Mike Mahoney

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 available observations; Ideally the World Population.

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

It is the subset of the population or the observations we are using for the total population (ideally randomly chosen)

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

Inferential statistics attempts to predict or infer using a sample, to make a that prediction for the population.

If one wanted to predict product spending in 25-35 US Males with children, you could get a diverse and accurate sample to infer ideas/ information on that demographic population.

Alternatively you could use that same demographic as a sample (very specifically chosen sample) of the whole US or Global population.

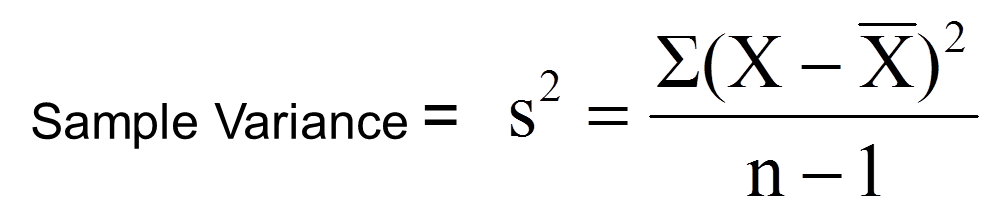
1. Name two common measures of central tendency (4 points)

Two measures include: 1) Mean: or the average of all samples (or observation being used); 2) Mode: The most common occurring value or observation.

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

When one has data that is highly skewed the median is shifted in the direction of the outliers (or skewedness), a smart data scientist would use the median or value(s) exactly in the middle

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 variance would decrease as the numerator shrinks against a constant denominator, would result in a lower value or Variance.

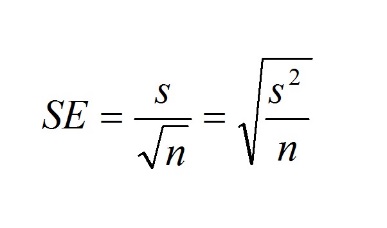
* 1. The sample standard deviation is closely related to the sample variance. How? (2 points)

The standard deviation is the square root of the variance, yes they are closely related.

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, several means could be close with in less than 1 SD, but an adequate capture of the population would have been obtained

* 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)

The standard error would be high(or increase) if the variation or s^2 was large

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

An increase of the sample size would reduce the standard error, if the variation was constant.

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)

The mean square error is similar.

* 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 increase exponentially.