Esercitazioni di statistica

Boxplot e numeri indici

Stefania Spina

Universitá di Napoli Federico II

stefania.spina@unina.it

14 Ottobre 2014

Il Box-plot

La descrizione di un carattere mediante un indice di posizione andrebbe sempre accompagnata da un indice di variabilità.

Il **grafico a scatola o box-plot** è un particolare tipo di diagramma che permette di rappresentare contemporaneamente la tendenza centrale, la variabilità e la forma di una data distribuzione.

Capisaldi di un box-plot

Dal punto di vista operativo, **i capisaldi** della rappresentazione di un box-plot sono:

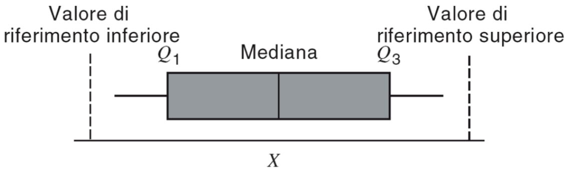
- $Q_0 = min(x_i)$
- $Q_1 = 1^\circ quartile$
- $Q_2 = mediana\ o\ 2^\circ quartile$
- $Q_3 = 3^\circ quartile$;
- $Q_4 = max(x_i)$
- $IQR = Q_3 - Q_1$ = campo di variazione interquartile

Elementi di un box-plot

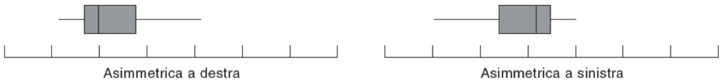
Gli elementi che caratterizzano un box-plot dal punto di vista grafico sono:

- una linea orizzontale interna alla scatola, che individua il valore dell'indice di posizione (media o mediana);
- un rettangolo (box) la cui altezza misura la variabilità della distribuzione del carattere intorno ad una media. Di solito contiene il 50% centrale della distribuzione (dal 1° al 3° quartile);
- due segmenti (**i baffi**) che individuano gli intervalli in cui sono posizionati i valori rispettivamente minori di Q_1 e maggiori di Q_3;
- il **valore di riferimento inferiore** di un box-plot è posizionato a $Q_1 - 1.5(IQR)$ e quello **superiore** a $Q_3 + 1.5(IQR)$.

Box - plot



Box-plot per dati asimmetrici



Box - plot, forma di una distribuzione e valori anomali

Dal box-plot si legge l'asimmetria di una distribuzione se:

- la distanza tra il primo quartile e la mediana è diversa dalla distanza tra mediana e terzo quartile;

- la distanza tra estremo inferiore e primo quartile è diversa dalla distanza tra terzo quartile ed estremo superiore.

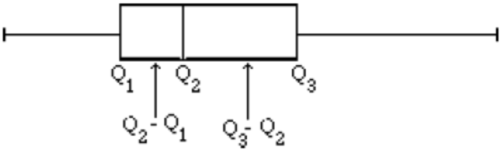
Dal box-plot si rileva la presenza di valori anomali.

Dal box-plot si rileva l'entità della variabilità, dal confronto tra i campi di variazione.

Dal box-plot paralleli si possono rilevare differenze in variabilità in due distribuzioni.

Come si legge l'asimmetria in un box-plot

Se la **differenza interquartile > 0** vuole dire che risulta (Q3 - Q2) > (Q2 - Q1) e quindi il boxplot assume una forma tipo quella riportata nella figura seguente ed in tal caso si dice che la distribuzione ha una **asimmetria positiva**: la coda di destra della distribuzione è più lunga, marcata della coda di sinistra.



Come si legge l'asimmetria in un box-plot

Se la **differenza interquartile < 0** vuole dire che (Q3 - Q2) < (Q2 - Q1) e quindi il box-plot ha una struttura come quella qui di sotto riportata. In tal caso si dice che la **distribuzione è asimmetrica negativa**: la coda di sinistra della distribuzione è più marcata di quella di destra.

