

Semestre 2  
LICENCE 1 mention ÉCONOMIE et GESTION  
LICENCE 1 mention ÉCONOMIE et GESTION parcours Européen  
LICENCE 1 mention ÉCONOMIE et DROIT  
LICENCE 1 mention ÉCONOMIE et DROIT parcours Européen

**STATISTIQUE DESCRIPTIVE**  
(durée 1h30)

Lundi 6 mai 2013 ~ 15h00 – 16h30

M. LATOURELLE

=====

**CONSIGNES**

- TOUT document est interdit, seule une calculatrice **non graphique, non programmable**, type casio-fx92 est autorisée.
- La rédaction sera prise en compte dans la notation. Chaque réponse devra être soigneusement justifiée, sauf mention contraire de l'énoncé.
- En ce qui concerne les calculs, les résultats seront arrondis à  $10^{-2}$ , sauf pour les densités de proportion à arrondir à  $10^{-3}$ .
- Le barème est indicatif, il pourra être modifié.
- Le sujet est composé de 3 feuilles, La feuille 3 est à rendre avec la copie.
- Durée de l'épreuve : **1 heure 30**.

**Questions de cours (2 points)**

Soient  $X$  et  $Y$  deux variables quantitatives continues. On observe la série statistique suivante :  $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$  où  $(x_i, y_i)$  sont les valeurs de  $X$  et  $Y$  observées pour le  $i$ -ème individu.

Montrer que la covariance de  $X$  et  $Y$  peut s'écrire sous la forme simplifiée :

$$Cov(X, Y) = \left( \frac{1}{n} \sum_{i=1}^n x_i y_i \right) - \bar{x} \bar{y}$$

### Régression linéaire (6 points)

Un hypermarché dispose de 20 caisses. On s'intéresse au temps moyen d'attente en fonction du nombre de caisses ouvertes. Le tableau ci-dessous donne le nombre  $X$  de caisses ouvertes et le temps moyen  $Y$  d'attente correspondant :

Nombre $X$ de caisses ouvertes	3	4	5	6	8	10	12
Temps moyen d'attente $Y$ en minutes	14	12	12.5	7.9	8.5	8.2	4

Le nuage de points associé aux variables  $X$  et  $Y$  est représenté en figure 1, page 3.

1. Commenter ce nuage de points.
2. Calculer la moyenne et la variance des variables  $X$  et  $Y$ .
3. Calculer la covariance de  $X$  et  $Y$ , puis le coefficient de corrélation entre  $X$  et  $Y$ .
4. Au vu du résultat de la question précédente, est-il judicieux d'effectuer une régression linéaire de  $Y$  sur  $X$  ?  
Si oui, calculer les coefficients de la droite de régression de  $Y$  sur  $X$ , puis tracer cette droite sur le nuage de points de la figure 1.
5. Estimer le nombre minimum de caisses à ouvrir, si on souhaite que le temps d'attente soit inférieur à 5,5 minutes.

### Liaison entre deux variables qualitatives (6 points)

Source des données : Ministère de l'Education nationale. Repères et références statistiques 2012.

Pour l'année 2011-2012, les effectifs (en milliers) d'étudiants dans les universités françaises se répartissaient selon le cursus et la discipline de la façon suivante :

Discipline \ Cursus	Licence	Master	Doctorat	
Droit, sciences politiques	115.00	72.34	7.75	
Sciences économiques, gestion	81.02	60.84	3.67	
AES	32.15	6.63	0.02	

On note  $X$  la variable discipline et  $Y$  la variable cursus.

1. De quel type sont les variables  $X$  et  $Y$  ?
2. (a) Quel nom donne-t-on au tableau ci-dessus ?  
(b) Compléter le tableau avec les distributions marginales des variables  $X$  et  $Y$ .
3. Proposer une représentation graphique des données et la faire.
4. Calculer la valeur du  $\chi^2$ . Commenter.
5. Calculer le coefficient  $C$  de Cramer. Commenter.

### Questionnaire à Choix Multiples (6 points)

Voir page suivante

### Questionnaire à Choix Multiples (6 points)

Dans le QCM suivant, seulement six réponses sont exactes. Toute bonne réponse cochée rapportera +1 point et toute mauvaise réponse cochée est pénalisée par -0,3 point. La note finale de cet exercice sera comprise entre 0 et 6 points. Attention à la note concernant les arrondis dans la consigne, située en début de sujet.

Le tableau suivant indique la distribution des salaires, en centaines d'euros nets par mois, des 100 salariés d'une entreprise :

Salaire net mensuel	[11 ; 15[	[15,17[	[17 ; 19[	[19 ; 21[	[21,35[
Effectifs	14	23	16	36	11

Cocher les cases qui conviennent :

- $F(x)$  correspond à la proportion des individus dont le salaire est égal à  $x$ .
- Environ 50% des individus gagnent moins de 1863 euros net par mois.
- 50% des individus gagnent entre 1596 euros et 2022 euros.
- $\sum_{i=1}^n \frac{1}{n} (x_i - \bar{x})^3 = \frac{1}{n} \sum_{i=1}^n x_i^3 - \bar{x}^3$ .
- Environ 50% de la masse salariale correspond à des salaires supérieurs à 1863 euros.
- L'indice de Gini est égal à 0,1.
- La série est parfaitement égalitaire.
- Les densités de proportions ne sont pas proportionnelles aux fréquences.
- La variable centrée réduite associée à  $X$  est la variable  $Y$  définie par  $Y = \frac{X-18,66}{16,42}$ .
- Le salaire égal à 1500 euros est situé à 0,9 écart-types de  $\bar{x}$ .
- Environ 36% des individus gagnent plus de 1900 euros nets par mois.
- Le moment simple d'ordre 2 est égal à 16,42 (centaines d'euros au carré).
- La médiane est toujours supérieure à la médiale.
- On pourrait représenter la série à l'aide d'un diagramme en tuyaux d'orgues.
- L'écart relatif entre la médiane et la médiale est d'environ 0,03 centaines d'euros.
- Le coefficient d'asymétrie de Fisher est égal à  $\gamma_1 = 0,91$ .
- D'après la boîte à moustaches de la série, la série est très étalée à gauche, vers les salaires les plus petits.
- Si tous les salariés bénéficient d'une prime exceptionnelle de 100€, le nouveau salaire moyen est 1966€ et la variance 174200 €<sup>2</sup>.

### Figure de l'exercice 2

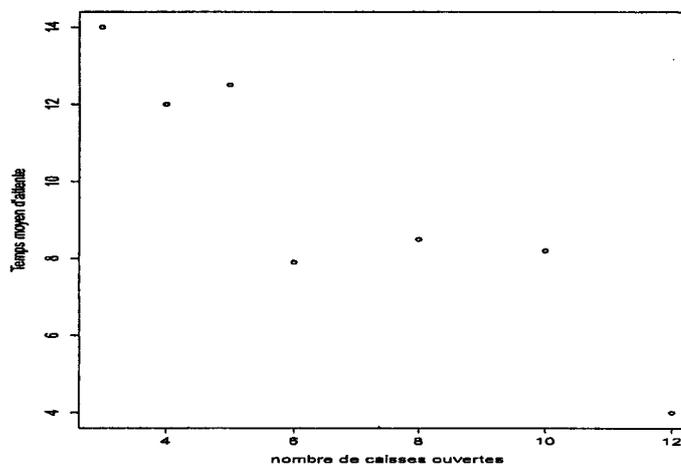


FIGURE 1 – Nuage de points des variables  $X$  et  $Y$ .