

Metodi Quantitativi per Economia, Finanza e Management

Lezione n°3

Campionamento, Dati di tipo qualitativo e dati di tipo quantitativo

Quantitative Market Research

Execution

Riepilogando, l'esecuzione di una Ricerca di Mercato può essere sintetizzata nei seguenti punti:

- *business goals*
- *target population*
- *channel*
- *questionnaire set-up*
- *sampling*
- *data matrix*
- *basic treatment of collected data*
- *data analysis*
- *business presentation of the results*

Il campionamento

- Elementi introduttivi
- Metodi di campionamento
- La dimensione del campione

Il campionamento

Elementi introduttivi

La ricerca quantitativa si basa:

- sulla rilevazione censuaria delle unità della popolazione target;
- su un sottoinsieme delle unità (*campione*) della popolazione target.

Il fine principale di un'indagine campionaria è *generalizzare i risultati ottenuti sul campione all'intera popolazione di riferimento.*

Il problema di tale approccio consiste nell'approssimazione con cui è possibile descrivere la popolazione attraverso il campione.

Il campionamento

Elementi introduttivi

- Oggetto di studio di una indagine campionaria è una popolazione finita $\Omega = \{1, 2, \dots, N\}$, le cui unità presentano tutte un medesimo carattere U
- y_1, y_2, \dots, y_n indicheranno i valori che la caratteristica U assume rispettivamente nelle unità $1, 2, \dots, N$ della popolazione.
- Si definisce *campione* di dimensione n della popolazione Ω un suo qualsiasi sottoinsieme $c = \{i_1, i_2, \dots, i_n\}$, contenente n unità
- Per ottenere il campione da analizzare, si possono estrarre dalla popolazione n unità secondo due modalità principali:
 - *Estrazione con ripetizione* ogni volta viene reintrodotta l'unità già estratta dalla popolazione
 - *Estrazione senza ripetizione* tutte le unità già selezionate non possono più essere estratte

Il campionamento

Metodi di campionamento

I metodi di campionamento si dividono in :

- *Probabilistici* ogni unità della popolazione ha una probabilità nota e diversa da zero di essere selezionata, e quindi di entrare nel campione. Richiedono un insieme di regole note a priori, utilizzabili matematicamente, per formare il campione (*piano di campionamento*)
- *Non probabilistici* la selezione delle unità avviene in base a criteri soggettivi (presenza di particolari esigenze conoscitive).

Il campionamento

Metodi di campionamento - Probabilistici

Campionamento casuale semplice

A ogni estrazione ogni elemento della popolazione ha la stessa probabilità di essere selezionato, attraverso un meccanismo che garantisce la casualità delle estrazioni. La probab. di estrazione di un elemento è:

- Camp. *senza* ripetizione:

$$P = 1 / N * (N-1) * (N-2) * \dots * (N-i+1)$$

- Camp. *Con* ripetizione:

$$P = 1 / N$$

Il campionamento

Metodi di campionamento - Probabilistici

Campionamento casuale semplice

Stratificato



Se si hanno informazioni suppletive sulla popolazione di riferimento, è possibile dividerla in strati, al cui interno le unità sono omogenee secondo un determinato criterio. Da ogni strato viene estratto, in modo indipendente, un campione casuale.

