STATISTICA PSICOMETRICA a.a. 2004/2005

Corsi di laurea

Scienze e tecniche neuropsicologiche

Modulo 1

Statistica descrittiva Monovariata

Modulo 1

Statistica descrittiva Monovariata

- Distribuzioni semplici di frequenza e loro rappresentazioni
- Operatori di tendenza centrale
- Operatori di dispersione
- Momenti omogenei ed indici di forma (simmetria/curtosi)
- Standardizzazione di variabili cardinali

Statistica descrittiva Monovariata

Distribuzioni semplici di frequenza e loro rappresentazioni

Dalla distribuzione unitaria

alla distribuzione semplice di frequenze

(sconnessa,ordinata,seriazione)

ident	genere
1	1
2	2
3	1
4	1
5	2
6	1
7	1
8	2
9	1
10	2
11	1
12	1
13	2

Matrice CxV: tante righe quanti sono i casi (N) !

genere	n
1	8
2	5

Distribuzione di frequenza : tante righe quante sono le modalità della variabile (K) !

Statistica descrittiva Monovariata

Distribuzioni semplici di frequenza e loro rappresentazioni

Frequenze:

Assolute	n_k	$\sum_{k=1}^{K} n_k = N$
Relative	$f_k = \frac{n_k}{N}$	$\sum_{k=1}^{K} f_k = 1$
Percentuali	$q_k = f_k \bullet 100$	$\sum_{k=1}^{K} q_k = 100$

Statistica descrittiva Monovariata

Distribuzioni semplici di frequenza e loro rappresentazioni

ident	residenza
1	1
2	2
3	1
4	1
5	2
6	3
7	4
8	2
9	1
10	3
11	4
12	3
13	1

residenza	n	f	q
1	5	0,38	38,5
2	3	0,23	23,1
3	3	0,23	23,1
4	2	0,15	15,4
totale	*13*	*1*	*100*

Statistica descrittiva Monovariata

Distribuzioni semplici di frequenza e loro rappresentazioni

Se la variabile è ordinale parliamo di serie ordinata di frequenze ➔ posso calcolare le frequenze cumulate

ident	titolo di studio
1	1
2	2
3	1
4	2
5	3
6	2
7	1
8	2
9	3
10	3
11	2
12	1
13	2

titolo di studio	n	f	q	n'	f'	q'
1=obbligo	4	0,31	30,8	4	0,31	30,8
2=med. Sup.	6	0,46	46,2	10	0,77	76,9
3=laurea	3	0,23	23,1	13	1,00	100,0

… quanti soggetti hanno al massimo …?

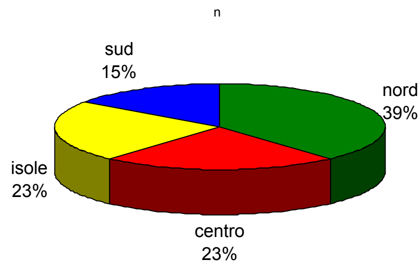
Statistica descrittiva Monovariata

Distribuzioni semplici di frequenza e loro rappresentazioni

Rappresentazioni Grafiche

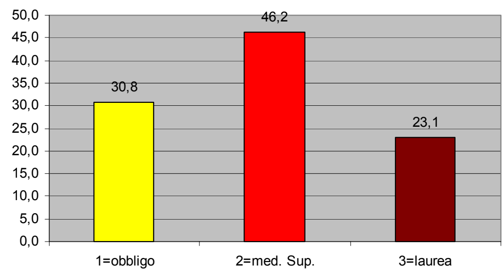
Serie sconnessa

➔diagramma a torta (o a barre)



Serie ordinata

➔istogramma



Statistica descrittiva Monovariata

Operatori monovariati

Un operatore statistico monovariato è:

un procedimento di calcolo che produce una statistica con le seguenti caratteristiche

1. È costituita da un unico scalare
2. Si riferisce ad una singola variabile
3. <u>È appropriata al livello di scala</u>
4. È insensibile all'ordine in cui i dati vengono registrati

Statistica descrittiva Monovariata

Operatori monovariati

Un operatore statistico monovariato può essere

1. Un operatore di tendenza centrale (il valore che rappresenta al meglio la distribuzione intera)
2. Un operatore di dispersione (il valore che informa circa la diversità esistente tra le osservazioni)
3. Un operatore di forma (relativi alla simmetria o alla curtosi della distribuzione)

Statistica descrittiva Monovariata

Operatori di tendenza centrale

➔ Moda

➔ Mediana

➔ Media

Statistica descrittiva Monovariata

Operatori di tendenza centrale: Moda

Moda= modalità della variabile con frequenza più elevata

Idonea per scale nominali, ordinali e cardinali

moda →

residenza	n
nord	5
centro	3
isole	3
sud	2

La moda è "Nord" , perché è la modalità della variabile RESIDENZA che si presenta con la frequenza più elevata (n=5)

Statistica descrittiva Monovariata

Operatori di tendenza centrale: Mediana

Mediana= modalità a cui appartiene il caso (caso mediano) che divide esattamente in due la distribuzione

