What are type I and type II errors?

When you do a hypothesis test, two types of errors are possible: type I and type II. The risks of these two errors are inversely related and determined by the level of significance and the power for the test. Therefore, you should determine which error has more severe consequences for your situation before you define their risks.

No hypothesis test is 100% certain. Because the test is based on probabilities, there is always a chance of drawing an incorrect conclusion.

**Type I error**

When the null hypothesis is true and you reject it, you make a type I error. The probability of making a type I error is α, which is the level of significance you set for your hypothesis test. An α of 0.05 indicates that you are willing to accept a 5% chance that you are wrong when you reject the null hypothesis. To lower this risk, you must use a lower value for α. However, using a lower value for alpha means that you will be less likely to detect a true difference if one really exists.

**Type II error**

When the null hypothesis is false and you fail to reject it, you make a type II error. The probability of making a type II error is β, which depends on the power of the test. You can decrease your risk of committing a type II error by ensuring your test has enough power. You can do this by ensuring your sample size is large enough to detect a practical difference when one truly exists.

The probability of rejecting the null hypothesis when it is false is equal to 1–β. This value is the power of the test.

|  |  |  |
| --- | --- | --- |
|  | **Null Hypothesis** | |
| **Decision** | True | False |
| Fail to reject | Correct Decision (probability = 1 - α) | **Type II Error** - fail to reject the null when it is false (probability = β) |
| Reject | **Type I Error** - rejecting the null when it is true (probability = α) | Correct Decision (probability = 1 - β) |

Example of type I and type II error

To understand the interrelationship between type I and type II error, and to determine which error has more severe consequences for your situation, consider the following example.

A medical researcher wants to compare the effectiveness of two medications. The null and alternative hypotheses are:

* Null hypothesis (H0): μ1= μ2

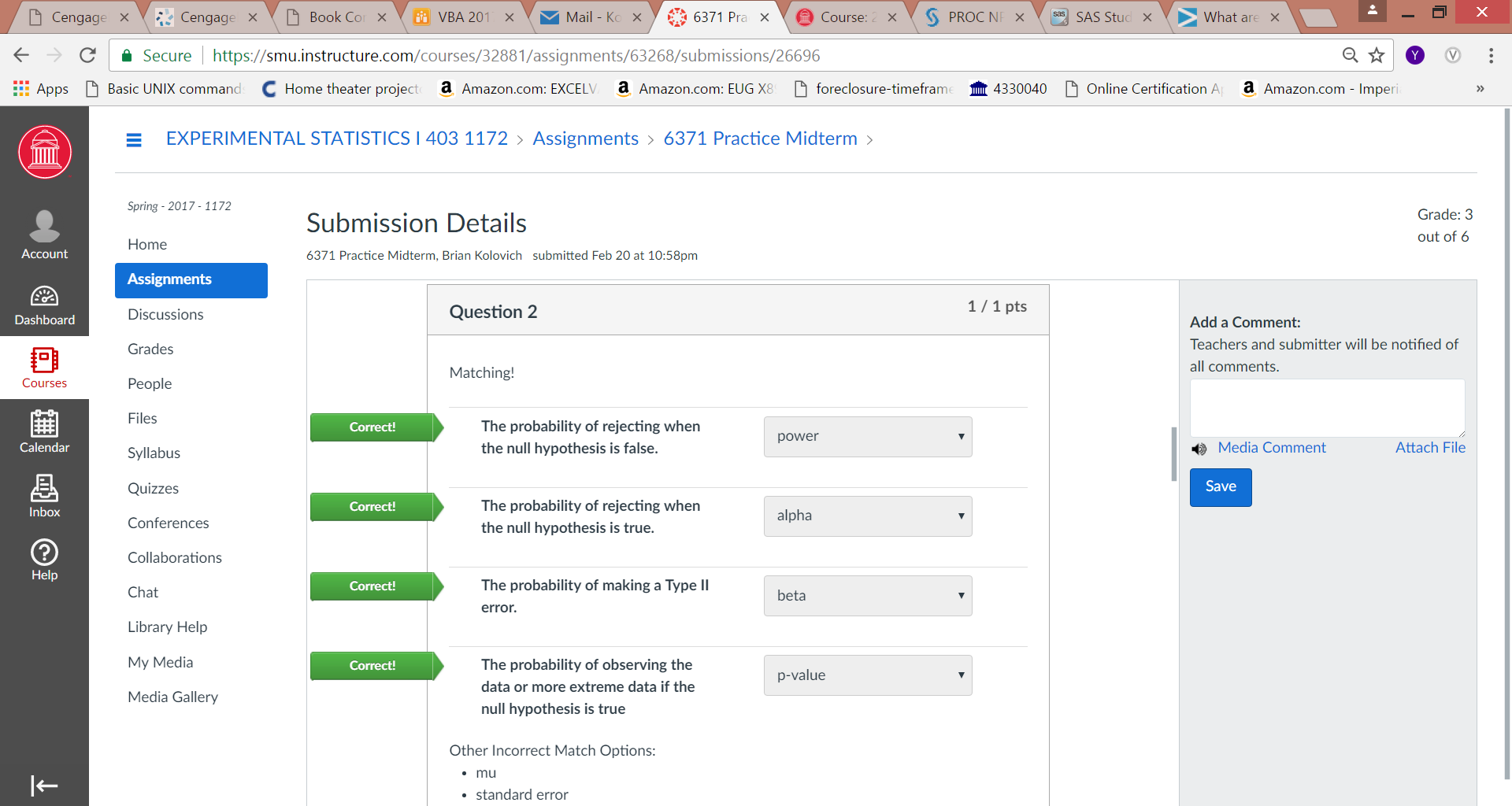
The two medications are equally effective.

