

TA: Ruiyan Luo

Office: 4268 CSSC

Office hours: TW 2:30–3:30pm

Phone number: 262-8182

E-mail: rluo@stat.wisc.edu

1. Measures of center:

- sample mean (\bar{y}):

$$\bar{y} = \frac{\sum y_i}{n}$$

where the y_i 's are the observations in the sample and n is the sample size (that is, the number of y_i 's). Mean is sensitive to outlying values.

- median:

the value that mostly nearly lies in the middle of the sample. For ordered values, when sample size is odd, median=the value for the middle observation; when sample size is even, median = the average of the middle two.

2. Boxplot

- quartiles:

The median splits the distribution into two parts. The first quartile (Q_1), is the median of the data values in the lower half of the data set. The third quartile (Q_3), is the median of the data values in the upper half of the data set.

A common alternative definition for quartiles is: Q_1 has rank position $(.25)(n + 1)$ and Q_3 has rank position $(.75)(n+1)$.

Interquartile range (IQR): $IQR = Q_3 - Q_1$

- boxplot: a visual representatin of the five-number summary: the mininum, the maximum, the median and the quartiles. Note: The scale on the numeric axis should consistent without breaks and with the same scale throughout.

- outliers:

$$\text{lower fence} = Q_1 - 1.5 \times \text{IQR}$$

$$\text{upper fence} = Q_3 + 1.5 \times \text{IQR}$$

An outlier is a data point that falls outside of the fences.

- parallel boxplot and modified boxplot

3. Measures of Dispersion

- the range=maximum-minimum
- deviation=observation- \bar{y}
- the standard deviation (s):

$$s = \sqrt{\frac{\sum (y_i - \bar{y})^2}{n - 1}}$$

In this formula, the expression $\sum (y_i - \bar{y})^2$ denotes the sum of the squared deviations.

- variance (s^2): the squared standard deviation.
- the coefficient of variation= $\frac{s}{\bar{y}} \times 100\%$