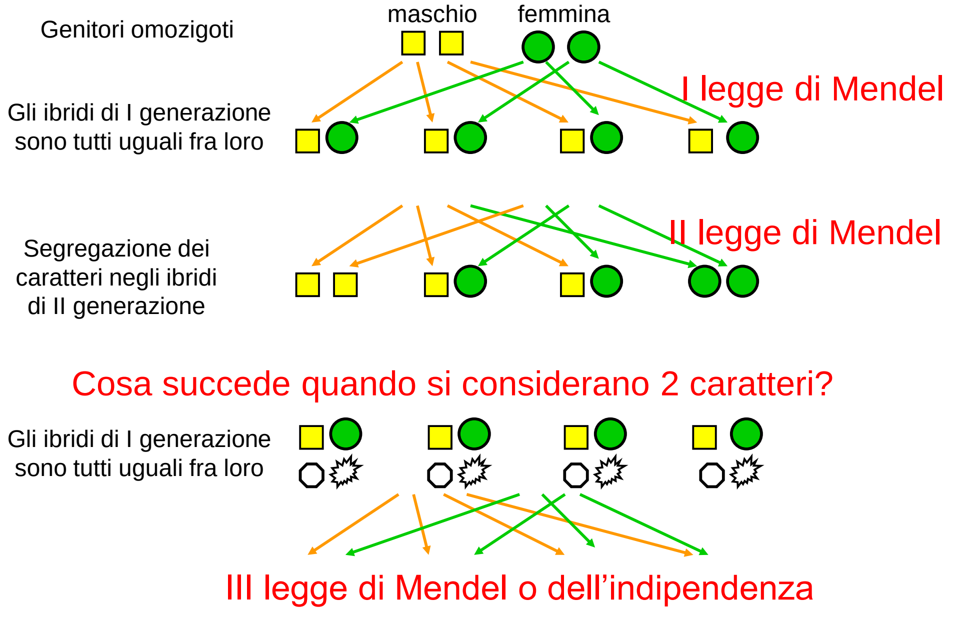
Test d'ipotesi: confronto fra proporzioni

- Prof. Giuseppe Verlato
- Sezione di Epidemiologia e Statistica Medica, Università di Verona

Richiami di genetica



Esperimento di Mendel:
incrocio di piselli lisci e gialli (caratteri dominanti) e rugosi e verdi (caratteri recessivi),
e incrocio degli ibridi di I generazione.

	giallo	verde	
Liscio	315	108	423
Rugoso	101	32	133
	416	140	556

Quanti sono gli ATTESI sotto l'ipotesi di indipendenza statistica?

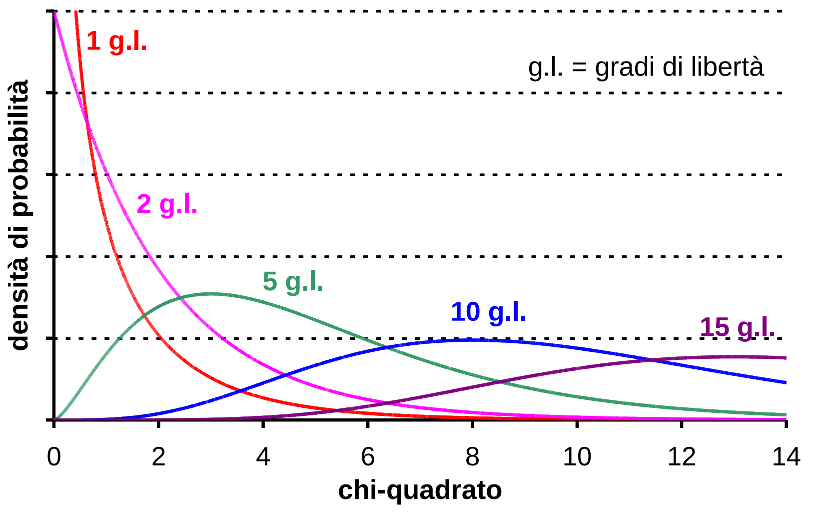
Attesi nella prima cella = p (liscio ∩ giallo) * N =
= p(liscio) * p(giallo) * N = (423/556) * (416/556) * 556 =
= 423 * 416 / 556 = 316,5

OSSERVATI	giallo	verde	
Liscio	315	108	423
Rugoso	101	32	133
	416	140	556

ATTESI	giallo	verde	
Liscio	316,5	106,5	423
Rugoso	99,5	33,5	133
	416	140	556

A occhio l'ipotesi di indipendenza statistica tra le caratteristiche della superficie (liscia / rugosa) e il colore (giallo / verde) è verificata: i caratteri si segregano indipendentemente

Per rispondere a questo quesito **IN MODO SCIENTIFICO**, si deve ricorrere al test del chi-quadrato, basato sulla distribuzione omonima.



Test del chi-quadrato

$$\chi^2 = \Sigma \text{ (osservati – attesi)}^2 \text{ / attesi}$$

$$\begin{cases} H_0\text{: le due variabili sono statisticamente indipendenti} \\ H_1\text{: le due variabili sono statisticamente dipendenti} \end{cases}$$

Livello di significatività = 5%

Gradi di libertà =(n° righe - 1) * (n° colonne - 1) =(2-1)*(2-1) =1*1 =1

Soglia critica = $\chi^2_{1,\,0,05}$ = 3,84

$$\chi^2 = \frac{(315-316,5)^2}{316,5} + \frac{(108-106,5)^2}{106,5} + \frac{(101-99,5)^2}{99,5} + \frac{(32-33,5)^2}{33,5}$$

