IL TEST DEL CHI^2 (χ^2)

Consente di verificare ipotesi su:
a) **relazioni tra variabili nella popolazione**
b) **differenze tra popolazioni**

relative a: **distribuzioni di frequenza**

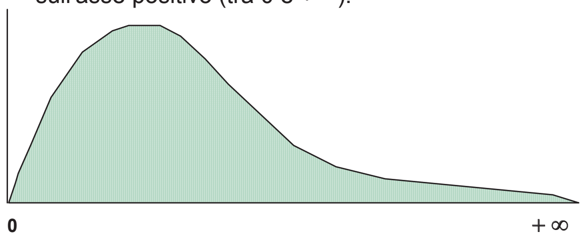
Livello di misura dei dati: **scala nominale o ordinale**

Ipotesi nulla: la distribuzione delle frequenze tra le diverse categorie è determinata dal caso
Ipotesi alternativa: la distribuzione delle frequenze tra le diverse categorie dipende dagli effetti delle variabili considerate

Statistica del test: $$\text{Chi}^2 = \sum \frac{(f_o - f_a)^2}{f_a}$$

dove: f_o = frequenza osservata
e f_a = frequenza attesa per effetto del caso (cioè se è vera l'ipotesi nulla)

Distribuzione di probabilità: funzione continua di probabilità che si definisce solo sull'asse positivo (tra 0 e + ∞):



La forma della curva varia al variare dei "gradi di libertà"

Limiti di applicabilità:
a) non più di due variabili nel disegno sperimentale
b) nessuna frequenza attesa inferiore a 1, non più del 20% delle frequenze attese inferiori a 5 (in caso contrario il test perde di **potenza,** cioè è più difficile rifiutare H_0).

DISEGNI CON UNA SOLA VARIABILE

Esempio 1

In una ricerca d'archivio è stato considerato il numero di incidenti gravi avvenuti in una determinata regione durante le ore notturne del sabato, nell'arco dell'ultimo anno. La distribuzione di frequenza degli incidenti nelle fasce orarie considerate è risultata la seguente:

N° INCIDENTI PER FASCIA ORARIA				
22 - 24	**24 - 02**	**02 - 04**	**04 - 06**	**TOTALE**
8	**5**	**13**	**10**	**36**

Si può sostenere che vi sono più incidenti in determinate fasce orarie?

(H_0): nella popolazione la distribuzione di frequenza degli incidenti tra le diverse fasce orarie è casuale

(H_1): nella popolazione la frequenza degli incidenti varia in funzione della fascia oraria (dipende dalla fascia oraria)

Calcolo delle frequenze attese

Le frequenze attese sono quelle che ci aspettiamo di ottenere **se è vera l'ipotesi nulla**. Se il numero degli incidenti non dipende dalla fascia oraria, allora la probabilità che accada un incidente è esattamente la stessa per ogni fascia oraria.
Essendo 4 le **categorie** considerate (4 diverse fasce orarie), la probabilità che accada un incidente sarà 1/4 (0.25). Essendo stati riscontrati in tutto 36 incidenti, il numero di incidenti per fascia oraria "atteso" in base all'ipotesi nulla è:
36 x 1/4 = 36/4 = **9**.

Quanto si discostano le frequenze osservate da quelle attese in base all'ipotesi nulla?

Le differenze tra frequenze osservate e attese sono abbastanza grandi per poter rifiutare l'ipotesi nulla?

Per rispondere alle precedenti domande bisogna calcolare la **statistica del test** e verificare se il risultato è **significativo**. Prima però bisogna stabilire il **livello di significatività** ed individuare la **zona di rifiuto**.

Livello di significatività e zona di rifiuto (o regione critica)

