

Devoir « Compléments de statistique » du 9 Janvier 2012

Utilisation du logiciel

Consignes

- Les réponses sont à donner directement sur le fichier de réponses au format LibreOffice (odt) ou Word 2003 (doc) : renommez le fichier Devoir-NOM.odt (ou Devoir-NOM.doc) en remplaçant « NOM » par votre nom. Je récupère tous les devoirs sur clé USB en fin d'épreuve : merci de faire attention à **ne pas me remettre de fichier au format docx**, que je lis de manière très imparfaite.
- **Matériel autorisé (à l'exclusion de toute autre chose)** : crayons, ordinateur, notes de cours et livres. **Les téléphones portables et les clés USB sont formellement interdits sur les tables, sur vos genoux, dans vos poches** : ils doivent être déposés, avec vos sacs, à côté de mon bureau. **L'utilisation d'internet est également formellement interdite** : tout étudiant que je surprendrais avec un navigateur ouvert ou toute autre application permettant l'échange de données sur internet sera immédiatement suspendu et recevra la note de 0 avant d'être jugé en conseil de discipline même si il n'utilise pas cette application.
- Il est formellement interdit de parler (même en langage des signes) ou d'échanger du matériel.
- **N'oubliez pas d'enregistrer régulièrement votre travail !** Je décline toute responsabilité pour le mauvais fonctionnement des machines, notamment si vous n'avez pas enregistré votre script de manière régulière.

Description des données : Les données utilisées dans ce devoir proviennent d'une enquête réalisée par les étudiants de 1ère année de DUT STID en 2010/2011 sur le devenir des anciens diplômés. Elle a été réalisée à l'aide du logiciel LimeSurvey®.

Instructions pour commencer : Les fichiers à utiliser pour le devoir se situent dans le répertoire DevoirNVV, généralement situé sur le bureau (sauf mention contraire). Vous y trouverez le fichier Devoir-NOM.odt (ou Devoir-NOM.doc), qui est le fichier OpenOffice.org (traitement de texte) que vous devez **renommer** puis remplir et rendre en fin de devoir, et les fichiers

- Surveydata_syntax.R : script R d'importation des données de LimeSurvey ;
- survey_43847_data_file.csv : données, au format csv (« comma separated values ») ;
- donnees-simplifiees.Rdata : données simplifiées au format Rdata.

Les deux premiers fichiers sont à utiliser pour la première partie du devoir. Le dernier pour la seconde. Les deux parties sont indépendantes.

Table des matières

1 Importation et mise en forme de données depuis LimeSurvey.....	2
2 Analyse du fichier de données simplifié.....	2
a) Mise en forme des données.....	2
b) Analyse univariée de la variable typeBAC.....	3
c) Analyse bivariée des variables typeBAC et ThemeStage_1.....	3

1 Importation et mise en forme de données depuis LimeSurvey

On rappelle que l'importation des données depuis LimeSurvey se fait avec la commande :

```
source("Surveydata_syntax.R", encoding = "UTF-8")
```

Cette ligne de commande importe les données dans un tableau nommé **data**.

1. Importer les données puis utiliser la commande **summary** pour obtenir les principales caractéristiques du jeu de données. Coller ci-dessous les résultats de cette commande :

Résultats obtenus :

2. Supprimer les variables **submitdate**, **lastpage**, **startlanguage**, **token**, **datestamp**, **startdate**, **ipaddr**, **refurl** et **num** du jeu de données. Coller le script R permettant cette opération ci-dessous :

COLLER ICI LE SCRIPT R

3. Utiliser la variable **id** comme identifiant de la ligne. Coller le script R permettant cette opération ci-dessous :

COLLER ICI LE SCRIPT R

2 Analyse du fichier de données simplifié

Charger le fichier de données **donnees-simplifiees.Rdata**. Celui-ci contient une partie, filtrée et mise en forme, des données précédentes, dans un tableau de données noté **d**. N'oubliez pas de sélectionner ce tableau de données comme tableau courant.

a) *Mise en forme des données*

La variable **CP_BAC** est le code postal de l'étudiant au moment où il a passé son baccalauréat : c'est donc un code qui est compris par R comme une variable numérique. Transformer cette variable en variable qualitative (facteur) sans changer ses valeurs. Coller le script R permettant cette

opération ci-dessous :

COLLER ICI LE SCRIPT R

b) Analyse univariée de la variable typeBAC

La variable `typeBAC` indique le type de baccalauréat qu'avait obtenu l'étudiant avant d'entrer en DUT STID. Effectuer le diagramme circulaire de la distribution de cette variable. Coller ci-dessous le script R permettant de générer ce graphique, le graphique lui-même et commenter les résultats obtenus.

On rappelle que si jamais la fonction d'exportation graphique du menu ne fonctionnait pas, la commande

```
dev.print(jpeg,file="monbeaugraphe.jpeg",width=500)
```

permet d'exporter le graphique courant dans le répertoire courant.

COLLER ICI LE SCRIPT R

Graphique :

Commentaires :

c) Analyse bivariée des variables typeBAC et ThemeStage_1

La variable `ThemeStage_1` contient les réponses à la question « Avez-vous effectué votre stage de deuxième année en Statistique ? ». On cherche à savoir si le type de baccalauréat de l'étudiant influence le fait de faire, ou non, un stage en statistique en deuxième année.

1. Déterminer la distribution de la variable `ThemeStage_1` conditionnellement à la variable

typeBAC. Coller le script R permettant d'obtenir cette distribution ainsi que le tableau de la distribution conditionnelle et commenter les résultats obtenus.

COLLER ICI LE SCRIPT R

Résultats obtenus : (faire un tableau proprement)

Commentaires :

2. Calculer le C de Cramer. Coller le script R permettant d'obtenir cette valeur, donner la valeur et commenter les résultats obtenus.

COLLER ICI LE SCRIPT R

$C =$

Commentaires :

3. Faire le graphique de la distribution conditionnelle de la variable **ThemeStage_1** conditionnellement à la variable **typeBAC**. Le graphique devra comporter une légende qui sera insérée, comme vu en TP, sur une partie séparée du graphique. Coller le script R permettant d'obtenir ce graphique ainsi que le graphique lui-même.

COLLER ICI LE SCRIPT R

Graphique :