Appendix B: Examples of Assessment Items

Examples of Assessment Items with Problems and Commentary About the Nature of the Difficulty

We first give some examples of assessment items with problems and commentary about the nature of the difficulty.

Assessment items to avoid using on tests: True/False, pure computation without a context or interpretation, items with too much data to enter and compute or analyze, items that only test memorization of definitions or formulas

Item 1

A teacher taught two sections of elementary statistics last semester, each with 25 students, one at 8:00 a.m. and one at 4:00 p.m. The means and standard deviations for the final exams were 78 and 8 for the 8:00 a.m. class and 75 and 10 for the 4:00 p.m. class. In examining these numbers, it occurred to the teacher that the better students probably sign up for 8:00 a.m. class. So she decided to test whether the mean final exam scores were equal for her two groups of students. State the hypotheses and carry out the test.

Critique: The teacher has all the population data so there is no need to do statistical inference.

Item 2

An economist wants to compare the mean salaries for male and female CEOs. He gets a random sample of 10 of each and does a t-test. The resulting p-value is .045.

1. State the null and alternative hypotheses.

2. Make a statistical conclusion.

3. State your conclusion in words that would be understood by someone with no training in statistics.

Critique: The question doesn’t address the conditions necessary for a t-test, and with the small sample sizes, they are almost surely violated here. Salaries are almost surely skewed.
**Item 3**

Which of the following gives the definition of a p-value?

A. It’s the probability of rejecting the null hypothesis when the null hypothesis is true.

B. It’s the probability of not rejecting the null hypothesis when the null hypothesis is true.

C. It’s the probability of observing data as extreme as that observed.

D. It’s the probability that the null hypothesis is true.

**Examples Showing Ways to Improve Assessment Items**

*True/False items, even when well-written, do not provide much information about student knowledge because there is always a 50% chance of getting the item right without any knowledge of the topic. One current approach is to change the items into forced-choice questions with three or more options.*

**Item 4**

The size of the standard deviation of a data set depends on where the center is. True or False

**CHANGED TO:**

Does the size of the standard deviation of a data set depend on where the center is located?

A. Yes, the higher the mean, the higher the standard deviation.

B. Yes, because you have to know the mean to calculate the standard deviation.

C. No, the size of the standard deviation is not affected by the location of the distribution.

D. No, because the standard deviation only measures how the values differ from each other, not how they differ from the mean.
**Item 5**

A correlation of +1 is stronger than a correlation of −1. True or False

**REWRITTEN AS:**

A recent article in an educational research journal reports a correlation of +0.8 between math achievement and overall math aptitude. It also reports a correlation of −0.8 between math achievement and a math anxiety test. Which of the following interpretations is the most correct?

A. The correlation of +0.8 indicates a stronger relationship than the correlation of −0.8.

B. The correlation of +0.8 is just as strong as the correlation of −0.8.

C. It is impossible to tell which correlation is stronger.

*Context is important for helping students see and deal with statistical ideas in real-world situations.*

**Item 6**

Once it is established that X and Y are highly correlated, what type of study needs to be done to establish that a change in X causes a change in Y?

**A CONTEXT IS ADDED:**

A researcher is studying the relationship between an experimental medicine and T4 lymphocyte cell levels in HIV/AIDS patients. The T4 lymphocytes, a part of the immune system, are found at reduced levels in patients with the HIV infection. Once it is established that the two variables, dosage of medicine, and T4 cell levels are highly correlated, what type of study needs to be done to establish that a change in dosage causes a change in T4 cell levels?

A. correlational study

B. controlled experiment

C. prediction study

D. survey

*Try to avoid repetitious/tedious calculations on exams that may become the focus of the problem for the students at the expense of concepts and interpretations.*
Item 7

A first-year program course used a final exam that contained a 20-point essay question asking students to apply Darwinian principles to analyze the process of expansion in major league sports franchises. To check for consistency in grading among the four professors in the course, a random sample of six graded essays were selected from each instructor. The scores are summarized in the table below. Construct an ANOVA table to test for a difference in means among the four instructors.

<table>
<thead>
<tr>
<th>Instructor</th>
<th>Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affinger</td>
<td>18</td>
</tr>
<tr>
<td>Beaulieu</td>
<td>14</td>
</tr>
<tr>
<td>Cleary</td>
<td>19</td>
</tr>
<tr>
<td>Dean</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>16</td>
</tr>
</tbody>
</table>

A first-year program course ... (same intro as above) ... The scores are summarized in the table below, along with some descriptive statistics for the entire sample and a portion of the one-way ANOVA output.

