

# *Le ricerche di mercato*

## **OBIETTIVO:**

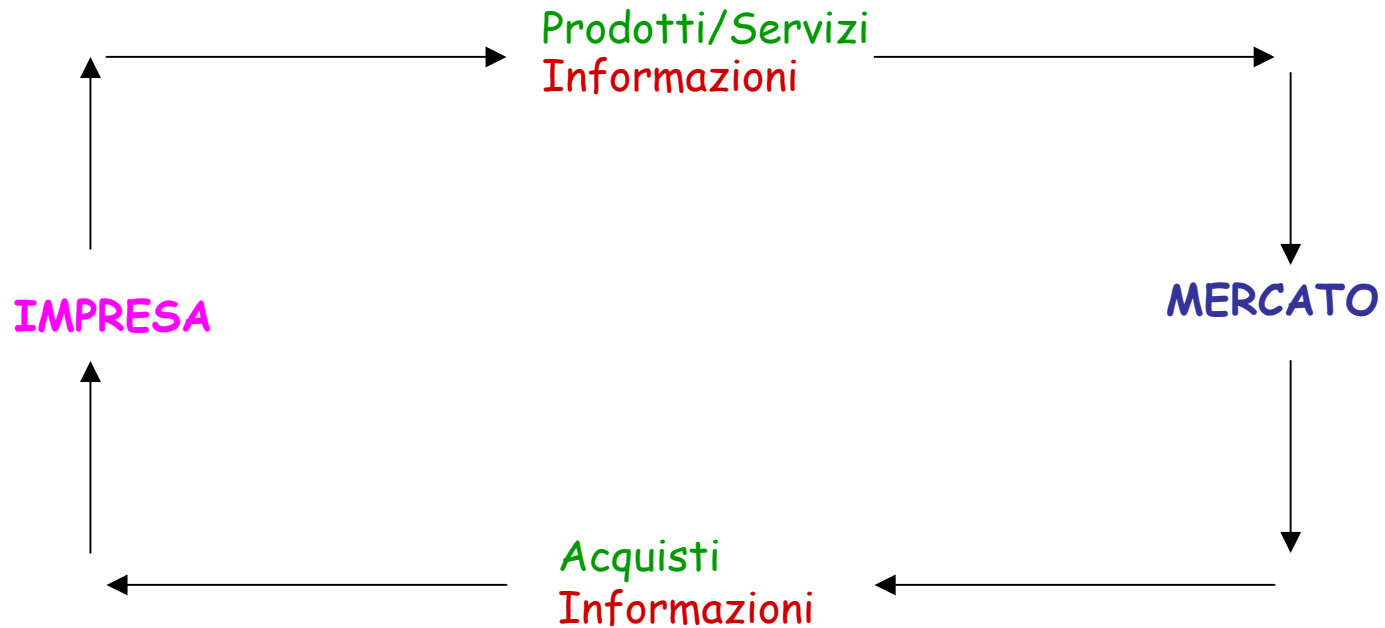
fornire informazioni alla dirigenza della funzione marketing per riconoscere le **opportunità** e i **problemi** del mercato e reagire convenientemente

## **RISULTATO:**

soddisfacimento dei **bisogni informativi** generati dalla necessità di sviluppare azioni e strategie all'interno di un'organizzazione

## **STRUMENTI:**

individuazione ed analisi sistematica e oggettiva delle informazioni rilevanti per la soluzione di un problema o l'individuazione di un'opportunità



**Noi ci concentreremo su come ottenere dal mercato (attuale o potenziale) le informazioni che l'impresa ritiene necessarie al punto da investire nel loro reperimento**

## Conoscenza dei mercati: alcuni esempi

- Risposta dei vari segmenti di consumatori a particolari caratteristiche del nostro prodotto o a particolari azioni
- Posizione della distribuzione verso i prodotti delle principali marche
- Valutazione della soddisfazione del cliente

<< Chi sono i nostri clienti? Ci sono fedeli? Credono in noi? >>

<< Quanti e quali prodotti / servizi richiedono? >>

<< Sono soddisfatti da quello che diamo loro? >>

# TIPI DI RICERCHE DI MERCATO

## 1. **CROSS-SEZIONALI** (esplorative/confermative)

### STRUMENTI:

INTERVISTE DA QUESTIONARIO  
FOCUS GROUPS  
USO DI DATI SECONDARI

## 2. **LONGITUDINALI**

### STRUMENTI:

INDAGINI RIPETUTE  
CAMPIONI RUOTATI  
*PANEL*

*Il **PANEL** è costituito da un gruppo di intervistati che si dichiara disponibile ad essere intervistato per un periodo di tempo*

Es. panel *continuativo*: campione fisso nel tempo su cui si rilevano le **stesse variabili in tempi diversi**

panel *omnibus*: un **campione fisso** nel tempo su cui si rilevano **variabili differenti in tempi diversi**

# ALTRI TIPI DI RICERCHE DI MERCATO

3. **SITUAZIONI ANALOGHE**: casi studio; simulazioni

4. **INDAGINI A TAVOLINO**: ricorso a fonti secondarie, ossia dati raccolti con obiettivi differenti.

## FONTI DI DATI PRIMARIE E SECONDARIE

Le indagini sul campo (**fonti primarie**) devono essere precedute da un'attenta indagine a tavolino, per raccogliere le informazioni disponibili sul fenomeno che si vuole analizzare.

Il ricorso a **fonti secondarie** offre numerosi vantaggi:

- è molto più economico di un'indagine sul campo
- è molto utile nella definizione del problema
- aiuta a progettare la raccolta di dati primari
- aiuta nella definizione della popolazione e del campione

---

In 100 countries around the world, Nielsen provides clients the most complete understanding of what consumers watch and buy.

---

As a global leader in measurement and information, we believe providing our clients a precise understanding of the consumer is the key to making the right decisions -- decisions that can lead to profitable growth. At Nielsen, we're always innovating to keep pace with emerging market trends and the increasingly diverse, demanding and connected consumer.

