Université Paris I, Panthéon - Sorbonne

Première Année Master M.A.E.F. 2021 – 2022

Econométrie II

Examen final, mai 2022

Examen de 2h00. Tout document ou calculatrice est interdit.

1. **(18 points)** Soit $Y_i,\ i = 1, \cdots, N$ le chiffre d'affaire de N magasins de même taille appartenant au même groupe. On s'intéresse à l'impact sur ce chiffre d'affaire de 2 variables: $Z^{(1)}$ qui vaut 1 si le magasin est en centre-ville et 0 sinon et $Z^{(2)}$ qui vaut 1 si le magasin est ouvert le dimanche et 0 sinon.

   (a) Lors d'une première étude, on choisit les magasins de telle manière que pour $i = 1, \cdots, n_1$, les magasins sont tous au centre-ville et fermés le dimanche, et pour $i = n_1 + 1, \cdots, N$, les magasins sont tous ouverts le dimanche sans qu'aucun ne soit au centre-ville. On pose le modèle:

   $$Y_i = \theta_1 Z_i^{(1)} + \theta_2 Z_i^{(2)} + \varepsilon_i \qquad \text{pour } i = 1, \cdots, N,$$

   où (ε_i) est une suite de variables aléatoires indépendantes gaussiennes centrées de même variance $\sigma^2 > 0$ inconnue.

      i. Déterminer explicitement et non vectoriellement les estimateurs $\widehat{\theta}_1$ et $\widehat{\theta}_2$ de θ_1 et θ_2 par moindres carrés ordinaires et un estimateur $\widehat{\sigma}^2$ de σ^2 **(2pts)**.

      ii. On veut tester H_0: $\theta_1 = 0$. Donner explicitement la statistique du test de Student permettant de tester cette hypothèse **(1pt)**. Quelle est sa loi **(0.5pts)**? Lorsque N est grand, quelle est asymptotiquement la loi de ce test **(0.5pts)**?

      iii. On veut tester si l'effet des 2 variables $Z^{(1)}$ et $Z^{(2)}$ est le même, c'est-à-dire si $\theta_1 = \theta_2$ (H_0'), contre l'hypothèse $\theta_1 > \theta_2$ (H_1'). Pour cela on considère la statistique:

      $$\widehat{T}_N = \frac{\widehat{\theta}_1 - \widehat{\theta}_2}{\sqrt{\frac{N}{n_1(N-n_1)}\,\widehat{\sigma}^2}}.$$

      Déterminer, en justifiant, la loi de $\widehat{T}_N$ **(1.5pts)**. Pour quel seuil franchi par $\widehat{T}_N$ allez-vous choisir H_0' plutôt que H_1' avec un risque de 5% **(0.5pts)**?

      iv. Si on suppose que pour $i = 1, \cdots, n_1$, la variance de ε_i est σ_1^2 et pour $i = n_1 + 1, \cdots, N$, la variance de ε_i est σ_2^2, déterminer explicitement les estimateurs par moindres carrés généralisés de θ_1 et de θ_2. Que remarquez-vous? **(2pts)** Quels sont les estimateurs de σ_1^2 et σ_2^2 par maximum de vraisemblance **(2pts)**? A partir de la loi de $\widehat{\theta}_1 - \widehat{\theta}_2$, en déduire une autre statistique permettant de tester H_0' contre H_1' **(1pt)**.

   (b) Pour chaque magasin i de l'étude décrite en (a), on considère X_i la variable telle que $X_i = 1$ si le magasin a un chiffre d'affaire croissant et $X_i = 0$ sinon. On veut expliquer X_i à l'aide de $Z_i^{(1)}$ et $Z_i^{(2)}$ en utilisant une régression logistique, avec la fonction logit. Présenter le modèle et donner explicitement les estimateurs des paramètres de cette régression **(2pts)**. Si un nouveau magasin numéroté $N+1$ est tel que $Z_{N+1}^{(1)} = 0$ $Z_{N+1}^{(2)} = 1$, comment prédire X_{N+1} **(2pts)**?

   (c) Désormais on a choisi N magasins pour lesquels il est possible d'avoir $Z^{(1)} = p$ et $Z^{(2)} = q$ pour tout $(p, q) \in \{0, 1\}^2$. Ainsi, pour $(p, q) \in \{0, 1\}^2$, on note maintenant $Z^{(pq)}$ la variable telle que $Z_i^{(pq)} = 1$ si $Z_i^{(1)} = p$ et $Z_i^{(2)} = q$, $Z_i^{(pq)} = 0$ sinon. On suppose que:

      - pour $i = 1, \cdots, n_{00}$, $Z_i^{(00)} = 1$ et 0 sinon;
      - pour $i = n_{00} + 1, \cdots, n_{00} + n_{01}$, $Z_i^{(01)} = 1$ et 0 sinon;
      - pour $i = n_{00} + n_{01} + 1, \cdots, n_{00} + n_{01} + n_{10}$, $Z_i^{(10)} = 1$ et 0 sinon;
      - pour $i = n_{00} + n_{01} + n_{10} + 1, \cdots, N = n_{00} + n_{01} + n_{10} + n_{11}$ alors $Z_i^{(11)} = 1$ et 0 sinon.

