Ekonometrija 1

Prvi seminar:
Uvod v programski paket Stata.

Na prvem seminarju bomo najprej spoznali zasnovo in osnove dela s programskim paketom Stata. Srečali se bomo z dvema pristopoma k analizi podatkov. Na konkretnem primeru si bomo pogledali pregled in urejanje podatkov, kreiranje različnih diagramov, delo s skalarji in matrikami, transformiranje spremenljivk, uporabo statističnih porazdelitev in testiranje ničelnih hipotez. Nato si bomo pogledali še primer časovno serije, kjer se bomo osredotočili na opredelitev časovne dimenzije, kreiranje periodičnih komponent, uporabo nepravih spremenljivk in uporabo odlogov.

♣-♣-♣

Primer 1: Na voljo imamo podatke za spremenljivke y, x_1, x_2 in x_3. Za navedene spremenljivke imamo po 8 opazovanj, kot je prikazano v tabeli. Podatki se že nahajajo v podatkovni datoteki `osnove_state.dta`, programska koda, ki jo boste potrebovali, pa v datoteki `osnove_state-ukazi.do`.

i	1	2	3	4	5	6	7	8
y_i	2	2	1	5	–4	1	4	1
x_{1i}	1	1	1	1	1	1	1	1
x_{2i}	1	2	0	–1	1	–1	–2	0
x_{3i}	–1	–1	2	–4	3	0	2	–1

a) Proučite podatke s pomočjo različnih ukazov za pregled podatkov. Kako bi najlažje ročno uredili podatke za konkretne spremenljivke in konkretna opazovanja v vaši bazi?
b) Proučite podatke še grafično s pomočjo različnih diagramov. Uporabite razsevni diagram, linijski diagram in histogram.
c) Na osnovi obstoječih spremenljivk iz podatkovne baze z različnimi transformacijami generirajte nekaj novih spremenljivk. Uporabite množenje, absolutne vrednosti, logaritmiranje, antilogaritmiranje ter standardiziranje.
d) Prikličite iz Statinega spomina rezultate izvedbe enostavnejšega ukaza `summarize` ter kompleksnejšega ukaza `regress`. Kako bi jih shranili za kasnejšo uporabo?
e) Prikažite kovariančno in korelacijsko matriko spremenljivk y, x_2 in x_3. Ugotovite tudi statistično značilnost izračunanih korelacijskih koeficientov.

Izpis rezultatov obdelav v programskem paketu Stata:

a) Pregled podatkov

. describe

```
Contains data
  obs:             8
 vars:             5
-------------------------------------------------------------------------------
              storage  display     value
variable name   type   format      label      variable label
-------------------------------------------------------------------------------
obs             float   %9.0g                 Opazovanje
y               float   %9.0g                 Spremenljivka y
x1              float   %9.0g                 Spremenljivka x1
x2              float   %9.0g                 Spremenljivka x2
x3              float   %9.0g                 Spremenljivka x3
-------------------------------------------------------------------------------
Sorted by:
     Note:  dataset has changed since last saved
```

. inspect x2

```
x2:  Spremenljivka x2                         Number of Observations
---------------------                    ------------------------------
                                         Total   Integers   Nonintegers
|      #   #   #            Negative         3          3           -
|      #   #   #            Zero             2          2           -
|      #   #   #            Positive         3          3           -
|  #   #   #   #   #                     -----      -----       -----
|  #   #   #   #   #        Total            8          8           -
|  #   #   #   #   #        Missing          -
+----------------------                  -----
-2                    2                     8
   (5 unique values)
```

. sum y x1 x2 x3

```
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
           y |         8         1.5    2.672612        -4          5
          x1 |         8           1           0         1          1
          x2 |         8           0    1.309307        -2          2
          x3 |         8           0    2.267787        -4          3
```

. sum y, detail

```
                        Spremenljivka y
-------------------------------------------------------------
      Percentiles      Smallest
 1%           -4             -4
 5%           -4              1
10%           -4              1       Obs                   8
25%            1              1       Sum of Wgt.           8

50%          1.5                      Mean                1.5
                        Largest       Std. Dev.      2.672612
75%            3              2
90%            5              2       Variance       7.142857
95%            5              4       Skewness          -.864
99%            5              5       Kurtosis         3.5344
```