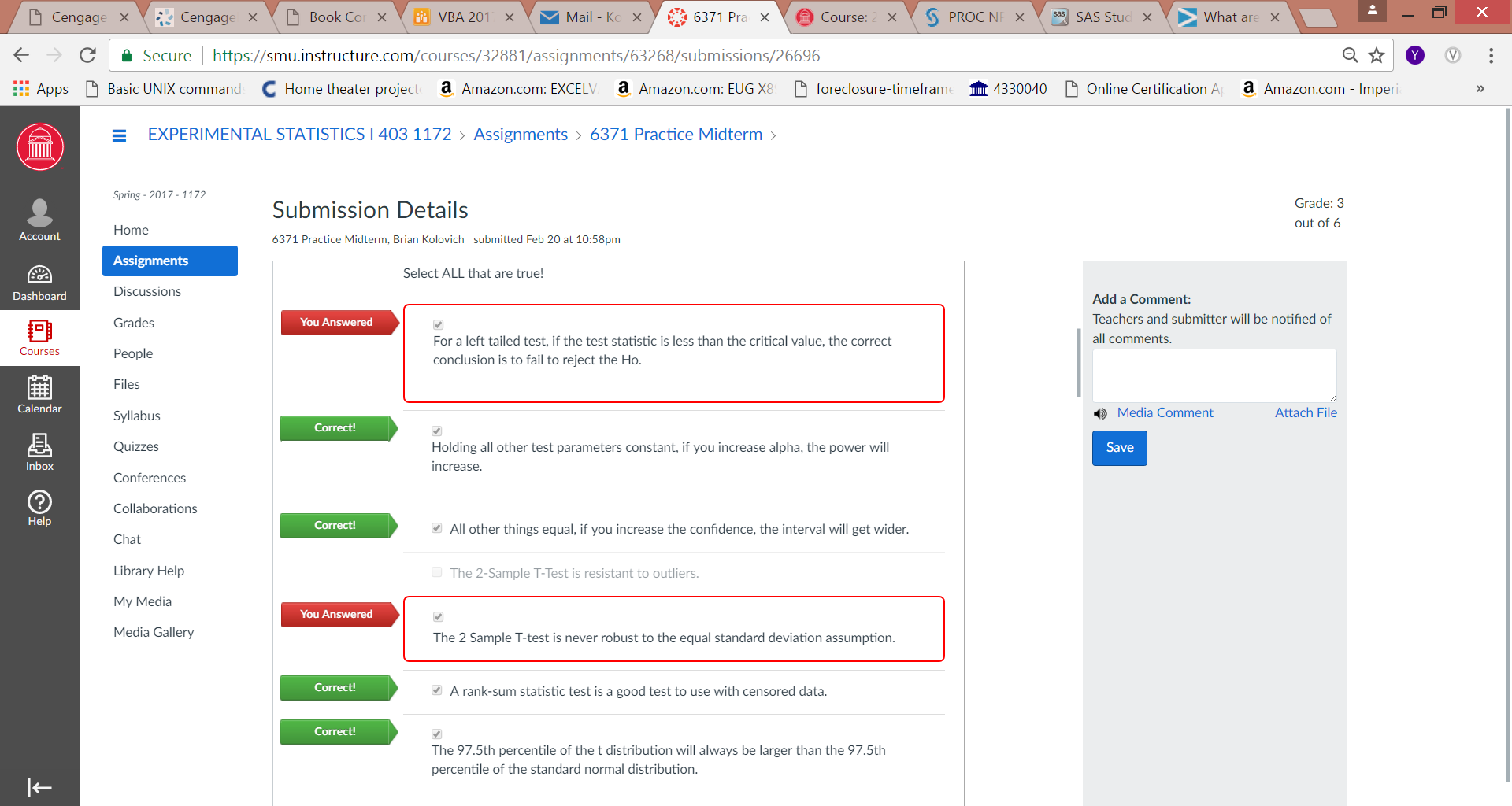
* Alternative hypothesis (H1): μ1≠ μ2

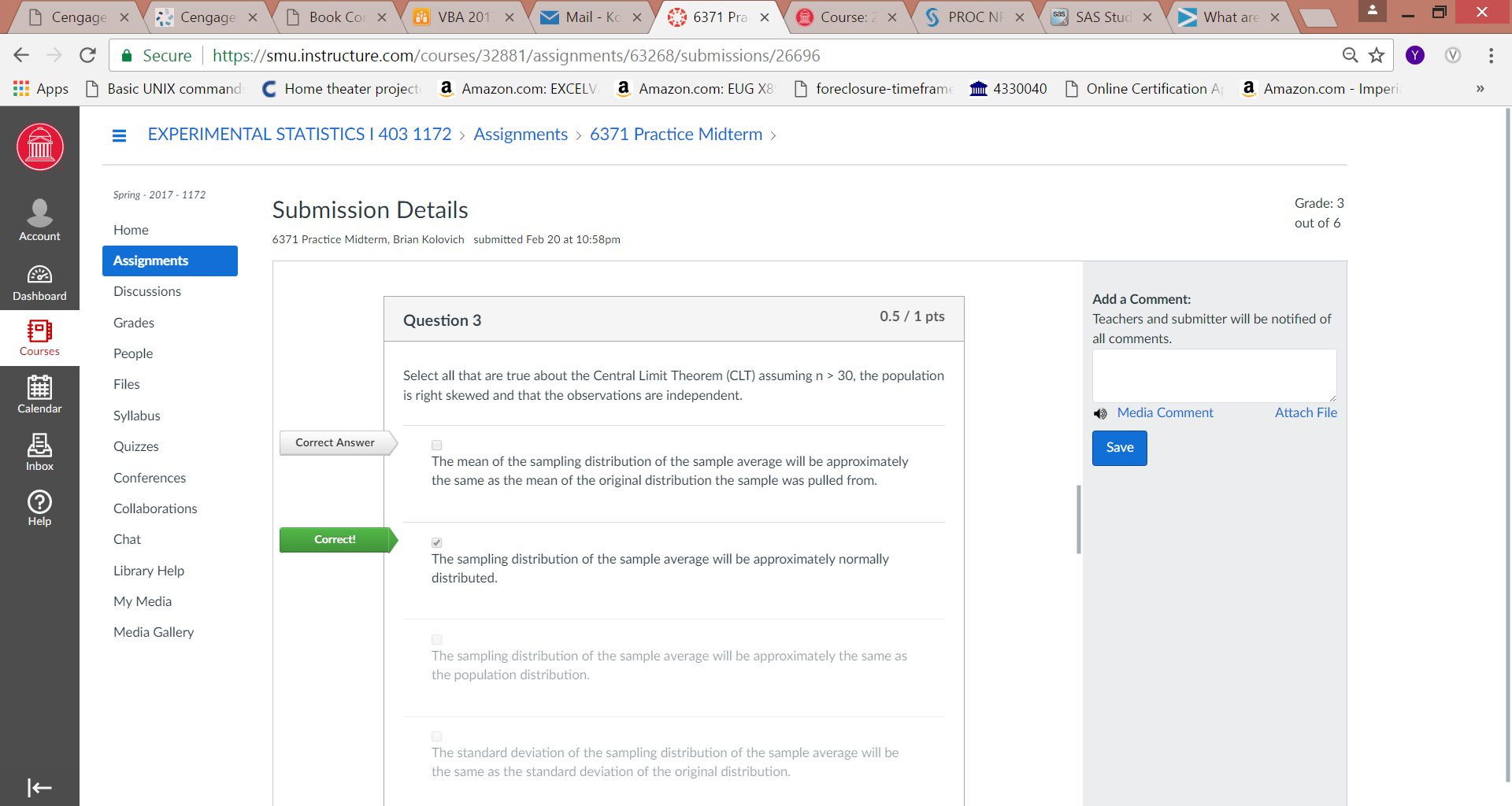
The two medications are not equally effective.

A type I error occurs if the researcher rejects the null hypothesis and concludes that the two medications are different when, in fact, they are not. If the medications have the same effectiveness, the researcher may not consider this error too severe because the patients still benefit from the same level of effectiveness regardless of which medicine they take. However, if a type II error occurs, the researcher fails to reject the null hypothesis when it should be rejected. That is, the researcher concludes that the medications are the same when, in fact, they are different. This error is potentially life-threatening if the less-effective medication is sold to the public instead of the more effective one.

