

Measures of Variability or Dispersion

- **Measure of dispersion**

It assesses the magnitude of departure from the average values (central tendency).

- **Range**

✓ Difference between maximum and minimum values

- **Interquartile Range**

✓ Difference between third and first quartile ($Q_3 - Q_1$)

- **Variance**

✓ Average* of the squared deviations from the mean

- **Standard Deviation**

✓ Square root of the variance. (positive value is taken).

* Definitions of population variance and sample variance differ slightly.



Variance and Standard Deviation

Population Variance

$$\begin{aligned}\sigma^2 &= \frac{\sum_{i=1}^N (x - \mu)^2}{N} \\ &= \frac{\sum_{i=1}^N x^2 - \frac{\left(\sum_{i=1}^N x\right)^2}{N}}{N} \\ \sigma &= \sqrt{\sigma^2}\end{aligned}$$

Sample Variance

$$\begin{aligned}s^2 &= \frac{\sum_{i=1}^n (x - \bar{x})^2}{(n-1)} \\ &= \frac{\sum_{i=1}^n x^2 - \frac{\left(\sum_{i=1}^n x\right)^2}{n}}{(n-1)} \\ s &= \sqrt{s^2}\end{aligned}$$

Standard deviation for grouped data

$$• S = \sqrt{\frac{\sum f(x - \bar{x})^2}{n-1}}$$

$$\text{Where } \bar{X} = \frac{\sum fx}{n}$$

n is the sample size = $\sum f$, X = midpoint of the class.



- Coefficient of variation (CV)
- Also called Relative dispersion.

$$cv = \frac{\overline{S}}{X} \quad \text{for sample data}$$

$$cv = \frac{\overline{\sigma}}{\mu} \quad \text{for population data}$$

cv is the measure to use when you want to see the relative spread across groups. It also measures the extent of spread in a distribution as a percentage to the mean.



Further Discussions

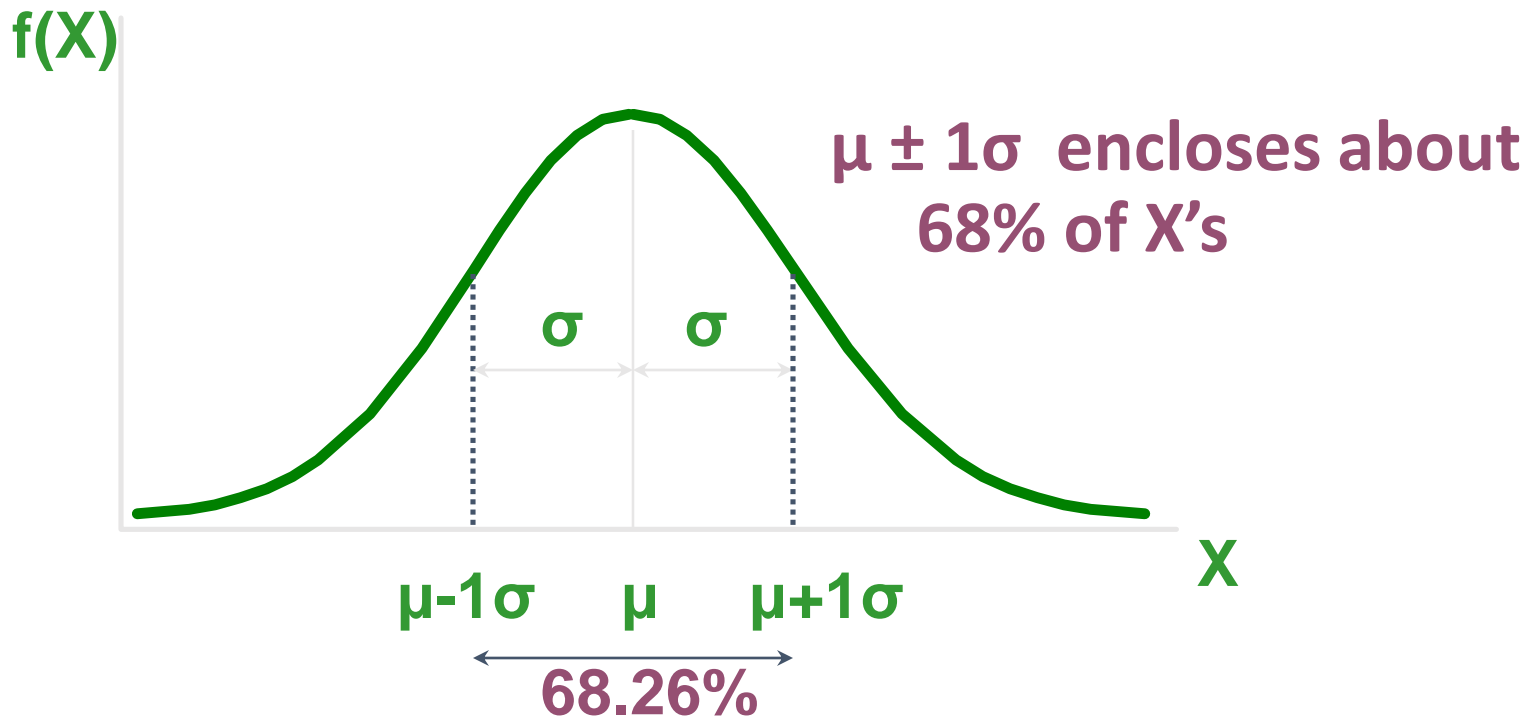
- Range (affected by extreme values)
- MAD (ignores the sign of the deviation)
- Standard deviation (an elegant measures).
- Coefficient of variation (it indicates the % risk of deviation from mean).