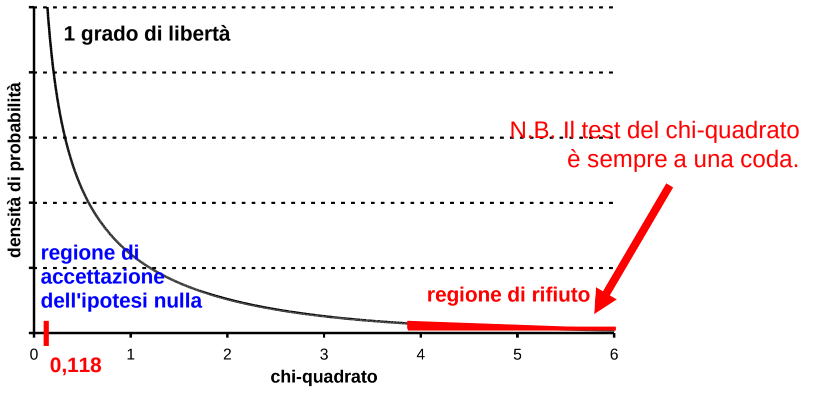
$$= 0,007 + 0,021 + 0,023 + 0,067 = 0,118$$

χ^2 osservato < soglia critica

0,118 3,84

Accetto H_0

I caratteri “caratteristiche della superficie” e “colore” si segregano indipendentemente l’uno dall’altro (**III legge di Mendel**)



Calcolo dei GRADI DI LIBERTA'
nel test del CHI-QUADRATO

?	?	423
?	111	133
140	416	556

		416-111	
	?	305	423
133-111	22	111	133
	140	416	556

	140-22		
423-305	118	305	423
	22	111	133
	140	416	556

In una TABELLA 2*2
GRADI DI LIBERTA' = 1

Calcolo dei GRADI DI LIBERTA' nel test del CHI-QUADRATO

?	?	?	463
?	72	56	193
100	140	416	656

	100-65	140-72	416-56	
463-68-360	35	68	360	463
193-72-56	65	72	56	193
	100	140	416	656

In una TABELLA 3*2
GRADI DI LIBERTA' = 2

Calcolo dei GRADI DI LIBERTA' nel test del CHI-QUADRATO

?	?	423
?	111	133
140	416	556

gradi di liberta'= 1

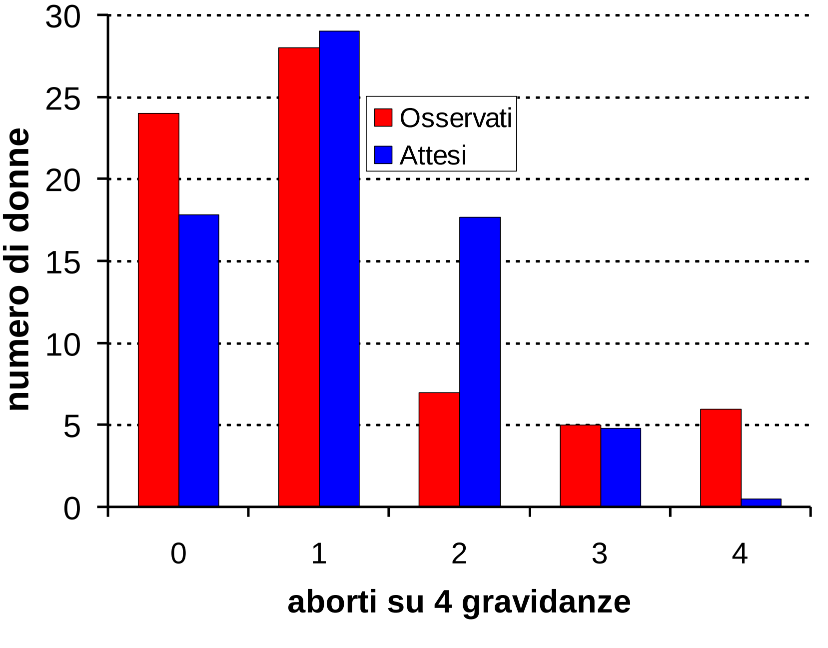
?	?	?	463
?	72	56	193
100	140	416	656

gradi di liberta'= 2

Esiste una formula unificante?
GRADI DI LIBERTA' =
=(n RIGHE - 1) (n COLONNE - 1)

- Il test del chi-quadrato è un test approssimato.
- In una tabella 2*2 può essere utilizzato quando nessun valore atteso è inferiore a 5.
- Altrimenti va effettuato il test esatto di Fisher.

Test del chi-quadrato per la bontà dell'adattamento



Test del chi-quadrato per la bontà dell'adattamento

$$\chi^2 = \Sigma \text{ (osservati – attesi)}^2 \text{ / attesi}$$

H_0: i dati osservati seguono la distribuzione binomiale
H_1: i dati osservati non seguono la distribuzione binomiale

Livello di significatività = 5%
Gradi di libertà =n° celle - n° parametri = 5-2 = 3

Soglia critica = $\chi^2_{3,\ 0,05}$ = 7,81

$$\chi^2 = \frac{(24-17,9)^2}{17,9} + \frac{(28-29,1)^2}{29,1} + \frac{(7-17,7)^2}{17,7} + \frac{(5-4,8)^2}{4,8} + \frac{(6-0,5)^2}{0,5}$$

$$= 2,12 + 0,04 + 6,48 + 0,01 + 61,96 = 70,60$$

χ^2 osservato > soglia critica → Rifiuto H_0
70,60 7,81

DATI APPAIATI

E quando la stessa variabile viene misurata più volte ? Ad esempio, presenza/assenza di rinite allergica in primavera/estate o in autunno/inverno ?

Test di McNemar = test per variabili nominali dicotomiche per 2 campioni dipendenti (2 misurazioni)

Test Q di Cochran = test per variabili nominali dicotomiche per 2 o più campioni dipendenti