Your Test 1 will cover Sections 1.1-5.1 & 5.5. These review problems include all problems over these sections from the released exams (Spring 2015 and Fall 2015) posted to the course website. You can review the posted grading guidelines to see how the free response questions were scored. The last free response problem is a question from the textbook over correlation and regression, as this topic was covered after Test 3 in the previous two semesters.

Thus, this set of review problems does not necessarily reflect the length and/or difficulty of Test 1. This review cannot be guaranteed to be inclusive or exclusive of all that is covered on Test 1, it is merely meant to aid you in the review process.

Part I: Multiple Choice. Circle the letter corresponding to the best answer of the choices given.

1. A sample of U.S. adults was surveyed and asked, “Do you get most of your information about current events from newspapers, magazines, the Internet, television, radio, or some other source?” A bar chart of the survey results is shown below.

For what proportion of these respondents is the Internet their primary news source?
(A) 0.1667
(B) 0.1875
(C) 6
(D) 32
2. According to the Empirical Rule, we know that:
   
   (A) Approximately 65% of the area under the normal curve falls within one standard deviation of the mean.
   
   (B) Approximately 98% of the area under the normal curve falls within two standard deviations of the mean.
   
   (C) Approximately 99.7% of the area under the normal curve falls within three standard deviations of the mean.
   
   (D) All of the above.

3. The personal income per capita of a state is the total income of all adults in the state, divided by the number of adult residents in the state. The histogram below summarizes the distribution of personal income per capita (in thousands of dollars) for each of the 50 states and the District of Columbia.

Which of the following best describes the shape of this distribution and the most likely relationship between the mean and median of this distribution?

(A) Skewed left; mean < median

(B) Skewed left; mean > median

(C) Skewed right; mean < median

(D) Skewed right; mean > median
4. Below is a stem and leaf display of the distribution of the percentage of the population aged 65 and over for each state and the District of Columbia.

```
  8 | 1
  9 | 0
 10 | 2 5 7
 11 | 5 5
 12 | 0 2 3 4 4 4 6 6 7 8 8
 13 | 0 1 3 4 5 6 6 7 7 8 8 9 9
 14 | 0 1 1 1 3 3 3 4 6 9
 15 | 0 3 5 6
 16 | 0
 17 | 8
```

Key: 8|1 = 8.1%

Based on this graph, which of the following combination of summary statistics is most appropriate for measuring the center and the spread of this distribution?

(A) The sample mean and the sample standard deviation
(B) The sample mean and the interquartile range
(C) The sample median and the sample standard deviation
(D) The sample median and the interquartile range

5. Given below are selected summary measures for the closing prices of Microsoft stock for each trading day in October 2010 and for each trading day in October 2014.

<table>
<thead>
<tr>
<th>October 2010</th>
<th>October 2014</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>25.181905</td>
</tr>
<tr>
<td>Std Dev</td>
<td>0.7128507</td>
</tr>
<tr>
<td>N</td>
<td>21</td>
</tr>
<tr>
<td>Variance</td>
<td>0.5081562</td>
</tr>
</tbody>
</table>

Based only on the provided summary measures, which of the following statements most accurately compares the distributions of closing prices for October 2010 and the distribution of closing prices for October 2014?

(A) The distribution of closing prices for October 2010 has a higher center and greater variability than the distribution of closing prices for October 2014.

(B) The distribution of closing prices for October 2010 has a higher center but less variability than the distribution of closing prices for October 2014.

(C) The distribution of closing prices for October 2010 has a lower center but greater variability than the distribution of closing prices for October 2014.

(D) The distribution of closing prices for October 2010 has a lower center and less variability than the distribution of closing prices for October 2014.
6. Joni drives to work every weekday morning. She kept track of the time it takes her to drive to work, in minutes, for 35 days. The variance of these 35 commute times is 31.36. Which of the following gives the standard deviation (with correct units) of these commute times?

