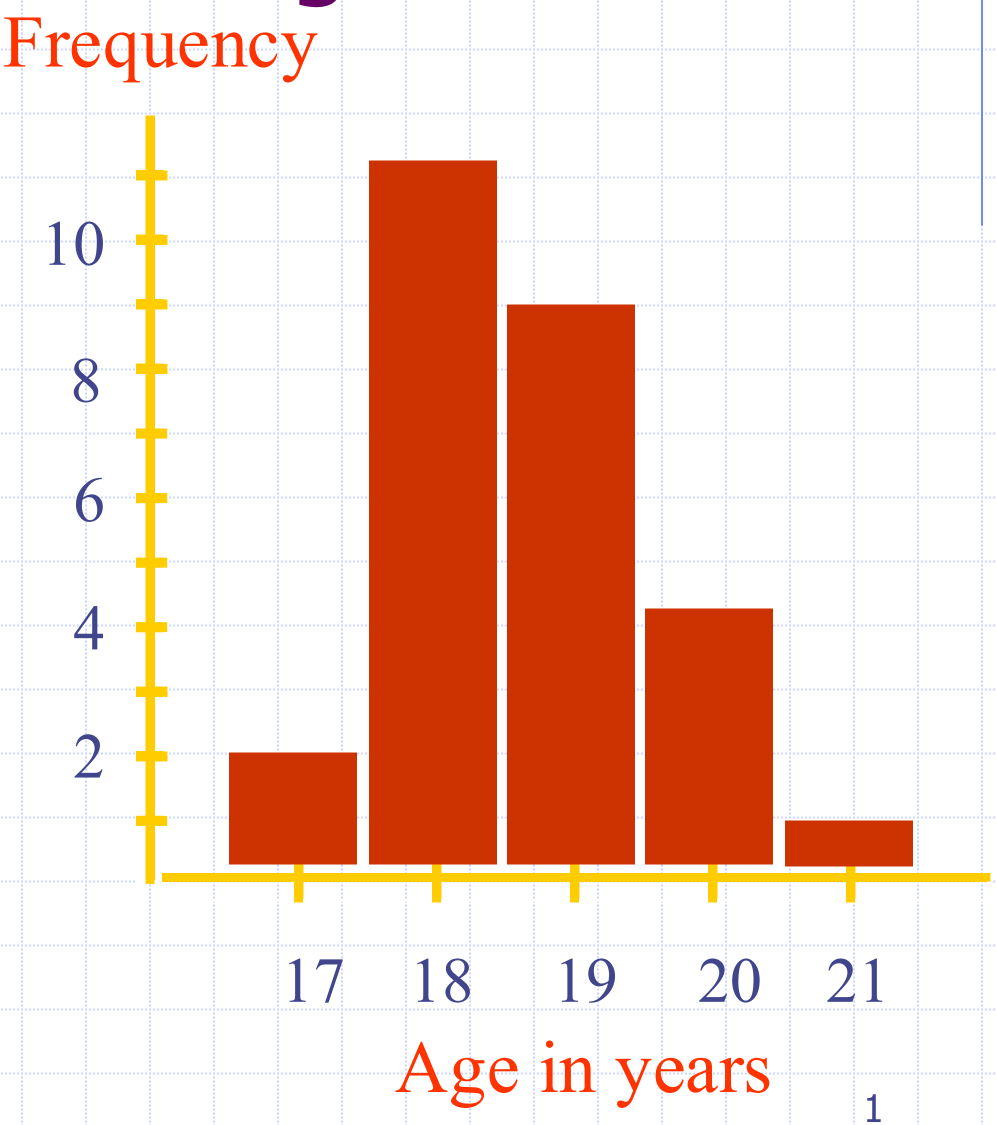
Ungrouped Frequency Distribution and Histogram

