

Elementi di Psicometria

3-I percentili

vers. 1.1 (10 novembre 2011)

versione per stampa

Germano Rossi¹

`germano.rossi@unimib.it`

¹Dipartimento di Psicologia, Università di Milano-Bicocca

2011-2012

Proporzioni e percentuali

Esprimono il rapporto fra una parte rispetto ad un tutto (o ad un valore di riferimento)

Proporzione

$$\frac{f}{N}$$

la somma dev'essere 1 e nessuna proporzione può superare 1

Esempio

proporzione di femmine

$$\frac{13}{30} = 0.43$$

Percentuale

$$\frac{f}{N} \times 100$$

la somma dev'essere 100 e nessuna percentuale può superare 100

Esempio

percentuale di maschi

$$\frac{17}{30} \times 100 = 56.67$$

Percentuali

- Le proporzioni e le percentuali si possono calcolare a tutti i livelli (N, O, I/R), anche se è preferibile non usarli per variabili intervallo/rapporto continue e con molti valori
- Quando la N su cui si calcolano le percentuali è molto piccola, la percentuale perde di significato
- In genere non si calcolano le percentuali se la N è inferiore a 20 (ma anche se è 22!)
- Infatti 1 unità su 20 equivale al 5%, mentre 1 unità su 100 equivale all'1%
- Se per un qualunque motivo, si è verificato un errore di misurazione e una unità è stata “categorizzata” male, se N è basso, le percentuali cambieranno parecchio dando una visione falsata della realtà

Percentuali

- Le percentuali non hanno senso se non si conosce la N su cui sono calcolate

Marca	% sodio	Residuo fisso in mg/l	Sodio in mg/l
Sales	.00004	198.5	0.45
Flavia	.00007	264.1	0.70
Sant'Anna	<.0001	39.2	0.90
Vitasnella	<.0002	380.0	3.00
Bagolino	.0002	130.0	2.30
San Benedetto	<.0007	250.0	6.80
Ferrarelle		1245.0	48.00

Tabella delle frequenze

Una **tabelle delle frequenze** indica quanti valori abbiamo per ogni categoria della variabile.

Sesso	freq	prop	%
Maschi	17	.57	56.67
Femmine	13	.43	43.33
Totale	30	1.00	100

- Di solito si indicano **frequenze e percentuali** oppure **frequenze e proporzioni**
- A livello ordinale, si possono calcolare anche le *percentuali cumulate*

Frequenze cumulate

D12

		Frequenza	Percentuale	Percentuale valida	Percentuale cumulata
Validi	1	8	6,8	6,9	6,9
	2	22	18,6	19,0	25,9
	3	18	15,3	15,5	41,4
	4	14	11,9	12,1	53,4
	5	15	12,7	12,9	66,4
	6	13	11,0	11,2	77,6
	7	5	4,2	4,3	81,9
	8	8	6,8	6,9	88,8
	9	2	1,7	1,7	90,5
	10	11	9,3	9,5	100,0
	Totale	116	98,3	100,0	
Mancanti	Mancante di sistema	2	1,7		
	Totale	118	100,0		

Il 53.4% dei soggetti ha dato come risposta valori inferiori a 5.

Frequenze cumulate

Autoritarismo

		Frequenza	Percentuale	Percentuale valida	Percentuale cumulata
Validi	20,00	1	,8	,9	,9
	24,00	1	,8	,9	1,8
	26,00	1	,8	,9	2,7
	27,00	2	1,7	1,8	4,5
	31,00	3	2,5	2,7	7,1
	32,00	2	1,7	1,8	8,9
	33,00	2	1,7	1,8	10,7
	34,00	1	,8	,9	11,6
	35,00	1	,8	,9	12,5
	36,00	2	1,7	1,8	14,3
	37,00	4	3,4	3,6	17,9
	38,00	3	2,5	2,7	20,5

Su variabili intervallo/rapporto non hanno molto senso le tabelle di frequenza (perché ci sono molti valori).

Rango percentile

- Percentuale di casi che ha ottenuto un punteggio pari o inferiore a quello considerato
- Qual è il rango percentile (RP) del punteggio 28?

