

# Agenda

---

- ✦ Censimento e campionamento
- ✦ Campionamento ripetuto: distribuzione campionaria delle statistiche
- ✦ Media campionaria e frequenza campionaria
- ✦ Inferenza statistica: stima puntuale
- ✦ Popolazioni finite
- ✦ Disegni di campionamento probabilistici:
  - ◆ Piano di campionamento Casuale Semplice
  - ◆ Piano di campionamento Sistemático
  - ◆ Piano di campionamento Stratificato
- ✦ Disegni di campionamento non probabilistici: campionamento per Quote
- ✦ Errore non campionario

# Censimento e campionamento (1)

Si chiama campione ogni sottoinsieme di unità della popolazione oggetto di studio, selezionate allo scopo di trarne conoscenze sull'intera popolazione.

## ✦ Vantaggi del censimento

- ◆ Consente di conoscere interamente una popolazione rispetto ad un insieme di caratteri e di ricavare il valore esatto (?) dei parametri descrittivi di interesse.
- ◆ Consente l'analisi della distribuzione di uno o più caratteri in una qualsiasi sottopopolazione possa interessare, piccola quanto si vuole.

## ✦ Svantaggi del censimento

- ◆ Tempi di svolgimento molto lunghi e costi estremamente elevati.
  - ◆ Bassa **qualità dei dati** (intendendo con questo termine la corrispondenza tra i dati raccolti e la realtà effettiva)
-

# Censimento e campionamento (2)

---

## ☀ Vantaggi del campionamento

- ◆ Tempi ridotti di esecuzione e costi contenuti.
- ◆ Possibile elevata qualità dei dati (accuratezza)

## ☀ Svantaggi del campionamento

- ◆ Non si possono ricavare risultati attendibili per sottopopolazioni più piccole di quelle considerate nello stabilire la dimensione del campione.
  - ◆ Presenza dell'**errore campionario** (scostamento tra stima del parametro e valore dello stesso nella popolazione )
-

# Fonti di errore nelle rilevazioni statistiche

---

L'**errore totale** – scostamento tra i valori calcolati dei parametri di interesse ed il valore effettivo nella popolazione considerata – è somma delle diverse fonti di **errore non campionario** e dell'**errore campionario**

## ✦ Fonti di errore non campionario

- ◆ **Progettazione lacunosa dell'indagine.** (pertinenza dei caratteri da rilevare rispetto al problema da studiare, la corretta individuazione della popolazione, etc.).
  - ◆ **Difettosità delle liste** da cui viene estratto il campione (nominativi mancanti, nominativi ripetuti, inesistenti o incompleti, etc.)
  - ◆ **Errori di misurazione** (generati dal rispondente, dall'intervistatore, dal questionario, da mancata osservazione, da errori di trattamento dei dati).
-

# Errore campionario e campione rappresentativo

---

- ✦ Il controllo dell'errore campionario richiede che il campione sia rappresentativo
- ✦ Campione rappresentativo (ideale): nel campione la distribuzione di frequenza relativa dei caratteri di interesse è uguale a quelle dell'intera popolazione

Uno schema di campionamento si dice probabilistico se ogni unità della popolazione ha una probabilità *positiva e conosciuta* di entrare a far parte del campione.

Un campionamento probabilistico ben fatto garantisce la rappresentatività del campione che ne scaturisce.

---

# Il principio del campionamento ripetuto

---

## POPOLAZIONE DI 4 UNITA'

Unità	$u_1$	$u_2$	$u_3$	$u_4$
Intensità	12	12	15	27

## Distribuzione del carattere nella popolazione

Modalità	$f_i$
12	0,50
15	0,25
27	0,25
	1,00

$$\mu = 16,5; \quad \sigma^2 = 38,25, \quad \sigma = 6,18.$$

# Campionamento casuale semplice CON ripetizione

---

Si estraggano 3 unità con ripetizione e siano

$X_1$  = risultato della prima **osservazione**

$X_2$  = risultato della seconda **osservazione**

$X_3$  = risultato della terza **osservazione**

Funzioni delle osservazioni

$$\bar{X} = \frac{X_1 + X_2 + X_3}{3}, \text{ media campionaria}$$

$$S^2 = \frac{1}{3} \sum_{i=1}^3 (X_i - \bar{X})^2, \text{ varianza campionaria}$$

$$S_c^2 = \frac{1}{2} \sum_{i=1}^3 (X_i - \bar{X})^2, \text{ varianza campionaria corretta}$$

Si chiama statistica campionaria ogni funzione delle osservazioni che non dipende da quantità incognite ed è calcolabile quale che sia il campione estratto.