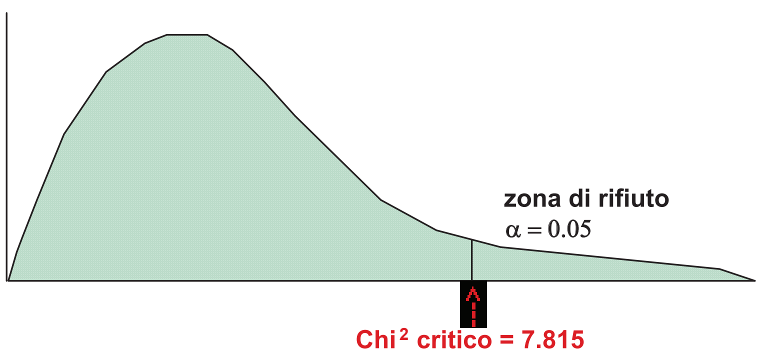
Scegliamo un livello di significatività α **= 0.05**. Rifiuteremo l'ipotesi nulla se il valore del Chi^2 calcolato avrà **probabilità di verificarsi per caso inferiore o uguale a 0.05**.
Utilizziamo quindi la distribuzione teorica di probabilità del Chi^{2}riportata sulle tavole. Come detto precedentemente, però, la curva del Chi^{2}varia al variare dei **"gradi di libertà"** (si tratta di una "famiglia" di curve), dobbiamo quindi calcolare i "gradi di libertà" del nostro disegno sperimentale per individuare la particolare distribuzione teorica di riferimento.

Calcolo dei gradi di libertà

Gradi di libertà: numero di valori **indipendenti** che generano una distribuzione.

Nel caso del nostro esempio abbiamo 4 valori (le frequenze di incidenti nelle 4 fasce orarie) la cui somma è 36. Di questi 4 valori, 3 sono liberi di variare (cioè sono indipendenti), mentre l'ultimo è vincolato dal fatto che la somma deve dare 36.
Es. 10+10+10+ ..... = 36 (l'ultimo è necessariamente 6)
9+12+15 + ... = 36 (l'ultimo è necessariamente 0)
Nei disegni con una sola variabile, il numero dei gradi di libertà è quindi uguale a: **numero di categorie - 1**.

Nel nostro caso g.d.l. = 4 - 1 = 3; sulla tavola del Chi^2, per 3 gradi di libertà, troviamo che il valore critico allo 0.05 è **7.815**. Qualsiasi valore calcolato che cada "al di là" di 7.815 avrà probabilità di verificarsi per caso inferiore a 0.05 e porterà al rifiuto dell'ipotesi nulla.

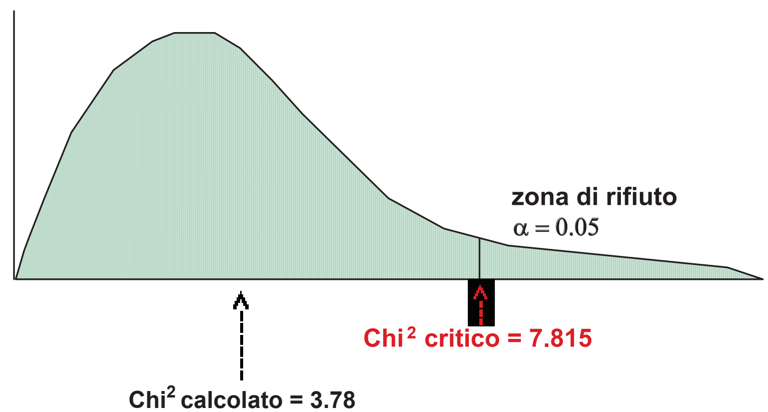


Calcolo della statistica del test

$$\text{Chi}^2 = \sum \frac{(f_o - f_a)^2}{f_a}$$

Fasce orarie:	22 - 24	24 - 02	02 - 04	04 - 06	TOTALI
f_o	8	5	13	10	36
f_a	9	9	9	9	36
$f_o - f_a$	-1	-4	4	1	0
$(f_o - f_a)^2$	1	16	16	1	
$(f_o - f_a)^2/f_a$	0.11	1.78	1.78	0.11	3.78

$$\text{Chi}^2 = \sum \frac{(f_o - f_a)^2}{f_a} = 0.11 + 1.78 + 1.78 + 0.11 = 3.78$$



Il valore ottenuto non cade nella zona di rifiuto (**non è significativo**).

Conclusione

Non possiamo rifiutare l'ipotesi nulla. Dobbiamo quindi concludere che nella popolazione la distribuzione degli incidenti nelle diverse fasce orarie è casuale (oppure: la probabilità che accada un incidente non dipende dalla fascia oraria).

