Psicometria

2-Ripasso di statistica
vers. 1.2

Germano Rossi[1]
germano.rossi@unimib.it
Giovanni Battista Flebus[1]
giovannibattista.flebus@unimib.it

[1] Dipartimento di Psicologia, Università di Milano-Bicocca

2008-2008

Livelli di misurazione

- Scala nominale (qualitativa, relazioni di uguaglianza/disuguaglianza)
- Scala ordinale (qualitativa, relazioni di uguaglianza/disuguaglianza e di ordine)
- Scala a intervallo (quantitativa, relazioni di uguaglianza/disuguaglianza e di ordine ed esistenza di una scala di misura unitaria valida per quella variabile con zero relativo)
- Scala a rapporto (quantitativa, relazioni di uguaglianza/disuguaglianza e di ordine ed esistenza di una scala di misura unitaria valida per quella variabile con zero assoluto)
- Ogni livello di scala include le caratteristiche della precedente

Frequenze, proporzioni e percentuali

Frequenza: conteggio del numero di occorrenze di un determinato valore di una variabile.

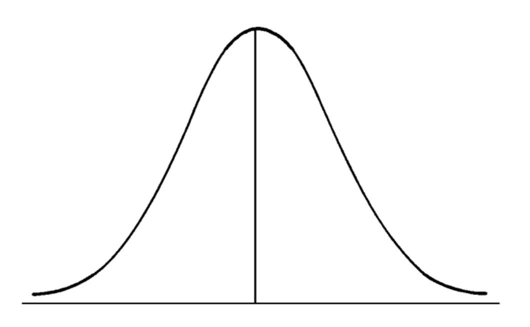
Proporzione ($0 \leq x \leq 1$)	$\frac{f}{N}$
Percentuale ($0 \leq x \leq 100$)	$\frac{f}{N} \cdot 100$

Una proporzione o una % può non avere senso se non si conosce l'N su cui è stata calcolata

Media

- E' il punto di equilibrio di una distribuzione di dati, coincide con il valore atteso e, se la distribuzione è normale, è anche il valore più probabile, cioè la “speranza matematica”, solitamente indicata con E(x).

$$\overline{X} = \frac{\sum X_i}{N}$$

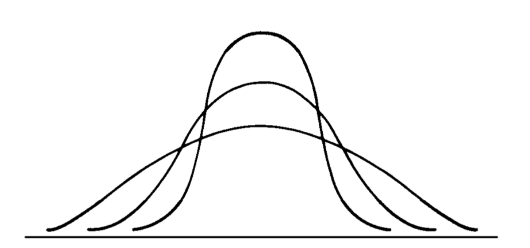


Varianza, deviazione standard

- Indicano la dispersione attorno alla media

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

$$s = \sqrt{s^2}$$



Covarianza e varianza

- La covarianza è:

$$cov(XY) = \frac{\sum(X - \overline{X})(Y - \overline{Y})}{N}$$

- Mentre la varianza è:

$$var(X) = \frac{\sum(X - \overline{X})(X - \overline{X})}{N} = \frac{\sum(X - \overline{X})^2}{N}$$

Notate la somiglianza fra le due formule
Esistono anche le versioni con $N - 1$ (usate da SPSS)

Correlazione

- E' un indice statistico che misura l'associazione fra due variabili
- Misura come le due variabili si muovono assieme, ossia come correlano
- Viene espresso come un valore che oscilla fra -1 e 1
- Esistono diverse formule per il calcolo della correlazione
  - La correlazione lineare prodotto momento di Bravais-Pearson è la più conosciuta e si applica a due variabili quantitative
  - La correlazione di Spearman si applica quando una delle due variabili è ordinale (l'altra può essere ordinale o intervallo/a rapporto)
  - Altre correlazioni (biseriale, punto-biseriale, policorica...) si applicano con variabili ai livelli inferiori

Correlazione di Pearson

$$r = \frac{\sum z_x z_y}{N} = \frac{\frac{\sum xy}{N} - \bar{X}\bar{Y}}{s_x s_y} = \frac{cov(X,Y)}{s_x s_y} = \frac{cov(X,Y)}{\sqrt{var(X)var(Y)}}$$

$$r = \frac{\sum XY - \frac{\sum X \sum Y}{N}}{\sqrt{(\sum X^2 - \frac{(\sum X)^2}{N})(\sum Y^2 - \frac{(\sum Y)^2}{N})}}$$

$$r = \frac{N\sum XY - \sum X \sum Y}{\sqrt{[N\sum X^2 - (\sum X)^2][N\sum Y^2 - (\sum Y)^2]}}$$

Varianza/Covarianza

Possiamo costruire la matrice di varianza/covarianza e di correlazione fra queste 5 variabili

	X_1	X_2	X_3	X_4	Y
	1	2	1	1	1
	2	3	1	3	2
	4	3	2	2	3
	5	5	2	3	4
	2	4	2	2	2
$\sum$	14	17	8	11	12
$\sum^2$	50	63	14	27	34