A grappoli

A due o più stadi

Sistematico

Il campionamento

Metodi di campionamento - Probabilistici

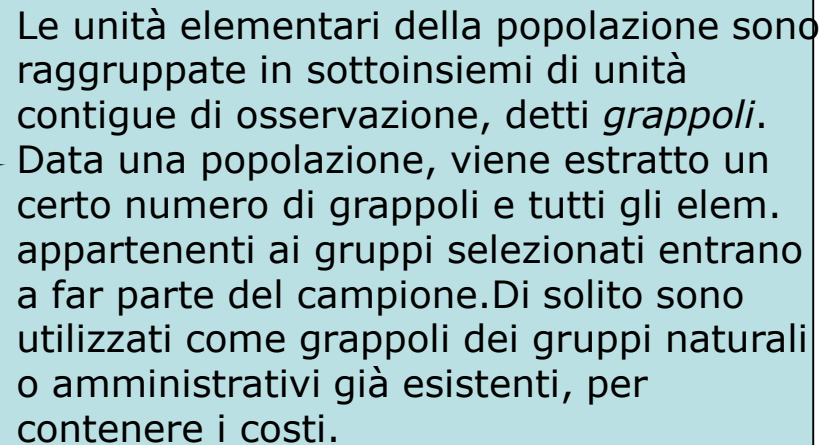
Campionamento casuale semplice

Stratificato

A grappoli

A due o più stadi

Sistematico



Le unità elementari della popolazione sono raggruppate in sottoinsiemi di unità contigue di osservazione, detti *grappoli*. Data una popolazione, viene estratto un certo numero di grappoli e tutti gli elem. appartenenti ai gruppi selezionati entrano a far parte del campione. Di solito sono utilizzati come grappoli dei gruppi naturali o amministrativi già esistenti, per contenere i costi.

Il campionamento

Metodi di campionamento - Probabilistici

Campionamento casuale semplice

Stratificato

A grappoli

A due o più stadi

Sistematico



Data una popolazione le cui unità elemen. sono riunite in gruppi, dapprima si selez. un campione casuale di gruppi, e poi si estrae un certo numero di unità elementari dai gruppi selezionati. Vi sono allora due livelli di campionamento: al primo vengono scelti i gruppi o le unità di primo stadio, al secondo vengono scelte le unità elementari, chiamate unità secondarie

Il campionamento

Metodi di campionamento - Probabilistici

Campionamento casuale semplice

Stratificato

A grappoli

A due o più stadi

Sistematico



È necessario che le unità siano ordinate secondo un criterio qualsiasi; solo la prima unità verrà estratta in modo casuale dalla popolazione, le altre verranno selezionate in modo automatico secondo un criterio prefissato (es. una ogni 3).

Il campionamento

Metodi di campionamento – Non Probabilistici

Per quote



La popolazione viene suddivisa in gruppi omogenei in base ad alcune variabili strutturali (sesso, età, residenza...). Si individua il numero di osservazioni da raccogliere in ogni gruppo (*quote*): gli intervistatori scelgono il campione allo interno delle classi, in modo da rispettare le proporzioni prescelte. L'intervistatore ha totale arbitrarietà nella scelta delle unità da intervistare nelle quote assegnate.

A scelta ragionata

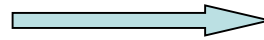
Per convenienza

Il campionamento

Metodi di campionamento – Non Probabilistici

Per quote

A scelta ragionata



Le unità campionarie sono selezionate solo in determinate aree di analisi, sulla base di informazioni preliminari circa la popolaz. indagata. È quindi un campionamento basato sulla conoscenza del carattere oggetto di studio, particolarmente adatto per campioni di piccole dimensioni.

Per convenienza

Il campionamento

Metodi di campionamento – Non Probabilistici

Per quote

A scelta ragionata

Per convenienza



La selezione degli elementi è basata su criteri di convenienza temporale, economica o di altro genere.

Target population definition

- We plan to target Milanese inhabitants (Italian and foreign) from 18 to 60 years old
- We chose to focus on the geographical area rather than on the origin as the French concept is internationally appealing
- We personally foresee that our real target population will be the people between 25 and 55 years old but in order not to bias our study, we prefer to extend this scale from 18 to 60 years old
- We assume that aspirations and financial means evolve along life cycle, therefore our study is based on four different categories pertaining to four life stages:

➤ 18-25

➤ 26-35

➤ 36-45

➤ 46-60

	M	F	Total
18- 25	44 390	41 036	85 426
26- 35	83 446	81 204	164 650
36- 45	117 411	115 740	233 151
46- 60	127 402	136 367	263 769
Total	372 649	374 347	746 996

Source: demo.istat.it/The overall Milanese targeted population

Choice of sample

- Only 100 questionnaires because we are foreign students and we are aware that our network is not as large as that of Italian students
- If the output of our project reveals a significant business opportunity, we recommend to carry out a new and wider-ranging market research in order to gain more accuracy and statistical relevance

	M	F	Total
18- 25	6	6	12
26- 35	11	11	22
36- 45	16	15	31
46- 60	17	18	35
Total	50	50	100

Sampling

E-COMMERCE ANALYSIS



▶ GROUP NAME - GOOGOL

2. DESCRIPTION of the DATABASE

- ▶ Universe: people of age between 20 and 25
- ▶ Period of collecting data: May 2008
- ▶ Size of the Database: 178 cases
- ▶ Type of variables:
 - Quantitative
 - Qualitative
- ▶ The latter ones have been converted from
into numeric



