

Third Data Assignment

This assignment will be due by midnight on Wednesday, October 27. Please attach a Word-compatible document to an e-mail to me with subject heading, "Data 3: YourLastName"

Part I. Methodology

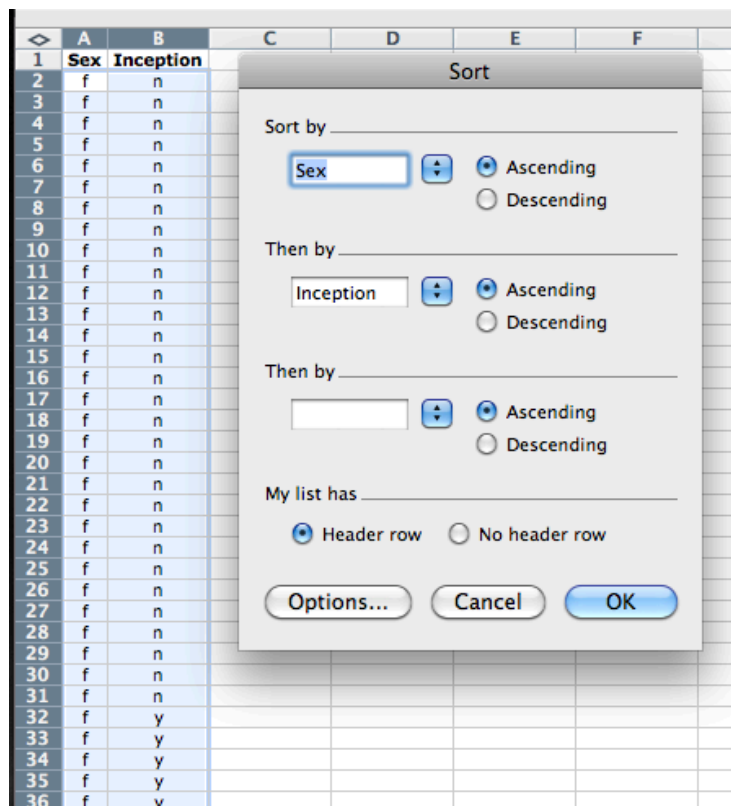
Choose a categorical question for which you are interested in seeing if there is a difference between the responses of men and the responses of women. Briefly explain your choice. Open the Big Survey spreadsheet and delete all columns except for the Question on gender and the question that you chose in above. Choose a particular possible response for your question for which you will analyze the proportion of all SU students who would answer _____ to the question you chose. For example, if you chose the political view question from the Small Survey, you might choose the "moderate" response and so your analysis would be of p = the proportion of SU students who would answer "moderate" to this question. For purposes of accuracy, you need to study a fairly common response to the question!

Part II. Create a confidence interval.

Construct a 95% confidence interval for p , the proportion of the population of all SU students who would have answered the question you chose with the category that you chose in Part I. To show your work, show p -hat, standard error, margin of error, and finally, the confidence interval given in a grammatically correct sentence that you could put in the newspaper (i.e., no jargon or referring to variables).

Part III. Two proportion hypothesis test

Compare the answer to your question by doing a hypothesis test for males vs. females at the 5% significance level. The first step for this should be a contingency table as shown in the sample solution. To make this easier to count, you might consider sorting your data. For example, if the two questions' data are in columns A and B, then you should select all of the data, choose "Custom Sort" under the Data tab, and the sort the data first on Column A and then on Column B. The screen shot on the right shows what this looks like on my Mac – the dialog box on a Windows machine is slightly different but uses the same terminology. Use Google to find an online tutorial on sorting data in Excel if



you need more direction.

The hypothesis test can be performed using the TI calculator's 2PropZTest function, which will require as input the numbers from your contingency table. Give exactly the input you put in the calculator and write down exactly the output that you see, and then make a conclusion about whether the evidence supports the claim, " p_1 is not equal to p_2 ."

Part IV. Discussion

For your results in Parts III and IV to be valid, it must be the case that our sample was randomly chosen from the population. Obviously it was not, but your conclusions might still be valid if you cannot think of hidden variables that might be producing a biased result. Write at least one sentence discussing the validity of your conclusions in Parts III and IV.

Grading Rubric (20 points total)

Methodology (2 points) _____

Confidence Interval (6 points) _____

Contingency Table (4 points) _____

Hypothesis Test (5 points) _____

Discussion (3 points) _____