

Metodi statistici per le ricerche di mercato

Prof.ssa Isabella Mingo
A.A. 2016-2017



SAPIENZA
UNIVERSITÀ DI ROMA

Facoltà di Scienze Politiche, Sociologia, Comunicazione

Corso di laurea Magistrale in «Organizzazione e marketing per la comunicazione d'impresa»

Glossario:

Il campione

Parte della popolazione statistica di riferimento, estratto mediante determinate tecniche

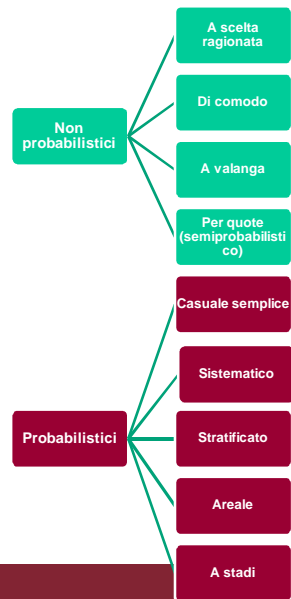
1. **Tecniche di campionamento probabilistici (probability sample):**
 - Consentono di estrarre le unità di campionamento in modo casuale
 - Ogni unità della popolazione è estratta con una probabilità nota

I campioni estratti con questi metodi consentono di effettuare "stime dei valori della popolazione" puntuali o ad intervalli mediante i metodi dell'inferenza statistica che consentono di calcolare il grado di attendibilità dei risultati dell'indagine nei confronti della popolazione da cui è stato selezionato il campione.
2. **Metodi di campionamento non probabilistici (haphazard sample):**
 - Effettuano la selezione delle unità di campionamento secondo criteri non casuali, ma a scelta ragionata, a casaccio, in base alla disponibilità di tali unità.

I campioni estratti con questi metodi consentono di ottenere indicazioni di massima sui valori della popolazione

Mingo 2016-2017

Tipi di campione



a.a. 2016-2017

Concetti preliminari Campioni non probabilistici

- **A scelta ragionata:** le unità vengono scelte in base alle loro caratteristiche sulla base delle conoscenze e del giudizio del ricercatore
 - Opinion leaders
 - Unità tipiche: elementi che rappresentano l'estremizzazione di alcune situazioni (massaia, donna in carriera, operaio modello..)
 - Unità estratte da *aree barometro* che rappresentano il comportamento medio della popolazione
- **A valanga:** si seleziona un soggetto e si individuano gli altri grazie alle indicazioni del primo. È utile nel caso di indagini su gruppi chiusi negli studi di comunità, nelle analisi di rete.
- **Di comodo** (ad adesione volontaria): le unità vengono scelte dagli intervistatori e generalmente esse si offrono volontariamente.
- **Per quote:** Si suddivide la popolazione in strati sulla base di un criterio e si decide quante unità rilevare per ciascuno strato. La selezione dell'unità non è però casuale, ma affidata all'intervistatore.
 - Viene definito anche «quasi probabilistico»

a.a. 2016-2017

Esercizio : campione per quote proporzionale

Sapendo che la distribuzione per sesso e ripartizione della popolazione di riferimento è quella riportata nella tabella seguente, fissata la dimensione campionaria $n=350$, per ottenere un campione rappresentativo di queste due caratteristiche della popolazione, quanti soggetti per ogni quota occorre intervistare?

Tavola di contingenza SESSO * RIPARTIZIONE

Conteggio

		RIPARTIZIONE					Totale
		1 Italia Nord-Occidentale	2 Italia Nord-Orientale	3 Italia Centrale	4 Italia Meridionale	5 Italia Insulare	
SESSO	1 Maschio	4648	4776	4040	6570	2438	22472
	2 Femmina	4997	5020	4347	6976	2652	23992
Totale		9645	9796	8387	13546	5090	46464

$$w_h = \frac{N_h}{N}$$

f_{ij} sul totale

	Italia Nord-Occidentale	Italia Nord-Orientale	Italia Centrale	Italia Meridionale	Italia Insulare	Totale
Maschio	0,1008	0,1028	0,0869	0,1414	0,0525	0,4836
Femmina	0,1075	0,1080	0,0936	0,1501	0,0571	0,5164
Totale	0,2076	0,2108	0,1805	0,2915	0,1095	1,0000

