OLLSCOIL TEICNEOLAÍOCHTA
BHAILE ÁTHA CLIATH
DUBLIN
TECHNOLOGICAL
UNIVERSITY DUBLIN

Measures of Location

Dr. John S. Butler

Measures of Locations

First moment

1. Mean (average)
2. Median
3. Mode

Second Moment (Spread)

1. Standard deviation
2. Variance

Third Moment

1. Skewness

First Moment of Data (Middle)

Mean (average)

Definition

- The point around which the sum of the deviations is o

https://youtu.be/ukqunhWvDQk

- $\sum_{i=1}^{n}(x_i - \bar{x}) = 0$

- $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

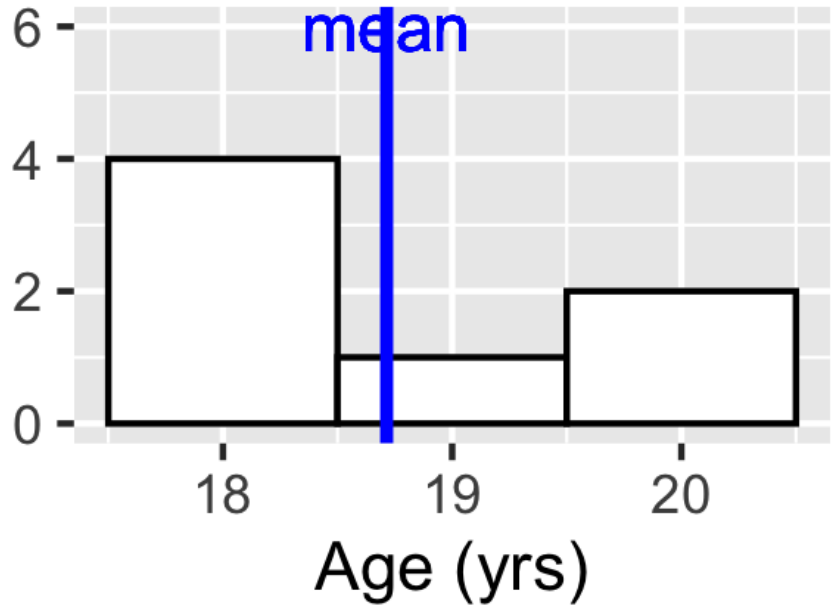
Average (Mean): Pros and Cons

Pros

- Easy to calculate and well known and understood
- Uses all the data

Cons

- Is sensitive to extreme values
- For example
- 18, 18, 18, 18, 19, 19, 20, 20, 58
- Mean=23.111



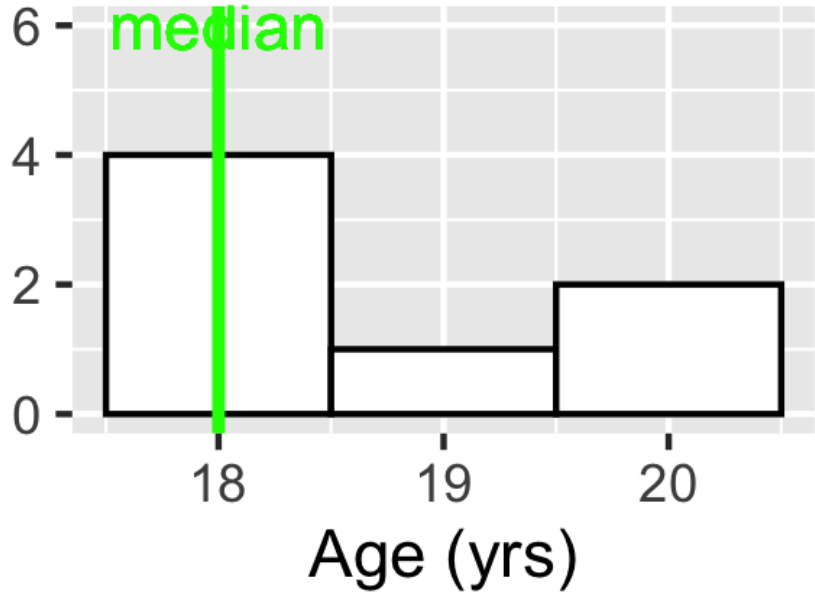
Median

- Definition
- The median is the middle values of the ordered set of values
- 18, 18, 18, 18, **19**, 19, 20, 20, 58
- 9 values the median value is the fifth value