---

# UNIVERSO DEI CAMPIONI

Campi.	$X_1$	$X_2$	$X_3$	$\bar{X}$	$S^2$	$S_c^2$	Camp.	$X_1$	$X_2$	$X_3$	$\bar{X}$	$S^2$	$S_c^2$
$u_1, u_1, u_1$	12	12	12	<b>12</b>	0	0	$u_3, u_1, u_1$	15	12	12	<b>13</b>	2	3
$u_1, u_1, u_2$	12	12	12	<b>12</b>	0	0	$u_3, u_1, u_2$	15	12	12	<b>13</b>	2	3
$u_1, u_1, u_3$	12	12	15	<b>13</b>	2	3	$u_3, u_1, u_3$	15	12	15	<b>14</b>	2	3
$u_1, u_1, u_4$	12	12	27	<b>17</b>	50	75	$u_3, u_1, u_4$	15	12	27	<b>18</b>	42	63
$u_1, u_2, u_1$	12	12	12	<b>12</b>	0	0	$u_3, u_2, u_1$	15	12	12	<b>13</b>	2	3
$u_1, u_2, u_2$	12	12	12	<b>12</b>	0	0	$u_3, u_2, u_2$	15	12	12	<b>13</b>	2	3
$u_1, u_2, u_3$	12	12	15	<b>13</b>	2	3	$u_3, u_2, u_3$	15	12	15	<b>14</b>	2	3
$u_1, u_2, u_4$	12	12	27	<b>17</b>	50	75	$u_3, u_2, u_4$	15	12	27	<b>18</b>	42	63
$u_1, u_3, u_1$	12	15	12	<b>13</b>	2	3	$u_3, u_3, u_1$	15	15	12	<b>14</b>	2	3
$u_1, u_3, u_2$	12	15	12	<b>13</b>	2	3	$u_3, u_3, u_2$	15	15	12	<b>14</b>	2	3
$u_1, u_3, u_3$	12	15	15	<b>14</b>	2	3	$u_3, u_3, u_3$	15	15	15	<b>15</b>	0	0
$u_1, u_3, u_4$	12	15	27	<b>18</b>	42	63	$u_3, u_3, u_4$	15	15	27	<b>19</b>	32	48
$u_1, u_4, u_1$	12	27	12	<b>17</b>	50	75	$u_3, u_4, u_1$	15	27	12	<b>18</b>	42	63
$u_1, u_4, u_2$	12	27	12	<b>17</b>	50	65	$u_3, u_4, u_2$	15	27	12	<b>18</b>	42	63
$u_1, u_4, u_3$	12	27	15	<b>18</b>	42	63	$u_3, u_4, u_3$	15	27	15	<b>19</b>	32	48
$u_1, u_4, u_4$	12	27	27	<b>22</b>	50	75	$u_3, u_4, u_4$	15	27	27	<b>23</b>	32	48
$u_2, u_1, u_1$	12	12	12	<b>12</b>	0	0	$u_4, u_1, u_1$	27	12	12	<b>17</b>	50	75
$u_2, u_1, u_2$	12	12	12	<b>12</b>	0	0	$u_4, u_1, u_2$	27	12	12	<b>17</b>	50	75
$u_2, u_1, u_3$	12	12	15	<b>13</b>	2	3	$u_4, u_1, u_3$	27	12	15	<b>18</b>	42	63
$u_2, u_1, u_4$	12	12	27	<b>17</b>	50	75	$u_4, u_1, u_4$	27	12	27	<b>22</b>	50	75
$u_2, u_2, u_1$	12	12	12	<b>12</b>	0	0	$u_4, u_2, u_1$	27	12	12	<b>17</b>	50	75
$u_2, u_2, u_2$	12	12	12	<b>12</b>	0	0	$u_4, u_2, u_2$	27	12	12	<b>17</b>	50	75
$u_2, u_2, u_3$	12	12	15	<b>13</b>	2	3	$u_4, u_2, u_3$	27	12	15	<b>18</b>	42	63
$u_2, u_2, u_4$	12	12	27	<b>17</b>	50	75	$u_4, u_2, u_4$	27	12	27	<b>22</b>	50	75
$u_2, u_3, u_1$	12	15	12	<b>13</b>	2	3	$u_4, u_3, u_1$	27	15	12	<b>18</b>	42	63
$u_2, u_3, u_2$	12	15	12	<b>13</b>	2	3	$u_4, u_3, u_2$	27	15	12	<b>18</b>	42	63
$u_2, u_3, u_3$	12	15	15	<b>14</b>	2	3	$u_4, u_3, u_3$	27	15	15	<b>32</b>	32	48
$u_2, u_3, u_4$	12	15	27	<b>18</b>	42	63	$u_4, u_3, u_4$	27	15	27	<b>32</b>	32	48
$u_2, u_4, u_1$	12	27	12	<b>17</b>	50	75	$u_4, u_4, u_1$	27	27	12	<b>22</b>	50	75
$u_2, u_4, u_2$	12	27	12	<b>17</b>	42	63	$u_4, u_4, u_2$	27	27	12	<b>22</b>	50	75
$u_2, u_4, u_3$	12	27	15	<b>18</b>	50	75	$u_4, u_4, u_3$	27	27	15	<b>23</b>	32	48
$u_2, u_4, u_4$	12	27	27	<b>22</b>	2	3	$u_4, u_4, u_4$	27	27	27	<b>27</b>	0	0