Esempio

	f	fc	
16-20	10	10	
21-25	5	15	$15/25 = 0,6$
26-30	5	20	$5/25 = 0,2$
31-35	2	22	
36-40	2	24	
41-45	1	25	

- Troviamo dove cade il nostro punteggio (26-30)
- Trasformiamo in % la frequenza dell'intervallo
- Sommiamo assieme le frequenze degli intervalli inferiori e trasformiamoli in %
- Il nostro rango percentile è almeno superiore a 60 (.6)

Rango percentile

- Qual è il rango percentile (RP) del punteggio 28?

Esempio

	f	fc	
16-20	10	10	
21-25	5	15	$15/25 = 0,6$
26-30	5	20	$5/25 = 0,2$
31-35	2	22	
36-40	2	24	
41-45	1	25	

- Cumuliamo le % ($.6 + .2$)
- Il nostro RP è comunque inferiore a 80
- È un valore intermedio
- Ricordiamo che le f di un intervallo occupano uno spazio uguale
- Le 5 f occupano $1/5$ dell'intervallo che è ampio 5, quindi occupano 1
- 28 è il terzo valore dell'intervallo e la sua percentuale corrisponde quindi a $3/25 = .12$
- In realtà l'intervallo non inizia a 26 ma a 25.5 (LRI)

Rango percentile

- Qual è il rango percentile (RP) del punteggio 28?

Esempio

	f	fc	
16-20	10	10	
21-25	5	15	$15/25 = 0,6$
26-30	5	20	$5/25 = 0,2$
31-35	2	22	
36-40	2	24	
41-45	1	25	

- Quindi dobbiamo calcolare dove si posiziona 28 all'interno dell'intervallo
- $28 - 25.5$ cioè $P_{unt} - LRI$
- Adesso trasformiamo la posizione in % ($2.5/25 = .1$)
- e sommiamo questa percentuale a quella cumulata degli intervalli inferiori ($.6 + .1 = .7$)

- Il rango percentile del punteggio 28 è $.7 \times 100 = 70$

Rango percentile

- Qual è il rango percentile (RP) del punteggio 28?

Possiamo riassumere questo ragionamento in una formula

Esempio

	f	fc	
16-20	10	10	
21-25	5	15	15/25 = 0,6 = L%
26-30	5	20	5/25 = 0,2 = I%
31-35	2	22	
36-40	2	24	
41-45	1	25	
Punt=28	h=5	LRI=25.5	

$$RP = L\% + \left(I\% \times \frac{Punt - LRI}{h} \right)$$

$$RP = .6 + \left(.2 \times \frac{28 - 25.5}{5} \right)$$

$$RP = .7 = 70$$

Percentile

- Punteggio al di sotto di cui c'è una determinata percentuale di casi
- A quale punteggio corrisponde il 70imo percentile?

Esempio

	f	fc
16-20	10	10
21-25	5	15
26-30	5	20
31-35	2	22
36-40	2	24
41-45	1	25

- Trasformiamo il percentile in una posizione ovvero “a cosa corrisponde il 70% di 25?”
- $p \times N = .7 \times 25 = 17.5$
- Il punteggio che cade nella posizione 17.5 è il punteggio che ha sotto di sé il 70 per cento dei valori
- A quale valore corrisponde questa posizione?
- Usando la *fc* troviamo l'intervallo in cui cade la posizione 17.5 (cioè l'intervallo 26-30)

Percentile

- A quale punteggio corrisponde il 70imo percentile?

Esempio

	f	fc
16-20	10	10
21-25	5	15
26-30	5	20
31-35	2	22
36-40	2	24
41-45	1	25

- Ma nell'intervallo 26-30 abbiamo 5 f . Qual è la posizione 17.5?
- Sappiamo che l'intervallo precedente arriva fino alla posizione 15
- Quindi ci serve il valore che sta a $2.5 = 17.5 - 15$ posizioni dall'inizio dell'intervallo
- Ma l'intervallo è ampio 5 e ha 5 frequenze, quindi $2.5/5 \times 5 = 2.5$
- Che sommeremo al limite inferiore dell'intervallo: $2.5 + 25.5 = 28$

Percentile

- A quale punteggio corrisponde il 70imo percentile?

