STATISTICA DESCRITTIVA

L’asimmetria

La forma della distribuzione, la relazione tra media, mediana e moda, gli indici di Pearson, Bowley e Fisher, con esercizi svolti

Asimmetria positiva e negativa | Media, mediana e moda | Indici di Pearson | Indice di Bowley | Indice di Fisher

Fabrizio Antolini · Università di Teramo

PERCHÉ L’ASIMMETRIA

Stessa media, stessa varianza, forma diversa

Distribuzione A

2 · 3 · 3 · 4 · 8



μ = 4 · σ² = 4,4 · Me = 3

Distribuzione B

0 · 4 · 5 · 5 · 6

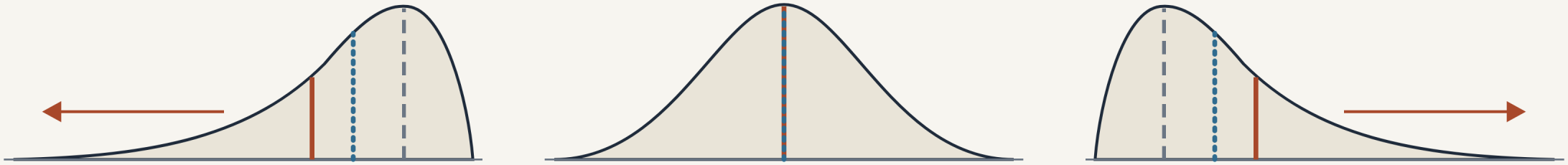


μ = 4 · σ² = 4,4 · Me = 5

Posizione e variabilità non bastano: serve una misura della **forma**. L’asimmetria dice da che parte la distribuzione si allunga.

ASIMMETRIA · TEORIA

Il nome viene dalla coda, non dalla gobba



Asimmetria negativa

coda a sinistra · μ < Me < Mo

Simmetrica

nessuna coda prevale · μ = Me = Mo

Asimmetria positiva

coda a destra · Mo < Me < μ

media μ · mediana Me · moda Mo · direzione della coda

ASIMMETRIA · TEORIA

La media segue la coda, la mediana no

Spesa giornaliera di 11 turisti (euro), ordinata

40 · 45 · 50 · 50 · 50 · 55 · 60 · 65 · 75 · 90 · **135**

Caso	Massimo	Media μ	Mediana Me
Dati originali	135	65,0	55
Massimo portato a 235	235	74,1	55

Media = baricentro

Ogni valore pesa per la sua distanza: pochi valori estremi la spostano verso la coda.

Mediana = posizione

Conta solo l'ordine: il valore centrale resta lo stesso, quanto lontano vada il massimo.

In una distribuzione asimmetrica media e mediana si separano, e la loro distanza misura la forza della coda.

ASIMMETRIA · TEORIA

L’ordine di moda, mediana e media

Asimmetria negativa

$$\mu < Me < Mo$$

La coda sinistra trascina la media verso il basso

Simmetria

$$\mu = Me = Mo$$

Le tre misure coincidono al centro

Asimmetria positiva

$$Mo < Me < \mu$$

La coda destra trascina la media verso l’alto

Relazione empirica di Pearson

$$\mu - Mo \approx 3\,(\mu - Me)$$

Nelle distribuzioni unimodali moderatamente asimmetriche la mediana sta tra moda e media, a circa due terzi della distanza dalla moda.

Una regolarità, non un teorema

Con caratteri discreti, poche modalità o più mode l’ordine può saltare. Per questo servono indici veri e propri.

INDICI DI PEARSON · TEORIA

Quanto si allontana la media dal centro

Prima formula · con la moda

$$Sk_1 = (\mu - Mo) / \sigma$$

Richiede una moda unica e ben definita.

Seconda formula · con la mediana

$$Sk_2 = 3\,(\mu - Me) / \sigma$$

Usa la relazione empirica; varia tra −3 e +3.

$Sk > 0$	$Sk = 0$	$Sk < 0$
asimmetria positiva	simmetria	asimmetria negativa

Dividere per σ rende l’indice un numero puro: la distanza è misurata in scarti quadratici medi e si possono confrontare caratteri con unità di misura diverse.