# Distribuzione delle statistiche campionarie

$X_1, X_2, X_3$  sono variabili casuali INDIPENDENTI e IDENTICAMENTE DISTRIBUITE

$$E(X_1) = E(X_2) = E(X_3) = 16,5;$$

$$V(X_1) = V(X_2) = V(X_3) = 38,25.$$

*Distribuzione di probabilità di  $\bar{X}$*

$\bar{X}$	12	13	14	15	17
$p(\bar{x})$	0,1250	0,1875	0,0938	0,0156	0,1874
$\bar{X}$	18	19	22	23	27
$p(\bar{x})$	0,1875	0,0469	0,0938	0,0469	0,0156

$$E(\bar{X}) = 12 \cdot 0,1250 + 13 \cdot 0,1875 + 14 \cdot 0,0938 + \dots + 27 \cdot 0,0156 = \underline{16,5}$$

$$V(\bar{X}) = (12-16,5)^2 \cdot 0,1250 + (13-16,5)^2 \cdot 0,1875 + \dots + (27-16,5)^2 \cdot 0,0156 = \underline{12,75} \quad \Rightarrow \quad \sigma_{\bar{X}} = \underline{3,57}$$

*Distribuzione di probabilità di  $S^2$*

$S^2$	0	2	32	42	50
$p(s^2)$	0,1562	0,2813	0,0937	0,1875	0,2813

$$E(S^2) = \underline{25,5}.$$

*Distribuzione di probabilità di  $S_c^2$*

$S_c^2$	0	3	48	63	75
$p(s_c^2)$	0,1562	0,2813	0,0937	0,1875	0,2813

$$E(S_c^2) = \underline{38,25}.$$

# La distribuzione campionaria della media del campione

Si consideri una popolazione ove  $X$  ha media  $\mu$  e varianza  $\sigma^2$   
 $\mu$  e  $\sigma^2$  parametri della popolazione

Dato un campione di osservazioni  $X_1, X_2, \dots, X_n$  si consideri la

$$\text{Media campionaria: } \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

## RISULTATI

$$1. E(\bar{X}) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} \sum_{i=1}^n \mu = \mu$$

$$2. V(\bar{X}) = V\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n V(X_i) = \frac{1}{n^2} \sum_{i=1}^n \sigma^2 = \frac{\sigma^2}{n}$$

$$3. \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

$$4. \bar{X} \stackrel{\text{asint}}{\sim} N\left[\mu, \frac{\sigma^2}{n}\right]$$

5. Se  $X \sim N(\mu; \sigma^2)$  allora  $\bar{X}$  è esattamente normale.

---

# La distribuzione campionaria della frequenza relativa del campione

Si consideri una popolazione in cui un **attributo** ha frequenza relativa  $\pi$ , con  $0 \leq \pi \leq 1$ .

$\pi$  **parametro** della popolazione

Data la variabile casuale dicotoma  $X$  indicatrice dell'attributo e un campione di osservazioni  $X_1, X_2, \dots, X_n$  si consideri la

**Frequenza relativa** dell'attributo nel campione  $F = \frac{\sum_{i=1}^n X_i}{n}$ .

## RISULTATI

1.  $E(F) = \pi$

2.  $V(F) = \frac{\pi(1-\pi)}{n}$

3.  $\Pr(F = \frac{k}{n}) = \binom{n}{k} \pi^k (1-\pi)^{n-k}$

4.  $F \stackrel{\text{asint}}{\sim} N[\pi; \frac{\pi(1-\pi)}{n}]$ .