$$2438/46464=0,0525$$

Dimensione delle quote

		RIPARTIZIONE					Totale
		1 Italia Nord-Occidentale	2 Italia Nord-Orientale	3 Italia Centrale	4 Italia Meridionale	5 Italia Insulare	
SESSO	1 Maschio	35	36	30	49	18	169
	2 Femmina	38	38	33	53	20	181
Totale		73	74	63	102	38	350

$$n_h = w_h \times n$$

$$0,0525 \times 350 = 18$$

Esercizio : campione per quote proporzionale

Sapendo che la distribuzione per sesso e titolo di studio della popolazione di riferimento è quella riportata nella tabella seguente, fissata la dimensione campionaria $n=200$, per ottenere un campione rappresentativo di queste due caratteristiche della popolazione, quanti soggetti per ogni quota occorre intervistare?

Tavola di contingenza sesso * titolo di studio (4 cat)

Conteggio

		titolo di studio (4 cat)				Totale
		Nessun titolo -elementare	Media inferiore	Media superiore	Laurea-dottorato	
sesso	maschio	3860	6388	6781	2071	19100
	femmina	4593	5759	6682	2626	19660
Totale		8453	12147	13463	4697	38760

Campionamento probabilistico

- ✗ Un campione casuale affida al caso la scelta delle unità statistiche da campionare
- ✗ Ogni campione di pari numerosità ha la stessa probabilità di essere estratto

Il campionamento casuale:

- ✓ garantisce le migliori proprietà statistiche degli stimatori
- ✓ garantisce la possibilità di stimare la bontà dei risultati ottenuti
- ✓ a parità di numerosità campionaria (e di costo) fornisce risultati più attendibili
- ✓ semplifica la costruzioni di modelli statistici, che spesso assumono che i dati siano stati raccolti in modo casuale
- ✓ gli intervistatori non hanno discrezionalità su quali unità statistiche intervistare
- ✓ la procedura di selezione del campione è ben definita e prevede l'introduzione e l'uso pianificato del caso

Pagina 7

Alcuni tipi di Campioni probabilistici

- **Casuale semplice:** tutte le N unità della popolazione hanno la stessa probabilità di essere incluse nel campione.

Le n unità del campione vengono estratte mediante sorteggio che può essere effettuato:

- Con reinserimento : le unità estratte vengono reinserite e dunque possono essere estratte nuovamente
- Senza reinserimento : le unità estratte non vengono reinserite, dunque possono essere estratte solo una volta.

Nella pratica si fa ricorso a [tavole di numeri casuali](#) o a opportuni software.

- Si deve disporre della lista completa delle unità della popolazione.

- **Sistematico:** le unità vengono selezionate da una lista ordinata casualmente, una ogni dato intervallo, stabilito con la relazione:

- $(k=N/n)$.

Si inizia estraendo a sorte un numero compreso tra 1 e k e considerando l'unità statistica corrispondente al numero estratto.

- **Stratificato:** vengono estratti tanti campioni casuali semplici, ciascuno nell'ambito di una sottopopolazione o [strato](#) ritenuto omogeneo secondo un criterio dato. I campioni così ottenuti vengono poi uniti in un campione globale.



Campioni probabilistici (2/2)

A grappoli: Vengono estratti casualmente gruppi di unità (famiglie, classi, unità territoriali) .

Areale: la procedura consiste nel suddividere una macro-area territoriale in su-aree omogenee da cui vengono estratte le unità

A stadi: dopo aver estratto casualmente unità dette di primo ordine, non si rilevano tutte le unità elementari in esse contenute, ma solo un loro campione (di secondo ordine)