Esercizio 1

Assumendo che la società Gamma s.p.a. in occasione di un'altra ricerca di personale qualificato abbia rilevato i seguenti livelli minimi di reddito desiderati da ulteriori 20 candidati:

$$\begin{array}{lcccccccccc} V2: & 4.4 & 5.2 & 2.9 & 2.9 & 2.9 & 4.1 & 1.5 & 2.9 & 2.9 & 0.7 \\ & 4.8 & 1.5 & 2.9 & 1.5 & 3.4 & 5.9 & 0.7 & 5.9 & 8.7 & 2.9 \end{array}$$

Rappresentare il box-plot della variabile Livello minimo di reddito desiderato

Esercizio 1

Ordiniamo i dati in ordine crescente

	Reddito desiderato
1	0.7
2	0.7
3	1.5
4	1.5
5	1.5
6	2.9
7	2.9
8	2.9
9	2.9
10	2.9
11	2.9
12	2.9
13	3.4
14	4.1
15	4.4
16	4.8
17	5.2
18	5.9
19	5.9
20	8.7

Esercizio 1

Mediana
Poichè N=20 si calcoleranno le posizioni $\frac{n}{2}$ e $\frac{n}{2}+1$

$$x_{(\frac{n}{2})} = x_{10} = 2.9 \quad x_{(\frac{n}{2}+1)} = x_{11} = 2.9$$
$$Me = \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2} = \frac{2.9+2.9}{2} = 2.9$$

1 Quartile

$$x_{(\frac{n}{4})} = x_5 = 1.5 \quad x_{(\frac{n}{4}+1)} = x_6 = 2.9$$
$$Q_1 = \frac{x_5 + x_6}{2} = \frac{1.5+2.9}{2} = 2.2$$

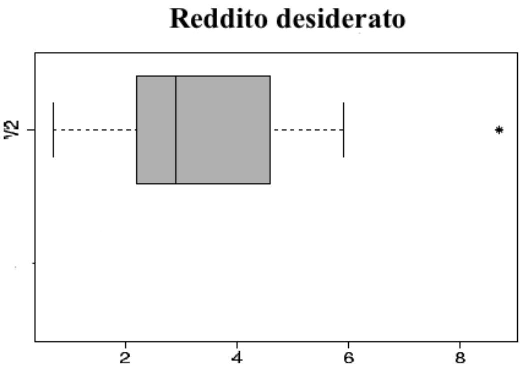
3 Quartile

$$x_{(\frac{3n}{4})} = x_{15} = 4.4 \quad x_{(\frac{3n}{4}+1)} = x_{16} = 4.8$$
$$Q_3 = \frac{x_{15} + x_{16}}{2} = \frac{4.4+4.8}{2} = 4.6$$

Valori Anomali

$$X < Q_1 - 1.5 * (Q_3 - Q_1); X < -1.4$$
$$X > Q_3 + 1.5 * (Q_3 - Q_1); X > 8.2 \Rightarrow 8.7$$

Esercizio 1



Esercizio 2

La tabella seguente riporta la distribuzione dei prezzi degli affitti mensili per un campione di 10 monolocali nel centro e in una zona periferica di una grande città.
Rappresentare ciascuna distribuzione attraverso un boxplot e commentare i risultati

	Prezzi affitti in centro	**Prezzi affitti in periferia**
1	955	750
2	1000	775
3	985	725
4	980	705
5	940	694
6	975	725
7	965	690
8	999	745
9	1247	575
10	1119	800

Esercizio 2

Rispetto ai prezzi affitti in centro, si mettono in ordine crescente

	Prezzi affitti in centro
1	940
2	955
3	965
4	975
5	900
6	985
7	999
8	1000
9	1119
10	1247

Mediana
Poichè N=10 si calcoleranno le posizioni $\frac{n}{2}$ e $\frac{n}{2}+1$

$$x_{(\frac{n}{2})} = x_5 = 980 \quad x_{(\frac{n}{2}+1)} = x_6 = 985$$

$$Me = \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2} = \frac{980+985}{2} = 982.5$$

