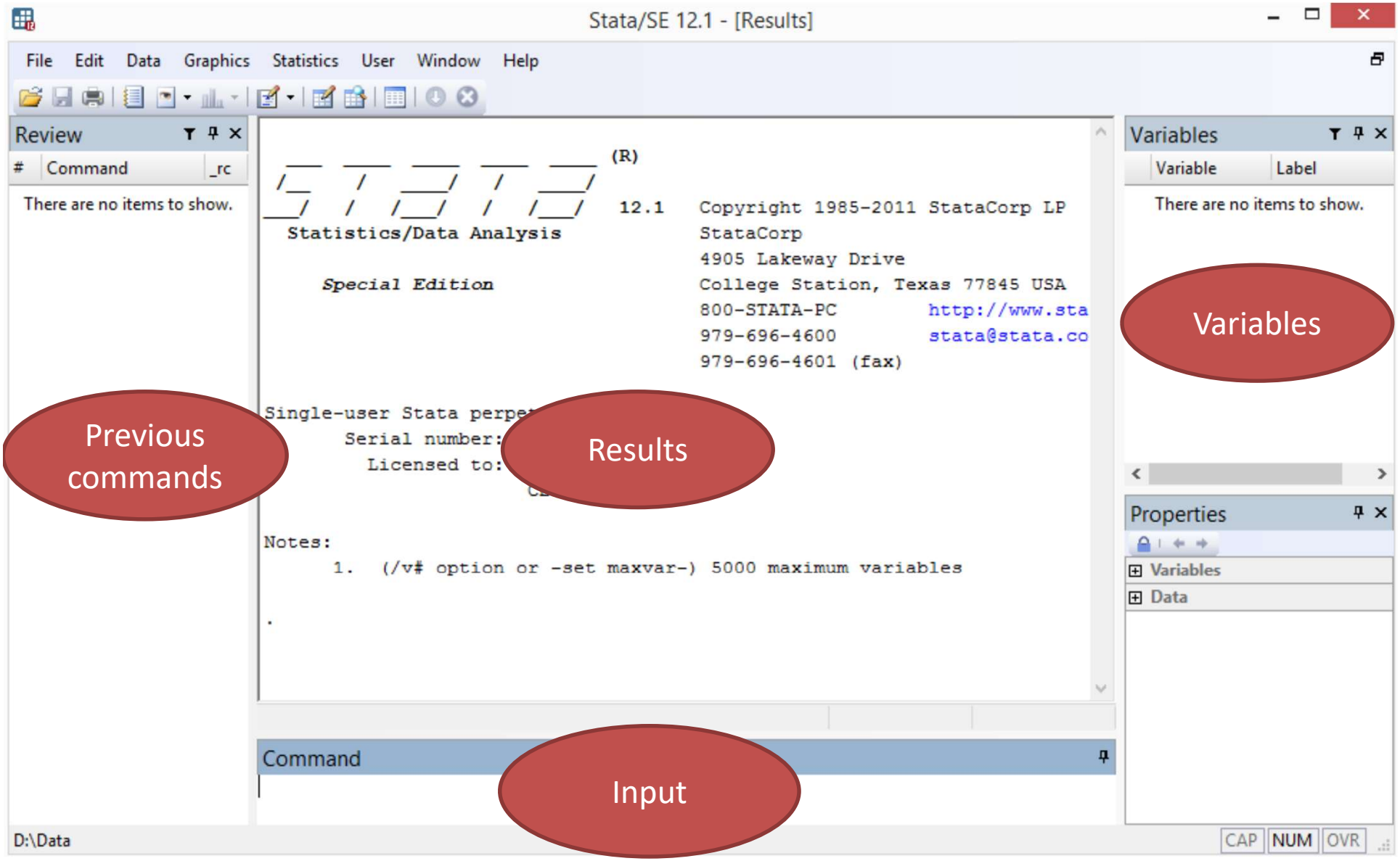
Data Analysis in Accounting

Illustrations

Esmeralda A. Ramalho

Stata – Main Windows

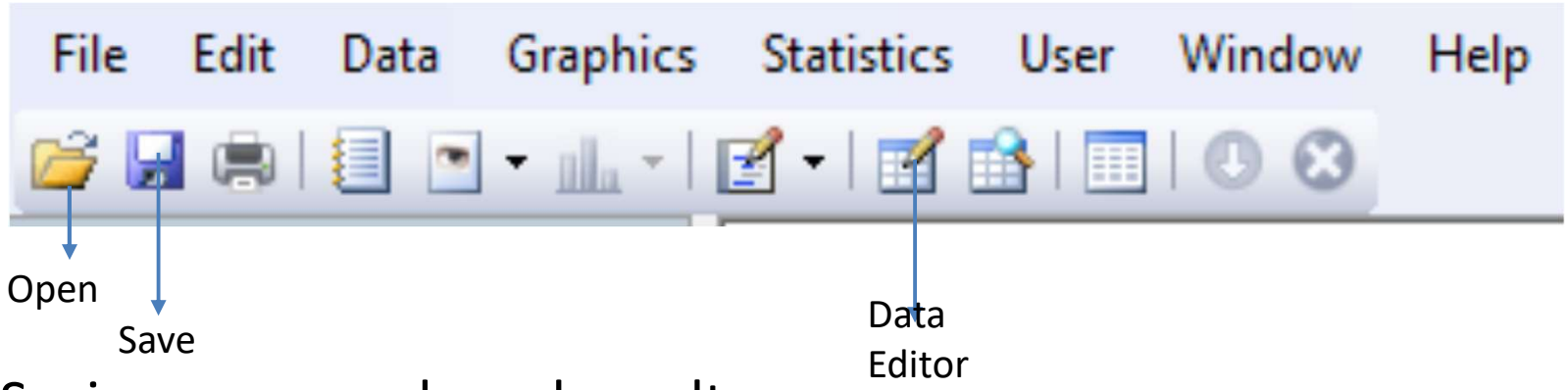


Using Stata

Approach to run commands: write the command line

Saving information:

- Saving the dataset
  - Open – open dataset
  - Save – save dataset



- Saving commands and results
  - Copy/past to Word
  - Select the font Courrier New, 9 (to obtain a convenient format)

Stata – Some Commands

describe [*varlist*]

- Produces a summary of the dataset

summarize [*varlist*] [, detail]

- Displays a variety of univariate summary statistics

generate *newvar* = *formula*

- Creates a new variable

drop [*varlist*] [if *expressao*] [in #/#]

- Drops variables or observations

Stata – Others

Missing data:

- .

Conditions:

- Equality: ==
- Different: !=
- Or: |
- And: &

Illustration 1 – Question 1

```
. describe
Contains data from H:\S1-19-20-ADF\CentralBalancos-BP.dta
  obs:        32,226
 vars:            16                          2 Sep 2019 18:02
 size:     1,482,396
-------------------------------------------------------------------------------
              storage   display    value
variable name   type    format     label      variable label
-------------------------------------------------------------------------------
id              int     %8.0g                 Firm id
YEAR            int     %8.0g                 Year
LEV_ST          float   %9.0g                 Short-term debt / (STD+LTD+Equity)
LEV_LT          float   %9.0g                 Long-term debt / (STD+LTD+Equity)
LEV             float   %9.0g                 Total debt / (STD+LTD+Equity)
COLLAT          float   %9.0g                 Tangible assets / Total assets
SIZE            float   %9.0g                 Log(Total assets)
PROF            float   %9.0g                 EBIT / Total assets
GROWTH          float   %9.0g                 Sales growth rate
AGE             int     %8.0g                 YEAR - Foundation year
LE              byte    %8.0g                 =1 if large firm
MicE            byte    %8.0g                 =1 if micro firm
SE              byte    %8.0g                 =1 if small firm
