

# Descrizione archivi ed esempi di utilizzo del Sondaggio congiunturale sulle Imprese Industriali e dei Servizi

30 giugno 2023

Per informazioni: [statistiche@bancaditalia.it](mailto:statistiche@bancaditalia.it)  
[www.bancaditalia.it/statistiche/index.html](http://www.bancaditalia.it/statistiche/index.html)

## 1. Indicazioni generali

Le informazioni raccolte nell'ambito del Sondaggio congiunturale sulle imprese industriali e dei servizi (di seguito Sondaggio), condotto annualmente tra settembre e ottobre, sono disponibili a partire dal **1993** per le imprese dell'industria in senso stretto e dei servizi e dal **2007** per le imprese delle costruzioni.

Il disegno campionario del Sondaggio riflette sostanzialmente quello dell'Indagine sulle imprese industriali e dei servizi (INVIND), condotta nella primavera dello stesso anno. Le imprese contattate sono in principio le stesse ed eventuali differenze nella composizione del campione per le due rilevazioni sono dovute esclusivamente all'attrito campionario e all'eventuale uscita di imprese dalla popolazione di riferimento dell'indagine durante l'estate.

Nel corso del Sondaggio, alle imprese sono posti quesiti di natura prevalentemente qualitativa sull'andamento delle principali variabili economiche e di aggiornamento delle informazioni rilevate nell'indagine INVIND dello stesso anno. Poiché una consistente parte del questionario varia di anno in anno, limitatamente alle imprese dell'industria in senso stretto e dei servizi, gli utenti hanno a disposizione due differenti archivi:

- **db\_sondstor**, in cui sono presenti solo le variabili rilevate in maniera continuativa nel corso degli anni. I vari anni sono identificati dalla variabile **annoril**. Si noti che nel Sondaggio, la variabile **annoril** indica l'anno in cui viene effettuata la rilevazione (ad esempio: 2007 per quello condotto tra settembre e ottobre 2007), mentre nell'Indagine sulle imprese industriali e dei servizi (INVIND) indica l'anno oggetto di rilevazione, ovvero l'anno precedente l'effettiva raccolta dei dati (ad esempio: 2007 per l'indagine condotta nei primi mesi del 2008);
- **db\_sondxxxx**, in cui sono presenti tutte le variabili rilevate nell'anno xxxx.

Nei database **db\_sondxxxx**, nel corso degli anni, alcune variabili che rilevano lo stesso fenomeno potrebbero avere un nome diverso e una scala di risposte parzialmente differenti. Nel dataset **db\_sondstor**, alle variabili aventi lo stesso significato, nel corso del tempo è stata attribuita una codifica univoca ed è stata uniformata la scala di valori in modo da garantire comparabilità temporale. Ad esempio, il nome della variabile "Investimenti programmati per l'anno successivo" è s3 nell'archivio **db\_sond1993** e x3 nel **db\_sond1994**, mentre nell'archivio **db\_sondstor** il nome della variabile è stg3 in tutti gli anni. Per le imprese delle costruzioni, l'archivio **db\_sondcost** è unico e contiene tutte le variabili misurate nei vari anni, con nomi che potrebbero variare nel corso del tempo.

Per un elenco completo delle variabili rilevate nei singoli anni e della loro ricodifica nell'archivio storico si consulti il *Database delle variabili*, [disponibile nella pagina internet dell'Istituto](#).

Ogni impresa ha un codice identificativo (la variabile **ident**) che consente l'aggancio delle informazioni nel solo database **db\_sondstor** nel caso in cui sia stata oggetto di rilevazione in più anni. Questo codice, generato in modo casuale, serve esclusivamente per le analisi longitudinali, non può essere tabulato ed è del tutto incorrelato con le variabili identificative delle imprese.

Negli archivi **db\_sondstor** e **db\_sondcost** la coppia (**annoril**, **ident**) identifica le informazioni riguardanti una data impresa in un dato anno. Si noti che, se la stessa impresa è rilevata in occasione dell'Indagine sulle imprese industriali e dei servizi (INVIND) e del Sondaggio, essa presenta la stessa **ident** in tutti gli archivi.

A puro scopo esemplificativo, la struttura logica degli archivi **db\_sondstor** e **db\_sondcost** (che l'utente non può visualizzare), è quindi la seguente:

ANNORIL	IDENT	A5	STG1	STG3	...	...	...	...	...	...
1993	1	40	1	4	...	...	...	...	...	...
1993	2	190	2	9	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...
2020	1	35	3	1	...	...	...	...	...	...
2020	2	240	9	2	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...

