Introduction to Stata: Part II

Working with Stata

Looking at the data

- Let us
  . *sysuse auto*
  to begin our session with the data.

- What kind of data is there? Type
  . *describe price*
  to get an overall picture.

- to get a listing of existing variables and observations
  . *list price*

Looking at the data

- Summary statistics
  . *summarize price*
  . *summarize price, d*

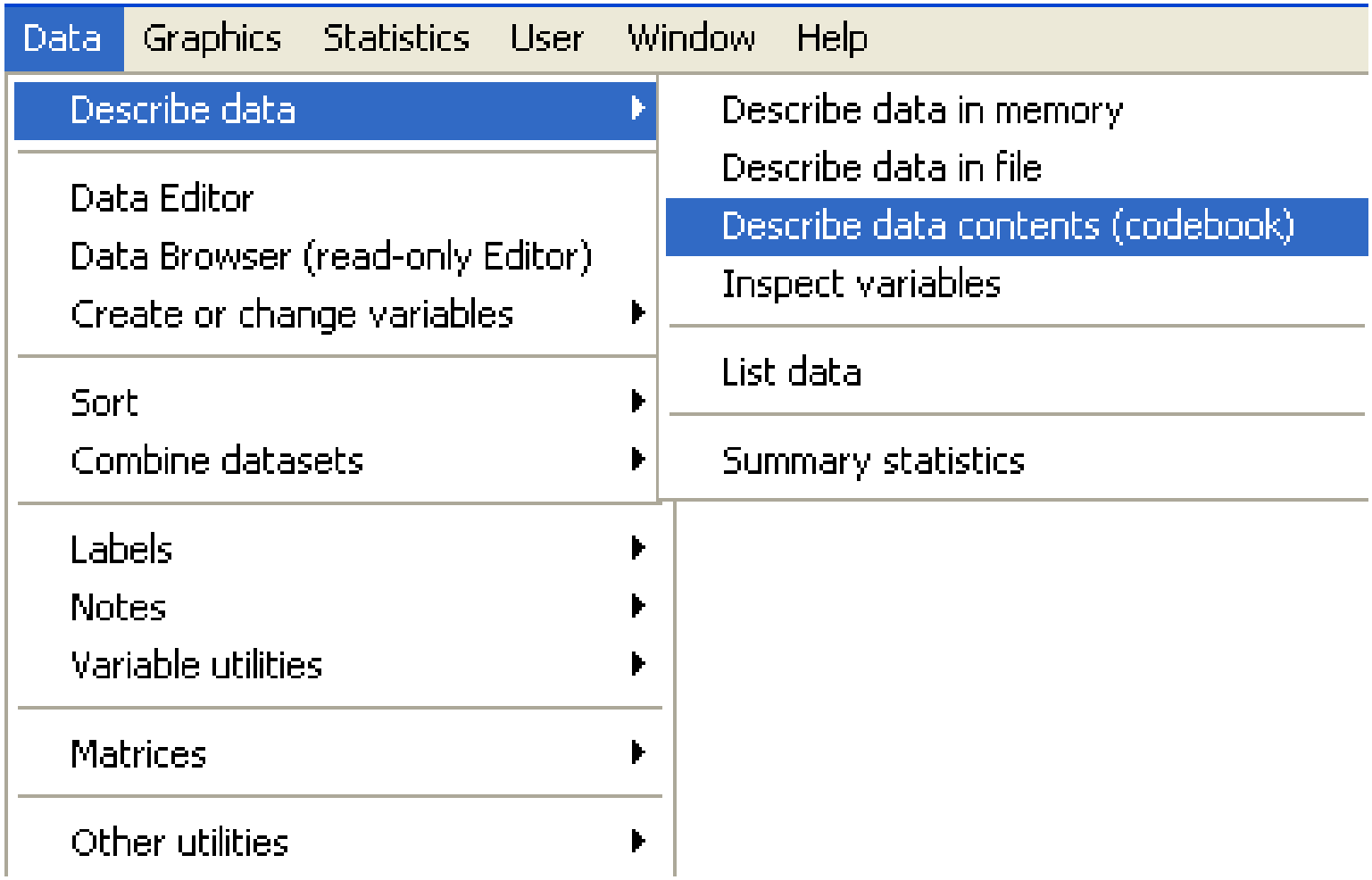
- Two-way tables

  . use this procedure when you want to look at 2-way frequencies of your categorical data

