THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

Online Review Course of
Undergraduate Probability and Statistics

Review Lecture 3
Descriptive Statistics, part 2

Chris A. Mack
Adjunct Associate Professor

Course Website: www.lithoguru.com/scientist/statistics/review.html
Data sets accompanying this lecture: StatReview_Lecture2&3.xlsx

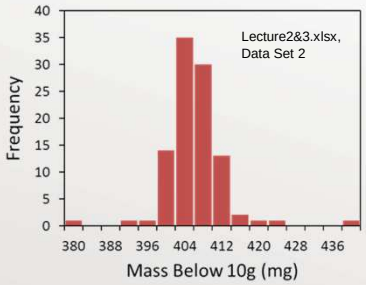
THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

Descriptive Statistics

- Descriptive statistics
  - Describe or summarize a large set of data with a graph and/or just a few numbers
  - Applies mainly to univariate data
  - The statistics that can be used depend on the measurement scale (nominal, ordinal, interval, or ratio)

THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

Descriptive Statistics

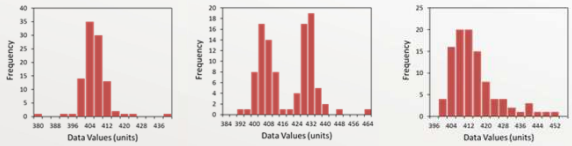


- How do we describe this data?
  - Shape
  - Central tendency
  - Spread
- Other measures could be important as well (e.g., skew)
- Aggregating data and looking only at summary statistics can hide important features of the data

THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

Distribution Shape

N = 100



Symmetric, Unimodal | Symmetric, Bimodal | Right Skewed

It is hard to see the shape from a histogram without many data points

THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

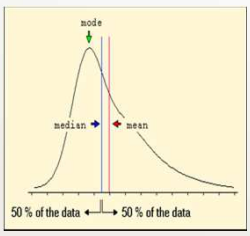
Measures of Central Tendency

- Mean: $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$
- Median:
  - rank order the data, pick out the middle value (odd number of data points) or the average of the two middle values (even number of data points)
- Mode:
  - Most frequent data value
- 10% Trimmed Mean:
  - Rank order the data, throw away the 10% of data with the largest values and the 10% of data with the smallest values, then calculate mean of the rest

THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

Measures of Central Tendency

- For a perfectly symmetrical data set, each of these central measures will be the same
- Comparing measures of central tendency tells us something about asymmetries



THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

Measures of Spread

- Range: max - min
  - An extremely unstable metric!
- Interquartile Range (IQR):
  - rank order the data, divide it in half
  - pick out the 25th percentile value (lower quartile, Q1)
  - pick out the 75th percentile value (upper quartile, Q3)
  - IQR = Q3 – Q1
- Mean absolute deviation $= \frac{1}{N}\sum_{i=1}^{N}|x_i - \bar{x}|$
- Standard deviation

THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

Measures of Spread

- Sample Variance (mean squared deviation from the mean):

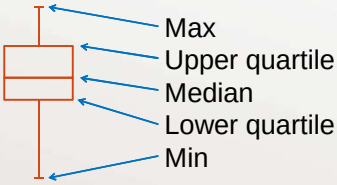
$$s^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2$$

- Standard deviation is the square root of the variance
- N-1 is the number of degrees of freedom (DoF), since the sample mean uses up one DoF
  - This results in an unbiased estimator for the variance
- For a population variance, use N rather than N-1

THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

Summary Statistics

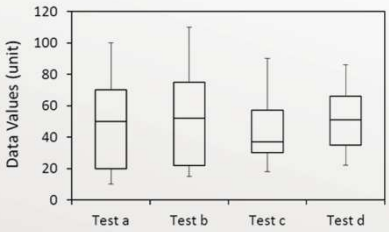
- The five-number summary:



Display graphically (box & whiskers plot) and/or in table form

THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

Box and Whiskers Plot



This makes for very quick comparison between the four tests.

THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

Practice

- Using the two data sets in Lecture2&3.xlsx, practice using Excel to calculate a variety of statistics
  - Measures of central tendency
  - Measures of spread
- Create a box & whiskers plot for data set 1 (or data set 2)

THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

Review #3: What have we learned?

- Name four measures of central tendency
- Name four measures of spread
- What are the five numbers generally used to make a box & whiskers plot?

