

**Matched groups estratti da popolazioni 2 x 2 limitate.
Condizionamento dei parametri sul risultato di un test e
intervalli di confidenza della differenza tra proporzioni.**

Enzo D'Arcangelo - Francesco Gori

1 - Premessa

In un precedente lavoro ([1] F. Gori), nel caso di due campioni indipendenti estratti da popolazione dicotomica limitata, veniva mostrato come il risultato di un test poteva essere condizionato dallo "stato di natura" π della popolazione stessa. In un successivo lavoro ([2] E. D'Arcangelo, F. Gori) veniva evidenziato il condizionamento, da parte del parametro π , sul risultato di alcuni test basati su confronti multipli tra risultati sperimentali, ottenuti da k gruppi indipendenti estratti dalla suddetta popolazione.

Il lavoro attuale fa seguito ad uno precedente ([3] F. Gori, L. Leone) e tratta la differenza tra proporzioni nel caso di campioni correlati, di ampiezza n limitata, utilizzando coppie di osservazioni su scala nominale, estratte da popolazioni bivariate di grandezza N limitata e per ρ e π variabili.

E' una procedura alternativa al test di Mc Nemar⁽¹⁾ che, come noto, è applicabile sia in ricerche aventi un aspetto descrittivo con variabile indipendente non manipolabile dal ricercatore, sia nella sperimentazione vera e propria in cui tale variabile (fattore e trattamenti) è manipolabile.

Una proporzione (n° successi/totale osservazioni) è un caso particolare di media, pertanto, analogamente alla differenza tra medie⁽²⁾, viene considerata la variabile $d = n(p_x - p_y) = n_{1.} - n_{.1}$, differenza tra i numeri dei successi (numeratori delle proporzioni).

Definita la popolazione tetracorica (due variabili x e y con alternative 0 e 1) sotto $H_0: \pi_x = \pi_y | \rho$, e data la numerosità n del campione, viene costruita la distribuzione campionaria delle differenze d : modello probabilistico di riferimento. La distribuzione risulta discontinua. Fissato α , e data una differenza sperimentale d_s , se ne determina la significatività.

In una prima fase del presente lavoro si vuole evidenziare il ruolo svolto dai parametri ρ e π caratterizzanti la popolazione in merito al loro condizionamento sul risultato del test, e per riflessioni sulle indicazioni che detto risultato può dare circa i valori dei predetti parametri.

⁽¹⁾ Il test di Mc Nemar si basa sulla significatività dei cambiamenti nelle frequenze in una tabella di contingenza: o delle risposte Prima-Dopo o tra caratteristiche, opinioni, di coppie di soggetti *legati* tra loro, quindi risposte *legate* tra loro; confronta le due frequenze (assolute) dei cambiamenti con la frequenza ottenuta dalla loro semisomma e utilizza il chi quadro per determinarne la significatività. L'unità statistica può essere costituita da un solo soggetto sul quale si rilevano i punteggi di uno stesso carattere prima e dopo un trattamento, e si hanno due punteggi: x ed y .

Oppure l'unità statistica può essere costituita da una coppia di soggetti (coppie appaiate) simili quanto più possibile secondo alcune caratteristiche essenziali, non oggetto di indagine ma che potrebbero influire sulla "risposta", come età, genere, istruzione, ambiente, ad es.: fratelli gemelli, scolari di pari capacità, genitore-figlio, moglie-marito e così via. Su ciascuno dei soggetti della coppia si rileva il punteggio del carattere che interessa, quindi per ogni unità statistica, anche in questo caso, si hanno due punteggi x ed y .

⁽²⁾ Nell'estrazione di campioni da una popolazione normale bivariata, sotto $H_0: (\mu_x = \mu_y) | \rho$, si parte dalla distribuzione di tutte le possibili differenze $\bar{x} - \bar{y}$, da cui $z = \frac{\bar{x} - \bar{y}}{\sigma_{\bar{x}-\bar{y}}}$, essendo

$\sigma_{\bar{x}-\bar{y}}^2 = \sigma_x^2 + \sigma_y^2 - 2\rho_{xy} \sigma_x \sigma_y$ (parametri noti), oppure $t = \frac{\bar{x} - \bar{y}}{S_{\bar{x}-\bar{y}}}$ (parametri stimati).

Al variare dei due parametri vengono presentati esempi, con relative Tabelle del p-value, ottenuti dalla distribuzione della d . Tali risultati vengono confrontati con quelli dati dal test di Mc Nemar. Inoltre, partendo da una tabella Prima - Dopo, viene determinata l'espressione della variabile ${}_r t_{p_1-p_2}$ (v. Sez. 4 -) e il corrispondente p-value nel continuo, calcolato utilizzando la distribuzione della t di Student. I risultati vengono confrontati e commentati (v. Sez. 5 -).

Infine, utilizzando la $dp = p_x - p_y$, si affronta il problema della significatività mediante l'intervallo di confidenza della differenza tra i parametri: $\delta = \pi_x - \pi_y$, condizionata da ρ e, come in tutti i casi, al variare di n . Se l'intervallo contiene $\delta = 0$ non si può rifiutare H_0 .

Sono stati utilizzati programmi in Fortran ($2 \leq n \leq 30$), fornibili gratuitamente con spiegazioni.

2 - Aspetti teorici

Definiamo la popolazione in termini di frequenze (si intende relative) o proporzioni, valutazioni delle probabilità degli eventi considerati:

Tab. 1

	y	0	1	Tot
x				
0		π_{00}	π_{01}	$1 - \pi_x$
1		π_{10}	π_{11}	π_x
Tot		$1 - \pi_y$	π_y	1

Le proporzioni dei successi in totale sono π_x, π_y , e $\rho = (\pi_{11} - \pi_x \pi_y) [\pi_x (1 - \pi_x) \pi_y (1 - \pi_y)]^{-0.5}$.

Definiamo la popolazione in termini di frequenze, sotto H_0 .

$H_0: \pi_x = \pi_y = \pi \mid \rho = \rho_0$, in cui $-1 < \rho_0 < 1$ (in un precedente lavoro: [3] F. Gori, L. Leone, per $\rho_0 = 0$) e $0 < \pi < 1$, contro $H_1: \pi_x \neq \pi_y$. Pertanto l'operatore dovrebbe conoscere o stabilire le caratteristiche dell'ambiente in cui si svolge l'indagine, ossia lo stato di natura: ρ e π .

Dato un esito sperimentale, mostreremo, mediante alcuni esempi, come risulti differente il p-value al variare dei due parametri, aspetto non evidenziabile con i test di Mc Nemar e ${}_r t_{p_1-p_2}$.

Per motivi di compatibilità tra ρ e π ⁽³⁾ è necessario fissare il valore di ρ , e successivamente il

⁽³⁾ Sotto H_0 , $\pi_x = \pi_y = \pi$ e $\pi_{10} = \pi_{01}$. Dalla Tab. 1, risulta: $\rho = (\pi_{00} \pi_{11} - \pi_{10}^2) / \pi(1 - \pi)$. Per la non negatività delle frequenze interne deve essere: $\pi_{00} \leq (1 - \pi)$, $\pi_{10} \leq (1 - \pi)$, $\pi_{10} \leq \pi$, $\pi_{11} \leq \pi$. Dall'espressione: $\rho \pi(1 - \pi) = \pi_{00} \pi_{11} - \pi_{10}^2$, da cui $\pi_{10}^2 = (1 - \pi - \pi_{10})(\pi - \pi_{10}) - \rho \pi(1 - \pi)$, dopo alcuni passaggi si ha: $\pi(1 - \pi)(1 - \rho) = \pi_{10} \leq \pi$ (condizione); ed infine $\pi \geq -\rho/(1 - \rho)$.

Se ρ è positivo non ci sono problemi, ma se ρ è negativo, ad es. -0.5, abbiamo $\pi \geq 0.5/1.5$; quindi π non può essere inferiore a 0.3333.

Oppure: $\pi(1 - \pi)(1 - \rho) \leq \pi$; $\pi - \pi^2 - \rho \pi + \rho \pi^2 \leq \pi$; $-\pi^2 - \rho \pi + \rho \pi^2 \leq 0$; $\pi^2(\rho - 1) - \rho \pi \leq 0$; $(1 - \rho)\pi^2 + \rho \pi \geq 0$; con radici $\pi = 0$ e $\pi \geq -\rho/(1 - \rho)$.

Lo stesso discorso si può fare per $1 - \pi$, infatti la espressione di partenza è simmetrica rispetto a π e $1 - \pi$. E ad es. in definitiva per $\rho = -0.5$ si ha: $0.3333 \leq \pi \leq 0.6667$.

valore di $\pi|\rho$, in sostanza se ρ è negativo il valore di π è vincolato ad un intervallo.

La procedura fa riferimento alla distribuzione campionaria data la popolazione sotto H_0 ; ed occorre costruire tutte le possibili tabelle di contingenza in termini di frequenze assolute, del tipo in Tab. 2:

Tab. 2

		Campione		
		Risposte figli		
x \ y		NO	SI	Tot.
Risposte genitori	NO	n_{00}	n_{01}	$n_{0.}$
	SI	n_{10}	n_{11}	$n_{1.}$
Tot.		$n_{.0}$	$n_{.1}$	n

Per comodità di scrittura, nell'indicare i caratteri vengono utilizzate le lettere x ed y (anziché x_1 ed x_2), ad es. x per le risposte dei genitori: NO o SI, ed y per le risposte dei figli: NO o SI. Oppure risposta ad un quesito da parte delle mogli (x) e dei mariti (y). Presenza della madre, per provare paura di bambini nel cimentarsi in una nuova esperienza (x: NO o SI); non presenza della madre, per gli stessi bambini, per provare paura (y: NO o SI) ecc. In generale, risposte Prima-Dopo un trattamento. Nel caso di confronti tra proporzioni o numero di successi, il carattere ovviamente deve essere lo stesso, altrimenti la differenza non ha senso (nel contesto attuale, NO vale 0 e SI vale 1). Ad ogni valore di $d = n_{1.} - n_{.1}$ corrispondono vari valori di $r = (n_{00}n_{11} - n_{10}n_{01}) / (n_{1.}n_{2.}n_{.1}n_{.2})^{0.5}$.

