

Progettare e condurre una ricerca educativa

Corso di
Pedagogia
Sperimentale ed
Evidence-based
education

4. Dalla costruzione della base
empirica all'interpretazione dei risultati

Corso di
Metodologia della
Ricerca
Educativa

Roberto Trinchero
roberto.trinchero@unito.it

Ricevimento e altre info:

www.edurete.org/ps1

Costruendo la base empirica abbiamo:

- prodotto una matrice dei dati, derivante dal caricamento su calcolatore (su Excel o altro foglio elettronico) dei dati dei questionari / test / check list / scale di valutazione, utilizzati per la rilevazione dei dati;
- prodotto testi derivanti dalla sbobinatura delle interviste libere o semistrutturate, dalla digitazione delle risposte aperte a prove di conoscenza/abilità e dei resoconti di osservazione (diari, diari di bordo, episodi aneddotici, ecc.).

Analisi di dati ad alta strutturazione

si serve di

Statistica Monovariata e Bivariata

può essere

per

Descrittiva o inferenziale

per

per

Spiegare gli stati
assunti da un dato
fattore sulla base di
quelli assunti da un
altro fattore

avviene mediante

Controllo della presenza di
una relazione significativa
tra due variabili

Non dovuta al caso ← ossia

Descrivere una data
realità educativa
attraverso parametri
quantitativi ricavati
dal campione

Inferire parametri
della popolazione a
partire da parametri
campione

JsStat: Pacchetto statistico on line

www.edurete.org/jsstat

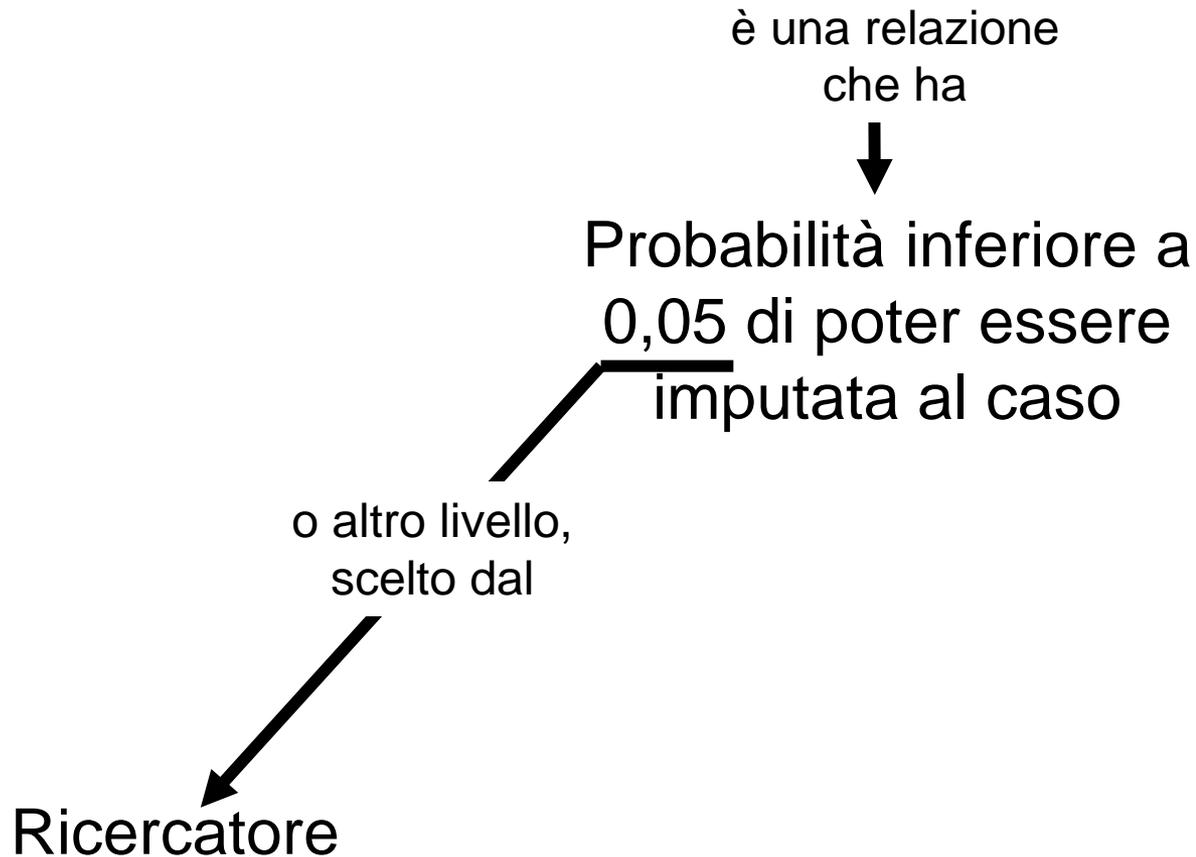
Cosa vuol dire «relazione significativa»?



Se qualcuno vi invita a giocare a testa e croce (testa perdete e croce vincete), dopo quanti lanci in cui perdetevi potreste affermare che il risultato dei lanci non è dovuto al caso (e quindi il lanciatore è un baro?)

1 lancio		Probabilità $P=1/2=0,5$	
2 lanci		$P=1/2*1/2=1/4=0,25$	<div data-bbox="1161 465 1914 679" style="border: 1px solid black; padding: 5px; display: inline-block;">0,05 → livello di fiducia (scelto dal ricercatore): «Da qui in poi non mi fido più!»</div>
3 lanci		$P=1/2*1/2*1/2=1/8=0,125$	
4 lanci		$P=1/2*1/2*1/2*1/2=1/16=0,0625$	
5 lanci		$P=$	 $=0,03125$

Cosa vuol dire «relazione significativa»?



Descrivere una realtà educativa con parametri quantitativi

significa dare



- **Distribuzione di frequenza** semplice e cumulata e relative rappresentazioni grafiche (*come* si distribuisce il campione);
- Indici di **tendenza centrale** (dicono *dove* è localizzata la distribuzione);
- Indici di **dispersione** (dicono *quanto* è ampia la distribuzione);
- Indici di **posizione di singoli soggetti** (dicono *dove* si trovano alcuni soggetti chiave rispetto al resto della distribuzione).

Distribuzione di frequenza semplice

Variabile: Genere

Modalità	Frequenza semplice	Percentuale semplice
Maschio	14	$=14/30*100$
Femmina	16	$=16/30*100$
totale	30	

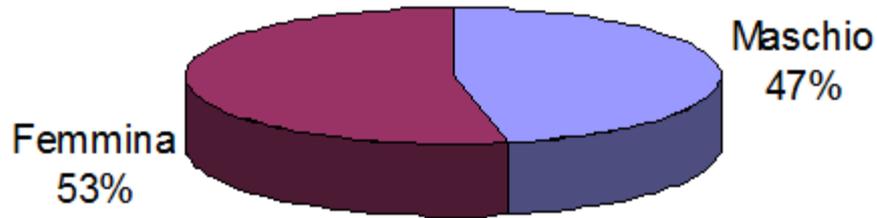


Grafico: a torta (mette in evidenza la ripartizione dei casi nelle varie modalità della variabile)

Distribuzione di frequenza semplice

Variabili: Genere 2010 e Genere 2011

Modalità	Frequenza semplice 2010	Frequenza semplice 2011
Maschio	14	11
Femmina	16	21
totale	30	32

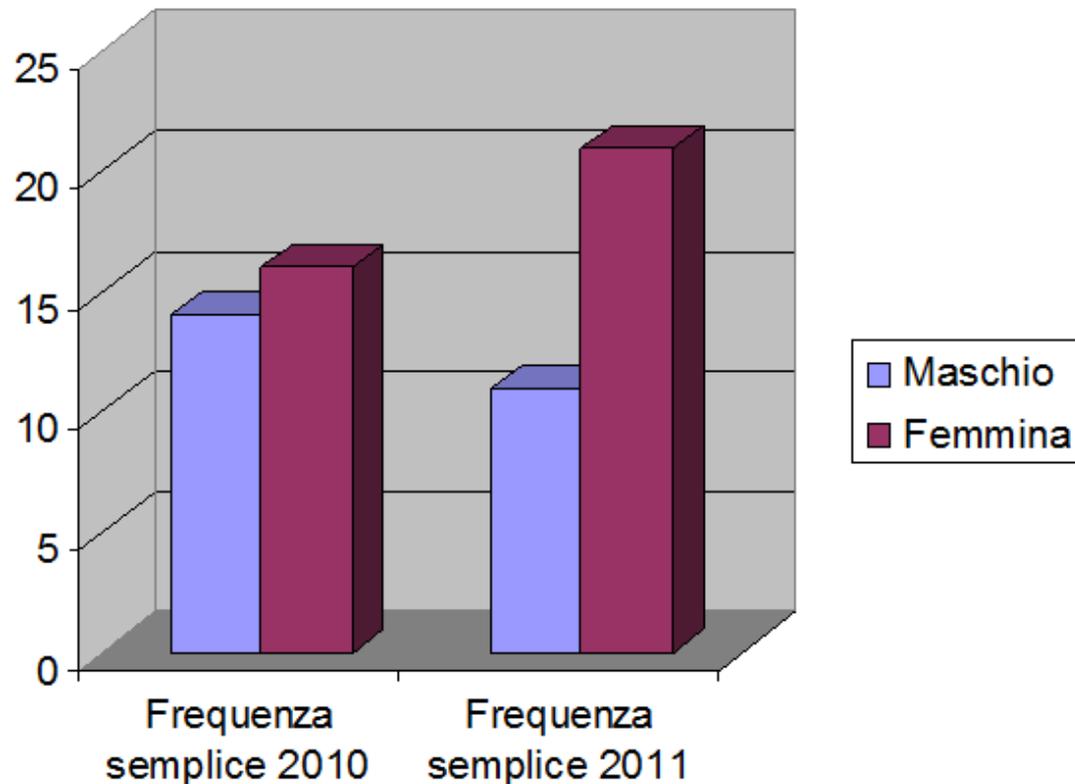


Grafico: a barre
(rende agevole
comparare dati
differenti)

Distribuzione di frequenza cumulata

Variabile: Giudizio in uscita dalla scuola media

Modalità	Freq. sempl.	Perc. Sempl.	Freq. cumulata	Perc. Cumulata
Sufficiente	6	=6/26 %	6	6/26*100
Buono	13	=13/26 %	6+13	(6+13)/26*100
Distinto	3	=3/26 %	6+13+3	(6+13+3)/26*100
Ottimo	4	=4/26 %	6+13+3+4	(6+13+3+4)/26*100
Totale	26	100	-	-

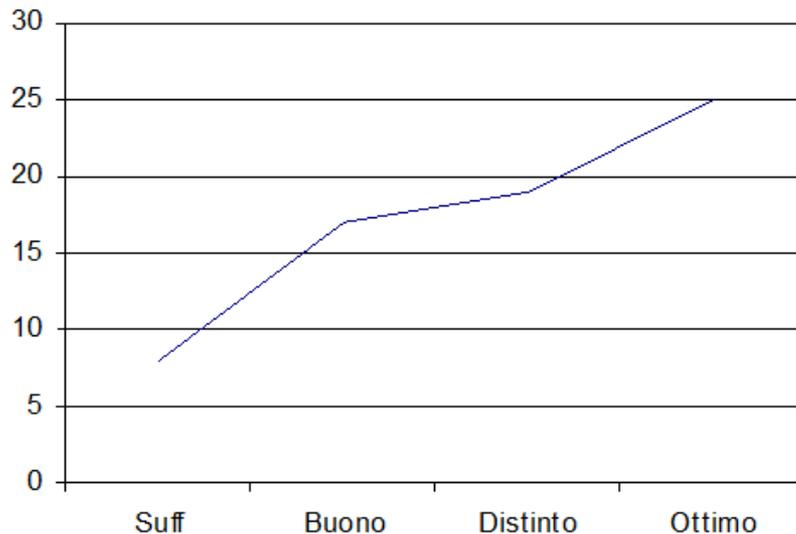


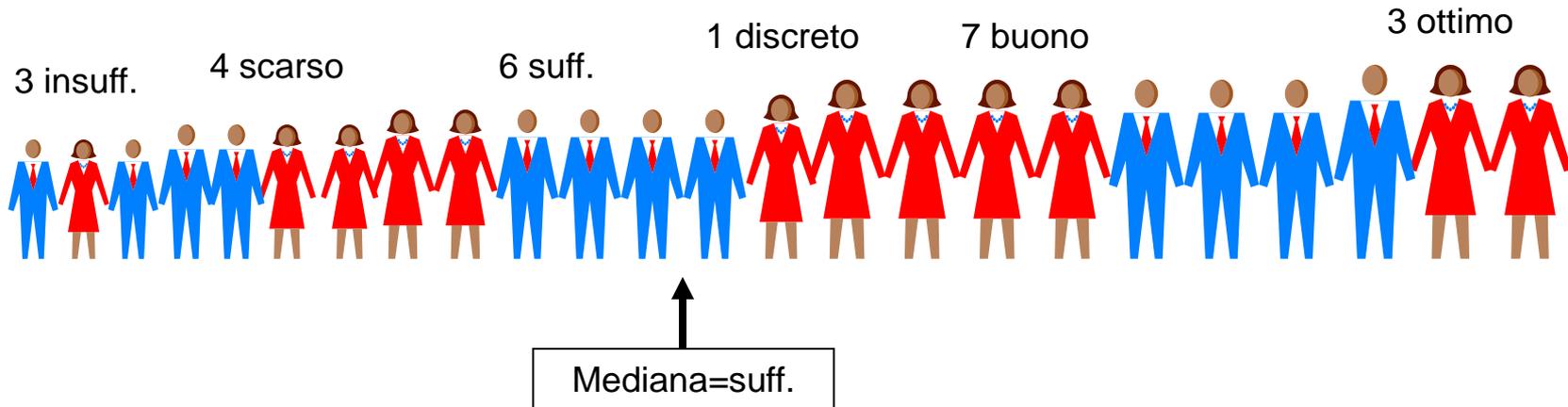
Grafico: ogiva di frequenze cumulate (mette in evidenza i punti in cui le frequenze cumulate crescono di più o di meno).

