Test di ipotesi

Che cosa sono le ipotesi e i test di ipotesi?

- Un’ipotesi è una formulazione di una congettura/ un’assunzione sulla popolazione che deve essere verificata e può essere vera o falsa
- Le conclusioni che traiamo sulla popolazione vengono valutate esaminando un campione estratto da quella popolazione
- Le ipotesi fatte sono relative a parametri di quella popolazione
- Il **test di ipotesi** è la procedura che ci permette di stabilire se l’ipotesi è vera o falsa

Ipotesi

Nei test di ipotesi le ipotesi sono due: l’**ipotesi nulla** e l’**ipotesi alternativa**.

L’ipotesi nulla è quella di uguaglianza o mancanza dell’effetto, quella alternativa è quella di differenza/diversità

Ipotesi

L'ipotesi nulla in genere è la negazione dell'ipotesi di ricerca (che coincide perciò con quella alternativa)

Es. Se vogliamo valutare se una moneta è truccata (perciò la probabilità che esca testa non è uguale alla probabilità che esca croce) allora formuleremo l'ipotesi come:

Ipotesi nulla: $p_{testa} = p_{croce}$ Ipotesi alternativa: $p_{testa} \neq p_{croce}$

Oppure

Ipotesi nulla: $p_{testa} = 0.5$ Ipotesi alternativa: $p_{testa} \neq 0.5$

Ipotesi

bilaterale

$$\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu \neq \mu_0 \end{cases}$$

unilaterale

$$\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu < \mu_0 \end{cases} \qquad \begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu > \mu_0 \end{cases}$$

H_0 è l'ipotesi nulla

H_1 è l'ipotesi alternativa

→

Preferiamo rifiutare l'ipotesi nulla a favore di quella alternativa

Ipotesi

bilaterale

$$\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu \neq \mu_0 \end{cases}$$

unilaterale

$$\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu < \mu_0 \end{cases} \qquad \begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu > \mu_0 \end{cases}$$

H_0 è l'ipotesi nulla

H_1 è l'ipotesi alternativa

Non rifiutare H_0 non significa che la condizione descritta dall'ipotesi nulla è vera, ma significa che non abbiamo sufficiente evidenza per rifiutarla!

La statistica test

- Per stabilire se rifiutare o non rifiutare l'ipotesi nulla dobbiamo calcolare la **statistica test**

- In generale la statistica test è calcolata come:

$$T = \frac{stima - valore\ supposto}{errore}$$

- Ogni statistica test ha una distribuzione nota

La statistica test della media

Consideriamo la seguente ipotesi:

«la pressione arteriosa media negli anziani è maggiore di 120»

Come è strutturato il test di ipotesi?

$$\begin{cases} H_0: \mu = 120 \\ H_1: \mu > 120 \end{cases}$$

La statistica test della media

Consideriamo la seguente ipotesi:

«la pressione arteriosa media negli anziani è maggiore di 120»

Che cosa facciamo per condurre la verifica di questa ipotesi??

1. Estraggo un campione dalla popolazione
2. Su tutte le unità valuto la pressione arteriosa
3. Calcolo la statistica test della media che è data da:

$$T = \frac{stima - valore\ supposto}{standard\ error} \rightarrow T = \frac{\bar{X} - 120}{\frac{\sigma}{\sqrt{n}}}$$

4. La statistica test per la media ha una distribuzione Gaussiana di media 0 e deviazione standard 1

La regione di accettazione e rifiuto

- Per capire se rifiutare o non rifiutare l'ipotesi nulla devo vedere se il valore assunto dalla statistica test appartiene alla **regione di accettazione** o alla **regione di rifiuto**
- La regione di accettazione è l'insieme di valori per cui non rifiuto H_0
- La regione di rifiuto è l'insieme di valori per cui rifiuto H_0

