

Tests d'hypothèses

Compétence 2 - Ressource 3.06

Fiche TP n°1

Objectif. Le but de ce TP est de vous présenter les outils numériques permettant de réaliser un **test d'hypothèses sur une moyenne** au travers du langage **R** et du logiciel **RStudio**. On va distinguer deux approches : l'approche paramétrique et l'approche non-paramétrique.

Approche paramétrique

Contexte. Un médecin souhaite démontrer l'impacte d'un régime riche en fibre sur le poids de personnes en situation d'obésité. Autrement dit, il souhaite démontrer que ce régime amène à une baisse significative du poids du patient.

Dans ce cadre, un échantillon de 25 patients volontaires en situation d'obésité avec un poids égal à 100 kilogrammes a été constitué, et pour lequel une portion spécifique de nourriture a été donnée (matin, midi et soir) pendant 6 mois à chacune des personnes le composant. A la fin de cette période, le poids de chacun des participant a été mesuré et enregistré dans un fichier de données nommé `Etude.fibre.csv` dont voici le contenu :

```
Ident;Poids
1;101.59
2;92.88
3;95.84
4;90.54
5;91.07
6;94.65
7;91
8;95.08
9;90.41
10;98.08
11;102.89
12;96.14
13;94.55
14;98.41
15;93.85
16;93.06
17;93.58
18;90.02
19;101.47
20;94.47
21;95.44
22;92.85
23;95.92
24;98.07
25;100.64
```



Ce fichier est disponible sur la plateforme E-Campus

<https://ecampus.unicaen.fr>



Réalisation de l'étude

• Étape 1 : préambule théorique à l'étude

1. Identifier la population d'intérêt dans le cas présent.
2. Identifier la variable d'intérêt dans le cas présent. Quel est le type de cette variable ?
3. Le médecin propose de réaliser le test d'hypothèses suivant :

$$H_0 : \mu = \mu_0 \quad \text{versus} \quad H_1 : \mu < \mu_0$$

Que représente μ_0 dans le cas présent ? Justifier l'écriture de l'hypothèse alternative selon le contexte. Comment se nomme ce test d'hypothèses ?

• Étape 2 : préliminaires à l'étude

1. Créer sur votre espace de travail un dossier nommé **Etude_Poids**, puis créer les sous-dossiers **Data** et **RStudio**.
2. Télécharger le fichier **Etude_fibre.csv**, puis le placer dans le sous-dossier **Data**.
3. Ouvrir le logiciel **RStudio**, puis créer un projet en choisissant le sous-dossier **RStudio** comme dossier de travail par défaut.
4. Ouvrir un fichier script (raccourci clavier **Ctrl+Shift+N**), puis indiquer l'en-tête suivant :

```
##### |  
# Formation : BUT Science des Données - IUT Grand Ouest Normandie  
# Niveau : SD2 - Ressource Tests d'hypothèses  
# Thématique : test d'hypothèses sur une moyenne  
# Auteur : Alain Lucas  
# Date : 25/10/2023  
# Source : Données obtenues par simulation  
##### |
```

5. Enfin, télécharger (si nécessaire) et charger en mémoire les librairies suivantes :

```
# Chargement des librairies  
  
library(dplyr)  
library(ggplot2)  
  
library(LearningStats) # Mean.test()  
library(report) # report()  
library(ggpubr) # ggqqplot()  
library(rstatix) # t_test()
```

Exécuter, puis vérifier le succès de l'opération.

• Étape 3 : importation des données

1. Utiliser la fonction **readLines()** pour afficher les 10 premières lignes du fichier, et ainsi déterminer sa structure.

```
[1] "Ident;Poids" "1;101.59"    "2;92.88"      "3;95.84"      "4;90.54"      "5;91.07"  
[7] "6;94.65"     "7;91"         "8;95.08"      "9;90.41"
```

Identifier en particulier : la présence ou non de commentaires, la présence ou non du nom des champs et enfin le type de séparateur de champs.

2. A l'aide de la fonction **read.table()**, importer les données dans un objet nommé **data**.
3. Afin de s'assurer que les données ont été correctement lues et que le type intrinsèque des variables est convenable, afficher à l'aide de fonctions spécifiques les informations suivantes :

	Ident	Poids
1	1	101.59
2	2	92.88
3	3	95.84
4	4	90.54
5	5	91.07
6	6	94.65

```
'data.frame': 25 obs. of 2 variables:
 $ Ident: int 1 2 3 4 5 6 7 8 9 10 ...
 $ Poids: num 101.6 92.9 95.8 90.5 91.1 ...
```

A-t-on besoin de modifier le type des données dans le cas présent ? Justifier votre réponse.