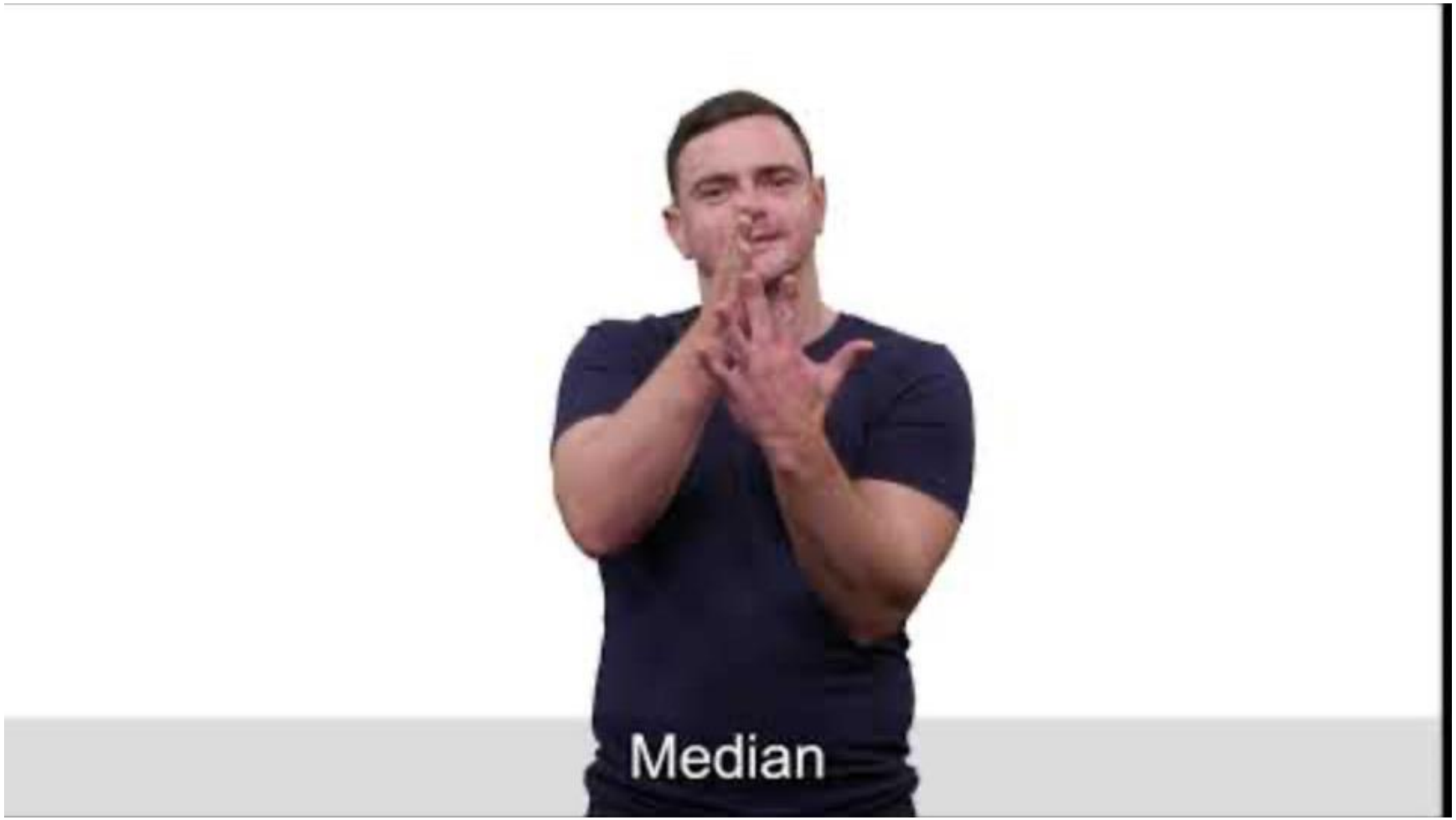
- If the is an even number of the median is usually the average of the two middle values
- 18, 18, 18, **18, 19**, 19, 20, 20
- 18.5

Median: Pros and Cons

- Pros
- Median five the centre of the data - very intuitive
- Not as sensitive to extreme values
- Cons
- Does not use all the data



Median: Irish Sign Language



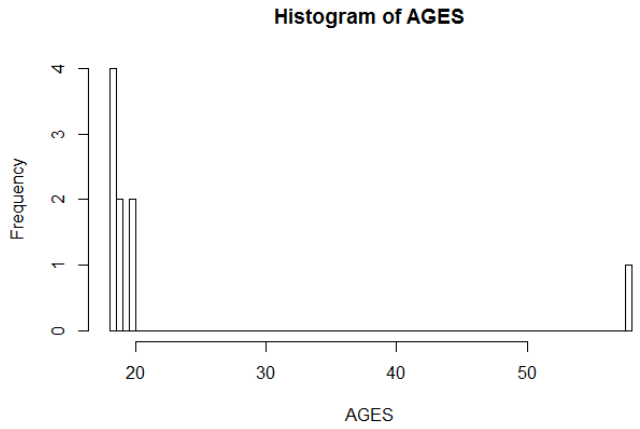
Irish Sign Language STEM Glossary https://www.dcu.ie/islstem/index.shtml

OLLSCOIL TEICNEOLAÍOCHTA BHAILE ÁTHA CLIATH
TECHNOLOGICAL UNIVERSITY DUBLIN

Mode

- Definition
- The most commonly occurring value in the distribution
- **18, 18, 18, 18**, 19, 19, 20, 20, 58

