

Variabili casuali continue e la distribuzione normale

Esercizi svolti — punteggi z, tabella normale standard, distribuzione del QI, test d'ipotesi e approssimazione normale alla binomiale.

Livello

Università

Esercizi

7, con soluzione ragionata

Prerequisiti

Distribuzione binomiale, test d'ipotesi

FORMULE DI RIFERIMENTO

PUNTEGGIO Z (POPOLAZIONE)

$$z = \frac{X - \mu}{\sigma}$$

MEDIA E DEV. STANDARD CAMPIONARIE

$$\bar{x} = \frac{\sum X}{n} \quad s = \sqrt{\frac{\sum (X - \bar{x})^2}{n - 1}}$$

APPROSSIMAZIONE NORMALE ALLA BINOMIALE

$$z = \frac{X - np}{\sqrt{np(1-p)}}$$

PUNTEGGIO Z (CAMPIONE)

$$z = \frac{X - \bar{x}}{s}$$

DAL PUNTEGGIO Z AL DATO ORIGINALE

$$X = z \cdot \sigma + \mu$$

PROBABILITÀ FRA DUE PUNTEGGI Z

$$P(a < Z < b) = \Phi(b) - \Phi(a)$$

ESERCIZIO 1

Confrontare punteggi su scale diverse con lo z-score

A due soggetti (A e B) sono misurate ansia e depressione somministrando dei test standardizzati che hanno queste caratteristiche: il test che misura l'ansia ha media 50 e deviazione standard 10; il test che misura la depressione ha media 70 e deviazione standard 15. Il soggetto A ottiene i seguenti punteggi: 65 e 65 nei due test rispettivamente. Il Soggetto B ottiene 40 e 55.

A) Confrontando i punteggi nei test di ansia e depressione del soggetto A, possiamo concludere che i livelli di ansia e depressione sono identici?

B) Confrontando i punteggi nei test di ansia e depressione del soggetto B, possiamo concludere che i livelli di ansia e depressione sono diversi?

C) Calcola i punteggi z associati a ciascuno dei quattro risultati della somministrazione dei test.

SOLUZIONE

Test su scale diverse (medie e deviazioni standard diverse) non si possono confrontare direttamente: bisogna prima **standardizzare** entrambi i punteggi in unità z, poi confrontare.

$$z_A^{ansia} = \frac{65 - 50}{10} = 1.5 \quad z_A^{dep} = \frac{65 - 70}{15} \approx -0.33$$

$$z_B^{ansia} = \frac{40 - 50}{10} = -1 \quad z_B^{dep} = \frac{55 - 70}{15} = -1$$

A – SOGGETTO A

No, non identici

Ansia +1,5 σ sopra la media, depressione -0,33 σ sotto: posizioni relative molto diverse.

B – SOGGETTO B

No, non diversi

Entrambi i punteggi sono esattamente 1 σ sotto la media: la posizione relativa è identica.

C – RIEPILOGO PUNTEGGI Z

Soggetto A: 1,5 / -0,33 – Soggetto B: -1 / -1

L'IDEA CHIAVE

Il punteggio grezzo da solo non basta per confrontare due misure: 65 su un test e 65 su un altro possono rappresentare posizioni molto diverse rispetto alla rispettiva popolazione di riferimento. Il punteggio z toglie l'unità di misura originale e permette il confronto diretto.

ESERCIZIO 2

Media, deviazione standard e punteggi z di un campione

Dati i seguenti dati: 10, 20, 25, 30, 45 calcola:

A) Media e deviazione standard (a 2 cifre decimali) dell'insieme di dati.

B) Il punteggio z associato a ciascun dato.