• Étape 4 : analyse exploratoire des données

1. On se propose dans un premier temps de déterminer les caractéristiques numériques des données. Pour cela, écrire par exemple les instructions suivantes :

```
# Analyse exploratoire des données

data %>%
  summary()
```

Exécuter, puis visualiser la sortie textuelle.

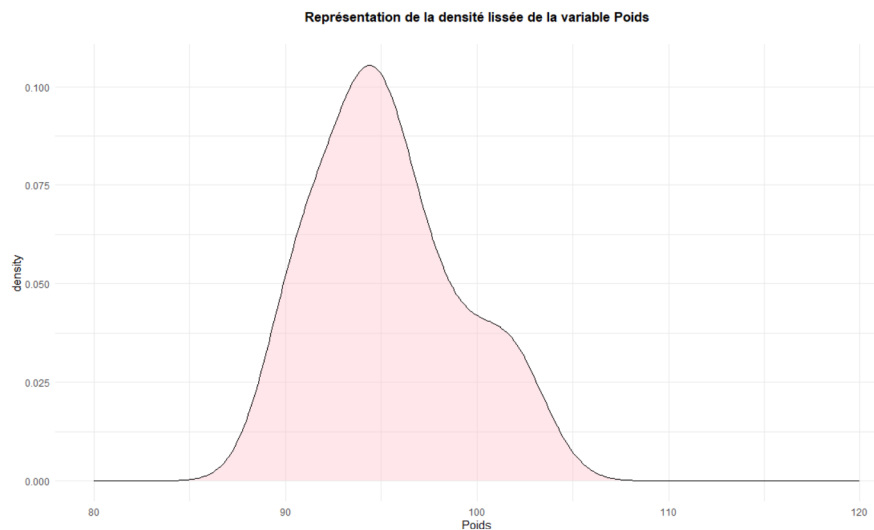
Ident	Poids
Min. : 1	Min. : 90.02
1st Qu.: 7	1st Qu.: 92.88
Median :13	Median : 94.65
Mean :13	Mean : 95.30
3rd Qu.:19	3rd Qu.: 98.07
Max. :25	Max. :102.89

Les indicateurs numériques de la variable **Ident** ont-ils un intérêt ? Justifier votre réponse. Pour la variable d'intérêt, observe-t-on la présence de données manquantes ? Pourquoi ? Quelle est le poids moyen observé sur l'échantillon ? Cela permet-il de justifier de l'intérêt de mener le test d'hypothèses ? Justifier votre réponse.

2. La taille de l'échantillon étant inférieure à 30, le médecin souhaite s'orienter vers un test paramétrique, i.e. avec l'hypothèse d'une distribution gaussienne de la variable **Poids**. Dans un premier temps, il se propose de représenter la densité lissée pour cette variable.

```
data %>%
  ggplot(mapping = aes(x = Poids))+
  geom_density(fill = "pink",
    alpha = 0.4)+
  scale_x_continuous(limits = c(80,120))+
  theme_minimal()+
  theme(plot.title = element_text(hjust = 0.5,
    face = "bold"))+
  labs(title = "Représentation de la densité lissée de la variable Poids\n",
    xlab = "Poids en kilogramme",
    ylab = "Densité")
```

Exécuter ce code, puis visualiser la sortie graphique.



Selon vous peut-on, au vu de cette courbe, rejeter l'hypothèse d'une distribution gaussienne ? Justifier votre réponse.

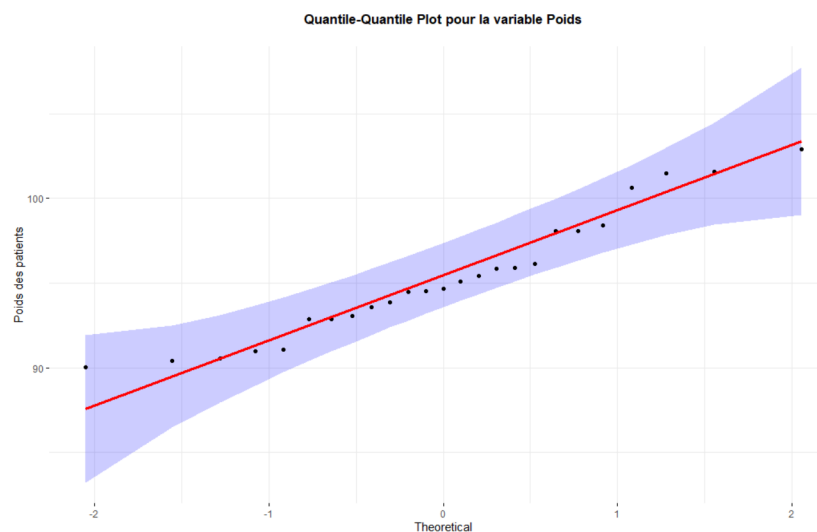
Alternative. On pourrait également faire usage de la fonction `ggdensity()` de la librairie `ggpubr` de la manière suivante :

```
data %>%
  ggdensity(x = "Poids",
            fill = "pink",
            alpha = 0.4)+
  scale_x_continuous(limits = c(80,120))+
  theme_minimal()+
  theme(plot.title = element_text(hjust = 0.5,
                                   face = "bold"))+
  labs(title = "Représentation de la densité lissée de la variable Poids\n",
       xlab = "Poids en kilogramme",
       ylab = "Densité")
```