MedE            byte    %8.0g                 =1 if medium firm
-------------------------------------------------------------------------------
```

Illustration 1 – Question 2

```
. summarize
    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
          id |     32,226     2714.09    1597.178         1       5514
        YEAR |     32,226     1998.01    1.989032      1995       2001
      LEV_ST |     32,226    .1583845    .2185988         0   .9997105
      LEV_LT |     32,226    .0768912    .1678684         0   .9982489
         LEV |     32,226    .2352757    .2589438         0   .9997105
-------------+---------------------------------------------------------
      COLLAT |     32,226    .3175925      .22441         0   .9982307
        SIZE |     32,226    13.66189    1.922785   4.976734   22.38044
        PROF |     32,226    .0671496    .1138881   -.822547   10.11724
      GROWTH |     32,226    27.98725    559.0894  -99.89224   73055.81
         AGE |     32,226    20.25119     15.6708         1        212
-------------+---------------------------------------------------------
          LE |     32,226    .0596413    .2368247         0          1
        MicE |     32,226    .3063986    .4610044         0          1
          SE |     32,226    .4156892    .4928481         0          1
        MedE |     32,226     .218271    .4130787         0          1
```

Illustration 1 – Question 2

```
. summarize LEV_ST LEV_LT , d
            Short-term debt / (STD+LTD+Equity)
-------------------------------------------------------------
      Percentiles      Smallest
 1%            0             0
 5%            0             0
10%            0             0       Obs                32,226
25%            0             0       Sum of Wgt.        32,226

50%     .0373733                     Mean           .1583845
                      Largest        Std. Dev.      .2185988
75%       .26672      .9975002
90%     .4921742      .9976246       Variance       .0477854
95%     .6274608      .9990628       Skewness       1.481064
99%     .8764215      .9997105       Kurtosis       4.530042
…
```