After nearly a century, we're more focused and skilled than ever at providing the complete view of what consumers watch and buy through powerful insights that clarify the relationship between content and commerce. Whether our clients are in media, consumer packaged goods, telecom or advertising, our expansive data and measurement capabilities provide market context and confidence through our long history of innovation and integrity.

Informazioni integrate Nielsen fornisce ai propri clienti la più completa visione dei loro mercati e una migliore conoscenza dei propri consumatori attraverso ricerche, analisi, esperienze, e tecnologie avanzate.

# Il Codice Nielsen

Arthur C. Nielsen Senior Presidente, A. C. Nielsen Company, 1931

Nel 1931, Arthur C. Nielsen Senior ha redatto il Codice Nielsen, che definisce i principi guida di ACNielsen per raggiungere la leadership mondiale nel campo delle ricerche di mercato. Il Codice conserva a tutt'oggi la sua importanza e il suo ruolo di guida.

## Imparzialità

Non essere influenzati da nulla se non dagli interessi del proprio cliente. Dire sempre al cliente la verità.

## Minuziosità

Un incarico va accettato solo a un prezzo che consenta di essere minuziosi. A questo punto, l'incarico deve essere svolto in modo accurato, indipendentemente dal suo costo per l'azienda.

## Accuratezza

Controllare ogni dettaglio che possa influire sull'accuratezza del lavoro.

## Integrità

Mantenere la riservatezza dei problemi dei clienti e dei potenziali clienti. Divulgare le informazioni esclusivamente con il loro consenso.

## Economia

Utilizzare qualsiasi criterio di economia compatibile con i criteri di minuziosità, accuratezza e affidabilità.

## Prezzo

Specificare nei preventivi i prezzi che consentiranno di ottenere un profitto equo. Non modificare mai i prezzi se non a seguito di una modifica delle specifiche.

## Consegna

Offrire ai clienti la più rapida consegna possibile, compatibilmente con il criterio di qualità e indipendentemente dal disturbo che possa arrecare a noi.

## Assistenza

Non lasciare nulla di intentato per aiutare i clienti a realizzare il massimo profitto dal loro investimento.

---

**Nel corso della sua storia, Nielsen ha sempre cercato di fornire ai clienti informazioni di marketing obiettive e affidabili e analisi sui consumatori in grado di aggiungere valore alla loro azienda. Il nostro obiettivo è quello di aiutare i clienti a prendere le decisioni di marketing migliori possibili per creare fatturato e profitto.**



## What we do – Market Research

TNS is one of the world's leading market research groups. We provide Industry Sector Insight, and Innovative Market Research Solutions, to many of the world's leading companies.



# *Qualità & Quantità*

Tipicamente gli strumenti di analisi cui si ricorre nell'ambito delle ricerche di mercato (e più in generale delle scienze politiche e sociali) vengono classificati in qualitativi e quantitativi, anche se oggi sempre più si è diffusa la convinzione dei vantaggi dei cosiddetti *mixed methods*, ovvero degli utilizzi congiunti dei due approcci

Esistono numerosi punti di vista per distinguere fra gli uni e gli altri. I due principali sono il numero di individui coinvolti e le discipline che forniscono gli strumenti metodologici per affrontare la ricerca

Uno sguardo alla storia ci aiuta a comprendere ...

# Breve storia delle ricerche di mercato (1)

- **Ottocento:**
  - **Stati Uniti:** applicazioni non sistematiche negli, finalizzate, in genere alla definizione di *campagne pubblicitarie*
  - **Gran Bretagna:** grande rilievo indagini campionarie di notevoli dimensioni, in ambito politico-sociale (non di mercato)
- **Nel 1908** nasce la prima **organizzazione** per le ricerche di mercato negli **Stati Uniti**
- **Il 1919** è considerata la data di nascita delle Ricerche di Mercato come disciplina, con la pubblicazione del libro *Commercial Research: an Outline of Working Principles* di da parte di C. S. Duncan
- **Primi anni Venti del Novecento**, l'Unilever, già nei, realizza ricerche statistiche sui consumatori

## *Breve storia delle ricerche di mercato (2)*

Sono gli anni in cui le ricerche di mercato ricorrono alle *tecniche statistiche*, anche per il diffondersi dei sondaggi di opinione ed elettorali, legati a figure quali *George Gallup*



*Nel 1935, George Gallup fondò l'American Institute of Public Opinion e diviene famoso per i sondaggi elettorali per le presidenziali del 1936: predisse sorprendentemente, per l'epoca, la vittoria di Franklin D. Roosevelt con uno scarto del -6,8%*

## *Breve storia delle ricerche di mercato (3)*

A partire dal Secondo Dopoguerra, la percezione diffusa è che le discipline quantitative si siano dimostrate ormai **inadeguate**

La **Sociologia** fornisce contributi sia in termini teorici, attraverso l'introduzione di concetti quali la **classe sociale**, il **nucleo familiare**, gli **stili di vita**, sia attraverso lo sviluppo di metodi di analisi e di osservazione basate più sulle capacità tecniche del ricercatore, che sulla quantità di osservazione disponibile

La **Psicologia** allo stesso modo sposta l'attenzione del ricercatore sull'individuo e orienta la ricerca alla comprensione ed all'approfondimento dei **comportamenti di acquisto** e delle **motivazioni** che ne sono alla base, fornendo strumenti metodologici quali le **interviste in profondità** che ricalcano gli schemi del colloquio clinico

## *Breve storia delle ricerche di mercato (4)*

Anche la **Statistica** sviluppa metodologie sempre più sofisticate ed adatte al trattamento di **situazioni complesse**, rese possibili dalle nuove disponibilità di strumenti di calcolo

Si sviluppano anche tecniche di analisi finalizzate al trattamento di dati diversi, quali i testi e le immagini, che grande importanza hanno nell'ambito delle ricerche di mercato

E' il trattamento automatico di ingenti moli di dati in tempi impensabili fino a pochi anni fa che ha reso possibile tecniche di *marketing one-to-one*, il *profiling* della clientela