3. La deuxième étape va consister à représenter le **Quantile-Quantile Plot** dont le nuage de points devrait, selon un cadre gaussien, être proche d'un alignement selon une droite. Pour cela, on va faire usage de la fonction `ggqqplot()` de la librairie `ggpubr`

```
data %>%
  ggqqplot(data = .,
            x = "Poids",
            color = "blue",
            conf.int = TRUE,
            conf.int.level = 0.95,
            title = "Quantile-Quantile Plot pour la variable Poids\n",
            ylab = "Poids des patients",
            ggtheme = theme_minimal())+
  geom_qq(color = "black") + # Personnaliser la couleur des points
  geom_qq_line(color = "red",
               linewidth = 1.2)+
  theme(plot.title = element_text(hjust = 0.5,
                                   face = "bold"))
```

Exécuter, puis visualiser la sortie graphique.



Selon vous peut-on, au vu de cette courbe, rejeter l'hypothèse d'une distribution gaussienne? Justifier votre réponse.

4. Enfin, on peut réaliser une approche plus objective avec un test de normalité de Shapiro-Wilk dont les hypothèses sont les suivantes :

H_0 : la distribution de X est gaussienne versus H_1 : la distribution de X n'est pas gaussienne

Pour cela, on pourra faire usage de la fonction `Shapiro.test()` comme suit :

```
data %>%
  pull(Poids) %>%
  shapiro.test(x = .)
```

Exécuter, puis visualiser la sortie textuelle.

Shapiro-wilk normality test

```
data: .  
W = 0.9449, p-value = 0.192
```

Peut-on avec un niveau de signification de 5% rejeter l'hypothèse d'une distribution gaussienne de la variable Poids? Justifier votre réponse.

• Étape 5 : réalisation du test d'hypothèses

Maintenant que l'hypothèse gaussienne est validée, on peut se tourner vers la mise en œuvre du test d'hypothèses d'intérêt selon un cadre paramétrique. Pour cela, plusieurs fonctions sont disponibles sous **R** dans différentes librairies.

1. La réalisation du test peut se faire avec la fonction de base `t.test()` comme suit :

```
data %>%  
  pull(Poids) %>%  
  t.test(x = .,  
         alternative = "less",  
         mu = 100)
```

Exécuter, puis visualiser la sortie textuelle.

One Sample t-test

```
data: .  
t = -6.3888, df = 24, p-value = 6.602e-07  
alternative hypothesis: true mean is less than 100  
95 percent confidence interval:  
-Inf 96.55864  
sample estimates:  
mean of x  
95.3
```

Peut-on rejeter l'hypothèse nulle avec un niveau de signification de 5%? Justifier votre réponse. En déduire une conclusion contextuelle sur le régime à base de fibre pour la perte de poids chez les personnes en situation d'obésité.

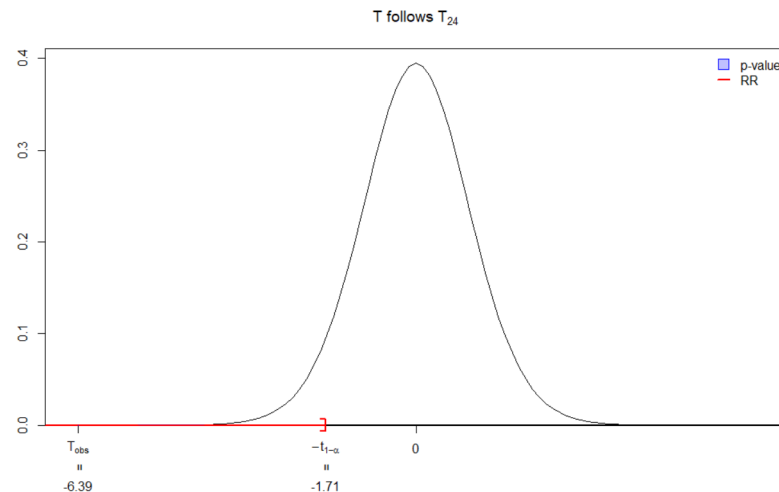
2. Une alternative consiste à faire usage de la fonction `Mean.test()` de la librairie **LearningStats**

```
data %>%  
  pull(Poids) %>%  
  Mean.test(x = .,  
            mu0 = 100,  
            sigma = NULL,  
            alternative = "less",  
            alpha = 0.05,  
            plot = TRUE)
```

