

Corso di

Statistica

Gli indici di variabilità

Lezione tre

Cristina Davino

cdavino@unimc.it



Gli indici sintetici

- Tendenza centrale
- Forma
- Variabilità

Qualche considerazione

- ✓ La scelta dell'indice di tendenza centrale/posizione dipende dal tipo e dalle caratteristiche della distribuzione;
- ✓ Più che individuare l'indice "migliore in assoluto" (che non esiste), è importante anche valutare le differenze tra le diverse misure, che possono fornire ulteriori, importanti informazioni anche, ad esempio, sulla *forma* della distribuzione;

Gli indici sintetici

- Posizione
- Forma
- Variabilità

Qualche considerazione

- ✓ Volendo comunque definire delle caratteristiche dei diversi indici di tendenza centrale/posizione, possiamo dire che: (*Piccolo, 2003*)
 - ☞ La moda è utile quando occorre "minimizzare gli scontenti", e quindi in tutte quelle situazioni in cui il consenso e il numero delle singole unità ha significato per la decisione. In breve, la moda è un indice per governare;
 - ☞ La mediana minimizza i costi complessivi ed è resistente ai valori estremi. Quindi, la mediana è un indice per decisioni che implicano costi elevati nei casi estremi;
 - ☞ La media aritmetica è il baricentro dei dati e propone, quindi, un valore che equi-ripartisce il fenomeno tra le unità statistiche, pervenendo così a decisioni nelle quali contano, a parità numerica, gli estremi molto più dei valori centrali. Quindi, la media aritmetica è un indice di equilibrio generale.

Gli indici sintetici

- Posizione
- Forma
- Variabilità

- ✓ La variabilità di un fenomeno è la sua attitudine ad assumere differenti modalità.
- ✓ Un indice di variabilità è una misura di tale attitudine, e dovrebbe possedere almeno le seguenti caratteristiche:
 - ☞ E' nullo se e solo se tutte le unità presentano la stessa modalità del carattere;
 - ☞ Aumenta all'aumentare della diversità tra le unità.
- ✓ La variabilità può essere misurata
 - ↳ Rispetto a un centro (Dispersione)
 - ↳ Come misura delle differenze tra tutte le possibili coppie di unità osservate (Disuguaglianza)

Gli indici sintetici

- Posizione
- Forma
- Variabilità

✓ Indici assoluti

→ Dipendono dalla natura della variabile che si sta esaminando e sono espressi nella stessa unità di misura della variabile.

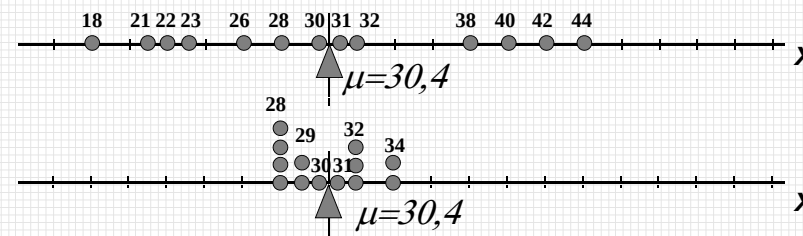
✓ Indici relativi

→ Sono svincolati dall'unità di misura perché costruiti come rapporti tra indici assoluti o tra indici assoluti e loro valori estremi. Sono, quindi, numeri puri, utili per confrontare fenomeni omogenei.

✓ Indici normalizzati

→ Sono particolari indici relativi che variano in un intervallo finito, generalmente in $[0, 1]$ oppure in $[-1, +1]$. Sono, quindi, di immediata interpretazione.

