# Partiel TISD - Master 1 Pro
# Jeudi 14 Janvier 2010

*Durée 2 h. Calculatrices autorisées. Seul un formulaire sur feuille double est autorisé.*
**Le barème est prévisionnel**

**Tran Viet Chi**, `chi.tran@univ-lille1.fr`, bureau 316 (bâtiment M3).

**Exercice 1 (Engrais, 16 points)**
Afin de tester l'efficacité d'un nouvel engrais, un biologiste a retenu $n = 30$ plantes, réparties en 3 groupes de tailles inégales. Pour chaque groupe, il teste une dose différente d'engrais. Les résultats sont les suivants :

| Groupe $j$ | Nombre de plantes $n_j$ | Taille moyenne des plantes $\bar{Y}_j$ | Dose moyenne administrée $\bar{X}_j$ |
|---|---|---|---|
| 1 | 12 | 1,7 | 5,8 |
| 2 | 10 | 1,8 | 6,2 |
| 3 | 8 | 1,9 | 6,4 |

Tab. 1 – Résultats des tests du biologiste

où $\bar{Y}_j$ et $\bar{X}_j$ représentent respectivement les moyennes des observations $Y_i$ et $X_i$ pour $i$ appartenant au groupe $j$. On suppose que :

$$Y_i = b_0 + b_1 X_i + \varepsilon_i, \quad i \in \{1, \cdots, n\} \tag{1}$$

où les $\varepsilon_i$ sont supposés être Gaussiens, centrés, de même variance $\sigma^2$ et indépendants.

**1.** Calculer à partir du Tableau 1 les moyennes $\bar{Y}$, $\bar{X}$. Donner des estimations de $\mathrm{Var}(X)$ et $\mathrm{Cov}(X, Y)$ (on négligera les variances intra-groupes).

**2.** En déduire l'estimateur MCO : $\widehat{b}_{MCO} = (\widehat{b}_{0,MCO}, \widehat{b}_{1,MCO})$ de $b$. On rappellera la formule du cours, puis on fera l'application numérique.

**3.** Montrer que (1) implique que :

$$\bar{Y}_j = \beta_0 + \beta_1 \bar{X}_j + \eta_j, \quad j \in \{1, \cdots, 3\} \tag{2}$$

où on précisera $\beta = (\beta_0, \beta_1)$ et $\eta_j$ en fonction de $b = (b_0, b_1)$ et des $\varepsilon_i$. Dans la suite, on notera

$$\mathbb{Y} = \begin{pmatrix} \bar{Y}_1 \\ \bar{Y}_2 \\ \bar{Y}_3 \end{pmatrix}, \quad \mathbb{X} = \begin{pmatrix} 1 & \bar{X}_1 \\ 1 & \bar{X}_2 \\ 1 & \bar{X}_3 \end{pmatrix}, \quad \eta = \begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{pmatrix}$$

de sorte que (2) se réécrive matriciellement $\mathbb{Y} = \mathbb{X}\beta + \eta$.

**4.** Donner l'espérance et la matrice de variance-covariance du vecteur $(\eta_1, \eta_2, \eta_3)$. On notera :

$$V = \begin{pmatrix} 1/n_1 & 0 & 0 \\ 0 & 1/n_2 & 0 \\ 0 & 0 & 1/n_3 \end{pmatrix}$$

**5.** Le modèle est hétéroscédastique. Quelle correction vue en cours nous permet-elle de nous ramener à un cas homoscédastique ? Montrer que pour le modèle corrigé, l'estimateur MCO de $\beta$ est :

$$\widehat{\beta} = \left({}^t\mathbb{X}V^{-1}\mathbb{X}\right)^{-1}\left({}^t\mathbb{X}V^{-1}\mathbb{Y}\right)$$

(cet estimateur s'appelle estimateur des moindres carrés généralisés).

**6.** Calculer en fonction des $\bar{X}_j$, $\bar{Y}_j$ et $n_j$ les matrices ${}^t\mathbb{X}V^{-1}\mathbb{X}$ et ${}^t\mathbb{X}V^{-1}\mathbb{Y}$ puis $({}^t\mathbb{X}V^{-1}\mathbb{X})^{-1}$ et $\widehat{\beta}$. Commentaires ?