Esempio 2 (Test di bontà di adattamento)

In uno studio su un caso singolo di dislessia evolutiva, viene somministrata al bambino in esame una prova di dettato di brani. Gli errori commessi vengono classificati in base alla categoria grammaticale, considerando le seguenti 5 categorie: Nomi, Verbi, Aggettivi, Avverbi, Preposizioni.

La distribuzione degli errori di scrittura commessi dal bambino è la seguente:

Categorie di errori:	Nomi	Verbi	Aggettivi	Avverbi	Preposiz.	Totale
Frequenze:	11	16	15	24	6	72

La distribuzione degli errori di scrittura fornita dal manuale del test, relativa al campione normativo della stessa età del bambino, è la seguente:

Categorie di errori:	Nomi	Verbi	Aggettivi	Avverbi	Preposiz.	Totale
Percentuali	32%	45%	6%	15%	2%	100%

Si vuole verificare se la distribuzione degli errori di scrittura tra le diverse categorie, osservata nel bambino, si discosta significativamente da quella del campione normativo. In altre parole, la distribuzione del campione di errori considerato, si adatta a quella della popolazione di riferimento? (test di **bontà di adattamento**).

Si vuole inoltre sapere quali sono le categorie grammaticali dove il bambino commette un numero di errori significativamente più elevato rispetto al campione normativo.

(H_0): Nella popolazione da cui il bambino è estratto, la distribuzione degli errori è uguale a quella della popolazione di riferimento (qualsiasi differenza è dovuta al caso).

(H_1): Nella popolazione da cui il bambino è estratto, la distribuzione degli errori è significativamente diversa da quella della popolazione di riferimento (la differenza è significativa).

Calcolo delle frequenze attese

Se è vera l'ipotesi nulla e le due distribuzioni di errori (quella del campione e quella della popolazione di riferimento) non si differenziano tra loro, allora dovremmo aspettarci, nel campione, che la percentuale di errori per categoria sia la stessa riscontrata nella popolazione.

Categorie di errori:	Nomi	Verbi	Aggettivi	Avverbi	Preposiz.	Totale
Freq. oss.:	**11**	**16**	**15**	**24**	**6**	**72**
%popolaz.	**32%**	**45%**	**6%**	**15%**	**2%**	**100%**
Freq.attese	**23.04**	**32.4**	**4.32**	**10.8**	**1.44**	**72**

N.B. Ogni frequenza attesa è calcolata sulla base di una proporzione. Es. categoria Verbi: su 72 errori, in base all'ipotesi nulla il 45% dovrebbe appartenere alla categoria Verbi, quindi 45 : 100 = X : 72; X = 45 x 72 / 100 = **32.4**.

Livello di significatività e zona di rifiuto

α = 0.01 (vogliamo limitare il rischio di **errore di I tipo**)
Gradi di libertà = n. di categorie - 1 = 5 - 1 = **4**
Chi^2 critico = 13.277
Si rifiuta l'ipotesi nulla se il valore calcolato sarà $Chi^2 \geq 13.277$

Calcolo della statistica del test

La formula e le procedure di calcolo sono esattamente uguali a quelle dell'esempio precedente.

$$\text{Chi}^2 = \sum \frac{(f_o - f_a)^2}{f_a}$$

= (11 - 23.04)^2 / 23.04 + (16 - 32.4)^2 / 32.4 + (15 - 4.32)^2 / 4.32 + (24 - 10.8)^2 / 10.8 + (6 - 1.44)^2 / 1.44 =
= 6.29 + 8.30 + 26.40 + 16.13 + 14.44 = 71.56

Conclusione
Poichè il valore del Chi^2 calcolato (71.56) supera il valore critico del Chi^2 allo 0.01 (13.277) rifiutiamo l'ipotesi nulla e concludiamo, con una probabilità d'errore dell'1%, che la distribuzione degli errori di scrittura commessi dal bambino dislessico è significativamente diversa da quella del campione normativo.

N.B. Più del 20% delle frequenze attese avevano un valore inferiore a 5 (4.32 e 1.44). Un elevato numero di frequenze attese inferiori a 5 **riduce la potenza del test**, cioè rende **più difficile rifiutare l'ipotesi nulla**. Tuttavia dato che il risultato ottenuto è significativo (nonostante la ridotta potenza del test), questo aspetto non costituisce un problema nel nostro caso.