Ages of students in a statistics class

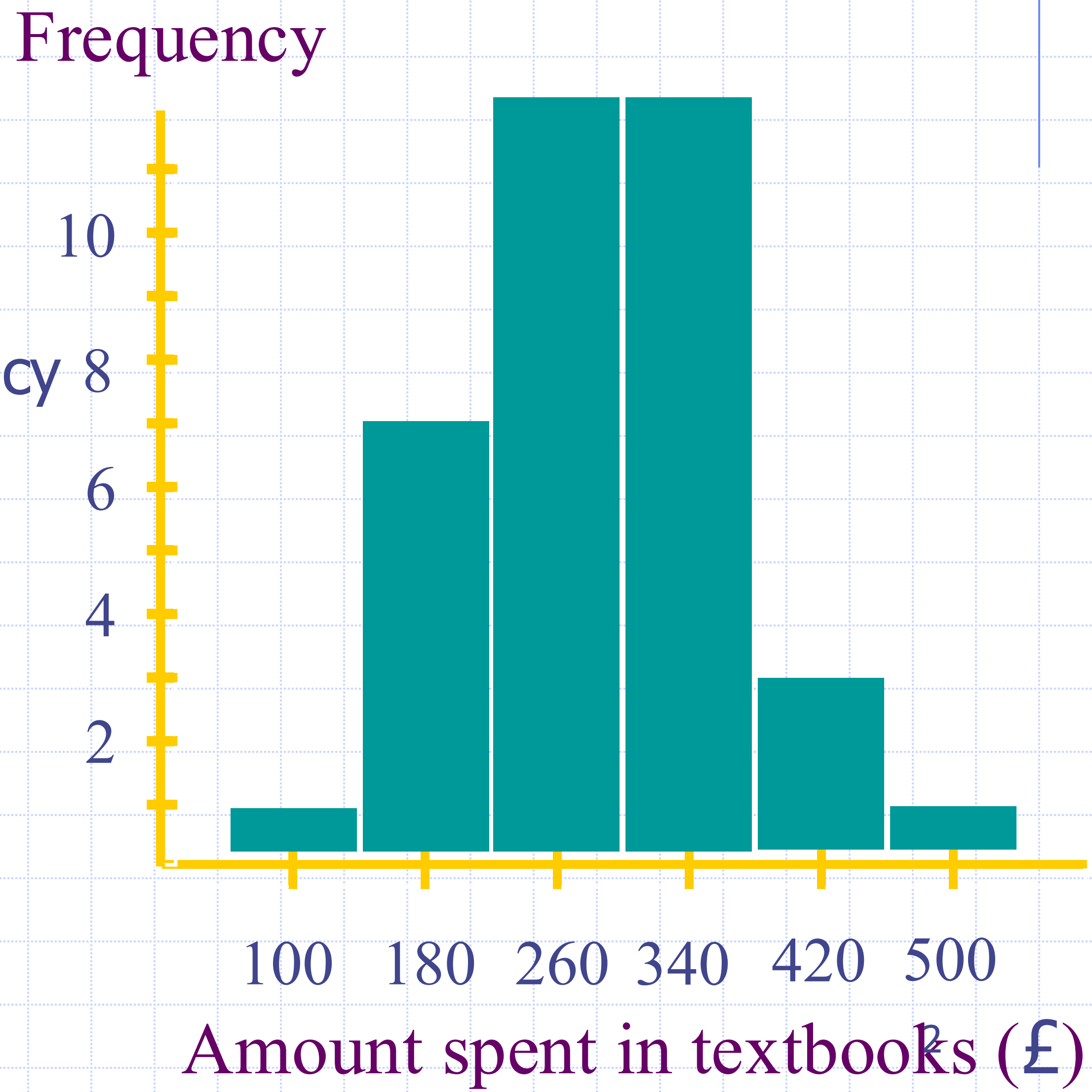
Age	Frequency
17	2
18	11
19	9
20	4
21	1



Grouped Frequency Distribution and Histogram

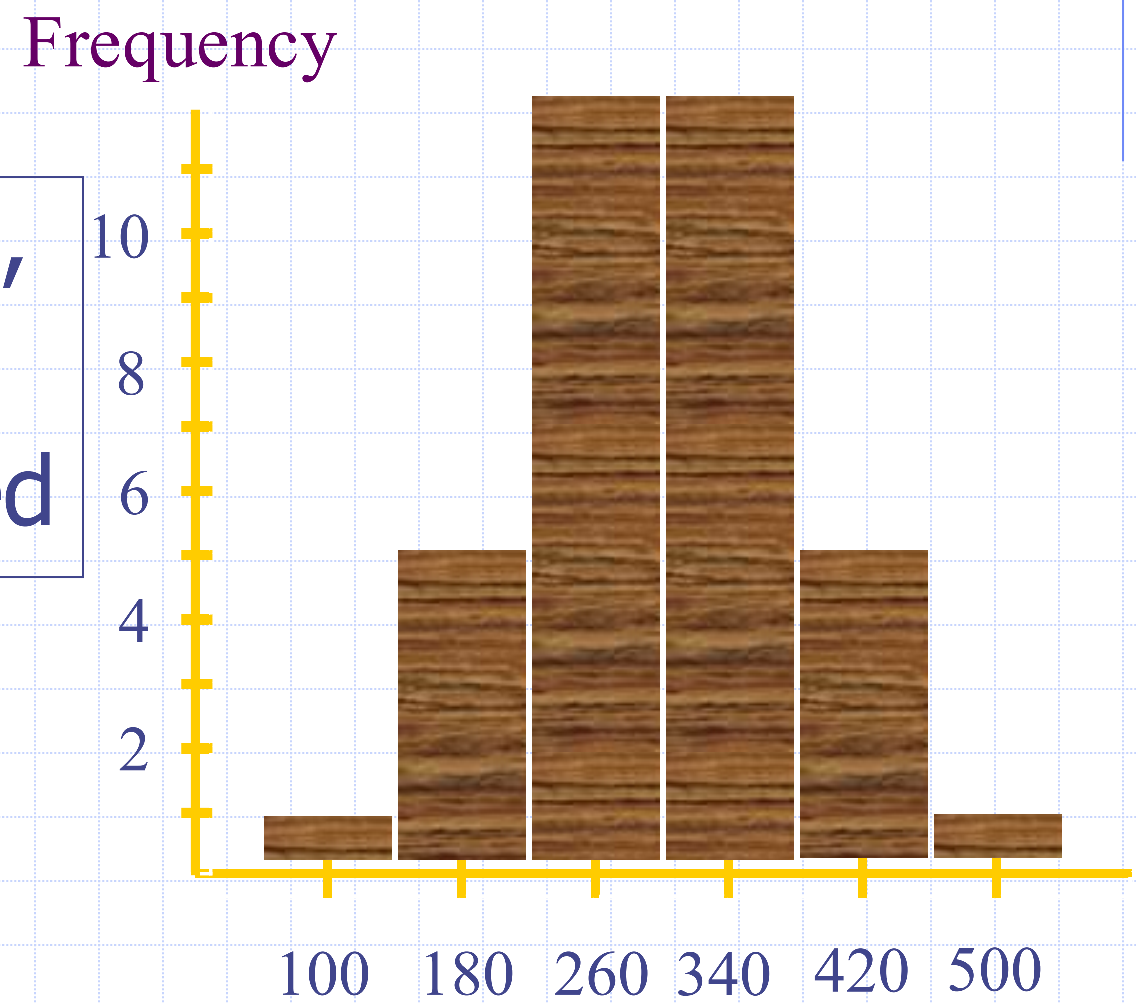
Amount spent on textbooks per student:

