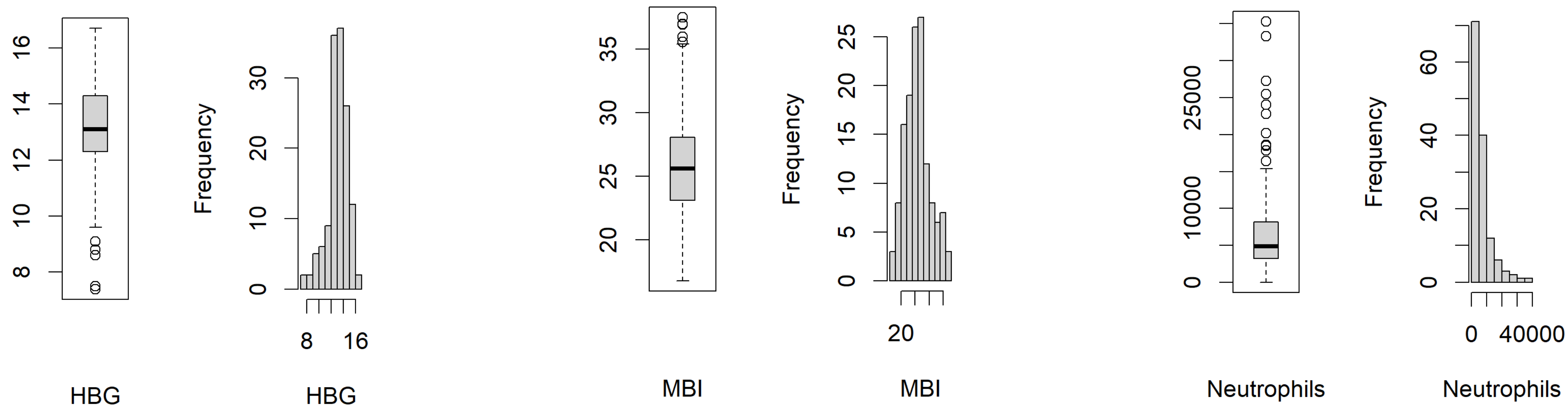
Distribution shape

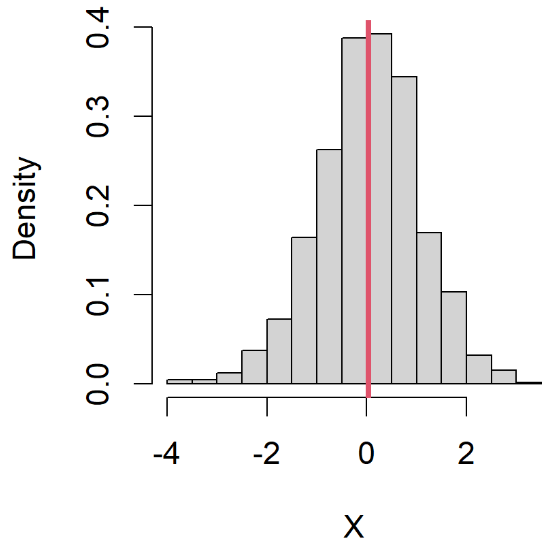
Shape of the frequency distribution

The shape of the frequency distribution can appear different for several aspects over and above location and variability



Symmetric versus asymetric distribution

Geometrically, a frequency distribution is symmetric if the right half of the distribution is the mirror image of the left half.



This implies that, for symmetric distribution we get

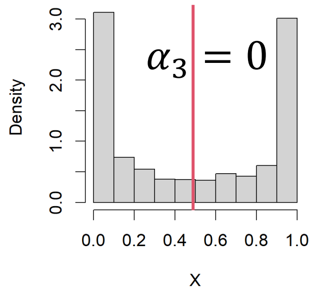
$$\bar{X} = Me$$

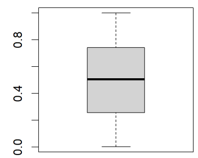
Hence an index for skewness is $\alpha_3 = \frac{(\bar{X}-Me)^3}{\tilde{S}^3}$