. tabstat y x1 x2 x3, stat(N mean sd median sum min max)

```
   stats |         y        x1        x2        x3
---------+------------------------------------------
       N |         8         8         8         8
    mean |       1.5         1         0         0
      sd |  2.672612         0  1.309307  2.267787
     p50 |       1.5         1         0       -.5
     sum |        12         8         0         0
     min |        -4         1        -2        -4
     max |         5         1         2         3
--------------------------------------------------
```

. tab y

```
Spremenljivka y |      Freq.     Percent        Cum.
----------------+-----------------------------------
             -4 |          1       12.50       12.50
              1 |          3       37.50       50.00
              2 |          2       25.00       75.00
              4 |          1       12.50       87.50
              5 |          1       12.50      100.00
----------------+-----------------------------------
          Total |          8      100.00
```

. tab y x3

```
Spremenljivka y |                    Spremenljivka x3
                |        -4         -1          0          2          3 |     Total
----------------+---------------------------------------------------------+----------
             -4 |         0          0          0          0          1 |         1
              1 |         0          1          1          1          0 |         3
              2 |         0          2          0          0          0 |         2
              4 |         0          0          0          1          0 |         1
              5 |         1          0          0          0          0 |         1
----------------+---------------------------------------------------------+----------
          Total |         1          3          1          2          1 |         8
```

. list, N mean sum

```
      +--------------------+
      |   y   x1   x2   x3 |
      |--------------------|
   1. |   2    1    1   -1 |
   2. |   2    1    2   -1 |
   3. |   1    1    0    2 |
   4. |   5    1   -1   -4 |
   5. |  -4    1    1    3 |
      |--------------------|
   6. |   1    1   -1    0 |
   7. |   4    1   -2    2 |
   8. |   1    1    0   -1 |
      |--------------------|
 Mean | 1.5    1    0    0 |
  Sum |  12    8    0    0 |
    N |   8    8    8    8 |
      +--------------------+
```

. list in 4/8

```
     +-------------------+
     |  y   x1   x2   x3 |
     |-------------------|
  4. |  5    1   -1   -4 |
  5. | -4    1    1    3 |
     +-------------------+
```

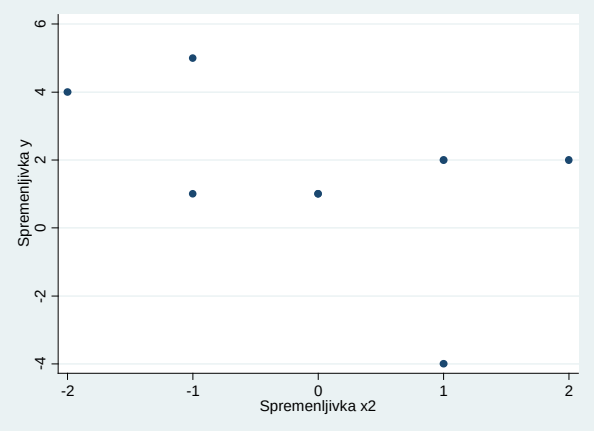
```
  6. |  1    1   -1    0 |
  7. |  4    1   -2    2 |
  8. |  1    1    0   -1 |
     +-------------------+
```

```
. list x1 x2 if x2>=0
```

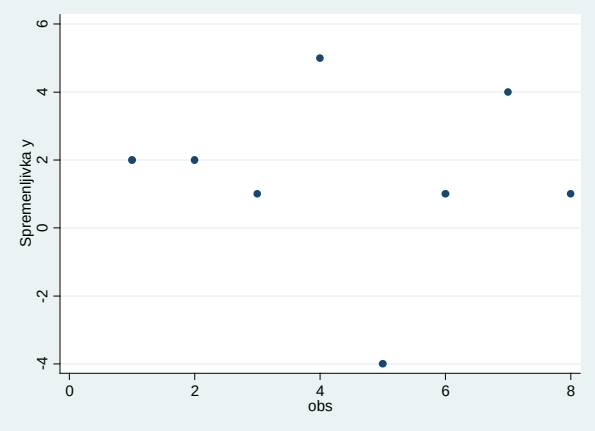
```
     +---------+
     | x1   x2 |
     |---------|
  1. |  1    1 |
  2. |  1    2 |
  3. |  1    0 |
  5. |  1    1 |
  8. |  1    0 |
     +---------+
```

