Mahoney, Mike

DA 6823

Kilger

Module 2: Part #2 (55 points)

**Standard Error of the Estimate + Confidence Intervals + the Logic of Hypothesis Testing + Type 1 and Type II errors**

**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. Explain in your own words what a type I error is (4 points)

When you assert something that is not there, or a false positive.

1. Explain in your own words what a type II error is (4 points)

When you deny something that exist, or false negative.

1. Imagine that you are a cancer researcher who has developed a new test for cancer. Think about what a type I and type II error means for this kind of test. Argue for what you think is the most egregious error – a type I or type II error in this case. (Hint – you can logically argue for either case, just explain why). (6 points)

I believe Type 2 error because you are missing cancer, any time lost in detection could result in the cancer advancing in stages, which could result in a lower survival rate. With a type 1 error, at least they could get a second opinion or use another method, where in the converse they most likely would not follow up.

1. Explain in your own words what the power of a statistical test means (4 points)

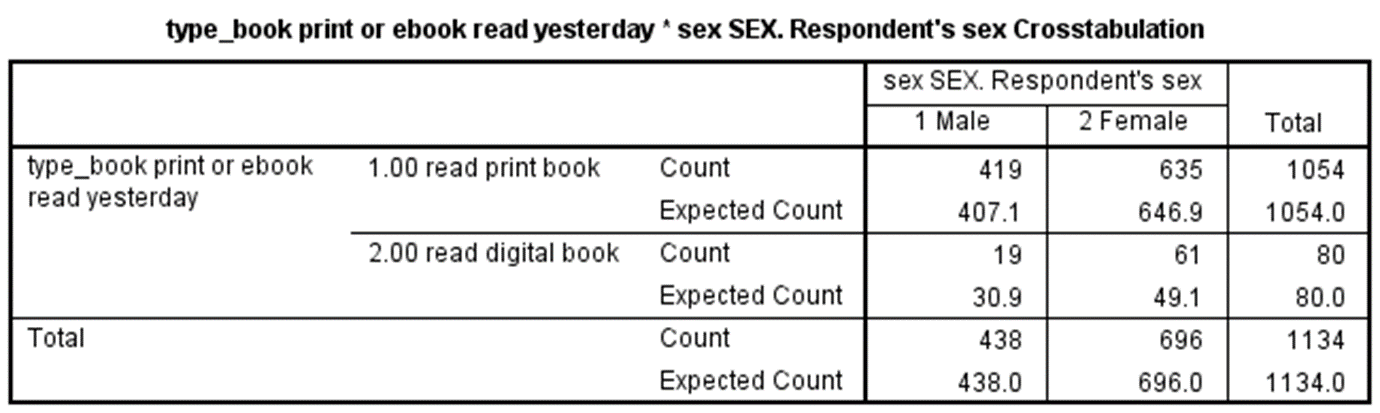
A statistical means test would tell you if the two samples are related, could have come from the same source (population) or are equal. If two means are similar they could have been sampled from the same population.

1. Name two things that can affect the power of a statistical test (4 points)

1)Skewed data: will give misinformation, moving the mean away from its true location. Outliers in this case should be observed and removed; 2) Sample Size: A low sample size will give inaccuracies, it should be larger than 35.

1. Here are the six steps of hypothesis testing:
   * 1. State the null (Hnull) and alternative (Halt) hypotheses
     2. State the assumptions of the test
     3. Determine the critical value for the test statistic
     4. Calculate the value of the test statistic from the data
     5. Compare the calculated and critical values for the test statistic
     6. Apply the decision rule and interpret the result of the test

We will use a simple chi-square test as our example in this module. Here is the data that examines if there is a relationship between gender and format of book read:



The questions on the next page take you through each step applying the chi-square test to this data. Use a stats book and/or the Internet to help you with this but write your answers in your own words, not copy and paste.