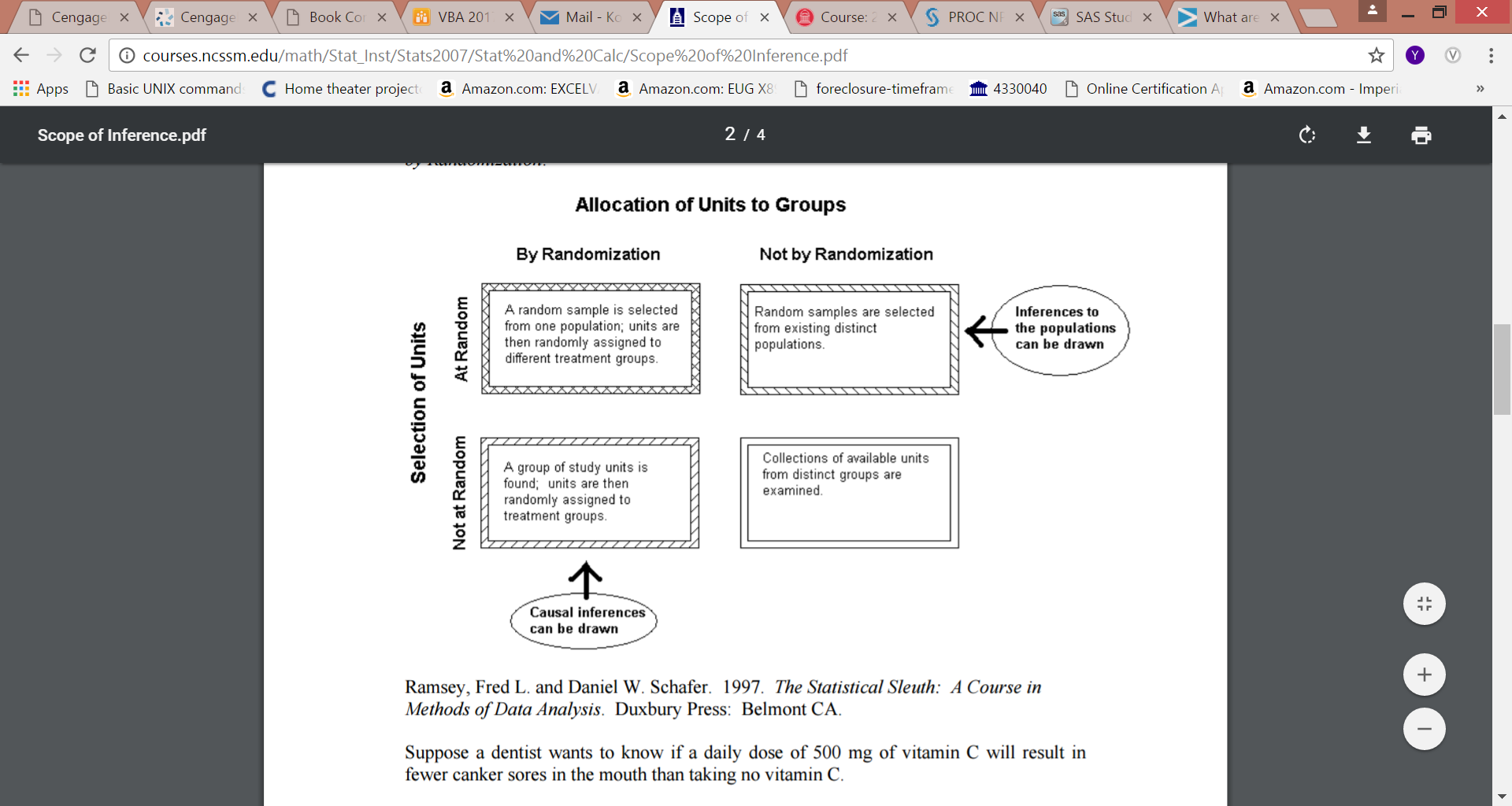
As you conduct your hypothesis tests, consider the risks of making type I and type II errors. If the consequences of making one type of error are more severe or costly than making the other type of error, then choose a level of significance and a power for the test that will reflect the relative severity of those consequences.







**SCOPE OF INFERENCE**



**Case 1) No Randomization in Selection of Units and No Randomization in Allocation of Units to Treatments** The dentist, working through the local dental society, convinces all of the dental patients in town with appointments the first two weeks in December to be subjects in an experiment. He divides them into two groups, those who take at least 500 mg of vitamin C each day and those who don't. He then asks them how often they have canker sores in their mouth and checks their patients records to see who has complained about canker sores. He compares the proportion of those who take vitamin C daily and complain of canker sores with the proportion of those who don't take vitamin C and complain of canker sores. There is a significant difference in the two proportions, with a significantly smaller proportion of those taking vitamin C having canker sores. What can we conclude?

**Case 2) No Randomization in Selection of Units but Randomization in Allocation of Units to Treatments** A dentist, working through the local dental society, convinces all of the dental patients in town with appointments the first two weeks in December to be subjects in an experiment. He randomly assigns half of them to take 500 mg of vitamin C each day and the other half to abstain from taking vitamin C for three months. At the end of this time he determines the proportion of each group that has suffered from canker sores during those three months. There is a significant difference in the two proportions, with a significantly smaller proportion of those taking vitamin C having canker sores. What can we conclude?

