



Esercitazione
6

A. Iodice

La forma di
una
distribuzione

Indici di
asimmetria

Il box-plot

box-plot e
valori anomali

Confronto
grafico
istogramma/
box-plot

Indice di
asimmetria
per variabili
qualitative
ordinate

Esercitazione 6

Statistica

Alfonso Iodice D'Enza
iodicede@gmail.com

Università degli studi di Cassino



Outline

Esercitazione 6

A. Indice

La forma di
una
distribuzione

Indici di
asimmetria

Il box-plot

box-plot e
valori anomali

Confronto
grafico
istogramma/
box-plot

Indice di
asimmetria
per variabili
qualitative
ordinate

- 1 La forma di una distribuzione
- 2 Indici di asimmetria
- 3 Il box-plot
- 4 box-plot e valori anomali
- 5 Confronto grafico istogramma/ box-plot
- 6 Indice di asimmetria per variabili qualitative ordinate



La forma di una distribuzione

Esercitazione
6

A. Iodice

La forma di
una
distribuzione

Indici di
asimmetria

Il box-plot

box-plot e
valori anomali

Confronto
grafico
istogramma/
box-plot

Indice di
asimmetria
per variabili
qualitative
ordinate

Due distribuzioni aventi stessa posizione e variabilità possono differire per **forma**. In altre parole, la forma dipende dal valore delle modalità più piccole (o più grandi) del valore centrale della distribuzione.



La forma di una distribuzione

Esercitazione
6

A. Iodice

La forma di
una
distribuzione

Indici di
asimmetria

Il box-plot

box-plot e
valori anomali

Confronto
grafico
istogramma/
box-plot

Indice di
asimmetria
per variabili
qualitative
ordinate

Due distribuzioni aventi stessa posizione e variabilità possono differire per **forma**. In altre parole, la forma dipende dal valore delle modalità più piccole (o più grandi) del valore centrale della distribuzione.

Aspetti che caratterizzano la forma di una distribuzione sono

- **asimmetria**
- **curtosi**



La forma di una distribuzione

Esercitazione
6

A. Iodice

La forma di
una
distribuzione

Indici di
asimmetria

Il box-plot

box-plot e
valori anomali

Confronto
grafico
istogramma/
box-plot

Indice di
asimmetria
per variabili
qualitative
ordinate

Due distribuzioni aventi stessa posizione e variabilità possono differire per **forma**. In altre parole, la forma dipende dal valore delle modalità più piccole (o più grandi) del valore centrale della distribuzione.

Aspetti che caratterizzano la forma di una distribuzione sono

- **asimmetria**
- **curtosi**



Asimmetria

Esercitazione 6

A. Iodice

La forma di una distribuzione

Indici di asimmetria

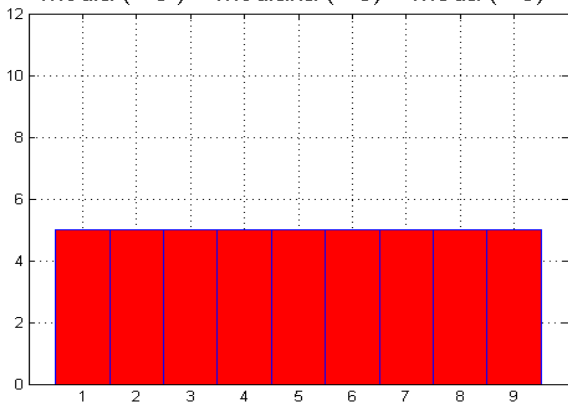
Il box-plot

box-plot e valori anomali

Confronto grafico istogramma/box-plot

Indice di asimmetria per variabili qualitative ordinate

distribuzione simmetrica rettangolare
media (= 5) = mediana (= 5) = moda (= 5)





