1 (8 Points)

An educational psychologist does a survey to estimate the average length of time that college students study each week. She uses students enrolled in her classes for the survey.

**For what population is this a representative sample?**

- [1] Students who participated in the survey.
- [2] All students at the school.
- [4] The average length of time that college students study each week.
- [0] no answer or skip this item

Submit Answer

2 (8 Points)

You are conducting a survey for a marketing firm at a major chain store. You position yourself just inside the store, questioning the first person you encounter, then the next person, and so on, until you have questioned 100 shoppers.

**Which of the following statements is correct?**

- [1] The responses you receive should not be considered representative of all shoppers in the store because you spoke with everyone you encountered, rather than just a random sample.
- [2] The responses you receive should be considered representative of all shoppers in chain store because people who were cooperative about the survey responded.
- [3] The responses you receive should be considered representative of all shoppers in the chain store because you conducted the survey at a time when the store was busy.
- [4] The responses you receive should not be considered representative of all shoppers in the chain because you spoke only with shoppers who were willing to participate in the survey.
- [0] no answer or skip this item

Submit Answer

3 (8 Points)

A study is done to compare the cholesterol levels in people who follow a particular diet plan, to the cholesterol levels in people who do not follow this plan. The researcher is able to study 200 individuals of each type.

Other factors that may affect cholesterol levels are an individual's weight, exercise habits and stress levels. The weight characteristics of the two groups of people are similar, their habits are different and their stress levels are unknown.

**In this study, the response variable is:**

- [1] exercise
- [2] cholesterol level
- [3] weight
- [4] diet plan
Suppose you were interested in the average number of hours on the Internet per day for the 40 students in a Statistics class, and you asked each student how many hours he or she spends on the Internet per day.

**What is the population of measurements?**

- [1] Whether or not students participants would respond to the question.
- [2] The 40 students in this class.
- [3] All students currently enrolled at this university.
- [4] The set of 40 answers you receive when you ask the students how many hours they spend on the Internet per day.
- [0] no answer or skip this item

---

A data set contains the variables **HW** and **Exams**. HW contains the overall grade percentage on homework assigned from a text; Exams contains the overall grade percentage on three in-class closed-book exams. Use this WebStat output to answer the following question:

<table>
<thead>
<tr>
<th>Summary Statistics:</th>
<th>Variable</th>
<th>n</th>
<th>Mean</th>
<th>Variance</th>
<th>Std. Dev.</th>
<th>Median</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>HW</td>
<td>107</td>
<td>81.82</td>
<td>138.9</td>
<td>11.79</td>
<td>84.08</td>
<td>67.75</td>
<td></td>
</tr>
<tr>
<td>Exams</td>
<td>107</td>
<td>82.89</td>
<td>85.39</td>
<td>9.241</td>
<td>84.9</td>
<td>43.87</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Min</th>
<th>Max</th>
<th>Q1</th>
<th>Q3</th>
</tr>
</thead>
<tbody>
<tr>
<td>HW</td>
<td>30.47</td>
<td>98.22</td>
<td>77.89</td>
<td>89.47</td>
</tr>
<tr>
<td>Exams</td>
<td>53.55</td>
<td>97.42</td>
<td>76.76</td>
<td>89.64</td>
</tr>
</tbody>
</table>

**Which statement best characterizes the similarities and differences in scores for the variables HW and Exams?**

- [1] Focusing on the medians, upper quartiles, and maximum for HW and Exams, one can observe that there is very little or no difference between the scores. That suggests that the top half of Homework and top half of Exams have very similar distributions of scores.
- [2] Focusing on the lower quartiles and minimum shows that the Exams were better than the Homework. That the lower Homework’s scores were worse than the lower Exam’s scores pulled the Homework’s means slightly lower.
- [3] Both preceding statements are true.
- [4] Neither of the two statements is correct.
- [0] no answer or skip this item
The next plot has $X = \text{Weight}$ and $Y = \text{numbers of brothers}$ for 120 students.

**What is the best description of the most obvious outlier?**

- [1] The outlier is a person who is unusually heavy.
- [2] The outlier is a person who has about 4 brothers.
- [3] The outlier is a person who has about 8 brothers and weighs less than 100 pounds.
- [4] The outlier is a person who has about 8 brothers and weighs about 160 pounds.
- [0] no answer or skip this item

---

A manufacturer of electronic communication equipment, wishing to market a new product, contacts a random sample of its current customers to determine how many now use a DVD player or a large screen TV.