SOLUZIONE

$$\bar{x} = \frac{10 + 20 + 25 + 30 + 45}{5} = \frac{130}{5} = 26$$

$$s = \sqrt{\frac{\sum (X - \bar{x})^2}{n - 1}} = \sqrt{\frac{670}{4}} \approx 12.94$$

MEDIA

$$\bar{x} = 26,00$$

DEVIAZIONE STANDARD CAMPIONARIA

$$s \approx 12,94$$

B) Punteggio z di ciascun dato:

$$z = \frac{X - 26}{12.94}$$

DATO (X)	PUNTEGGIO Z
10	-1,24
20	-0,46
25	-0,08
30	0,31
45	1,47

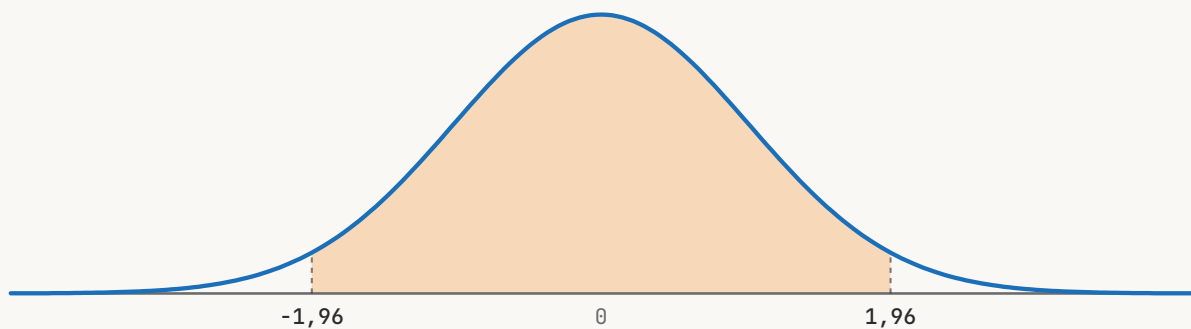
ESERCIZIO 3

Probabilità dalla tabella dei punteggi z

Utilizzando la tabella dei punteggi z calcola la probabilità dei seguenti eventi:

- A) Ottenere un punteggio z compreso fra $-3,54$ e 0 .
- B) Ottenere un punteggio z compreso fra $-1,96$ e $1,96$.
- C) Ottenere un punteggio z maggiore di 0 .
- D) Ottenere un punteggio z maggiore di $1,881$.
- E) Ottenere un punteggio z compreso fra $-2,054$ e $2,326$.

SOLUZIONE



Esempio: area sotto la curva normale standard fra $z = -1,96$ e $z = 1,96$ (punto B) – il 95% centrale della distribuzione

A
0,4998

B
0,9500

C
0,5000

D
0,0300

E
0,9700

COME SI LEGGONO

La tabella dei punteggi z riporta di norma l'area fra 0 e z. Per un intervallo simmetrico come B ed E si sommano le due metà; per una coda singola come C e D si usa la simmetria della curva rispetto a 0 (l'area totale è sempre 1).

ESERCIZIO 4

Il QI di 32 studenti: estremi, probabilità e proiezioni

I dati riportati sotto sono i punteggi di QI di 32 studenti di terza media ottenuti usando un test standardizzato (media 100 e deviazione standard 15).

140	106	126	105
116	119	97	115
123	108	117	144
125	122	89	130
89	135	120	127
113	95	109	111
132	87	87	134
98	106	93	114

A) Quanti dei punteggi estremi corrispondono a punteggi z minori di $-1,96$ e maggiori di $1,96$? Quali sono?

B) Qual è la probabilità di osservare dei punteggi compresi fra 100 e 125?

C) Se venissero testati altri 100 studenti dalla stessa popolazione, quanti ci aspettiamo abbiano un QI compreso fra 100 e 125?

D) E superiore a 115?

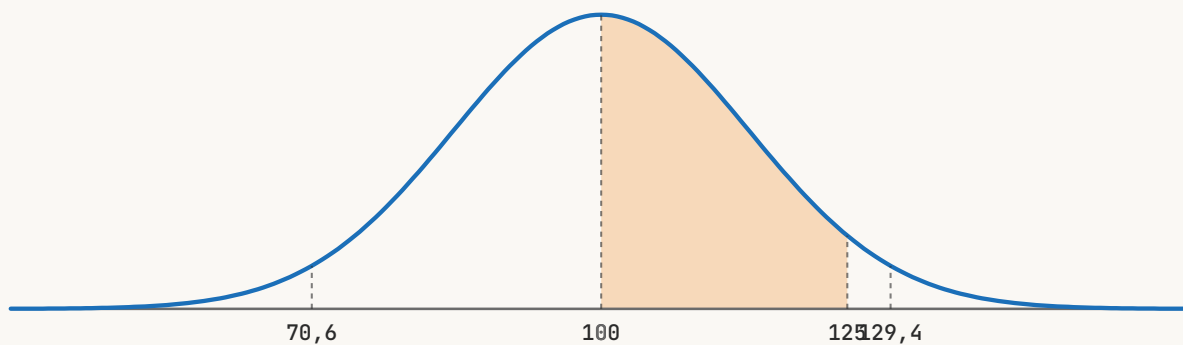
$$X = 100 \pm 1.96 \cdot 15 \Rightarrow 70.6 \text{ e } 129.4$$