   On pose le modèle:

   $$Y_i = \theta_{00} Z_i^{(00)} + \theta_{01} Z_i^{(01)} + \theta_{10} Z_i^{(10)} + \theta_{11} Z_i^{(11)} + \varepsilon_i, \qquad \text{pour } i = 1, \cdots, N = n_{00} + n_{01} + n_{10} + n_{11},$$

   où (ε_i) est une suite de variables aléatoires indépendantes gaussiennes centrées de variance $\sigma^2 > 0$. Déterminer les estimateurs par moindres carrés ordinaires des θ_{pq} **(1pt)**. Proposer un test permettant de tester si $\theta_{00} = \theta_{01}$ **(1pt)**. Si on rejette cette hypothèse, que conclure sur l'influence mutuelle des variables $Z^{(1)}$ et $Z^{(2)}$ **(1pt)**?

2. **(10 points)** Exercice de TP utilisant le logiciel `R`

   (a) On a tapé les commandes suivantes:

```
t=c(1:100)
Z1=t
Z2=sqrt(t)*(2+sin(5*t))
Z3=1/t
epsilon=4*rnorm(100)
X=6+0.05*Z1-3*Z2+4*Z3+epsilon
reg=lm(X~Z1+Z2+Z3)
summary(reg)
```

   Voici les résultats:

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.93171    1.30383   5.316 6.88e-07 ***
Z1           0.03387    0.02218   1.527    0.130
Z2          -2.98914    0.09127 -32.750  < 2e-16 ***
Z3          -1.30067    4.55615  -0.285    0.776

Residual standard error: 4.615 on 96 degrees of freedom
Multiple R-squared: 0.9492,     Adjusted R-squared: 0.9476
F-statistic:   598 on 3 and 96 DF,  p-value: < 2.2e-16
```

   Question 1: Qu'a-t-on fait et est-ce satisfaisant? Formellement, expliquer ce que sont les valeurs `0.03387 0.02218 1.527 0.130`*. Pouvait-on s'attendre à la valeur* `-2.98914`*? Comment expliquer la valeur* `0.776`*?* **(3pts)**

   (b) On tape ensuite les commandes:

```
M=cbind(1+0*t,Z1,Z2,Z3)
solve(t(M)%*%M)
```

   Voici les résultats:

```
                             Z1            Z2           Z3
    0.0798087578 -5.180381e-04 -2.637467e-03 -0.163034696
Z1 -0.0005180381  2.310538e-05 -5.419362e-05  0.001432305
Z2 -0.0026374665 -5.419362e-05  3.910900e-04  0.003009123
Z3 -0.1630346962  1.432305e-03  3.009123e-03  0.974554159
```

   Question 2: Quel intérêt d'avoir ainsi effectué ces calculs? A quelle conclusion arrive-t-on? **(2pts)**

   (c) On a ensuite tapé les commandes:

```
Z4=t^2
Z5=runif(100,5,9)
Y=6+0.05*Z1-3*Z2+epsilon
Z=matrix(c(Z1,Z2,Z3,Z4,Z5),ncol=5);
colnames(Z)=c("Z1","Z2","Z3","Z4","Z5");
library(MASS)
ZZ=as.data.frame(Z);
y.lm=lm(Y~.,data=ZZ);
y.bic=stepAIC(y.lm,k=log(100))
```

   Voici les résultats:

```
Start:  AIC=292.72
Y ~ Z1 + Z2 + Z3 + Z4 + Z5

Step:  AIC=288.13
Y ~ Z1 + Z2 + Z4 + Z5

Step:  AIC=283.88
Y ~ Z1 + Z2 + Z5

Step:  AIC=282.39
```

```
Y ~ Z1 + Z2
       Df Sum of Sq     RSS    AIC
<none>               1466.9 282.39
- Z1    1      99.4  1566.3 284.34
- Z2    1   22283.2 23750.1 556.23
```

Question 3: Expliquer ce qui a été fait. Expliquer ce que représente la valeur `284.34`. **(2pts)**

(d) Enfin on a tapé les commandes:

```
zero=0*c(1:100)
W=1+zero
W[Y>-40]=zero[Y>-40]
W
logitW=glm(W~Z1+Z2,family=binomial(link="logit"),na.action=na.pass)
summary(logitW)
library(aod)
wald.test(b=coef(logitW),Sigma=vcov(logitW),Terms=1:2)
```

Voici ce que l'on obtient:

```
W
  [1] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 1 0 0 0 0
 [38] 1 0 0 0 0 1 0 0 0 1 1 0 0 0 1 1 0 0 0 1 1 0 0 0 1 1 0 0 0 1 1 0 0 0 1 1 0
 [75] 0 1 1 1 0 0 1 1 1 0 0 1 1 1 0 0 1 1 1 0 0 1 1 1 0 0

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -1241.031 185243.018  -0.007    0.995
Z1              -3.887    596.555  -0.007    0.995
Z2              86.720  12942.124   0.007    0.995

    Null deviance: 1.2382e+02  on 99  degrees of freedom
Residual deviance: 1.0577e-07  on 97  degrees of freedom

Chi-squared test:
X2 = 4.5e-05, df = 2, P(> X2) = 1.0
```

Question 4: Expliquez ce qui a été fait. Comment expliquer le 0.995 *que l'on retrouve* 3 *fois? Que conclure du test? Pouvait-on s'y attendre?* **(3pts)**