- The mode is 18



Mode: Pros and Cons

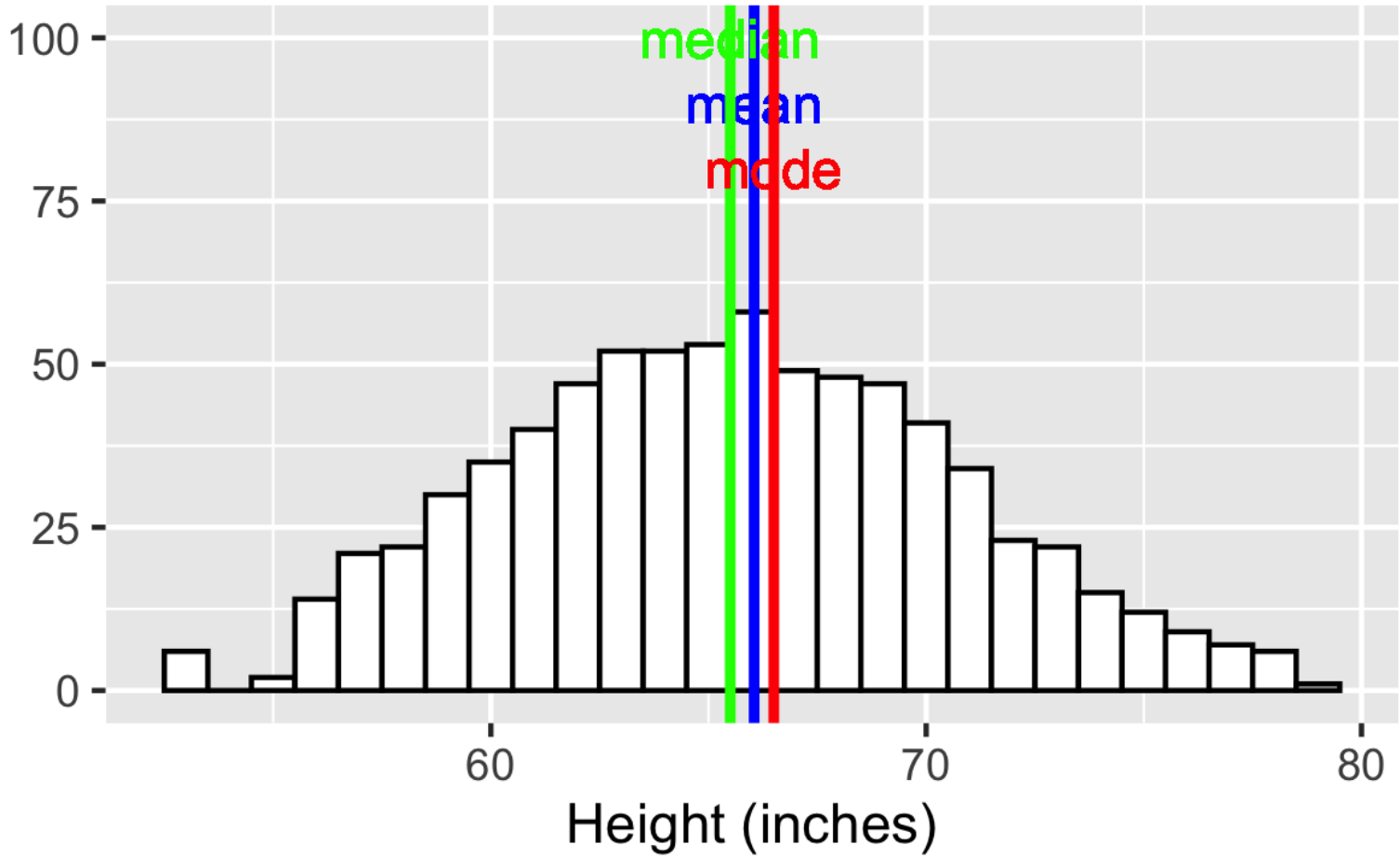
Pros

- Shows where the data is concentrated

Cons

- More than one Mode in a dataset
- Does not use the data

All in One Plot



Second Moment

The spread of the data

Variance

- Is the spread of the data around the mean
- $\mathrm{Var}(x) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$

Standard deviation

- *Measure of the average amount by which observations deviate from the mean. The square root of the variance.*

- $s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$

Standard deviation: Pros and Cons

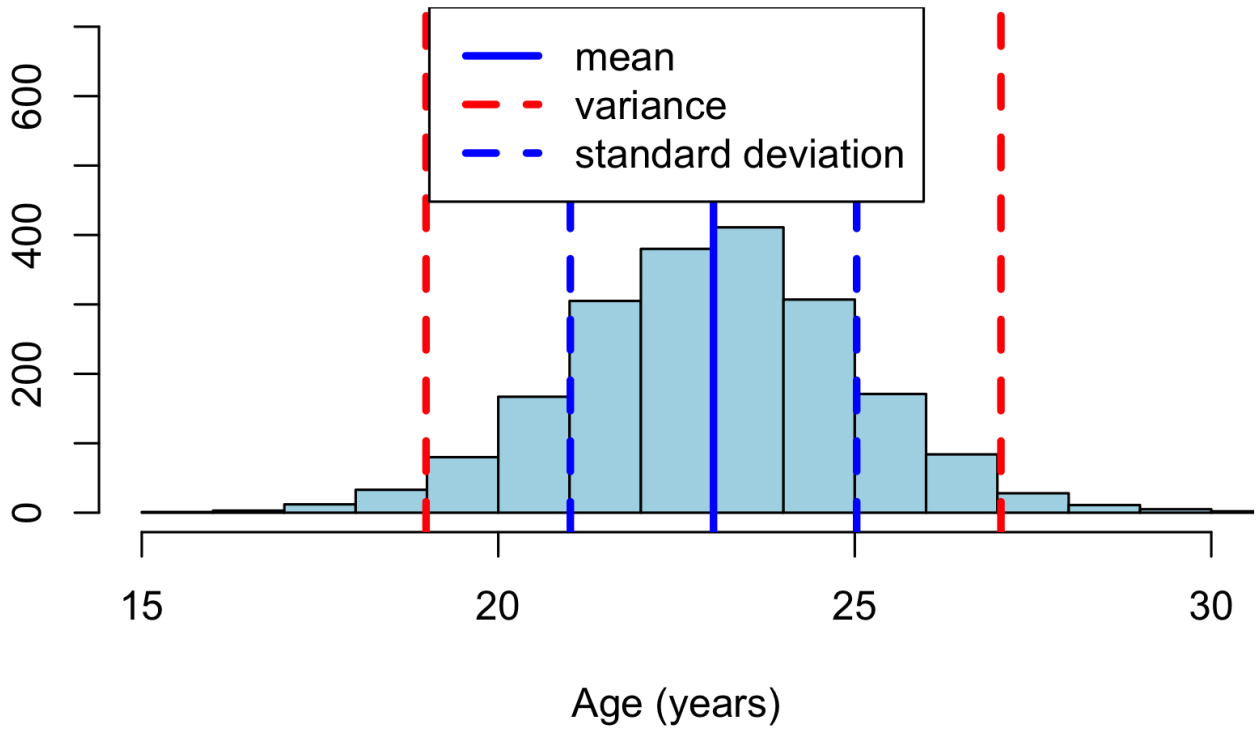
Pros

- Takes all data into account.
- Lends itself to computation of other stable measures (and is a prerequisite for many of them).

Cons

- Hard to interpret.
- Can be influenced by extreme scores.

Standard deviation - Plot



Standard Deviation - Irish Sign Language

Irish Sign Language STEM Glossary
https://www.dcu.ie/islstem/index.shtml

OLLSCOIL TEICNEOLAÍOCHTA BHAILE ÁTHA CLIATH
TECHNOLOGICAL UNIVERSITY DUBLIN

Range

- Difference between the smallest and largest observations.