Amount (£)	Frequency
60-139	1
140-219	7
220-299	12
300-379	12
380-459	3
460-539	1



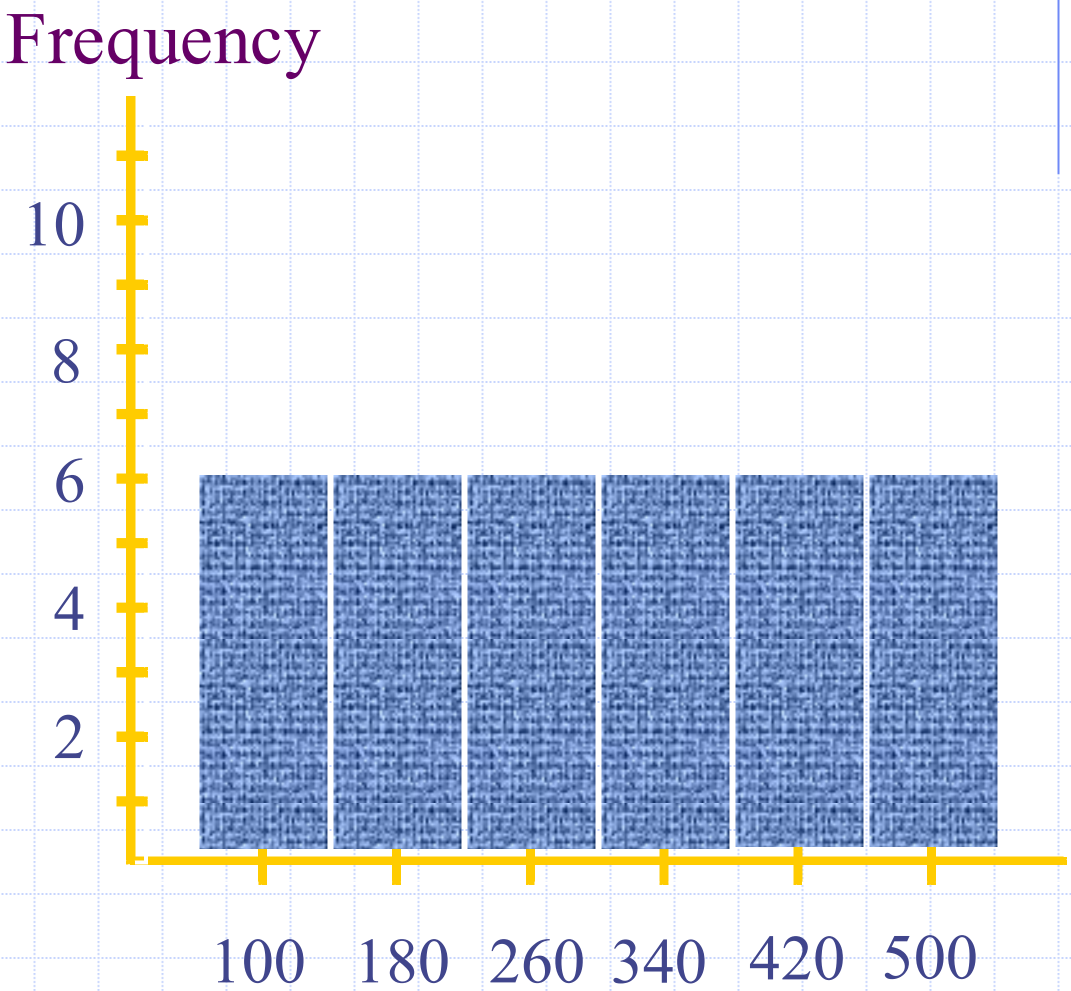
Shapes of Histograms I

Symmetrical, normal, or bell-shaped



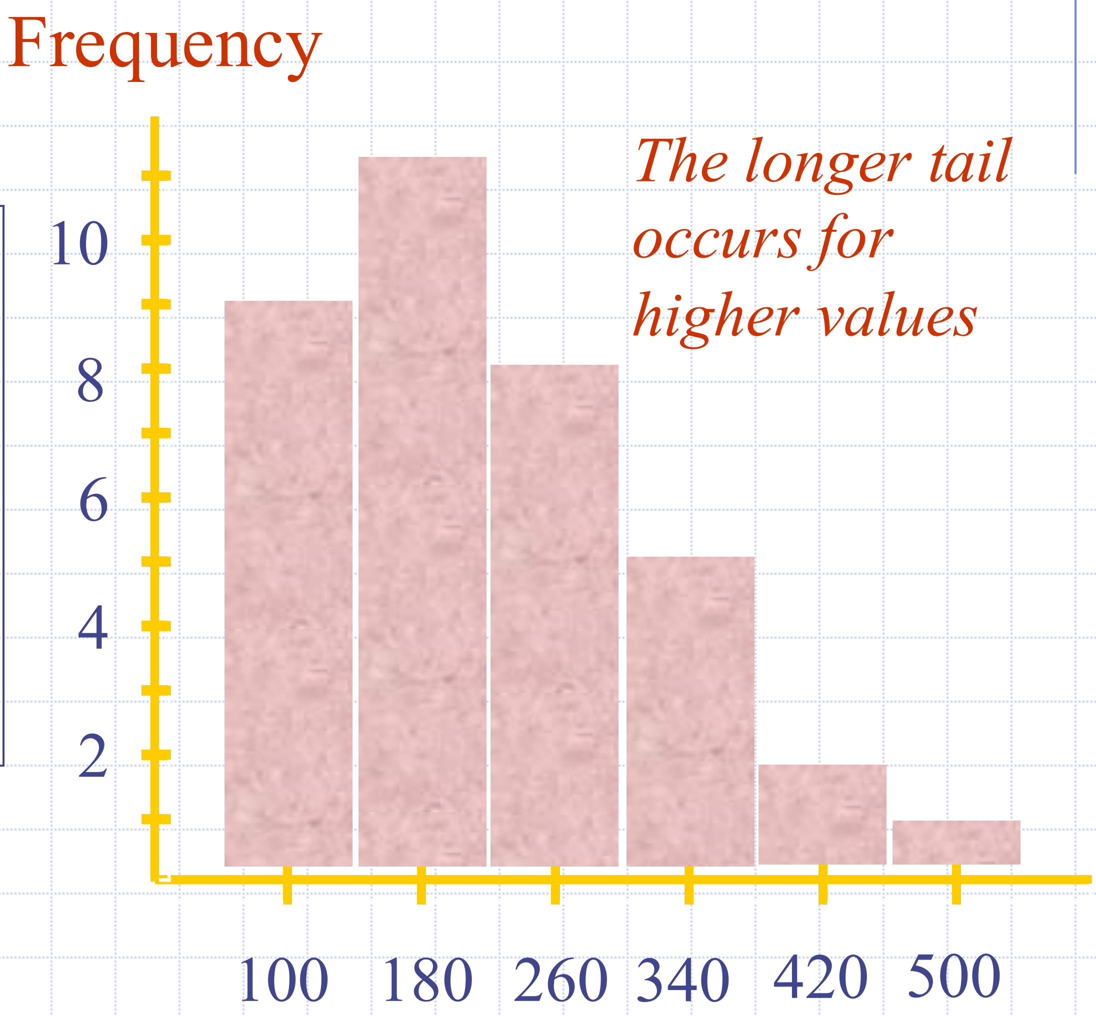
Shapes of Histograms II

Uniform
or
rectangular



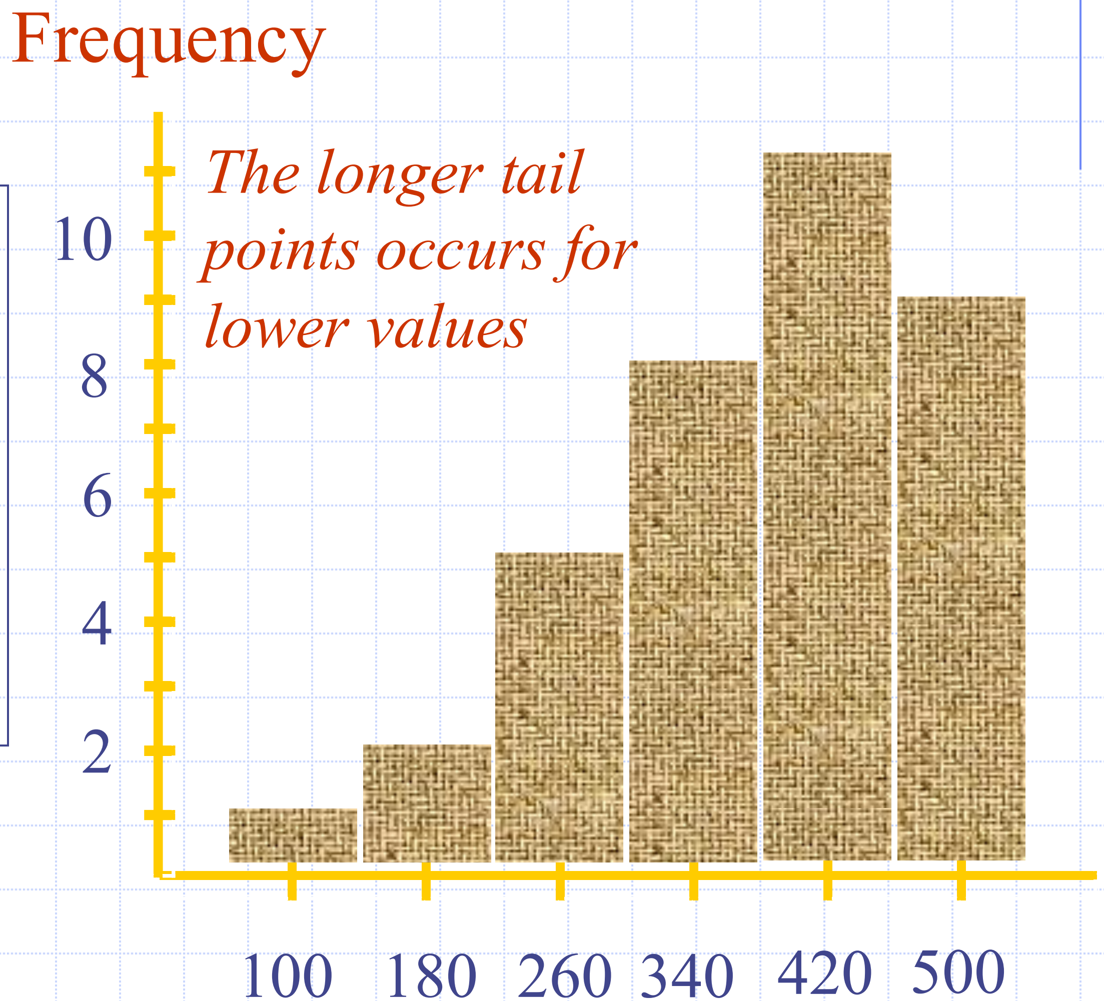
Shapes of Histograms III

Skewed right
or
Positively
