1. Consider the scatterplot below of annual income and years of education of several men.

(a) (3 pts) Describe the relationship between Income and Education in these men. Comment on the following: form, direction, and strength of the association.

Income increases with increased education with one outlier. Form is linear, direction is positive, and strength is moderate-to-strong.

(b) (3 pts) An outlier is in the upper left hand corner of the scatterplot. Explain how this individual differs from the others?

The outlier represents an individual whose income is much higher than the relationship between income and education suggests.

(c) (3 pts) Omitting the outlier, would the correlation between these two variables be closest to -0.8, -0.2, 0.2, or 0.8? Explain your answer.

+0.8 because the relationship is fairly strong and in a positive direction.

(d) (3 pts) Including the outlier, explain if the correlation would increase, decrease, or stay the same from your answer in part c above?

The outlier would weaken the relationship.
2. Measures of waist circumference and intra-abdominal fat from CT scans were collected on 25 men. Researchers are interested in predicting abdominal fat from measures of waist circumference. A scatterplot of the two variables showed a linear form.

(a) (5 pts) Use the summary statistics below to find the equation of the least-squares regression line for predicting abdominal fat from waist circumference:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y Intra-abdominal fat (cm²)</td>
<td>96</td>
<td>27</td>
</tr>
<tr>
<td>X Waist circumference (cm)</td>
<td>87</td>
<td>9.2</td>
</tr>
</tbody>
</table>

The correlation is \( r = 0.86 \)

\[
\begin{align*}
& b = r \left( \frac{s_y}{s_x} \right) = 0.86 \left( \frac{27}{9.2} \right) = 2.52 \\
& a = \bar{Y} - b \bar{X} = 96 - (2.52 \times 87) = -123.24 \\
& \hat{Y} = a + bX
\end{align*}
\]

\[\text{Intra-ab. Fat} = -123.24 + 2.52 \text{ Waist Circumference}\]

(b) (3 pts) Calculate \( r^2 \) and explain what this value means.

\[
\begin{align*}
& r^2 = (0.86)^2 = 0.7396 \\
& r^2 = 73.96\% \\
& \text{73.96\% of the variation in intra-abdominal fat is accounted for by its relationship with waist circumference.}
\end{align*}
\]

(c) (3 pts) One man had his waist circumference measured at 110 cm and his intra-abdominal fat measured at 125 cm². What is the predicted intra-abdominal fat for this man? What is the residual for this man?

\[
\begin{align*}
& \hat{Y} = -123.24 + 2.52 \times 110 = 153.92 = \text{predicted} \\
& \text{Residual} = Y - \hat{Y} = 125 - 153.92 = -28.92 = \text{residual}
\end{align*}
\]

(d) (1 pt) If researchers wanted to predict waist circumference from intra-abdominal fat, what would be the slope of the least squares regression line?

\[
b = r \left( \frac{s_y}{s_x} \right) = 0.86 \left( \frac{27}{9.2} \right) = 2.52
\]
3. Each of the following statements contains a blunder. Explain in each case what is wrong.

(a) (1 pts) "A correlation of _1 means that two variables have no association whatsoever."

\[ r = -1 \text{ indicates a perfect negative correlation.} \]

(b) (1 pts) "The correlation between atmospheric CO₂ and Sea Surface temperature was found to be .83 degrees Celsius."

\[ r \text{ is unitless} \]

(c) (1 pts) "There is a high correlation between ethnicity of American workers and their income."

\[ r \text{ is a measure between two quantitative variables. Ethnicity is categorical.} \]

(d) (1 pts) "We found a correlation of 1.02 between exercise and life span in the elderly."

\[ r \text{ takes on values between -1 and +1, so an } r = 1.02 \text{ is too big.} \]

4. (4 pts) A professor wants to send four of her students to a conference. In an effort to be fair, she decides to choose the students who will go by random selection. A list of the 20 students appears below.

01 Nakamaye 05 Embid 08 Christensen 13 Lorenz 17 Zhang
02 Herlan 06 Dudley 10 Nitsche 14 Stone 18 Umland
03 Guindani 07 Pereyra 11 Storlie 15 Loring 19 Simanca
04 Bolli 08 Briand 12 Huerta 16 Kauffman 20 Nitsche

Use the excerpt from Table B below to choose a simple random sample of four students. Label each student and circle the names of those sampled.

<table>
<thead>
<tr>
<th>Line</th>
<th>145 19887 12635 57857 95803 09991 02150 43163 58636</th>
</tr>
</thead>
</table>

n = 19, 06, 09, 10

Note: to choose, look at line 145 and circle the letters L, D, H, and B.
5. A group wants to know how much exercise the average adult New Mexican receives each week. A booth was set-up in front of a local gym, and 27 of the 45 adults asked, agreed to fill out a small survey.

(a) (1 pt) What is the population?  Adults in New Mexico

(b) (1 pt) What type of sampling is this?  Convenience

(c) (1 pt) What is the non-response rate?  \[
\frac{45 - 27}{45} = \frac{18}{45} = 40\%
\]

(d) (2 pts) As an estimate of the amount of exercise received each week by the average Albuquerque adult, will the sample mean most likely be too high, too low, or just right? Explain your answer.

Too high because the sample consists of people going to get exercise; it doesn't represent all adults in New Mexico.

6. A researcher conducts a study to investigate the effect of exercise and diet on mood. She selects three types of exercise: yoga, weight lifting, and aerobics, and two diet types: low sugar, and low fat. A 12-week experiment will compare all combinations of exercise and diet. Each treatment will have 10 different participants. The mood of each participate will be scored before and after each treatment.

(a) (2 pts) List each factor?  Exercise, Diet

(b) (2 pts) List each treatment?

- Yoga - low sugar
- Lift - low sugar
- Aerobics - low sugar
- Yoga - low fat
- Lift - low fat
- Aerobics - low fat

(c) (2 pts) How many participants does the experiment require?

10\times 10 \times 6 treatments = 600

(d) (2 pts) What is the response variable(s)?

Difference between mood scores before and after treatment
7. (4 pts) A study of elementary school children, ages 6 to 11, finds a high positive correlation between pant size and score on a test of math comprehension. Suggest a lurking variable that best explains the observed correlation. Explain your answer.

8. (4 pts) A health club is interested in finding out which of two brands of aerobic exercise equipment provides a more vigorous workout. They purchase 10 machines of each type and for five days between 9 and 12 AM they measure the average pulse rate of each person who is working out on one of these 20 machines. Is this an observational study or an experiment? Explain your answer.