Poniamo che le tabelle-campione, Tab. 2, abbiano senso e significato, in sostanza siano valide, se è possibile calcolare d ed r (relazione tra x ed y). Ciò premesso **non si ritiene di escludere** i valori di d nei casi in cui sembra non sia possibile calcolare r ($= 0/0$, forma indeterminata); ad es. se in Tab. 2 gli n soggetti figli danno tutti la stessa risposta dei genitori, o tutti risposta diversa, una sola casella interna risulta $\neq 0$ e non è possibile calcolare r con la nota espressione. In Appendice I viene mostrato che per i quattro casi in cui un $n_{ij} = n$, con un passaggio al limite, i valori di r sono: 0.5, -0.5, 0.5, -0.5. Come poteva essere valutato anche al livello **intuitivo**, dal momento che solo una casella delle diagonali è interessata. Ed anche i casi in cui: Prima sono tutti concordi per il SI, Dopo alcuni cambiano opinione, o viceversa: con il passaggio al limite risulta $r = 0$.

Si potrebbe discutere sul valore numerico dei risultati in Appendice I (o meglio sulle velocità di convergenza nelle varie caselle), ma comunque sarebbe dannoso per il ricercatore escludere il valore di d fornito da tali tabelle in quanto non è possibile determinare r con la nota formula, trascurando informazioni così ricche di significato.

Dato $n = 30$ il totale delle tabelle di contingenza nell'universo dei campioni non supera:

$$\binom{n+4-1}{n} = \binom{33}{30} = \binom{33}{3} = 5456$$

Come nel caso di estrazione con reimmissione.

Per ciascuna tabella campionaria viene calcolata d e la frequenza (ed anche la r , stima di ρ e la $p = (n_{1.} + n_{.1})/2n$, stima di π), come segue.

3 - Frequenza dei campioni nel caso di estrazione con e senza reimmissione

Nel caso di estrazione con reimmissione, d'ora in poi e. c. r., la frequenza (nell'universo dei campioni) corrispondente ad una Tab. 2, utilizza i dati della popolazione in Tab. 1 ed è data dalla seguente espressione della multinomiale:

$$fr(n_{00}, n_{01}, n_{10}, n_{11}) = \frac{n!}{n_{00}!n_{01}!n_{10}!n_{11}!} \pi_{00}^{n_{00}} \pi_{01}^{n_{01}} \pi_{10}^{n_{10}} \pi_{11}^{n_{11}}, \quad (1)$$

Nel caso di estrazione senza reimmissione, d'ora in poi e. s. r., definiamo la popolazione in termini di frequenze assolute:

Tab. 3

	y	0	1	Tot
x				
0		$N\pi_{00}$	$N\pi_{01}$	$N-N\pi_x$
1		$N\pi_{10}$	$N\pi_{11}$	$N\pi_x$
Tot		$N-N\pi_y$	$N\pi_y$	N

sotto $H_0 : N\pi_x = N\pi_y$. La frequenza del campione, corrispondente ad una Tab. 2, segue la distribuzione ipergeometrica multivariata:

$$fr(n_{00}, n_{01}, n_{10}, n_{11}) = \frac{\binom{N\pi_{00}}{n_{00}} \binom{N\pi_{01}}{n_{01}} \binom{N\pi_{10}}{n_{10}} \binom{N\pi_{11}}{n_{11}}}{\binom{N}{n}}. \quad (2)$$

Le distribuzioni - siamo nel discontinuo - della differenza tra i numeri dei successi $d = n_{1.} - n_{.1}$ si ottengono sommando le frequenze (1) oppure (2) a parità di d , nei due casi di estrazione.

L'e. s. r. è solitamente usata nelle scienze del comportamento. Comunque, sperimentalmente si ha che per popolazioni aventi numerosità $N \geq 10000$ i valori critici della variabile d , nel caso di e. s. r. sono quasi sempre coincidenti con quelli dell'e. c. r. (v. Esempio 1).

Un caso di popolazione sotto H_0 , limitatamente alle frequenze interne della Tab. 1 (qui vengono scritte una di seguito all'altra), potrebbe essere: 0.81, 0.09, 0.09, 0.01, di totale 1; un caso come in Tab. 3 potrebbe essere: 810, 90, 90, 10, di totale 1000. Per esse $\rho = 0$ e $\pi_x = \pi_y = \pi = .1$ ⁽⁴⁾.

Sotto H_0 , la distribuzione della variabile d è simmetrica e, se $\rho = 0$, nell'e. c. r. coincide con quella del caso di due campioni indipendenti, mentre ciò non accade nell'e. s. r. ([4] F. Gori). Nei due casi la media: $E(d) = E(n_{1.} - n_{.1}) = n\pi - n\pi = 0$. Nell'e. c. r. la varianza è: $\sigma_d^2 = 2n[\pi(1-\pi)][1-\rho]$, mentre nell'e. s. r. occorre ulteriormente moltiplicare per: $(N-n)/(N-1)$, in tal caso la standardizzata risulta: $z_d = (d-0)/\sigma_d = (n_{1.} - n_{.1}) / [2n(\pi - \pi^2)(1-\rho)(N-n)/(N-1)]^{-0.5}$, non trattata nel seguito.

⁽⁴⁾ In generale: $\pi_{11} = \pi_x \times \pi_y + \rho \sqrt{\pi_x \pi_y (1-\pi_x)(1-\pi_y)}$, e dati ρ , π_x e π_y , per determinare la struttura di una popolazione basta calcolare π_{11} .

I valori critici, strettamente legati all'andamento delle frequenze nelle code delle distribuzioni, assumeranno valori conseguenti a tali variabilità. Sono riportati alcuni esempi eseguiti mediante i programmi citati nella Premessa.

4 - La t nel caso di due proporzioni

Partiamo dalla nota espressione del confronto fra medie $\bar{x} - \bar{y}$ (scale ad intervalli) nel caso di due campioni correlati, posto $d_i = x_i - y_i$, sotto $H_0: \mu_x - \mu_y = 0$, la t di Student con $gl = n - 1$, è:

$$t = \frac{\bar{d}}{\sqrt{\frac{\sum d_i^2 - n\bar{d}^2}{n(n-1)}}} = \frac{\sum d_i / n}{\sqrt{\frac{\sum d_i^2 / n - (\sum d_i / n)^2}{n-1}}}; \quad 3)$$

Nel caso di piccoli campioni possiamo confrontare il p-value del discontinuo, ottenuto mediante le formule (1) o (2), con il p-value corrispondente al seguente test. Estendendo la formula 3) al caso di punteggi 0, 1; NO, SI; -, +; ecc., in un esperimento Prima - Dopo, elencati nella parte sinistra del prospetto (l'esempio è limitato a 12 soggetti per motivi di spazio), si ha:

Tab. 3 a
Risultati soggetto per soggetto

Soggetti			differenze	
	Prima	Dopo	d	d ²
1	0	1	-1	1
2	0	1	-1	1
3	1	0	1	1
4	0	1	-1	1
5	0	0	0	0
6	1	1	0	0
7	0	1	-1	1
8	1	0	1	1
9	0	1	-1	1
10	1	1	0	0
11	0	0	0	0
12	1	1	0	0

Tab. 3 b
Risultati raggruppati in tabella 2 x 2

		Dopo		Tot
		0	1	
Prima	0	a = 2	b = 5	7 a+b
	1	c = 2	e = 3	5 c+e
Tot		4 a+c	8 b+e	12 n

$$\begin{aligned} c+e = 5 & \quad b+e = 8 \\ \sum d_i = -3 & \quad \sum d_i^2 = 7, \text{ la media delle differenze, } \bar{d} = -3/12 = -0.25 \\ = c+e-b-e = c-b & \quad = c+b \end{aligned}$$

dalla 3) si ha:

$$t = \frac{-0.25}{\sqrt{\frac{0.5833 - 0.0625}{11}}} = \frac{-0.25}{\sqrt{0.04735}} = -1.1489.$$

Con i simboli della tabella Prima – Dopo, parte destra del prospetto, abbiamo:
 $p_1 = (c + e) / n$, $p_2 = (b + e) / n$ e $dp = p_1 - p_2 = (c - b) / n$. E dalla 3), dopo alcuni passaggi:

$${}_r t_{p_1-p_2} = \frac{(c-b)/n}{\sqrt{\frac{(c+b)/n - [(c-b)/n]^2}{n-1}}} = \frac{c-b}{\sqrt{\frac{n(c+b) - (c-b)^2}{n-1}}}; \quad 4)$$

Sostituendo i dati si ha: ${}_r t_{p_1-p_2} = \frac{-3}{\sqrt{\frac{84-9}{11}}} = -3 / 2.6112 = -1.1489$.

Dalla seconda delle 4), per n grande, $n - 1 \cong n$, e $(c - b)^2 / (n - 1)$ trascurabile, ${}_r t_{p_1-p_2}^2$ coincide con il $\chi^2 = (c-b)^2 / (c+b)$ del test di Mc Nemar. Il denominatore della prima delle 4), se anche $c \cong b$, dà l'errore standard: $(c + b)^{0.5} / n$, utilizzabile per determinare i limiti di confidenza approssimati, della differenza tra le frequenze relative ([5] P. Armitage, pag. 130, applicando la teoria normale).

5 - La significatività al variare di ρ e π

Per trattare tale argomento facciamo riferimento ad alcuni esempi evidenziando di volta in volta l'aspetto del condizionamento. Per semplicità si considerano test unidirezionali.

Come accennato (v. Sez. 1 -), nel continuo consideriamo il test di Mc Nemar che utilizza il χ^2 , basato sulla differenza $(c - b)^2$ e la ${}_r t_{p_1-p_2}$, basata sulla differenza $(c - b)$; mentre nel discontinuo viene usata la $d (= n_{1.} - n_{.1} = c - b)$, al variare di ρ e π .

