Université Paris I, Panthéon - Sorbonne

Première Année Master M.A.E.F. 2016 – 2017

Econométrie I

Contrôle continu n°2, 13 décembre 2016

Examen de 1h30. Tout document ou calculatrice est interdit.

Exercice 1 (Sur 12 points)

On effectue une étude de marketing sur l'impact du flacon sur la vente d'un parfum. On conçoit donc I différents flacons et pour chaque flacon on demande à n individus différents de noter (de 0 à 10) l'appréciation du parfum. On notera N_{ik} la note obtenue pour le flacon i par le k-ème individu.

1. Si $(x_1, \ldots, x_m)$ est une famille de m réels, que vaut $\widehat{a} = \text{argmin}_{a \in \mathbf{R}} \sum_{j=1}^{m} (x_j - a)^2$ **(1pt)** ?
2. On suppose que l'on peut écrire que $N_{ik} = \mu_i + \varepsilon_{ik}$ pour tout $i = 1, \ldots, I$ et $k = 1, \ldots, n$ où on suppose que les erreurs (ε_{ik}) forment une famille de v.a.i.i.d. centrées de variance σ^2. Montrer que l'estimation par moindres carrés des μ_i revient à minimiser $\sum_{i=1}^{I} \sum_{k=1}^{n} (N_{ik} - \mu_i)^2$ **(0.5pts)** et en déduire que pour tout $i = 1, \ldots, I$, $\widehat{\mu}_i = \frac{1}{n} \sum_{k=1}^{n} N_{ik}$ **(0.5pts)**. Déterminer en fonction de $\widehat{\mu}_i$, σ^2 et n, un intervalle de confiance asymptotique à 95% pour μ_i **(2pts)**.
3. Expliciter un estimateur non biaisé $\widehat{\sigma}^2$ de σ^2, en fonction des N_{ik}, des $\widehat{\mu}_i$, de n et de I **(1pt)**. En déduire (justifier !) un intervalle de confiance asymptotique à 95% pour μ_i en fonction de $\widehat{\mu}_i$, $\widehat{\sigma}^2$ et n **(2pts)**.
4. On veut tester si le flacon a un réel impact sur l'appréciation du parfum. Expliquer pourquoi cela revient à tester $H_0 : \mu_1 = \mu_2 = \cdots = \mu_I$ contre sa contraposée H_1 **(0.5pts)**. On note N le vecteur colonne tel que $N = {}^t\big(N_{11}, \ldots, N_{1n}, N_{21}, \ldots, N_{In}\big)$. Écrire le modèle sous forme matricielle, en notant $\mu = {}^t(\mu_1, \ldots, \mu_I)$ **(0.5pts)**. Déterminer une statistique de Fisher $\widehat{F}$ permettant de réaliser ce test **(1pt)**. Montrer qu'asymptotiquement $\widehat{F} \xrightarrow[n \to \infty]{\mathcal{L}} \frac{1}{I-1} \chi^2(I-1)$ **(3pts)**.

Exercice 2 (Sur 7 points)

On considère le modèle de régression suivant :

$$y_i = \beta_0 + \beta_1 x_{i,1} + \beta x_{i,2} + \varepsilon_i, \quad 1 \le i \le n,$$

les $x_{i,j}$ étant des variables explicatives observées du modèle et les ε_i des v.a.i.i.d. de loi $\mathcal{N}(0, \sigma^2)$. On note et on calcule :

$$X = \begin{bmatrix} 1 & x_{1,1} & x_{1,2} \\ \vdots & \vdots & \vdots \\ 1 & x_{n,1} & x_{n,2} \end{bmatrix}, \quad Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \quad \Longrightarrow \quad {}^tXX = \begin{bmatrix} 30 & 20 & 0 \\ 20 & 20 & 0 \\ 0 & 0 & 10 \end{bmatrix}, \quad {}^tXY = \begin{bmatrix} 15 \\ 20 \\ 10 \end{bmatrix}, \quad \text{et} \quad {}^tYY = 59.5.$$

1. Déterminer n, la moyenne de $(x_{i,2})_i$, le coefficient de corrélation des $(x_{i,1})_i$ et des $(x_{i,2})_i$ **(1.5pts)**.
2. Calculer numériquement les estimateurs par moindres carrés ordinaires $\widehat{\theta}$ et $\widehat{\sigma}^2$ de $\theta = {}^t(\beta_0, \beta_1, \beta_2)$ et de σ^2 (on montrera que $\|Y - X\widehat{\theta}\|^2 = \|Y\|^2 - \|X\widehat{\theta}\|^2$) **(2pts)**.
3. Calculer pour β_1 un intervalle de confiance à 95% (on utilisera des valeurs approchées des quantiles d'une loi de Student) **(1pt)**. Tester également l'hypothèse $\beta_2 = 0.8$ au niveau 10% **(1pt)**.
4. Déterminer la moyenne empirique des y_i : $\bar{y}_n$ et en déduire le coefficient de détermination R^2 **(1.5pt)**.

Exercice 3 (Sur 3 points)

Considérons le jeu de données suivant : $X = (1, 2, 3, 4, 5)$, $Y = (9, 13, 2, 8, 1)$, $Z = (3, 4, 5, 6, 8)$.

1. Considérons le modèle à deux variables explicatives où Z est expliqué par X et Y. Quelles commandes faut-il taper dans `R` pour obtenir le résultat de la Figure 1 **(0.5pts)** ?
2. Quel est le modèle estimé en utilisant les résultats de la Figure 1 **(0.5pts)** ? Ce modèle est-il cependant statistiquement satisfaisant ? Que doit-on faire ? **(1pt)**
3. Calculer et comparer les coefficients de détermination partielle associés à X et à Y respectivement, à partir des Figure 2 et 3 et conclure **(1pt)**.

```
Call:
lm(formula = Z ~ X + Y)

Residuals:
      1       2       3       4       5 
 0.1559  0.1054 -0.3128 -0.3142  0.3657 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)  
(Intercept)  1.91629    0.85672   2.237   0.1548  
X            1.14851    0.18001   6.380   0.0237 *
Y           -0.02452    0.05659  -0.433   0.7071  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4276 on 2 degrees of freedom
Multiple R-squared:  0.9753,	Adjusted R-squared:  0.9506 
F-statistic: 39.47 on 2 and 2 DF,  p-value: 0.02471
```

Figure 1 –

```
Analysis of Variance Table

Response: Z
          Df Sum Sq Mean Sq F value Pr(>F)
Y          1 6.9917  6.9917  2.6863 0.1997
Residuals  3 7.8083  2.6028
```

```
Analysis of Variance Table

Response: Z
          Df Sum Sq Mean Sq F value   Pr(>F)   
X          1   14.4 14.4000     108 0.001901 **
Residuals  3    0.4  0.1333                    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 2 –

```
Analysis of Variance Table

Response: Z
          Df  Sum Sq Mean Sq F value  Pr(>F)  
X          1 14.4000 14.4000 78.7586 0.01246 *
Y          1  0.0343  0.0343  0.1877 0.70706  
Residuals  2  0.3657  0.1828                  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 3 –