Stat 439: Homework 2

Due Thur 2/10/22 by 5pm in D2L

Your name here

Instructions

You are strongly encouraged to use R Markdown to complete your Homework assignments, starting with this file as a template. If using R for your answer, include all relevant R code and output. Your homework submission in D2L should be a single pdf file (do not turn in the .Rmd file).

Part I: Book Exercises

Agresti Exercises 2.8, 2.30ab

Part II

1.

A study was conducted to test the claim that the intake of one gram of vitamin C per day substantially reduces the frequency of colds. Volunteers were sought from a variety of occupations and age groups. Each subject was randomly assigned to either a 250 mg Vitamin C tablet or a placebo tablet per day for one month. At the end of the month, researchers recorded whether or not each subject developed a cold. Of the 407 subjects assigned to vitamin C, 302 developed a cold; of the 411 subjects assigned to the placebo, 335 developed a cold.

a. Assume the sample size was fixed by the researchers. What type of sampling was used in this study: "binomial" or "multinomial." Justify your answer.
b. Is this a randomized experiment or an observational study? Justify your answer.
c. Enter these data into R as a 2x2 table and display the output. The response variable should be entered as the columns, with a "success" as column 1; the explanatory variable should be entered as rows with the base line group as column 1. (This will make the use of the `OddsRatio` function easier.)
d. Create an appropriate plot that displays the relationship between whether the subject took vitamin C or the placebo and whether the subject developed a cold or not. Write one or two sentences describing what the plot tells you.
e. Calculate the estimated relative risk of developing a cold for the vitamin C group compared to the placebo. Fill in the blank in the following interpretation: "The risk of developing a cold for the vitamin C group was _____% less than the risk of developing a cold in the placebo group."
f. Calculate the estimated relative risk of developing a cold for the placebo group compared to the vitamin C group. Fill in the blank in the following interpretation: "The risk of developing a cold for the placebo group was _____% less than the risk of developing a cold in the vitamin C group." Explain mathematically why the value in the blank does not equal the value in the blank from part c.
g. Calculate the estimated odds ratio with the vitamin C group in the numerator. Find and interpret an approximate 95% confidence interval for this odds ratio. (Calculations should be done by hand, though you may check your confidence interval calculation using the `oddsRatio` function.)

2.

Psychologists performed a randomized experiment on 50 male bank supervisors attending a management institute, to investigate biases against women in personnel decisions. The supervisors were asked to make a decision on whether to promote a hypothetical applicant based on a personnel file. For 26 of them, the application file described a female candidate; for the others it described a male. The files were identical in all other respects. Results on the promotion decisions for the two groups are shown below.

```
bank_table <- as.table(
                  matrix(c(21, 3, 14, 12),
                      nrow = 2, ncol = 2,
                      byrow = TRUE,
                      dimnames = list(
                        applicant = c("male", "female"),
                        promoted = c("yes","no")
                        )
                      )
              )
bank_table
```

```
##          promoted
## applicant yes no
##    male    21  3
##    female  14 12
```

a. Carry out a randomization test for these data using the odds ratio for the statistic. Include: (i) definition of the parameter, (ii) hypotheses, (iii) value of the estimated odds ratio, (iv) p-value, and (v) conclusion. Your conclusion should also include the scope of inference for this study; that is, to which population can you generalize? and can you conclude causation?

b. Find the p-value for the Fisher's Exact Test of these data using only the `dhyper` and/or `phyper` functions in R (though you may check your work with the `fisher.test` function). Does the p-value from this exact test match the p-value for the randomization test?

Part III: Cite Sources

Write the sources you used to complete this assignment at the end of your homework submission, adhering to the "Guidance on Citing Sources" bullet points in the collaboration policy section on our course syllabus.