- Find the max and min of a range
- [18, 58]

Range: Pros and Cons

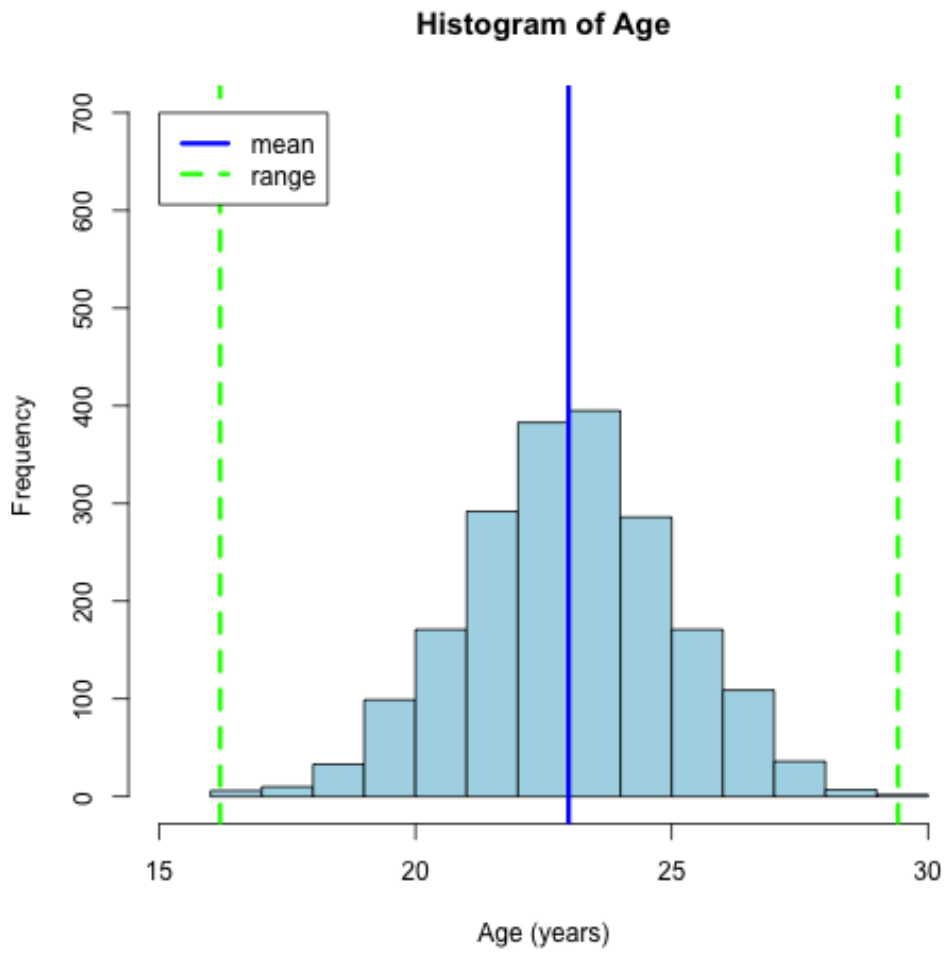
Pros

- Very easy to compute.
- Scores exist in the data set.

Cons

- Value depends only on two scores.
- Very sensitive to outliers.
- Influenced by sample size (the larger the sample, the larger the range).

Range Plot



Interquartile Range

- The inter quartile range is Q3-Q1
- 50% of the observations in the distribution are in the inter quartile range.
- The following figure shows the interaction between the quartiles, the median and the inter quartile range.

Interquartile Range

Quartiles:

$$Q_1 = \frac{n+1}{4}\ \text{th}$$

$$Q_3 = \frac{3(n+1)}{4}\ \text{th}$$

Inter quartile :

$$IQR = Q_3 - Q_1$$

Interquartile: Range Pros and Cons

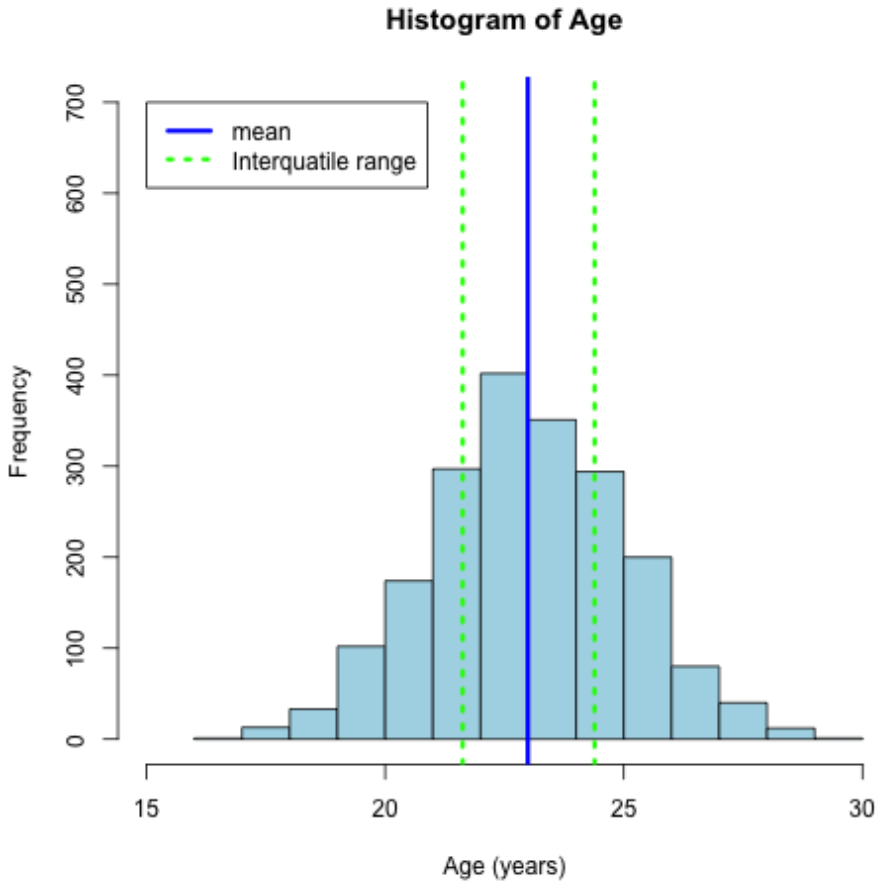
Pros

- Fairly easy to compute.
- Scores exist in the data set.
- Eliminates influence of extreme scores.

Cons

- Discards much of the data.

