Analisi dei dati:

Analisi bivariata

Analisi bivariata dei dati

- Primo passo:
- Analisi bivariata dei dati quando le due variabili analizzate sono entrambe categoriali
- Analisi delle **TABELLE DI CONTINGENZA**
- Le avevamo già viste in una delle prime lezioni

Tabella di contingenza

- Se si dispone di due variabili (o caratteri) su scala nominale, A e B, le frequenze assolute, conteggiate SIMULTANEAMENTE per i due caratteri, si possono organizzare in una tabella di questo tipo, che è detta TABELLA DI CONTINGENZA:

Tabella di contingenza

	B_1	B_2	…	B_j	…	B_k	
A_1	n_{11}	n_{12}	…	n_{1j}		n_{1k}	$n_{1\cdot}$
A_2	n_{21}	n_{22}	…	n_{2j}		n_{2k}	$n_{2\cdot}$
…							
…							
A_i	n_{i1}	n_{i2}	…	n_{ij}		n_{ik}	$n_{i\cdot}$
…							
…							
…							
A_h	n_{h1}	N_{h2}	…	n_{hj}	…	n_{hk}	$n_{h\cdot}$
	n_1	$n_{\cdot 2}$	…	$n_{\cdot j}$	…	$n_{\cdot k}$	N

Riprendiamo l'esempio già visto

- Abbiamo valutato, in un **campione** di studenti iscritti al III anno di **quattro licei scientifici**, se esista (o meno) una relazione tra il **rendimento scolastico** (espresso dal voto medio di fine anno in tutte le materie) e il **livello di istruzione del capofamiglia**.

- La ricerca, svoltasi su un campione di **180** studenti, ha dato i seguenti risultati (espressi appunto da una tabella di contingenza):

	Lic.media o inferiore	Diploma	Laurea o sup.	TOTALE
Rendimento scarso-insufficiente (3-4-5)	16	12	6	**34**
Rendimento sufficiente (6-7)	30	55	13	**98**
Rendimento buono-ottimo (8-9-10)	3	23	22	**48**
TOTALE	**49**	**90**	**41**	**180**

C’è relazione tra i due fenomeni?

- Il nostro problema è quello di valutare la **relazione** (**connessione**) tra i due fenomeni (**rendimento scolastico** degli allievi e **livello di istruzione della famiglia**), ovvero di valutare in che misura le due variabili qualitative considerate ***non sono indipendenti***.

- A tale scopo, si rende necessario stabilire come <u>*sarebbe*</u> conformata la distribuzione nel caso, <u>teorico</u>, di ***indipendenza*** tra le due variabili.

La regola del prodotto

- Una regola del calcolo delle probabilità, detta “regola del prodotto”, afferma che la probabilità del <u>verificarsi contemporaneo</u> di due eventi tra loro <u>**indipendenti**</u> è pari al prodotto delle singole probabilità dei due eventi:
- $Pr(A \cap B) = Pr(A)*Pr(B)$

Regola del prodotto e tabelle di contingenza

- Possiamo applicare la <u>**regola del prodotto**</u> alle tabelle di contingenza: il verificarsi contemporaneo di due modalità di A e B, **se le variabili A e B sono indipendenti**, ha probabilità pari al prodotto delle probabilità del verificarsi delle singole modalità di A e B.

- Come conseguenza dell'applicazione della regola del prodotto, è possibile ricavare le **frequenze assolute <u>teoriche</u>** (attese) nel caso di <u>***indipendenza stocastica***</u> tra le due variabili.