Indici di tendenza centrale

Dicono qual è il punto intorno al quale si trova la distribuzione dei dati

Moda: Categoria con la frequenza più alta.

Mediana (solo per variabili categoriali ordinate e cardinali): Punto della distribuzione ordinata dei casi che lascia alla sua sinistra e alla sua destra lo stesso numero di casi.



Media aritmetica (solo per variabili cardinali): Punto della distribuzione che azzera la somma degli scarti di ciascun dato da essa.

Es. ho preso 27 in un esame da 6 cfu, 30 in un esame da 9 cfu la media è ...

$$\text{Media} = (27 \cdot 6 + 30 \cdot 9) / (6 + 9)$$

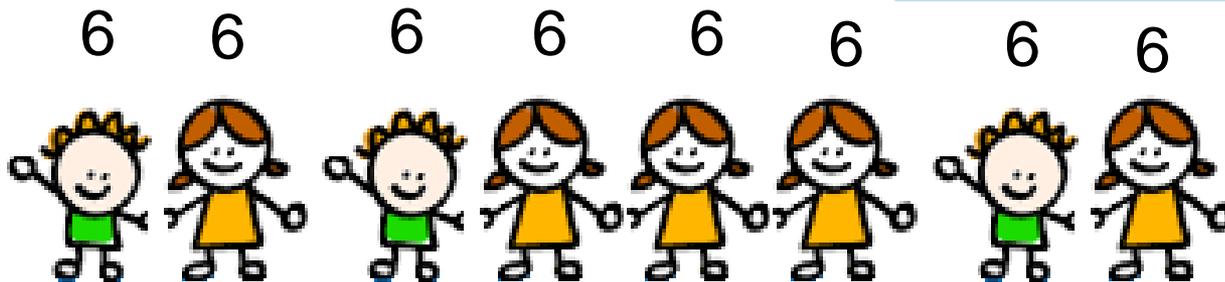
Indici di dispersione

Dicono quanto è concentrata la distribuzione intorno al suo punto centrale

LA STATISTICA
di Trilussa

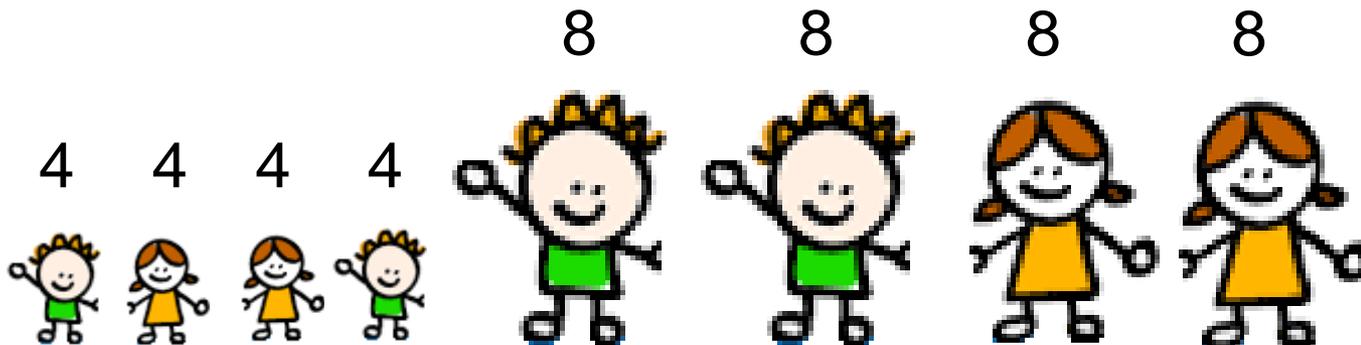
Sai ched'è la statistica? È na' cosa che serve pe fà un conto in generale de la gente che nasce, che sta male, che more, che va in carcere e che spósa. Ma pè me la statistica curiosa è dove c'entra la percentuale, pè via che, lì, la media è sempre eguale puro co' la persona bisognosa. Me spiego: da li conti che se fanno seconno le statistiche d'adesso risurta che te tocca un pollo all'anno: e, se nun entra nelle spese tue, t'entra ne la statistica lo stesso perch'è c'è un antro che ne magna due.

GA



Media=6

GB



Media=6

Indici di dispersione (variabili cardinali)

Scarto semplice = somma delle differenze di ciascun punto della distribuzione dalla media:

$$GA \rightarrow (6-6)+(6-6)+(6-6)+(6-6)+(6-6)+(6-6)+(6-6)+(6-6)=0$$

$$GB \rightarrow (6-4)+(6-4)+(6-4)+(6-4)+(6-8)+(6-8)+(6-8)+(6-8)=0$$

Devianza (Sum of Square, SS) = somma delle differenze di ciascun punto della distribuzione dalla media, elevate al quadrato:

$$GA \rightarrow (6-6)^2+(6-6)^2+(6-6)^2+(6-6)^2+(6-6)^2+(6-6)^2+(6-6)^2+(6-6)^2=0$$

$$GB \rightarrow (6-4)^2+(6-4)^2+(6-4)^2+(6-4)^2+(6-8)^2+(6-8)^2+(6-8)^2+(6-8)^2=32$$

Varianza = Devianza diviso il numero dei casi:

$$GA \rightarrow 0/8 = 0$$

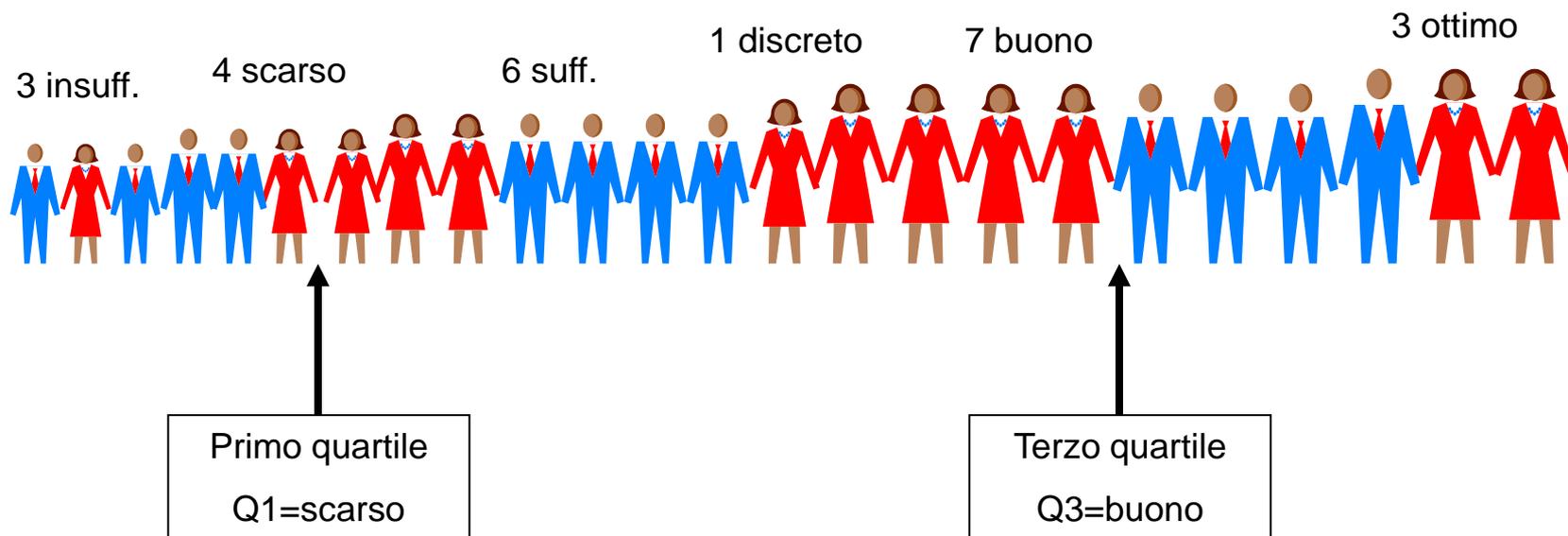
$$GB \rightarrow 32/8 = 4$$

Scarto tipo (o scarto quadratico medio, o deviazione standard) = radice quadrata della Varianza:

$$GA \rightarrow \text{radq}(0 \text{ Kg}) = 0 \text{ Kg} \rightarrow \text{se è vicino a 0 non vi è dispersione}$$

$$GB \rightarrow \text{radq}(4 \text{ Kg}^2) = 2 \text{ Kg} \rightarrow \text{se è vicino a } 1/3 \text{ della media vi è dispersione alta}$$

Indici di dispersione (variabili categoriali ordinate o cardinali)



Differenza interquartilica
= distanza tra Q3 e Q1
= $Q3 - Q1$

In questo caso «da scarso a buono»

Analisi dei dati con JsStat (www.edurete.org/jsstat)

Matrice dei dati, copiata e incollata da Excel

Selezione variabili da elaborare

Etichette variabili e modalità

Esclusione modalità indesiderate dall'elaborazione

Selezione tipo di elaborazione

Università degli studi di Torino
Dipartimento di Scienze dell'Educatione e della Formazione
JsStat - Analisi quantitativa dei dati e Data mining
di Roberto Trinchero - ver. 2.11 beta - Richiede Explorer 6.0 o successivi

Matrice dei dati da analizzare [Selezionare l'area dati in Excel, copiarla ed incollarla in questa casella. La prima riga deve contenere i nomi delle variabili, la prima colonna i codici dei soggetti. Eventuali celle vuote devono essere sostituite con un trattino.]

codice	età	genere	provnas	vivigen	annoisc	matlacun	matprep	s_esempi	s_schmi	s_ripesto
a01	28	m	to	1	5	matematica	italiano	3	3	3
a02	22	m	al	2	2	-	-	-	-	-
a03	23	f	to	2	2	matematica	storia	-	2	-

Etichette [E' possibile omettere...]
Codice Soggetto

Variabile indipendente
-

Variabile dipendente
-

Tipo di elaborazione
Analisi monovariata

Escludi i casi con valori: (es. 2 > 5 o < 3)
Raggruppa i casi con valori: (es. 2 3 4 o 3-6)

Escludi i casi con valori: (es. 2 > 5 o < 3)
Raggruppa i casi con valori: (es. 2 3 4 o 3-6)

Analisi dei dati introdotti Cancellala

Variabile moderatrice

Illustra processo (N<100 casi)

Illustrazione passo-passo del processo di elaborazione

Esempi di matrici dati: [Questionario Test Data Mining](#)

Elaborazioni disponibili:

Analisi monovariata. [Distribuzione di frequenza, indici di tendenza centrale, di dispersione, di posizione.](#)

Punteggi standardizzati. [Punteggi z e le loro ricodifiche.](#)

Analisi bivariata. Controllo delle ipotesi bivariate:

Variabile dipendente-> Variabile indipendente	Categoriale non ordinata	Categoriale ordinata	Cardinale
Categoriale non ordinata	Tabella a doppia entrata	Analisi della varianza non parametrica	Analisi della varianza
Categoriale ordinata		Cograduazione	Cograduazione Analisi della varianza
Cardinale			Correlazione

Creazione gruppi di modalità

Esempi di elaborazioni

Analisi multivariata. [Estrazione regolarità con procedura Rough data set.](#)

Analisi monovariata con JsStat: distribuzione di frequenza

Come si distribuiscono i casi del campione nelle varie categorie della variabile?

