

Exercises VII - Hypothesis Testing for Categorical Data

1. Load the births data set from the OpenIntro package. Using the `table()` command, make a table of the numbers of premature babies and normal babies by whether the mother smokes. Do a proportion test with **`prop.test()`** to see if the proportion of premature babies differs between smokers and non-smokers. Explain.

2. Looking at the same data set, you notice a variable for the number of doctor visits a mother has during pregnancy. You wonder if this affects prematurity (better pre-natal care leads to fewer premature births). Use the `cut` command to cut visits (a numeric variable) into a factor with three levels. Then test whether the proportion of premature births varies across this factor, using an appropriate hypothesis test. Explain. **You can use `prop.test()` or `chisq.test()`.**

3. Load the email data set from OpenIntro. Examine it and design a proportion test for a table you create of interest regarding spam versus not spam and one other variable. Explain. Illustrate your results with a mosaic plot. **You can use `prop.test()` or `chisq.test()`.**

4. A person claims to be able to throw a die and make a 1 come up more often than chance ($1/6$). The die is thrown 100 times and a one comes up 18 times. Is this proportion significantly different from $1/6$? Explain. **Use `binom.test()`.**

5. A recent experiment investigated the relationship between smoking and urinary incontinence. Of 322 subjects in the study who were incontinent, 113 were smokers, 51 were former smokers, and 158 had never smoked. Of 284 control subjects who were not incontinent, 68 were smokers, 23 were former smokers, and 193 had never smoked. Do a significance test to see if there is a relationship between smoking and incontinence. **Use `chisq.test()`.**

6. At a school pep rally, a group of sophomore students organized a free raffle for prizes. They claim that they put the names of all of the students in the school in the basket and that they randomly drew 36 names out of this basket. Of the prize winners, 6 were freshmen, 14 were sophomores, 9 were juniors, and 7 were seniors. The results do not seem that random to you. You think it is a little fishy that sophomores organized the raffle and also won the most prizes. Your school is composed of 30% freshmen, 25% sophomores, 25% juniors, and 20% seniors. (a) What are the expected frequencies of winners from each class? (b) Conduct a significance test to determine whether the winners of the prizes were distributed throughout the classes as would be expected based on the percentage of students in each group. Report your Chi Square and p values. (c) What do you conclude? **Use `chisq.test()`.**