**introductory biostatISTICS:**

**Analysis of Variance (ANOVA)**

**Goals**

• Learn how to use ANOVA to assess group differences in means.

• Learn how to compute subgroup means and how to perform multiple comparisons of means.

• Interpret the results of ANOVA.

**Reading Assignment**

• Read Chapter 12 in the textbook*.*

**Introduction**

So far in this course we have been examining different ways to compare two groups of subjects on the distributions of variables of interest. The T-test can be used to test the null hypothesis that 2 groups have the same mean value of some variable. Analysis of variance (ANOVA) is another useful tool, and it has many advantages over the 2-sample T test. ANOVA can be used to generate a single P-value to test the null hypothesis that 3 or more groups have the same mean value of the study variable (as opposed to doing multiple T-tests, each with its own significance test at the .05 level, among the pairs of groups means, leading to an increased "risk" of finding a significant result purely by chance). The collection of variables that go into an ANOVA is referred to as the "model": there is a single dependent variable (also called the outcome variable) and one or more independent variables. ANOVA can generate adjusted group means that take into account the influence of multiple factors in the model.

It is always important to think about the underlying assumptions of statistical tests. For ANOVA, there must be independent observations, normal distributions in the continuous variables, and homogeneity of variance among the groups. When one or more of these assumptions is not met, one can turn to alternative statistics that do not have the same restricting assumption, or "transform" non-normal variables into more normal approximations. The usual data exploration techniques (data plots, means, medians, and standard errors, for example) can be used to examine certain of these assumptions.

**Homework**

**PART (A):** A study in the city of Baltimore examined blood lead levels of children between the ages of 2 and 4 in relation to the location of the child's residential neighborhood. Due to past deposition of inorganic lead from leaded gasoline into the environment, neighborhoods near major roads and highways ("A" neighborhoods) were hypothesized to have somewhat higher levels of lead in soil and houses, compared to neighborhoods located between 1 and 5 miles away from the highways ("B" neighborhoods) and neighborhoods located more than 5 miles away from the highway ("C" neighborhoods). Blood lead levels are a marker of environmental exposure and body burden of this toxic metal. There is some suggestion in the literature of gender differences in blood lead levels and the investigators want to explore this relationship in their data.

**Your assignment is to use one-way ANOVA models to separately evaluate if there are gender differences on blood lead levels and also if there are neighborhood differences regarding blood lead levels.**

You should:

- run a one-way ANOVA model to test if there are neighborhood differences in blood lead levels;

- run a one-way ANOVA model to test if there are gender differences in blood lead levels; do the same using the t-test and compare the results;

            - answer the questions on the worksheet.

**Data**

The variables are neighborhood (1=A, 2=B, 3=C, as defined above), gender (1=male, 2=female), and blood lead (micrograms per deciliter of whole blood). Import the Excel file named Lead.xls into STATA.

**PART (B):** A clinical trial examined the efficacy of 3 types of medications in lowering total blood cholesterol levels. All participants in the trial were adults with high (>250 mg/dL) cholesterol levels, previously untreated. There was no placebo group, as it was felt that all subjects should have some chance to receive an effective medication to treat their condition. The cholesterol variable is the final measurement taken at the end of the trial. **Your assignment is to perform ANOVA to evaluate which medication is most efficacious for blood cholesterol control (assuming that lower levels are best!).**

- run a one-way ANOVA model to evaluate the effects of medications;

- compute the mean cholesterol levels;

- determine which medication was best, using the Bonferroni method for the one-way ANOVA model;

- answer the questions on the worksheet.

**Data**

The variables are medication (1, 2, or 3), age, body mass index, and cholesterol. Import the Excel file named Cholesterol.xls into STATA.

**HOMEWORK WORKSHEET – ANOVA**

Name\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# PART A {Blood Lead Data}

1. What are the assumptions for this model, and were they true?

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

2. What is the overall F-test result from the one-way ANOVA model for neighborhood? F=\_\_\_\_\_, P=\_\_\_\_\_. Interpretation:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

3. What is the overall F-test result from the one-way ANOVA model for gender? F =\_\_\_\_\_, P=\_\_\_\_\_.

Interpretation: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

4. What is the t-test result for gender? T statistic =\_\_\_\_\_, P=\_\_\_\_\_.

Compare this result with the result from the one-way ANOVA model: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

## PART B - {Cholesterol Drug Trial Data}

1. What is the SS for treatment in the medication one-way ANOVA model? \_\_\_\_ Degrees of freedom for treatment? \_\_\_\_\_\_\_ What is the SS error? \_\_\_\_\_\_\_\_

2. Report and interpret the overall F-test from the one-way ANOVA model for medication.

F=\_\_\_\_\_\_\_ P=\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

3. Provide the descriptive statistics below.

Medication mean cholesterol standard deviation cholesterol

1

2

3

4. Interpret the results of the Bonferroni method for the three medications. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_