Oggi esiste sempre più la consapevolezza che i **mixed methods**, con l'utilizzo coordinato di tecniche di ambiti differenti, siano indispensabili, anche se questo rende sempre più

*complesso e sofisticato il compito dei ricercatori di mercato*



## *Progettazione di una ricerca di mercato*

- 1. Decidere di quali informazioni abbiamo bisogno*
- 2. Scegliere il metodo per ottenerle*
- 3. Progettare e Scrivere un buon questionario*
- 4. Estrarre un campione*

*e poi, costruire la base di dati, raccogliarli, analizzarli,  
interpretare i risultati, comunicare i risultati, ...*

## *Rapidi cenni a metodi qualitativi utilizzati nelle ricerche di mercato*

Dal punto di vista dello statistico, la maggiore discriminante fra metodi qualitativi e metodi quantitativi è data, forse, dalla natura degli elementi osservati

Sempre nel linguaggio dello statistico, potremmo dire se ci troviamo di fronte ad un campione casuale (sulla cui definizione ritorneremo in seguito) o su di un campione non casuale

Nelle ricerche di mercato, tipicamente, ci troviamo di fronte a campioni non casuali:

- *a scelta ragionata*
- *per quote*
- *a testimoni privilegiati*

Oppure ad *interviste di gruppo* (Focus, Delphi,...)

## **Campione a *scelta ragionata***

### **Quando?**

**Necessità di includere elementi rilevanti per la totalità del fenomeno. Es. Aree di prova**

### **Come?**

**Sulla base di scelte effettuate da esperti**

### **Inconvenienti**

**Risente del punto di vista degli esperti**

**Non consente di avvalersi di procedimenti statistici di tipo inferenziale**

## *Interviste a testimoni privilegiati*

(soggetti, ma anche consumatori, ritenuti, per caratteristiche individuali o ruoli sostenuti, portatori di una informazione rilevante)

- **Interviste in profondità, semi-strutturate**

Simulano i **colloqui clinici** psicoanalitici. L'intervistatore segue una schema ed interviene quando necessario  
l'intervistato parla liberamente

**Scopo:** far emergere motivazioni razionali e non, motivazioni individuali e influenze del gruppo di appartenenza

Registrate, durata circa un'ora

- **Intervistatore qualificato** psicologo o esperto

**Es. Studi di atteggiamento o motivazionali**

## *Interviste a testimoni privilegiati*

### ❖ **Story- telling** (interviste non direttive)

All'intervistato viene chiesto di raccontare una storia in cui compaia l'oggetto dell'intervista

**Scopo:** capire se il consumatore accetta o meno determinati prodotti

durata libera (poco utilizzata)

**Ricercatore:** psicologo

### ❖ **Etnometodologia, Videoanalisi** (interviste non direttive)

L'intervistatore osserva (direttamente o in filmato) individui nella loro dimensione naturale, a loro insaputa

**Scopo:** indagare consumi fruiti nel sociale, abitudini

## *Selezione per quote*

**VANTAGGI** (rispetto ad uno schema casuale)

- **Minori costi**
- Se la rilevazione è effettuata in maniera rigorosa è **comparabile ad un campione statistico**
- **Minori tempi di rilevazione**

## *Selezione per quote*

### **INCONVENIENTI** (rispetto ad uno schema casuale)

- **Distorsione** rispetto alle abitudini degli individui che compongono la popolazione
- Necessità di un rigoroso **controllo** sull'attività degli intervistatori
- Attenzione non solo alle quote definite rispetto alle distribuzioni marginali, ma anche a quelle congiunte (quando necessario)

## *Cenni sui metodi di gruppo*



**Frequentemente nelle indagini di mercato di tipo qualitativo si riuniscono gli intervistati in un gruppo, cui, di regola, non viene posta una batteria di domande predefinite, ma spunti di discussione, proposti dal conduttore del gruppo**

**I *vantaggi* delle interviste di gruppo:**

- **rapidità**
- **economicità**
- **dinamiche** di gruppo (interazioni, maggior spontaneità, una caduta delle resistenze dei partecipanti, un maggior confronto, verifiche dirette)

## *Il focus group (1)*



***Il focus group nasce negli Stati Uniti negli anni '40 del Novecento, ad opera di due sociologi (K. Levin e R. Merton)***

***«Un focus group è una forma di ricerca qualitativa, in cui un gruppo di persone è interrogato riguardo all'atteggiamento personale nei confronti di un prodotto, di un concetto, di una pubblicità, di un'idea, o di un imballaggio ...***

***Nel mondo del marketing, i focus group sono uno strumento importante per l'acquisizione di feedback riguardo ai nuovi prodotti. In particolare, i focus group permettono alle aziende che desiderano sviluppare, nominare, o esaminare un nuovo prodotto, di discutere, osservare e/o esaminare il nuovo prodotto prima che esso sia messo a disposizione del pubblico. Ciò può fornire informazioni inestimabili sull'accettazione del prodotto da parte del suo mercato potenziale »***

**(Wikipedia)**

## *Il focus group (2)*



Nelle ricerche di mercato si usa il focus group per individuare **atteggiamenti** e **comportamenti** di consumatori (attuali e potenziali) nei confronti di un determinato prodotto/servizio, e per approfondire e spiegarne le **motivazioni**

L'esperienza mostra come:

- siano particolarmente utili per ottenere **reazioni** dirette a proposte di prodotti/servizi
- vadano preferiti alle interviste in profondità se l'argomento concerne prodotti o servizi fruiti in un **contesto sociale**
- hanno anche una fondamentale importanza nella **progettazione di questionari** per successive indagini a carattere quantitativo

## *Struttura del focus group*



**Durata:** il *focus group* non deve durare meno di 90 minuti e non oltre i 120 minuti

**Condizione:** il *focus group* viene normalmente effettuato da due persone: un *animatore/moderatore* che conduce la discussione e un osservatore che esamina le dinamiche di relazione del gruppo

