Measures of Location: Third Moment

[John S Butler (TU Dublin)](https://johnsbutler.netlify.com)

# Introduction

A different aspects of a distribution of data can be summarised by the measures of location:

1. The First Moment: Middle.
2. The Second Moment: Spread.
3. **The Third Moment: Symmetry.**

All that being said, I would always recommend plotting the data first before anything else.

**A picture (histogram) is worth a thousand words.**

# Third Moment: Symmetry

## Skewness

### Definition 1:

Skewness is a measure of symmetry (or not symmetry) of a distribution. Pearson’s Coefficient of Skewness number 1 uses the mode to calculate skewness, given by the formula is:

### Definition 2:

Pearson’s Coefficient of Skewness number 2 uses the median to calculate skewness, given by the formula is:

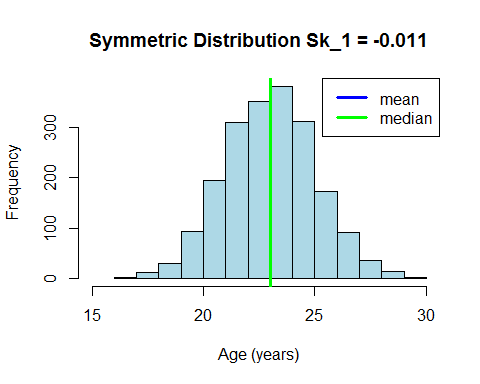
, where is the average of the elements, is the standard deviation, is the mode of the elements and is the median of the elements.

# Interpretation of Skewness

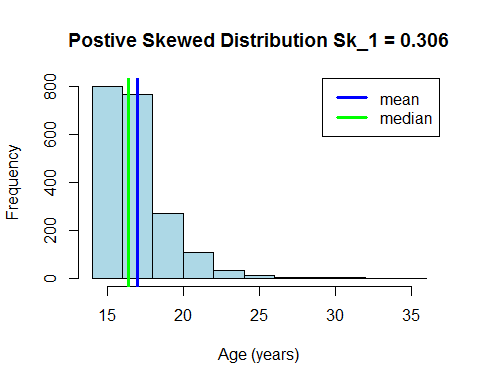
How to interpret Skewness:

* A skewness near means the distribution is symmetric.
* A Negative skewness means the distribution is right skewed
* A Positive skewness means the distribution is left skewed

# Symmetric Example

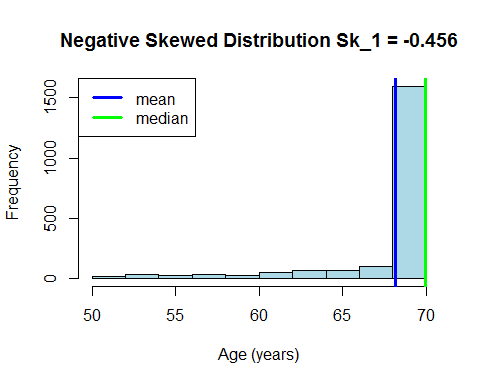
The figure below shows a symmetric histogram of 2000 concert attendees ages observations at a concert:  The age distribution is equally distributed around the mean, hence the skewness is more or less 0.

# Positive Skewness Example

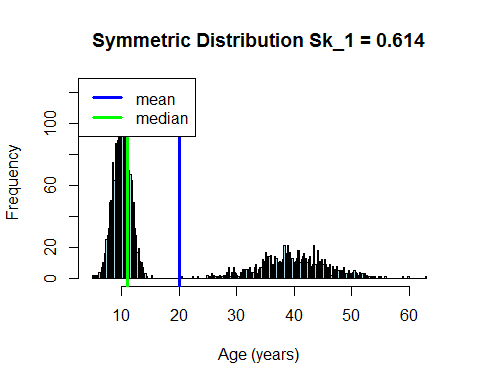
The figure below shows a positive skewness histogram of 2000 attendees age observations at a concert:  The age distribution shows that the concert attendees are mostly young, hence the skewness is positive.

## Negative Skewed Example

The figure below shows a negative skewness histogram of 2000 attendees age observations at a Andre Rieu concert:

 The age distribution is skewed to the right side as the concert has mostly older adults, hence the skewness is negative.

## Bi-modal distribution

The figure below shows a histogram of 3000 attendees age observations at a Wiggles concert:  The distribution is bimodal as there are 2000 excited children with 1000 parents wear earplugs, the skewness is positive but this is misleading.

### All in one plot

library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.6.1

library(ggpubr)

## Warning: package 'ggpubr' was built under R version 3.6.1

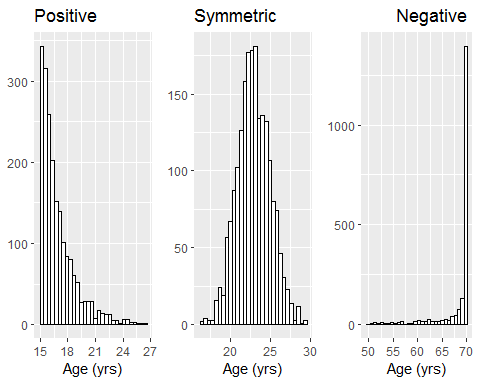
## Loading required package: magrittr

## Warning: package 'magrittr' was built under R version 3.6.1

Sym<-rnorm(2000,23,2)  
Neg<-rbeta(2000,1,0.1)\*20+50  
Pos<-rchisq(2000,2)+15  
df <- data.frame(Neg,Sym,Pos)  
p.neg<-ggplot(df, aes(x=Neg)) + geom\_histogram(color="black", fill="white")+ggtitle("Negative ")+xlab("Age (yrs)")+ylab(NULL)+ theme(plot.title = element\_text(hjust=1.0))  
p.sym<-ggplot(df, aes(x=Sym)) + geom\_histogram(color="black", fill="white")+ggtitle("Symmetric")+xlab("Age (yrs)")+ylab(NULL)  
p.pos<-ggplot(df, aes(x=Pos)) + geom\_histogram(color="black", fill="white")+ggtitle("Positive")+xlab("Age (yrs)")+ylab(NULL)  
ggarrange( p.pos,p.sym, p.neg, ncol = 3, nrow = 1)

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



ggsave("skew.png",dpi=300, width = 4, height = 2.5)