Exercise involving the Normal distribution and the sample distribution

L Collado-Torres

28 December 2016

In this exercise we will use data from the Normal distribution. It's the main distribution used in Statistics and has important properties such as symmetry around the mean and it's bell shape. This exercise shows this distribution by building upon the exploratory data analysis graphics and commands that we learned previously. Then we will use data to learn properties of the sample distribution and the Central Limit Theorem.

Variables in this data set

name	description
observation	An observation from the Normal distribution
size	The number of observations in the sample
variance	Variance of the Normal distribution
group	The group identifier for a given size and variance. There are 15 groups for each size and variance

Normal distribution

First lets learn about the Normal distribution using graphics from Exploratory Data Analysis. Complete the following tasks:

1. Load the file normal_2016.dta into Stata.
2. Which are the unique `variance` values?
3. How many observations do we have for each unique `variance`?
4. What is the sum of the unique `size` values?
5. Plot the distribution of the `observation` variable. Does it look bell shaped?
6. Plot the distribution of the `observation` variable separately for each value of the `variance` variable. Does these plots look different to the previous one? What properties do you observe?
7. Compare the distribution of the `observation` variable against the theoretical Normal distribution. What would you conclude from this `qnorm` plot?
8. Make a separate `qnorm` plot for only the observations with variance equal to 1 (or 5 or 10). Does it look closer to the theoretical Normal distribution?

Solutions

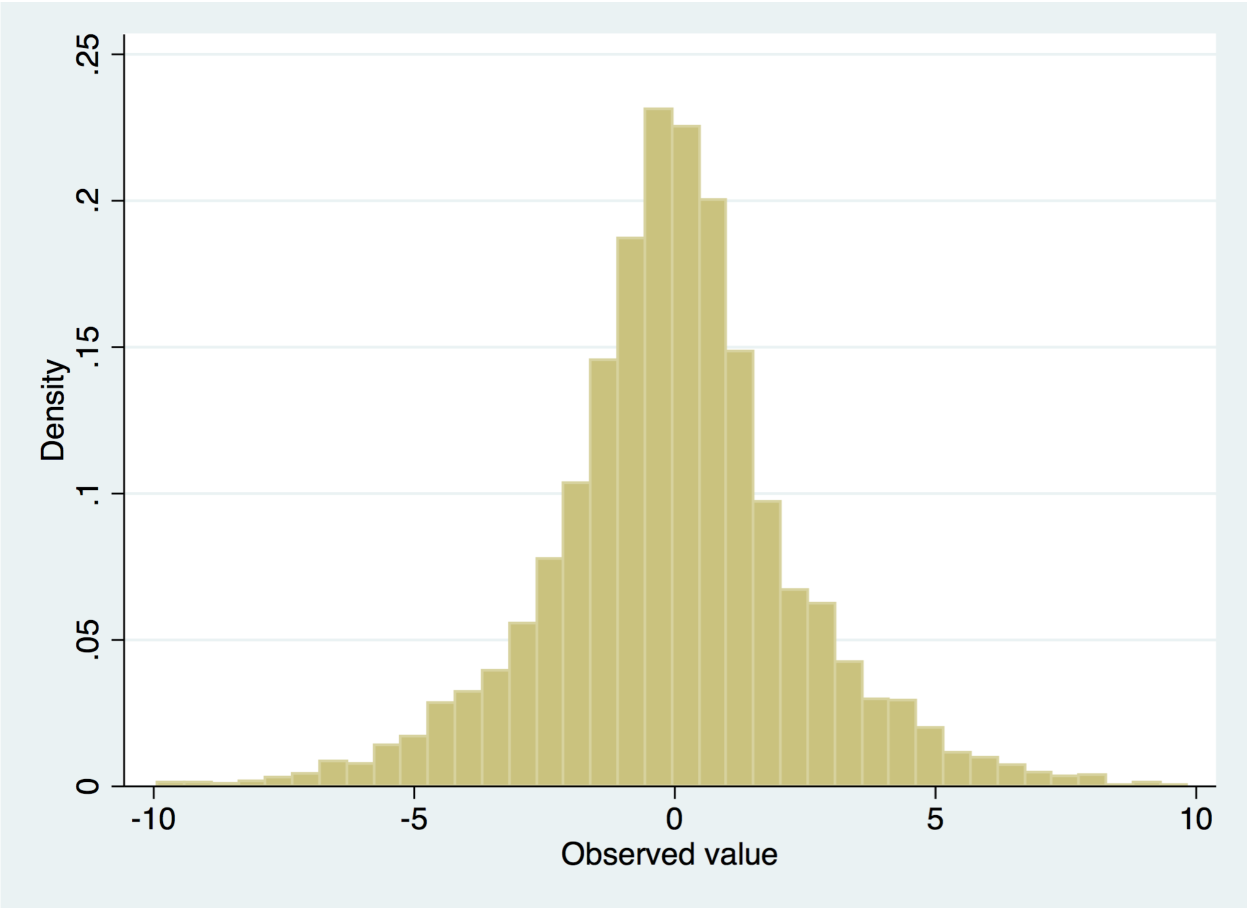
```r
library('knitr')
statapath <- '/Applications/Stata/Stata.app/Contents/MacOS/Stata'
```

```
* Load the data
use normal_2016.dta, clear

* Unique variance values
codebook variance

* Observations for each unique variance
codebook variance

* Sum of the unique size values
codebook observation
* 5 + 15 + 25 + 100 = 145
* Note 145 * 15 = 2175

* Distribution of observation
histogram observation
graph export "hist1.png", replace
graph box observation
graph export "box1.png", replace

* Separately
histogram observation, by(variance)
graph export "hist2.png", replace
graph box observation, by(variance)
graph export "box2.png", replace

* Global qnorm
qnorm observation
graph export "qnorm1.png", replace

* For each variance
qnorm observation if variance == 1
graph export "qnorm-var1.png", replace
qnorm observation if variance == 5
graph export "qnorm-var5.png", replace
qnorm observation if variance == 10
graph export "qnorm-var10.png", replace
```

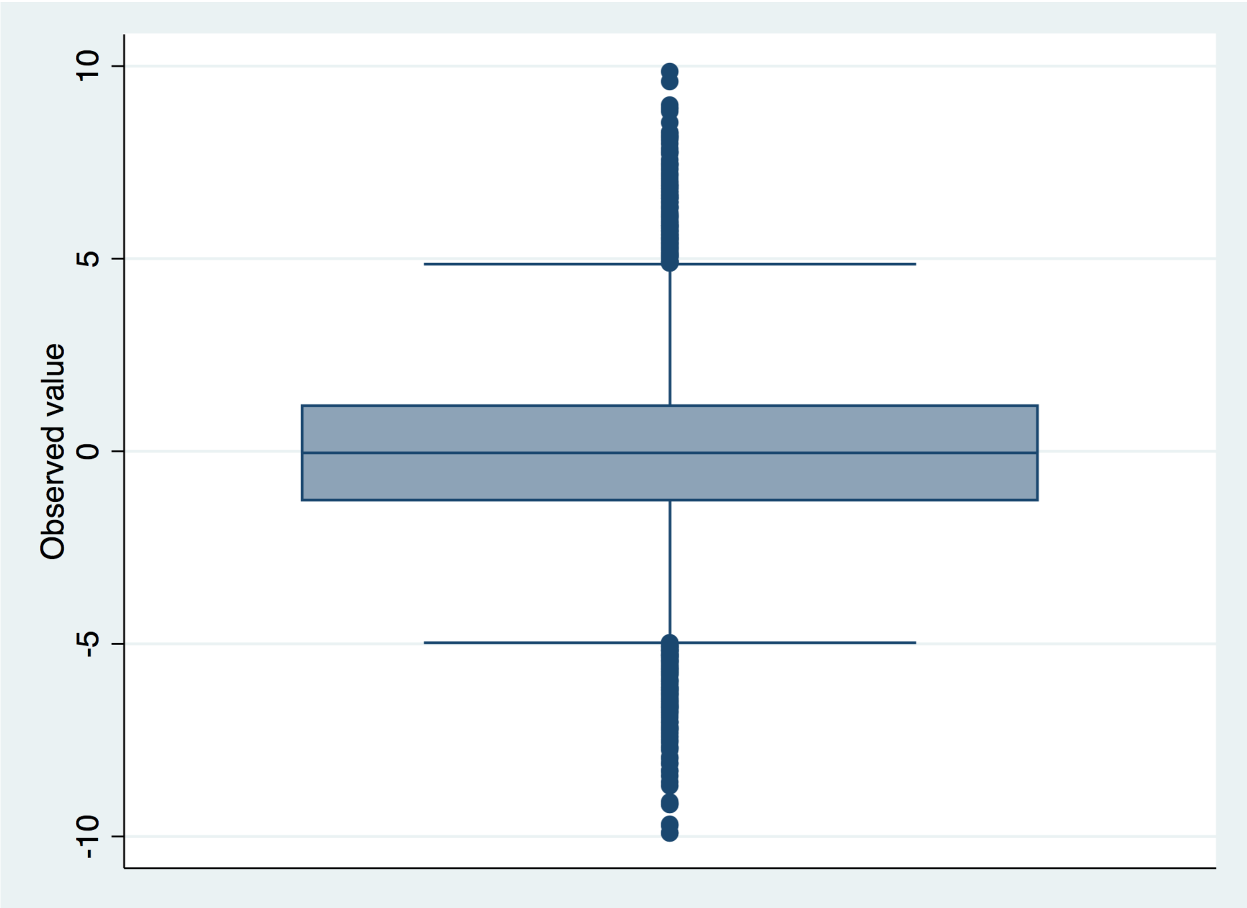
```
. * Load the data
. use normal_2016.dta, clear
(Written by R.              )

.
. * Unique variance values
. codebook variance

-------------------------------------------------------------------------------
```

```
variance                     Variance of the normal distribution for the group
-------------------------------------------------------------------------------

                  type:  numeric (double)

                 range:  [1,10]                       units:  1
         unique values:  3                        missing .:  0/6,525

            tabulation:  Freq.  Value
                         2,175  1
                         2,175  5
                         2,175  10

.
. * Observations for each unique variance
. codebook variance

-------------------------------------------------------------------------------
variance                     Variance of the normal distribution for the group
-------------------------------------------------------------------------------

                  type:  numeric (double)

                 range:  [1,10]                       units:  1
         unique values:  3                        missing .:  0/6,525

            tabulation:  Freq.  Value
                         2,175  1
                         2,175  5
                         2,175  10

.
. * Sum of the unique size values
. codebook observation

-------------------------------------------------------------------------------
observation                                                      Observed value
-------------------------------------------------------------------------------

                  type:  numeric (double)

                 range:  [-9.9257767,9.8362535]       units:  1.000e-12
         unique values:  6,525                    missing .:  0/6,525

                  mean:  -.039081
              std. dev:   2.35833

           percentiles:        10%       25%       50%       75%       90%
                          -2.81559  -1.28552   -.04104   1.17471   2.80592

. * 5 + 15 + 25 + 100 = 145
. * Note 145 * 15 = 2175
.
. * Distribution of observation
```

