Université de Nice — L1MASS, année 2014-2015
Département de Mathématiques — Statistique
NOM : Date : .
PRENOM : Groupe : .

Feuille-réponse 1
Prise en main de R

1 Prise en main de R

Recherchez le programme R et lancez-le. La fenêtre qui s'ouvre s'appelle la *console*. Elle sert essentiellement à l'affichage des résultats. On peut y saisir une commande, telle la demande d'affichage de la valeur d'une variable (on tape le nom de la variable, puis "return"), mais il convient d'éviter de l'utiliser pour des commandes plus compliquées telles les boucles, etc. car il n'est guère possible d'en demander la réexecution.

Pour saisir une succession de commandes, il convient d'ouvrir une *fenêtre d'édition*. Pour cela utiliser la commande de menu **Fichier** $->$ **Nouveau script**. Il s'agit d'un assez simple éditeur de texte Comme pour tout éditeur, il convient de commencer par donner un nom au fichier édité. Pour cela, utiliser la commande de menu **Fichier** $->$ **Sauvegarder-sous** et choisir le nom `Stat1.r` et éventuellement le disque et répertoire où vous voulez sauvegarder votre séance (votre clé USB par exemple). Il est essentiel de **garder les fichiers** de vos séances : cela ne vous sera pas seulement utile pour vos révisions ; dès la semaine suivante vous pourrez repartir des idées de la semaine précédente pour développer de nouvelles questions. Le nom choisi se compose de `Stat` car il s'agit du cours de Statistique, `1` car il s'agit du premier TP de la série `1`, `2`, ..., `10`, et enfin de `.r`, qui est l'extension reservée pour les *sources* de R. Par la suite il vous suffira de double-cliquer sur un tel fichier-source pour à la fois lancer Scilab (si ce n'est pas déjà fait) puis ouvrir le source, ou *script*, dans l'éditeur de R.

A présent, vous allez taper les instructions successives dans votre fenêtre d'édition, puis marquer (en tirant la souris tout en enfonçant le bouton gauche, ou en cliquant en début de texte et par un Maj-click-gauche en fin de zône à marquer), puis demander l'execution de la zône marquée par la commande de menu **Édition** $->$ **Exécuter la ligne ou sélection** (ou simplement ne tapant Ctrl+R). Ainsi vous construisez progressivement une succession d'instructions et en testez l'effet en lisant dans la **console** les résultats produits. Vous pourrez ensuite réexécuter tout ou partie de votre code, dans lequel vous aurez peut-être modifié la valeur de quelques constantes.

Avant d'exécuter tout ou partie de votre code, tapez **Ctrl-S** (ou exécutez la commande de menu **Fichier-Sauvegarder**) car il arrive que R plante(!) (ou déjante(!!)) c'est-à-dire s'arrête complètement et il ne vous reste plus qu'à repartir, en double-cliquant sur le fichier que vous avez fort heureusement sauvegardé ! Notez que la fenêtre d'édition s'affiche souvent par dessus la fenêtre console, ce qui gêne pour lire les résultats sur cette console. Pour cela, modifiez la place et la taille des deux fenêtres : à gauche une console étroite et à droite un éditeur plus large, les deux s'étirant du haut en bas de votre écran. Évitez tout chevauchement. Faites cela systématiquement (et immédiatement) : vous n'aurez plus à le faire lorsque vous rencontrez une difficulté et qu'il vous faudra toute votre attention sur elle.

2 Le fichier `survey`

Nous allons travailler sur un fichier appelé "survey" comportant 12 variables relevées sur un échantillon de 236 individus [1]. Pour charger ce fichier, charger la librairie MASS puis ce fichier qui s'appelle survey :

```
library(MASS)
data(survey)
```

Exercice 1 : le fichier survey

1. Pour connaitre l'usage d'une commande de R, il suffit de taper dans la console son nom précédé d'un point d'interrogation. Une page de l'aide en ligne apparait alors. En utilisant au besoin l'aide en ligne, expliquer en face de chacune des lignes suivantes ce que fait la ligne.

   ```
   head(survey, n=10)
   names(survey)
   ```

1. Cette séance est fortement inspirée d'un TD de A.B. Dufour et J. Lobry qu'on trouve en ligne à `publi.univ-lyon/R/pdf/tdr601.pdf`

```
fix(survey)
essai<-survey[1:10,2:5] ; fix(essai)
```

2. A l'aide de le commande `summary(survey)` indiquer quelles sont, parmi les 12 variables, celles qui sont catégoriques et celles qui sont numériques.

3. Pour la variable catégorique `smoke`, indiquer combien elle a de modalités et pour combien d'individus cette variable n'est pas relevée (données dites NA pour *Non Accessible*).

4. Pour créer un nouveau fichier n'ayant plus de données manquantes, taper
   `survey.cc <- survey[complete.cases(survey), ]`
   Quelle est la première ligne du fichier `survey` qui a été supprimée dans ce nouveau fichier "survey.cc" ?

5. Pour la variable numérique `Age`, indiquer les ages extrêmes relevés sur cet échantillon (minimum et maximum.

6. Dans la suite, nous allons nous intéresser plus particulièrement à 3 des variables de survey, l'empan [2] de la main d'écriture dite main dominante, celui de la main qui n'écrit pas, et la taille des individus. Ces trois variables sont numériques et exprimées en cm. Exécuter les lignes suivantes et expliquer ce que vous obtenez :
   ```
   mesures <- survey.cc[, c("Wr.Hnd", "NW.Hnd", "Height")]
   names(mesures) <- c("Empan1", "Empan2", "Taille")
   ```

3 Histogramme

Former le caractère `x` constitué des valeurs de `Wr.Hnd` (l'empan1). Représenter l'histogramme de `x` au moyen de la commande `hist(x)`. Répondre aux questions suivantes, en veillant à faire une phrase.

1. Combien y a-t-il de classes et quelle sont les bornes de ces classes ?

2. Quelles est (sont) la classe d'effectif maximal, et celle d'effectif minimal ?

2. l'empan est la distance entre les extrémités du pouce et du petit doigt lorsqu'ils sont écartés au maximum.

3. Exécuter le code suivant et expliquer ce qu'il permet d'obtenir :

```
longueur=length(x)
for (i in 1:longueur) {if ((x[i]<=14)&(x[i]>13)) {print(c(i,x[i]))} }
```

4. Quel est le numéro k de la classe des $x_i \in]22, 23] =]b_{k-1}, b_k]$? Quel est l'*effectif* de cette classe ? Quels sont les individus i de cette classe et leur caractère x_i ?

5. Que retourne les commande `hist(x,breaks=c(min(x),20,22,23,max(x)))` et `hist(x,breaks=seq(13,25,1))`

4 Le caractère "stature"

Former le caractère `u` constitué des valeurs de `Height` (la stature) et représenter son histogramme, comme précédemment.

- combien y a-t-il de classes ? Quelle est la largeur $b_k - b_{k-1}$ des classes de cet histogramme-ci ? Quelles sont les valeurs extrêmes du caractère `u` ?

- Calculer et donner la stature moyenne `mu` de cet échantillon. Appartient-elle à la classe d'effectif maximal ?

- Calculer et donner l'écart-type `sigma`. Y a-t-il des individus en dehors de l'intervalle $[\mu - 3\sigma, \mu + 3\sigma]$? Si oui, quelle est leur stature ?