

Metodi e tecniche di analisi dei dati nella ricerca psico-educativa Parte V

Laura Palmerio
Università Tor Vergata

A.A. 2005/2006

Variabili nominali

- **Moda** = **Punteggio più frequente**
 - sensibile a cambiamenti anche minimi
 - l'unico utilizzabile con variabili categoriali, ovvero con variabili misurabili solo su **scala nominale**

Variabili ordinali

Le misure di tendenza centrale

- **Mediana** = il valore che separa la metà superiore dei punteggi dalla metà inferiore

– N dispari $\Rightarrow (n + 1)/2$

– N pari $\Rightarrow$

Coppia mediana

- Posizione primo punteggio: $n/2$
- Posizione secondo punteggio: $n/2 + 1$



semisomma dei punteggi della coppia mediana:
$$\frac{\text{Posizione primo punteggio} + \text{Posizione secondo punteggio}}{2}$$

- Può essere calcolata su variabili misurate su **scala ordinale**
- Scarsa sensibilità ai valori numerici corrispondenti

Le misure di tendenza centrale

TAB. 2.11. Distribuzione di frequenza della variabile ordinale «autocollocazione sulla scala sinistra-destra»

	PRIMA DISTRIBUZIONE			SECONDA DISTRIBUZIONE		
	V.A.	%	% CUM.	V.A.	%	% CUM.
1. Estrema sinistra	52	2,2	2,2	153	10,1	10,1
2. Sinistra	531	22,0	24,2	254	16,8	26,9
3. Centro-sinistra	742	30,8	55,0	181	12,0	38,9
4. Centro	313	13,0	68,0	379	25,1	64,0
5. Centro-destra	505	20,9	88,9	156	10,3	74,3
6. Destra	243	10,1	99,0	207	13,7	88,0
7. Estrema destra	24	1,0	100	181	12,0	100
Totale	2.410	100		1.511	100	
Moda	Centro-sinistra			Centro		
Mediana	Centro-sinistra			Centro		
Q_1	3			2		
Q_3	5			6		
$Q = Q_3 - Q_1$	2			4		

Variabili cardinali

- **Media** = Somma dei valori / Numero dei valori
 - Può essere calcolata su variabili misurate su **scala a intervalli e a rapporti**
 - Notevole sensibilità ai valori estremi

Misure di variabilità: variabili nominali

Indice di omogeneità (O)

- **Distribuzione omogenea**: le frequenze sono concentrate su una o poche modalità
- **Distribuzione eterogenea**: le frequenze sono distribuite in modo equilibrato su tutte le modalità (k = numero delle modalità)
- Come si calcola: somma dei quadrati delle proporzioni (le frequenze relativizzate a 1: p_i)

$$\text{Indice di omogeneità } O = p_1^2 + p_2^2 + \dots + p_k^2$$

- Valore massimo = massima omogeneità = **1**
- Valore minimo = massima eterogeneità = **$1/k$**
- Per neutralizzare l'influenza del numero di modalità si può calcolare l'**Indice di omogeneità relativa (O_{rel})**:

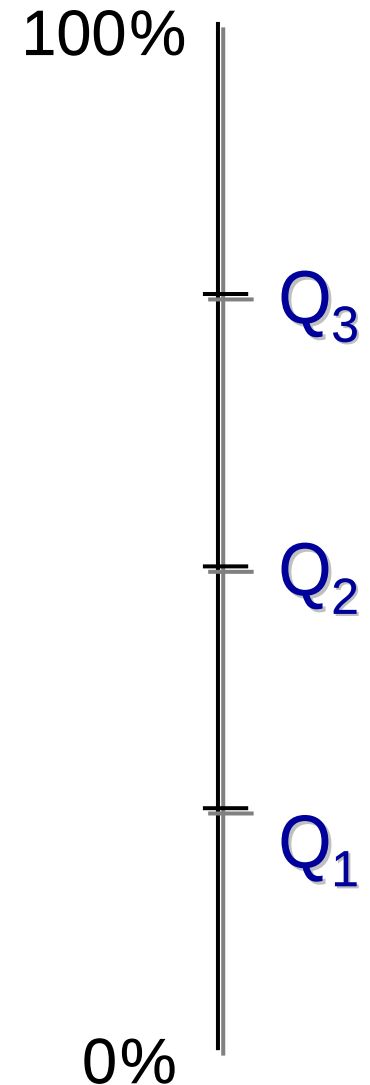
$$O_{rel} = \frac{k * O - 1}{k - 1}$$

- Valore massimo = massima omogeneità = **1**
- Valore minimo = massima eterogeneità = **0**

Misure di variabilità: variabili ordinali

Differenza interquartile

- Dividendo la distribuzione in 4 parti di uguale numerosità, i valori che segnano i confini fra i quattro quarti sono detti *quartili*, che come la mediana sono *valori di posizione*.
- **Primo quartile**: ha sotto di sé il 25% e sopra di sé il 75% della distribuzione
- **Secondo quartile**: separa il 50% inferiore dal 50% superiore, quindi coincide con la mediana
- **Terzo quartile**: ha il 75% dei casi sotto di sé e il 25% sopra di sé.
- **Differenza interquartile** = $Q_3 - Q_1$



Misure di variabilità: variabili cardinali

- **Campo di variazione (*Range*)** = Punteggio più alto – Punteggio più basso
- **Scostamento semplice medio** =
$$\frac{\sum |X_i - \bar{X}|}{n}$$
- **Varianza** = Somma degli scarti dalla media elevati al quadrato / Numero dei punteggi
$$\frac{\sum (X_i - \bar{X})^2}{n}$$
- **Deviazione standard** (o scarto quadratico medio) = radice quadrata della varianza

$$\sqrt{\frac{\sum (X_i - \bar{X})^2}{n}}$$

Misure di variabilità: variabili cardinali

- Poiché la D.S. risente della grandezza della media, per confrontare la variabilità di diverse distribuzioni $\Rightarrow$ **Coefficiente di variazione**