**Discussione:** non è strutturata (o è poco strutturata ed il conduttore incoraggia il flusso libero delle idee)

**Partecipanti:** i partecipanti al *focus group* devono essere almeno 6/7 (un numero inferiore potrebbe inficiare le dinamiche di gruppo) e non più di 12/13 persone (un numero superiore tende a censurare l'intervento delle opinioni contrarie o deboli e non permette a tutti i partecipanti di esprimere al meglio le proprie idee)

## *Inconvenienti dei Focus group*



- **la presenza di più persone può costituire un freno inibitorio all'espressione di alcune sensazioni personali**
- **il colloquio di gruppo consente di ottenere un maggior numero di spunti rispetto all'intervista in profondità, ma a un livello meno analitico e meno profondo**
- **non è sempre facile impedire che la conversazione sia almeno in parte dominata da persone estroverse e dotate di facilità di parola**

## *Il Metodo Delphi*

**Il Metodo Delphi consente, tramite la somministrazione ripetuta di questionari, di ottenere opinioni singole e di sollevare un confronto, una sorta di **dibattito “virtuale”** tra gli esperti selezionati per il campione, intorno all’oggetto di una ricerca**



**Si tratta di un metodo **qualitativo, partecipativo, previsionale****

Il suo utilizzo è particolarmente adatto alla realizzazione di più scopi:

- realizzare il confronto tra più esperti appartenenti a “categorie” diverse;
- far emergere il “conoscere tacito” sotto ad ogni professione;
- disegnare il quadro dell’oggetto, evento, problema su cui si sta indagando;
- delineare un probabile sviluppo dell’oggetto di studio

Il metodo Delphi consiste in fasi successive di sollecitazione e consultazione, tramite questionari ed incontri, di un campione di “testimoni privilegiati” riguardo alla tematica oggetto dello studio

*L'INDAGINE CAMPIONARIA come  
principale strumento quantitativo per  
le  
RICERCHE DI MERCATO*

## *Le FASI di una ricerca di mercato*

1. Stabilire i **FABBISOGNI** di **INFORMAZIONE**
2. Specificare gli **OBIETTIVI** della ricerca in relazione ai fabbisogni informativi
3. Determinare il **PIANO** della ricerca e le **FONTI** dei **DATI**
4. Sviluppare le **PROCEDURE DI RACCOLTA** dei dati
5. Predisporre il **PIANO DI CAMPIONAMENTO**
6. Effettuare la **RILEVAZIONE** dei dati
7. Elaborare ed **ANALIZZARE** i dati
8. Presentare i **RISULTATI** della ricerca

**CIASCUNA DI QUESTE FASI  
PUO' ESSERE **FONTE DI ERRORE****



L'errore è spesso un problema di **COMUNICAZIONE**

**COMMITTENTE**



**RICERCATORE**

### **COMMITTENTE (O DECISORE)**

- DETERMINA IL PROBLEMA E DEFINISCE GLI OBIETTIVI (DELL'ORGANIZZAZIONE E PERSONALI)
- COMUNICA CHIARAMENTE PROBLEMI E OBIETTIVI AL RICERCATORE

### **RICERCATORE DI MERCATO**

- STABILISCE I FABBISOGNI INFORMATIVI
- DEFINISCE IN TERMINI CHIARI IL PROBLEMA
- SVILUPPA GLI OBIETTIVI SPECIFICI
- DEFINISCE CON IL MANAGER LE SPECIFICHE

## »FORMULA IL MODELLO

1. INDIVIDUA LE VARIABILI DI INTERESSE
2. INDIVIDUA LE VARIABILI CONTROLLABILI E QUANTO CONTROLLO PUO' ESSERE ESERCITATO
3. DETERMINA LE RELAZIONI FUNZIONALI FRA LE VARIABILI

## »FORMALIZZA IL PROGETTO

### SCELTE DI BASE:

1. FARE LE DOMANDE GIUSTE
2. FARLE ALLE PERSONE GIUSTE
3. FARE LE DOMANDE NEL MODO GIUSTO
4. OTTENERE LE RISPOSTE GIUSTE RISPETTO ALLE ESIGENZE CONOSCITIVE DEL COMMITTENTE

## »SCEGLIE LA TECNICA DI RILEVAZIONE

INTERVISTE TELEFONICHE

INTERVISTE POSTALI (Internet, fax, ecc.)

INTERVISTE PERSONALI

**FARE LE DOMANDE “GIUSTE”**

**FARLE ALLE PERSONE “GIUSTE”**

**FARE LE DOMANDE NEL MODO “GIUSTO”**

**OTTENERE LE RISPOSTE “GIUSTE”**



**Ovvero, cercare di evitare i 4 *maggiori* pericoli di un'indagine:**

**L'errore di *copertura*: la lista che utilizziamo non include tutti gli individui che si vorrebbero studiare**

**L'errore di *campionamento*: quando si osserva un campione, piuttosto che la popolazione**

**L'errore di *misurazione*: quando la risposta fornita da un rispondente ad una specifica domanda non è accurata, precisa, o comunque comparabile con quelle fornite dagli altri rispondenti**

**L'errore di *mancata risposta*: quando un numero “rilevante” di soggetti non risponde e possiamo ritenere che siano “diversi” da coloro che ci hanno risposto**

# L'ERRORE IN UN'INDAGINE: definizione

## ERRORE STATISTICO

E' la discrepanza tra **valore vero** e **valore disponibile** dall'indagine statistica

Può essere riferito alla singola **unità statistica**: la discrepanza è dovuta al complesso degli errori legati alla rilevazione dell'informazione dall'unità statistica

Può essere riferito ad una **statistica di sintesi** dei dati rilevati mediante indagine campionaria. In questo caso si aggiunge l'errore legato al campionamento a causa del quale, anche nell'ipotesi puramente teorica che i valori di ogni unità fossero rilevati con esattezza, la statistica campionaria differirebbe comunque dal valore osservabile in un'indagine completa