**Case 3) Randomization in Selection of Units but No Randomization in Allocation of Units to Treatments** The dentist, working through the local dental society, selects a random sample of dental patients in town and convinces them to be subjects in an experiment. He divides them into two groups, those who take at least 500 mg of vitamin C each day and those who don't. He then asks them how often they have canker sores in their mouth and checks their patients records to see who has complained about canker sores. He compares the proportion of those who take vitamin C daily and complain of canker sores with the proportion of those who don't take vitamin C and complain of canker sores. There is a significant difference in the two proportions, with a significantly smaller proportion of those taking vitamin C having canker sores. What can we conclude?

**Case 4) Randomization in Selection of Units and Randomization in Allocation of Units to Treatments** The dentist, working through the local dental society, selects a random sample of dental patients in town and convinces them to be subjects in an experiment. He randomly assigns half of them to take 500 mg of vitamin C each day and the other half to abstain from taking vitamin C for three months. At the end of this time he determines the proportion of each group that has suffered from canker sores during those three months. There is a significant difference in the two proportions, with a significantly smaller proportion of those taking vitamin C having canker sores. What can we conclude?

**Conclusions**

**Case 1)** Since the patients do not represent a random sample from any population, it is not possible to make any inference about this result holding for a larger population. Since the study was observational, with subjects not randomly assigned to treatments, no causal inference can be made. We just know that for these patients, those who take vitamin C have fewer canker sores than those who don't. We don't know why, and we don't know if this result would be consistent with another group.

**Case 2)** Since the patients do not represent a random sample from any population, it is not possible to make any inference that this result would hold for a larger population. However, the treatments were randomly assigned to the subjects, so (assuming other factors were controlled or randomized) the difference in proportions having canker sores can be attributed to the vitamin C. We don't know if this result would be consistent with another group, but we believe we know why, for this group, the proportions differ.

**Case 3)** Since the patients selected were a random sample of dental patients in town, we can infer that the results observed in this experiment would be consistent with results from the whole population of dental patients in this town. However, since the study was observational, with subjects not being randomly assigned to treatments, no causal inference can be made. We believe that for the population of dental patients in this town, that those taking vitamin C have fewer canker sores than those who didn't. We don't know if it is the vitamin C that causes this reduction or some other confounding variable. We cannot conclude that for the general population, those taking vitamin C have fewer canker sores, since the sample was only of dental patients. To the extent that the dental patients in this town are representative of dental patients in general, we can infer that dental patients who take vitamin C tend to have fewer canker sores than those who don't.

**Case 4)** Since the patients selected were a random sample of dental patients in town, we can infer than the results observed in this experiment would be consistent with results from the whole population of dental patients in this town. Moreover, the treatments were randomly assigned to the subjects, so (assuming other factors were controlled or randomized) the difference in proportions having canker sores can be attributed to the vitamin C. We believe that for the population of dental patients in this town, that those taking vitamin C have fewer canker sores than those who don't. Also, we believe that the reduction in canker sores is a consequence of taking the vitamin C. We cannot conclude that for the general population, those taking vitamin C have fewer canker sores, since the sample was only of dental patients. To the extent that the dental patients in this town are representative of dental patients in general, we can infer that dental patients who take vitamin C tend to have fewer canker sores than those who don't, as a result of taking the vitamin C.

**DESCRIPTIVE STATISTICS**

writing score

-------------------------------------------------------------

Percentiles Smallest**i**

1%**e** 31 31

5% 35.5 31

10% 39 31 Obs**b** 200

25%**f** 45.5 31 Sum of Wgt.**k** 200

50%**g** 54 Mean**c** 52.775

Largest**j** Std. Dev.**d** 9.478586

75%**h** 60 67

90% 65 67 Variance**l** 89.84359

95% 65 67 Skewness**m** -.4784158

99% 67 67 Kurtosis**n** 2.238527

e.  **1%** - This is the first percentile.  Percentiles are calculated by ordering the values of a variable from lowest to highest, and then finding the value that corresponds to whatever percent you are interested in, in this case, 1%.  Hence, 1% of the values of the variable **write** are equal to or less than 31.

f.  **25%** - This is the 25th percentile, also known as the first quartile.

g.  **50%** - This is the 50th percentile, also known as the median.  If you order the values of the variable from lowest to highest, the median would be the value exactly in the middle.  In other words, half of the values would be below the median, and half would be above.  This is a good measure of central tendency if the variable has outliers.

h.  **75%** - This is the 75th percentile, also known as the third quartile.

i.  **Smallest** - This is a list of the four smallest values of the variable.  In this example, the four smallest values are all 31.

