

STAT 213 (SUMMER 2021): HW1

DUE ELECTRONICALLY BY THE START OF CLASS, FRIDAY 6/4/21

1. **Scooping ice cream.** In a study reported in the *Journal of Preventative Medicine*, 85 nutrition experts were asked to scoop themselves as much ice cream as they wanted. Some of them were randomly given a large bowl (34 ounces) as they entered the line, and the others were given a smaller bowl (17 ounces). Similarly, some were randomly given a large spoon (3 ounces) and the others were given a smaller spoon (2 ounces). Researchers then recorded how much ice cream each subject scooped for themselves. Their conjecture was that those given a larger bowl would tend to scoop more ice cream, as would those given a larger spoon.
 - (a) Identify the cases (observational units) in this study.
 - (b) Is this an observational study or a controlled experiment? Explain how you know.
 - (c) Identify the response variable in this study, and classify it as quantitative or categorical.
 - (d) Identify the explanatory variable(s) in this study, and classify it(/them) as quantitative or categorical.
2. **Roller coasters.** The Roller Coaster Database (rrdb.com) contains lots of information about roller coasters all over the world. The following statistical model for predicting the top speed (in miles per hour) of a coaster is based on 100 roller coasters in the United States and data displayed on the database in November 2003:

$$\text{TopSpeed}_i = 54 + 7.6 \cdot \text{TypeCode}_i + \varepsilon_i$$

where $\text{TypeCode} = 1$ for steel roller coasters and $\text{TypeCode} = 0$ for wooden roller coasters.

- (a) What top speed does this model predict for a wooden roller coaster?
- (b) What top speed does this model predict for a steel roller coaster?
- (c) Determine the difference in predicted speeds in miles per hour for the two types of coasters. Also identify where this number appears in the model equation, and explain why that makes sense.

Date: May 28, 2021.

3. **Computing a residual.** Consider the fitted regression equation

$$\hat{Y} = 25 + 7 \cdot X$$

If $X_1 = 10$ and $Y_1 = 100$, what is the residual for the first data point?

4. **Pythagorean theorem of baseball.** Renowned baseball statistician Bill James devised a model for predicting a team's winning percentage. Dubbed the "Pythagorean Theorem of Baseball," this model predicts a team's winning percentage as

$$\text{WinPct} = \frac{\text{RunsScored}^2}{\text{RunsScored}^2 + \text{RunsAllowed}^2} \times 100 + \varepsilon$$

- (a) Use this model to predict the winning percentage for the 2009 New York Yankees, who scored 915 runs and allowed 753 runs.
- (b) The Yankees actually won 103 games and lost 59 in 2009. Determine the winning percentage corresponding to this record, and also determine the residual winning percentage from the Pythagorean model (by taking the observed winning percentage minus the predicted winning percentage).
- (c) Interpret what this residual value means for the 2009 Yankees. (Hint: Did the team do better or worse than expected given their runs scored and runs allowed? By how much?)
- (d) Repeat (a-c) for the 2009 San Diego Padres, who scored 638 runs and allowed 769, while winning 75 games and losing 87.
- (e) Which team (Yankees or Padres) exceeded their Pythagorean expectations by more?