimension du modèle.
2. Rappeler la définition du critère des moindres carrés.
3. Montrer que la fonction obtenue à la question 2 admet un minimum.
4. En dérivant selon $\alpha$ la fonction obtenue à la question 2, montrer que dans ce cas l'estimateur des moindres carrés de $\alpha$ est
$$a = \frac{\sum_{i=1}^{n} x_i (Y_i - \beta)}{\sum_{i=1}^{n} x_i^2} \tag{2}$$
5. A l'aide des équations (1) et (2) montrer que
$$a = \alpha + \frac{\sum_{i=1}^{n} x_i \varepsilon_i}{\sum_{i=1}^{n} x_i^2}$$
6. En déduire que $a$ est un estimateur sans biais de $\alpha$.
7. Montrer que
$$Var(a) = \frac{\sigma^2}{\sum_{i=1}^{n} x_i^2}$$
8. En déduire que $Var(a)$ est inférieure ou égale à la variance de l'estimateur des moindres carrés de $\alpha$ dans le cas “classique”. Commenter ce résultat.

**Exercice 5** ***Plus théorique II***

On suppose que l'on est dans le cas d'un modèle de régression linéaire simple de la forme

$$Y_i = \alpha x_i + \beta + \varepsilon_i \quad \text{où} \quad \forall i = 1 \dots n \quad \varepsilon_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2). \tag{3}$$

On suppose connaître $\alpha$, mais pas $\beta$. Le but de cet exercice est donc de trouver un estimateur de $\beta$ sachant que $\underline{\alpha \text{ est connu.}}$

1. Calculer $\mathbb{E}(Y_i)$ et en déduire la dimension du modèle.
2. Rappeler la définition du critère des moindres carrés.
3. Montrer que la fonction obtenue à la question 2 admet un minimum.
4. En dérivant selon $\beta$ la fonction obtenue à la question 2, montrer que dans ce cas l'estimateur des moindres carrés de $\beta$ est
$$b = \bar{Y} - \alpha \bar{x} \tag{4}$$
5. A l'aide des équations (3) et (4) montrer que
$$b = \beta + \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i$$
6. En déduire que $b$ est un estimateur sans biais de $\beta$.
7. Montrer que
$$Var(b) = \frac{\sigma^2}{n}$$
8. En déduire que $Var(b)$ est inférieure ou égale à la variance de l'estimateur des moindres carrés de $\beta$ dans le cas "classique". Commenter ce résultat.

**Exercice 6** ***Plus théorique III***

On suppose que l'on est dans le cas d'un modèle de la forme

$$Y_i = \beta + \varepsilon_i \quad \text{où} \quad \forall i = 1 \dots n \quad \varepsilon_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2). \tag{5}$$

Le but de cet exercice est de trouver et d'étudier les propriétés d'un estimateur de $\beta$.

1. Calculer $\mathbb{E}(Y_i)$ et en déduire la dimension du modèle.
2. Rappeler la définition du critère des moindres carrés.
3. Montrer que la fonction obtenue à la question 2 admet un minimum.
4. En dérivant selon $\beta$ la fonction obtenue à la question 2, montrer que dans ce cas l'estimateur des moindres carrés de $\beta$ est
$$b = \bar{Y} \tag{6}$$

5. A l'aide des équations (5) et (6) montrer que

$$b = \beta + \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i$$

6. En déduire que $b$ est un estimateur sans biais de $\beta$.
7. Montrer que

$$Var(\hat{B}) = \frac{\sigma^2}{n}.$$

8. En déduire que $Var(b)$ est inférieure ou égale à la variance de l'estimateur des moindres carrés de $\beta$ dans le cas "classique". Commenter ce résultat.