SOLUZIONE

A – 6 PUNTEGGI ESTREMI

130, 132, 134, 135, 140, 144

Tutti oltre la soglia superiore di 129,4 ($z = 1,96$); nessun punteggio scende sotto la soglia inferiore di 70,6.



Area evidenziata: probabilità di un QI compreso fra 100 e 125 (punto B)

$$P(100 < X < 125) = \Phi(1.667) - \Phi(0) \approx 0.452$$

$$P(X > 115) = 1 - \Phi(1) \approx 0.1587$$

B
 $\approx 0,4522$

C
 ≈ 45 studenti

D
 ≈ 16 studenti



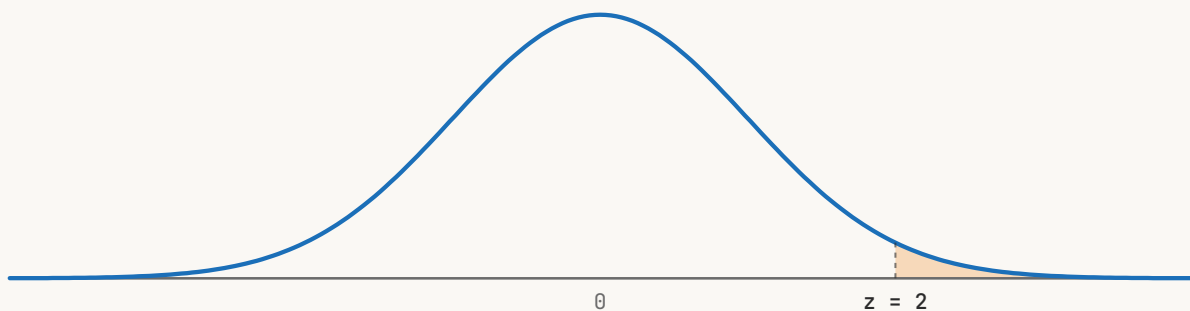
ESERCIZIO 5

Test d'ipotesi su un singolo punteggio

Un test per valutare le abilità cognitive viene somministrato ad una persona anziana per valutare se essa abbia problemi cognitivi (per esempio una forma iniziale di demenza senile). Il test somministrato è standardizzato con media 70 e deviazione standard 12 (a punteggi maggiori corrispondono problemi cognitivi maggiori). Il soggetto testato ottiene un punteggio di 94. Possiamo rigettare l'ipotesi nulla che indica che il soggetto è stato estratto da una popolazione di anziani senza problemi cognitivi?

SOLUZIONE

$$z = \frac{94 - 70}{12} = 2$$



Area evidenziata: probabilità di ottenere un punteggio $z \geq 2$ sotto l'ipotesi nulla

$$P(Z \geq 2) = 1 - \Phi(2) \approx 0.0228$$

CONCLUSIONE

$0,0228 < 0,05 \rightarrow$ si rigetta H_0

La probabilità di osservare, per puro caso, un punteggio così alto (o più alto) in una popolazione di anziani senza problemi cognitivi è inferiore al 5%. Sulla base di questo singolo dato, si conclude che il soggetto testato mostra problemi di carattere cognitivo.

Tre concetti chiave del test d'ipotesi

- A) Se si commette un errore di Tipo I a cosa ci si riferisce?
- B) Cosa si intende per potere statistico?
- C) Posso accettare l'ipotesi nulla se un test statistico non è significativo?

SOLUZIONE

A) Errore di Tipo I

Si commette quando, pur essendo l'ipotesi nulla corretta, questa viene erroneamente rigettata sulla base dei dati campionari. È il rischio di "vedere un effetto che non c'è".