## ERRORE DI RILEVAZIONE (Statistico)

Si ha un errore statistico ogni qual volta il valore osservato differisce dal valore **vero**. Si manifesta, quindi, sia nelle rilevazioni **campionarie** che in quelle **esaustive**. Di regola viene misurato in termini di ERRORE QUADRATICO MEDIO:

$$EQM(\hat{\vartheta}) = E(\hat{\vartheta} - \vartheta)^2$$

# L'ERRORE IN UN'INDAGINE: quantificazione

**ERRORE TOTALE** (statistico) =  
**ERRORE CAMPIONARIO** +  
**ERRORE EXTRA-CAMPIONARIO**

## **ERRORE CAMPIONARIO:**

e' l'errore variabile che deriva dal fatto che noi stimiamo qualcosa circa la nostra popolazione di interesse a partire da un campione

può essere misurato dalla

*ampiezza di un intervallo di confidenza*

(fissato il livello  $\alpha$ )

## **ERRORE EXTRA-CAMPIONARIO:**

tutti gli altri possibili errori

Ad esempio:

- errore nell'**identificazione della popolazione**
- errore nell'**individuazione delle unita'** appartenenti alla popolazione
- **non risposte sistematiche**
- errori nell' **estrazione** del campione
- errori di **rilevazione**

.....

# Qualità di un'indagine

La **qualità di un'indagine** è strettamente legata alla varianza di risposta, poiché è possibile operare dei controlli, mentre quella di campionamento è determinata dalle caratteristiche del fenomeno ed alle scelte di campionamento

Se si concepisce un' *INDAGINE* come un **PRODOTTO**, allora, l'**ERRORE EXTRA-CAMPIONARIO** può essere preso come **misura della qualità del prodotto**, in termini di:

**ATTENDIBILITA'** = accuratezza delle stime

**ADEGUATEZZA** = raggiungimento degli obiettivi + diffusione dei risultati

# L'ERROR PROFILE

Il **profilo dell'errore** (*error profile*) di un'indagine è:

<< la **descrizione completa** e **ordinata** delle operazioni e delle **potenziali fonti di errore** nonché dell'**effetto** dell'errore di ciascuna operazione sull'errore complessivo >>

## Vantaggi dell'*error profile*

- \* dimensionamento dell'errore complessivo
- \* sensibilizzazione degli utilizzatori sugli **effetti** dei diversi tipi di errore
- \* stimolo a predisporre programmi per il controllo della qualità dei dati

# Meta-dati e Meta-analisi

## *Metadati :*

informazioni di carattere qualitativo e quantitativo riguardanti le caratteristiche di un'indagine prendono il nome di *metadati* e consentono di qualificare i dati che si vanno ad utilizzare

## *Meta-analisi :*

analisi dei metadati, per porre a confronto le risultanze di esperimenti e di analisi, pratica comune in epidemiologia e statistica medica

*Gli analisti di mercato (e più in generale gli utilizzatori di dati statistici, ma anche gli statistici) dovrebbero sempre più acquisire la consapevolezza e la capacità critica di giudicare il dato su cui operano, prima ancora di interpretare i risultati della propria analisi*

# Le fonti di errore extra campionario

## 1. errori dovuti al ricercatore

sono gli errori commessi nella fase di progettazione dell'indagine

## 2. errori dovuti al rilevatore

## 3. errori dovuti al rispondente

2. e 3. sono gli errori compiuti in fase di rilevazione una delle insidie maggiori è "l'effetto rilevatore", generato dall'interazione dei due soggetti, ovvero il condizionamento indotto sul rispondente dal rilevatore

## 4. errori dovuti all'operatore di registrazione

molto ridotti oggi, grazie ai sistemi di intervista e di registrazione assistite dal computer

# le fonti di errore extra campionario 1. :

## errori dovuti al ricercatore

- non chiarezza sugli obiettivi dell'indagine
- non corretta individuazione della tecnica d'indagine
- non corretta predisposizione del questionario
- non corrette istruzioni per i rilevatori
- imprecisioni nel sistema di codifica
- non corretta identificazione della popolazione
- non corretta identificazione della lista
- ...

## le fonti di errore extra campionario 2. :

### errori dovuti al rilevatore

- impreparazione sul questionario
- impreparazione sulla conduzione dell'intervista
- scarsa motivazione

# le fonti di errore extra campionario 3. :

## errori dovuti al rispondente

- problemi di memoria
- mancanza d'informazione
- scarsa motivazione
- fraintendimento involontario di alcuni quesiti
- distorsione volontaria della risposta
- condizionamento dovuto alla presenza di altre persone
- effetto proxy

# le fonti di errore extra campionario 4. :

## errori dovuti all'operatore di registrazione

- Errore di registrazione del **valore** del dato
- Errore di registrazione della **posizione** del dato nel record finale

# il dato mancante

*L'errore determinato dalla mancanza d'espressione del dato può essere visto come la manifestazione di un errore non campionario*

*Questa "mancanza" si manifesta sotto forma di non partecipazione alla rilevazione di unità statistiche appartenenti al campione oppure nell'assenza di un valore espresso per una singola variabile su una osservazione*

In un'indagine statistica, è possibile identificare differenti categorie di "incompletezza":

1

se nessuna delle variabili oggetto d'interesse è misurata si ha una *unit nonresponse*



In questo caso l'assenza di una risposta causerebbe una riga di meno all'interno della matrice dei dati

2

se l'unità selezionata risponde a molte delle domande proposte ma non a tutte, o se la risposta è giudicata non corretta e cancellata in fase di editing si ha una *item nonresponse*



In questo caso l'assenza di una risposta causa un vuoto all'interno della matrice dei dati

Dati Mancanti

		VARIABILI				
		1	2	3	...	p
Unità	1					
	2	?				
	3					?
	.	?		?		
	.		?			?
n	?		?			

La **non risposta parziale** può avere origini diverse



- quando un'unità rispondente fornisce un'informazione incompleta perché non è disponibile a collaborare, o perché non è in grado di fornire tutte le informazioni richieste
- alle non risposte parziali possono essere associati anche i valori non ammissibili e le incongruenze logiche tra risposte a domande differenti, che si manifestano come contraddizioni rispetto a **relazioni sostanziali** (implicite nella realtà esaminata) o **relazioni formali** (costruite insieme al questionario) tra i valori assunti dalle variabili relative.