**If 54% own DVD players, 23% own large screen TVs and 12% own both, what is the probability that a randomly selected customer will own a DVD player or a large screen TV?**

- [1] about 0.89
- [2] about 0.65
- [3] about 0.43
- [4] about 0.12
- [0] no answer or skip this item
Lie detector tests have often been criticized as unreliable. It is suspected that false positives, results indicating the subjects were lying when they were not, occur far too frequently. Suppose 300 people submitted to a lie detector test, with some being told to always tell the truth and others to lie to certain questions. The individuals administering the lie detector test were not told who was truthful and who was not truthful. Positive results indicate the subject is lying.

<table>
<thead>
<tr>
<th>Lie Detector</th>
<th>Lied</th>
<th>Truthful</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive Test</td>
<td>59</td>
<td>18</td>
<td>77</td>
</tr>
<tr>
<td>Negative Test</td>
<td>7</td>
<td>216</td>
<td>223</td>
</tr>
<tr>
<td>Total</td>
<td>66</td>
<td>234</td>
<td>300</td>
</tr>
</tbody>
</table>

Among people that lied, what is the probability that a negative test result is reported for this person?

- [1] about 0.8939
- [2] about 0.7662
- [3] about 0.1061
- [4] about 0.0233
- [0] no answer or skip this item

Suppose the amount of detergent delivered through a high pressure sprayer is a uniform distribution. If $X = \text{gallons of detergent per minute}$ and $X$ has values between 5.5 and 6.5 gallons which are equally probable. The probability density function is $f(x) = 1$ for $5.5 < x < 6.5$, $f(x) = 0$ otherwise.

What is the probability that more than 5.8 gallons of detergent per minute are dispensed?

- [1] 0.1
- [2] 0.2
- [3] 0.3
- [4] 0.7
- [0] no answer or skip this item

The following table shows the probability distribution for the number of heads that can result when a fair coin is flipped three times.

<table>
<thead>
<tr>
<th>Heads</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
</table>
The expected value of this distribution is 1.5 heads.

**Calculate the variance of the number of heads in three flips of the coin.**

- [1] 1.00
- [2] 0.866
- [3] 0.75
- [4] 0.5
- [0] no answer or skip this item

11 (8 Points)

Suppose that the probability of a prior conviction for driving under the influence of alcohol is .01 and that the police stop 3 motorists who are driving under the influence of alcohol.

**What is the probability that at least one of these 3 motorists has a prior conviction?**

- [1] about 0.000001
- [2] about 0.0297
- [3] about 0.9703
- [4] about 0.999999
- [0] no answer or skip this item

12 (8 Points)

Imagine a histogram of the sample means of ideal weights from 1,000 different random samples of $n = 25$ students.

**Would the center and spread of the histogram of $n = 25$ differ from the center and spread of the histogram for $n = 9$?**

- [1] For each sample size, the center of the distribution is different. For $n = 25$, there is more variation among the different sample means than when $n = 9$.
- [2] For each sample size, the distribution is centered at about the same point, which also is approximately the population mean. For $n = 25$, there is more variation among the different sample means than when $n = 9$.
- [3] For each sample size, the distribution is centered at about the same point, which also is approximately the population mean. For $n = 25$, there is less variation among the different sample means than when $n = 9$.
- [4] For each sample size, the center of the distribution is different. For $n = 25$, there is less variation among the different sample means than when $n = 9$.
- [0] no answer or skip this item
A bowl contains a huge number of coins. One third of the coins are pennies, one third are nickels and one third are dimes. The population mean of these coins is $0.0533 (5.33 c), and the standard deviation is $0.028 (2.8 c). Imagine choosing a sample of 400 coins from the bowl, where each coin has an equal chance of being picked.

**The standard deviation of the sum of the coins is closest to:**

- [1] $0.028
- [2] $0.056
- [3] $0.56
- [4] $5.60
- [5] $11.20
- [0] no answer or skip this item

Assume that 38% of all faculty employed at universities are female. The females have value '1', and males have value '0.' Thus the population mean is 0.38. The population standard deviation is 0.485. Assume we are choosing a sample of size 100 employees.

