1. A test was conducted to determine the effectiveness of using an anti-inflammatory cream on delayed-onset muscle soreness. A random sample of 10 patients is treated with the cream on one arm and with a placebo on the other (control) arm. After 4 days, a measure of muscle soreness is then taken for each patient on each arm. The results are listed in the table below and in the Minitab worksheet p:\data\math\stats\muscle.mtw. Retrieve this worksheet and answer the questions below.

<table>
<thead>
<tr>
<th>Control Arm</th>
<th>Treated Arm</th>
</tr>
</thead>
<tbody>
<tr>
<td>46</td>
<td>2</td>
</tr>
<tr>
<td>22</td>
<td>32</td>
</tr>
<tr>
<td>10</td>
<td>30</td>
</tr>
<tr>
<td>14</td>
<td>3</td>
</tr>
<tr>
<td>26</td>
<td>14</td>
</tr>
<tr>
<td>29</td>
<td>32</td>
</tr>
<tr>
<td>29</td>
<td>2</td>
</tr>
<tr>
<td>47</td>
<td>39</td>
</tr>
<tr>
<td>20</td>
<td>18</td>
</tr>
<tr>
<td>13</td>
<td>2</td>
</tr>
</tbody>
</table>

a. List two measures of center for the measurements on the control arm. (5)

\[ \text{mean} = 25.6, \; \text{median} = 29, \; \text{trimmed mean} = 24.87 \]

b. List two measures of spread for the measurements on the treated arm. (5)

\[ s = 19.84, \; IQR = 32 - 2 = 30, \; \text{Range} = 39 - 2 = 37 \]

c. Compare the muscle soreness values for treated and control arms using appropriate descriptive statistics and at least one graphical display. Do you think the cream reduces muscle soreness? Explain. (10)

Use side-by-side boxplots

Although there is considerably more variation in the muscle soreness measurements for the treated arms \((s_T = 14.84, \; s_c = 12.82)\), both the mean and the median are lower: \((\bar{x}_T = 17.4, \; \bar{x}_c = 25.6, \; \bar{x}_T = 16 \; \text{vs} \; \bar{x}_c = 24)\). There is some evidence that the cream reduces muscle soreness.
2. Mark each of the following statements as True or False. (5 each part)

T F a. The major difference between an observational study and an experiment is that one uses randomization and the other does not.

T F b. When a histogram is strongly skewed to the left, the mean will be larger than the median.

F c. Blocks are used in designs to reduce variation.

F d. A pie chart is most useful for categorical data.

F e. A z-score tells you how many standard deviations your observation is away from the mean.

T F f. Self-selection bias can be eliminated by carefully wording each question on the survey.

3. A psychologist wants to study the behavior patterns of the 8,563 students at State U. She decides to start by obtaining a random sample of 30 students and asking the average number of hours each member of the sample sleeps on a weekday night. Match each of the following questions to its correct sample type. (5 each)

<table>
<thead>
<tr>
<th>SAMPLE-TYPE</th>
<th>QUESTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Simple random</td>
<td>a. Each student is assigned a number from 0001 to 8563. A number from 1 to 285 is randomly selected, and every 285th student on the list from that point on is then included in the sample.</td>
</tr>
<tr>
<td>2. Stratified</td>
<td>b. Students are separated into academic majors, and a proportional number of students are selected from each academic major.</td>
</tr>
<tr>
<td>3. Systematic</td>
<td>c. Students are listed by number (0001 to 8563) and a computer is used to randomly generate a list of 30 numbers representing the students to be used in the sample.</td>
</tr>
<tr>
<td>4. Cluster</td>
<td>d. Students are listed by their school residence locations (dormitories or apartment buildings). Three residence locations are randomly selected. Then students from each of these locations are chosen for the psychologist's sample.</td>
</tr>
</tbody>
</table>
4. A 1996 paper from *Civil Engineering* describes the compressive strength of concrete for freshwater exhibition tanks as having a mean of 6,000 psi and a standard deviation of 240 psi. Assuming the compressive strength is normally distributed,

a. What is the chance that the compressive strength of a sample of concrete is below 6,340 psi? (5)

b. What is the chance that the compressive strength of a sample of concrete is between 5,000 and 5,900 psi? (5)

c. What is the chance that the compressive strength of a sample of concrete is above 5,800 psi? (5)

d. Above what value will 70 percent of the compressive strengths of samples of concrete be? (5)

e. Below what value will 85 percent of the compressive strengths of samples of concrete be? (5)