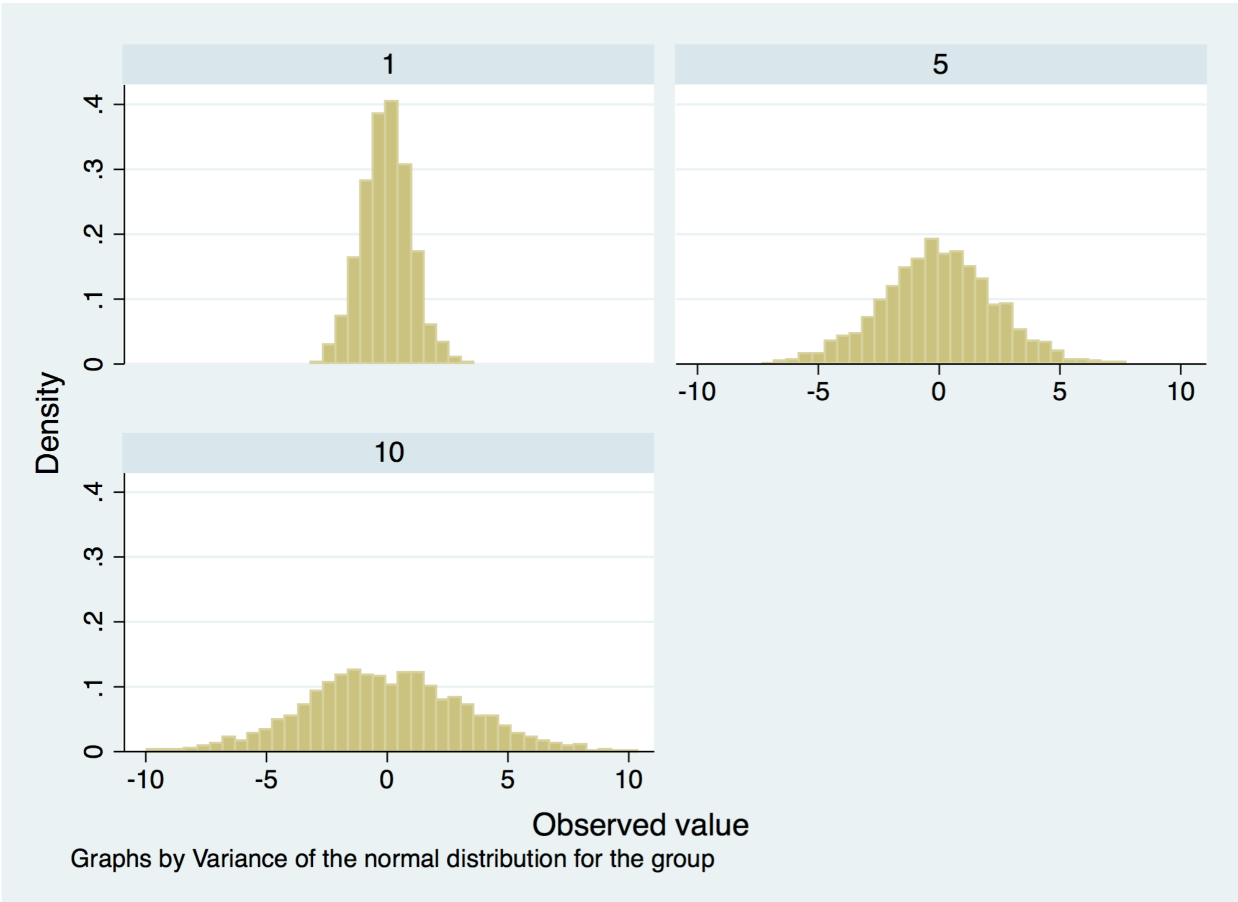
```
. histogram observation
(bin=38, start=-9.9257767, width=.52005342)

. graph export "hist1.png", replace
(file hist1.png written in PNG format)

. graph box observation

. graph export "box1.png", replace
(file box1.png written in PNG format)

.
. * Separately
. histogram observation, by(variance)

. graph export "hist2.png", replace
(file hist2.png written in PNG format)

. graph box observation, by(variance)

. graph export "box2.png", replace
(file box2.png written in PNG format)

.
. * Global qnorm
. qnorm observation

. graph export "qnorm1.png", replace
(file qnorm1.png written in PNG format)

.
. * For each variance
. qnorm observation if variance == 1

. graph export "qnorm-var1.png", replace
(file qnorm-var1.png written in PNG format)

. qnorm observation if variance == 5

. graph export "qnorm-var5.png", replace
(file qnorm-var5.png written in PNG format)

. qnorm observation if variance == 10

. graph export "qnorm-var10.png", replace
(file qnorm-var10.png written in PNG format)
```



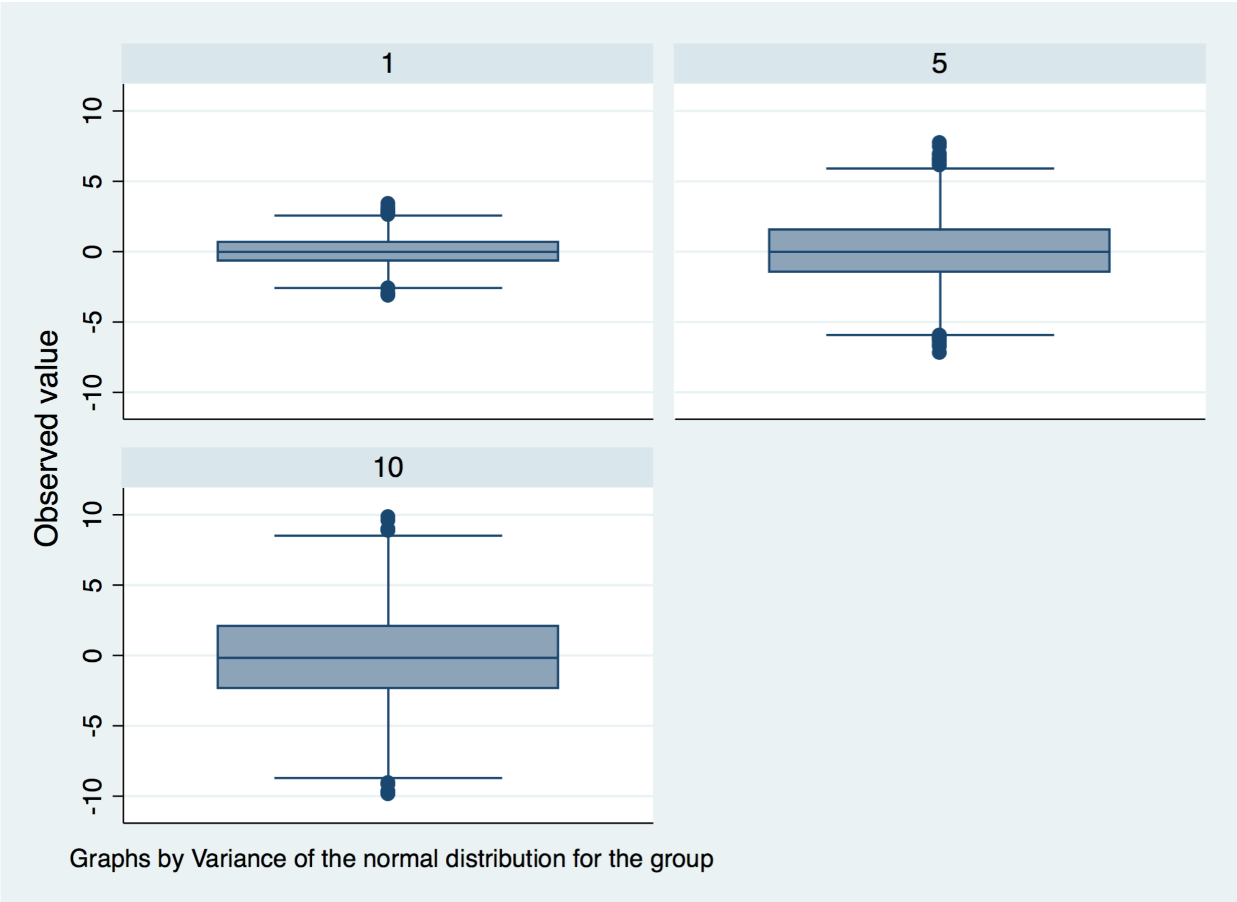
Density
0
.05
.1
.15
.2
.25
-10
-5
0
5
10
Observed value