Illustration 1 – Question 2

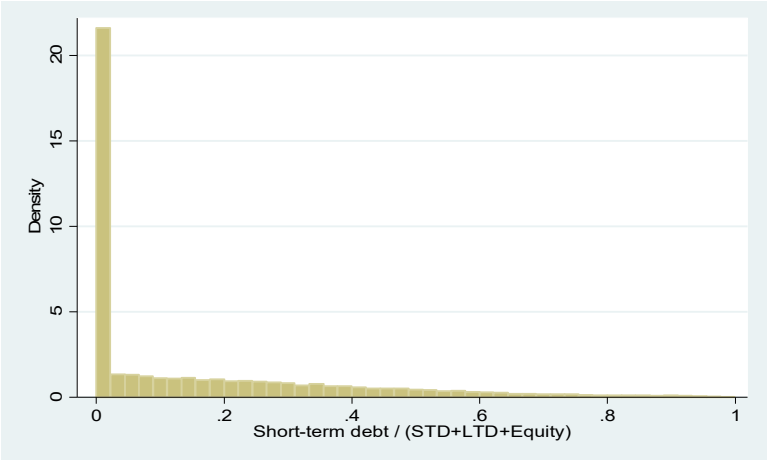
```
                Long-term debt / (STD+LTD+Equity)
-------------------------------------------------------------
      Percentiles      Smallest
 1%             0             0
 5%             0             0
10%             0             0       Obs                32,226
25%             0             0       Sum of Wgt.        32,226

50%             0                     Mean            .0768912
                        Largest       Std. Dev.       .1678684
75%     .0282531       .9862829
90%     .3221912       .9871089       Variance        .0281798
95%       .46735       .9912307       Skewness         2.52465
99%     .7566051       .9982489       Kurtosis        9.260746

.
```

Illustration 1 – Question 3

```
. histogram LEV_ST
(bin=45, start=0, width=.02221579)
```




```
. histogram LEV_LT
(bin=45, start=0, width=.02221579)
```

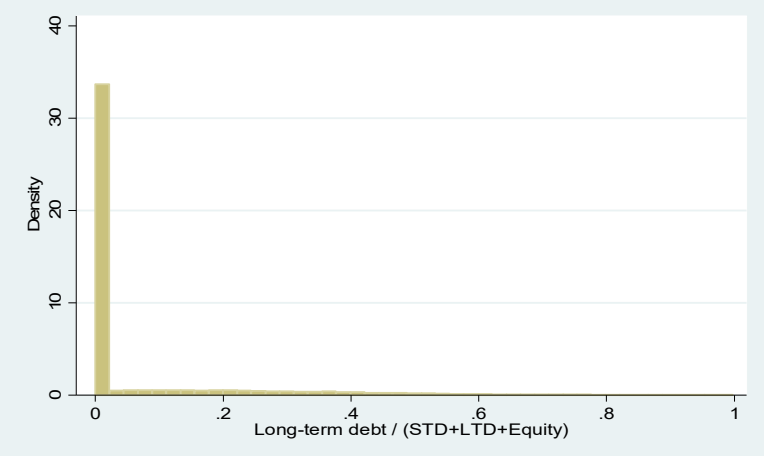


Illustration 1 – Question 4

```
. gen GROUPS="1-Micro" if MicE==1
. replace GROUPS="2-Small" if SE==1
. replace GROUPS="3-Medium" if MedE==1
. replace GROUPS="4-Large" if LE==1

. graph box LEV_ST, over(GROUPS)
```

