Problem set 12

S520

Upload your typed answers to questions 1 to 6 through the Assignments tab on Canvas by 11:59 pm, Thursday 12th December.

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. The psychologists Daniel Kahneman and Amos Tversky described the following situation:

   > The instructors in a flight school adopted a policy of consistent positive reinforcement recommended by psychologists. They verbally reinforced each successful execution of a flight maneuver. After some experience with this training approach, the instructors claimed that contrary to psychological doctrine, high praise for good execution of complex maneuvers typically results in a decrement of performance on the next try.[1]

   Is there a simpler explanation for the decreased performance following praise? What does this have to do with chapter 15?

2. Trosset chapter 15.7 exercise 5, parts (a), (b), and (c).

3. Trosset chapter 15.7 exercise 8

4. For adult women in the U.S., height in centimeters and $\log_e$ of weight in kilograms follow an approximately bivariate normal distribution (except at the tails.) Summary statistics are:

   - Height: Mean 162.1 cm, SD 7.3 cm
   - Log weight: 4.289, SD 0.253
   - Correlation between height and log weight: 0.3075

   (a) Find the equation of the regression like to predict log weight from height.
   (b) Complete the following: "According to the bivariate normal, adult women who are 165 cm tall have a mean log weight of _____ and an SD of log weight of _____."
   (c) What percentage of 165 cm adult women are less than 80 kg in weight?

5. A Major League Baseball team plays 162 games each season. There are 30 teams. Each season, the number of wins by Major League Baseball teams has an approximately normal distribution with mean 81 and standard deviation 11.7. The correlation between a team's wins one season and their wins the next season is 0.54.

---

[1] Reprinted in *Judgement Under Uncertainty: Heuristics and Biases* (1982).

(a) Suppose a baseball executive believed the best prediction of a team's wins in 2015 should be equal to their wins in 2014. For example, he predicts that the Los Angeles Angels, who had the most wins in 2014 with 98, would have 98 wins in 2015. Using the data given, explain to the baseball executive (who knows very little statistics) why this particular prediction is likely too high.

(b) Use regression to predict the Los Angeles Angels' 2015 wins using only the above data.

(c) The executive looks at the regression predictions for all MLB players and sees that no team is predicted to win more than 91 games. The executive suspects the predictions are too low, because in every full season since 1961, at least one team has won at least 96 games. Explain to the executive, who knows very little statistics, why his suspicions are misplaced.

6. The file `examanxiety.txt` on Canvas contains information on a number of variables:

   - `Exam`: score on a math exam
   - `Revise`: hours spend revising for the math exam
   - `Anxiety`: "math anxiety" on a scale from 0 to 100 (100 is most anxious)

(a) Find the regression line to predict exam score from anxiety. Write down your answer as an equation (do not just paste R output.)

(b) Which of the following regression assumptions are met?
    i. Linearity
    ii. Independence
    iii. Equal variance (homoskedasticity)
    iv. Normality of errors

(c) Suppose we want to make probabilistic predictions of a student's exam score given their math anxiety. Should we use the bivariate normal? Why or why not?