Homework for Chapter 4: Describing Relationships

1. What is a conditional distribution?
2. The following figure (using fictional data) describes the relationship between Income level and rating on a scale testing for signs of Depression.  
   Chart, scatter chart

   Description automatically generated
   1. How does the conditional mean of Depression change as Income increases?
   2. Does the graph indicate that lower income causes depression? Why or why not?
3. The below fictional table depicts data collected from 3000 university students on their classification (Freshman, Sophomore, Junior, Senior) and whether or not they receive financial aid. The table shows a cross tabulation of classification and receipt of financial aid.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Financial Aid | Freshman | Sophomore | Junior | Senior |
| Yes | 508 | 349 | 425 | 288 |
| No | 371 | 337 | 384 | 338 |

* 1. Calculate the probability of receiving financial aid given that a student is a Senior.
  2. Calculate the probability that a student is a Senior given that they receive financial aid.
  3. Calculate the probability of receiving financial aid given that a student is a Freshman.

1. Describe two advantages and one disadvantage of using line-fitting methods as opposed to calculating local means.
2. Consider the line of best fit: .
   1. What is the conditional predicted mean of when ?
   2. What is the conditional predicted mean of when ?
3. Which of the following terms describes a measurement of how much two variables vary with each other?
   1. Variance
   2. Conditional mean
   3. Covariance
   4. Local mean
4. What is the difference between covariance and correlation?
5. Figure A and Figure B below depict the (fictional-data) relationship between scores on a math exam and an intelligence measure from data collected from a fictional sample of 100 students.   
   Chart, scatter chart

   Description automatically generated
   1. What type of shape is fitted in Figure A?
   2. What kind of shape is fitted in Figure B?
   3. Which shape is a better fit for the data, and how can you tell?
   4. For Figure A, describe the residuals for different ranges of math exam scores. Are the observed data below or above the line/curve? Are the residuals evenly scattered around the line/curve?
6. The below table contains fictional data on 5 employees from a company, repotting on how well they get along with their coworkers (GetAlong) and their level of job satisfaction (Satisfaction). The Prediction variable is the predicted satisfaction level, or the conditional mean of satisfaction, for each value of GetAlong derived after fitting a line of best fit using ordinary least squares estimation.

|  |  |  |  |
| --- | --- | --- | --- |
| GetAlong | Satisfaction | Prediction | Residual |
| 4.7 | 5.07 | 4.72 |  |
| 4.21 | 4.05 | 4.28 |  |
| 5.42 | 5.33 | 5.38 |  |
| 4.14 | 4.02 | 4.22 |  |
| 3.3 | 3.59 | 3.45 |  |

* 1. Fill in the “residual” column.
  2. Describe how ordinary least squares uses residuals when fitting a line.

1. We’ll be thinking here about the process of controlling for a variable. Consider the example: What is the relationship between first generation status and graduation rate in a population of college students?
   1. Give an example of at a variable that might explain why first generation status and graduation rate are correlated other than one causing the other.
   2. Describe in five steps how you would subtract out the part of the first-generation/graduation-rate relationship that is explained by the variable you listed in part a.
   3. How would you interpret the first-generation/graduation-rate relationship you get after performing the steps in part b?