1 Quartile

$$Pos\ Q_1 = \frac{(\frac{n}{4}) + (\frac{n}{4}+1)}{2} = \frac{2.5+3.5}{2} = 3 \rightarrow Q_1 = x_3 = 965$$

Esercizio 2

	Prezzi affitti in centro
1	940
2	955
3	965
4	975
5	980
6	985
7	999
8	1000
9	1119
10	1247

3 Quartile

$$Pos\ Q_3 = \frac{(\frac{3n}{4}) + (\frac{3n}{4} + 1)}{2} = \frac{7.5 + 8.5}{2} = 8 \rightarrow Q_3 = x_8 = 1000$$

Valori Anomali

$$X < Q_1 - 1.5 * (Q_3 - Q_1); X < 912.5$$

$$X > Q_3 + 1.5 * (Q_3 - Q_1); X > 1052.5 \Rightarrow 1119, 1247$$

Esercizio 2

Rispetto ai prezzi affitti in periferia, si mettono in ordine crescente

	Prezzi affitti in periferia
1	575
2	690
3	694
4	705
5	725
6	725
7	745
8	750
9	775
10	800

Mediana
Poichè N=10 si calcoleranno le posizioni $\frac{n}{2}$ e $\frac{n}{2}+1$

$$x_{(\frac{n}{2})} = x_5 = 725 \quad x_{(\frac{n}{2}+1)} = x_6 = 725$$

$$Me = \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2} = \frac{725+725}{2} = 725$$

1 Quartile

$$Pos\ Q_1 = \frac{(\frac{n}{4}) + (\frac{n}{4}+1)}{2} = \frac{2.5+3.5}{2} = 3 \rightarrow Q_1 = x_3 = 694$$

Esercizio 2

	Prezzi affitti in periferia
1	575
2	690
3	694
4	705
5	725
6	725
7	745
8	750
9	775
10	800

3 Quartile

$$Pos\ Q_3 = \frac{(\frac{3n}{4}) + (\frac{3n}{4} + 1)}{2} = \frac{7.5 + 8.5}{2} = 8 \rightarrow Q_3 = x_8 = 750$$

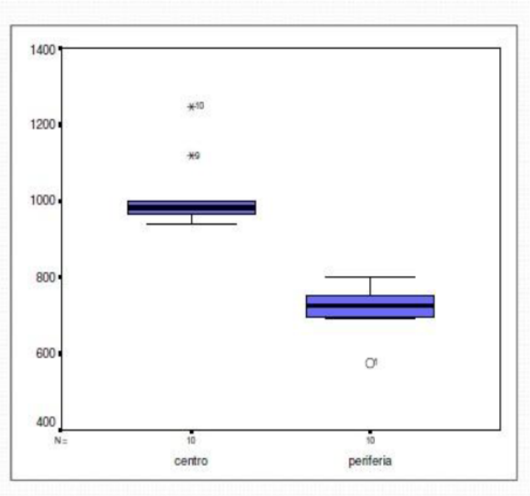
Valori Anomali

$$X < Q_1 - 1.5 * (Q_3 - Q_1); X < 610 \Rightarrow 575$$

$$X > Q_3 + 1.5 * (Q_3 - Q_1); X > 834$$

Esercizio 2

Una volta individuati gli elementi principali si possono costruire i due box-plot



Introduzione

Il **numero indice** è un rapporto che permette di confrontare le intensità o frequenze di un fenomeno in situazioni temporali o spaziali differenti.
Si costruisce ponendo al denominatore un'intensità (detta base) della stessa natura del fenomeno che è al numeratore.
Si può distinguere tra:
- **numeri indici a base fissa**, se il periodo di riferimento è costante al variare del tempo;
- **numeri indici a base mobile**, se per ciascuno di essi si fa riferimento al periodo precedente.

Introduzione

Un'altra distinzione si può fare tra:
- i **numeri indici semplici** che sono costituiti dal rapporto fra singole grandezze economiche riferite a beni omogenei;
- i **numeri indici composti** che sono costituiti dal rapporto fra medie di grandezze economiche eterogenee.