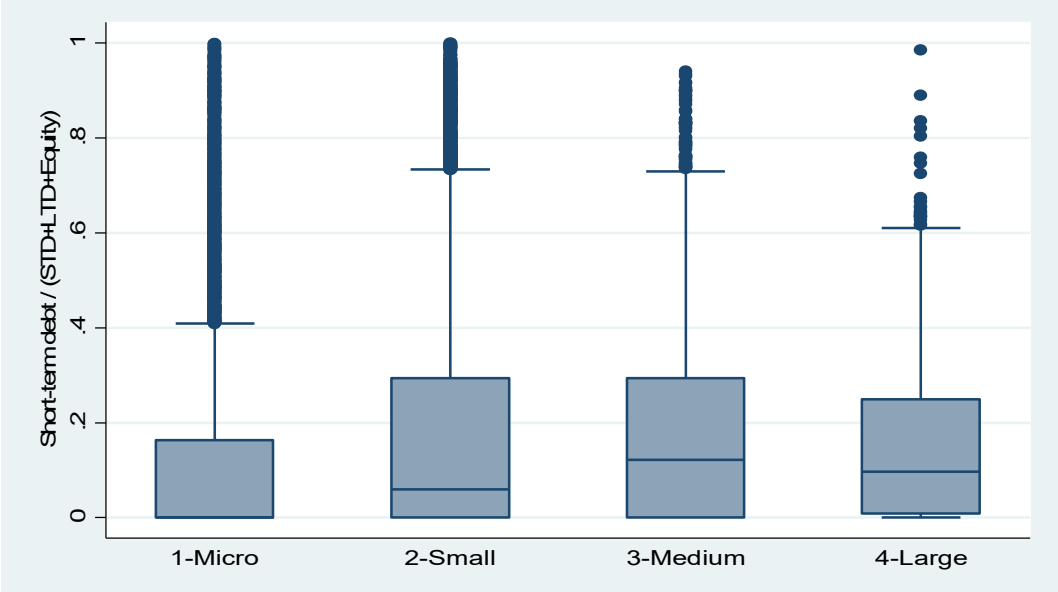


Illustration 1 – Question 4

```
. graph box LEV_LT, over(GROUPS)
```

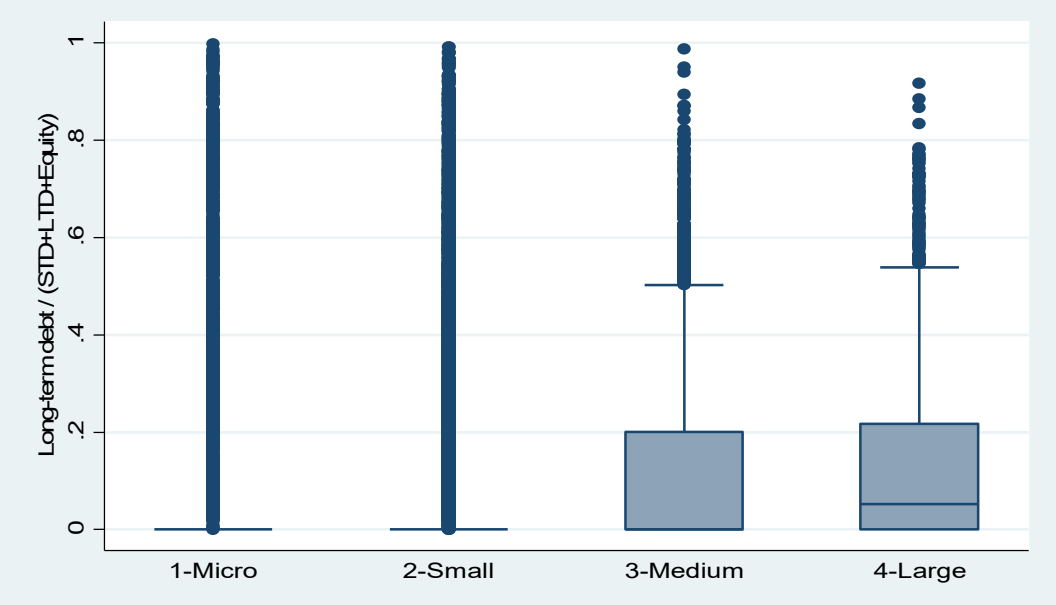


Illustration 1 – Question 5

```
. tabulate GROUPS

     GROUPS |      Freq.     Percent        Cum.
------------+-----------------------------------
    1-Micro |      9,874       30.64       30.64
    2-Small |     13,396       41.57       72.21
   3-Medium |      7,034       21.83       94.04
    4-Large |      1,922        5.96      100.00
------------+-----------------------------------
      Total |     32,226      100.00
```

Illustration 1 – Question 6

```
. tabulate GROUPS YEAR
```

GROUPS	Ano 1995	1996	1997	1998	Total
1-Micro	1,418	1,441	1,414	1,431	9,874
2-Small	1,845	1,893	1,907	1,946	13,396
3-Medium	967	968	1,006	1,030	7,034
4-Large	279	267	271	277	1,922
Total	4,509	4,569	4,598	4,684	32,226

GROUPS	Ano 1999	2000	2001	Total
1-Micro	1,446	1,392	1,332	9,874
2-Small	1,951	1,950	1,904	13,396
3-Medium	1,024	1,019	1,020	7,034
4-Large	271	283	274	1,922
Total	4,692	4,644	4,530	32,226

Illustration 1 – Question 7.1

```
. table GROUPS, contents(mean LEV_ST mean LEV_LT mean LEV)

----------------------------------------------------------
GROUPS      | mean(LEV_ST)   mean(LEV_LT)      mean(LEV)
------------+---------------------------------------------
   1-Micro  |      .1288624       .0417675       .1706299
   2-Small  |      .1721035       .0750359       .2471394
  3-Medium  |      .1752302       .1147087        .289939
   4-Large  |      .1527809       .1318627       .2846436
----------------------------------------------------------
```

Illustration 1 – Question 7.3

```
. gen DEBT_ST=LEV_ST>0
. gen DEBT_LT=LEV_LT>0
. gen DEBT=LEV>0

. table GROUPS if DEBT_ST==1, contents(mean LEV_ST)
. table GROUPS if DEBT_LT==1, contents(mean LEV_LT)
. table GROUPS if DEBT==1, contents(mean LEV)
```

```
--------------------------
GROUPS      | mean(LEV_ST)
------------+-------------
   1-Micro  |    .3578141
   2-Small  |    .2961462
  3-Medium  |    .2380397
   4-Large  |    .1882339
--------------------------
```

```
--------------------------
GROUPS      | mean(LEV_LT)
------------+-------------
   1-Micro  |    .4429778
   2-Small  |    .3261457
  3-Medium  |    .2472001
   4-Large  |    .2215386
--------------------------
```

```
----------------------
GROUPS      | mean(LEV)
------------+---------
   1-Micro  |  .4065635
   2-Small  |  .3703635
  3-Medium  |  .3566685
   4-Large  |  .3323724
----------------------
```

Illustration 1 – Question 8

```
. cor LEV_LT COLLAT SIZE PROF GROWTH
(obs=32,226)

             |   LEV_LT   COLLAT     SIZE     PROF   GROWTH
-------------+------------------------------------------------
      LEV_LT |   1.0000
      COLLAT |   0.1071   1.0000
        SIZE |   0.2716   0.1349   1.0000
        PROF |  -0.0779  -0.1060  -0.1281   1.0000
      GROWTH |   0.0076  -0.0248  -0.0049   0.0325   1.0000
```

