Gov. 1010 Statistics Basics

November 2019

Why statistics?

- As a requirement for this class, you will be have to carry out and interpret a test of statistical significance **in your final paper**
- You will likely have to answer some questions regarding basic statistical tests and their interpretations **on the exam**

Previous experience or knowledge?

- No previous statistical knowledge is assumed for this class
- No previous statistical software experience is assumed

- We will teach you everything required to complete the statistics requirement so you can choose to do so using a calculator or Excel. (If you have experience with statistical software you are, of course, allowed to use your preferred software)

Describing a quantitative variable: Mean

A good sense of where the values are centered is the mean of a variable

Mean $= \frac{\text{Sum of all data values}}{\text{Number of data values}}$

Mean $= \frac{x_1+x_2+x_3+\cdots+x_n}{n} = \frac{\sum x}{n}$ where n is the total number of observations, and x_1, x_2 represent each observation; this is simply the average

The mean of a sample is often written as $\bar{x}$ (pronounced "x-bar")
In Excel: `=AVERAGE(A2:A50)` [1] where A indicates the column and the numbers indicate row numbers.

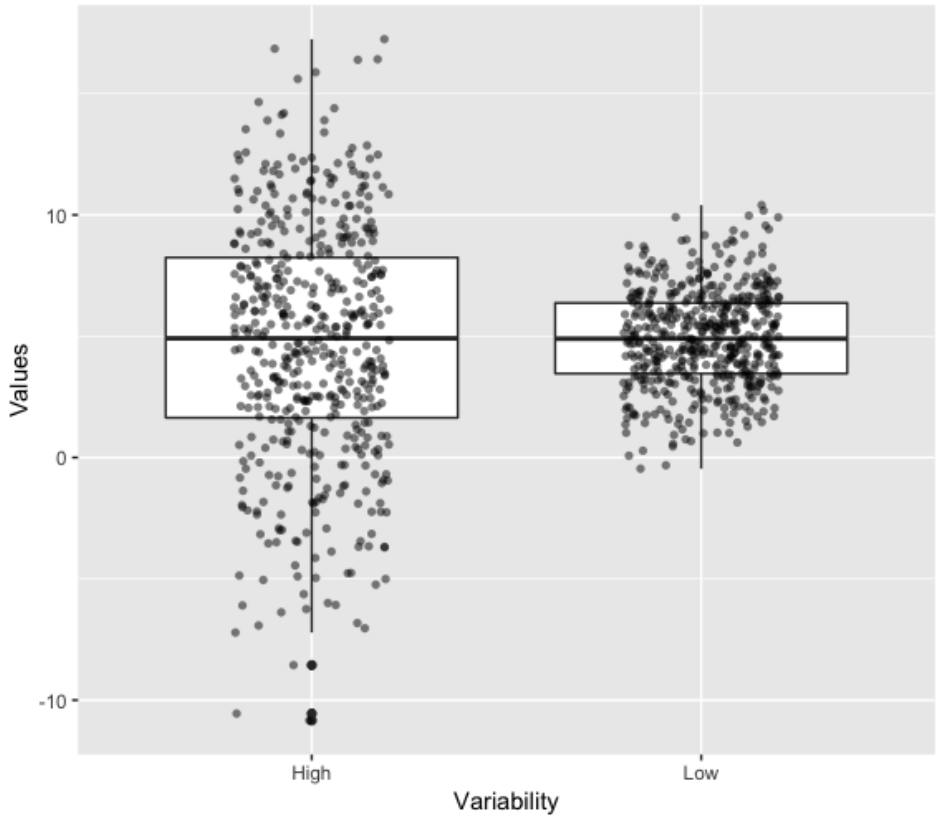
[1] The = sign must be included at the beginning of an Excel formula

Describing a quantitative variable: Spread

Another important thing to know is the spread or variability of a variable. This helps understand if the values are somewhat close together, near the mean, or spread out

Describing a quantitative variable: Spread

Another important thing to know is the spread or variability of a variable. This helps understand if the values are somewhat close together, near the mean, or spread out



Describing a quantitative variable: Spread

Measures of spread include variance and standard deviation. The square root of the variance is equal to the standard deviation. The standard deviation of a sample measures how spread out the data are from the mean. So higher variability = higher standard deviation

Standard deviation $= \sqrt{\frac{\sum (x-\bar{x})^2}{n-1}}$

In words, subtract the sample mean, $\bar{x}$ from each observation, square each of these, and then add all together. Divide this value by the number of observations n minus 1, and then take the square root.

