**Text Exercise Set 10**

**NAME:**

10-1 An instrument consisting of 15 questions is constructed to measure a person's 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 30 indicates a normal level of stress, and a score of 50 indicates an unusually high 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.
(d) Freshmen with lower grade point averages tend to have lower levels of stress.
(e) Working for a higher grade point average causes lower levels of stress.
(f) Freshmen with lower grade point averages tend to have higher 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. If 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 poorer eating habits.
(d) Freshmen with lower grade point averages tend to have poorer eating habits.
(e) Working for a higher grade point average causes better 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 to perform a particular task and the actual time to perform the task is the same as describing the difference between the distribution of training time and the distribution of performance time. True
(b) Describing the relationship between hair color (light, dark) and eye color (light, dark) is the same as describing the difference in distribution of eye colors between light-haired and dark-haired individuals. False
(c) Describing the relationship between length of a fish (inches) and which of three lakes the fish came from (North Lake, Blue Lake, Harvey Lake) is the same as describing the difference in length of fish among the three lakes. False

10-4 Circle true or false for each of the following statements:

(a) Describing the relationship between light bulb lifetime (hours) and light bulb brand (Brite, Softlite, Nodark) is the same as describing the difference in bulb lifetime among the three brands. False
(b) Describing the relationship between height and stride length (distance between footprints) is the same as describing the difference between the distribution of height and the distribution of stride length. False
(c) Describing the relationship between sex (male, female) and riding a bicycle regularly (yes, no) is the same as describing the difference in distribution of regular and non-regular bike riders between males and females. False

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.

<table>
<thead>
<tr>
<th>Biotin (per kilogram)</th>
<th>Weight (grams)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>196</td>
</tr>
<tr>
<td>25</td>
<td>212</td>
</tr>
<tr>
<td>50</td>
<td>215</td>
</tr>
<tr>
<td>75</td>
<td>296</td>
</tr>
<tr>
<td>100</td>
<td>316</td>
</tr>
<tr>
<td>125</td>
<td>383</td>
</tr>
<tr>
<td>150</td>
<td>363</td>
</tr>
<tr>
<td>175</td>
<td>418</td>
</tr>
<tr>
<td>200</td>
<td>417</td>
</tr>
<tr>
<td>225</td>
<td>444</td>
</tr>
</tbody>
</table>

(b) Find the covariance and correlation between the two variables.
10-5 - continued

(c) Based on the value of the correlation, which of Figures 9-1a to 9-1j (reproduced here) would you expect a scatter plot of this data to resemble? Why?

(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.

(c) Based on the value of the correlation, which of Figures 9-1a to 9-1j (reproduced here) would you expect a scatter plot of this data to resemble? Why?

(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 created with the variable “number of churches” scaled on the horizontal axis and the variable “number of bars” scaled on the vertical axis. A strong positive correlation is found.
(a) Is it appropriate to conclude that more churches will be built if more bars are opened in a city? Why or why not?

(b) Why do you think this correlation might be considered to be a spurious correlation?

10-9 Data is gathered on hair color and eye color from randomly selected individuals. The results are summarized in the contingency table displayed on the right.
(a) Explain why it makes no sense to talk about finding the correlation $r$ between 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 proportion of light-eyed people between people with light hair and people with dark hair? Why or why not?

(d) Is it proper to say that want to study the relationship between light eyes and dark eyes? Why or why not?

(e) Is it proper to say that want to study the relationship between light hair and dark eyes? Why or why not?
10-9 - continued

(f) Complete the construction of the contingency table of relative frequencies to compare the distribution of hair colors between eye colors, then, complete the construction of the corresponding stacked bar chart, and scale each bar to a height representing 100%.

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

(h) Complete the construction of the contingency table of relative frequencies to compare the distribution of eye colors between hair colors, then, complete the construction of the corresponding stacked bar chart, and scale each bar to a height representing 100%.

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

(j) The stacked bar charts constructed in parts (f) and (h) are two different ways of displaying the same data. Do you prefer one over the other? If yes, why? If no, why not?

(k) Would it be appropriate to construct a scatter plot or contiguous box plots with this data? Why or why not?

10-10 In a certain state, teenagers are selected randomly, after which the sex of the teenager and whether or not the teenager rides a bicycle on a regular basis is recorded. The results are summarized in the contingency table displayed on the right.

<table>
<thead>
<tr>
<th>Sex</th>
<th>Ride a Bicycle</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Regularly</td>
</tr>
<tr>
<td></td>
<td>No</td>
</tr>
<tr>
<td>Female</td>
<td>623</td>
</tr>
<tr>
<td>Male</td>
<td>605</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>111</td>
</tr>
<tr>
<td></td>
<td>196</td>
</tr>
</tbody>
</table>
10-10 - continued
(a) Explain why it makes no sense to talk about finding the correlation $r$ between sex of a teenager and bike riding habits.

(b) Is it proper to say that we want to study the difference between sex of teenager and bike riding habits? Why or why not?

(c) Is it proper to say that we want to study the relationship between sex of teenager and bike riding habits? Why or why not?

(d) Is it proper to say that we want to study the difference in proportion of regular bike riders between male and female teenagers? Why or why not?

(e) Is it proper to say that we want to study the relationship between male and female teenage bike riders? Why or why not?

(f) Complete the construction of the contingency table of relative frequencies to compare the distribution of bike-riding habits between male and female teenagers; then, complete the construction of the corresponding stacked bar chart, and scale each bar to a height representing 100%.

(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 of relative frequencies to compare the distribution of sexes between regular bike riding teenagers and non-regular bike riding teenagers, then, complete the construction of the corresponding stacked bar chart, and scale each bar to a height representing 100%.

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

(j) The stacked bar charts constructed in parts (f) and (h) are two different ways of displaying the same data. Do you prefer one over the other? If yes, why? If no, why not?

(k) Would it be appropriate to construct a scatter plot or contiguous box plots with this data? Why or why not?

10-11 For each pair of quantitative variables listed, indicate whether or not there is likely to be a linear relationship and whether the correlation would be negative, close to zero, or positive, if data were taken.

(a) The variable "test score" is measured from 0 to 100 on a history chapter test, and the variable "study hours" is measured from 0 to 20, for each of several high school students.

(b) The variable "test score" is measured from 0 to 100 on a history chapter test, and the variable "study hours" is measured from 0 to 80, for each of several high school students.

(c) The variable "test score" is measured from 0 to 100 on a history chapter test, and the variable "height" is measured in inches, for each of several high school students.
10-11 - continued

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

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

10-12 For each pair of quantitative variables listed, indicate whether or not there is likely to be a linear relationship and whether the correlation would be negative, close to zero, or positive, if data were taken.

(a) The variable "right hand grip strength" is measured in pounds of force and the variable age is measured in years, for each of several right handed males from grades 1 through 12.

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

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

(d) The variable "number of bars" and the variable "number of churches" are measured for each of several cities.

(e) The variable "shoe size" and the variable "IQ score" are measured for each of several high school students.
(f) A multiple choice test is administered to several sections of a college course. The variable "number of questions answered correctly" and the variable "number of questions answered incorrectly" are measured for each student.

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

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