10
5
0
-5
-10
Observed value



1
5
10
Density
0
.1
.2
.3
.4
-10
-5
0
5
10
Observed value
Graphs by Variance of the normal distribution for the group


1
5
10
Observed value
10
5
0
-5
-10
Graphs by Variance of the normal distribution for the group

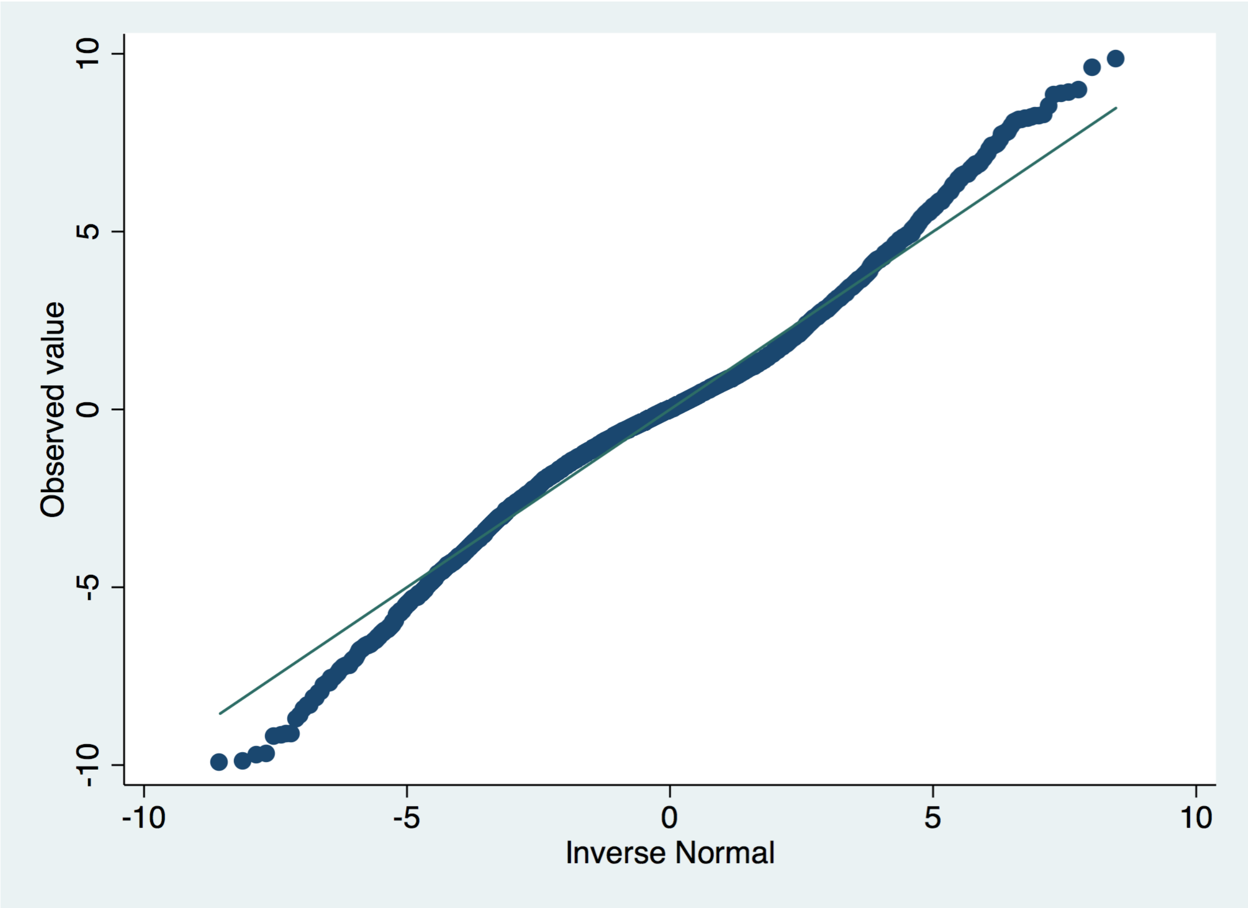
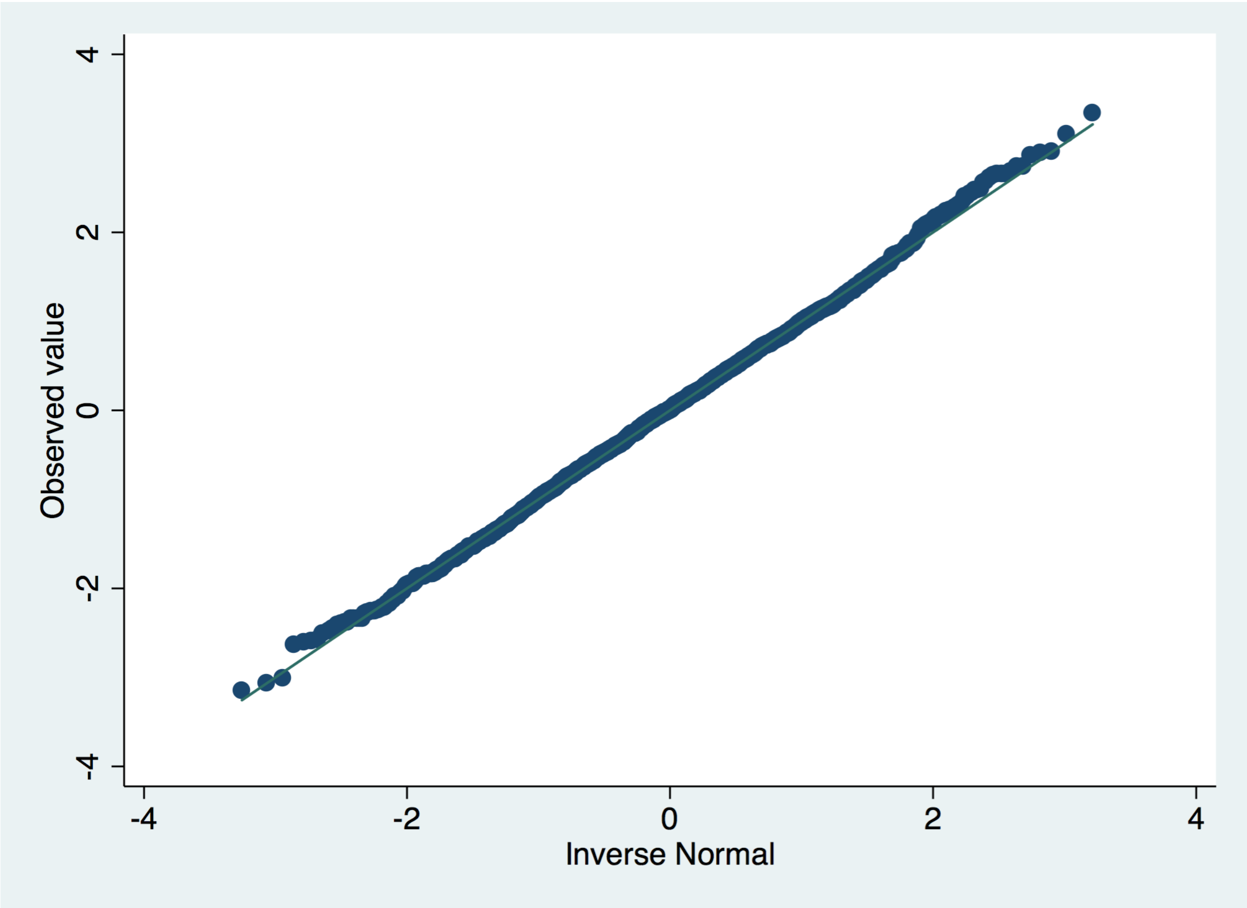
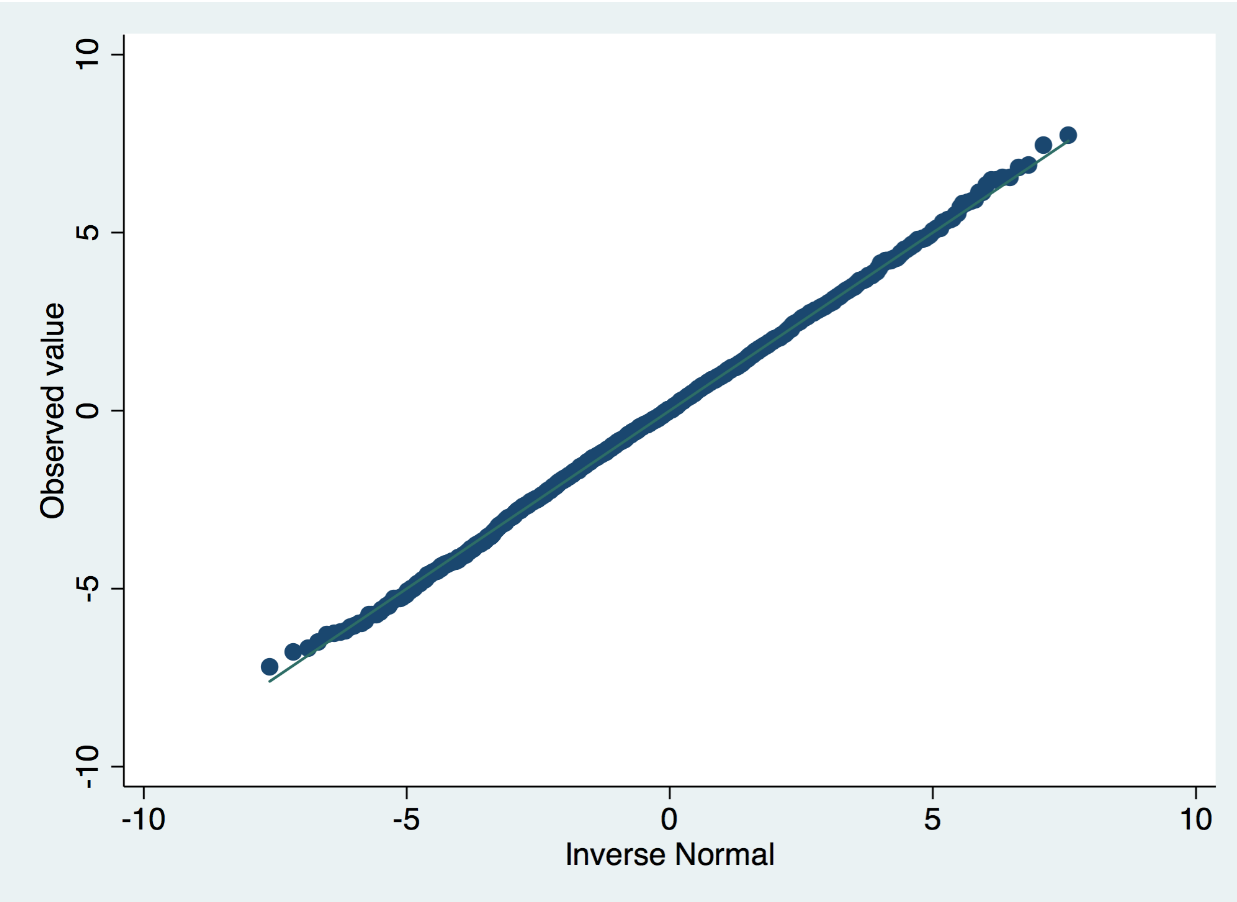
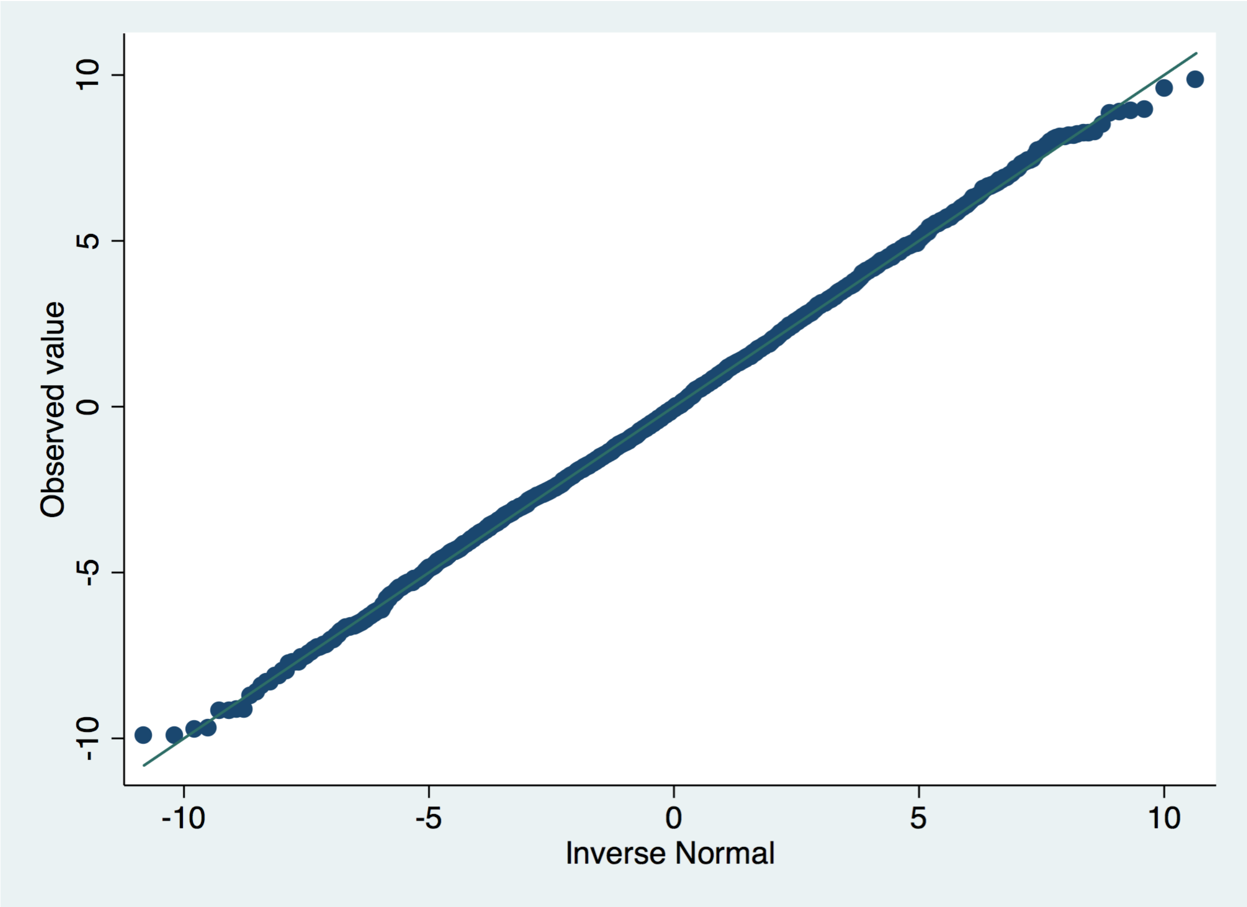