skewed



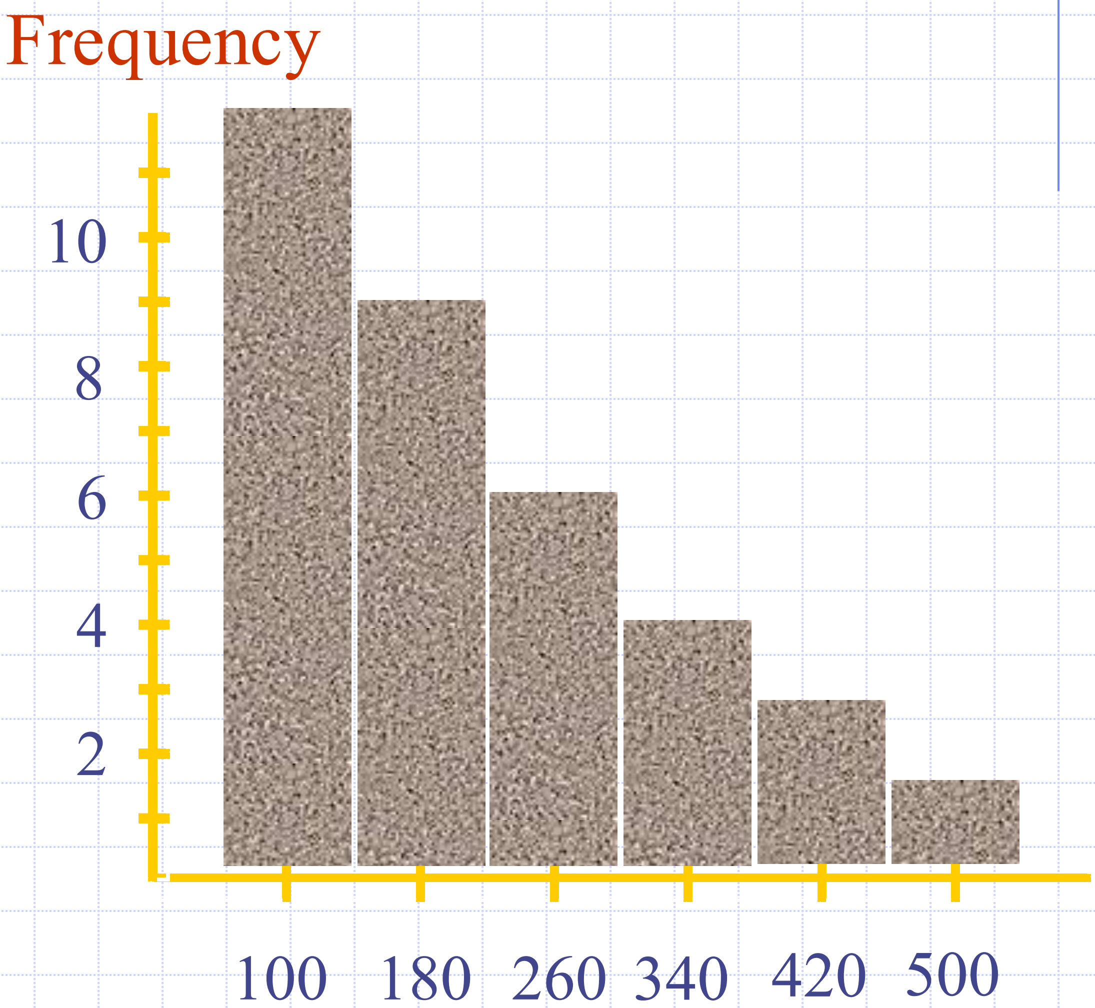
Shapes of Histograms IV

Skewed left
or
Negatively
skewed



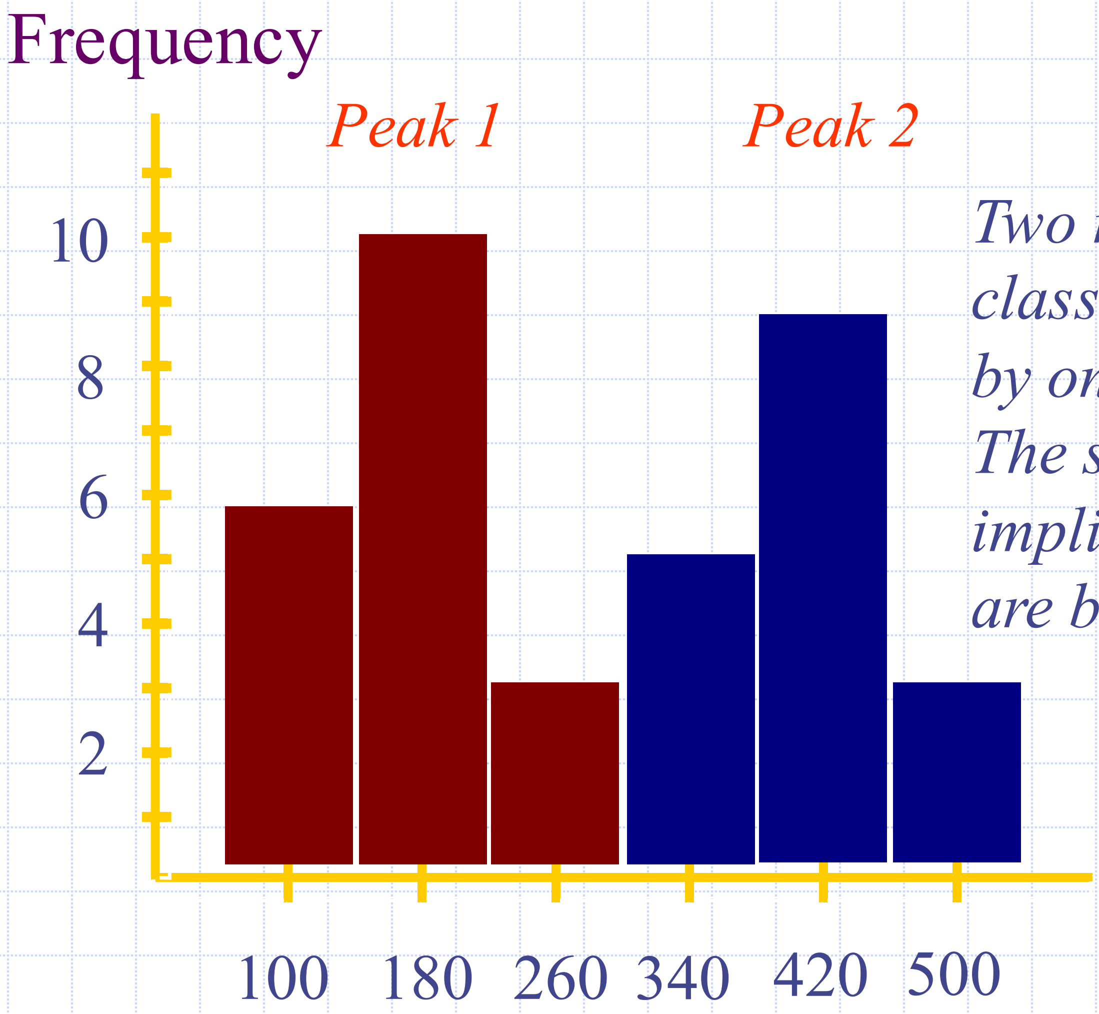
Shapes of Histograms V

J-shaped



Shapes of Histograms VI

Bimodal



Two most populous classes are separated by one or more classes. The situation often implies that 2 populations are being sampled.

