Skewed and Symmetric

- **Symmetric**

  distribution of data is symmetric if the left half of its histogram is roughly a mirror image of its right half

- **Skewed**

  distribution of data is skewed if it is not symmetric and extends more to one side than the other

Skewed Left or Right

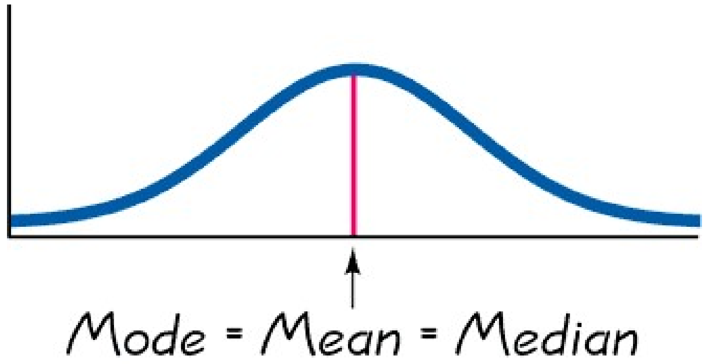
- **Skewed to the left**

  (also called negatively skewed) have a longer left tail, mean and median are to the left of the mode

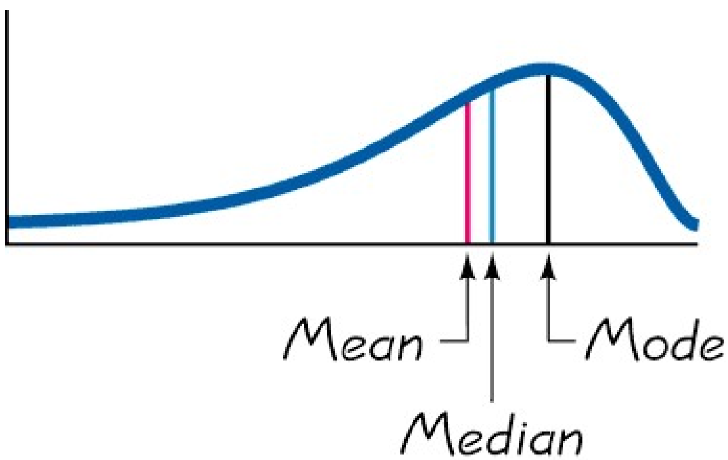
- **Skewed to the right**

  (also called positively skewed) have a longer right tail, mean and median are to the right of the mode

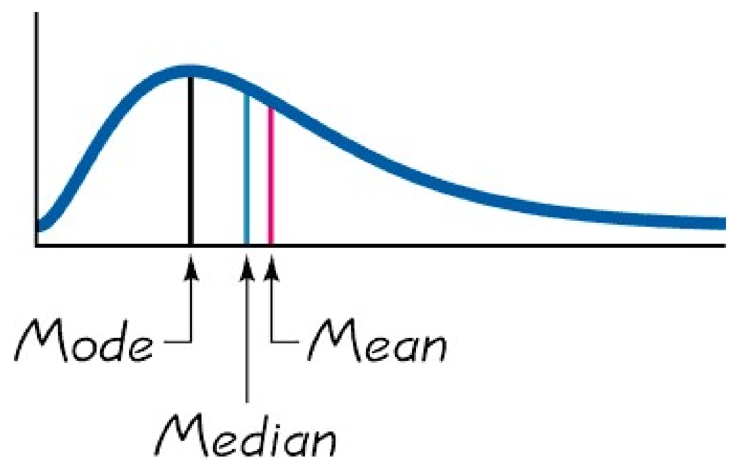
Skewness



(b) Symmetric



(a) Skewed to the Left (Negatively)