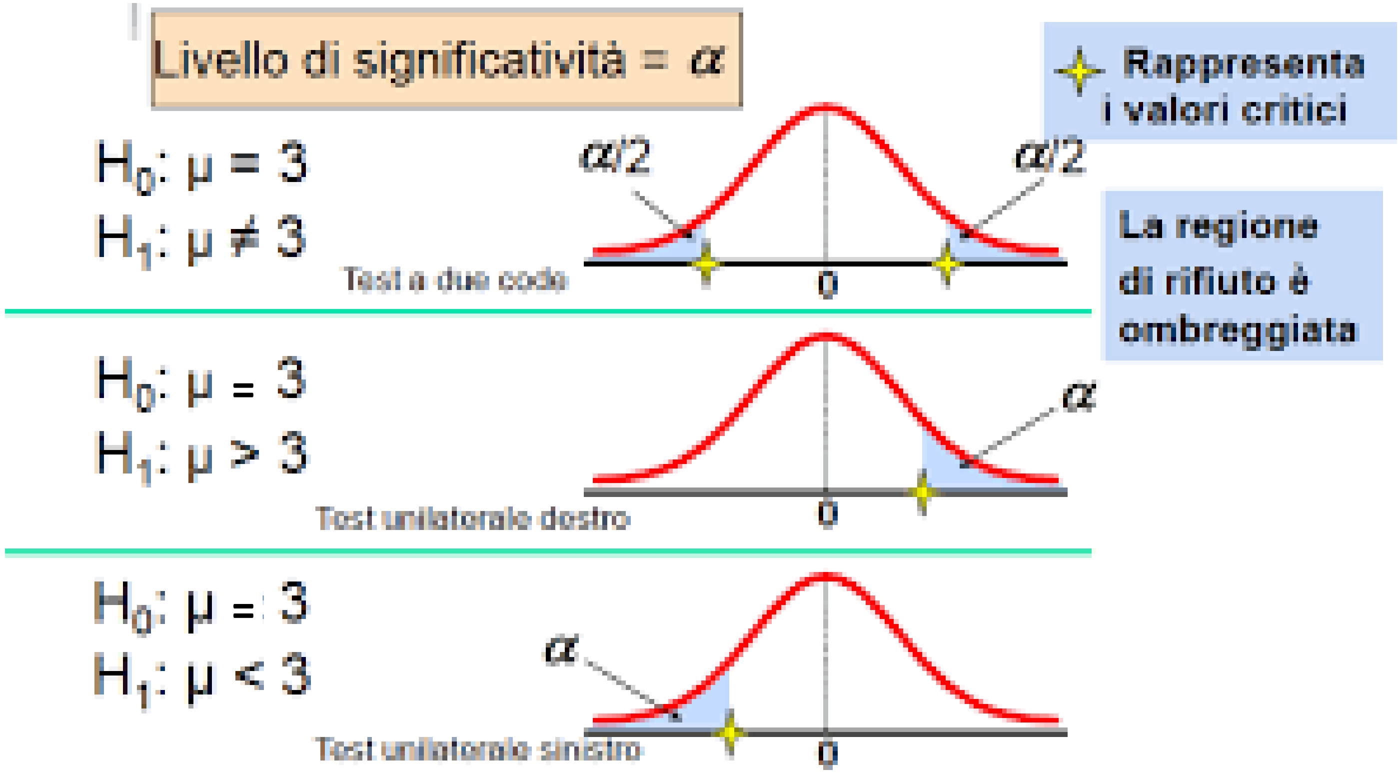
Livello di significatività α

- L'insieme dei valori che appartengono alla regione di rifiuto/di accettazione è definito dal **livello di significatività (α)**

- **Il livello di significatività è una probabilità, cioè la probabilità di rifiutare l'ipotesi nulla quando questa è vera**

- α viene detto anche **errore del 1°tipo**, e in genere assume valore 0.05 ma può assumere anche valore 0.10 o 0.01 in base all'entità dell'errore che siamo disposti a commettere

Livello di significatività α



La collocazione di α segue l'ipotesi alternativa, perciò se nell'ipotesi alternativa troviamo il ≠ posizioneremo metà errore nella coda sinistra e metà nella coda destra, se è presente il < allora α è tutto a sinistra, con il a destra come nell'immagine.

Se il valore della statistica test ricade nell'area bianca allora non rifiuto l'ipotesi nulla, se ricade nella zona celeste rifiuto l'ipotesi nulla

Errore del 2° tipo

- **errore del secondo tipo** viene indicato con **β**

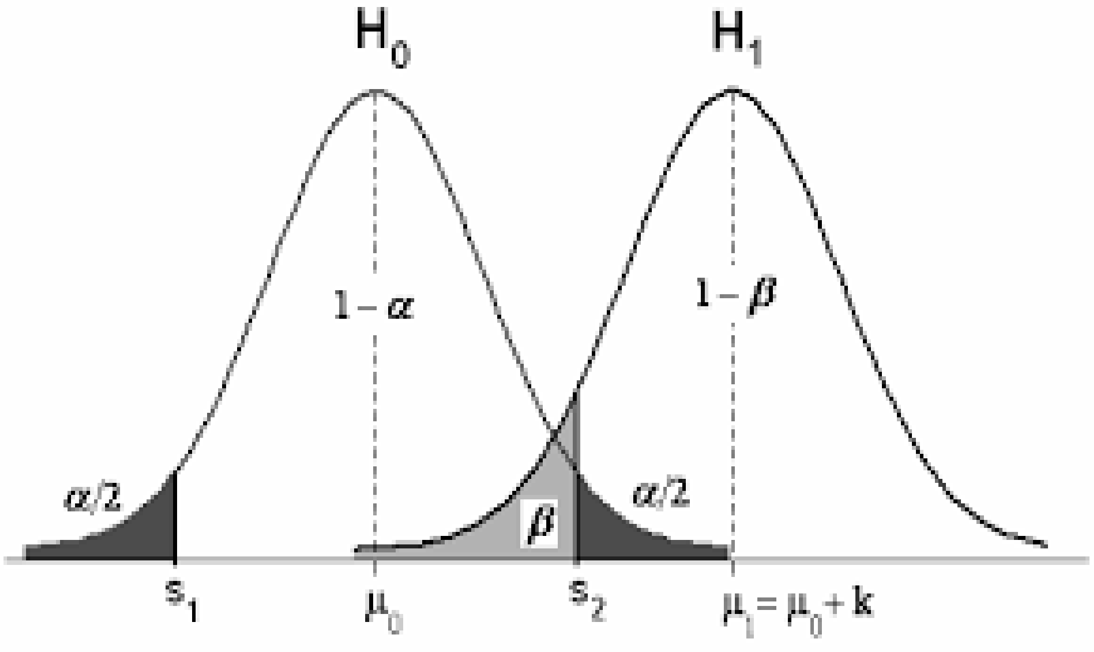
- indica la probabilità di accettare l'ipotesi nulla quando andrebbe rifiutata

- è un errore meno grave di quello del primo tipo

- È in genere più elevato rispetto a quello del primo tipo

Errore del 2° tipo

- in genere viene settato a 0.20, 0.15 o 0.10. Nella ricerca scientifica 0.20 è il valore più frequentemente utilizzato. Non dovremmo considerare valori più elevati di 0.20
- **β** non è noto quando conduciamo un test (infatti è un parametro che non interviene direttamente)
- 1-**β** viene detta **potenza** e questo è un parametro fondamentale da considerare nella stima della numerosità campionaria



Riassumendo..

