**Example 1:** We gathered a sample of data for 32 European countries for the year 2003. We have the following variables available (the data are provided in Stata Data file `health.dta`, while the programming code is given in Stata Do file `health-commands-l08.do`):

- life expectancy at birth (*LIFE*; in years);
- health expenditure per capita (*EXP*; in U.S. dollars);
- percentage of smokers among adults (*TOBACCO*);
- consumption of alcohol per capita (*ALCO*; in litres of distilled spirits).

a) Estimate the linear regression model: $LIFE_i = \beta_1 + \beta_2 EXP_i + \beta_3 TOBACCO_i + u_i$ and check validity of the assumption on normality of the disturbances.

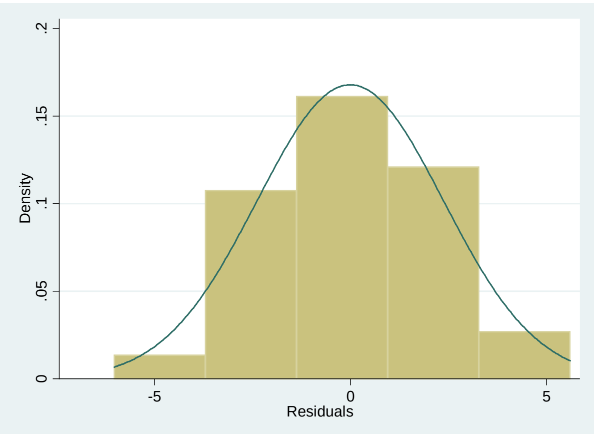
b) Estimate the linear regression model: $LIFE_i = \beta_1 + \beta_2 EXP_i + \beta_3 TOBACCO_i + \beta_4 ALCO_i + u_i$ and check validity of the assumption on (absence of) multicollinearity.

### *Computer printout of the results in Stata:*

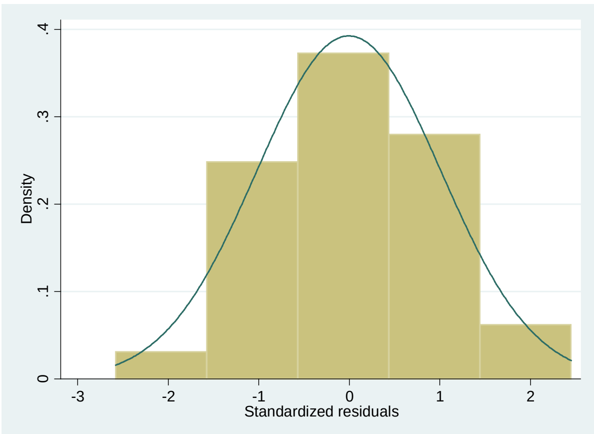
*Normality of the disturbances:*

```
. regress life exp tobacco

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  2,    29) =   31.97
       Model |  385.751827     2  192.875914           Prob > F      =  0.0000
    Residual |   174.97295    29  6.03354999           R-squared     =  0.6880
-------------+------------------------------           Adj R-squared =  0.6664
       Total |  560.724777    31   18.087896           Root MSE      =  2.4563

------------------------------------------------------------------------------
        life |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         exp |   .0023323   .0003709     6.29   0.000      .0015736    .0030909
     tobacco |  -.2503555   .0889983    -2.81   0.009     -.4323774   -.0683335
       _cons |   79.62409   2.796632    28.47   0.000      73.90433    85.34384
------------------------------------------------------------------------------

. predict elife, resid
. predict elifestd, rstandard
```

