Exam 1 will cover the material found in chapters 1-4 and 6, and in lectures. It will consist of short answer, calculation, and interpretation problems. The questions below represent the kind of question to expect on the exam, however there will be fewer questions on the exam because of time limitations. Please bring a calculator with you to the exam and Moore's green insert listing important formulas.

For calculation problems, please express your answers in complete sentences and make reference to the variables in the problem.

1. A teacher records the values of several variables for each student in her class. These include the variables listed below. Classify each variable below as quantitative or categorical.
   a. score on the final exam (out of 200 points) \textbf{quantitative}
   b. final grade for the course (A, B, C, D, F) \textbf{categorical}
   c. the number of classes the student missed \textbf{quantitative}

2. A sample of five recent births at a local hospital yielded birth weights (in ounces) of
   \[89 \quad 122 \quad 137 \quad 144 \quad 98\]
   a. Find the mean birth weight.

   \[
   \text{Mean birth weight} = \frac{\sum x}{n} = \frac{89+122+137+144+98}{5} = \frac{590}{5} = 118
   \]

   \text{The mean birth weight is 118 ounces.}

   b. Find the median birth weight.

   \[
   \text{Location of the median in an ordered distribution} = \frac{n+1}{2} = \frac{6}{2} = 3^{rd} \text{ case}
   \]

   \text{The 3^{rd} case is 122}

   \text{The median birth weight is 122 ounces.}

   c. The standard deviation is 23.95. Use the value of the standard deviation and a comparison of the mean and median birth weight to describe the shape of the distribution.

   \text{The variation in birth weights is quite large. On average, birth weights vary 23.95 ounces above and below the mean of 118 ounces (94.05 to 141.95 ounces). The smaller mean birth weight of 118 ounces, in comparison to the median of 122 ounces, shows that the distribution is negatively skewed because of the presence of the lower birth weights of 89 and 98 ounces. However, the difference of 4 ounces is not remarkable.}
3. The household income in a community is normally distributed with a mean of $42,000 and a standard deviation of $5,000. Find the proportion of households with incomes exceeding $38,000.

Given: N ($42,000, $5,000)

Step 1: Find Z when x = 38000.

\[ Z = \frac{x - \mu}{\sigma} = \frac{38000 - 42000}{5000} = -0.80 \]

Step 2: Use Table A to find the proportion for area under the normal curve above Z = - 0.80.

\[ \text{Area} = .2119 \]

Step 3: Because values in the body of the table are for areas to the left of Z scores, subtract .2119 from 1.0.

\[ \text{Proportion} = 1.0 - .2119 = .7881 \]

The proportion of households with incomes exceeding $38,000 is .7881.

4. A researcher wishes to study the relationship between years of education completed among GSS respondents (Y) and how many years of education their mothers completed (X). She finds that the relationship is linear and decides to fit a least-squares regression line. She computes the following values:

<table>
<thead>
<tr>
<th>Measure</th>
<th>Years of education completed by respondent’s mother</th>
<th>Years of education completed by respondent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>11.52</td>
<td>13.39</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>3.39</td>
<td>2.96</td>
</tr>
<tr>
<td>Correlation</td>
<td>+0.39</td>
<td>+0.39</td>
</tr>
</tbody>
</table>

a. Find the slope of the least-squares regression line.

\[ b = r \left( \frac{s_y}{s_x} \right) = .39 \left( \frac{2.96}{3.39} \right) = .34 \]

b. Find the constant of the least-squares regression line.

\[ a = \bar{Y} - b\bar{X} = 13.39 - .34(11.52) = 9.47 \]

c. Interpret the values of the constant and slope in the regression equation. Make reference to the units of measurement of each variable in your answer.

When the respondent’s mother has completed zero years of education, a respondent will have completed 9.47 years of education, and for every additional year of education completed by the respondent’s mother, the respondent will complete an additional .34 years of education.
d. Predict how many years of education someone will complete if his/her mother completed 16 years of education.

\[ \hat{Y} = a + bX = 9.47 + .34X = 9.47 + .34(16) = 9.47 + 5.44 = 14.91 \]

If the respondent’s mother completes 16 years of education, the respondent is predicted to complete 14.91 years of education.

e. How much variation in years of education completed is explained by how many years of education are completed by the mother?

\[ r^2 = \left( \hat{r} \right)^2 = (.39)^2 = .15 \]

The number of years of education completed by a respondent’s mother explains 15% of the variation in the respondent’s years of education completed.

5. Questionnaires were sent to 200 randomly selected businesses of different sizes, for a total of 600 questionnaires. The response rate was 41%.

