

```

* my commands Intro to Econometrics - Summer 2013.

** Ex. Session 1

display "Hello World"

display 2*3+1

display 5+5-3

display 5>0

display 5<0

* creating variable
set obs 500
gen var1 = uniform()

gen var2 = 2*var1 + 3

gen var3 = invnorm(uniform())

* describing data
hist var1
kdensity var1
tway scatter var1

* how to delete variables
drop var2

* how to change variable name
rename var1 prva

* helping tool
** help (...)
help gen

*Basic summaries
sum prva var3 var2
sum prva var3 var2, detail

* Creating dummy variables
gen d_var4=0
replace d_var4=1 if var3>0

tab d_var4

*** Ex. Session 2
* How to import data from Excel
** note: command "clear" deletes all the data

**importing data from Excel
insheet using "C:\Users\Admin\Desktop\Ekonometrija\data.csv"

**locating where we work
cd"C:\Users\Admin\Desktop\Ekonometrija\Dragana"

```

```

** Label variables
label var totwage "Wage per week"
label var educat "Education level"
label var exper "Experience"
label var marst "Marital Status"

** Simple Statistics
sum totwage educat exper marst sex

sum totwage educat exper marst sex, detail

** Inspecting variables
tab marst

tab marst sex

tab totwage

*****
** Ex. Session 3
*****

** setting memory in STATA
set mem 500m

** locating the directory where we work
cd"C:\Users\Admin\Desktop\Ekonometrija\Dragana"

** importing data file
insheet using "C:\Users\Admin\Desktop\Ekonometrija\data.csv"

*I) Dropping outliers/extremes
** We always want to exclude outliers to have robust results.

** Inspecting wage
sum totwage, detail

drop if totwage>10000

*** we have to exclude those who do not work.
drop if totwage==0

drop if totwage>6000

*** Inspecting education
sum educat, detail

*** Inspecting experience
sum exper, detail

drop if exper>68

* II. Wage equation 1.

reg totwage educat exper marst sex

```

```

* III. Problems with Wage equation 1.

*** 1. negative beta zero
*** 2. marital status is categoric variable.
*** 3. gender is also a categoric variable.
*** 4. concave relationship between wage and experience.
*** 5. exponential function of wage. (wage se povecava nelinearno)

* file C:\Users\Admin\Desktop\Ekonometrija\Dragana\ExSession1.dta saved

*****
** Ex. Session 4
*****
** setting memory in STATA
set mem 500m

** locating the directory where we work
cd"C:\Users\Admin\Desktop\Ekonometrija\Dragana"

** importing data file
use "C:\Users\Admin\Desktop\Ekonometrija\Dragana\ExSession.dta", clear

* III. Addressing problems with Wage equation 1.

**1. change depenedent variable into log
gen logtotwage = log(totwage)
sum logtotwage, detail

**2. concave relationship between experience and wage
**** we need to creat experience square - ref. concave marginal returns.
gen exper2 = exper^2
sum exper2, detail

**3. categoric variables
** we have to choose one group relative to other - ref. wage and marriage.
gen married = 0
replace married = 1 if marst==1
tab married

** gender dummy
rename sex male
tab male

***
reg logtotwage educat exper exper2 married male

*****
*** Ex. Session 5
*****

**I. Introducing interaction terms - ref. x1 * x2

**** for example, we assume that rural or urban area will have different
levels of education.

tab locality

```

```

** 0 is rural, 1 is urban
rename locality urban

**** first we need to generate interaction term.
gen urban_educat = urban*educat

reg logtotwage educat exper exper2 married male urban urban_educat

pwcrr logtotwage educat exper exper2 married male urban urban_educat, sig

*****
*** Ex. Session 6
*****

* note: working memory can only be set before uploading data.
clear
set mem 500m

* working directory
cd"C:\Users\Admin\Desktop\Ekonometrija\Dragana"

* upload .dta
use "C:\Users\Admin\Desktop\Ekonometrija\Dragana\ExSession.dta", clear

* I. Wage equation 2.

reg logtotwage educat exper exper2 married male urban urban_educat

* note: we calculated that 35 years is maximum experience return on wage.
* note: interaction term between urban and education gives higher return on
** education in rural areas.

* II. Manipulation with residuals. (u)

reg logtotwage educat exper exper2 married male urban urban_educat, vce(robust)

reg logtotwage educat exper exper2 married male urban urban_educat,
cluster(state)

**notes: 1. India has 32 clusters (states).
***** 2. Std. Err. of coefficients increased, because we have variations
within the state.
***** 3. Urban_educat is no longer significant. Assumption: because states
are either
***** very urban or very rural.

*****
*** Ex. Session 7
*****

* setting working memory
set mem 500m

* working directory
cd"C:\Users\Admin\Desktop\Ekonometrija\Dragana"

```

```

* upload .dta
use "C:\Users\Admin\Desktop\Ekonometrija\Dragana\ExSession.dta", clear

*** Main wage model
reg logtotwage educat exper exper2 married male urban urban_edu, vce(robust)

* I. comparing results for different groups: gender

sort male
by male: reg logtotwage educat exper exper2 married urban urban_edu,
vce(robust)

** note: we get estimation of the model by gender.

* II. output tables in STATA

quietly reg logtotwage educat exper exper2 married urban urban_edu if male==1
estimates store men

quietly reg logtotwage educat exper exper2 married urban urban_edu if male==0
estimates store women

estimates table men women

**note: STATA gives us two comperable columns of men and women coefficients.
est tab men women, b(%5.3f) se(%5.3f)

* III. Predicted values

reg logtotwage educat exper exper2 married urban urban_edu male, vce(robust)
predict yhat
predict resid, residuals

sum yhat, detail

sum resid, detail

**

*****
SUBOTA 6.07.
*****

** Setting memory in STATA UVEK PRE UPLOADA PODATAKA!!!!
set mem 500m

** Locating the directory where we work

cd"C:\Users\statistika\Desktop\Ekonometrija\Viktorija"

***Breusch-Pagan Test for Homoscedasticity