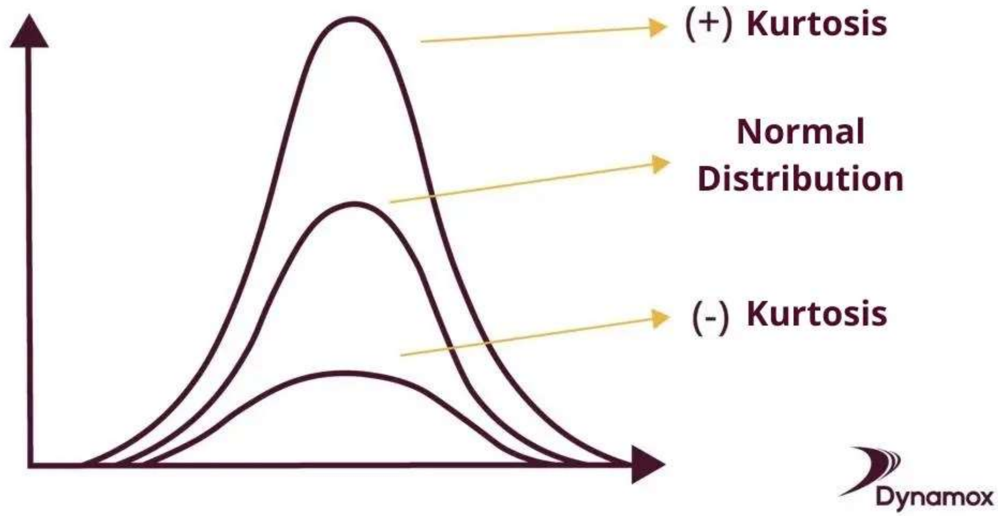


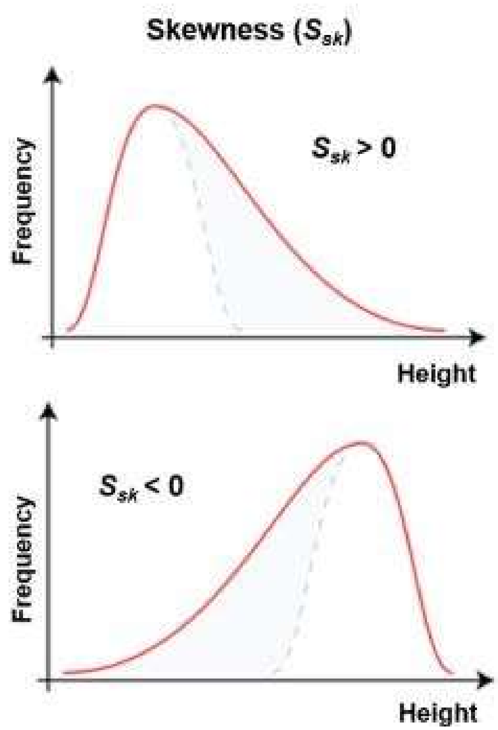
(c) Skewed to the Right (Positively)

