

SCI03 Automne 2014

TP 1: Statistique descriptive, probabilités

1 Familiarisation avec R

1.1 Utilisation comme une calculatrice

Pour vous familiariser avec R, taper les instructions suivantes :

```
8*5/(2*2)
x=4
y=-5
z=x*y
sin(pi*x)
ls()
```

1.2 Création d'une variable quantitative

Les nombres de baleines échouées au Texas dans les années 1990 ont été les suivants : 74, 122, 235, 111, 292, 111, 211, 133, 156, 79. Ces données peuvent être entrées dans R à l'aide de la commande suivante :

```
x=c(74, 122, 235, 111, 292, 111, 211, 133, 156, 79)
```

Entrez les instructions suivantes et observez le résultat :

```
y=sort(x)
mean(x)
var(x)
sd(x)
summary(x)
stem(x)
boxplot(x)
stripchart(x)
```

1.3 Création d'une variable qualitative

On a observé les couleurs de 20 objets, représentées dans un fichier avec le codage suivant : 0=blanc, 1=vert, 2=bleu. Les données codées de cette façon sont les suivantes : 0 1 1 0 2 1 2 0 0 1 2 2 2 0 0 1 1 0 0 0. Les instructions suivantes ont pour effet de stocker ces données dans R et de les représenter comme des observations d'une variable qualitative (facteur), en indiquant explicitement les noms des modalités :

```
x=c(0,1,1,0,2,1,2,0,0,1,2,2,2,0,0,1,1,0,0,0)
couleur=factor(x,levels=0:2)
levels(couleur)=c("blanc","vert","bleu")
```

La variable "couleur" est maintenant traitée comme une variable qualitative :

```
couleur
summary(couleur)
T=table(couleur)
barplot(T)
```

2 Analyse descriptive d'un fichier de données

2.1 Description des données

Le jeu de données `brain.txt` contient des données relatives à des résultats de tests de QI et aux poids du cerveau de 40 étudiants. Voici une description de ce jeu de données :

Description : Willerman et al. (1991) collected a sample of 40 right-handed Anglo introductory psychology students at a large southwestern university. Subjects took four subtests (Vocabulary, Similarities, Block Design, and Picture Completion) of the Wechsler (1981) Adult Intelligence Scale-Revised. The researchers used Magnetic Resonance Imaging (MRI) to determine the brain size of the subjects. Information about gender and body size (height and weight) are also included. The researchers withheld the weights of two subjects and the height of one subject for reasons of confidentiality.

Number of cases : 40

Variable Names :

- Gender : Male or Female
- FSIQ : Full Scale IQ scores based on the four Wechsler (1981) subtests
- VIQ : Verbal IQ scores based on the four Wechsler (1981) subtests
- PIQ : Performance IQ scores based on the four Wechsler (1981) subtests
- Weight : body weight in pounds
- Height : height in inches
- MRI_Count : total pixel Count from the 18 MRI scans

2.2 Travail demandé

Les données peuvent être chargées par l'instruction suivante :

```
brain<-read.table("../Data/brain.txt",header=T) # lecture des données
attach(brain)
Gender=factor(Gender) # déclaration de la variable Gender comme facteur
```

En utilisant les outils présentés en cours (résumés numériques, graphiques), décrivez les distributions de ces variables ainsi que leurs relations. Discutez les résultats.

3 Probabilités

3.1 Génération de nombres aléatoires

1. Générez un échantillon de $n = 50$ nombres issus d'une loi normale d'espérance $\mu = 1$ et d'écart-type $\sigma = 2$. Calculez la moyenne et l'écart-type empirique de l'échantillon. Faites un histogramme de ces valeurs en superposant la fonction de densité de la loi normale correspondante. Répétez plusieurs fois cette opération. Recommencer avec $n = 1000$. Commentez les résultats.
2. Générez un échantillon de $n = 50$ nombres issus d'une loi binomiale $\mathcal{B}(20, 0.7)$. Calculez la moyenne et l'écart-type empirique de l'échantillon. Représentez graphiquement les fréquences relatives dans l'échantillon, et comparez les aux fréquences théoriques (probabilités). Répétez plusieurs fois cette opération. Recommencer avec $n = 1000$. Commentez les résultats.

3.2 Calculs de probabilité

Calculez avec R les probabilités des événements suivants :

1. Une variable normale centrée-réduite est supérieure à 3 ;
2. Une variable normale d'espérance 35 et d'écart-type 6 est supérieure à 42 ;
3. Une variable normale d'espérance 35 et d'écart-type 6 est comprise entre 40 et 50 ;
4. Obtenir $n - 1$ faces sur n lancers d'une pièce de monnaie équilibrée, avec $n = 5, 10, 30$;
5. Obtenir plus de 15 faces sur 20 lancers d'une pièce de monnaie équilibrée ;
6. Obtenir entre 10 et 15 faces sur 20 lancers d'une pièce de monnaie équilibrée.