

Solutions TD1

Statistique descriptive

Série statistique simple

Exercice 1

Au cours d'une étude du métabolisme chez le rat, des mesures de masse de carcasse, de proportion de lipide, protéine, glucide et d'eau dans leur alimentation sont effectuées sur plusieurs lot d'animaux des deux sexes.

1) Charger et lire les données dans R (fichier Rat.csv)

```
# Lecture des données dans R avec stockage
rat<-read.table('Rat.txt',header=TRUE)
```

On vérifie que l'effectif de l'échantillon et le nombre de variables est correct (152 individus X 8 variables).

```
> dim(rat)
[1] 152 8
```

On vérifie que les variables sont bien numériques

```
>is.numeric(rat$PPROT)
TRUE
```

Tableau 1

Nom	Définition	Exemple
Individu	Sujet étudié	Un rat
Échantillon	Sous-ensemble des individus	158 rats
Variable(s) /modalités	Ce qui est étudié chez les individus (et qui varie d'un individu à l'autre)/Liste de toutes les valeurs possibles pour une variable	7 variables ¹ Sexe /1 et 2; LOT 1 à 6;PLIP / 5,8 à 67,9; PGLU 7,1 à 82,1; Eau / 21,2 à 71,9 ;PPROT / 7,6 à 43,8 ;CarcassW / 34 à 220
Nature de la variable	Type de variables	Nominales : Sexe, Lot Quantitatives continues : PLIP ; PGLU ; Eau;PPROT;CarcassW
Variable(s) dépendante(s) /expliquée(s)	Type de variables	PLIP ; PGLU ; Eau;PPROT;CarcassW ²

¹Le numéro de l'individu est ici considéré comme un index.

²C'est la seule variable dépendante de toutes les autres. Cela dépend de l'hypothèse de travail pour cette étude.

3) Donner l'effectif par sexe avec R (tableau d'effectifs)

```
> table(rat$SEXE)
```

```
 1 2  
74 78
```

4) Déterminer les paramètres statistiques de chaque variable

On calcule des principaux paramètres de position ce qui nous permet de vérifier qu'il n'y a pas d'erreur de saisie dans le tableau de données « Rat ».

```
> summary(rat)
```

```
  Animaux      SEXE      LOT      PPROT  
Min.   : 1.00  Min.   :1.000  Min.   :1.000  Min.   : 7.653  
1st Qu.: 40.75 1st Qu.:1.000  1st Qu.:2.000  1st Qu.:23.842  
Median :114.50 Median :2.000  Median :4.000  Median :28.849  
Mean   :100.07 Mean   :1.513  Mean   :3.605  Mean   :28.488  
3rd Qu.:152.25 3rd Qu.:2.000  3rd Qu.:5.000  3rd Qu.:33.210  
Max.   :191.00 Max.   :2.000  Max.   :6.000  Max.   :43.840  
  PLIP      PGLU      Eau      CarcassW  
Min.   : 5.847  Min.   : 7.112  Min.   :21.20  Min.   : 33.64  
1st Qu.:18.292 1st Qu.:34.384 1st Qu.:33.73 1st Qu.: 95.57  
Median :25.553 Median :45.776  Median :39.89  Median :115.57  
Mean   :26.999 Mean   :44.513  Mean   :41.32  Mean   :135.81  
3rd Qu.:34.196 3rd Qu.:53.528 3rd Qu.:46.64 3rd Qu.:176.86  
Max.   :67.963 Max.   :82.142  Max.   :71.98  Max.   :220.14
```

5) Tracer le box plot pour toutes les variables expliquées

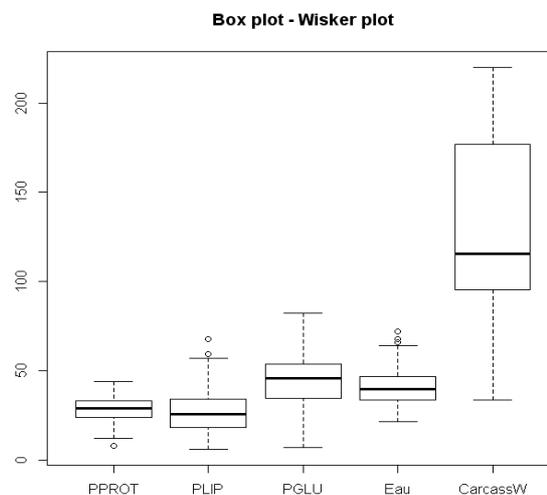
Avec toutes les lignes ; les colonnes de 4 à 8

```
> boxplot(rat[,4:8])
```

```
> title("Box plot - Wisker plot")
```

Avec la fonction aide afin de tenter de tracer un boxplot horizontal par exemple

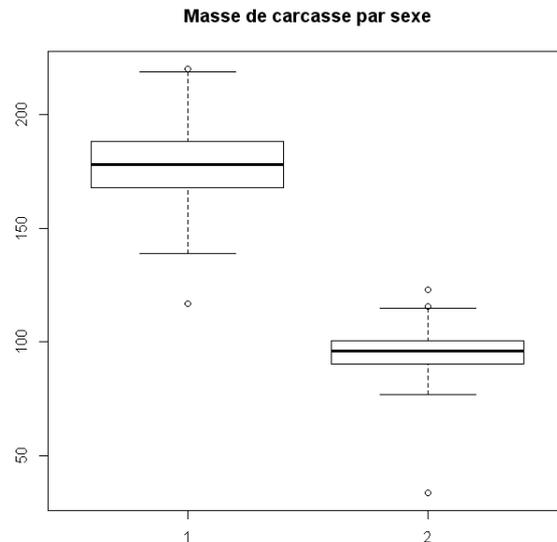
On remarque quelques valeurs aberrantes pour PLIP et l'eau. La variation la plus importante est pour le poids de carcasse. Cela dépend peut-être du sexe.



6) Tracer le box plot de la masse de carcasse en fonction du sexe de l'animal

On remarque que le sexe 1 a une masse de carcasse plus importante que le sexe 2. La variabilité individuelle (moustache) est réduite par rapport au cas précédent (tous sexes confondus)

```
> boxplot(rat$CarcassW~rat$SEXE, data=rat)
> title("Masse de carcasse par sexe")
```



Avec la fonction `aid` afin de tenter de tracer un boxplot avec un argument titre `boxplot(rat$CarcassW~$SEXE,data=rat, main='Masse de carcasse par sexe')`

7) Tracer l'histogramme d'une variable expliquée

Cette représentation nous permet de vérifier intuitivement que la distribution de la variable suit une loi normale. Ce sera à vérifier avec le test de Shapiro par exemple.

```
> hist(rat$PLIP,main="Histogramme",xlab="Proportion de lipide")
```

