Économétrie 1, M1 MAEF, Partiel
17/01/2022, Durée : deux heures
Documents et téléphones portables ne sont pas autorisés.

Nom :	**Prénom :**

Exercice 1.

Considérons le modèle linéaire simple $Y_i = \mu + \beta X_i + \varepsilon_i$, $i = 1, \ldots, n$ où ε_i vérifie l'hypothèse $\{\varepsilon_i, i = 1, \ldots, n\} \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$.

On a lancé les codes suivants :

```
X = c(10, 15, 20, 25, 30); Y = c(29, 44, 53, 60, 70); summary(lm(Y~X))
```

Voici les résultats :

```
Residuals:
  1    2    3    4    5
-2.6  2.6  1.8 -1.0 -0.8
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  12.0000     3.3226   3.612   0.0365 *
X             1.9600     0.1566  12.513   0.0011 **
Residual standard error: 2.477 on 3 degrees of freedom
Multiple R-squared:  0.9812,Adjusted R-squared:  0.9749
F-statistic: 156.6 on 1 and 3 DF,  p-value: 0.0011
```

Pouvez vous remplir le tableau suivant?

1.	$\hat{\mu} =$	12
2.	$\hat{\beta} =$	1,96
3.	$\widehat{\sigma^2} =$	2.477^2
4.	$\text{Var}(\hat{\mu}) =$	$3{,}3226^2$
5.	$\text{Var}(\hat{\beta}) =$	$0{,}1566^2$

Nous pouvons obtenir la valeur numérique 0.0011 en utilisant les autres résultats de la sortie en R. Écrire le code R qui renvoie 0.0011 en utilisant les fonctions pt() et pf().

6.	code R avec pt() :	pt(12.513, 3, lower.tail=F)*2
7.	code R avec pf() :	pf(156.6, 1, 3, lower.tail=F)

8. Est il surprenant que la somme des résidus vaut zéro, pourquoi ?

$\bar{\hat{\varepsilon}} = \overline{Y - \hat{Y}} = \bar{Y} - \overline{(\hat{\mu} + \hat{\beta} X)} = \bar{Y} - (\hat{\mu} + \hat{\beta}\bar{X}) = \bar{Y} - \hat{\beta}\bar{X} - \hat{\mu} = \hat{\mu} - \hat{\mu}$
$= 0$

Oui.

Exercice 2.

On a lancé les codes suivants.

```
Y = c(10, 15, 20, 25, 30); X1 = c(29, 44, 53, 60, 70); X2 = c(29, 39, 49, 61, 66);
summary(lm(Y~X1+X2))
```

Les sorties sont présentées dessous.

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -5.4937     1.6533  -3.323   0.0799 .
X1            0.1949     0.1804   1.081   0.3928
X2            0.3179     0.1848   1.721   0.2275
Residual standard error: 0.9734 on 2 degrees of freedom
Multiple R-squared:  0.9924,Adjusted R-squared:  0.9848
F-statistic: 130.9 on 2 and 2 DF,  p-value: 0.00758
```

1. Observez vous des anomalies ? Pouvez vous les expliquer ?

"colinéarité" : Parce que X_1 et X_2 sont fortement corrélées

Oui, anomalies : aucune variable est significative (0,3928 > 0,05, 0,2275 > 0,05)

le modèle est globalement significatif (0.00758 < 0,05)

On tape maintenant les codes suivants.