Regola del prodotto e tabelle di contingenza

- Le frequenze assolute <u>attese</u> si possono ottenere coi seguenti passaggi:

$$f_{ij}^* = f_{i\cdot} \times f_{\cdot j} = \frac{n_{i\cdot} \times n_{\cdot j}}{n^2}$$

$$n_{ij}^* = n f_{ij}^* = n\left(\frac{n_{i\cdot} \times n_{\cdot j}}{n^2}\right) = \frac{n_{i\cdot} \times n_{\cdot j}}{n}$$

Calcolo delle frequenze attese

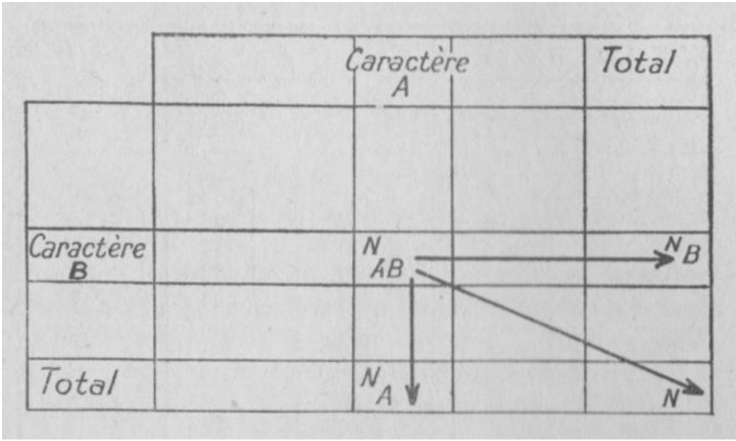
- Si ha pertanto:

$$FA_{ij} = \frac{FO_{i\cdot} \times FO_{\cdot j}}{n}$$

Calcolo delle frequenze attese

- Dove $FO_{i\cdot}$ e $FO_{\cdot j}$ sono, rispettivamente, i totali **di riga** e **di colonna** (i cosiddetti **totali marginali**) della tabella di contingenza, mentre ***n*** è la numerosità totale del campione.
- Il calcolo delle frequenze "attese" secondo l'ipotesi di indipendenza stocastica avviene moltiplicando tra loro i due totali marginali, dividendo il tutto per *n*; può essere utile lo schema rappresentato nella figura che segue:

Calcolo delle frequenze attese



Calcolo delle frequenze attese

- Utilizzando la formula appena vista, si può "riscrivere" la tabella di contingenza, come se fosse la tabella "attesa" che si otterrebbe nel caso di TOTALE INDIPENDENZA (indipendenza stocastica) tra le due variabili nominali.

- Quella che segue è la tabella teorica, calcolata nel modo appena visto.

Tabella delle frequenze attese: i valori in rosso sono i valori "attesi"

	Lic.media o inferiore	Diploma	Laurea o sup.	TOTALE
Rendimento scarso-insufficiente (3-4-5)	9,3	17,0	7,7	**34**
Rendimento sufficiente (6-7)	26,7	49,0	22,3	**98**
Rendimento buono (8-9)	13,1	24,0	10,9	**48**
TOTALE	**49**	**90**	**41**	**180**

Una misura della dipendenza

- Quella che stiamo cercando è una misura della "dipendenza in distribuzione" (stocastica) tra i due caratteri da noi considerati.
- Una misura di questo tipo deve essere basata sulla <u>differenza</u> tra le frequenze – <u>osservate</u> – della tabella empirica e le frequenze – <u>attese</u> – della tabella teorica.
- È stata messa a punto una misura di questo tipo che si chiama "chi quadrato": se si dispone di una tabella di contingenza formata da *p* righe e *k* colonne, tale misura si ottiene applicando la formula che segue :

Il "chi quadrato"

$$\chi^2 = \sum_{i=1}^{p} \sum_{j=1}^{k} \frac{\left(FO_{ij} - FA_{ij}\right)^2}{FA_{ij}} =$$

$$= \sum_{ij}^{p \times k} \frac{\left(FO_{ij} - FA_{ij}\right)^2}{FA_{ij}}$$

- Dove :
- FO_{ij} è la frequenza osservata (empirica) che si trova nella ij-esima casella della tabella;
- FA_{ij} è la frequenza attesa (teorica) che si trova nella ij-esima casella.