  . *tab foreign rep78*

- We can get the correlation:
  . *correlate* price mpg

Options within the *Data* menu

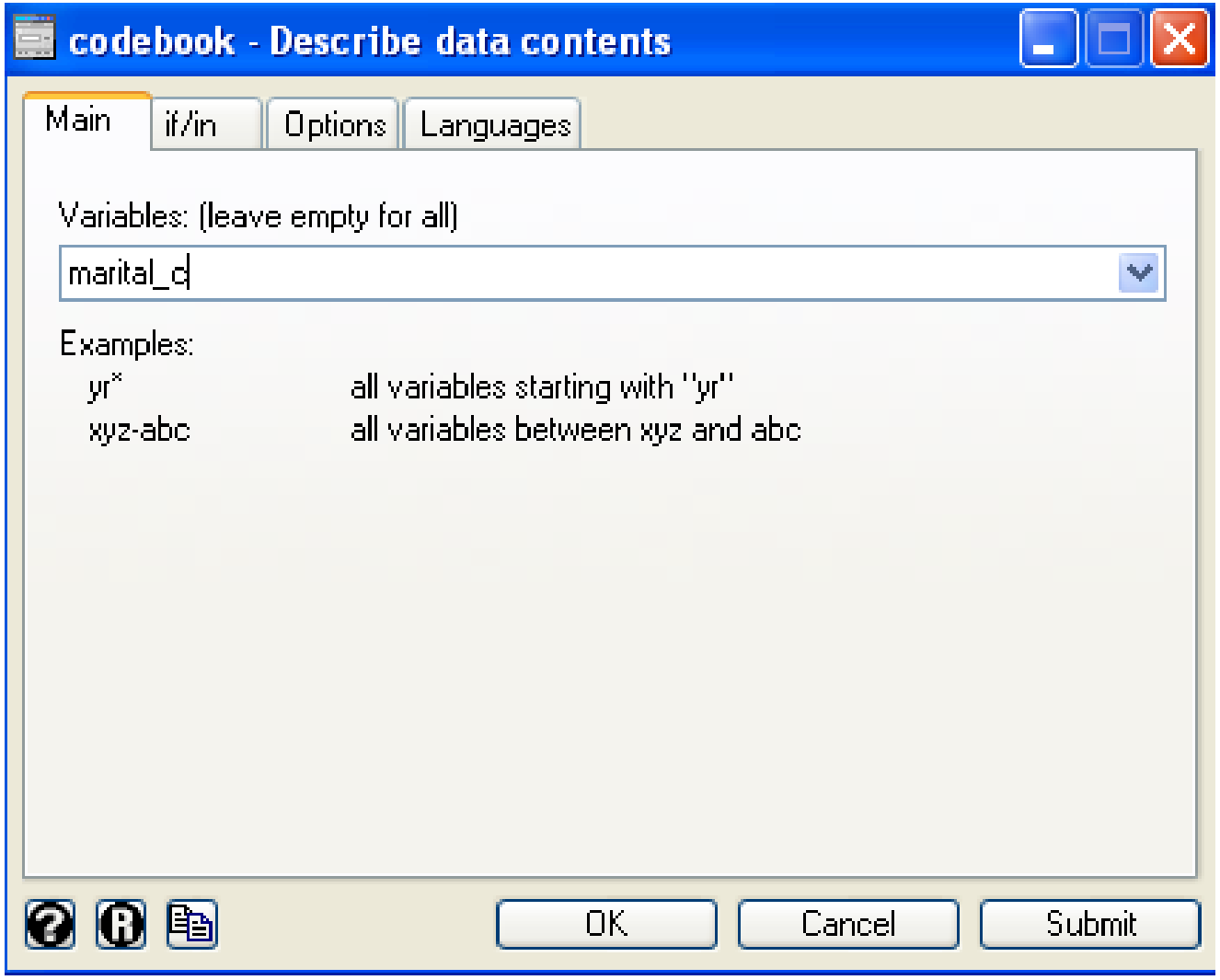


Results from codebook and summarize

```
. codebook marital_c

-------------------------------------------------------------------------------
marital_c
-------------------------------------------------------------------------------

                  type:  string (str13), but longest is str12

        unique values:  5                        missing "":  0/321

           tabulation:  Freq.  Value
                          15  "Divorced/Sep"
                         236  "Married Mono"
                          10  "Married Poly"
                          17  "Single"
                          43  "Widowed"
```

```
. summarize age

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
         age |        321    48.44548    17.20786        19         99
```

