

# BIOS 600 · Computer Skills and Data Analysis

8 November 2011

**Submission Instructions:** You will submit three items:

1. A paper copy of your final product,
2. An electronic copy of your final product (in pdf or word format),
3. A single electronic file with well annotated computer code (a .sas, .R, etc. file).

You should submit the paper copy in class by 10:50am on Tuesday, 22 November 2011. You should submit the electronic copy and the single electronic file to [tgs@email.unc.edu](mailto:tgs@email.unc.edu) by the end of the day Tuesday, 22 November 2011.

You must submit all three items.

**Honor Code:** You are allowed to use your notes, reference texts, or the internet when completing this assignment. You are not permitted to collaborate with classmates. You are not permitted to seek help from individuals other than the instructor or TAs.

**Project Goal:** BIOS 600 introduces students to a wide variety of statistics concepts and data analysis methods. Your final product is a self-written reference manual for many of these concepts and methods. If done well, this manual can be a valuable resource in the classes and projects that follow BIOS 600.

**Instructions:** Complete the reference-manual; a skeleton version is posted on Blackboard. In the statistical concepts section, provide a brief summary of the concept. The skeleton version includes a number of prompts written in red. Feel free delete the prompts. In the code section, provide the code snippets needed to complete the example. In the output section, include the parts of the output relevant to the specific example. If it is not relevant, do not include it.

Use the dataset `example.txt` when creating examples. Some of the variables in `example.txt` come from Stefanski (2007). The linear regression entry is an example, and does not need to be included in the final product.

See the data dictionary for a description of the variables.

**Format:** Every page should start with the entry name. Every entry should start on a new page. Computer code and output should be typeset in a fixed width font like courier new. See the linear regression entry for an example.

You are encouraged to use any other formatting conventions that will aid you in the future.

**Computer Code File:** The computer code file should contain all the code needed to produce the examples in your reference manual. Other than changing the path location of `example.txt`, the code should run as is.

For your own benefit, annotate the code so that you can refer to it at a later date.

**Data Dictionary** The dataset `example.txt` includes 8 variables. They are all fictitious.

Variable	Description
OBS	Observation Number
X	HbA1c blood concentration (recorded as a percent)
Y	Average number of hours spent exercising per day
H	A health score, ranges from 0 to 30
G	Gender, coded as follows: 1 = Males 0 = Females
P	Political Party, coded as follows: D = Democrat R = Republican U = Unaffiliated O = Other
T	Time period
R	Income (in \$10 K)

**References** Stefanski, L. A. (2007), "Residual (Sur)Realism," *The American Statistician*, 61, 163177.