j.  **Largest** - This is a list of the four largest values of the variable.  In this example, the four largest values are all 67.

b.  **Obs** - This column tells you the number of observations (or cases) that were valid (i.e., not missing) for that variable.  If you had 200 observations in your data set, but you had 10 missing values for the variable **female,** then the number in this column would be 190.

k.  **Sum of Wgt**. - This is the sum of the weights.  In Stata, you can use different kinds of weights on your data.  By default, each case (i.e., subject) is given a weight of 1.  When this default is used, the sum of the weights will equal the number of observations.

c.  **Mean** - This is the arithmetic mean across the observations. It is the most widely used measure of central tendency. It is commonly called the average. The mean is sensitive to extremely large or small values.

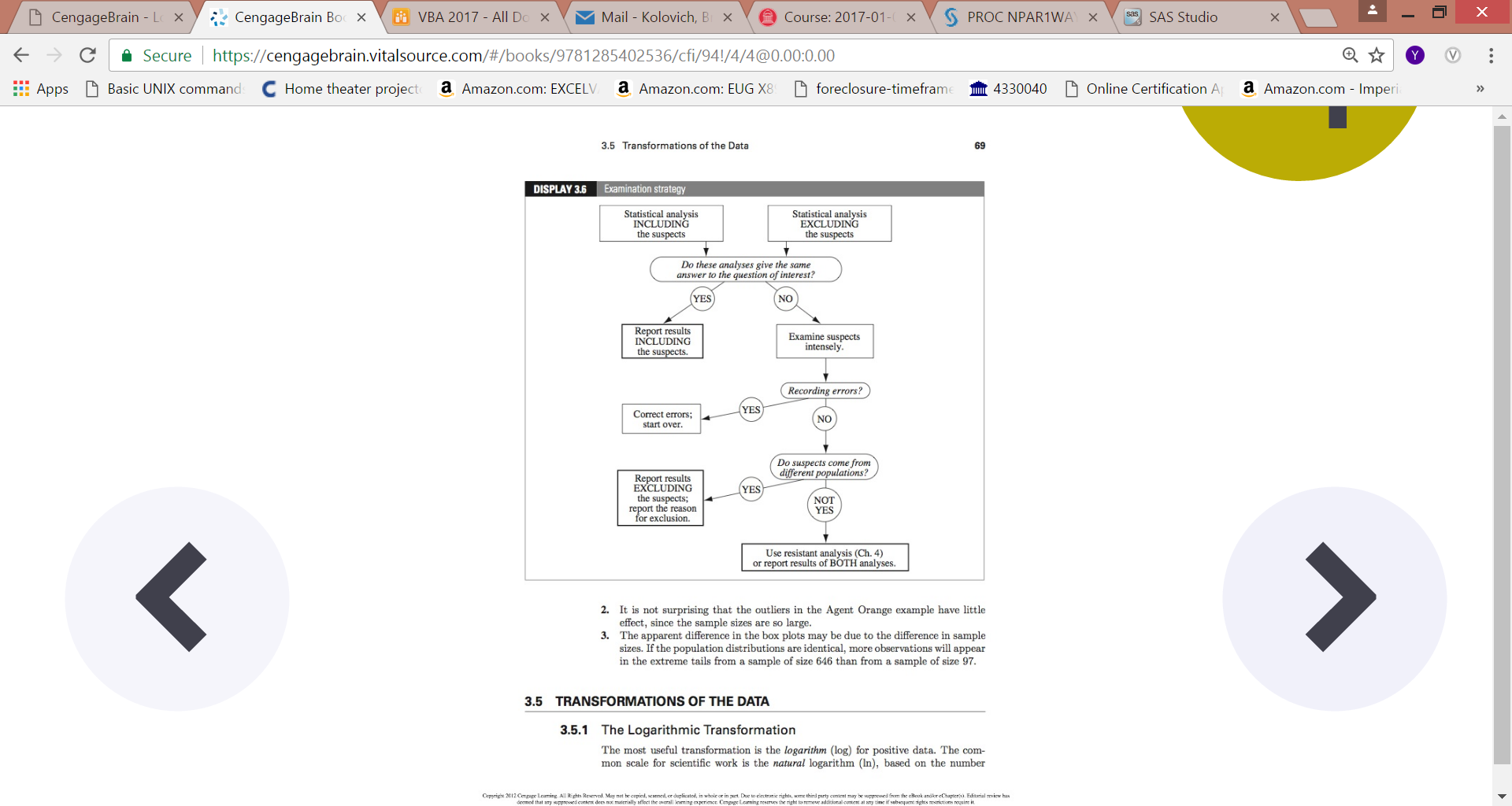
d.  **Std. Dev.** - This is the standard deviation of the variable.  This gives information regarding the spread of the distribution of the variable.

l.  **Variance** - This is the standard deviation squared (i.e., raised to the second power).  It is also a measure of spread of the distribution.

