

Project Homework 5

STAT011 with Prof Suzy

Due by Tuesday, 12/5/23 11:59pm

Introduction and Purpose

The purpose of this assignment is for you to practice using software for statistical inference and to practice assessing which of the chi-squared procedures to use for different problem settings. Problems 1 and 2 require you to analyse data of your choice from any of the previous data sets we've used in the first four PHW assignments. These first two problems will be graded for completion. Problems 3, 4, and 5 are all related to one another and do not require any data analysis. These problems will be graded for correctness.

Instructions

Choose which data set you would like to analyze. If you are using Excel, you can download the data sets from our Stat 11 Github Data page. Do this by right clicking on the link "View Raw" and save the link with the name XXX.csv.

If you are using R, you can import any of our previous data sets with the following commands:

```
Burger_King_items <- read.csv(  
  "https://raw.githubusercontent.com/dr-suz/Stat11/main/Data/Burger_King_items.csv")
```

```
EnvoyAir_flights <- read.delim(  
  "https://raw.githubusercontent.com/dr-suz/Stat11/main/Data/EnvoyAir_flights.txt",  
  sep=",")
```

```
arthritis <- read.delim(  
  "https://raw.githubusercontent.com/dr-suz/Stat11/main/Data/arthritis.csv",  
  sep=",")
```

```
titanic <- read.delim(  
  "https://raw.githubusercontent.com/dr-suz/Stat11/main/Data/titanic.csv",  
  sep=",")
```

```
gardasil <- read.delim(  
  "https://raw.githubusercontent.com/dr-suz/Stat11/main/Data/gardasil_data.txt",  
  sep="\t")
```

You are encouraged to work with your classmates (in particular, with your final project group mates) on this assignment but you must hand in your own, unique write up of the solutions. In a Word document, clearly label each problem's solution. Most solutions will include graphics which can be copied from Excel or RStudio and pasted into your solution document. All solutions require a written component. When you are ready to submit your assignment, save the Word document as a PDF and upload it to the Moodle link for Project Hw #5.

Problem 1

Define your population of interest and define a relevant proportion related to this population. Conduct a hypothesis test about this unknown proportion using the data. (Don't repeat Problem 5 from PHW #4.) Make sure your answer includes:

- 1) the definition of your parameter and your significance level;
- 2) your null and alternative hypotheses;
- 3) an assessment of the required assumptions and conditions;
- 4) the calculated test statistic;
- 5) the p-value and conclusion of your test in context.

Problem 2

Define your population of interest and define a relevant population mean that you will estimate with a confidence interval. (Don't repeat Problem 1 from PHW #4.) Make sure your answer includes:

- 1) the definition of your parameter and your confidence level;
 - 2) the sample mean;
 - 3) an assessment of the required assumptions and conditions;
 - 4) the critical value corresponding to your confidence level;
 - 5) the standard error of your sample mean;
 - 6) the lower and upper bounds of your interval interpreted in context.
-

Problem 3

Assess whether or not the problems listed below deal with count data for a categorical variable. (Hint: There are four problems that do not deal with count data.)

Scenario 1. A brokerage firm wants to see whether the type of account a customer has (Silver, Gold, or Platinum) affects the type of trades that customer makes (in person, by phone, or on the Internet). It collects a random sample of trades made for its customers over the past year and performs a test.

Scenario 2. The brokerage firm from (1) also wants to know if the type of account affects the size of the account (in dollars). It performs a test to see if the mean size of the account is the same for the three account types.

Scenario 3. The academic research office at a large community college wants to see whether the distribution of courses chosen (Humanities, Social Science, or Science) is different for its residential and nonresidential students. It assembles last semester's data and performs a test.

Scenario 4. A medical researcher wants to know if blood cholesterol level is related to heart disease. She examines a database of patients, testing whether the cholesterol level (in milligrams) is related to whether or not a person has heart disease.

Scenario 5. Is the quality of a car affected by what day it was built? A car manufacturer examines a random sample of the warranty claims filed over the past two years to test whether defects are randomly distributed across days of the workweek.

Scenario 6. A student wants to find out whether political leaning (liberal, moderate, or conservative) is related to choice of major. He surveys randomly chosen students and performs a test.

Scenario 7. A sales representative who is on the road visiting clients thinks that, on average, he drives the same distance each day of the week. He keeps track of his mileage for several weeks and discovers that he averages miles on Mondays, miles on Tuesdays, miles on Wednesdays, miles on Thursdays, and miles on Fridays. He wonders if this evidence contradicts his belief in a uniform distribution of miles across the days of the week.

Scenario 8. A study was performed examining epidurals as one factor that might inhibit successful breastfeeding of newborn babies. Suppose a broader study included several additional issues, including whether the mother drank alcohol, whether this was a first child, and whether the parents occasionally supplemented breastfeeding with bottled formula.

Scenario 9. Two different professors teach an introductory statistics course. The table shows the distribution of final grades they reported. We wonder whether one of these professors is an “easier” grader.

Scenario 10. A student studying the music preferences of her classmates wants to test her theory that men and women have different music preferences. For each respondent, she records their gender and what their favorite music genre is (hip hop, trap, rock, rap, folk, indie, other).

Problem 4

For the six scenarios in Problem 3 that do deal with count data, determine which chi-square test is appropriate for each.

Problem 5

State the null and alternative hypotheses for each of the scenarios in problem 4.