b) Diagrami v Stati

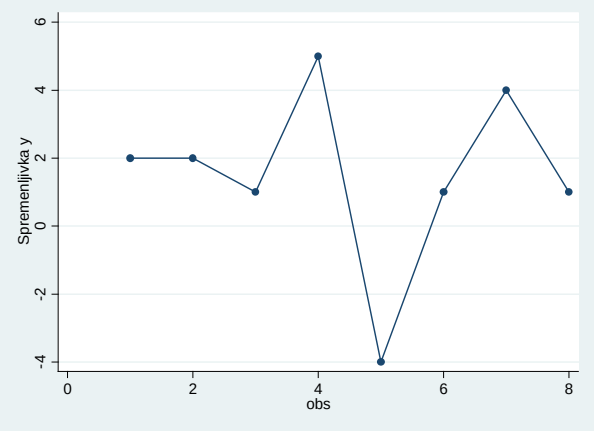
```
. twoway scatter y x2
```



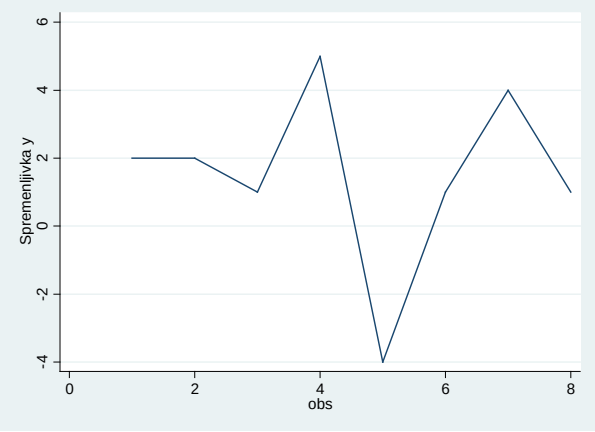

```
. twoway scatter y obs
```



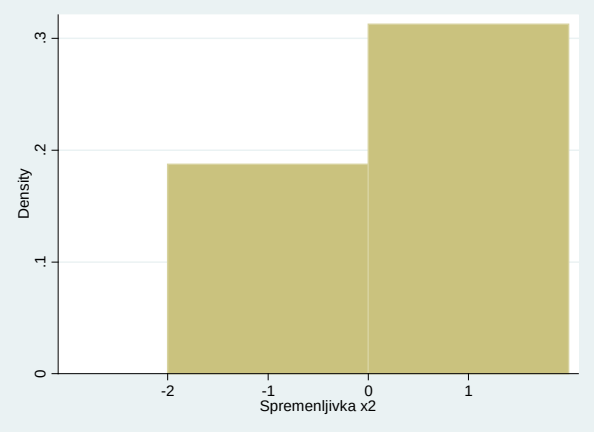

```
. twoway connected y obs
```



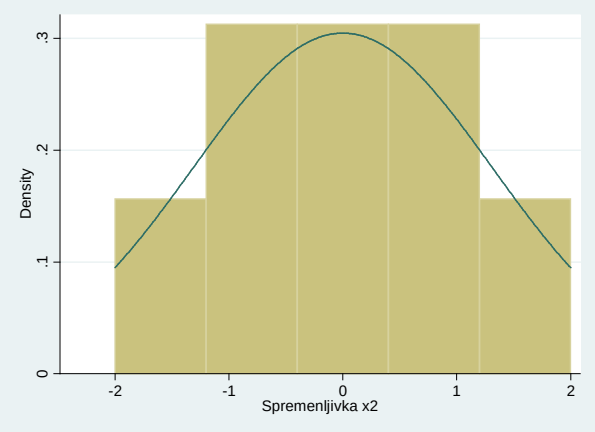

```
. twoway line y obs
```



```
. hist x2
(bin=2, start=-2, width=2)
```




```
. hist x2, bin(5) normal
(bin=5, start=-2, width=.8)
```