Possiamo riassumere questo ragionamento in una formula

Esempio

	f	fc	
16-20	10	10	
21-25	5	15	SFI=15
26-30	5	20	
31-35	2	22	
36-40	2	24	
41-45	1	25	

p=.7

N=25

h=5

f=5

$$P_{70} = LRI + \left(h \times \frac{pN - SFI}{f} \right)$$

$$P_{70} = 25.5 + \left(5 \times \frac{.7 \times 25 - 15}{5} \right)$$

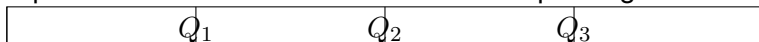
$$P_{70} = 28$$

Indici di posizione - Quantili

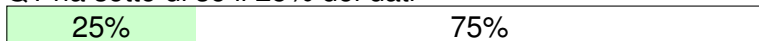
- Dopo aver ordinato i valori di una distribuzione, possiamo suddividere l'intera distribuzione di frequenza in n parti uguali.
- Se divisa in 100 parti, **Centili** (C_1, C_2, \dots, C_{99}) o **Percentili** (P_1, P_2, \dots)
- se divisa in 10 parti, **Decili** (D_1, D_2, \dots, D_9)
- se divisa in 4 parti, **Quartili** (Q_1, Q_2, Q_3)
- se divisa in 3 parti, **Terzili**
- Notate che
 - $D_1 = P_{10}$, e così via
 - $Q_1 = P_{25}$,
 - $Q_2 = P_{50} = D_5$ è anche chiamato "Mediana",
 - $Q_3 = P_{75}$

Quartili

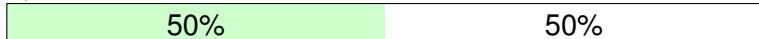
- I quartili suddividono la distribuzione in 4 parti uguali



- Si usano solitamente il primo e il terzo quartile (Q_1 e Q_3)
- Q_1 ha sotto di sé il 25% dei dati



- Q_2 ha sotto di sé il 50% dei dati



- Q_3 ha sotto di sé il 75% dei dati



Quartili: formule

- La posizione in cui cadono i quartili si trova con:

$$Q_1 = \frac{1}{4}(N + 1) = \frac{(N + 1)}{4}$$

$$Q_2 = \frac{2}{4}(N + 1) = \frac{2(N + 1)}{4} = \frac{N + 1}{2}$$

$$Q_3 = \frac{3}{4}(N + 1) = \frac{3(N + 1)}{4}$$

- Se la posizione trovata non è un intero, si tronca (ovvero si usa l'intero inferiore)
- Una volta trovata la posizione si identifica il valore del quartile (**il valore che corrisponde alla posizione**)

Quartili

Esempio

2 4 6 8 10 12 14 16 18 21 22 24 26 28 30

- $Q1 = (1/4) * (15+1) = 16/4 = 4 \Rightarrow Q1=8$
- $Q2 = (15+1)/2 = 8 \Rightarrow Q2=16$
- $Q3 = (3/4) * (15+1) = 48/4 = 12 \Rightarrow Q3=24$

Esercizio

2 4 5 9 10 12

- 1** $Q1 = ?$
- 2** $Q2 = ?$
- 3** $Q3 = ?$

Soluzione

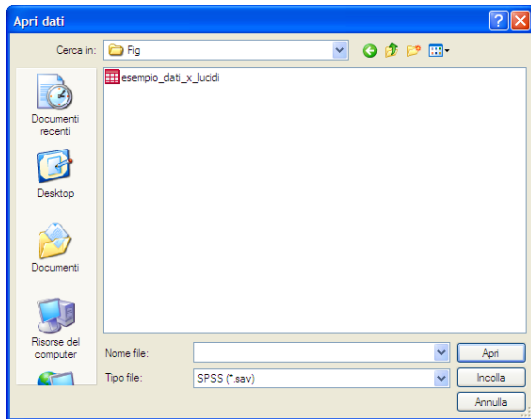
2 4 5 9 10 12

- 1** $N=6; \text{pos}=1.75; Q1=2$
- 2** $N=6; \text{pos}=3.5; Q2=5$
- 3** $N=6; \text{pos}=5.25; Q3=10$