Probabilità		Realtà	
		Ipotesi nulla vera	ipotesi alternativa vera
Decisione	non rifiuto l'ipotesi nulla	1-α	errore di II tipo β
	rifiuto l'ipotesi nulla	errore di I tipo α	potenza 1-β

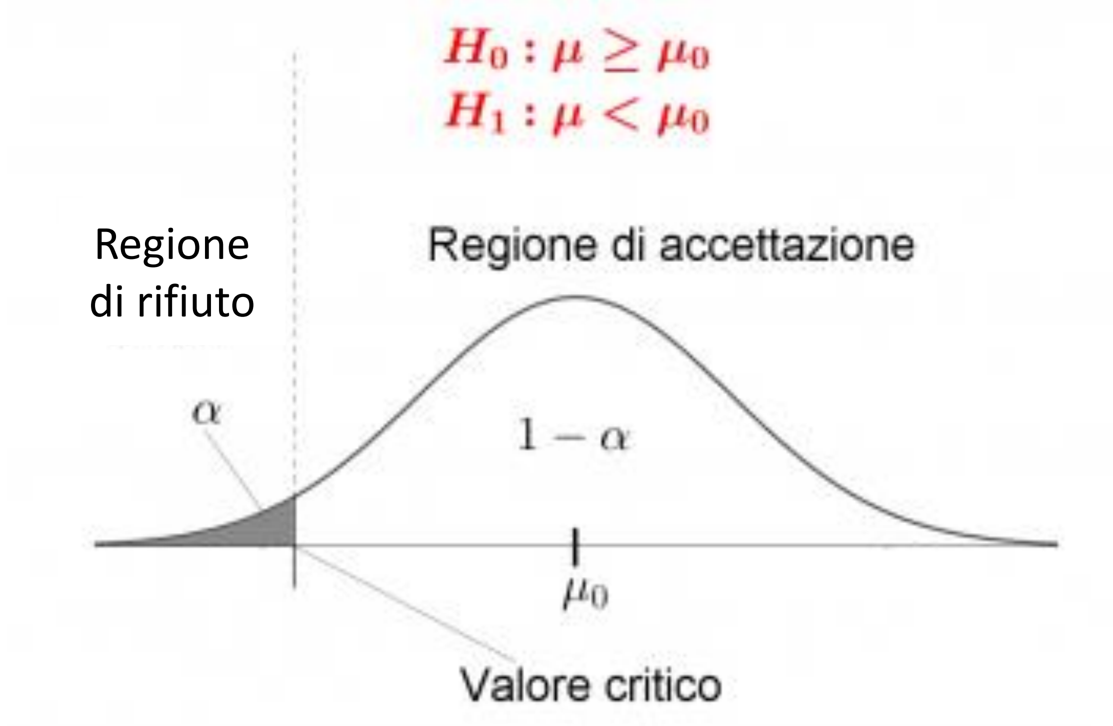
Regione di accettazione e rifiuto

Indichiamo con *z* il percentile che individua α, allora:

- Per ipotesi unilaterale a sinistra (<)

  Regione di accettazione: valori > z

  Regione di rifiuto: valori < z



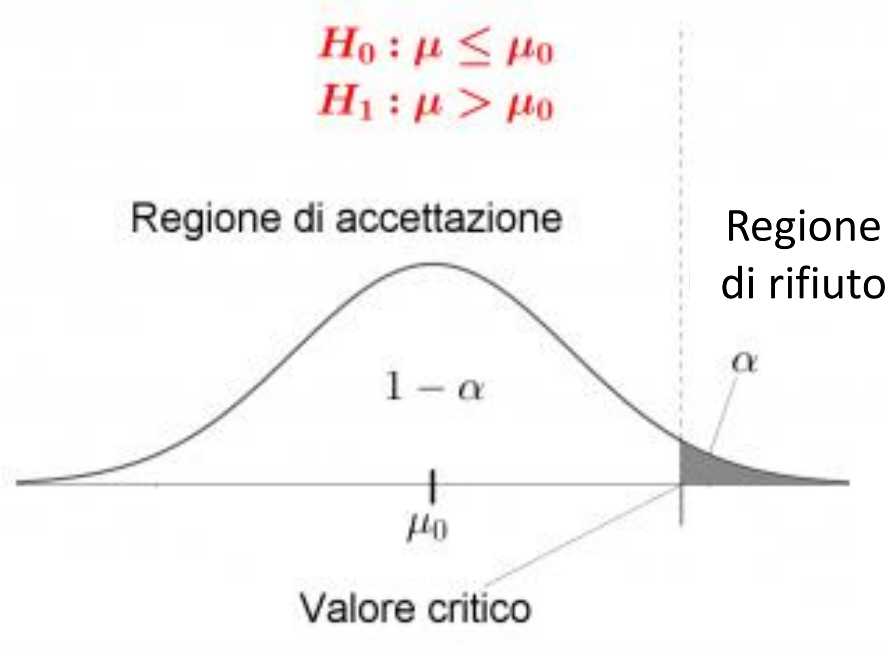
Regione di accettazione e rifiuto

Indichiamo con *z* il percentile che individua α, allora:

- Per ipotesi unilaterale a destra (>)

