

Elementi di Psicometria

5-Indici di variabilità
vers. 1.0a (5 dicembre 2011)
versione per stampa

Germano Rossi¹

`germano.rossi@unimib.it`

¹Dipartimento di Psicologia, Università di Milano-Bicocca

2011-2012

Misure di variabilità

Gli indici di variabilità ci dicono quanto i valori sono dispersi attorno alla tendenza centrale.

Esempio

<i>valori</i>	$\sum X/N = M$
7 7 6 5 4 4 4 3	40/8=5
10 10 9 7 5 4 3 2 0 0	50/10=5

A livello di scala intervallo/rapporto ci sono diversi indici di variabilità:

- Campo di variazione o gamma (di oscillazione) o *range*
- Differenza interquartilica (DI o IQR)
- Deviazione media o scostamento semplice medio
- Varianza e deviazione standard

Misure di variabilità: campo di variazione

- Il **campo di variazione** o **gamma** (di oscillazione) o **range** o **intervallo** (per SPSS) è la differenza fra il valore massimo e quello minimo

$$\text{gamma} = \text{max} - \text{min}$$

Esempio

<i>valori</i>	<i>campo var.</i>
7 7 6 5 4 4 4 3	7-3=4
10 10 9 7 5 4 3 2 0 0	10-0=10

Misure di variabilità: differenza interquartilica

- La **differenza interquartilica** (DI, IQR) è la differenza fra il terzo e il primo quartile

$$IQR = Q_3 - Q_1$$

e corrisponde al 50% centrale dei valori della distribuzione

Esempio

<i>valori</i>	<i>IQR</i>
7 7 6 5 4 4 4 3	6.25-4=2.25
10 10 9 7 5 4 3 2 0 0	8.50-2.25=6.25

- La semi-differenza interquartilica è la metà dell'IQR e corrisponde al 25% dei valori sopra o sotto la mediana

Misure di variabilità

- Gli scarti dalla media potrebbero essere una misura di variabilità, sennonché abbiamo visto che
- la somma degli scarti dalla media è sempre pari a 0 (zero)
- Alcune soluzioni sono:
 - **Deviazione media (DM)** o **scostamento semplice medio (SSM)**: considerare gli scarti senza il segno (in valore assoluto) e fare la loro media

$$DM = \frac{\sum_{i=1}^N |X_i - \bar{X}|}{N}$$

- **Varianza (s^2 , var)**: elevare gli scarti a quadrato e fare la loro media

$$s^2 = var = \frac{\sum_{i=1}^N (X_i - \bar{X})^2}{N}$$

Misure di variabilità

- La scelta generale è caduta sulla varianza, perché minimizza le piccole differenze e massimizza le grandi differenze
- Però la varianza è un quadrato (un'area) e quindi si introduce anche una versione lineare (che è una distanza), lo scarto quadratico medio.
- Lo **scarto quadratico medio (sqm)** o **deviazione standard (ds, s)** è la radice quadrata della varianza

$$s = \sqrt{s^2} = \sqrt{var} = \sqrt{\frac{\sum (X_i - \bar{X})^2}{N}}$$

Misure di variabilità

Esempio

$$\text{Var}(1,2,3,4,5) =$$

$$\frac{[(1-3)^2 + (2-3)^2 + (3-3)^2 + (4-3)^2 + (5-3)^2]}{5} =$$

$$\frac{[(-2)^2 + (-1)^2 + (0)^2 + (1)^2 + (2)^2]}{5} = \frac{4 + 1 + 0 + 1 + 4}{5} =$$

$$\frac{10}{5} = 2$$

$$\text{DS}(1,2,3,4,5) = \sqrt{2} = 1.41$$

Misure di variabilità

- la varianza finora vista è calcolata sul campione

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

- possiamo però usare il campione per stimare la varianza della popolazione, in tal caso la formula diventa:

$$s^2 = \frac{\sum_{i=1}^N (X_i - \bar{X})^2}{N - 1}$$

- analogamente per la deviazione standard

$$\sigma = \sqrt{\frac{\sum (X_i - \bar{X})^2}{N}} \quad s = \sqrt{\frac{\sum (X_i - \bar{X})^2}{N - 1}}$$

Misure di variabilità: formula alternativa

Calcolare gli scarti dalla media, *nella maggior parte dei casi*, produce valori decimali che possono generare imprecisioni nei calcoli. Esiste quindi una formula alternativa da usare con i dati grezzi:

	X	$X - M$	$(X - M)^2$	X^2	
	2	-0,8	0,64	4	$Var =$
	2	-0,8	0,64	4	
	3	0,2	0,04	9	
	3	0,2	0,04	9	
	4	1,2	1,44	16	
Somma	14		2,8	42	$= \frac{42}{5} - 2.8^2$
Media	2,8		0,56	8,4	$= 8.4 - 7.84 =$
					$= 0.56$
					$= \frac{\sum X^2}{N} - \bar{X}^2$

La varianza è quindi uguale a... **la media dei quadrati meno il quadrato della media**

