

Examples of data use: Stata platform¹

To obtain the results of calculations more rapidly, you should limit the number of variables included in the datasets used. Please remember that permanent datasets cannot be saved.

The list headings contain an example of authentication data needed to submit the program to the BIRD system. **Obviously, users will have to substitute the content of the first three fields, which begin with the character “*”, with their own personal authentication data.**

The statement ***package =** specifies the program language, Stata in this case.

The Stata commands are written in lower-case: please note that this language is case-sensitive.

All the examples assume that the character marking the end of commands is the semi-colon (“;”).

1. Examples using data from the Survey of industrial and service firms

In each of these examples a file is opened in Stata format, containing the survey data. This file is identified with the conventional symbolic name: **\$db_001**. You are shown how to restrict the analysis to a single sector (for example, the industrial sector, `survey==1`) or year (for example, 2005, `annoril==2005`). The first five examples show calculations exclusively for industrial firms in the year 2005.

Example n. 1

- We estimate a logit model in which the dichotomous dependent variable is membership of a group of firms. The explicative variables are the average number of workers (`v24`) and the variables relative to the geographical area of the headquarters and the sector of economic activity. These last two variables are devised so that they are suitable for treatment as dummies.

```
*user = user
*password = password
*project = project
*package = stata
#delimit;
use annoril indagine peso areag4 settor11 v521 v24
    if annoril==2005 & indagine == 1
using $db_001;
/* creation of the geographical area and sector of economic activity dummies */
tab areag4,gen(areag4d);
tab settor11,gen(settor11d);
/* this creates 4 geographical area dummies and 7 sector dummies */
/* estimation of the logit, in which one dummy is omitted for both area and
sector, which acts as a reference for the others */
```

¹ Stata is a registered trademark of StataCorp LP, 4905 Lakeway Drive, College Station, TX 77845 USA. The examples were prepared by Leandro D'Aurizio.

```
logit v521 v24 areag4d1-areag4d3 settor11d1-settor11d6 [pweight=peso];
```

Example n. 2

- The aim is to calculate, for industrial firms only (indagine==1) the percentage change in the average number of workers and the number of firms belonging to a group as a proportion of the total and divided by geographical area. To obtain the weighted estimates correctly, you must perform the following steps (note that the creation of the variable var_occ serves merely to obtain estimates referred to a percentage change).

```
*user = user
*password = password
*project = project
*package = stata
#delimit;
use annoril indagine peso popstr strato areag4 settor11 v521 v15 v24
    if annoril == 2005 & indagine == 1
using $db_001;
svyset _n[pw=peso],strata(strato) fpc(popstr);
gen var_occ=(v24-v15)*100;
svy:ratio var_occ/v15;
svy:ratio var_occ/v15, over(areag4);
svy:proportion v521;
svy:proportion v521, over(areag4);
```

Example n. 3

- In a similar way to the previous example, we want to calculate the percentage change in investment at constant prices. This variable is first treated to limit the effect of anomalous values (outliers), using the Type II Winsorization procedure, which is used to calculate the estimates of investment published in the Supplements to the Statistical Bulletin dedicated to surveys.

```
*user = user
*password = password
*project = project
*package = stata
#delimit;
use annoril indagine peso strato popstr areag4 v200cos
    v202cos v810cos v811cos v24
    if annoril == 2005 & indagine == 1;
using $db_001;
/* creation of the variable total investment at constant prices for 2004 */
gen i0tot=v200cos+v810cos;
/* creation of the variable total investment at constant prices for 2005 */
gen i1tot=v202cos+v811cos;
/* Type II winsorization (on the basis of the 5th and 95th percentiles */
gen diffe=(i1tot-i0tot)/v24;
```



```

gen f=1/peso;
su diffe [w=peso],de;
sca pp5=r(p5);
sca pp95=r(p95);
gen diffe_p5=pp5;
gen diffe_p95=pp95;
replace diffe=f*diffe+(1-f)*diffe_p95 if diffe !=.
& diffe>diffe_p95;
replace diffe=diffe_p95 if diffe != . & diffe>diffe_p95 & f==1
& v24<5000;
replace diffe=f*diffe+(1-f)*diffe_p5 if diffe !=.
& diffe<diffe_p5;
replace diffe=diffe_p5 if diffe != . & diffe<diffe_p5 & f==1
& v24<5000;
/* creation of a new variable i1totw containing total 2005 investment, which
attenuates the effect of outliers */
gen i1totw=i0tot+diffe*v24;
svyset _n[pw=peso],strata(strato) fpc(popstr);
gen var_inv=(i1totw-i0tot)*100;
svy:ratio var_inv/i0tot;
svy:ratio var_inv/i0tot, over(areag4);

