
**Sintesi della distribuzione
di un carattere:
la variabilità (prima parte)**

Introduzione alla variabilità

- Nelle lezioni precedenti abbiamo visto come sia possibile riassumere tutte le informazioni disponibili attraverso l'utilizzo di una misura sintetica (come la media aritmetica o la mediana).
- I fenomeni (sociali, economici, ecc...) che vengono studiati tendono a variare. Tale elemento giustifica l'esistenza stessa delle metodologie quantitative; pertanto, sarà necessario comprendere in che modo tale variabilità si manifesta all'interno dell'analisi che stiamo effettuando.

Introduzione alla variabilità

ATTENZIONE!!!!

□ Quando sintetizziamo una distribuzione con una sola informazione quantitativa guadagniamo qualcosa in termini di *capacità informativa*, ma perdiamo qualcos'altro rispetto all'informazione disaggregata. Sarà dunque necessario capire se la sintesi effettuata è “buona”, in altri termini se l'informazione ottenuta riassume bene i dati di partenza, altrimenti possono risultare errate le conclusioni alle quali perveniamo (proprio perché la misura utilizzata non rappresenta bene i dati disponibili)

Introduzione alla variabilità

A)	40	50	60	$\bar{X}=50$
B)	50	50	50	$\bar{X}=50$
C)	0	0	150	$\bar{X}=50$

**Distribuzioni con media aritmetica uguale ma
differente variabilità**

Il Campo di Variazione

$$R = X_n - X_1$$

A)	40	50	60
B)	50	50	50
C)	0	0	150

$$A) R_A = X_n - X_1 = 60 - 40 = 20$$

$$B) R_B = X_n - X_1 = 50 - 50 = 0$$

$$C) R_C = X_n - X_1 = 150 - 0 = 150$$

Difetti: Ha un minimo, ma non ha un massimo definito;
è difficilmente interpretabile nel caso di *outliers*

Alcune considerazioni

Le misure appena presentate:

- ❑ **Prendono in considerazione solo due soggetti**
- ❑ **Non hanno estremi definiti**
- ❑ **Sono spesso di difficile interpretazione**
- ❑ **Sono misure assolute rendono difficile il confronto**

La Varianza

La **VARIANZA** è data dalla sommatoria di tutti gli scarti tra le singole modalità e la media aritmetica, elevate al quadrato e rapportata alla numerosità totale:

$$\text{Var} = \sigma^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}$$

X	X ²
1	1
2	4
3	9
5	25

$$\text{s.q.m.} = \sigma = \sqrt{\sigma^2}$$

Esempio 1 – Varianza e S.Q.M.

In un'azienda lavorano 5 dipendenti, per i quali l'azienda stessa sostiene un costo annuo pari, rispettivamente, a:

25.000€ 37.000€ 19.000€ 21.000€ 28.000€

Si calcoli:

A) il costo medio annuo per dipendente

B) la varianza

C) lo scarto quadratico medio

A) $(25.000+37.000+19.000+21.000+28.000)/5=26.000$

Esempio 1 – seconda parte

$$\text{B) } \text{Var} = \sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

$$[(25.000-26.000)^2 + (37.000-26.000)^2 + (19.000-26.000)^2 + (21.000-26.000)^2 + (28.000-26.000)^2]/5 =$$

$$= [(-1.000)^2 + (+11.000)^2 + (-7.000)^2 + (-5.000)^2 + (+2.000)^2]/5 =$$

$$= (1.000.000 + 121.000.000 + 49.000.000 + 25.000.000 + 4.000.000)/5 = (200.000.000 / 5) = 40.000.000$$

$$\text{C) } \sigma = \sqrt{\sigma^2} = \sqrt{40.000.000} = 6.324,5$$

Esempio 2 – Varianza e S.Q.M.

x_i	n_i	$x_i \cdot n_i$	$(x_i - \bar{X})$	$(x_i - \bar{X})^2$	$(x_i - \bar{X})^2 \cdot n_i$
1	11	11	-2,33	5,43	59,73
2	21	42	-1,33	1,77	37,17
3	32	96	-0,33	0,11	3,52
4	28	112	0,67	0,45	12,60
5	16	80	1,67	2,79	44,64
6	7	42	2,67	7,13	49,91
	115	383			207,57

$$\bar{X} = 383/115 = 3,33$$

MAI NEGATIVA

$$\text{Var} = \sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{X})^2 \cdot n_i}{n} = \frac{207,57}{115} = 1,80$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{1,80} = 1,34$$

Esempio 3 – Varianza e S.Q.M.

x_i	n_i	c_i	$c_i \cdot n_i$
1-5	45	3	135,0
6-9	35	7,5	262,5
10-19	26	14,5	377,0
20-29	20	24,5	490,0
30-39	15	34,5	517,5
40-49	10	44,5	445,0
	151		2.227,0

a) $\bar{X} = 2.227 / 151 = 14,75$

b) $R = X_n - X_1 = 49 - 1 = 48$

Calcolare:

a) La media aritmetica

b) Il campo di variazione

c) La varianza e lo scarto quadratico medio

Esempio 3 – seconda parte

x_i	n_i	c_i	$c_i \cdot n_i$	$(c_i - \bar{X})$	$(c_i - \bar{X})^2$	$(c_i - \bar{X})^2 \cdot n_i$
1-5	45	3	135,0	-11,75	138,06	6.212,70
6-9	35	7,5	262,5	-7,25	52,56	1.839,60
10-19	26	14,5	377,0	-0,25	0,06	1,56
20-29	20	24,5	490,0	9,75	95,06	1.901,20
30-39	15	34,5	517,5	19,75	390,06	5.850,90
40-49	10	44,5	445,0	29,75	885,06	8.850,60
	151		2.227,0			24.656,56

c)

$$\text{Var} = \sigma^2 = \frac{\sum_{i=1}^n (c_i - \bar{X})^2 \cdot n_i}{n} = \frac{24.656,56}{151} = 163,29$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{163,29} = 12,78$$

