

Année universitaire 2016–2017  
Session 2 - Semestre 1

Licence 1 mention Economie, parcours Economie – Mathématiques et Informatique Appliquées

**EPREUVE : STATISTIQUE**

Enseignant : E. LECONTE

Date de l'épreuve : 16/06/2017

Durée de l'épreuve : 1h30

Liste des documents autorisés : aucun.

Liste des matériels autorisés : calculatrice de type Casio FX 92.

Nombre de pages : 2.

**Justifier et détailler les calculs.**

**Avant tout calcul numérique, donner d'abord la formule littérale que vous allez utiliser. La note tiendra compte de la qualité de la rédaction.**

**Sauf mention contraire, on arrondira les calculs à  $10^{-2}$  près.**

**Barème indicatif : 3 – 8,5 – 8,5.**

---

**Question de cours**

Soient  $X$  et  $Y$  deux variables quantitatives. On note  $\{x_i\}_{i=1}^n$  et  $\{y_i\}_{i=1}^n$  les observations de ces variables sur un échantillon de taille  $n$ . Soit  $Z = X + Y$ .

1. Rappeler (sans la démontrer) la relation liant la moyenne de  $Z$  aux moyennes de  $X$  et  $Y$ .
2. Rappeler et démontrer la relation qui existe entre la variance de  $Z$ , les variances de  $X$  et de  $Y$  et la covariance de  $X$  et de  $Y$ .

**Exercice 1**

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

1. Donner le type des variables  $X$  et  $Y$ .
2. Faire un graphique pour illustrer la liaison entre les variables  $X$  et  $Y$ .
3. Calculer la moyenne et la variance des variables  $X$  et  $Y$ .
4. Calculer la covariance de  $X$  et  $Y$  puis le coefficient de corrélation linéaire de  $X$  et  $Y$ .
5. 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  $X$ , donner l'équation de la droite obtenue puis la tracer sur le graphique de la question 2.
6. D'après ce modèle, quel nombre minimum de caisses faut-il ouvrir si l'on souhaite que le temps d'attente soit inférieur à 5 minutes?

## Exercice 2

On dispose du salaire annuel en milliers d'euros et de la catégorie d'emploi (employé, cadre stagiaire ou cadre) de salariés d'une entreprise. Utiliser les commandes et résultats obtenus avec le logiciel R pour répondre aux questions ci-dessous.

```
> summary(salaire)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
15.75  24.00   28.50   33.98  35.70  135.00
> table(cat_emploi)
      Cadre Cadre_stagiaire      Employe
         81             27         353
> tapply(salaire,cat_emploi,mean)
      Cadre Cadre_stagiaire      Employe
 64.98             30.94         27.10
> tapply(salaire^2,cat_emploi,sum)
      Cadre Cadre_stagiaire      Employe
367343.11         25961.06        271755.58
```

1. Quelle est la population étudiée? De combien d'individus dispose-t-on dans le fichier de données?
2. Quelles sont les deux variables étudiées? Donner leur type.
3. En vous basant uniquement sur le résultat de l'instruction `summary(salaire)`, donner le nom et la valeur numérique de deux indicateurs de position et de deux indicateurs de dispersion pour la variable salaire.
4. Représenter graphiquement la distribution de la variable catégorie d'emploi et donner le nom du graphique.
5. Donner le nom d'un graphique qui permet de visualiser la liaison entre les deux variables. Pouvez-vous le tracer? Justifier votre réponse.
6. Rappeler la définition de la variance inter-groupes en précisant toutes les notations que vous utilisez et la calculer.
7. Rappeler la définition du rapport de corrélation  $\eta^2$ . Le calculer et conclure sur l'intensité de la liaison entre les deux variables.