```

Example n. 4

- Let's assume we want to assess a linear model where the number of workers (variable v24) is the dependent, and the co-variants are turnover (variable v210) and the geographical area where the firm has its headquarters; this last variable is used as a dummy.

```

*user = user
*password = password
*project = project
*package = stata
#delimit;
use annoril indagine peso areag4 v210 v24
    if annoril ==2005 & indagine == 1
    using $db_001;
/* creation of the geographical area dummies */
tab areag4,gen(areag4d);
/* this creates 4 geographical area dummies */
/* estimation of the regression, in which one dummy is omitted for the area,
which acts as a reference for the others */
reg v24 v210 areag4d1 areag4d2 areag4d3 [pweight=peso];

```

Example n. 5

- The program below replicates the regression of the previous example, but restricts it to those firms having a number of workers within the 1st and 99th percentiles of the distribution.

```
*user = user
*password = password
*project = project
*package = stata
#delimit;
use annoril indagine peso areag4 v210 v24
    if annoril == 2005 & indagine == 1
        using $db_001;
/* creation of the geographical area dummies */
tab areag4,gen(areag4d);
/* this creates 4 geographical area dummies */
/* creation of the two variables pc1_v24 and pc99_v24
containing the first and 99th percentiles of the variable v24 */
egen pc1_v24=pctile(v24),p(1);
egen pc99_v24=pctile(v24),p(99);
/* estimation of the regression, in which one dummy is omitted for the area,
which acts as a reference for the others, and the units with v24 outside the
percentiles are omitted */
reg v24 v210 areag4d1 areag4d2 areag4d3 [pweight=peso]
if v24>=pc1_v24 & v24<=pc99_v24;
```

Example n. 6

- The program below presents an example of panel assessment with random effects on a group of firms that have always been present in the years considered by the model. The analysis is restricted exclusively to the industrial sector (survey==1) in the years 2001-2006. We use turnover as a dependent variable (v210), and the average number of workers (v24) and operating result as co-variants (v545). Before being used as a dummy the variable v545 is re-codified.

```
*user = user
*password = password
*project = project
*package = stata
#delimit;
use annoril indagine ident areag4 v545 v210 v24
    if annoril >=2001 & annoril <=2006 & indagine == 1
        using $db_001;
/* selection of the firms covered by the survey in the 6 years from 2001 to 2006
*/
generate one=1;
sort ident;
```

```

by ident: egen conta=sum(one);
keep if conta==6;
/* creation of the result for the year dummies */
tab v545,gen(v545d);
/* estimation of the regression model on the panel, in which one dummy is
omitted for the result for the year, which acts as a reference for the others */
iis ident;
tis annoril;
xtreg v210 v24 v545d1 v545d2 v545d3 v545d4,re;

```

2. Examples using data from the Business Outlook Survey of industrial and service firms

Example n. 7

- From the time series database a tabulation is created that, for all available years, shows the frequency distribution of variable stg3 (planned investment for the subsequent year) only for manufacturing firms with 50+ workforce (indag3==1).

```

*user = user
*password = password
*project = project
*package = stata
#delimit;
use annoril indag3 cc2 stg3 pesorisc using $db_sondstor;
gen one=1;
sort annoril;
by annoril: tabulate stg3 one [aweight=pesorisc] if cc2>=2
& indag3==1;

```

Example n. 8

- This program presents a merge between the time series database of the Business Outlook Survey of Industrial and Service Firms and the time series database of the Survey of Industrial and Service Firms, with the aim of comparing investment plans surveyed in the 2006 Outlook Survey to corresponding realizations (surveyed as continuous variables and then discretized) in the 2007 Annual Survey. Only firms having participated in both surveys are considered; those that answered “9” (“don’t know, no answer”) when asked for plans are excluded.

```

*user = user
*password = password
*project = project
*package = stata
#delimit;
/* load data from Business Outlook survey 2006 */
use annoril ident stg3 using $db_sondstor if annoril==2006;
sort ident;
save sond2006,replace;

```

```
clear;
/* load data from Annual Survey 2007 */
use annoril ident v200 v810 v202 v811 peso using $db_001
if annoril==2007;
/* compute changes in investment */
generate i0tot=v200+v810;
generate i1tot=v202+v811;
/* substitute zeroes with small positive values
to obtain a valid rate of change even if the
value of i0tot or both terms are zero */
replace i0tot=0.1 if i0tot==0;
replace i1tot=0.1 if i1tot==0;
generate varinv=(i1tot/i0tot-1)*100;
/* discretize continuous variable varinv */
generate varinvd=1 if varinv<-10;
replace varinvd=2 if (varinv>=-10 & varinv<-3);
replace varinvd=3 if (varinv>=-3 & varinv<=3);
replace varinvd=4 if (varinv>3 & varinv<=10);
replace varinvd=5 if varinv>10;
sort ident;
merge ident using sond2006;
/* keep only firms appearing in both datasets */
keep if _merge==3;
tabulate stg3 varinvd [aweight=peso] if stg3 !=9,cell;
```