**7.** Comparer $\widehat{\beta}$ à l'estimateur MCO de (2), $\widehat{\beta}_{MCO}$, obtenu à partir de $\mathbb{Y} = \mathbb{X}\beta + \eta$ (cet estimateur n'est pas $\widehat{b}_{MCO}$ calculé à la question **2**). Quels sont les qualités et les défauts de cet estimateur $\widehat{\beta}_{MCO}$ dans ce cas ?

**8.** Montrer que $\widehat{\beta}$ suit une loi normale d'espérance $\beta$ et dont on calculera la matrice de variance-covariance $\Sigma$.

**9.** Montrer que $\widehat{\eta} := \mathbb{Y} - \mathbb{X}\widehat{\beta}$ se réécrit $\widehat{\eta} = V^{+1/2}MV^{-1/2}\eta$ où $M$ est la matrice de projection orthogonale sur l'orthogonal de l'espace engendré par les colonnes de $V^{-1/2}\mathbb{X}$. En déduire que l'estimateur $\widehat{\sigma^2} = {}^t\widehat{\eta}V^{-1}\widehat{\eta}/(n-2)$ de $\sigma^2$ satisfait :

$$(n-2)\frac{\widehat{\sigma^2}}{\sigma^2} \rightsquigarrow \chi^2(n-2). \tag{3}$$

Calculer numériquement $\widehat{\sigma}^2$ avec les données du Tab. 1.

**10.** Tester si l'engrais a une influence sur la taille des plantes : écrire $H_0$ et $H_a$, proposer une statistique de test, expliquer son comportement sous $H_0$ et $H_1$, donner la région de rejet, conclure. On donne les quantiles à 2,5% et 97,5% de la loi de Student à 28 degrés de libertés : -0,67 et 2,05 respectivement.

**Exercice 2 (Equations de vraisemblance pour un modèle Tobit, 8 points)**
On considère le modèle suivant :

$$Y_i = Y_i^* \mathbf{1}_{Y_i^*>0}, \qquad \text{où } Y_i^* = X_i\beta + \varepsilon_i \tag{4}$$

où pour $i \in \{1, \cdots, n\}$, $X_i = (1, X_{i1}, \cdots X_{iK}) \in \mathcal{M}_{1,K+1}(\mathbb{R})$ est le vecteur ligne des variables explicatives caractérisant l'individu $i$ (pour simplifier, on pourra noter $X_{i0} = 1$ dans les calculs). Le vecteur des coefficients est $\beta = {}^t(\beta_0, \cdots \beta_K)$. On suppose que les $\varepsilon_i$ sont des variables i.i.d. de loi $\mathcal{N}(0, \sigma^2)$. La variable $Y_i^* \in \mathbb{R}$ est une variable continue latente, ce sont des $Y_i$ dont on dispose. Comme d'habitude, on notera matriciellement $Y^* = X\beta + \varepsilon$. On notera $\varphi$ et $\Phi$ les densité et fonction de répartition de la loi $\mathcal{N}(0,1)$, on rappelle bien-sûr que :

$$\varphi(x) = \frac{1}{\sqrt{2\pi}}\exp(-x^2/2).$$

On notera $n_1 = \text{Card}\ \{i \,|\, Y_i > 0\}$ et $n_0 = \text{Card}\ \{i \,|\, Y_i = 0\}$.

**1.** Calculer $\mathbb{P}(Y_i = 0 \mid X)$.

**2.** Ecrire la log-vraisemblance $\log \mathcal{L}(Y_1, \cdots, Y_n; \beta, \sigma^2)$, conditionnellement à $X$, des expliquées $(Y_1, \cdots Y_n)$.

**3.** Dériver la log-vraisemblance par rapport à $\beta_k$ ($k \in \{0, \ldots, K\}$) et $\sigma^2$. En déduire que le gradient de la log-vraisemblance par rapport à $\beta$ est

$$\nabla_\beta \log \mathcal{L} = \sum_{i,\, Y_i=0} \frac{-\varphi(-X_i\beta/\sigma)\, {}^tX_i}{\sigma\Phi(-X_i\beta/\sigma)} + \sum_{i,\, Y_i>0} \frac{(Y_i - X_i\beta)\, {}^tX_i}{\sigma^2} \tag{5}$$

