STAT:5400 Computing in Statistics
Fall 2017
Final Exam

- You may submit a text file containing your code (R and SAS) and R output. If you prefer to use LaTeX and Sweave, that is fine; you would submit your .Rnw and .pdf files.
- 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.
- SAS output may be submitted either as part of your LaTeX or text 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.
- You may use notes, computer files, on-line resources; in short, anything except other people!

1 SAS

A file called `soccho2002.dat` is available under Datasets on the public course web page. Use SAS to do the following. Submit all of your SAS code and output.

1. Read the file into a SAS dataset. In your data step, also create two new variables: `year` and `month` extracted from the date variable.
2. Print the first 12 observations in your dataset. The listing should look like this:

The SAS System

Obs	account	unitval	date	year	month
1	CREFsoci	90.1582	10/28/2001	2001	10
2	CREFsoci	89.0058	10/29/2001	2001	10
3	CREFsoci	88.2204	10/30/2001	2001	10
4	CREFsoci	88.4410	10/31/2001	2001	10
5	CREFsoci	89.5331	11/01/2001	2001	11
6	CREFsoci	89.4553	11/02/2001	2001	11
7	CREFsoci	89.4553	11/03/2001	2001	11
8	CREFsoci	89.4553	11/04/2001	2001	11
9	CREFsoci	90.2549	11/05/2001	2001	11
10	CREFsoci	91.1156	11/06/2001	2001	11
11	CREFsoci	91.2097	11/07/2001	2001	11
12	CREFsoci	91.1165	11/08/2001	2001	11

3. Use a SAS procedure to create the following table:

The SAS System

		unitval	
	N	Mean	StdDev
year	month		
2001	10	4	88.96	0.87
	11	30	91.09	0.75
	12	31	91.41	0.70
2002	1	31	91.70	0.68
	2	28	90.13	0.59
	3	31	92.29	0.52
	4	30	91.05	0.74
	5	31	90.02	0.76
	6	30	87.20	1.22
	7	31	81.68	2.46
	8	31	82.64	1.71
	9	30	81.34	1.55
	10	28	80.13	1.99

4. Now write a SAS macro to create similar tables. Your SAS macro should enable to user to specify arguments that control

   - the row definitions
   - which observations to include in the table
   - the title of the table (if any)

5. Call your macro three times to create the following tables.

Year = 2002

	unitval		
	N	Mean	StdDev
month			
1	31	91.70	0.68
2	28	90.13	0.59
3	31	92.29	0.52
4	30	91.05	0.74
5	31	90.02	0.76
6	30	87.20	1.22
7	31	81.68	2.46
8	31	82.64	1.71
9	30	81.34	1.55
10	28	80.13	1.99

2001-02 Academic Year

		unitval		
		N	Mean	StdDev
year	month			
2001	10	4	88.96	0.87
	11	30	91.09	0.75
	12	31	91.41	0.70
2002	1	31	91.70	0.68
	2	28	90.13	0.59
	3	31	92.29	0.52
	4	30	91.05	0.74
	5	31	90.02	0.76
	6	30	87.20	1.22
	7	31	81.68	2.46
	8	31	82.64	1.71
	9	30	81.34	1.55
	10	28	80.13	1.99

	unitval		
	N	Mean	StdDev
year			
2001	65	91.11	0.93
2002	301	86.85	4.80

2 R and the bootstrap

A file called `bootresults.RData` is available under Datasets on the public course web page. It was created with the R function `save` and contains partial results from running a nonparametric bootstrap. It contains a list called `bootresults` of two items: a scalar called `orig` (the statistic calculated from the real dataset) and a vector called `bootstats` (the statistic values from the bootstrap datasets).

Write your own R code to do the following:

1. Load the contents of `bootresults.RData` into the R workspace. You may download the file from the web and read it from your directory, read it directly from the web, or access it in my Datasets directory using this path: `/mnt/nfs/clasnetappvm/homepage/kcowles/Datasets`.
2. Compute the bootstrap estimate of bias in the original statistic.
3. Compute the bootstrap estimate of standard error of the original statistic.
4. Compute a 95% bootstrap percentile interval for the population parameter.

Include your R code and output. (Don't print the entire `bootresults` object.)