La variabilità rispetto a un centro



$$\sum_i (x_i - \mu) = 0$$

$$\sum_i (x_i - \mu)^2 \rightarrow \text{devianza} \xrightarrow{\text{Distribuzioni di frequenza}} \sum_i (x_i - \mu)^2 \cdot n_i$$

$$\frac{\sum_i (x_i - \mu)^2}{n} \rightarrow \text{varianza} \xrightarrow{\text{Distribuzioni di frequenza}} \frac{\sum_i (x_i - \mu)^2 \cdot n_i}{n}$$

La varianza

$$\text{Var}(X) = \frac{\sum_i (x_i - \mu)^2 \cdot n_i}{n}$$

Età studenti del Corso	Frequenze assolute (n_i)
18	2
19	44
20	66
21	32
22	18
23	13
24	9
25	6
Totale	190
Età media	20,6579

La varianza

$$\text{Var}(X) = \frac{\sum_i (x_i - \mu)^2 \cdot n_i}{n}$$

Età studenti del Corso	Frequenze assolute (n_i)	$(x_i - \mu) n_i$
18	2	-5,32
19	44	-72,95
20	66	-43,42
21	32	10,95
22	18	24,16
23	13	30,45
24	9	30,08
25	6	26,05
Totale	190	0,00
Età media	20,6579	

La Varianza e lo Scarto quadratico medio

$$\text{Var}(X) = \sigma^2 = \frac{\sum_i (x_i - \mu)^2 \cdot n_i}{n}$$

Età studenti del Corso	Frequenze assolute (n _i)	(x _i - μ) n _i	(x _i - μ) ² n _i
18	2	-5,32	14,13
19	44	-72,95	120,94
20	66	-43,42	28,57
21	32	10,95	3,75
22	18	24,16	32,42
23	13	30,45	71,31
24	9	30,08	100,53
25	6	26,05	113,12
Totale	190	0,00	484,76
Età media	20,6579		

Il problema dell'unità di misura



Lo Scarto quadratico medio

$$\text{Sqm}(X) = \sigma = \sqrt{\frac{\sum_i (x_i - \mu)^2 \cdot n_i}{n}}$$

$$\text{Var}(X) = \frac{484,76}{190} = 2,55 \Rightarrow \text{Sqm}(X) = \sqrt{2,55} = 1,60$$

La Varianza

- La varianza è sempre non negativa (perché media di quadrati)
- E' un indice assoluto espresso nell'unità di misura del fenomeno al quadrato
- Ha come valore minimo 0 (se tutte le unità presentano lo stesso valore che è pari alla media)

Lo Scarto quadratico medio

- È espresso nella stessa unità di misura del carattere
- E' un indice assoluto espresso
- Ha come valore minimo 0 (se tutte le unità presentano lo stesso valore che è pari alla media)

La Varianza e lo Scarto quadratico medio

$$\sigma^2 = M(X^2) - [\mu]^2$$

Un metodo alternativo per il calcolo di σ^2

Età studenti del Corso	Frequenze assolute (n _i)
18	2
19	44
20	66
21	32
22	18
23	13
24	9
25	6
	190
μ	20,66
Varianza	2,5514

La Varianza e lo Scarto quadratico medio

$$\sigma^2 = M(X^2) - [\mu]^2$$

Un metodo alternativo per il calcolo di σ^2

Età studenti del Corso	Frequenze assolute (n _i)	X ²
18	2	324
19	44	361
20	66	400
21	32	441
22	18	484
23	13	529
24	9	576
25	6	625
	190	
μ	20,66	
Varianza	2,5514	

La Varianza e lo Scarto quadratico medio

$$\sigma^2 = M(X^2) - [\mu]^2$$

Un metodo alternativo per il calcolo di σ^2

Età studenti del Corso	Frequenze assolute (n _i)	X ²	X ² n _i
18	2	324	648
19	44	361	15884
20	66	400	26400
21	32	441	14112
22	18	484	8712
23	13	529	6877
24	9	576	5184
25	6	625	3750
	190		81567
μ	20,66		
Varianza	2,5514		

$$M(X^2) = \frac{81567}{190} = 429,30$$

$$[\mu]^2 = 20,66^2 = 426,75$$

$$\sigma_x^2 = 429,30 - 426,75 = 2,55$$

Il coefficiente di variazione

$$CV = \frac{\sigma}{\mu}$$



.) Non è definito per $\mu < 0$

.) Tende ad "esplodere" per $\mu \approx 0$



.) Confronto tra distribuzioni espresse in diverse unità di misura

.) Confronto tra distribuzioni espresse nella stessa unità di misura ma con valori medi molto differenti