  Regione di accettazione: valori < z

  Regione di rifiuto: valori > z



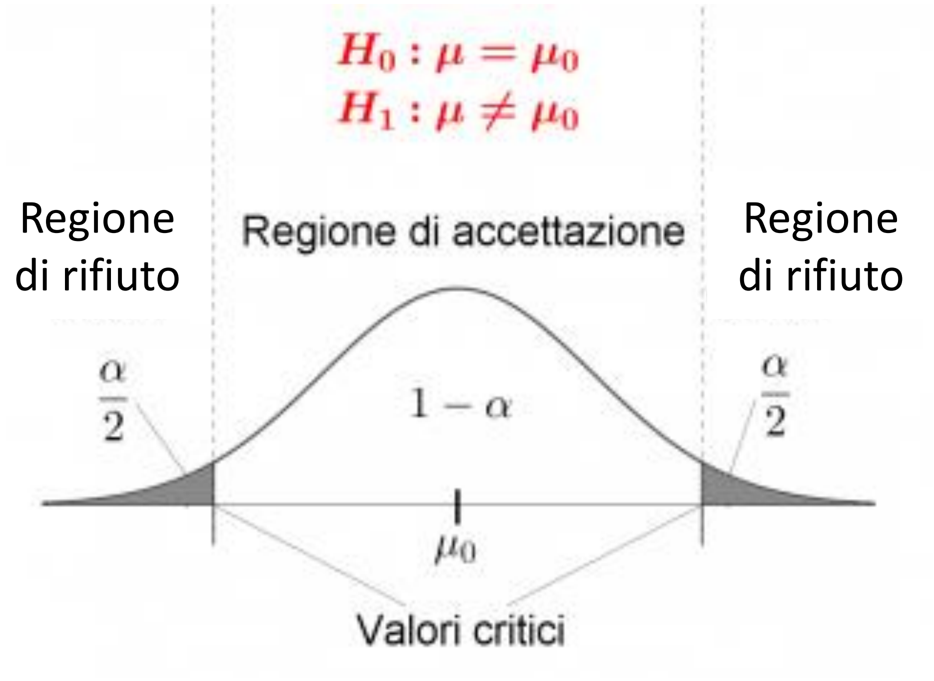
Regione di accettazione e rifiuto

Indichiamo con *z* il percentile che individua α, allora:

- Per ipotesi bilaterale (≠)

  Regione di accettazione: $z_1 <$ valori < z_2

  Regione di rifiuto: valori < z_1 e valori > z_2



La statistica test della media

Consideriamo la seguente ipotesi:

«la pressione arteriosa media negli anziani è maggiore di 120»

Che cosa facciamo per condurre la verifica di questa ipotesi??

1. Stabilire il livello di significatività ad esempio 0.05
2. Individuare il valore critico

z	1.282	1.645	1.960	2.326	2.576	3.090	3.291	3.891	4.417
$\Phi(z)$	0.90	0.95	0.975	0.99	0.995	0.999	0.9995	0.99995	0.999995
$2[1-\Phi(z)]$	0.20	0.10	0.05	0.02	0.01	0.002	0.001	0.0001	0.00001

3. Costruire le due regioni di accettazione/rifiuto e vedo a quale delle due appartiene la statistica test:
   - Se $z<1.645$ non rifiuto l'ipotesi nulla
   - Se $z\geq1.645$ rifiuto l'ipotesi nulla

p-value e decisione del test

- Per stabilire se accettare o rifiutare l’ipotesi nulla possiamo anche calcolare il ***p-value***

- Il p-value è la probabilità di osservare quel risultato ottenuto o un valore più estremo

- Il p-value è l’area individuata dal valore calcolato della statistica test. Anche il p-value viene posizionato in base all’ipotesi alternativa.

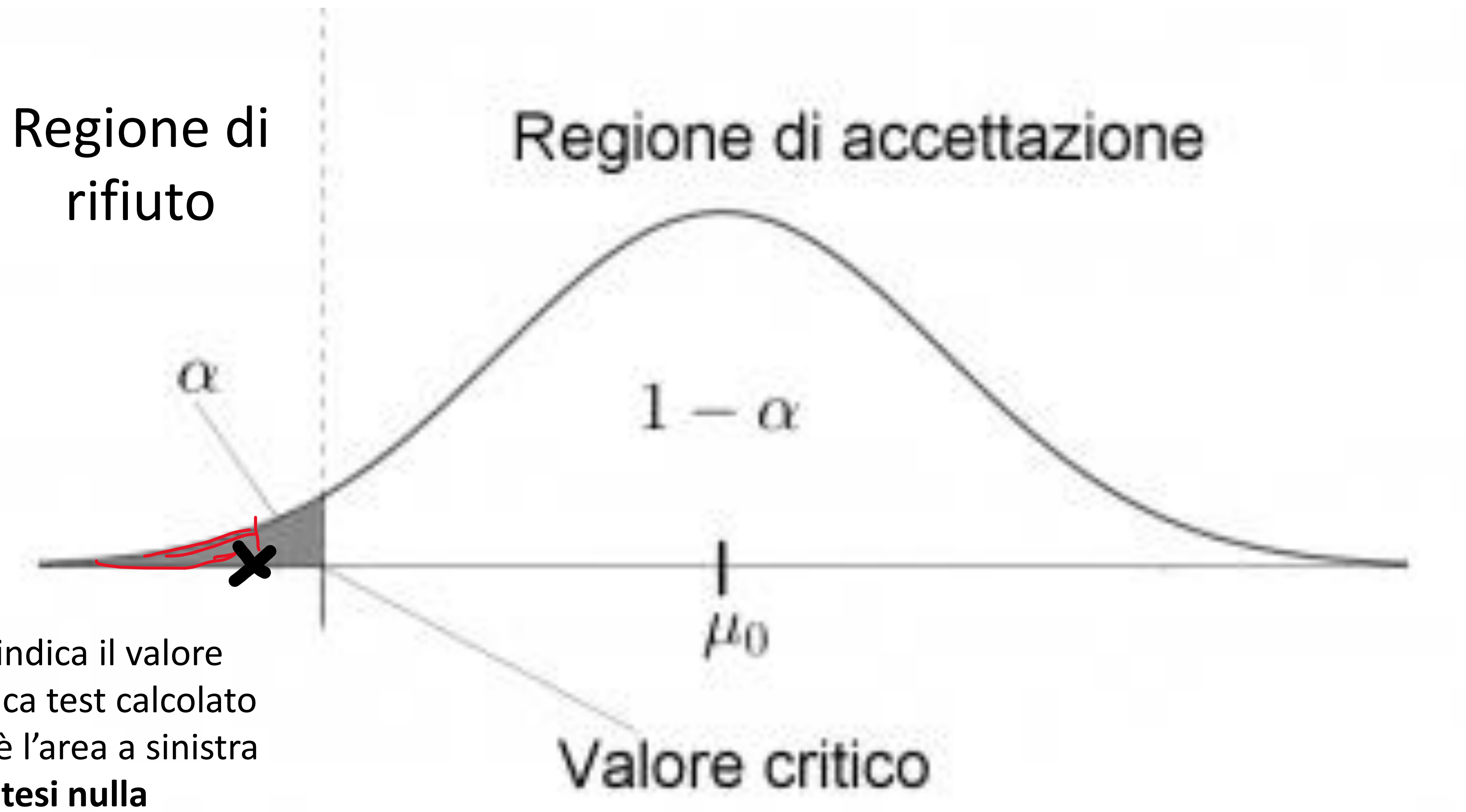
p-value e decisione del test

- Il p-value deve essere confrontato con l’errore del primo tipo/livello di significatività e se:

$p<\alpha \quad \rightarrow \quad$ rifiuto l’ipotesi nulla

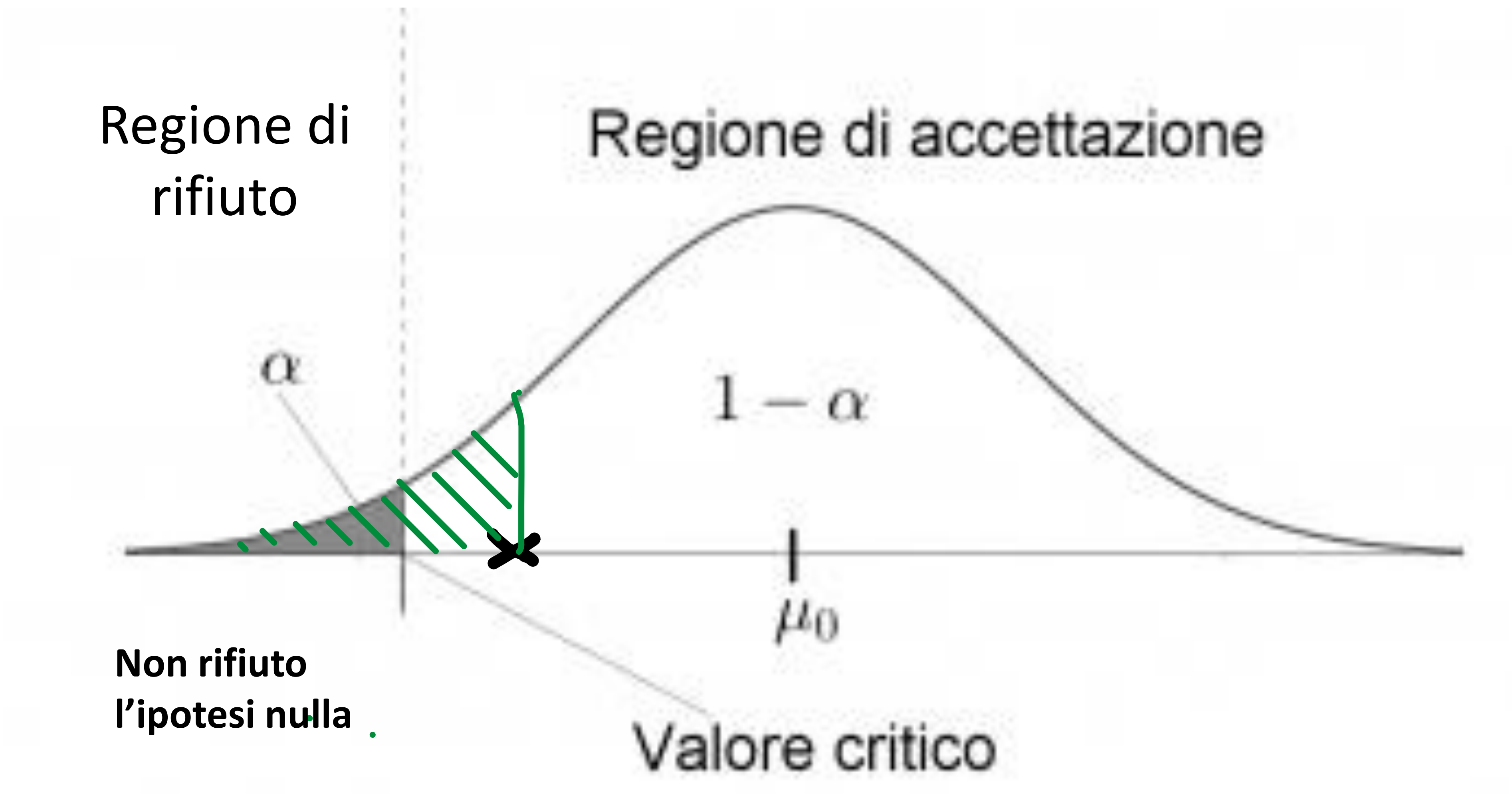
$p\geq \alpha \quad \rightarrow \quad$ non rifiuto l’ipotesi nulla