Descriptive Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>Median</th>
<th>TrMean</th>
<th>StDev</th>
<th>SEMean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score</td>
<td>24.00</td>
<td>15.00</td>
<td>15.00</td>
<td>15.00</td>
<td>2.92</td>
<td>0.60</td>
</tr>
</tbody>
</table>

One-way Analysis of Variance

**ANOVA TABLE OMITTED**

<table>
<thead>
<tr>
<th>Level</th>
<th>N</th>
<th>Mean</th>
<th>StDev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affinger</td>
<td>6</td>
<td>13.00</td>
<td>2.97</td>
</tr>
<tr>
<td>Beaulieu</td>
<td>6</td>
<td>13.00</td>
<td>1.55</td>
</tr>
<tr>
<td>Cleary</td>
<td>6</td>
<td>18.00</td>
<td>2.00</td>
</tr>
<tr>
<td>Dean</td>
<td>6</td>
<td>16.00</td>
<td>1.55</td>
</tr>
</tbody>
</table>

Pooled StDev = 2.098
### Individual 95% CIs for Mean
**Based on Pooled StDev**

<table>
<thead>
<tr>
<th></th>
<th>Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affinger</td>
<td><img src="scores_affinger.png" alt="Scores" /></td>
</tr>
<tr>
<td>Beaulieu</td>
<td><img src="scores_beaulieu.png" alt="Scores" /></td>
</tr>
<tr>
<td>Cleary</td>
<td><img src="scores_cleary.png" alt="Scores" /></td>
</tr>
<tr>
<td>Dean</td>
<td><img src="scores_dean.png" alt="Scores" /></td>
</tr>
</tbody>
</table>

1. Unfortunately, we are missing the ANOVA table from the output. Use the information given above to construct the ANOVA table and conduct a test (5% level) for any significant differences among the average scores assigned by the four instructors. Be sure to include hypotheses and a conclusion. If you have trouble getting one part of the table that you need to complete the rest (or the next question), make a reasonable guess or ask for assistance (for a small point fee).

2. After completing the ANOVA table, construct a 95% confidence interval for the average score given by Dr. Affinger. **Note:** Your answer should be consistent with the graphical display.

### Additional Examples of Good Assessment Items

**Item 8**

Let $Y$ denote the amount a student spends on textbooks for one semester. Suppose Nancy, who is statistically savvy, wants to know how fall, semester 1, and spring, semester 2, compare. In particular, suppose she is interested in the averages $\mu_1$ and $\mu_2$. You may assume that Nancy has taken several statistics courses and knows a lot about statistics, including how to interpret confidence intervals and hypothesis tests. You have random samples from each semester and are to analyze the data and write a report. You seek advice from four persons:

1. **Rudd says**, “Conduct an $\alpha = .05$ test of $H_0 : \mu_1 = \mu_2$ vs. $H_A : \mu_1 \neq \mu_2$ and tell Nancy whether you reject $H_0$.”

2. **Linda says**, “Report a 95% confidence interval for $\mu_1 - \mu_2$. ”
3. **Steve says,** “Conduct a test of $H_0 : \mu_1 = \mu_2$ vs. $H_A : \mu_1 \neq \mu_2$ and report to Nancy the $p$-value from the test.”

4. **Gloria says,** “Compare $\bar{y}_1$ to $\bar{y}_2$. If $\bar{y}_1 > \bar{y}_2$, then test $H_0 : \mu_1 = \mu_2$ vs. $H_A : \mu_1 > \mu_2$ using $\alpha = .05$ and tell Nancy whether you reject $H_0$. If $\bar{y}_1 < \bar{y}_2$, then test $H_0 : \mu_1 = \mu_2$ vs. $H_A : \mu_1 < \mu_2$ using $\alpha = .05$ and tell Nancy whether you reject $H_0$.”

Rank the four pieces of advice from worst to best and explain why you rank them as you do. That is, explain what makes one better than another.

**Item 9**

Researchers took random samples of subjects from two populations and applied a Wilcoxon-Mann-Whitney test to the data; the $P$-value for the test, using a nondirectional alternative, was .06. For each of the following, say whether the statement is true or false and why.