(A) 983.45 minutes\(^2\)
(B) 31.36 minutes
(C) 5.6 minutes\(^2\)
(D) 5.6 minutes

**Use the following information to answer questions 7 – 9.**

Suppose that for a course you need to take next semester, the time slot that you would prefer is taught by a professor with a reputation for being extremely difficult. To determine if there is any merit to this reputation, you would like to estimate the proportion of students that took this course from this professor last semester and made an A. To do this, you randomly select 10 students that took the course from this professor last semester, who freely tell you their final later grade (A, B, C, D, or F) in the course.

7. For this example, what is the population of interest?

(A) The 10 students you surveyed.
(B) All students that have ever taken this course.
(C) All students that took this course last semester.
(D) All students that took this course from this professor last semester.

8. The variable being measured is:

(A) quantitative because the observations are numeric.
(B) qualitative because the observations are numeric.
(C) quantitative because it assigns each observation to a particular category.
(D) qualitative because it assigns each observation to a particular category.

9. Suppose that 2 of the 10 students surveyed made an A. Which of the following is correct?

(A) The value 0.20 is a population parameter.
(B) The value 0.20 is a sample statistic.
(C) The value 0.20 is a population statistic.
(D) The value 0.20 is a sample parameter.
Use the following information to answer questions 10 – 11.

The histogram below displays the distribution of the age at retirement for a random sample of 100 retirees from a Fortune 500 company.

10. What proportion of these retirees retired before the age of 60?
   (A) 0.07
   (B) 0.14
   (C) 0.70
   (D) Cannot be determined from the information given.

11. Based on this histogram, what is the most likely relationship between the mean retirement age and the median retirement of age of these 100 retirees?
   (A) The mean age is less than the median age.
   (B) The mean age is greater than the median age.
   (C) The mean age is approximately equal to the median age.
   (D) The mean age is exactly equal to the median age.
12. Parking Services randomly selected 50 commuting student permit holders to determine their level of satisfaction (very satisfied, somewhat satisfied, neutral, somewhat dissatisfied, or very dissatisfied) with the availability of parking on campus. Which of the following is the most appropriate graphical display for the survey results?

(A) Bar chart
(B) Dot plot
(C) Stem-and-leaf plot
(D) Histogram

13. Consider a data set of 30 observations which has a distribution that is approximately symmetric and unimodal with no outliers. If an outlier that is much larger than all other values is added to the data set, what effect will this have on the standard deviation and the interquartile range?

(A) Both the standard deviation and the interquartile range will greatly increase.
(B) Both the standard deviation and the interquartile range will remain close to their original values.
(C) The standard deviation will increase noticeably but the interquartile range will remain close to its original value.
(D) The standard deviation will remain close to its original value but the interquartile range will increase noticeably.

14. The areas of the 46 counties in South Carolina, in square miles, are listed in order below. The areas range from 392 to 1228 with \( Q_1 = 507 \), \( M = 660.5 \), and \( Q_3 = 795 \).

<table>
<thead>
<tr>
<th>392</th>
<th>394</th>
<th>395</th>
<th>397</th>
<th>407</th>
<th>411</th>
<th>413</th>
<th>462</th>
<th>463</th>
<th>485</th>
<th>494</th>
<th>507</th>
</tr>
</thead>
<tbody>
<tr>
<td>511</td>
<td>512</td>
<td>516</td>
<td>555</td>
<td>557</td>
<td>563</td>
<td>567</td>
<td>576</td>
<td>577</td>
<td>586</td>
<td>647</td>
<td>674</td>
</tr>
<tr>
<td>682</td>
<td>696</td>
<td>696</td>
<td>700</td>
<td>710</td>
<td>724</td>
<td>740</td>
<td>757</td>
<td>758</td>
<td>772</td>
<td>795</td>
<td>804</td>
</tr>
<tr>
<td>806</td>
<td>814</td>
<td>819</td>
<td>916</td>
<td>937</td>
<td>1080</td>
<td>1128</td>
<td>1133</td>
<td>1134</td>
<td>1228</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

According to the 1.5 × IQR rule, are there any potential outliers in the data set?