p-value e decisione del test

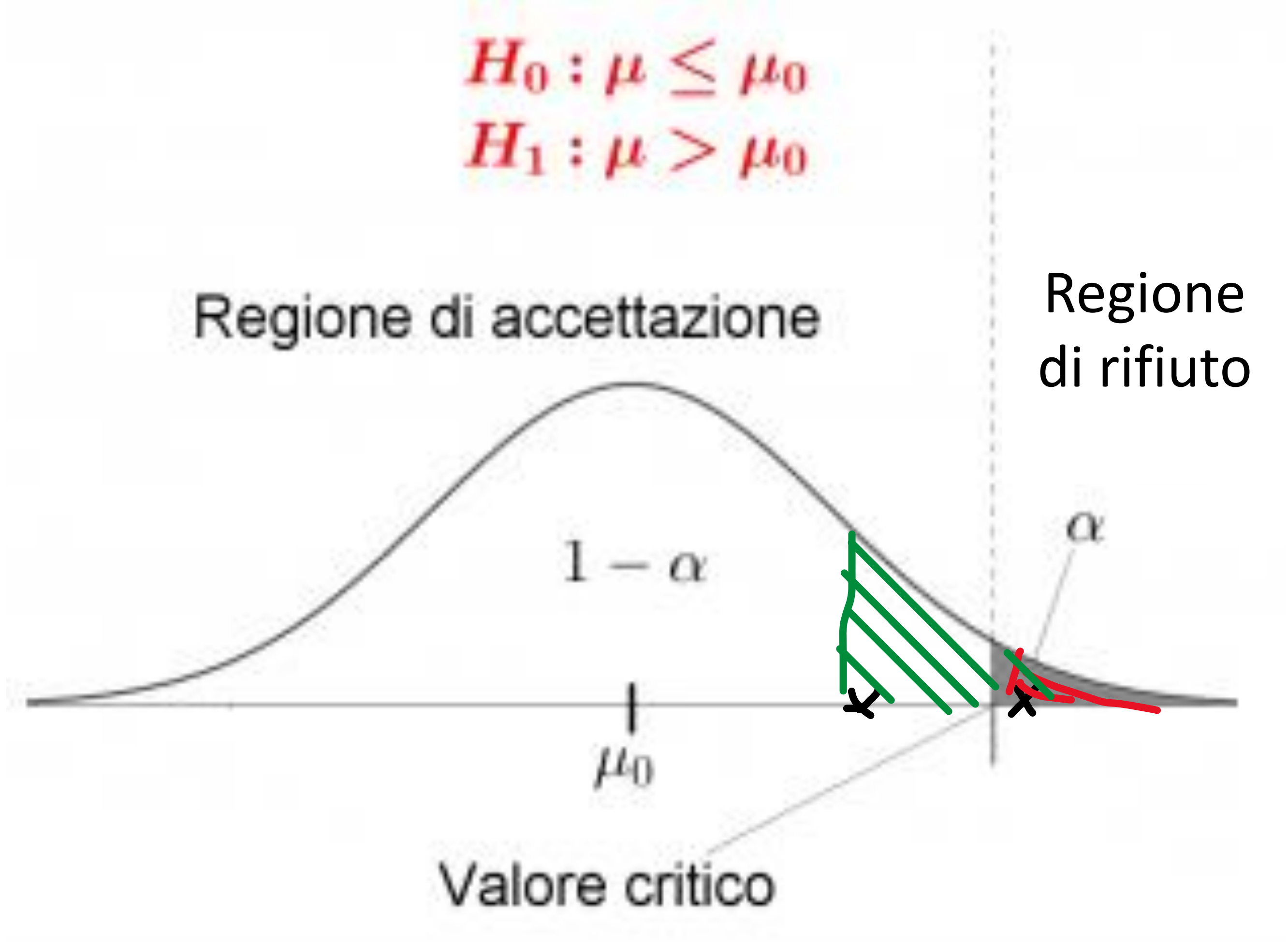


La crocetta indica il valore della statistica test calcolato e il p-value è l’area a sinistra
Rifiuto l’ipotesi nulla

p-value e decisione del test



p-value e decisione del test



Principali ipotesi:

- **Ipotesi sulla distribuzione della variabile**
- **Ipotesi sulle varianze**
- **Ipotesi riguardanti le «frequenze»:**
  - Ipotesi sulle proporzioni
    - → Vogliamo valutare se il verificarsi di un evento segue una certa proporzione
  - Ipotesi di associazione
    - → Vogliamo valutare la relazione tra due variabili qualitative
- **Ipotesi riguardanti le medie**
  - Ipotesi sulla media
    - → Vogliamo vedere se la media di una variabile è uguale o meno a un valore prefissato
  - Ipotesi sulla differenza tra medie
    - → Vogliamo vedere se le medie di due o più gruppi sono uguali oppure no
- **Ipotesi di correlazione**

Test di ipotesi

I test di ipotesi possono essere suddivisi in due gruppi:

Test parametrici

Test non parametrici o distribution free

I test parametrici sono particolari test per i quali, prima di applicarli, devono essere verificate delle ipotesi stabilite a priori generalmente sulla distribuzione di una variabile.

I test non parametrici invece non hanno ipotesi a priori da verificare e possono essere sempre applicati.

Test di ipotesi

Tuttavia nel caso in cui le ipotesi dei test parametrici siano soddisfatte è preferibile utilizzare questi test data la maggior potenza, cioè a parità di numerosità campionaria sono in grado di evidenziare differenze più piccole come significative.

Ipotesi sulla distribuzione

- Per capire preliminarmente quale potrebbe essere la distribuzione di una variabile potremmo utilizzare l’istogramma

- La distribuzione di una variabile viene verificata attraverso il test di **Kolmogorov-Smirnov.** In genere questo test viene utilizzato per valutare se la distribuzione ha una forma Gaussiana oppure no (ipotesi preliminare di tutti i test parametrici che studieremo).

Ipotesi sulla distribuzione

- L’ipotesi nulla è: la distribuzione empirica/osservata segue la distribuzione teorica presupposta

- L’ipotesi alterntiva è: la distribuzione empirica/osservata non segue la distribuzione teorica presupposta

Ipotesi sulle varianze

L'omogeneità/omoschedasticità delle varianze viene valutata tramite il test di Bartlett o di Levene oppure attraverso una regola pratica

$$\frac{\sigma_{max}^2}{\sigma_{min}^2} < 2$$

Anche questa ipotesi viene utilizzata come ipotesi preliminare di alcuni test parametrici

Ipotesi sulle varianze

- L’ipotesi nulla è: le varianze sono omogenee

- L’ipotesi alternativa è: le varianze non sono omogenee

Ipotesi sulle frequenze

1. Ipotesi sulle proporzioni
2. Ipotesi di associazione

Ipotesi sulle frequenze

Proporzioni

- Uguaglianza a una data proporzione
- Differenza di proporzioni

Associazione tra due variabili

- Test di McNemar
  (tabelle di contingenza 2x2)
- Chi-quadrato
  (tabelle di contingenza 2x2 più grandi di 2x2)
- Chi-quadrato con correzione di Yates
  (tabelle di contingenza 2x2)
- Test esatto di Fisher
  (tabelle di contingenza 2x2)
- Chi-quadrato per il trend
  (tabelle di contingenza 2xK)

Ipotesi sulla proporzione

Vogliamo valutare se il verificarsi di un evento segue una certa proporzione prefissata

L'evento è dicotomico perciò può essere sintetizzato come successo/insuccesso, si/no, avvenuto/non avvenuto

Questa proporzione viene calcolata come:

$$\frac{n° \; di \; volte \; in \; cui \; si \; verifica \; l'evento}{n}$$

Ipotesi sulla proporzione

Per verificare se l'evento rispetta una proporzione prefissata abbiamo due possibilità in base al fatto che il campione è piccolo o grande:

- Se il campione è piccolo possiamo utilizzare il **test binomiale** (confronto diretto del numero di eventi osservato con la distribuzione binomiale)
- Se il campione è sufficientemente grande allora possiamo utilizzare la seguente formula:

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$

Ipotesi sulla proporzione

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$

$\rightarrow$ Confrontiamo questa statistica test con i valori critici della normale standardizzata!