A differenza delle non risposte totali, le non risposte parziali possono essere evidenziate solo nella fase di *screening*, cioè dopo la fase di raccolta delle informazioni, e quindi devono essere trattate con tecniche specificatamente approntate.

## Dati Mancanti

		Variabili				
		1	2	3	...	p
Unità	1					
	2		?			
	3					?
	.			?		
	.	?			?	
	.			?		?
n		?		?		

Capire se la presenza di valori mancanti è strutturale e, quindi, se è possibile definire un modello (**pattern**) che descriva i meccanismi che li determinano consente di individuare le procedure più adatte per trattare questi dati mancanti, richiedendo i diversi approcci una differente strategia e conseguente impegno computazionale

I **pattern** possono dipendere dalla natura delle variabili, ma anche dal fenomeno che si vuole investigare.

*Software* quali SPSS consentono di identificare **pattern** di dati mancanti

Dati Mancanti

		Variabili				
		1	2	3	...	p
Unità	1					
	2		?			
	3					?
	.			?		
	.	?			?	
	.			?		?
n	?		?			

In letteratura si considerano tre specifici pattern di dati:

1. **unit missing**: mancano sistematicamente interi blocchi di risposte, ad esempio, nel corso di studi longitudinali, frequentemente una parte dei soggetti decide di abbandonare l'indagine;
2. **pattern monotono**: quando per un insieme di individui, si identificano variabili maggiormente soggette a non risposta);
3. **missing by design** : quando agli intervistati non vengono posti tutti i quesiti, ma un sottoinsieme. In questo caso è il ricercatore a determinare la mancanza di dati, ad esempio per questioni di ordine pratico (tempo insufficiente per portare a termine l'intervista, scarsa volontà di collaborazione, e così via), oppure perché alcuni temi vengono approfonditi soltanto per soggetti che presentano alcune caratteristiche (domande "filtro")

## Dati Mancanti

		VARIABILI				
		1	2	3	...	p
USCITA	1					
	2		?			
	3					?
	.			?		
	.	?			?	
	.			?		?
	n	?			?	

La conoscenza dei meccanismi per i quali accade che in un'indagine alcuni dati non sono disponibili o sono incompleti, è un elemento cruciale per la successiva **analisi** dei dati raccolti e **l'interpretazione dei risultati**

## Dati Mancanti

		Variabili				
		1	2	3	...	p
Unità	1					
	2	?				
	3					?
	.			?		
	.	?			?	
	.			?		?
n	?			?		

Si parla di meccanismo *ignorabile*, quando, nell'ambito di un campionamento casuale si ritiene che non esistano elementi strutturali che determinano la partecipazione di un individuo all'indagine, assimilando questa situazione a quella che determina la sua non estrazione a fare parte del campione

Si tratta, evidentemente di un'ipotesi felice, difficilmente ipotizzabili nelle indagini reali

È però possibile che il disegno di campionamento scelto dai ricercatori non sia efficiente, o che sia riscontrabile un alto tasso di non rispondenti tra le unità selezionate: in questi casi è difficile tenere sotto controllo la presenza di dati mancanti e non è possibile trascurare le cause che hanno prodotto tale mancanza



Data una certa popolazione di ampiezza  $N$ , si estragga un campione casuale di numerosità  $n < N$   
Indicando con  $y_i$  (con  $i=1, \dots, N$ ) il valore assunto dalla variabile  $Y$  nella  $i$ -esima unità, se si registrano valori validi solo per  $m$  unità, si avrà un campione di ampiezza ridotta, almeno per  $Y$ , poiché si avranno  $(n-m)$  valori mancanti.  
Gli analisti possono decidere di effettuare ugualmente l'indagine utilizzando i valori registrati, ma se la probabilità che  $y_i$  sia osservata dipende dal valore assunto dalla stessa variabile (accade ad esempio per il reddito), allora non è possibile ignorare il meccanismo che ha causato una perdita d'informazione, perché le stime che si otterranno saranno poco affidabili

## Dati Mancanti

		Variabili				
		1	2	3	...	p
Unità	1					
	2		?			
	3					?
	.			?		
	.	?			?	
.			?		?	
.						
n	?		?			

Si consideri, oltre alla variabile  $Y$ , una variabile  $X$  non soggetta a non risposta, per la quale sia, quindi, possibile registrare  $n$  determinazioni campionarie: si supponga, ad esempio, che la variabile  $X$  sia una variabile di struttura (sesso, età, titolo di studio, e così via), mentre la variabile  $Y$  rappresenti il reddito annuo dell'intervistato

## Dati Mancanti

		Variabili				
		1	2	3	...	p
Unità	1					
	2		?			
	3					?
	...			?		
	...	?			?	
	...			?		?
n	?			?		

Quando si è in presenza di dati con questo pattern è utile classificare il meccanismo che provoca la mancanza di dati a seconda che la probabilità di risposta sia indipendente dalla variabile  $X$  e dalla variabile  $Y$ , dipenda da  $X$  ma non da  $Y$ , oppure dipenda da  $X$  ed eventualmente anche da  $Y$

# Missing completely at random

Se la probabilità di risposta è indipendente sia da  $X$  che da  $Y$  si dice che i dati mancanti sono **missing at random** e che i dati osservati sono **observed at random**, o più semplicemente che i dati mancanti sono **missing completely at random**

Questo caso è quello "felice", in cui possiamo ritenere che i valori osservati della variabile  $Y$  formino un sottocampione casuale dei valori campionati

# Missing at random

Se la probabilità di risposta dipende da  $X$  ma non da  $Y$ , il meccanismo di generazione della non risposta è ricostruibile sulla base di altre variabili rilevate (e non di quella per la quale mancano le determinazioni)

Si parla, allora, di dati **missing at random**

Si supponga di effettuare un'indagine telefonica sul consumo di un bene. Accade che le giovani mamme che lavorano rifiutino di partecipare, poiché non hanno tempo. In questo caso, il fenomeno che genera la non risposta non è dato dalla variabile relativa, ma da elementi indirettamente legati all'indagine.