Spss

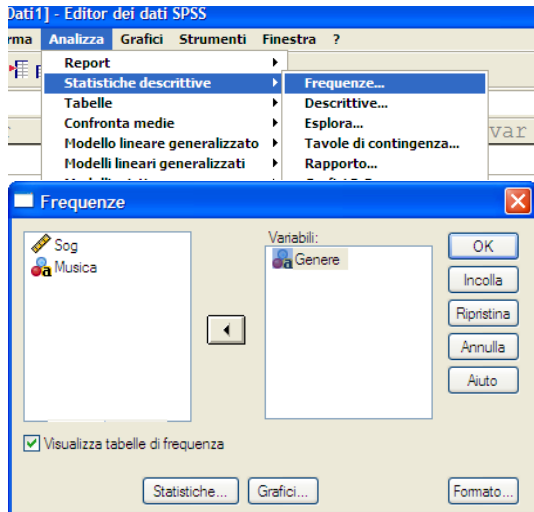
Usando il file `esempio1.sav` vediamo le statistiche presentate.

- Usando File | Apri | Dati cerchiamo e apriamo il file



Spss: frequenze

- Calcoliamo le frequenze con il comando
Analizza | Statistiche descrittive | Frequenze...
- Poi spostiamo Genere fra le Variabili
- E premiamo OK



Spss: frequenze

Frequenze

[InsiemeDati1] C:\Documenti\TeX\lucidi\Elem\Fig\esempio_dati_x_lucidi.sav

Statistiche

Genere

N	Validi	30
	Mancanti	0

Genere

		Frequenza	Percentuale	Percentuale valida	Percentuale cumulata
Validi	F	13	43,3	43,3	43,3
	M	17	56,7	56,7	100,0
Totale		30	100,0	100,0	

Spss: n-tili (Frequenze)

- In **Statistiche...**
- **Quartili** calcola i quartili
- **Punti di divisione** divide in n parti uguali
- **Percentili**: scrivete il percentile che volete e aggiungete

Fasce di età

N	Validi	100
	Mancanti	0
Percentili	25	1,00
	33,33333333	2,00
	45	2,00
	50	2,00
	66,66666667	3,00
	75	3,00

Valori percentili

☒ Quartili

☒ Punti di divisione per: gruppi uguali

☒ Percentili:

Aggiungi

Cambia

Rimuovi

45,0

Qui abbiamo chiesto: i quartili, i
terzili e il 45esimo percentile

Spss: Esplora...

- Dopo aver scelto le variabili, metterle in Variabili dipendenti



- poi

- click-are su

Statistiche...

e

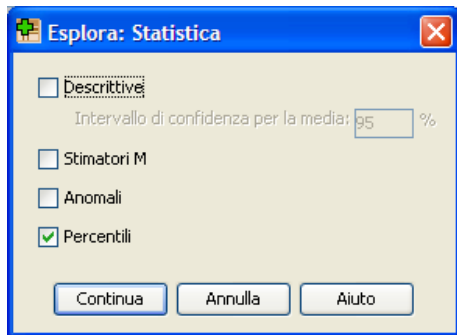
scegliere Percentili

- Quindi, click-are su

Continua

- Poi

OK



Spss: Esplora...

Esplora non permette di scegliere, ma fornisce alcuni n-tili notevoli

Percentili

		Percentili						
		5	10	25	50	75	90	95
Media ponderata (Definizione 1)	Fasce di età	1,00	1,00	1,00	2,00	3,00	3,00	3,00
Cardini di Tukey	Fasce di età			1,00	2,00	3,00		

I tre quartili e i valori corrispondenti al 5% e 10% su entrambi i lati. Questi valori hanno particolarmente senso con variabili “normali” (capiremo più avanti).

Riepilogo

- **con variabili nominali**: frequenze, proporzioni, percentuali; tabelle di frequenze
- **con variabili ordinali**: frequenze, proporzioni e percentuali cumulate; tabelle di frequenze; quantili (quartili; percentili e ranghi percentili con molte categorie)
- **con variabili intervallo/rapporto**: si possono fare tutte queste statistiche; *le più utilizzate sono*: quantili (in particolare quartili e terzili; percentili e ranghi percentili con molte categorie)