Esempio 1

Un'iniziativa, rivolta a nuclei familiari, suscita consensi in seno a coppie Madri-Figlie? Poniamo $\alpha = 0.05$.

Tab. 4

		Figlie		Tot
		0	1	
Madri	0	14	6	20
	1	2	3	5
Tot		16	9	25

Applicando Mc Nemar si ha: $\chi_s^2 = 1.125$ (corretto), $gl = 1$, $p\text{-value} = 0.2890$ oppure $\chi_s^2 = 2$ (non corretto) e $p\text{-value} = 0.1574$.

Dalla (4) nel continuo si ha: ${}_r t_{p_1-p_2} = -1.4446$ con un $p\text{-value} = 0.0808$ (distribuzione t di Student).

Nel discontinuo, data la differenza sperimentale $d_s = -4$ (la coda interessata è la sinistra) abbiamo:

Tab. 5

Tabella del p-value al variare di ρ e π , per $n = 25$ (14, 6, 2, 3), data una popolazione come in Tab. 1 (e. c. r), e come in Tab. 3 (e. s. r.) per $N = 10000$, rispettivamente nella prima e seconda riga. I valori riportati in grassetto rappresentano i casi di significatività.

$\rho \setminus \pi$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.9	0.0001 0.0001	0.0005 0.0004	0.0011 0.0011	0.0018 0.0017	0.0020 0.0020	0.0018 0.0017	0.0011 0.0011	0.0005 0.0004	0.0001 0.0001
0.8	0.0007 0.0007	0.0042 0.0042	0.0090 0.0089	0.0127 0.0126	0.0140 0.0139	0.0127 0.0126	0.0090 0.0089	0.0042 0.0042	0.0007 0.0007
0.7	0.0025 0.0025	0.0127 0.0126	0.0240 0.0238	0.0316 0.0315	0.0343 0.0341	0.0316 0.0315	0.0240 0.0238	0.0127 0.0126	0.0025 0.0025
0.6	0.0059 0.0058	0.0248 0.0246	0.0424 0.0422	0.0534 0.0532	0.0570 0.0568	0.0534 0.0532	0.0424 0.0422	0.0248 0.0246	0.0059 0.0058
0.5	0.0108 0.0107	0.0388 0.0386	0.0615 0.0613	0.0747 0.0745	0.0790 0.0788	0.0747 0.0745	0.0615 0.0613	0.0388 0.0386	0.0108 0.0107
0.4	0.0169 0.0168	0.0534 0.0532	0.0799 0.0796	0.0945 0.0943	0.0992 0.0989	0.0945 0.0943	0.0799 0.0796	0.0534 0.0532	0.0169 0.0168
0.3	0.0240 0.0238	0.0678 0.0675	0.0969 0.0966	0.1124 0.1122	0.1173 0.1170	0.1124 0.1122	0.0969 0.0966	0.0678 0.0675	0.0240 0.0238
0.2	0.0316 0.0315	0.0815 0.0813	0.1124 0.1122	0.1285 0.1283	0.1335 0.1333	0.1285 0.1283	0.1124 0.1122	0.0815 0.0813	0.0316 0.0315
0.1	0.0397 0.0395	0.0945 0.0943	0.1266 0.1263	0.1430 0.1427	0.1481 0.1478	0.1430 0.1427	0.1266 0.1263	0.0945 0.0943	0.0397 0.0395
0	0.0479 0.0477	0.1067 0.1064	0.1395 0.1393	0.1561 0.1558	0.1611 0.1608	0.1561 0.1558	0.1395 0.1393	0.1067 0.1064	0.0479 0.0477
-0.1	0.0561 0.0559	0.1180 0.1177	0.1513 0.1510	0.1679 0.1676	0.1729 0.1726	0.1679 0.1676	0.1513 0.1510	0.1180 0.1177	0.0561 0.0559
-0.2	----- -----	0.1285 0.1283	0.1621 0.1618	0.1786 0.1783	0.1836 0.1833	0.1786 0.1783	0.1621 0.1618	0.1285 0.1283	----- ----
-0.3	----- -----	----- -----	0.1720 0.1717	0.1884 0.1881	0.1933 0.1931	0.1884 0.1881	0.1720 0.1717	----- -----	----- -----
-0.4	----- -----	----- -----	0.1811 0.1808	0.1974 0.1971	0.2023 0.2020	0.1974 0.1971	0.1811 0.1808	----- -----	----- -----
-0.5	----- -----	----- -----	----- -----	0.2056 0.2054	0.2105 0.2102	0.2056 0.2054	----- -----	----- -----	----- -----
-0.6	----- -----	----- -----	----- -----	0.2133 0.2130	0.2181 0.2178	0.2133 0.2130	----- -----	----- -----	----- -----
-0.7	----- -----	----- -----	----- -----	----- -----	0.2251 0.2248	----- -----	----- -----	----- -----	----- -----
-0.8	----- -----	----- -----	----- -----	----- -----	0.2315 0.2312	----- -----	----- -----	----- -----	----- -----
-0.9	----- -----	----- -----	----- -----	----- -----	0.2355 0.2353	----- -----	----- -----	----- -----	----- -----

In corrispondenza dei tratteggi non è possibile definire la popolazione per il vincolo a pag. 2 nota⁽³⁾. In questo esempio, dato $N = 10000$, nei due casi di estrazione i risultati non si discostano molto. Usando la variabile d si osserva come i valori di ρ e π possano condizionare l'esito dell'esperimento, aspetto ignorato nel caso dei test di Mc Nemar e della ${}_r t_{p_1-p_2}$. La stessa ${}_r t_{p_1-p_2}$ ha

un p-value = 0.0808 intermedio fra i valori della Tab. 5 e prossimo a un arco di p-value (colore blu): 0.0813, 0.0796, 0.0788, 0.0796, 0.0813 (e. s. r.) relativi alla d al variare di ρ e π .

Ma nella Tab. 5 è anche evidente come, al crescere di ρ ⁽⁵⁾ e al discostarsi di π dal valore .5 ⁽⁶⁾, il test sia più sensibile al trattamento. Ad es., dato $\pi = .7$ per $\rho = .6$ il test è significativo, mentre per $\rho = .3$ non lo è.

Facciamo un commento strettamente legato all'aspetto statistico: il risultato sembra dipendere dalla variabilità delle distribuzioni, funzione dei valori numerici dei parametri, quindi l'esito del test, oltre al tipo di trattamento proposto (obiettivamente efficace o meno, cioè tale da stimolare reazioni concordi o non nei soggetti), è anche condizionato dai valori dei parametri.

Interessanti anche i commenti legati all'aspetto interpretativo dal punto di vista dei contributi scientifici dell'esperimento, in quanto evidenzia il ruolo dei parametri. Il ricercatore potrebbe chiedersi e voler chiarire perché per alcuni valori di ρ il risultato sia maggiormente significativo.

Ma non avendo nessuna informazione a priori sullo stato di natura dovrebbe procurarsi qualche notizia orientativa, oppure stabilirlo per sue esigenze e conoscenze scientifiche. Esiste realmente qualche nesso con la sua teoria? Nel senso che il risultato sperimentale risulta significativo in un sottoinsieme di valori di ρ e π previsti o di valori non previsti?

Il ricercatore, ad es., ha constatato che il trattamento somministrato in quella forma o dose, per ρ molto elevato (escludendo i valori di π che ritiene poco sensati o estremi), ha efficacia più di quanto ritenuto in base ad indagini pregresse e sulle quali è basata la sua teoria. E' il caso di rivedere e modificare le sue convinzioni? O in altre parole, sapendo per le stesse esperienze precedenti che quel trattamento e quella dose danno esiti positivi e sono quindi efficaci, il prospetto (Tab. 5) potrebbe dare informazioni sulla effettiva natura della popolazione su cui lavora e sui veri valori di ρ e π che intervengono.

Un altro aspetto riguarda confronti fra risultati dello stesso test (sia descrittivo che sperimentale), applicato in due diverse popolazioni quando sono caratterizzate da parametri di riferimento di valore diverso l'una dall'altra e gli esiti del test risultano condizionati da detti parametri.

Supponiamo che l'Esempio 1 sia il risultato (identico in termini di frequenze della tabella di contingenza Madri-Figlie) di una indagine che sarebbe stato possibile applicare a due diversi gruppi etnici, religiosi o più in generale due tipi di società, nel senso che potrebbe riferirsi ad una prima popolazione, ma anche ad una seconda, noi non sappiamo quale, sappiamo solo che quel campione (con quel risultato) potrebbe provenire dall'una o dall'altra. Poniamo che detto campione di coppie sia rappresentativo della composizione della massa nei due differenti casi e, per semplicità, che da dati pregressi si sappia che mediamente l'accoglienza (i SI) per l'iniziativa sia $\pi = .7$, parametro condizionante (per due valori di π diversi, occorrerebbe considerare due colonne della Tab. 5!).

Sapendo che in un gruppo etnico c'è prevalenza di famiglie con figlie che continuano a vivere sotto lo stesso tetto, è probabile ed implicita l'importanza che le figlie stesse diano ai tradizionali nuclei

⁽⁵⁾ Cioè della proprietà che dovrebbe avere la massa dei soggetti: mantenere tra il prima e il dopo le stesse caratteristiche o essere quanto più possibile simili relativamente alle variabili estranee che potrebbero influenzare il risultato della ricerca.

