

Statistical Methods and Computing, STAT:2010

Instructor: Cowles

Lab 2

Jan. 25, 2019

1 Copying data directly into the data step

We will work with the billionaire dataset again today. You may either download it and read it in from the file on disk, or you may copy it directly into your data step.

Here is the code for including the data in the data step.

```
data billion ;                * gives dataset a name for SAS ;
input wlth age region $ ;     * names the variables in each row ;
                              * $ after region identifies character vbl;

datalines ;
37 50 M
24 88 U
14 64 A
.
.
.
1 59 E
1 . E
1 . 0
;
run ;                          * end of data step;
```

2 Using SAS procedures to list and tabulate the dataset

Once the dataset is created, you may run SAS *procedures* to analyze it. To list the entire dataset:

```
proc print data = billion ;
run ;
```

To get a frequency distribution of the regions in which billionaires lived:

```
proc freq data = billion ;
tables region ;              * tables is a keyword; region is the name of
                              * the variable for which you want counts ;
run ;
```

The output is:

The FREQ Procedure				
region	Frequency	Percent	Cumulative Frequency	Cumulative Percent
A	38	16.31	38	16.31
E	80	34.33	118	50.64
M	22	9.44	140	60.09
O	29	12.45	169	72.53
U	64	27.47	233	100.00

3 Proc univariate: SAS workhorse of descriptive statistics

Use *proc univariate* for quantitative variables when you want the following:

- means
- medians
- quartiles
- 5-number summary
- stem plots (for small datasets) or histograms (large datasets)
- boxplots

```
proc univariate plot data = billion ;
var wlth ;
run ;
```

The output is:

The UNIVARIATE Procedure			
Variable: wlth			
Moments			
N	233	Sum Weights	233
Mean	2.68154506	Sum Observations	624.8
Std Deviation	3.31884032	Variance	11.0147011
Skewness	6.57544276	Kurtosis	56.9655987
Uncorrected SS	4230.84	Corrected SS	2555.41064
Coeff Variation	123.765972	Std Error Mean	0.21742446

Basic Statistical Measures			
Location		Variability	
Mean	2.681545	Std Deviation	3.31884
Median	1.800000	Variance	11.01470
Mode	1.000000	Range	36.00000
		Interquartile Range	1.70000

Tests for Location: Mu0=0

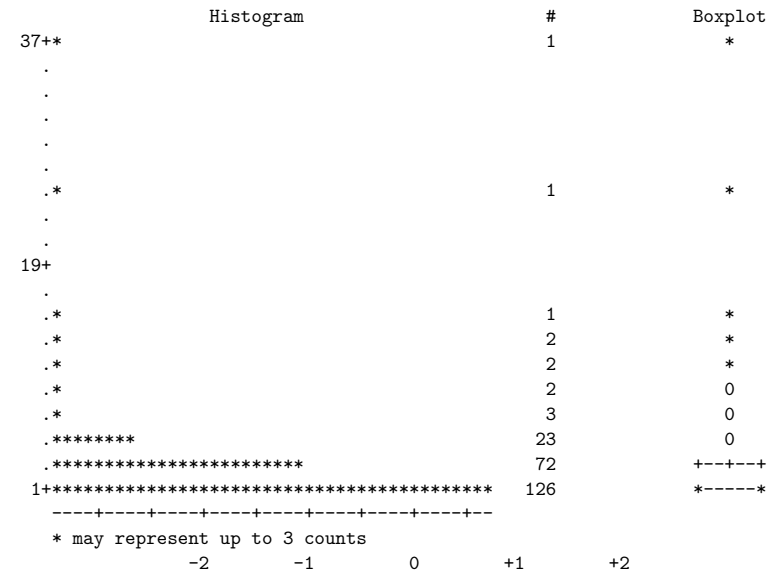
Test	-Statistic-	-----p Value-----		
Student's t	t 12.33323	Pr > t	<.0001	
Sign	M 116.5	Pr >= M	<.0001	
Signed Rank	S 13630.5	Pr >= S	<.0001	

Quantiles (Definition 5)

Quantile	Estimate
100% Max	37.0
99%	14.0
95%	6.2
90%	4.5
75% Q3	3.0
50% Median	1.8
25% Q1	1.3
10%	1.1
5%	1.0
1%	1.0
0% Min	1.0

Extreme Observations

----Lowest----		----Highest---	
Value	Obs	Value	Obs
1	233	13	4
1	232	13	5
1	231	14	3
1	230	24	2
1	229	37	1



4 Bar graphs and pie charts

```
goptions device = win ;
pattern v = solid color = gray ;

proc gchart data = billion ;
vbar region ;
title 'Billionaires in 1992; Regions ' ;
run ;
```

```
proc gchart data = billion ;
pie region ;
title 'Billionaires in 1992; Regions ' ;
run ;
```

5 Printing and Saving Files

Copying output from SAS windows into Microsoft Word will enable you to edit the SAS output and incorporate it into your homework writeups. You can then print from Word. When you highlight a block of text in the SAS output window in order to copy it, do not highlight all the way to the right margin of the last line. Due to a bug in SAS, that prevents the copy from working.

To save a file, click in the window whose contents you want to save. Go to the file menu and choose “Save as”. Navigate to where you wish to save the file. Your H: drive is a good choice, since SAS can see it from the Virtual Desktop. SAS will automatically give the file extension “.sas” to SAS commands and programs. For example, to name a SAS program “myprog,” you would type

myprog

in the box for the name of the file.

6 When you have finished...

Be sure to exit from SAS using the File menu, and to log out of the Virtual Desktop in your browser, and to log out of the computer using the icon on the desktop.