

Pour ce TP, on utilisera le RStudio, environnement de développement multiplateforme gratuit et open source pour R, langage de programmation utilisé pour le traitement de données et l'analyse statistique.

## I Traitements statistiques "basiques"

1. Création d'un vecteur  $x$  associé à des données

```
> x=c(8,10,12,14,16)
> x
```

```
[1] 8 10 12 14 16
```

2. Création d'un vecteur associé à une séquence

```
> y=seq(from=10,to=14,by=1)
> y
```

```
[1] 10 11 12 13 14
```

3. Calcul de moyenne, variance et écart-type

```
> mean(x)
> var(x)
> sqrt(var(x))
> sd(x)
```

A noter que la variance obtenue correspond à la variance estimée et non calculée.

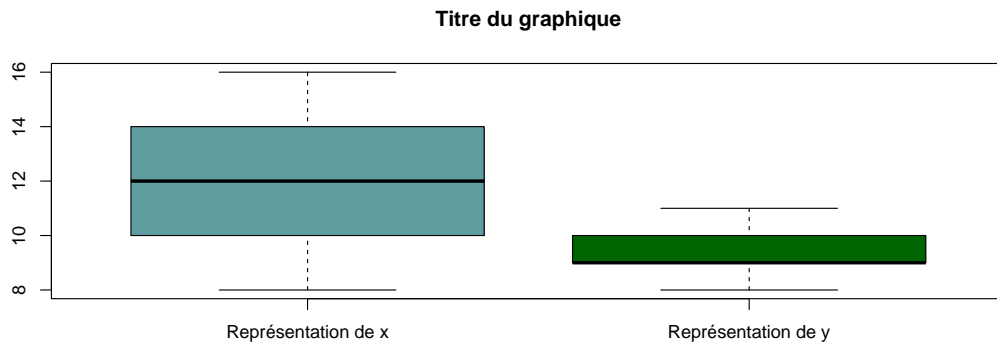
On peut vérifier cela en calculant  $s^2 = \frac{SCE}{n}$  et  $\hat{\sigma}^2 = \frac{SCE}{n-1}$ .