**To answer this question, you should use a normal curve calculator or a normal table of your choice.**

**What is the probability that the proportion is in the interval from 35% to 41%? (Choose the closest answer!)**

- [1] almost 0
- [2] about 0.5
- [3] about 0.95
- [4] almost 1
- [0] no answer or skip this item

In wanting to estimate the average travel time between customers, a manager measured the travel time for a random sample of 84 customers. Using MINITAB, she calculated a 95% confidence interval for the average travel time:
TravelTime  84  22.27  3.02  .327  (21.63, 22.91)

Which statement correctly summarizes the manager's findings?

- [1] The manager can be 95% confident that 95% of the travel times between customers is between 21.63 and 22.91 minutes.
- [2] The manager can be 95% confident that the true average travel time between customers is between 21.63 and 22.27 minutes.
- [3] The manager can be 90% confident that the true average travel time between customers is between 21.63 and 22.91 minutes.
- [4] The manager can be 95% confident that the true average travel time between customers is between 21.63 and 22.91 minutes.

[0] no answer or skip this item

16 (8 Points)

A coin is tossed 10,000 times to see if it is fair (i.e., 'heads' and 'tails' are equally likely). In particular, the investigator thought that a head came up less than it should. Let \( p \) be the probability of a head. If the coin is fair, then \( p = 1/2 \).

What is the alternative hypothesis?

- [1] \( H_a: p = 1/2 \)
- [2] \( H_a: p < 1/2 \)
- [3] \( H_a: p > 1/2 \)
- [4] \( H_a: p \neq 1/2 \)

[0] no answer or skip this item

17 (8 Points)

Researchers are interested in testing whether there are an excessive number of rat hairs in jars of peanut butter produced at a particular factory. They examine a random sample of 169 jars, and find an average of 6.3 rat hairs in each jar. The sample standard deviation is 2. They would like to do a one-sided z-test of whether the population average is equal to five (the maximum permitted by law) versus the alternative that it is greater than five.

What is the standard error of the sample mean?

- [1] about 1.3
- [2] about 0.1667
- [3] about 0.1538
- [4] about 0.014
- [5] about 0.0118

[0] no answer or skip this item
The recommended dietary allowance of folic acid for adult females is 400 mcg. Folic acid is found naturally in leafy dark green vegetables, legumes (dried beans and peas), citrus fruits and juices, and most berries. A vitamin supplement is supposed to contain $\mu = 400$ mcg of folic acid.

A random sample of 100 such vitamin tablets was obtained and the amount of folic acid contained in each tablet was determined. The sample mean was 399.92 mcg, and the sample standard deviation was $s = .5$ mcg. The question of interest is whether this supplement provides less folic acid than what it is supposed to do.

**What are the null and the alternative hypotheses?**

- [1] $H_0$: Population mean = 400; $H_a$: Population mean = 399.92
- [2] $H_0$: Population mean $\neq$ 400; $H_a$: Population mean = 400
- [3] $H_0$: Population mean = 400; $H_a$: Population mean > 400
- [4] $H_0$: Population mean = 400; $H_a$: Population mean < 400
- [5] $H_0$: Population mean = 400; $H_a$: Population mean $\neq$ 400
- [0] no answer or skip this item

A random sample of 18 female Penn College students and 18 male Penn College was selected. The stress level of each student was measured on a scale from 0 to 9 and recorded in the following data sets:

**Females**
6 1 8 5 5 4 7 8 7 7 9 5 8 4 9 3 4 7

**Males**
1 0 7 3 8 0 5 5 3 4 4 3 9 4 2 7 6 8

**WebStat** was used to calculate 95% confidence intervals for the stress levels of both female and male Penn College students:

<table>
<thead>
<tr>
<th>Females: 95% Confidence interval results:</th>
<th>Variable</th>
<th>Estimate</th>
<th>Std. Err.</th>
<th>DF</th>
<th>L. Lim</th>
<th>U. Lim</th>
</tr>
</thead>
<tbody>
<tr>
<td>Females</td>
<td></td>
<td>5.944</td>
<td>0.5206</td>
<td>17</td>
<td>4.846</td>
<td>7.0429</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Males: 95% Confidence interval results:</th>
<th>Variable</th>
<th>Estimate</th>
<th>Std. Err.</th>
<th>DF</th>
<th>L. Lim</th>
<th>U. Lim</th>
</tr>
</thead>
<tbody>
<tr>
<td>Males</td>
<td></td>
<td>4.389</td>
<td>0.6424</td>
<td>17</td>
<td>3.034</td>
<td>5.744</td>
</tr>
</tbody>
</table>

**Based on the confidence intervals for female and male Penn College students,** can
you be confident that the average stress level for females differs from the average stress levels for males?