La seconda domanda posta dai ricercatori è: quali sono le categorie grammaticali dove il bambino commette più errori?

Si applica la formula di Haberman sui residui standardizzati:

$$\mathbf{R} = \frac{f_o - f_a}{\sqrt{f_a}}$$

La differenza è significativa se il risultato è superiore a 2.

Applichiamo la formula solo alle celle dove si riscontra un numero di errori superiore rispetto alle attese.
Aggettivi: 15 - 4.32/2.08 = **5.13**
Avverbi: 24 - 10.8/3.29 = **4.01**
Preposizioni: 6 - 1.44/1.20 = **3.80**

In tutti e tre i casi il numero di errori commesso è significativamente più elevato rispetto a quello atteso.

DISEGNI CON DUE VARIABILI

Viene condotto un sondaggio di opinione su una nuova proposta di legge. Le persone intervistate appartengono a tre diverse fasce di età: 18 - 30 anni (giovani), 31 - 55 anni (adulti), 56 - 70 anni (anziani). Dall'esame delle risposte del campione intervistato (N = 300), si ottiene la seguente **tabella di contingenza** (**tabella a doppia entrata**):

	FASCE D'ETA'			
OPINIONE	**Giovani**	**Adulti**	**Anziani**	**Totale**
Favorevole	76	40	24	140
Indifferente	10	68	14	92
Contrario	12	24	32	68
Totale	98	132	70	300

Si vuole verifcare se esiste una relazione tra l'età dei soggetti e l'opinione espressa nei confronti della proposta di legge
(α **= 0.05**).

(H_0): Non esiste nessuna relazione, nella popolazione, tra le opinioni sulla proposta di legge e l'età (la distribuzione delle frequenze è casuale)

(H_1): Esiste una relazione, nella popolazione, tra le opinioni sulla proposta di legge e l'età (le opinioni espresse sono influenzate dall'età)

Gradi di libertà, livello di significatività e zona di rifiuto

Nel test del Chi^2 su due variabili i gradi di libertà sono dati da: (numero di righe - 1) x (numero di colonne - 1) **senza tener conto dei totali**. Nel nostro caso:
(r - 1) x (c - 1) = (3 - 1) x (3 - 1) = 2 x 2 = **4**
Per α **= 0.05 e g.d.l. = 4**, sulla tavola del Chi^2 troviamo **Chi^2 critico = 9.488**.

Calcolo delle frequenze attese e della statistica del test

Nelle tabelle di contingenza il calcolo delle frequenze attese si effettua secondo la seguente formula:
Totale di riga x totale di colonna / totale generale

	FASCE D'ETA'			
OPINIONE	**Giovani**	**Adulti**	**Anziani**	**Totale**
Favorevole	76 **45.7**	40 **61.6**	24 **32.7**	140
Indifferente	10 **30.1**	68 **40.5**	14 **21.4**	92
Contrario	12 **22.2**	24 **29.9**	32 **15.9**	68
Totale	98	132	70	300

per $f_o = 76$ ---> $f_a = 140 \times 98 / 300 =$ **45.7**
per $f_o = 40$ ---> $f_a = 140 \times 132/300 =$ **61.6**
per $f_o = 10$ ---> $f_a = 92 \times 98 / 300 =$ **30.1**
per $f_o = 68$ ---> $f_a = 92 \times 132/300 =$ **40.5**

N.B. In una tabella con 4 gradi di libertà, le frequenze attese delle prime 4 celle (quelle bordate in neretto) devono essere calcolate nel modo indicato, le restanti (che sono vincolate dai totali marginali), possono essere determinate per sottrazione.
Si applica la consueta formula di calcolo:

$$\text{Chi}^2 = \sum \frac{(f_o - f_a)^2}{f_a} = \mathbf{86.9}$$

Poichè il valore del Chi^2 calcolato (86.9) è superiore al valore del Chi^2 critico allo 0.05 (9.488), possiamo rifiutare l'ipotesi nulla ed affermare (con una probabilità d'errore del 5%), che esiste una relazione, nella popolazione, tra la fascia d'età e l'opinione nei confronti della proposta di legge.

Volendo verificare quali celle hanno il peso maggiore nel determinare la significatività del risultato, possiamo applicare la formula di Haberman sui residui standardizzati (vedi esempio precedente).