

Stats First Quarter Project

In this project, you'll run a χ^2 -independence test on a contingency table of your choosing (see your notes and/or pages 523-526 in the book for more info). Choose something which is important to you... music, social justice, football, physics, the environment, etc. First, you need to gather your data. You can use pre-existing data from the web, but you must cite your sources. You will be testing to see whether two variables are independent. For example, testing whether or not higher income is independent of your life expectancy. The contingency table will be a two way frequency table in which the rows represent different values of one variable (for example, intervals of incomes: 0-\$20,000, \$20,000-\$40,000, etc) and the columns represent different values of the other variable (for example, intervals of life expectancy 50-60 years, 60-70 years, 70-80 years, 80-90 years, etc). The entries in the table then are the frequencies observed from the data (for example, the entry in the \$20,000-\$40,000 row and the 60-70 years column would be the number individuals in the study which had income between 20 and 40 thousand, and died between the ages of 60 and 70).

Once you have your contingency table, you'll run a χ^2 -test as follows: (1) Enter in the table to your TI calculator as a matrix (as we did in class), (2) run the χ^2 -independence test as found in the TESTS tab of the STAT menu (make sure that the observed matrix corresponds to the one you entered in the previous part), (3) Look at the p -value and decide whether or not we should believe the variables are independent with 95% confidence, (4) find the means and standard deviations of each row and each column.

OK, the descriptive stats of this project are done. Now we interpret what this tells us. First of all, make a claim. For example, "In the United States, Income is not independent of life expectancy". Identify population and sample: the population is the group we're making the claim about, e.g., Americans, while the sample is the group where the table came from, e.g., a survey found on some website or in a magazine. Interpret the p -value in words. List all possible confounding variables (including how the collection of the data might have affected the results [watch out for convenience sampling, self-selection, etc]). In the example above, just because income and life expectancy are dependent doesn't mean that having a low income implies a low life expectancy, it just says that there is some relationship, but they could both be, say, consequences of education. The point is that things can be related without being cause and effect. Interpret why certain columns have larger or smaller means/standard deviations than others.

Write up your results in essay form. Write complete sentences and use your own words. You had a question or interest; did the test support your claim or not? Maybe you can't be 95% confident but you could be 90% confident (this goes back to the p -value). This is your chance to be creative and give your grade a big boost at the same time. If you do well on this project, it will certainly make-up for doing poorly on the quizzes.