An example of a dialogue box



Dialogue results from using the menu sequence

Data,

⇒ **Describe data,**

⇒ **Describe data contents** (codebook)

Looking at the data

- Graph

  . *histogram* price

- sysuse uslifeexp2
  - scatter le year
  - scatter le year, connect(l)
  - scatter le year, connect(l) msymbol(i)

Creating a new variable

- Let us

  . *sysuse auto*

- A natural step further in your data management is to create some new variables

  . *gen gpm = 100/mpg*

  . *gen lgpm = ln(mpg)*

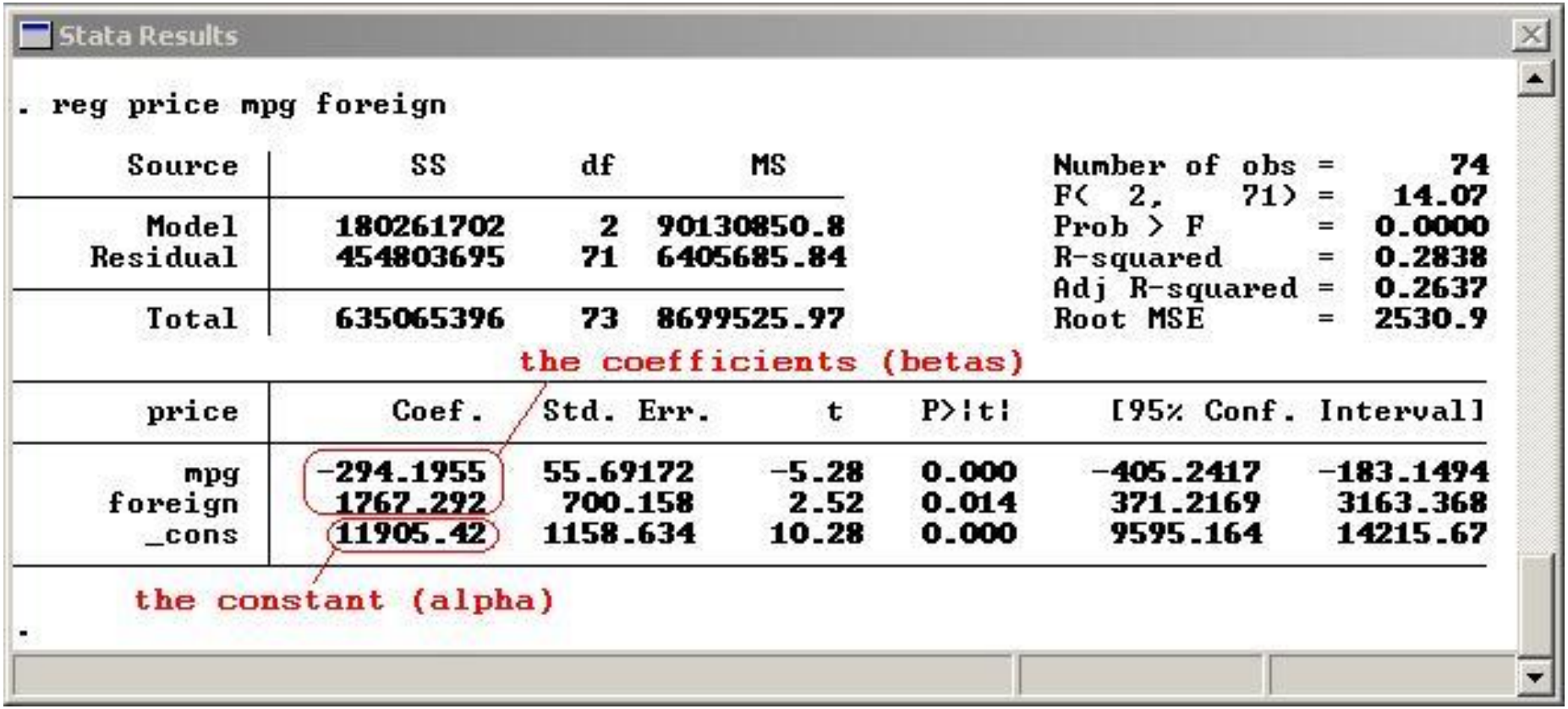
Important issues

- Formatting output
  - Highlight results
  - Edit→Copy table
- Missing data
- Rename variables

Estimation

- OLS model: Interpreting Regression Output
  - reg price mpg foreign
  - Coefficients
  - Standard Error of the coefficients
  - CI
  - R^2