4. Calcul d'une SCE (Somme des Carrés des Ecartés) du type  $\sum_{i=1}^n (x_i - \bar{x})^2$ .

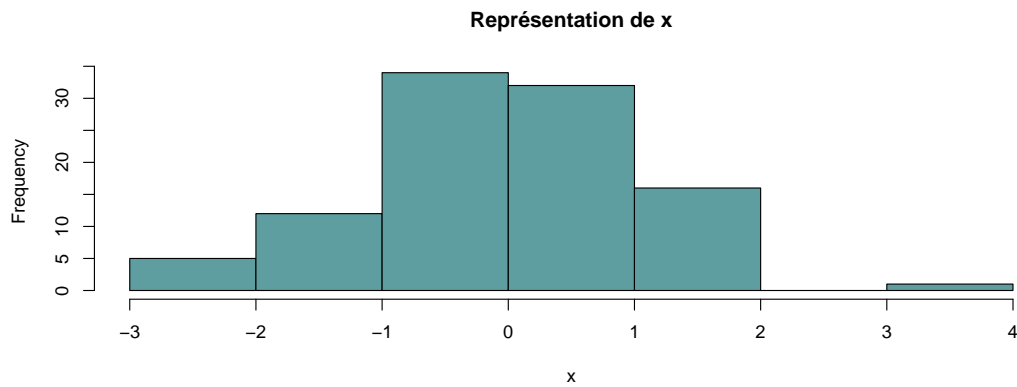
```
> sum( (x-mean(x) )^2)
```

5. Représentation graphique à l'aide d'un box plot, d'un histogramme ou d'un diagramme en barres.

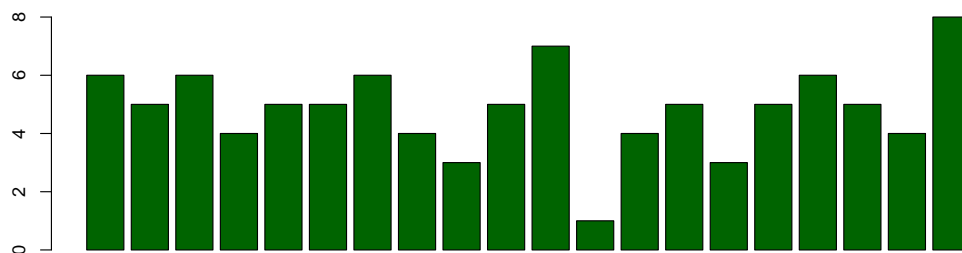
```
> x=c(8,10,12,14,16)
> y=c(8,9,9,10,11)
> boxplot(x,y,main = "Titre du graphique",
+ names=c("Représentation de x","Représentation de y"),col=c("cadetblue","darkgreen")
```



```
> # Génération d'un échantillon de 100 valeurs issues d'une population de loi N(0;1)
> x=rnorm(100,0,1)
> hist(x,col="cadetblue",main="Représentation de x")
```



```
> # Génération d'un échantillon de 20 valeurs issues d'une population de loi
> # binomiale B(10,0.5)
> x=rbinom(n = 20, size = 10, prob = 0.5)
> barplot(x, col="darkgreen")
```



## II Calculs de probabilité

### II.1 Présentation de commandes R

Pour chaque loi, on associe :

- La fonction de densité de probabilité pour les lois continues ou la probabilité d'un évènement élémentaire du type  $P(X = k)$  pour une loi discrète : préfixe d.
- La fonction de répartition : préfixe p.

On rappelle que la fonction de répartition  $F_X$  associée à une variable  $X$  est définie par :

$$F_X(x) = P(X \leq x)$$

- La fonction quantile qui est la fonction "réciproque" de la fonction de répartition : préfixe q. Ainsi, dans le cas d'une loi continue, si  $F_X(x) = p$  avec  $F_X$  strictement croissante sur  $\mathbb{R}$  alors

$$x = F_X^{-1}(p)$$

$x$  est obtenu avec la fonction quantile associée à  $p$ .

- La génération de nombres pseudo-aléatoires : préfixe r.

### II.2 Exemples de lois continues

1. Lois normales : [dpqr]norm( $x$  , mean = ..., sd = ....)

Exemple : Supposons que la variable aléatoire  $X$  suit la loi  $\mathcal{N}(100; 5)$  ( $\sigma = 5$ ).

Déterminer :

1.  $P(X \leq 105)$  et  $P(X \geq 110)$ .
2. la valeur  $x$  telle que :  $P(X \leq x) = 0,975$ .
3. un échantillon de 10 valeurs issues de la loi  $\mathcal{N}(100; 5)$ .

1.

```
> pnorm(105,mean=100,sd=5)
```

```
[1] 0.8413447
```

```
> 1-pnorm(110,mean=100,sd=5)
```

```
[1] 0.02275013
```

2.

```
> qnorm(p=0.975,mean=100,sd=5)
```

```
[1] 109.7998
```

3.

```
> rnorm(n=10,mean=100,sd=5)
```

```
[1] 92.48807 89.91148 92.37504 113.15070 98.59043 101.52364 102.42313
[8] 101.71487 93.43028 100.75702
```

2. Lois de Student : `[dpqr]t(x , ddl)`

Exercice : Supposons que la variable aléatoire  $T$  suit la loi  $\mathcal{T}(10)$ .

Déterminer :

1.  $P(T \geq 2)$ .
2. la valeur  $t$  telle que  $P(T \geq t) = 0.95$ .

1.

```
> pt(2,df=10)
```

```
[1] 0.963306
```

2.  $P(T \geq t) = 0.95 \Leftrightarrow P(T \leq t) = 0.05$ .

```
> qt(p=0.05,df=10)
```