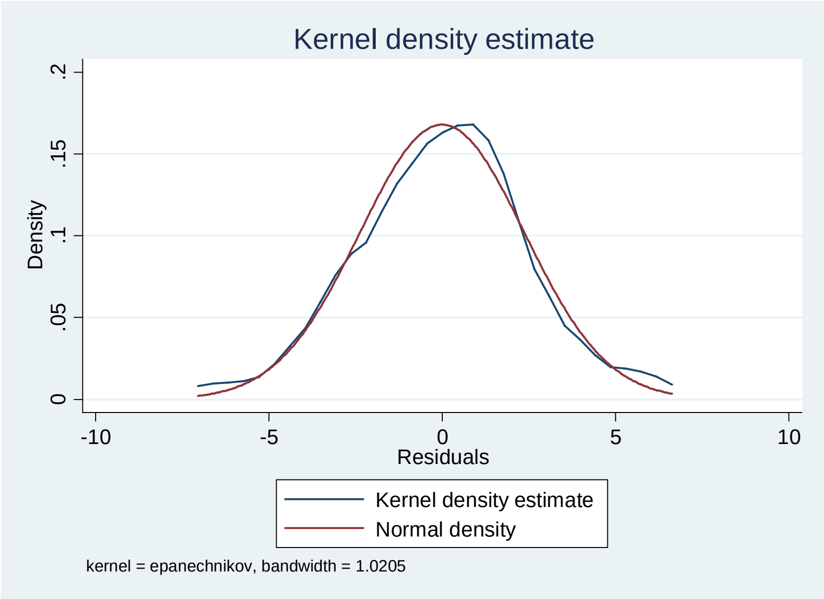
```
. histogram elife, normal
(bin=5, start=-6.0237689, width=2.3258684)
```




```
. histogram elifestd, normal
(bin=5, start=-2.5803783, width=1.0055467)
```



```
. kdensity elife, normal
(n() set to 32)
```




```
. sum elife, detail

                          Residuals
-------------------------------------------------------------
      Percentiles      Smallest
 1%    -6.023769      -6.023769
 5%    -3.561575      -3.561575
10%    -2.820123      -2.855744       Obs                  32
25%    -1.700531      -2.820123       Sum of Wgt.          32

50%     .2013203                      Mean           1.28e-08
                        Largest       Std. Dev.      2.375771
75%     1.358591        2.43306
90%      2.43306       2.645627       Variance       5.644289
95%     4.427194       4.427194       Skewness      -.0833723
99%     5.605573       5.605573       Kurtosis       3.410826

. return list

scalars:
                  r(N) =  32
              r(sum_w) =  32
               r(mean) =  1.28056854010e-08
                r(Var) =  5.644288676812403
                 r(sd) =  2.37577117517921
           r(skewness) =  -.0833723127091076
           r(kurtosis) =  3.410825630496252
                r(sum) =  4.09781932831e-07
                r(min) =  -6.023768901824951
                r(max) =  5.605573177337647
                 r(p1) =  -6.023768901824951
                 r(p5) =  -3.561574935913086
                r(p10) =  -2.820123434066773
                r(p25) =  -1.700530529022217
                r(p50) =  .2013202682137489
                r(p75) =  1.358591318130493
                r(p90) =  2.433059930801392
                r(p95) =  4.427193641662598
                r(p99) =  5.605573177337647
```

```
. scalar obs=r(N)
. scalar s=r(skewness)
. scalar k=r(kurtosis)

. scalar jb=obs*(s^2/6+(k-3)^2/24)
. display jb
.26210863

. display chi2tail(2,jb)
.87717013

. jb6 elife
Jarque-Bera normality test:  .2621 Chi(2)  .8772
Jarque-Bera test for Ho: normality: (elife)
```

