AMS7: WEEK 1. CLASS 3

Frequency distributions. Visualizing data. Measures of Center

Friday April 3rd, 2015
MEASURES OF CENTER

• Value at the center or middle of the data set. The most used measures are:
  - Mean
  - Median
  - Mode
  - Midrange

Statistic: Sample Mean
- \( \bar{X} = \frac{\sum X}{n} \)

Parameter: Population Mean
- \( \mu = \frac{\sum X}{N} \)

- \( n \) is the number of values in the sample
- \( N \) is the population size
Measure of Center (Cont.)

- DISADVANTAGE OF THE MEAN: It is too sensible to extreme values
- MEDIAN: Is the middle value of the original data values when they are arranged in increasing order.
  - Case $n$ is odd: The median is exactly the center value
  - Case $n$ is even: The median is the average of the two middle values
- MODE: Value that occurs most frequently

**Note 1:** You can have two modes or more (bi-modal or multi-modal) if you have more than one value with the same frequency.

**Note 2:** Normally you provide more than one measure of center
Example: Monitoring lead in Air

- Maximum value established by EPA: 1.5 \mu g/m^3
- Five measured levels of lead taken on Sept. 11/2001

$$\bar{X} = \frac{\sum x}{n} = \frac{5.40 + 1.10 + 0.42 + 0.73 + 0.48 + 1.10}{6} = 1.538 \mu g/m^3$$

Q: Is 5.4 and outlier??

Mean when excluding the 5.4 value is 0.9575
• **Median**: is the middle value when the original data is arranged in order.
  - If the number of values is odd the median is exactly the middle value
  - If the number of values is even the median is the average of the two middle values

• Notation for the median: \( \tilde{X} \)

• Example: Monitoring lead in the Air

SORTED VALUES: (increasing order)

- \( n=6 \) (even number)

- Median = \( \tilde{X} = \frac{0.73 + 1.10}{2} = 0.915 \, \mu g/m^3 \)
MODE: Value that occurs more frequently

EXAMPLES:
1. 5.40, 1.10, 0.42, 0.73, 0.48, 1.10
2. 27, 27, 27, 55, 55, 55, 88, 88, 99
3. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10

Mode example 1 = 1.10 (single mode)
Mode example 2 = 27 and 55 (bimodal)
Mode example 3 = There is no mode
• MIDRANGE: \( \frac{\text{Max. Value} + \text{Min. Value}}{2} \)

EXAMPLE: Monitoring lead
Midrange = \( \frac{5.40 + 0.42}{2} = 2.910 \, \mu g/m^3 \)

• WEIGHTED MEAN: Each value has a different level of importance:

\[
\bar{X} = \frac{\sum w \cdot X}{\sum w}
\]

EXAMPLE: Grades of three test scores: 85, 90, 75; with weights: 20%, 30%, 50% respectively.

\[
\bar{X} = \frac{(20 \times 85) + (30 \times 90) + (50 \times 75)}{20 + 30 + 50} = 81.5
\]
Skewness

- A distribution of data is not necessarily symmetric