(A) No
(B) Yes; 1228 is a potential outlier
(C) Yes; 392 and 1228 are potential outliers
(D) Yes; 1080, 1128, 1133, 1134, and 1228 are potential outliers
15. A biologist wishes to study the relationship between aspect (the compass direction that a slope faces) and growth of a certain wildflower that lives in a mountainous area. She obtains a sample of 28 plants and records two variables for each plant: the biomass (in kilograms) and the aspect (in degrees, where 0 degrees is due East and 180 degrees is due West). Which of the following statements is false?

(A) Biomass and aspect are both quantitative.

(B) The level of measurement for biomass is ratio.

(C) The level of measurement for aspect is ratio.

(D) This study is an observational study.

Use the following information to answer questions 16 – 17.

Environmental toxicologists are interested in determining the amount of PCB (polychlorinated biphenyls) found in striped bass in Lake Hartwell. Researchers collect samples of fish at a single location of interest in the northern part of the lake. They use an instrument that generates an electrical charge that temporarily stuns all fish in the surrounding water, and the shocked fish float to the surface to be collected. They randomly select 10 bass that weigh less than 1 pound, 10 that weigh between 1 and 2 pounds, 10 that weigh between 2 and 3 pounds, and 10 fish that weigh more than 3 pounds. The amount of PCBs are measured and recorded along with the weight of each fish (mg/lb).

16. The sampling method used in the PCB study is:

(A) cluster sampling because the location is a cluster.

(B) systematic sampling because the researchers systematically chose striped bass to measure PCB contamination.

(C) simple random sampling because random fish float to the surface.

(D) stratified sampling because the fish are stratified by weight in the sampling process.

17. The PCB study is:

(A) a well-designed experiment because the researchers specifically chose to have 10 replications per each fish weight range.

(B) a well-designed experiment because the researchers included control groups to compare PCB content.

(C) an observational study because the researchers are measuring PCB content of fish in situ (in their natural habitat).

(D) an observational study because the researchers used a completely randomized design.
18. A reporter for the university student newspaper wants to estimate the proportion of all university students who plan to vote for Candidate A in the upcoming election. In order to perform this analysis, the reporter obtains a sample of 127 students by polling students as they walk by the library. Which of the following statements is true?

(A) The population of interest is the 127 university students the reporter polls.
(B) The proportion that the reporter wishes to estimate is a statistic.
(C) The sample of 127 students that the reporter obtains is a simple random sample.
(D) None of the above are true.

19. A news organization polls adults who attend the fourth Republican debate in January in Charleston, SC in order to determine which Republican presidential candidate is likely to win the South Carolina primary in February. They are able to randomly distribute 2,500 ballots among attendees that look identical to those seen when actual primary voting occurs. All 2,500 ballots have responses and are collected. What can be said about bias in this study?

(A) This study is likely to include a strong nonresponse bias.
(B) This study is likely to include a strong sampling bias.
(C) This study is likely to include sampling, response, and nonresponse bias.
(D) There is probably no bias in this study.

20. In the Olympic men's long jump event, the winning distances have an average of 7.96 meters with a standard deviation of 0.61 meters. In the Olympic women's long jump event, the winning distances have an average of 6.80 meters with a standard deviation of 0.42 meters. At the 2012 Olympic Games, Greg Rutherford of Great Britain won the men's long jump event with a distance of 8.31 meters and Brittney Reese from the U.S. won the women's long jump event with a distance of 7.12 meters. Which athlete had the more impressive performance, relative to their respective events?

