



Indagine sulle Indagine sulle  
imprese industriali e dei servizi  
(INVIND)  
Sondaggio congiunturale sulle  
Imprese Industriali e dei Servizi  
(SONDTEL)

Esempi di utilizzo dati: Piattaforma Stata



## Sommario

Esempi di utilizzo dei dati: piattaforma Stata .....	3
1. Esempi relativi ai dati dell'indagine sulle imprese industriali e dei servizi.....	3
Esempio n. 1: regressione logistica .....	3
Esempio n. 2: distribuzioni di frequenza .....	3
Esempio n. 3: regressione lineare .....	4
Esempio n. 4: regressione lineare (winsorizzata) .....	4
Esempio n. 5: regressione con effetti casuali per dati panel .....	5
2. Esempi relativi ai dati del sondaggio congiunturale sulle imprese industriali e dei servizi .....	5
Esempio n. 6: distribuzione di frequenza .....	5
Esempio n. 7: integrazione delle due indagini .....	5



## 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 utilizzate nelle elaborazioni.

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

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

Negli esempi che seguono viene importato un file CSV contenente i dati dell'indagine. Tale file viene individuato dal menu dei dataset disponibili. 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 2021, `annoril==2021`). I primi quattro esempi presentano delle elaborazioni sulle sole imprese industriali per l'anno 2021, il quinto mostra un'analisi panel, il sesto sui dati storici del sondaggio congiunturale, e l'ultimo mostra come operare un merge tra i dati dell'archivio annuale e quelli del sondaggio congiunturale.

#### Esempio n. 1: regressione logistica

- 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*.

```
import delimited "indann_completo_csv.csv"
keep annoril indagine peso areag4 settor11 v521 v24
keep if annoril==2021 & indagine == 1
/* creo le dummy per l'area geografica e il settore di attività economica */
tabulate areag4, gen(areag4d)
tabulate settor11, gen(settor11d)
/* 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: distribuzioni di frequenza

- 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).

<sup>1</sup> Stata è un marchio registrato della StataCorp LP, 4905 Lakeway Drive, College Station, TX 77845 USA.  
Banca d'Italia



```
import delimited "indann_completo_csv.csv"
keep annoril indagine peso popstr strato areag4 settor11 v521 v15 v24
keep if annoril ==2021 & indagine == 1
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: regressione lineare

- 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*.

```
import delimited "indann_completo_csv.csv"
keep annoril indagine peso areag4 v210 v24
keep if annoril ==2021 & indagine == 1
/* 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. 4: regressione lineare (winsorizzata)

- 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.

```
import delimited "indann_completo_csv.csv"
keep annoril indagine peso areag4 v210 v24
keep if annoril ==2021 & indagine == 1
/* 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. 5: regressione con effetti casuali per dati panel

- 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 2016-2021. 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*.

```
import delimited "indann_completo_csv.csv"
keep annoril indagine ident areag4 v545 v210 v24
keep if annoril >=2016 & annoril <=2021 & indagine == 1
/* seleziono le sole imprese presenti nei 6 anni dal 2016 al 2021 */
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. 6: distribuzione di frequenza

- 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).

```
import delimited "sondstor.csv"
keep annoril indag3 cc2 stg3 pesorisc
sort annoril
by annoril: tabulate stg3 [iweight=pesorisc] if cc2>=2 & indag3==1
```

### Esempio n. 7: integrazione delle due indagini

- 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 2021 e le corrispondenti realizzazioni (rilevate in forma continua nell'indagine annuale sul 2021 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").

```
/* carico i dati relativi al sondaggio 2021*/
import delimited "sondstor.csv"
keep annoril ident stg3
keep if annoril==2021
sort ident
save sond2021, replace
clear
/* carico i dati relativi all'indagine annuale 2021 */
import delimited "indann_completo_csv.csv"
keep annoril ident v200 v810 v202 v811 peso
keep if annoril==2021
/* 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 sond2021
/*conservo solo le unità presenti in entrambi gli archivi*/
keep if _merge==3
tabulate stg3 varinvd [iweight=peso] if stg3 !=9, cell
```