1. Since the two confidence intervals overlap, we can be confident in concluding that the average stress level for females differs from the average stress levels for males.

2. Since the two confidence intervals overlap, we can be confident in concluding that the average stress level for females does not differ from the average stress levels for males.

3. We can conclude that the female average differs from the male average.

4. We can conclude that the female average is less than the male average.

[0] no answer or skip this item

---

20 (8 Points)

What is the average age at the time a person is married for the first time? This question was posed to a random selection of 36 people. Use this Minitab output to answer the question.

<table>
<thead>
<tr>
<th>Test of mu = 23 vs mu &gt; 23</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
</tr>
<tr>
<td>36</td>
</tr>
</tbody>
</table>

How can we interpret the p-value of this test?

1. If the population mean is really 25.16, the probability is 0.0003 that the mean of the 36 ages could be different from 23.

2. If the population mean is really 25.16, the probability is 0.0003 that the mean of the 36 ages is 23.

3. If the population mean is really 23, the probability is 0.0003 that the mean of the 36 ages could be different from 23.

4. If the population mean is really 23, the probability is 0.0003 that the mean of the 36 ages could be more than 23.

[0] no answer or skip this item

---

21 (8 Points)

The table below shows the results of four random samples conducted by four different instructors who want to determine whether students brush their teeth after every meal:

<table>
<thead>
<tr>
<th>Sample</th>
<th>Brush After Every Meal</th>
<th>Don’t Brush After Every Meal</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. 1</td>
<td>400</td>
<td>500</td>
</tr>
<tr>
<td>No. 2</td>
<td>70</td>
<td>112</td>
</tr>
<tr>
<td>No. 3</td>
<td>12</td>
<td>25</td>
</tr>
<tr>
<td>No. 4</td>
<td>4</td>
<td>9</td>
</tr>
</tbody>
</table>

For which of the above samples would it be inappropriate to use the z-interval to estimate the true proportion of students who brush their teeth after every meal?
What statistical technique should be used to answer the question Is there a difference between the mean life expectancy of men and the mean life expectancy of women?

- [1] A significance test of the null hypothesis that the means are equal.
- [2] A significance test of the null hypothesis that the means are different.

A chi-square test for independence in a 2x2 table based on 1000 observations resulted in a p-value of 0.02. If the sample size were to be decreased to 100 observations, and the proportions in each cell remained the same, it is likely that:

- [1] The p-value would increase.
- [3] The p-value would remain about the same.
- [4] There is no way to determine whether the p-value would increase or decrease.

The height and weight of a group of basketball players are related by the linear model

\[ Y = -208 + 5 \times X + e \]

where \( X \) is height (in inches) and \( Y \) is weight (in pounds).

What is the expected weight difference between a person 68 inches tall and a
**25 (8 Points)**

In data on 60 locations, the least squares line for \( X = \text{Rain} \) versus \( Y = \text{Mortality} \) is

\[
\text{Mortality} = 850 + 2.37 \times (\text{Rain}).
\]

**What is the least squares estimate of the slope \( b \)?**

- [ ] 2.37
- [ ] 358.6
- [ ] 850
- [ ] 2014.5
- [ ] no answer or skip this item

**Submit Answer**

---

**26 (12 Points)**

Suppose that your school's administration is doing a survey to see if students are in favor of building a new basketball arena.

**Describe how they might choose a sample that is representative of all students.**

**Submit Answer**

---

**27 (12 Points)**

Below are **Heights** in inches and **Weights** in pounds of 132 professional male athletes, in two sports. Also included are their **body mass index** numbers, which are defined by

\[
\text{BMI} = \text{Body Mass Index} = (\text{Weight in Pounds}) \times 703/((\text{Height in inches})^2)
\]

BMI is supposed to measure how overweight or underweight a person is. A value in the range
20-25 is fine; more is deemed overweight; under 20 is deemed underweight. It is fairly easy to be overweight under this measure. Here are boxplots for the three variables:

(a) What is the median weight, approximately?