Questo li rende **meno pericolosi**

## Missing NOT at random

Se la probabilità di risposta dipende da entrambe le variabili il meccanismo che causa la non risposta non è trascurabile; se un partecipante ad uno studio sulla perdita di peso non si presenta ad una delle verifiche periodiche, i dati relativi potrebbero mancare per fattori non trascurabili: in contrasto con il caso precedente, il meccanismo che determina la perdita di informazioni può essere spiegato soltanto dalla variabile che presenta mancanze

## Missing Data Strategie

		Variabili				
		1	2	3	...	p
Unità	1					
	2	?				
	3					?
	.			?		
	.	?			?	
	.			?		?
	.					
	n	?			?	

Le azioni da perseguire per il trattamento dei dati mancanti comportano spesso costi e tempi di realizzazione elevati: se il numero di dati incompleti o mancanti è poco rilevante, allora l'esiguità del problema non giustifica l'introduzione di complesse procedure correttive; per contro, la frequenza di tassi di mancata risposta più elevati evidenzia una non trascurabilità del fenomeno e delle conseguenze che esso comporta



Se è possibile identificare le variabili che spiegano il perché della mancanza (**mechanism variables**) allora si ha il controllo del meccanismo che tale mancanza genera

Le variabili che sono collegate più frequentemente al meccanismo di non risposta riguardano il **livello d'istruzione**, lo **status sociale**, il **bisogno di approvazione** o la **motivazione**: se il ricercatore ha sotto controllo tutte le variabili che possono spiegare il meccanismo di non risposta allora questo è trascurabile, nel senso che i risultati della ricerca non saranno soggetti a distorsioni per la presenza di non risposte

## Missing Data Strategie

		Variabili				
		1	2	3	...	p
Unità	1					
	2		?			
	3					?
	.			?		
	.	?			?	
	.			?		?
n	?			?		

Nessun procedimento analitico è in grado di affrontare il problema delle non risposte, e dei dati mancanti in generale, nella fase di raccolta dei dati; non essendo gli strumenti a disposizione dei ricercatori sufficienti a tenere sotto controllo il fenomeno delle mancate risposte, è necessario ricorrere a forme d'intervento successivo



**Obiettivo** non è quello di eliminare completamente il fenomeno, non potendo far affidamento su alcun metodo statistico che corregga tutte le differenze, ma ottenere una riduzione della non risposta

Le metodologie utilizzate per l'analisi dei dati parzialmente mancanti sono molteplici, e la varietà dei metodi proposti in letteratura è cruciale per operare la scelta più adeguata: ad ogni modo, tutti i metodi per il trattamento dei dati incompleti sono basati sull'assunzione che non è possibile verificarne la funzionalità utilizzando i soli dati a disposizione

# Le principali strategie proposte per il trattamento dei dati mancanti

1.

## *Procedure basate sull'analisi delle unità completamente registrate*

non si considerano nell'analisi le unità per le quali mancano, in alcuni campi, i valori

- sono facili da attuare
- adatte ai casi in cui si abbiano piccole quantità di dati mancanti
- possono condurre a forti distorsioni

### Missing Data Strategie

		Variabili				
		1	2	3	...	p
Unità	1					
	2		?			
	3					?
	.			?		
	.	?			?	
.			?		?	
.						
n	?		?			

## Missing Data Strategie

		Variabili				
		1	2	3	...	p
Unità	1					
	2		?			
	3					?
	.			?		
	.		?		?	
	.			?		?
n		?		?		

Si consideri una matrice rettangolare nella quale sono registrati i valori relativi a  $p$  variabili per le  $n$  osservazioni campionarie selezionate: in presenza di dati incompleti, una delle strategie adottabili è la **complete-case analysis**: si eliminano le unità con valori mancanti.

Questa eliminazione può essere effettuata, principalmente in 2 modi:

la **listwise deletion**, in base alla quale si escludono le unità per le quali manchi anche la registrazione di uno solo dei valori delle variabili utilizzate nell'indagine: tale tecnica per trattare i dati incompleti è implementata di default dai più diffusi software statistici, come SAS o SPSS;

la **pairwise deletion**, più complessa, che prevede l'inclusione di una data unità solo se vengono registrati i valori relativi ad una predeterminata coppia di variabili delle quali si vuole stimare la correlazione: tale soluzione fornisce la miglior stima per ogni correlazione, in altre parole utilizza tutte le informazioni disponibili

## 2.

### *Procedure basate sull'imputazione*

si adottano delle tecniche d'imposizione di codici plausibili alle celle vuote, in modo da creare un set di dati completo da analizzare con le tecniche standard;

talvolta tale procedimento produce un ulteriore errore, poiché il codice "forzato" è soltanto plausibile

esiste una vasta letteratura statistica in tema di imputazione di dati mancanti

## Missing Data Strategie

		Variabili				
		1	2	3	...	p
Unità	1					
	2		?			
	3					?
	.			?		
	.	?			?	
	.			?		?
n	?			?		

Per *imputazione* s'intende la sostituzione dei valori mancanti o errati di un dato record con alternative coerenti e plausibili ottenute dai dati stessi (campionari e non), da fonti esterne all'indagine o dalla combinazione di entrambi, in conformità a regole e metodi prestabiliti

Le procedure d'imputazione possono favorire la riduzione delle distorsioni introdotte dalla presenza di *missing data*, offrendo inoltre maggiori garanzie sulla coerenza dei risultati derivati dalle analisi applicate, ma al contempo non confermano l'universalità di tale riduzione, amplificando in taluni casi le distorsioni esistenti

L'immissione di dati effettuata ricorrendo a codici casuali, o ottenuti formulando congetture, comporta notevoli rischi il sistema migliore per trattare i dati mancanti, secondo alcuni studiosi, è quello di utilizzare il codice "nessuna risposta", rimandando la valutazione del fenomeno alla fase d'interpretazione dei risultati

*Imputazione della media*      *Imputazione della mediana*

*Hot Deck imputation*

## Missing Data

### Hot deck imputation

		Variabili				
		1	2	3	...	p
Unità	1					
	2		?			
	3					?
	.			?		
	.	?			?	
	.			?		?
n	?			?		