```
Y = Y-mean(Y); X1 = X1-mean(X1); X2 = X2-mean(X2); c = coefficients(lm(Y~X1+X2)
```

Pouvez vous prédire les valeurs des sorties des commandes suivantes?

2.	`c[1]` ≈	0
3.	`c[2]` ≈	0,1949
4.	`c[3]` ≈	0,3179

Exercice 3.

On dispose d'un échantillon $(Y_1, \ldots, Y_n)$ de variables aléatoires réelles dans lequel Y_i est défini par la relation

$$Y_i = \frac{\mu}{p_i} + \varepsilon_i$$

où $(\varepsilon_1, \ldots, \varepsilon_n)$ est une suite des variables aléatoires i.i.d. de loi $\mathcal{N}(0, \sigma^2)$ avec σ^2 inconnu et $(p_1, \ldots, p_n)$ sont des nombres réels non nuls supposés connus. On cherche à construire des estimateurs de μ et σ^2

1. On note $M = \frac{1}{n}\sum_{i=1}^{n} p_i Y_i$. Montrer que M est un estimateur sans biais de μ et calculer la variance de cet estimateur. Quelle est la loi de M?

2. Rappelons que $P_{[v]}(u) = \frac{<u,v>}{\|v\|^2} \cdot v$ et $P_{[v]^\perp}(u) = u - P_{[v]}(u)$. On note E le sous-espace vectoriel de $\mathbb{R}^n$ engendré par le vecteur $e = (\frac{1}{p_1}, \ldots, \frac{1}{p_n})^T$. Calculer $P_E(Y)$ et $P_{E^\perp}(Y)$.

3. On note ν le vecteur $(\frac{\mu}{p_1}, \ldots, \frac{\mu}{p_n})^T$. Calculer $P_{E^\perp}(\nu)$.

4. Déduire des questions précédentes et du théorème de Cochran la loi de $\|P_{E^\perp}(Y - \nu)\|^2$. Proposer un estimateur sans biais de σ^2. Calculer la variance de cet estimateur. Comment s'écrit l'estimateur lorsque $p_1 = \cdots = p_n = 1$?

5. Calculer les estimateurs de μ et σ^2 en utilisant la méthode des moindres carrés, c.a.d. $\hat{\mu} = (Z^T Z)^{-1} Z^T Y$ et $\widehat{\sigma^2} = \text{SCR}/(n-k)$. Comparer les espérances et variances de $\hat{\mu}$ et $\widehat{\sigma^2}$ avec celles des estimateurs obtenus dans les questions **1** et **4**.

Exo 3. (17/01)

1. $\mathbb{E}M = \frac{1}{n}\sum P_i \mathbb{E}Y_i = \frac{1}{n}\sum P_i \cdot \frac{\mu}{P_i} = \mu \Rightarrow$ sans biais

$\mathbb{V}M = \frac{1}{n^2}\sum P_i^2 \mathbb{V}Y_i = \frac{1}{n^2}\sum P_i^2 \cdot \sigma^2 = \frac{\sigma^2}{n^2}\sum P_i^2$

