Lecture 3.3 - Practicing Hypotheses Tests

Student

2025-11-20

Hypothesis Testing

Setting Expectations

Included is a dataset regarding student drinking habits and other student characteristics from a 2008 survey of Portugese high school students.

Some basic facts about Europe:

- According to a paper titled, "Youth Drinking Rates and Problems: A Comparison of European Countries and the United States" by Friese and Grube, the average problem drinking rate in Europe is about 19%.
- According to https://www.statista.com/statistics/377585/household-internet-access-in-eu28/, in 2008, the number of households in Europe with Internet access was about 60%
- According to http://strongerfamilies.eu/about-us-2/one-parent-families-in-europe/, about 10% of families in Europe are single-parent families
- According to https://www.pewinternet.org/2015/10/01/basics-of-teen-romantic-relationships/, about 19% of US students are in a romantic relationship

Setup and Data Exploration

1. What are the variables that map to these outcomes in your dataset?

For the following questions, make a small table:

2. What is the percentage of problem drinkers in the sample?

   Note that student drinking is measured on a 1-5 scale. You can make a two-category version by using the `mutate` verb and `case_when()` as follows:

```
student.drinking.cleaned <- student.drinking.cleaned %>%
  mutate(problem.drinking = case_when(alcohol.use < 4 ~ 0,
                                      alcohol.use >= 4 ~ 1))
```

   Think about what would happen if you reclassified this by defining 4 and above as problem drinking – make a note about this.

3. What is the percentage of students with Internet in their home in the sample?
4. What is the percentage of students who have single family homes in the sample?

   Note that A is Apart and T is Together

5. What is the percentage of students who are in a romantic relationship in the sample?

Planning

Pick one of the variables from questions 2-5. For this variable, consider the following questions:

6. What, in your opinion, would be a substantively significant difference in this mean/proportion? How large would a difference need to be for you to consider it meaningful? Make some notes about this.
7. Are the conditions satisfied for conducting a hypothesis test? Why or why not?
8. What, in your opinion, should Portuguese policymakers do if you judge the the mean/proportion in your data are significantly different than world averages? Make some notes about this.

Calculation

9. Write a fully specified hypothesis using a 95% cutoff.
10. Calculate the p value for your hypothesis and and indicate whether you reject or fail to reject the null hypothesis

Interpretation

11. Write a paragraph describing your results and also what kind of policy reforms you would recommend to policymakers based on the results of your hypotheses tests

Extra

Repeat the process for any of the other variables from questions 2-5 that you did not consider in the first round.