Calcolo delle differenze

	Lic.media o inferiore	Diploma	Laurea o sup.	TOTALE
Rendimento scarso-insufficiente (3-4-5)	16-9,3	12-17	6-7,7	**34**
Rendimento sufficiente (6-7)	30-26,7	55-49	13-22,3	**98**
Rendimento buono-ottimo (8-9-10)	3-13,1	23-24	22-13,9	**48**
TOTALE	**49**	**90**	**41**	**180**

Con i dati del nostro esempio, il valore del "chi quadrato" è pari a **30,82**.

Il "phi quadrato" di Pearson

- Questa misura, però, ha il difetto di dipendere fortemente dalla <u>**numerosita'**</u> del campione esaminato.
- Una misura <u>relativa</u> della dipendenza distributiva, che permette di confrontare distribuzioni di diversa numerosità (ad esempio, due campioni provenienti da regioni o paesi diversi), è la seguente :

Il "phi quadrato" di Pearson

$$\phi^2 = \frac{\chi^2}{n} = \frac{1}{n}\sum_{ij}^{p\times k}\frac{\left(FO_{ij} - FA_{ij}\right)^2}{FA_{ij}}$$

- Con i dati del nostro esempio, abbiamo:
- Φ^2= 30,82 / 180 = **0,1712**
- Possiamo, quindi, concludere che <u>**vi è dipendenza**</u> tra rendimento scolastico e livello di istruzione del capofamiglia della famiglia di origine.
- Il "phi quadrato" assume valore 1 in caso di dipendenza perfetta solo con tabelle di dimensioni pX2 o 2Xk;
- Diversamente, il valore in caso di dipendenza perfetta è pari a:
- min[(p-1);(k-1)]

La T di Tschuprov

Un'altra misura <u>relativa</u> della connessione si ottiene nel seguente modo:

$$T = \frac{\phi^2}{\sqrt{(p-1)(k-1)}}$$

La T di Tschuprov ha un vantaggio: assume valore **1** in caso di dipendenza perfetta quando la tabella di contingenza è QUADRATA.

Il rapporto tra odds (Odds Ratio: OR) o "rapporto crociato"

- Una misura di associazione molto importante è il Rapporto tra Odds (**OR**), in alcuni testi chiamato anche "**<u>rapporto crociato</u>**".
- Dato un certo insieme di individui, suddiviso **<u>dicotomicamente</u>** in due parti: coloro che sono ammalati, (o hanno una certa <u>condizione</u>, A) e coloro che non sono ammalati (NA), si dice **<u>ODDS</u>**, o **<u>rapporto di probabilità</u>**, il rapporto tra la probabilità di essere ammalati (o di avere una certa <u>condizione</u>, A) e la probabilità dell'evento complementare, cioè di non essere ammalati (NA, o di NON avere una certa <u>condizione</u>, A):

Il rapporto di probabilità (Odds)

- Odds = P(A)/P(NA)=P(A)/(1-P(A))
- Poiché, con semplici passaggi, si ha:

$$Odds = \frac{P(A)}{P(NA)} = \frac{\dfrac{\text{Ammalati}}{\text{Popolazione a rischio}}}{\dfrac{\text{Non ammalati}}{\text{Popolazione a rischio}}} = \frac{\text{Ammalati}}{\text{Non ammalati}}$$

- Risulta evidente che per calcolare l'odds non è necessario conoscere la consistenza numerica della popolazione a rischio, ma è sufficiente sapere quanti sono ammalati e quanti no.

Il rapporto tra odds (Odds Ratio: OR) o "rapporto crociato"

- Questo per quanto riguarda gli odds; se, però, vogliamo confrontare ciò che accade a un insieme di individui (es. gli ESPOSTI a un certo Fattore Di Rischio) con ciò che accade a un altro insieme di individui (es. i NON ESPOSTI), possiamo fare il rapporto tra i due ODDS (Odds Ratio):

$$OR = \frac{Odds(ESPOSTI)}{Odds(NonESPOSTI)} = \frac{[\text{Ammalati/Non ammalati}]_{\text{esposti}}}{[\text{Ammalati/Non ammalati}]_{\text{non esposti}}}$$