La loi de M est $\mathcal{N}(\mu, \frac{\sigma^2}{n^2}\sum P_i^2)$

2. $\langle Y, e\rangle = \sum \frac{Y_i}{P_i}$, $\|e\|^2 = \sum \frac{1}{P_i^2}$, $P_E(Y) = \frac{\langle Y, e\rangle}{\|e\|^2}\cdot e = \frac{\sum Y_i/P_i}{\sum 1/P_i^2}\left(\frac{1}{P_1}, \cdots, \frac{1}{P_n}\right)^T$

$P_{E^\perp}(Y) = Y - P_E(Y) = (Y_1, \cdots, Y_n)^T - \frac{\sum Y_i/P_i}{\sum 1/P_i^2}\left(\frac{1}{P_1}, \cdots, \frac{1}{P_n}\right)^T$

3. $P_E(\gamma) = \frac{\sum \mu/P_i^2}{\sum 1/P_i^2}\left(\frac{1}{P_1}, \cdots, \frac{1}{P_n}\right)^T = \mu\left(\frac{1}{P_1}, \cdots, \frac{1}{P_n}\right)^T$

$P_{E^\perp}(\gamma) = \gamma - P_E(\gamma) = (0, \cdots, 0)^T$

4. Puisque $Y - \gamma \sim \mathcal{N}(\vec{0}, \sigma^2 I)$, $\|P_{E^\perp}(Y-\gamma)\|^2 \sim \sigma^2\chi^2(k)$ avec $k = \dim(E^\perp) = n-1$. Donc la loi de $\|P_{E^\perp}(Y-\gamma)\|^2$ est $\sigma^2 \cdot \chi^2(n-1)$.

$\|P_{E^\perp}(Y-\gamma)\|^2 = \|P_{E^\perp}Y - P_{E^\perp}\gamma\|^2 \overset{P_{E^\perp}(\gamma) = \vec{0}}{=\!=\!=} \|P_{E^\perp}Y\|^2$

$\mathbb{E}\|P_{E^\perp}Y\|^2 = \mathbb{E}\|P_{E^\perp}(Y-\gamma)\|^2 = \mathbb{E}(\sigma^2 \cdot \chi^2(n-1)) = \sigma^2(n-1)$

$\frac{\|P_{E^\perp}Y\|^2}{n-1}$ est un estimateur sans biais de σ^2.

Notons $\hat{\sigma}^2 = \frac{\|P_{E^\perp}Y\|^2}{n-1}$. $\mathbb{V}\hat{\sigma}^2 = \frac{1}{(n-1)^2}\mathbb{V}(\sigma^2\chi^2(n-1)) = \frac{2\sigma^4}{n-1}$

Lorsque $P_1 = \cdots = P_n = 1$, on a

$\hat{\sigma}^2 = \left((Y_1 - \frac{\sum Y_i}{n})^2 + \cdots + (Y_n - \frac{\sum Y_i}{n})^2\right)\frac{1}{n-1} = \frac{\|Y - \bar{Y}\|^2}{n-1}$

5. Le modèle $Y_i = \frac{\mu}{p_i} + \varepsilon_i$ peut s'écrire en forme matricielle

$$Y = Z\mu + \varepsilon \quad \text{avec } Z = \left(\frac{1}{p_1}, \cdots, \frac{1}{p_n}\right)^T.$$

On a $Z^T Z = \sum \frac{1}{p_i^2}$, $Z^T Y = \sum \frac{Y_i}{p_i}$, $(Z^T Z)^{-1} Z^T Y = \frac{\sum Y_i / p_i}{\sum 1/p_i^2} = \hat{\mu}$.

On a $SCR = \|Y - \hat{Y}\|^2 = \|Y - Z\hat{\mu}\|^2 = \|Y - \left(\frac{1}{p_1}, \cdots, \frac{1}{p_n}\right)^T \cdot \frac{\sum Y_i/p_i}{\sum 1/p_i^2}\|^2 = \|P_{E^\perp}(Y)\|^2$.

