STA 207 HW-1 Solution

Problem-1 [38 points]

For each of the sets of variables, identify the response variable (Y) and independent variables (X) in the study. Then, identify if the variables are quantitative (numerical: continuous or discrete) or qualitative (categorical: ordinal or nominal).

a.) [12 points] A study is conducted to understand how the state average SAT Math scores vary based on the average SAT Verbal scores. The variables are SAT Math scores, SAT verbal scores, percentage of eligible students taking the SAT, percentage of adult population without a high school education, average annual teacher salary, and state population.

As we discussed in the class, we are trying to report Y and X envisioning what would be measured/asked/recorded on each sampling unit. We are trying (our best!) to see what might the data look like. If you explained a different type based on how you envision that variable being measured, that is ok.

Solution:

Y: state Math score, numeric continuous variable
X1: SAT Verbal scores, numeric continuous variable
X2: was the sampled student eligible to take the SAT or not? categorical (yes/no) nominal variable
X3: did the sampled adult completed a high school education or not? Categorical, nominal variable
X4: annual salary of sampled teacher, numeric continuous variable, and
X5: state population, numeric discrete variable

b.) [12 points] To see how shooting percentage, average total points, rebounds, assists, and turnover percentage affect a team's winning percentage.

Solution:

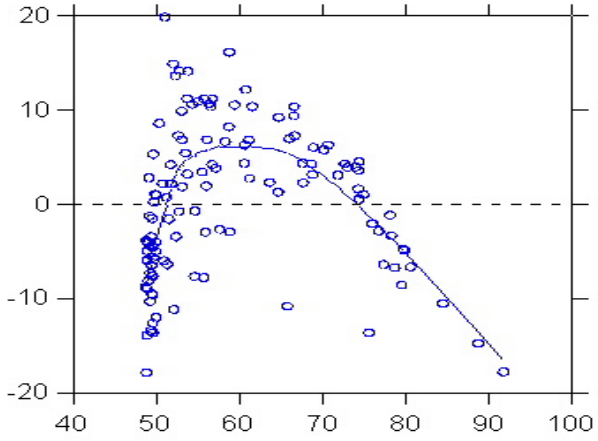
Y: number of wins and ties will be used to calculate the winning %, numeric discrete variable
X1: shooting percentage is obtained using the # of points scored, numeric discrete variable
X2: total points are computed using total points scores, numeric discrete variable
X3: # of rebounds for a player, numeric, discrete variable
X4: # of assists, numeric discrete variable, and
X5: turnover percentage is computed using the # of turnovers divided by the # of possessions, so I am considering the final quantity as the divided # which would be a numeric continuous variable

c.) [14 points] To evaluate the effect of lifestyle factors on sleep efficiency, an experiment is conducted and variables including caffeine consumption, alcohol consumption, smoking status, exercise frequency is the # of times exercise in a week, age, gender, and duration of sleep are recorded.

Solution:
Y: sleep duration, numeric continuous variable
X1: caffeine consumption, numeric continuous variable
X2: alcohol consumption, numeric continuous variable
X3: smoking status, categorical, ordinal variable
X4: exercise frequency, numeric discrete variable,
X5: age, numeric discrete variable considering an integer is used to describe the age, and
X6: gender, categorical ordinal variable

Problem-2 [5 points]

Discuss what might be the value of the correlation between the X and Y variable using this graph. Explain why.



Solution: We would expect the correlation to be approximately 0 as a non-linear relation exists between the X and Y variables.

Problem-3 [7 points]

Share an interesting study (with reference to the article/blog) where correlation is used inaccurately to imply causation and write a summary.

Answer varies.

Example Study: Chocolate Consumption, Cognitive Function, and Nobel Laureates

Summary: The study explores the correlation between the number of Nobel prizes won by a country (adjusting for population) and the per capita chocolate consumption. To measure the cognitive function, the number of Nobel laureates per capita is used as a as a proxy for superior cognitive function at the national level.