Variabile indipendente Quanti anni hai?	Variabile dipendente -	Tipo di elaborazione Analisi monovariata	
Escludi i casi con valori: (es. 2 o >5 o <3) [] [] [] []	Escludi i casi con valori: (es. 2 o >5 o <3) [] [] [] []	Analisi dei dati introdotti	Cancel
Raggruppa i casi con valori: (es. 2 3 4 o 3-6) [] [] [] []	Raggruppa i casi con valori: (es. 2 3 4 o 3-6) [] [] [] []	Variabile moderatrice -	<input checked="" type="checkbox"/> Illustra prc (N<100 casi)

Distribuzione di frequenza:
Quanti anni hai?

Modalità	Frequenza semplice	Percent. semplice	Frequenza cumulata	Percent. cumulata	Int. Fid. 95%
19	1	3%	1	3%	0%:12%
20	7	21%	8	24%	7%:35%
21	8	24%	16	48%	10%:39%
22	5	15%	21	64%	3%:27%
23	4	12%	25	76%	1%:23%
24	2	6%	27	82%	0%:14%
25	3	9%	30	91%	0%:19%
28	3	9%	33	100%	0%:19%

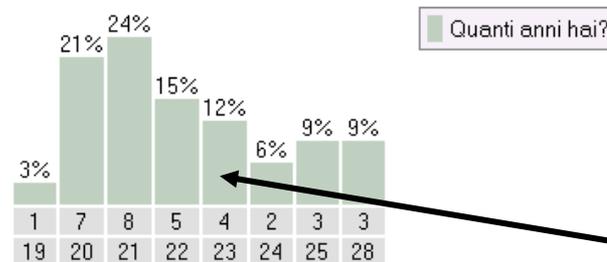


Grafico a barre

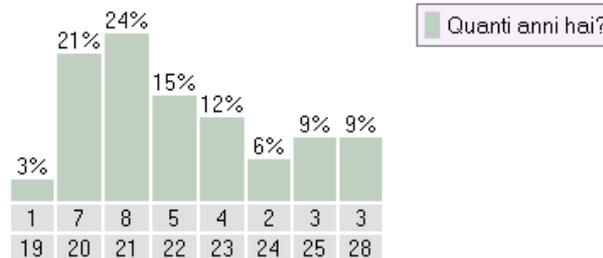
Proiezione sulla popolazione: Intervallo di fiducia del 95% per la proporzione dei casi nella categoria

Distribuzione di frequenza semplice e cumulata, con percentuali

Qual è il punto centrale intorno al quale si distribuiscono i valori della variabile?

Distribuzione di frequenza:
Quanti anni hai?

Modalità	Frequenza semplice	Percent. semplice	Frequenza cumulata	Percent. cumulata	Int. Fid. 95%
19	1	3%	1	3%	0%:12%
20	7	21%	8	24%	7%:35%
21	8	24%	16	48%	10%:39%
22	5	15%	21	64%	3%:27%
23	4	12%	25	76%	1%:23%
24	2	6%	27	82%	0%:14%
25	3	9%	30	91%	0%:19%
28	3	9%	33	100%	0%:19%



Campione:

Numero di casi = 33

Indici di tendenza centrale:

Moda = 21
Mediana = 22
Media = 22.3

Indici di dispersione:

Squilibrio = 0.16
Campo di variazione = 9
Differenza interquartilica = 2
Scarto tipo = 2.39

Popolazione:

Parametro	Int. Fid. 95%
Media	da 21.49 a 23.12
Scarto tipo	da 1.92 a 3.17

Tendenza centrale:

Moda = Categoria con la frequenza più alta (per variabili categoriali e cardinali)

Mediana = Punto centrale della distribuzione ordinata dei casi: lascia il 50% dei casi sotto e il 50% sopra (solo per variabili categoriali ordinate e cardinali)

Media aritmetica = Somma dei valori di tutti i casi diviso il numero dei casi (solo per variabili cardinali)

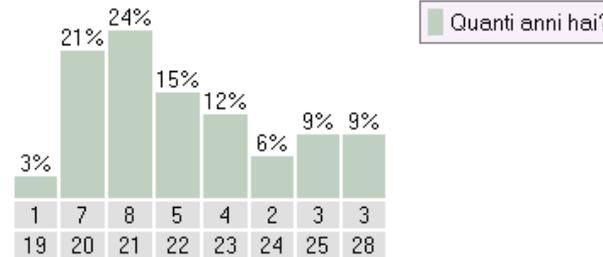
$$\frac{19 + 20 + 20 + 20 + 20 + 20 + 20 + 20 + 21 + 21 + 21 + 21 + 21 + 21 + 21 + 21 + 22 + 22 + 22 + 22 + 22 + 23 + 23 + 23 + 23 + 23 + 24 + 24 + 25 + 25 + 25 + 28 + 28 + 28}{33} = 22.3$$

Analisi monovariata con JsStat: dispersione (variabili categoriali)

I valori della variabile sono concentrati intorno al punto centrale della distribuzione o ne sono lontani?

Distribuzione di frequenza:
Quanti anni hai?

Modalità	Frequenza semplice	Percent. semplice	Frequenza cumulata	Percent. cumulata	Int. Fid. 95%
19	1	3%	1	3%	0%:12%
20	7	21%	8	24%	7%:35%
21	8	24%	16	48%	10%:39%
22	5	15%	21	64%	3%:27%
23	4	12%	25	76%	1%:23%
24	2	6%	27	82%	0%:14%
25	3	9%	30	91%	0%:19%
28	3	9%	33	100%	0%:19%



$$\sum_{j=1}^k p_j^2$$

Campione:

Numero di casi= 33

Indici di tendenza centrale:

Moda = 21

Mediana = 22

Media = 22.3

Indici di dispersione:

Squilibrio = 0.16

Campo di variazione = 9

Differenza interquartilica = 2

Scarto tipo = 2.39

Popolazione:

Parametro	Int. Fid. 95%
Media	da 21.49 a 23.12
Scarto tipo	da 1.92 a 3.17

Dispersione:

Squilibrio = Somma delle proporzioni al quadrato per ciascuna modalità della variabile (per variabili categoriali e cardinali)

$$0.03^2 + 0.21^2 + 0.24^2 + 0.15^2 + 0.12^2 + 0.06^2 + 0.09^2 + 0.09^2 = 0.16$$

Minimo: $1/k$ → Tutte le k categorie hanno la stessa frequenza

Massimo: 1 → Una sola delle k categorie contiene il totale dei casi

Analisi monovariata con JsStat: dispersione (variabili categoriali ordinate e cardinali)

I valori della variabile sono concentrati intorno al punto centrale della distribuzione o ne sono lontani?

Distribuzione di frequenza:
Quanti anni hai?

Modalità	Frequenza semplice	Percent. semplice	Frequenza cumulata	Percent. cumulata	Int. Fid. 95%
19	1	3%	1	3%	0%:12%
20	7	21%	8	24%	7%:35%
21	8	24%	16	48%	10%:39%
22	5	15%	21	64%	3%:27%
23	4	12%	25	76%	1%:23%
24	2	6%	27	82%	0%:14%
25	3	9%	30	91%	0%:19%
28	3	9%	33	100%	0%:19%

Campione:

Numero di casi = 33

Indici di tendenza centrale:

Moda = 21

Mediana = 22

Media = 22.3

Indici di dispersione:

Squilibrio = 0.16

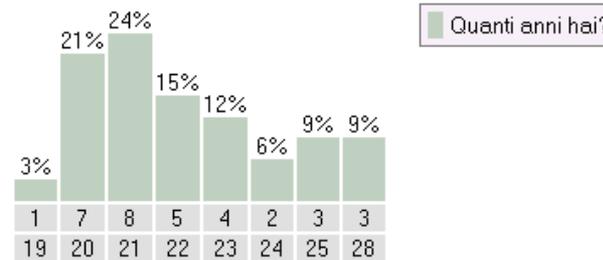
Campo di variazione = 9

Differenza interquartile = 2

Scarto tipo = 2.39

Popolazione:

Parametro	Int. Fid. 95%
Media	da 21.49 a 23.12
Scarto tipo	da 1.92 a 3.17



Dispersione:

Campo di variazione = Distanza tra **valore minimo** e **valore massimo**
(solo per variabili categoriali ordinate e cardinali)

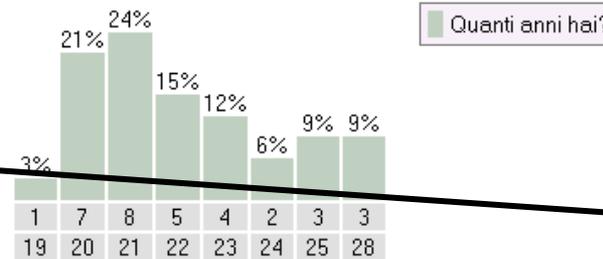
La distanza tra 19 e 28 vale: "da 19 a 28" ossia 9

Analisi monovariata con JsStat: dispersione (variabili categoriali ordinate e cardinali)

I valori della variabile sono concentrati intorno al punto centrale della distribuzione o ne sono lontani?

Distribuzione di frequenza:
Quanti anni hai?

Modalità	Frequenza semplice	Percent. semplice	Frequenza cumulata	Percent. cumulata	Int. Fid. 95%
19	1	3%	1	3%	0%:12%
20	7	21%	8	24%	7%:35%
21	8	24%	16	48%	10%:39%
22	5	15%	21	64%	3%:27%
23	4	12%	25	76%	1%:23%
24	2	6%	27	82%	0%:14%
25	3	9%	30	91%	0%:19%
28	3	9%	33	100%	0%:19%



Campione:

Numero di casi = 33
Indici di tendenza centrale:
Moda = 21
Mediana = 22
Media = 22.3
Indici di dispersione:
Squilibrio = 0.16
Campo di variazione = 9

Differenza interquartile = 2
Scarto tipo = 2.39

Popolazione:

Parametro	Int. Fid. 95%
Media	da 21.49 a 23.12
Scarto tipo	da 1.92 a 3.17

Dispersione:

Differenza interquartile = Distanza tra valore posizionato sul 75% della distribuzione e valore posizionato sul 25% (solo per variabili categoriali ordinate e cardinali)

La distanza tra 23 e 21 vale 2.

Terzo quartile

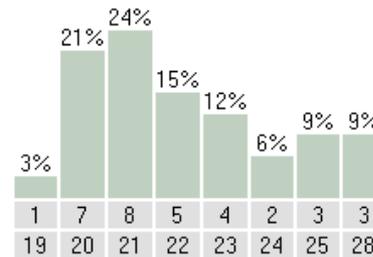
Primo quartile

Analisi monovariata con JsStat: dispersione (variabili cardinali)

I valori della variabile sono concentrati intorno al punto centrale della distribuzione o ne sono lontani?

Distribuzione di frequenza:
Quanti anni hai?

Modalità	Frequenza semplice	Percent. semplice	Frequenza cumulata	Percent. cumulata	Int. Fid. 95%
19	1	3%	1	3%	0%:12%
20	7	21%	8	24%	7%:35%
21	8	24%	16	48%	10%:39%
22	5	15%	21	64%	3%:27%
23	4	12%	25	76%	1%:23%
24	2	6%	27	82%	0%:14%
25	3	9%	30	91%	0%:19%
28	3	9%	33	100%	0%:19%



Quanti anni hai?