⁽⁶⁾ $\sigma_d^2 = 2n[\pi(1-\pi)][1-\rho] = 2n[\pi(1-\pi)]\{1 - (\pi_{11} - \pi^2)/[\pi(1-\pi)]\}$ e $\sigma_x^2 = \sigma_y^2 = \pi(1-\pi)/n$.
 π influisce sulla quantità $\pi(1-\pi)$ ed anche su ρ . Se π si discosta dallo 0.5, $\pi(1-\pi)$ diminuisce mentre ρ in valore assoluto tende ad aumentare; quindi per ρ positivo σ_d^2 diminuisce per la duplice azione. Pertanto la distribuzione si raccoglie intorno al valor medio, e uno stesso valore d_s si trova spostato verso la coda dando luogo ad un p-value inferiore e tale da rendere più facilmente significativo il test.

familiari, conseguentemente che esista una concordanza di opinioni sulla opportunità dell'iniziativa, valutiamo $\rho = .6$, secondo parametro condizionante. In tal caso il ricercatore, oltre che per la bontà dell'iniziativa, si aspetta che prevalga l'accettazione. Infatti il test è significativo: p-value = 0.0424. L'opposto vale se ci troviamo nel caso dell'altro gruppo, se preesiste una situazione non idilliaca, figlie andate a vivere da sole, quindi meno interessate all'iniziativa, esisterà minore concordanza tra madri e figlie, poniamo $\rho = .3$. Il risultato non è significativo: p-value = 0.0969.

Nei due casi lo stesso campione porta a conclusioni diverse: nel primo la proposta ha efficacia, nel secondo non suscita consensi statisticamente significativi. Pur avendo le stesse frequenze nella Tabella Madri - Figlie, i risultati dell'iniziativa dipendono anche da caratteristiche, dalla struttura dei gruppi etnici di riferimento, vista in termini di quei parametri condizionanti ρ e π ⁽⁷⁾.

Pertanto, confronti ed estensioni dei risultati di test ad altri ambienti o differenti stati di natura hanno senso se, dati certi parametri condizionanti, questi non sono molto differenti da una popolazione all'altra. E' il ricercatore che deve stabilire se sono presenti, individuarli e valutarli. Nell'esempio ne abbiamo data una interpretazione intuitiva, orientativa e approssimata.

Le attuali conclusioni non sono generali, dipendono dal test e dal tipo di popolazione trattata, resta il fatto che i risultati di alcune indagini potrebbero andar bene quando riferite all'ambiente in cui sono state eseguite e non ad ambiente diverso.

Il discorso si potrebbe estendere anche ad altri test.

Esempio 2

Consideriamo un altro esempio ([3] F. Gori, L. Leone), un gruppo di $n = 25$, per $\alpha = 0.05$.

Tab. 6

Voting intentions for the majority party before last elections and two years after elections

		<i>After</i>		Tot
		0	1	
<i>Before</i>	0	14	9	23
	1	2	0	2
Tot		16	9	25

Applicando Mc Nemar abbiamo $\chi_s^2 = 3.2727$ (corretto), gl = 1 e p-value = 0.0705, test non significativo, oppure $\chi_s^2 = 4.4545$ (non corretto) e p-value = 0.0348 test significativo.

Nel continuo la ${}_r t_{p_1-p_2} = -2.2811$, con p-value = 0.0159, inferiore ad, il test è significativo.

La differenza sperimentale risulta $d_s = -7$, la coda interessata è la sinistra.

Costruiamo la tabella dei risultati nel discontinuo al variare dei parametri ρ e π . Il test riguarda una popolazione di $N = 10000$ soggetti, e. s. r. ed $n = 25$ (Prima - Dopo, gli stessi soggetti).

⁽⁷⁾ Il test di Mc Nemar (χ^2) e la ${}_r t_{p_1-p_2}$ non distinguono le due situazioni e danno p-value non significativi.

Tab. 7

Tabella del p-value al variare di ρ e π della popolazione come in Tab. 3 per $\alpha = 0.05$, $n = 25$ (14, 9, 2, 0) ed e. s. r., $N = 10000$.

$\rho \setminus \pi$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.9	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.8	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.7	0.0000	0.0000	0.0000	0.0005	0.0005	0.0005	0.0000	0.0000	0.0000
0.6	0.0000	0.0003	0.0009	0.0017	0.0020	0.0017	0.0009	0.0003	0.0000
0.5	0.0000	0.0007	0.0024	0.0040	0.0046	0.0040	0.0024	0.0007	0.0000
0.4	0.0001	0.0017	0.0047	0.0075	0.0085	0.0075	0.0047	0.0017	0.0001
0.3	0.0002	0.0031	0.0080	0.0120	0.0135	0.0120	0.0080	0.0031	0.0002
0.2	0.0005	0.0050	0.0120	0.0174	0.0193	0.0174	0.0120	0.0050	0.0005
0.1	0.0008	0.0075	0.0167	0.0233	0.0256	0.0233	0.0167	0.0075	0.0008
0	0.0013	0.0104	0.0218	0.0296	0.0323	0.0296	0.0218	0.0104	0.0013
-0.1	0.0019	0.0137	0.0272	0.0361	0.0391	0.0361	0.0272	0.0137	0.0019
-0.2	----	0.0174	0.0328	0.0427	0.0461	0.0427	0.0328	0.0174	----
-0.3	----	----	0.0386	0.0494	0.0529	0.0494	0.0386	----	----
-0.4	----	----	0.0444	0.0559	0.0597	0.0559	0.0444	----	----
-0.5	----	----	----	0.0624	0.0664	0.0624	----	----	----
-0.6	----	----	----	0.0688	0.0729	0.0688	----	----	----
-0.7	----	----	----	----	0.0793	----	----	----	----
-0.8	----	----	----	----	0.0855	----	----	----	----
-0.9	----	----	----	----	0.0926	----	----	----	----

Dato $\alpha = 0.05$, possiamo notare che per valori di ρ positivi il test risulta significativo, mentre non lo sarebbe nei casi di ρ fortemente negativo.

Esempio 3

Un campione di ampiezza $n = 7$ e frequenze: 1, 2, 3, 1. $\alpha = 0.05$.

Mc Nemar dà $\chi_s^2 = 0.0$ (corretto) con p-value = 0.999, o $\chi_s^2 = 0.2$ (non corretto) con p-value = 0.6551.

Seguendo l'espressione (4), nel continuo si ha: $r t_{p_1-p_2} = 0.4201$ e p-value = 0.3445.

La differenza sperimentale è $d_s = 1$, positiva, il p-value pertanto è calcolato sulla coda destra.

Nel discontinuo al variare dei due parametri abbiamo la seguente tabella.

Tab. 8

Tabella del p-value al variare di ρ e π della popolazione come in Tab. 1, e. c. r., $\alpha = 0.05$, $n = 7$.

$\rho \setminus \pi$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
-0.9	----	----	----	----	0.4591	----	----	----	----
-0.8	----	----	----	----	0.4375	----	----	----	----
-0.7	----	----	----	----	0.4263	----	----	----	----
-0.6	----	----	----	0.4173	0.4201	0.4173	----	----	----
-0.5	----	----	----	0.4139	0.4160	0.4139	----	----	----
-0.4	----	----	0.4040	0.4173	0.4125	0.4173	0.4040	----	----
-0.3	----	----	0.4001	0.4069	0.4089	0.4096	0.4001	----	----
-0.2	----	0.3790	0.3957	0.4029	0.4050	0.4029	0.3957	0.3790	----
-0.1	0.3197	0.3729	0.3907	0.3983	0.4005	0.3983	0.3907	0.3729	0.3197
0	0.3085	0.3658	0.3848	0.3929	0.3953	0.3929	0.3848	0.3658	0.3085
0.1	0.2953	0.3572	0.3779	0.3866	0.3891	0.3866	0.3779	0.3572	0.2953
0.2	0.2798	0.3466	0.3695	0.3790	0.3817	0.3790	0.3695	0.3466	0.2798
0.3	0.2616	0.3333	0.3589	0.3695	0.3725	0.3695	0.3589	0.3333	0.2616
0.4	0.2401	0.3162	0.3451	0.3572	0.3606	0.3572	0.3451	0.3162	0.2401
0.5	0.2148	0.2937	0.3264	0.3404	0.3444	0.3404	0.3264	0.2937	0.2148
0.6	0.1848	0.2638	0.2999	0.3162	0.3209	0.3162	0.2999	0.2638	0.1848
0.7	0.1493	0.2237	0.2616	0.2798	0.2852	0.2798	0.2616	0.2237	0.1493
0.8	0.1075	0.1698	0.2053	0.2237	0.2294	0.2237	0.2053	0.1698	0.1075
0.9	0.0581	0.0972	0.1222	0.1362	0.1406	0.1362	0.1222	0.0972	0.0581

In Appendice II sono riportati alcuni grafici illustrativi dell'Esempio 3.

Esempio 4

Data una popolazione del tipo come rappresentato in Tab. 3, e. s. r., $N = 1000$, e un campione di ampiezza $n = 12$ che, in un esperimento Prima-Dopo un trattamento, presenta frequenze interne:

$n_{00} = 2, n_{01} = 5, n_{10} = 2, n_{11} = 3$.

Il test di Mc Nemar dà $\chi_s^2 = 0.5714$ (corretto) con p-value = 0.4499, o $\chi_s^2 = 1.2857$ (non corretto) con p-value = 0.2569.

Seguendo la (4) si ha ${}_r t_{p_1-p_2} = -1.1489$, con p-value = 0.1375 (continuo).

Nel discontinuo consideriamo il caso di e. s. r. La differenza sperimentale $d_s = -3$ si trova sulla coda sinistra della distribuzione.

Tab. 9

Tabella del p-value al variare di ρ e π della popolazione, strutturata come in Tab. 3, $N = 1000$, campione e. s. r., $n = 12$ (2, 5, 2, 3), differenza sperimentale $d_s = -3$.

