

# Ekonometrija 1

## **Prve vaje:**

### *Uvod v programski paket Stata.*

Na prvih vajah 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 nepravilnih 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

- Proučite podatke s pomočjo različnih ukazov za pregled podatkov. Kako bi najlažje ročno uredili podatke za konkretne spremenljivke in konkretna opazovanja v vaši bazi?
- Proučite podatke še grafično s pomočjo različnih diagramov. Uporabite razsevni diagram, linijski diagram in histogram.
- Na osnovi obstoječih spremenljivk iz podatkovne baze z različnimi transformacijami generirajte nekaj novih spremenljivk. Uporabite množenje, absolutne vrednosti, logaritmiranje, antilogitmiranje ter standardiziranje.
- Prikličite iz Statinega spomina rezultate izvedbe enostavnejšega ukaza `summarize` ter kompleksnejšega ukaza `regress`. Kako bi jih shranili za kasnejšo uporabo?
- 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:     4
size:    128
```

variable name	storage type	display format	value label	variable label
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	-		
+-----						-----	-----	
						8		

(5 unique values)

**. sum**

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	
			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					Total
	-4	-1	0	2	3	
-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
```

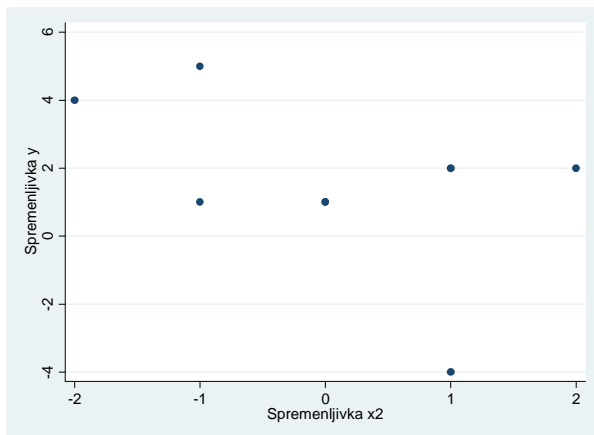
```

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

```

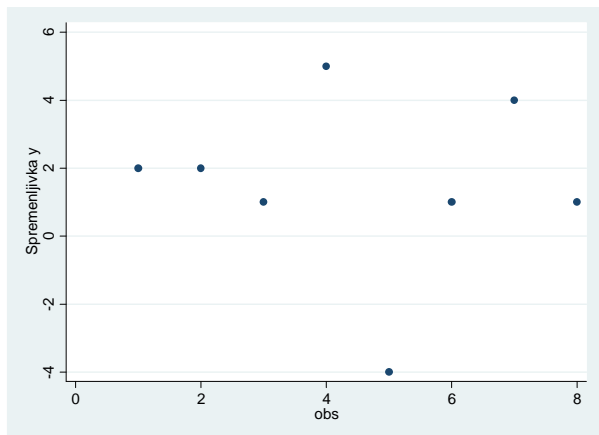
b) Diagrami v Stati

```
. twoway scatter y x2
```