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$$

Dispersione:

Scarto tipo = Radice della somma delle differenze di ciascun valore rispetto alla media elevate al quadrato e rapportate al numero dei casi (solo per variabili cardinali)

$$\text{radq}(((19-22.3)^2+(20-22.3)^2+(20-22.3)^2+(20-22.3)^2+(20-22.3)^2+(20-22.3)^2+(20-22.3)^2+(20-22.3)^2+(21-22.3)^2+(21-22.3)^2+(21-22.3)^2+(21-22.3)^2+(21-22.3)^2+(21-22.3)^2+(21-22.3)^2+(21-22.3)^2+(21-22.3)^2+(22-22.3)^2+(22-22.3)^2+(22-22.3)^2+(22-22.3)^2+(22-22.3)^2+(22-22.3)^2+(23-22.3)^2+(23-22.3)^2+(23-22.3)^2+(23-22.3)^2+(23-22.3)^2+(24-22.3)^2+(24-22.3)^2+(25-22.3)^2+(25-22.3)^2+(25-22.3)^2+(25-22.3)^2+(28-22.3)^2+(28-22.3)^2+(28-22.3)^2)/33) = 2.39$$

Minimo: 0 → Tutti i valori della distribuzione coincidono con la media

Quando è vicino ad 1/3 della media → I valori della distribuzione sono distanti dalla media → La dispersione è alta

Campione:

Numero di casi= 33
Indici di tendenza centrale:
Moda = 21
Mediana = 22
Media = 22.3
Indici di dispersione:
Squilibrio = 0.16
Campo di variazione = 9
Differenza interquartile = 2
Scarto tipo = 2.39

Popolazione:

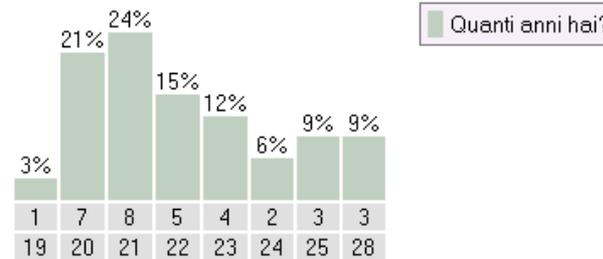
Parametro	Int. Fid. 95%
Media	da 21.49 a 23.12
Scarto tipo	da 1.92 a 3.17

Analisi monovariata con JsStat: intervalli di fiducia per la media e lo scarto tipo

Se il campione è estratto con campionamento casuale, quanto valgono media e scarto tipo nella popolazione?

Distribuzione di frequenza:
Quanti anni hai?

Modalità	Frequenza semplice	Percent. semplice	Frequenza cumulata	Percent. cumulata	Int. Fid. 95%
19	1	3%	1	3%	0%:12%
20	7	21%	8	24%	7%:35%
21	8	24%	16	48%	10%:39%
22	5	15%	21	64%	3%:27%
23	4	12%	25	76%	1%:23%
24	2	6%	27	82%	0%:14%
25	3	9%	30	91%	0%:19%
28	3	9%	33	100%	0%:19%



Campione:

Numero di casi= 33

Indici di tendenza centrale:

Moda = 21

Mediana = 22

Media = 22.3

Indici di dispersione:

Squilibrio = 0.16

Campo di variazione = 9

Differenza interquartilica = 2

Scarto tipo = 2.39

Popolazione:

Parametro	Int. Fid. 95%
Media	da 21.49 a 23.12
Scarto tipo	da 1.92 a 3.17

Media della popolazione: nel 95% percento dei campioni estratti con campionamento casuale dalla popolazione, sarà compresa tra 21,49 e 23,12

Scarto tipo della popolazione: nel 95% percento dei campioni estratti con campionamento casuale dalla popolazione, sarà compreso tra 1,92 e 3,17

Indici di posizione: i punteggi standardizzati

E' possibile definire la posizione di un allievo all'interno della sua classe in termini di "quanti scarti tipo sopra o sotto la media della classe" l'allievo si trova. L'indice ricavato in tal modo si chiama "punteggio z" o "punteggio standardizzato".

Gli studenti con voto 5 in questo esempio hanno un punteggio z di:

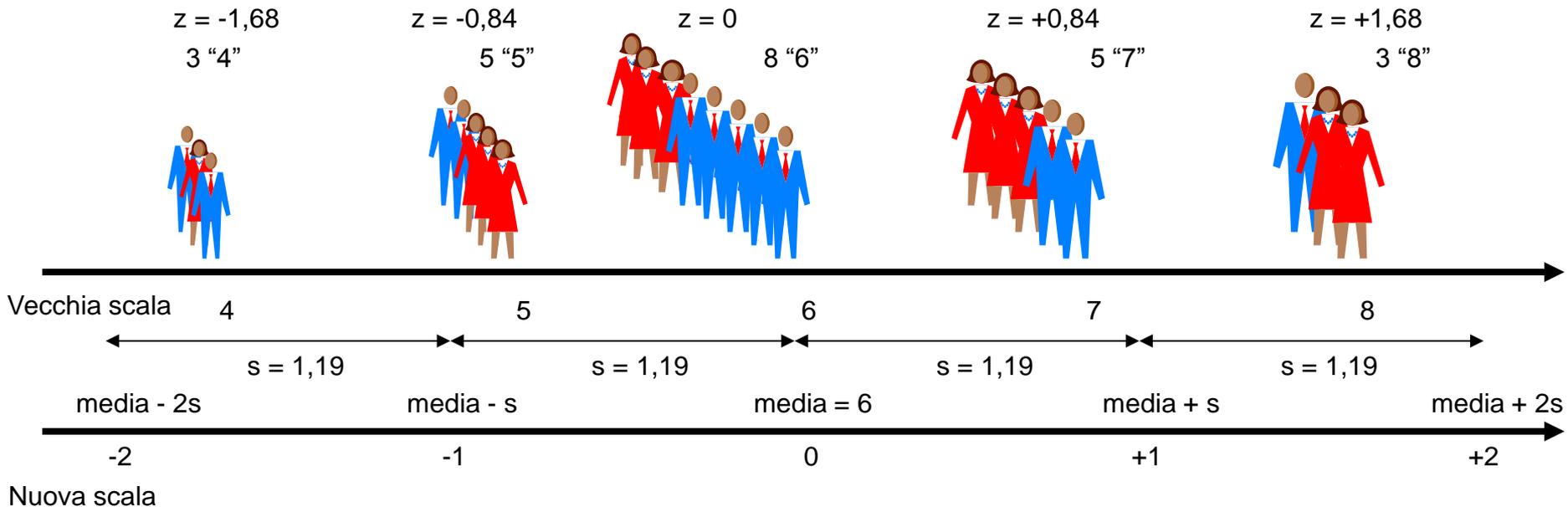
$$Z = (5-6)/1,19 = -0,84$$

$$z = \frac{X - \bar{X}}{S}$$

Essi si trovano a 0,84 scarti tipo al di sotto della media.

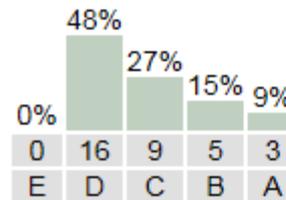
Se z è positivo il voto "grezzo" (non standardizzato) dello studente è superiore alla media, viceversa per z negativo. Se z è vicino allo zero il voto "grezzo" dello studente è vicino alla media della classe.

Voto	Frequenza semplice
4	3
5	5
6	8
7	5
8	3



Indici di posizione con JsStat → Comparazione tra casi (punteggi z)

Soggetto	Punti grezzi	Punto z	Punto C di Guilford	Punto pentenaria	Punto T	Punto L	Centile
a01	28	2.38	10	A	73.8	9.6	97
a29	28	2.38	10	A	73.8	9.6	94
a33	28	2.38	10	A	73.8	9.6	91
a11	25	1.13	7	B	61.3	7.7	88
a21	25	1.13	7	B	61.3	7.7	85
a28	25	1.13	7	B	61.3	7.7	82
a07	24	0.71	6	B	57.1	7.1	79
a27	24	0.71	6	B	57.1	7.1	76
a03	23	0.29	6	C	52.9	6.4	73
a09	23	0.29	6	C	52.9	6.4	70
a10	23	0.29	6	C	52.9	6.4	67
a20	23	0.29	6	C	52.9	6.4	64
a02	22	-0.13	5	C	48.7	5.8	61
a08	22	-0.13	5	C	48.7	5.8	58
a22	22	-0.13	5	C	48.7	5.8	55
a24	22	-0.13	5	C	48.7	5.8	52
a31	22	-0.13	5	C	48.7	5.8	48
a06	21	-0.54	4	D	44.6	5.2	45
a12	21	-0.54	4	D	44.6	5.2	42
a13	21	-0.54	4	D	44.6	5.2	39
a18	21	-0.54	4	D	44.6	5.2	36
a23	21	-0.54	4	D	44.6	5.2	33
a26	21	-0.54	4	D	44.6	5.2	30
a30	21	-0.54	4	D	44.6	5.2	27
a32	21	-0.54	4	D	44.6	5.2	24
a04	20	-0.96	3	D	40.4	4.6	21
a05	20	-0.96	3	D	40.4	4.6	18
a14	20	-0.96	3	D	40.4	4.6	15
a15	20	-0.96	3	D	40.4	4.6	12
a16	20	-0.96	3	D	40.4	4.6	9
a17	20	-0.96	3	D	40.4	4.6	6
a25	20	-0.96	3	D	40.4	4.6	3
a19	19	-1.38	2	D	36.2	3.9	0



■ Quanti anni hai? (pentenaria)

Campione:

Numero di casi= 33

Indici di tendenza centrale:

Moda = 21

Mediana = 22

Media = 22.3

Indici di dispersione:

Squilibrio = 0

Campo di variazione = 9

Differenza interquartilica = 2

Scarto tipo = 2.39

$Z = (\text{punteggio del soggetto} - \text{media del gruppo}) / \text{scarto tipo del gruppo}$

Centile = Posizione ordinale del soggetto su una scala da 0 a 100

Punto standardizzato	Punto pentenaria
da -2,5 a -1,5	A
da -1,5 a -0,5	B
da -0,5 a 0,5	C
da 0,5 a 1,5	D
da 1,5 a 2,5	E

Punto standardizzato	Punto C di Guilford
da -2,75 a -2,25	0
da -2,25 a -1,75	1
da -1,75 a -1,25	2
da -1,25 a -0,75	3
da -0,75 a -0,25	4
da -0,25 a 0,25	5
da 0,25 a 0,75	6
da 0,75 a 1,25	7
da 1,25 a 1,75	8
da 1,75 a 2,25	9
da 2,25 a 2,75	10

$$T = (z * 10) + 50$$

$$L = (z * 1,5) + 6$$

	Tendenza centrale	Dispersione	Inferenza	Posizione dei soggetti
Variabili categoriali non ordinate	Moda: categoria con la frequenza più alta	Squilibrio: somma delle proporzioni elevate al quadrato	Intervallo di fiducia per la proporzione	Collocazione del soggetto nella categoria
Variabili categoriali ordinate	Idem + Mediana: punto centrale della distribuzione ordinata	Idem + Campo di variazione: distanza tra il valor minimo e il valor massimo + Differenza interquartilica: distanza tra il punto al 25% della distribuzione e il punto al 75% della distribuzione	Idem	Centile del soggetto
Variabili cardinali	Idem + Media aritmetica: somma di tutti i punti divisa per il numero dei casi	Idem + Scarto tipo: radice della somma delle differenze di ciascun valore rispetto alla media elevate al quadrato e rapportate al numero dei casi	Idem + Intervallo di fiducia per la media e per lo scarto tipo	Punto z del soggetto e sue ricodifiche

Le variabili quasi cardinali si possono considerare cardinali se è possibile supporre che la distanza percepita tra i punti della scala sia la stessa per tutti i soggetti, ordinali altrimenti

Spiegare gli stati assunti da un dato fattore sulla base di quelli assunti da un altro fattore

significa



Controllare la presenza di relazioni significative tra **ciascuna** variabile generata dal fattore indipendente e **ciascuna** variabile generata dal fattore dipendente

ognuna di esse è



Una **conferma** (se la relazione va nello stesso verso) o una **confutazione** (se la relazione va nel verso opposto) dell'ipotesi di partenza

es.



Cresce l'indicatore di profitto scolastico al calare dell'indicatore di motivazione allo studio

usando le tecniche



- Tabella a doppia entrata (X quadro)
- Analisi della varianza (eta quadro)
- Correlazione (r)

Analisi bivariata: tecniche fondamentali

Variabile dipendente		Categoriale	Cardinale
Variabile indipendente		Categoriale	Cardinale
Categoriale		Tabella a doppia entrata (X quadro)	Analisi della varianza (eta quadro)
Cardinale		Invertire le variabili e applicare l'Analisi della varianza (eta quadro)	Correlazione (coefficiente r)

Tabella a doppia entrata

O = frequenze osservate sul campione

A = frequenze attese in caso di assenza di relazione

	Basso profitto scolastico	Alto profitto scolastico	<i>Marginali di riga</i>
Bassa motivazione allo studio	O=6 A=11*14/24=6,4	O=8 A=13*14/24=7,6	14
Alta motivazione allo studio	O=5 A=11*10/24=4,6	O=5 A=13*10/24=5,4	10
<i>Marginali di colonna</i>	11	13	24

Le frequenze attese A derivano da una semplice proporzione: «in caso di assenza di relazione la stessa proporzione di allievi con basso profitto scolastico presente nella popolazione deve esserci anche nel sottocampione di allievi con bassa motivazione allo studio»: A sta a 14 come 11 sta a 24, quindi $A : 14 = 11 : 24$, quindi $A = 11 \cdot 14 / 24$

Se imponendo un valore ad una cella gli altri sono automaticamente determinati dal vincolo di rispettare i marginali, si dice che la tabella ha 1 solo grado di libertà (k).