Glossario: Tavola di numeri casuali

- E' composta da una serie di cifre generate ed elencate casualmente

ESEMPIO DI TAVOLA DI NUMERI CASUALI

8966	4502	4212	3823	9379	2769	4171	4532	4396	8245
7696	4571	2687	1746	1025	8219	3327	8770	4519	1141
2876	0813	3178	2656	0690	1574	5372	8523	7683	9302
5783	6711	4842	3206	7816	5713	3766	1089	1043	0481
3292	4223	2418	1250	9077	9677	2164	4339	8770	4653
5463	8195	5500	6858	4448	8753	4618	1394	4601	3757
2903	2172	4041	9143	8353	4025	5604	3510	5117	5264
7494	7119	2588	2651	3733	2718	1234	1368	5439	5009
1434	0441	2694	3474	6611	4083	9086	8908	9716	3892
9295	5133	8685	8077	6872	9474	9440	1248	8846	7883
5216	1653	2632	4104	3317	0636	2287	5567	3781	7634
7210	7101	6499	7741	0520	7367	0256	8041	0498	9314
3814	1620	0987	0790	7880	7792	4313	7298	1222	0109
3380	0481	3502	0650	9957	7314	7622	1329	6934	1790
1395	7128	9993	5948	0286	4253	8044	7892	5201	2015
8872	7536	5521	5924	8138	4187	1175	6661	2001	5994
0430	8098	6550	4087	4176	6572	3950	3516	4675	2575
3597	5707	3074	1896	8046	2543	2263	2986	9237	6778

Per usare la tavola:

- si assegna ad ogni unità della popolazione un codice numerico.
- Si stabilisce un criterio di lettura dei numeri
- Si legge la tavola a partire da una qualsiasi cella scelta a caso
- si selezionano le unità il cui codice coincide con il numero casuale.



Campione sistematico : esempio

- Campione sistematico di 350 unità estratto tra tutti gli utenti che usufruiscono del servizio di assistenza clienti dal 1 gennaio al 31 marzo.
 - Supponiamo di sapere in base all'andamento delle presenze osservate in periodi precedenti che gli utenti saranno circa 1750 e che la dimensione del campione è di 350 .

L'intervallo di campionamento sarà:

- $(k=N/n) \quad k=1750/350=5$

Si estrarrà un numero compreso tra 1 e 5: l'estratto sarà il primo elemento del campione. Si procederà poi selezionando un utente ogni cinque.



Campione a grappoli: esempio

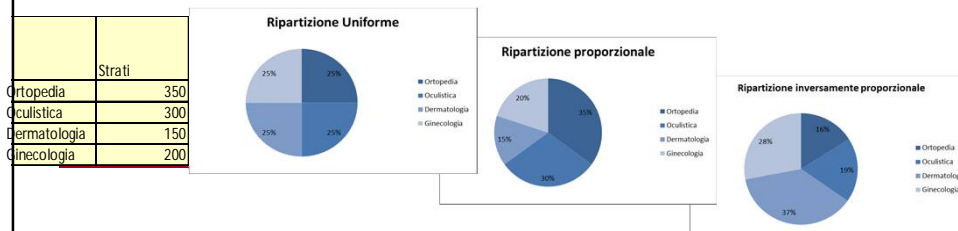
- Selezione di famiglie del comune, che costituisce il bacino di utenza del punto vendita, oggetto di indagine, e intervista a tutti i componenti di età maggiore di 15.
 - La famiglia costituisce il grappolo (unità primaria), i componenti le unità secondarie.



Campione stratificato: esempio

- Campione stratificato di utenti di un Centro di servizi sanitari in relazione al tipo di visita prenotata.
- Il campione potrà:
 - Ripartire le unità campionarie in modo uniforme in tutti gli strati
 - Rispettare le proporzioni della popolazione (stratificato proporzionale)
 - Non rispettare tali proporzioni, sovra-rappresentando gli strati più piccoli (ripartizione inversamente proporzionale).

Questo campionamento richiederà una ponderazione in sede di elaborazione per correggere il sovra o sottodimensionamento dei singoli strati.