B) Potere statistico

È la probabilità di rigettare correttamente l'ipotesi nulla quando questa è effettivamente falsa. Un test con potere statistico basso rischia di non rilevare un effetto reale (errore di Tipo II); il potere cresce tipicamente con la dimensione campionaria, come si è visto confrontando gli Esercizi 9 e 10 del set sulla distribuzione binomiale.

C) Non rigettare H_0 significa poterla accettare come vera?

No. L'ipotesi nulla è per convenzione assunta vera in partenza; un test statistico può solo fornire prove sufficienti per rigettarla, non per dimostrarla. Il fatto di non rigettarla non dimostra la sua correttezza: può darsi che l'ipotesi nulla sia in realtà falsa, ma che i dati raccolti non siano sufficienti (per numerosità campionaria o per grandezza dell'effetto) a rilevarlo con certezza statistica.

IN SINTESI

"Non significativo" non equivale a "nessun effetto" — equivale a "non abbiamo prove sufficienti per escludere il caso". La distinzione è la stessa emersa negli Esercizi 8 e 9 del set precedente, dove un risultato vicino alla soglia non permetteva di rigettare H_0 pur suggerendo una tendenza reale.

ESERCIZIO 7

Approssimazione normale alla binomiale: due etichette a confronto

In una ricerca di mercato a 100 persone è dato da scegliere fra 2 diverse etichette per la stessa bottiglia di olio. 30 scelgono l'etichetta A e 70 l'etichetta B. Decidi se vi sia una differenza significativa nella preferenza per le etichette.

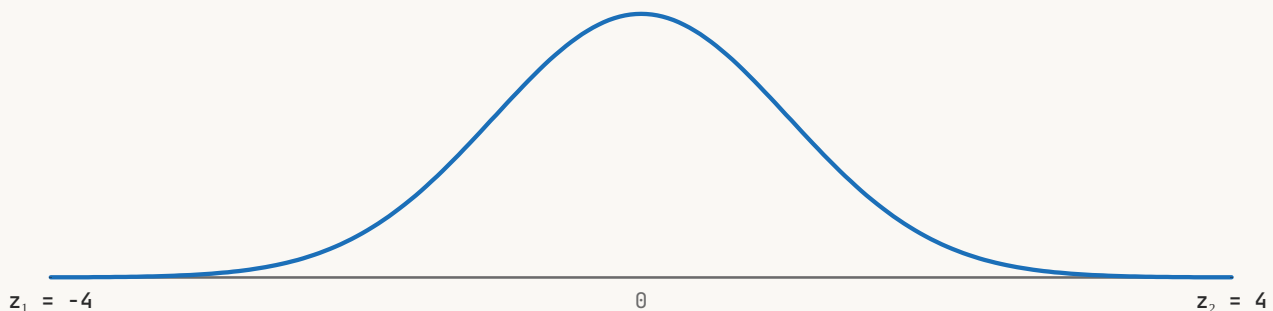
SOLUZIONE

Si usa la distribuzione normale come approssimazione della binomiale. Sotto l'ipotesi nulla di equiprobabilità ($p = 0,5$), con $N = 100$ (senza correzione di continuità):

$$z = \frac{X - np}{\sqrt{np(1-p)}}$$

$$z_1 = \frac{30 - 50}{\sqrt{100 \cdot 0.5 \cdot 0.5}} = \frac{-20}{5} = -4$$

$$z_2 = \frac{70 - 50}{5} = 4$$



Entrambi i risultati osservati cadono a $z = \pm 4$: code estremamente improbabili sotto l'ipotesi di equiprobabilità

$$P(Z \leq -4) + P(Z \geq 4) \approx 0.00003 + 0.00003 = 0.00006$$

CONCLUSIONE

$0,00006 < 0,05 \rightarrow$ si rigetta H_0

L'ipotesi di equiprobabilità di scelta fra le due etichette è rigettata: l'etichetta B è preferita a quella A in una percentuale significativamente superiore. Una divisione 30/70 su 100 persone è praticamente impossibile da osservare per puro caso se le due etichette fossero davvero scelte con la stessa probabilità.