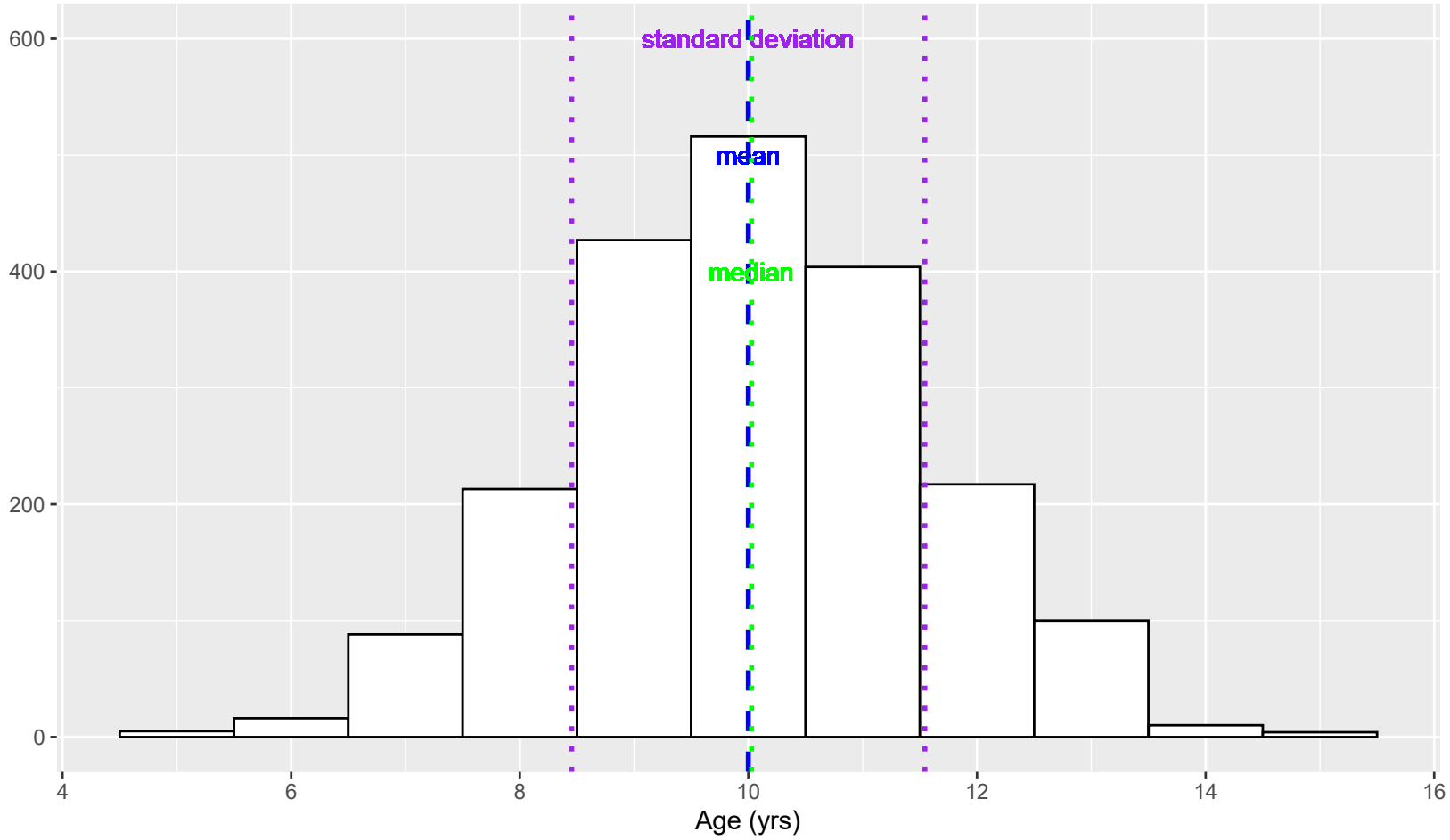
Interquartile - Plot



Mean and Standard Deviation

- Using the mean and standard deviation together:
  - Is an efficient way to describe a distribution with just two numbers.
  - Allows a direct comparison between distributions that are on different scales.

Mean and Standard Deviation



Coefficient of Variation

- Uses both the mean and standard to describe the distribution
- $CV = \frac{standard\ deviation}{mean} = \frac{s}{\bar{x}} = \frac{\sigma}{\mu}$