c) Generiranje novih spremenljivk

```
. gen yx2=100*y*x2

. gen x2kv=x2^2

. gen x2a=abs(x2)

. gen lx2=log(x2)
(5 missing values generated)

. gen ex2=exp(lx2)
(5 missing values generated)

. list
```

	y	x1	x2	x3	obs	yx2	x2kv	x2a	lx2	ex2
1.	2	1	1	-1	1	200	1	1	0	1
2.	2	1	2	-1	2	400	4	2	.6931472	2
3.	1	1	0	2	3	0	0	0	.	.
4.	5	1	-1	-4	4	-500	1	1	.	.
5.	-4	1	1	3	5	-400	1	1	0	1
6.	1	1	-1	0	6	-100	1	1	.	.
7.	4	1	-2	2	7	-800	4	2	.	.
8.	1	1	0	-1	8	0	0	0	.	.

```
. replace lx2=0 if lx2==.
(5 real changes made)

. egen x2s=std(x2)

. list lx2 x2s
```

	lx2	x2s
1.	0	.7637626
2.	.6931472	1.527525
3.	0	0

```
  4. |        0   -.7637626 |
  5. |        0    .7637626 |
     |----------------------|
  6. |        0   -.7637626 |
  7. |        0   -1.527525 |
  8. |        0           0 |
     +----------------------+
```

. drop obs yx2 x2kv x2a lx2 ex2 x2s

d) Priklic podatkov iz Statinega spomina

. sum y

```
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
           y |         8         1.5    2.672612         -4          5
```

. return list

```
scalars:
                  r(N) =  8
              r(sum_w) =  8
               r(mean) =  1.5
                r(Var) =  7.142857142857143
                 r(sd) =  2.672612419124244
               r(min) =  -4
               r(max) =  5
               r(sum) =  12
```

. regress y x2 x3

```
      Source |       SS       df       MS              Number of obs =       8
-------------+------------------------------           F(  2,     5) =    4.54
       Model |       32.25     2      16.125           Prob > F      =  0.0751
    Residual |       17.75     5        3.55           R-squared     =  0.6450
-------------+------------------------------           Adj R-squared =  0.5030
       Total |          50     7  7.14285714           Root MSE      =  1.8841

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x2 |         -1   .5439056    -1.84   0.125    -2.398154    .3981539
          x3 |       -.75   .3140241    -2.39   0.063    -1.557225    .0572245
       _cons |        1.5   .6661456     2.25   0.074    -.2123819    3.212382
------------------------------------------------------------------------------
```

. ereturn list

```
scalars:
                  e(N) =  8
               e(df_m) =  2
               e(df_r) =  5
                  e(F) =  4.542253521126761
                 e(r2) =  .645
               e(rmse) =  1.884144368141677
                e(mss) =  32.25
                e(rss) =  17.75
               e(r2_a) =  .503
                 e(ll) =  -14.53928416260374
               e(ll_0) =  -18.68183412063062
               e(rank) =  3
```

```
macros:
            e(cmdline) : "regress y x2 x3"
              e(title) : "Linear regression"
          e(marginsok) : "XB default"
                e(vce) : "ols"
             e(depvar) : "y"
                e(cmd) : "regress"
         e(properties) : "b V"
            e(predict) : "regres_p"
              e(model) : "ols"
          e(estat_cmd) : "regress_estat"

matrices:
                  e(b) :  1 x 3
                  e(V) :  3 x 3

functions:
             e(sample)
```

```
. scalar r2=e(r2)
. display r2
.645

. matrix varcov=e(V)
. matrix list varcov

symmetric varcov[3,3]
              x2         x3      _cons
   x2  .29583333
   x3          0  .09861111
_cons          0          0     .44375

. scalar varcov22=varcov[2,2]
. display varcov22
.09861111
```

e) Kovariance in korelacija

```
. correlate y x2 x3, covariance
(obs=8)

             |        y       x2       x3
-------------+---------------------------
           y |  7.14286
          x2 | -1.71429  1.71429
          x3 | -3.85714        0  5.14286

. correlate y x2 x3
(obs=8)

             |        y       x2       x3
-------------+---------------------------
           y |   1.0000
          x2 |  -0.4899   1.0000
          x3 |  -0.6364   0.0000   1.0000
```

```
. pwcorr y x2 x3, sig

             |        y       x2       x3
-------------+---------------------------
           y |   1.0000
             |
             |
          x2 |  -0.4899   1.0000
             |   0.2178
             |
          x3 |  -0.6364   0.0000   1.0000
             |   0.0898   1.0000
             |

. clear all
```

