



STATISTICA INFERENZIALE: TEST DI IPOTESI

La statistica inferenziale è un ramo della statistica che si occupa di trarre conclusioni su una popolazione basandosi sull'analisi di un campione di dati. Essa si basa sull'idea che l'esame di un sottoinsieme (campione) di una popolazione possa fornire informazioni sufficienti per fare inferenze sulla popolazione stessa. La statistica inferenziale si basa su principi di probabilità, teoria della stima e test di ipotesi per fornire strumenti per analizzare i dati e raggiungere conclusioni valide.

Campioni e popolazione: concetti chiave

In statistica, la popolazione è l'insieme completo di individui, oggetti o eventi che si desidera studiare. Un campione è un sottoinsieme della popolazione, selezionato in modo da rappresentare la popolazione stessa. La statistica inferenziale si basa sulla capacità di generalizzare i risultati ottenuti dal campione alla popolazione.

La scelta del campione è un aspetto fondamentale della statistica inferenziale. Un campione rappresentativo è essenziale per ottenere risultati affidabili e generalizzabili. Esistono diversi metodi di campionamento, come il campionamento casuale semplice, il campionamento stratificato e il campionamento a grappoli. La scelta del metodo di campionamento dipende dagli obiettivi della ricerca e dalle caratteristiche della popolazione.

Campione

Sottoinsieme di una popolazione selezionato per l'analisi.

- Rappresentazione della popolazione.
- Generalizzazione dei risultati alla popolazione.

Popolazione

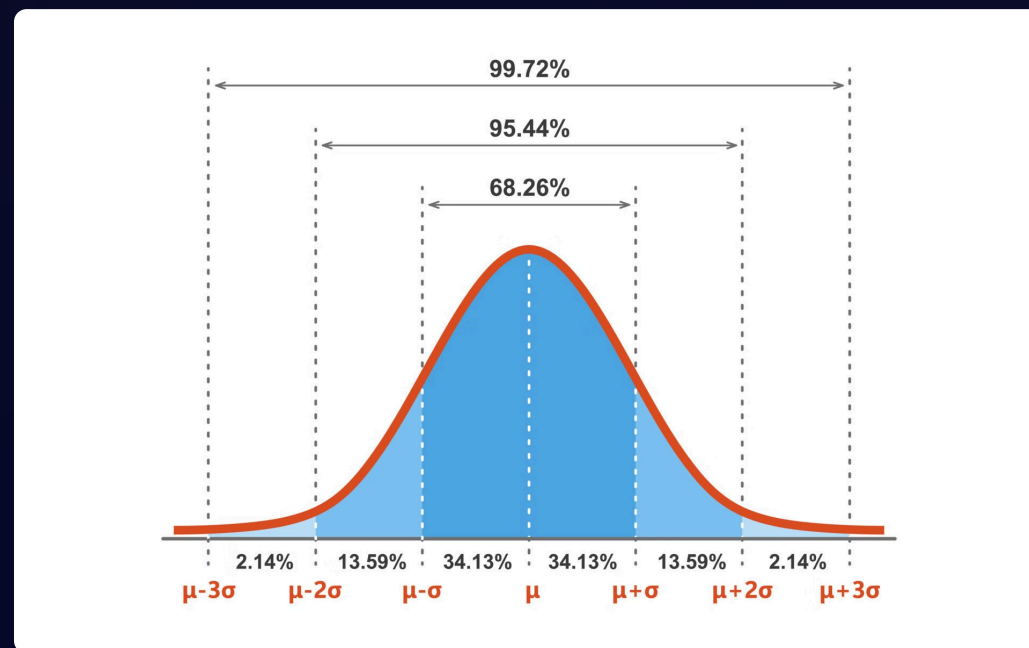
Insieme completo di individui, oggetti o eventi che si desidera studiare.

- Spesso troppo grande per essere analizzata direttamente.
- L'analisi di un campione permette di trarre conclusioni sulla popolazione.

Distribuzione normale: caratteristiche e proprietà

La distribuzione normale $N(\mu, \sigma^2)$, o curva a campana, è una distribuzione di probabilità continua simmetrica con una forma a campana. È una delle distribuzioni più importanti in statistica e trova applicazione in molti contesti, come la statistica inferenziale, la teoria della probabilità e l'analisi dei dati.