Illustration 1 – Question 9

```
. ttest LEV_LT==0.075

One-sample t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
  LEV_LT |  32,226    .0768912    .0009351    .1678684    .0750583     .078724
------------------------------------------------------------------------------
    mean = mean(LEV_LT)                                           t =   2.0224
Ho: mean = 0.075                                 degrees of freedom =    32225

  Ha: mean < 0.075            Ha: mean != 0.075            Ha: mean > 0.075
 Pr(T < t) = 0.9784         Pr(|T| > |t|) = 0.0431          Pr(T > t) = 0.0216
```

Illustration 1 – Question 10

```
. ttest LEV_LT , by(LE)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       0 |  30,304    .0734047    .0009585    .1668595    .0715259    .0752834
       1 |   1,922    .1318627    .0039709    .1740878     .124075    .1396505
---------+--------------------------------------------------------------------
combined |  32,226    .0768912    .0009351    .1678684    .0750583     .078724
---------+--------------------------------------------------------------------
    diff |           -.0584581    .0039352               -.0661713   -.0507449
------------------------------------------------------------------------------
    diff = mean(0) - mean(1)                                      t = -14.8551
Ho: diff = 0                                     degrees of freedom =    32224

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 1.0000
```

Illustration 1 – Question 11

```
. oneway LEV_LT GROUPS, bonferroni

                        Analysis of Variance
    Source              SS         df      MS             F     Prob > F
------------------------------------------------------------------------
Between groups      28.0952331      3    9.3650777    342.91     0.0000
 Within groups      879.998687  32222    .027310492
------------------------------------------------------------------------
    Total            908.09392  32225    .028179796

Bartlett's test for equal variances:  chi2(3) = 196.0970  Prob>chi2 = 0.000
```

Illustration 1 – Question 11

Comparison of Long-term debt / (STD+LTD+Equity) by GROUPS
(Bonferroni)

Row Mean- Col Mean	1-Micro	2-Small	3-Medium
2-Small	.033268 0.000		
3-Medium	.072941 0.000	.039673 0.000	
4-Large	.090095 0.000	.056827 0.000	.017154 0.000

Illustration 1 – Question 12

```
. kwallis LEV_LT, by(GROUP)

Kruskal-Wallis equality-of-populations rank test

  +-------------------------------+
  |   GROUPS |    Obs | Rank Sum |
  |----------+--------+----------|
  |  1-Micro |  9,874 | 1.34e+08 |
  |  2-Small | 13,396 | 2.10e+08 |
  | 3-Medium |  7,034 | 1.35e+08 |
  |  4-Large |  1,922 | 4.03e+07 |
  +-------------------------------+

chi-squared =  2050.033 with 3 d.f.
probability =     0.0001

chi-squared with ties =  3434.319 with 3 d.f.
probability =     0.0001
```

Illustration 2 – Question 1

```
. summarize item13-item24

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
      item13 |      1,419    4.450317    .7374944         1          5
      item14 |      1,424    4.516854     .709139         1          5
      item15 |      1,424    4.434691    .7478835         1          5
      item16 |      1,420    4.270423    .8387034         1          5
      item17 |      1,423    4.158819    .8969815         1          5
-------------+---------------------------------------------------------
      item18 |      1,424    3.924157    1.032095         1          5
      item19 |      1,420    4.072535    .9665034         1          5
      item20 |      1,396    3.770774    .9137137         1          5
      item21 |      1,422    3.769339    .9863042         1          5
      item22 |      1,414    3.592645    1.122807         1          5
-------------+---------------------------------------------------------
      item23 |      1,423    3.800422    .9639492         1          5
      item24 |      1,417    3.653493    .9308223         1          5
```

Illustration 2 – Question 2

```
 . correlate item13-item24
(obs=1,365)

             |   item13   item14   item15   item16   item17   item18   item19   item20   item21   item22
-------------+---------------------------------------------------------------------------------------------
      item13 |   1.0000
      item14 |   0.6615   1.0000
      item15 |   0.6000   0.6346   1.0000
      item16 |   0.5663   0.5000   0.5053   1.0000
      item17 |   0.5769   0.5515   0.5866   0.5865   1.0000
      item18 |   0.4090   0.4331   0.4571   0.4048   0.5547   1.0000
      item19 |   0.2863   0.3204   0.3587   0.3354   0.4493   0.6266   1.0000
      item20 |   0.3042   0.3148   0.3557   0.3168   0.4168   0.5205   0.4465   1.0000
      item21 |   0.4755   0.4490   0.5090   0.4524   0.5953   0.5542   0.4992   0.4248   1.0000
      item22 |   0.3325   0.3331   0.3688   0.3626   0.4498   0.5361   0.4840   0.3830   0.5065   1.0000
      item23 |   0.5640   0.5646   0.5823   0.4588   0.6130   0.5695   0.4440   0.4096   0.5975   0.4932
      item24 |   0.4536   0.4428   0.4348   0.4297   0.5206   0.4738   0.3738   0.3572   0.4998   0.4444

             |   item23   item24
-------------+------------------
      item23 |   1.0000
      item24 |   0.7046   1.0000
```