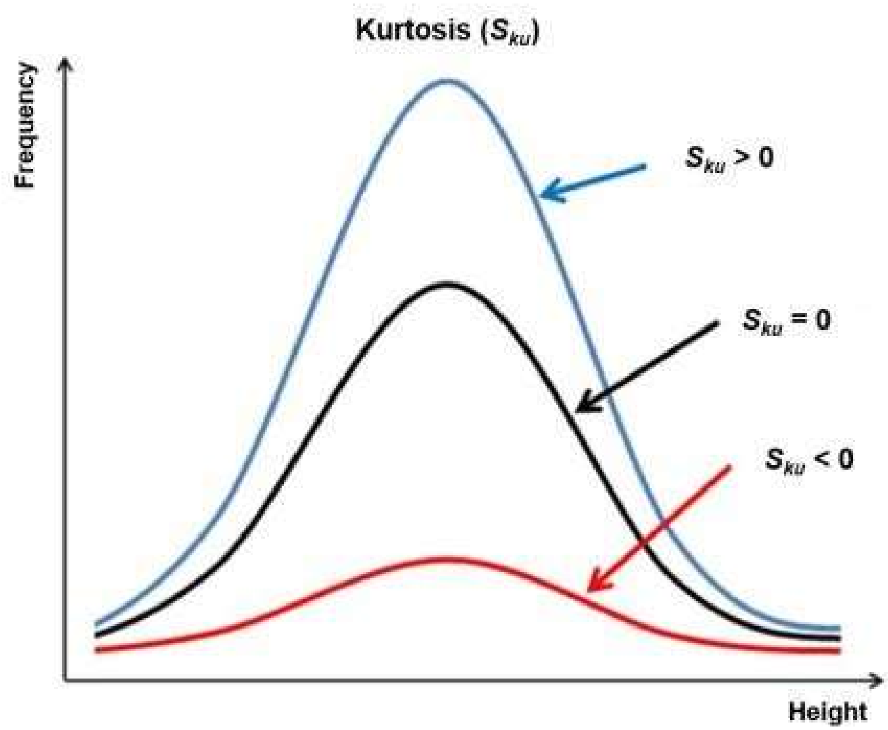
What is Kurtosis?

- Kurtosis is a statistical measure that describes the "tailedness" of a distribution.
- It indicates how much data resides in the tails versus the center of the distribution.

What is Kurtosis?



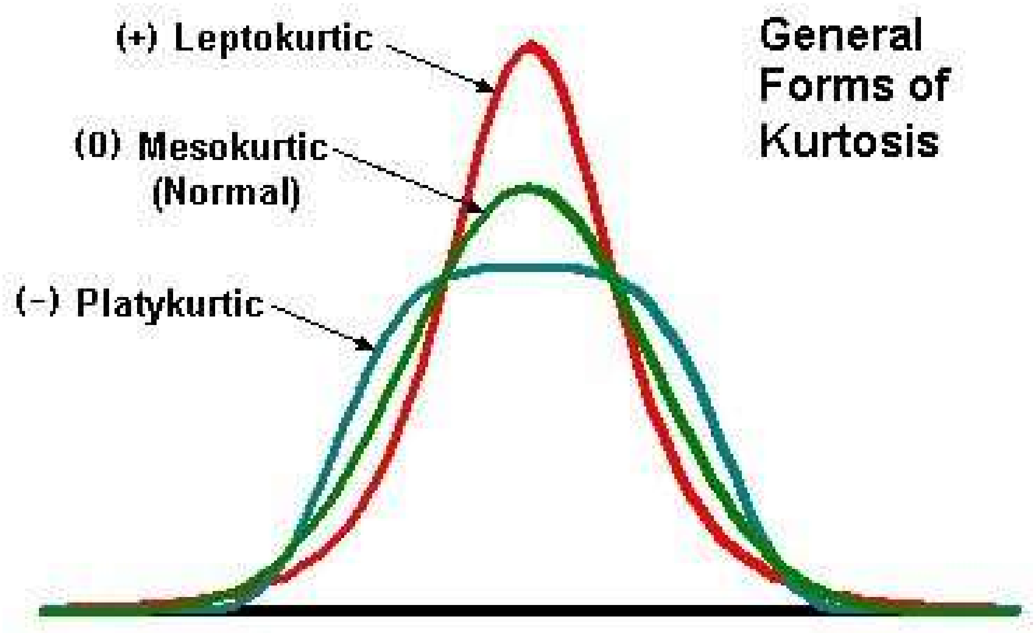




Key Takeaways: Kurtosis

- Describes the “fatness” of the tails in probability distributions.
- High kurtosis means data extends farther from the mean.
- Three categories: mesokurtic (normal), platykurtic (less than normal), leptokurtic (more than normal).
- Kurtosis risk measures how often an investment’s price moves dramatically. Indicates the level of risk in an investment.

What is Kurtosis?