1. There is a 6% chance that the two population distributions really are the same.

2. If the two population distributions really are the same, then a difference between the two samples as extreme as the difference that these researchers observed would only happen 6% of the time.

3. If a new study were done that compared the two populations, there is a 6% probability that $H_0$ would be rejected again.

4. If $\alpha = .05$ and a directional alternative were used, and the data departed from $H_0$ in the direction specified by the alternative hypothesis, then $H_0$ would be rejected.

**Item 10**

An article on the CNN web page ([www.cnn.com/HEALTH/9612/16/fai...](http://www.cnn.com/HEALTH/9612/16/fai...)) on Monday begins with the sentence, “Family doctors overwhelmingly believe that religious faith can help patients heal, according to a survey released Monday.” Later, the article states, “Medical researchers say the benefits of religion may be as simple as helping the immune system by reducing stress,” and Dr. Harold Koenig is reported to say that “people who regularly attend church have half the rate of depression of infrequent churchgoers.”

Use the language of statistics to critique the statement by Dr. Koenig and the claim, suggested by the article, that religious faith and practice help people fight depression. You will want to select some of the following words in your critique: observational study,
experiment, blind, double-blind, precision, bias, sample, spurious, confounding, causation, association, random, valid, reliable.

**Item 11**

Francisco Franco (Class of ’98) weighed 100 Hershey’s Kisses (with almonds). He found that the sample average was 4.80 grams and the SD was .28 grams. In the context of this setting, explain what is meant by the sampling distribution of an average.

**Item 12**

A gardener wishes to compare the yields of three types of pea seeds—type A, type B, and type C. She randomly divides the type A seeds into three groups and plants some in the east part of her garden, some in the central part of the garden, and some in the west part of the garden. Then, she does the same with the type B seeds and type C seeds.

1. What kind of experimental design is the gardener using?
2. Why is this kind of design used in this situation? (Explain in the context of the situation.)

**Item 13**

The scatterplot shows how divorce rate, y, and marriage rate, x, are related for a collection of 10 countries. The regression line has been added to the plot.

1. The U.S. is not one of the 10 points in the original collection of countries. It happens that the U.S. has a higher marriage rate than any of the 10 countries. Moreover, the divorce rate for the U.S. is higher than one would expect, given the pattern of the other countries. How would adding the U.S. to the data set affect the regression line? Why?
2. Think about the scatterplot and regression line after the U.S. has been added to the data set. Provide a sketch of the residual plot. Label the axes and identify the U.S. on your plot with a triangle.
**Item 14**

Researchers wanted to compare two drugs, formoterol and salbutamol, in aerosol solution to a placebo for the treatment of patients who suffer from exercise-induced asthma. Patients were to take a drug or the placebo, do some exercise, and then have their “forced expiratory volume” measured. There were 30 subjects available. (Based on A.N. Tsoy, et al., *European Respiratory Journal* 3 (1990): 235; via Berry, *Statistics: A Bayesian Perspective*.)

1. Should this be an experiment or an observational study? Why?
2. Within the context of this setting, what is the placebo effect?
3. Briefly explain how to set up a randomized blocks design (RBD) here.
4. How would an RBD be helpful? That is, what is the main advantage of using an RBD in a setting like this?

**Item 15**

I noticed that eight students from the 114 class attended the review session prior to the second exam (in April). The average score among those eight students was lower than the average for the 21 students who did not attend the review session. Suppose I want to use this information in a study of the effectiveness of review sessions.

1. What kind of study is this, observational or experimental? Why?
2. What kind(s) of sampling error(s) or bias(es) might be of concern here?
3. (Hypothetical) I gave the data for the eight who attended and for the 21 who did not attend to my friend George. George used the data to conduct a hypothesis test. Does a hypothesis test make sense? If so, what is \( H_0 \)? If not, why not?
**Item 16**

For each of the following three settings, state the type of analysis you would conduct (e.g., one-sample $t$-test, regression, chi-square test of independence, chi-square goodness-of-fit test, etc.) if you had all the raw data and specify the roles of the variable(s) on which you would perform the analysis, but do not actually carry out the analysis.

1. Elizabeth Larntz (Class of ’02) measured the effect of exercise on pulse for each of 13 students. She measured pulse before and after exercise (doing 30 jumping jacks) and found that the average change was 55.1 and the SD of the changes was 18.4. How would you analyze the data?