Illustration 2 – Question 3

```
. quietly factor item13-item24
. estat kmo
Kaiser-Meyer-Olkin measure of sampling adequacy
    ----------------------
        Variable |     kmo
    -------------+--------
          item13 |  0.9225
          item14 |  0.9236
          item15 |  0.9449
          item16 |  0.9409
          item17 |  0.9547
          item18 |  0.9251
          item19 |  0.9157
          item20 |  0.9579
          item21 |  0.9611
          item22 |  0.9571
          item23 |  0.9119
          item24 |  0.9094
    -------------+--------
         Overall |  0.9344
    ----------------------
```

Illustration 2 – Question 4

```
. factor item13-item24, pcf
(obs=1,365)
Factor analysis/correlation                       Number of obs    =      1,365
    Method: principal-component factors           Retained factors =          2
    Rotation: (unrotated)                         Number of params =         23
```

Factor	Eigenvalue	Difference	Proportion	Cumulative
Factor1	6.24915	5.01967	0.5208	0.5208
Factor2	1.22948	0.51049	0.1025	0.6232
Factor3	0.71899	0.10586	0.0599	0.6831
Factor4	0.61313	0.05196	0.0511	0.7342
Factor5	0.56116	0.05817	0.0468	0.7810
Factor6	0.50299	0.03173	0.0419	0.8229
Factor7	0.47126	0.08245	0.0393	0.8622
Factor8	0.38882	0.02091	0.0324	0.8946
Factor9	0.36790	0.03970	0.0307	0.9252
Factor10	0.32820	0.01082	0.0274	0.9526
Factor11	0.31738	0.06584	0.0264	0.9790
Factor12	0.25154	.	0.0210	1.0000

```
LR test: independent vs. saturated:  chi2(66) = 8683.10 Prob>chi2 = 0.0000
```

Illustration 2 – Question 4

Factor loadings (pattern matrix) and unique variances

Variable	Factor1	Factor2	Uniqueness
item13	0.7273	-0.4492	0.2693
item14	0.7238	-0.4077	0.3100
item15	0.7462	-0.3081	0.3482
item16	0.6851	-0.2814	0.4515
item17	0.8065	-0.1053	0.3385
item18	0.7551	0.3659	0.2959
item19	0.6410	0.4970	0.3421
item20	0.5927	0.3779	0.5059
item21	0.7634	0.1345	0.3992
item22	0.6514	0.3645	0.4429
item23	0.8191	-0.0403	0.3274
item24	0.7137	0.0047	0.4906

.