Donc $\hat{\sigma}^2$ et $\tilde{\sigma}^2$ sont identiques.

$\mathbb{E}\hat{\mu} = \mu$, $\mathbb{V}\hat{\mu} = \frac{\sigma^2}{\sum 1/p_i^2}$

Si $p_1 = \cdots = p_n = p$, alors $\mathbb{V}M = \mathbb{V}\hat{\mu} = \frac{\sigma^2 \cdot p^2}{n}$.

Sinon $\frac{\mathbb{V}M}{\mathbb{V}\hat{\mu}} = \frac{\sum p_i^2 \cdot \sum 1/p_i^2}{n^2} = \frac{n + \sum_{i \neq j} p_i^2/p_j^2}{n^2} > 1$, si $\sum_{i\neq j} p_i^2/p_j^2 > A_n^2$.

Exo3 (14/02)

1. $\mathbb{E}(M) = \frac{1}{n}\sum \frac{\mu p_i}{p_i} = \mu \Rightarrow$ sans biais. $\mathbb{V}M = \frac{\sigma^2}{n^2}\sum \frac{1}{p_i^2}$. $M \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n^2}\sum\frac{1}{p_i^2}\right)$

2. $P_E(X) = \frac{\sum p_i X_i}{\sum p_i^2}(p_1, \cdots, p_n)^T$. $P_{E^\perp}(X) = (X_1, \cdots, X_n)^T - \frac{\sum p_i X_i}{\sum p_i^2}(p_1, \cdots, p_n)^T$.

3. $P_{E^\perp}(\gamma) = \vec{0}$.

4. $\|P_{E^\perp}(X - \gamma)\|^2 \sim \sigma^2 \cdot \chi^2(n-1)$

$\tilde{\sigma}^2 = \frac{\|P_{E^\perp} X\|^2}{n-1}$ est un estimateur sans biais de σ^2.

$\mathbb{V}\tilde{\sigma}^2 = \frac{2\sigma^4}{n-1}$. Lorsque $p_1 = \cdots = p_n = p$, on a $\tilde{\sigma}^2 = \frac{\|X - \bar{X}\|^2}{n-1}$.

5. $\hat{\mu} = \frac{\sum p_i X_i}{\sum p_i^2}$, $\hat{\sigma}^2 = \tilde{\sigma}^2$

$\mathbb{E}\hat{\mu} = \mu$, $\mathbb{V}\hat{\mu} = \frac{\sigma^2}{\sum p_i^2}$

Si $p_1 = \cdots = p_n = p$, alors $\mathbb{V}M = \mathbb{V}\hat{\mu} = \frac{\sigma^2}{np^2}$

Sinon $\frac{\mathbb{V}M}{\mathbb{V}\hat{\mu}} > 1$ si $\sum_{i\neq j} p_i^2/p_j^2 > A_n^2$

Économétrie 1, M1 MAEF, Partiel
14/02/2022, Durée : deux heures
Documents et téléphones portables ne sont pas autorisés.

Nom :	**Prénom :**

Exercice 1.
Considérons le modèle gaussien

$$Y_{ij} = \mu_i + \varepsilon_{ij} \quad \text{avec} \quad i = 1, 2, 3 \quad \text{et} \quad j = 1, \ldots, 6,$$

où $\mathbb{E}(\varepsilon_{ij}) = 0$, $\mathbb{V}(\varepsilon_{ij}) = \sigma^2$, $\forall (i,j)$ et les ε_{ij} sont indépendants.
On a lancé les codes suivants :