In Excel: `=STDEV(A2:A50)`

Describing a quantitative variable: Standard error

Once you have standard deviation, you can simply calculate the standard error by dividing by the square root of the number of observations.

Standard error $= \frac{s.d.}{\sqrt{n}}$

A standard error is the standard deviation of the sample statistic. What does this mean in plain English? Well, if you were to repeat your sampling process 1000 times you would have 1000 estimates of the mean of one variable. The standard error is an estimate of the standard deviation of these 1000 estimates of the mean. This helps you know how accurate your one estimate is since is likely not feasible to take 1000 different samples.

Describing a quantitative variable: Confidence intervals

With an estimated statistic, we're using the mean in this case, and its accompanying standard error, we can calculate a confidence interval.

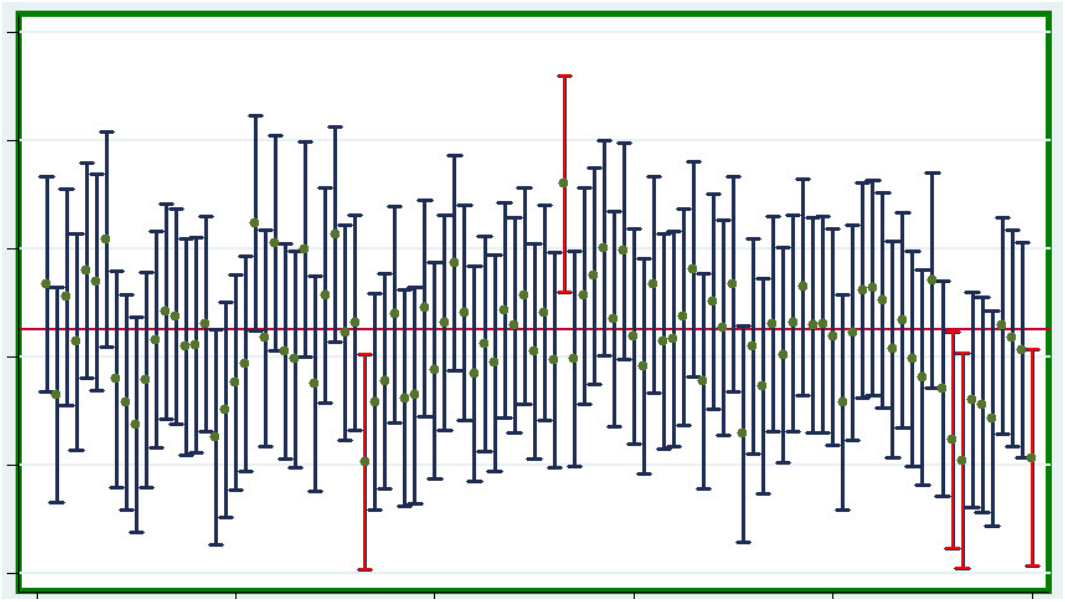
95% confidence interval $=$ estimated statistic $\pm 1.96 * S.E.$

The confidence interval of a parameter is an interval computed from sample data by a method that will capture a specified proportion of all samples. Or in English, when we calculate an X% confidence interval the way we do, the true population mean will fall within the calculated interval X% of the time.

Describing a quantitative variable: Confidence intervals

So for a 95% confidence interval, if we took random 100 samples from a population and calculated 100 different 95% confidence intervals, we would expect 95 of these to include the population mean.

Visually:



Describing data: Categorical variables

Proportion $= \frac{\text{Number in that category}}{\text{Total number}}$

Response	**Frequency**
Agree	735
Disagree	1812
I don't know	78
Total	2625

$$\text{Proportion who agree} = \frac{\text{Number who agree}}{\text{Total number}} = \frac{735}{2625} = 0.28$$