Estimation


Stata Results
. reg price mpg foreign
Source | SS df MS
Model | 180261702 2 90130850.8
Residual | 454803695 71 6405685.84
Total | 635065396 73 8699525.97
Number of obs = 74
F(2, 71) = 14.07
Prob > F = 0.0000
R-squared = 0.2838
Adj R-squared = 0.2637
Root MSE = 2530.9
the coefficients (betas)
price | Coef. Std. Err. t P>|t| [95% Conf. Interval]
mpg -294.1955 55.69172 -5.28 0.000 -405.2417 -183.1494
foreign 1767.292 700.158 2.52 0.014 371.2169 3163.368
_cons 11905.42 1158.634 10.28 0.000 9595.164 14215.67
the constant (alpha)
.

Predictions

- calculate the fitted values and residuals
  - fitted values: predict pprice
  - residuals: predict resid, residuals

Plotting the Data and a Linear Fit

- We can inspect the quality of the fit
  - graph twoway (scatter price mpg) (lfit price mpg)

Estimation

- Hypotheses testing
  - t-test: in the table
  - F-test: in the table
  - We could also run F-test:
    - reg rep78 price weight
    - test price weight

Testing for Heteroskedasticity

- White test :
  - reg price mpg foreign
  - estat imtest, white
- Breusch-Pagan test :
  - reg price mpg foreign
  - estat hettest, normal

Robust standard error

- We add the option *robust* at the end of the regression command:
  - reg price mpg foreign, robust

WLS estimator

- reg price mpg foreign
- reg price mpg foreign, robust
- vwls price mpg foreign

probit and logit models

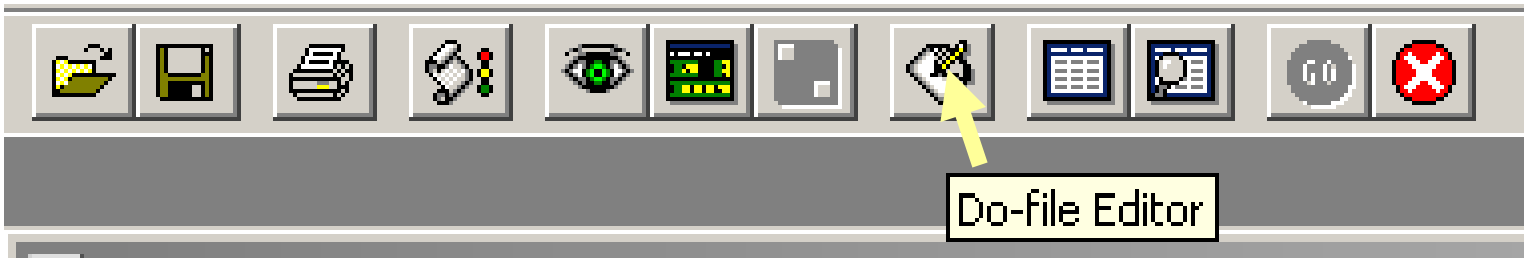
- probit foreign weight mpg
- logit foreign weight mpg

