Text Exercise Set 10

10-1 An instrument consisting of 15 questions is constructed to measure typical stress level with a score ranging from 10 to 50, where a score of 10 indicates an unusually low level of stress, a score of 50 indicates an unusually high level of stress, and a score of 30 indicates a moderate level of stress. The stress score and freshmen grade point average are recorded for each of many college freshmen. Suppose a very strong negative correlation was found between the stress score and freshmen grade point average. Circle true or false for each of the following statements:

(a) Freshmen with higher grade point averages tend to have higher levels of stress.
(b) Working for a higher grade point average causes higher levels of stress.
(c) Freshmen with higher grade point averages tend to have lower levels of stress.

10-2 An instrument consisting of 30 items is constructed to measure the quality of a person’s eating habits with a score ranging from 0 to 30, where a score of 0 indicates extremely poor eating habits, a score of 15 indicates satisfactory eating habits, and a score of 30 indicates excellent eating habits. The eating habits score and freshmen grade point average are recorded for each of many college freshmen. Suppose a very strong positive correlation was found between the eating habits score and freshmen grade point average. Circle true or false for each of the following statements:

(a) Freshmen with higher grade point averages tend to have better eating habits.
(b) Working for a higher grade point average causes poorer eating habits.
(c) Freshmen with higher grade point averages tend to have better eating habits.
(d) Freshmen with lower grade point averages tend to have better eating habits.
(e) Working for a higher grade point average causes poorer eating habits.
(f) Freshmen with lower grade point averages tend to have better eating habits.

10-3 Circle true or false for each of the following statements:

(a) Describing the relationship between the training time for a particular task and the actual time to perform the task is the same as describing the difference between the distribution of training times and the distribution of performance times.

(b) Describing the relationship between hair color (light, dark) is the same as describing the difference in distributions of eye colors between light-haired and dark-haired individuals.

(c) Describing the relationship between length of a fish and which of three lakes the fish came from (North Lake, South Lake, Harvey Lake) is the same as describing the difference in distributions of fish among the three lakes.

(c) Describing the relationship between sex (male, female) bicycle regularly (yes, no) is the same as describing the difference in distributions of regular and non-regular bike riders by sex.

10-5 The relationship between the amount of biotin added per kilogram of feed and the weight in grams of newly hatched turkeys after three weeks is being studied, and the data displayed as Table 10-5 is recorded for several newly hatched turkeys each given the same diet and care with varying amounts of biotin.

(a) Find the mean for each variable and the variance for each variable.
Based on the value of the correlation, which of Figure (here) would you expect a scatter plot of this data to

(d) Complete the construction of the scatter plot of the data with biotin on the horizontal axis, and decide whether you think the relationship should be considered linear or non-linear.
10-6 Table 10-6 displays data taken to study the relationship between a person's height and a person's length of stride (distance between footprints).

(a) Find the mean for each variable and the variance for each variable.

(b) Find the covariance and correlation between the two variables.

(d) Complete the construction of the scatter plot of the data with length of stride on the horizontal axis, and decide whether you think the relationship should be considered linear or non-linear.

10-7 Data is recorded on several people concerning how much orange juice they chose to drink last year and the number of days they experienced cold symptoms last year. A scatter plot is created with the variable “amount of orange juice drunk” scaled on the horizontal axis and the variable “number of days with cold symptoms” scaled on the vertical axis. A strong negative correlation is found.

(a)

(b) Why do you think this correlation might be considered to be a spurious correlation?
10-8 From data recorded about many cities, a scatter plot is “number of churches” scaled on the horizontal axis and “bars” scaled on the vertical axis. A strong positive correlation is observed. (a) Is it appropriate to conclude that more churches will open in a city? Why or why not?

(b) Why do you think this correlation might be considered reversed correlation?

10-9 Data is gathered on hair color and eye color.

(b) Is it proper to say that we want to study the relationship between hair color and eye color? Why or why not?

(c) Is it proper to say that we want to study the difference in the number of people with blue eyes between people with light hair and people with dark hair? Why or why not?

(d) Is it proper to say that we want to study the relationship between hair color and eye color? Why or why not?

(e) Is it proper to say that we want to study the relationship between eye color and height? Why or why not?
10-9 - continued

(f) Complete the construction of the contingency table to compare the distribution of hair colors between eye colors. Construction of the corresponding stacked bar chart, with height representing 100%.

(g) What does the stacked bar chart constructed in part (f) tell us about the comparison of hair colors between eye colors?

(h) Complete the construction of the contingency table to compare the distribution of eye colors between hair colors. Construction of the corresponding stacked bar chart, and scale each bar.

(i) What does the stacked bar chart constructed in part (h) tell us about the comparison of eye colors between hair colors?

(j) The stacked bar charts constructed in parts (f) and (h) display the same data. Do you prefer one over the other, why not?

(k) Would it be appropriate to construct a scatter plot or this data? Why or why not?
(a) Explain why it makes no sense to talk about finding sex of a teenager and bike riding habits.

(b) Is it proper to say that we want to study the differences and bike riding habits? Why or why not?

(c) Is it proper to say that we want to study the relation: teenager and bike riding habits? Why or why not?

(d) Is it proper to say that we want to study the difference in riders between male and female teenagers? Why or

(e) Is it proper to say that we want to study the relationship:

Ride a Bicycle
Regularly

<table>
<thead>
<tr>
<th>Sex</th>
<th>No</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(g) What does the stacked bar chart constructed in part (f) appear to suggest about the comparison of bike-riding habits between male and female teenagers?
10-10 - continued

(h) Complete the construction of the contingency table to compare the distribution of sexes between regular and non-regular bike riding teenagers; then, complete the corresponding stacked bar chart, and scale each bar to 100%.

(i) What does the stacked bar chart constructed in part (h) tell about the comparison of the sexes between regular bike riders and non-regular bike riding teenagers?

(j) The stacked bar charts constructed in parts (f) and (h) display the same data. Do you prefer one over the other? Why not?

10-11 For each pair of quantitative variables listed, indicate whether you are likely to see a linear relationship and whether the correlation is close to zero or positive, if data were taken.

(a) The variable "test score" is measured from 0 to 100 and the variable "study hours" is measured from 0 to 100 for high school students.

(b) The variable "test score" is measured from 0 to 100 and the variable "study hours" is measured from 0 to 100 for high school students.
10-11 - continued

(d) The length of time to perform a certain complicated, measured in minutes and the length of time spent in measured in hours, for each of several employees.

(e) The variable "height in inches" and the variable "height measured for each of several grade school students.

10-12 For each pair of quantitative variables listed, indicate

(b) The variable "right hand grip strength" is measured in variable age is measured in years, for each of several the ages of 5 to 50 years old.

(c) The variable "weekly TV hours" and the variable "measured for each of several high school students.

(d) The variable "number of bars" and the variable "number measured for each of several cities."
(f) A multiple choice test is administered to several sect.
The variable "number of questions answered correctly"
"number of questions answered incorrectly" are mean

(g) The variable "gas mileage" and the variable "weight
each of several different cars.

(h) The variable "gas mileage" and the variable "average
several different trips with the same automobile, who
ranged between 35 and 70 miles per hour."