2. Three HIV treatments were tested for their effectiveness in preventing progression of HIV in children. Of 276 children given drug A, 259 lived and 17 died. Of 281 children given drug B, 274 lived and seven died. Of 274 children given drug C, 264 lived and 10 died. How would you analyze the data?

3. A researcher was interested in the relationship between forearm length and height. He measured the forearm lengths and heights of each of 16 women. How would you analyze these data?

**Item 17**

I had Data Desk construct parallel dotplots of the data from four samples. I then conducted a test of $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ and rejected $H_0$ at the $\alpha = .05$ level. I also tested $H_0 : \mu_1 = \mu_2 = \mu_3$ and rejected $H_0$ at the $\alpha = .05$ level. However, when I tested $H_0 : \mu_2 = \mu_3$ using $\alpha = .05$, I did not reject $H_0$. Likewise, when I tested $H_0 : \mu_1 = \mu_4$ using $\alpha = .05$, I did not reject $H_0$.

1. Your job is to sketch a graph of the parallel dotplots of the data. That is, based on what I told you about the tests, you should have an idea of how the data look. Use that idea to draw a graph. Indicate the sample means with triangles that you add to the dotplots.

2. It is possible to get data with the same sample means that you graphed in part 1, but for which the hypothesis $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ is not rejected at the $\alpha = .05$ level. Provide a graph of this situation. That is, keep the same sample means (triangles) you had from part 1, but show how the data would have been different if $H_0$ were not to be rejected.
**Item 18**

Atley Chock (Class of ’02) collected data on a random sample of 12 breakfast cereals. He recorded $x =$ fiber (in grams/ounce) and $y =$ price (in cents/ounce). A scatterplot of the data shows a linear relationship. The fitted regression model is

$$\hat{y} = 17.42 + 0.62x$$

The sample correlation coefficient, $r$, is 0.23. The SE of $b_1$ is .81. Also, $s_{y|x} = 3.1$.

1. Find $r^2$ and interpret $r^2$ in the context of this problem.
2. Suppose a cereal has 2.63 grams of fiber/ounce and costs 17.3 cents/ounce. What is the residual for this cereal?
3. Interpret the value of $s_{y|x}$ in the context of this problem. That is, what does it mean to say that $s_{y|x} = 3.1$?
4. In the context of this problem, explain what is meant by “the regression effect.”

**Item 19**

Give a rough estimate of the sample correlation for the data in each of the scatterplots below.
Item 20

A matched pairs experiment compares the taste a regular cheese pizza of Pizza Joe’s to Domino’s. Each subject tastes two unmarked pieces of pizza, one of each type, in random order and states which he or she prefers. Of the 50 subjects who participate in the study, 21 prefer Pizza Joe’s.

1. Find a 96% confidence interval for the proportion of the population that prefers Pizza Joe’s to Domino’s.

2. How large a sample is required if the desired margin of error of the confidence interval is 4%?

Item 21

It was claimed that 1 out of 5 cardiologists takes an aspirin a day to prevent hardening of the arteries. Suppose the claim is true. If 1,500 cardiologists are selected at random, what is the probability that at least 275 of the 1,500 take an aspirin a day?

Item 22

Identify whether a scatterplot would or would not be an appropriate visual summary of the relationship between the variables. In each case, explain your reasoning.

1. Blood pressure and age

2. Region of country and opinion about stronger gun control laws

3. Verbal SAT and math SAT score

4. Handspan and gender (male or female)

Item 23

The paragraphs that follow each describe a situation that calls for some type of statistical analysis. For each, you should:

1. Give the name of an appropriate statistical procedure to apply (from the list below). You may use the same procedure more than once, and some questions might have more than one correct answer.