I numeri indici semplici

I **numeri indici semplici a base fissa** si hanno quando rimane fissa la base in tutta la serie, si ottiene un indice che consente di analizzare un fenomeno rispetto al tempo o al luogo usato come base:

$$_{b}I_{i} = \frac{p_i}{p_b} * 100$$

Se la base non rimane fissa, si ottengono i **numeri indici semplici a base mobile**

$$_{i-1}I_{i} = \frac{p_i}{p_{i-1}} * 100 \; per \quad i = 1, \ldots, T$$

Essi sono ottenuti rapportando ogni manifestazione al tempo corrente (t) alla circostanza immediatamente precedente (t-1).

Esercizio

Sono dati i prezzi in euro di due diversi tipi di olio da cucina A e B dal 2000 al 2003. Calcolare per ognuna delle due serie storiche, la serie dei numeri indici dei prezzi a base fissa 2000 e la serie dei numeri indici dei prezzi a base mobile.

	Olio A	**Olio B**
2000	3.2	4.2
2001	3.5	4.3
2002	3.6	4.8
2003	4.6	5.8

Esercizio

Sono dati i prezzi in euro di due diversi tipi di olio da cucina A e B dal 2000 al 2003. Calcolare per ognuna delle due serie storiche, la serie dei numeri indici dei prezzi a base fissa 2000 e la serie dei numeri indici dei prezzi a base mobile.

	Olio A	**B2000**	**Bmob.**	**Olio B**	**B2000**	**Bmob.**
2000	3.2	100	-	4.2	100	-
2001	3.5	109.4	109.4	4.3	102.4	102.4
2002	3.6	112.5	102.9	4.8	114.3	111.6
2003	4.6	143.8	127.8	5.8	138.1	120.8

I numeri indici composti

Se i confronti temporali o territoriali riguardano un fenomeno che risulta dal concorso di più componenti, allora è necessario effettuare una sintesi delle informazioni elementari relative alle componenti medesime: **i numeri indici composti**.

Costituiscono indicatori delle variazioni relative di più fenomeni eterogenei, resi confrontabili dal riferimento a prezzi, quantità, ecc. Si ottengono come sintesi di numeri indici semplici
Possiamo distinguere:

a. media di rapporti
b. rapporti tra medie

I numeri indici composti

Tra i numeri indici composti i due più importanti sono:

a. Indice di Laspeyres
b. Indice di Paasche

Indice di Laspeyres

E' un indice che misura le variazioni dei prezzi del tempo 0 (p_{j0}) al tempo i (p_{ji}) e usa come pesi le quantità relative al tempo iniziale (q_{j0}).
L'ipotesi è che le quantità consumate di ogni bene nella situazione i siano uguali a quelle consumate nella situazione base.

$$_0I_L = \frac{\sum_{j=1}^{s} p_{ji} q_{j0}}{\sum_{j=1}^{s} p_{j0} q_{j0}}$$

N.B. Tale indice, mantenendo costante la quantità del bene a quella del periodo iniziale, presenta una tendenziosità positiva, tende cioè a sopravvalutare gli aumenti dei prezzi e a sottovalutarne le diminuzioni.

Indice di Paasche

E' un indice che utilizza i pesi relativi al tempo finale (q_{ji}), quindi assume come quantità costanti quelle relative alla situazione i.

$$ {}_0I_i^P = \frac{\sum_{j=1}^{s} p_{ji} q_{ji}}{\sum_{j=1}^{s} p_{j0} q_{ji}} $$

N.B. Tale indice presenta una tendenziosità negativa

Esercizio

I numeri indici di Laspeyres e Paasche con anno base 2000 sono calcolati utilizzando le seguenti formule:

$$_0I_L = \frac{\sum_{j=1}^{s} p_{ji} q_{j0}}{\sum_{j=1}^{s} p_{j0} q_{j0}}$$

$$_0I_i^P = \frac{\sum_{j=1}^{s} p_{ji} q_{ji}}{\sum_{j=1}^{s} p_{j0} q_{ji}}$$

pertanto le corrispondenti serie sono calcolate utilizzando i dati nella seguente tabella dove sono prima calcolati i singoli termini della sommatoria e successivamente è calcolato l'indice.