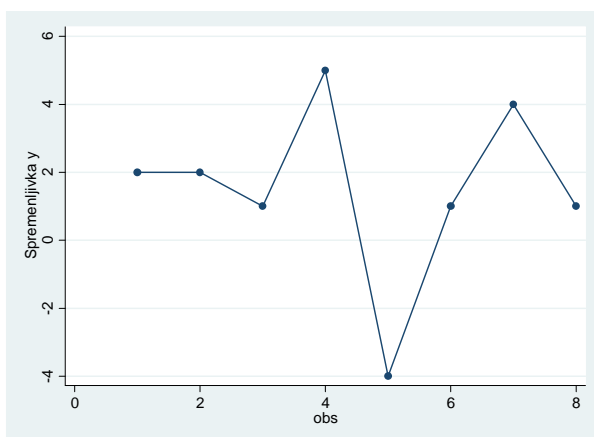


```
. gen obs=_n
```

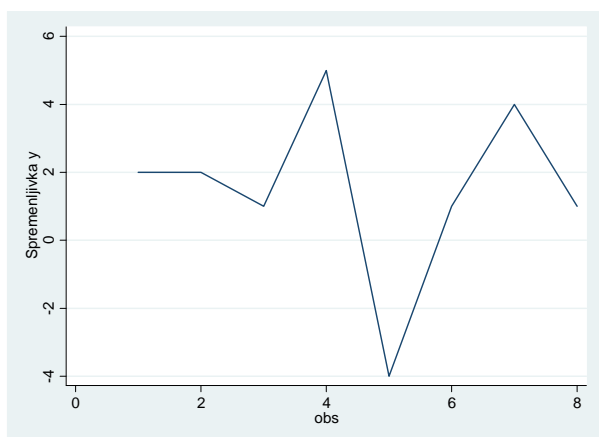
```
. twoway scatter y obs
```



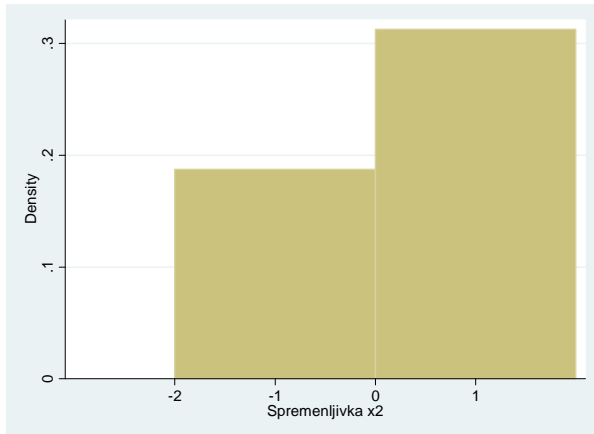
```
. twoway connected y obs
```



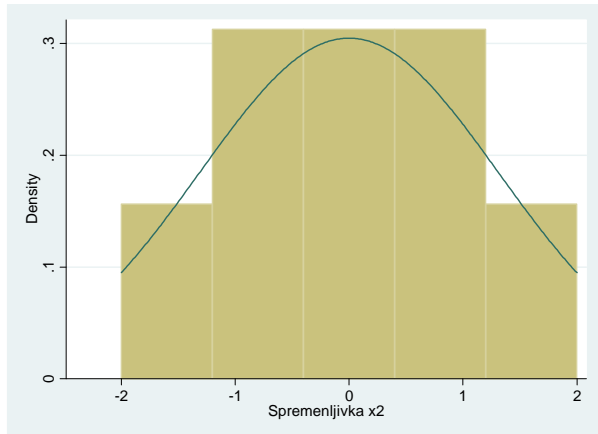
```
. twoway line y obs
```



```
. 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
Model	32.25	2	16.125	F( 2, 5) =	4.54
Residual	17.75	5	3.55	Prob > F =	0.0751
Total	50	7	7.14285714	R-squared =	0.6450
				Adj R-squared =	0.5030
				Root MSE =	1.8841

	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 0.2178	1.0000	
x3	-0.6364 0.0898	0.0000 1.0000	1.0000

```
. clear all
```



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

- Odprite v programskem paketu Stata podatkovno datoteko. Proučite podatke s pomočjo različnih ukazov za pregled podatkov.
- Kreirajte spremenljivko, ki bo opredeljevala časovno dimenzijo, če je časovna serija opredeljena na letni, kvartalni oziroma mesečni ravni. Za podatke iz naše podatkovne baze je pravilna četrletna dimenzija, ki jo zato obdržimo.
- Sortirajte podatkovno bazo po časovni dimenziji ter zamenjajte vrstni red spremenljivk. Generirajte trend, nepravne spremenljivke za četrletja in ciklično komponento.
- Dodajte opazovanje, ki naj ima vrednost spremenljivke enako, kot v zadnjem obstoječem opazovanju. Nato iz podatkovne baze izbrišite zadnjih pet opazovanj.
- 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.
- 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: 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)							



```
. sum, 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  
75%              6312.85         Largest  
90%              7621.9          9102.5         Mean                4562.646  
95%              8442.9          9229.4         Std. Dev.           2113.962  
99%              9229.4          9303.9         Variance            4468837  
                                         Skewness             .4680201  
                                         Kurtosis              2.133958
```