INDICE DI BOWLEY · TEORIA

I quartili attorno alla mediana

$$B = (Q3 + Q1 - 2\,Me) / (Q3 - Q1)$$

equivalente a [(Q3 – Me) – (Me – Q1)] / (Q3 – Q1): la differenza tra le due metà della scatola, rapportata alla sua ampiezza

Asimmetria positiva: la metà destra è più lunga



Q3 – Me > Me – Q1, quindi B > 0

Proprietà

- Varia tra –1 e +1
- Usa solo il 50% centrale: robusto ai valori anomali
- Non vede le code oltre i quartili

INDICE DI FISHER · TEORIA

Il momento terzo standardizzato

Distribuzione semplice

$$m3 = \Sigma\,(xi - \mu)^3 / N$$

$$\gamma 1 = m3 / \sigma^3$$

Distribuzione di frequenze

$$m3 = \Sigma\,(xi - \mu)^3 \cdot ni / N$$

$$\gamma 1 = m3 / \sigma^3$$

Perché il cubo

Il quadrato cancella il segno degli scarti, il cubo lo conserva e amplifica quelli grandi. Se la coda lunga è a destra prevalgono i cubi positivi e $\gamma 1 > 0$; se è a sinistra prevalgono quelli negativi.

Formula di calcolo

$$m3 = M3 - 3\,\mu\,M2 + 2\,\mu^3$$

con $Mk = \Sigma\, xi^k \cdot ni / N$, i momenti dall'origine: non servono gli scarti, come per la varianza.

ASIMMETRIA · CAUTELE

Indici diversi, risposte diverse

Caratteri discreti, poche modalità

Q1, Me e Q3 cadono su modalità consecutive ed equidistanti: Bowley vale 0 anche se la distribuzione è asimmetrica.

Moda e mediana “a gradini”

Con poche modalità Sk1 e Sk2 possono essere molto diversi tra loro, e la moda può non essere unica.

Scale non confrontabili

Bowley sta tra −1 e +1, Sk2 tra −3 e +3, γ1 non ha limiti: si confrontano distribuzioni con lo stesso indice.

Il segno è la prima informazione. Per l’intensità il riferimento è γ1, che usa tutte le osservazioni.

SINTESI

I quattro indici a confronto

Indice	Formula	Campo	Valori usati	Valori anomali
Pearson 1	(μ − Mo) / σ	non limitato	μ, Mo, σ	sensibile
Pearson 2	3 (μ − Me) / σ	da −3 a +3	μ, Me, σ	sensibile
Bowley	(Q3 + Q1 − 2 Me) / (Q3 − Q1)	da −1 a +1	Q1, Me, Q3	robusto
Fisher γ1	m3 / σ³	non limitato	tutti	molto sensibile

> 0 · coda a destra	= 0 · simmetria	< 0 · coda a sinistra

Coppie naturali, come per la variabilità: **mediana e quartili** con Bowley, **media e momenti** con Fisher.

SECONDA PARTE

Esercizi svolti

Quattro distribuzioni, due con coda a destra e due con coda a sinistra. Per ciascuna: prima γ1, poi gli indici di Pearson e Bowley sugli stessi dati.

1 · Spesa giornaliera · semplice | 2 · Punteggi · semplice | 3 · Notti di soggiorno · frequenze | 4 · Soddisfazione · frequenze | Riepilogo

ESERCIZIO SVOLTO 1 · DISTRIBUZIONE SEMPLICE · Γ1

Spesa giornaliera di 11 turisti (euro)

xi	xi − μ	(xi − μ)²	(xi − μ)³
40	−25	625	−15.625
45	−20	400	−8.000
50	−15	225	−3.375
50	−15	225	−3.375
50	−15	225	−3.375
55	−10	100	−1.000
60	−5	25	−125
65	0	0	0
75	+10	100	+1.000
90	+25	625	+15.625
135	+70	4.900	+343.000
715	0	7.450	+324.750