Exécuter, puis visualiser la sortie textuelle et la sortie graphique.

Test for the mean of a Normal population with unknown variance

```
H0:  $\mu \geq 100$  units  
Ha:  $\mu < 100$  units  
T = (bar.x -  $\mu_0$ ) / (Sc /  $\sqrt{n}$ )  
T  $\in$  T_{n-1}  
 $\alpha$  = 0.05  
T_obs = -6.38878  
RR = (- $\infty$ , -1.71088]  
p-value = 0
```



Observer que vous obtenez évidemment les mêmes résultats et par conséquent la même conclusion. Que représente le segment **RR** représenté en rouge sur le graphique ?

- Une autre alternative, intéressante pour afficher les indicateurs numériques sur un graphique, consiste à faire usage de la fonction `t_test()` de la librairie **rstatix**

```
data %>%
  t_test(data = .,
         formula = Poids ~ 1,
         alternative = "less",
         mu = 100,
         conf.level = 0.95) -> result3

print(result3)
```

Exécuter, puis visualiser la sortie textuelle.

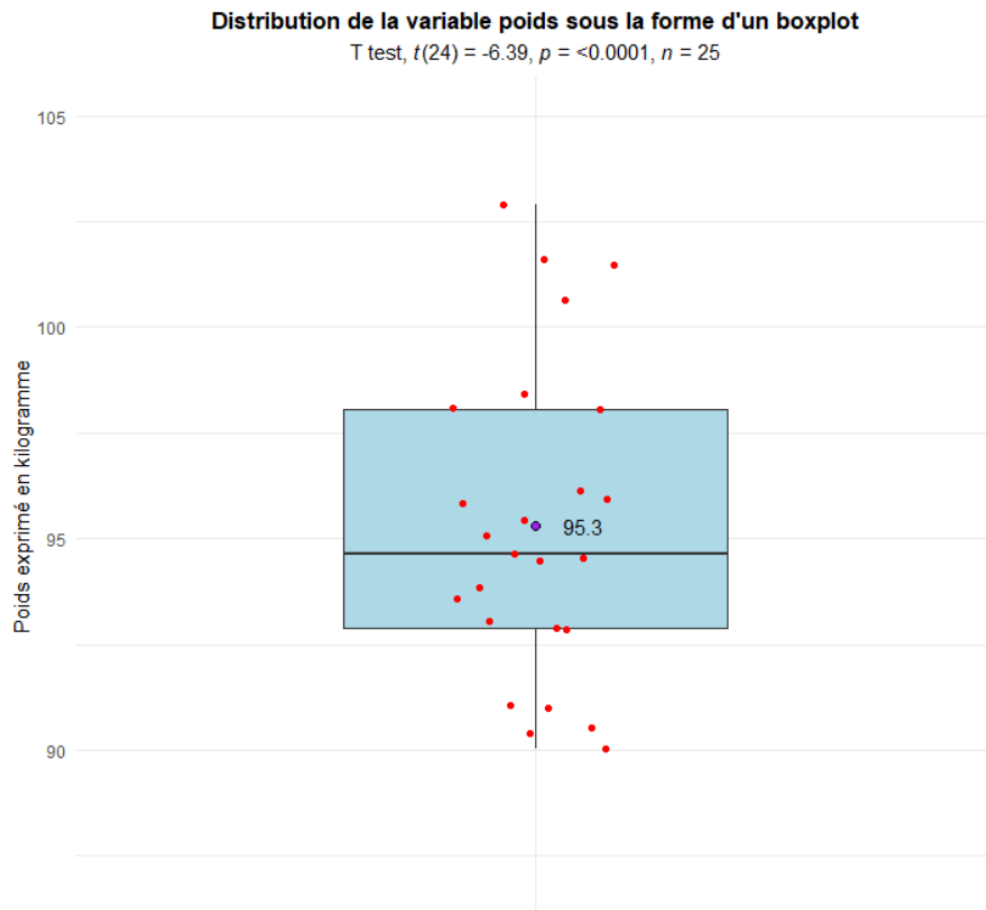
```
# A tibble: 1 x 7
  .y. group1 group2      n statistic    df      p
*   <chr> <chr>   <chr>   <int>   <dbl> <dbl>   <dbl>
1 Poids 1      null model    25    -6.39    24 0.00000066
```

Remarque. L'intérêt de cette sortie est que les indicateurs sont contenus dans un objet de classe **tibble** lequel est très facile à manipuler comme on va le voir par la suite ...