Misure di variabilità: formula alternativa

Con $N - 1$ la formula non è così semplice, ma bisogna “aggiustarla”

$$s^2 = \frac{N}{N-1} \frac{\sum X^2}{N} - \bar{X}^2$$

$$s = \frac{N}{N-1} \sqrt{\frac{\sum X^2}{N} - \bar{X}^2}$$

Uso della distribuzione di frequenza

In questo caso, tutte le formule finora viste si possono riscrivere tenendo conto dei valori e delle relative frequenze

$$M = \frac{\sum X_i}{N} = \frac{\sum X_i f_i}{N}$$
$$\sigma^2 = \frac{\sum (X_i - \bar{X})^2}{N} = \frac{\sum (X_i - \bar{X})^2 f_i}{N}$$

dove

- \sum significa sommatoria
- X_i è un singolo valore della variabile
- f_i è la frequenza di comparsa di quel valore X_i
- N è la numerosità

Proprietà della varianza (e dev. st)

Esercizio

Soluzione

1 $\text{var}(1,2,3,4,5)$

2 $\text{var}(3,4,5,6,7)$

3 $\text{var}(2,4,6,8,10)$

1 $5/4 * (55/5 - 3^2) = 2.5, s = 1.58$

2 $5/4 * (135/5 - 5^2) = 2.5, s = 1.58$

3 $5/4 * (220/5 - 6^2) = 10, s = 3.16$

1 i numeri da 1 a 5

2 i numeri della prima serie sommati a 2

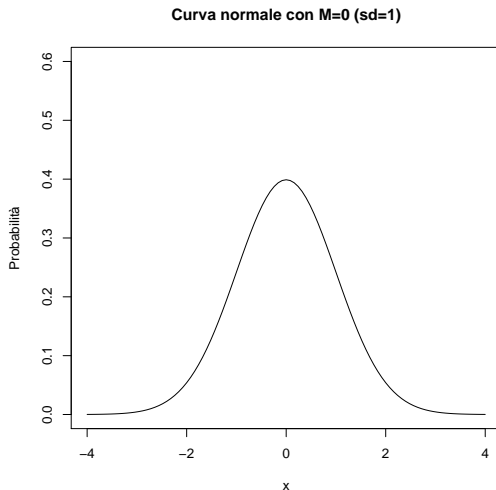
3 i numeri della prima serie moltiplicati per 2

Proprietà della var 1: Aggiungendo, sottraendo, una costante a tutti i dati della distribuzione, la varianza non subisce trasformazioni

Proprietà della var 2: Moltiplicando o dividendo per una costante, la varianza cambia ma la dev. st. subisce la stessa trasformazione

Misure di variabilità: curtosi e asimmetria

- **Curtosi:** è il grado di piattezza della curva della densità delle frequenze rispetto ad una curva particolare che è la **normale** (valore di riferimento = 0);
- **Asimmetria:** è il grado di asimmetria della curva della densità delle frequenze rispetto ad una curva particolare che è la **normale** (valore di riferimento = 0);



Spss: Variabilità

Anche varianza e dev. st. sono visualizzate da Spss (sempre N-1) in molte procedure. Quelle specifiche sono:

- `Analizza | Statistiche descrittive | Frequenze...` (fra le varie statistiche vi è anche quelle di variabilità)
- `Analizza | Statistiche descrittive | Descrittive...` (è la procedura specifica per le statistiche descrittive)
- `Analizza | Statistiche descrittive | Esplora...` (stampa le misure di variabilità come parte delle diverse statistiche per capire l'andamento e la distribuzione di una variabile)

Spss: Variabilità con Frequenze...

- Dopo aver scelto le variabili, click-are su **Statistiche...** e selezionare quelle che servono

Dispersione	Distribuzione
<input type="checkbox"/> Deviazione stand.	<input type="checkbox"/> Asimmetria
<input type="checkbox"/> Varianza	<input type="checkbox"/> Curtosi
<input type="checkbox"/> Intervallo	
<input type="checkbox"/> Minimo	
<input type="checkbox"/> Massimo	
<input type="checkbox"/> Errore standard della media	

- Quindi, click-are su **Continua**
- Con variabili quantitative conviene selezionare anche

☒ Visualizza tabelle di frequenza

☐ Sopprimi le tabelle con più di n categorie

Numero massimo di categorie:

- oppure in **Formato...**

- Poi **OK**

Spss: Variabilità con Descrittive...

- Dopo aver scelto le variabili, click-are su

Opzioni...

Dispersione

☐ Deviazione stand. ☐ Minimo

☐ Varianza ☐ Massimo

☐ Intervallo ☐ Errore standard della media

Distribuzione

☐ Curtosi ☐ Asimmetria

- Poi

Continua

e

OK

Statistiche descrittive

	N	Intervallo	Minimo	Massimo	Deviazione std.	Varianza	Asimmetria		Curtosi	
	Statistica	Statistica	Statistica	Statistica	Statistica	Statistica	Statistica	Errore std	Statistica	Errore std
gruppo	100	1	0	1	,499	,249	,245	,241	-1,980	,478
Antisemitismo	100	16,00	42,00	58,00	3,64499	13,286	-,092	,241	-,523	,478
Validi (listwise)	100									

Confronto fra statistiche

- **Campo di variazione** (Intervallo/Rapporto): è l'indice più grossolano
- **Differenza interquartilica** (Intervallo/Rapporto): poco usato in psicologia
- **Semi-differenza interquartilica** (Intervallo/Rapporto): pochissimo usato in psicologia
- **Deviazione media** (Intervallo/Rapporto): per nulla usato
- **Varianza, Deviazione standard** (Intervallo/Rapporto): i più usati
- **Curtosi** (Intervallo/Rapporto): per verificare la piattezza della distribuzione rispetto alla normale
- **Asimmetria** (Intervallo/Rapporto): per verificare se la distribuzione è simmetrica