Il rapporto tra odds (Odds Ratio: OR)

- In altri termini, se abbiamo una tabella a doppia entrata del tipo:

	CASI (ammalati)	CONTROLLI (non ammalati)	TOTALE
Esposti al FDR	*A*	*B*	A+B
Non esposti al FDR	*C*	*D*	C+D
TOTALE	A + C	B + D	A+B+C+D

- Si ha che:
- Odds(E) = A/B
- Odds(NE) = C/D
- OR = Odds(E)/Odds(NE)= (A/B)/(C/D)= A*D/B*C

Il rapporto tra odds (Odds Ratio: OR)

- L'aspetto più interessante dell'OR consiste nel fatto che è possibile calcolarlo anche se NON SI CONOSCE l'entità numerica della <u>popolazione a rischio</u> di riferimento;
- L'OR trova uno dei suoi utilizzi più interessanti negli studi epidemiologici di tipo "caso-controllo".

Un esempio di OR.
Uno studio caso-controllo: esiste una relazione tra vaccinazione contro il morbillo (durante l'infanzia) e malattie infiammatorie dell'intestino, IBD (in età adulta)?

A case-control study of measles vaccination and inflammatory bowel disease

*Mark Feeney, Andrew Clegg, Paul Winwood, Jonathon Snook, for the East Dorset Gastroenterology Group**

Summary

Background The cause of inflammatory bowel disease (IBD) remains to be established. Evidence has linked measles infection in early childhood with the subsequent risk of developing IBD, particularly Crohn's disease. A cohort study raised the possibility that immunisation with live attenuated measles vaccine, which induces active immunity to measles infection, might also predispose to the later development of IBD, provoking concerns about the safety of the vaccine.

Method We report a case-control study of 140 patients with IBD (including 83 with Crohn's disease) born in or after 1968, and 280 controls matched for age, sex and general practitioner (GP) area, designed to assess the influence of measles vaccination on later development of IBD. Documentary evidence of childhood vaccination history was sought from GP and community health records.

Findings Crude measles vaccination rates were 56·4% in patients with IBD and 57·1% among controls. Matched odds ratios for measles vaccination were 1·08 (95% CI 0·62–1·88) in patients with Crohn's disease, 0·84 (0·44–1·58) in patients with ulcerative colitis, and 0·97 (0·64–1·47) in all patients with IBD.

Interpretation These findings provide no support for the hypothesis that measles vaccination in childhood predisposes to the later development of either IBD overall or Crohn's disease in particular.

Lancet 1997; **350:** 764–66

Introduction

The possibility that the inflammatory bowel diseases (IBDs)—ulcerative colitis and Crohn's disease—are caused by a transmissible agent such as a virus is an attractive hypothesis.[1] Wakefield and colleagues have suggested that Crohn's disease might be the late result of measles virus infection at a critical time during early childhood. This "measles hypothesis" is based on a series of pathological and epidemiological studies.[2-4]

Wild-type measles infection is associated with substantial morbidity and mortality. In the developed world, complications occur in about one in 15 infections, and most deaths result from the development of pneumonia, acute encephalitis, or the rare but relentlessly progressive subacute sclerosing panencephalitis.[5,6] Live attenuated measles vaccine was introduced in the UK in 1968, and as a result of the vaccination campaign the incidence of measles infection and complications has fallen strikingly.[5,6]

The measles hypothesis has been embellished with evidence from a cohort study suggesting an increased risk of IBD in individuals given live attenuated measles vaccine in early childhood.[7] This report has led to concern about the safety of measles vaccination and resulted in some parents declining an effective vaccine. Counselling has been particularly difficult because the evidence on which to base reassurance on this issue is very limited.[8]

The present investigation was devised to assess the risk of IBD associated with vaccination against measles in early childhood. A case-control design was used because of the relative rarity of the disease and the frequency of vaccine exposure in the population.[9,10] The study focuses on the period between 1968 and 1991, during which measles vaccination was routinely offered and immunisation status

