Test 1 Information

Time, Location, Coverage

Test 1 will be given in class on Wednesday, March 2. The test covers chapters 1–3 and 5–7 except for sections 6.6 and 7.3, which are about confidence intervals. You are responsible for material covered in the text, in the problem sets, and in class.

Format

Test questions will be designed to try to see how well you understand the material, not how well you can perform various procedures mindlessly.

A variety of question formats may be used. You may be required to compute numerical statistics; produce graphs by hand or with the computer (you will not need to print anything, however); or to analyze data or numerical or graphical summaries of data. Some items may be be tested using "short answers" (a couple sentences to a paragraph), multiple choice, or true/false.

You will be allowed to use your calculator or RStudio (but no other computer software or web sites) during the exam.

Instructions

Read through these prior to coming to the test and follow them when you take your test.

1. Always show your work and explain your reasoning. Answers without work or reasoning will not receive full credit.
   - Use mathematical notation (especially the equals sign) correctly.
   - Don’t be afraid to use words in your explanations.
   - If you get an unreasonable answer, be sure to say so. Give a brief explanation about how you know your answer is wrong (for example, the mean I calculated is less than 10, but I can see from the data that it should be at least 20). Then go on to other problems and come back and try to fix the error if you have time at the end of the test period.
   - Even if you cannot do a problem completely, show me what you do know.

2. You may use your calculator or RStudio during the exam, but for each number you write on the exam, it must be clear where it came from. For example, if you got .25 by multiplying .5 by .5, I want to see .5 · .5 = .25 on your paper (or words indicating the same).

3. Short answer questions will be graded based on truth, accuracy, clarity, significance, and brevity. In short, I’m looking for high quality answers. (Example: If you are asked to give an example of something, pick the best example you can think of, one that makes the issue especially clear.)

4. Test restrictions.
   - The test is closed book No notes are allowed.
   - Do not write in purple on the exam. (The exam will be graded in purple.)
Content

Here is a list of things you should be sure you know how to do. It is not intended to be an exhaustive list, but it is an important list.

You should be able to:

- Understand, use and explain the statistical vocabulary/terminology.
- Compute numerical summary statistics (mean, median, mode, range, standard deviation, quartiles, 5-number summary, correlation coefficient, etc.) and know what they tell you – and don’t tell you – about a data set.
- Understand how to make and interpret graphical representations of data (stemplot, histogram, boxplot, bar graph, mosaic plot) and when each might be appropriate or inappropriate to use.
- Use R to compute numerical summaries, make plots, conduct a binomial test, etc.
- Understand the issues involved in collecting good data and the design of studies, including the distinctions between observational studies and experiments.
- Make use of the rules of probability to determine probabilities of events.
- Compute and interpret conditional probabilities.
- Read, write, and understand random variable notation ($P(X \geq 3)$, etc.).
- Use the 4-step process for conducting a hypothesis test.
- Express the logic of a p-value in words.

Note that the test will be a sample from the possible topics; it is not possible to cover everything on the test.

Example Problems

The following problems come from old tests or quizzes I have given. They are provided here to give you some examples of types of questions I have asked in the past. The problems sets (including the extra problems) also provide a good source of problems to use when preparing for the exam.

1. Below are 5 boxplots and 5 histograms from 5 data sets. Match them up. (Boxplots and histograms were given.)

2. Each of the distributions below has the same scale. (Histograms were given.)
   a) Which has the largest standard deviation?
   b) Which has the smallest mean?

3. Compute the mean and standard deviation of the following small data set: 1, 2, 4, 5, 8. Do this "by hand" **showing all of your work** and using a calculator only to perform arithmetic operations like addition, subtraction, multiplication, division, square roots, etc. That is, do not use any statistical functions on your calculator or RStudio.
4. What is bias? Give one example (you can make it up) of a situation where poor statistical
design could lead to bias and explain what could be done to eliminate or reduce the bias in that
situation. [Similar types of questions could be asked about other important words.]

5. Consider the probability chart below which shows the probabilities for sex and favorite color in
Mr. Ortez’s class.

<table>
<thead>
<tr>
<th></th>
<th>Red</th>
<th>Yellow</th>
<th>Blue</th>
</tr>
</thead>
<tbody>
<tr>
<td>boy</td>
<td>.15</td>
<td>.05</td>
<td>.25</td>
</tr>
<tr>
<td>girl</td>
<td>.10</td>
<td>.25</td>
<td>.20</td>
</tr>
</tbody>
</table>

Let $M$ be the event that a student is a boy (male), $F$ be the event that a student is a girl
(female), $R$ be the event that a student’s favorite color is red, $Y$ be the event that a student’s
favorite color is yellow, and $B$ be the event that a student’s favorite color is blue. For each of the
following, express the probability using probability notation and determine the numerical value.
If a student is selected at random from Mr. Ortez’s class:

a) What is the probability of selecting a boy?

b) What proportion of the girls chose blue as their favorite color?

c) What is the probability that the student is a girl, given that the favorite color is blue?