Formula and Calculation

Calculating Kurtosis:

The formula for sample kurtosis is:

$$\text{Kurtosis} = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \times \sum \left(\frac{x_i - \bar{x}}{s}\right)^4 - \frac{3(n-1)^2}{(n-2)(n-3)}$$

Where:

- n is the number of observations.
- x_i is each individual observation.
- $\bar{x}$ is the mean of the observations.
- s is the standard deviation of the observations.

Percentiles

are measures of location. There are 99 **percentiles** denoted P_1, P_2, . . . P_{99}, which divide a set of data into 100 groups with about 1% of the values in each group.

Finding the Percentile of a Data Value

$$\text{Percentile of value } x = \frac{\text{number of values less than } x}{\text{total number of values}} \cdot 100$$

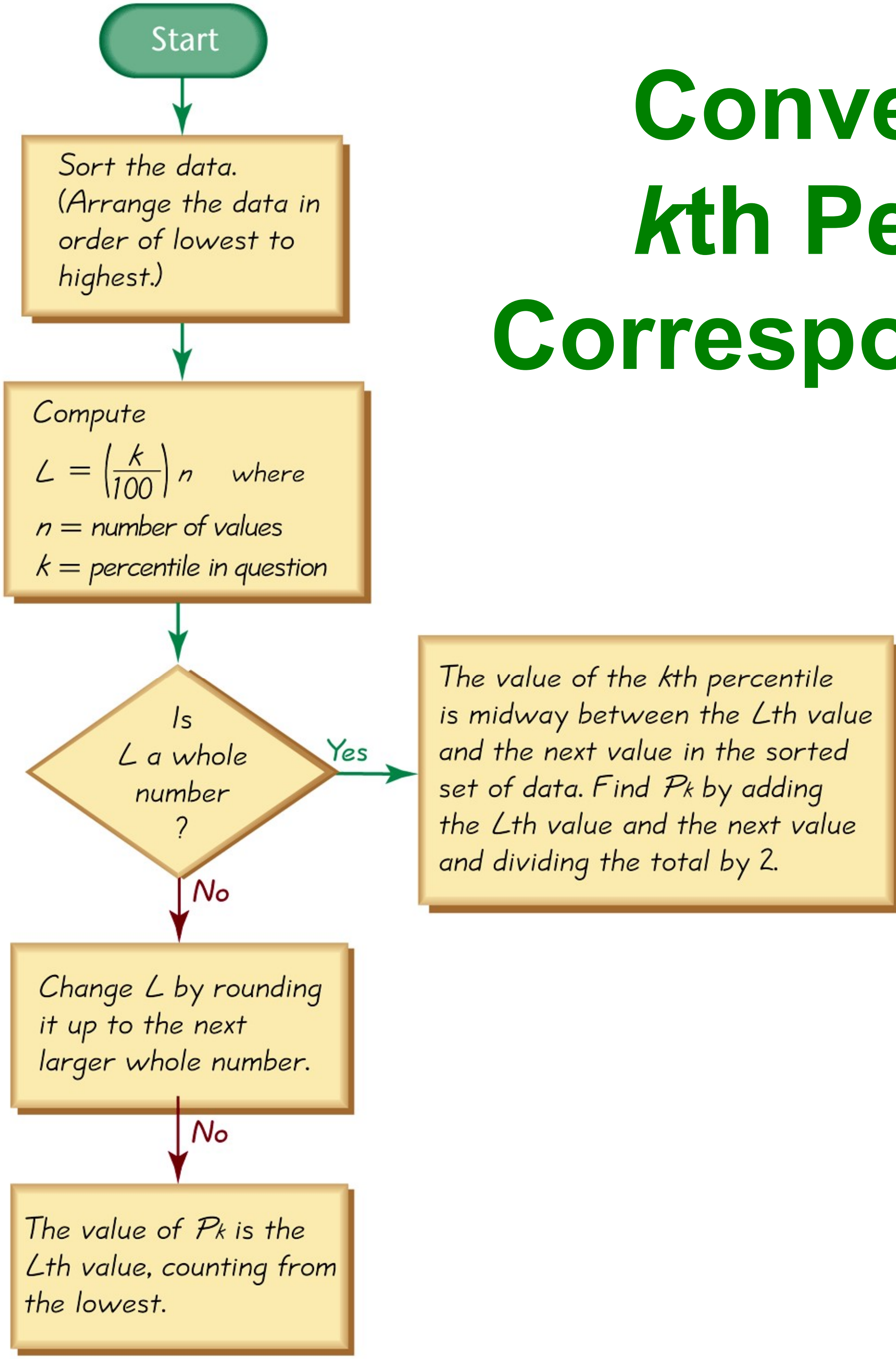
Converting from the *k*th Percentile to the Corresponding Data Value

