STAT:5400 Computing in Statistics
Fall 2016
Final Exam

- Run R and LaTeX on the Linux machines. You may not use your own laptop. You may run SAS either under Linux or on the Virtual Desktop.
- You are required to use Sweave or a similar platform for incorporating R code and output into a document.
- SAS output may be submitted either as part of your LaTeX document, or as a separate pdf, text, or html file produced directly from SAS. Your SAS code may be submitted as a separate text file.
- I will give you 3 percentage points extra credit if you use Linux SAS and put your SAS code and output into verbatim environments in your Sweave document.
- You may use notes, computer files, on-line resources; in short, anything except other people!

1 SAS

A dataset called `jaws.dat` is on the course web page under Datasets. Documentation is provided in `jaws.info`.

1. Read the dataset into SAS. Submit your code.
2. Display all records in which the jaw measurement at age 9.5 is greater than 52.5. You may use any appropriate SAS procedure to do this. Your output should look like this (your variable names may be different from mine). Submit your code and output.

age8	age85	age9	age95
52.5	53.2	53.3	53.7
51.2	53	54.3	54.5
49.8	50	50.3	52.7
48.1	50.8	52.3	54.4
48.5	49.2	53	55.5
52.1	52.8	53.7	55
50.7	51.7	52.7	53.3
53.3	54.6	55.1	55.3

3. Reformat the dataset so that (a) there is a new variable – an i.d. number for each boy; (b) there is one record for each measurement; and (c) the three variables are i.d., age of measurement, and measured value. Display the first 8 observations in the restructured dataset. Your output should look like this (again, different variable names are ok):

Obs	id	age	size
1	1	8.0	47.8
2	1	8.5	48.8

```
3     1    9.0    49.0
4     1    9.5    49.7
5     2    8.0    46.4
6     2    8.5    47.3
7     2    9.0    47.7
8     2    9.5    48.4
```

Submit your code and output.

4. Use an appropriate SAS procedure to get the following table from your restructured dataset. (If you weren't able to create the restructured dataset, you may read in the dataset called `jawtrans.dat` from the course web page and work with it.) Submit your code and output.

	size	
	N	Mean
age		
8	20	48.66
8.5	20	49.63
9	20	50.57
9.5	20	51.45

5. Now write a SAS macro to create a table of statistics of jaw size for each age. Your macro should allow the user to specify (a) which statistics they want the procedure to calculate and (b) labels for the two variables.

   Call your macro so as to produce the table below:

	Ramus height	
	Mean	StdDev
Age in Years		
8	48.66	2.52
8.5	49.63	2.54
9	50.57	2.63
9.5	51.45	2.73

Submit your code and output.

2 R and the jackknife

Read the information describing a dataset at `http://www.statsci.org/data/oz/kiama.html`. The data itself is at: `http://www.statsci.org/data/oz/kiama.txt`. (Although the web site says this is a tab-delimited file, it's actually just a single column of data so delimiting is irrelevant.)

The coefficient of variation is a measure of the spread of a distribution relative to its mean. The coefficient of variation is calculated from a set of data values as

$$CV = \frac{s}{\bar{x}} \times 100\%$$

where s is the sample standard deviation and $\bar{x}$ is the sample mean.

1. Read the Kiama blowhole data into R. Submit your code.
2. Write a function to compute the CV given a vector argument. Submit your function code, and the output from applying it to the Kiama blowhole data.
3. Write your own function to perform the jackknife. Your function should accept as its arguments
   - a vector of a data
   - a function definition.

   Your function should return a list containing
   - the result of applying the function to the original data
   - the jackknife estimate of bias in the original estimate
   - the jackknife estimate of standard error

   Submit your code.
4. Use your jackknife function to estimate the bias and standard error of your CV estimate from the previous question. Submit your code and its output.