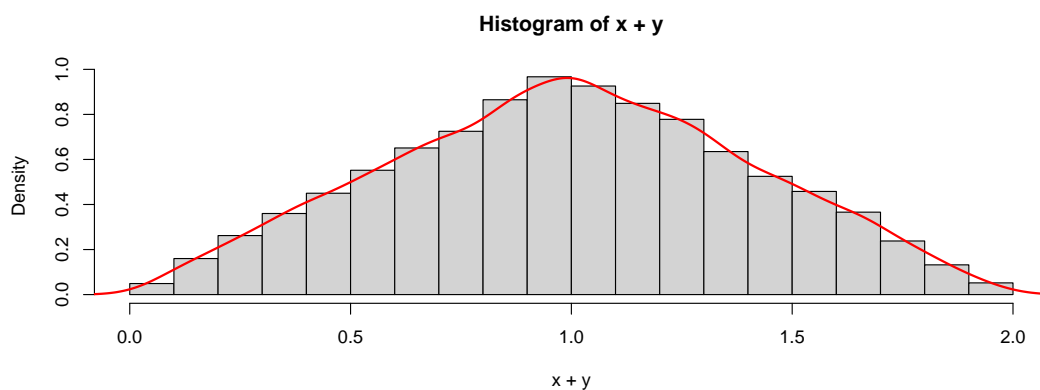
```
[1] -1.812461
```

3. Lois uniformes : `[dpqr]unif(x , min = ..., max = ...)`

Exercice : On considère deux variables aléatoires  $X$  et  $Y$  indépendantes de même loi uniforme sur  $[0; 1]$ .

Déterminer deux échantillons de taille 10000 associés à la loi uniforme sur  $[0; 1]$  et en déduire une conjecture concernant la loi de  $X + Y$ .

```
> x<-runif(n=10000,min=0,max=1)
> y<-runif(n=10000,min=0,max=1)
> densite <- density(x+y)
> hist(x+y,xlim=c(0,2),freq=FALSE)
> lines(densite$x, densite$y, lwd = 2,col="red")
```



On peut conjecturer que la loi de  $X + Y$  est une loi triangulaire de support  $[0; 2]$ .

## II.3 Exemples de lois discrètes

1. Lois binomiales : `[dpqr]binom(k= ... , size = ... , prob=... )`

Exemple : On suppose que  $X$  suit la loi binomiale  $\mathcal{B}(20; 0, 1)$ . Déterminer :

1.  $P(X = 2)$ .
2.  $P(X \leq 2)$ .

1.  $P(X = 2)$  correspond à :

```
> dbinom(x=2,size=20,prob=0.1)
```

```
[1] 0.2851798
```

2.  $P(X \leq 2)$  correspond à :

```
> pbinom(2,size=20,prob=0.1)
```

```
[1] 0.6769268
```

2. Lois de Poisson : `[dpqr]pois(k= ... ,lambda=... )`

Exemple : Si  $X$  suit la loi de Poisson  $\mathcal{P}(1)$  alors  $P(X \leq 2)$  correspond à :

```
> ppois(2,lambda=1)
```

```
[1] 0.9196986
```

### III Exercices

#### Exercice 1 : \*\*

Cet exercice traite de la simulation et tend à illustrer un célèbre théorème : le Théorème Central Limite.

On considère cinq variables indépendantes de loi uniforme sur  $[0; 5]$  notées  $X_1, \dots, X_5$ .

1. Générer un échantillon aléatoire  $a$  de  $n = 10000$  valeurs issues d'une distribution uniforme sur  $[0; 5]$  en utilisant `runif()`.
2. (a) Représenter  $a$  à l'aide d'un histogramme.  
(b) Générer 4 autres échantillons notés  $b, c, d$  et  $e$ .  
(c) Représenter l'échantillon  $a + b + c + d + e$ . Que remarque-t-on ?  
(d) Représenter la densité obtenue avec R via la fonction `density()`.

3. En utilisant la simulation, donner une valeur approchée de  $\mathbb{E} \left( \sum_{i=1}^5 X_i \right)$ .

Quelle est la valeur exacte ?

*On pourra utiliser que l'espérance de la loi uniforme sur  $[a; b]$  est  $\frac{a+b}{2}$ .*

4. En utilisant la simulation, donner une valeur approchée de  $\mathbb{V} \left( \sum_{i=1}^5 X_i \right)$ .

Quelle est la valeur exacte ?

*On pourra utiliser (et éventuellement démontrer) que la variance de la loi uniforme sur  $[a; b]$  est  $\frac{(b-a)^2}{12}$ .*

#### Exercice 2 : \*\*

Une grande entreprise organise un voyage de groupe à Marrakech pour récompenser certains de ses collaborateurs.

On admet que la probabilité qu'une personne invitée se rende à ce séjour est égale à 0,9.

1. L'entreprise a décidé de récompenser 10 salariés de la région bordelaise.
  - (a) Quelle est la probabilité qu'ils se rendent tous à ce séjour ?
  - (b) Quelle est la probabilité qu'au moins 8 personnes se rendent à ce séjour ?
2. L'entreprise a décidé de récompenser 200 salariés sur toute la France.
  - (a) Donner une estimation du nombre de salariés qui participeront à ce séjour.
  - (b) Déterminer la probabilité qu'il y ait entre 170 et 190 participants à ce séjour.

