
TP 1 : *Introduction au langage R et Statistique descriptive*

Le logiciel utilisé pendant les séances de TP est le logiciel libre R. Une présentation de R est disponible sur le web à l'adresse suivante : <http://cran.r-project.org>

Exercice 01 :Sans utiliser le logiciel On définit :

x=c(1,2,4,5,7,9)

y=c(3,4,5,7,11,12,13)

Donner les résultats des commandes R suivantes

1) x+1

2) y*2

3) length(x) et length(y)

4) x+y

5) sum(x>5) et sum(x[x>5])

6) sum(x>5|x<3)

7) y[3]

8) y[-3]

9) y[x]

10) (y>7) et y |y>7

Exercice 02 :Statistique descriptive Les données 'sleep' donnent l'augmentation ou la diminution du temps de sommeil (variable extra) chez deux groupes de patients traités par deux médicaments (la variable 'group' prend les valeurs 1 ou 2). On dispose de 20 données.

```
data(sleep)
sleep
  extra group
1  0.7   1
2 -1.6   1
3 -0.2   1
4 -1.2   1
...
15 -0.1   2
16  4.4   2
17  5.5   2
18  1.6   2
19  4.6   2
20  3.4   2
```

1- Créez un vecteur "g1" contenant la variable "extra" du groupe 1. puis, un vecteur "g2" contenant

```
g1= sleep$extra[1:10]
#plusieurs syntaxes
#ou sleep$extra[sleep$group==1]
#ou sleep$extra[which(sleep$group==1)]
```

la variable "extra" du groupe 2. 2- Représenter l'histogramme (en utilisant la fonction "hist" avec l'option "proba=TRUE", des données "g1" puis ajouter sur ce graphique des droites verticales de couleurs différentes correspondant

- 1) à la moyenne (fonction mean)
- 2) à la médiane (fonction median)
- 3) aux quantile Q_1 et Q_3 (fonction quantile)

Ajouter une légende. 3) Refaire le même graphique avec les données g2 4) Tracer sur un même gra-

```
hist(g1)
abline(v=mean(g1))
-----etc-----
legend(-----à-complète-----)
```

phique les densités pour les deux groupes

```
>plot(density(g1))
>lines(density(g2),col=2)
```

Ajuster les axes en utilisant les options `xlim=c(a,b)` et `ylim` de la fonction `plot`. 5) Tracer les boxplots des deux séries sur un même graphique

```
>boxplot(data.frame(g1,g2))
```

6) Commenter les résultats obtenus.

Exercice 03 : Étude de la série "nottem" : relevé de températures à Nottingham pendant 20 ans (une donnée par mois)

1) Tracer la série de données

```
>plot(nottem)
```

et commenter l'allure de la courbe.

2) On organise les données sous la forme d'une matrice (12 colonnes, 20 lignes) telle que la première colonne contient les données du mois de janvier.

```
>DT = matrix(nottem, ncol=12 , byrow =T)
```

3) Représenter les différents mois sur un même graphique. Commenter

```
>matplot(DT, type="b", main="les différents mois")
```

4) Représenter les différentes années sur un même graphique.

5) Expliquer et commenter le graphique suivant

```
>boxplot(data.frame(DT))
```

Exercice 04 : Nombre de classes dans un histogramme

L'option "nclass" (ou `breaks`) de la fonction "hist" permet d'ajuster le nombre de classes de l'histogramme. Par défaut, il est fixé par la formule de Sturges $n_c = \lceil 1 + \log_2(n) \rceil$ où $\lceil \rceil$ désigne la partie entière.

1) Simuler $n = 100$ variables aléatoires iid suivant la loi gaussienne standard.

- 2) Sur une même page (par(mfrow=c(3,3)), par exemple) : tracer l'histogramme de l'échantillon simulé en faisant varier le nombre de classes. Prendre par exemple nclass= 3 ; 5 ; 8 ; 10 ; 15 ; 20 ; 25 ; 30 ; 50.
- 3) Refaire la question précédente avec un échantillon simulé suivant la loi exponentielle de paramètre 1, puis avec un échantillon simulé suivant la loi de Cauchy.
- 4) Commenter les résultats.

Exercice 05 :

Ponte d'une poule pendant huit ans Le nombre d'œufs pondus en un an par une poule a été relevé pendant 8 ans. Les résultats sont les suivants :

Année (t)	1	2	3	4	5	6	7	8
Nombre d'œufs pondus (n)	160	140	122	112	96	88	72	62

- a) Représenter graphiquement les données. Quelle relation entre n et t vous suggère cette représentation ?
- b) Calculer la moyenne et la variance du nombre d'œufs pondus.

Z. DJERIDI