Asimmetria

Esercitazione
6

A. Iodice

La forma di
una
distribuzione

Indici di
asimmetria

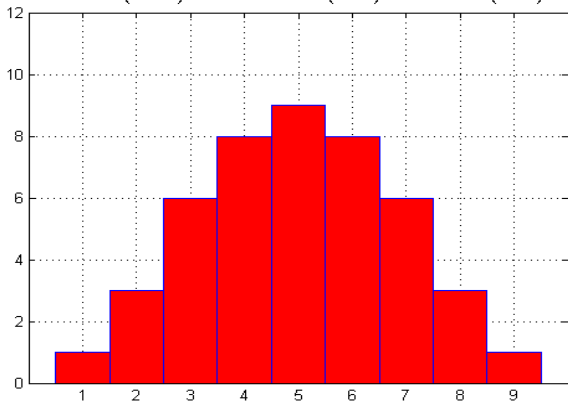
Il box-plot

box-plot e
valori anomali

Confronto
grafico
istogramma/
box-plot

Indice di
asimmetria
per variabili
qualitative
ordinate

distribuzione simmetrica campanulare
media (= 5) = mediana (= 5) = moda (= 5)





Asimmetria

Esercitazione
6

A. Iodice

La forma di
una
distribuzione

Indici di
asimmetria

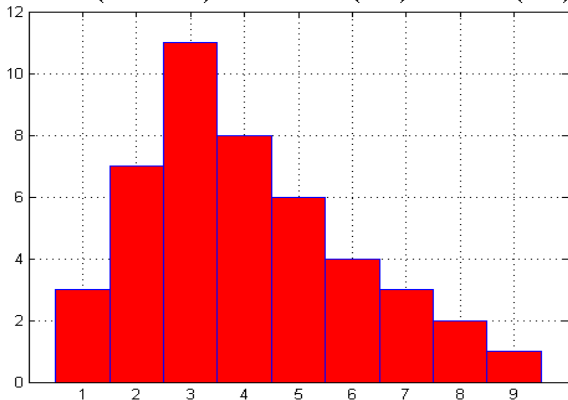
Il box-plot

box-plot e
valori anomali

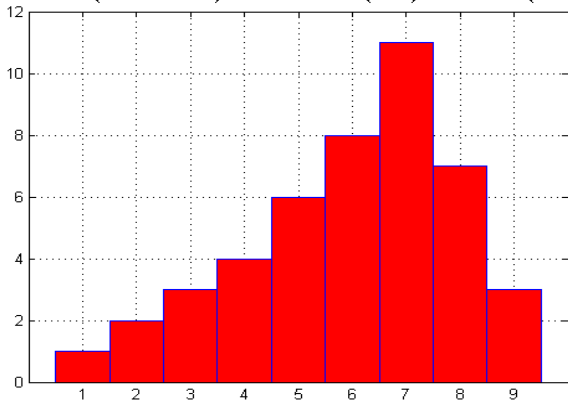
Confronto
grafico
istogramma/
box-plot

Indice di
asimmetria
per variabili
qualitative
ordinate

distribuzione asimmetrica positiva
media (= 4.044) > mediana (= 4) > moda (= 3)



distribuzione asimmetrica negativa
media (= 5.9556) < mediana (= 6) < moda (= 7)





Esercizio: confronto tra distribuzioni

Esercitazione 6

A. Indice

La forma di
una
distribuzione

Indici di
asimmetria

Il box-plot

box-plot e
valori anomali

Confronto
grafico
istogramma/
box-plot

Indice di
asimmetria
per variabili
qualitative
ordinate

Si considerino tre studenti, X , Y e Z che nei primi 9 esami hanno riportato i seguenti voti:

X	Y	Z
18	22	21
20	23	22
21	24	23
23	24	24
25	25	25
26	27	26
27	29	27
27	29	28
30	30	29

- misurare la tendenza centrale, la variabilità e la forma della distribuzione dei voti di ciascun studente;
- utilizzare un opportuno strumento grafico per confrontare le tre distribuzioni.



Indici di asimmetria

Esercitazione

6

A. Indice

La forma di una distribuzione

Indici di asimmetria

Il box-plot

box-plot e valori anomali

Confronto grafico istogramma/box-plot

Indice di asimmetria per variabili qualitative ordinate

indice normalizzato di asimmetria \mathcal{A}

Tale indice la versione normalizzata della differenza tra media e mediana, dal momento che σ risulta essere in qualunque caso tale che $\sigma \geq \mu - Me$

$$\mathcal{A} = \frac{\mu - Me}{\sigma}$$

Tale indice varia nell'intervallo $[-1, 1]$.