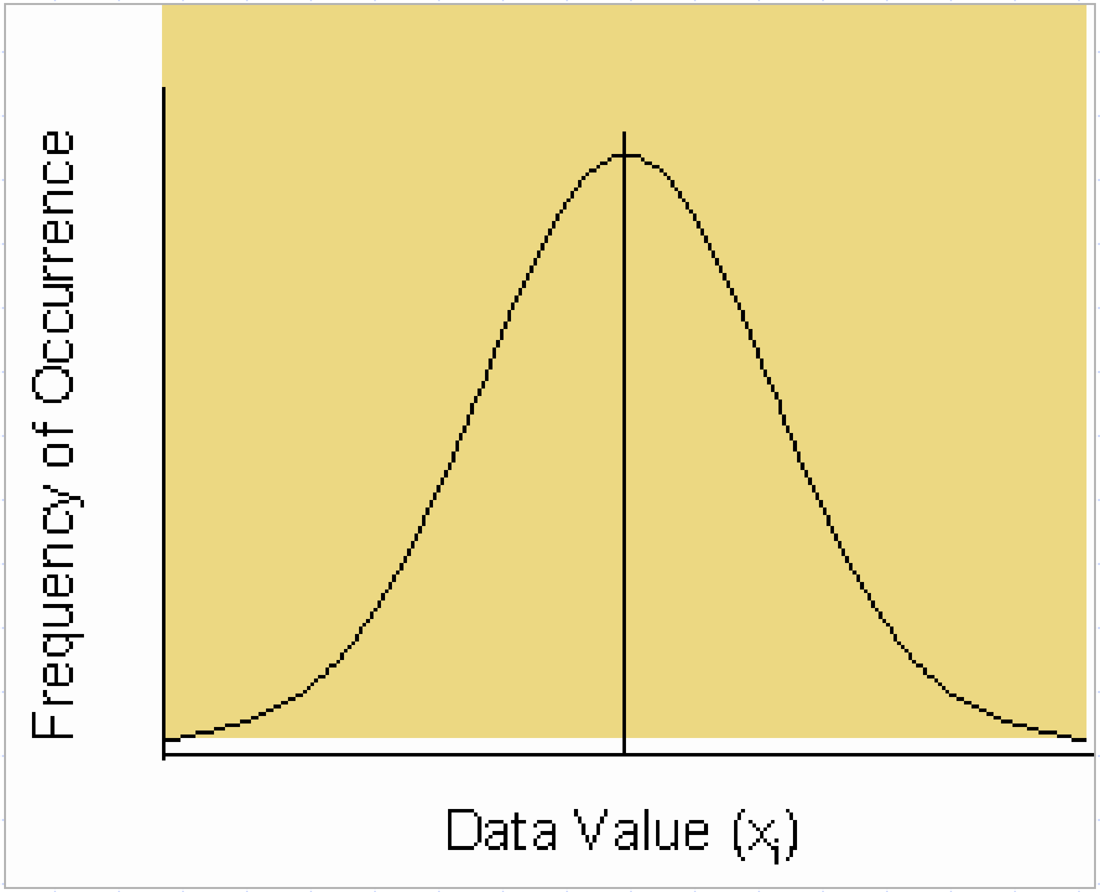
Shapes of distributions

- Symmetrical - normal
- Symmetrical - not normal
- Skewed right or left
- More than one peak

Frequency Distributions

- A frequency distribution is a model that indicates how the entire population is distributed based on sample data.
- Since the entire population is rarely considered, sample data and frequency distributions are used to estimate the shape of the actual distribution.
- This estimate allows inferences to be made about the population from which the sample data were obtained.
- It is a representation of how data points are distributed.
- It shows whether the data are located in a central location, scattered randomly or located uniformly over the whole range.