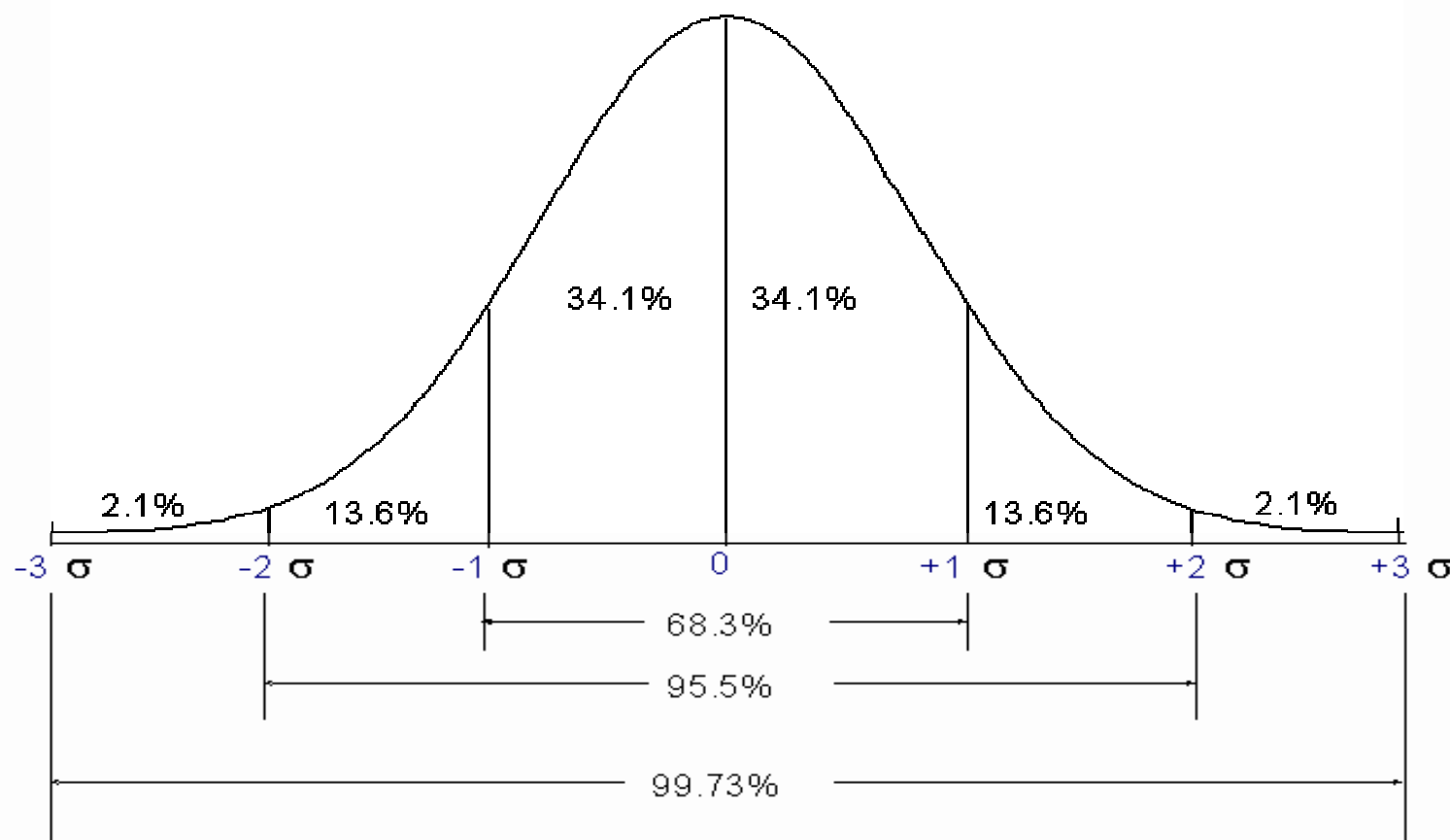
$$C_v = \frac{DS}{\bar{X}}$$

TAB. 2.17. Deviazione standard e coefficiente di variazione (voti alle elezioni politiche del 1983)

	Dc	PlI		Dc	PlI
Piemonte	27,6	6,6	Molise	55,5	2,2
Lombardia	33,4	3,8	Campania	36,2	2,4
Veneto	42,6	2,8	Puglia	36,3	2,1
Liguria	27,3	4,7	Basilicata	46,0	0,8
Friuli-V.G.	34,5	2,2	Calabria	36,8	0,9
Trentino-A.A.	27,6	1,6	Sardegna	31,7	1,5
Emilia-Romagna	22,8	2,3	Sicilia	26,9	1,7
Toscana	25,3	1,4			
Umbria	26,2	1,2			
Marche	33,4	1,6	$\bar{X}$	33,9	2,3
Lazio	31,1	2,7	S	8,01	1,41
Abruzzo	42,2	1,7	C_v	0,24	0,61

La curva normale

detta anche **distribuzione normale**, **curva gaussiana**, **curva a campana**



La curva normale

Proporzione dei casi nella curva normale

Limiti nei quali si collocano i casi	% dei casi
Fra la media e $+1.0$ DS o -1.0 DS	34.1
Fra la media e $+2.0$ DS o -2.0 DS	47.7
Fra la media e $+3.0$ DS o -3.0 DS	49.9
Fra $+1.0$ DS e -1.0 DS	68.3
Fra $+2.0$ DS e -2.0 DS	95.5
Fra $+3.0$ DS e -3.0 DS	99.8

- La standardizzazione dei punteggi (trasformazione in punti z o standard) consente di leggere la nostra distribuzione nei termini della curva normale
- Per trasformare un punteggio in un punteggio z:

$$\text{Punteggio } z = \frac{\text{Punteggio} - \text{Media}}{DS}$$