Le principali caratteristiche della distribuzione normale sono la sua forma simmetrica, la sua posizione definita dalla media (μ) e la sua dispersione definita dalla deviazione standard (σ). La distribuzione normale è completamente definita da questi due parametri.



Alcune proprietà importanti della distribuzione normale sono:

1 Simmetria

La distribuzione è simmetrica rispetto alla media. La metà dei dati si trova a sinistra della media e l'altra metà a destra.

2 Media, mediana e moda coincidono

Nella distribuzione normale, la media, la mediana e la moda hanno lo stesso valore.

3 Regola empirica

Circa il 68% dei dati si trova entro 1 deviazione standard dalla media, il 95% entro 2 deviazioni standard e il 99,7% entro 3 deviazioni standard.

Intervalli di confidenza per la media

Gli intervalli di confidenza forniscono un intervallo di valori che è probabile contenga il vero valore di un parametro della popolazione, come la media. Gli intervalli di confidenza sono costruiti utilizzando i dati del campione e un livello di confidenza, che indica la probabilità che l'intervallo contenga il vero valore del parametro.

La costruzione di un intervallo di confidenza per la media si basa sul teorema del limite centrale, che afferma che la distribuzione campionaria delle medie tende ad avvicinarsi alla distribuzione normale all'aumentare della dimensione del campione. L'intervallo di confidenza è dato dalla formula:

$$\text{Intervallo di confidenza} = \text{Media del campione} \pm \text{Margine di errore}$$

Il margine di errore dipende dalla deviazione standard del campione e dalla dimensione del campione. Maggiore è la deviazione standard, maggiore è il margine di errore. Maggiore è la dimensione del campione, minore è il margine di errore.

Test di ipotesi: principi e metodologia

Un test di ipotesi è una procedura statistica utilizzata per valutare la plausibilità di una dichiarazione sulla popolazione, detta ipotesi nulla, rispetto a un'ipotesi alternativa. In pratica, si valuta, in termini percentuali (livello di affidabilità), la validità di una certa ipotesi sulla base dei dati riscontrati empiricamente, in modo da capire se l'ipotesi stessa possa essere valida alla luce dei dati osservati.

La procedura di test di ipotesi prevede i seguenti passaggi:

1

Formulare l'ipotesi nulla (H_0) e l'ipotesi alternativa (H_1)

L'ipotesi nulla è la dichiarazione che si desidera verificare. L'ipotesi alternativa è la dichiarazione che si accetta se l'ipotesi nulla viene rifiutata.

2

Selezionare un livello di significatività (α)

Il livello di significatività indica la probabilità di rifiutare l'ipotesi nulla quando è vera. Un livello di significatività comune è $\alpha = 0,05$ (i.e. probabilità di ipotesi vera pari al 95%).

3

Calcolare la statistica test

Misurare di quanto i dati del campione sono in accordo con H_0 . La statistica test, calcolata attraverso la cosiddetta **quantità pivotale**, è una misura di quanto i dati del campione sono in accordo con l'ipotesi nulla.

4

Calcolare il valore p

Il valore p è la probabilità di ottenere risultati del campione estremi come quelli osservati, assumendo che l'ipotesi nulla sia vera. Esso corrisponde all'ordinata nel grafico della funzione densità di probabilità.

5

Rifiutare o non rifiutare H_0

Se il valore p è inferiore al livello di significatività, l'ipotesi nulla viene rifiutata. Se il valore p è maggiore o uguale al livello di significatività, l'ipotesi nulla non viene rifiutata.