$\rho \setminus \pi$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.9	0.0001	0.0005	0.0013	0.0119	0.0021	0.0019	0.0013	0.0005	0.0001
0.8	0.0008	0.0041	0.0081	0.0111	0.0122	0.0111	0.0081	0.0041	0.0008
0.7	0.0026	0.0111	0.0204	0.0269	0.0292	0.0269	0.0204	0.0111	0.0026
0.6	0.0055	0.0211	0.0363	0.0461	0.0494	0.0461	0.0363	0.0211	0.0055
0.5	0.0095	0.0331	0.0536	0.0660	0.0701	0.0660	0.0536	0.0331	0.0095
0.4	0.0146	0.0461	0.0709	0.0859	0.0898	0.0859	0.0709	0.0461	0.0146
0.3	0.0204	0.0594	0.0875	0.1030	0.1079	0.1030	0.0875	0.0594	0.0204
0.2	0.0269	0.0725	0.1030	0.1192	0.1243	0.1192	0.1030	0.0725	0.0269
0.1	0.0339	0.0852	0.1173	0.1339	0.1390	0.1339	0.1173	0.0852	0.0339
0	0.0411	0.0972	0.1304	0.1472	0.1524	0.1472	0.1304	0.0972	0.0411
-0.1	0.0486	0.1086	0.1424	0.1593	0.1645	0.1593	0.1424	0.1086	0.0486
-0.2	-----	0.1192	0.1534	0.1703	0.1754	0.1703	0.1534	0.1192	-----
-0.3	-----	-----	0.1635	0.1804	0.1855	0.1804	0.1635	-----	-----
-0.4	-----	-----	0.1729	0.1896	0.1947	0.1896	0.1729	-----	-----
-0.5	-----	-----	-----	0.1981	0.2031	0.1981	-----	-----	-----
-0.6	-----	-----	-----	0.2060	0.2108	0.2060	-----	-----	-----
-0.7	-----	-----	-----	-----	0.2176	-----	-----	-----	-----
-0.8	-----	-----	-----	-----	0.2220	-----	-----	-----	-----
-0.9	-----	-----	-----	-----	0.2189	-----	-----	-----	-----

Fissando $\alpha = 0.05$ possiamo vedere per quali valori dei due parametri la differenza $d_s = -3$ dà luogo ad un p-value (in grassetto) inferiore a 0.05, cioè risulta significativa.

Esempio 5

Consideriamo una popolazione $N = 1000$, e. s. r. di un campione $n = 30$ (7, 2, 4, 17). $\alpha = 0.025$.

Mc Nemar dà $\chi_s^2 = 0.1667$ (corretto) con p-value = 0.6835 o $\chi_s^2 = 0.6667$ con p-value = 0.4144.

La ${}_r t_{p_1-p_2} = 0.8118$, p-value = 0.2118.

Nel discontinuo, abbiamo $d_s = 2$, giace sulla coda destra della distribuzione, e $dp_s = 0.0667$.

Tab. 10

Tabella del p-value al variare di ρ e π della popolazione, strutturata come in Tab. 3, $N = 1000$, campione e. s. r., $n = 30$ (7, 2, 4, 17), differenza sperimentale $d_s = 2$.

$\rho \setminus \pi$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.9	0.0216	0.0538	0.0775	0.0912	0.0956	0.0912	0.0775	0.0538	0.0216
0.8	0.0633	0.1245	0.1589	0.1761	0.1813	0.1761	0.1589	0.1245	0.0633
0.7	0.1043	0.1761	0.2106	0.2269	0.2318	0.2269	0.2106	0.1761	0.1043
0.6	0.1391	0.2125	0.2450	0.2599	0.2644	0.2599	0.2450	0.2125	0.1391
0.5	0.1678	0.2393	0.2696	0.2834	0.2875	0.2834	0.2696	0.2393	0.1678
0.4	0.1912	0.2599	0.2883	0.3012	0.3050	0.3012	0.2883	0.2599	0.1912
0.3	0.2106	0.2764	0.3031	0.3152	0.3188	0.3152	0.3031	0.2764	0.2106
0.2	0.2269	0.2899	0.3152	0.3266	0.3300	0.3266	0.3152	0.2899	0.2269
0.1	0.2408	0.3012	0.3253	0.3362	0.3393	0.3362	0.3253	0.3012	0.2408
0	0.2528	0.3109	0.3339	0.3443	0.3473	0.3443	0.3339	0.3109	0.2528
-0.1	0.2633	0.3192	0.3414	0.3513	0.3542	0.3513	0.3414	0.3192	0.2633
-0.2	-----	0.3266	0.3479	0.3575	0.3603	0.3575	0.3479	0.3266	-----
-0.3	-----	-----	0.3537	0.3629	0.3656	0.3629	0.3537	-----	-----
-0.4	-----	-----	0.3589	0.3678	0.3704	0.3678	0.3589	-----	-----
-0.5	-----	-----	-----	0.3722	0.3747	0.3722	-----	-----	-----
-0.6	-----	-----	-----	0.3761	0.3786	0.3761	-----	-----	-----
-0.7	-----	-----	-----	-----	0.3822	-----	-----	-----	-----
-0.8	-----	-----	-----	-----	0.3854	-----	-----	-----	-----
-0.9	-----	-----	-----	-----	0.3897	-----	-----	-----	-----

Il test risulta sostanzialmente non significativo in tutti i casi.

Esempio 6

Popolazione $N = 1000$, e. s. r. di un campione $n = 30$ (5, 3, 12, 10). $\alpha = 0.025$. $d_s = 9$, $dp_s = 0.3$.

Mc Nemar dà $\chi_s^2 = 4.2667$ (corretto) con p-value = 0.0389, o $\chi_s^2 = 5.4$ con p-value = 0.0201.

Per la $r t_{p_1-p_2} = 2.523$, abbiamo p-value = 0.009.

Tab. 11

Tabella del p-value al variare di ρ e π della popolazione in Tab. 3, $N = 1000$, per $n = 30$ (5, 3, 12, 10)

$\rho \setminus \pi$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.9	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.8	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.7	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.6	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.5	0.0000	0.0001	0.0003	0.0007	0.0008	0.0007	0.0003	0.0001	0.0000
0.4	0.0000	0.0002	0.0009	0.0016	0.0019	0.0016	0.0009	0.0002	0.0000
0.3	0.0000	0.0005	0.0018	0.0031	0.0037	0.0031	0.0018	0.0005	0.0000
0.2	0.0000	0.0009	0.0031	0.0053	0.0061	0.0053	0.0031	0.0009	0.0000
0.1	0.0001	0.0016	0.0050	0.0079	0.0091	0.0079	0.0050	0.0016	0.0001
0	0.0001	0.0026	0.0072	0.0111	0.0125	0.0111	0.0072	0.0026	0.0001
-0.1	0.0002	0.0038	0.0099	0.0147	0.0164	0.0147	0.0099	0.0038	0.0002
-0.2	-----	0.0053	0.0128	0.0185	0.0206	0.0185	0.0128	0.0053	-----
-0.3	-----	-----	0.0161	0.0226	0.0249	0.0226	0.0161	-----	-----
-0.4	-----	-----	0.0195	0.0269	0.0295	0.0269	0.0195	-----	-----
-0.5	-----	-----	-----	0.0313	0.0341	0.0313	-----	-----	-----
-0.6	-----	-----	-----	0.0358	0.0388	0.0358	-----	-----	-----
-0.7	-----	-----	-----	-----	0.0436	-----	-----	-----	-----
-0.8	-----	-----	-----	-----	0.0483	-----	-----	-----	-----
-0.9	-----	-----	-----	-----	0.0436	-----	-----	-----	-----

6 – Approccio alla significatività basato sull'intervallo di confidenza di $\delta = \pi_x - \pi_y$

Come noto ([6] C. J. Clopper, E. S. Pearson e [7] T. J. Stantner, Altri), data l'ampiezza campionaria n , per la costruzione dell'intervallo di confidenza nel discontinuo occorre preventivamente determinare, al variare del parametro di interesse, i limiti (inferiori e superiori) dei sottoinsiemi, di dimensioni (probabilità) $1-\alpha$, dei valori della variabile campionaria. Quindi, dato un valore sperimentale della variabile campionaria, si determina l'intervallo dei valori del parametro in corrispondenza ai quali tale valore appartiene ai detti sottoinsiemi. Questo è l'intervallo cercato.

Nel nostro caso il parametro è $\delta = \pi_x - \pi_y$, la variabile campionaria è la differenza tra proporzioni $dp = p_x - p_y$. Quindi, dato un valore sperimentale dp_s , si osserverà da quale δ' a quale δ'' è contenuto entro i limiti dei sottoinsiemi precedentemente determinati.

δ' e δ'' delimitano l'intervallo di confidenza.

Nel caso di due campioni indipendenti ([3] F. Gori, L. Leone e [7] T. J. Stantner, Altri) sono già note curve che danno l'ampiezza, soluzione grafica, dell'intervallo di confidenza per un dato dp_s .

Nel caso attuale di campioni correlati abbiamo lo stesso obiettivo, ma c'è la presenza di ρ .

Per valori di $\delta = \pi_x - \pi_y$ ($-1 < \delta < 1$) e ρ ($-1 < \rho < 1$) compatibili, si inizierà determinando i limiti dei sottoinsiemi dei valori dp ($-1 \leq dp \leq 1$), dp' (inferiore) e dp'' (superiore), per ogni valore del parametro δ . Pertanto l'intervallo di confidenza di δ , basato sulle distribuzioni della dp , dipende: dalla dimensione campionaria n , dai valori dei π_{ij} (vedere (1) e (2)), e implicitamente anche dal parametro ρ . Vedremo inoltre che ρ gioca un ruolo fondamentale circa la costruzione delle popolazioni al variare di δ e con riferimento alla Tab. 1 in cui la differenza $\delta = \pi_x - \pi_y$ (Tab. 3 in cui $N\delta = N\pi_x - N\pi_y$), può assumere valori negativi, nullo, o positivi. Infatti, dati ρ , π_x e π_y , si ha⁽⁸⁾: $\pi_{11} = \pi_x \pi_y + \rho \sqrt{\pi_x \pi_y (1 - \pi_x)(1 - \pi_y)}$ e quindi la popolazione come già visto.

Daremo la priorità a ρ , e per ogni valore di ρ determineremo le popolazioni al variare di δ .

Per ρ assumeremo valori da -.9 a .9 con passo .1, ed anche per δ valori da -.9 a .9 con lo stesso passo; ma per maggior precisione il passo per ambedue potrebbe anche essere inferiore.