Things to Remember

1. The more spread out, or dispersed, the data are, the larger will be the range, the interquartile range, the variance, the standard deviation.
2. The more homogeneous the data are, the smaller will be the range, the interquartile range, the variance, the standard deviation.
3. If the observations are all the same (so that there is no variation in the data), the range, the interquartile range, the variance, the standard deviation will all be zero.
4. None of the measures of variation (the range, the interquartile range, the variance, the standard deviation) can ever be negative.

Empirical Rules

What can we say about the distribution of values around the mean? There are some general rules:

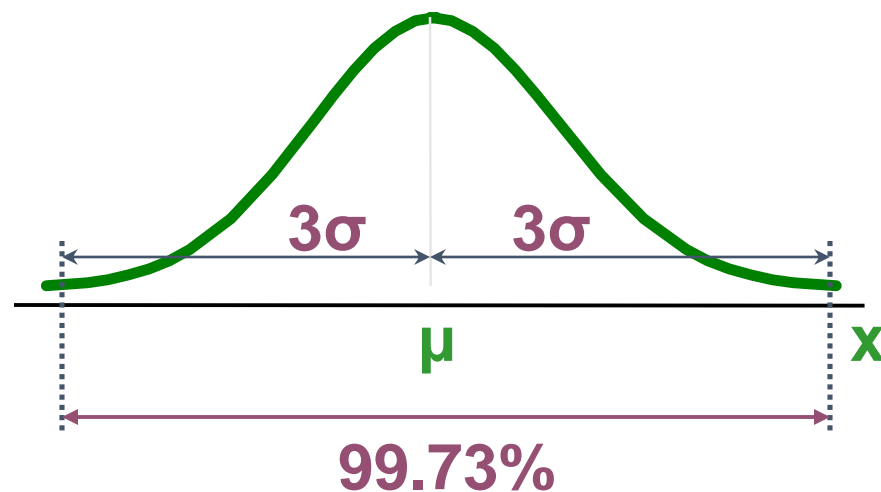
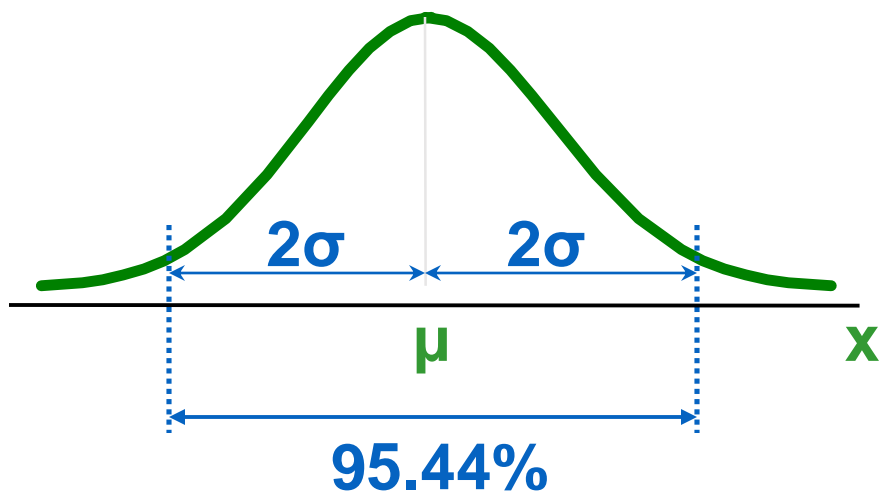