# Inferenza Statistica

---

Calcolo delle probabilità      =>    Il problema diretto  
Inferenza Statistica          =>    Il problema inverso

Procedure inferenziali      **stima puntuale dei parametri**  
   **stima per intervallo**  
   **verifica delle ipotesi**

Campione casuale: osservazioni indipendenti e identicamente distribuite

---

## Stima puntuale di una media $\mu$

---

Sia  $X \sim (\mu, \sigma^2)$  la distribuzione del carattere nella popolazione

**Stimatore  $\bar{X}$**   $\bar{X} \stackrel{\text{asint}}{\sim} N\left[\mu, \frac{\sigma^2}{n}\right].$

**Errore standard**  $\sigma_{\bar{X}} = \sqrt{\frac{\sigma^2}{n}}$

Nel caso in cui  $\sigma^2$  non sia nota, suo stimatore di è  $S_c^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$   
in quanto  $E(S_c^2) = \sigma^2$ .

**Stimatore dell'errore standard:**  $\hat{\sigma}_{\bar{X}} = \sqrt{\frac{S_c^2}{n}} = \frac{S_c}{\sqrt{n}}.$

ESEMPIO: Si vuole stimare il tempo medio in minuti dedicato dalle persone ad attività di volontariato nel corso di una settimana. Da un campione casuale di 75 persone si è ricavato  $\bar{X} = 90$  e  $S_c^2 = 80$ .

Stima del tempo medio = 90 minuti - Errore standard  $\hat{\sigma}_{\bar{X}} = \sqrt{\frac{80}{75}} = 1,03.$

---

## Stima puntuale di una frequenza relativa $\pi$

---

Sia  $X \sim (\pi; \pi(1 - \pi))$  la distribuzione del carattere nella popolazione

**Stimatore  $F$**   $F \stackrel{\text{asint}}{\sim} N\left[\pi, \frac{\pi(1 - \pi)}{n}\right]$ .

**Errore Standard**  $\sigma_F = \sqrt{\frac{\pi(1 - \pi)}{n}}$

**Stimatore dell'errore standard**  $\hat{\sigma}_F = \sqrt{\frac{F(1 - F)}{n}}$

**ESEMPIO:** In un campione casuale di 500 elettori di una regione italiana 95 hanno espresso l'intenzione di votare per il partito A alle prossime elezioni.

Stima puntuale della frequenza relativa degli elettori di A:  $F = 95/500 = 19,0\%$ .

$\hat{\sigma}_F = \sqrt{\frac{0,19(1 - 0,19)}{500}} = 0,0175 = 1,75\%$ . Errore standard

---

# Elementi di teoria della stima puntuale

---

Si chiama parametro della popolazione ogni indice statistico calcolato sulla popolazione.

Si chiama stimatore di un parametro incognito della popolazione ogni statistica campionaria utilizzata per assegnare un valore numerico ad esso.

Il parametro  $\theta$

Lo stimatore  $\hat{\theta}$

$$\hat{\theta} = f(X_1, X_2, \dots, X_n).$$

# Non distorsione o correttezza

$$E(\hat{\theta}) = \theta.$$

## ESEMPI di stimatori corretti

$\bar{X}$	perché	$E(\bar{X}) = \mu,$
$F$	perché	$E(F) = \pi,$
$S_c^2$	perché	$E(S_c^2) = \sigma^2$

## ESEMPIO di stimatore non corretto

$$S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \quad \text{perché} \quad E(S^2) = \frac{n-1}{n} \sigma^2.$$

Si chiama distorsione dello stimatore di un parametro la differenza tra il suo valore atteso ed il parametro.

$$B(\hat{\theta}) = E(\hat{\theta}) - \theta.$$

# Efficienza

-----  
L'efficienza di uno stimatore è inversamente proporzionale al suo errore standard

$$\hat{\theta}_1 \text{ è pi\`u efficiente di } \hat{\theta}_2 \text{ se} \\ \sqrt{V(\hat{\theta}_1)} \leq \sqrt{V(\hat{\theta}_2)} , \text{ per ogni } \theta.$$

ESEMPIO : parametro da stimare la media della popolazione.

Campione estratto da una popolazione normale

Errore standard della media

$$\sigma / \sqrt{n}$$

Errore standard della mediana

$$1,253\sigma / \sqrt{n}$$

---

# Consistenza

La proprietà della consistenza assicura che al crescere della dimensione del campione causale lo stimatore tende ad assumere valori sempre più vicini al parametro che intende stimare, sino a coincidere con esso nei campioni di ampiezza infinita.