Osserviamo che per ogni δ ($=\pi_x - \pi_y$) possiamo avere più popolazioni bivariate, utilizzeremo quelle che forniranno gli estremi dp' e dp'' meno distanti.

Fissato ρ e per un dato n , utilizzando tali popolazioni, al variare del parametro δ (nel discontinuo, per alcuni valori prestabiliti) determineremo i sottoinsiemi di probabilità $1 - \alpha$. Operativamente, *per trovare la coppia di estremi dp' e dp'' in corrispondenza di un dato valore di δ è sufficiente determinare i valori critici ottenuti in base ad $\alpha/2$ su ogni coda della distribuzione della dp partendo dalla (1) o dalla (2). Tali valori critici sono determinati per difetto (area $\leq \alpha/2$), quindi il sottoinsieme avrà dimensione $\geq 1-\alpha$ ⁽⁹⁾. Ciò è stato ripetuto al variare di δ . Infine, dato un valore sperimentale di dp_s , si osserva da quale valore δ' a quale valore δ'' corrispondono sottoinsiemi che contengono il valore dp_s . $\delta' | \text{---} | \delta''$ è l'intervallo di confidenza.*

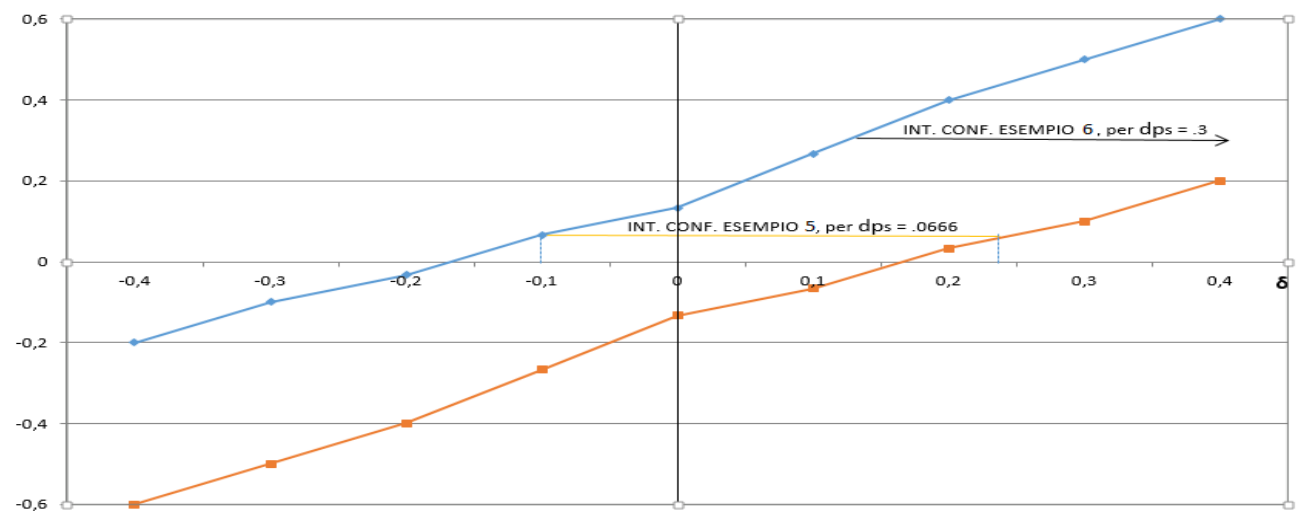
⁽⁸⁾ Dati ρ , π_x e π_y , si ottiene la struttura della generica popolazione, quindi dato n , la distribuzione campionaria della variabile dp , dalla quale i valori critici dp' e dp'' , per $\alpha/2 = .025$ (compresi), su ogni coda della generica distribuzione campionaria, estremi dei sottoinsiemi di dimensione $\geq 1-\alpha$.

⁽⁹⁾ Il programma lavora sulla coda sinistra e determina direttamente il valore dp' . Per determinare dp'' , basta invertire la distribuzione della dp , scambiando righe e colonne interne della tabella della popolazione.

Graficamente, riportati in ascisse i valori del parametro, ad es. δ : -.9, -.8,..-.1, 0, .1, .,8, .9, e in ordinate i valori dp' e dp'' , si individuano due linee (spezzate), una superiore ed una inferiore. Data l'ordinata dp_s si traccia in corrispondenza una linea orizzontale che interseca le due linee spezzate, prima la superiore poi la inferiore, ai due punti di intersezione corrispondono le ascisse δ' e δ'' . Dati tali limiti di confidenza del parametro δ , se non contengono $\delta = 0$ si rifiuta H_0 .

Riprendiamo l'Esempio 5 di pag. 11. $n = 30$ (7, 2, 4, 17), $dp_s = 0.0667$, test unidirezionale, con probabilità di rifiuto sotto H_0 su una coda pari a 0.025. Nell'attuale contesto $\alpha = 0.05$ su due code. Scegliamo il valore di $\rho = .4$, in Appendice III è riportato il dettaglio per δ da -.4 a .4 (per altri valori non è stato possibile determinare popolazioni). Come già osservato, poiché un dato valore di δ può essere determinato da più coppie di valori $\pi_x - \pi_y$, quindi da più popolazioni bivariate, ciascuna di queste dà luogo ad una distribuzione dello stimatore dp , diversa, tali distribuzioni non concordano circa i limiti dp' e dp'' dei sottoinsiemi. Per ogni valore di δ sono stati preferiti i limiti meno distanti fra loro (colorati rossi e azzurri sia nel grafico seguente che nell'Appendice III in Tab. 12) con conseguente intervallo di confidenza meno ampio - linee minimali -; soluzione meno conservativa perché gli intervalli includono lo 0 meno facilmente. Al variare di δ si è ottenuta la seguente serie di coppie di valori dp' e dp'' , come risulta nel grafico.

Linee minimali ed intervalli di confidenza di $\delta = \pi_x - \pi_y$, per $n = 30$ e $\rho = .4$.



Si può osservare, con $dp_s = 0.0667$, che l'intervallo di confidenza si estende da circa $\delta' = -.01$ a $\delta'' = .235$, comprende $\delta = 0$, quindi non si può rifiutare l'ipotesi nulla H_0 . D'altra parte c'era da sospettarlo in quanto, osservando la Tab. 10 del p-value, per $\rho = .4$ al variare di π sotto H_0 , quindi almeno per $\delta = 0$, i risultati sono tutti **non significativi**.

Riportiamo, dall'Appendice III per $\delta = 0$, i valori π_x e π_y , i p-value (Tab. 10) e i valori critici di Tab. 12, a ciascuno dei quali corrisponde una frequenza (effettiva, nel discreto) $\leq \alpha/2 = 0.025$:

$\delta = \pi_x - \pi_y$.1 -.1	.2 -.2	.3 -.3	.4 -.4	.5 -.5	.6 -.6	.7 -.7	.8 -.8	.9 -.9
p-value	0.1912	0.2599	0.2883	0.3012	0.3050	0.3012	0.2883	0.2599	0.1912
val. critici	-.133 .133	-.2 .2	-.2 .2	-.233 .233	-.233 .233	-.233 .233	-.2 .2	-.2 .2	-.133 .133
freq. $\leq \alpha/2$	0.0237	0.0099	0.0207	0.0121	0.0136	0.0121	0.0207	0.0099	0.0237

Per $\alpha/2 = 0.025$, la $dp_s = 0.0667$ non supera mai i valori critici e con un p-value sempre $> \alpha/2$.

Per l'Esempio 6 (5, 3, 12, 10) di pag. 12, $n = 30$ (5, 3, 12, 10) $dp_s = 0.3$, l'intervallo di confidenza si estende da $\delta' = 0.14$ ad oltre 0.4, non comprende $\delta = 0$ quindi si può rifiutare H_0 . In questo caso, osservando la Tab. 11 del p-value per $\rho = .4$ e π variabile, i risultati sono tutti **significativi**.

Per quanto riguarda i valori critici, per $\delta = 0$ e $\rho = .4$ sono ancora gli stessi: da -0.133 a 0.133 e da -0.233 a 0.233 $< dp_s = 0.3$, sempre superiore, mentre i p-value, sono tutti $< \alpha/2$, come segue, dalla Tab. 11 dell'Esempio 6:

0.0000 0.0002 0.0009 0.0016 0.0019 0.0016 0.0009 0.0002 0.0000

Commento riferito agli Esempi 5 e 6. Sono considerate le stesse popolazioni, $N = 1000$ per $\rho = .4$, al variare di π e per $\delta = 0$ (sotto H_0). La numerosità campionaria $n = 30$ è la stessa⁽¹⁰⁾, pertanto si hanno le stesse distribuzioni (e gli stessi valori critici dato α). Ciò che cambia (per $\rho = .4$) è il risultato sperimentale - diverso - (2 e 9) che nel primo caso non cade nelle zone di rifiuto mentre nel secondo caso cade nelle zone di rifiuto.

Rimandiamo all'Appendice III per altri commenti.

Grafici (ed intervalli) analoghi potrebbero essere determinati considerando valori di $\delta = \pi_x - \pi_y$ più dettagliati in termini di valori π_x e π_y .

Inoltre il grafico precedente vale per $\rho = .4$, ma potrebbe essere determinato anche per altri valori di ρ immaginando un terzo asse di riferimento uscente perpendicolare al piano del grafico. Per ogni valore di ρ avremo le due linee di ogni caso particolare come quello sopra visto.

Ed al variare di ρ si hanno due superfici in cui giacciono le linee.

Conclusioni

Sia nel precedente (per gruppi indipendenti) sia nell'attuale Rapporto Tecnico, abbiamo mostrato come, nel caso di popolazioni limitate e di piccoli gruppi, la struttura di una popolazione in termini di parametri, possa inserirsi nella dinamica di un trattamento ed in conseguenza sul risultato di un test. Ciò pone il ricercatore in primo piano, sia nel caso in cui determinati parametri della popolazione obiettivo gli siano noti o gli interessi comunque sperimentarli, sia nel caso in cui più prudentemente ne determini preventivamente l'entità, ad esempio mediante un sondaggio e/o una indagine pilota. In ogni caso, rimane il fatto che egli deve stabilire quali parametri possano influire sul risultato di un test e valutarli. Altro punto è la ricerca di nuove variabili – test e delle loro distribuzioni sensibili ai parametri individuati.