The Empirical Rule

(continued)

- $\mu \pm 2\sigma$ covers about **95%** of X 's
- $\mu \pm 3\sigma$ covers about **99.7%** of X 's





Skewness and Kurtosis

- **Skewness**

- Measure of asymmetry of a frequency distribution
 - Skewed to left
 - Symmetric or unskewed
 - Skewed to right

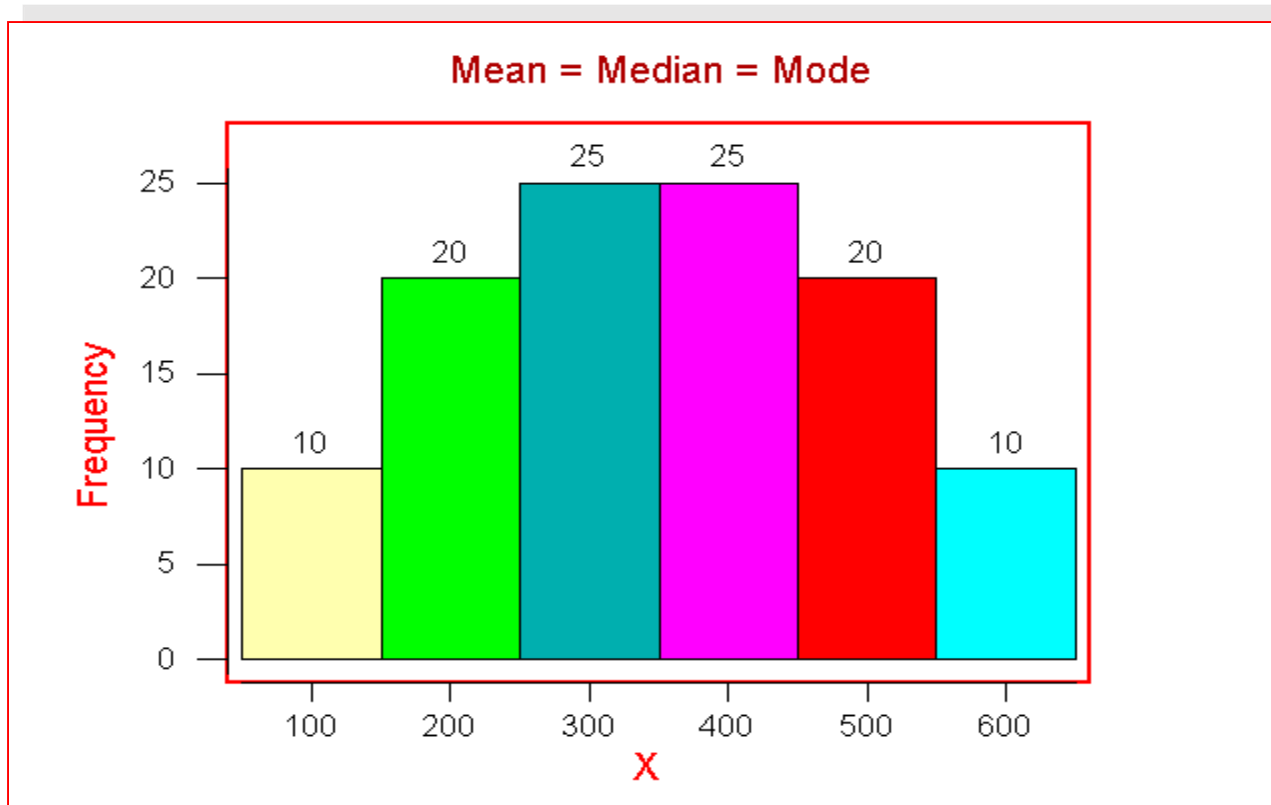
- **Kurtosis**

- Measure of flatness or peakedness of a frequency distribution
 - **Platykurtic** (relatively flat)
 - **Mesokurtic** (normal)
 - **Leptokurtic** (relatively peaked)