Condizione sufficiente perché uno stimatore sia consistente è

$$\lim_{n \rightarrow \infty} B(\hat{\theta}_n) = 0$$

$$\lim_{n \rightarrow \infty} V(\hat{\theta}_n) = 0$$

ESEMPI di stimatori consistenti

$\bar{X}$  ed  $F$

# Riprendiamo esempio di pop. finita

Unità	$u_1$	$u_2$	$u_3$	$u_4$
Intensità	12	12	15	27

## Campionamento casuale semplice SENZA ripetizione

Camp.	$X_1$	$X_2$	$X_3$	$\bar{X}$	Camp.	$X_1$	$X_2$	$X_3$	$\bar{X}$
$u_1, u_2, u_3$	12	12	15	<b>13</b>	$u_3, u_1, u_2$	15	12	12	<b>13</b>
$u_1, u_2, u_4$	12	12	27	<b>17</b>	$u_3, u_1, u_4$	15	12	27	<b>18</b>
$u_1, u_3, u_2$	12	15	12	<b>13</b>	$u_3, u_2, u_1$	15	12	12	<b>13</b>
$u_1, u_3, u_4$	12	15	27	<b>18</b>	$u_3, u_2, u_4$	15	12	27	<b>18</b>
$u_1, u_4, u_2$	12	27	12	<b>17</b>	$u_3, u_4, u_1$	15	27	12	<b>18</b>
$u_1, u_4, u_3$	12	27	15	<b>18</b>	$u_3, u_4, u_2$	15	27	12	<b>18</b>
$u_2, u_1, u_3$	12	12	15	<b>13</b>	$u_4, u_1, u_2$	27	12	12	<b>17</b>
$u_2, u_1, u_4$	12	12	27	<b>17</b>	$u_4, u_1, u_3$	27	12	15	<b>18</b>
$u_2, u_3, u_1$	12	15	12	<b>13</b>	$u_4, u_2, u_1$	27	12	12	<b>17</b>
$u_2, u_3, u_4$	12	15	27	<b>18</b>	$u_4, u_2, u_3$	27	12	15	<b>18</b>
$u_2, u_4, u_1$	12	27	12	<b>17</b>	$u_4, u_3, u_1$	27	15	12	<b>18</b>
$u_2, u_4, u_3$	12	27	15	<b>18</b>	$u_4, u_3, u_2$	27	15	12	<b>18</b>

Distribuzione di probabilità di  $\bar{X}$

$\bar{X}$	13	17	18
$p(\bar{x})$	0,2500	0,2500	0,5000

$$E(\bar{X}) = 16,5 \quad V(\bar{X}) = \underline{4,25} \quad \sigma_{\bar{X}} = \underline{2,06}$$

Ciascuna unità ha probabilità  $\frac{3}{4}$  di far parte del campione

# Popolazioni Finite e loro caratteristiche

Si definisce popolazione finita ogni insieme di unità di uno stesso tipo e di numerosità limitata di cui interessa studiare una o più caratteristiche comuni.

- ☀ Una popolazione finita è **identificabile o etichettabile** =>  
Lista delle unità:  $U=\{1,2,\dots,N\}$
- ☀ Distribuzione semplice in forma unitaria o disaggregata

<i>Unità</i>	1	2	3	...	<i>i</i>	...	<i>N</i>
<i>Intensità di y</i>	$Y_1$	$Y_2$	$Y_3$	...	$Y_i$	...	$Y_N$
<i>Intensità di x</i>	$X_1$	$X_2$	$X_3$	...	$X_i$	...	$X_N$
<i>Intensità di z</i>	$Z_1$	$Z_2$	$Z_3$	...	$Z_i$	...	$Z_N$

- ☀ Parametri descrittivi: costanti che descrivono aspetti della distribuzione di uno o più caratteri statistici. Ad esempio:

$$\text{Il totale : } \tau = \sum_{i=1}^N X_i \quad \text{La media: } \mu = \frac{1}{N} \sum_{i=1}^N X_i$$

# La stima delle quantità di interesse

---

☀ Stimatore corretto del totale è:

$$\hat{\tau} = \sum_s w_i X_i$$

dove  $w_i = 1 / \pi_i$  è il *PESO DI RIPORTO ALL'UNIVERSO* (numero delle unità che la unità  $i$ -esima rappresenta nella popolazione) e  $\pi_i$  è la probabilità che una unità ha di far parte del campione.