Indici di variabilità relativi

- **Giudicare la rappresentatività della media aritmetica**

Devo capire se la media che ho calcolato sintetizza bene le informazioni

- **Confrontare la variabilità di due distribuzioni**

Varianza e scarto quadratico medio dipendono dall'unità di misura e dal "livello"; non posso fare confronti

Rappresentatività della media

Distribuzione dei clienti di un'agenzia di viaggi per numero di viaggi effettuati nell'ultimo anno

x_i	n_i
1	8
2	5
3	6
4	1
	20

$$\bar{x} = \frac{\sum_{i=1}^n x_i \cdot n_i}{n} = \frac{40}{20} = 2$$

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot n_i}{n}$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{0,9} = 0,95$$

x_i	n_i	$x_i \cdot n_i$	$(x_i - \bar{x})^2 \cdot n_i$
2	20	40	$(2 - 2)^2 \times 40 = 0$
	20	40	0

MINIMA VARIABILITA'

$$\sigma = \sqrt{\sigma^2} = \sqrt{0} = 0$$

Rappresentatività della media - 2

x_i	n_i	$x_i \cdot n_i$	$(x_i - \bar{x})^2 \cdot n_i$
0	19	0	$(0 - 2)^2 \times 19 = 76$
40	1	40	$(40 - 2)^2 \times 1 = 1444$
	20	40	1.520

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot n_i}{n}$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{76} = 8,72$$

MASSIMA VARIABILITA'

$$\sigma_{\max} = \bar{x} \sqrt{(n-1)} = 2 \cdot \sqrt{(20-1)} = 2 \cdot \sqrt{19} = 2 \cdot 4,36 = 8,72$$

Rappresentatività della media - 3

$$\frac{\sigma}{\sigma_{\max}}$$

- + si avvicina a 0
- + Media rappresentativa
- + si avvicina a 1
- + Media NON rappresentativa

□ Nel nostro caso:

$$\frac{\sigma}{\sigma_{\max}} = \frac{0,95}{8,72} = 0,11$$

La Media rappresenta bene i dati

Confrontare la variabilità

□ Confrontare la variabilità di due distribuzioni

Varianza e scarto quadratico medio dipendono dall'unità di misura e dal "livello"; non posso fare confronti

□ Esempio: confronto tra peso adulti e neonati

78,4 65,8 72,2 85,6 75,2 58,7 69,9

3,875 2,954 3,458 4,512 2,722 3,158

Confrontare la variabilità - 2

Adulti

$$\bar{x} = \frac{(78,4 + 65,8 + 72,2 + 85,6 + 75,2 + 58,7 + 69,9)}{7} = 72,26$$

$$\sigma = \sqrt{\frac{\left[(78,4 - 72,26)^2 + (65,8 - 72,26)^2 + (72,2 - 72,26)^2 + (85,6 - 72,26)^2 + \right.}{7} \left. + (75,2 - 72,26)^2 + (58,7 - 72,26)^2 + (69,9 - 72,26)^2 \right]}{7}} = 8,066$$

Neonati

$$\bar{x} = \frac{(3,875 + 2,954 + 3,458 + 4,512 + 2,722 + 3,158)}{6} = 3,447$$

$$\sigma = \sqrt{\frac{\left[(3,875 - 3,447)^2 + (2,954 - 3,447)^2 + (3,458 - 3,447)^2 + \right.}{6} \left. + (4,512 - 3,447)^2 + (2,722 - 3,447)^2 + (3,158 - 3,447)^2 \right]}{6}} = 0,601$$

Confrontare la variabilità - 3

$$CV = \frac{\sigma}{x} \cdot 100$$

Coefficiente di Variazione

Più è grande CV, maggiore è la variabilità

$$CV_{\text{adu}} = \frac{\sigma}{x} \cdot 100 = \frac{8,066}{72,26} \cdot 100 = 0,112 \cdot 100 = 11,2$$

$$CV_{\text{neo}} = \frac{\sigma}{x} \cdot 100 = \frac{0,601}{3,447} \cdot 100 = 0,174 \cdot 100 = 17,4$$

$CV_{\text{neo}} > CV_{\text{adu}}$ - La variabilità è maggiore nei neonati

Variabilità – competenze acquisite

Cosa abbiamo imparato?

- Cos'è la variabilità
- Misurare la variabilità
- Valutare la “bontà” della media aritmetica
- Confrontare la variabilità di differenti distribuzioni

Quali strumenti usare?

□ Per misurare la variabilità

→ **Varianza**

□ Per la rappresentatività della media

→ σ_{\max}

□ Per confrontare differenti distribuzioni

→ **CV**