Handout seminar 6, ECON4150

Herman Kruse*

March 17, 2013

Introduction - list of commands

This week, we need a couple of new commands in order to solve all the problems.

- hist var1 if var2, options - creates a histogram of var1 with var2 as reference. Options are many and can be found using "findit"
- su(mmarize) - summarizes some key properties of the specified variable. Option "detail" is added to get the more detailed properties (such as kurtosis and skewness)
- scalar - creates a scalar product

```
gen lprice = ln(price)
reg lprice sqft
```

```
      Source |       SS       df       MS              Number of obs =     880
-------------+------------------------------           F(  1,   878) = 2143.38
       Model |  88.3556977     1  88.3556977           Prob > F      =  0.0000
    Residual |  36.1934444   878  .041222602           R-squared     =  0.7094
-------------+------------------------------           Adj R-squared =  0.7091
       Total |  124.549142   879  .141694132           Root MSE      =  .20303

------------------------------------------------------------------------------
      lprice |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        sqft |    .000596   .0000129    46.30   0.000     .0005707    .0006212
       _cons |   10.59379     .02185   484.84   0.000      10.5509    10.63667
------------------------------------------------------------------------------
```

```
predict ehat1, residuals

mean sqft
mean price
```

```
Mean estimation                    Number of obs    =     880

--------------------------------------------------------------
             |       Mean   Std. Err.     [95% Conf. Interval]
-------------+------------------------------------------------
        sqft |   1611.968   17.93339      1576.771    1647.165
--------------------------------------------------------------
```

*Thanks to Erling Skancke for excellent suggestions to this document

```
Mean estimation                   Number of obs    =     880

--------------------------------------------------------------
             |       Mean   Std. Err.     [95% Conf. Interval]
-------------+------------------------------------------------
       price |   112810.8   1780.356      109316.6    116305.1
--------------------------------------------------------------
```

```
gen lsqft = ln(sqft)
reg lprice lsqft
```

```
      Source |       SS       df       MS              Number of obs =     880
-------------+------------------------------           F(  1,   878) = 1993.88
       Model |  86.4716562     1  86.4716562           Prob > F      =  0.0000
    Residual |  38.0774859   878  .043368435           R-squared     =  0.6943
-------------+------------------------------           Adj R-squared =  0.6939
       Total |  124.549142   879  .141694132           Root MSE      =  .20825

------------------------------------------------------------------------------
      lprice |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lsqft |   1.006582   .0225423    44.65   0.000     .9623386    1.050825
       _cons |   4.170677   .1655084    25.20   0.000     3.845839    4.495515
------------------------------------------------------------------------------
```

```
predict ehat2, residuals

reg price sqft
```

```
      Source |       SS       df       MS              Number of obs =     880
-------------+------------------------------           F(  1,   878) = 1799.75
       Model |  1.6479e+12     1  1.6479e+12           Prob > F      =  0.0000
    Residual |  8.0391e+11   878   915618929           R-squared     =  0.6721
-------------+------------------------------           Adj R-squared =  0.6717
       Total |  2.4518e+12   879  2.7893e+09           Root MSE      =   30259

------------------------------------------------------------------------------
       price |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        sqft |   81.38899   1.918489    42.42   0.000     77.62363    85.15435
       _cons |  -18385.65   3256.424    -5.65   0.000    -24776.94   -11994.37
------------------------------------------------------------------------------
```

```
predict ehat3, residuals

hist ehat1 if lprice, bin(35) start (-1)

hist ehat2 if lprice, bin(35) start (-1)

hist ehat3 if price, bin(35) start (-110000)
```

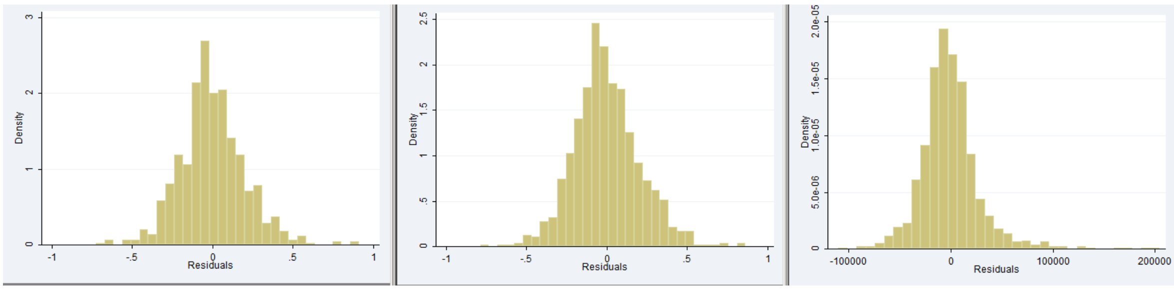

Figure 1: Histograms of residuals

For *Jarque-Bera-testing*, we need to know how to construct the test-statistic. It has the following formula:

$$\text{JB} = \frac{n}{6} \cdot [(\text{skewness}^2) + \frac{(\text{kurtosis} - 3)^2}{4}]$$

And what we really test is the hypothesis about normality in the residuals. If the observed value is above some critical value, we reject the hypothesis and conclude that the residuals are not compatible with an assumption about normality. The skewness refers to how symmetric the residuals are around zero, while kurtosis refers to the peakedness of the distribution. For a normal distribution, the skewness is equal to zero, while the kurtosis is equal to three. So we need to check whether the skewness is sufficiently different from zero and kurtosis sufficiently different from three in order to conclude that the residuals are *not* normally distributed. When the residuals are normally distributed, the Jarque-Bera statistic has a chi-square distribution with two degrees of freedom. So we reject the null-hypothesis when we have a test statistic exceeding $\chi^2_{2,0.95} = 5.99$ with a 5% significance-level.