☀ Stimatore corretto della media è  $\hat{\mu} = \hat{\tau} / N = \sum_s w_i X_i / N$

# Il campionamento casuale semplice SENZA ripetizione

---

**Un CCS e' caratterizzato dall'attribuire probabilità uguali ad ogni possibile campione di dimensione  $n$  da una popolazione di dimensione  $N$**

Come estrarre un campione casuale semplice di dimensione  $n$  :

- ☀ Lo schema dell'urna.
- ☀ Alternativa: se si dispone di una lista su supporto magnetico (cioè contenuta in un archivio elettronico), si assegna mediante il computer un numero casuale compreso tra 0 ed 1 ad ogni unità della popolazione e si includono nel campione le  $n$  unità a cui corrispondono i numeri casuali più piccoli (oppure i più grandi, etc.).
- ☀ Probabilità di inclusione della generica unità  $\Rightarrow \frac{n}{N}$
- ☀ Peso per il riporto all'universo della unità campionaria  $i$ -esima:  $\frac{N}{n} = w_i$

# Il campionamento sistematico

---

## ☀ Modalità di estrazione

1) Si determini l'intero  $k$  che risulta prendendo la parte intera del rapporto  $N/n$ .

2) Si estragga un numero casuale intero  $r$  compreso tra 1 e  $k$ .

3) Il campione sistematico è formato dalle unità che nella lista occupano le posizioni

$$r, r+k, r+2k, r+3k, \dots$$

dove  $k$  è il passo di campionamento ed  $r$  è il punto di partenza

ESEMPIO:  $N = 50$ ,  $n = 8$ ,  $N/n = 6,25 \Rightarrow k = 6$ .

Se  $r = 3 \Rightarrow s = \{3, 9, 15, 21, 27, 33, 39, 45\}$ .

# Il campionamento sistematico: caratteristiche e stima

---

- ☀ Applicabile anche quando non si dispone della lista ma le unità sono in sequenza (es. supermercati)
- ☀ Probabilità di inclusione è pari a  $1/k$ ; di conseguenza  $w_i = k$ . ( $w_i = N/n$  nel caso  $N$  sia un multiplo di  $n$ ).
- ☀ Efficiente nel caso di popolazioni ordinate casualmente o secondo qualche variabile ausiliaria.
- ☀ Può risultare inefficiente nel caso di popolazioni periodiche.

# Il campionamento stratificato 1.

---

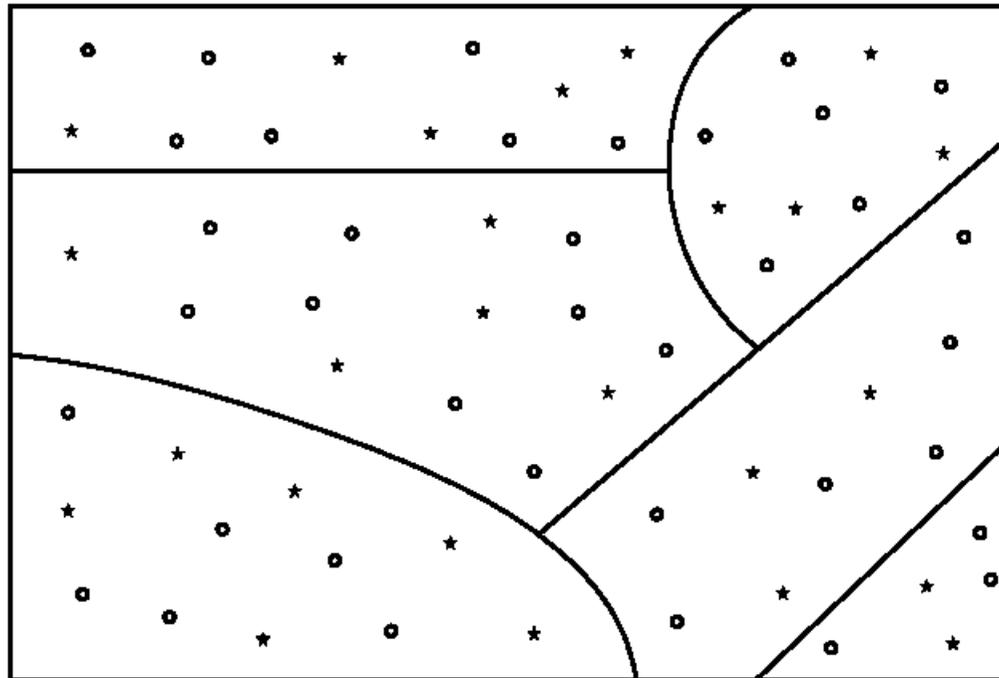
Esempio: E' noto che la propensione all'acquisto di un certo prodotto è maggiore tra le persone di sesso maschile che tra quelle di sesso femminile. Quindi, se in un campione casuale di dimensione  $n = 1000$  viene estratta una percentuale di maschi (ad es. il 60%) superiore a quella che esiste nella popolazione (circa il 50%) si tenderà a sovrastimare la propensione all'acquisto.