**4.** Montrer que :

$$\widehat{\sigma}^2 = \frac{1}{n_1} \sum_{i \mid Y_i>0} Y_i(Y_i - X_i\widehat{\beta}).$$

**5.** Montrer que $\widehat{\beta}$ satisfait :

$$\frac{1}{\widehat{\sigma}} \sum_{i \mid Y_i>0} (Y_i - X_i\widehat{\beta})\, {}^tX_i = \sum_{i \mid Y_i=0} \frac{\varphi(-X_i\widehat{\beta}/\widehat{\sigma})\, {}^tX_i}{\Phi(-X_i\widehat{\beta}/\widehat{\sigma})}. \tag{6}$$

**6.** On définit par $X^1$ la matrice obtenue en rassemblant les explicatives pour les individus $i$ tels que $Y_i > 0$ et $X^0$ celle obtenue pour les individus $i$ tels que $Y_i = 0$. On note également $Y^1$ le vecteur des $Y_i$ pour les $i$ tels que $Y_i > 0$. On définit par $\gamma^0$ le vecteur de composantes $\varphi(-X_i\beta/\sigma)/\Phi(-X_i\beta/\sigma)$ pour les $i$ tel que $Y_i = 0$. Montrer en réécrivant matriciellement (6) que l'estimateur du maximum de vraisemblance de $\beta$ est :

$$\widehat{\beta} = ({}^tX^1X^1)^{-1}\, {}^tX^1Y^1 - \widehat{\sigma}({}^tX^1X^1)^{-1}\, {}^tX^0\gamma^0.$$

**7.** Qu'en déduire quant à la méthode qui consiste à estimer $\beta$ en faisant une régression MCO sur le sous-échantillon des individus $i$ tels que $Y_i = Y_i^*$ ? Citer une méthode d'estimation préférable vue en cours.

**Exercice 3 (Modèle de marché. 4 points)**

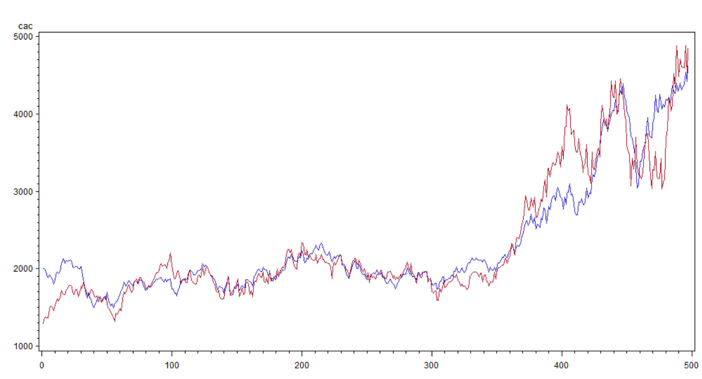


FIG. 1 – *Evolution hebdomadaire du CAC40 et de l'action d'Elf (renormalisée par le facteur multiplicatif 2325/70) entre janvier 1990 et juillet 1999.*

On considère le modèle simple suivant pour le rendement d'une action :

$$R_t = \alpha + \beta RM_t + \varepsilon_t \tag{7}$$

où $R_t$ est le rendement de l'action au temps $t$, $RM_t$ est le rendement du portefeuille de marché (un indicateur de santé de l'économie tel que le CAC40 pour la France). Si $\beta > 1$, les variations du portefeuille de marché sont amplifiées par l'action. Elle est considérée comme agressive. Si $\beta < 1$, l'action est défensive. On considère l'action d'Elf. Entre janvier 1990 et juillet 1999, nous avons 497 dates.

**1.** Dans l'annexe, pour la régression 1, des informations ont été perdues. Recalculer le coefficient de détermination, les statistiques de Student pour le test de significativité des coefficients, la borne supérieure de l'intervalle de confiance pour le coefficient $\beta$.

**2.** La régression est-elle globalement significative ?

**3.** Faire un test (à 5%) pour dire s'il y a une rupture de tendance à la date 300. Pour ce faire, on utilisera les régressions de l'annexe. On donne (en langage **R**) :

```
qf(0.95,2,493)=3.01401
qf(0.975,2,493)=3.71662
qf(0.025,2,493)=0.02532
qf(0.95,298,195)=1.24296
qf(0.975,298,195)=1.296165
qf(0.025,298,195)=0.776873
```