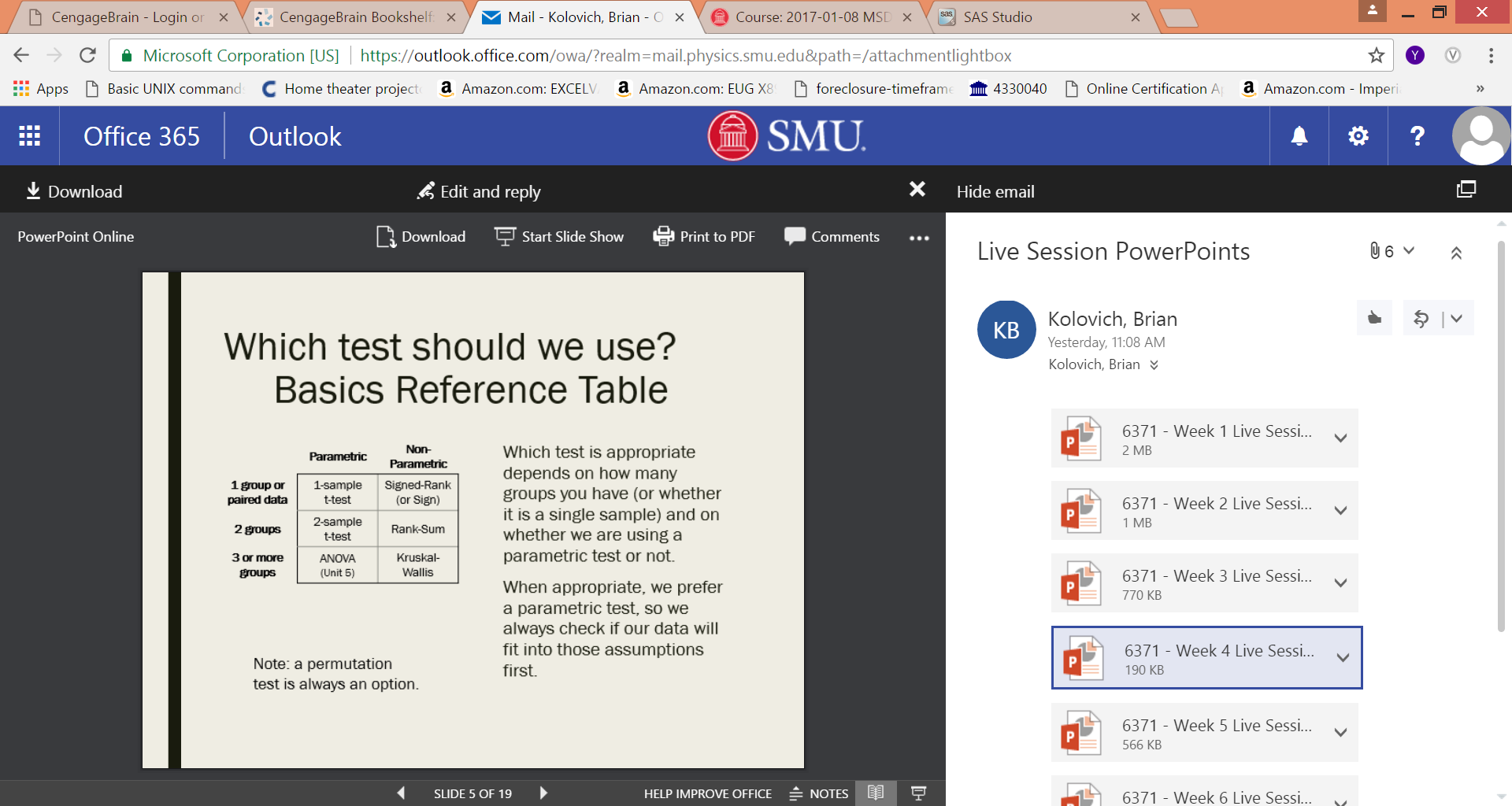
m.  **Skewness** - Skewness measures the degree and direction of asymmetry.  A symmetric distribution such as a normal distribution has a skewness of 0, and a distribution that is skewed to the left, e.g., when the mean is less than the median, has a negative skewness.

