

Assignment 6

Released: 10/10/2020

Due: 10/18/2020

The objective of this homework is to identify appropriate statistical methods for given problems, apply and interpret statistical techniques, evaluate applications of statistical techniques and use a statistical application package.

Evaluation: Since this is practice, actually having the right final answer will not carry much weight in evaluating your work. You will be evaluated primarily on your process and on making a good faith effort to apply course material and interpret your results. Neatness and professionalism in preparing your submission also matter.

Data

Use the provided file “supplement.csv” to work with in Stata. The file contains data for a (fictional) block randomized trial aimed at testing the effectiveness of a dietary supplement in reducing the incidence and severity of colds. Thirty male and thirty females were randomly assigned to take the supplement or a placebo for 2 months. The data records whether each subject experienced a cold and for how many days they experienced symptoms.

Variable Definitions

Subject: Individual subject identifier.

Female: 1 if the subject is female, 0 otherwise.

Supplement: 1 if the subject was given the supplement, 0 otherwise.

Cold: 1 if the subject experienced a cold over the trial, 0 otherwise.

Days: Number of days the subject experienced cold symptoms over the trial.

Weight: Subject’s weight in pounds.

Use the above data to answer the following in Stata. Create a do file and a log file of the run to attach in the appendix.

Summary Statistics

1. Calculate summary statistics for the proportion that caught a cold and for the number of days with a cold among those that caught a cold. Do this overall, separately by supplement use, separately by gender, and then separately by gender and supplement use.

Hint 1: If you just do this with repeated applications of the summarize command, which is fine, you will have to use “if var==X”, and sometimes “var==X and var==X” where “var” is the variable(s) you need to isolate certain cases of and X is the value(s) you need to isolate.

Hint 2: While not required, if you look into the command tabulate (help tabulate) with the option summarize used for two-way tabulations, there is a much more elegant way to do this than brute forcing it with repeated summarize commands, and it creates a very informative table for each variable summarized (cold and days) overall and by gender.

Quantitative Methods

2. Using the command prtest, test the association between getting a cold and both taking the supplement and gender. (Hint: Type “help prtest” to learn the syntax, or use the menus)

- a. Test whether the proportion that get a cold is influenced by the supplement.
- b. Test whether the proportion that get a cold is influenced by the supplement for males and females separately.
- c. Test whether the proportion that get a cold is influenced by gender.
- d. Test whether the proportion that get a cold is influenced by gender for those taking the supplement and those not taking it separately.

3. For those who got a cold, test the association between number of days with symptoms and both taking the supplement and gender using the command ttest. (Hint: Type “help ttest” to learn the syntax or use the menus)

- a. For those that got a cold, test whether average days with cold symptoms is influenced by the supplement.
- b. For those that got a cold, test whether average days with cold symptoms is influenced by the supplement separately for males and females.
- c. For those that got a cold, test whether average days with cold symptoms is influenced by gender.
- d. For those that got a cold, test whether average days with cold symptoms is influenced by gender for those taking the supplement and those not taking it separately.

Report Requirements

4. Prepare a brief write up summarizing, interpreting and analyzing your results from 1-3. Don't write more than 3 pages of text. Write as if you are a member of the team conducting the trial explaining the results to a funding agency or to your boss. Refer to results from 1-3 that you use as you need them. For example, you might want to create a table of means by category to refer to from #1 and place it in your write up. Also, you will want to refer to confidence intervals, p-value, t-statistics, etc... that are relevant to your explanation. But you don't have to include every table generated in your output—just what you use. What matters is that is that your

write up is easy to read, concise, and thoroughly explains the results of your analysis. Please make sure your report includes the following at a minimum

- cover page
- table of contents
- list of figures and tables
- Executive Summary (similar to abstract)
- Professional headings for each section
- captions for tables and figures
- Conclusion
- Appendix (includes Stata do-file)

Grading Rubric

Analysis – 50 points

Write-Up – 25 points

Report organization and formatting – 25 points.