

STAT:5400 Computing in Statistics  
 Fall 2015  
 Final Exam Review

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`).

Note that these files may be read into SAS on the Linux system using the following paths.

```
'/mnt/nfs/netapp1/homepage/kcowles/Datasets/states.txt'  

'/mnt/nfs/netapp1/homepage/kcowles/Datasets/regions.txt'
```

1. Read the `states.txt` and `regions.txt` data into SAS. Submit your SAS data steps.
2. 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.
3. 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|  

-----
```

4. 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.
5. Call the macro in such a way that it produces the same table as above. Submit your code and output.
6. 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

7. Use `proc sql` to produce the following table. The variable `avgincome` is the mean income in the region.

regname	avgincome
North Central	4611.083
Northeast	4570.222
South	4011.938
West	4702.615

8. Would it work to use `proc transpose` to create a dataset with one observations for each region and containing the values the income variable for every state in the region? Why or why not?