<table>
<thead>
<tr>
<th>Company Size</th>
<th>Questionnaire Response</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Small</td>
<td>125</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td>(62.50%)</td>
<td>(37.50%)</td>
</tr>
<tr>
<td>Medium</td>
<td>81</td>
<td>119</td>
</tr>
<tr>
<td></td>
<td>(40.50%)</td>
<td>(59.50%)</td>
</tr>
<tr>
<td>Large</td>
<td>40</td>
<td>160</td>
</tr>
<tr>
<td></td>
<td>(20.00%)</td>
<td>(80.00%)</td>
</tr>
<tr>
<td></td>
<td>246</td>
<td>354</td>
</tr>
</tbody>
</table>

a. Did the size of a company affect the response rate to this survey? Use cell percentages to support your answer.

Yes, the size of a company did affect the response rate to the survey. Small companies were most likely and large companies were least likely to return the questionnaire. 62.5% of small companies compared to 40.5% of medium and 20.0% of large companies returned the questionnaire. Although an equal number of small, medium, and large companies received the survey, small companies accounted for about half of the companies that returned the survey (125/246).

b. What would be the best way to display this data graphically? Why?

A simple bar chart of companies that responded to the questionnaire by company size would work well. Another alternative is to display a bar chart with both categories of questionnaire response to highlight the differences by company size. A pie chart of the category “yes” would work too, but requires finding column percentages in the table above.
6. In order to determine if smoking causes cancer, researchers surveyed a large sample of adults. For each adult they recorded whether the person had smoked regularly at any period in his or her life and whether the person had cancer. They then compared the proportion of cancer cases among those who had smoked regularly at some time in their lives with the proportion of cases among those who had never smoked regularly at any point in their lives. The researchers found there were a higher proportion of cancer cases among those who had smoked regularly than among those who had never smoked regularly.

   a. What kind of study does this research exemplify and why?

   **This is an observational study. Researchers observed the behavior of individuals (smoking or not) and the purported consequence of that behavior (number of cancer cases). Researchers did not introduce any treatment and did not attempt to control the behavior of any of the study participants.**

   **NOTE: Understand the difference between and know how to identify an observational study and an experiment.**
7. I need a sample of 1,200 students for a campus study. I divide the population into men and women of different class years and then take simple random samples of those groups of students, resulting in the sample presented in the table below.

<table>
<thead>
<tr>
<th>Class Year</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>First</td>
<td>160</td>
<td>165</td>
</tr>
<tr>
<td>Second</td>
<td>158</td>
<td>167</td>
</tr>
<tr>
<td>Third</td>
<td>130</td>
<td>145</td>
</tr>
<tr>
<td>Fourth</td>
<td>135</td>
<td>140</td>
</tr>
</tbody>
</table>

a. Identify the sampling technique. Support your answer.

This is a stratified sampling technique. The population is divided by gender and by class year. It is not clear if the size of the sampling strata are in proportion to the size of the strata in the population. If so, this is a proportionate stratified sampling technique. If not, it is a disproportionate stratified sampling technique.

8. Administrators at a large urban university wonder if commuter students are different from the general student body in terms of academic achievement. They gather a random sample of 200 commuter students and learn from the registrar that the mean GPA for all students is 2.00, but the standard deviation of the population has never been computed. They find that the mean GPA of commuter students is 2.58 (s = 1.23). Is the GPA of commuter students significantly higher than the mean of the population of all students? Set \( \alpha = .05 \).

\[
\begin{align*}
H_0: \mu &= 2.00 \\
H_a: \mu &> 2.00
\end{align*}
\]

\[
Z = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} = \frac{2.58 - 2.00}{\frac{1.23}{\sqrt{200}}} = 6.67
\]

\[Z^* = 1.645; \ Z = 6.67; \ p-value < .0002; \ \alpha = .05\]

\[Z > Z^*, \ \text{so reject the null hypothesis} \ – \ OR – \ p-value < \alpha, \ \text{so reject the null hypothesis.}\]

Decision: Reject the null hypothesis. The population mean GPA of commuter students is significantly higher than 2.00, the GPA of the general student population.

9. Use the data in question 8 above to construct a 99% confidence interval for the population mean GPA of commuter students.

If 99% confidence interval and \( Z^* = 2.576 \).

\[
c.i. = \bar{X} \pm Z^* \left( \frac{s}{\sqrt{n}} \right) = 2.58 \pm 2.576 \left( \frac{1.23}{\sqrt{200}} \right) = 2.58 \pm .22 = 2.36 \text{ to } 2.80
\]

We are 99% confident that the population mean GPA of commuter students is between 2.36 and 2.80.