Esercizio

Un macellaio ha venduto nel mese di maggio del 2008 una certa quantità di carne e vuole analizzare le variazioni di prezzo al mese di maggio del 2009. Calcolare l'indice di Laspeyres e Paasche.

TABELLA
Carne venduta dal macellaio (maggio 2008-maggio 2009)

Tipo di carni	Quantità venduta (in kg) maggio 2008	Quantità venduta (in kg) maggio 2009	Prezzo (in €) maggio 2008	Prezzo (in €) maggio 2009
Bovino adulto	200	150	2,9	2,94
Agnello	50	70	6,64	8,04
Coniglio	34	10	3,85	4,02
Maiale	200	150	2,66	2,81
Pollo	100	200	2,17	2,96
Tacchino	40	45	2,49	3,44
Vitello	170	160	5,54	5,39

Esercizio

TABELLA
Carne venduta dal macellaio (maggio 2008-maggio 2009)

Tipo di carni	q_{j0} Quantità venduta (in kg) maggio 2008	q_{ji} Quantità venduta (in kg) maggio 2009	p_{j0} Prezzo (in €) maggio 2008	p_{ji} Prezzo (in €) maggio 2009
Bovino adulto	200	150	2,9	2,94
Agnello	50	70	6,64	8,04
Coniglio	34	10	3,85	4,02
Maiale	200	150	2,66	2,81
Pollo	100	200	2,17	2,96
Tacchino	40	45	2,49	3,44
Vitello	170	160	5,54	5,39

Esercizio

TABELLA
Schema di calcolo per i numeri di Laspeyres e di Paasche

Tipo di carni	$p_{j0}q_{j0}$	$p_{ji}q_{ji}$	$p_{ji}q_{j0}$	$p_{j0}q_{ji}$
Bovino adulto	580,0	441,0	588,0	435,0
Agnello	332,0	562,8	402,0	464,8
Coniglio	130,9	40,2	136,7	38,5
Maiale	532,0	421,5	562,0	399,0
Pollo	217,0	592,0	296,0	434,0
Tacchino	99,6	154,8	137,6	112,1
Vitello	941,8	862,4	916,3	886,4
Totale	*2833,3*	*3074,7*	*3038,6*	*2769,8*

Esercizio

Indice di Laspeyres

$$_0I_L = \frac{\sum_{j=1}^{s} p_{ji}q_{j0}}{\sum_{j=1}^{s} p_{j0}q_{j0}} = \frac{3038.6}{2833.3} = 1.0725$$

Indice di Paasche

$$_0I_i^P = \frac{\sum_{j=1}^{s} p_{ji}q_{ji}}{\sum_{j=1}^{s} p_{j0}q_{ji}} = \frac{3074.7}{2769.8} = 1.1101$$

L'incremento dei prezzi risulta essere del 7.25% secondo l'indice di Laspeyres mentre è dell'11% secondo l'indice di Paasche.

Indice di Fisher

Per considerare contemporaneamente l'informazione fornita dai due numeri indici, si può calcolare la media geometrica dei due indici come suggerito da Fisher:

$$I_F = \sqrt{I_L * I_P}$$

Considerato l'esercizio precedente:

$$I_F = \sqrt{I_L * I_P} = \sqrt{1.0725 * 1.1101} = 1.0911$$

L'incremento medio dei prezzi è del 9.11%.