The graph of the frequency distribution will display the general variability and the symmetry of the data.



The frequency distribution may be represented in the form of an equation and as a graph

Frequency Distribution

- When using a frequency distribution, the interest is rarely in the particular set of data being investigated.
- In virtually all cases, the data are samples from a larger set or population.
- Sometimes, it is wrongfully assumed that data follow the pattern of a known distribution such as the normal.
- The data should be tested to determine if this is true.
- Goodness of Fit tests are used to compare sample data with known distributions.
- The inferences made from a frequency distribution apply to the entire population.

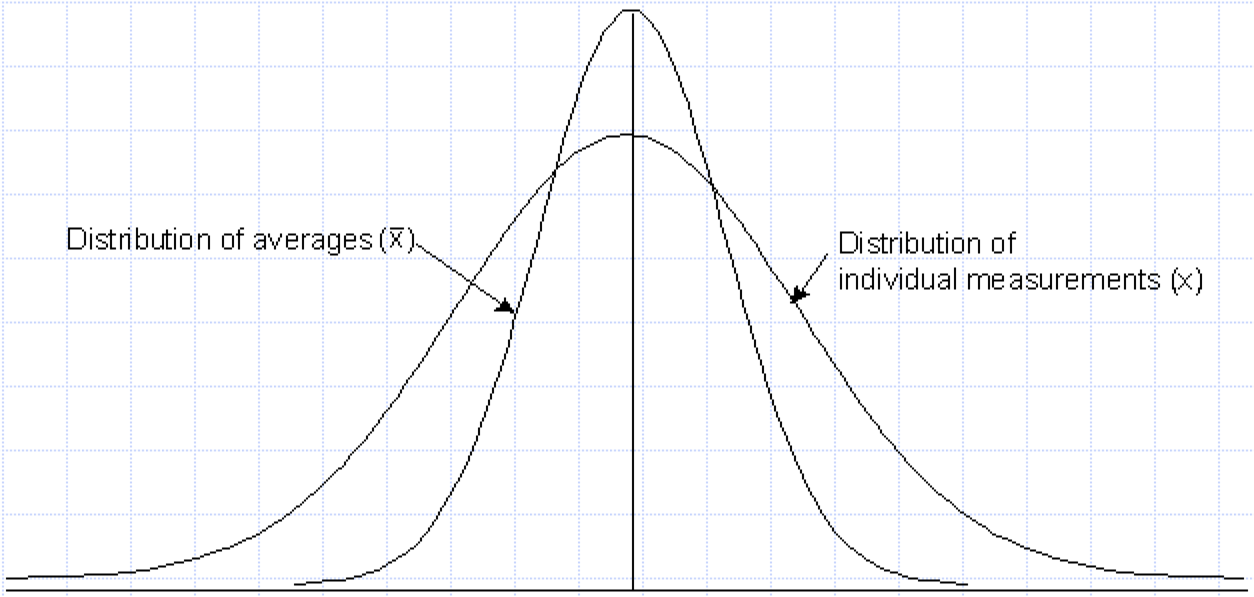
Central limit theory

- statisticians deal with distributions formed from individual measurements as well as distributions formed by sets of averages.

- **If the data are taken from the same population, there is a relationship between the distribution of individual measurements and the distribution of averages.**
- The means will be equal to $\bar{X} = \bar{\bar{X}}$
- If the standard deviation for individual measurements is s, then the standard error for the distribution of averages is $s/\sqrt{n}$.

- If a sample of 100 parts is divided into 20 subsets of 5 parts each, then n is 100 when calculating the variance and standard deviation of individual measurements and n is 5 when calculating the standard error using .
- standard error = the standard deviation for a set of averages

Distribution of individual measurements versus averages



Pattern of distribution of data

- Some distributions have more than one point of concentration and are called multimodal.

- When multimodal distributions occur, it is likely that portions of the output were produced under different conditions.

- A distribution with a single point of concentration is called unimodal.

  - **A distribution is symmetrical if the mean, median and mode are at the same location.**

Pattern of distribution of data

- **The symmetry of variation is indicated by skewness.**
- **If a distribution is asymmetrical it is considered to be skewed.**
- **The tail of a distribution indicates the type of skewness.**
- **If the tail goes to the right, the distribution is skewed to the right and is positively skewed.**
- **If the tail goes to the left, the distribution is skewed to the left and is negatively skewed.**
- **A symmetrical distribution has no skewness**