Contestualmente, la struttura logica ad esempio degli archivi **db\_sond1993** e **db\_sond2020** (che l'utente non può visualizzare), sarà invece la seguente<sup>1</sup>:

#### DB\_SOND1993

IDENT	a5m	s1	s3	...	...	...	...	...	...
1	40	1	4	...	...	...	...	...	...
2	190	2	9	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...

#### DB\_SOND2020

IDENT	a5m	p1	p3	...	...	...	...	...	...
1	35	3	1	...	...	...	...	...	...
2	240	9	2	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...

Si noti che il questionario del Sondaggio viene considerato valido solo se completamente compilato. A tal fine tutte le variabili hanno tra le modalità di risposta il "9 = non so, non intendo rispondere" e, quando necessario, l'"8 = non applicabile". I valori missing non vanno quindi interpretati come un rifiuto dell'impresa a fornire una data risposta, ma come quesiti posti solo ad un sottoinsieme di imprese; in alcuni casi, il gruppo di imprese che non riceve tali quesiti può avere il relativo campo valorizzato con il numero "99".

Gli archivi contengono alcune variabili relative al riporto all'universo delle stime campionarie. L'adozione del peso campionario consente di allineare la struttura del campione a quella dell'universo secondo le variabili di stratificazione<sup>2</sup>; se ne raccomanda l'uso nelle analisi per ottenere stime non distorte che riportino alla popolazione di riferimento.

## 2. Variabili disponibili negli archivi

Per motivi di riservatezza, negli archivi **non** sono disponibili le variabili del questionario che permetterebbero l'identificazione dell'impresa rispondente, solitamente presenti nelle prime pagine dei questionari, tra le quali: Codice Fiscale, Ragione sociale, Filiale di rilevazione e Gruppo di appartenenza. Non sono inoltre rese disponibili le risposte testuali di tipo "Altro specificare".

Al contrario, gli archivi contengono alcune variabili non presenti nei questionari ma utili ai fini delle elaborazioni. Tra queste, oltre l'anno di riferimento dell'indagine (variabile **annoril**) ci sono:

<sup>1</sup> Le variabili ident, mostrate qui al solo scopo di chiarire la relazione tra gli archivi, non sono presenti nei dataset annuali.

<sup>2</sup> Poiché la numerosità della popolazione è nota con ritardo, talvolta anche di uno o due anni, i pesi sono calcolati adoperando le numerosità della popolazione negli strati dell'anno più recente. I piccoli scostamenti tra l'effettiva numerosità della popolazione di riferimento e la somma dei pesi sono dovuti al processo di post-stratificazione.

a) Variabili di classificazione rispetto al settore di attività economica<sup>3</sup>

Nome	Valori	Descrizione	ATECO 2002	ATECO 2007
<b>settore11</b>	SS1	Industrie alimentari, bevande e tabacco	DA	10, 11, 12
	SS2	Industrie tessili, dell'abbigliamento, pelli cuoio e calzature	DB, DC	13, 14, 15
	SS3	Fabbricazione di coke, industria chimica, gomma e plastica	DF, DG, DH	19, 20, 21, 22
	SS4	Industria della lavorazione dei minerali non metalliferi	DI	23
	SS5	Industria metalmeccanica	DJ, DK, DL, DM	24, 25, 26, 27, 28, 29, 30, 33
	SS6	Altre industrie manifatturiere	DD, DE, DN	16, 17, 18, 31, 32
	SS7	Altre industrie in senso stretto	CA, CB, CE	05, 06, 07, 08, 09, 35, 36, 37, 38, 39
	SS8	Commercio ingrosso e dettaglio, riparazione di autoveicoli e motocicli	G	45, 46, 47
	SS9	Servizi di alloggio e ristorazione	H	55, 56
	SS10	Trasporto e magazzinaggio, servizi di informazione e comunicazione	I	49, 50, 51, 52, 53, 58, 59, 60, 61, 62, 63
	SS11	Attività immobiliari, professionali, scientifiche e tecniche, amministrative e di servizi di supporto	K	68, 69, 70, 71, 72, 73, 74, 75, 77, 78, 79, 80, 81, 82
<b>indag3</b>	1	Industria manifatturiera	D	C
	2	Industria energetico – estrattiva	C, E	B, D, E
	3	Servizi	G, I, H, K	G, I, H, J, L, M, N
<b>indagine</b>	1	Industria in senso stretto	C, D, E	C, B, D, E
	2	Servizi	G, I, H, K	G, I, H, J, L, M, N