Figure 1: qnorm1



4
2
0
-2
-4
Observed value
-4
-2
0
2
4
Inverse Normal


Observed value
10
5
0
-5
-10
-10
-5
0
5
10
Inverse Normal



Sampling

1. Load the file normal_2016.dta into Stata.
2. What is the standard deviation of the `observation` variable?
3. What is the standard deviation of the `observation` variable when `variance` is equal to 1? Is it equal to 1? If not, how close is it to 1?
4. Repeat the same process for each unique value of the `group` variable when `variance` is equal to 1. Are the means equal to 0? Are the standard deviations equal to 1? What differences do you notice with respect to the previous bullet point?

To differentiate the properties of the sample distribution and the theoretical distribution, we say that:

- μ is the mean of the theoretical distribution
- σ is the standard deviation of the theoretical distribution
- $\bar{x}$ is the mean of the sample distribution, also known as the *sample mean*
- $s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}$ is the standard deviation of the sample distribution

Solutions

```
* Load the data
use normal_2016.dta, clear
```

```
* SD of observation
summarize observation

* SD of observation when variance is equal to 1
summarize observation if variance == 1

* SD of observation when variance is equal to 1 for each group
bys group: summarize observation if variance == 1
```

```
. * Load the data
. use normal_2016.dta, clear
(Written by R.              )

.
. * SD of observation
. summarize observation

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |      6,525   -.0390809    2.358333  -9.925777   9.836253

.
. * SD of observation when variance is equal to 1
. summarize observation if variance == 1

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |      2,175   -.0213126    .9758493  -3.145141   3.332773

.
. * SD of observation when variance is equal to 1 for each group
. bys group: summarize observation if variance == 1

----------------------------------------------------------------------------
-> group = 1

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |        145   -.0035318    .9033008  -2.488066   2.395979

----------------------------------------------------------------------------
-> group = 2

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |        145    .0815653    .9707447  -2.246286   2.489026

----------------------------------------------------------------------------
-> group = 3

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
```

```
 observation |       145   -.2412682    .9133215  -2.580764   2.640297

-------------------------------------------------------------------------------
-> group = 4

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
 observation |       145   -.0059926    .9485665   -2.37961   2.107302

-------------------------------------------------------------------------------
-> group = 5

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
 observation |       145   -.0226683    1.017992   -2.19802    2.89616

-------------------------------------------------------------------------------
-> group = 6

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
 observation |       145   -.0448871    .9687357  -2.587038   3.332773

-------------------------------------------------------------------------------
-> group = 7

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
 observation |       145    .0147168    .9272922  -1.965008   2.566479

-------------------------------------------------------------------------------
-> group = 8

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
 observation |       145    .1171013     .994055    -2.2517   2.859635

-------------------------------------------------------------------------------
-> group = 9

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
 observation |       145   -.0288215    1.034813  -3.061025   2.659089

-------------------------------------------------------------------------------
-> group = 10

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
 observation |       145    -.012625    .9142583  -2.630021   2.396376

-------------------------------------------------------------------------------
-> group = 11
```

```
    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |        145      -.1366    1.035556  -2.611554   2.467293

-----------------------------------------------------------------------------
-> group = 12

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |        145   -.0781324    .9741902  -2.253071   2.317058

-----------------------------------------------------------------------------
-> group = 13

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |        145    .0359654    .9722664  -2.275057   2.478941

-----------------------------------------------------------------------------
-> group = 14

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |        145    .1516833    .9738071  -2.513651   3.099097

-----------------------------------------------------------------------------
-> group = 15

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |        145    -.146194    1.045504  -3.145141   2.651.
end of do-file
```

Central Limit Theorem

In many applications of biostatistics we study the *sample mean* and the Central Limit Theorem (a theoretical result) provides us with a incredibly powerful tool. The CLT tells us that as the sample size n increases, the distribution of the sample mean $barx$ gets closer to the Normal distribution with mean μ and variance $\frac{s}{\sqrt{n}}$.

This sounds great, but lets see it in action. Complete the following tasks:

1. Load the file normal_2016.dta into Stata.
2. For each group of observations that have sample size equal to 5 and variance 10, calculate the mean. Open a text editor or Excel and write down the 15 sample means in column 1 (name it mean) and write 5 as the sample size in the second column (name it size).
3. Repeat this process for each group of observations that have sample size equal to 15 and variance 10. Add the means to the same table from the previous step.
4. Repeat this process for each group of observations that have sample size equal to 25 and variance 10.
5. Repeat this process for each group of observations that have sample size equal to 100 and variance 10.