2. In some problems, you will also be given a p-value. Use it to reach a conclusion for that specific problem. Be sure to say something more than just Reject $H_0$ or Fail to Reject $H_0$. (Assume a 5% significance level.)
Confidence interval (for a mean, \( p, \ldots \)) & Normal distribution \\
Determining sample size & Correlation \\
Test for a mean & Simple linear regression \\
Test for proportion & Multiple regression \\
Difference in means (paired data) & Two-way table (chi-square test) \\
Difference in means (two independent samples) & ANOVA for difference in means \\
Difference in proportions & Two-way ANOVA for means \\

Some statistical procedures you might choose:

A. Anthropologists have found two burial mounds in the same region. They know several tribes lived in the region and that the tribes have been classified according to different lengths of skulls. They measure a random sample of skulls found in each burial mound and wish to determine if the two mounds were made by different tribes. (\( p \)-value = 0.0082)

B. The Hawaiian Planters Association is developing three new strains of pineapple (call them A, B, and C) to yield pulp with higher sugar content. Twenty plants of each variety (60 plants in all) are randomly distributed into a two-acre field. After harvesting, the resulting pineapples are measured for sugar content and the yields are recorded for each strain. Are there significant differences in average sugar content between the three strains? (\( p \)-value = 0.987)

C. Researchers were commissioned by the Violence In Children’s Television Investigative Monitors (VICTIM) to study the frequency of depictions of violent acts in Saturday morning TV fare. They selected a random sample of 40 shows that aired during this time period over a 12-week period. Suppose 28 of the 40 shows in the sample were judged to contain scenes depicting overtly violent acts. How should they use this information to make a statement about the population of all Saturday morning TV shows?

D. The Career Planning Office is interested in seniors’ plans and how they might relate to their majors. A large number of students are surveyed and classified according to their MAJOR (Natural Science, Social Science, Humanities) and FUTURE plans (Graduate School, Job, Undecided). Are the type of major and future plans related? (\( p \)-value = 0.047)
E. *Sophomore Magazine* asked a random sample of 15 year olds if they were sexually active (yes or no). They would like to see if there is a difference in the responses between boys and girls. *(p-value = 0.029)*

F. Every week during the Vietnam War, a body count (number of enemy killed) was reported by each army unit. The last digits of these numbers should be fairly random. However, suspicions arose that the counts might have been fabricated. To test this, a large random sample of body count figures was examined and the frequency with which the last digit was a 0 or a 5 was recorded. Psychologists have shown that people making up their own random numbers will use these digits less often than random chance would suggest (i.e., 103 sounds like a more “real” count than 100). If the data were authentic counts, the proportion of numbers ending in 0 or 5 should be about 0.20. *(p-value=0.002)*

G. In one of his adventures, Sherlock Holmes found footprints made by the criminal at the scene of a crime and measured the distance between them. After sampling many people, measuring their height and length of stride, he confidently announced that he could predict the height of the suspect. How?
**Item 24**

How accurate are radon detectors of a type sold to homeowners?
To answer this question, university researchers placed 12 detectors in a chamber that exposed them to 105 picocuries per liter of radon. The detector readings found are below. A printout of the descriptive statistics from Minitab follows.

| 91.9 | 97.8 | 111.4 | 122.3 | 105.4 | 95.0 |
| 103.8 | 99.6 | 96.6 | 119.3 | 104.8 | 101.7 |

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>Median</th>
<th>TrMean</th>
<th>StDev</th>
<th>SE Mean</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Q1</th>
<th>Q3</th>
</tr>
</thead>
<tbody>
<tr>
<td>readings</td>
<td>12</td>
<td>104.13</td>
<td>102.75</td>
<td>103.54</td>
<td>9.40</td>
<td>2.71</td>
<td>91.90</td>
<td>122.30</td>
<td>96.90</td>
<td>109.90</td>
</tr>
</tbody>
</table>

1. Is there convincing evidence that the mean 20 readings of all detectors of this type differs from the true value of 105? Perform the appropriate hypothesis test with α = .05.

2. What is the Type I error associated with this problem?

3. What is the Type II error associated with this problem?

4. What is the probability of a type II error if the reading of the detectors is too low by 5 picocuries (really 100 when it should read 105)?

**Item 25**

According a U.S. Food and Drug Administration (FDA) study, a cup of coffee contains an average of 115 mg of caffeine, with the amount per cup ranging from 60 to 180 mg depending on the brewing method. Suppose you want to repeat the FDA experiment to obtain an estimate of the mean caffeine content to within 5 mg with 95% using your favorite brewing method. In problems such as this, we can estimate the standard deviation of the population to be 1/4 of the range. How many cups of coffee must you brew?