b) Variabili di classificazione rispetto alla classe dimensionale<sup>4</sup>

Nome	Valori	Descrizione
<b>cldimet</b>	0	20 - 49 addetti
	1	50 - 99 addetti
	2	100 - 199 addetti
	3	200 - 499 addetti
	4	500 - 999 addetti
	5	1.000 addetti e oltre
<b>cc2</b>	1	20 - 49 addetti
	2	50 addetti e oltre

c) Variabili di classificazione rispetto all'area geografica<sup>5</sup>

Nome	Valori	Descrizione
<b>areag4</b>	1	Nord ovest
	2	Nord est
	3	Centro
	4	Sud e Isole
<b>areag2</b>	1	Nord, Centro
	2	Sud e isole

<sup>3</sup> Fino al 2009 tali variabili erano ottenute dall'aggregazione di alcune sottosezioni ATECO 2002; dal 2010 sono basate sulle prime due cifre della classificazione ATECO 2007. Le variabili di questa tabella sono presenti nei soli archivi relativi alle imprese dell'industria in senso stretto e dei servizi; tutte le imprese dell'archivio costruzioni hanno invece un unico codice ateco (2002, 2007).

<sup>4</sup> Fino all'anno di riferimento 2003 la classe dimensionale è riferita al numero di addetti a fine anno; dal 2004 in poi al numero di addetti medi nell'anno.

<sup>5</sup> Per motivi di riservatezza non sono disponibili le classificazioni relative alle singole regioni e province, ma solo quelle relative alle macroaree geografiche.

d) *Variabili relative al disegno campionario e al sistema di ponderazione*<sup>6</sup>

<b>strato:</b>	Formato dalle 68 combinazioni di <b>settor11</b> e <b>cldimet</b> . Si tenga presente che le imprese con almeno 5000 addetti hanno peso unitario e sono considerate, separatamente per industria e servizi, facenti parte di due strati a parte (rispettivamente 67 e 68).
<b>poststrato:</b>	Formato dalle 48 combinazioni di <b>areag4</b> , <b>cc2</b> e una riaggregazione dei settori di attività economica in 6 gruppi: 1) indag3=1; 2) indag3=2; 3) settor11=SS8; 4) settor11=SS9; 5) settor11=SS10; 6) settor11=SS11.
<b>pesorisc:</b>	Peso campionario (di espansione): a livello di strato e poststrato, la somma dei pesi riproduce la distribuzione delle imprese della popolazione di riferimento dell'indagine e somma al numero di imprese del campione. Alle imprese con più di 5000 addetti (c.d. autorappresentative) e ad un numero limitato di imprese che non si ritengono rappresentative dello strato a cui appartengono si attribuisce peso unitario.
<b>pesoadd:</b>	Peso campionario (di espansione): a livello di strato e poststrato, la somma dei pesi riproduce la distribuzione degli addetti della popolazione di riferimento dell'indagine e somma al numero degli addetti delle imprese del campione. La variabile pesoadd è raccomandata se si desidera produrre una distribuzione di frequenze delle modalità di risposta per una variabile che tenga conto della diversa dimensione delle imprese e di conseguenza attribuisca maggiore importanza alla risposta fornita dalle imprese più grandi in termini di addetti.
<b>popstr:</b>	Numerosità della popolazione a livello di strato.
<b>poppost:</b>	Numerosità della popolazione a livello di post-strato.

a) *Variabili di classificazione della quota del fatturato esportato*

Nome	Valori	Descrizione
<b>a6</b>	1	impresa non esportatrice
	2	meno di 1/3 di fatturato esportato
	3	tra 1/3 e 2/3 di fatturato esportato
	4	oltre 2/3 di fatturato esportato
<b>qexp</b>	1	meno di 1/3 di fatturato esportato o nessuna esportazione
	2	tra 1/3 e 2/3 di fatturato esportato
	3	oltre 2/3 di fatturato esportato

### 3. Esempi di utilizzo degli archivi

#### 3.1. Esempi basati sul software R<sup>7</sup>

Per ottenere più rapidamente i risultati delle proprie elaborazioni si suggerisce di limitare il numero di variabili incluse nei dataset utilizzati per le stime. Si ricorda che R è un linguaggio case-sensitive.

