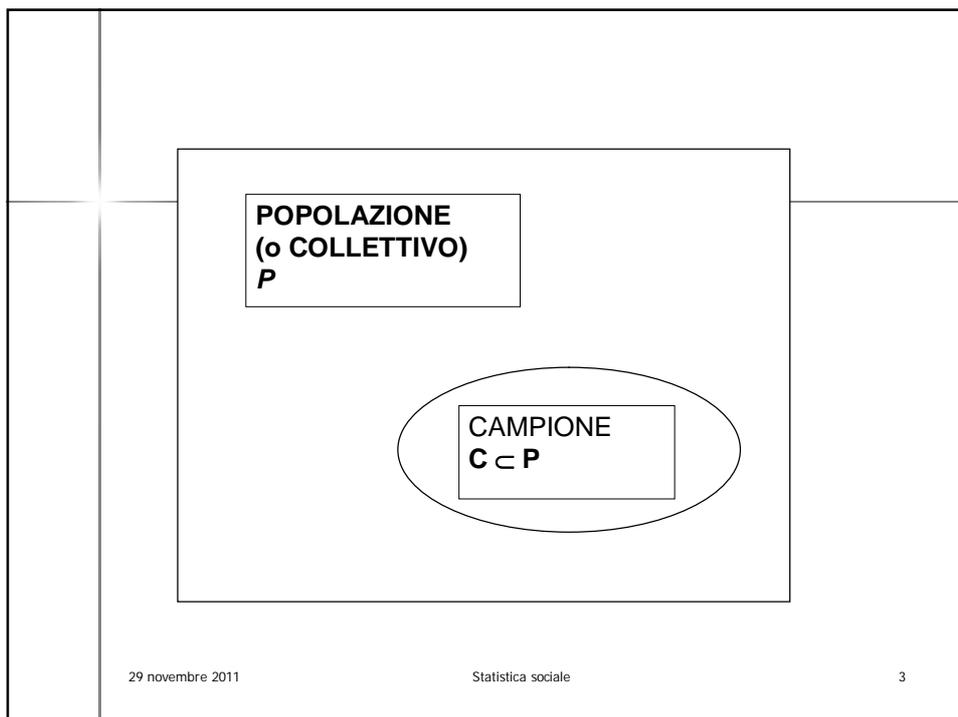


## Le indagini campionarie: i principali schemi di campionamento

## Le indagini campionarie

- Il reperimento dei DATI STATISTICI attraverso una RILEVAZIONE è un'operazione che ha dei COSTI, sia in termini di TEMPO IMPIEGATO che in termini ECONOMICI.
- Nei primi anni del '900, integrandosi con alcuni risultati fondamentali del calcolo delle probabilità, la statistica ha cominciato ad interessarsi ai CAMPIONI.
- Come abbiamo visto, un CAMPIONE è un SOTTOINSIEME del collettivo (popolazione) sul quale si sta studiando un certo fenomeno: di solito, è un sottoinsieme di numerosità MOLTO INFERIORE a quella del collettivo (si può avere, ad esempio, un campione la cui numerosità è pari anche all'UNO PER MILLE della numerosità del collettivo originario).
- Come abbiamo visto, l'**inferenza statistica** si basa sull'applicazione di misure statistiche in contesti campionari.



## Le indagini campionarie

- Le **TECNICHE CAMPIONARIE** (la cosiddetta **TEORIA DEI CAMPIONI**) permettono di **OTTIMIZZARE** i criteri di **ESTRAZIONE DEL CAMPIONE**, in modo tale da permetterci di **RICAVARE DAL CAMPIONE PRESSOCHÉ LE STESSA INFORMAZIONI CHE SI SAREBBERO RICAVATE DISPONENDO DELL'INTERO COLLETTIVO**.
- In questo modo si possono ottenere le stesse informazioni, con **COSTI, PERO', NETTAMENTE INFERIORI** e, cosa che spesso è molto importante, con **TEMPI ESTREMAMENTE PIU' RAPIDI** (si pensi al caso, in questo senso eclatante, delle **PROIEZIONI ELETTORALI**).
- Altro pregio dell'indagine campionaria: **IL DATO RISULTA MOLTO PIU' ACCURATO** (QUESTIONARI PIU' DETTAGLIATI, INTERVISTATORI PIU' PREPARATI, ECC...), proprio perché il minor tempo impiegato permette di dedicare maggiore cura ai particolari.

