



# Survey of Industrial and service Firms (INVIND) Business Outlook Survey of Industrial and Service Firms (SONDTEL)

Stata Examples



## Table of contents

Examples of data use: Stata platform .....	3
1. Examples using data from the Survey of industrial and service firms .....	3
Example n. 1: logistic regression .....	3
Example n. 2: frequency distributions .....	3
Example n. 3: linear regression .....	4
Example n. 4: linear regression (winsorized) .....	4
Example n. 5 panel data regression with causal effects.....	5
2. Examples using data from the Business Outlook Survey of Industrial and Service Firms .....	6
Example n. 6 frequency distribution .....	6
Example n.7 merging the two surveys.....	6



## Examples of data use: Stata platform<sup>1</sup>

To obtain the results of calculations more rapidly, you should limit the number of variables included in the datasets used.

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

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

In each of these examples a CSV file is imported in Stata format, containing the survey data. The file name is identified in the section "Available datasets". You are shown how to restrict the analysis to a single sector (for example, the industrial sector, `indagine==1`) or year (for example, 2021, `annoril==2021`). The first four examples show calculations only for industrial firms for 2021, the fifth example shows a panel analysis, the sixth example shows an analysis of time series data from the Business Outlook Survey, and the last example shows how to merge information from the annual database and the Business Outlook Survey.

#### Example n. 1: logistic regression

- We estimate a logit model, for industrial firms only (`indagine==1`), 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.

```
import delimited "indann_completo_csv.csv"
keep annoril indagine peso areag4 settor11 v521 v24
keep if annoril==2021 & indagine == 1
/* 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 */
logit v521 v24 areag4d1-areag4d3 settor11d1-settor11d6 [pweight=peso]
```

#### Example n. 2: frequency distributions

- 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).

<sup>1</sup> Stata is a registered trademark of StataCorp LP, 4905 Lakeway Drive, College Station, TX 77845 USA.  
Bank of Italy



```

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)
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: linear regression

- 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.

```

import delimited "indann_completo_csv.csv"
keep annoril indagine peso areag4 v210 v24
keep if annoril ==2021 & indagine == 1
/* 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. 4: linear regression (winsorized)

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

```

import delimited "indann_completo_csv.csv"
keep annoril indagine peso areag4 v210 v24
keep if annoril == 2021 & indagine == 1
/* 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. 5 panel data regression with causal effects

- 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 (indagine==1) in the years 2016-2021. 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.

```
import delimited "indann_completo_csv.csv"
keep annoril indagine ident areag4 v545 v210 v24
keep if annoril >=2016 & annoril <=2021 & indagine == 1
/* selection of the firms covered by the survey in the 6 years from 2016 to 2021
*/
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. 6 frequency distribution

- 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).

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

### Example n.7 merging the two surveys

- 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 2021 Outlook Survey to corresponding realizations (surveyed as continuous variables and then discretized) in the 2021 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.

```
/* load data from Business Outlook survey 2021 */
import delimited "sondstor.csv"
keep annoril ident stg3
keep if annoril==2021
sort ident
save sond2021, replace
clear
/* load data from Annual Survey 2021 */
import delimited "indann_completo_csv.csv"
keep annoril ident v200 v810 v202 v811 peso
keep if annoril==2021
/* 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 sond2021
/* keep only firms appearing in both datasets */
keep if _merge==3
tabulate stg3 varinvd [iweight=peso] if stg3 !=9, cell
```

