

Numerical Measures of Relative Standing

z-score – the distance between a measurement x and the mean, expressed in standard units

Use of standard units allows comparison across data sets

$$z = \frac{x - \mu}{\sigma} \qquad z = \frac{x - \bar{x}}{s}$$

28

Methods for Detecting Outliers

Box Plots

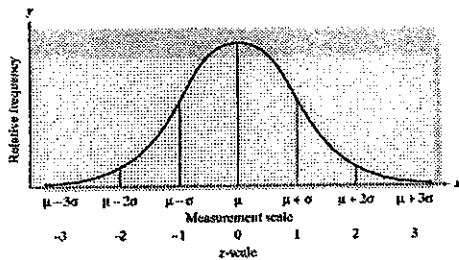
- based on quartiles, values that divide the dataset into 4 groups
- Lower Quartile Q_L – 25th percentile
- Middle Quartile - median
- Upper Quartile Q_U – 75th percentile
- Interquartile Range (IQR) = $Q_U - Q_L$

32

Numerical Measures of Relative Standing

More on z-scores

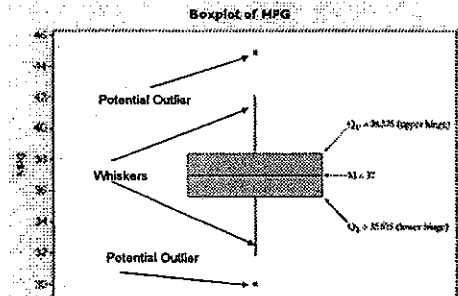
Z-scores follow the empirical rule for mound distributions



30

Methods for Detecting Outliers

Box Plots



Not on plot – inner and outer fences, which determine potential outliers

33

Methods for Detecting Outliers

Outlier – an observation that is unusually large or small relative to the data values being described

Causes

- Invalid measurement
- Misclassified measurement
- A rare (chance) event

2 detection methods

- Box Plots
- z-scores

31

Methods for Detecting Outliers

Rules of thumb

• Box Plots

- measurements between inner and outer fences are suspect
- measurements beyond outer fences are highly suspect

• Z-scores

- Scores of ± 3 in mound distributions (± 2 in highly skewed distributions) are considered outliers