■

Primer 2: V priloženi podatkovni datoteki `osnove_casovnih_vrst.dta` se nahaja časovna vrsta z začetkom v letu 1950. Programska koda, ki jo boste potrebovali, se nahaja v datoteki `osnove_casovnih_vrst-ukazi.do`.

a) Odprite v programskem paketu Stata podatkovno datoteko. Proučite podatke s pomočjo različnih ukazov za pregled podatkov.
b) Opredelite časovno dimenzijo podatkov. Nato sortirajte podatkovno bazo po časovni dimenziji ter zamenjajte vrstni red spremenljivk. Generirajte trend, neprave spremenljivke za četrtletja in ciklično komponento.
c) Generirajte nepravo spremenljivko, ki zavzame vrednost 1, če ima naša spremenljivka vrednost, ki je večja ali enaka 80 % njene mediane ali pa manjša od dveh tretjin njene aritmetične sredine, v ostalih primerih pa zavzame vrednost 0.
d) Generirajte prve in četrte odloge naše spremenljivke ter druge vodeče odloge. Generirajte še prve diference naše spremenljivke.

Izpis rezultatov obdelav v programskem paketu Stata:

a) Pregled podatkov

```
. inspect spr

spr:  Casovna spremenljivka (v enotah mere)     Number of Observations
------------------------------------------- ------------------------------
                                              Total   Integers   Nonintegers
|  #                          Negative            -          -          -
|  #                          Zero                -          -          -
|  #   #                      Positive          204         18        186
|  #   #   #   #                                -----      -----      -----
|  #   #   #   #              Total             204         18        186
|  #   #   #   #   #          Missing             -
+----------------------                        -----
1610.5           9303.9                         204
(More than 99 unique values)
```

```
. tabstat spr, stat(N mean sd median sum min max)

    variable |        N      mean        sd       p50       sum       min       max
-------------+--------------------------------------------------------------------------
         spr |      204  4562.646  2113.962    4142.2  930779.7    1610.5    9303.9
----------------------------------------------------------------------------------------
```