Indici di asimmetria

Esercitazione

6

A. Indice

La forma di una distribuzione

Indici di asimmetria

Il box-plot

box-plot e valori anomali

Confronto grafico istogramma/box-plot

Indice di asimmetria per variabili qualitative ordinate

indice normalizzato di asimmetria \mathcal{A}

Tale indice la versione normalizzata della differenza tra media e mediana, dal momento che σ risulta essere in qualunque caso tale che $\sigma \geq \mu - Me$

$$\mathcal{A} = \frac{\mu - Me}{\sigma}$$

Tale indice varia nell'intervallo $[-1, 1]$.

- se $\mathcal{A} > 0$ allora la distribuzione asimmetrica positiva
- se $\mathcal{A} < 0$ allora la distribuzione asimmetrica negativa
- se $\mathcal{A} = 0$ allora la distribuzione simmetrica



Indici di asimmetria

Esercitazione
6

A. Iodice

La forma di
una
distribuzione

Indici di
asimmetria

Il box-plot

box-plot e
valori anomali

Confronto
grafico
istogramma/
box-plot

Indice di
asimmetria
per variabili
qualitative
ordinate

Standardizzazione di una variabile

L'operazione di standardizzazione consiste nel sottrarre a ciascuna modalità x_i la media μ , dividendo poi per lo scarto quadratico medio σ . Tale operazione consente il confronto tra distribuzioni con medie e varianze diverse.

$$z_i = \frac{x_i - \mu}{\sigma}$$



Indici di asimmetria

Esercitazione 6

A. Indice

La forma di
una
distribuzione

Indici di
asimmetria

Il box-plot

box-plot e
valori anomali

Confronto
grafico
istogramma/
box-plot

Indice di
asimmetria
per variabili
qualitative
ordinate

indice di asimmetria di Fisher γ

$$\gamma = \frac{1}{n} \sum_{i=1}^n (z_i)^3 = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right)^3$$



Indici di asimmetria

Esercitazione

6

A. Iodice

La forma di una distribuzione

Indici di asimmetria

Il box-plot

box-plot e valori anomali

Confronto grafico istogramma/box-plot

Indice di asimmetria per variabili qualitative ordinate

indice di asimmetria di Fisher γ

$$\gamma = \frac{1}{n} \sum_{i=1}^n (z_i)^3 = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right)^3$$

- se $\gamma > 0$ allora la distribuzione asimmetrica positiva
- se $\gamma < 0$ allora la distribuzione asimmetrica negativa
- se $\gamma = 0$ allora la distribuzione simmetrica



Esercizio: confronto tra distribuzioni

Esercitazione
6

A. Iodice

La forma di
una
distribuzione

Indici di
asimmetria

Il box-plot

box-plot e
valori anomali

Confronto
grafico
istogramma/
box-plot

Indice di
asimmetria
per variabili
qualitative
ordinate

Si considerino tre studenti, X , Y e Z che nei primi 9 esami hanno riportato i seguenti voti:

- misurare la **tendenza centrale**, la **variabilità** e la forma della distribuzione dei voti di ciascun studente;

x_i	y_i	z_i	$sc_{x_i} = (x_i - \mu_x)^2$	$sc_{y_i} = (y_i - \mu_y)^2$	$sc_{z_i} = (z_i - \mu_z)^2$
18	22	21	37.346	15.123	16
20	23	22	16.901	8.346	9
21	24	23	9.679	3.568	4
23	24	24	1.235	3.568	1
25	25	25	0.790	0.790	0
26	27	26	3.568	1.235	1
27	29	27	8.346	9.679	4
27	29	28	8.346	9.679	9
30	30	29	34.679	16.901	16
217	233	225	120.89	68.89	60

- $\mu_x = \frac{\sum(x_i)}{n} = \frac{217}{9} = 24.11$, $\mu_y = \frac{\sum(y_i)}{n} = \frac{233}{9} = 25.89$, $\mu_z = \frac{\sum(z_i)}{n} = \frac{225}{9} = 25$