- On peut enfin faire apparaître simultanément la distribution de la variable **Poids** et le résultat du test de moyenne sur un même graphique. On peut soit utiliser l'approche standard suivante, où l'on trouvera néanmoins l'utilisation de la fonction `get_test_label()` de la librairie **rstatix** :

```
data %>%
  ggplot(mapping = aes(y = Poids,
                       x = NA))+
  geom_boxplot(width = 0.5,
               fill = "lightblue")+
  geom_jitter(width = 0.12,
              colour = "red")+
  stat_summary(fun = "mean",
               geom = "point",
               color = "black",
               fill = "purple",
               shape = 21,
               size = 2)+
  stat_summary(fun = "mean",
               colour = "black",
               geom = "text",
               vjust = +0.5,
               hjust = -0.7,
               mapping = aes(label = round(after_stat(y),
                                           digits = 1)))+
  scale_y_continuous(limits = c(87,105))+
  theme_minimal()+
  theme(axis.text.x=element_blank(),
        axis.ticks.x=element_blank(),
        plot.title = element_text(hjust = 0.5,
                                   face = "bold"),
        plot.subtitle = element_text(hjust = 0.5))+
  labs(title = "Distribution de la variable poids sous la forme d'un boxplot",
       subtitle = get_test_label(stat.test = result3,
                                detailed = TRUE),
       x = NULL,
       y = "Poids exprimé en kilogramme")
```

Exécuter, puis visualiser la sortie graphique.