*Multicollinearity:*

```
. regress life exp tobacco alco

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   26.30
       Model |  413.850212     3  137.950071           Prob > F      =  0.0000
    Residual |  146.874565    28  5.24552017           R-squared     =  0.7381
-------------+------------------------------           Adj R-squared =  0.7100
       Total |  560.724777    31   18.087896           Root MSE      =  2.2903

------------------------------------------------------------------------------
        life |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         exp |   .0018569   .0004023     4.62   0.000     .0010329    .0026809
     tobacco |  -.2238391   .0837702    -2.67   0.012    -.3954346   -.0522436
        alco |  -.6493606   .2805689    -2.31   0.028     -1.22408   -.0746412
       _cons |   81.42053   2.720683    29.93   0.000     75.84746    86.99359
------------------------------------------------------------------------------

. pwcorr exp tobacco alco, sig

             |      exp  tobacco     alco
-------------+---------------------------
         exp |   1.0000 
             |
             |
     tobacco |  -0.3017   1.0000 
             |   0.0933
             |
        alco |  -0.5510   0.2751   1.0000 
             |   0.0011   0.1276

. regress exp tobacco alco

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  2,    29) =    7.08
       Model |    15822811     2  7911405.51           Prob > F      =  0.0031
    Residual |    32415407    29  1117772.65           R-squared     =  0.3280
-------------+------------------------------           Adj R-squared =  0.2817
       Total |    48238218    31  1556071.55           Root MSE      =  1057.2

------------------------------------------------------------------------------
         exp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     tobacco |  -38.96759   37.98675    -1.03   0.313    -116.6592    38.72404
        alco |  -356.1345   111.3588    -3.20   0.003    -583.8887   -128.3802
       _cons |   3602.096    1062.97     3.39   0.002     1428.078    5776.114
------------------------------------------------------------------------------
```

```
. scalar R2exp=e(r2)
. scalar Fexp=(R2exp/(4-2))/((1-R2exp)/(32-4+1))
. scalar pFexp=Ftail(4-2,32-4+1,Fexp)
. scalar vifexp=1/(1-R2exp)
. scalar toleranceexp=1/vifexp

. display Fexp, pFexp, vifexp, toleranceexp
7.0778306 .00313851 1.4881262 .671986

. regress tobacco exp alco

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  2,    29) =    1.76
       Model |  90.5229799     2    45.26149           Prob > F      =  0.1906
    Residual |  747.496685    29  25.7757477           R-squared     =  0.1080
-------------+------------------------------           Adj R-squared =  0.0465
       Total |  838.019664    31  27.0328924           Root MSE      =   5.077

------------------------------------------------------------------------------
     tobacco |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         exp |  -.0008986    .000876    -1.03   0.313    -.0026902     .000893
        alco |    .458065   .6160993     0.74   0.463    -.8019995    1.718129
       _cons |   28.65308   2.839478    10.09   0.000     22.84569    34.46046
------------------------------------------------------------------------------

. scalar R2tobacco=e(r2)
. scalar Ftobacco=(R2tobacco/(4-2))/((1-R2tobacco)/(32-4+1))
. scalar pFtobacco=Ftail(4-2,32-4+1,Ftobacco)
. scalar viftobacco=1/(1-R2tobacco)
. scalar tolerancetobacco=1/viftobacco

. display Ftobacco, pFtobacco, viftobacco, tolerancetobacco
1.7559719 .19061119 1.1211015 .89197989

. regress alco exp tobacco

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  2,    29) =    6.72
       Model |  30.8785495     2  15.4392748           Prob > F      =  0.0040
    Residual |  66.6360968    29  2.29779644           R-squared     =  0.3167
-------------+------------------------------           Adj R-squared =  0.2695
       Total |  97.5146463    31  3.14563375           Root MSE      =  1.5158

------------------------------------------------------------------------------
        alco |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         exp |  -.0007321   .0002289    -3.20   0.003    -.0012003   -.0002639
     tobacco |   .0408345   .0549226     0.74   0.463    -.0714948    .1531638
       _cons |   2.766477   1.725856     1.60   0.120    -.7632956     6.29625
------------------------------------------------------------------------------

. scalar R2alco=e(r2)
. scalar Falco=(R2alco/(4-2))/((1-R2alco)/(32-4+1))
. scalar pFalco=Ftail(4-2,32-4+1,Falco)
. scalar vifalco=1/(1-R2alco)
. scalar tolerancealco=1/vifalco

. display Falco, pFalco, vifalco, tolerancealco
6.7191656 .00400199 1.4633907 .6833445
```

```
. qui regress life exp tobacco alco
. estat vif

    Variable |       VIF       1/VIF
-------------+----------------------
         exp |      1.49    0.671986
        alco |      1.46    0.683345
     tobacco |      1.12    0.891980
-------------+----------------------
    Mean VIF |      1.36

. collin exp tobacco alco, corr rinv
(obs=32)

  Collinearity Diagnostics

                        SQRT                   R-
  Variable      VIF     VIF    Tolerance    Squared
----------------------------------------------------
       exp      1.49    1.22    0.6720      0.3280
   tobacco      1.12    1.06    0.8920      0.1080
      alco      1.46    1.21    0.6833      0.3167
----------------------------------------------------
  Mean VIF      1.36

                           Cond
        Eigenval          Index
---------------------------------
    1     1.7677          1.0000
    2     0.7842          1.5014
    3     0.4481          1.9862
---------------------------------
 Condition Number         1.9862
 Eigenvalues & Cond Index computed from deviation sscp (no intercept)
 Det(correlation matrix)    0.6211

Inverse of correlation matrix

                exp     tobacco        alco
    exp   1.4881262
tobacco   .24169899   1.1211015
   alco   .75351706  -.17517811   1.4633907
```

■

**Example 2:** We analyse production functions for 81 manufacturing companies in the computer manufacturing industry for a given year (the data are provided in Stata Data file `production.dta`, while the programming code is given in Stata Do file `production-commands-l08.do`). We have cross-section data available for the following variables:

- value added as a proxy for the product ($Q$; in 1,000 monetary units);
- average number of employed workers as a proxy for labour ($L$);
- sum of tangible and intangible assets as a proxy for capital ($K$; in 1,000 monetary units).

a) Check validity of the assumption on normality of the disturbances by applying the Jarque–Bera test in the model of linear and log-linear production function.
b) Check validity of the assumption on (absence of) multicollinearity in the model of linear and log-linear production function.