La ragione primaria per ricorrere a procedure hot deck è la necessità di ridurre le distorsioni prodotte dal verificarsi del fenomeno delle mancate risposte; l'imputazione hot deck a tal scopo utilizza frequentemente dei *processi di classificazione* di tutte le unità in gruppi distinti e quanto più omogenei all'interno, in base a caratteristiche precise stabilite di volta in volta in relazione al contesto dell'analisi: per ogni valore mancante si imputa un valore presente nello stesso subset, sotto l'assunzione che all'interno dei gruppi definiti i non rispondenti seguono la stessa distribuzione dei rispondenti.



### hot deck imputation within adjustment cells

una volta costruite le celle d'aggiustamento, s'imputano i valori mancanti in ogni singola cella scegliendo tra i valori registrati nelle stesse



### nearest neighbor hot deck imputation

prevede la definizione di una metrica per misurare la distanza tra le diverse unità, e scegliendo quindi come valori da imputare quelli relativi alle unità rispondenti più vicine a quelle che sono invece affette da incompletezza

### 3.

#### *Procedure basate su modelli*

questa classe generale di procedure prevede che sia generato un modello per i dati parzialmente mancanti, il procedimento inferenziale è quindi basato sulla verosimiglianza sotto quel particolare modello

# II QUESTIONARIO



▪ **CONTENUTI:** fatti (variabili socio-demografiche), conoscenze (economiche, sociali, politiche), opinioni, (giudizi e conoscenze) atteggiamenti (opinioni non sempre esplicite) motivazioni (a volte inconsce)

▪ **TIPO DI DOMANDE:** aperte, chiuse, semichiose, filtro

## COSE DA FARE

▪ Accertare che situazioni, conoscenze, comportamenti considerati rientrino nella cultura degli intervistati

▪ Il significato di ogni domanda deve essere univocamente comprensibile a tutti gli intervistati

- Le domande devono richiedere all'intervistato soltanto un limitato sforzo di memoria
- E' necessario limitare le indagini ai temi centrali, ed evitare che il questionario diventi eccessivamente lungo

## COSE DA EVITARE

- Formulazioni generiche
- Formulazioni che creino uno stato di tensione psicologica
- Formulazione di domande orientate
- Attenzione ai quesiti su argomenti riservati

# la rilevazione

strumenti

Esiste un'ampia gamma di possibilità tecniche per l'acquisizione delle informazioni, la cui scelta dipende da numerosi parametri connessi all'estensione degli scopi perseguiti, all'ampiezza del collettivo di riferimento, al costo ed al budget disponibili per la raccolta dei dati, al grado di precisione desiderato, e così via

La scelta del metodo ha conseguenze determinanti per tutte le successive fasi di un'indagine

Una netta discriminante, in termini di costi, tempi e qualità dei risultati è quella della presenza (telefonica, personale) o meno dell'intervistatore (postale, fax, web)

Lo sviluppo dei **sistemi assistiti da computer** (CAI, nelle versioni CAPI, CATI, CASI, CAWI) ha migliorato sensibilmente la qualità dei dati

i sistemi CAI

# Il questionario nelle indagini

Il **questionario** è lo strumento di supporto alla raccolta dei dati nelle indagini campionarie, è il veicolo preferenziale tramite il quale raccogliere informazioni, ma al contempo rappresenta lo specchio degli obiettivi dell'indagine

La costruzione del questionario è una delle fasi più delicate della progettazione di un'indagine ed è regolata da norme che devono contemperare problemi di natura psicologica e sociale agli obiettivi conoscitivi e agli strumenti di analisi statistica che verranno successivamente applicati ai dati raccolti

La formulazione delle domande e delle risposte (aperte/chiuse; dicotomiche/multiple; cartellini; a scale; ordinamenti, ... ) è fondamentale e dipende dalla tecnica di rilevazione prescelta

I sistemi CAI consentono, inoltre, la strutturazione di un "questionario intelligente", che cerca di controllare la **coerenza** formale e sostanziale delle risposte fornite dagli intervistati

# il questionario

struttura

1 Da un punto di vista generale può essere considerato come un insieme di domande cui è imposta una struttura logica, definita in modo che il soggetto intervistato possa fornire risposte rilevanti e consistenti ai quesiti posti

vincoli

2 I problemi dell'intervista e della formulazione del questionario richiedono indubbiamente un patrimonio di conoscenze tecniche, ma esigono soprattutto uno sforzo costante d'analisi empirica e verifica applicativa, in condizioni di rilevazione sempre nuove

Formulazione  
dei quesiti



Linguaggio

Ogni **domanda** è uno **stimolo** che sollecita una **risposta**: affinché ci sia omogeneità e confrontabilità nelle risposte, è necessario che i quesiti siano formulati in modo da produrre in soggetti simili un medesimo stimolo

Il **linguaggio** usato svolge un ruolo fondamentale: è necessario utilizzare un **vocabolario** ed un **lessico** che permettano la **comunicazione**, usando termini **semplici, non ambigui e comprensibili** a tutte le unità rispondenti (**wording**) e costruendo quesiti non lunghi o "pesanti" per le troppe precisazioni (**phrasing**)

Quesiti  
condizionanti



Attenzione alle **domande "condizionanti"**:

nelle domande riguardanti argomenti come onestà, giustizia, dovere, spesso all'intervistato è indirettamente suggerita l'esistenza di una risposta socialmente accettabile

Informazioni  
sensibili



Attenzione alle **domande su temi "sensibili"**:

i soggetti intervistati in genere tendono a eludere i quesiti riguardanti la *sessualità*, il *reddito*, il *patrimonio personale*, le *malattie* e così via

# Schema di campionamento:

*« L'insieme delle decisioni prese per formare il campione »*

- *Regole per la selezione*
- *Struttura del campione*
- *Determinazione della numerosità campionaria*
- *Probabilità di inclusione (nei campionamenti probabilistici)*