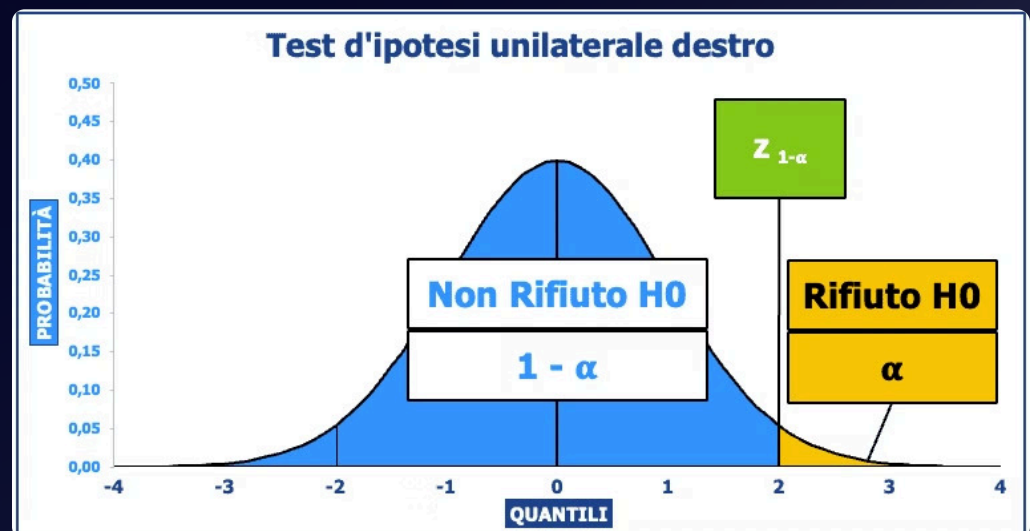
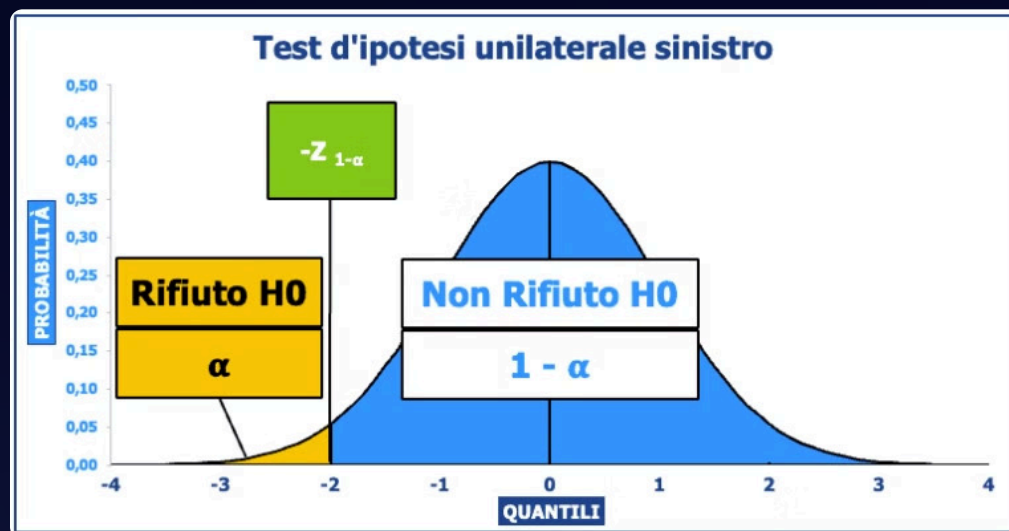
Valutazione del test

Il valore pivot (la quantità pivotale) che otteniamo è relativo al cosiddetto "quantile" del test, corrispondente ai valori in ascissa del grafico di distribuzione di probabilità. Ad ogni quantile corrisponde, in ordinata, un valore di probabilità che rappresenta il livello di affidabilità α della coda della campana (valori tabulati in tabelle di riferimento). Questo valore sarà tanto maggiore quanto più il nostro quantile si avvicina allo zero (sarà quindi piccolo in valore assoluto).

Ricordando che l'area totale della funzione dev'essere unitaria (i.e. probabilità totale pari al 100%), si ottiene così il livello di affidabilità dell'ipotesi H_0 .

Perciò, dato un α di riferimento ed il quantile relativo, se il valore pivot trovato è maggiore a quest'ultimo, in valore assoluto, vorrà dire che ci troviamo verso la coda della campana, ovvero nella zona a più bassa probabilità e l'ipotesi non sarà quindi accettabile con quel livello di affidabilità α (tipicamente si parla di 90%, 95% e 99%).

In caso contrario, se ci troviamo verso il centro del grafico, l'ipotesi sarà accettabile.