Symmetric versus asymmetric distribution

$\alpha_3 = 0$ ➔ symmetric distribution

in this case $Me - Q_1 \approx Q_3 - Me$

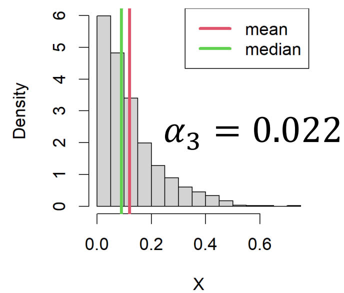


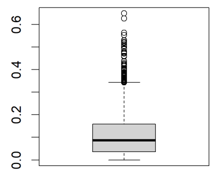


$\alpha_3 > 0$ ➔ *positive* skewed distribution

➔ $\bar{X} > Me$

In this case $Me - Q_1 < Q_3 - Me$

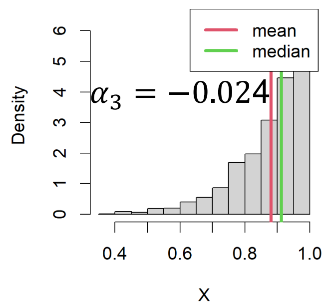


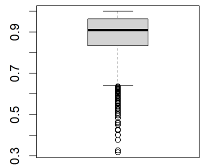


$\alpha_3 < 0$ ➔ *negative* skewed distribution

➔ $\bar{X} < Me$

In this case $Me - Q_1 > Q_3 - Me$

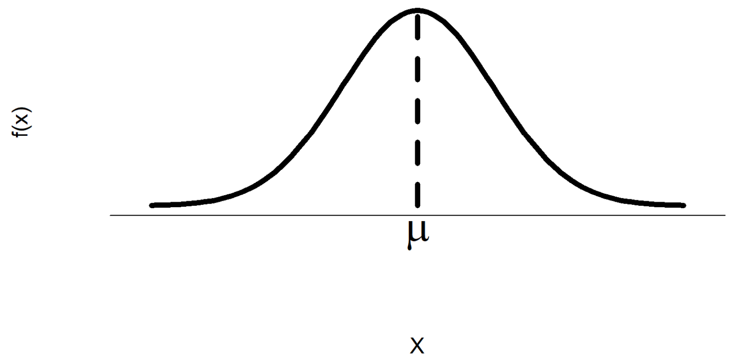




Kurtosis

Kurtosis measures how much the distribution peaks around its mode.

It is defined as the degree of departure of the observed frequency distribution from the normal or *Gaussian distribution*, which is assumed as a benchmark.



The *normal distribution* is a mathematical model often adopted to describe many phenomena in real life.

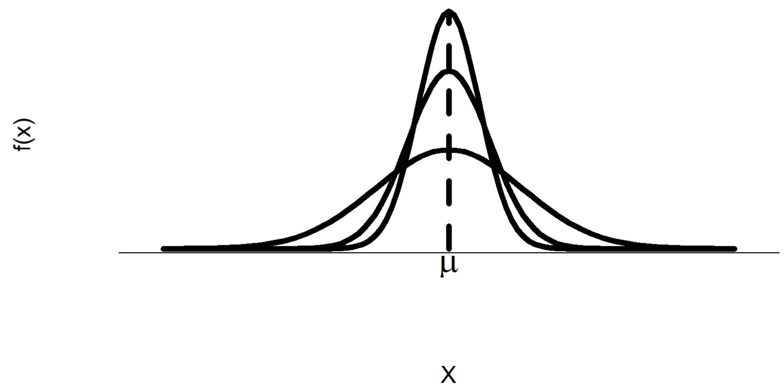
The Normal distribution is defined by the following equation

$$f(x;\ \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} \quad x \in \mathbb{R}, \mu \in \mathbb{R}, \sigma > 0$$

Kurtosis cont'd

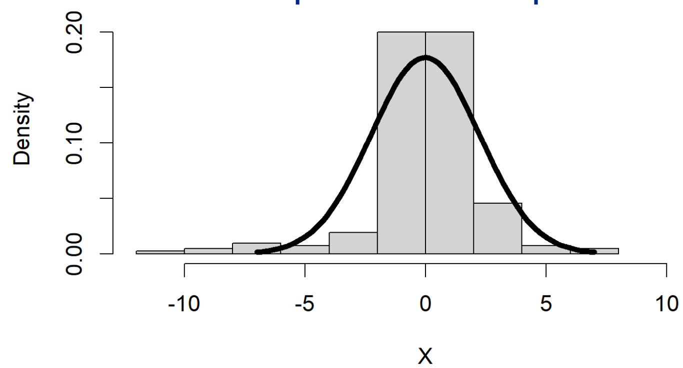
μ is the maximum of the function.

σ determines the narrowness of the curve as σ decreases, the curve becomes more peaked around μ.



The observed distribution is called *leptokurtic* if, compared to the normal, it has fatter tails, whereas it is called *platykurtic* if it has thinner tails.

Leptokurtic shape

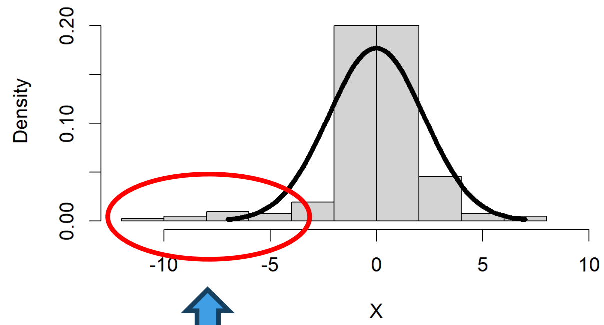


Kurtosis cont'd

The kurtosys index is defined by

$$\gamma = \frac{1}{\tilde{S}^4}\left[\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{X})^4 - 3\right]$$

γ =9.4



$\gamma \approx 0$ ➔ the distribution is close to the normal shape

$\gamma > 0$ ➔ the distribution is leptokurtic (heavy tails)

$\gamma < 0$ ➔ the distribution is platykurtic (thin tails)

Summarising…description of a variable: BMI

Min.	0
Q1	23.10
Median	25.58
Mean	25.78
Q3	28.07
Max.	37.50
sd	4.97
asimmetry	0
kurt	7.26
NA	2