Lo studio dimostra che non c'è nessuna relazione tra i due fenomeni.
Andiamo, infatti, a vedere gli OR (*odds ratio*):

	Crohn's disease	Ulcerative colitis	IBD overall
Measles	1·08 (0·62–1·88)	0·84 (0·44–1·58)	0·97 (0·64–1·47)
Pertussis	··	··	1.00 (0·62–1·62)
Diphtheria/tetanus	··	··	0.94 (0·39–2·26)

Table 3: **Odds ratios (95% CI) for IBD in vaccine recipients**

Un’applicazione pratica degli Odds: il mondo delle scommesse

- È probabile che abbiate già incontrato gli odds;
- Sono infatti **odds**, ovvero rapporti di probabilità, quelle che i bookmakers chiamano “quote”;
- Vediamo il seguente esempio, tratto da un sito di “comparazioni” di quote:

Un’applicazione pratica degli Odds: il mondo delle scommesse

Quote Calcio Scommesse

BetClic.it 20€ offerti !

Per vincere di più con le tue scommesse Calcio, confronta le quote di tutti i bookmakers sugli incontri sportivi di Calcio. Scommetti con la quota migliore!

Tutti | In diretta | Oggi | Domani | Risultati

Calcio | Gli ultimi risultati | Tutte le partite

Calcio > Italia > Serie A

Martedì 25 ottobre	Quote migliori	Migliori vincite	
20:45 Juventus - Fiorentina			
Juventus	1,75 su BetClic.it	Vinci 17,5€	Confronta su 7 bookmakers
Pareggio	3,60 su UNIBET	Vinci 36€	
Fiorentina	5,50 su betfair	Vinci 55€	
Mercoledì 26 ottobre	**Quote migliori**	**Migliori vincite**	
20:45 Napoli - Udinese			
Napoli	2,15 su intralot	Vinci 21,5€	Confronta su 7 bookmakers
Pareggio	3,25 su BetClic.it	Vinci 32,5€	
Udinese	4,00 su betfair	Vinci 40€	
20:45 Milan - Parma			
Milan	1,35 su BetClic.it	Vinci 13,5€	Confronta su 7 bookmakers
Pareggio	5,00 su betfair	Vinci 50€	
Parma	10,00 su betfair	Vinci 100€	
			Vedi più partite

Un’applicazione pratica degli Odds: il mondo delle scommesse

- Che cosa significano quei numeri: **1,75**; **3,60**; **5,50** ecc.?
- Forse lo sapete: se avete scommesso 10 Euro, nel caso si verifichi l’evento sul quale avete puntato vincerete, rispettivamente, 17,5; 36; 55 Euro.
- Se, invece, l’evento non si verificherà (come spessissimo accade), avrete perso i vostri 10 Euro (ricordate che il banco vince sempre!)

Un’applicazione pratica degli Odds: il mondo delle scommesse

- Ma come sono nate quelle “quote”?
- Gli esperti (esiste una apposita figura professionale: il “quotista”) hanno valutato le due **probabilità**:
- Quella **dell’evento avverso** (la Fiorentina non vince: 0,8461)
- Quella **dell’evento favorevole** (la Fiorentina vince: sarà 1-0,8461= 0,1539)

Un’applicazione pratica degli Odds: il mondo delle scommesse

- Il rapporto tra le due probabilità:
- Evento avverso (la Fiorentina perde o pareggia)
- **P=0,8461**
- Evento favorevole (la Fiorentina vince)
- **1-P=1-0,8461=0,1536**
- È un odds:
- Odds(evento avverso)=0,8461/0,1536=5,5
- Che è appunto la “quota” che il bookmaker assegna alla vittoria della Fiorentina

Ci sono anche situazioni più equilibrate (tra i due eventi)

