Test 2 will be given in class on Wednesday, April 6.

Material Covered

The test is cumulative, but will emphasize the recent material (Sections 6.1–12.2). Note that we omitted a few sections (8.6, 9.5, 11.5). 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. Some items may be be tested using "short answers" (a couple sentences to a paragraph), multiple choice, or true/false.

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. Test restrictions.
   - The test is closed book. No notes are allowed.
   - You may use RStudio (bring your own laptop) or your calculator. There may be portions of the test where you are not allowed to use technology.
   - 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. (Some examples: population, sample, parameter, statistic, sampling distribution, standard error, p-value, confidence interval, confidence level, critical value, significance level, z-score, observational study, experiment, paired design, independent samples design, etc.)

- Work with random variables. This includes:
  - Using the basic rules of probability to determine probabilities of events.
  - Computing the mean and standard deviation of a random variable and knowing what they tell you.
  - Interpreting area as probability in a graph of a distribution.
  - Recognizing situations that are described by binomial, normal and t distributions.
  - Being able to use rules for means and variances to determine the mean and variance of a more complicated random variable from means and variances of simpler random variables.

- Understand the issues involved in collecting good data and the design of studies, including
  - the distinctions between sample surveys, observational studies, and randomized experiments.
  - matching study designs with appropriate analysis methods.

- Work with normal, t, and binomial distributions. This includes being able to use the 68-95-99.7 Rule and/or technology to find percentages, z-scores, critical values, etc.

- Understand the basic framework for hypothesis testing and how to interpret p-values.

- Understand the basic framework for confidence intervals and how to interpret the confidence level.

- Perform and interpret all of the confidence intervals and hypothesis tests covered so far. (You should be able to do these using RStudio and “by hand”.)

- Be aware of the assumptions that must be true to make use of various statistical procedures and the degree to which the procedures are robust.

- Understand how to make and interpret graphical representations of data (stemplot, histogram, boxplot, pie chart, bar graph) and when each might be appropriate or inappropriate to use.

Note that the test will be a sample from the possible topics, it will not be exhaustive.
Example Problems

A number of extra problems have been assigned with each problem set. The following problem is one that I have used very frequently on tests.

1. What do I do? In each of the following situations, pretend you want to know some information and you are designing a statistical study to find out about it. Give the following THREE pieces of information for each: (i), what variables you would need to have in your data set (ii) whether they are categorical or quantitative, and (iii) what statistical procedure you would use to analyze the results.

Select your procedures from the following list: 1-proportion (a.k.a. binomial test), Chi-squared goodness of fit, Chi-squared for 2-way tables, 1-sample $t$, Paired $t$, 2-sample $t$, none of these.

Record your answers in the table. Part a) has been done as an example.

(a) You want to know if boys or girls score better on reading tests in Kent County grade schools.

Answer Table:

<table>
<thead>
<tr>
<th>Variable(s)</th>
<th>Procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>reading score on standardized test [quant]</td>
<td>2-sample $t$</td>
</tr>
<tr>
<td>gender (male or female) [cat]</td>
<td></td>
</tr>
</tbody>
</table>

Notes:

- Often more than one design is possible, so there may be multiple correct answers. But you should not choose a design that is clearly inferior to another design we have already studied.

- You should not choose “none of these” if there is a reasonable design that can be analyzed by a method we already know about. “None of these” should mean that none of the listed procedures will suffice.