Un ufficio studi di una certa assicurazione ha constatato che nella località A, dove conta 25 automobili assicurate, vi sono stati 5 furti d'auto; nella località B, a fronte di 45 auto assicurate, vi sono stati 8 furti d'auto. Vi sono evidenze statistiche per affermare all'1% che le due località hanno differente rischiosità?

$\alpha = 0{,}01$

CASO: TEST PER DUE FREQUENZE (POP. BERNOULLIANE)

$X_A \sim B(p_A) \qquad \bar{x}_A = \frac{5}{25} = \frac{1}{5} \qquad n_A = 25$

$X_B \sim B(p_B) \qquad \bar{x}_B = \frac{8}{45} \qquad n_B = 45$

CONDIZIONI DI APPLICABILITÀ:

$n_A \bar{x}_A = 5 \qquad \Rightarrow n_A(1-\bar{x}_A) = 25 - 5 = 20$

$n_B \bar{x}_B = 8 \qquad \Rightarrow n_B(1-\bar{x}_B) = 45 - 8 = 37$

$\Rightarrow$ OK!

IPOTESI: $H_0: p_A = p_B \qquad H_1: p_A \neq p_B$

STATISTICA: VEDIAMO DUE DIVERSE STATISTICHE.

LA PRIMA, LEGGERMENTE MENO ACCURATA:

$$z_0' = \frac{\bar{x}_A - \bar{x}_B}{\sqrt{\frac{\bar{x}_A(1-\bar{x}_A)}{n_A} + \frac{\bar{x}_B(1-\bar{x}_B)}{n_B}}} = \frac{1/5 - 8/45}{\sqrt{\frac{\frac{4}{25}}{25} + \frac{\frac{8}{45}\cdot\frac{37}{45}}{45}}} \simeq 0{,}226$$

LA SECONDA, QUELLA UFFICIALE, PRESENTATA A LEZIONE:

$$z_0 = \frac{\bar{x}_A - \bar{x}_B}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}} \qquad \text{CON } \hat{p} = \frac{n_A \bar{x}_A + n_B \bar{x}_B}{n_A + n_B} = \frac{13}{70}$$

DA CUI
$$z_0 = \frac{1/5 - 8/45}{\sqrt{\frac{13 \cdot 57}{70^2}\left(\frac{1}{25} + \frac{1}{45}\right)}} = \frac{5}{9}\sqrt{\frac{42}{247}} \approx 0{,}2291$$

COME SI NOTA $z_0' \approx z_0$. PROSEGUIAMO CON z_0

REGIONE RIFIUTO: $R = I^C$ CON $I = [-z_\beta ; +z_\beta]$

$\beta = 1 - \frac{\alpha}{2} = 0,995$ $z_{0,995} = 2,5758$

RIFIUTO SE $|z_0| > z_\beta \Rightarrow 0,2291 > 2,5758$ FALSO

$\Rightarrow$ NON POSSO RIFIUTARE H_0

La compagnia assicurativa non vuole correre il rischio di "sottovalutare gli effetti negativi della criminalità nella località B rispetto alla località A." Proporre un test adeguato, svolgerlo al 10% e stimarne il p-value

$\alpha = 0{,}1$

DI CONSIDERARE H FALSA SE VERA

$\hookrightarrow p_B > p_A$

CASO: TEST PER DUE FREQUENZE

I DATI SONO GLI STESSI:

$X_A \sim B(p_A) \qquad \bar{x}_A = \frac{5}{25} = \frac{1}{5} \qquad n_A = 25$

$X_B \sim B(p_B) \qquad \bar{x}_B = \frac{8}{45} \qquad n_B = 45$

CONDIZIONI DI APPLICABILITÀ: $\Rightarrow$ OK! (STESSI DATI)

IPOTESI: METODO ERRORE DI I SPECIE: $H \to H_0$

QUINDI $H_0: p_B \geq p_A \qquad H_1: p_B < p_A$

STATISTICA: $z_0 = \dfrac{\bar{x}_B - \bar{x}_A}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}} \approx -0{,}2291$

REGIONE RIFIUTO: $R = (-\infty ; +z_\beta]$

$\beta = \alpha = 0{,}1$

0.88	1.1749868	1
0.9	1.2815516	1
0.91	1.340755	1

$z_\beta = z_{0,1} = -z_{0,9} = -1{,}28$

RIFIUTO SE: $z_0 < z_\beta \Rightarrow -0{,}2291 < -1{,}28$ FALSO

$\Rightarrow$ NON POSSO RIFIUTARE H_0

p-VALUE: $\bar{\alpha} = \Phi(z_0) = \Phi(-0{,}2291) \approx 1 - \Phi(0{,}23) \approx 1 - 0{,}591 \approx 0{,}409$

	0.00	0.01	0.02	0.03
0.0	0.5000	0.5040	0.5080	0.5120
0.1	0.5398	0.5438	0.5478	0.5517
0.2	0.5793	0.5832	0.5871	0.5910

Nel 1856 il biologo Gregor Mendel condusse un esperimento di ibridazione sui piselli. Secondo la teoria sulla trasmissione dei caratteri ereditari da lui stesso proposta, l'esperimento avrebbe dovuto produrre piselli del tipo **RY**, **RG**, **WY** e **WG** rispettivamente con frequenze relative 9/16, 3/16, 3/16 e 1/16. I risultati dell'esperimento furono:

Piselli	RY	RG	WY	WG
Frequenze	315	108	101	32

Questi risultati confermano la teoria di Mendel?