(A) Greg had the more impressive performance.
(B) Brittney had the more impressive performance.
(C) Their performances were equally impressive.
(D) There is not enough information for comparison.
21. The following stem-and-leaf plot displays the distribution of the percentage of workers who carpool to work for the 50 states plus Washington, D.C.

```
7 | 2 8 8 9
8 | 1 3 5 6 6 6 7 8
9 | 0 1 2 2 2 4 4 6 7 7 9 9
10| 0 0 0 1 2 3 3 3 3 7 7 9
11| 2 2 2 3 3 3 5 5 7
12| 4 5
13| 6 8
14| 4
15|
16| 4
```

Key: 7|2 = 7.2%

Which of the following gives the range of this data set? Note that your answer will be in percentage points, the unit for the arithmetic difference of two percentages.

(A) 8.5 percentage points
(B) 9.2 percentage points
(C) 16.4 percentage points
(D) 92 percentage points

22. A university researcher is interested in studying the effectiveness of two brands of over the counter allergy medications. At the beginning of the fall semester, she finds 20 students who use Brand A for their fall allergies, and another 20 students who use Brand B for their fall allergies. At the end of the semester she totals the number of days that each student experienced allergy symptoms. She performs a statistical analysis, and concludes that students who used Brand A experienced fewer days with allergy symptoms versus the students who used Brand B. Which of the following statements is true?

(A) In this study, the factor is the number of days that each student experienced allergy symptoms.
(B) The relationship that the researcher found could be due to a lurking variable.
(C) The researcher used a control group and randomization in her design.
(D) None of the above are true.
Part II: Free Response. Show all your work. Indicate clearly the methods you use, because you will be graded on the correctness of your methods as well as on the accuracy and completeness of your results and explanations. **Answers with no justification will receive no credit.**

1. The number of grams of fat in 9 items from the McDonald’s breakfast menu are given below.

<table>
<thead>
<tr>
<th>Menu Item</th>
<th>Fat (g)</th>
<th>Menu Item</th>
<th>Fat (g)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Egg McMuffin</td>
<td>13</td>
<td>Bacon, Egg &amp; Cheese Bagel</td>
<td>29</td>
</tr>
<tr>
<td>Sausage Biscuit</td>
<td>27</td>
<td>Sausage McGriddles</td>
<td>22</td>
</tr>
<tr>
<td>Southern Style Chicken Biscuit</td>
<td>20</td>
<td>Hotcakes &amp; Sausage</td>
<td>24</td>
</tr>
<tr>
<td>Bacon, Egg &amp; Cheese Biscuit</td>
<td>26</td>
<td>Sausage Burrito</td>
<td>16</td>
</tr>
<tr>
<td>Steak, Egg &amp; Cheese Biscuit</td>
<td>32</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(a) Calculate the five-number summary for this distribution and enter your responses in the table below.

<table>
<thead>
<tr>
<th>MIN</th>
<th>Q₁</th>
<th>MEDIAN</th>
<th>Q₃</th>
<th>MAX</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(b) Are there any outliers in the data set? Show your calculations.

(c) Draw a boxplot of the data using the number line below.

(d) McDonald’s Big Breakfast with Hotcakes has 56 grams of fat. If this value was added to the data set of 9 numbers above, how would the mean and median of the data set above compare to the mean and median of the new data set with the 10 numbers? Explain how this comparison could be made without performing any calculations.
2. Three horses are running a race. Their names are Xerxes (X), Yellowtail (Y), and Zoinks (Z). They are equally likely to win the race.

   (a) List all the possible orders in which these horses could finish the race.

   (b) Carlos bets that the horses will finish in the order YXZ. What is the probability that at least one of the horses will finish in the position that he predicts?
3. A clinical trial to determine the effectiveness of zinc lozenges at reducing the duration of the common cold randomly assigned the 100 study participants to begin taking zinc lozenges or placebo lozenges within 24 hours of developing cold symptoms. The side-by-side boxplots below display the distributions of time to recovery (in days) for the two groups.