La Varianza e lo Scarto quadratico medio

$$Var(X) = \frac{\sum (x_i - \mu)^2 \cdot n_i}{n}$$

Il confronto fra misure di variabilità

$$Sqm(X) = \sqrt{\frac{\sum (x_i - \mu)^2 \cdot n_i}{n}}$$

X Mamme	
Peso (Kg.)	n
45 - 50	4
50 - 55	12
55 - 60	22
60 - 65	40
65 - 70	19
70 - 75	3
	100

$$\mu(X) = 60,85$$

$$Var(X) = 31,528$$

$$Sqm(X) = 5,615$$

$$CV(X) = 0,092$$

Y Neonati	
Peso (Kg.)	n
1,5 - 2,0	5
2,0 - 2,5	12
2,5 - 3,0	25
3,0 - 3,5	35
3,5 - 4,0	18
4,0 - 4,5	5
	100

$$\mu(Y) = 3,07$$

$$Var(Y) = 0,358$$

$$Sqm(Y) = 0,598$$

$$CV(Y) = 0,195$$

Coefficiente di variazione

$$CV = \frac{\sigma}{\mu}$$



.) Non è definito per $\mu < 0$

.) Tende ad "esplodere" per $\mu \approx 0$

Altri indici di variabilità

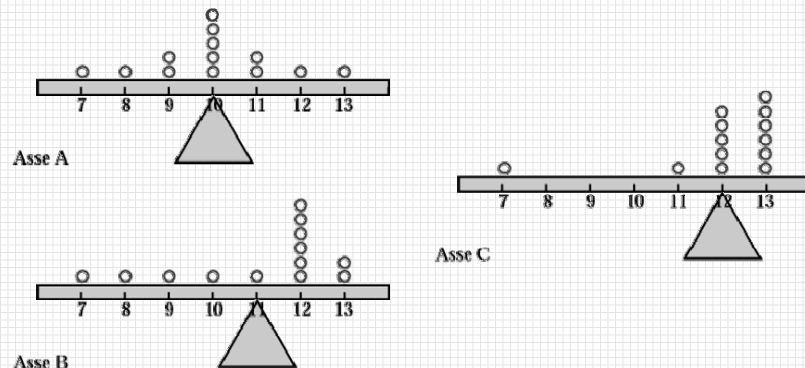
.) Campo di variazione (Range): $R = X_{\max} - X_{\min}$

.) Differenza interquartile: $Q_3 - Q_1$

.) Differenza semplice media: $\Delta = \frac{\sum_{i \neq j=1}^n |x_i - x_j|}{n \cdot (n-1)}$
 (Indice di mutua variabilità)

Altri indici di variabilità

.) Campo di variazione (*Range*): $R = x_{\max} - x_{\min}$



Dove e come studiare



- Libro di testo: D. Piccolo (2004) – Statistica per le decisioni – Il Mulino.
Cap. 5 (escluso paragrafi 5.4, 5.5, 5.6)
- Libro di testo: S. Borra, A. Di Ciaccio (2008) – Statistica – Metodologie per le scienze economiche e sociali – McGraw-Hill.
Cap. 4 (escluso paragrafi 4.4)



File “**esercizi indici sintetici.pdf**”

T2: Test di autovalutazione

(domande 5, 6, 9, 10, 12, 16, 20, 26)

- Esercizio n. 1 – punto 7
- Esercizio n.3 – punto e
- Esercizio n. 5 – punto b
- Esercizio n. 6 – punto b
- Esercizio n. 8 – punto b
- Esercizio n. 10 – punto b
- Esercizio n. 11 – punto a
- Esercizio n. 12 – punto a

Riepilogo

Gli indici sintetici: la variabilità

- ✓ Scarto quadratico medio, varianza, devianza (nel caso di una distribuzione unitaria semplice, di una distribuzione di frequenza, di una distribuzione in classi)
- ✓ Indici di variabilità relativi: coefficiente di variazione
- ✓ Campo di variazione
- ✓ Differenza interquartile
- ✓ Differenza semplice media