(b) What is the weight of the heaviest person, approximately?

(c) About what percentage of these athletes have a weight of more than 330 pounds?

The next interactivity is based on 11 observations with 6 variables. Use it to construct some plots. Choose among the following choices for the best description of the plot:

1. Linear
2. Linear, but with an outlier
3. Curved
4. None of the above

Use the Interactive Tool below.

Best description for X = Y1 and Y = Y4?
In 1989, scientists identified the genetic cause of cystic fibrosis (CF): an error on chromosome 7. This has helped identify carriers of the gene within high-risk families. The genetic test for CF is 85% accurate in identifying carriers.

Suppose a husband and wife are both carriers. What is the chance that at least one of them tests positive and will be identified as carrier?

The heights of women have a distribution that is approximated by a normal curve with a mean of 64 inches and a standard deviation equal to 2.6 inches.

To answer this question, you should use a normal curve calculator or a normal table of your choice.

About what proportion of women are between 62 and 70 inches tall?

The heights of women have a distribution that is approximated by a normal curve with a mean of 64 inches and a standard deviation equal to 2.6 inches.

To answer this question, you should use a normal curve calculator or a normal table of your choice.

For women, approximately what is the 70th percentile of heights?
Suppose that 30 students will participate in an experiment in which the effectiveness using a Web-based approach to teaching statistics is compared to the effectiveness of a textbook-based approach.

**Describe how the researcher could assign participants to the two different approaches.**

In the 1970 Draft Lottery, numbers from 1 to 366 were randomly assigned to the days of the year. A young man's draft lottery number was then the number assigned to his birthday. Some people thought the lottery was not completely random.

The average lottery number for the first half of the year (first 183 days) was 205.7. If the lottery is truly random, the average should be 183.5. An approximate 95% confidence interval for the population mean of the first half of the year, using the randomization method that was used, is (194.9, 216.5).

Based on this result, what would you conclude about the null hypothesis that the theoretical mean is 183.5?

A null hypothesis is that the mean systolic blood pressure of women is 115. The alternative hypothesis is that the mean is higher than 115. Based on a sample of 100 women, a t-test is performed and the p-value is 0.035.

**What is the appropriate conclusion?**
Suppose you were given a 95% confidence interval for the difference in two population means.

What could you conclude about the two population means if the confidence interval contained the number 0?

Use **WebStat**. Load from "Data > Sample Data" the data set "Weight_loss_obese_patients.dat". This is one out of 8 questions that will work with this data set.

Calculate (and report) the mean for the variables treat-A, treat-B, and treat-C, representing treatment A, treatment B, and treatment C.

Use **WebStat**. Load from "Data > Sample Data" the data set "Weight_loss_obese_patients.dat". This is one out of 8 questions that will work with this data set.

Calculate (and report) the median for the variables treat-A, treat-B, and treat-C.
38 (12 Points)

Use WebStat. Load from "Data > Sample Data" the data set "Weight_loss_obese_patients.dat". This is one out of 8 questions that will work with this data set.

**Compare the means and medians for the variables treat-A, treat-B, and treat-C. Are they similar or do they differ much? Just based on this comparison, what could we conclude about the symmetry of the data for the three variables?**

Submit Answer

39 (12 Points)

Use WebStat. Load from "Data > Sample Data" the data set "Weight_loss_obese_patients.dat". This is one out of 8 questions that will work with this data set.

**Construct boxplots of the variables treat-A, treat-B, and treat-C, using fences to identify outliers. Are there any outliers - if so, specify them and indicate for which of the variables they have been observed. Do the boxplots suggest that the data is fairly symmetric or whether there is skewness? Answer this for each of the variables separately.**

Submit Answer

40 (12 Points)

Use WebStat. Load from "Data > Sample Data" the data set "Weight_loss_obese_patients.dat". This is one out of 8 questions that will work with this data set.

**The manufacturers of weight loss treatment C claim that their product results in an average weight loss of 35 pounds. Conduct a test for the mean weight loss under treat-C (assuming the data originates from a random sample from all customers using treatment C), where H0: mean = 35 versus H1: mean less than 35. Can we use a z-test here (if yes, do so) or do we have to use a t-test (in this case, indicate the degrees of freedom). Explain. Report the p-value for the test you choose and provide a verbal conclusion.**
41 (12 Points)

Use **WebStat**. Load from "Data > Sample Data" the data set "Weight_loss_obese_patients.dat". This is one out of 8 questions that will work with this data set.