Tutti gli esempi presuppongono che su ogni riga ci sia un solo comando e che lo stesso comando possa essere esteso su più righe, se troppo lungo.

#### Esempio 1: 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).

<sup>6</sup> Per maggiori dettagli sul disegno campionario, la costruzione dei pesi, dei deflatori e per tutti gli altri aspetti riguardanti la metodologia utilizzata in ogni anno di indagine, si rimanda alla **Nota metodologica** pubblicata congiuntamente a ciascuno dei **Fascicoli** dedicati alle indagini e reperibili sul sito internet della Banca d'Italia.

<sup>7</sup> R è un ambiente open-source per l'analisi statistica dei dati; se si desiderano ulteriori informazioni sul linguaggio, si consiglia di visitare il sito <http://cran.r-project.org/>.

```

##funzioni per La manipolazione dei dati
library(dplyr)
library(data.table)

##funzioni per La gestione del disegno campionario
library(survey)

##funzioni per statistiche descrittive di dati
library(gtsummary)

##Lettura dei dati
dati <- fread("sondstor.csv")

oggetto <- dati %>% filter(cc2>=2, indag3 == 1)

##Lettura delle variabili di interesse
oggetto <- oggetto %>% select(annoril, indag3, cc2, stg3, pesorisc)

out_svy <- svydesign(id= ~1, weights= ~pesorisc,
                   data=oggetto)
summary(out_svy)

##calcolo delle frequenze relative pesate
out_svy %>% tbl_svysummary(by = annoril,
                          percent = "column",
                          include = stg3,
                          statistic = list( stg3 ~ "{p}%" )) %>%
  modify_footnote(c(all_stat_cols()) ~ NA) %>%
  modify_header(c(all_stat_cols()) ~ "***{level}***", label = "")

```

## Esempio 2: 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").

```

##funzioni per La manipolazione dei dati
library(dplyr)

library(data.table)

##Lettura dei dati storici del sondaggio
dati <- fread("sondstor.csv")

##Lettura delle variabili di interesse
##creazione del data frame con le variabili di interesse e filtro
##per anno=2006 e stg3<>9
sond2021 <- dati %>% filter(annoril == 2021, stg3 !=9) %>%
  select(annoril, ident, stg3)

##Lettura dei dati sull'indagine INVIND 2021
dati <- fread("indann_completo_csv.csv")

```

```

invind2021 <- dati %>% filter(annoril == 2021) %>%
  select(annoril, ident, v200, v810, v202, v811, peso)%>%
  mutate(
    i0tot=v200+v810,
    i1tot=v202+v811,
    i0tot=if_else(i0tot==0, 0.1, i0tot),
    i1tot=if_else(i1tot==0, 0.1, i1tot),
    varinv=(i1tot/i0tot-1)*100,
    varinvd= cut(varinv,
                 breaks = c(-Inf, -10, -3, 3, 10,Inf),
                 labels = c(1,2,3,4,5)))

##merge dei due dataframe in modo da mantenere solo le imprese
##presenti in entrambi
out_merge=inner_join(sond2021,invind2021) %>% select(stg3, peso, varinvd)

out_svy <- svydesign(id= ~1, weights= ~peso,
                  data=out_merge)
summary(out_svy)

##calcolo delle frequenze relative pesate
out_svy %>% tbl_svysummary(by = stg3,
                          percent = "cell",
                          include = varinvd,
                          statistic = list( stg3 ~ "{p}%", varinvd ~ "{p}%" )) %>%
  modify_footnote(c(all_stat_cols()) ~ NA) %>%
  modify_header(c(all_stat_cols()) ~ "**{level}**", label = "stg3")

```

### 3.2. Esempi basati sul software Stata<sup>8</sup>

Essendo anche questo software case-sensitive, tutti i comandi vanno scritti in minuscolo.

#### Esempio n. 1

- 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 ( $indag3==1$ ).

```

#delimit;
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. 2

- 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").

<sup>8</sup> Stata è un marchio registrato della StataCorp LP, 4905 Lakeway Drive, College Station, TX 77845 USA.

```

#delimit;
/* carico i dati relativi al sondaggio 2021 */
import delimited "sond2021.csv";
keep annoril ident stg3 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 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 v
alida 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;

```