Esercizio

Il proprietario di un hotel chiede al suo consulente contabile alcune informazioni sulle spese sostenute per l'acquisto di quattro beni negli ultimi 5 anni. A tale scopo gli fornisce alcuni dati relativi al costo medio unitario (in Euro) ed al numero di unità di beni acquistati nei 5 anni di riferimento:

	Televisori		*Condizionatori*		*Frigo Bar*		*Impianti Stereo*	
Anni	prezzo (×100)	quantità	prezzo (×100)	quantità	prezzo (×100)	quantità	prezzo (×100)	quantità
1999	2.5	2	4	3	2.8	10	2.6	11
2000	2.7	7	4.8	6	3.1	2	2.9	5
2001	2.8	6	5.2	1	3.3	4	3.6	4
2002	3.1	15	4.9	4	3.5	1	2.8	3
2003	2.9	9	4.2	7	3.4	3	2.5	6

Il proprietario dell'hotel, allo scopo di avere dati di sintesi, chiede:

1. *La serie dei numeri indici a base fissa 2001 dei prezzi dei* Televisori
2. *La serie dei numeri indici a base mobile dei prezzi dei* Televisori
3. *Le serie dei numeri indici di Laspeyres e di Paasches con anno base 2000.*

Esercizio

1. La costruzione della serie dei numeri indici a base fissa 2001 dei Televisori è effettuata utilizzando i seguenti rapporti:

$$_bI_i = \frac{p_i}{p_b}$$

pertanto la serie richiesta è:

Anno	1999	2000	2001	2002	2003
$_{01}I_t$	0.893	0.964	1.00	1.107	1.036

2. La serie dei numeri indici a base mobile è invece costruita come

$$_{i-1}I_i = \frac{p_i}{p_{i-1}} \; per \quad i = 1, \ldots, T$$

e quindi:

Anno	1999	2000	2001	2002	2003
$_{t-1}I_t$	-	1.080	1.037	1.107	0.935

Esercizio

Numeri indici di Laspeyres

	Televisori	*Condizionatori*	*Frigo Bar*	*Impianti Stereo*		*Indice di Laspeyres*
Anni	$p_{ji} \times q_{jo}$	$p_{ji} \times q_{jo}$	$p_{ji} \times q_{jo}$	$p_{ji} \times q_{jo}$	$\sum_{i=1}^{k} p_{ji}\, q_{jo}$	${}_{00}I_t^L$
1999	17.5	24	5.6	13	60.1	0.879
2000	18.9	28.8	6.2	14.5	68.4	1.000
2001	19.6	31.2	6.6	18	75.4	1.102
2002	21.7	29.4	7	14	72.1	1.054
2003	20.3	25.2	6.8	12.5	64.8	0.947

$\sum_{j=1}^{s} p_{j0} q_{j0}$	2000	18.9	8.8	6.2	14.5	68.4

In maniera simile è costruita la serie dei numeri indici di Paasches che, a differenza del'indice di Laspeyres, richiede maggiori calcoli come evidenziato dalle seguenti tabelle.

Esercizio

Numeratori dei numeri indici di Paasches

	Televisori	Condizionatori	Frigo Bar	Impianti Stereo	
Anni	$p_{ji} \times q_{ji}$	$p_{ji} \times q_{ji}$	$p_{ji} \times q_{ji}$	$p_{ji} \times q_{ji}$	$\sum_{i=1}^{k} p_{ji}\, q_{ji}$
1999	5	12	28	28.6	73.6
2000	18.9	28.8	6.2	14.5	68.4
2001	16.8	5.2	13.2	14.4	49.6
2002	46.5	19.6	3.5	8.4	78
2003	26.1	29.4	10.2	15	80.7

Denominatori dei numeri indici di Paasches

	Televisori	Condizionatori	Frigo Bar	Impianti Stereo	
Anni	$p_{j0} \times q_{ji}$	$p_{j0} \times q_{ji}$	$p_{j0} \times q_{ji}$	$p_{j0} \times q_{ji}$	$\sum_{i=1}^{k} p_{j0}\, q_{ji}$
1999	5.4	14.4	31	31.9	82.7
2000	18.9	8.8	6.2	14.5	68.4
2001	16.2	4.8	12.4	11.6	45
2002	40.5	19.2	3.1	8.7	71.5
2003	24.3	33.6	9.3	17.4	84.6

e quindi la serie dei numeri indici di Paasches è:

Anno	1999	2000	2001	2002	2003
${}_{00}I_t^P$	0.890	1.000	1.102	1.091	0.954