That is fill out the following table.

mean	size
	5

mean	size
	5
	5
	5
	5
	5
	5
	5
	5
	5
	5
	5
	5
	5
	5
	15
	15
	15
	15
	15
	15
	15
	15
	15
	15
	15
	15
	15
	15
	15
	25
	25
	25
	25
	25
	25
	25
	25
	25
	25
	25
	25
	25
	25
	25
	100
	100
	100
	100
	100
	100
	100
	100

mean	size
	100
	100
	100
	100
	100
	100
	100

6. Save your new table and import it with Stata. Alternatively use sample_means_2016.txt.
7. Plot the distribution of the sample mean separately for each unique sample size (*size* variable). What happens to the distribution of the sample mean as you increase the sample size?
8. You can optionally repeat this process for the observations drawn from a Normal distribution with variance 1 or 5 (don't mix them!). Can be useful too if you have a large class and want to divide them into 3.

Solutions

```
* Load the data
use normal_2016.dta, clear

* Calculate the means but then register them manually
bys group: summarize observation if variance == 10 & size == 5
bys group: summarize observation if variance == 10 & size == 15
bys group: summarize observation if variance == 10 & size == 25
bys group: summarize observation if variance == 10 & size == 100

* Calculate the means with programatic code (this is called looping)
* and save the results to the sample_means_2016.txt file.
* For moe details to understand this code check
* http://www.ats.ucla.edu/stat/stata/faq/filewrite.htm
file open myresults using "sample_means_2016.txt", write replace
file write myresults "mean" _tab "size" _n
set more off
foreach samplesize of numlist 5 15 25 100 {
    foreach groupid of numlist 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 {
        summarize observation if variance == 10 & size == `samplesize' & group == `groupid'
        file write myresults (r(mean)) _tab "`samplesize'" _n
    }
}
file close myresults
set more on


* Load the sample means table
import delimited "sample_means_2016.txt", clear

* Make the graphics
histogram mean, by(size)
graph export "hist3.png", replace
graph box mean, by(size)
graph export "box3.png", replace
```

```
* Optional qnorm plots
qnorm mean if size == 5
graph export "qnorm_mean_size5.png", replace
qnorm mean if size == 15
graph export "qnorm_mean_size15.png", replace
qnorm mean if size == 25
graph export "qnorm_mean_size25.png", replace
qnorm mean if size == 100
graph export "qnorm_mean_size100.png", replace
```

```
. * Load the data
. use normal_2016.dta, clear
(Written by R.              )

.
. * Calculate the means but then register them manually
. bys group: summarize observation if variance == 10 & size == 5

-----------------------------------------------------------------------------
-> group = 1

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |          5    .2589072    1.425755  -1.648349   1.861652

-----------------------------------------------------------------------------
-> group = 2

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |          5   -1.026144    3.231694  -5.056384   2.442126

-----------------------------------------------------------------------------
-> group = 3

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |          5   -.1037298     2.94163  -4.149247   3.451327

-----------------------------------------------------------------------------
-> group = 4

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |          5    .9424025    3.402822  -4.129304   4.411133

-----------------------------------------------------------------------------
-> group = 5

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
```

```
 observation |          5      -.3513    2.728499  -2.991597   4.036663

-------------------------------------------------------------------------------
-> group = 6

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |          5   -.0983589    3.102556  -2.976197   3.934591

-------------------------------------------------------------------------------
-> group = 7

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |          5    -1.05307    1.599494  -2.660201   1.078001

-------------------------------------------------------------------------------
-> group = 8

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |          5   -3.633132    3.384315  -9.172001  -.0727315

-------------------------------------------------------------------------------
-> group = 9

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |          5    .6918162    3.217179  -2.668756   5.420033

-------------------------------------------------------------------------------
-> group = 10

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |          5    .4025495     4.19843  -6.270437   4.448832

-------------------------------------------------------------------------------
-> group = 11

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |          5   -1.625534    .9935249  -2.809964  -.0869543

-------------------------------------------------------------------------------
-> group = 12

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |          5   -.3757984    1.047461  -1.804874   .5804903

-------------------------------------------------------------------------------
-> group = 13
```

```
    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |          5    1.999273    2.863914  -1.778234     6.0721

-----------------------------------------------------------------------------
-> group = 14

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |          5    .5663508    2.005641   -1.64762   3.094349

-----------------------------------------------------------------------------
-> group = 15

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |          5    .8156639     3.47456  -2.588775    6.61. bys group: summarize observation

-----------------------------------------------------------------------------
-> group = 1

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |         15     .158506    3.694453  -5.336938    8.13184

-----------------------------------------------------------------------------
-> group = 2

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |         15    .3319717    3.830569  -9.122918   5.582776

-----------------------------------------------------------------------------
-> group = 3

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |         15    1.167319    3.629757  -6.140796   8.877068

-----------------------------------------------------------------------------
-> group = 4

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |         15    .0111664    2.893321  -4.243984   6.554693

-----------------------------------------------------------------------------
-> group = 5

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |         15   -1.030764     3.63934  -9.709027   4.823821

-----------------------------------------------------------------------------
```

```
-> group = 6

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
 observation |        15    .0140114    2.937627  -7.251023    4.63186

-------------------------------------------------------------------------------
-> group = 7

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
 observation |        15   -.2918965    2.854115  -6.517677   4.405031

-------------------------------------------------------------------------------
-> group = 8

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
 observation |        15     .111334    3.038959  -5.193361    4.62143

-------------------------------------------------------------------------------
-> group = 9

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
 observation |        15   -.1541863    2.896933  -4.935158   4.383737

-------------------------------------------------------------------------------
-> group = 10

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
 observation |        15   -.5667987    3.327492  -4.874509   4.482293

-------------------------------------------------------------------------------
-> group = 11

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
 observation |        15    .1455095    3.666581  -7.201359   6.894349

-------------------------------------------------------------------------------
-> group = 12

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
 observation |        15    .0024407    3.304709  -5.488784   7.140138

-------------------------------------------------------------------------------
-> group = 13

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
 observation |        15     .306066    3.929371  -6.626971   5.038344
```

```
-------------------------------------------------------------------------------
-> group = 14

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |         15    .0196183    3.002673  -5.330605   4.478324

-------------------------------------------------------------------------------
-> group = 15

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |         15    .3361137    3.932817  -9.918619    5.04154


. bys group: summarize observation if variance == 10 & size == 25

-------------------------------------------------------------------------------
-> group = 1

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |         25   -1.156035    3.230117  -7.048663   6.613952

-------------------------------------------------------------------------------
-> group = 2

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |         25   -.5435422    3.478905  -5.793712    6.98085

-------------------------------------------------------------------------------
-> group = 3

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |         25    .3093568    3.280309  -6.810264   7.966505

-------------------------------------------------------------------------------
-> group = 4

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |         25    .4389188    3.088024  -4.856911   7.297013

-------------------------------------------------------------------------------
-> group = 5

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |         25   -1.166001    3.340921  -9.188413   5.386401

-------------------------------------------------------------------------------
```

```
-> group = 6

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
 observation |        25    .5979134    3.313374  -6.395189   8.264582

-------------------------------------------------------------------------------
-> group = 7

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
 observation |        25    .5478882    3.072026  -5.974334    6.27983

-------------------------------------------------------------------------------
-> group = 8

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
 observation |        25   -.1541815    4.551343  -8.713041   9.584764

-------------------------------------------------------------------------------
-> group = 9

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
 observation |        25     .886669    3.869363  -7.518301   6.874853

-------------------------------------------------------------------------------
-> group = 10

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
 observation |        25   -.9920868    2.826544  -6.322566    3.63698

-------------------------------------------------------------------------------
-> group = 11

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
 observation |        25    -.718968    3.872108   -9.71908   6.369649

-------------------------------------------------------------------------------
-> group = 12

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
 observation |        25   -.6490565    2.740573  -5.476446   4.895412

-------------------------------------------------------------------------------
-> group = 13

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
 observation |        25   -1.015911    3.490921  -6.750055   5.913665
```

```
------------------------------------------------------------------------------
-> group = 14

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
 observation |        25     .3622803    3.639054  -7.990776   7.393661

------------------------------------------------------------------------------
-> group = 15

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
 observation |        25   -.2420888     3.75885  -7.558431   7.056714


. bys group: summarize observation if variance == 10 & size == 100

------------------------------------------------------------------------------
-> group = 1

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
 observation |       100   -.4383494    3.147817   -8.32965   6.735005

------------------------------------------------------------------------------
-> group = 2

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
 observation |       100   -.2713836      2.8653   -8.11684   6.329414

------------------------------------------------------------------------------
-> group = 3

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
 observation |       100   -.1207133    3.082861  -7.415754   8.970666

------------------------------------------------------------------------------
-> group = 4

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
 observation |       100    .1647736    2.960644  -7.550699   8.227775

------------------------------------------------------------------------------
-> group = 5

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
 observation |       100   -.1129649    3.181103  -7.347415   6.563672

------------------------------------------------------------------------------
```

```
-> group = 6

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |        100   -.4950483    3.691278  -9.151102   9.836253

-----------------------------------------------------------------------------
-> group = 7

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |        100    .1835885     3.43671  -8.431155   8.177943

-----------------------------------------------------------------------------
-> group = 8

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |        100   -.2842181    3.109602  -6.662359   8.513903

-----------------------------------------------------------------------------
-> group = 9

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |        100   -.2147722    3.160636  -6.871425   7.787577

-----------------------------------------------------------------------------
-> group = 10

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |        100    .4189958    3.202346  -9.925777   6.724799

-----------------------------------------------------------------------------
-> group = 11

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |        100    .1742163    3.468608  -7.731831   8.916285

-----------------------------------------------------------------------------
-> group = 12

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |        100    .4067078    3.203779  -5.795245   8.824605

-----------------------------------------------------------------------------
-> group = 13

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |        100   -.3901959    2.845062  -8.121987   8.125437
```

```
--------------------------------------------------------------------------------
-> group = 14

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |        100   -.1594001     3.16435  -8.615282   7.327055

--------------------------------------------------------------------------------
-> group = 15

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |        100    .3011528    3.201704  -7.442526    7.15992


.
. * Calculate the means with programatic code (this is called looping)
. * and save the results to the sample_means_2016.txt file.
. * For moe details to understand this code check
. * http://www.ats.ucla.edu/stat/stata/faq/filewrite.htm
. file open myresults using "sample_means_2016.txt", write replace

. file write myresults "mean" _tab "size" _n

. set more off

. foreach samplesize of numlist 5 15 25 100 {
  2.     foreach groupid of numlist 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 {
  3.         summarize observation if variance == 10 & size == `samplesize' &
> group == `groupid'
  4.         file write myresults (r(mean)) _tab "`samplesize'" _n
  5.     }
  6. }

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |          5    .2589072    1.425755  -1.648349   1.861652

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |          5   -1.026144    3.231694  -5.056384   2.442126

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |          5   -.1037298     2.94163  -4.149247   3.451327

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |          5    .9424025    3.402822  -4.129304   4.411133

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |          5      -.3513    2.728499  -2.991597   4.036663
```