Calcolo

Iniziamo con le varianze (N-1)

$$s^2_{xx} = \frac{\sum X^2 - \frac{\sum(X)^2}{N}}{N-1}$$

$$s^2_{x_1x_1} = \frac{50 - \frac{14^2}{5}}{5-1} = 2.70$$

Varianza/Covarianza

Varianze

	X_1	X_2	X_3	X_4	Y
X_1	2.70				
X_2		1.30			
X_3			0.30		
X_4				.70	
Y					1.30

Coprodotti

	X_1	X_2	X_3	X_4	Y
X_1					
X_2	53				
X_3	25	29			
X_4	34	40	18		
Y	41	45	21	29	

Calcolo

Ora le covarianze (N-1)

$$s^2_{xy} = \frac{\sum XY - \frac{\sum X \sum Y}{N}}{N-1}$$

$$s^2_{X_1 X_2} = \frac{53 - \frac{14 * 17}{5}}{5-1} = 1.08$$

Varianza/Covarianza

E dalle varianze/covarianze alle correlazioni

	X_1	X_2	X_3	X_4	Y
X_1	2.70				
X_2	1.35	1.30			
X_3	0.65	0.45	0.30		
X_4	0.80	0.65	.10	.70	
Y	1.85	1.05	.45	.65	1.30

	X_1	X_2	X_3	X_4	Y
X_1	1				
X_2	.72	1			
X_3	.72	.72	1		
X_4	.58	.68	.22	1	
Y	.99	.81	.72	.68	1

Calcolo

$$\frac{1.35}{\sqrt{2.7 * 1.3}} = 0.72$$

Correlazione

	X_1	X_2
X_1	1	
X_2	.72	1

Inferenza statistica

- Oltre alla statistica descrittiva, avete studiato l'inferenza statistica
- ovvero la possibilità di scoprire se un'ipotesi di lavoro è probabilisticamente accettabile
- Per l'inferenza è necessario derivare, dall'ipotesi di lavoro, due ipotesi statistiche (H_0 e H_1)
- H_0 è l'ipotesi nulla, conservativa, quella che non ipotizza nessun tipo di cambiamento ed è anche l'ipotesi che si può falsificare
- H_1 è l'ipotesi che noi facciamo e che vorremmo verificare, ma che possiamo accettare solo dimostrando che H_0 è falsa
- L'accettazione dell'ipotesi nulla (o il suo rifiuto) dipendono anche dal livello α che utilizziamo e che possiamo fissare arbitrariamente (ma in genere è fissato a .05, .01, .001)

Analisi dei dati

- L'applicazione di una tecnica statistica ad un insieme di dati (scelta perché può rispondere alla nostra ipotesi di lavoro) produce un indice statistico (ad es. il $\mathtt{t}$ o lo $\mathtt{z}$)
- A questo indice statistico è associata una probabilità (eventualmente anche in base ai gradi di libertà)
- Confrontando questa probabilità (spesso chiamata *significatività*) con il livello α da noi scelto possiamo prendere una decisione inferenziale.
- Ogni tipo di analisi dei dati ha il suo modo di impostare l'ipotesi nulla e quella alternativa; non bisogna confondere le tecniche tra di loro

Scelta inferenziale

- Dall'applicazione di una tecnica statistica di analisi dei dati si ottiene un indice statistico. La scelta inferenziale può essere fatta in due modi

Tramite il **valore critico**: usando le tabelle della statistica, troviamo il valore critico per un certo livello di α; confrontiamo quindi il valore trovato e il valore critico e, su questa base, accettiamo o rifiutiamo l'ipotesi nulla;

$$r < r_c \Rightarrow H_0$$

Tramite la **significatività** diretta: usando altri procedimenti troviamo la probabilità (asintotica o esatta) associata a quell'indice statistico. Confrontiamo quindi questa probabilità con il livello α di riferimento e, su questa base, accettiamo o rifiutiamo l'ipotesi nulla

$$P(r) > \alpha \Rightarrow H_0$$

Dimostrazione (su richiesta)

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

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

$$\frac{1}{N}\sum(X_i^2 + \bar{X}^2 - 2X_i\bar{X}) = \frac{1}{N}(\sum X_i^2 + \sum \bar{X}^2 - \sum 2X_i\bar{X}) =$$

$$\frac{1}{N}(\sum X_i^2 + N\bar{X}^2 - 2\bar{X}\sum X_i) = \frac{1}{N}\sum X_i^2 + \frac{N}{N}\bar{X}^2 - 2\bar{X}\frac{\sum X_i}{N} =$$

$$\frac{\sum X_i^2}{N} + \bar{X}^2 - 2\bar{X}\bar{X} = \frac{\sum X_i^2}{N} + \bar{X}^2 - 2\bar{X}^2 = \frac{\sum X_i^2}{N} - \bar{X}^2$$