In generale $k = (\text{numero di righe} - 1) \times (\text{numero di colonne} - 1)$

Indici di forza della relazione tra variabili categoriali

X = somma delle differenze tra le frequenze osservate e le frequenze attese di ciascuna cella:

$$= (6-6,4)+(8-7,6)+(5-4,6)+(5-5,4) = 0$$

X quadro = somma delle differenze tra le frequenze osservate e le frequenze attese di ciascuna cella, elevate al quadrato (per evitare i segni negativi) e rapportate alle frequenze attese (per tenere conto del «peso» della differenza):

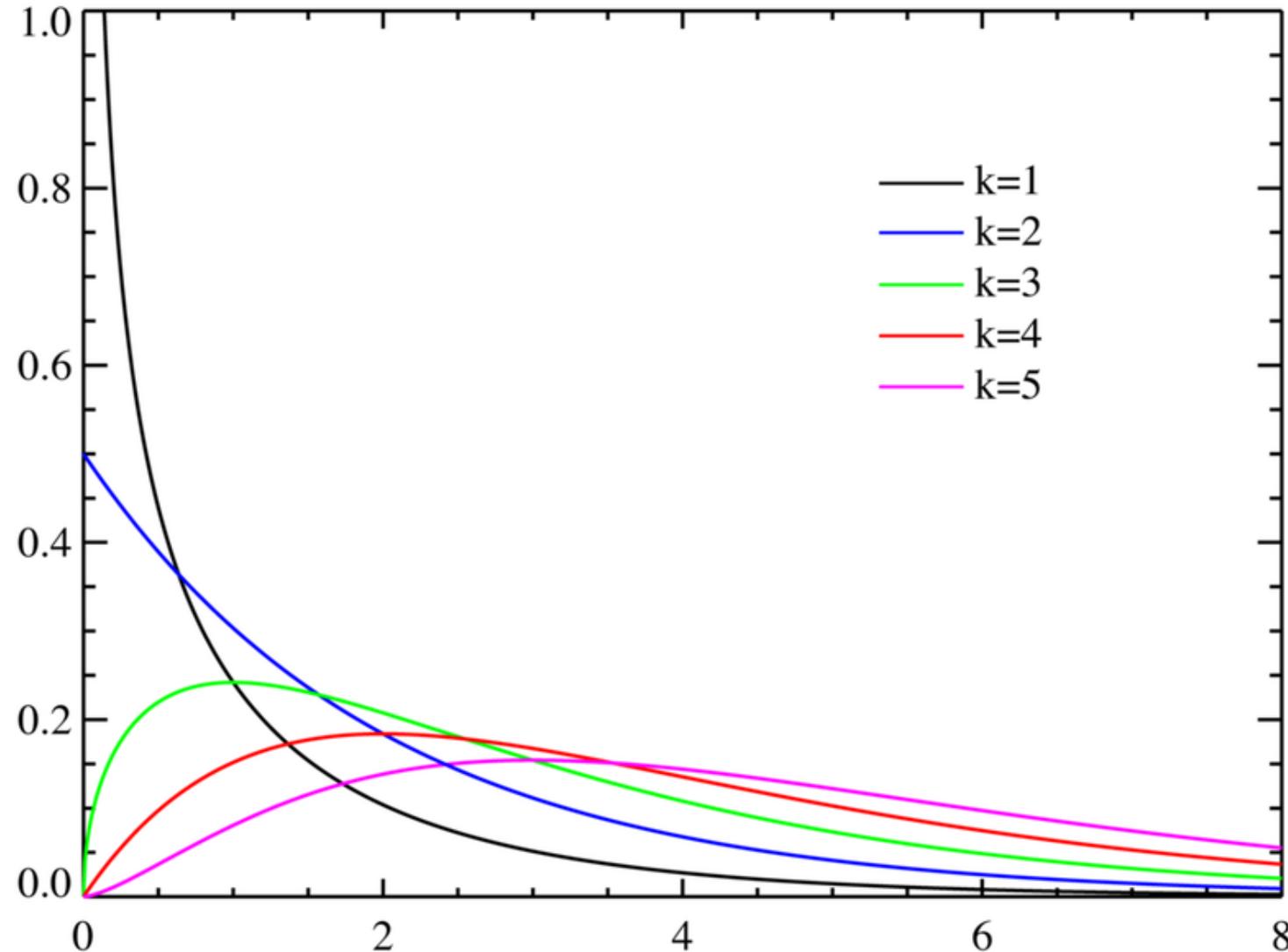
$$= (6-6,4)^2/6,4+(8-7,6)^2/7,6+(5-4,6)^2/4,6+(5-5,4)^2/5,4 = 0,11$$

Se X quadro è:

vicino a 0 → non vi è differenza tra frequenze osservate e attese → non vi è relazione
 vicino a N (numero dei casi) → vi è un'alta differenza tra frequenze osservate e attese
 → vi è una fortissima relazione

La relazione inizia ad essere forte quando X quadro è vicino ad $1/3$ del numero dei casi. 28

Significatività dell'X quadro



E' la probabilità che X quadro sia diverso da zero per effetto del caso. Corrisponde all'area della distribuzione di probabilità Chi-quadro con k gradi di libertà per l'ascissa X. Si calcola con Excel con `DISTRIB.CHI(Xquadro, k)`

Residui standardizzati di cella (rs)

O = frequenze osservate sul campione

A = frequenze attese in caso di assenza di relazione

	Basso profitto scolastico	Alto profitto scolastico	<i>Marginali di riga</i>
Bassa motivazione allo studio	<p>O=6 A=6,4 $rs=(6-6,4)/\text{radq}(6,4)=-0,16$</p>	<p>O=8 A=7,6 $rs=(8-7,6)/\text{radq}(7,6)=0,15$</p>	14
Alta motivazione allo studio	<p>O=5 A=4,6 $rs=(5-4,6)/\text{radq}(4,6)=0,19$</p>	<p>O=5 A=5,4 $rs=(5-5,4)/\text{radq}(5,4)=-0,17$</p>	10
<i>Marginali di colonna</i>	11	13	24

I residui standardizzati di cella si calcolano con la formula $(O-A)/\text{radq}(A)$ e si leggono come dei **punteggi z**.

Se sono **positivi**, le frequenze osservate sono superiori alle frequenze attese, quindi vi è **attrazione** tra le modalità corrispondenti delle due variabili.

Se sono **negativi**, le frequenze osservate sono inferiori alle frequenze attese, quindi vi è **repulsione** tra le modalità corrispondenti delle due variabili.

Se sono vicini a zero l'attrazione/repulsione non è significativa. Se sono **superiori a 1,96** in modulo l'attrazione/repulsione è **significativa** a livello di fiducia 0,05.

Analisi bivariata con JsStat: relazione tra variabili categoriali non ordinate

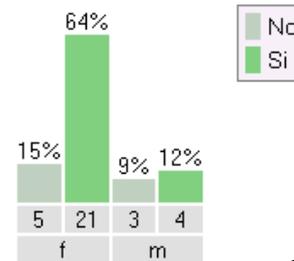
L'ipotesi "Esiste relazione tra il genere dei soggetti e il fatto di abitare con i genitori." è confermata dai dati?

Tabella a doppia entrata:

Maschio o femmina? x Vivi con i genitori?				
Vivi con i genitori?-> Maschio o femmina?		No	Si	Marginale di riga
f	5	21	26	
	6.3	19.7		
	-0.5	0.3		
m	3	4	7	
	1.7	5.3		
	1	-0.6		
Marginale di colonna	8	25	33	

Frequenza **osservata** nel campione sotto esame

Frequenza **attesa** se tra le due variabili non vi fosse relazione



Più è alta la differenza, più è probabile che esista una relazione significativa tra le variabili

χ^2 quadro = 100. Significatività = 0.2
V di Cramer = 0.23

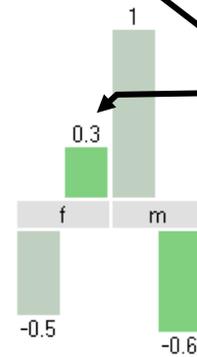
Probabilità esatta (dal test di Fisher) = 0.17

Nelle celle della tabella sono indicati:

- la frequenza osservata O
- la frequenza attesa A
- il residuo standardizzato di cella, ossia lo scarto tra frequenza osservata e attesa rapportato alla radice quadrata della frequenza attesa $(O-A)/\sqrt{A}$

Significatività: Se è inferiore a 0,05 si opta per la presenza di una relazione significativa

Residuo standardizzato di cella: Se è superiore a **+1,96** si opta per l'esistenza di un'**attrazione** significativa tra le corrispondenti modalità delle due variabili. Se è inferiore a **-1,96**, si opta per una **repulsione** significativa.

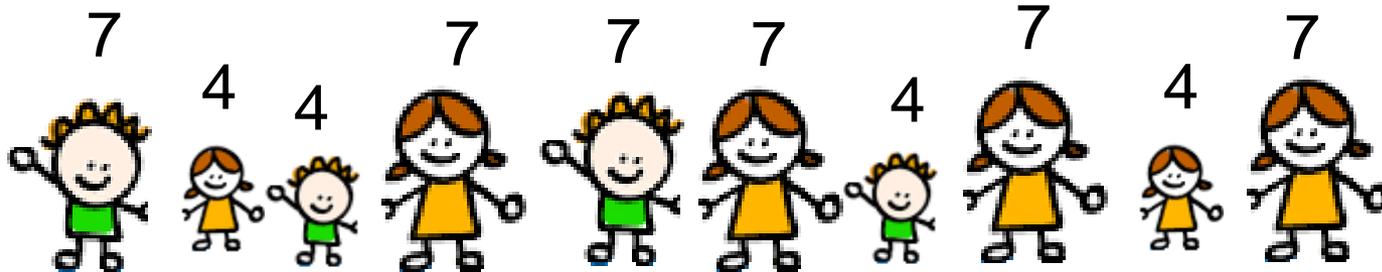


V di Cramer: Se è vicino a 0 → relazione debole, se è vicino a 1 → relazione forte

Analisi della varianza

E' una procedura che serve per determinare l'esistenza di differenze significative tra le medie di due o più gruppi, che equivale a dire l'esistenza di una relazione significativa tra appartenenza ad uno dei due gruppi e valore associato al soggetto

Ho 10 bambini da dividere in due gruppi



Media generale = 5,8
SS generale = $6 \cdot (7 - 5,8)^2 + 4 \cdot (4 - 5,8)^2 = 21,6$

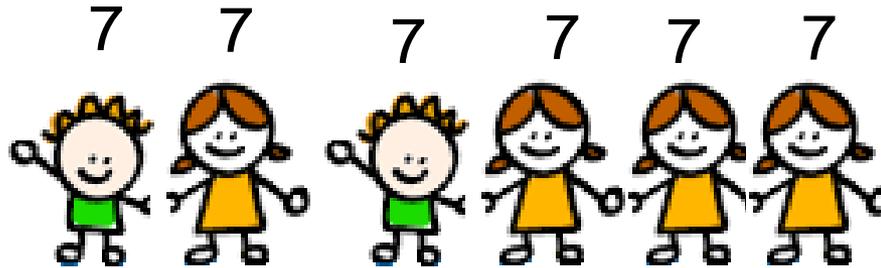
Gruppo A

Gruppo B

Analisi della varianza

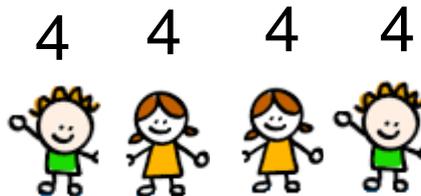
Caso 1

Gruppo A



Media gruppo A = 7
SS gruppo A = 0

Gruppo B



Media gruppo B = 4
SS gruppo B = 0

Media generale = 5,8
SS generale = $6 \cdot (7 - 5,8)^2 + 4 \cdot (4 - 5,8)^2 = 21,6$

Non vi sono differenze all'interno dei gruppi (*within groups*) e vi sono differenze rilevanti tra un gruppo e l'altro (*between groups*).