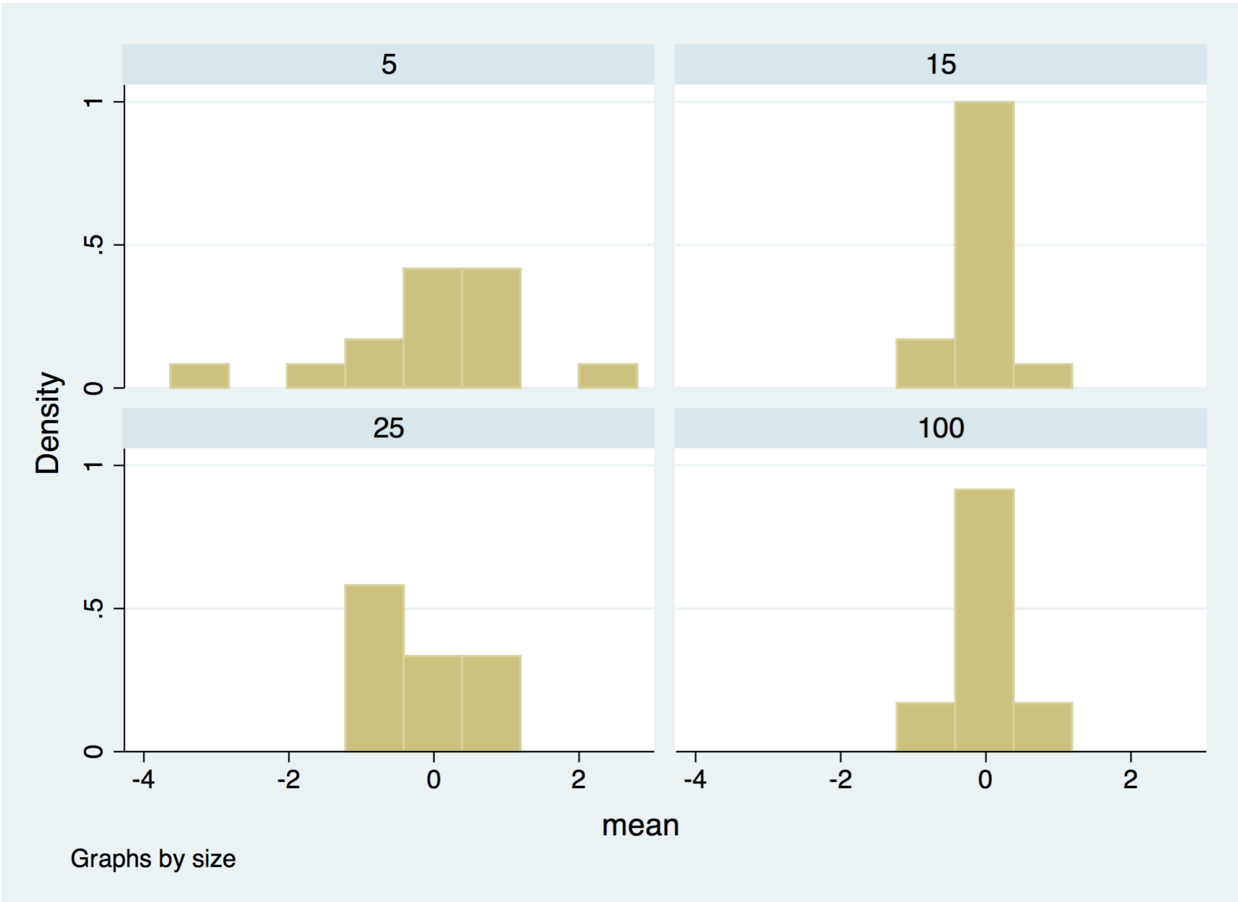
```
    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |          5   -.0983589    3.102556  -2.976197   3.934591

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |          5    -1.05307    1.599494  -2.660201   1.078001

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |          5   -3.633132    3.384315  -9.172001  -.0727315

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |          5    .6918162    3.217179  -2.668756   5.420033

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |          5    .4025495     4.19843  -6.270437   4.448832

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |          5   -1.625534    .9935249  -2.809964  -.0869543

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |          5   -.3757984    1.047461  -1.804874   .5804903

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |          5    1.999273    2.863914  -1.778234     6.0721

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |          5    .5663508    2.005641   -1.64762   3.094349

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |          5    .8156639     3.47456  -2.588775    6.61666

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |         15     .158506    3.694453  -5.336938    8.13184

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |         15    .3319717    3.830569  -9.122918   5.582776

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |         15    1.167319    3.629757  -6.140796   8.877068

    Variable |        Obs        Mean    Std. Dev.       Min        Max
```

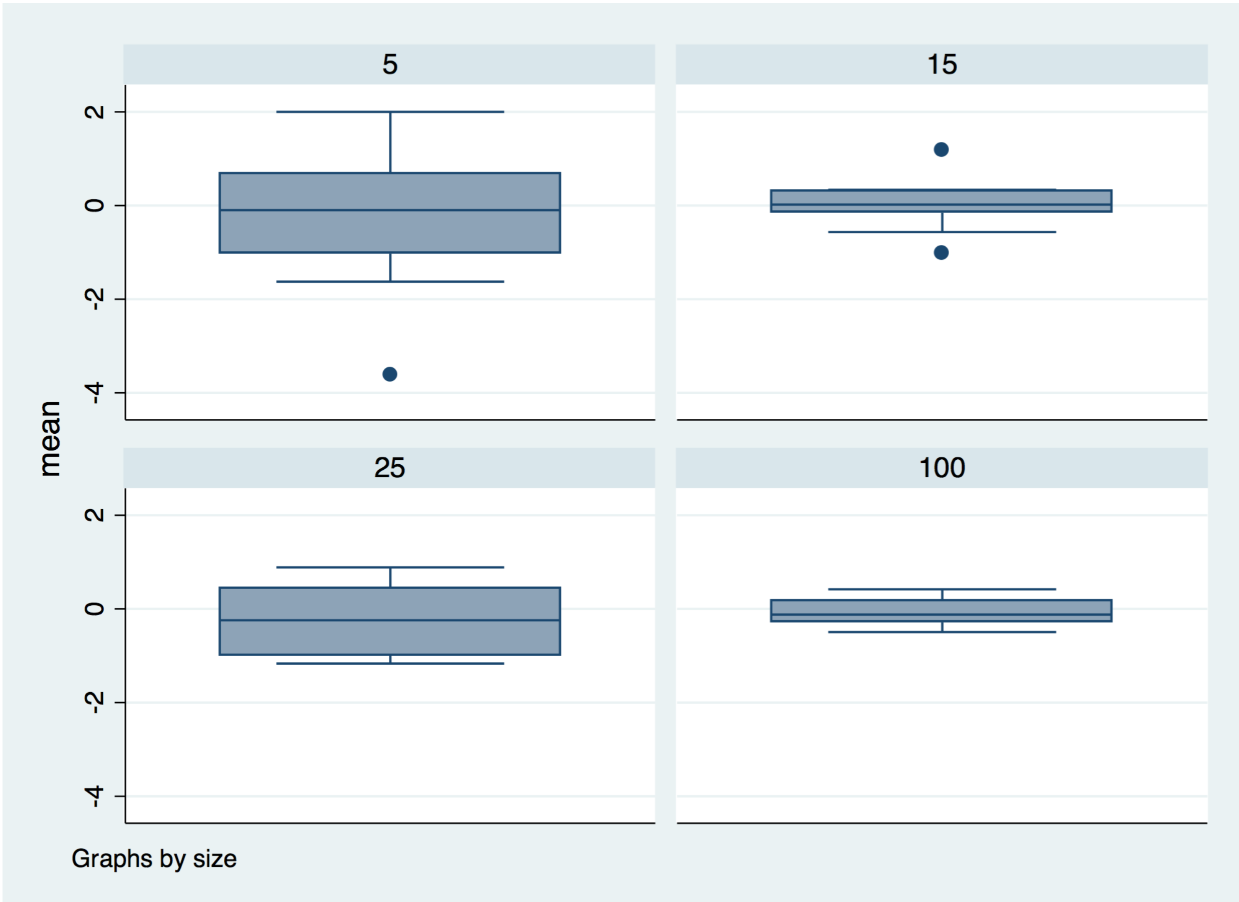
```
-------------+--------------------------------------------------------
 observation |        15    .0111664    2.893321  -4.243984   6.554693

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
 observation |        15   -1.030764     3.63934  -9.709027   4.823821

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
 observation |        15    .0140114    2.937627  -7.251023    4.63186

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
 observation |        15   -.2918965    2.854115  -6.517677   4.405031

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
 observation |        15     .111334    3.038959  -5.193361    4.62143

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
 observation |        15   -.1541863    2.896933  -4.935158   4.383737

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
 observation |        15   -.5667987    3.327492  -4.874509   4.482293

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
 observation |        15    .1455095    3.666581  -7.201359   6.894349

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
 observation |        15    .0024407    3.304709  -5.488784   7.140138

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
 observation |        15     .306066    3.929371  -6.626971   5.038344

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
 observation |        15    .0196183    3.002673  -5.330605   4.478324

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
 observation |        15    .3361137    3.932817  -9.918619    5.04154

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
 observation |        25   -1.156035    3.230117  -7.048663   6.613952

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
 observation |        25   -.5435422    3.478905  -5.793712    6.98085
```