- $\sigma_x = \sqrt{\frac{\sum(x_i - \mu_x)^2}{n}} = \sqrt{\frac{120.89}{9}} = \sqrt{13.43} = 3.66$,

$$\sigma_y = \sqrt{\frac{\sum(y_i - \mu_y)^2}{n}} = \sqrt{\frac{68.89}{9}} = \sqrt{7.65} = 2.76,$$

$$\sigma_z = \sqrt{\frac{\sum(z_i - \mu_z)^2}{n}} = \sqrt{\frac{60}{9}} = \sqrt{6.67} = 2.58$$



Indici di asimmetria

Esercitazione
6

A. Iodice

La forma di
una
distribuzione

Indici di
asimmetria

Il box-plot

box-plot e
valori anomali

Confronto
grafico
istogramma/
box-plot

Indice di
asimmetria
per variabili
qualitative
ordinate

calcolo dell'indice normalizzato di asimmetria \mathcal{A}

$$\mathcal{A} = \frac{\mu - Me}{\sigma}$$

- $\mathcal{A}_x = \frac{\mu_x - Me_x}{\sigma_x} = \frac{24.11 - 25}{3.66} = -0.242$
- $\mathcal{A}_y = \frac{\mu_y - Me_y}{\sigma_y} = \frac{25.89 - 25}{2.76} = 0.321$
- $\mathcal{A}_z = \frac{\mu_z - Me_z}{\sigma_z} = \frac{25 - 25}{2.58} = 0$



Indice di asimmetria di Fisher γ

Esercitazione 6

A. Indice

La forma di una distribuzione

Indici di asimmetria

Il box-plot

box-plot e valori anomali

Confronto grafico istogramma/box-plot

Indice di asimmetria per variabili qualitative ordinate

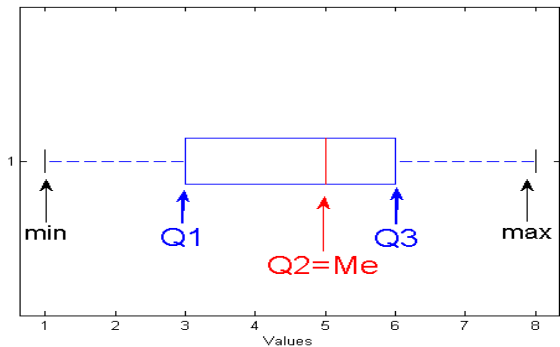
calcolo indice γ

$$\gamma = \frac{1}{n} \sum_{i=1}^n (z_i)^3 = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right)^3$$

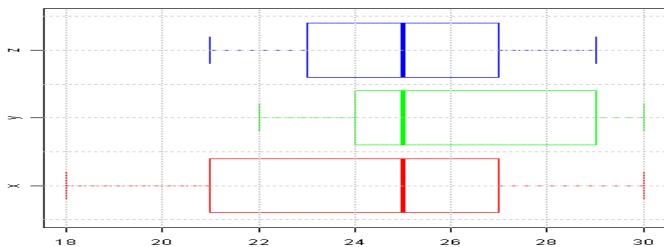
S_x	S_y	S_z	S_x^3	S_y^3	S_z^3
-1.67	-1.41	-1.55	-4.64	-2.78	-3.72
-1.12	-1.04	-1.16	-1.41	-1.14	-1.57
-0.85	-0.68	-0.77	-0.61	-0.32	-0.46
-0.30	-0.68	-0.39	-0.03	-0.32	-0.06
0.24	-0.32	0.00	0.01	-0.03	0.00
0.52	0.40	0.39	0.14	0.06	0.06
0.79	1.12	0.77	0.49	1.42	0.46
0.79	1.12	1.16	0.49	1.42	1.57
1.61	1.49	1.55	4.15	3.28	3.72
			-1.41	1.6	0

- $\gamma_x = \frac{\sum(x_i - \mu_x)}{n} = \frac{-1.41}{9} = -0.156$
- $\gamma_y = \frac{\sum(y_i - \mu_y)}{n} = \frac{1.604}{9} = 0.178$
- $\gamma_z = \frac{\sum(z_i - \mu_z)}{n} = \frac{0}{9} = 0$

Il grafico a scatola (box-plot) è una particolare rappresentazione di una distribuzione: gli elementi utilizzati per *costruire* la scatola sono i **quantili** e gli estremi della distribuzione.