Write a few sentences comparing the distribution of time to recovery for the treatment group taking the zinc lozenges and the distribution of time to recovery for the placebo group.
4. The data below give the weight, in grams, recorded from a sample of 25 U.S. nickels.

4.97  5.03  5.00  4.96  5.05  4.98  5.07  5.01  4.98  4.93  5.04  5.04  5.00
4.98  4.99  5.03  5.08  4.96  5.01  4.99  4.94  5.00  5.02  5.00  5.00

(a) Use the number line below to construct a dot plot of these weights.

4.90  4.92  4.94  4.96  4.98  5.00  5.02  5.04  5.06  5.08  5.10

(b) Based on the dot plot in part (a), is it appropriate to apply the Empirical Rule to the distribution of nickel weights? Explain.

(c) If the Empirical Rule applies to a given distribution, approximately what percentage of observations will fall within two standard deviations of the mean?

(d) For the sample of nickel weights above, the mean is $\bar{x} = 5.002$ grams and the standard deviation is $s = 0.037$ grams. Determine the percentage of nickel weights in this sample of 25 observations that actually fall within two standard deviations of the mean.

(e) How close is your answer in part (d) to your answer in part (c)? Is this what you would expect based on your answer in part (b)?
José, Kishawn, and Lloyd are running a race. They are all equally likely to win. A gold and silver medal will be awarded for first and second place, respectively.

(a) Use a counting rule to determine the number of ways that two of these three runners could finish in first and second place. Show all work.

(b) List all the ways that two of these three runners could finish in first and second place. Use the letters J, K, and L.

(c) What is the probability that José finishes in first place or Kishawn finishes in second place, or both?
6. Your Fair Isaacs Corporation (FICO) credit score is used to determine your creditworthiness. It is used to determine whether you qualify for a mortgage or credit card and is even used to determine interest rates. FICO scores have a range of 300 to 850, with a higher score indicating a better credit history. The given data represent the interest rate (in percent) a bank would offer on a 36-month auto loan for various FICO scores.

<table>
<thead>
<tr>
<th>Credit Score</th>
<th>Interest Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>545</td>
<td>19.0</td>
</tr>
<tr>
<td>595</td>
<td>18.0</td>
</tr>
<tr>
<td>640</td>
<td>12.2</td>
</tr>
<tr>
<td>675</td>
<td>8.6</td>
</tr>
<tr>
<td>705</td>
<td>6.7</td>
</tr>
<tr>
<td>750</td>
<td>5.2</td>
</tr>
</tbody>
</table>

Source: www.myfico.com

(a) Which variable do you believe is likely the explanatory variable and which is the response variable?

(b) Use the graph below to draw a scatter diagram of the data.
(c) Does credit score have a positive, negative, or no association with interest rate? Explain.

(d) The following JMP output is the result of fitting a least-squares regression line to the FICO data.

![Bivariate Normal Ellipse P=0.950](image)

Record the correlation between credit score and interest rate below. Interpret this value.

(e) Does a linear relation exist between credit score and interest rate? Explain.
(f) Use the JMP output to determine the regression equation and record it in the space below. Use appropriate notation and round the slope and the $y$-intercept to three decimal places.

(g) Interpret the $y$-intercept, if appropriate. If it is not appropriate, explain why not.

(h) Interpret the slope of the regression equation.

(i) According to the regression equation, what is the predicted interest rate for a person with a credit score of 595?

(j) According to the regression equation, what is the predicted interest rate for a person with a credit score of 700?

(k) Add the $(x, \hat{y})$ points for credit score and predicted interest rate for the credit scores of 595 and 700 to your scatter diagram. Draw a line connecting these two points. This line is the least-squares regression line.

(l) What is the residual for a person with a credit score of 595?

(m) On your scatter diagram, draw a line from the predicted interest rate for a person with a credit score of 595 to the actual interest rate for a person with a credit score of 595 to illustrate the residual.