

Math 58 / 58B - Introduction to (Bio)Statistics

your name here

due Mar 5, 2020

Homework 6

Assignment Summary (Goals)

- two sample proportion
- experiments vs observational studies
- causation

1. Diabetes and unemployment¹.

A Gallup poll surveyed Americans about their employment status and whether or not they have diabetes. The survey results indicate that 1.5% of the 47,774 employed (full or part time) and 2.5% of the 5,855 unemployed 18-29 year olds have diabetes.

- (a) Create a two-way table presenting the results of this study. [Hint: in R create the data frame in the way we've done prior to `infer ...` do **not** apply `infer` here. Use the `table()` command to see a 2x2 summary of the data.]
- (b) State appropriate hypotheses to test for difference in proportions of diabetes between employed and unemployed Americans.
- (c) Is it appropriate to use the normal distribution to perform the test? Why or why not?
- (d) The sample difference is about 1%. If we completed the hypothesis test (you don't need to do so), we would find that the p-value is very small (about 0), meaning the difference is statistically significant. Use this result to explain the difference between statistically significant and practically significant findings.

2. Eat better, feel better?².

In a public health study on the effects of consumption of fruits and vegetables on psychological well-being in young adults, participants were randomly assigned to three groups: (1) diet-as-usual, (2) an ecological momentary intervention involving text message reminders to increase their fruits and vegetable consumption plus a voucher to purchase them, or (3) a fruit and vegetable intervention in which participants were given two additional daily servings of fresh fruits and vegetables to consume on top of their normal diet. Participants were asked to take a nightly survey on their smartphones. Participants were student volunteers at the University of Otago, New Zealand. At the end of the 14-day study, only participants in the third group showed statistically significant improvements to their psychological well-being across the 14-days relative to the other groups.

- (a) What type of study is this?
- (b) Identify the explanatory and response variables.
- (c) Comment on whether the results of the study can be generalized to the population.
- (d) Comment on whether the results of the study can be used to establish causal relationships.
- (e) A newspaper article reporting on the study states, "The results of this study provide proof that giving young adults fresh fruits and vegetables to eat can have psychological benefits, even over a brief period of time." How would you suggest revising this statement so that it can be supported by the study?

¹From OpenIntro Statistics, exercise 6.45. Data taken from: Gallup Wellbeing, Employed Americans in Better Health Than the Unemployed, data collected Jan. 2, 2011 - May 21, 2012. <https://news.gallup.com/poll/155408/Employed-Americans-Better-Health-Unemployed.aspx>

²From OpenIntro Statistics, exercise 1.41. Tamlin S Conner et al. "Let them eat fruit! The effect of fruit and vegetable consumption on psychological well-being in young adults: A randomized controlled trial". In: PloS one 12.2 (2017), e0171206.

3. **Mandatory vaccinations?**³ In Fall 2014, the Pew Research Center surveyed 2,002 adults (“The survey of the general public was conducted by landline and cellular telephone August 15-25, 2014 with a representative sample of 2,002 adults nationwide”) and a representative sample 3,748 AAAS (American Association for the Advancement of Science) U.S. based members (the survey was conducted online from Sept. 11 to Oct. 13, 2014) to compare the “public” and “scientists” views on science and society. One of the common survey questions was about whether there should be mandatory vaccination against childhood decisions (e.g., MMR). For this question, 68% of the public favored vaccinations compared to 86% of the scientists.
- Identify and classify the two variables in this study. Which variable is being treated as the response variable and which as the explanatory variable?
 - Would you classify this study as independent random samples, one random sample with two-variables, or a randomized experiment?
 - Create a two-way table of counts to summarize these data.
 - Set up null and alternative hypotheses to assess whether there is convincing evidence that “scientists see world differently than the general public” on this issue. Be sure to include an interpretation of the parameter of interest.
 - Are the technical conditions for the normal approximation to be valid met for this study? Explain.
 - Use the normal approximation to calculate the test statistic and approximate the p-value. Interpret the p-value in the context of the scientific results.
 - Calculate and interpret a 95% confidence interval for the parameter. Be sure to specify the population(s) you are willing to generalize your conclusions to.
 - Suppose you had defined the parameter by subtracting in the other direction (e.g., scientists - public instead of public - scientists or vice versa). How would that change:
 - The observed statistic?
 - The test statistic?
 - The p-value?
 - The confidence interval?
 - Another question asked whether humans evolved over time, with 65% of the public agreeing and 98% of the scientists agreeing. Would the normal approximation be valid for this study? Explain.

³From ISCAM, HW 2.2