The manufacturers of weight loss treatment A claim that their product is as good as weight loss treatment C. To verify their claim, they tell us to construct (and report) 99% confidence intervals (based on t-intervals) for the mean weight loss under treatment A and for the mean weight loss under treatment C of all customers of these two treatments (assuming the data originates from random samples from all customers using treatments A and C, respectively). Do these two confidence intervals confirm the claim? Explain your answer.

42 (12 Points)

Use **WebStat**. Load from "Data > Sample Data" the data set "Weight_loss_obese_patients.dat". This is one out of 8 questions that will work with this data set.

Assume you manually have to construct a 99% confidence interval for the mean weight loss under treatment A of all customers under this treatment (assuming the data originates from a random sample from all customers using treatment A), based on a t-interval. The general formula is "sample mean +/- multiplier * standard error". Indicate the value for the multiplier and show how you would calculate the standard error.

43 (12 Points)

Use **WebStat**. Load from "Data > Sample Data" the data set "Weight_loss_obese_patients.dat". This is one out of 8 questions that will work with this data set.

Conduct a test for the variance of the weight loss under treatment A of all customers under this treatment (assuming the data originates from a random sample from all customers using treatment A), where $H_0$: variance = 9 versus $H_1$: variance > 9. Explain what type of a test we are using and how many degrees of
freedom we have. Report the p-value and provide a verbal conclusion.

44 (12 Points)
Use WebStat. Load from "Data > Sample Data" the data set "SAT_GPA_25Students.dat". This is one out of 7 questions that will work with this data set.

Construct a scatterplot of SAT scores (x) versus GPA scores (y). Describe the overall appearance of this plot. Is there an apparent trend? Are there any obvious outliers?

Submit Answer

45 (12 Points)
Use WebStat. Load from "Data > Sample Data" the data set "SAT_GPA_25Students.dat". This is one out of 7 questions that will work with this data set.

Calculate (and indicate) the correlation between SAT scores and GPA scores. What does this value indicate?

Submit Answer

46 (12 Points)
Use WebStat. Load from "Data > Sample Data" the data set "SAT_GPA_25Students.dat". This is one out of 7 questions that will work with this data set.

Construct (and report) the regression line, predicting GPA scores from SAT scores. Do we have a significant slope? And can we use the regression equation to predict GPA scores from SAT scores?
Use WebStat. Load from "Data > Sample Data" the data set "SAT_GPA_25Students.dat". This is one out of 7 questions that will work with this data set.

Based on your regression calculations above, what would be the predicted GPA scores for someone with SAT scores of 1000, 1100, and 1200, respectively? Which two of these three predicted values are least reliable? Explain why.

Use WebStat. Load from "Data > Sample Data" the data set "SAT_GPA_25Students.dat". This is one out of 7 questions that will work with this data set.

Construct a histogram of the GPA scores. To guarantee that the 4.00 GPA scores still fall into the highest class (rather than forming a new class), set "Start bins at" to 1.001 and "Binwidth" to 1. Describe the overall shape of the histogram (i.e., is it fairly symmetric, skewed towards the higher values, skewed towards the lower values)? Are there any outliers? What is the modal bar?

Use WebStat. Load from "Data > Sample Data" the data set "SAT_GPA_25Students.dat". This is one out of 7 questions that will work with this data set.

According to the book "1600 Perfect Score" by Tom Fischgrund (HarperCollins Publishers, 2003), "every year roughly 2.3 million high school students take the SAT; of those, however, only 650 students on average achieve a perfect score of 1600." Assuming independence, how many of Utah State University’s approximately 20,000 students would you expect have had a perfect SAT score when they took this test as high school students?
Use WebStat. Load from "Data > Sample Data" the data set "SAT_GPA_25Students.dat". This is one out of 7 questions that will work with this data set.

Using all your knowledge on GPA scores, SAT scores, and your results from the previous questions related to this data set, would it be possible to predict the GPA score for all applicants to Utah State University based on your regression equation above? Explain why or why not.

When you are done answering all questions above to your satisfaction, press the button below to complete your test.

Mark Test Completed