b) Opredelitev casovne dimenzije in generiranje periodicnih komponent

```
. tsset kvartal
        time variable:  kvartal, 1950q1 to 2000q4
                delta:  1 quarter

. sort kvartal
. order kvartal, last
. order kvartal, first

. gen t=_n

. gen q=quarter(dofq(kvartal))

. tabulate q, gen(d)

          q |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |         51       25.00       25.00
          2 |         51       25.00       50.00
          3 |         51       25.00       75.00
          4 |         51       25.00      100.00
------------+-----------------------------------
      Total |        204      100.00

. drop q

. gen t2=t^2
. gen t3=t^3

. list

     +-----------------------------------------------------------------+
     | kvartal      spr     t   d1   d2   d3   d4      t2        t3 |
     |-----------------------------------------------------------------|
  1. |  1950q1   1610.5     1    1    0    0    0       1         1 |
  2. |  1950q2   1658.8     2    0    1    0    0       4         8 |
  3. |  1950q3     1723     3    0    0    1    0       9        27 |
  4. |  1950q4   1753.9     4    0    0    0    1      16        64 |
  5. |  1951q1   1773.5     5    1    0    0    0      25       125 |
     |-----------------------------------------------------------------|
  6. |  1951q2   1803.7     6    0    1    0    0      36       216 |
  7. |  1951q3   1839.8     7    0    0    1    0      49       343 |
  8. |  1951q4   1843.3     8    0    0    0    1      64       512 |
  9. |  1952q1   1864.7     9    1    0    0    0      81       729 |
 10. |  1952q2   1866.2    10    0    1    0    0     100      1000 |
     |-----------------------------------------------------------------|
 11. |  1952q3     1878    11    0    0    1    0     121      1331 |
 12. |  1952q4   1940.2    12    0    0    0    1     144      1728 |
 13. |  1953q1     1976    13    1    0    0    0     169      2197 |
 14. |  1953q2   1992.2    14    0    1    0    0     196      2744 |
 15. |  1953q3   1979.5    15    0    0    1    0     225      3375 |
     |-----------------------------------------------------------------|
 16. |  1953q4   1947.8    16    0    0    0    1     256      4096 |
 17. |  1954q1   1938.1    17    1    0    0    0     289      4913 |

...
```

```
189. |  1997q1   8016.4   189    1    0    0    0   35721   6751269 |
190. |  1997q2   8131.9   190    0    1    0    0   36100   6859000 |
     |-----------------------------------------------------------------|
191. |  1997q3   8216.6   191    0    0    1    0   36481   6967871 |
192. |  1997q4   8272.9   192    0    0    0    1   36864   7077888 |
193. |  1998q1   8396.3   193    1    0    0    0   37249   7189057 |
194. |  1998q2   8442.9   194    0    1    0    0   37636   7301384 |
195. |  1998q3   8528.5   195    0    0    1    0   38025   7414875 |
     |-----------------------------------------------------------------|
196. |  1998q4   8667.9   196    0    0    0    1   38416   7529536 |
197. |  1999q1   8733.5   197    1    0    0    0   38809   7645373 |
198. |  1999q2   8771.2   198    0    1    0    0   39204   7762392 |
199. |  1999q3   8871.5   199    0    0    1    0   39601   7880599 |
200. |  1999q4   9049.9   200    0    0    0    1   40000   8000000 |
     |-----------------------------------------------------------------|
201. |  2000q1   9102.5   201    1    0    0    0   40401   8120601 |
202. |  2000q2   9229.4   202    0    1    0    0   40804   8242408 |
203. |  2000q3   9260.1   203    0    0    1    0   41209   8365427 |
204. |  2000q4   9303.9   204    0    0    0    1   41616   8489664 |
     +-----------------------------------------------------------------+
```

. keep kvartal spr

c) Generiranje neprave spremenljivke

. sum spr, detail

```
            Casovna spremenljivka (v enotah mere)
-------------------------------------------------------------
      Percentiles      Smallest
 1%         1723         1610.5
 5%         1878         1658.8
10%       2058.1           1723       Obs                 204
25%      2600.05         1753.9       Sum of wgt.         204

50%       4142.2                      Mean           4562.646
                        Largest       Std. dev.      2113.962
75%      6312.85         9102.5
90%       7621.9         9229.4       Variance        4468837
95%       8442.9         9260.1       Skewness       .4680201
99%       9229.4         9303.9       Kurtosis       2.133958
```

. return list

```
scalars:
                  r(N) =  204
              r(sum_w) =  204
               r(mean) =  4562.645581413718
                r(Var) =  4468836.741843408
                 r(sd) =  2113.962332172314
           r(skewness) =  .4680200721224719
           r(kurtosis) =  2.133957959862034
                r(sum) =  930779.6986083984
                r(min) =  1610.5
                r(max) =  9303.900390625
                 r(p1) =  1723
                 r(p5) =  1878
                r(p10) =  2058.10009765625
                r(p25) =  2600.050048828125
                r(p50) =  4142.199951171875
                r(p75) =  6312.849853515625
                r(p90) =  7621.89990234375
                r(p95) =  8442.900390625
                r(p99) =  9229.400390625
```

```
. gen d=0
. replace d=1 if spr>=0.8*r(p50) | spr<(2/3)*r(mean)
(195 real changes made)

. tab d

          d |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |          9        4.41        4.41
          1 |        195       95.59      100.00
------------+-----------------------------------
      Total |        204      100.00

. drop d
```

d) Generiranje odlozenih in vodecih spremenljivk

```
. sort kvartal

. gen spr_lag1=l.spr
(1 missing value generated)

. gen spr_lag4=l4.spr
(4 missing values generated)

. gen spr_lead2=f2.spr
(2 missing values generated)

. gen spr_diff1=d.spr
(1 missing value generated)

. list
```