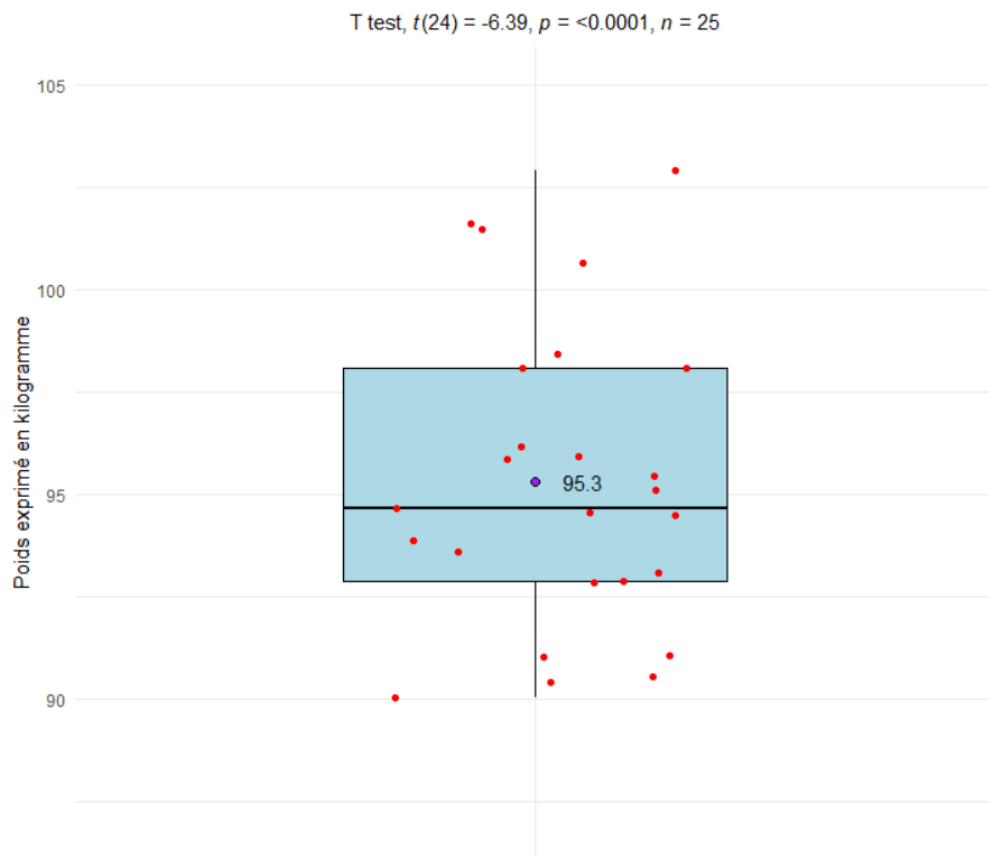


Une autre solution consiste à faire usage de la fonction `ggboxplot()` de la librairie `ggpubr`

```
data %>%
  mutate(y = NA) %>%
  ggboxplot(x = "y",
            y = "Poids",
            width = 0.5,
            fill = "lightblue",
            add = c("jitter"),
            add.params = list(color = c("red")),
            ylab = "Poids exprimé en kilogramme")+
  stat_mean(geom = "point",
            colour = "black",
            fill = "purple",
            shape = 21,
            size = 2)+
  stat_mean(geom = "text",
            colour = "black",
            vjust = +0.5,
            hjust = -0.7,
            mapping = aes(label = round(after_stat(y),
                                      digits = 1)))+
  scale_y_continuous(limits = c(87,105))+
  theme_minimal()+
  theme(plot.title = element_text(hjust = 0.5,
                                  face = "bold"),
        plot.subtitle = element_text(hjust = 0.5),
        axis.text.x=element_blank(),
        axis.ticks.x=element_blank())+
  labs(title = "Distribution de la variable poids sous la forme d'un boxplot\n",
        subtitle = get_test_label(stat.test = result3,
                                  detailed = TRUE),
        x = NULL)
```

Exécuter, puis visualiser la sortie graphique.

Distribution de la variable poids sous la forme d'un boxplot



On obtient bien sur dans l'un et l'autre cas une sortie graphique similaire et cohérente.

Approche non paramétrique

Contexte. On dispose sur la plateforme **kaggle** d'un fichier de données en relation avec le poids et la taille de 500 individus exprimés respectivement en kg et en cm.

Dans la suite, on se propose de faire usage de ce jeu de données pour réaliser le tirage aléatoire de 40 individus, puis de réaliser le test d'hypothèses suivant :

$$H_0 : \mu = \mu_0 \quad \text{versus} \quad H_1 : \mu > \mu_0$$

où μ représentera le poids théorique moyen de l'ensemble des 500 individus supposés être la population d'intérêt, et où μ_0 sera égale à la valeur 100 kg (cliquer sur l'illustration ci-dessous pour accéder à la page web)



500 Person Gender-Height-Weight-Body Mass Index

Height and Weight random generated, Body Mass Index Calculated



Les données relatives à cette étude sont disponibles sur la plateforme E-Campus de l'Université de Caen Normandie au sein d'un fichier nommé `500_Person_Gender_Height_Weight_Index`.

<https://ecampus.unicaen.fr>

Réalisation de l'étude

• Étape 1 : préambule théorique à l'étude

1. Identifier la population d'intérêt dans le cas présent.
2. Identifier la variable d'intérêt dans le cas présent. Quel est le type de cette variable ?
3. Comment se nomme le test d'hypothèses que l'on est amené à réaliser ?

• Étape 2 : préliminaires à l'étude

1. Créer sur votre espace de travail un dossier nommé `Etude_kaggle`, puis créer les sous-dossiers `Data` et `RStudio`.
2. Télécharger le fichier `500_Person_Gender_Height_Weight_Index`, puis le placer dans le sous-dossier `Data`.
3. Ouvrir le logiciel **RStudio**, puis créer un projet en choisissant le sous-dossier `RStudio` comme dossier de travail par défaut.
4. Ouvrir un fichier script (raccourci clavier `Ctrl+Shift+N`), puis indiquer l'en-tête suivant :

```
##### |
# Formation : BUT Science des Données - IUT Grand Ouest Normandie
# Niveau : SD2 - Ressource Tests d'hypothèses
# Thématique : test d'hypothèses sur une moyenne [cadre non paramétrique]
# Auteur : Alain Lucas
# Date : 25/10/2023
# Source : kaggle - 500 Person Gender-Height-Weight-Body Mass Index
##### |
```

5. Enfin, télécharger (si nécessaire) et charger en mémoire les librairies suivantes :

```
# Chargement des librairies

library(dplyr)
library(ggplot2)

library(rstatix) # get_summary_stats()
library(ggpubr) # ggqqplot()
```

Exécuter, puis vérifier le succès de l'opération.

• Étape 3 : importation des données

1. Utiliser une fonction adéquate pour afficher les 10 premières lignes du fichier, et ainsi déterminer sa structure.

```
[1] "Gender,Height,Weight,Index" "Male,174,96,4" "Male,189,87,2"
[4] "Female,185,110,4" "Female,195,104,3" "Male,149,61,3"
[7] "Male,189,104,3" "Male,147,92,5" "Male,154,111,5"
[10] "Male,174,90,3"
```

Identifier en particulier : la présence ou non de commentaires, la présence ou non du nom des champs et enfin le type de séparateur de champs.

