

Statistica e Analisi dei dati

Matteo Mascherpa

a.a. 2024/2025

Contents

Lezione 1

Lezione 2

Lezione 3

Lezione 4

Lezione 5

Lezione 6

Mediana Campionaria

Sia il campione: 23, 04, 02, 2, 110, 5, 7, 6, 7, 3 la *media campionaria* è $\bar{x} = 140/7 = 20$. Se si vuole avere un valore che identifichi il centro del campione serve la *mediana campionaria* per indicarla uso m .

Dati i valori di un campione ordinati in ordine crescente. Se la cardinalità del campione è **dispari** allora m è il valore intermedio della lista altrimenti, se la cardinalità è **pari** allora m è la media dei due valori intermedi.

Al contrario della *media campionaria* che prende in considerazione tutti i valori degli insiemi dati la *mediana campionaria* non è influenzata dai valori estremi.

Percentili campionari

La media campionaria è un caso di statistica nota come: *100p-esimo percentile campionario* con $p \in [0, 1]$. Esso è un valore che è maggiore del(di almeno) $100p\%$ dei valori del campione e minore del(di almeno) $100(1 - p)\%$ dei valori. Nel caso della mediana $p = 0,5$.

Come trovare il $100p - \text{esimo}$ percentile

1. Ordina i dati in ordine crescente
2. Se np non è intero, trova il più piccolo \geq di np . Il valore è quello nella posizione trovata.
3. Se np è intero, allora il valore è la media tra i valori in posizione np e $np + 1$

Quartili

I quartili suddividono il campionario dei dati in quattro parti 25% l'una.

1. Primo quartile: Il 25-esimo percentile.

2. Secondo quartile: Il 50-esimo percentile.
3. Terzo quartile: Il 75-esimo percentile.

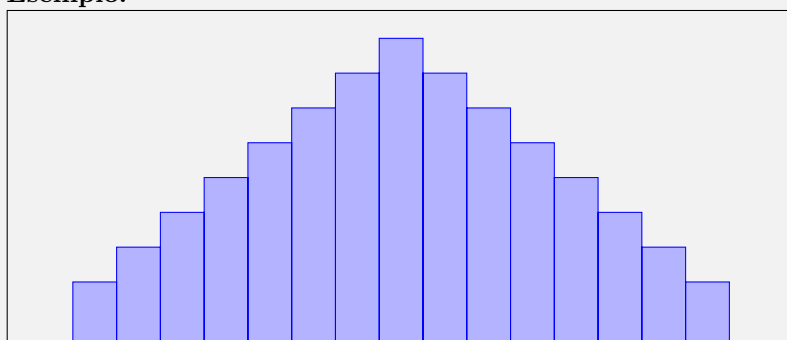
Gli insiemi di dati normali e la regola empirica

La maggior parte degli istogrammi hanno un simile aspetto. Sono spesso simmetrici sulla frequenza massima e assumono una forma a campana. L'insieme di questi istogrammi si dice *istogrammi normali*.

Un insieme si dice *normale* se il rispettivo istogramma:

1. Ha punto di massima in corrispondenza dell'intervallo centrale.
2. Spostandosi dal centro in una qualsiasi direzione l'altezza cala in modo da creare una forma a campana.
3. L'istogramma è simmetrico a rispetto all'intervallo centrale.

Esempio:



Regola empirica

Se un insieme è approssimativamente normale con media \bar{x} con una deviazione standard s .

1. Approssimativamente 68% dei dati si trovano nell'intervallo: $[\bar{x} - s, \bar{x} + s]$
2. Approssimativamente 95% dei dati si trovano nell'intervallo: $[\bar{x} - 2s, \bar{x} + 2s]$
3. Approssimativamente 99,7% dei dati si trovano nell'intervallo: $[\bar{x} - 3s, \bar{x} + 3s]$

22	372
23	512, 688, 941
24	706
25	020, 057, 128, 400, 446, 575
26	183, 894, 982
27	671, 711, 744
28	345, 764, 913, 967

Un modo efficiente di rappresentare un insieme di dati di dimensioni medie consiste nell'utilizzare il *diagramma ramo-foglia* (o a stelo). Tale grafico si ottiene dividendo ciascun valore dei dati in due parti, chiamati appunto rami e foglie.

La scelta dei rami dovrebbe essere fatta in modo che il diagramma ramo-foglia che ne risulta sia informativo sui dati. Questi diagrammi sono particolarmente adatti a descrivere insiemi di dati dimensioni ridotte.

Lezione 7

Lezione 8