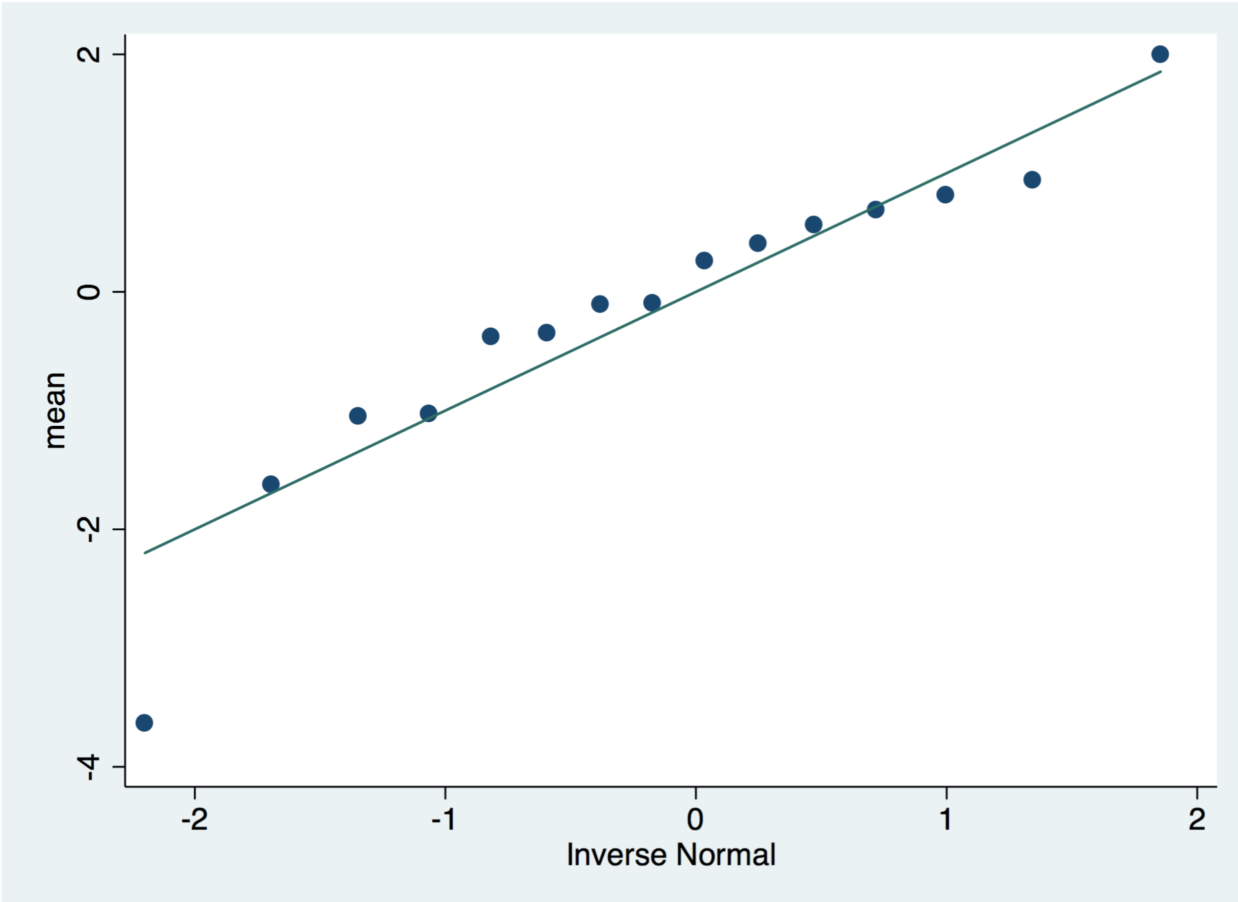
```
    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |         25    .3093568    3.280309  -6.810264   7.966505

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |         25    .4389188    3.088024  -4.856911   7.297013

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |         25   -1.166001    3.340921  -9.188413   5.386401

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |         25    .5979134    3.313374  -6.395189   8.264582

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |         25    .5478882    3.072026  -5.974334    6.27983

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |         25   -.1541815    4.551343  -8.713041   9.584764

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |         25     .886669    3.869363  -7.518301   6.874853

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |         25   -.9920868    2.826544  -6.322566    3.63698

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |         25    -.718968    3.872108   -9.71908   6.369649

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |         25   -.6490565    2.740573  -5.476446   4.895412

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |         25   -1.015911    3.490921  -6.750055   5.913665

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |         25    .3622803    3.639054  -7.990776   7.393661

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |         25   -.2420888     3.75885  -7.558431   7.056714

    Variable |        Obs        Mean    Std. Dev.       Min        Max
```

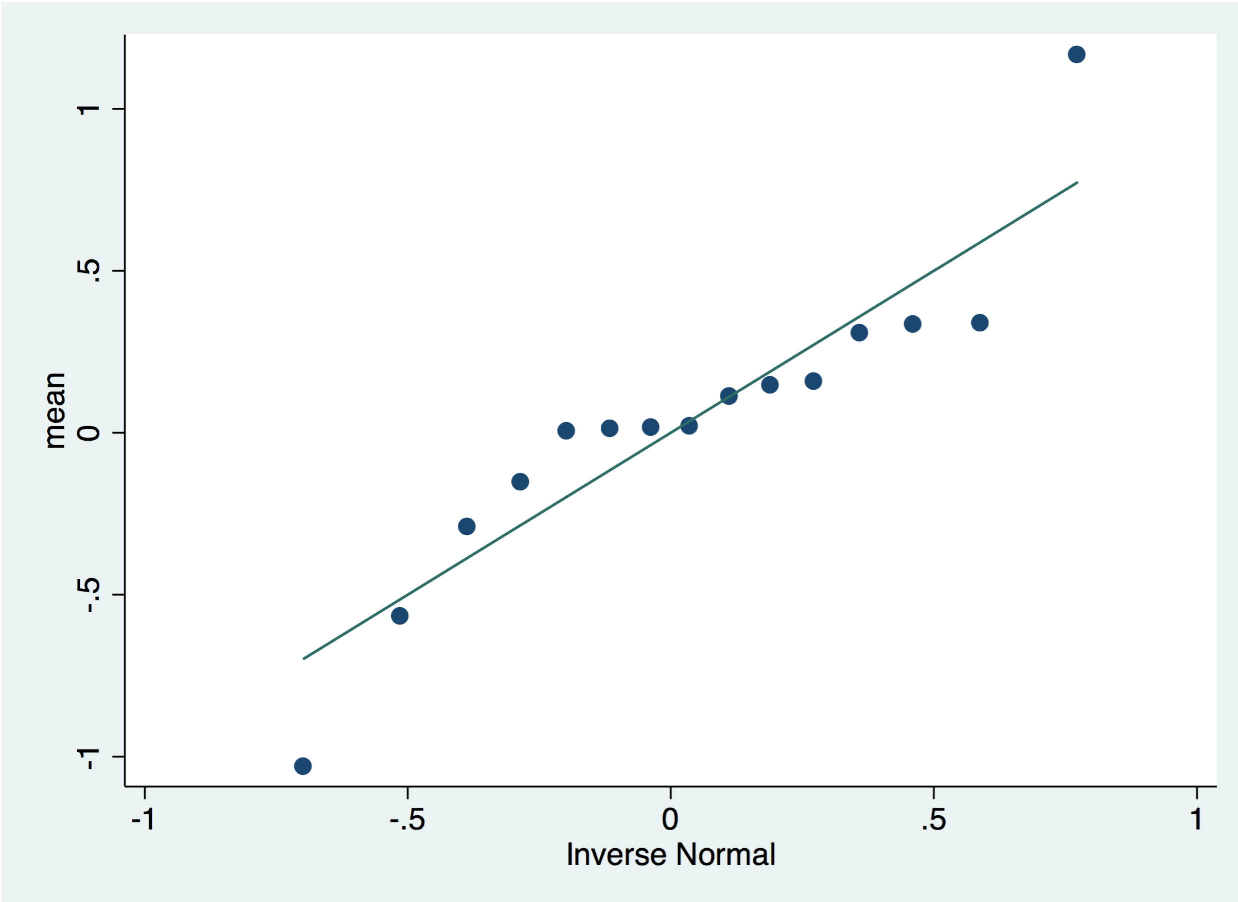
```
-------------+---------------------------------------------------------
 observation |        100   -.4383494    3.147817   -8.32965   6.735005

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |        100   -.2713836      2.8653   -8.11684   6.329414

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |        100   -.1207133    3.082861  -7.415754   8.970666

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |        100    .1647736    2.960644  -7.550699   8.227775

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |        100   -.1129649    3.181103  -7.347415   6.563672

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |        100   -.4950483    3.691278  -9.151102   9.836253

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |        100    .1835885     3.43671  -8.431155   8.177943

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |        100   -.2842181    3.109602  -6.662359   8.513903

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |        100   -.2147722    3.160636  -6.871425   7.787577

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |        100    .4189958    3.202346  -9.925777   6.724799

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |        100    .1742163    3.468608  -7.731831   8.916285

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |        100    .4067078    3.203779  -5.795245   8.824605

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |        100   -.3901959    2.845062  -8.121987   8.125437

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |        100   -.1594001     3.16435  -8.615282   7.327055
```

