Collin Real (yhi267)

DA 6823

Kilger

Module 1: Part #2 (35 points)

**The Power of Statistics + the Levels of Measurement + the Different Classes of Variables and Determining Appropriate Statistical Technique + Basic Descriptive Measures**

**General Instructions:** In your own words, answer each of the following questions - don’t copy (e.g. cut and paste) some definition out of a book word for word. This is not a group project – you are expected to complete this module on your own. You may refer to text books, online or other sources but not your fellow classmates. If you don’t understand the question, feel free to ask the instructor in class, in office hours or in an email.

1. The first couple of questions deal with the concepts of population and sample.
   1. What is a population? (3 points)

The entire set of data used to draw conclusions in a statistical analysis. A population can be a group of people, organizations, events, etc.

* 1. What is a sample? (3 points)

A subset of a population that represents the larger group. Samples are smaller, more manageable, and an unbiased representation of the population. Samples are used when the population data is too large.

* 1. What is the objective of inferential statistics in terms of sample and population? (4 points)

The objective is to draw conclusions and make predictions (i.e. inferences) about a population using a smaller subset of sample data. Statistical models and probability theory are used to extrapolate information from samples to make these inferences.

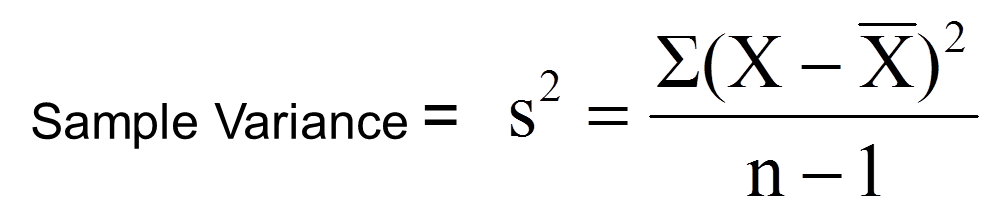
1. Name two common measures of central tendency (4 points)

Mean and median

1. Describe one situation where one measure of central tendency might be better than another measure (2 points)

If a dataset does not have a symmetrical distribution (i.e. the data is skewed), the mean can be significantly influenced by extremely small/large values, so the median would be the preferred measure of central tendency in this situation.

1. Variance and standard deviation are two of the most commonly used measures of variability. Take a look at the formula for variance below:



* 1. Looking at the variance formula, if the data points (X) are closely packed around the sample mean, what happens to the sample variance? (3 points)

The sample variance would decrease. Data points closely packed around the mean reduce the distance between a data point and the mean, resulting in less error.

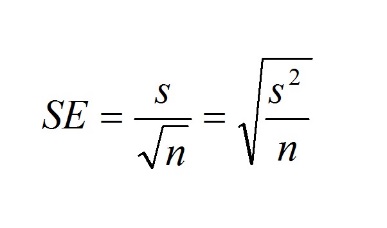
* 1. The sample standard deviation is closely related to the sample variance. How? (2 points)

Sample standard deviation = square root of sample variance

1. Many sample statistics you encounter have standard errors associated with them. Imagine that you are studying the heights of the undergraduate student body of UTSA. The total number of students is 30,000 (e.g. the population) and you randomly pull 10 samples of 100 students each from that population.
   1. Will the sample means from each of the 10 samples be the same? (2 points)

No

* 1. The standard error of the mean represents the variation in sample means that you find in different samples. The formula below is the standard error of the mean.



1. What happens to the standard error of the mean if there is a lot of variation in the data? (3 points)

More variation in the data causes the standard error of the mean to increase.

ii. What happens to the standard error of the mean as sample size increases? (3 points)

As the sample size increases, the standard error of the mean decreases because larger samples are more representative of the population.

1. The sum of squares is a statistical concept that measures variation in data that you will find in many different statistical techniques. Here is the formula for sum of squares below:



* 1. This formula should look familiar. What other statistical measures of variation are related to this? (3 points)

Standard deviation and variance are other statistical measures of variation related to the sum of squares. The variance is the average of the sum of squares, and the standard deviation is the square root of the variance.

* 1. As the data points (X) get spread out farther and farther from the sample mean, what happens to the sum of squares? (3 points)

The sum of squares becomes larger as data points spread out farther from the sample mean.