- confrontare graficamente le tre distribuzioni



- $\min_x = 18, \max_x = 30, Q_1 = 21, Q_2 = 25, Q_3 = 27$
- $\min_y = 22, \max_y = 30, Q_1 = 24, Q_2 = 25, Q_3 = 27$
- $\min_z = 21, \max_z = 29, Q_1 = 23, Q_2 = 25, Q_3 = 29$



Costruzione di un box plot

Esercitazione 6

A. Iodice

La forma di
una
distribuzione

Indici di
asimmetria

Il box-plot

box-plot e
valori anomali

Confronto
grafico
istogramma/
box-plot

Indice di
asimmetria
per variabili
qualitative
ordinate

Si consideri la seguente distribuzione di frequenze

Modalità	Frequenze
23	3
24	6
25	8
26	5
27	5
28	3
88	1

6



Costruzione di un box plot

Esercitazione 6

A. Iodice

La forma di una distribuzione

Indici di asimmetria

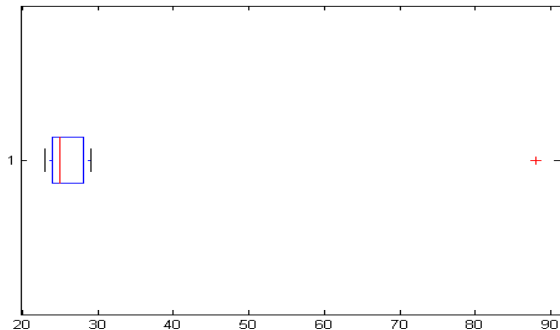
Il box-plot

box-plot e valori anomali

Confronto grafico istogramma/box-plot

Indice di asimmetria per variabili qualitative ordinate

La rappresentazione evidenzia la presenza di un valore anomalo





Costruzione di un box plot

Esercitazione 6

A. Iodice

La forma di una distribuzione

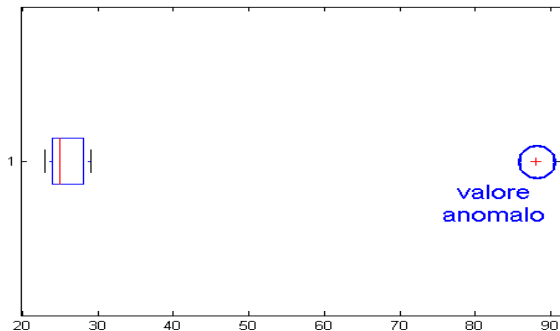
Indici di asimmetria

Il box-plot

box-plot e valori anomali

Confronto grafico istogramma/box-plot

Indice di asimmetria per variabili qualitative ordinate





Costruzione di un box plot

Esercitazione 6

A. Iodice

La forma di
una
distribuzione

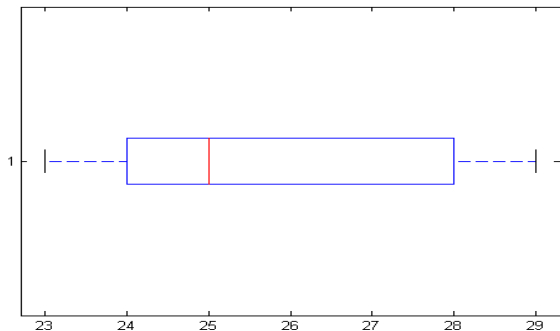
Indici di
asimmetria

Il box-plot

**box-plot e
valori anomali**

Confronto
grafico
istogramma/
box-plot

Indice di
asimmetria
per variabili
qualitative
ordinate





Costruzione di un box plot

Esercitazione 6

A. Iodice

La forma di
una
distribuzione

Indici di
asimmetria

Il box-plot

**box-plot e
valori anomali**

Confronto
grafico
istogramma/
box-plot

Indice di
asimmetria
per variabili
qualitative
ordinate

I valori anomali, che quindi non partecipano alla costruzione della scatola, vengono determinati dal confronto con il campo di variazione interquartile. In particolare vengono considerate due **soglie**:



Costruzione di un box plot

Esercitazione
6

A. Iodice

La forma di
una
distribuzione

Indici di
asimmetria

Il box-plot

box-plot e
valori anomali

Confronto
grafico
istogramma/
box-plot

Indice di
asimmetria
per variabili
qualitative
ordinate

I valori anomali, che quindi non partecipano alla costruzione della scatola, vengono determinati dal confronto con il campo di variazione interquartile. In particolare vengono considerate due **soglie**:

$$Q_1 - 1,5 \times (Q_3 - Q_1)$$

rappresenta il valore al di sotto del quale una modalità considerata *outlier*



Costruzione di un box plot

Esercitazione 6

A. Iodice

La forma di
una
distribuzione

Indici di
asimmetria

Il box-plot

box-plot e
valori anomali

Confronto
grafico
istogramma/
box-plot

Indice di
asimmetria
per variabili
qualitative
ordinate

I valori anomali, che quindi non partecipano alla costruzione della scatola, vengono determinati dal confronto con il campo di variazione interquartile. In particolare vengono considerate due **soglie**:

$$Q_1 - 1,5 \times (Q_3 - Q_1)$$

rappresenta il valore al di sotto del quale una modalità considerata *outlier*

$$Q_3 + 1,5 \times (Q_3 - Q_1)$$

rappresenta il valore al di sopra del quale una modalità considerata *outlier*



Box plot e istogramma

Esercitazione 6

A. Iodice

La forma di una distribuzione

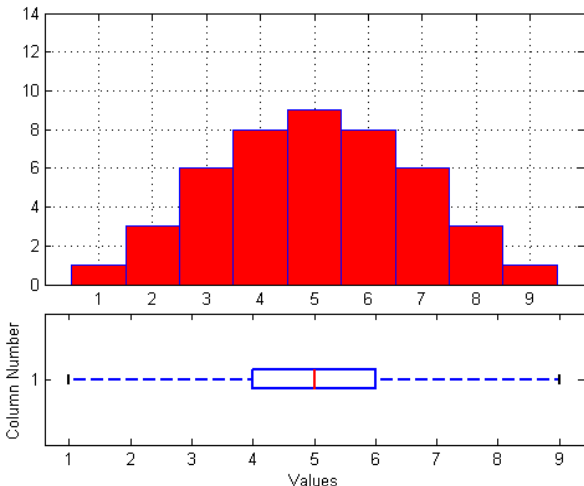
Indici di asimmetria

Il box-plot

box-plot e valori anomali

Confronto grafico istogramma/box-plot

Indice di asimmetria per variabili qualitative ordinate





Box plot e istogramma

Esercitazione 6

