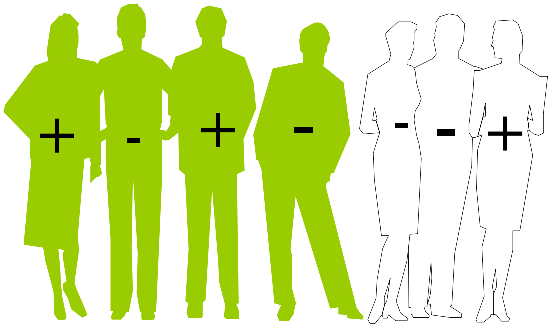
Il chi-quadro di Mantel-Haenszel

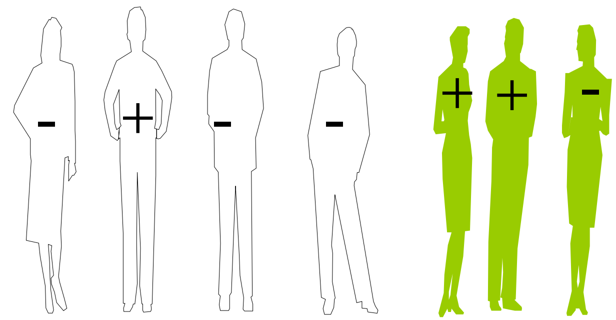
- Obiettivo: confrontare due proporzioni, studiare il legame in
  - **presenza di un fattore di stratificazione**

- Assunzioni:
  - Campioni indipendenti
  - Distribuzione binomiale

Quesito



Gruppo 1
N=7
X=numero di positivi=3



Gruppo 2
N=7
X=numero di positivi=3

La proporzione di positivi è uguale nel gruppo 1 e nel gruppo 2 considerando anche il numero di "bianchi" e il numero di "verdi"?

Ipotesi
$$\begin{cases} H_0 \quad p_1 = p_2 \\ H_1 \quad p_1 \neq p_2 \end{cases}$$

Con il metodo di Mantel-Haenszel si "aggiusta" il confronto tra i due gruppi per una terza variabile.

L'aggiustamento è ottenuto mediante la determinazione del chi-quadro delle tavole 2x2 di ogni strato, pesato per la dimensione del campione di ciascuno strato.

Strato i- esimo

Criterio 1

Criterio2	gruppo1	gruppo2	Totale
Positivo	a_i	b_i	$(a_i+b_i)=R1_i$
Negativo	c_i	d_i	$(c_i+d_i)=R2_i$
Totale	$(a_i+c_i)=C1_i$	$(b_i+d_i)=C2_i$	N_i

Si calcola il valore atteso per la cella a_i

$$E(a_i) = (R1_i \times C1_i) / N_i$$

Si calcola la varianza per la cella a_i

$$Var(a_i) = \frac{R1_i \times R2_i \times C1_i \times C2_i}{N_i^2 (N_i - 1)}$$

in ogni strato

La statistica test valuta contemporaneamente i risultati di tutti gli strati

$$X^2_{CMH} = \frac{\left(\left|\Sigma\, a_i - \Sigma\, E(a_i)\right| - \frac{1}{2}\right)^2}{\Sigma Var(a_i)}$$

La <u>distribuzione</u> della statistica test è X^2 con 1 grado di libertà

<u>Regola di decisione</u>:

Individuo il valore tabulato del X^2 con 1 grado di libertà e livello di significatività 0,05.
Se il $X^2_{cmh} > X^2_{0,05}$ (valore tabulato) **<u>rifiuto H_0</u>**

Il test di Mantel-Haenszel può essere utilizzato anche per la verifica dell'ipotesi sul odds-ratio

$$H_0 \; : \; OR = 1$$

$$H_1 \; : \; OR \neq 1$$

In uno studio caso – controllo sono state rilevate informazioni per valutare l'associazione tra l'assunzione di alcool e tumore all'esofago in soggetti fumatori e non fumatori.

Come cambia il rischio associato al consumo di alcool considerando che un individuo può essere fumatore?