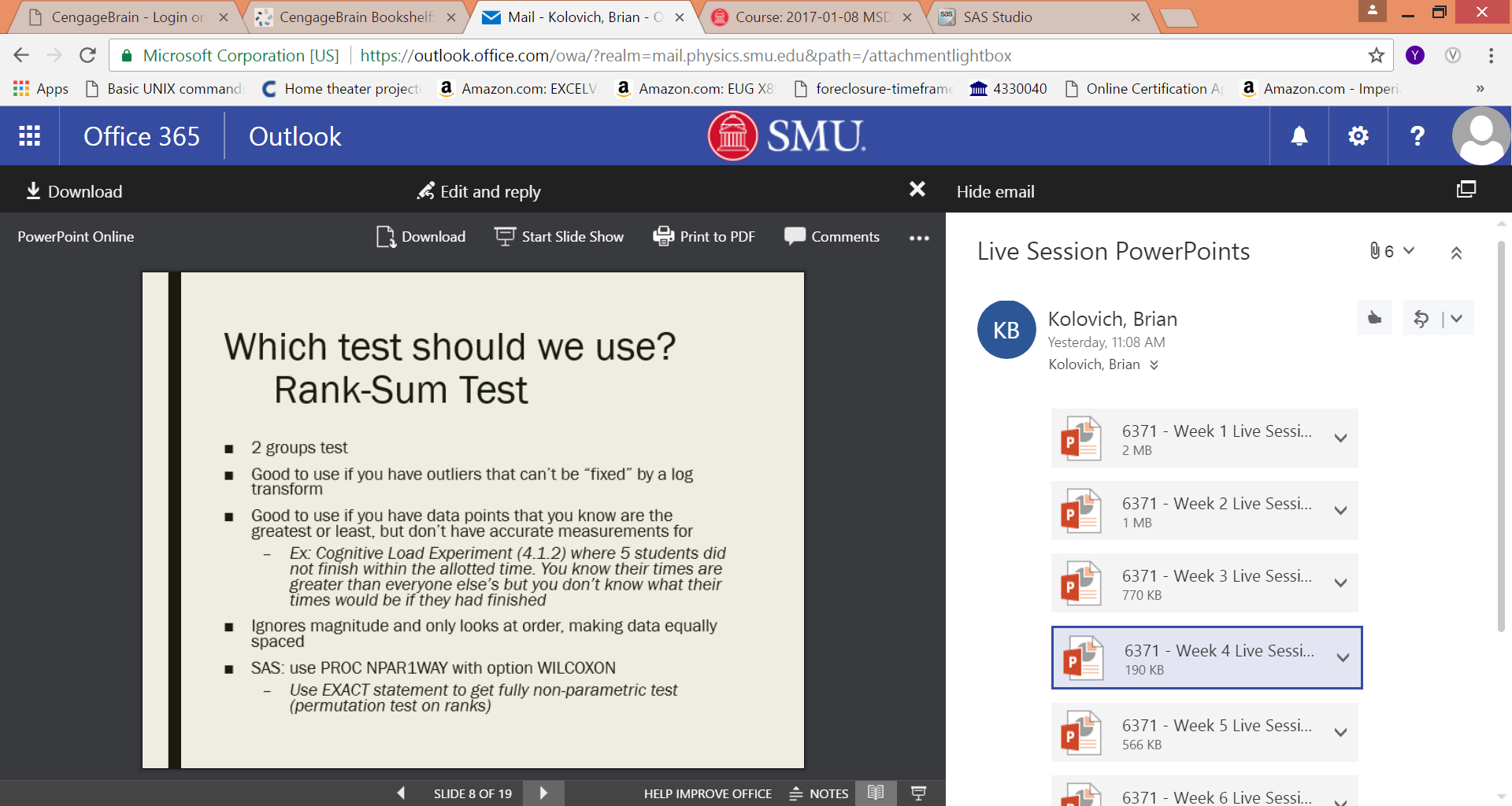
n.  **Kurtosis** - Kurtosis is a measure of the heaviness of the tails of a distribution. A normal distribution has a kurtosis of 3. Heavy tailed distributions will have kurtosis greater than 3 and light tailed distributions will have kurtosis less than 3.

**OUTLIERS**



**Which Test to Use?**





**Critical Values Code in SAS**

**SAS PROGRAM TO FIND CRITICAL VALUES**

data findt; \* 'CRITICAL VALUES'; \* These functions find the critical value for a specified probability 'p'; \*

Nvalue=PROBIT (0.95);

Tvalue=TINV (0.95,20);

Chivalue=CINV (0.95,20);

Fvalue=FINV (0.95,15,14);

\* 'PROBABILITY'; \* These functions return the pb that an observation is less or equal to x;

Nprob=PROBNORM (1.96);

Tprob=PROBT (2.086,20);

Chiprob=PROBCHI (34.17,20);

Fprob=PROBF (2.94932,15,14);

Given a specified value for *α* :

1. For a two-sided test, find the column corresponding to 1-*α*/2 and reject the null hypothesis if the absolute value of the test statistic is greater than the value of *t*1-*α*/2,*ν* in the table below.
2. For an upper, one-sided test, find the column corresponding to 1-*α* and reject the null hypothesis if the test statistic is greater than the table value.
3. For a lower, one-sided test, find the column corresponding to 1-*α* and reject the null hypothesis if the test statistic is less than the negative of the table value.