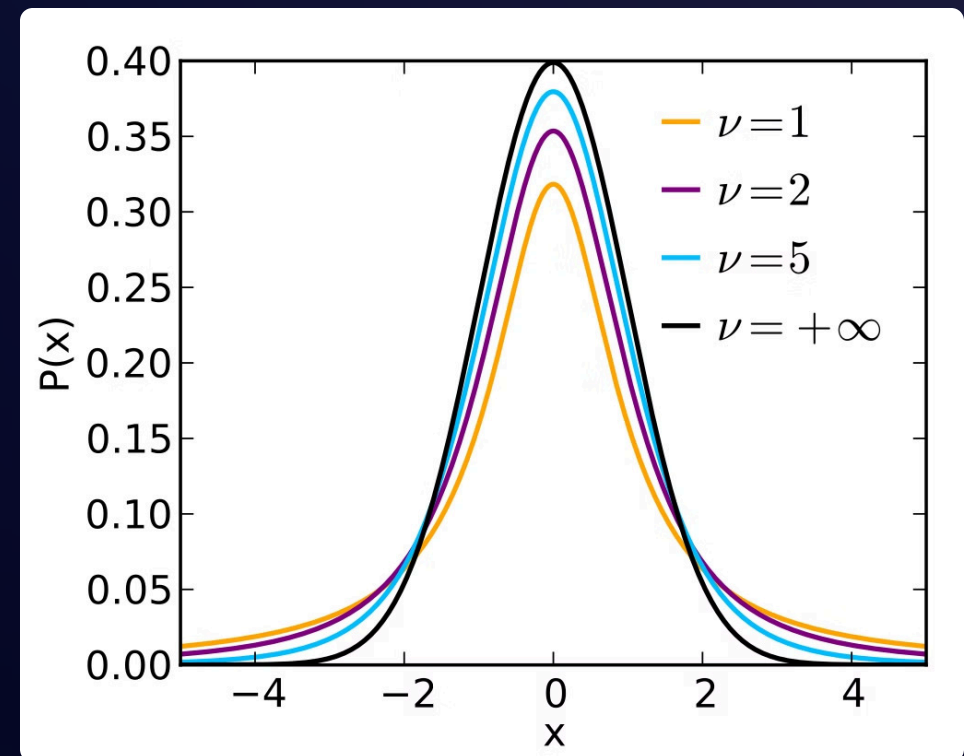


Test per la media: Normale z e t-Student

I test per la media vengono utilizzati per verificare ipotesi su la media di una popolazione. Esistono diversi test per la media, a seconda della conoscenza della deviazione standard della popolazione e della dimensione del campione.

- Se la deviazione standard della popolazione è nota e la dimensione del campione è grande ($n > 30$), si utilizza il test z. Il test z si basa sulla distribuzione normale standard.
- Se la deviazione standard della popolazione è sconosciuta o la dimensione del campione è piccola ($n \leq 30$), si utilizza il test t. Il test t si basa sulla distribuzione t di Student (grafico a destra) che dipende dai gradi di libertà $\nu = n - 1$.

Il test z e il test t sono entrambi test parametrici, ovvero presuppongono che la popolazione da cui è stato estratto il campione abbia una distribuzione normale. Se la popolazione non è distribuita normalmente, si possono utilizzare test non parametrici.



TEST	Deviazione standard	Dimensione del campione	Distribuzione	Gradi di libertà	Quantità pivotale
Test z	Nota	Grande ($n > 30$)	Normale standard	-	$Z^* = \frac{x - \mu}{\sigma} \sqrt{n}$
Test t	Sconosciuta o $n \leq 30$	Piccola ($n \leq 30$)	t di Student	$n - 1$	$T_{n-1}^* = \frac{ x - \mu }{s} \sqrt{n}$

Valori di riferimento per i test di ipotesi

A destra vengono riportate le tabelle relative ai test di ipotesi che tratteremo in questo corso.

Per validare o meno l'ipotesi nulla, sarà necessario confrontare il valore pivot calcolato con quelli di riferimento relativi a livelli di affidabilità del 90%, 95%, 99% e 99.9%.

Per i test T e χ^2 , i valori dipendono inoltre dal cosiddetto grado di libertà v , per cui vi è una serie di valori soglia per ciascun grado di libertà.

Perché l'ipotesi nulla non sia negabile con un certo grado di affidabilità, il valore pivot dovrà essere inferiore al valore soglia relativo al livello di significatività corrispondente.