Illustration 2 – Question 5

```
. screeplot
```

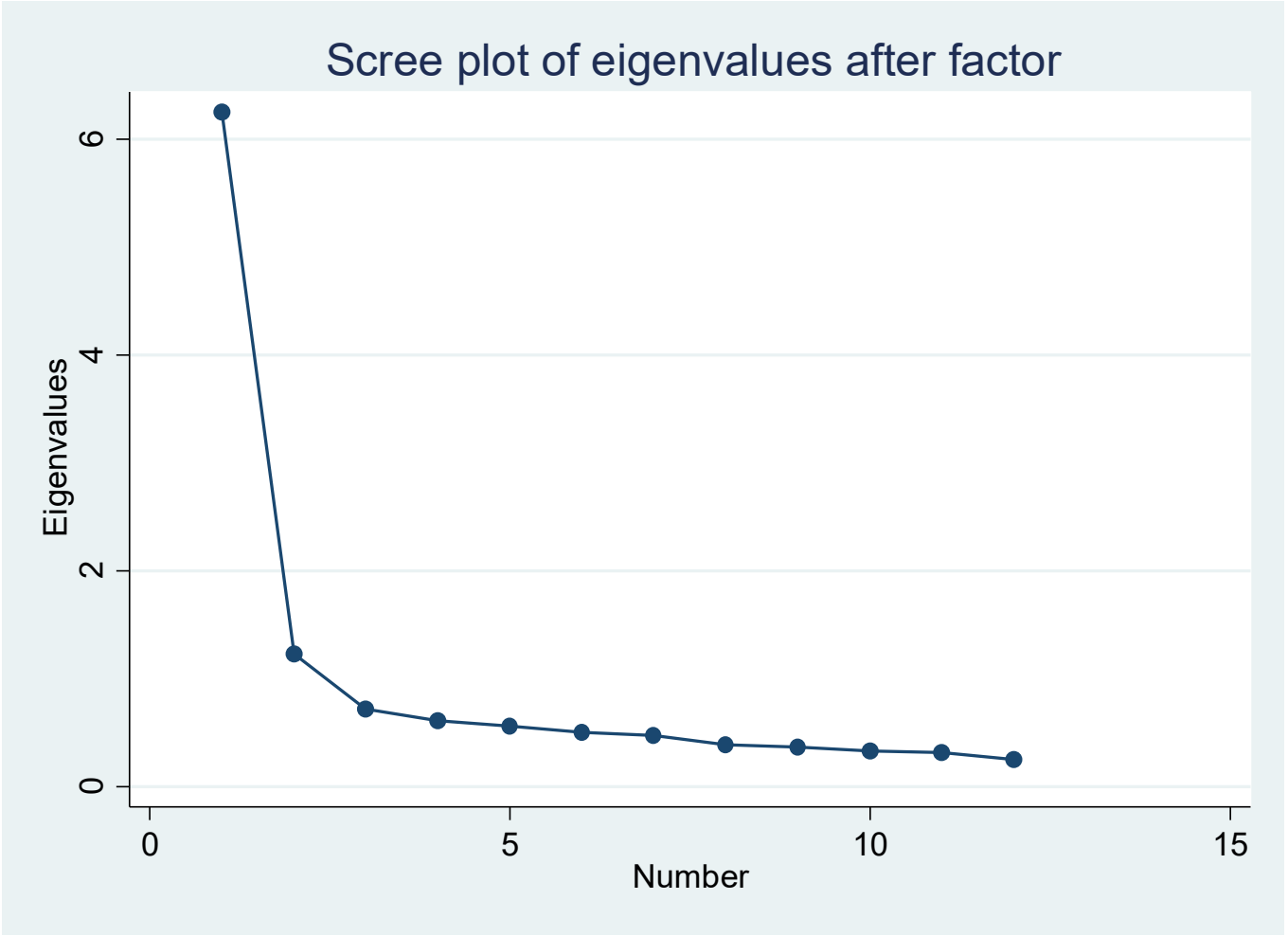


Illustration 2 – Question 5

```
. loadingplot
```

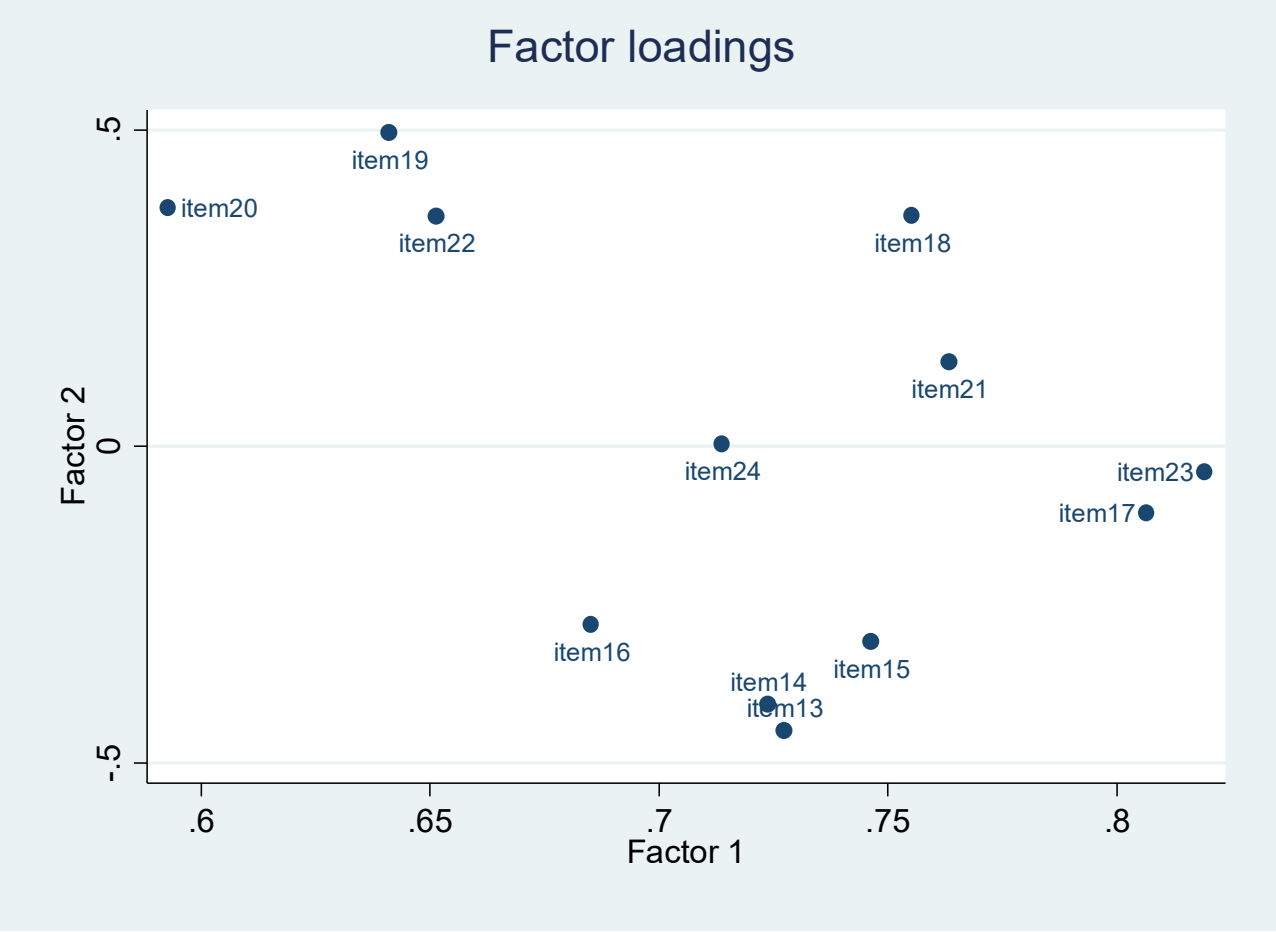


Illustration 2 – Question 6

```
. rotate
Factor analysis/correlation                       Number of obs    =      1,365
    Method: principal-component factors           Retained factors =          2
    Rotation: orthogonal varimax (Kaiser off)     Number of params =         23
    ------------------------------------------------------------------------------
         Factor  |     Variance   Difference        Proportion   Cumulative
    -------------+----------------------------------------------------------------
        Factor1  |      4.02937      0.58011            0.3358       0.3358
        Factor2  |      3.44926            .            0.2874       0.6232
    ------------------------------------------------------------------------------
    LR test: independent vs. saturated:  chi2(66) = 8683.10 Prob>chi2 = 0.0000

Rotated factor loadings (pattern matrix) and unique variances
    ------------------------------------------------
        Variable |  Factor1   Factor2 |   Uniqueness
    -------------+--------------------+--------------
          item13 |   0.8419    0.1482 |      0.2693
          item14 |   0.8116    0.1769 |      0.3100
          item15 |   0.7622    0.2661 |      0.3482
          item16 |   0.6988    0.2455 |      0.4515
          item17 |   0.6723    0.4577 |      0.3385
          item18 |   0.3206    0.7755 |      0.2959
          item19 |   0.1482    0.7974 |      0.3421
...
```

Illustration 2 – Question 6

```
...
          item20 |   0.1913    0.6764 |      0.5059
          item21 |   0.4806    0.6081 |      0.3992
          item22 |   0.2441    0.7054 |      0.4429
          item23 |   0.6386    0.5146 |      0.3274
          item24 |   0.5299    0.4781 |      0.4906
    -------------------------------------------------

Factor rotation matrix

    --------------------------------
                 | Factor1  Factor2
    -------------+------------------
         Factor1 |  0.7468   0.6650
         Factor2 | -0.6650   0.7468
    --------------------------------
```

Illustration 2 – Question 7

```
. predict factor1 factor2
(regression scoring assumed)

Scoring coefficients (method = regression; based on varimax rotated factors)
```

Variable	Factor1	Factor2
item13	0.32987	-0.19547
item14	0.30699	-0.17061
item15	0.25584	-0.10776
item16	0.23407	-0.09801
item17	0.15331	0.02188
item18	-0.10768	0.30264
item19	-0.19221	0.37011
item20	-0.13357	0.29264
item21	0.01846	0.16296
item22	-0.11929	0.29072
item23	0.11969	0.06269