⁽¹⁰⁾ Per ottenere le distribuzioni vengono utilizzate nella (2) le stesse (tutte le possibili) 4 - uple di n_{ij} .

BIBLIOGRAFIA

- [1] F.Gori. Il test della differenza tra proporzioni per piccoli campioni. Critiche all'approssimazione mediante la normale, significatività condizionata, considerazioni sulla logica inferenziale e sulla rappresentatività campionaria. A.D.S.S.C.C., Università di Bari, 2011.
- [2] E. D'Arcangelo, F. Gori. La funzione di densità campionaria per procedure di confronti multipli nel caso di estrazioni senza reimmissione da popolazioni dicotomiche limitate. Test per evidenziare il condizionamento dei parametri sul risultato sperimentale. Rapporto Tecnico n. 4, Dip. Scienze Stat., Sapienza, 2015.
- [3] F. Gori, L.Leone. Difference between proportions and correlations in small paired groups sampled with and without replacemeny. TPM, PADOVA, 2004.
- [4] F. Gori. Il coefficiente di correlazione r nel caso di caratteri dicotomici e la differenza tra proporzioni, negli universi con e senza ripetizione per piccoli campioni. Serie A – Ricerche, N. 20, La Sapienza, Roma, 1989.
- [5] P. Armitage. Statistica Medica. Feltrinelli, Milano, settembre 1979.
- [6] Clopper C.J. & Pearson E.S. The use of confidence or fiducial limits illustrated in the case of the binomial. BIOMETRICA XXVI, 1934, pp. 404, 413
- [7] T.J.Santner, V. Pradham, P. Senchaudhuri, C. R. Mehta, A. Tamhane. Small-sample comparison of confidence intervals for the difference of two independent binomial proportions. Computational Statistics & Data Analysis 51 (2007), 5791-5799.

APPENDICE I

Come risulta nel testo, poniamo che le tabelle-campione, Tab. 2, abbiano senso e significato, in sostanza siano valide, se è possibile calcolare d ed r . Ciò premesso, non si ritiene di escludere i valori di d cui corrispondono i seguenti casi in cui sembra non sia possibile calcolare r (perché forma indeterminata $0/0$). Ad es. gli n soggetti danno tutti la stessa risposta, o tutti risposta diversa, come nei seguenti quattro casi:

0	0	0	0	0	0	0	0	0	0	0	0	0	0			
1	0	n_{11}	$n_{1.}$	1	0	0	0	1	0	0	0	1	n_{10}	0	$n_{1.}$	$n_{1.}$
T	0	$n_{.1}$	n	T	0	$n_{.1}$	n	T	$n_{.0}$	0	n	T	$n_{.0}$	0	n	n

Infatti il valore di r può essere determinato mediante un passaggio al limite.

Consideriamo la prima tabella in termini di proporzioni:

	0	1	T
0	α	β	$\alpha+\beta$
1	γ	$1-\delta$	$1+\gamma-\delta$
T	$\alpha+\gamma$	$1+\beta-\delta$	$1+\alpha+\beta+\gamma-\delta$

con $\delta = \alpha + \beta + \gamma$

da cui: $r = [1-\delta-(1+\gamma-\delta)(1+\beta-\delta)]/[(\alpha+\beta)(1+\gamma-\delta)(\alpha+\gamma)(1+\beta-\delta)]^{-0.5}$.

e ponendo $\alpha = \beta = \gamma$ (con la stessa velocità di convergenza a 0) abbiamo $\delta = 3\alpha$, quindi:

$$\lim_{\alpha \rightarrow 0} r = \lim_{\alpha \rightarrow 0} [(1-3\alpha)-(1-2\alpha)^2] / 2\alpha(1-2\alpha) = 1/2.$$

Per i quattro casi i valori di r sono: 0.5, -0.5, 0.5, -0.5; come poteva essere valutato anche al livello **intuitivo**, dal momento che è interessata solo una casella delle diagonali. E' il caso di madri e figlie che sono **completamente in accordo o in disaccordo** su una sola delle alternative.

Si potrebbe discutere sul valore numerico (o meglio sulle velocità di convergenza), ma comunque sarebbe dannoso per il ricercatore escludere il valore di d fornito da tali tabelle perché non è possibile determinare r con la nota formula, trascurando una informazione così ricca di significato.

Ed anche i casi in cui Prima sono tutti concordi per il SI, poi alcuni cambiano opinione, o viceversa:

	0	1	T				
0	0	n_{01}	$n_{0.}$	0	n_{00}	n_{01}	$n_{0.}$
1	0	n_{11}	$n_{1.}$	1	0	0	0
T	0	$n_{.1}$	n	T	$n_{.0}$	$n_{.1}$	n

	0	1	T				
0	n_{00}	0	$n_{0.}$	0	n_{00}	0	$n_{0.}$
1	n_{10}	0	$n_{1.}$	1	n_{10}	0	$n_{1.}$
T	$n_{.0}$	0	n	T	$n_{.0}$	0	n

	0	1	T				
0	0	0	0	0	0	0	0
1	n_{10}	n_{11}	$n_{1.}$	1	n_{10}	n_{11}	$n_{1.}$
T	$n_{.0}$	$n_{.1}$	n	T	$n_{.0}$	$n_{.1}$	n

Facendo riferimento alle proporzioni abbiamo:

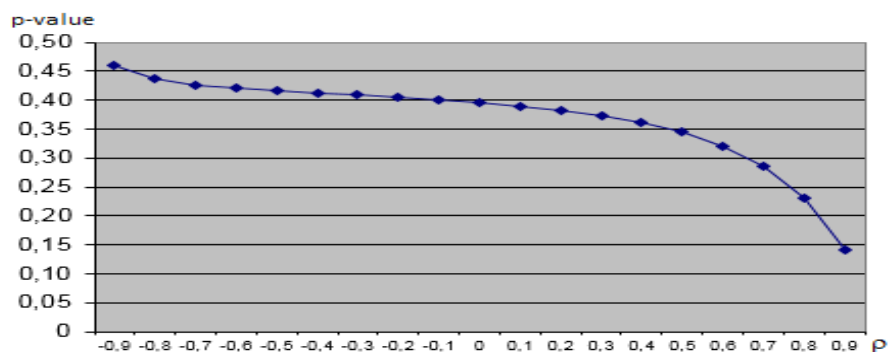
	0	1	T
0	ε	$A-\varepsilon$	A
1	ε	$1-A-\varepsilon$	$1-A$
T	2ε	$1-2\varepsilon$	1

da cui: $\lim_{\varepsilon \rightarrow 0} r = \lim_{\varepsilon \rightarrow 0} (1-2A)/[2(1/\varepsilon - 2)A(1-A)]^{0.5} = 0.$

APPENDICE II

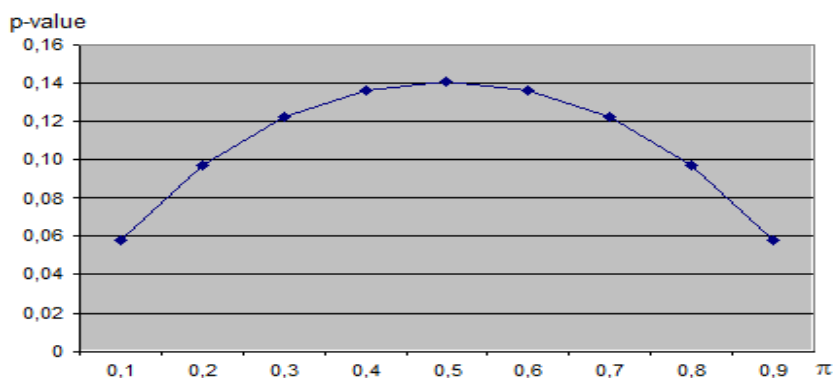
Dall'Esempio 3

Dalla Tab. 8, dato $\pi = 0.5$, si nota che la distribuzione del p-value ha valori massimi che diminuiscono al crescere di ρ , come nel grafico seguente:

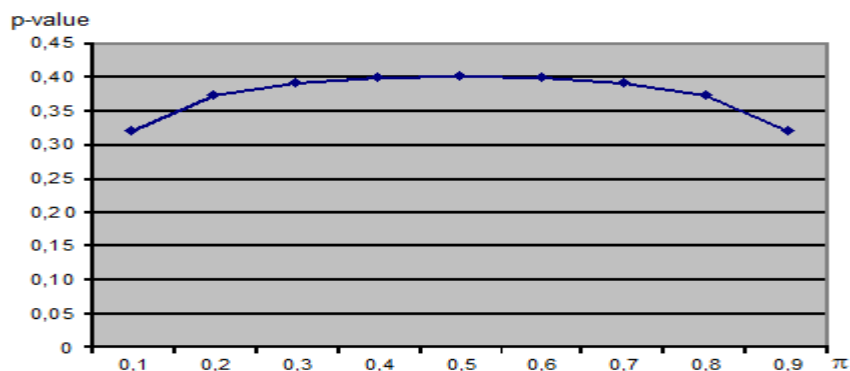


Dato ρ , il p-value ha andamento prima crescente e poi decrescente, simmetrico rispetto a $\pi = 0.5$. Ma con una curvatura che va attenuandosi al decrescere di ρ .

In particolare per $\rho = 0.9$ abbiamo, per π da 0.1 a 0.9:



Mentre per $\rho = -0.1$ abbiamo:



Per valori molto negativi di ρ , al variare di π i valori si riducono e tendono alla stazionarietà.

Dato $\alpha = 0.05$, se i valori di ρ e π ritenuti **sensati** dal ricercatore sono quelli rappresentati in Tab. 8, l'esempio è un caso di non significatività assoluta. Ma in realtà potrebbero esistere valori, ad es.