Pros

- It is unitless and therefore can be used to compare across different variables

Cons

- Loses some meaning

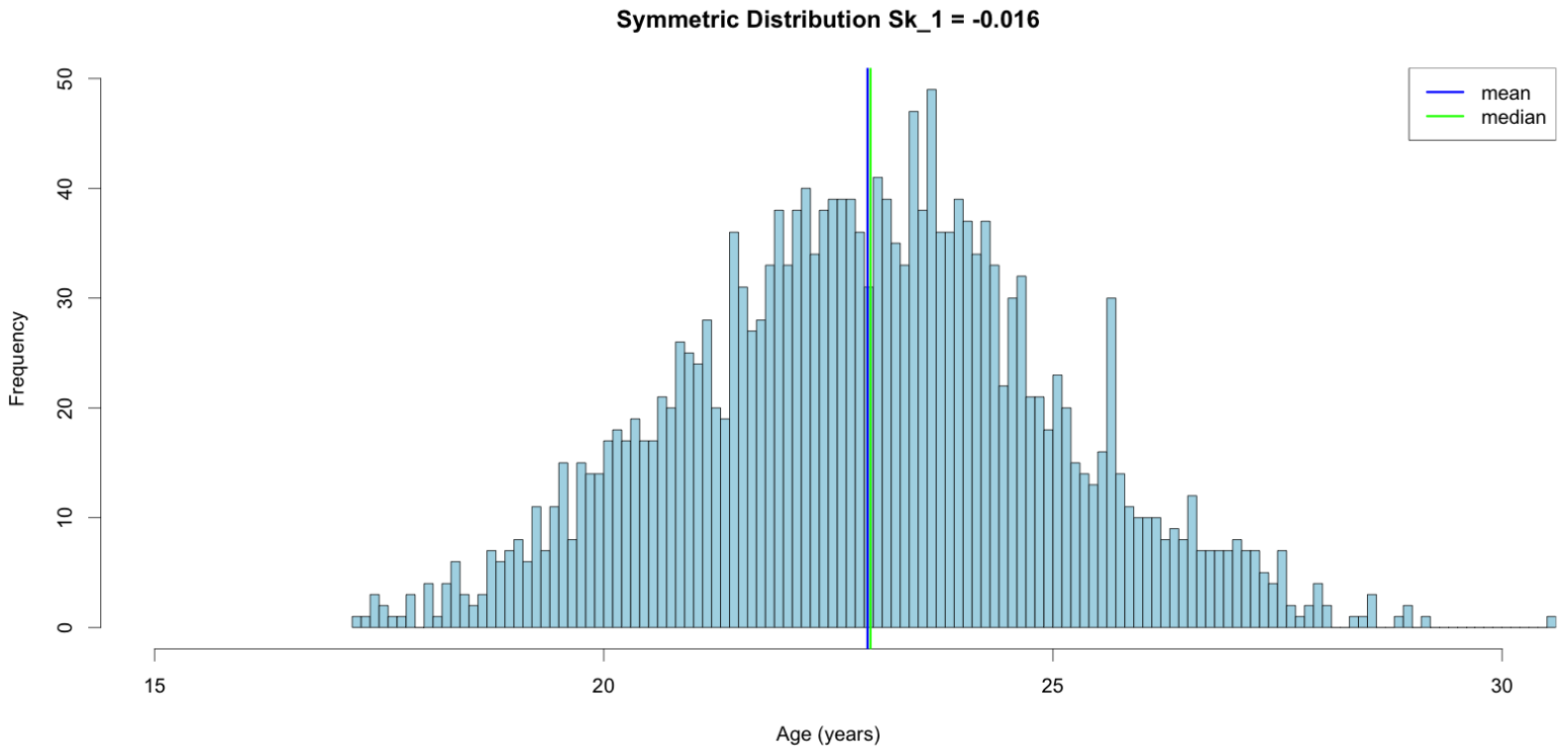
Third Moment of Data (Skewness)

Pearson's Skewness Coefficient

It is a measure of symmetry (or not symmetry) of a distribution

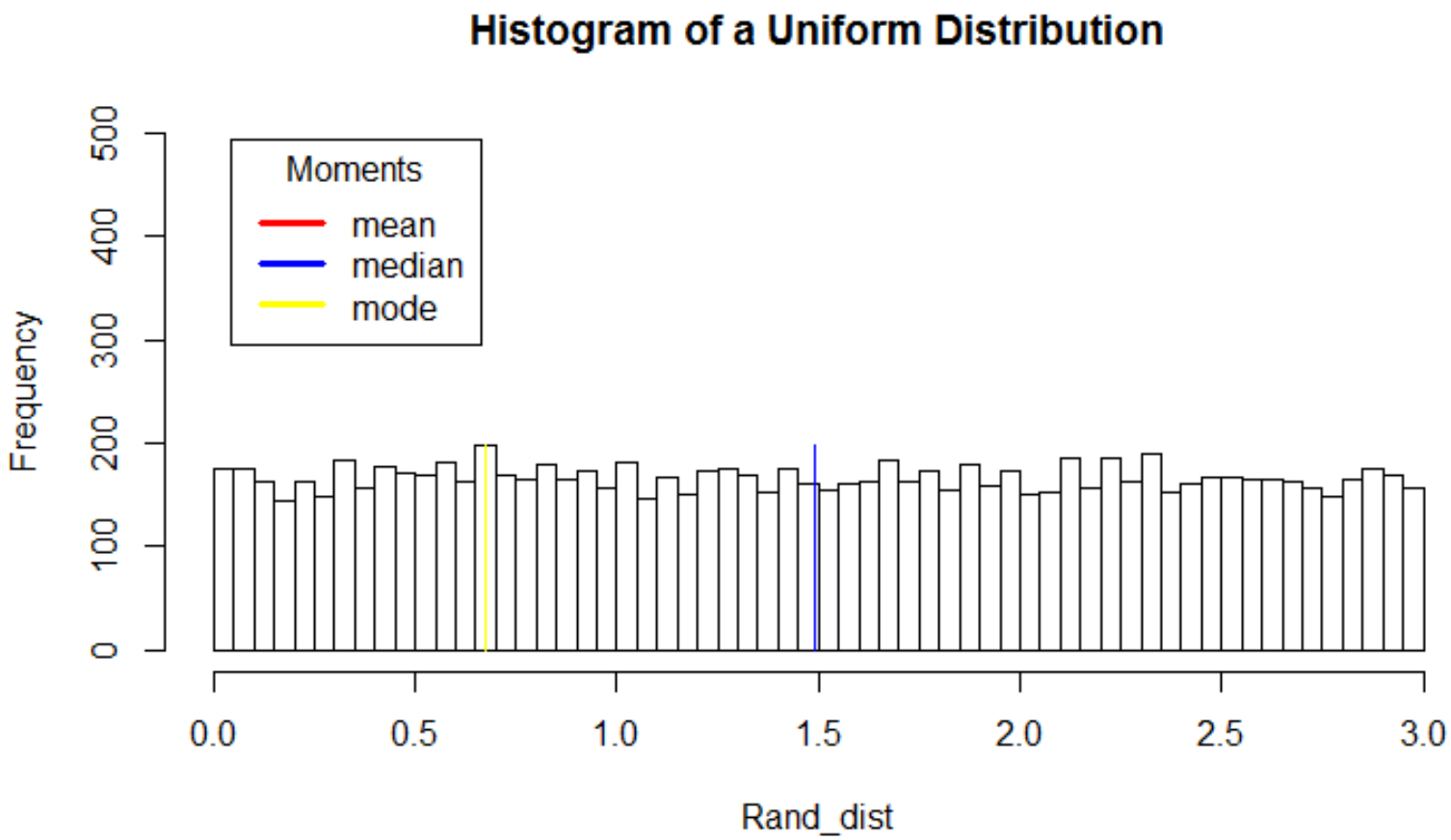
- $\frac{mean-mode}{standard\ deviation} = \frac{\mu-mode}{\sigma}$
- or
- $3\frac{mean-median}{standard\ deviation} = 3\frac{\mu-median}{\sigma}$
- o means no skewedness
- Negative numbers mean right skewed
- Positive numbers mean left skewed