1 · Media

μ = 715 / 11 = 65

2 · Varianza e scarto quadratico medio

σ^2 = 7.450 / 11 = 677,27 → σ = 26,02

3 · Momento terzo

m3 = 324.750 / 11 = 29.522,7

4 · Indice di Fisher

$\gamma 1$ = 29.522,7 / $26{,}02^3$ = 29.522,7 / 17.625,7 = **1,67**

ESERCIZIO SVOLTO 1 · DISTRIBUZIONE SEMPLICE · PEARSON E BOWLEY

Stessi dati: moda, mediana e quartili

40	45	50	50	50	55	60	65	75	90	135
1°	2°	Q1 · 3°	4°	5°	Me · 6°	7°	8°	Q3 · 9°	10°	11°

Mo = 50 (tre volte) · Me = 55 · μ = 65 · σ = 26,02 · Q1 = 50 · Q3 = 75

Pearson 1

Sk1 = (65 − 50) / 26,02

= 0,58

Pearson 2

Sk2 = 3 (65 − 55) / 26,02

= 1,15

Bowley

B = (75 + 50 − 110) / 25

= 0,60

Tutti positivi, come γ1: Mo < Me < μ. La coda destra tira la media lontano dalla mediana.

ESERCIZIO SVOLTO 2 · DISTRIBUZIONE SEMPLICE · Γ1

Punteggi di 11 studenti in una prova (su 100)

xi	xi − μ	(xi − μ)²	(xi − μ)³
45	−35	1.225	−42.875
60	−20	400	−8.000
70	−10	100	−1.000
75	−5	25	−125
80	0	0	0
85	+5	25	+125
90	+10	100	+1.000
90	+10	100	+1.000
90	+10	100	+1.000
95	+15	225	+3.375
100	+20	400	+8.000
880	0	2.700	−37.500

1 · Media

$\mu = 880 / 11 = 80$

2 · Varianza e scarto quadratico medio

$\sigma^2 = 2.700 / 11 = 245{,}45 \rightarrow \sigma = 15{,}67$

3 · Momento terzo

$m3 = -37.500 / 11 = -3.409{,}1$

4 · Indice di Fisher

$\gamma 1 = -3.409{,}1 / 15{,}67^3 = -3.409{,}1 / 3.845{,}5 =$ **−0,89**

ESERCIZIO SVOLTO 2 · DISTRIBUZIONE SEMPLICE · PEARSON E BOWLEY

Stessi dati: moda, mediana e quartili

45	60	70	75	80	85	90	90	90	95	100
1°	2°	Q1 · 3°	4°	5°	Me · 6°	7°	8°	Q3 · 9°	10°	11°

Mo = 90 (tre volte) · Me = 85 · μ = 80 · σ = 15,67 · Q1 = 70 · Q3 = 90

Pearson 1

Sk1 = (80 − 90) / 15,67

= −0,64

Pearson 2

Sk2 = 3 (80 − 85) / 15,67

= −0,96

Bowley

B = (90 + 70 − 170) / 20

= −0,50

Tutti negativi, come γ1: μ < Me < Mo. Il 45 isolato allunga la coda sinistra.

ESERCIZIO SVOLTO 3 · DISTRIBUZIONE DI FREQUENZE · Γ1

Notti di soggiorno di 100 turisti

xi	ni	Fi	xi · ni	$xi^2 \cdot ni$	$xi^3 \cdot ni$
1	8	0,08	8	8	8
2	22	0,30	44	88	176
3	30	0,60	90	270	810
4	18	0,78	72	288	1.152
5	10	0,88	50	250	1.250
6	7	0,95	42	252	1.512
7	5	1,00	35	245	1.715
Tot.	100		341	1.401	6.623

1 · Momenti dall'origine (totali / 100)