Sampling - *Sample size*

The major aspects that affect the choice of the sample size are the following:

- Cost of fieldwork
- Time of fieldwork
- Degree of accuracy of the estimates:
 - Variability of the character investigated in the targeted population
 - Sampling error
 - Sampling Design
 - Choice of the estimator

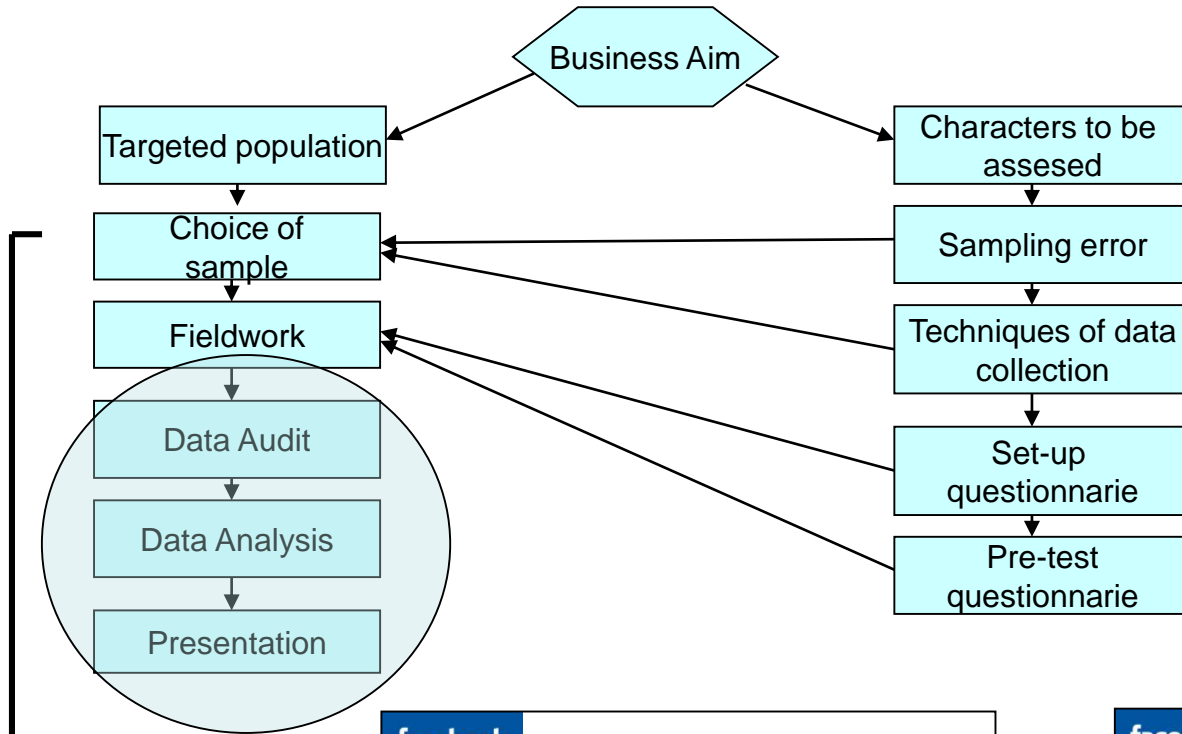
Sampling - *Sample size*

Practical tips:

- Sample picked by a Customer Database with regard to customer profiling analysis: 50.000-200.000 sample units (consumer)
- Sample picked with regard to ad hoc surveys: 1.000-3.000 sample units (consumer)
- Sample picked with regard to repeated surveys (Panel): 4.000-9.000 sample units (consumer)

Quantitative Market Research

Set-up Protocol



facebook

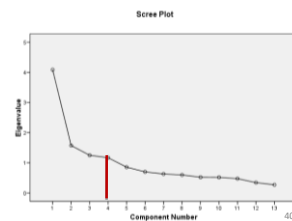
4 Factors

Our choice was consistent with the following criteria:

- The proportion between the maximum number of variables and the chosen factors is in the acceptable range (4/13 < 30%)
- The Eigenvalues are all bigger than 1
- The Cumulative Variance Explained is over 60%
- Communalities homogeneous values

As the Scree Plot confirms, only after 4 components the slope of the curve sensibly decreases.