309 **casi** sono fumatori di questi 265 assumono alcool
208 **controlli** sono fumatori di questi 151 assumono alcool
Sono anche bevitori:
63 **casi** non fumatori
107 **controlli** non fumatori

Fumatori (f)			
	Casi	Controlli	Totale
Alcool si	265	151	416
Alcool no	44	57	101
Totale	309	208	517
Non fumatori (nf)			
Alcool si	63	107	170
Alcool no	63	136	243
Totale	126	243	369

OR f = ad/bc =
265*57 / 151*44=
=2,27

OR nf = ad/bc =
63*136 / 107*63=
=1,27

E' possibile determinare un unico OR che ci dia informazioni sul rischio di essere bevitori e anche fumatori?

L'odds ratio di Mantel-Haenszel è dato da

$$\mathbf{OR_{MH}} = \frac{\Sigma(\mathbf{a_i d_i} / \mathbf{N_i})}{\Sigma(\mathbf{b_i c_i} / \mathbf{N_i})}$$

Nel nostro esempio:

$$OR_{MH} = \frac{[(265*57)/517] + [(63*136)/369]}{[(44*151)/517] + [(63*107)/369]} = \frac{52,44}{31,12} = 1,69$$

Il test di significatività per verificare l'ipotesi H_0 OR=1 è sempre un test X^2

$$X^2 = \frac{[\Sigma a_i - \Sigma E(a_i)]^2}{\Sigma Var(a_i)}$$

I valori attesi e la varianza sono:

$\Sigma E(a_i)=E(a_1)+E(a_2)=$

$=[(a_1+b_1)(a_1+c_1)/N_1]+[(a_2+b_2)(a_2+c_2)/N_2]=$

$=248{,}63+58{,}05=306{,}68$

$\Sigma Var(a_i)=Var(a_1)+Var(a_2)=$

$=\{[(a_1+b_1)(c_1+d_1)(a_1+c_1)(b_1+d_1)] / N_1^2 (N_1-1)\}+$

$+\{[(a_2+b_2)(c_2+d_2)(a_2+c_2)(b_2+d_2)] / N_2^2 (N_2-1)\} =$

$=19{,}58+20{,}67=40{,}25$

Applicando ai dati del nostro esempio il metodo di Mantel-Haenszel si ha:

$$X^2 = \frac{[\Sigma a_i - \Sigma E(a_i)]^2}{\Sigma Var(a_i)} = \frac{(328\text{-}306{,}68)^2}{40{,}25} = 10{,}77$$

p<0,001; l'OR è statisticamente significativo

L'intervallo di confidenza può essere calcolato con il metodo di Miettinen:

$$OR^{\left(1 \pm \frac{z_{1-\alpha/2}}{\sqrt{\chi^2}}\right)}$$

Limiti di confidenza al 95%

$$1{,}69^{(1 - 1{,}96\,/\,3{,}28)} = 1{,}23$$

$$1{,}69^{(1 + 1{,}96\,/\,3{,}28)} = 2{,}32$$

La probabilità di tumore all'esofago può anche essere predetta attraverso la regressione logistica

ln(***p*** avere tumore / ***q*** non avere tumore)=

b_0(parametro che indica il rischio di tumore senza altri fattori)+

+b_1(parametro che "pesa" la variabile)*fumo+

+b_2*alcool

Alcool e fumo non sono variabili quantitative!

Si trasformano in variabili apparentemente quantitative assegnando arbitrariamente ***valore 1*** alla ***presenza*** del fattore di rischio e *valore 0* all'*assenza* del fattore di rischio.
Queste variabili vengono comunemente chiamate variabili "dummy".

Il rischio di tumore in relazione all'abitudine all'alcool e al fumo può essere valutata con la ***regressione logistica***

$$y=b_0+b_1x_1+b_2x_2+\ldots+b_nx_n$$

Dove $$y = \ln\left(\frac{p}{q}\right)$$

Trasformazione logit

Questo rapporto è l'*odd*

Semplificando con una sola variabile: $y=b_0+b_1x_1$

$$p = \frac{\exp(b_0+b_1x_1)}{1+\exp(b_0+b_1x_1)} \Rightarrow \mathrm{OR} = \exp b_1$$