29 novembre 2011 Statistica sociale 4

## IN STATISTICA:

■ **RAPPRESENTATIVITÀ DEL CAMPIONE**

=

■ **ESTRAZIONE CASUALE DELLE UNITA' DEL CAMPIONE**

29 novembre 2011

Statistica sociale

5

## Campionamento a scelta ragionata

- Prima dell'affermazione dei criteri di campionamento casuale, il cosiddetto "campionamento ragionato" era il tipo di campionamento più diffuso.
- Si ha quando un campione NON VIENE ESTRATTO CASUALMENTE dalla popolazione, ma le unità del campione vengono SELEZIONATE, in modo ragionato, dalla popolazione, seguendo un certo criterio.

29 novembre 2011

Statistica sociale

6

## Campionamento a scelta ragionata

- Ad esempio, se è già noto a priori che una certa parte della popolazione "riproduce" in piccolo le caratteristiche della popolazione, il campionamento a scelta ragionata può dare buoni risultati. Un campionamento ragionato dà, invece, risultati scadenti se l'interesse ricade su **VARIABILI IL CUI COMPORTAMENTO NEL PASSAGGIO DALLA POPOLAZIONE AL CAMPIONE NON E' NOTO**.
- Il difetto principale di questo tipo di campionamento è, tuttavia, il fatto che **NON SI HANNO CRITERI PER STABILIRE LA BONTÀ DEI RISULTATI**.
- Il campionamento ragionato è spesso utilizzato nei **CAMPIONAMENTI A DUE (O PIU') FASI**.

29 novembre 2011

Statistica sociale

7

## Campionamento casuale semplice

- Le unità del campione vengono estratte dalla popolazione in modo casuale: condizione **FONDAMENTALE** è che ciascuna unità del collettivo abbia la **STESSA PROBABILITÀ** di essere estratta; che, cioè, non vi siano unità "favorite" rispetto ad altre (**PRINCIPIO DI EQUIPROBABILITÀ DELL'ESTRAZIONE**).

29 novembre 2011

Statistica sociale

8

## Campionamento casuale semplice

- Le fasi di un campionamento casuale semplice sono le seguenti:
- **A. INDIVIDUAZIONE DELLA BASE DEL CAMPIONAMENTO:** vengono elencate tutte le  $N$  unità che costituiscono la popolazione (la c.d. **popolazione di selezione**: vedremo più avanti il significato di questo termine).
- **IMPORTANTE!** Quasi mai le UNITA' DI RILEVAZIONE coincidono con le UNITA' STATISTICHE.
- Ad esempio, in quasi tutte le indagini ISTAT l'unità di rilevazione (e, quindi, anche **di campionamento**) è la FAMIGLIA, mentre l'unità statistica di interesse è la SINGOLA PERSONA. La lista delle famiglie è reperibile presso le **anagrafi comunali**.

29 novembre 2011

Statistica sociale

9

## Campionamento casuale semplice

- **B. DETERMINAZIONE DELLA NUMEROSITÀ DEL CAMPIONE.**
- La dimensione del campione,  $n$ , va decisa sia sulla base di esigenze di economicità dell'indagine, sia sulla base di valutazione probabilistiche sull'errore campionario.
- Al decrescere di  $n$ , infatti, decresce anche la PRECISIONE DELLE STIME che si possono ricavare dal campione.

29 novembre 2011

Statistica sociale

10

## La base del campionamento

- Per effettuare un campionamento casuale semplice è necessario disporre di una **lista** delle **unità** che compongono il collettivo che si intende osservare (non necessariamente di nominativi, ma anche di codici, come codici fiscali, numeri di partita Iva, codici interni, ecc.).
- Gli "identificativi" (ID) della lista devono corrispondere uno-a-uno alle unità della popolazione.
- La lista identifica la **popolazione di selezione**. Sono frequenti i casi in cui non esiste perfetta coincidenza tra popolazione di selezione e **popolazione obiettivo** dell'indagine (perché, per qualche motivo, la lista è incompleta).

