Quartiles

Quartiles are merely particular percentiles that divide the data into quarters, namely:

Q_1 = 1st quartile = 25th percentile (P_{25})
Q_2 = 2nd quartile = 50th percentile = median (P_{50})
Q_3 = 3rd quartile = 75th percentile (P_{75})

Quartile Example

Using the applicant (aptitude) data, the first quartile is:

$$n \bullet \frac{P}{100} = (50)(.25) = 12.5$$

Rounded up Q_1 = 13th ordered value = 46

Similarly the third quartile is:

$$n \bullet \frac{P}{100} = (50)(.75) = 37.5 \quad 38 \text{ and } \quad Q_3 = 75$$

Interquartile Range

The **interquartile range** (IQR) is essentially the middle 50% of the data set

$$\text{IQR} = Q_3 - Q_1$$

Using the applicant data, the IQR is:

IQR = 75 - 46 = 29

Z-Scores

- **Z-score** determines the ***relative position*** of any particular data value x and is ***based*** on the ***mean*** and ***standard deviation*** of the data set
- The Z-score is expresses the number of standard deviations the value x is from the mean
- A negative Z-score implies that x is to the left of the mean and a positive Z-score implies that x is to the right of the mean

Z Score Equation

$$z = \frac{x - \bar{x}}{s}$$

For a score of 83 from the aptitude data set,

$$z = \frac{83 - 60.66}{18.61} = 1.22$$

For a score of 35 from the aptitude data set,

$$z = \frac{35 - 60.66}{18.61} = -1.36$$

Standardizing Sample Data

The process of subtracting the mean and dividing by the standard deviation is referred to as **standardizing the sample data**.

The corresponding z-score is the standardized score.

Measures of Shape

- ***Skewness***
  - Skewness measures the tendency of a distribution to stretch out in a particular direction
- ***Kurtosis***
  - Kurtosis measures the peakedness of the distribution

Skewness

- In a symmetrical distribution the mean, median, and mode would all be the same value and Sk = 0
- A positive Sk number implies a shape which is skewed right and the
  mode < median < mean
- In a data set with a negative Sk value the
  mean < median < mode

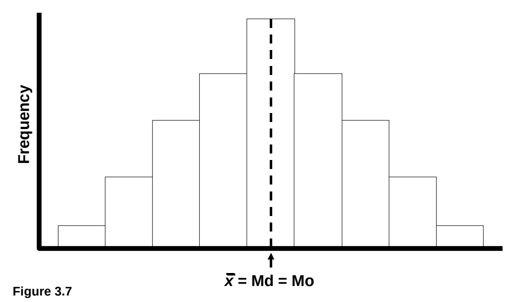
Skewness Calculation

Pearsonian coefficient of skewness

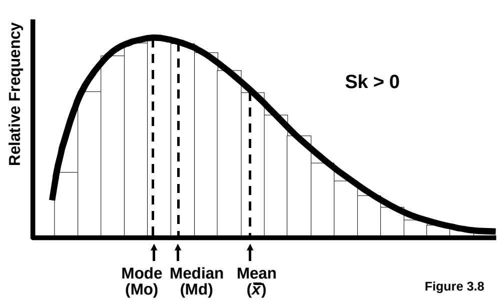
$$\mathbf{Sk} = \frac{3(\bar{x} - \mathrm{Md})}{s}$$

Values of Sk will always fall between -3 and 3

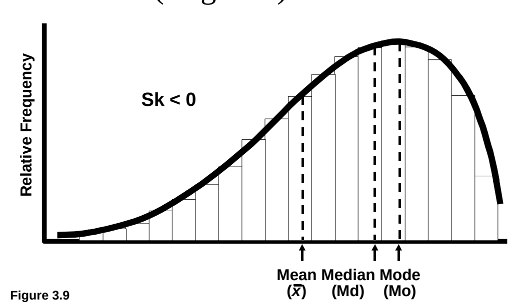
Histogram of Symmetric Data



Histogram with Right (Positive) Skew



Histogram with Left (Negative) Skew



Kurtosis

- Kurtosis is a measure of the peakedness of a distribution
- Large values occur when there is a high frequency of data near the mean and in the tails
- The calculation is cumbersome and the measure is used infrequently

Chebyshev's Inequality

1. At least 75% of the data values are between $\bar{x}$ - 2s and $\bar{x}$ + 2s, or
   At least 75% of the data values have a z-score value between -2 and 2
2. At least 89% of the data values are between $\bar{x}$ - 3s and $\bar{x}$ + 3s, or
   At least 75% of the data values have a z-score value between -3 and 3
3. In general, at least $(1-1/k^2)$ x 100% of the data values lie between $\bar{x}$ - ks and $\bar{x}$ + ks for any k>1

Empirical Rule

Under the assumption of a bell shaped population:

1. Approximately 68% of the data values lie between $\bar{x}$ - s and $\bar{x}$ + s (have z-scores between -1 and 1)
2. Approximately 95% of the data values lie between $\bar{x}$ - 2s and $\bar{x}$ + 2s (have z-scores between -2 and 2)
3. Approximately 99.7% of the data values lie between $\bar{x}$ - 3s and $\bar{x}$ + 3s (have z-scores between -3 and 3)

A Bell-Shaped (Normal) Population

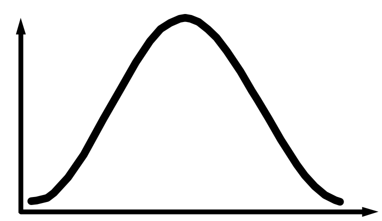


Figure 3.10

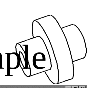
Chebyshev's Versus Empirical

Between	Actual Percentage	Chebyshev's Inequality Percentage	Empirical Rule Percentage
$\bar{x} - s$ and $\bar{x} + s$	66% (33 out of 50)	—	68%
$\bar{x} - 2s$ and $\bar{x} + 2s$	98% (49 out of 50)	75%	95%
$\bar{x} - 3s$ and $\bar{x} + 3s$	100% (50 out of 50)	89%	100%

Md **= 62**
Sk **= -.26**

Table 3.3

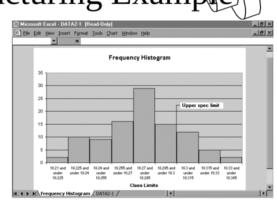
Allied Manufacturing Example



Is the Empirical Rule applicable to this data?

Probably yes.

Histogram is approximately bell shaped.



$\bar{x} - 2s$ = 10.275 and $\bar{x} + 2s$ = 10.3284

96 of the 100 data values fall between these limits closely approximating the 95% called for by the Empirical Rule