Ad esempio, ottenendo $T^* = 1.80$ con $v = 59$ (i.e. $n = 60$ campioni), l'ipotesi nulla sarebbe rifiutabile con un livello di significatività del 90% ma non del 95%, quindi la probabilità che H_0 sia vera si attesta tra il 5% e il 10%.

Valori di riferimento per i test di ipotesi

Test T				
α v	0.10	0.05	0.01	0.001
1	6.314	12.706	63.657	636.619
2	2.920	4.303	9.925	31.599
3	2.353	3.182	5.841	12.924
4	2.132	2.776	4.604	8.610
5	2.015	2.571	4.032	6.869
6	1.943	2.447	3.707	5.959
7	1.895	2.365	3.499	5.408
8	1.860	2.306	3.355	5.041
9	1.833	2.262	3.250	4.781
10	1.812	2.228	3.169	4.587
11	1.796	2.201	3.106	4.437
12	1.782	2.179	3.055	4.318
13	1.771	2.160	3.012	4.221
14	1.761	2.145	2.977	4.140
15	1.753	2.131	2.947	4.073
16	1.746	2.120	2.921	4.015
17	1.740	2.110	2.898	3.965
18	1.734	2.101	2.878	3.922
19	1.729	2.093	2.861	3.883
20	1.725	2.086	2.845	3.850
21	1.721	2.080	2.831	3.819
22	1.717	2.074	2.819	3.792
23	1.714	2.069	2.807	3.768
24	1.711	2.064	2.797	3.745
25	1.708	2.060	2.787	3.725
26	1.706	2.056	2.779	3.707
27	1.703	2.052	2.771	3.690
28	1.701	2.048	2.763	3.674
29	1.699	2.045	2.756	3.659
30	1.697	2.042	2.750	3.646
39	1.685	2.023	2.708	3.558
49	1.677	2.010	2.680	3.500
59	1.671	2.001	2.662	3.463
69	1.667	1.995	2.649	3.437
79	1.664	1.990	2.640	3.418
89	1.662	1.987	2.632	3.403
99	1.660	1.984	2.626	3.392

Test Z				
α	0.10	0.05	0.01	0.001
	1.645	1.960	2.576	3.291

Test χ^2				
α v	0.10	0.05	0.01	0.001
1	2.706	3.841	6.635	10.828
2	4.605	5.991	9.210	13.816
3	6.251	7.815	11.345	16.266
4	7.779	9.488	13.277	18.467
5	9.236	11.070	15.086	20.515
6	10.645	12.592	16.812	22.458
7	12.017	14.067	18.475	24.322
8	13.362	15.507	20.090	26.124
9	14.684	16.919	21.666	27.877
10	15.987	18.307	23.209	29.588
11	17.275	19.675	24.725	31.264
12	18.549	21.026	26.217	32.909
13	19.812	22.362	27.688	34.528
14	21.064	23.685	29.141	36.123
15	22.307	24.996	30.578	37.697
16	23.542	26.296	32.000	39.252
17	24.769	27.587	33.409	40.790
18	25.989	28.869	34.805	42.312
19	27.204	30.144	36.191	43.820
20	28.412	31.410	37.566	45.315
21	29.615	32.671	38.932	46.797
22	30.813	33.924	40.289	48.268
23	32.007	35.172	41.638	49.728
24	33.196	36.415	42.980	51.179
25	34.382	37.652	44.314	52.620
26	35.563	38.885	45.642	54.052
27	36.741	40.113	46.963	55.476
28	37.916	41.337	48.278	56.892
29	39.087	42.557	49.588	58.301
30	40.256	43.773	50.892	59.703

Esempio: Test Z

Immaginiamo di voler testare l'ipotesi che l'altezza media delle donne italiane sia di 1,65 metri (= ipotesi nulla, H_0).

Supponiamo di avere un campione di 100 donne italiane con un'altezza media di 1,63 metri e una deviazione standard nota di 5 cm. Utilizzando il test z, possiamo calcolare la statistica test e il valore p per verificare se l'ipotesi nulla è vera o meno. Quindi ci stiamo chiedendo se, dato il campione considerato, è plausibile che l'ipotesi nulla sia vera.

