

Problem set 11

S520

Upload your typed answers to questions 1 to 6 through the Assignments tab on Canvas by 11:59 pm, Thursday 21th November.

Trosset question numbers refer to the hardcover textbook. Draw all graphs in R and include all R code. You may work with others, but you must write up your homework independently — you should not have whole sentences in common with other students or other sources.

1. Trosset chapter 13.4 exercise 1.
2. Trosset chapter 13.4 exercise 3.
3. Trosset chapter 13.4 exercise 10.
4. Trosset chapter 13.4 exercise 11.
5. Is anger associated with heart disease? Williams et al. (2000) gave a sample of 8474 people with no heart disease and normal blood pressure a test to classify their anger as low, moderate, or high. The researchers then tracked the subjects for four years, recording which of them developed heart disease. The results:
 - Of 3110 people with low anger, 53 got heart disease.
 - Of 4731 people with moderate anger, 110 got heart disease.
 - Of 633 people with high anger, 27 got heart disease.
 - (a) Perform a chi-squared test, giving a test statistic, P -value, and conclusion.
 - (b) Does this analysis alone prove that anger affects the chance of getting heart disease? Explain why or why not.
6. In class we found that the number of goals scored in games of the 2000–01 English Premier League soccer season could be adequately modeled using a Poisson distribution. Does the same hold for the 2016–17 season? Data for this season can be readily found on the website `football-data.co.uk` as a .csv file, a common format for spreadsheet data. You can read it into R using `read.csv()`:

```
EPL201819 <- read.csv("http://www.football-data.co.uk/mmz4281/1819/E0.csv")
```

Within the data set, the column `FTHG` contains the number of home goals and the column `FTAG` contains the number of away team goals for each of the 380 games.

Taking each game as an experimental unit, is the Poisson model a good fit for the following total goals (home plus away)? Perform an appropriate chi-squared test as follows:

- (a) Create a new variable (of length 380) called `total.goals`.
- (b) Count the number of games with 0 goals, 1 goals, ..., 5 goals, 6 goals, and “7 or more goals.” These are your observed counts.
- (c) Find the mean number of total goals over all 380 games. Call this `m`.
- (d) Generate the expected counts as follows:

```
expected <- 380 * c(dpois(0:6, m), 1 - ppois(6, m))
```
- (e) Calculate a chi-squared statistic and get a P -value by comparing it to a χ^2 -distribution with 6 degrees of freedom (8 categories minus 1 total minus 1 estimated parameter), and state a conclusion.