Panel data

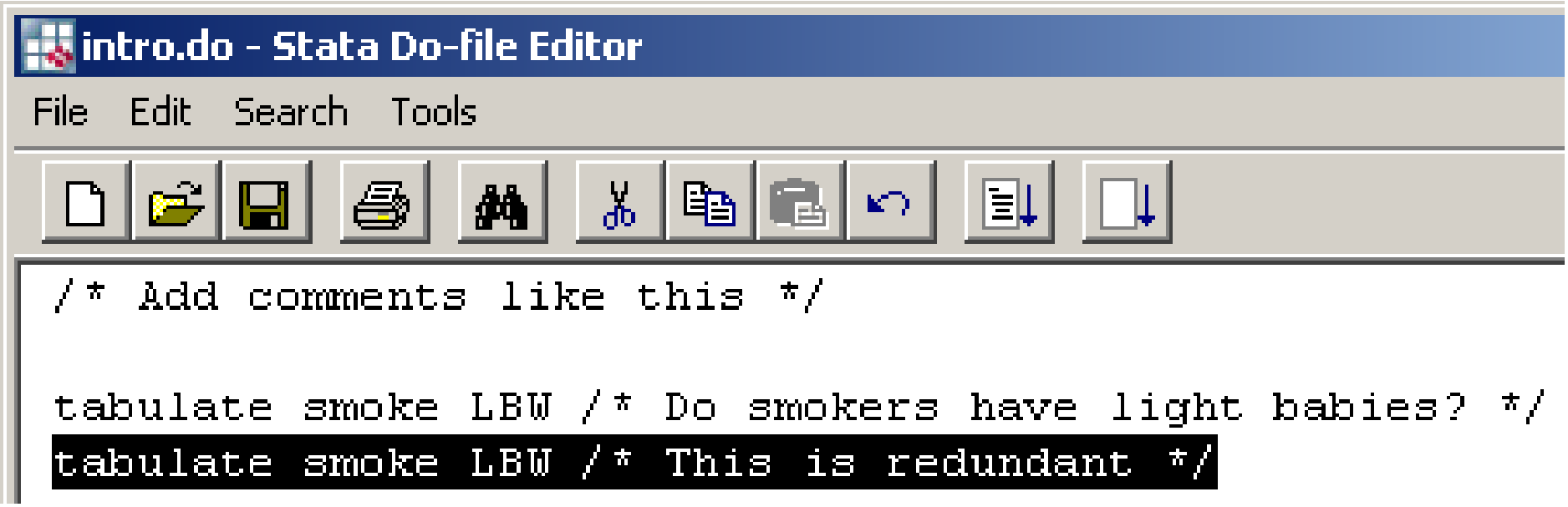
- We use nlswork.dta
- tsset
- Fixed effect:
  - xtreg hours birth_yr age race, fe
- Random effect:
  - xtreg hours birth_yr age race, re

Resuming Stata

- Launch Stata:
- Old commands are in the do-file. To see it,



  - In do-file editor
    - File→Open
    - Find X:\intro.do



Running a Do-file

- To run your old commands,
  - File→Do…
  - Find “X:\intro.do”

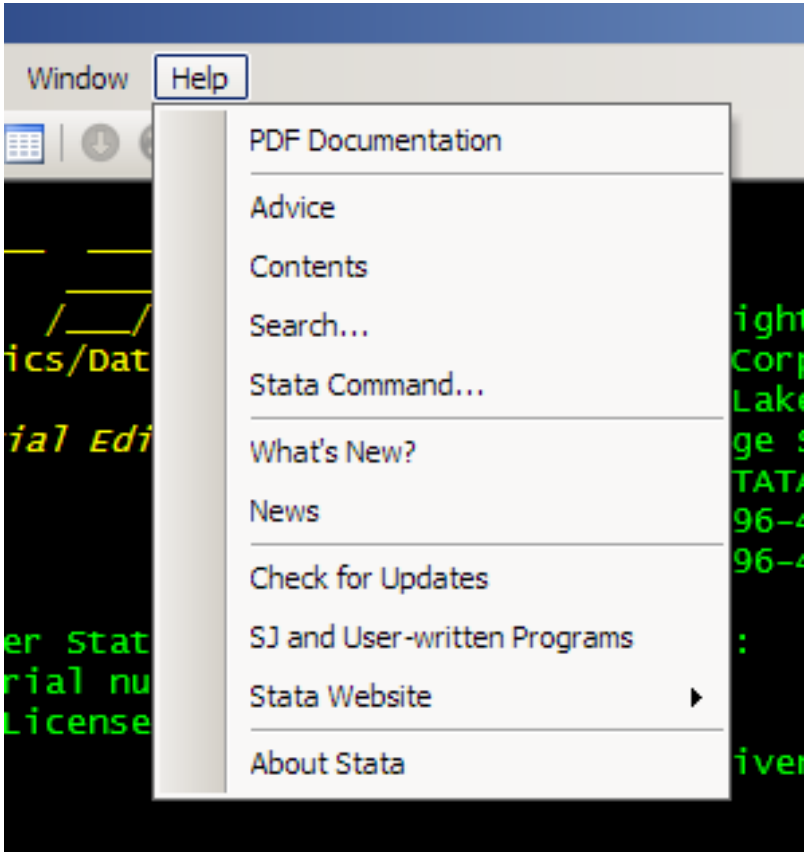
Labeling values

- To add labels to dummy values
  - **Data→Labels**
  - Define label name
  - Add values

- We can label the data direclty from the editor

To learn more

- Help menu (or help command)
  - e.g., in Command window type “help cmdlog”



Exercise

- Edit the do-file until it runs all the way through
- This will also familiarize you with the typed versions of some commands.
- http://www.reading.ac.uk/ssc/n/UBOS_DVD/Module_4/Module_4.htm