29 novembre 2011

Statistica sociale

11

## La base del campionamento

- Oltre a **popolazione obiettivo** e **popolazione di selezione**, si deve parlare anche di **popolazione di indagine**.
- Una volta selezionato il campione, spesso non sarà possibile osservare **tutte** le unità selezionate, o per l'impossibilità di contattarle, o per un loro rifiuto, esplicitamente dichiarato, di partecipazione all'indagine (non risposta).
- **La popolazione obiettivo differisce dalla popolazione di selezione a causa dell'incompletezza della lista.**
- **La popolazione di selezione, a sua volta, differisce da quella di indagine per due possibili cause:**
  - **Mancati contatti** (es. numeri di telefono sbagliati);
  - **Non risposte** o, più correttamente, **rifiuti di intervista**.

29 novembre 2011

Statistica sociale

12

■ L'entità di un campionamento si valuta sulla base della **FRAZIONE CAMPIONARIA** (O FRAZIONE DI CAMPIONAMENTO):

$$FC = \frac{n}{N} \cdot 100$$

**C. ESTRAZIONE DELLE UNITA' DEL CAMPIONE** con criteri computerizzati (con opportuni algoritmi), o ricorrendo alla tavola dei NUMERI CASUALI.

5965 2913 5612 6361 7075 5490 9626 4307 0840 7945  
 5801 9383 6173 8358 9236 5543 5811 5520 5814 7864  
 1223 5344 3649 6397 1678 4400 7715 7614 1209 7729  
 0220 2108 0784 8837 3916 0282 4490 3442 6471 6593  
 4131 9772 7594 8863 0874 1864 8117 6411 7012 2682  
 3074 5746 2723 5681 0989 8015 0818 5380 9981 3758  
 2939 6585 6658 7756 7916 9770 2868 2128 2665 2386  
 6003 5982 8829 2833 8160 2101 3365 4121 4522 8216  
 2039 2993 4362 6363 2914 4955 6364 5237 6456 5561  
 0176 2425 2968 3834 6077 4302 3499 9938 7231 2136  
 2161 1365 2764 7836 1584 2421 4247 2930 0783 9989  
 0407 1760 7048 1929 9034 0242 0753 4851 9465 0791  
 0055 7981 7760 2215 3323 4727 8884 8066 7965 3939  
 0726 2104 9164 6275 5464 4073 1715 3215 7883 8087  
 2475 9583 8713 1445 2702 4952 4307 5796 2913 0589  
 0686 1266 4341 9760 9608 5773 7394 9333 4752 8395  
 4223 4033 3734 8221 2055 5131 0065 1626 7742 5806  
 9596 5241 3230 3269 4836 9776 2894 5740 1557 2515  
 1581 5007 6906 8933 9981 3175 4979 4525 5334 6038  
 6558 6350 1273 6164 7125 1481 3084 1517 4748 0956

## Campionamento sistematico

- Se la base del campionamento NON ESISTE o NON È DISPONIBILE, non è possibile applicare il campionamento casuale semplice; l'unica forma di campionamento CASUALE che si può applicare in questi casi è il CAMPIONAMENTO SISTEMATICO.
- Il campionamento sistematico consiste, dopo aver effettuato una numerazione delle unità di rilevazione, nella scelta casuale della PRIMA unità del campione, mentre tutte le successive unità del campione vengono scelte SALTANDO un NUMERO COSTANTE di unità non scelte. Solitamente, una volta deciso quale deve essere la frazione campionaria, il numero costante di unità da saltare è dato dal **reciproco della frazione campionaria**:

$$K = \frac{1}{FC} = \frac{N}{n}$$

- La quantità K è detta *PASSO DEL CAMPIONAMENTO*.

## Campionamento sistematico

- Ad esempio, si vuole svolgere un'indagine campionaria sui turisti stranieri arrivati in Italia durante un certo periodo. Ovviamente, non può esistere un elenco completo della base campionaria, e si dovrà quindi procedere ad un **CAMPIONAMENTO SISTEMATICO**.
- Così, se si chiede una **FRAZIONE CAMPIONARIA** pari al **5%**, il passo di campionamento sarà pari a  $K = 1 / 0,05 = 20$ , cioè **UNO STRANIERO OGNI VENTI** andrà intervistato al passaggio della frontiera.