A. Iodice

La forma di una distribuzione

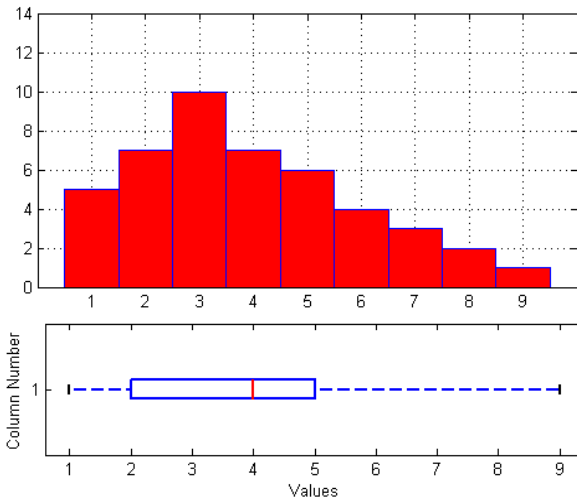
Indici di asimmetria

Il box-plot

box-plot e valori anomali

Confronto grafico istogramma/box-plot

Indice di asimmetria per variabili qualitative ordinate





Box plot e istogramma

Esercitazione 6

A. Iodice

La forma di una distribuzione

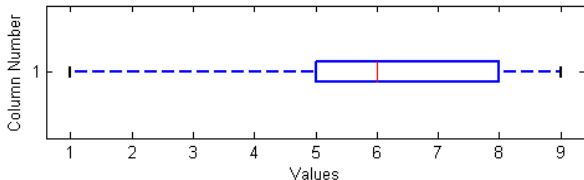
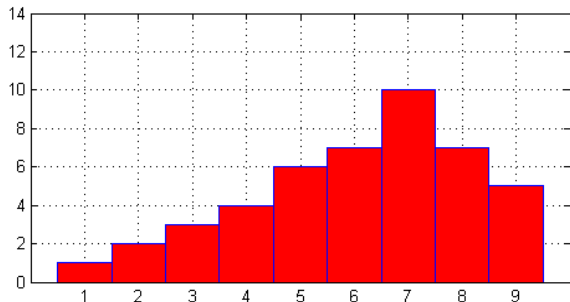
Indici di asimmetria

Il box-plot

box-plot e valori anomali

Confronto grafico istogramma/box-plot

Indice di asimmetria per variabili qualitative ordinate





Box plot e istogramma

Esercitazione
6

A. Iodice

La forma di
una
distribuzione

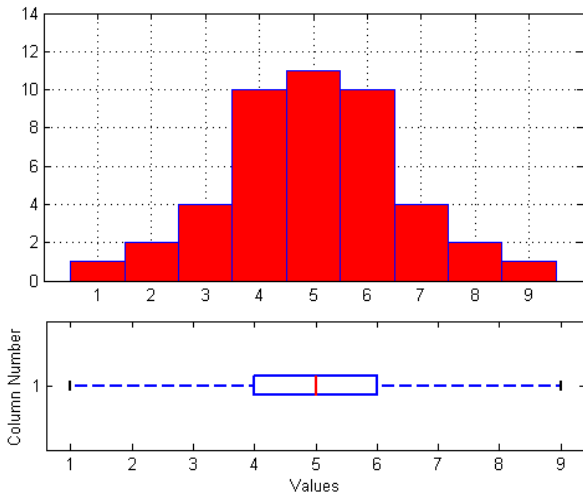
Indici di
asimmetria

Il box-plot

box-plot e
valori anomali

Confronto
grafico
istogramma/
box-plot

Indice di
asimmetria
per variabili
qualitative
ordinate





Box plot e istogramma

Esercitazione 6

A. Iodice

La forma di una distribuzione

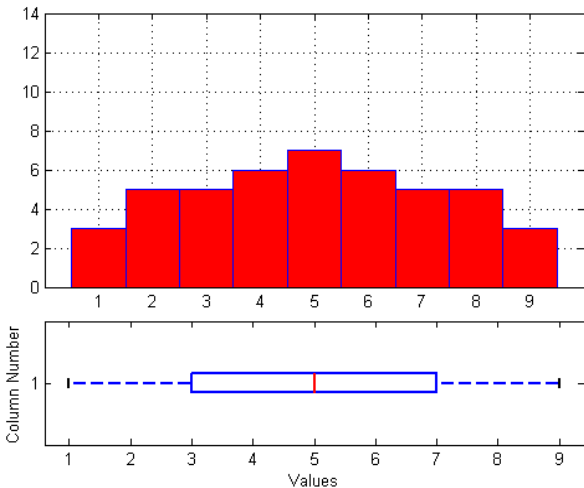
Indici di asimmetria

Il box-plot

box-plot e valori anomali

Confronto grafico istogramma/box-plot

Indice di asimmetria per variabili qualitative ordinate





Indice di asimmetria per variabili qualitative ordinate

Esercitazione 6

A. Iodice

La forma di una distribuzione

Indici di asimmetria

Il box-plot

box-plot e valori anomali

Confronto grafico istogramma/box-plot

Indice di asimmetria per variabili qualitative ordinate

Nel caso di variabili qualitative ordinali, è possibile misurare l'asimmetria attraverso un indice costruito sulla base della dispersione D .

