

## Esempi di utilizzo dei dati: piattaforma Stata<sup>1</sup>

Per ottenere più rapidamente i risultati delle proprie elaborazioni si suggerisce di limitare il numero di variabili incluse nei dataset utilizzati nelle elaborazioni. Si ricorda che non si possono memorizzare dataset permanenti.

I listati contengono, in testa, un esempio di informazioni di autenticazione necessarie per la sottomissione del programma al sistema BIRD. **Ciascun utente, ovviamente, dovrà sostituire il contenuto dei primi tre campi, che iniziano con il carattere "\*", con le proprie informazioni di autenticazione.**

Lo statement **\*package** = dichiara invece il linguaggio utilizzato nel programma: Stata nel nostro caso.

I comandi Stata sono scritti in minuscolo: va ricordato che questo linguaggio è case-sensitive.

Tutti gli esempi presuppongono che il carattere di fine comando sia il punto e virgola (";").

### 1. Esempi relativi ai dati dell'indagine sulle imprese industriali e dei servizi

Negli esempi che seguono viene aperto un file di formato dati Stata contenente i dati dell'indagine. Tale archivio viene individuato con un nome simbolico convenzionale che è **\$db\_001**. Vi si mostra come limitare l'analisi a un solo settore (ad esempio, il settore industriale, `indagine==1`) o a un solo anno (ad esempio al 2005, `annoril==2005`). I primi cinque esempi presentano delle elaborazioni sulle sole imprese industriali per l'anno 2005.

#### Esempio n.1

- Stimiamo, per le sole imprese industriali (`indagine==1`) un modello logit in cui la variabile dipendente dicotomica è l'appartenenza a un gruppo di imprese. Le variabili esplicative sono il numero medio di addetti (`v24`) e le variabili relative all'area geografica della sede amministrativa e al settore di attività economica. Queste ultime due variabili sono create in modo opportuno per essere trattate come *dummy*.

```
*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;
/* creo le dummy per l'area geografica e il settore di attività economica */
tabulate areag4,gen(areag4d);
tabulate settor11,gen(settor11d);
```

<sup>1</sup> Stata è un marchio registrato della StataCorp LP, 4905 Lakeway Drive, College Station, TX 77845 USA. Gli esempi sono a cura di Leandro D'Aurizio.

```
/* sono create in questo modo 4 dummy di area geografica e 7 di settore */
/* stimo il logit, in cui ometto una dummy ciascuna per area e settore, che funge
da riferimento per le altre */
logit v521 v24 areag4d1-areag4d3 settor11d1-settor11d6 [pweight=peso];
```

### **Esempio n.2**

- Per le sole imprese industriali (indagine==1) si vuole calcolare la variazione percentuale degli addetti medi e la frazione di imprese appartenenti a un gruppo, sul totale e distintamente per area geografica. Per ottenere delle stime ponderate in modo corretto occorre eseguire le seguenti istruzioni (si noti che la creazione della variabile var\_occ serve semplicemente a ottenere stime riferite a una variazione percentuale).

```
*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);
generate 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);
```

### **Esempio n.3**

- Analogamente al precedente esempio, si vuole calcolare la variazione percentuale degli investimenti a prezzi costanti. Essi sono precedentemente trattati per limitare l'effetto dei valori anomali (*outlier*) con un procedimento chiamato *winsorizzazione del secondo tipo*, utilizzato per il calcolo delle stime degli investimenti pubblicate nel Supplementi al Bollettino Statistico dedicati all'indagine.

```
*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;
/* creo la variabile investimenti totali a prezzi costanti per il 2004 */
```

```

generate i0tot=v200cos+v810cos;
/* creo la variabile investimenti totali a prezzi costanti per il 2005 */
generate i1tot=v202cos+v811cos;
/* procedimento di winsorizzazione del secondo tipo (sulla base del 5°
e 95° percentile*/
generate diffe=(i1tot-i0tot)/v24;
generate f=1/peso;
su diffe [w=peso],de;
scalar pp5=r(p5);
scalar pp95=r(p95);
generate diffe_p5=pp5;
generate 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;
/* creo una nuova variabile i1totw contenente gli investimenti totali
2005 che attenua l'effetto dei dati anomali */
generate i1totw=i0tot+diffe*v24;
svyset _n[pw=peso],strata(strato) fpc(popstr);
generate var_inv=(i1totw-i0tot)*100;
svy:ratio var_inv/i0tot;
svy:ratio var_inv/i0tot, over(areag4);

```

#### **Esempio n.4**

- Si supponga di voler stimare un modello lineare dove il numero di addetti (variabile v24) è la dipendente e le covariate sono il fatturato (variabile v210) e l'area geografica dove è localizzata la sede amministrativa dell'impresa, quest'ultima utilizzata come variabile *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;
/* creo le dummy per l'area geografica */

```

```

tabulate areag4,gen(areag4d);
/* sono create in questo modo 4 dummy di area geografica */
/* stimo la regressione, in cui ometto una dummy per l'area, che funge da
riferimento per le altre */
regress v24 v210 areag4d1 areag4d2 areag4d3 [pweight=peso];

```

### **Esempio n.5**

- Il seguente programma replica la stessa regressione dell'esempio precedente, ma la limita alle sole imprese con numero di addetti all'interno del primo e del 99-esimo percentile della distribuzione.