Il campionamento stratificato permette di estrarre esattamente 500 maschi e 500 femmine in modo che la percentuale dei maschi nel campione sia pari a quella della popolazione. In tal modo, si possono ottenere stime più precise

Il campionamento stratificato è una tecnica molto diffusa che consente di sfruttare l'informazione contenuta in una o più variabili ausiliarie conosciute per tutte le unità della popolazione (informazione presente nella lista di campionamento).

## Il campionamento stratificato 2.

- 
1. Si costruisce una partizione della popolazione (suddivisione in sottopopolazioni disgiunte e mutuamente esclusive), ciascun sottoinsieme della quale è chiamato strato.
  2. Da ogni strato si estrae un campione indipendente con un prefissato schema di campionamento
- Il campione complessivo si dice campione stratificato.



## Il campionamento stratificato 3.

---

- $h = 1, 2, \dots, H \rightarrow$  indice di strato,
  - $H$  numero degli strati [corrisponde al numero di modalità della variabile ausiliaria (ad es. 2 nel caso della variabile sesso) o al prodotto tra il numero di modalità delle variabili ausiliarie (ad es.  $2 \times 4$  nel caso delle variabili sesso e livello di istruzione – elementare/medio/superiore/laurea).
  - $N_h \rightarrow$  dimensione della popolazione nello strato  $h$
  - $N_1 + N_2 + \dots + N_H = N$
  - $n_h \rightarrow$  dimensione del campione nello strato  $h$
  - $n_1 + n_2 + \dots + n_H = n$  dimensione complessiva del campione
  - probabilità di inclusione  $n_h/N_h$
  - peso di riporto all'universo  $N_h/n_h$
- ☀ Incremento della precisione delle stime (tanto maggiore quanto maggiore è la omogeneità interna agli strati delle variabili rilevate).

# L'allocazione del campione negli strati

- PROPORZIONALE: L'allocazione del campione negli strati si dice proporzionale quando la quota di unità campionarie provenienti da uno strato è uguale al suo peso nella popolazione:

$$\frac{n_h}{n} = \frac{N_h}{N} \qquad n_h = \frac{N_h}{N} n$$

- EGUALE: L'allocazione del campione negli strati si dice eguale quando tutti i campioni di strato hanno la stessa dimensione:

$$n_h = \frac{n}{H}$$

- DI COMPROMESSO:  $n_h = \frac{1}{2} (n_h^{\text{eguale}} + n_h^{\text{prop}}) = \frac{1}{2} \left( \frac{n}{H} + \frac{N_h}{N} n \right)$

## L'allocazione del campione negli strati: un esempio

Si vuole condurre un'indagine di customer satisfaction in una Banca di piccole dimensioni. Si intende impiegare un piano campionamento stratificato in cui gli strati sono le filiali. La seguente tabella riporta la distribuzione di tutti i clienti per le cinque filiali. Avendo a disposizione un campione complessivo di 500 unità, si riporta anche la dimensione del campione per ciascuna filiale secondo i tre tipi di allocazione considerati.

Filiale	Clienti nella POP. $N_h$	Peso strato nella pop $N_h/N$	$n_h$ alloc. prop.	$n_h$ all. prop. (arrotondato)	peso strato nel campione	$n_h$ alloc. eguale	peso strato nel campione	$n_h$ alloc. compromesso	peso strato nel campione
AG. 1	7476	33.2%	165.765	166	33.2%	100	20.0%	133	26.6%
AG. 2	5470	24.3%	121.286	121	24.2%	100	20.0%	110.5	22.1%
AG. 3	5342	23.7%	118.448	118	23.6%	100	20.0%	109	21.8%
AG. 4	3188	14.1%	70.687	71	14.2%	100	20.0%	85.5	17.1%
AG. 5	1074	4.8%	23.814	24	4.8%	100	20.0%	62	12.4%
<b>Totale</b>	<b>22550</b>	<b>100.0%</b>		<b>500</b>		<b>500</b>		<b>500</b>	