Calcio > Italia > Serie B

Lunedì 24 ottobre	Quote migliori		Migliori vincite	
20:45 Gubbio - Torino				
Torino	1,85 su	EUROBET	Vinci 18,5€	Confronta su 7 bookmakers
Pareggio	3,40 su	bwin	Vinci 34€	
Gubbio	5,25 su	betfair	Vinci 52,5€	
Venerdì 28 ottobre	**Quote migliori**		**Migliori vincite**	
19:00 Ascoli - Modena				
Ascoli	2,15 su	bwin	Vinci 21,5€	Confronta su 1 bookmakers
Pareggio	2,95 su	bwin	Vinci 29,5€	
Modena	3,60 su	bwin	Vinci 36€	
21:00 Bari - Pescara				
Bari	2,10 su	bwin	Vinci 21€	Confronta su 1 bookmakers
Pareggio	3,30 su	bwin	Vinci 33€	
Pescara	3,30 su	bwin	Vinci 33€	
				Vedi più partite

Come posso confrontare tra loro due scommesse?

- È facile: posso calcolare il rapporto tra due "quote" (odds) relative a due possibili esiti di una partite:
- Che rapporto c'è, ad esempio, tra la vittoria del Parma (VS Milan) e la vittoria dell'Udinese (VS Napoli)?
- Calcoliamo il rapporto tra gli odds (Odds Ratio):
- OR = Odds(Parma)/Odds(Udinese)=
- = 10,00/4,00 = 2,5
- Tradotto: se scommetto sul Parma, guadagno 2,5 volte di più di quanto guadagnerei scommettendo sull'Udinese.

Un altro sito che propone quote di diversi bookmakers

Martedì 25 Ottobre 2011					1	X	2
■ Serie A : Giornata n.9							
20:45	Juventus	-	Fiorentina		1,70	3,50	5,25

Mercoledì 26 Ottobre 2011					1	X	2
■ Serie A : Giornata n.9							
20:45	Atalanta	-	Inter		3,40	3,10	2,15
	Cesena	-	Cagliari		2,60	3,10	2,65
	Chievo	-	Bologna		1,78	3,20	5,00
	Genoa	-	Roma		2,55	3,20	2,65
	Lazio	-	Catania		1,62	3,60	5,50
	Milan	-	Parma		1,31	4,75	9,50
	Novara	-	Siena		2,40	3,10	2,90
	Napoli	-	Udinese		2,05	3,20	3,60

TRAPPOLE NELLE TABELLE DI CONTINGENZA: IL PARADOSSO DI SIMPSON

- Vediamo il seguente esempio
- In due diversi ospedali, su due campioni di pazienti, sono stati sperimentati due diversi trattamenti farmacologici, A e B. Gli esiti possibili potevano essere: MIGLIORAMENTO (M) oppure NESSUNA VARIAZIONE O PEGGIORAMENTO (P)

Consideriamo la tabella di contingenza α, relativa al primo ospedale

Esito	Peggiorati o n.v.	Migliorati	Totale
Trattamento			
A	160	40	**200**
B	170	30	**200**
Totale	**330**	**70**	**400**

- **Consideriamo adesso la tabella di contingenza β, relativa al <u>secondo</u> ospedale:**

Esito	Peggiorati o n.v.	Migliorati	Totale
Trattamento			
A	15	85	**100**
B	100	300	**400**
Totale	**115**	**385**	**500**

- Vogliamo valutare, nel primo ospedale, se c'è differenza, nella quota di miglioramento, tra il trattamento A e il trattamento B

- Calcoliamo quindi le relative percentuali (di "miglioramenti" sul totale):
- TRATTAMENTO A : quota di miglioramenti = 40/200 = 0,2 = 20%
- TRATTAMENTO B : quota di miglioramenti = 30/200 = 0,15 = 15%

- Si ha quindi, nell'ospedale "ALFA":

QUOTA (A) = 20% > QUOTA (B) = 15%

- Facciamo ora la stessa valutazione per quanto riguarda l'ospedale "BETA"
- TRATTAMENTO A : quota di miglioramenti = 85/100 = 0,85 = 85%
- TRATTAMENTO B : quota di miglioramenti = 300/400 = 0,75 = 75%
- SI HA QUINDI, nell'ospedale "BETA":

QUOTA (A) = 85% > QUOTA (B) = 75%

- Possiamo quindi concludere che le indicazioni emerse dai due ospedali siano sostanzialmente **molto simili**:
- **In entrambi i casi la quota del trattamento "A" è stata più alta della quota del trattamento "B".**
- Visto che le indicazioni emerse dalle due tabelle sono simili, ci viene naturale l'idea di "sommare" le due tabelle di contingenza, cioè costruire la distribuzione bivariata risultante dalla somma delle due distribuzioni originarie, "α" e "β";
- La nuova tabella così costruita ci permetterà di dare una valutazione complessiva.