In questo caso è molto probabile che vi sia una differenza significativa tra i due gruppi (le medie sono molto diverse), ossia che vi sia una relazione tra appartenenza al gruppo e valore indicato.

Analisi della varianza

Caso 2

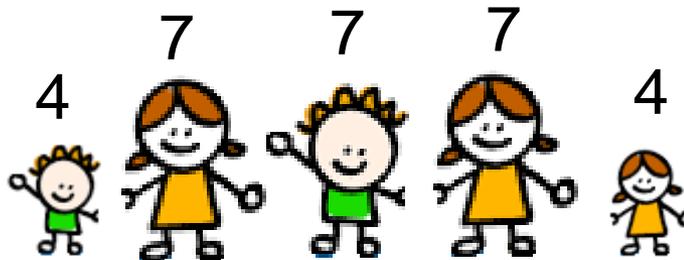
Gruppo A



Media gruppo A = 5,8

$$SS \text{ gruppo A} = 2 \cdot (4 - 5,8)^2 + 3 \cdot (7 - 5,8)^2 = 10,8$$

Gruppo B



Media gruppo B = 5,8

$$SS \text{ gruppo B} = 2 \cdot (4 - 5,8)^2 + 3 \cdot (7 - 5,8)^2 = 10,8$$

Media generale = 5,8

$$SS \text{ generale} = 6 \cdot (7 - 5,8)^2 + 4 \cdot (4 - 5,8)^2 = 21,6$$

Non vi sono differenze tra un gruppo e l'altro (*between groups*) e vi sono differenze rilevanti all'interno dei gruppi (*within groups*).

In questo caso è poco probabile che vi sia una differenza significativa tra i due gruppi (le medie sono uguali), ossia che vi sia una relazione tra appartenenza al gruppo e valore indicato.

Si può notare che:

- La devianza **totale** (TSS – Total Sum of Squares) è la somma delle devianze **interna ai gruppi** (WSS – Within Sum of Squares) e **tra un gruppo e l'altro** (BSS - Between Sum of Squares);
- Più è alta la quota di BSS rispetto al totale della TSS, più siamo vicini alla situazione di massima relazione;

Definisco l'indice *eta quadro* come rapporto tra la BSS e la TSS:

$$\text{Eta quadro} = \text{BSS} / \text{TSS}$$

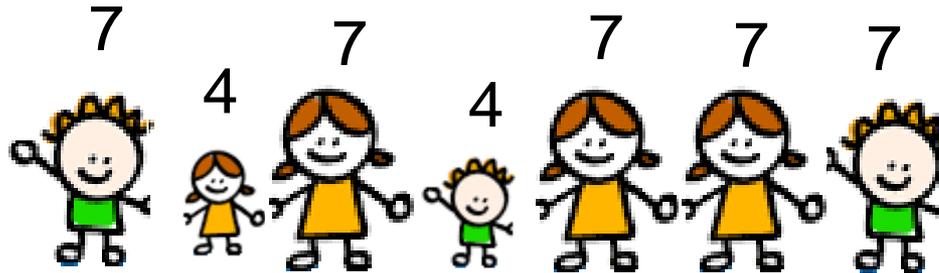
Se eta quadro è vicino a zero, non vi è relazione (tutta la devianza è all'interno dei gruppi).

Se eta quadro è vicino a uno, vi è relazione massima (tutta la devianza è tra un gruppo e l'altro).

Analisi della varianza

Caso 3

Gruppo A



$$\begin{aligned} \text{Media gruppo A} &= 6,14 \\ \text{SS gruppo A} &= 2 \cdot (4 - 6,14)^2 + 5 \cdot (7 - 6,14)^2 = 12,87 \end{aligned}$$

Gruppo B



$$\begin{aligned} \text{Media gruppo B} &= 5 \\ \text{SS gruppo B} &= 2 \cdot (4 - 5)^2 + 1 \cdot (7 - 5)^2 = 6 \end{aligned}$$

$$\begin{aligned} \text{Media generale} &= 5,8 \\ \text{SS generale} &= 6 \cdot (7 - 5,8)^2 + 4 \cdot (4 - 5,8)^2 = 21,6 \end{aligned}$$

Vi sono sia differenze tra un gruppo e l'altro (*between groups*) sia differenze all'interno dei gruppi (*within groups*).

$$\text{WSS} = \text{SS gruppo A} + \text{SS gruppo B} = 18,87$$

$$\text{TSS} = 21,6$$

Analisi della varianza

9

Caso 3, se non ci fosse WSS

Calcolo di BSS

Gruppo A 6,14 6,14 6,14 6,14 6,14 6,14 6,14

Media gruppo A =
6,14

Gruppo B 5 5 5

Media gruppo B
= 5

Media generale = 5,8
SS generale = $6 \cdot (7 - 5,8)^2 + 4 \cdot (4 - 5,8)^2 = 21,6$

Se azzero WSS imponendo che tutti i soggetti di ciascun gruppo abbiano valore pari alla loro media rimane solo BSS, che quindi vale

$$BSS = 7 \cdot (6,14 - 5,8)^2 + 3 \cdot (5 - 5,8)^2 = 2,73$$

$$\text{Eta quadro} = BSS / TSS = 2,73 / 21,6 = 0,13$$

Anche BSS ha una significatività ad essa associata → se è inferiore a 0,05 si considera non casuale il fatto che BSS sia diversa da zero.

Analisi bivariata con JsStat: relazione tra una variabile categoriale e una cardinale

L'ipotesi "Esiste relazione tra vivere con i genitori e l'età dei soggetti del campione." è confermata dai dati?

Analisi della varianza:

Vivi con i genitori? x Quanti anni hai?

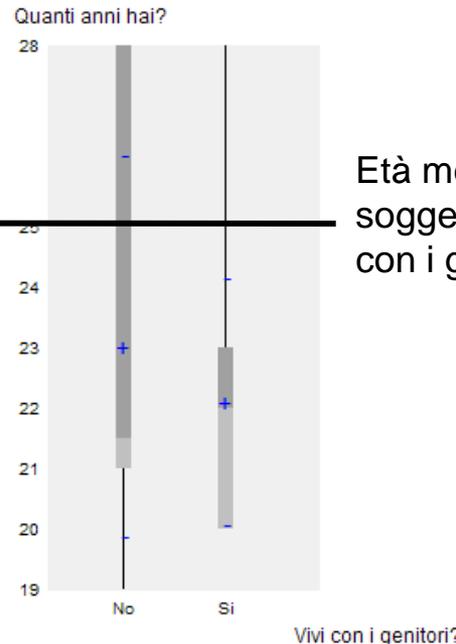
Categoria	Numero di casi	Media	Devianza	Scarto tipo
No	8	23	90	3.16
Si	25	22.08	103.84	2.04
Intero campione	33	22.3	188.97	2.39

Eta quadro = 0.03. Significatività = 0.36.

Eta quadro: Se è vicino a 0 → relazione debole, se è vicino a 1 → relazione forte

Significatività: Se è inferiore a 0,05 si opta per la presenza di una relazione significativa

Età media dei soggetti che **non** vivono con i genitori



Età media dei soggetti che vivono con i genitori

Più è alta la differenza, più è probabile che esista una relazione significativa tra le variabili

Analisi bivariata con JsStat: relazione tra una variabile categoriale non ordinata e una variabile categoriale ordinata

L'ipotesi "Esiste relazione tra il genere dei soggetti e il vedere il cartone animato i Bingson." è confermata dai dati?

Analisi della varianza non parametrica

Distribuzione di frequenza congiunta.

Maschio o femmina? x Ti capita di vedere il cartone animato i Bingson (1 mai .. 5 sempre)?

Ti capita di vedere il cartone animato i Bingson (1 mai .. 5 sempre)?->	1	2	3	4	Marginale di riga
Maschio o femmina?					
f		5	11	1	17
m	1	1	3	5	
Marginale di colonna	1	6	14	1	22

Frequenza **osservata** nel campione sotto esame

Più le frequenze dei maschi e quelle delle femmine si addensano ai due estremi opposti della scala, più è probabile che vi sia relazione

H di Kruskal & Wallis = 0.61. Significatività = 0.43

Significatività: Se è inferiore a 0,05 si opta per la presenza di una relazione significativa

H di Kruskal & Wallis: Se è vicino a 0 → relazione debole, se è grande → relazione forte

Analisi bivariata con JsStat: relazione tra due variabili categoriali ordinate

L'ipotesi "Esiste relazione tra il vedere il cartone i Bingson e il vedere il cartone GogoBall." è confermata dai dati?

Cograduazione

Distribuzione di frequenza congiunta:

Ti capita di vedere il cartone animato i Bingson (1 mai .. 5 sempre)? x Ti capita di vedere il cartone animato GogoBall (1 mai .. 5 sempre)?

Ti capita di vedere il cartone animato GogoBall (1 mai .. 5 sempre)?->	1	2	3	4	Marginale di riga
Ti capita di vedere il cartone animato i Bingson (1 mai .. 5 sempre)?					
1		1			1
2			4	2	6
3	1	2	7	3	13
4				1	1
Marginale di colonna	1	3	11	6	21

Frequenza **osservata** nel campione sotto esame

Più le frequenze si addensano sulla diagonale principale (dall'alto verso il basso), più è probabile che vi sia relazione

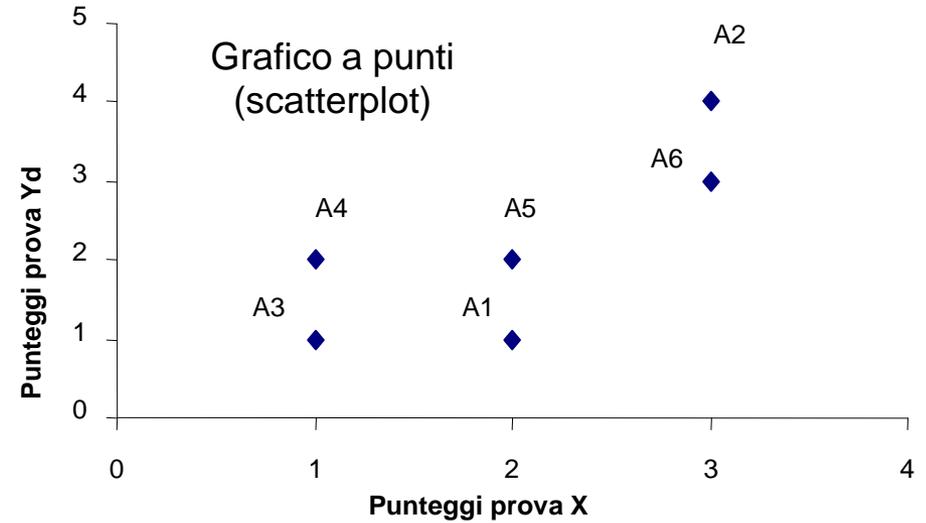
Tau c di Kendall (approssimato) = 0.07. Significatività = 0.66

Significatività: Se è inferiore a 0,05 si opta per la presenza di una relazione significativa

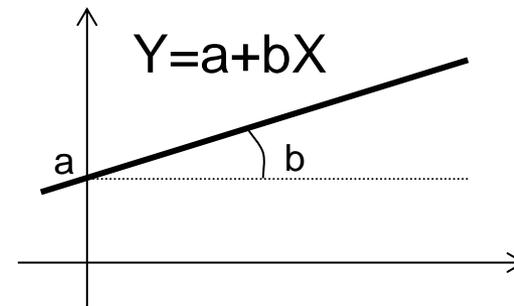
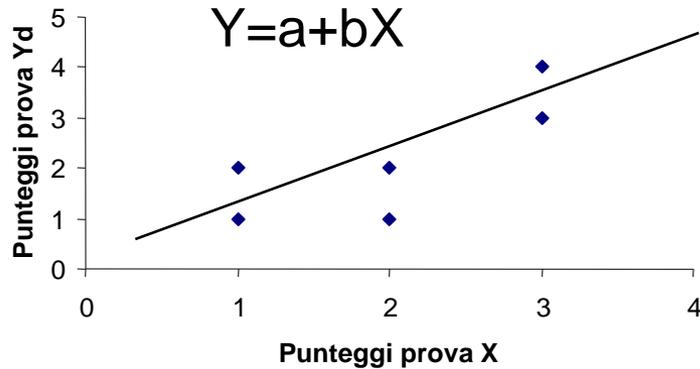
Tau c di Kendall: Se è vicino a 0 → relazione debole, se è vicino a +1 → relazione forte e positiva, se è vicino a -1 → relazione forte e negativa