Skewness

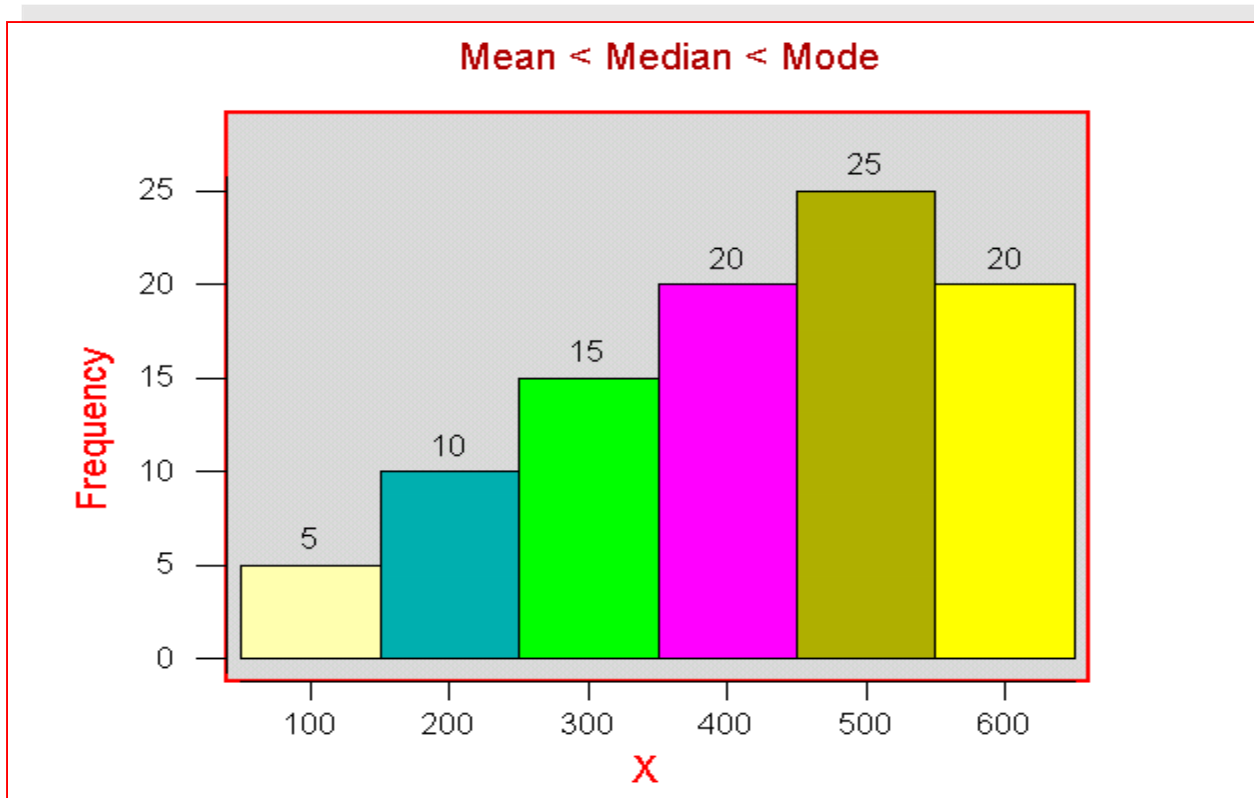
Symmetric





Skewness

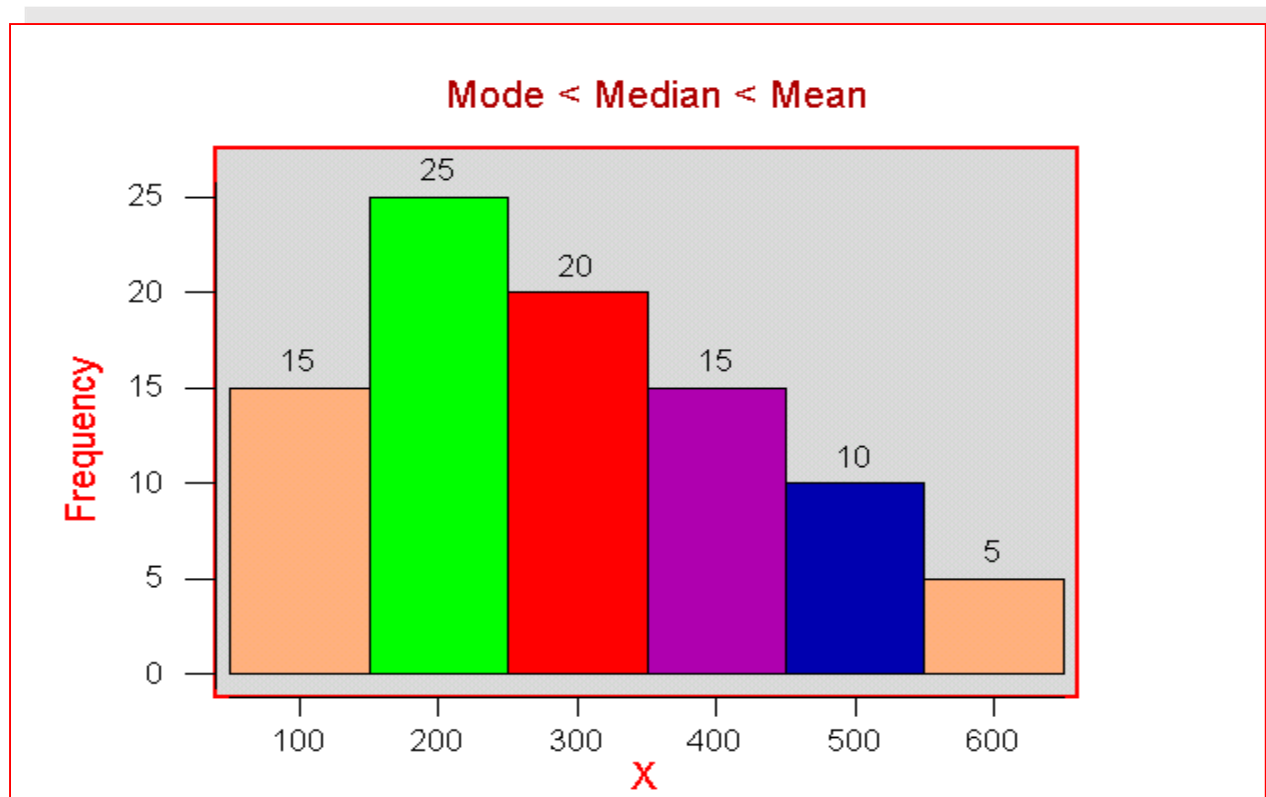
Skewed to left