L'indice D

L'indice D per il calcolo della dispersione in variabili qualitative ordinali si basa sulle frequenze **cumulate** F_j e **retrocumulate** RF_j , con $j = 1, \dots, k$, dove k è il numero di modalità della variabile. Ricordando che la frequenza relativa cumulata F_j e quella retrocumulata RF_j della j -esima modalità sono date rispettivamente da:

$$F_j = f_1 + f_2 + \dots + f_j \text{ e } RF_j = f_j + f_{j+1} + \dots + f_K;$$

L'indice D è il seguente:

$$D = \sum_{j=1}^k [F_j(1 - F_j) + RF_j(1 - RF_j)]$$

indice di asimmetria A

Tale indice si basa sul confronto tra la dispersione che caratterizza la distribuzione delle modalità che si trovano alla sinistra di quella centrale e quella delle variabili alla destra di quelle centrali.

$$A = \frac{D_d - D_s}{D_d + D_s}$$

dove D_d è la dispersione a dx e D_s quella a sx



Indice di asimmetria per variabili qualitative ordinate

Esercitazione 6

A. l'indice

La forma di una distribuzione

Indici di asimmetria

Il box-plot

box-plot e valori anomali

Confronto grafico istogramma/box-plot

Indice di asimmetria per variabili qualitative ordinate

<i>t.studio</i>	<i>absFreqs</i>	<i>relFreqs</i>	F_j	RF_j
<i>analfabeta</i>	5	0.25	0.25	1.00
<i>lic.elementare</i>	5	0.25	0.50	0.75
<i>lic.media</i>	3	0.15	0.65	0.50
<i>diploma</i>	3	0.15	0.80	0.35
<i>laurea</i>	4	0.2	1.00	0.20
<i>tot</i>	20	1.00		

Esempio di calcolo dell'indice A

Per effettuare il calcolo dell'indice bisogna individuare le modalità della parte destra della distribuzione e quelle della parte sinistra. Poichè N è pari, $\frac{N}{2} = 10$. Dunque le modalità della parte sinistra sono $\{1, 2\}$, quelle della parte destra sono $\{3, 4, 5\}$.

$$D_s = \sum_{j=1}^2 [F_j(1 - F_j) + RF_j(1 - RF_j)] =$$
$$= [0.25(1 - 0.25) + 1(1 - 1)] + [0.5(1 - 0.5) + 0.75(1 - 0.75)] = 0.625$$

$$D_d = \sum_{j=3}^5 [F_j(1 - F_j) + RF_j(1 - RF_j)] =$$
$$= [0.65(1 - 0.65) + 0.5(1 - 0.5)] + [0.8(1 - 0.8) + 0.35(1 - 0.35)] +$$
$$+ [1(1 - 1) + 0.2(1 - 0.2)] = 1.025$$



Indice di dispersione per variabili qualitative ordinate

Esercitazione 6

A. Indice

La forma di una distribuzione

Indici di asimmetria

Il box-plot

box-plot e valori anomali

Confronto grafico istogramma/box-plot

Indice di asimmetria per variabili qualitative ordinate

<i>t.studio</i>	<i>absFreqs</i>	<i>relFreqs</i>	F_j	RF_j
<i>analfabeta</i>	5	0.25	0.25	1.00
<i>lic.elementare</i>	5	0.25	0.50	0.75
<i>lic.media</i>	3	0.15	0.65	0.50
<i>diploma</i>	3	0.15	0.80	0.35
<i>laurea</i>	4	0.2	1.00	0.20
<i>tot</i>	20	1.00		

Esempio di calcolo dell'indice A

Per effettuare il calcolo dell'indice bisogna individuare le modalità della parte destra della distribuzione e quelle della parte sinistra. Poiché N è pari, $\frac{N}{2} = 10$. Dunque le modalità della parte sinistra sono $\{1, 2\}$, quelle della parte destra sono $\{3, 4, 5\}$.

$$A = \frac{D_d - D_s}{D_d + D_s} = \frac{1.025 - 0.625}{1.025 + 0.625} = \frac{0.4}{1.65} = 0.24$$