

22S:166 Computing in Statistics  
Fall 2010  
Final Exam

You may submit your answers either as:

1. a plain text file:  
include all R and SAS code and the output requested, and  
put your name as the first line of the file.
2. for 5 percentage points extra credit, as an .Rnw file and PDF file:  
embed R code into the .Rnw file, and  
just copy and paste SAS code and output into a `verbatim` environment.

Run R (and possibly  $\text{\LaTeX}$ ) on the Linux machines. You may run SAS wherever is most convenient for you, including the Virtual Desktop.

1. 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.

- (a) Read the Kiama blowhole data into R. Submit your code.
  - (b) 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.
  - (c) Use the jackknife to estimate the bias and standard error of your  $CV$  estimate from the previous question. Submit your code and its output.
2. Under Datasets on the course web page, there are two files called `states.txt` and `regions.txt`. Read their associated documentation files (`states.info` and `regions.info`).
    - (a) Read the `states.txt` and `regions.txt` data into SAS. Submit your SAS data steps.
    - (b) Merge the two datasets into a single dataset that contains all the variables from `states` plus the region name from `regions`. Submit your SAS code.
    - (c) Write a SAS procedure to produce the table below. (Don't worry if your column widths aren't exactly the same as mine on this or the other table.) Submit your SAS code and output.

```
-----  
|           |           income           | |
|           |-----|  
|           |   N   |   Mean   |  
|-----+-----+-----|  
| regname   |           |           |  
|-----|           |           |  
|North Central |   12.00|   4611.08|
```

Northeast	9.00	4570.22
South	16.00	4011.94
West	13.00	4702.62
All	50.00	4435.80

- (d) Now convert your procedure into a SAS macro that can accept two arguments: the numeric variable(s) to be summarized, and the statistic(s) to be calculated. Submit your SAS code.
- (e) Call the macro in such a way that it produces the same table as above. Submit your code and output.
- (f) Call the macro again in such a way that it produces the table below. Submit your code and output.

	area			pop	
	N	Mean	Sum	N	Mean
regname					
North Central	12.00	62652.00	751824.00	12.00	4803.00
Northeast	9.00	18141.00	163269.00	9.00	5495.11
South	16.00	54605.13	873682.00	16.00	4208.13
West	13.00	134463.00	1748019.00	13.00	2915.31
All	50.00	70735.88	3536794.00	50.00	4246.42

	pop
	Sum
regname	
North Central	57636.00
Northeast	49456.00
South	67330.00
West	37899.00
All	212321.00