$\rho = 0.95$ (non in Tab. 8) e $\pi = 0.1$, che danno un p-value $0.0303 < \alpha$, oppure $\rho = 0.9$ e $\pi = 0.05$, che danno un p-value $0.0319 < \alpha$, quindi per tali valori di ρ e π , il test sarebbe significativo.

Un eventuale grafico a tre dimensioni (qui non riportato) del p-value (in ordinate) in funzione dei due parametri ρ e π (sul piano) avrebbe ordinate **decrescenti** al crescere di ρ e all'allontanarsi di π dal valore 0.5; alle estremità della curva, per $\rho = 0.9$ e per $\pi = 0.1$ o 0.9 il p-value si avvicina al livello 0.05, pertanto in questi casi estremi sarebbe più facile cadere in zona di rifiuto.

E dando un valore di α sull'asse del p-value, che individua un piano parallelo al piano su cui giacciono ρ e π , si potrebbero ottenere anche graficamente indicazioni sulla significatività.

APPENDICE III

Dall'Esempio 5, $n = 30$ (7, 2, 4, 17), $dp_s = 0.0666$ ed $\alpha = 0.025$, test unidirezionale. Nel contesto dell'intervallo di confidenza, osserviamo che per il generico sottoinsieme di dimensione $\geq 1-\alpha$, abbiamo $\alpha = 0.05$ ed $\alpha/2 = 0.025$ su ogni coda. Vedremo appresso alcune coincidenze per $\delta = 0$.

Facciamo una esplorazione considerando l'utilizzazione di popolazioni costruite sulla base di π_{11} , per mostrare esempi di determinazione di estremi di sottoinsiemi di dimensione $1-\alpha$.

Nella Tab. 12, che segue, per ogni valore di δ sono riportati i diversi valori di π_x e π_y e quindi le popolazioni da cui possono essere ottenuti. Se un valore di π_{ij} risulta negativo la popolazione non è possibile. In rosso e celeste sono riportati i limiti dp' e dp'' dei sottoinsiemi utilizzati per il grafico.

Tab. 12

Popolazioni per $\rho = .4$ e per δ da $-.8$ a $.8$							$dp_s = 0.0666$	Limiti dei sottoinsiemi		
δ	π_x	π_y	π_{00}	π_{01}	π_{10}	π_{11}	p-value	dp'	fr. Rifiuto	dp''
-8	.1	.9				-.026		POPOL. IMPOSSIBILE		
-7	.1	.8				-.028		"		"
	.2	.9				-.028		"		"
-6	.1	.7				-.025		"		"
	.2	.8				-.024		"		"
	.3	.9				-.025		"		"
-5	.1	.6				-.019		"		"
	.2	.7				-.013		"		"
	.3	.8				-.013		"		"
	.4	.9				-.019		"		"
-4	.1	.5				-.010		"		"
	.2	.6	.398	.402	.002	.198		-6		-2
	.3	.7	.294	.406	.006	.294		-6		-2
	.4	.8	.198	.402	.002	.398		-6		-2
	.5	.9				-.010		POPOL. IMPOSSIBILE		
-3	.1	.4	.599	.301	.001	.099		-5		-1
	.2	.5	.480	.320	.020	.180		-5		-1
	.3	.6	.370	.330	.030	.270		-.5333		-1
	.4	.7	.270	.330	.030	.370		-.5333		-1
	.5	.8	.180	.320	.020	.480		-5		-1
	.6	.9	.099	.301	.001	.599		-5		-1
-2	.1	.3	.685	.215	.015	.085		-4		-.03333
	.2	.4	.558	.242	.042	.158		-4		0
	.3	.5	.442	.258	.058	.242		-.43333		.03333
	.4	.6	.336	.264	.064	.336		-.43333		.03333
	.5	.7	.242	.258	.058	.442		-.43333		.03333
	.6	.8	.158	.242	.042	.558		-4		0
	.7	.9	.085	.215	.015	.685		-4		-.03333

-1	.1	.2	.768	.132	.032	.068			-.26666	.06666
	.2	.3	.633	.167	.067	.133			-3	.1
	.3	.4	.510	.190	.090	.210			-.33333	.13333
	.4	.5	.398	.202	.102	.298			-.33333	.13333
	.5	.6	.298	.202	.102	.398			-.33333	.13333
	.6	.7	.210	.190	.090	.510			-.33333	.13333
	.7	.8	.133	.167	.067	.633			-3	.1
	.8	.9	.068	.132	.032	.768			-.26666	.06666
							Dalla Tab. 10			
0	.1	.1	.846	.054	.054	.046	0.1912	-.13333	.0237	.13333
	.2	.2	.704	.096	.096	.104	0.2599	-2	.0099	.2
	.3	.3	.574	.126	.126	.174	0.2883	-2	.0207	.2
	.4	.4	.456	.144	.144	.256	0.3012	-.23333	.0121	.23333
	.5	.5	.350	.150	.150	.350	0.3050	-.23333	.0136	.23333
	.6	.6	.256	.144	.144	.456	0.3012	-.23333	.0121	.23333
	.7	.7	.174	.126	.126	.574	0.2883	-2	.0207	.2
	.8	.8	.104	.096	.096	.704	0.2599	-2	.0099	.2
	.9	.9	.046	.054	.054	.846	0.1912	-.13333	.0237	.13333
.1	.2	.1	.768	.032	.132	.068		-.06666		.26666
	.3	.2	.633	.067	.167	.133		-1		.3
	.4	.3	.510	.090	.190	.210		-.13333		.33333
	.5	.4	.398	.102	.202	.298		-.13333		.33333
	.6	.5	.298	.102	.202	.398		-.13333		.33333
	.7	.6	.210	.090	.190	.510		-.13333		.33333
	.8	.7	.133	.067	.167	.633		-1		.3
	.9	.8	.068	.032	.132	.768		-.06666		.26666
.2	.3	.1	.685	.015	.215	.085		.03333		.4
	.4	.2	.558	.042	.242	.158		0		.4
	.5	.3	.442	.058	.258	.242		-.03333		.43333
	.6	.4	.336	.064	.264	.336		-.03333		.43333
	.7	.5	.242	.058	.258	.442		-.03333		.43333
	.8	.6	.158	.042	.242	.558		.0		.4
	.9	.7	.085	.015	.215	.685		.03333		.4
.3	.4	.1	.599	.001	.301	.099		.1		.5
	.5	.2	.480	.020	.320	.180		.1		.5
	.6	.3	.370	.030	.330	.270		.1		.53333
	.7	.4	.270	.030	.330	.370		.1		.53333
	.8	.5	.180	.020	.320	.480		.1		.5
	.9	.6	.099	.001	.301	.599		.1		.5
.4	.5	.1		-.010				POPOL. IMPOSSIBILE		
	.6	.2	.398	.002	.402	.198		.2		.6
	.7	.3	.294	.006	.406	.294		.2		.6
	.8	.4	.198	.002	.402	.398		.2		.6
	.9	.5		-.010				POPOL. IMPOSSIBILE		
.5	.6	.1		-.019				“		“
	.7	.2		-.013				“		“
	.8	.3		-.013				“		“
	.9	.4		-.019				“		“
.6	.7	.1		-.025				“		“
	.8	.2		-.024				“		“
	.9	.3		-.025				“		“
.7	.8	.1		-.028				“		“
	.9	.2		-.028				“		“
.8	.9	.1		-.026				“		“

E' stato fissato $\rho = .4$. Le distribuzioni sono simmetriche per $\delta = 0$, ciò dà luogo agli stessi valori assoluti di dp' e dp'' nelle due code. Circa i valori dei limiti degli altri sottoinsiemi, si nota una simmetria, inversa, per valori di δ negativi e positivi.

Per la costruzione delle due linee del grafico, una in base ai limiti superiori dp'' ed una ai limiti inferiori dp' , osserviamo che dato un valore di δ si possono avere più popolazioni, ciascuna delle quali dà luogo a una determinata distribuzione dello stimatore dp ; tali distribuzioni non sempre sono tra loro concordanti circa i limiti dp' e dp'' dei sottoinsiemi. Quali coppie di valori utilizzare per la costruzione del grafico al variare di δ ?

Se consideriamo quelle che danno luogo ai limiti più distanti tra loro, quelli in grassetto, soluzione massimale, avremo due linee in cui la dimensione degli insiemi (livello di confidenza) sarà $\geq 1-\alpha$, quindi intervalli δ^* --- δ^* più ampi, con conseguente maggior facilità di contenere lo 0, soluzione più prudentiale, conservativa, perché in tal caso non si rifiuta H_0 .

Se si preferisce una soluzione minimale, più vicina agli obiettivi della ricerca (rifiuto di H_0), si considerano le coppie meno distanti tra loro, quelle in rosso e celeste. Ciò comporta un livello di confidenza non superiore a quello stabilito: linee più vicine tra loro e in conseguenza un intervallo δ^* --- δ^* meno ampio. Offrono una soluzione meno prudentiale perché includono lo 0 meno facilmente.

In Tab. 12 per $\delta = 0$ sono riportati i p-value dalla Tab. 10 e le frequenze $\leq \alpha/2$, utilizzate per determinare i limiti (simmetrici) dei sottoinsiemi: dp^* e dp^{**} ⁽¹¹⁾, che sarebbero anche le frequenze effettive di rifiuto nel test⁽¹²⁾. E il test ($H_0: \pi_x = \pi_y$) fa riferimento, come noto, ad un particolare punto dell'intervallo di confidenza: per $\delta = \pi_x - \pi_y = 0$.

Naturalmente, lo stesso discorso non si può fare nel caso dell'Esempio 6, il cui intervallo non contiene $\delta = 0$.

⁽¹¹⁾ Come già visto, nei due casi, del test e dell'intervallo per $\delta = 0$, le popolazioni utilizzate sono le stesse e quindi (nei due casi $n = 30$) sono uguali anche le distribuzioni campionarie.

⁽¹²⁾ Con esse è stato confrontato il p-value.