No Skew



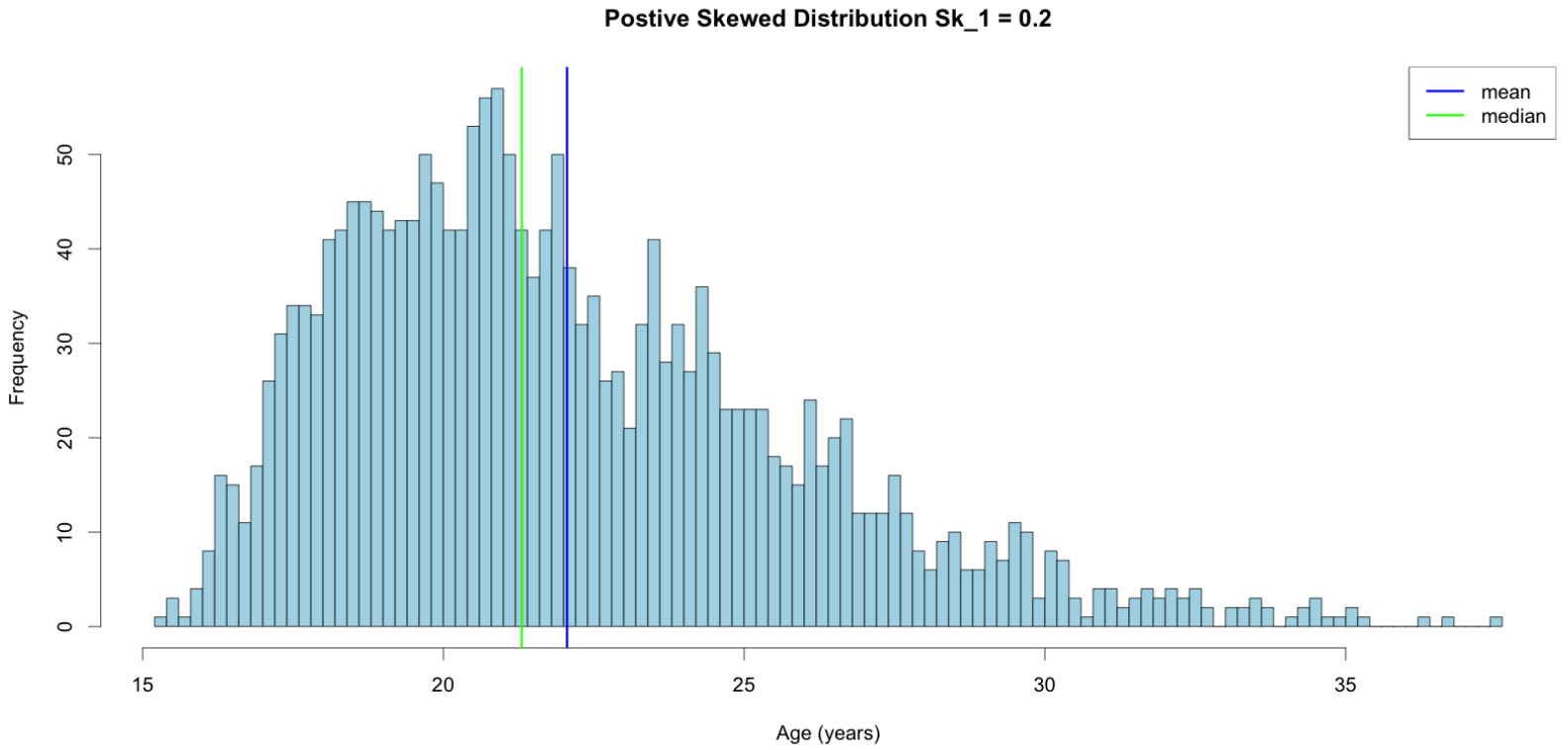
Guassian mean=0
s.d.=3

No Skew



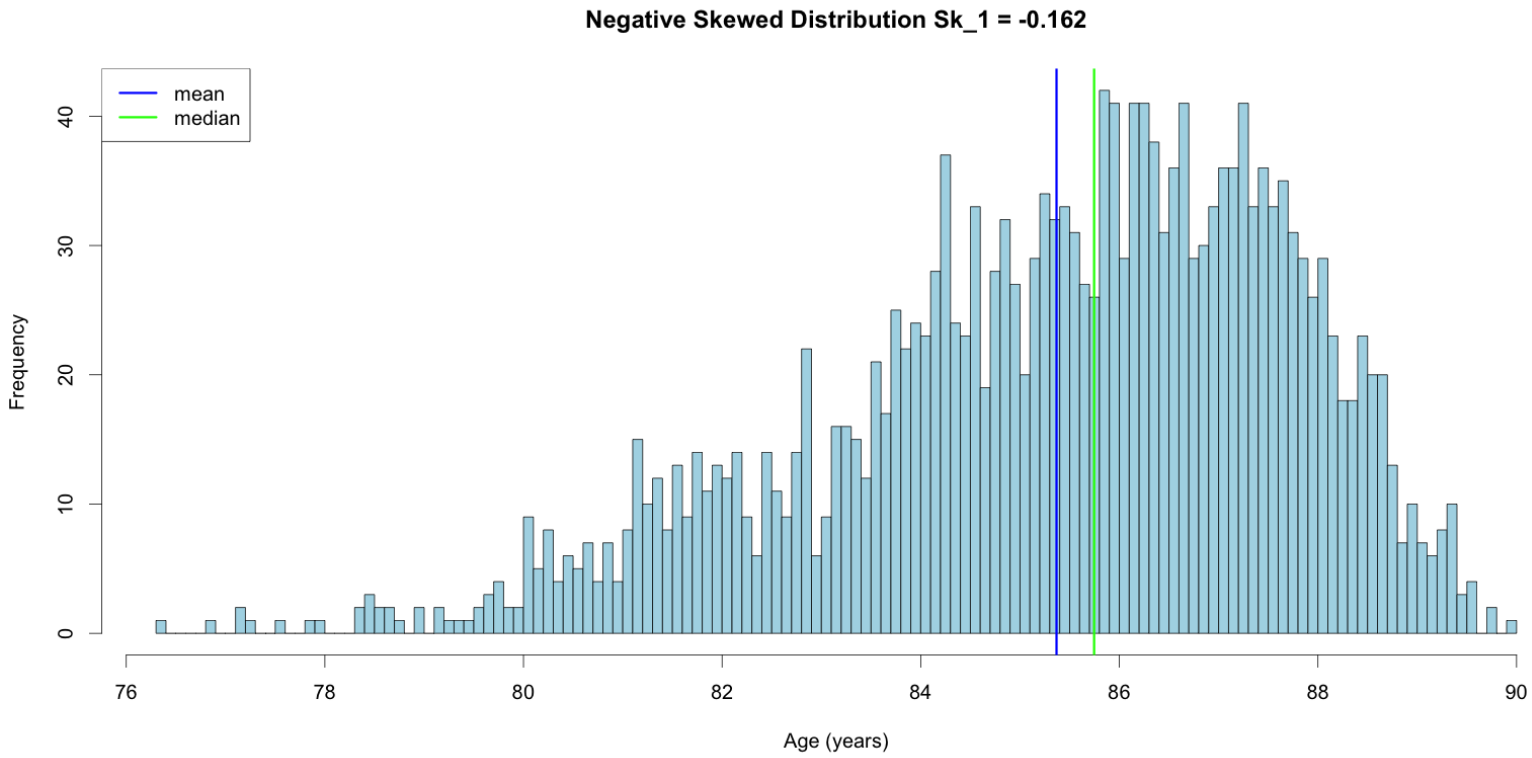
Uniform Distribution

Positive Skew (Beiber Concert)