## *Computer printout of the results in Stata:*

*Normality of the disturbances:*

```
. regress q l k

      Source |       SS       df       MS              Number of obs =      81
-------------+------------------------------           F(  2,    78) =   52.90
       Model |  6.9350e+12     2  3.4675e+12           Prob > F      =  0.0000
    Residual |  5.1130e+12    78  6.5551e+10           R-squared     =  0.5756
-------------+------------------------------           Adj R-squared =  0.5647
       Total |  1.2048e+13    80  1.5060e+11           Root MSE      =  2.6e+05

------------------------------------------------------------------------------
           q |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           l |   9687.383   3640.852     2.66   0.009     2439.003    16935.76
           k |    2.27941   .7553228     3.02   0.003      .775678    3.783142
       _cons |  -11875.29   34865.13    -0.34   0.734    -81286.43    57535.85
------------------------------------------------------------------------------

. predict eq, resid
. sum eq, detail

                          Residuals
-------------------------------------------------------------
      Percentiles      Smallest
 1%    -928124.7      -928124.7
 5%    -205781.2      -693306.2
10%    -149066.2        -520897       Obs                  81
25%     -30861.5      -258838.8       Sum of Wgt.          81

50%    -3095.372                      Mean          -.0010632
                        Largest       Std. Dev.      252809.1
75%     7945.411         350021
90%      60822.5       778072.3       Variance       6.39e+10
95%     166468.2       816731.7       Skewness        1.55663
99%      1310726        1310726       Kurtosis       14.98543

. scalar obs=r(N)
. scalar slin=r(skewness)
. scalar klin=r(kurtosis)

. scalar jblin=obs*(slin^2/6+(klin-3)^2/24)
. display jblin
517.53233

. display chi2tail(2,jblin)
4.16e-113

. jb6 eq
Jarque-Bera normality test:  517.5 Chi(2)  4.e-113
Jarque-Bera test for Ho: normality: (eq)

. gen lq=log(q)
. gen ll=log(l)
. gen lk=log(k)
```

```
. regress lq ll lk

      Source |       SS       df       MS              Number of obs =      81
-------------+------------------------------           F(  2,    78) =  190.75
       Model |  178.261263     2  89.1306313           Prob > F      =  0.0000
    Residual |    36.44752    78  .467275898           R-squared     =  0.8302
-------------+------------------------------           Adj R-squared =  0.8259
       Total |  214.708783    80  2.68385978           Root MSE      =  .68358

------------------------------------------------------------------------------
          lq |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          ll |   .9645479   .1199229     8.04   0.000     .7257997    1.203296
          lk |   .1885438   .0673358     2.80   0.006     .0544886     .322599
       _cons |   7.546026   .4617465    16.34   0.000      6.62676    8.465293
------------------------------------------------------------------------------

. predict elq, resid
. sum elq, detail

                          Residuals
-------------------------------------------------------------
      Percentiles      Smallest
 1%    -1.225007      -1.225007
 5%    -1.015222      -1.073147
10%    -.8845653      -1.022881       Obs                  81
25%    -.4654464      -1.018599       Sum of Wgt.          81

50%    -.0882534                      Mean          -1.24e-09
                        Largest       Std. Dev.       .674977
75%     .4299134       1.258059
90%     1.020899       1.329013       Variance        .455594
95%     1.214762       1.409848       Skewness       .4283599
99%     1.786791       1.786791       Kurtosis       2.619945

. scalar obs=r(N)
. scalar slog=r(skewness)
. scalar klog=r(kurtosis)

. scalar jblog=obs*(slog^2/6+(klog-3)^2/24)
. display jblog
2.9646371

. display chi2tail(2,jblog)
.22711051

. jb6 elq
Jarque-Bera normality test:  2.965 Chi(2)  .2271
Jarque-Bera test for Ho: normality: (elq)
```

*Multicollinearity:*

```
. collin l k, corr
(obs=81)

  Collinearity Diagnostics

                        SQRT                   R-
  Variable      VIF     VIF    Tolerance    Squared
----------------------------------------------------
         l      3.55    1.88    0.2817      0.7183
         k      3.55    1.88    0.2817      0.7183
----------------------------------------------------
  Mean VIF      3.55
```

```
                          Cond
        Eigenval         Index
---------------------------------
    1     1.8476         1.0000
    2     0.1524         3.4813
---------------------------------
 Condition Number        3.4813
 Eigenvalues & Cond Index computed from deviation sscp (no intercept)
 Det(correlation matrix)    0.2817
```

```
. collin ll lk, corr
(obs=81)

  Collinearity Diagnostics

                        SQRT                   R-
  Variable      VIF     VIF    Tolerance    Squared
----------------------------------------------------
        ll      3.45    1.86    0.2896      0.7104
        lk      3.45    1.86    0.2896      0.7104
----------------------------------------------------
  Mean VIF      3.45

                          Cond
        Eigenval         Index
---------------------------------
    1     1.8428         1.0000
    2     0.1572         3.4242
---------------------------------
 Condition Number        3.4242
 Eigenvalues & Cond Index computed from deviation sscp (no intercept)
 Det(correlation matrix)    0.2896
```

■