29 novembre 2011

Statistica sociale

15

## Campionamento sistematico

Da un elenco numerato delle unità che formano la popolazione, dopo l'estrazione casuale della **prima unità** (utilizzando una tavola di numeri casuali) si estraggono le unità a distanza costante (il **PASSO DI CAMPIONAMENTO**).

Es.: supponiamo di disporre di una lista di 100 unità, dalla quale vogliamo estrarre un campione di **10 unità**.

001	002	003	004	005	006	007	008	009	010
011	012	013	014	015	016	017	018	019	020
021	022	023	024	025	026	027	028	029	030
031	032	033	034	035	036	037	038	039	040
041	042	043	044	045	046	047	048	049	050
051	052	053	054	055	056	057	058	059	060
061	062	063	064	065	066	067	068	069	070
071	072	073	074	075	076	077	078	079	081
081	082	083	084	085	086	087	088	089	090
091	092	093	094	095	096	097	098	099	100

29 novembre 2011

Statistica sociale

16

## Campionamento sistematico

- La frazione campionaria richiesta è il 10%.
- Pertanto, il passo di campionamento richiesto è pari a
  - $K = 1 / 0,10 = 10$ .
- A questo punto, procedo nel modo seguente: scelgo casualmente una unità dalla prima riga, e poi prendo sistematicamente le unità distanti 10 numeri l'una dall'altra.
- Supponendo di aver scelto quale prima unità quella contraddistinta dal numero "005", il campione estratto sarà composto dalle unità corrispondenti ai numeri:
  - 005, 015, 025, 035, 045, 055, 065, 075, 085, 095.

29 novembre 2011

Statistica sociale

17

## Campionamento sistematico

- Il punto debole di questo tipo di campionamento sta nel fatto che può verificarsi un errore sistematico dovuto a certi fenomeni "concomitanti" con il passo di campionamento, i cui effetti sono, oltretutto, quasi sempre visibili SOLO A RILEVAZIONE AVVENUTA;
- Se, ad esempio, vogliamo intervistare con campionamento sistematico una comitiva di una GITA SCOLASTICA (composta da studenti e docenti), se il passo di campionamento è VENTI corriamo il rischio di intervistare solo (o prevalentemente) gli insegnanti.

29 novembre 2011

Statistica sociale

18

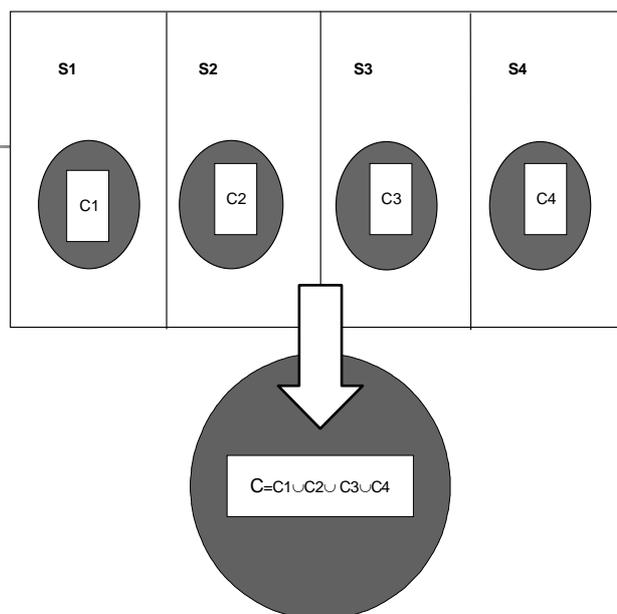
## Campionamento stratificato

- Come abbiamo visto nelle lezioni precedenti, un modo per aumentare la precisione delle stime ottenute da un campione è aumentare la dimensione del campione.
- C'è però anche un'altra possibilità:
- essa consiste nel formare, a priori, degli strati della popolazione di riferimento, in base ad un certo carattere che si ritiene fortemente correlato alla variabile oggetto di studio.
- Successivamente, si estrae DA CIASCUNO STRATO, con CAMPIONAMENTO CASUALE SEMPLICE, un "piccolo campione".
- L'**unione** di tutti questi "sottocampioni" andrà a formare il campione finale.