Campione stratificato: criteri

- Ripartizione uniforme:
 - A ciascuno strato viene assegnato lo stesso numero di elementi campionari

$$n_1 = n_2 = n_3 = \dots n_h$$

$$n_h = n/H$$
 con H= numero di strati
- Ripartizione proporzionale
 - A ciascuno strato viene assegnato un numero di elementi campionari proporzionali alla numerosità degli strati della popolazione. La frazione di campionamento sarà dunque proporzionale a ciascuno strato. Il subcampione dello strato h avrà dimensione

$$w_h = \frac{N_h}{N} \quad n_h = w_h \times n$$
- Ripartizione non proporzionale
 - Sovra-rappresentare gli strati più piccoli (ripartizione inversamente proporzionale).
$$w_h = \frac{N}{N_h} \quad w = \sum_{h=1}^H w_h \quad n_h = \frac{w_h}{w} \times n$$

Campionamento stratificato esercizio

Si vuole effettuare un sondaggio tra i negozi di una determinata zona. Si ritiene che la dimensione del negozio possa influire sulle risposte. La popolazione di riferimento è riportata nella tabella seguente.

Posto che il campione da estrarre è di 500 unità, individuare la numerosità di ciascuno strato applicando rispettivamente il criterio di ripartizione uniforme, proporzionale e inversamente proporzionale.

Negozi: popolazione (N=3000)		Campioni (n=500)				
Dimensioni	N _h	Ripartizione uniforme n _h =N/H	Ripartizione proporzionale w _h =N _h /N n _h =w _h *n		Ripartizione inversamente proporzionale w _h =N/N _h n _h =w _h /W*n	
Grande	500	166,6	0,17	85	6	273
Medie	1000	166,6	0,33	165	3	136
Piccole	1500	166,6	0,50	250	2	91
Totale	3000	500	1,00	500	11	500

$$500/3=166$$

$$1000/3000=0,33$$

$$0,33*500=165$$

$$3000/1000=3$$

$$3/11*500=136$$

2016-2017

Campionamento stratificato esercizio

- Si vuole rilevare la soddisfazione dei clienti riguardo i servizi di assistenza di alcuni punti vendita. Si ritiene che la tipologia di prodotto possa influire sulle risposte. La popolazione di riferimento è riportata nella tabella seguente.

Posto che il campione da estrarre è di 300 unità, individuare la numerosità di ciascuno strato applicando rispettivamente il criterio di ripartizione uniforme, proporzionale e inversamente proporzionale.

Tipologia prodotto	N _h
Grandi elettrodomestici	200
Piccoli elettrodomestici	900
Telefonia	532
	1632

2016-2017 Mingo

Campione stratificato ottimale

- Si ricorre a questo tipo di campione quando gli strati presentano una diversa variabilità (s_h): quindi alcuni sono omogenei al loro interno, altri più eterogenei.
- In tali casi, non si rispetta la proporzionalità degli strati ma si sovra-rappresentano quelli con gli strati più eterogenei, che presentano una maggiore variabilità.
- Il campione così costruito si definisce «ottimale»

$$n_h = n \frac{N_h s_h}{\sum_{h=1}^H N_h s_h}$$

Campione stratificato ottimale: esercizio

- Un istituto di credito decide di effettuare un'indagine campionaria sulla sua clientela. Disponendo di informazioni tratte dai suoi database e riportate nella tabella seguente, tenendo conto della eterogeneità dei depositi per categoria, si definisca un piano di campionamento stratificato di $n=300$.

Tipo di cliente	Strati (N_h)	Depositi in euro scarto quadratico medio (s_h)	$N_h s_h$	n_h
Titolari di impresa	150	50,7	7605	123
Liberi professionisti	300	23,2	6960	113
Impiegati	350	5,4	1890	31
Operai	200	10,5	2100	34
TOTALE	1000		18555	300

$$n_h = n \frac{N_h s_h}{\sum_{h=1}^H N_h s_h}$$

Campione stratificato ottimale: esercizio

- Una catena di supermercati decide di effettuare un'indagine campionaria sulla sua clientela. Disponendo di informazioni tratte dai suoi database e riportate nella tabella seguente, tenendo conto della eterogeneità delle famiglie in base al numero di componenti per zona di vendita, si definisca un piano di campionamento stratificato di $n=500$.

Zona di vendita	Numerosità delle famiglie clienti (N_h)	Numero dei componenti scarto quadratico medio (s_h)
Zona Nord Ovest	3400	0,3
Zona Nord-Est	1500	0,8
Zona Centro	2800	0,5
Zona Sud	1300	0,4
TOTALE	9000	

$$n_h = n \frac{N_h s_h}{\sum_{h=1}^H N_h s_h}$$