Pattern of distribution of data

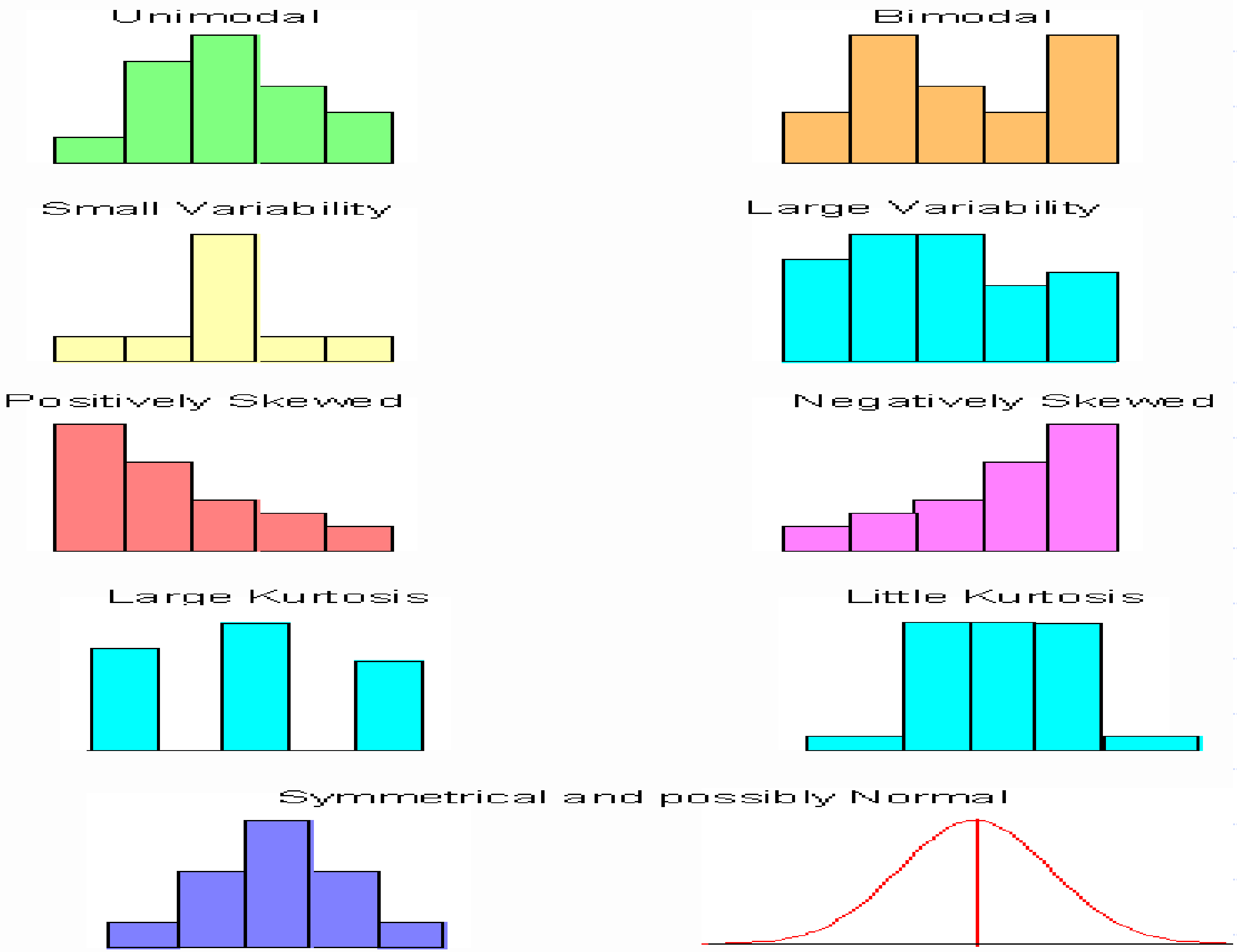
- **Kurtosis is defined as the state or quality of flatness or peakedness of a distribution.**

- **If a distribution has a relatively high concentration of data in the middle and out on the tails, but little in between, it has large kurtosis.**

- **If it is relatively flat in the middle and has thin tails, it has little kurtosis.**

- *If the frequencies of occurrence of a frequency distribution are cumulated from the lower end to the higher end of a scale, a cumulative frequency distribution is formed.*

SHAPES OF DISTRIBUTIONS



THE NORMAL CURVE

- **The normal curve is one of the most frequently occurring distributions in statistics.**
- **The pattern that most distributions form tend to approach the normal curve.**
- **It is sometimes referred to as the Gaussian curve named after Karl Friedrich Gauss (1777-1855) a German mathematician and astronomer.**
- **The normal curve is symmetrical about the average, but not all symmetrical curves are normal.**
- **For a distribution or curve to be normal, a certain proportion of the entire area must occur between specific values of the standard deviation.**

THE NORMAL CURVE

- There are two ways that the normal curve may be represented: The actual normal curve and the standard normal curve.

- [1] Actual Normal

The curve represents the distribution of actual data. The actual data points (xi) are represented on the abscissa (x-scale) and the number of occurrences are indicated on the ordinate (y-scale).

- [2] **Standard Normal**

The sample average and standard deviation are transformed to standard values with <u>A Mean Of Zero (0) And A Standard Deviation Of One (1).</u> The area under the curve represents the probability of being between various values of the standard deviation.

THE NORMAL CURVE

- By transforming the actual measurements to standard values, one table is used for all measurement scales.

- The abscissa on the actual normal curve is denoted **by x** and the abscissa on the standard normal curve is denoted **by Z.**

- The relationship between x and Z:

$$Z = \frac{(x_i - \overline{x})}{s}$$

- This is known as the **transformation formula**. It transforms the x value to its corresponding Z value.

- <u>A distribution of averages may also be represented with the normal curve.</u>

Normal curve

- The abscissa on the actual normal curve for a distribution of averages is denoted by $\bar{x}$
- The center is denoted by $\bar{\bar{x}}$ (the average of averages.)

Normal curve

$\frac{s}{\sqrt{n}}$

- The relationship between $\bar{x}$ and Z: $$Z = \frac{(\bar{x}_i - \bar{\bar{x}})}{s/\sqrt{n}}$$

- The statistic $s/\sqrt{n}$ is the standard error or the standard deviation for a set of **averages**.

Normal curve

- The standard normal curve areas are used to make certain forecasts and predictions about the population from which the data were taken.
- The standard normal curve areas are probability numbers. The area indicates the probability of being between two values on the Z scale.

Areas Under the Standard Normal Curve