Test 3 will be given in class on Monday, May 2.

**Material Covered**

The test is cumulative, but will emphasize the material in chapters 11 through 16.1. We omitted a few things along the way, including Sections 11.5, 12.5, 12.6, 12.7, and most of Sections 15.5 and 15.6. In addition, you should know in what situations linear regression is used (for the “You want to know” problem), but you will not need to know the details of linear regression for this test.

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**, but you may bring with you one **review sheet**, no larger than 8.5 × 5.5 inches (half sheet of paper).
   - You may use RStudio (bring your own laptop) or your calculator.
   - 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.
- Understand the issues involved in collecting good data and the design of studies, including
  - the distinctions between observational studies and randomized experiments.
  - how and when to use paired designs or blocking,
  - matching study designs with appropriate analysis methods.
- Work with normal, t, binomial, chi-squared, and F 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.
- Understand what power means and how it is used to determine sample sizes for studies as they are planned.
- Understand the issue of multiple comparisons and how adjustments are made to compensate for them (e.g., Tukey’s Honest Significant Differences)
- Perform and interpret all of the confidence intervals and hypothesis tests covered so far.
- Be aware of the assumptions for the various statistical procedures and the degree to which the procedures are robust.
- Know what steps can be taken when the assumptions of a statistical procedure are violated.
- Understand how to make and interpret graphical representations of data (stemplot, histogram, boxplot, normal-quantile plot) and when each might be appropriate or inappropriate to use.
- Know how to interpret a correlation coefficient.

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. I have also posted more extra problems on the test information page. Solutions are included for these problems as well.

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-sample $t$, Paired $t$, 2-sample $t$, 1-proportion, 2-proportion, chi-squared goodness of fit, chi-squared for 2-way tables, simple linear regression, ANOVA, linear regression, 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.

<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.