```

*user = user
*password = password
*project = project
*package = stata
#delimit;
use annoril indagine peso areag4 v210 v24
if annoril ==2005 & indagine == 1
using $db_001;
/* creo le dummy per l'area geografica */
tabulate areag4,gen(areag4d);
/* sono create in questo modo 4 dummy di area geografica */
/* creo le due variabili pc1_v24 e pc99_v24
contenenti il primo e 99-esimo percentile della variabile v24 */
egen pc1_v24=pctile(v24),p(1);
egen pc99_v24=pctile(v24),p(99);
/* stimo la regressione, in cui ometto una dummy per l'area,
che funge da riferimento per le altre e escludo dalla regressione
le unità con v24 all'esterno dei percentili */
regress v24 v210 areag4d1 areag4d2 areag4d3 [pweight=peso]
if v24 >= pc1_v24 & v24 <= pc99_v24;

```

### **Esempio n.6**

- Il seguente programma presenta un esempio di stima panel ad effetti casuali su un gruppo di imprese sempre presenti negli anni considerati nel modello. L'analisi è limitata al solo settore industriale (indagine=1) per gli anni 2001-2006. Utilizziamo come variabile dipendente il fatturato (v210) e come covariate il numero medio di addetti (v24) e il risultato di esercizio (v545). La variabile v545 è prima ricodificata per essere utilizzata come *dummy*.

```

*user = user
*password = password
*project = project
*package = stata
#delimit;

```

```

use annoril indagine ident areag4 v545 v210 v24
    if annoril >=2001 & indagine == 1
    using $db_001;
/* seleziono le sole imprese presenti nei 6 anni dal 2001 al 2006 */
generate one=1;
sort ident;
by ident: egen conta=sum(one);
keep if conta==6;
/* creo le dummy per il risultato di esercizio */
tabulate v545,gen(v545d);
/* stimo il modello di regressione sul panel, in cui ometto una dummy per il
risultato di esercizio, che funge da riferimento per le altre */
iis ident;
tis annoril;
xtreg v210 v24 v545d1 v545d2 v545d3 v545d4,re;

```

## 2. Esempi relativi ai dati del sondaggio congiunturale sulle imprese industriali e dei servizi

### **Esempio n.7**

- Si vuole tabulare nell'archivio storico, per tutti gli anni disponibili, la distribuzione di frequenza delle modalità di risposta alla variabile STG3 (investimenti programmati per l'anno successivo) per le sole imprese manifatturiere con 50 addetti e oltre (indag3==1).

```

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

```

### **Esempio n.8**

- Il seguente programma presenta un esempio di *merge* tra l'archivio storico del sondaggio e quello con i dati dell'indagine annuale, al fine di confrontare i piani di investimento rilevati nel corso del sondaggio del 2006 e le corrispondenti realizzazioni (rilevate in forma continua nell'indagine annuale sul 2007 e successivamente discretizzate). Sono elaborati i dati solo per le imprese che hanno partecipato a entrambe le indagini e che nel sondaggio hanno fornito un dato valido (diverso dalla modalità 9="non so, non intendo rispondere").

```

*user = user
*password = password
*project = project
*package = stata

```

```
#delimit;
/* carico i dati relativi al sondaggio 2006 */
use annoril ident stg3 using $db_sondstor if annoril==2006;
sort ident;
save sond2006,replace;
clear;
/* carico i dati relativi all'indagine annuale 2007 */
use annoril ident v200 v810 v202 v811 peso using $db_001
if annoril==2007;
/* costruisco la variazione degli investimenti */
generate i0tot=v200+v810;
generate i1tot=v202+v811;
/* sostituisco gli zeri con valori molto piccoli positivi per ottenere
una variazione valida anche nel caso che il termine i0tot o entrambi
i termini valgano zero */
replace i0tot=0.1 if i0tot==0;
replace i1tot=0.1 if i1tot==0;
generate varinv=(i1tot/i0tot-1)*100;
/* discretizzo la variabile continua 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;
/*conservo solo le unità presenti in entrambi gli archivi*/
keep if _merge==3;
tabulate stg3 varinvd [aweight=peso] if stg3 !=9,cell;
```