

## Devoir « Statistique descriptive » du 13 Janvier 2011

### Utilisation du logiciel

**Description des données :** Les données sont issues du « Mozilla Open Data Visualization Competition »<sup>1</sup>, version 2010 et, plus précisément, de la plateforme de recherche « Test Pilot » qui recueille des données utilisateur sur Firefox. Les données utilisées sont les données « A Week in the Life of a Browser - Version 2: Aggregated Data Samples » qui sont disponibles à l'adresse : <https://testpilot.mozillalabs.com/testcases/a-week-life-2/aggregated-data.html>. À partir des données originelles, nous avons conservé uniquement les données relatives aux réponses de l'enquête menée sur l'utilisation de Firefox (et pas les données événementielles ni les données utilisateur) et nous nous sommes restreints aux données relatives à des utilisateurs du système d'exploitation Mac OS X. Le fichier de données final, que vous allez charger et sur lequel vous allez travailler, contient 397 individus (des utilisateurs Firefox sous système d'exploitation sous Mac OS X) et 17 variables (les réponses aux 17 questions de l'enquête).

**Instructions pour commencer :** Les fichiers à utiliser pour le devoir se situent dans le répertoire utilisateur /home/std/DevoirNVV-NOM<sup>2</sup> où NOM est votre nom. Vous y trouverez le fichier Devoir-NOM.odt, qui est le fichier OpenOffice.org (traitement de texte) que vous devez remplir et rendre en fin de devoir, et le fichier FirefoxSurvey-Simplifie.RData qui est le fichier contenant les données. Pour commencer :

- Ouvrez le fichier Devoir-NOM.odt ;
- Lancer R et le package Rcmdr ;
- Définir le répertoire courant (c'est donc /home/std/DevoirNVV-NOM) ;
- Ouvrir le fichier de données : il contient le tableau de données, nommé surveymac, dans lequel sont situées les données décrites ci-dessus. Les colonnes de ce tableau sont nommées q1, q2 ... q17. Les variables que nous allons utiliser ont été préalablement recodées, si nécessaire, pour permettre leur utilisation directe.

**N'oubliez pas d'enregistrer régulièrement votre travail !**

**Orientation générale du devoir :** Le devoir portera sur la question générale suivante : « Existe-t-il des comportements différenciés selon le sexe concernant l'utilisation d'internet ? » En particulier, on s'intéressera au navigateur internet utilisé et à la perception de son niveau en informatique. Les questions de l'enquête étudiées seront donc :

- la question q4 : Si vous utilisez plusieurs navigateurs, lequel considérez-vous comme votre navigateur principal ? (réponses possibles : « Chrome », « Firefox », « Internet explorer », « Opera », « Safari », « Utilise uniquement Firefox ») ;
- la question q5 : « Quel est votre sexe ? » (réponses possibles : « Homme » et « Femme ») ;
- la question q8 : « Comment noteriez-vous votre niveau en informatique et internet ? » (réponses possibles : de 1 à 10, valeurs entières uniquement).

---

1 <http://design-challenge.mozillalabs.com/open-data/OpenDataCompetition.php>

2 Les étudiants utilisant leur ordinateur personnel trouveront le fichier DevoirNVV-NOM sur le bureau de leur machine.

## Table des matières

|   |   |
|---|---|
| 1Analyse univariée de la variable « Navigateur principal » (Question q4).....           | 2 |
| 2Analyse bivariée des variables « Navigateur principal » et « Sexe » (Question q5)..... | 3 |
| 3Analyse univariée de la variable « Niveau en informatique » (Question q8).....         | 3 |
| 4Analyse bivariée des variables « Niveau en informatique » et « Sexe ».....             | 4 |

### 1 Analyse univariée de la variable « Navigateur principal » (Question q4)

1. Donnez le tableau d'effectifs et de fréquences de la variable « Navigateur principal ». Commentez les résultats obtenus.

Tableaux d'effectifs et de fréquences :

Commentaires :

Code généré :

|  |
|--|
|  |
|--|

2. Construire un diagramme en tuyaux d'orgue pour représenter la distribution de la variable « Navigateur principal ».

Diagramme :

Code généré :

|  |
|--|
|  |
|--|

## 2 Analyse bivariée des variables « Navigateur principal » et « Sexe » (Question q5)

1. Déterminez la distribution de la variable « Navigateur principal » conditionnellement au sexe. Commentez cette distribution conditionnelle.

*Remarque : Si, lors de la réalisation de cette question, vous avez le message d'erreur « impossible de trouver la fonction "abind" », tapez library(abind) et recommencez.*

Tableau de la distribution conditionnelle :

Commentaires :

Code généré :

2. Déterminez le  $\chi^2$  et le C de Cramer. Commentez la valeur de ce dernier.

Résultat :

Commentaires :

Code généré :

## 3 Analyse univariée de la variable « Niveau en informatique » (Question q8)

1. Effectuez une boîte à moustaches représentant la distribution la note d'auto-évaluation du sondé en informatique et internet (question q8 : « Niveau en informatique »).

Diagramme :

Code généré :

2. Déterminez la médiane de la note « Niveau en informatique ». Que signifie concrètement cette valeur ?

Médiane :

Signification :

Code généré :

#### **4 Analyse bivariée des variables « Niveau en informatique » et « Sexe »**

1. Effectuez des boîtes à moustaches parallèles du « Niveau en informatique » selon le sexe. Commentez la différence entre les deux distributions (et attention à ne pas m'irriter dans vos commentaires).

Diagramme :

Commentaires :

Code généré :