**Item 26**

An advertisement claims that by applying a particular drug, hair is restored to bald-headed men. Outline the design of an experiment you would use to examine this claim. Assume you have money to use 20 bald men in this experiment.
Item 27

A study of iron deficiency among infants compared samples of infants following different feeding regimens. One group contained breast-fed infants, while the children in another group were fed a standard baby formula without any iron supplements. Here are the summary results on blood hemoglobin levels at 12 months of age:

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>$\bar{X}$</th>
<th>s</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breast-fed</td>
<td>23</td>
<td>13.3</td>
<td>1.7</td>
</tr>
<tr>
<td>Formula</td>
<td>19</td>
<td>12.4</td>
<td>1.8</td>
</tr>
</tbody>
</table>

Assume that the blood hemoglobin levels in children (both breast-fed and formula-fed) are normally distributed. Do a significance test to determine the statistical significance of the observed difference.

Item 28

Which implies a stronger linear relationship, a correlation of $+0.4$ or a correlation of $-0.6$? Briefly explain your choice.

Item 29

A group of physicians subjected the polygraph to the same careful testing given to medical diagnostic tests. They found that if 1,000 people were subjected to the polygraph and 500 told the truth and 500 lied, the polygraph would indicate that approximately 185 of the truth-tellers were liars and 120 of the liars were truth-tellers. In the application of the polygraph test, an individual is presumed to be a truth-teller until indicated that s/he is a liar. What is a type I error in the context of this problem? What is the probability of a type I error in the context of this problem? What is a type II error in the context of this problem? What is the probability of a type II error in the context of this problem?
Item 30

Audiologists recently developed a rehabilitation program for hearing-impaired patients in a Canadian program for senior citizens. A simple random sample of the 30 residents of a particular senior citizens home and the seniors were diagnosed for degree and type of sensorineural hearing loss which was coded as follows: 1 = hear within normal limits, 2 = high-frequency hearing loss, 3 = mild loss, 4 = mild-to-moderate loss, 5 = moderate loss, 6 = moderate-to-severe loss, and 7 = severe-to-profound loss. The data are as follows:

| 6 | 7 | 1 | 1 | 2 | 6 | 4 | 6 | 4 | 2 | 5 | 2 | 5 | 1 | 5 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 4 | 6 | 6 | 5 | 5 | 5 | 2 | 5 | 3 | 6 | 4 | 6 | 6 | 4 | 2 |

1. Create a boxplot of the data.
2. Give a good description of the data.
3. Find a 95% confidence interval for the mean hearing loss of senior citizens in this Canadian program. The mean and standard deviation of the above data are 4.2 and 1.808, respectively. Interpret the interval.

Item 31

A utility company was interested in knowing if agricultural customers would use less electricity during peak hours if their rates were different during those hours. Customers were randomly assigned to continue to get standard rates or to receive the time-of-day structure. Special meters were attached that recorded usage during peak and off-peak hours; the technician who read the meter did not know what rate structure each customer had.

1. Is this an observation study or experiment? Defend your answer.
2. What is the explanatory and response variable?
3. List a potential confounding variable in this work.
4. Is this a matched-pair design? Defend your answer.
**Item 32**

At the beginning of the semester, we measured the width of a page in our statistics book. Below is the scatterplot of the first measurement vs. the second measurement.

1. Describe the distribution.
2. Estimate the correlation with and without the starred point.

**Item 33**

A study in the *Journal of Leisure Research* investigated the relationship between academic performance and leisure activities. Each in a sample of 159 high-school students was asked to state how many leisure activities they participated in weekly. From the list, activities that involved reading, writing, or arithmetic were labeled “academic leisure activities.” Some of the results are in the table below:

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPA</td>
<td>2.96</td>
<td>0.71</td>
</tr>
<tr>
<td>Number of leisure activities</td>
<td>12.38</td>
<td>5.07</td>
</tr>
<tr>
<td>Number of academic leisure activities</td>
<td>2.77</td>
<td>1.97</td>
</tr>
</tbody>
</table>

Based on these numbers (and knowing that the GPA is a value between 0 and 4 and the number of activities cannot be negative), discuss the potential skewness of each of the above variables.

**Item 34**

Events A and B are disjoint. Discuss whether or not A and B can be independent.

**Item 35**

A sample of 200 mothers and a sample of 200 fathers were taken. The age of the mother when she had her first child and the age of the father when he had his first child were recorded.

1. Describe the data for the mothers’ age.
2. Describe the data for the fathers’ age.
3. Compare the distributions.
4. A suggestion is made to check the correlation between the ages if we wish to compare the two populations. Is this a good suggestion? Why or why not?