```

```

*** ocenimo model (bez vcr robust)

reg logtotwage educat exper exper2 married male urban urban_edu

*** predvidimo rezidualne

predict resid, residuals

*** rezidualne na kvadrat regresiramo na nezavisne varijable

gen resid2 = resid*resid
reg resid2 educat exper exper2 married male urban urban_edu

*** note R2 is very small 0.01 so probably solution with better
specification.
*** search of comands in STATA we do in help

estat hettest educat exper exper2

estat hettest

estat hettest, rhs

**** LPM, LOGIT, PROBIT

*** for example we want to test which factors affect selfemployment

tab selfemp

*** we need to generate dummy variable selfe for selfemployment

gen selfe = 0
replace selfe = 1 if selfemp == 1

*** determining variables influencing selfemployment: married, age, age2,
male, educat, urban

gen age2 = age*age

*** linear probability model LPM

reg selfe married age age2 male educat urban

estimates store LPM

*** PROBIT

probit selfe married age age2 male educat urban
estimates store PROBIT

mfx

**** LOGIT

logit selfe married age age2 male educat urban
estimates store LOGIT

```

mfx

**** very important the estimation results of probit and logit are NOT comparable with the marginal effects we get from LPM

```
est tab LPM PROBIT LOGIT, b(%5.3f) se(%5.3f)
```

```
*****  
search for panel data and shaping for stata  
*****
```

** Formatting data from wide to long

* a. download data from world bank

* b. in excel make sure all numeric cells are in numbers

* c. each variable (gdpgr) is on one sheet

* d. we add x before year

* e. copy paste from excel to data editor in stata

* f. note: country, variable are string (red); x1960...x2012 are numeric (black). empty cells in stata have dots.

* g. we need to generate id for every country. This ID will be the same for all years for one country.

```
gen id = _n
```

```
order id
```

* h. reshape wide to long

```
reshape long x, i(id) j(year)
```

* i. we need to drop "variable" and to rename x to name of the variable

```
drop variable
```

```
rename x gdpgr
```

```
*****
```

```
petak 12.07.
```

```
*****
```

*** unesemo novu promenljivu u excel life_exp i kopiramo u data editor (nowi cisti)

```
gen id = _n
```

```
order id
```

* h. reshape wide to long

```
reshape long x, i(id) j(year)
```

* i. we need to drop "variable" and to rename x to name of the variable

```
drop variable
```

```
rename x life_exp
```

*** unesemo novu promenljivu u excel rd i kopiramo u data editor (nowi cisti)

```

gen id = _n
order id

* h. reshape wide to long

reshape long x, i(id) j(year)

* i. we need to drop "variable" and to rename x to name of the variable

drop variable

rename x rd

*** unesemo novu promenljivu u excel nat_res i kopiramo u data editor (nowi
cisti)

gen id = _n
order id

* h. reshape wide to long

reshape long x, i(id) j(year)

* i. we need to drop "variable" and to rename x to name of the variable

drop variable

rename x nat_res

*** unesemo novu promenljivu u excel unemp i kopiramo u data editor (nowi
cisti)

gen id = _n
order id

* h. reshape wide to long

reshape long x, i(id) j(year)

* i. we need to drop "variable" and to rename x to name of the variable

drop variable

rename x unemp

*** spajanje svih varijabli iz 5 razlicitih filova u jedan
*** moramo biti u jednom filu, tako da cemo otvoriti gdpgr
*** pre spajanja podataka uvek izdamo komandu za sortiranje zemalja i godina,
u slucaju da su podaci haoticni i sacuvamo

sort country year

*** otvorimo novi prozor desnim klikom dole na ikonicu state u osnovnom baru
*** u novom prozoru otvorimo naredni file life_exp

```

```

*** sortiramo i ovu bazu podataka i sacuvamo u command window nove
baze

sort country year

*** the main data is gdpgr we merge all others to this
merge using life_exp, keep(life_exp)

cd"C:\Users\statistika\Desktop\Ekonometrija\Viktorija"

*** proveravamo kako nam je stata povezala podatke!

tab _merge

tab country if _merge ==1

tab country year if _merge ==1

tab year if _merge ==1

*** od 1960 do 1984 uopste nema podataka za life expectancy

*** drop all observation that are not matched

drop if _merge ==1
*** pre novog spajanja moramo da izbacimo _merge
drop _merge
*** dodajemo nat_res: otvorimo novi prozor state i u njemu ovaj file

merge using nat_res, keep(nat_res)

tab _merge

drop _merge

merge using rd, keep (rd)

tab _merge

drop if _merge ==1

drop _merge

merge using unemp, keep(unemp)

tab _merge

drop if _merge ==2

drop _merge

*** fixed effects
set mem 500m

iis id

```

```
xtreg gdpgr life_exp nat_res rd unemp, fe
*** randomeffects
iis id
xtreg gdpgr life_exp nat_res rd unemp, re
```