#### Exercice 3 : \*\*

Le département qualité d'une production en grande série de circuits électroniques a évalué à 3% le nombre de circuits non valides.

Un client reçoit un lot de 500 circuits.

1. En utilisant le TCL, déterminer la probabilité qu'il reçoive au plus 1% de circuits non valides dans son lot.  
Quelle est la valeur exacte ?

2. Par contrat, le client peut renvoyer le lot si celui-ci comporte plus de 5% de circuits non valides. Quelle est la probabilité qu'il soit conduit à renvoyer le lot ?  
On utilisera une loi exacte et une loi approchée.

**Exercice 4 : \*\* Lois et simulation**

On vous propose le jeu suivant :

On lance un dé.

Si le résultat est 5 ou 6, on gagne 3 €.

Si le résultat est 4, on gagne 1 € et si on obtient au plus 3, on perd 2,5 €.

1. Avant d'accepter la partie, simuler une partie puis 1000 parties.  
On pourra utiliser la fonction `sample(x=..., n=...)` pour générer un échantillon ainsi que la fonction conditionnelle `ifelse()`.
2. Proposer une représentation graphique.
3. Quelle est l'espérance de gain pour un joueur lors d'une partie ? lors de 1000 parties ?  
*On donnera des estimations puis les valeurs exactes.*

**Exercice 5 : \*\***

Une machine calibre des tomates avec une valeur cible de 150 grammes.

On admet que la variable aléatoire  $Y$  égale à la masse, en grammes, d'une tomate prise au hasard, est distribuée suivant une loi normale  $\mathcal{N}(150; \sigma)$ .

1. Déterminer la valeur de  $\sigma$  sachant que 84 % des tomates ont une masse inférieure à 160 g.
2. On prend, pour cette partie,  $\sigma = 10$ .
  - (a) Quelle est la probabilité qu'une tomate prise au hasard pèse moins de 130 g ?
  - (b) Quelle est la probabilité qu'une tomate prise au hasard pèse plus de 170 g ?
  - (c) Quelle est la probabilité qu'une tomate prise au hasard pèse entre 145 g et 160 g ?
3. Quelle devrait être la valeur de l'écart-type pour que 95 % des tomates aient un poids compris entre 145 g et 155 g ?

**Exercice 6 : \*\***

Dans une boulangerie industrielle, on produit des brioches dont la quantité nominale est égale à 210 grammes.

L'entreprise souhaite réaliser un contrôle statistique sur la masse moyenne des brioches d'une fabrication.

On suppose que la variable aléatoire  $X$  prenant pour valeur la masse, exprimée en grammes, d'une brioche prélevée au hasard dans cette fabrication est distribuée selon une loi normale de moyenne  $\mu$  et d'écart type  $\sigma$  connu égal à 4.

On extrait un échantillon aléatoire simple de taille 9.

La masse moyenne, exprimée en grammes, des brioches de cet échantillon est égale à 208,4 g.

Au vu de ces résultats, doit-on considérer, au seuil de risque 0,05, que la masse moyenne des brioches de cette fabrication est inférieure à 210 grammes ?

**Exercice 7 : \*\***

Le pH est un des éléments parmi les plus importants permettant de caractériser un vin (acidité, vivacité , ...). Afin de contrôler le pH de vins issus d'un même cépage (Merlot), on a effectué des mesures récapitulés ci-dessous :

3,40 3,44 3,46 3,46 3,49 3,54 3,45 3,45 3,47 3,49 3,5 3,51

1. Déterminer une estimation ponctuelle du pH moyen et de la variance associés à ce cépage.
2. On admet que la distribution des pH est normale.  
Donner une estimation par intervalle de confiance du pH moyen de la production dont est extrait l'échantillon considéré.  
Le pH moyen observé est-il significativement différent de 3,45?

**Exercice 8 : \*\***

Sur un échantillon de 1000 personnes, 72% d'entre elles déclarent qu'elles iront voter au second tour d'une élection présidentielle.

En outre, parmi ces dernières, 51,8% déclarent qu'elles vont voter pour le candidat *A* lors de cette élection.

1. Donner une estimation, par intervalle de confiance, de la proportion de personnes, dans la population dont est extrait cet échantillon, qui vont voter pour la candidat *A*.
2. Peut-on considérer, avec un risque d'erreur inférieur à 5%, que le candidat *A* sera élu?