```
. 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) *Opredefinitev casovne dimenzije*

```
. gen leto=1950+_n-1  
. format leto %ty  
. label variable leto "Leto opazovanja"  
. tsset leto  
    time variable: leto, 1950 to 2153  
    delta: 1 year  
  
. gen kvartal=tq(1950q1)+_n-1  
. format kvartal %tq  
. label variable kvartal "Cetrletje opazovanja"  
. tsset kvartal  
    time variable: kvartal, 1950q1 to 2000q4  
    delta: 1 quarter  
  
. gen mesec=tm(1950m1)+_n-1  
. format mesec %tm  
. label variable mesec "Mesec opazovanja"  
. tsset mesec  
    time variable: mesec, 1950m1 to 1966m12  
    delta: 1 month  
  
. keep spr kvartal  
  
. tsset kvartal  
    time variable: kvartal, 1950q1 to 2000q4  
    delta: 1 quarter
```

c) *Generiranje periodicnih komponent*

```
. sort kvartal  
. 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
```

```
. keep kvartal spr
```

d) Ravnanje z opazovanji

```
. set obs `=_N+1'
```

```
obs was 204, now 205
```

```
. replace kvartal=tq(2001q1) if _n==205
```

```
(1 real change made)
```

```
. replace spr=spr[204] if _n==205
```

```
(1 real change made)
```

```
. list in 201/205
```

```
+-----+
| kvartal      spr |
+-----+
201. | 2000q1    9102.5 |
202. | 2000q2    9229.4 |
203. | 2000q3    9260.1 |
204. | 2000q4    9303.9 |
205. | 2001q1    9303.9 |
+-----+
```

```
. drop in 201/205
```

```
(5 observations deleted)
```

```
. sum
```

Variable	Obs	Mean	Std. Dev.	Min	Max
kvartal	200	59.5	57.87918	-40	159
spr	200	4469.419	2028.071	1610.5	9049.9

e) Generiranje neprave spremenljivke

```
. sum spr, detail
```

```
Casovna spremenljivka (v enotah mere)
-----
```

Percentiles	Smallest		
1%	1690.9	1610.5	
5%	1872.1	1658.8	
10%	2029.5	1723	Obs 200
25%	2584.15	1753.9	Sum of Wgt. 200

```

50%      4124.85          Largest      Mean      4469.419
75%      6102.65          8733.5      Std. Dev. 2028.071
90%      7474.9          8771.2      Variance  4113071
95%      8174.25          8871.5      Skewness  .4342106
99%      8821.35          9049.9      Kurtosis  2.07264

```

**. return list**

scalars:

```

      r(N) = 200
      r(sum_w) = 200
      r(mean) = 4469.418991088867
      r(Var) = 4113071.201324418
      r(sd) = 2028.070807768905
      r(skewness) = .4342106439937721
      r(kurtosis) = 2.072640270733662
      r(sum) = 893883.7982177734
      r(min) = 1610.5
      r(max) = 9049.900390625
      r(p1) = 1690.900024414063
      r(p5) = 1872.099975585938
      r(p10) = 2029.500061035156
      r(p25) = 2584.150024414063
      r(p50) = 4124.849853515625
      r(p75) = 6102.650146484375
      r(p90) = 7474.900146484375
      r(p95) = 8174.249755859375
      r(p99) = 8821.35009765625

```

```

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

```

**. tab d**

d	Freq.	Percent	Cum.
0	9	4.50	4.50
1	191	95.50	100.00
Total	200	100.00	

*f) Generiranje odlozenih in vodecih spremenljivk*

**. sort kvartal**

```

. gen spr_lag1alt=1.spr
(1 missing value generated)

```

```

. gen spr_lag4alt=14.spr
(4 missing values generated)

```

```

. gen spr_lead2alt=f2.spr
(2 missing values generated)

```

```

. gen spr_diff1alt=d.spr
(1 missing value generated)

```

**. drop d**

```
. 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 |
+-----+
...
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      .    100.2998 |
200. | 1999q4  9049.9      8871.5    8667.9      .    178.4004 |
+-----+
```

```
. clear all
```