```
Y = c(23.4, 24.4, 24.6, 24.9, 25.0, 26.2, 22.5, 22.9, 23.7, 24.0, 24.4, 18.9,
 21.1, 21.2, 22.1, 22.5, 23.5, 24.5)
I = c(rep(1,6), rep(2,6), rep(3,6)); I = as.factor(I)
summary(lm(Y~I)); anova(lm(Y~I))
```

Voici les résultats :

```
Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)    24.7500     0.6060  40.838   <2e-16 ***
I2             -2.0167     0.8571  -2.353   0.0327 *
I3             -2.2667     0.8571  -2.645   0.0184 *

Response: Y
          Df Sum Sq Mean Sq F value  Pr(>F)
I          2 18.534  9.2672  4.2052 0.03549 *
Residuals 15 33.057  2.2038
```

1. Que signifient elles les sorties numériques 24.75, -2.0167, $< 2e-16$ et 0.0184 ?

1)	24.75	$\hat{\mu}_1$
2)	-2.0167	$\hat{\mu}_2 - \hat{\mu}_1$
3)	$< 2e-16$	P-value pour tester entre $H_0: \mu_1 = 0$ et $H_1: \mu_1 \neq 0$
4)	0.0184	P-value pour tester entre $H_0: \mu_3 = \mu_1$ et $H_1: \mu_3 \neq \mu_1$

2. Que signifient elles les valeurs numériques 18.53, 33.057 et 0.03549 ?

1)	18.53	$SCE = \sum_1^3 \sum_1^6 (Y_{i\cdot} - Y_{\cdot\cdot})^2$: intergroupe
2)	33.057	$SCR = \sum_1^3 \sum_1^6 (Y_{ij} - Y_{i\cdot})^2$: intra groupe
3)	0.03549	P-value pour testr entre $H_0: \mu_1 = \mu_2 = \mu_3$ et H_1 : au moins $\exists \mu_i$ t.q. $\mu_i \neq \mu_j$, $i \neq j$

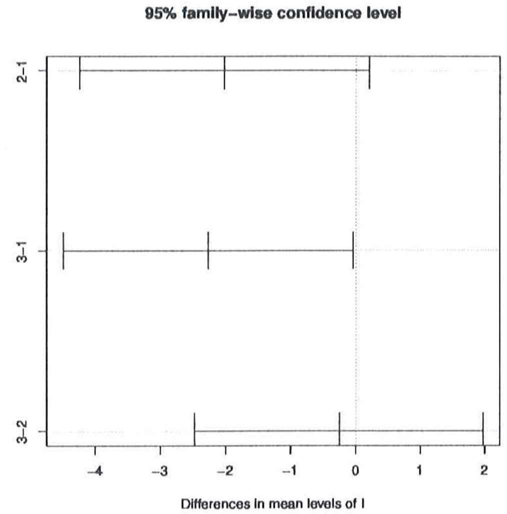
3. Pouvez vous prédire la valeur de la commande `sum((Y-mean(Y))^2)` ?

$18{,}534 + 33{,}057 = 51{,}591$

4. Nous pouvons obtenir la valeur numérique 0.03549 en utilisant la fonction `pf()` et les autres résultats de la sortie en R. Écrire le code R qui renvoie 0.03549 en utilisant `pf()`.

$1 - pf\left(\frac{18{,}534/2}{33.057/15}, 2, 15\right)$

5. On a tapé la commande `plot(TukeyHSD(aov(Y~I)))`. La sortie est présentée ci dessous. Pouvez vous commenter ce graphique ? Est il cohérent avec les sorties numériques ?



μ_1 et μ_3 sont significativement différents. μ_1 et μ_2 ne sont pas, μ_2 et μ_3 ne sont pas. Il est cohérent avec les sorties numériques : $0{,}0184 < 0{,}05 \Rightarrow \mu_1 \neq \mu_3$

Exercice 2.

On considère un échantillon $(X_1, \dots, X_n)$ de variables aléatoires réelles indépendantes dans lequel X_i suit une loi $\mathcal{N}(\mu p_i, \sigma^2)$ avec μ et σ^2 des paramètres inconnus et $(p_1, \dots, p_n)$ des nombres réels non nuls supposés connus. On cherche à construire des estimateurs de μ et σ^2

1. On note $M = \frac{1}{n}\sum_{i=1}^{n} \frac{X_i}{p_i}$. Montrer que M est un estimateur sans biais de μ et calculer la variance de cet estimateur. Quelle est la loi de M?

2. Rappelons que $P_{[v]}(u) = \frac{<u,v>}{\|v\|^2} \cdot v$ et $P_{[v]^\perp}(u) = u - P_{[v]}(u)$. On note E le sous-espace vectoriel de $\mathbb{R}^n$ engendré par le vecteur $e = (p_1, \dots, p_n)^T$. Calculer $P_E(Y)$ et $P_{E^\perp}(Y)$.

3. On note ν le vecteur $(\mu p_1, \dots, \mu p_n)^T$. Calculer $P_{E^\perp}(\nu)$.

4. Déduire des questions précédentes et du théorème de Cochran la loi de $\|P_{E^\perp}(Y - \nu)\|^2$. Proposer un estimateur sans biais de σ^2. Calculer la variance de cet estimateur. Comment s'écrit l'estimateur lorsque $p_1 = \dots = p_n = p$?

5. Calculer les estimateurs de μ et σ^2 en utilisant la méthode des moindres carrés, c.a.d. $\hat{\mu} = (Z^T Z)^{-1} Z^T Y$ et $\widehat{\sigma^2} = \mathrm{SCR}/(n-k)$ pour le modèle $X_i = \mu p_i + \varepsilon_i$. Comparer les espérances et variances de $\hat{\mu}$ et $\widehat{\sigma^2}$ avec celles des estimateurs obtenus dans les questions **1** et **4**.