1. State the null and alternative hypotheses for this test. (4 points)

The means in gender are equal

The means in type of book are equal

1. State at least one assumption for this test. (3 points)

The means in gender are not equal

The means in type of book are not equal

1. Determine the critical value of chi-square that your data will have to exceed in order to reject the null hypothesis. This involves calculating the degrees of freedom for our data as well as looking up the critical value in a chi-square table. Show your work for calculation degrees of freedom. (6 points)
2. Write out the chi-square formula and then using the data in the table provided above, calculate the chi-square value from the data. Show your work. (10 points)

Chi^2 = Sum\_of ((Observed – Expected)^2 )/Expected

(419-407.1)^2/ 407.1 == 141.61/407.1 == 0.348

(635-646.9)^2/646.9 == 141.61/646.9 == 0.219

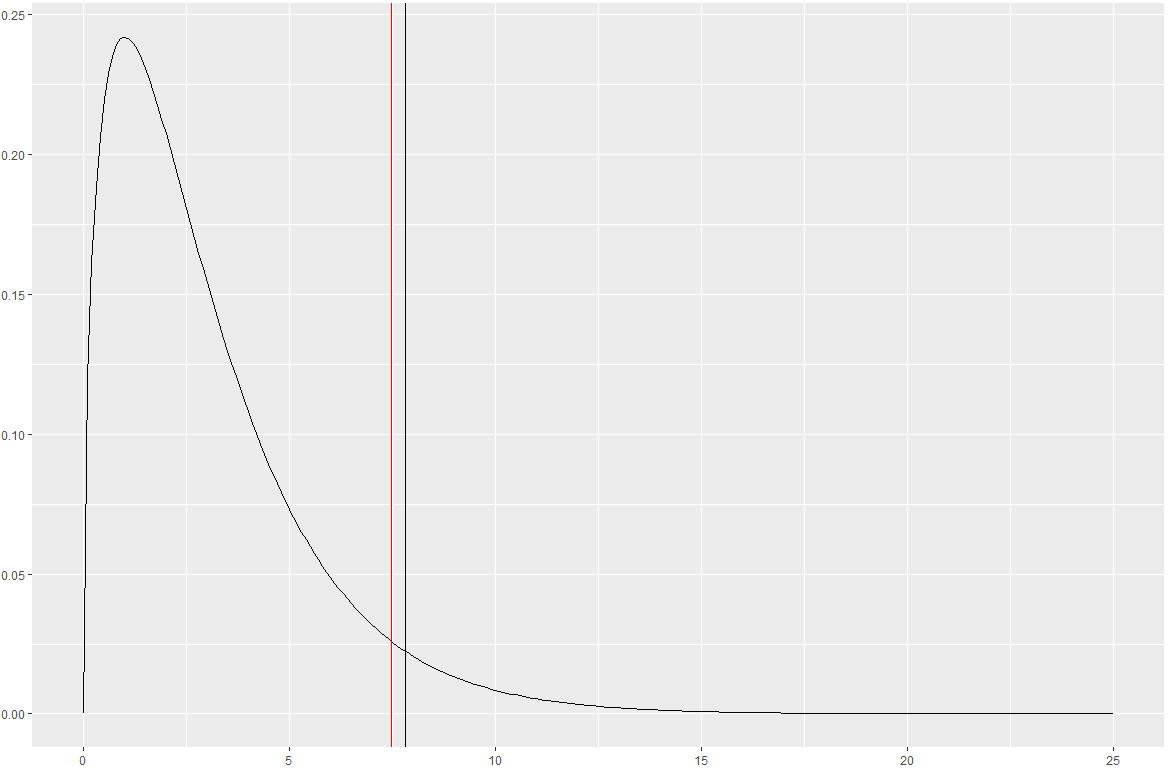
(19-30)^2/30 == 121/30 == 4.033

(61-49.1)^2/49.1 == 141.61/49.1 == 2.884

0.348 + 0.219 + 4.033 + 2.884 == 7.484 >> 7.48

1. Compare the chi-square critical value and the chi-square value calculated from the data and draw a rough sketch of a chi square curve and place those two values on the curve. (5 points)

Df = 3 , alpha = 0.05 , critical value = 7.81



The vertical black line represents the critical value, and the area to the right of the black line represents the critical area. Our Chi-Squared is denoted by the vertical red line.

1. Apply the decision rule for the chi-square test and interpret the result of your analysis. (5 points)

We fail to reject the null hypothesis because our Chi-Squared does not fall into the critical region. Therefore there is no difference between the observed and expected values.

Appendix >>

chi.values <- seq(0,25,.1)

critic\_area <- chi.values[76:251]

chi\_sq <- ggplot(NULL, aes(x=chi.values)) +

geom\_line(aes(y=dchisq(chi.values,3))) +

geom\_vline(xintercept = 7.48, color = "red") +

geom\_vline(xintercept = 7.81, color = "black")

chi\_sq