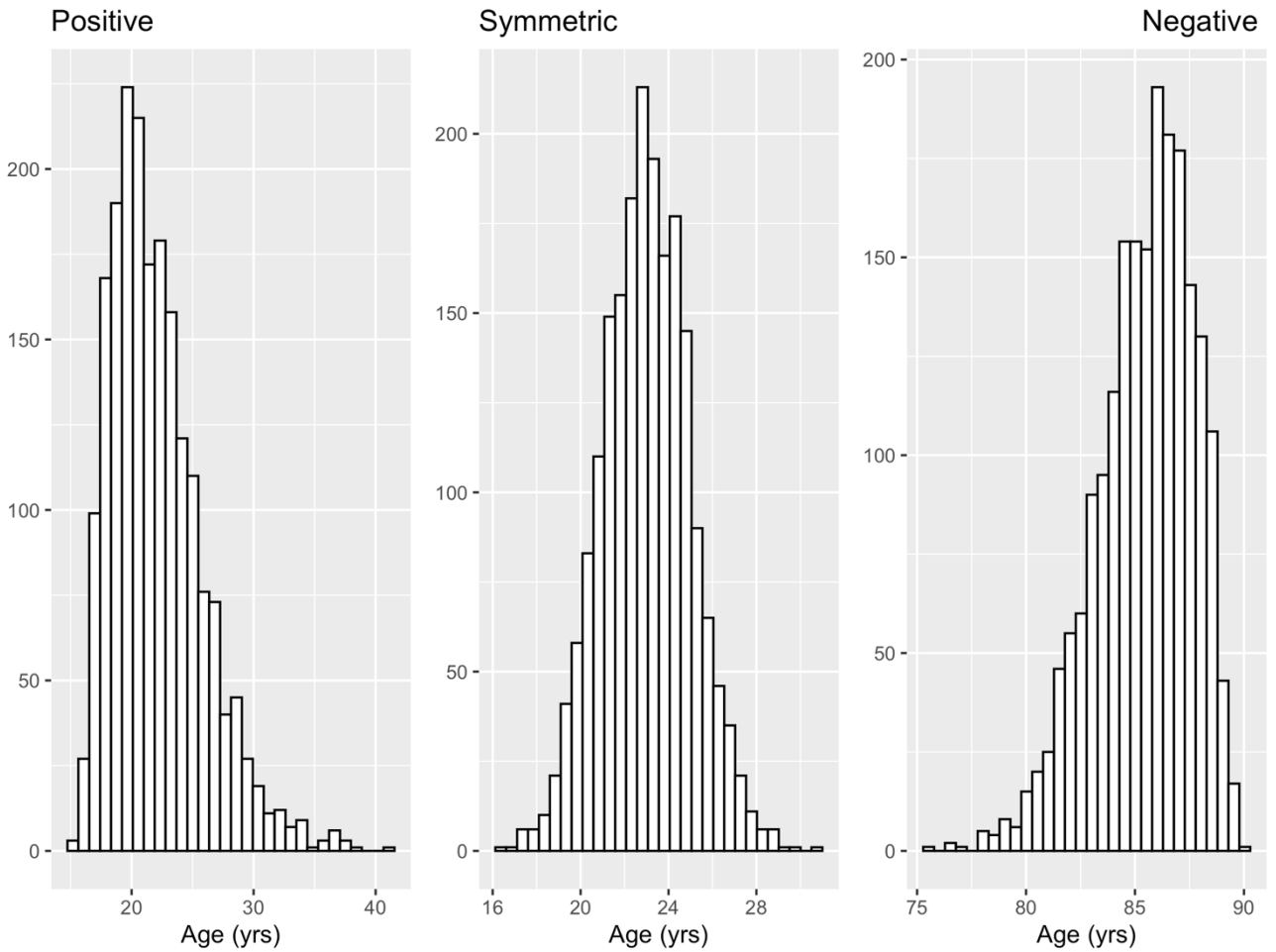
Chi-square DF=7

Negative Skew (Andre Rieu)



Beta, Alpha=10, beta=3

All in One



Problems with Moments of location