Skewness

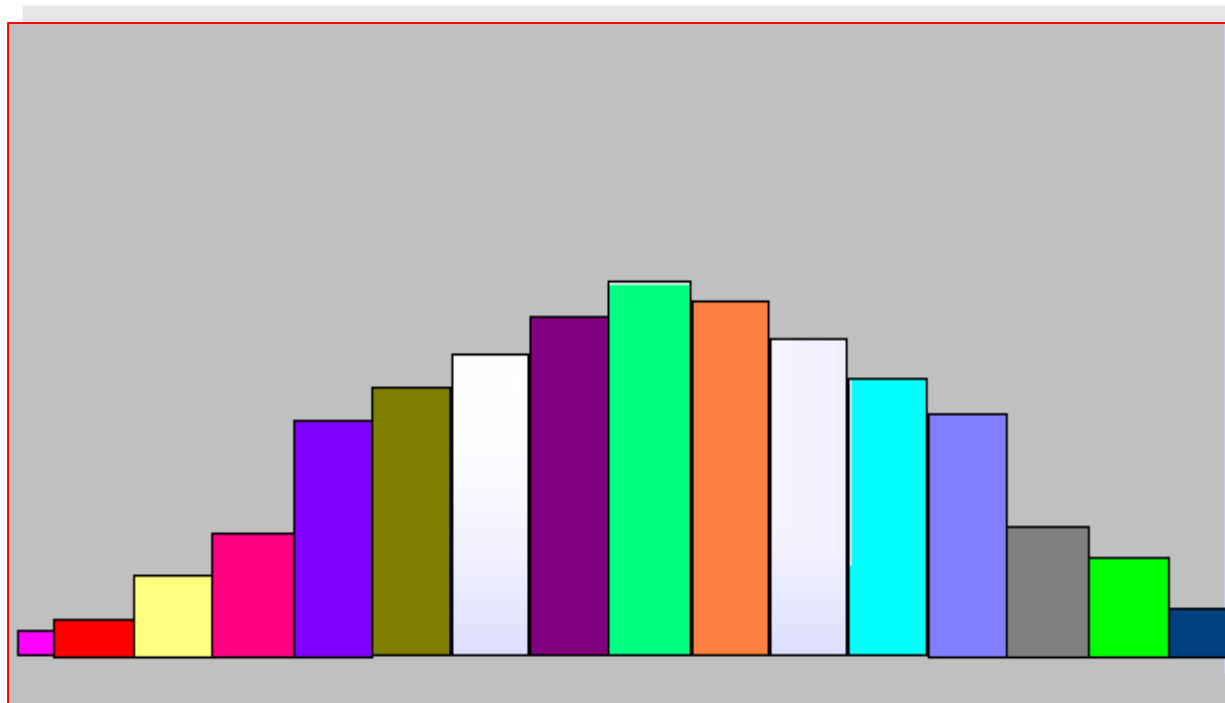
Skewed to right





Kurtosis

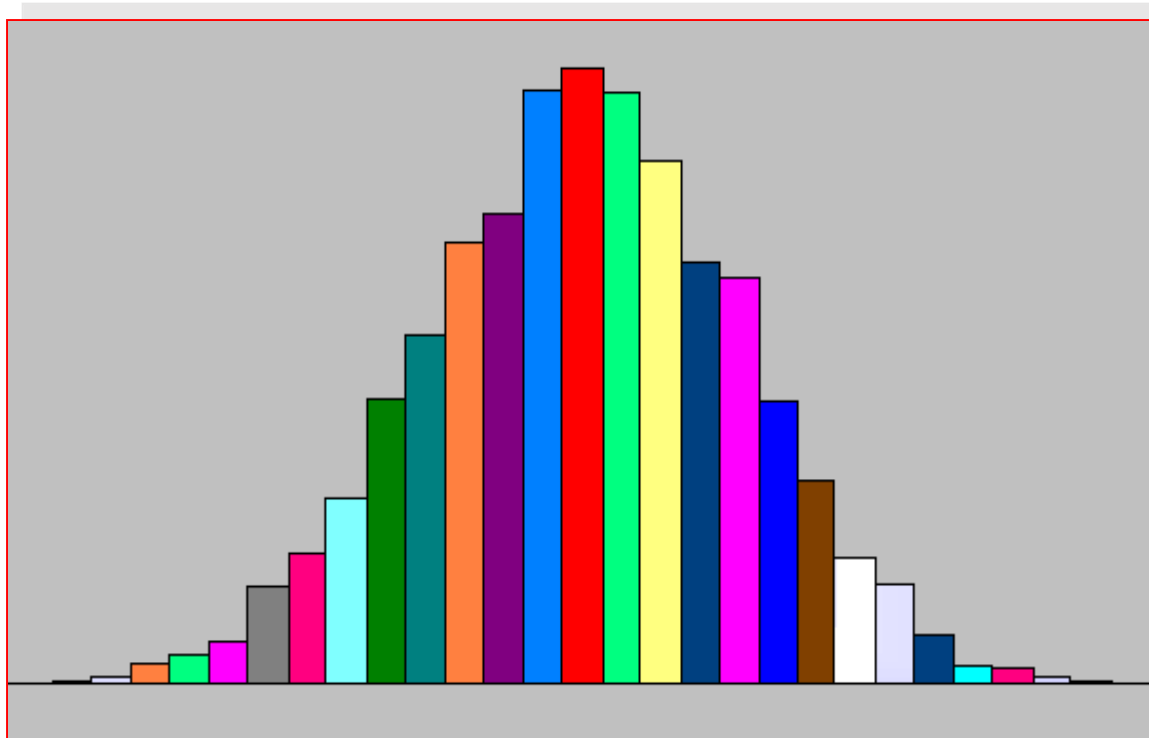
Platykurtic - flat distribution





Kurtosis

Mesokurtic - not too flat and not too peaked





Kurtosis

Leptokurtic - peaked distribution