NUMBER OF FACTORS (K/P)	OK
EINGENVALUE (> 1)	OK
% global explained variance (Between 60 and 75%)	62% - OK
Communalities - Low difference between highest and lowest (+-30%)	79% - 50% = 29% - OK



facebook

The 5 Clusters

- **Cool Hunters (28%)**: More than all, they are users absolutely interested on **Broadening**.
- **PR's (7%)**: Interested above all in **Public Relations** and express some attachment to **Spying**, but not related at all with **Keeping Up**.
- **Detached (20%)**: Apart from some light interest on **Broadening**, they do not express any involvement with the Facebook use (in particular with **Public Relations**).
- **Functional (18%)**: Above all, interested in **Keeping up** with their network of friends and use **Public Relations** inside this network. Besides, they do not care at all about **Spying** and **Broadening**.
- **Gossipers (27%)**: They are also interested in **Keeping up**, but above all in **Spying** their network. Furthermore, they are not interested in **Public Relations** and **Broadening**.

Each single Cluster was then crossed with socio-demographic and usage variables, through the contingency table tool, in order to better understand their main characteristics. The following slides sum-up the most relevant results of these crossings for each single cluster.

Tipologie di dati

- **Qualitativi** dati espressi in forma verbale, solitamente classificati in categorie
- **Quantitativi** dati espressi in forma numerica. si distinguono in:
 - **discreti** dati caratterizzati da una quantità finita o infinita numerabile di classi di misura
 - **continui** risposta numerica derivante da un processo di misurazione che fornisce indicazioni puntuali all'interno di un *continuum*
- **Territoriali**
- **Date**

Tipologie di dati *qualitativi*

- **Nominale** usato per dati *qualitativi*, che vengono così classificati in categorie distinte senza alcun ordine implicito (es. professione del cliente)

Where do you come from?	
a. North Italy	<input type="checkbox"/>
b. Center Italy	<input type="checkbox"/>
c. South Italy	<input type="checkbox"/>
d. Outside Italy	<input type="checkbox"/>

- **Ordinale** le categorie presentano un ordine implicito; consente di stabilire una relazione d'ordine tra le diverse categorie, ma nessuna asserzione numerica, ovvero si può dire che un determinato valore è più grande di un altro, ma non di quanto

Education level (Currently Attending)	
a. High School	<input type="checkbox"/>
b. Undergraduate	<input type="checkbox"/>
c. Graduate	<input type="checkbox"/>

Tipologie di dati *quantitativi*

- **Scala di rapporti** con questa tipologia si può dire di quanto una categoria è maggiore di un'altra; è fissato un valore "0" della scala.

es. Le variabili *spesa media* e *tempo impiegato* sono misurate a livello di rapporto, ovvero rientrano in una scala di valutazione comparativa

How long have you been a Facebook user for (Months):

Approximately, how many friends do you have on Facebook:

How many of these friends do you contact regularly:

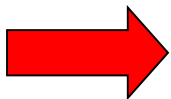
On average, how many times a week do you check Facebook:

How much time do you spend on each visit (in minutes):

Tipologie di dati *quantitativi*

- **Scala di intervalli** presenta le stesse caratteristiche della precedente, ma non possiede un valore “0” fissato.
es. In una indagine sui clienti di un supermercato, il loro livello di soddisfazione può essere adeguatamente rappresentato mediante una scala di valutazione compresa tra 1 e 9, ciò che posso asserire è che la differenza tra 2 e 3 è la medesima di quella tra 8 e 9, ma non che 8 sia il doppio di 4.

Where do you connect on Facebook more frequently?	Low			Medium			High		
	1	2	3	4	5	6	7	8	9
a. Home,									
b. Work/ University									
c. Other places (internet point, friends' houses ..)									



La tipologia di dati guida l'analisi

L'analisi statistica dei dati

Statistica descrittiva insieme dei metodi che riguardano la rappresentazione e sintesi di un insieme di dati al fine di evidenziarne le caratteristiche principali

Statistica inferenziale insieme dei metodi che permettono la *stima* di una caratteristica di una *popolazione* basandosi sull'analisi di un *campione*

Misura riassuntiva, calcolata sui dati campionari, utile per descrivere una caratteristica non nota della popolazione

parte di popolazione selezionata per l'analisi

Totalità degli elementi presi in esame dalla indagine