Correlazione (tra variabili cardinali)

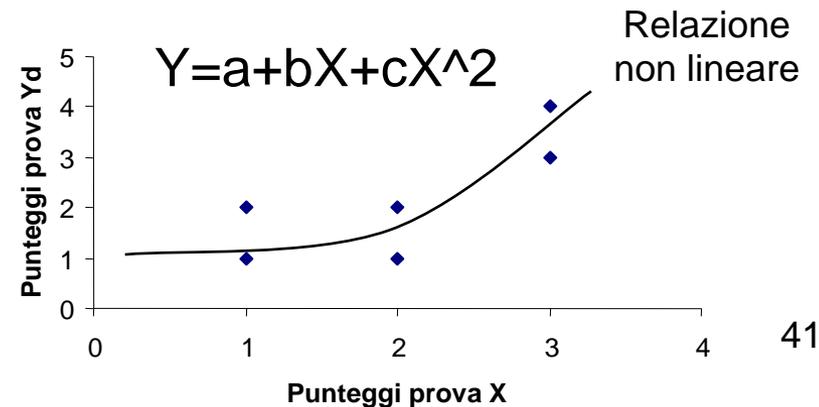
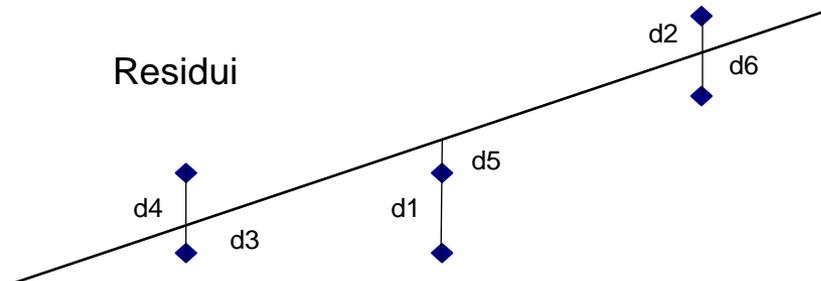
Soggetto	Punteggio grezzo nel test X	Punteggio grezzo nel test Y_d
A1	2	1
A2	3	4
A3	1	1
A4	1	2
A5	2	2
A6	3	3



Retta di regressione



Residui



Coefficiente di correlazione r (tra variabili cardinali)

Indice che indica la forza delle relazioni tra le due variabili.

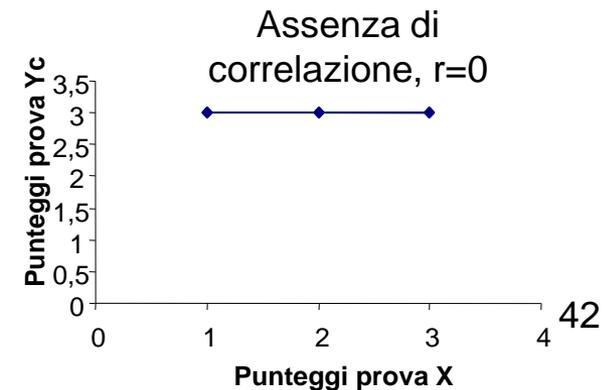
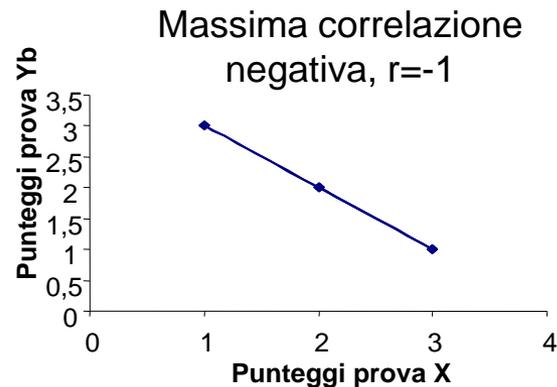
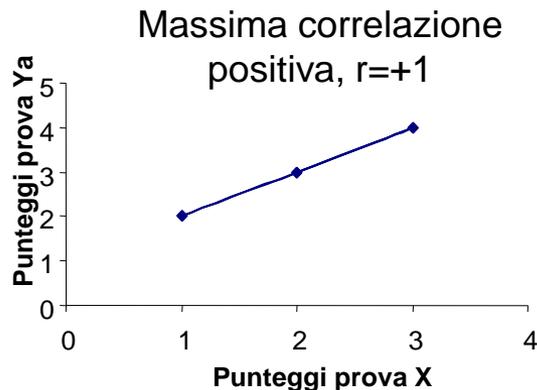
Se vale +1 → massima correlazione positiva (se una cresce, cresce anche l'altra).

Se vale -1 → massima correlazione negativa (se una cresce, l'altra diminuisce).

Se vale 0 → assenza di correlazione (se una cresce, l'altra non varia).

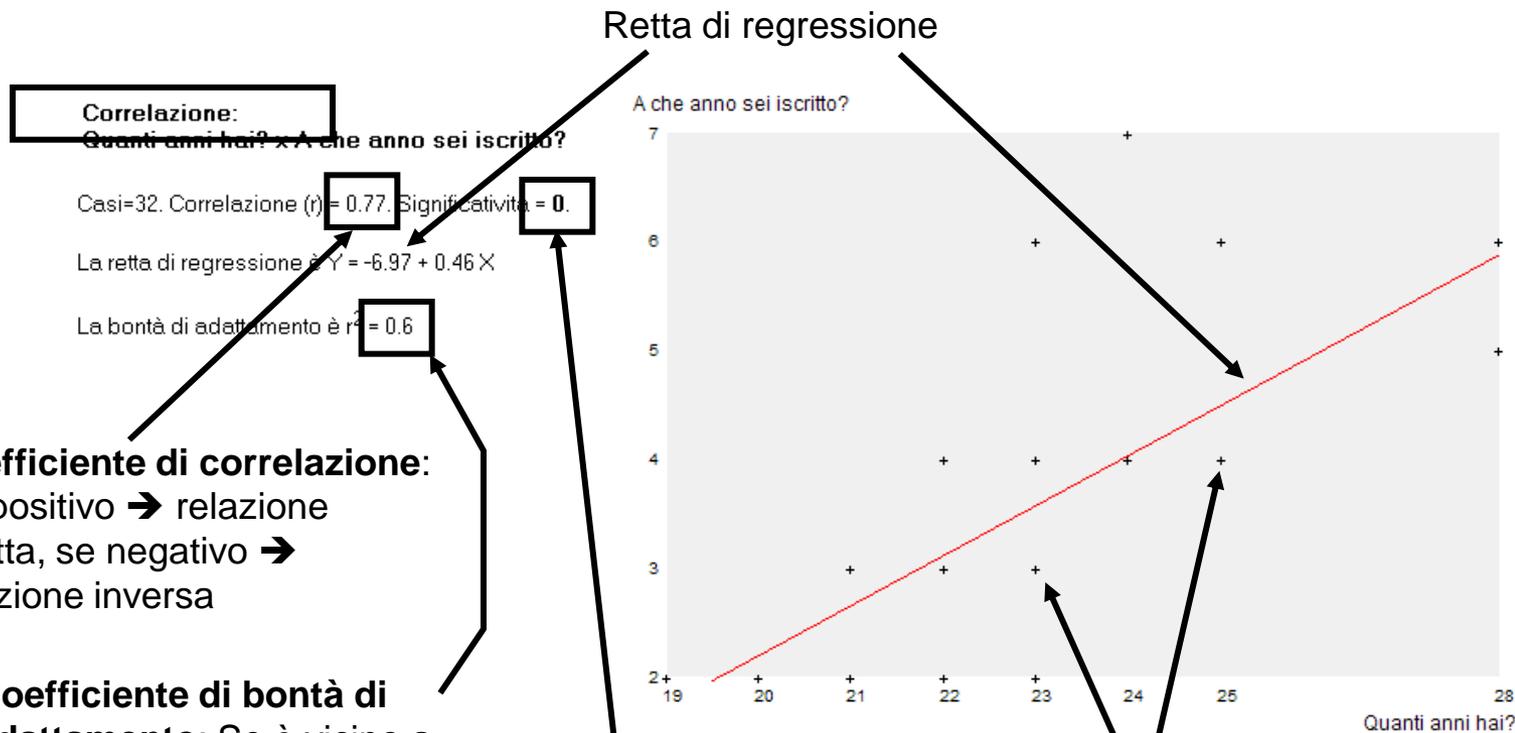
Anche r ha una significatività ad esso associata → se è inferiore a 0,05 si considera non casuale il fatto che r sia diverso da zero.

Soggetto	Punteggio grezzo nel test X	Punteggio grezzo nel test Ya	Punteggio grezzo nel test Yb	Punteggio grezzo nel test Yc
A1	2	3	2	3
A2	3	4	1	3
A3	1	2	3	3
A4	1	2	3	3
A5	2	3	2	3
A6	3	4	1	3



Analisi bivariata con JsStat: relazione tra due variabili cardinali

L'ipotesi "Esiste una relazione lineare tra l'età dei soggetti e l'anno di iscrizione all'Università." è confermata dai dati?



Più i soggetti sono vicini alla retta di regressione, più è probabile che esista una relazione lineare tra le variabili

Coefficiente di correlazione:
Se positivo → relazione diretta, se negativo → relazione inversa

Coefficiente di bontà di adattamento: Se è vicino a 0 → relazione debole, se è vicino a 1 → relazione forte

Significatività: Se è inferiore a 0,05 si opta per la presenza di una relazione significativa

Soggetti

Analisi bivariata: riepilogo

	Variabili categoriali non ordinate	Variabili categoriali ordinate	Variabili cardinali
Variabili categoriali non ordinate	Tabella a doppia entrata	Analisi della varianza non parametrica	Analisi della varianza
Variabili categoriali ordinate		Cograduazione	Analisi della varianza o Cograduazione
Variabili cardinali			Correlazione

Le variabili quasi cardinali si possono considerare cardinali se è possibile supporre che la distanza percepita tra i punti della scala sia la stessa per tutti i soggetti, ordinali altrimenti

Descrizioni ed esempi su: www.edurete.org/jsstat

Analisi esplorativa con JsStat: possibili relazioni in una matrice dei dati

Sono presenti nella matrice dei dati concordanze significative tra le modalità di due variabili?

Tipo di elaborazione
Esplora relazioni tra variabili

Matrice di probabilità degli incroci (Probabilità esatta di Fisher)

	età_28	età_22	età_23	età_20	età_21	età_24	età_25	età_19	genere_m	genere_f	provnas_to	provnas_al
età_28	0	0.6	0.67	0.48	0.42	0.82	0.74	0.91	0.01	0.01	0.21	0.91
età_22	0.6	0	0.5	0.28	0.22	0.72	0.6	0.85	0.23	0.23	0.23	0.15
età_23	0.67	0.5	0	0.37	0.31	0.77	0.67	0.88	0.37	0.37	0.36	0.88
età_20	0.48	0.28	0.37	0	0.11	0.62	0.48	0.79	0.38	0.38	0.28	0.79
età_21	0.42	0.22	0.31	0.11	0	0.57	0.42	0.76	0.33	0.33	0.02	0.76
età_24	0.82	0.72	0.77	0.62	0.57	0	0.82	0.94	0.62	0.62	0.36	0.94
età_25	0.74	0.6	0.67	0.48	0.42	0.82	0	0.91	0.48	0.48	0.21	0.91
età_19	0.91	0.85	0.88	0.79	0.76	0.94	0.91	0.03	0.79	0.79	0.39	0.97
genere_m	0.01	0.23	0.37	0.38	0.33	0.62	0.48	0.79	0	0	0.28	0.21
genere_f	0.01	0.23	0.37	0.38	0.33	0.62	0.48	0.79	0	0	0.28	0.21
provnas_to	0.21	0.23	0.36	0.28	0.02	0.36	0.21	0.39	0.28	0.28	0	0.39
provnas_al	0.91	0.15	0.88	0.79	0.76	0.94	0.91	0.97	0.21	0.21	0.39	0.03
provnas_ao	0.91	0.85	0.88	0.21	0.76	0.94	0.91	0.97	0.21	0.21	0.39	0.97
provnas_vc	0.82	0.72	0.42	0.62	0.38	0.88	0.82	0.94	0.62	0.62	0.15	0.94
provnas_cn	0.67	0.4	0.58	0.37	0.03	0.77	0.67	0.88	0.37	0.37	0.02	0.88
provnas_bi	0.74	0.35	0.67	0.42	0.44	0.82	0.74	0.91	0.48	0.48	0.05	0.91