# Un esempio di campionamento non probabilistico: il campionamento per quote

---

## ☀ Modalità di estrazione (per il campionamento di individui)

- 1) La popolazione è idealmente suddivisa in sottogruppi omogenei in base ad alcune caratteristiche (ad es. sesso, età, residenza) come nella stratificazione. Da informazioni ausiliarie (precedenti rilevazioni totali, dati di fonte amministrativa [www.demo.istat.it](http://www.demo.istat.it)) si ricava la distribuzione della popolazione secondo tali caratteristiche.
- 2) Si stabilisce il numero di interviste  $L$  che ogni intervistatore deve realizzare e il numero di esse da realizzare in ciascuno strato mediante la formula  $L \cdot N_h / N$  (**quota**)
- 3) Si demanda all'intervistatore la selezione delle unità campionarie in ciascuno strato.

# Esempio di campionamento per quote

-----  
Comune di Perugia – 1000 interviste telefoniche con quote per sesso e 4 classi di età. Popolazione complessiva con 15 anni e più al 1/1/2009 ([www.demo.istat.it](http://www.demo.istat.it)) : 143.500

## Distribuzione della popolazione

	Maschi	Femmine	Totale
15 -- 24	7.876	7.704	15.580
25 -- 44	24.403	24.715	49.118
45 -- 64	20.407	22.827	43.234
65 e oltre	14.863	20.705	35.568
Totale	67.549	75.951	143.500

## Distribuzione del campione

	Maschi	Femmine	Totale
15 – 24	55	54	109
25 – 44	170	172	342
45 – 64	142	159	301
65 e oltre	104	144	248
Totale	471	529	1.000

Supponendo di avere 4 intervistatori (L=250), queste saranno le loro quote

Intervistatore 1

	Maschi	Femmine	totale
15 -- 24	14	13	27
25 -- 44	43	43	86
45 -- 64	35	40	75
65 e oltre	26	36	62
Totale	118	132	250

Intervistatore 3

	Maschi	Femmine	totale
15 -- 24	14	13	27
25 -- 44	43	43	86
45 -- 64	35	40	75
65 e oltre	26	36	62
Totale	118	132	250

Intervistatore 2

	Maschi	Femmine	totale
15 -- 24	14	13	27
25 -- 44	43	43	86
45 -- 64	35	40	75
65 e oltre	26	36	62
Totale	118	132	250

Intervistatore 4

	Maschi	Femmine	totale
15 -- 24	13	15	28
25 -- 44	41	43	84
45 -- 64	37	39	76
65 e oltre	26	36	62
Totale	117	133	250

# Il campionamento per quote: pro e contro

---

## ☀ Confronto con il campionamento stratificato

### ☀ Vantaggi:

- ◆ Economico
- ◆ Veloce da realizzare
- ◆ Non servono liste

### ☀ Svantaggi:

- ◆ Distorsioni generate dai meccanismi di selezione dell'intervistatore o da meccanismi di autoselezione dei rispondenti
- ◆ Le proprietà degli stimatori non sono più garantite
- ◆ L'errore campionario non è misurabile tramite l'errore standard

# Errore non campionario

---

## ✦ Errori da lista

- ◆ elementi estranei,
- ◆ lista incompleta,
- ◆ nominativi ripetuti

## ✦ Non risposta parziale

- ◆ Imputazione

## ✦ Non risposta totale: mancato contatto o rifiuto

- ◆ Solleciti, tentativi ripetuti
  - ◆ Riserve
  - ◆ Sub campionamento dei non rispondenti
  - ◆ Ponderazione a posteriori
-

# Ponderazione

Attribuzione di un peso alle unità del campione per rendere la sua composizione più simile a quella della popolazione

- 1) Piano di campionamento (ad es. stratificazione con allocazione NON proporzionale)
- 2) Post-stratificazione: stratificazione fatta a posteriori su conoscenze a livello di popolazione ma non usate per il piano di campionamento
- 3) Aggiustamento della non-risposta

Esempio. Distribuzione % del campione per classi di età, informazione sulla popolazione, peso

	18 - 35	36 - 50	51 - 65	Oltre 65
% nel campione (a)	12	23	35	30
% nella pop. (b)	17	27	30	26
Peso = $b/a$	1,4	1,2	0,9	0,9