

Biostatistics 510
Final Project Winter 2006
Due Friday, April 21st 12:00 noon
(note changes in red in the document)

Instructions:

For the final project, you are asked to use both SAS and SPSS to answer questions about data taken from the Tecumseh Community Health Project (see the data description below). Your final project is worth a total of 300 points, which is 25% of your grade for the course. The point breakdown is shown below.

What to include:

Please include the following items in your project, in the order shown below:

1. **Title Page** (10 points).
 - Include a separate sheet with your name, e-mail address, and the date.
2. **Printout of command files** (100 points: 50 points each for the SAS and SPSS command file printouts).
 - Include *all* commands used for SAS and SPSS to get the results for your project.
 - Use a font size 10 or greater for the printouts of the command files.
 - These command files should immediately follow the title page.
3. **Printed Summary of results (Project Summary)** (150 points).
 - Include any requested tables, and an *interpretation* of the results for each question, plus any discussion that you feel would be helpful.
 - Make sure your responses are **complete and concise**.
 - Include any computer output that is necessary to interpret your results, or that is requested, along with your discussion of each question.
 - Use at least font size 12 for the written portions of your project summary writeup.
 - Place the project summary **after** your command file printouts.
 - Please make sure that any computer output is well-integrated into your project summary. Do not hand in output separately from your project summary.**
 - Make sure the page-breaks are in sensible places, so that all output is clearly visible.
 - You may have to adjust the margins of your document, or modify the size of some of the computer output, to make the output appear clear.
 - Use a font size of 8 for your SAS output.
4. **A diskette or CD containing your command files in SAS and SPSS, plus your raw data file.** (40 points: 10 points for the SAS command file, 10 points for the SPSS command file, 10 points for the raw data file, and 10 points for the commands running properly using each program).
 - Label the diskette or CD with your name and your unique name printed clearly.
 - Name your SAS command file with your unique name .sas and your spss command file with your unique name .sps.
 - For example, if your unique name were jansmith, your SAS commands would be jansmith.sas and your SPSS commands would be jansmith.sps.

- Include your raw data file (e.g. jansmith.dat) on the diskette.
- Be sure that both command files run without error in both programs.
 - Make sure your command files are set up to work properly with the raw data file on your diskette or CD
 - Your commands will be run on your data file to be sure they work properly.

How to write up your results:

The final project should be written up as if you were reporting to a knowledgeable investigator. Please explain what your results mean. Do not simply report the results without comment.

Tests of significance:

When you report test results, please specify the value of the test statistic, any degrees of freedom (if appropriate), and the p-value (significance) of the test. Please use $\alpha=0.05$ for the statistical tests for this project.

Grading:

Points for the project will be awarded based on the points listed by each component. Grading will involve both the technical component (i.e. use of the computer programs) and the substantive component (i.e. interpretation of results).

Getting help:

You must work independently and may not consult with classmates. If you have problems understanding the assignment or getting the programs to run, check with me or with one of the GSIs.

You can reach me at CSCAR most of the time. The phone number at CSCAR is 734-764-7828 (extension 0). Speak to the person on desk and tell them that you are a Biostat 510 student. They will let me know that I have a call. Please don't leave messages on my phone. You can reach me pretty effectively with e-mail (kwelch@umich.edu).

Handing in your project:

Projects are due by 4:00 p.m. on Tuesday, April 18th, 2006. Please turn in your project at Melissa Karby and Rhiannon Popa's office in the Biostatistics Department (Biostat department office, Room 4210, 4th floor SPH II).

Email messages regarding the project:

Email messages with clarification, or answering questions about the project will be sent to the bios510 e-mail group. You may send email questions to me (kwelch@umich.edu) or to the Biostat 510 instructors' email group (bios510inst@umich.edu).

Office hours:

I will be available during my regular office hours on April 5th, 10th, 12th and 17th.

Course Lectures:

There will be regular class lectures on April 6th, April 11th, April 12th, and April 18th.

Course Evaluations:

Course evaluations will be filled out at the end of class on the last day of class, April 18th. Please come to class that day to get your doughnut, and to fill out the course evaluation forms.

Data:

The data are on my web page at <http://www.umich.edu/~kwelch>. You will have your own randomly selected data set, with 800 cases, named with your unquename. **For example, if your unquename were jansmith, your data set would be called jansmith.dat.**

Downloading your data:

You should download your data set and be sure that you can open it and use it immediately. Let me know right away if your data set is not on my web page, or if you have difficulty getting it.

Note that there are periods (.) in the raw data to indicate missing values. This will cause SPSS to give messages in the output window, which can be ignored. The data will be read in properly by SPSS, in spite of the apparent problems in the output.

Description of the Tecumseh Community Health Study:

The Tecumseh Community Health Study is a longitudinal, ongoing prospective epidemiologic study of the health status of residents in the city of Tecumseh, Michigan, collected by researchers at the University of Michigan. Data were collected for a period of over 30 years, beginning with a 1957 canvass of all households and adding all newcomers to the community until 1970, after which only previous residents were followed.

The complete data, from which these data files were excerpted, are available through ICPSR at: <http://www.icpsr.umich.edu:8080/ICPSR-STUDY/08969.xml>. Data for Round I (CV I) were collected in 1959 and 1960, data for Round II (CV II) were collected between 1962 and 1965, and data for Round III (CV III) were collected between 1967 and 1969. The data for the final project include the 4397 participants who were between the ages of 20 and 69 years at CV I. We are using selected variables from round I and round III only. Note that all missing values in the data set are already set to the missing value code in SAS and to system-missing in SPSS.

Variable	Description	Cols.	Codes
ID	Case Number	1-4	
SEX	SEX	5	1=Male 2=Female
AGE1	Age at CV I	7-8	No Missing values for age in this data set.
MARITAL1	Marital Status at CVI	9	1=Married 2=Never Married 3=Widowed 4=Divorced 5=Separated
EDUC1	Education Level CV I	10	0=No Schooling