2. A l'aide d'une fonction usuelle de **R** base, importer les données dans un objet nommé **data**.
3. Afin de s'assurer que les données ont été correctement lues et que le type intrinsèque des variables est convenable, afficher à l'aide de fonctions spécifiques les informations suivantes :

	Gender	Height	Weight	Index
1	Male	174	96	4
2	Male	189	87	2
3	Female	185	110	4
4	Female	195	104	3
5	Male	149	61	3
6	Male	189	104	3

```
'data.frame': 500 obs. of 4 variables:
 $ Gender: chr "Male" "Male" "Female" "Female" ...
 $ Height: int 174 189 185 195 149 189 147 154 174 169 ...
 $ Weight: int 96 87 110 104 61 104 92 111 90 103 ...
 $ Index : int 4 2 4 3 3 3 5 5 3 4 ...
```

Le type proposée par défaut par **R** pour chacune des variables est-il cohérent avec le type intrinsèque ?

• Étape 4 : préparation des données

1. On se propose de modifier le type de la variable **Gender** afin de lui attribuer le type **factor** et non **character**. Dans ce but, écrire le code suivant :

```
# Préparation des données

data %>%
  mutate(Gender = factor(Gender)) %>% str()
```

Exécuter, puis vérifier le succès de l'opération.

```
'data.frame': 500 obs. of 4 variables:
 $ Gender: Factor w/ 2 levels "Female","Male": 2 2 1 1 2 2 2 2 2 1 ...
 $ Height: int 174 189 185 195 149 189 147 154 174 169 ...
 $ Weight: int 96 87 110 104 61 104 92 111 90 103 ...
 $ Index : int 4 2 4 3 3 3 5 5 3 4 ...
```

Modifier alors le code précédent afin d'enregistrer ce changement dans l'objet **data**.

2. On se propose maintenant d'obtenir quelques indicateurs statistiques sur les variables. Pour cela, on va faire usage de la fonction **get_summary_stats()** de la librairie **rstatix**

```
# Statistique Exploratoire

data %>%
  get_summary_stats()
```

Exécuter, puis visualiser la sortie textuelle.

```
# A tibble: 3 x 13
  variable n min max median q1 q3 iqr mad mean sd se ci
  <fct> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 Height 500 140 199 170. 156 184 28 20.0 170. 16.4 0.732 1.44
2 Weight 500 50 160 106 80 136 56 41.5 106 32.4 1.45 2.84
3 Index 500 0 5 4 3 5 2 1.48 3.75 1.36 0.061 0.119
```

Observe-t-on la présence de données manquantes ? Quelle est le poids moyen théorique pour la variable **Weight** dans le cas présent ?

3. On va maintenant réaliser un tirage aléatoire sans remise de 40 individus au sein de la population des 500 individus. Pour cela, écrire le code suivant :

```
# Sélection d'un échantillon de taille n = 40

set.seed(seed = 125)

data %>%
  sample_n(size = 40,
           replace = FALSE) -> sample

print(sample)
```

Exécuter, puis visualiser le succès de l'opération.

	Gender	Height	Weight	Index
1	Male	181	139	4
2	Female	179	150	5
3	Male	141	85	5
4	Female	167	85	4
5	Male	188	57	1
6	Male	165	62	2
7	Female	168	87	4
8	Male	147	84	4
9	Male	190	50	0
10	Female	161	145	5

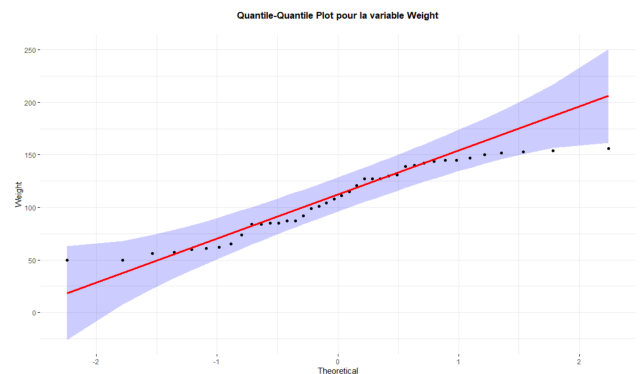
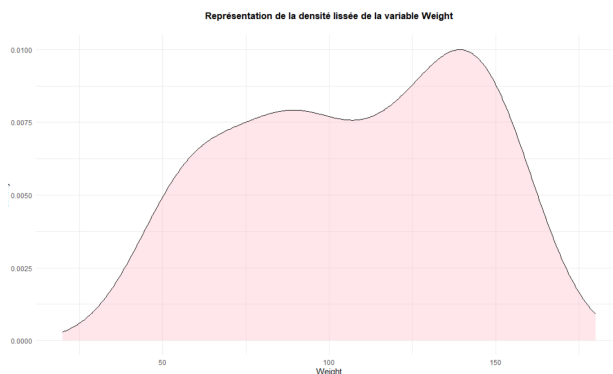
• Étape 5 : analyse exploratoire des données

