

# Heritability of BMI

## Set working environment

```
setwd("G:/Stockage/Bio_A3_S2/CoursOption/Projet_Math")
# data = read.csv("500_Person_Gender_Height_Weight_Index.csv", sep = ",", header=TRUE)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
## filter, lag
```

```
## The following objects are masked from 'package:base':
##
## intersect, setdiff, setequal, union
```

## Functions

As a general rule of thumb: If skewness is less than -1 or greater than 1, the distribution is highly skewed. If skewness is between -1 and -0.5 or between 0.5 and 1, the distribution is moderately skewed. If skewness is between -0.5 and 0.5, the distribution is approximately symmetric.

```
# KURTOSIS
kurtosis <- function(x) {
  num <- mean((x-mean(x))^4)
  den <- sd(x)^4
  return((num/den)-3)
}

# SKEWNESS
skewness <- function(x){
  num <- sum((x-mean(x))^3)
  den <- length(x)*sd(x)^3
  return(num/den)
}
```

## Load and clean data

```
data_2 = read.csv("phenofile.csv", sep = " ", header=TRUE)
data_clean = data_2 %>%
  filter(Height != -999)%>%
  filter(Weight != -999)%>%
  relocate(BMI, .after = last_col())

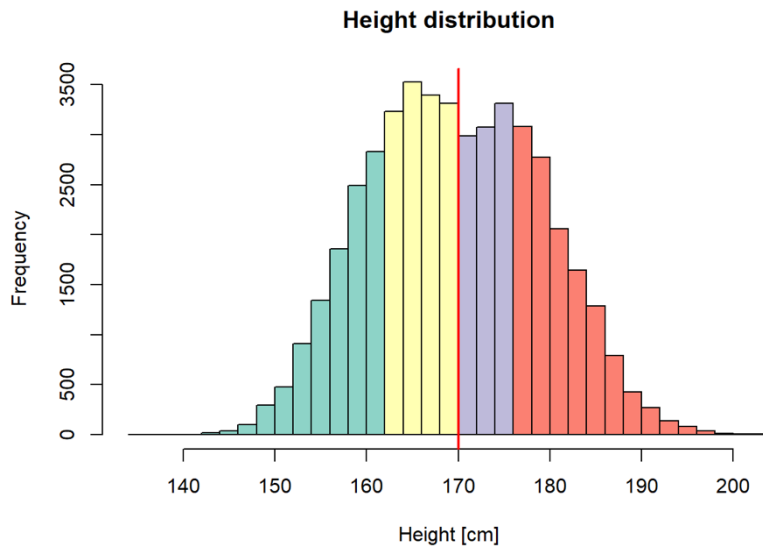
head(data_clean)
```

	Height <dbl>	Weight <dbl>	BMI <dbl>
1	157	63.50766	25.7648
2	168	95.06492	33.6823
3	192	113.82940	30.8782
4	159	47.29291	18.7069
5	168	82.90998	29.3757
6	162	54.29989	20.6904

6 rows

## Check height data #

```
h = hist(data_clean$Height, breaks=30, plot=F)
cols = c('grey', "#8DD3C7", "#FFFFB3", "#BEBADA", "#FB8072")
k = cols[findInterval(h$mids, quantile(data_clean$Height), rightmost.closed=T, all.inside=F) + 1]
plot(h, col=k, main = "Height distribution", xlab = "Height [cm]")
abline(v = median(data_clean$Height), col = "red", lwd = 2)
```



```
#Quartiles
summary(data_clean$Height)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 135.0  163.0  170.0  170.2  177.0  204.0
```

```
#Standard deviation
sd(data_clean$Height)
```

```
## [1] 9.417697
```

```
#Kurtosis
kurtosis(data_clean$Height)
```

```
## [1] -0.5085849
```

```
#Skewness
skewness(data_clean$Height)
```

```
## [1] 0.0737519
```

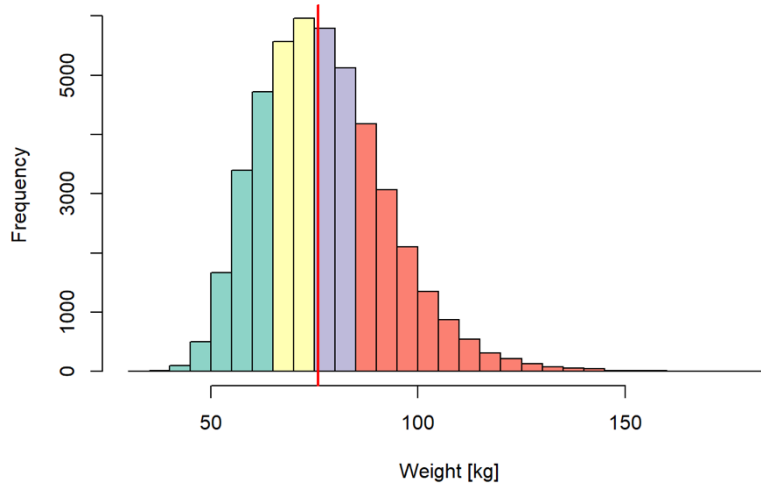
```
#Mode
sort(table(data_clean$Height),decreasing=TRUE)[1]
```

```
## 165
## 2037
```

### Check weight data

```
h = hist(data_clean$Weight, breaks=30, plot=F)
k = cols[findInterval(h$mids, quantile(data_clean$Weight), rightmost.closed=T, all.inside=F) + 1]
plot(h, col=k, main = "Weight distribution", xlab = "Weight [kg]")
abline(v = median(data_clean$Weight), col = "red", lwd = 2)
```

## Weight distribution



```
#Quartiles  
summary(data_clean$Weight)
```

```
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   
##  33.85  65.97   75.86   77.32  86.68  185.00
```

```
#Standard deviation  
sd(data_clean$Weight)
```

```
## [1] 15.65867
```

```
#Kurtosis  
kurtosis(data_clean$Weight)
```

```
## [1] 1.212756
```

```
#Skewness  
skewness(data_clean$Weight)
```

```
## [1] 0.7450064
```

```
#Node  
sort(table(data_clean$Weight),decreasing=TRUE)[1]
```

```
## 71.40002335
##          7
```

### Add ponderal index

```
data = data_clean %>%
  mutate(PI = Weight/((0.01*Height)**3))
```

### Add other indexes

```
data = data %>%
  mutate(BMI = (Weight**2)/((0.01*Height)**3))
head(data)
```

	Height <dbl>	Weight <dbl>	BMI <dbl>	PI <dbl>	Pi <dbl>
1	157	63.50766	25.7648	16.41070	1042.2051
2	168	95.06492	33.6823	20.04899	1905.9555
3	192	113.82940	30.8782	16.08240	1830.6494
4	159	47.29291	18.7069	11.76535	556.4175
5	168	82.90998	29.3757	17.48554	1449.7253
6	162	54.29989	20.6904	12.77185	693.5101

6 rows