29 novembre 2011

Statistica sociale

19



29 novembre 2011

Statistica sociale

20

## Campionamento stratificato

- La **STRATIFICAZIONE** migliora le stime perché RENDE IL CAMPIONE PIU' SIMILE ALLA POPOLAZIONE, ALMENO PER ALCUNE CARATTERISTICHE (IL CAMPIONE DIVIENE, IN QUALCHE MODO, UNA "MINIATURA" DEL COLLETTIVO).
- Con la stratificazione, è meno probabile che vengano estratte unità "ESTREME" rispetto alla variabile oggetto di studio.

29 novembre 2011

Statistica sociale

21

## Campionamento stratificato

- Una tipica stratificazione che si può applicare a una popolazione è quella per classi di età e sesso.
- Ad esempio, supponiamo che la nostra popolazione target sia la popolazione della provincia di Ferrara in età compresa tra 18 e 79 anni (popolazione al 1.1.2010).

29 novembre 2011

Statistica sociale

22

**Criterio dell'allocazione proporzionale: popolazione provincia di Ferrara 18-79 per classi di età e sesso.**

	Maschi	Femmine	<b>Totale</b>
18-34	29.815	29.255	<b>59.070</b>
35-49	43.591	42.817	<b>86.408</b>
50-64	36.514	39.373	<b>75.887</b>
65-79	28.787	35.543	<b>64.330</b>
<b>Totale</b>	<b>138.707</b>	<b>146.988</b>	<b>285.695</b>

29 novembre 2011

Statistica sociale

23

**Criterio dell'allocazione proporzionale**

- Supponiamo di voler estrarre da questa popolazione, con un campionamento stratificato, con allocazione proporzionale, un campione di dimensione
- $n = 2000$ .

29 novembre 2011

Statistica sociale

24

## Le proporzioni richieste dal campione per i singoli strati

	Maschi	Femmine
<b>18-34</b>	10,4%	10,2%
<b>35-49</b>	15,3%	15,0%
<b>50-64</b>	12,8%	13,8%

29 novembre 2011

Statistica sociale

25

## La composizione del campione

- Applicando le proporzioni appena viste al campione richiesto, con n=2000, si ottiene:

	Maschi	Femmine	Totale
18-34	208,7	204,8	414
35-49	305,2	299,7	605
50-64	255,6	275,6	531
65-79	201,5	248,8	450
Totale	971	1029	2000

29 novembre 2011

Statistica sociale

26

## Campionamento a grappoli (*cluster sampling*)

- Si tratta di un campionamento casuale semplice, nel quale però le "unità di campionamento" sono costituite da GRUPPI di unità elementari (GRAPPOLI).
- Vengono estratti, casualmente, i grappoli; successivamente vengono considerate, come facenti parte del CAMPIONE FINALE, tutte le unità comprese in ciascun grappolo.
- La scelta di questo tipo di campionamento avviene molto spesso per ragioni di **economicità**, perché esso consente di osservare le unità campionarie **CONCENTRATE IN POCHE AREE**, mentre con altre forme di campionamento le unità sarebbero risultate **SPARSE SUL TERRITORIO**.

29 novembre 2011

Statistica sociale

27

## Campionamento a grappoli (*cluster sampling*)

- Difetto principale di questo tipo di campionamento è che, **CONTRARIAMENTE A QUANTO AVVIENE CON IL CAMPIONAMENTO STRATIFICATO**, la precisione delle stime **DIMINUISCE** rispetto al C.C.S.; tale precisione diminuisce in modo tanto più forte quanto più sono **SIMILI TRA LORO** le unità contenute in ciascun grappolo.
- **ESEMPIO**: se vogliamo effettuare un'indagine sui clienti degli alberghi in un certo comprensorio turistico, ad un dato istante, naturalmente non sarà possibile conoscere con precisione la base del campionamento. Sarà quindi necessario effettuare un campionamento **A GRAPPOLI**, estraendo direttamente un campione di **ALBERGHI**, e intervistando poi **TUTTI** i clienti degli alberghi estratti.

29 novembre 2011

Statistica sociale

28

## **Campionamento a due (o più) stadi (campionamento complesso)**

- Il campione si forma a seguito di più estrazioni, che avvengono in DUE (O PIÙ) FASI, quasi sempre con UNITA' DI CAMPIONAMENTO DIVERSE; inoltre, i disegni di campionamento sono in generale DIVERSI nelle diverse fasi.
- Le indagini campionarie dell'ISTAT sono tutte effettuate con campionamenti complessi.