	kvartal	spr	spr_lag1	spr_lag4	spr_lead2	spr_diff1
1.	1950q1	1610.5	.	.	1723	.
2.	1950q2	1658.8	1610.5	.	1753.9	48.30005
3.	1950q3	1723	1658.8	.	1773.5	64.19995
4.	1950q4	1753.9	1723	.	1803.7	30.90002
5.	1951q1	1773.5	1753.9	1610.5	1839.8	19.59998
6.	1951q2	1803.7	1773.5	1658.8	1843.3	30.19995
7.	1951q3	1839.8	1803.7	1723	1864.7	36.1001
8.	1951q4	1843.3	1839.8	1753.9	1866.2	3.5
9.	1952q1	1864.7	1843.3	1773.5	1878	21.3999
10.	1952q2	1866.2	1864.7	1803.7	1940.2	1.5
11.	1952q3	1878	1866.2	1839.8	1976	11.80005
12.	1952q4	1940.2	1878	1843.3	1992.2	62.19995
13.	1953q1	1976	1940.2	1864.7	1979.5	35.80005
14.	1953q2	1992.2	1976	1866.2	1947.8	16.19995
15.	1953q3	1979.5	1992.2	1878	1938.1	-12.69995
16.	1953q4	1947.8	1979.5	1940.2	1941	-31.69995
17.	1954q1	1938.1	1947.8	1976	1962	-9.700073
18.	1954q2	1941	1938.1	1992.2	2000.9	2.900024
19.	1954q3	1962	1941	1979.5	2058.1	21
20.	1954q4	2000.9	1962	1947.8	2091	38.90002
21.	1955q1	2058.1	2000.9	1938.1	2118.9	57.20007
22.	1955q2	2091	2058.1	1941	2130.1	32.8999

...

```
183. |  1995q3   7561.4     7503.3     7370.2     7676.4     58.1001 |
184. |  1995q4   7621.9     7561.4     7461.1     7802.9        60.5 |
185. |  1996q1   7676.4     7621.9     7488.7     7841.9        54.5 |
     |-----------------------------------------------------------------|
186. |  1996q2   7802.9     7676.4     7503.3     7931.3       126.5 |
187. |  1996q3   7841.9     7802.9     7561.4     8016.4          39 |
188. |  1996q4   7931.3     7841.9     7621.9     8131.9     89.3999 |
189. |  1997q1   8016.4     7931.3     7676.4     8216.6     85.1001 |
190. |  1997q2   8131.9     8016.4     7802.9     8272.9       115.5 |
     |-----------------------------------------------------------------|
191. |  1997q3   8216.6     8131.9     7841.9     8396.3    84.69971 |
192. |  1997q4   8272.9     8216.6     7931.3     8442.9    56.30078 |
193. |  1998q1   8396.3     8272.9     8016.4     8528.5    123.3994 |
194. |  1998q2   8442.9     8396.3     8131.9     8667.9    46.60059 |
195. |  1998q3   8528.5     8442.9     8216.6     8733.5    85.59961 |
     |-----------------------------------------------------------------|
196. |  1998q4   8667.9     8528.5     8272.9     8771.2    139.4004 |
197. |  1999q1   8733.5     8667.9     8396.3     8871.5    65.59961 |
198. |  1999q2   8771.2     8733.5     8442.9     9049.9     37.7002 |
199. |  1999q3   8871.5     8771.2     8528.5     9102.5    100.2998 |
200. |  1999q4   9049.9     8871.5     8667.9     9229.4    178.4004 |
     |-----------------------------------------------------------------|
201. |  2000q1   9102.5     9049.9     8733.5     9260.1    52.59961 |
202. |  2000q2   9229.4     9102.5     8771.2     9303.9    126.9004 |
203. |  2000q3   9260.1     9229.4     8871.5          .    30.69922 |
204. |  2000q4   9303.9     9260.1     9049.9          .    43.80078 |
     +-----------------------------------------------------------------+
```

```
. clear all
```

■