$\mu = M1 = 3{,}41 \cdot M2 = 14{,}01 \cdot M3 = 66{,}23$

2 · Varianza

$\sigma^2 = 14{,}01 - 3{,}41^2 = 2{,}3819 \rightarrow \sigma = 1{,}5433$

3 · Momento terzo: M3 − 3 μ M2 + 2 μ³

$m3 = 66{,}23 - 143{,}32 + 79{,}30 = 2{,}2113$

4 · Indice di Fisher

$\gamma 1 = 2{,}2113 / 1{,}5433^3 = 2{,}2113 / 3{,}6761 =$ **0,60**

ESERCIZIO SVOLTO 3 · DISTRIBUZIONE DI FREQUENZE · PEARSON E BOWLEY

Quando Bowley non vede l’asimmetria

xi	ni	Fi	Ruolo
1	8	0,08	
2	22	0,30	Q1: prima F ≥ 0,25
3	30	0,60	Mo (n max) e Me: prima F ≥ 0,50
4	18	0,78	Q3: prima F ≥ 0,75
5	10	0,88	
6	7	0,95	
7	5	1,00	

B = 0 ma γ1 = 0,60: i quartili sono equidistanti dalla mediana, e l’asimmetria sta nella coda di 5, 6 e 7 notti, che Bowley non guarda.

Pearson 1

Sk1 = (3,41 − 3) / 1,5433

= 0,27

Pearson 2

Sk2 = 3 (3,41 − 3) / 1,5433

= 0,80

Bowley

B = (4 + 2 − 2 · 3) / (4 − 2)

= 0

ESERCIZIO SVOLTO 4 · DISTRIBUZIONE DI FREQUENZE · Γ1

Soddisfazione di 100 ospiti (scala 1–5)

xi	ni	Fi	xi · ni	xi² · ni	xi³ · ni
1	5	0,05	5	5	5
2	8	0,13	16	32	64
3	17	0,30	51	153	459
4	30	0,60	120	480	1.920
5	40	1,00	200	1.000	5.000
Tot.	100		392	1.670	7.448

1 · Momenti dall'origine (totali / 100)

μ = M1 = 3,92 · M2 = 16,70 · M3 = 74,48

2 · Varianza

$\sigma^2 = 16{,}70 - 3{,}92^2 = 1{,}3336 \rightarrow \sigma = 1{,}1548$

3 · Momento terzo: M3 − 3 μ M2 + 2 μ³

m3 = 74,48 − 196,39 + 120,47 = −1,4394

4 · Indice di Fisher

$\gamma_1 = -1{,}4394 / 1{,}1548^3 = -1{,}4394 / 1{,}5401 =$ **−0,93**

ESERCIZIO SVOLTO 4 · DISTRIBUZIONE DI FREQUENZE · PEARSON E BOWLEY

Stesso segno, intensità molto diverse

xi	ni	Fi	Ruolo
1	5	0,05	
2	8	0,13	
3	17	0,30	Q1: prima F ≥ 0,25
4	30	0,60	Me: prima F ≥ 0,50
5	40	1,00	Mo (n max) e Q3: prima F ≥ 0,75

$\mu = 3{,}92 < Me = 4 < Mo = 5 \cdot \sigma = 1{,}1548 \cdot \gamma 1 = -0{,}93$

Sk2 è piccolo perché media e mediana sono vicine (3,92 e 4), mentre la moda è lontana: con poche modalità le misure di posizione vanno “a gradini”.

Pearson 1

$Sk1 = (3{,}92 - 5) / 1{,}1548$

$= -0{,}94$

Pearson 2

$Sk2 = 3\,(3{,}92 - 4) / 1{,}1548$

$= -0{,}21$

Bowley

$B = (5 + 3 - 2 \cdot 4) / (5 - 3)$

$= 0$

RIEPILOGO

Quattro distribuzioni, quattro indici

Esercizio	Ordine	Sk1	Sk2	B	γ1
1 · Spesa (semplice)	Mo < Me < μ	0,58	1,15	0,60	1,67
2 · Punteggi (semplice)	μ < Me < Mo	−0,64	−0,96	−0,50	−0,89
3 · Notti (frequenze)	Mo = Me < μ	0,27	0,80	0	0,60
4 · Soddisfazione (frequenze)	μ < Me < Mo	−0,94	−0,21	0	−0,93

Il segno

Dice da che parte sta la coda, e concorda con l'ordine di moda, mediana e media.

L'intensità

Si confronta solo con lo stesso indice: nell'esercizio 4 Sk1 e Sk2 differiscono di oltre quattro volte.