TABELLA DI CONTINGENZA "$\alpha + \beta$"

Esito	Peggiorati o n.v.	Migliorati	Totale
Trattamento			
A	175	125	**300**
B	270	330	**600**
Totale	**445**	**455**	**900**

- Facciamo ora su questa tabella lo stesso conteggio fatto prima:
- TRATTAMENTO A : quota di miglioramenti = 125/300 = 0,42 = 42%
- TRATTAMENTO B : quota di miglioramenti = 330/600 = 0,55 = 55%

- SI HA QUINDI, prendendo la "somma" dei due ospedali:
- **QUOTA (B) = 55% > QUOTA (A) = 42%**

- **CHE È ESATTAMENTE <u>IL CONTRARIO</u> DEI DUE RISULTATI VISTI IN PRECEDENZA !**

- Questo effetto, apparentemente paradossale ("apparentemente", nel senso che tale risultato nasconde in realtà una "trappola" facilmente spiegabile) è noto come **"Paradosso di Simpson"**

ANCORA SUL PARADOSSO DI SIMPSON: UN ESEMPIO DI USO DISONESTO

- Vediamo adesso un altro esempio;
- Si vuole sperimentare l'effetto sulle vendite di un certo prodotto di una campagna pubblicitaria;
- Da un sondaggio effettuato, è noto che i contenuti del messaggio pubblicitario sono stati considerati "sgradevoli" da una certa quota delle persone intervistate (il messaggio era, o appariva, "politicamente scorretto").

La sperimentazione ha dato i seguenti risultati:

	HANNO ACQUISTATO	**NON HANNO ACQUISTATO**	***TOTALE***
CON PUBBLICITA'	200	200	**400**
SENZA PUBBLICITA'	160	240	**400**
TOTALE	**360**	**440**	**800**

% di acquisto con pubblicità: 50%
% di acquisto senza pubblicità: 40%

Quindi sembra che – nonostante la sgradevolezza del messaggio – la campagna abbia comunque avuto un certo successo.
Ora andiamo a vedere che cosa accade se separiamo il nostro campione in due parti: rispettivamente, in maschi e femmine:

MASCHI

	HANNO ACQUISTATO	NON HANNO ACQUISTATO	*TOTALE*
CON PUBBLICITA'	180	120	**300**
SENZA PUBBLICITA'	70	30	**100**
TOTALE	**250**	**150**	**400**

% di acquisto con pubblicità: 60%
% di acquisto senza pubblicità: 70%

FEMMINE

	HANNO ACQUISTATO	NON HANNO ACQUISTATO	*TOTALE*
CON PUBBLICITA'	20	80	**100**
SENZA PUBBLICITA'	90	210	**300**
TOTALE	**110**	**290**	**400**

% di acquisto con pubblicità: 20%
% di acquisto senza pubblicità: 30%

- A sorpresa, quindi, gli **stessi** dati – una volta considerati separatamente per maschi e femmine, danno risultati **completamente opposti** a quelli visti per il campione complessivo;

- Questo fenomeno è dovuto alla scelta – **evidentemente disonesta** – del "protocollo" sperimentale: durante la sperimentazione, i maschi sono stati sottoposti al messaggio pubblicitario **in misura molto maggiore** rispetto alle femmine (300 maschi **vs** 100 femmine);

- E questo perché ai disonesti sperimentatori era noto che, per i maschi, il messaggio pubblicitario era risultato essere **molto meno sgradevole** che per le femmine.