5. An experiment results in observations on a group of subjects for three variables, an explanatory variable (X), a response variable (Y), and gender (male or female). Is it possible for X and Y to be positively associated for both males and females, but when gender is ignored, the overall association between X and Y is negative? If so, sketch a possible scatterplot to illustrate this situation. If not, explain why not. (10)
6. The Per Capita Gross Domestic Product (GDP) and the Per Capita Health Care Spending (HCS) for 22 countries is provided in p:\data\math\stats\health.mtw.

a. Find and interpret the value of Pearson’s correlation coefficient for HCS and GDP. (5)

\[ r = 0.899 \Rightarrow \text{strong linear association between HCS and GDP} \]

b. What is the equation of the least squares regression line for predicting HCS using GDP? (5)

\[ \text{HCS} = -325 + 1.04 \times \text{GDP} \]

or

\[ \text{HCS} = -324.9 + 1.0429 \times \text{GDP} \]

c. Interpret the value of the slope parameter in the least squares regression line. (5)

For a unit increase in GDP, there is a

\[ 1.0429 \] unit increase in HCS on average.

d. The country of Auschtabeckwinstille has a GDP of $12000, but no data is available for the value of its HCS. What would you predict as the value of Auschtabeckwinstille's HCS? Show how you arrived at this value. (3)

Same GDP as Australia ⇒ Predicted value = 926.54

(or plug value into regression equation)

e. What is the value of the residual for the country with GDP of 11500 and HCS of 1050? (5)

\[ \text{France:} \quad y = 874.9 \]

\[ \text{Resid} = \text{obs} - \text{pred} = 175.604 \]

f. Plot the residuals against GDP. Comment on the appearance of the plot and any implications this may have to the adequacy of the linear model. (5)

Most of the residuals are in an unstructured band between -200 and 200. However, there is one unusually large residual. Also, the three residuals GDPs close to 15,000 are all negative. A model with some curvature should be considered, but the linear model does a good job.

g. Are the residuals normally distributed? Comment on the appearance of an appropriate graphical display. (5)

The normal probability plot is roughly linear, with the exception of the one unusually large value in the top right corner. The residuals are roughly normal.
7. The winning time in the Olympic men’s 500-meter speed skating race over the years 1924 to 1992 can be described by the regression equation: Winning Time = 255 - 0.1094*(Year).

a. Is the correlation between winning time and year positive or negative? Explain. (5)

Since the correlation coefficient is the only term in the slope, $b = \frac{-534}{50}$ that can be negative or positive and $b = -0.1094$, the correlation is negative.

b. Would you be willing to use this regression equation to predict the winning time in the 2050 Winter Olympics? Explain. (5)

No, the explanatory variable (year) ranges from 1924 to 1992. There is no guarantee that the linear regression model will be appropriate for winning times in the future especially in 2050! Do not use the regression equation for extrapolation.

8. A study published in the *Journal of the American Medical Association* described a sample of 46,395 postmenopausal women who were studied for 15 years. These women were asked about their use of hormone therapy and whether or not they had breast cancer. 2082 women developed the disease. One striking result was that “for each year of combined [estrogen and progestin] therapy, a woman’s risk of breast cancer was found to increase by 8% compared to a 1% increase in women taking only estrogen.”

a. Do you think this research was an observational study or an experiment? Explain. (5)

The women were asked about their use of hormone therapy since the researchers did not impose any treatments, this is an observational study.

b. What are the explanatory and response variables for this study? Are they quantitative or categorical? (10)

Response Variable: Breast Cancer {0 if not, 1 if yes} Categorical

Explanatory Variable: Type of therapy {0 none, 1 only one category, 2 combined} Categorical