Notation

$$L = \frac{k}{100} \cdot n$$

- n total number of values in the data set
- k percentile being used
- L locator that gives the **position** of a value
- P_k *k*th percentile

Converting from the *k*th Percentile to the Corresponding Data Value



Quartiles

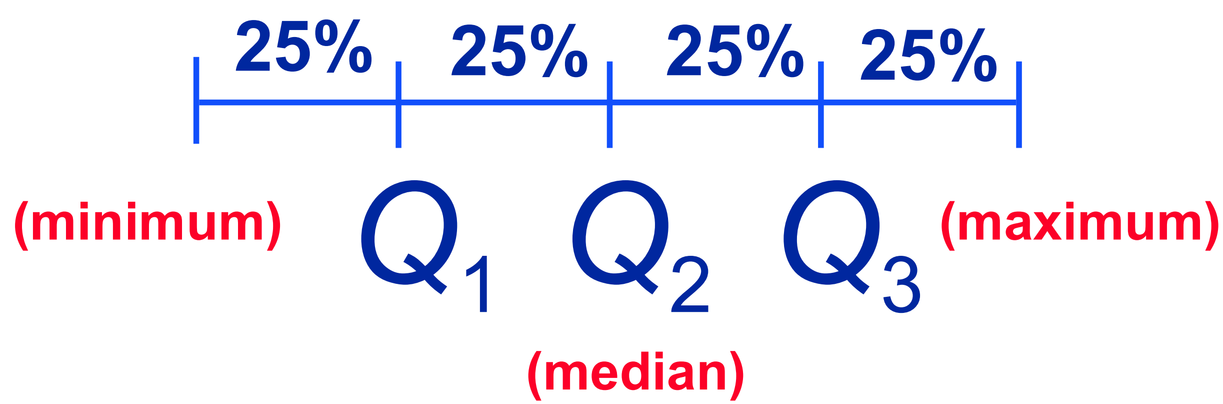
Are measures of location, denoted Q_1, Q_2, and Q_3, which divide a set of data into four groups with about 25% of the values in each group.

- Q_1 (First Quartile) separates the bottom 25% of sorted values from the top 75%.
- Q_2 (Second Quartile) same as the median; separates the bottom 50% of sorted values from the top 50%.
- Q_3 (Third Quartile) separates the bottom 75% of sorted values from the top 25%.

Quartiles

$$Q_1,\ \ Q_2,\ \ Q_3$$

divide ranked scores into four equal parts



Some Other Statistics

- **Interquartile Range (or IQR):** $Q_3 - Q_1$
- **Semi-interquartile Range:** $\dfrac{Q_3 - Q_1}{2}$
- **Midquartile:** $\dfrac{Q_3 + Q_1}{2}$
- **10 - 90 Percentile Range:** $P_{90} - P_{10}$

5-Number Summary

- For a set of data, the **5-number summary** consists of the minimum value; the first quartile Q_1; the median (or second quartile Q_2); the third quartile, Q_3; and the maximum value.