

Probabilité et statistique I (partim Statistique descriptive)

Premier Bachelier en sciences informatiques

Vendredi 10 juin 2016 – Examen pratique – Groupe 1a (14h-15h15)

Nom: Prénom: Matricule:.....

Indications

- L'examen comporte plusieurs questions spécifiques pour lesquelles des réponses précises doivent être écrites sur le questionnaire. Lorsqu'un graphique doit être représenté (ce qui est indiqué par le symbole \square), celui-ci doit être montré à G. Haesbroeck ou M. Ernst. Le graphique tel que montré est celui sur lequel la cote sera basée.

- Il est interdit de communiquer avec quiconque via Internet (ou GSM,...) sous peine d'annulation de l'examen. Il n'est pas permis non plus d'utiliser une machine à calculer.

- L'ensemble de données à considérer lors de cet examen (en ligne sur le site <http://www.statmath.ulg.ac.be/retard1/>) s'intitule `retard1.txt` et contient les informations suivantes concernant 176 voyages en avion ayant eu lieu en novembre 2015 entre deux aéroports américains, l'aéroport d'arrivée étant celui de Birmingham. Plus précisément les informations suivantes sont disponibles pour ces vols:

Jour: variable précisant le jour de la semaine

Origine: variable qualitative précisant l'aéroport d'origine; 4 modalités: MD (pour un aéroport de Baltimore), MI (aéroport de Detroit), NV (Las Vegas) et NY (New York)

Heure: variable précisant l'heure de décollage (l'heure étant exprimée en minutes comptées à partir de 0h)

Retard: nombre de minutes de retard à l'arrivée (les valeurs peuvent être négatives, ce qui correspond dans ce cas à une arrivée de l'avion avant l'heure prévue)

Ces données doivent être chargées dans R à l'aide des commandes suivantes:

```
data <- read.table("...",header=TRUE)
```

```
attach(data)
```

où ... doit être remplacé par le chemin d'accès au fichier.

1. On s'intéresse dans un premier temps à la variable **Jour**.

(a) Quel est le jour de départ le plus fréquent. Déterminer son effectif:.....

(b) Représenter la distribution des effectifs de cette variable à l'aide d'un diagramme en barres. \square

2. En étudiant conjointement la variable **Jour** avec celle précisant l'aéroport d'origine, déterminer l'effectif bivarié du couple (**Lundi**, **NY**) et préciser également la fréquence des vols en partance de New York sachant que le vol a lieu un lundi.

Effectif conjoint:.....

Fréquence conditionnelle:.....

3. Concentrons-nous à présent sur les nombres de minutes de retard.

(a) On souhaite tout d'abord comparer ceux en fonction du jour du vol. Pour cela, représenter des boîtes à moustaches de la variable **Retard** en fonction de la variable qualitative **Jour**. □

(b) Représenter la distribution de la variable **Retard** à l'aide d'un histogramme construit sur les classes suivantes: $[-50, -25]$, $]-25, 0]$, $]0, 25]$, $]25, 50]$ et $]50, 250]$. □

(c) Pour caractériser la distribution du nombre de minutes de retard, donner les formules d'un coefficient de dispersion et d'un coefficient de dissymétrie calculés à partir des quartiles et préciser leurs valeurs sur les minutes de retard:

Dispersion: nom: formule: Valeur=

Dissymétrie: nom: formule: Valeur=

(d) Quel est le pourcentage des vols qui sont arrivés avant l'heure prévue?

4. On peut s'interroger sur le lien entre l'heure de départ et le nombre de minutes de retard.

(a) Calculer la corrélation entre les variables **Heure** et **Retard**:.....

(b) Déterminer la pente (coefficient a) de la droite de régression de la variable **Retard** par rapport à la variable **Heure** lorsque ce paramètre est estimé par la technique des moindres carrés: $\hat{a} = \dots\dots\dots$

(c) Représenter le nuage de points de la variable **Retard** en fonction de la variable **Heure** et y ajouter la droite de régression calculée. □

(d) Que vaut le coefficient de détermination?.....

(e) Déterminer le résidu du 85ème vol repris dans la base de données:

(f) Calculer l'écart-interquartile de la série des résidus de la droite de régression de la variable **Retard** en fonction de la variable **Heure**: