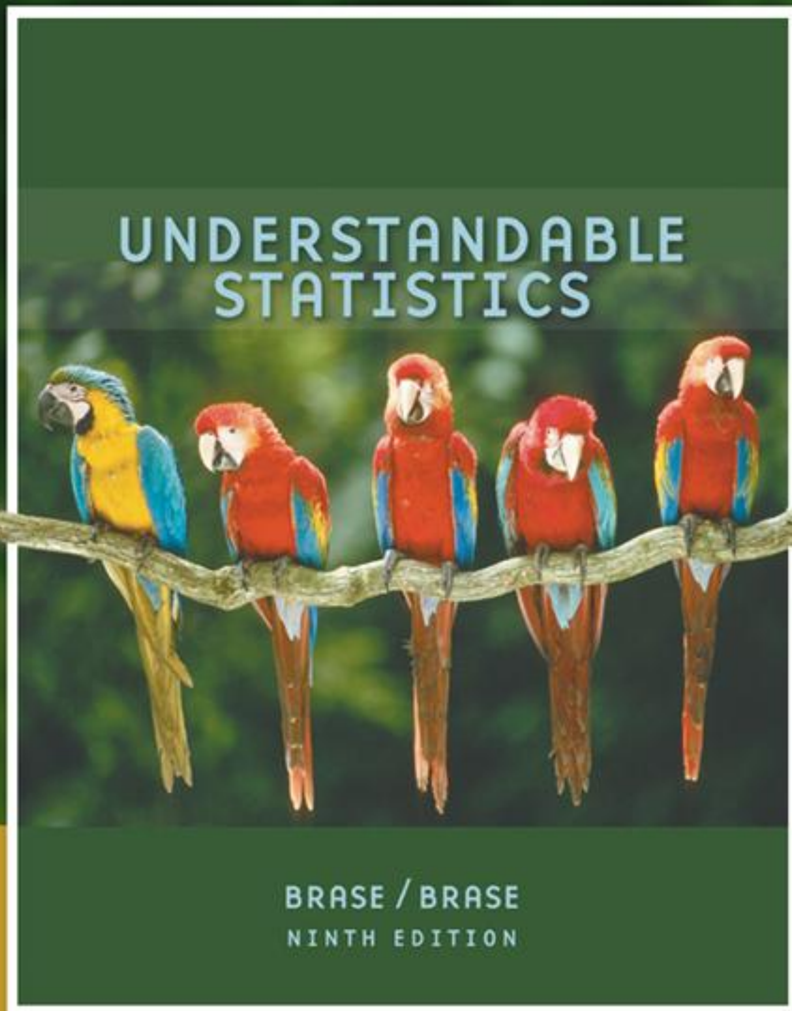


Chapter 3
Averages and Variation

Understandable Statistics
Ninth Edition
By Brase and Brase
Prepared by Yixun Shi
Bloomsburg University of Pennsylvania

Measures of Central Tendency

- We use the term “average” to indicate one number that gives a measure of center for a population or sample.
- This text investigates three “averages”:
  - Mode
  - Median
  - Mean

Mode

- The mode is the most frequently occurring value in a data set.
  - Example: Sixteen students are asked how many college math classes they have completed.

    {0, 3, 2, 2, 1, 1, 0, 5, 1, 1, 0, 2, 2, 7, 1, 3}

  - The mode is 1

Median

PROCEDURE

How to find the median

The median is the central value of an ordered distribution. To find it,

1. Order the data from smallest to largest.
2. For an *odd* number of data values in the distribution,

$$\text{Median} = \text{Middle data value}$$

3. For an *even* number of data values in the distribution,

$$\text{Median} = \frac{\text{Sum of middle two values}}{2}$$

For an ordered data set of size n,

$$\text{Position of the middle value} = \frac{n+1}{2}$$

Mean

$$\text{Sample mean} = \bar{x} = \frac{\Sigma x}{n}$$

- Read "*x-bar*"

$$\text{Population mean} = \mu = \frac{\Sigma x}{N}$$

- Read "*mu*"

Trimmed Mean

- Order the data and remove *k%* of the data values from the bottom and top.
- 5% and 10% trimmed means are common.
- Then simply compute the mean with the remaining data values.

Resistant Measures of Central Tendency

- A resistant measure will not be affected by extreme values in the data set.
- The mean is not resistant to extreme values.
- The median is resistant to extreme values.
- A *trimmed mean* is also resistant.

Critical Thinking

- Four levels of data – nominal, ordinal, interval, ratio
- Mode – can be used with all four levels.
- Median – may be used with ordinal level or above.
- Mean – may be used with interval or ratio level

Critical Thinking

- Mound-shaped symmetrical data – values of mean, median and mode are almost same.
- Skewed-left data – mean < median < mode.
- Skewed-right data – mean > median > mode.

Weighted Average

- At times, we may need to assign more importance to some of the data values.

$$Weighted\ \ Average = \frac{\sum xw}{\sum w}$$

- *x* is a data value.
- *w* is the weight assigned to that value.

Measures of Variation: Range

- Range = Largest value – smallest value
- Only two data values are used in the computation, so much of the information in the data is lost.

Sample Variance and Standard Deviation

- Sample Variance = s^2 = $$\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

- Sample Standard Deviation = s = $$\sqrt{s^2}$$

Population Variance and Standard Deviation

Population Variance = $$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}$$

Population Standard Deviation = $$\sigma = \sqrt{\sigma^2}$$

The Coefficient of Variation

For Samples

$$CV = \frac{s}{\bar{x}} \cdot 100$$

For Populations

$$CV = \frac{\sigma}{\mu} \cdot 100$$

Chebyshev's Theorem

Chebyshev's theorem

For *any* set of data (either population or sample) and for any constant k greater than 1, the proportion of the data that must lie within k standard deviations on either side of the mean is *at least*

$$1 - \frac{1}{k^2}$$

Chebyshev's Theorem

Results of Chebyshev's theorem

For *any* set of data:

- at *least* 75% of the data fall in the interval from $\mu - 2\sigma$ to $\mu + 2\sigma$.
- at *least* 88.9% of the data fall in the interval from $\mu - 3\sigma$ to $\mu + 3\sigma$.
- at *least* 93.8% of the data fall in the interval from $\mu - 4\sigma$ to $\mu + 4\sigma$.

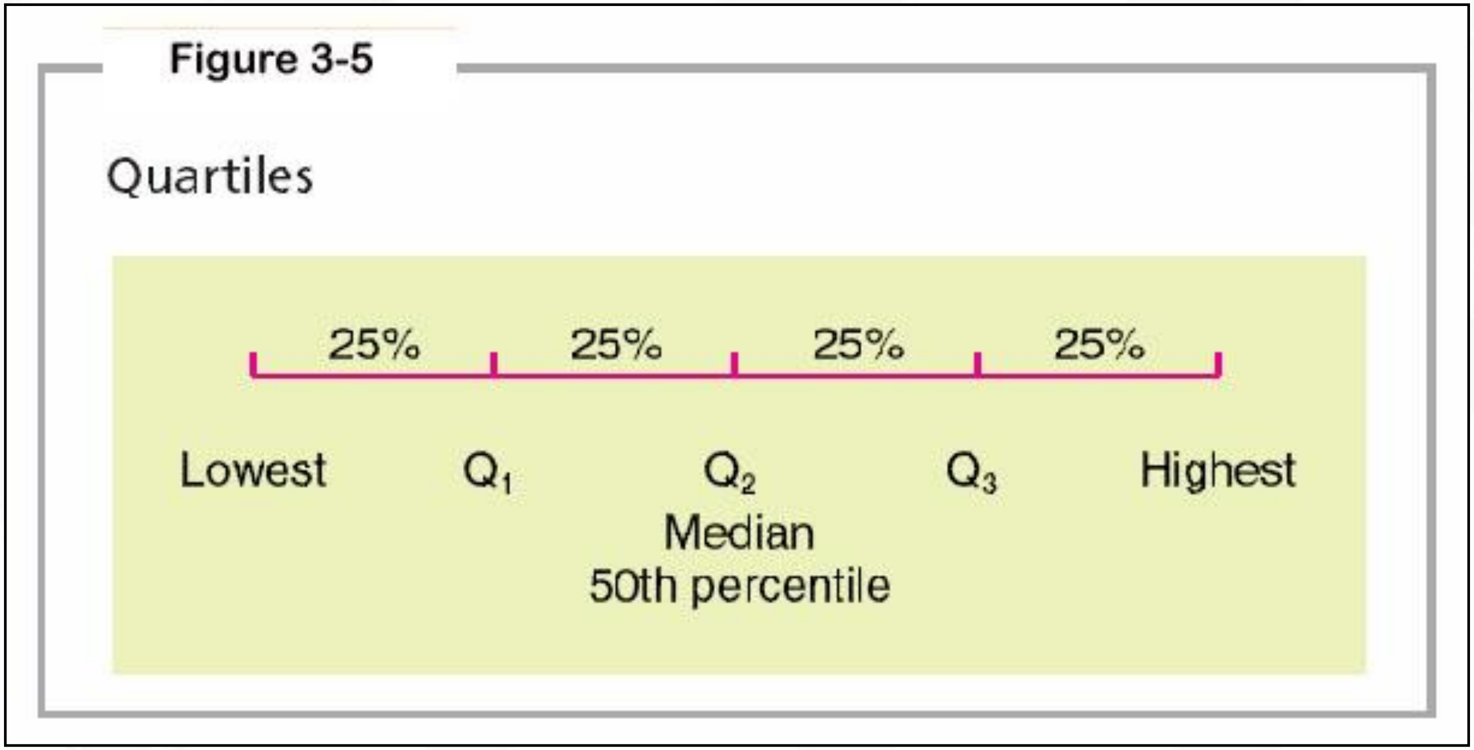
Critical Thinking

- Standard deviation or variance, along with the mean, gives a better picture of the data distribution.
- Chebyshev's theorem works for all kinds of data distribution.
- Data values beyond 2.5 standard deviations from the mean may be considered as outliers.

Percentiles and Quartiles

- For whole numbers *P*, $1 \le P \le 99$, the P^{th} percentile of a distribution is a value such that *P%* of the data fall below it, and *(100-P)%* of the data fall at or above it.
- Q_1 = 25th Percentile
- Q_2 = 50th Percentile = The Median
- Q_3 = 75th Percentile

Quartiles and Interquartile Range (IQR)



$$\text{Interquartile range} = Q_3 - Q_1$$

Computing Quartiles

PROCEDURE

How to compute quartiles

1. Order the data from smallest to largest,
2. Find the median. This is the 2nd quartile.
3. The first quartile Q_1 is then the median of the lower half of the data; that is, it is the median of the data falling *below* the Q_2 position (and not including Q_2).
4. The third quartile Q_3 is the median of the upper half of the data; that is, it is the median of the data falling *above* the Q_2 position (and not including Q_2).

Five Number Summary

- A listing of the following statistics:
  - Minimum, Q_1, Median, Q_3, Maximum

- Box-and-Whisder plot – represents the five-number summary graphically.

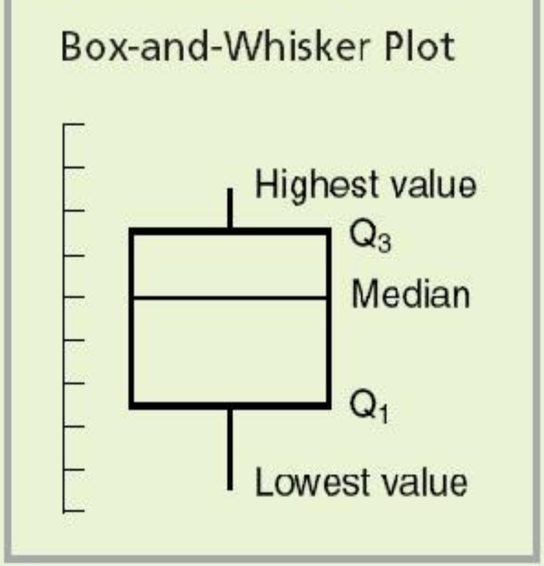
Box-and-Whisker Plot Construction

PROCEDURE

How to make a box-and-whisker plot

1. Draw a vertical scale to include the lowest and highest data values.
2. To the right of the scale draw a box from Q_1 to Q_3.
3. Include a solid line through the box at the median level.
4. Draw solid lines, called *whiskers*, from Q_1 to the lowest value and from Q_3 to the highest value.

Figure 3-6



Tutorial de como hacer un Boxplot

http://math.uprag.edu/boxplot/boxplot.htm

Critical Thinking

- Box-and-whisker plots display the spread of data about the median.
- If the median is centered and the whiskers are about the same length, then the data distribution is symmetric around the median.
- Fences – may be placed on either side of the box. Values lie beyond the fences are outliers.