1. On se propose de déterminer les indicateurs statistiques usuels de cet échantillon. Pour cela, écrire un code permettant d'obtenir la sortie textuelle suivante :

```
# A tibble: 3 × 13
  variable     n  min  max median   q1   q3  iqr  mad  mean  sd  se   ci
  <fct>   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 Height     40   141   195   178.   164   186.  22.2  14.8  174.   15.3  2.41  4.88
2 Weight     40    50   156   110.    84   140.  56.5  44.5  108.   34.5  5.45  11.0
3 Index      40     0     5     4     3     5     2   1.48   3.68   1.49  0.236  0.477
```

L'information apportée par cette sortie textuelle vient-elle confirmer l'intérêt du test d'hypothèses? Justifier votre réponse.

2. On souhaite maintenant étudier la distribution de la variable **Weight** afin de savoir si elle pourrait être assimilée à une loi gaussienne. Dans ce but, écrire un code **R** permettant d'obtenir les deux graphiques suivants :



A-t-on ici des arguments permettant de rejeter l'hypothèse d'une distribution gaussienne? Afin d'avoir une approche plus objective, écrire des instructions permettant de réaliser un test de normalité de Shapiro-Wilk

```
Shapiro-wilk normality test

data: .
W = 0.92274, p-value = 0.009401
```

Que peut-on en déduire dans le cas présent avec un niveau de signification de 5%?

• Étape 6 : réalisation du test d'hypothèses

En l'état l'application de l'approche paramétrique avec le **t-test** ne semble pas cohérent car l'hypothèse d'une distribution gaussienne n'est pas validée. On se propose dans de réaliser l'approche non paramétrique laquelle utilise le **Théorème de la Limite Centrale** (cf. cours).

A ma connaissance, il n'existe pas de fonction sous **R** permettant de réaliser cette approche. On va donc la réaliser à la main!!

1. La statistique de test est ici

$$W = \sqrt{n} \left(\frac{\bar{X}_n - \mu_0}{S_{n-1}} \right)$$

laquelle converge en loi vers une loi gaussienne centrée-réduite. On sera amené dans 9le cas présent à rejeter l'hypothèse nulle su cette statistique est plus grande que la valeur critique ou bien si la p -value est inférieure à la valeur de référence 5%. Pour déterminer ces quantités, écrire le code suivant :

```
# Statistique inférentielle : test d'hypothèses

mu_0 = 100
alpha = 0.05

sample %>%
  summarize(W = sqrt(length(weight))*(mean(weight) - mu_0)/sd(weight)) %>%
  mutate(Critical_Value = qnorm(p = 1-alpha),
         p_value = 1-pnorm(W))
```

Exécuter ce code, puis visualiser la sortie textuelle.

	W	Critical_Value	p_value
1	1.408487	1.644854	0.07949349

Comparer la valeur de la statistique de test avec la valeur critique. Que doit-on décider dans le cas présent ? Justifier votre réponse. Que doit-on décider à partir de la valeur de la p -value pour un niveau de signification de 5% ? Justifier votre réponse. Les deux approches sont-elles cohérentes ?

2. La loi de Student à n degrés de liberté converge en loi vers la loi gaussienne standard lorsque n tend vers l'infini. En conséquence, pour une taille d'échantillon n assez grand (typiquement supérieur ou égal à 30), on peut considérer que les deux lois sont quasiment identiques. Selon ce résultat de probabilité, on peut alors envisager malgré tout de faire usage de la fonction **t.test()** pour mener le test d'hypothèses. Écrire alors un code permettant de réaliser le test et d'obtenir la sortie suivante :

```
One Sample t-test

data: .
t = 1.4085, df = 39, p-value = 0.08345
alternative hypothesis: true mean is greater than 100
95 percent confidence interval:
 98.49393      Inf
sample estimates:
mean of x
 107.675
```

Observer ici d'une part que les valeurs de la statistique de test sont identiques, et que d'autre part les p -values sont presque identiques malgré un écart relatif de moins de 5%.