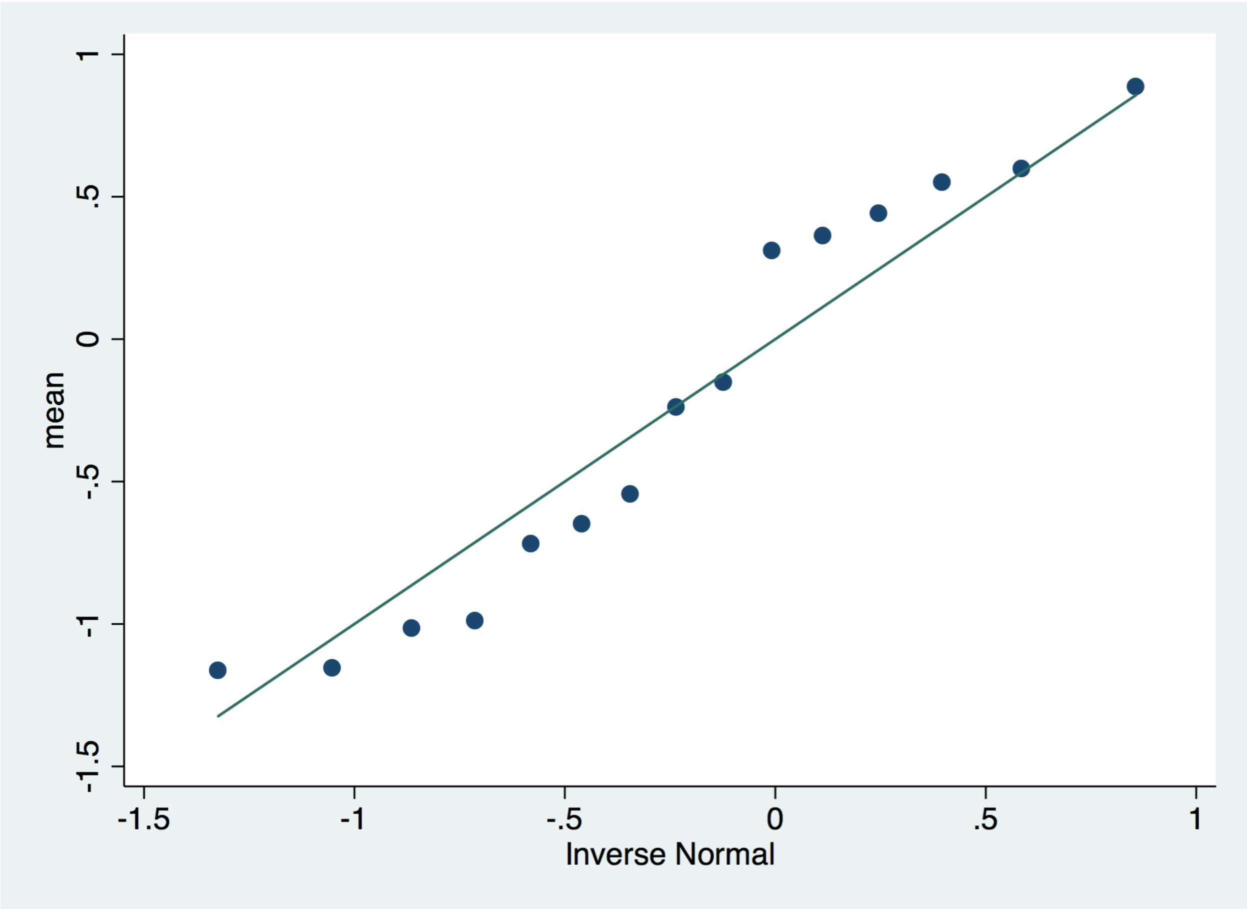
```
    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 observation |        100    .3011528    3.201704  -7.442526    7.15992

. file close myresults

. set more on

. 
. 
. * Load the sample means table
. import delimited "sample_means_2016.txt", clear
(2 vars, 60 obs)

. 
. * Make the graphics
. histogram mean, by(size)

. graph export "hist3.png", replace
(file hist3.png written in PNG format)

. graph box mean, by(size)

. graph export "box3.png", replace
(file box3.png written in PNG format)

. 
. * Optional qnorm plots
. qnorm mean if size == 5

. graph export "qnorm_mean_size5.png", replace
(file qnorm_mean_size5.png written in PNG format)

. qnorm mean if size == 15

. graph export "qnorm_mean_size15.png", replace
(file qnorm_mean_size15.png written in PNG format)

. qnorm mean if size == 25

. graph export "qnorm_mean_size25.png", replace
(file qnorm_mean_size25.png written in PNG format)

. qnorm mean if size == 100

. graph export "qnorm_mean_size100.png", replace
(file qnorm_mean_size100.png written in PNG format)
```

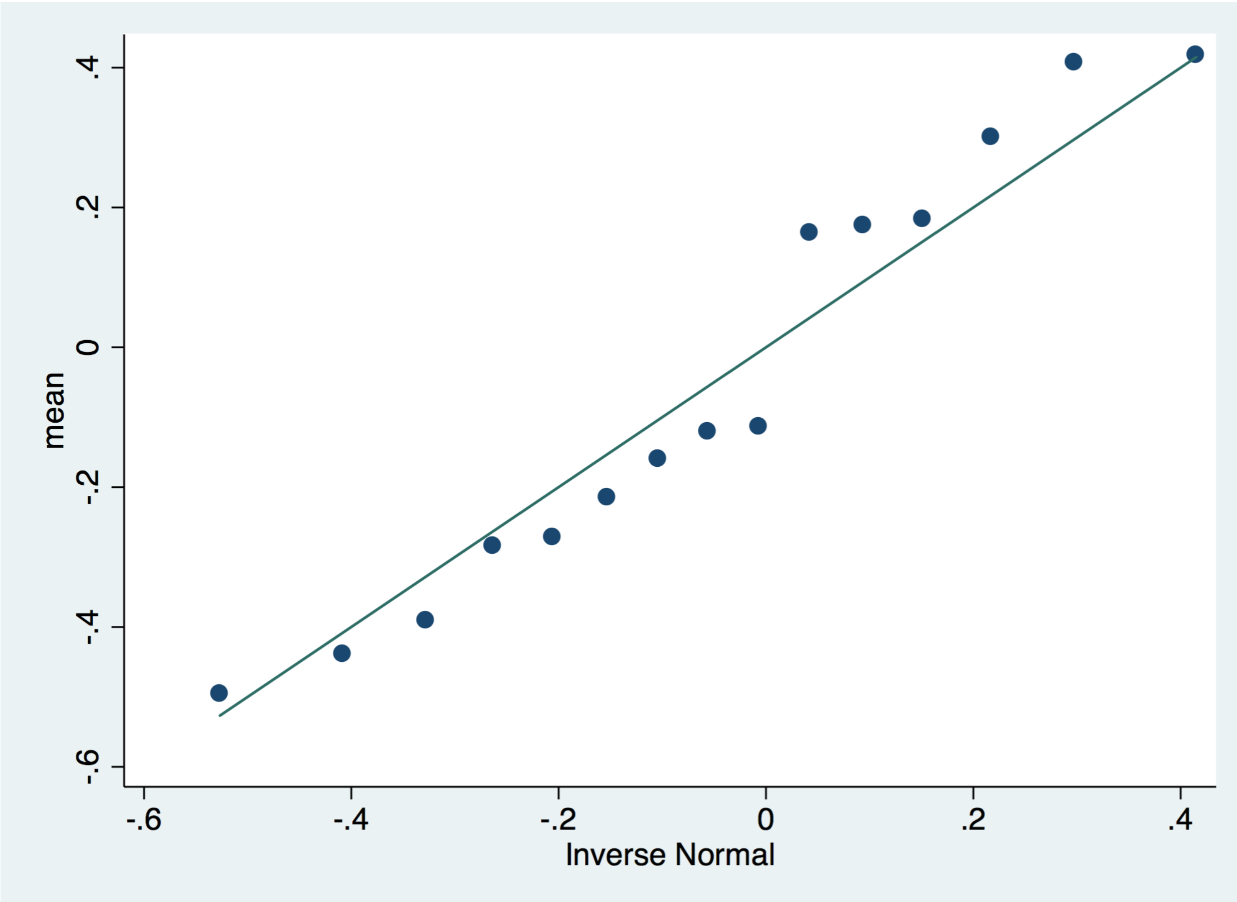

5
15
25
100
Density
1
.5
0
-4
-2
0
2
mean
Graphs by size


5
15
25
100
mean
2
0
-2
-4
Graphs by size


2
0
-2
-4
mean
-2
-1
0
1
2
Inverse Normal


mean
1
.5
0
-.5
-1
-1
-.5
0
.5
1
Inverse Normal


mean
1
.5
0
-.5
-1
-1.5
-1.5
-1
-.5
0
.5
1
Inverse Normal



mean
.4
.2
0
-.2
-.4
-.6
-.6
-.4
-.2
0
.2
.4
Inverse Normal