In questo caso, la statistica test z sarebbe pari a:

$$z^* = [(1,63 - 1,65) / 0,05] \sqrt{100} = -4 \rightarrow |z^*| = 4.$$

Questo valore va confrontato con quelli noti estrapolati dalla relativa tabella che lega ad ogni livello di significatività α il valore pivot di riferimento. Se il valore pivot ottenuto è minore o uguale a quello tabulato per un certo α , l'ipotesi H_0 viene accettata con probabilità $1 - \alpha$, altrimenti viene rigettata.

Nel test Z, il valore soglia per accettare l'ipotesi con livello di confidenza $\alpha = 0.001$ (99.9%) è pari a 3.291, che è minore del nostro z^* . Se avessimo ottenuto, ad esempio, $z^* = 2$, l'ipotesi sarebbe stata rifiutabile con un livello di confidenza del 95% ma non del 99%.

Rigettiamo quindi l'ipotesi nulla con un livello di confidenza del 99.9%, ovvero H_0 ha una probabilità di essere vera inferiore allo 0.1%.

Questo significa che sarebbe altamente improbabile ottenere un risultato del campione con una media di 1,63 metri o inferiore, assumendo che l'altezza media delle donne italiane sia di 1,65 metri.

Test per la varianza: "Chi-quadro" χ^2

Il test chi-quadro è un test non parametrico utilizzato per verificare ipotesi sulla varianza di una popolazione composta da n campioni.

In particolare, questo test serve per verificare la seguente ipotesi H0: "Gli n esperimenti sono coerenti con la probabilità attesa p_i di ciascun evento i -esimo".

Il test chi-quadro può essere utilizzato anche per confrontare le varianze di due popolazioni.

Innanzitutto, è necessario soddisfare le seguenti condizioni:

- eventi indipendenti tra loro
- $f_i = n * p_i > 5$ (i.e.: frequenze assolute > 5)

Se k è il numero di risultati possibili (variabile singola), i gradi di libertà v saranno k - 1.

Nel caso di due variabili, v = (numero di righe - 1)*(numero di colonne - 1)

Si calcolino le f_i osservate, come numero di ricorrenze sul totale, e quelle attese dal punto di vista probabilistico. Nel caso di popolazione multipla, sarà necessario definire la tabella delle distribuzioni congiunte.

Infine, se f_i sono le frequenze dei dati, il valore pivot per, rispettivamente, popolazione singola e doppia sarà il seguente:

$$\chi^2 = \sum_{i=1}^k \frac{(f_i^{\text{expected}} - f_i^{\text{observed}})^2}{f_i^{\text{expected}}}$$

$$\chi^2 = \sum_i \sum_j \frac{(f_{ij}^a - f_{ij}^o)^2}{f_{ij}^o}$$

X	Y						Totale
	y_1	y_2	...	y_j	...	y_c	
x_1	n_{11}	n_{12}	...	n_{1j}	...	n_{1c}	$n_{1.}$
x_2	n_{21}	n_{22}	...	n_{2j}	...	n_{2c}	$n_{2.}$
⋮	⋮	⋮		⋮		⋮	⋮
x_i	n_{i1}	n_{i2}	...	n_{ij}	...	n_{ic}	$n_{i.}$
⋮	⋮	⋮		⋮		⋮	⋮
x_r	n_{r1}	n_{r2}	...	n_{rj}	...	n_{rc}	$n_{r.}$
Totale	$n_{.1}$	$n_{.2}$...	$n_{.j}$...	$n_{.c}$	n

Esempio test χ^2 (1)

Nella tabella a fianco, si riportano i dati relativi ad un trattamento medico veterinario. È evidente che i dati grezzi indicano che la xmicina è più efficace della streptomicina, ma la superiorità della xmicina potrebbe essere dovuta al caso.

Tabella 1. Dati ottenuti

Trattamento ↓	Esito →	guariti	non-guariti	totale
xmicina		52 a	10 b	62
streptomicina		40 c	21 d	61
totale		92	31	123

→ Dei 62 cani trattati con xmicina, ne sono guariti 52 (84%)
 $52/62 = 0.84$

→ Dei 61 cani trattati con streptomicina, ne sono guariti 40 (66%)
 $40/61 = 0.66$

Prima di giungere a una conclusione affrettata, occorre rispondere alla seguente domanda: "Supponendo che, in realtà, NON esistano differenze nell'efficacia dei due trattamenti, che probabilità c'è di osservare, in uno studio di dimensioni simili a questo, differenze nell'efficacia dei due antibiotici uguali o superiori a quelle che si è osservato?". La risposta a questa domanda dipende da quanto i dati ottenuti si discostano dai dati che sarebbe lecito attendersi se i trattamenti avessero la stessa efficacia e se i dati fossero influenzati soltanto dalla variazione casuale.