CASO: TEST χ^2 DI BUON ADATTAMENTO

IPOTESI H_0: I DATI PROVENGONO DALLA DISTRIBUZIONE TEORICA DI MENDEL

STATISTICA $Q = n \sum_{i=1}^{4} \frac{(p_i - f_i)^2}{p_i} \simeq 0{,}47$

	Piselli	RY	RG	WY	WG	Totale
n*f_i	Fr. assolute osservate	315	108	101	32	556
f_i	Fr. relative osservate	0,567	0,194	0,182	0,058	1,000
p_i	Fr. relative teoriche	0,5625	0,1875	0,1875	0,0625	1,000
n*p_i	verifica condizioni	312,75	104,25	104,25	34,75	556
p_i - f_i		-0,004	-0,007	0,006	0,005	
(p_i - f_i)^2		0,000016	0,000045	0,000034	0,000024	
Q_i		0,000029	0,000243	0,000182	0,000391	0,000845
					Q	0,47

REGIONE $R = \left(\chi^2_\beta (k-1); +\infty \right)$

$\beta = 1 - \alpha = 0{,}95$ $\quad k = 4 \quad \chi^2_{0,95}(3) \simeq 7{,}815$

9	0,95	
1	3,841455	5
6	5,991476	7
4	7,814725	9

RIFIUTO SE $Q > \chi^2_{0,95}(3)$ $\quad 0{,}47 > 7{,}815$ FALSO

$\Rightarrow$ NON POSSO RIFIUTARE H_0

P-VALUE $\bar{\alpha}$: $\quad Q = \chi^2_{1-\bar{\alpha}}(3)$

```
> 1-pchisq(0.47,3)
[1] 0.9254311
```

$\Rightarrow \bar{\alpha} = 1 - F_{\chi^2(3)}(Q) \simeq 0{,}9254$

0,35185	0,05	
0,47	0,07541	0,92459
0,58438	0,1	

$\Rightarrow$ H_0 NON PUÒ ESSERE RIFIUTATA PER QUASI QUALSIASI α

Da uno studio condotto su un campione di 1423 individui di 32 anni, di cui 849 primogeniti, si è appurato che ben 115 sono in possesso di una laurea e 543 si sono fermati al diploma di scuola superiore. Dei rimanenti, 84 possiedono una laurea, mentre il numero di coloro che si sono fermati al diploma di scuola superiore è 403. Da questi dati si può desumere l'indipendenza tra primogenitura e grado di istruzione?

CASO: TEST χ^2 DI INDIPENDENZA

IPOTESI H_0: I CAMPIONI PROVENGONO DA DISTRIBUZIONI INDIPENDENTI

STATISTICA:
$$Q = \sum_{i,j} \frac{\left(\frac{F_i F_j}{n} - F_{ij}\right)^2}{\frac{F_i F_j}{n}} = \sum_{i,j} \frac{n^2\left(\frac{F_i}{n}\frac{F_j}{n} - \frac{F_{ij}}{n}\right)^2}{n\,\frac{F_i}{n}\frac{F_j}{n}} =$$
$$= n \sum_{i,j} \frac{(f_i f_j - f_{ij})^2}{f_i f_j} \simeq 11{,}75$$

REGIONE $R = \left(\chi^2_p(h) \, ; +\infty\right)$ $\quad p = 1-\alpha = 0{,}95$

$$h = (j-1)(k-1) = (2-1)(3-1) = 2 \qquad \chi^2_{0,95}(2) \simeq 5{,}9915$$

Fr. ass. osservate	primo	non primo	Totale
no diploma	191	87	278
diploma	543	403	946
laurea	115	84	199
Totale	849	574	1423

Fr. rel. osservate	primo	non primo	Totale
no diploma	0,134	0,061	0,195
diploma	0,382	0,283	0,665
laurea	0,081	0,059	0,140
Totale	0,597	0,403	1,000

Fr. rel. teoriche	primo	non primo	Totale
no diploma	0,117	0,079	0,195
diploma	0,397	0,268	0,665
laurea	0,083	0,056	0,140
Totale	0,597	0,403	1,000

Fr. ass. teoriche	primo	non primo
no diploma	166	112
diploma	564	382
laurea	119	80

Statistica Q	primo	non primo
no diploma	0,0027	0,0040
diploma	0,0006	0,0008
laurea	0,0001	0,0001
	Q	11,75
	p-value	0,0028

RIFIUTO SE $Q > \chi^2_{0,95}(2)$

$11{,}75 > 5{,}9915$ VERO

RIFIUTO L'INDIPENDENZA

P-VALUE $\bar{\alpha}$: $\quad Q = \chi^2_{1-\bar{\alpha}}(2) \Rightarrow \bar{\alpha} = 1 - F_{\chi^2(2)}(Q) \simeq 0{,}0028$

$\Rightarrow$ H_0 RIFIUTATA PER QUASI QUALSIASI α

```
> 1-pchisq(11.75,2)
[1] 0.002808794
```