Dove:

$\hat{p}$= probabilità/proporzione stimata

p_0= proporzione supposta

$\hat{p}(1 - \hat{p})$ = varianza della proporzione

Ipotesi sulla proporzione

Esempio: voglio valutare se la moneta è truccata e la probabilità che esce testa è pari a 0.25; considerando che ho lanciato la moneta 22 volte ed è uscita testa 5 volte.

- Con il test binomiale inserisco questi valori nella funzione della distribuzione binomiale e calcolo i p-value
- Con la formula

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \rightarrow \hat{p} = \frac{5}{22} = 0.23 \rightarrow \frac{0.23 - 0.25}{\sqrt{\frac{0.23(1-0.23)}{22}}}$$

Ipotesi sulla proporzione

Nel caso in cui volessimo comparare due proporzioni e capire se c'è differenza allora potremmo utilizzare la formula:

$$z = \frac{(\widehat{p_1} - \widehat{p_2}) - (p_1 - p_2)}{\sqrt{\frac{\widehat{p_1}(1-\widehat{p_1})}{n_1} + \frac{\widehat{p_2}(1-\widehat{p_2})}{n_2}}}$$

Esempio: vogliamo verificare se la probabilità che esca testa dipende dal fatto che i lanci della moneta vengono fatti con una moneta d'oro o di argento

Ipotesi di associazione

- Coinvolgono due variabili di tipo qualitativo

- La tecnica descrittiva adeguata per sintetizzare i dati è la tabella di contingenza

- La tecnica grafica più adatta è il diagramma a barre

Ipotesi di associazione

Il fumo è associato con al cancro al polmone?

	Cancro al polmone		
Fumo	Si	No	Totale
Si	**52**	**27**	79
No	**34**	**77**	111
Totale	86	104	190

Ipotesi di associazione

Ipotesi nulla: non c’è associazione

La variabile che ha funzione di exposure non produce alcun cambiamento nell’outcome.

Ipotesi alternativa: c’è associazione

La variabile che ha funzione di exposure comporta cambiamenti nelle frequenze dell’outcome.

Ipotesi di associazione

Per capire quale test utilizzare per verificare un’ipotesi di associazione dobbiamo tenere in considerazione i seguenti aspetti:

- Tipologia di variabili qualitative coinvolte
- Misura della tabella di contingenza (numero di categorie per variabile)
- Frequenze attese
- Appaiamento dei dati (i dati vengono definiti appaiati quando sulle stesse unità di analisi viene ripetuta la misurazione in tempi diversi)

Test di Mc Nemar

Test che si applica solamente a tabelle di dimensione 2x2

Test non parametrico

Viene applicato quando i dati sono appaiati, perciò la valutazione della variabile qualitativa viene ripetuta prima e dopo un certo evento/trattamento

Nel caso di frequenze attese basse è preferibile utilizzare il test binomiale

Test di Mc Nemar

Esempio: consideriamo 200 persone che vengono classificate come fumatori/non fumatori prima dell'uso della sigaretta elettronica. Siamo interessati a vedere se dopo un periodo di uso della sola sigaretta elettronica i pazienti tornano o meno a fumare sigarette

	Fumo post	
Fumo pre	SI	NO
SI	90	30
NO	20	60

La somma delle numerosità di queste celle non deve essere piccola

Test del chi-quadrato generico

Le due variabili possono essere di qualunque tipo (policotomiche o ordinali), è sufficiente che la tabella di contingenza associata sia di dimensione maggiore di 2x2

I dati non devono essere appaiati

È piuttosto sensibile alle frequenze attese basse. Le regole pratiche sono:

- L’80% delle frequenze attese deve essere più grande di 5
- Tutte le frequenze attese devono essere maggiori di 1

Test del chi-quadrato generico

Esempio: il tipo di trattamento (combinazione di morfina con antidolorifici o solo morfina o solo antidolorifici) è associato con il dolore (basso, medio o alto)?

Frequenze osservate	**Dolore**			
Terapia	Basso	Medio	Alto	Totale
Morfina	12	34	27	73
Antidolorifici	16	10	3	29
Morfina+ antidolorifici	14	27	31	72
Totale	42	71	61	174

Frequenze attese	**Dolore**			
Terapia	Basso	Medio	Alto	Totale
Morfina	18	30	26	73
Antidolorifici	7	12	10	29
Morfina+ antidolorifici	17	29	25	72
Totale	42	71	61	174

Test del chi-quadrato generico

Esempio: il tipo di trattamento (combinazione di morfina con antidolorifici o solo morfina o solo antidolorifici) è associato con il dolore (basso, medio o alto)?

Frequenze osservate	**Dolore**			
Terapia	Basso	Medio	Alto	Totale
Morfina	2	2	8	12
Antidolorifici	10	10	3	23
Morfina+ antidolorifici	5	7	1	13
Totale	17	19	12	48

Frequenze attese	**Dolore**			
Terapia	Basso	Medio	Alto	Totale
Morfina	4	5	3	12
Antidolorifici	8	9	6	23
Morfina+ antidolorifici	5	5	3	13
Totale	17	19	12	48

Test del chi-quadrato generico

Esempio: il tipo di trattamento (combinazione di morfina con antidolorifici o solo morfina o solo antidolorifici) è associato con il dolore (basso, medio o alto)?