item24	0.08277	0.07878

- See Factor 1 and Factor 2 in the database

Illustration 2 – Question 8

```
. factor item13-item24, ml
(obs=1,365)
…
Factor analysis/correlation                       Number of obs    =      1,365
    Method: maximum likelihood                    Retained factors =          7
    Rotation: (unrotated)                         Number of params =         63
                                                  Schwarz's BIC    =    456.529
    Log likelihood = -.8686265                    (Akaike's) AIC   =    127.737

    Beware: solution is a Heywood case
            (i.e., invalid or boundary values of uniqueness)
    ------------------------------------------------------------------------------
         Factor  |   Eigenvalue   Difference        Proportion   Cumulative
    -------------+----------------------------------------------------------------
        Factor1  |      4.21187      2.15154            0.5276       0.5276
        Factor2  |      2.06033      1.18994            0.2581       0.7857
        Factor3  |      0.87039      0.45198            0.1090       0.8948
        Factor4  |      0.41841      0.22126            0.0524       0.9472
        Factor5  |      0.19715      0.06378            0.0247       0.9719
        Factor6  |      0.13337      0.04225            0.0167       0.9886
        Factor7  |      0.09112            .            0.0114       1.0000
    ------------------------------------------------------------------------------
    LR test: independent vs. saturated:  chi2(66) = 8683.10 Prob>chi2 = 0.0000
    LR test:   7 factors vs. saturated:  chi2(3)  =    1.73 Prob>chi2 = 0.6314
    (tests formally not valid because a Heywood case was encountered)
```

Illustration 2 – Question 8

Factor loadings (pattern matrix) and unique variances

Variable	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7	Uniqueness
item13	0.5769	0.5103	-0.4464	0.1041	-0.1767	-0.0115	-0.0372	0.1638
item14	0.5515	0.4687	-0.2638	0.0664	0.1126	-0.0754	0.0416	0.3820
item15	0.5866	0.4566	-0.1898	0.0938	0.3669	0.0204	0.0243	0.2669
item16	0.5865	0.2939	-0.1539	0.0928	-0.0504	0.1558	0.1977	0.4715
item17	1.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	0.0000
item18	0.5547	0.4450	0.3928	0.2596	-0.0523	-0.1396	0.0214	0.2499
item19	0.4493	0.3489	0.4064	0.2587	-0.0274	0.0350	0.0008	0.4423
item20	0.4168	0.3131	0.2581	0.1715	-0.0074	-0.0058	0.0374	0.6307
item21	0.5953	0.4010	0.1766	0.0506	0.0100	0.2381	-0.1427	0.3739
item22	0.4498	0.3745	0.2983	0.0742	-0.0532	0.1436	0.0056	0.5395
item23	0.6130	0.5763	0.1133	-0.2951	0.0325	-0.0597	-0.0728	0.1822
item24	0.5206	0.4840	0.1508	-0.3562	-0.0924	0.0352	0.1442	0.3145