Modalità "28 anni" della variabile Età

La significatività viene calcolata attraverso il Test esatto di Fisher

Significatività: Se è inferiore a 0,05 si opta per la presenza di una relazione significativa

Nel campione, esiste una concordanza tra "avere 28 anni" ed "avere genere maschile"

Analisi di dati semistrutturati in testi con concetti separati (pre-interpretati) con JsStat

Elenca i lavori precedentemente svolti:

cameriera, baby_sitter,
commessa_supermercato

JsStat

Tipo di elaborazione

Estrai lessico da variabile

Singoli termini
rappresentanti un concetto

Estrazione del **lessico**, conteggio
frequenze dei termini, costruzione
variabili aggiuntive nella matrice dei
dati (0 = termine non presente / 1 =
termine presente)

Lessico della variabile: Esperienze lavorative

Termine/concetto	Frequenza
baby_sitter	7
animatore	3
ripetizioni	2
cassiere	2
cameriere	2
promotore_cosmetici	2
barista	2
insegnante_elementare	1
commesso	1
maschera	1
insegnante_materna	1
assistente_disabili	1
supplente_elementari	1
rilevatore_ricerche_mercato	1
assistente_dentista	1

esplav	esplav_baby_sitter	esplav_cameriere	esplav_cassiere	esplav_commesso
insegnante_elementare	0	0	0	0
ripetizioni	0	0	0	0
cassiere commesso baby_sitter	1	0	1	1
-	0	0	0	0
baby_sitter cameriere	1	1	0	0
maschera	0	0	0	0
insegnante_materna	0	0	0	0
assistente_disabili	0	0	0	0
ripetizioni	0	0	0	0
supplente_elementari	0	0	0	0
baby_sitter	1	0	0	0
animatore	0	0	0	0

Analisi di dati a bassa strutturazione (o semi-strutturati, ma non pre-interpretati)

Costruzione di categorie a posteriori

espresse da

Concetti e/o asserti riassumono

Intervista: Anna, 27 anni

Intervistatore: Descrivimi il rapporto con la tua famiglia.

Allora, quando ero in casa, il tipo di rapporto con la mia famiglia non è mai stato fantastico, idilliaco, in quanto mia madre e mio padre erano molto severi con me ... volevano che vivessi una vita che avevano scelto loro per me ... quindi volevano che io frequentassi gente di un certo tipo, andassi in una scuola di un certo tipo, facessi solo cose che a loro sembrava andassero bene e hanno dato per scontato che a me queste cose piacessero, e invece a me non piacevano, non mi piaceva neanche stare con la gente che loro volevano che io frequentassi.

Quindi, diciamo che io sono stata una bambina abbastanza ribelle, perché io queste cose non le accettavo e...

es.

Rapporto difficile con la famiglia

Mancanza di possibilità di decidere la propria vita

Non-accettazione delle regole famigliari

es.

I genitori erano severi con lei

I genitori decidevano come si doveva comportare

Anna dichiara di essere stata una bambina ribelle

Unità naturali di significato presenti nel testo

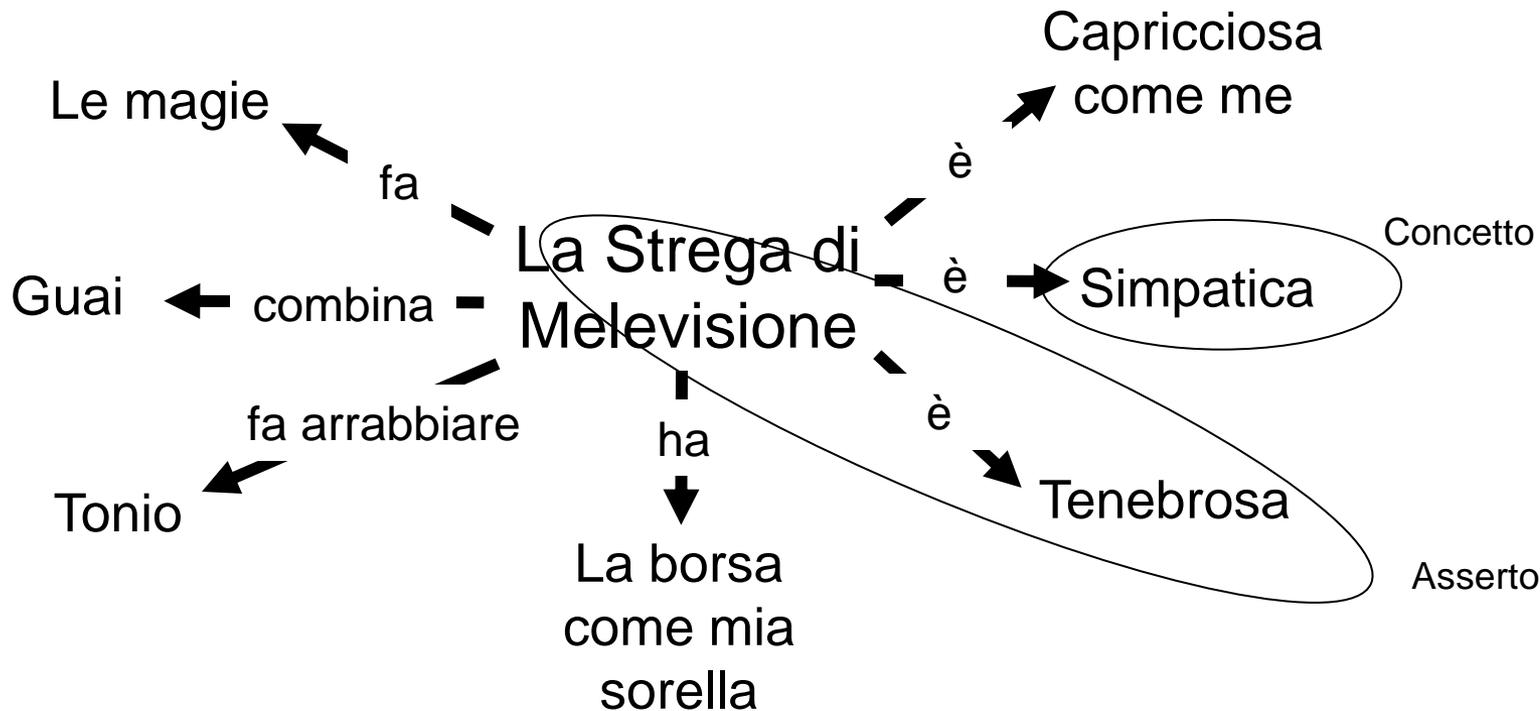
Regole per la costruzione di categorie a posteriori

... se uscivo, se potevo, alle 11 dovevo essere a casa ... anche se sono sicura che tutto questo l'hanno fatto per il mio bene, cioè fin troppo bene ...	Distinguere i fatti dalle interpretazioni
... mio padre amava leggere libri ... era molto severo con me ...	Distinguere ciò che è pertinente all'obiettivo di ricerca da ciò che non lo è
... mio padre e mia madre andavano molto d'accordo tra di loro [ride] ...	Tenere conto del comportamento non verbale
... non potevo mai fare quello che le mie amiche facevano ... quando i miei genitori hanno scoperto che fumavo in fondo non l'hanno presa così male ...	Mettere in luce coerenze e incoerenze
... quell'imbecille della mia amica ... quanto ci siamo divertite con lei ...	Tenere conto del contesto per assegnare significato alle affermazioni

Descrizione qualitativa

Quali sono le caratteristiche della Strega di Melevisione percepite dai bambini?

Le categorie costruite possono essere riassunte da una **mappa concettuale**



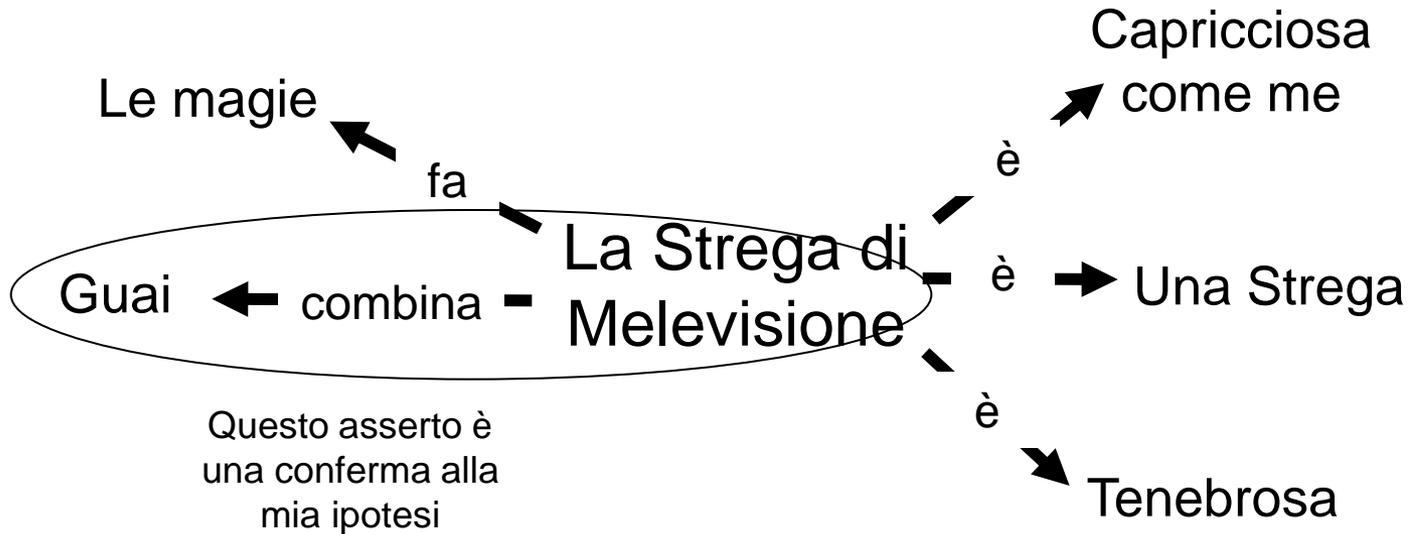
Ricostruzione delle “buone ragioni” soggettive alla base delle scelte

Perché piace la Strega di Melevisione?



Controllo di ipotesi nella ricerca interpretativa

Ipotesi: La Strega di Melevisione piace ai bambini perché è una figura negativa



Interpretazione dei risultati

Consiste nel:

- Inserire nel rapporto di ricerca tutte le relazioni significative trovate incrociando **tutte** le variabili generate dal fattore indipendente con **tutte** le variabili generate dal fattore dipendente (se non vi sono relazioni significative inserire tutte quelle che si avvicinano alla significatività) e commentarle;
- Dire se queste relazioni significative confermano o confutano l'ipotesi e dire secondo voi **perché** (se non vi sono relazioni significative l'ipotesi non è né confermata, né confutata dai dati);
- Formulare possibili **interpretazioni alternative** o effetti di possibili **fattori intervenienti** (es. non abbiamo trovato relazione perché ...).

Riflessione sull'esperienza di ricerca 10

Domande guida per l'autoriflessione:

- Cosa si è imparato?
- Cosa si rifarebbe allo stesso modo, se si potesse rifare la ricerca?
- Cosa si farebbe in modo diverso?
- Dal confronto con i rapporti di ricerca presentati dagli altri gruppi emergono margini di miglioramento?
- Quali sono a vostro avviso i punti di forza del vostro lavoro?
- Quali i punti di debolezza?

Raccomandazione. Compilare il **modulo di Autoscoring** della ricerca: <http://www.edurete.org/public/obj/autoscoringps.rtf>

Attività 4: Dai dati grezzi ai risultati di ricerca

1. Partendo dagli esiti delle Attività 1, 2 e 3, analizzate i dati che avete raccolto, in relazione agli obiettivi della vostra ricerca. Costruite poi un rapporto di ricerca finale in cui date conto delle scelte fatte durante l'intero processo di ricerca e in cui proponete delle conclusioni, in risposta al problema di ricerca da cui siete partiti.

5. Rivedete il vostro elaborato per renderlo compatibile con i criteri definiti nel punto precedente.

4. Sulla base dei confronti fatti e dei punti di forza emersi elencate i criteri che deve soddisfare un elaborato "ottimale".

2. Confrontate il vostro elaborato con quello dei vostri colleghi del gruppo di formazione. Valutatelo sulla base dei criteri proposti.

3. Il vostro elaborato rispetta i criteri proposti? Quali sono i punti di forza del vostro elaborato? Quali i punti di debolezza?