Note that if we do not reject the null-hypothesis, this does *not* directly imply normality in the residuals. There are more distributions with skewness 0 and kurtosis 3 (or so-called symmetric and mesokurtosic distributions). So the Jarque-Bera test will, if we reject, say we have strong evidence about a skewed distribution, or a sharply peaked distribution.

```
su ehat1, detail
su ehat2, detail
su ehat3, detail
```

```
                            Residuals
-------------------------------------------------------------
      Percentiles      Smallest
 1%    -.4598814      -.710299
 5%    -.3142879     -.6619065
10%    -.2410653     -.6477303       Obs                 880
25%    -.1200507     -.6345798       Sum of Wgt.         880

50%    -.0139173                     Mean          -1.73e-10
                        Largest      Std. Dev.       .202918
75%     .1158161       .7670023
```

```
90%     .2606355       .7681834       Variance       .0411757
95%     .3558994       .8957195       Skewness       .3239307
99%     .5422422       .9086631       Kurtosis       4.315611


                         Residuals
-------------------------------------------------------------
      Percentiles      Smallest
 1%       -.49264      -.7518981
 5%     -.3091487      -.6739541
10%     -.2441997      -.6184166       Obs                 880
25%     -.1303633      -.5487044       Sum of Wgt.         880

50%      -.018237                      Mean           -2.38e-10
                        Largest        Std. Dev.       .2081324
75%      .1182303       .7180918
90%      .2746109       .7190305       Variance        .0433191
95%      .3657138       .8612714       Skewness        .3488042
99%      .5302507       .8624387       Kurtosis        3.975605


                         Residuals
-------------------------------------------------------------
      Percentiles      Smallest
 1%     -68089.01      -101224.1
 5%     -41894.24      -91337.09
10%     -30454.09      -84395.79       Obs                 880
25%     -16140.87      -76857.88       Sum of Wgt.         880

50%     -2667.711                      Mean           -9.86e-06
                        Largest        Std. Dev.       30241.98
75%      12093.73       166920.7
90%      28794.58         168413       Variance        9.15e+08
95%      50104.84       186850.3       Skewness         1.59206
99%      112023.8       204279.8       Kurtosis        10.53922
```

```
scalar jb = (880/6)*(0.3239307)^2 + (880/6)*((4.315611-3)^2)/4
di "Jarque-Bera Statistic = " jb
scalar chic = invchi2tail(2,.05)
di "Chi-square(2) 95th percentile = " chic
scalar pvalue = chi2tail(2,jb)
di "Jarque-Bera p-value = " pvalue

-------------------------

scalar jb = (880/6)*(0.3488042)^2 + (880/6)*((3.975605-3)^2)/4
di "Jarque-Bera Statistic = " jb
scalar chic = invchi2tail(2,.05)
di "Chi-square(2) 95th percentile = " chic
scalar pvalue = chi2tail(2,jb)
di "Jarque-Bera p-value = " pvalue

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

```
scalar jb = (880/6)*(1.59206)^2 + (880/6)*((10.53922-3)^2)/4
di "Jarque-Bera Statistic = " jb
scalar chic = invchi2tail(2,.05)
di "Chi-square(2) 95th percentile = " chic
scalar pvalue = chi2tail(2,jb)
di "Jarque-Bera p-value = " pvalue
```

```
Jarque-Bera Statistic = 78.853746
Chi-square(2) 95th percentile = 5.9914645
Jarque-Bera p-value = 7.536e-18

------------------------

Jarque-Bera Statistic = 52.743629
Chi-square(2) 95th percentile = 5.9914645
Jarque-Bera p-value = 3.523e-12

------------------------

Jarque-Bera Statistic = 2455.8768
Chi-square(2) 95th percentile = 5.9914645
Jarque-Bera p-value = 0
```

```
scatter ehat1 sqft

scatter ehat2 sqft

scatter ehat3 sqft
```

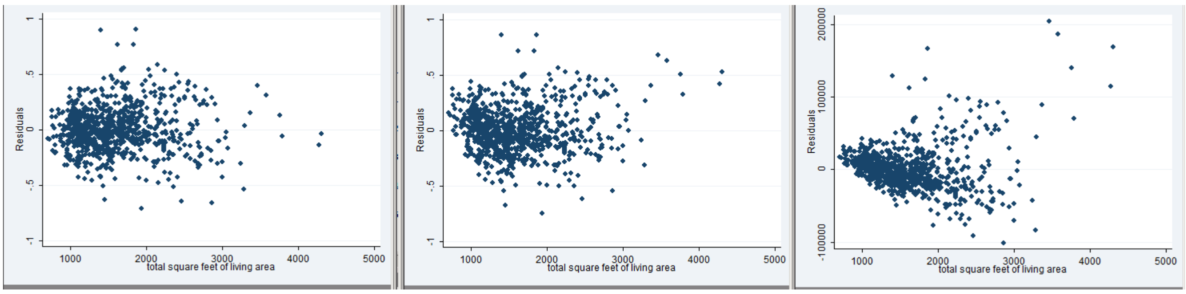

Figure 2: Scatter plot residuals on sqft

—

Mis-calculation:

```
di exp(10.59379+0.000596*2700)
di exp(4.170677)*2700^1.006582
di (81.38899*2700-18385.65)
```

```
199384.42
184183.62
201364.62
```

Correct calculation:

```
di exp(10.59379+0.000596*2700+0.041222602/2)
di exp(4.170677+1.006582*ln(2700)+0.043368435/2)
di (81.38899*2700-18385.65)
```

```
203536.64
188221.12
201364.62
```