Dobbiamo costruire la Tabella 2: il trattamento è risultato efficace nel 74.8% dei casi, infatti sono guariti, sempre complessivamente e indipendentemente dall'antibiotico utilizzato, $52+40=92$ animali (a+c) su 123 trattati. Applicando questa percentuale di successo ($92/123=74.8\%$) a ciascuno dei due gruppi di cani in esame (gruppo xmicina - 0.748×62 , gruppo streptomicina - 0.748×61 e "b" e "d" per differenza), si ricavano i dati della Tabella 2, che illustra la situazione che ci saremmo

aspettati se i due antibiotici avessero avuto la stessa efficacia:

Tabella 2. Dati attesi

Trattamento ↓	Esito →	guariti	non-guariti	totale
xmicina		46.37 a	15.63 b	62.00
streptomicina		45.63 c	15.37 d	61.00
totale		92.00	31.00	123.00

Esempio test χ^2 (2)

Calcoliamo il valore pivot:

$$\chi^2 = \frac{(52 - 46.37)^2}{46.37} + \frac{(10 - 15.63)^2}{15.63} + \frac{(40 - 45.63)^2}{45.63} + \frac{(21 - 15.37)^2}{15.37} = 5.46$$

In questo caso, il valore ottenuto è un chi-quadrato con 1 grado di libertà, uguale a (numero di righe - 1)*(numero di colonne - 1). Ora, confrontando il valore di chi-quadrato (5.46) con quelli tabulati, notiamo che esso è > 3.841 e < 6.635 . Ciò consente di ritenere che la differenza fra i due gruppi sia significativa al livello di probabilità 5% ma non al livello di probabilità 1%.

Perciò, la differenza tra animali trattati con xmicina e quelli trattati con streptomina è statisticamente significativa al livello di probabilità 5%.

Questo risultato significa che, ammettendo che i due antibiotici abbiano pari efficacia e ripetendo l'esperimento infinite volte, potremmo osservare piuttosto raramente (ossia 5 volte su 100 o meno) dati simili a quelli ottenuti o più favorevoli a xmicina.

In sostanza: in base ai risultati del test del chi-quadrato, l'affermazione "la xmicina è più efficace di streptomina" ha il 95% di probabilità di essere vera (e quindi ha il 5% di probabilità di essere falsa).

Si può quindi concludere come segue: "In base ai risultati ottenuti, la xmicina è risultata più efficace della streptomina con livello di probabilità P inferiore al 5%", dove il valore P indica la probabilità di respingere una ipotesi zero vera.

Tabella dei valori di χ^2

Gradi di libertà	Probabilità	
	5%	1%
1	3.841	6.635
2	5.991	9.210
3	7.815	11.345
4	9.488	13.277
5	11.070	15.086
6	12.592	16.812
7	14.067	18.475
ecc.

Conclusioni e applicazioni pratiche

La statistica inferenziale è uno strumento potente per analizzare i dati e trarre conclusioni valide sulla popolazione. Le tecniche di inferenza statistica sono ampiamente utilizzate in molti campi, come la medicina, l'ingegneria, l'economia e la scienza sociale.

La statistica inferenziale consente ai ricercatori di trarre conclusioni significative dai dati raccolti e di ottenere una comprensione più profonda dei fenomeni che studiano. Ad esempio, in ambito sanitario, la statistica inferenziale viene utilizzata per valutare l'efficacia di nuovi trattamenti e per determinare il rischio di effetti collaterali.

In ambito economico, la statistica inferenziale viene utilizzata per prevedere le tendenze del mercato e per valutare l'impatto delle politiche economiche. In generale, la statistica inferenziale offre una serie di strumenti per analizzare i dati e per prendere decisioni informate in base ai risultati ottenuti.