L’80% di 9 celle è 7 ma in questo caso sono solamente 6 le celle con frequenze più elevate di 5

→ Non possiamo utilizzare il chi-quadrato generico

→ L’unica alternativa che abbiamo è accorpare le categorie di significato simile

Frequenze attese	**Dolore**			
Terapia	Basso	Medio	Alto	Totale
Morfina	4	5	3	12
Antidolorifici	8	9	6	23
Morfina+ antidolorifici	5	5	3	13
Totale	17	19	12	48

Test del chi-quadrato generico

In questo esempio non sarebbe stato sensato accorpare le terapie perché molto diverse.

Abbiamo accorpato le categorie del dolore e abbiamo risolto il problema delle frequenze attese.

Frequenze attese	**Dolore**		
Terapia	Basso	Medio-Alto	Totale
Morfina	4	8	12
Antidolorifici	8	15	23
Morfina+ antidolorifici	5	8	13
Totale	17	31	48

Test del chi-quadrato con correzione di Yates

Questo test viene applicato solamente per tabelle di contingenza di dimensione 2x2, perciò le variabili sono entrambe dicotomiche.

I dati non sono appaiati

Questo test in genere viene utilizzato quando le osservazioni non sono inferiori a 40

Test esatto di Fisher

Questo test viene applicato per tabelle di contingenza di dimensione 2x2 , perciò le variabili sono entrambe dicotomiche.

I dati non sono appaiati

Questo test può essere sempre applicato perché non è influenzato dalle frequenze attese basse.

Test esatto di Fisher

Questo test si definisce «esatto» perché calcola la probabilità esatta di ottenere la tabella osservata fra tutte quelle che hanno gli stessi valori di totali di riga e colonna (è per questo che viene utilizzato per frequenze osservate basse)

Es

13	23	**36**
7	12	**19**
20	**35**	

11	25	**36**
9	10	**19**
20	**35**	

Queste due tabelle hanno frequenze assolute diverse ma i totali di riga e colonna uguali

Ne esiste un’approssimazione per tabelle di contingenza più grandi di 2x2

Test del chi quadrato per il trend

È detto anche test di Armitage

Le variabili sono una di tipo dicotomico e l’altra di tipo ordinale

La tabella di contingenza associata ha dimensione 2xk

Vogliamo valutare se esiste un trend lineare nelle proporzioni/frequenze dell’evento all’aumentare/diminuire della variabile ordinale

Test del chi quadrato per il trend

Esempio: verificare se il numero di cavie morte aumenta dopo somministrazione di dosi crescenti di un certo farmaco sperimentale

	0.1g/l	0.15g/l	0.20g/l	0.25g/l
Morto	19	29	24	20
Vivo	497	560	269	147

0.037 0.049 0.082 0.120

Ci sono altri tipi di analisi diverse dall'associazione che vengono svolte sulle tabelle di contingenza (di dimensione 2x2 o più grandi ma sempre quadrate.)

Queste analisi dipendono dall'obiettivo di ricerca:

- Se vogliamo valutare la performance di uno strumento o di una classificazione, la capacità di individuare delle condizioni, dobbiamo fare analisi di sensibilità e specificità (gradi di errore)
  Es. confronto tra diagnosi clinica del melanoma e l'istologia

- Se invece vogliamo capire l'affidabilità di una classificazione o di uno strumento utilizziamo analisi di concordanza.
  Es. vogliamo capire se due ricercatori hanno lo stesso giudizio sulla benignità/malignità di un nodulo utilizzando un'ecografia

Concordanza

Valuta l’affidabilità di una classificazione.

Valuta il **grado di accordo** tra due valutatori o tra un valutare/medico e uno strumento o tra due strumenti o valutazioni dello stesso soggetto fatte in tempi differenti.

Gli strumenti/valutatori devono dare una valutazione sullo stesso fenomeno.

Concordanza

Il coefficiente che misura il grado di accordo è la **kappa di Cohen**

In base al suo valore viene misurato il grado di accordo

Se:

- $k<0$ non c’è concordanza
- $0<k<0.4$ concordanza scarsa
- $0.4<k<0.6$ concordanza discreta
- $0.6<k<0.8$ concordanza buona
- $0.8<k<1$ concordanza ottima

Concordanza

$$k = \frac{accordo(osservato) - accordo(atteso)}{1 - accordo(atteso)}$$

Frequenze osservate

		Osservatore 2		Totale
		SÌ	**NO**	
Osservatore 1	SÌ	a	b	**a + b**
	NO	c	d	**c + d**
Totale		**a + c**	**b + d**	**N**

Accordo osservato $A_o = \frac{a+d}{N}$

Frequenze attese

		Osservatore 2		Totale
		SÌ	**NO**	
Osservatore 1	SÌ	$\frac{(a+c)*(a+b)}{N}$	$\frac{(b+d)*(a+b)}{N}$	**a + b**
	NO	$\frac{(a+c)*(c+d)}{N}$	$\frac{(b+d)*(c+d)}{N}$	**c + d**
Totale		**a + c**	**b + d**	**N**

Accordo atteso $A_a = \frac{[(a+c)*(a+b)]+[(b+d)*(c+d)]}{N^2}$

Concordanza

Esempio. Valutazione della presenza/assenza di lesioni addominali valutate da due medici sulla base di tomografie computerizzate effettuate su 300 pazienti

	Frequenze osservate		
	Presenza	Assenza	Totale
Presenza	14	20	34
Assenza	24	242	266
Totale	38	262	300

	Frequenze attese		
	Presenza	Assenza	Totale
Presenza	4	30	34
Assenza	34	232	266
Totale	38	262	300

accordo osservato=(14+242)/300=0.85

accordo atteso=(4+232)/300=0.79

$$k = \frac{0.85-0.79}{1-0.79} = 0.31$$