Idonea per scale ordinali e cardinali

Se N (numero totale dei casi) è dispari il caso mediano sarà uno solo

$$C_{Mdn} = \frac{(N+1)}{2}$$

Se N (numero totale dei casi) è pari avremo due casi mediani

$$1°C_{Mdn} = \frac{N}{2} \qquad 2°C_{Mdn} = \frac{N}{2} + 1$$

Statistica descrittiva Monovariata

Operatori di tendenza centrale: Mediana

I casi sono dispari (13) quindi c'è solo un caso mediamo

$$C_{Mdn} = \frac{(N+1)}{2} \qquad C_{Mdn} = \frac{(13+1)}{2} = \frac{14}{2} = 7$$

titolo di studio	n	f	q	n'
1=obbligo	4	0,31	30,8	4
2=med. Sup.	6	0,46	46,2	10
3=laurea	3	0,23	23,1	13

Calcolando le frequenze cumulate vediamo che il 7° caso cade nella categoria Media superiore.

Media superiore rappresenta quindi la mediana della distribuzione

Statistica descrittiva Monovariata

Operatori di tendenza centrale: Media Aritmetica

Media (aritmetica): la media si ottiene sommando tutti i valori di X (da 1 a N) e dividendo tale somma per il numero dei casi (N)

Idonea per scale cardinali

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

Statistica descrittiva Monovariata

Operatori di tendenza centrale: Media Aritmetica

ident	peso
1	72
2	58
3	65
4	78
5	49

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

$$\bar{x} = \frac{(72+58+65+78+49)}{5} = \frac{322}{5} = 64{,}4$$

Statistica descrittiva Monovariata

Operatori di tendenza centrale: Media Aritmetica

Proprietà della media

➔ **La somma algebrica degli scarti dei valori dalla loro media aritmetica è uguale a zero**

$$\sum_{i=1}^{N}(x_i - \bar{x}) = 0$$

ident	peso	scarto
1	72	7,6
2	58	-6,4
3	65	0,6
4	78	13,6
5	49	-15,4
	somma	0,0

Statistica descrittiva Monovariata

Operatori di tendenza centrale: Media Aritmetica

Proprietà della media

➔ **La somma algebrica dei quadrati degli scarti dei valori dalla loro media aritmetica è minima**

$$\sum_{i=1}^{N}(x_i - a)^2 = \min \qquad \text{se} \qquad a = \bar{x}$$

Statistica descrittiva Monovariata

Operatori di tendenza centrale: Altri indici di posizione

Così come la mediana è la modalità nella quale cade il caso che divide in 2 parti uguali la distribuzione possiamo pensare ad indici che dividono in più parti uguali la distribuzione

Ad esempio i quartili sono le categorie nelle quali cadono i 3 casi che dividono in 4 parti uguali la distribuzione

Idonea per scale ordinali e cardinali

Quartili – decili – centili

quantili

Operatori di dispersione

➔ Mutabilità Categoriale

➔ Varibilità non metrica Ordinale

➔ Variabilità metrica Cardinale

Operatori di dispersione

Un operatore di dispersione produce uno scalare con cui si valuta sinteticamente la diversità esistente tra le osservazioni

Campione 1

ID	età
1	27
2	25
3	32
4	27
5	25
6	27
7	31
8	30
9	32
10	27
11	26
12	26
13	33
14	30
15	25
16	33
17	27
18	34
19	30
20	26

Campione 2

ID	età
1	45
2	34
3	67
4	34
5	25
6	18
7	17
8	6
9	21
10	8
11	24
12	39
13	41
14	15
15	26
16	45
17	10
18	24
19	63
20	11

Statistica descrittiva Monovariata

Operatori di dispersione

I due campione hanno media identica

ID	età
1	27
2	25
3	32
4	27
5	25
6	27
7	31
8	30
9	32
10	27
11	26
12	26
13	33
14	30
15	25
16	33
17	27
18	34
19	30
20	26

Media Campione 1

28,65

Media Campione 2

ID	età
1	45
2	34
3	67
4	34
5	25
6	18
7	17
8	6
9	21
10	8
11	24
12	39
13	41
14	15
15	26
16	45
17	10
18	24
19	63
20	11

Statistica descrittiva Monovariata

Operatori di dispersione

Sono molto diversi invece per quanto riguarda il campo di variazione

ID	età
1	27
2	25
3	32
4	27
5	25
6	27
7	31
8	30
9	32
10	27
11	26
12	26
13	33
14	30
15	25
16	33
17	27
18	34
19	30
20	26

Campione 1

25 — 34

Campione 2

6 — 67

ID	età
1	45
2	34
3	67
4	34
5	25
6	18
7	17
8	6
9	21
10	8
11	24
12	39
13	41
14	15
15	26
16	45
17	10
18	24
19	63
20	11

Statistica descrittiva Monovariata

Operatori di dispersione: mutabilità

Per variabili categoriali

Eterogeneità ⟷ **Omogeneità**

Ciascuna modalità della variabile ha la medesima frequenza (N/K)

Tutte le osservazioni si riferiscono ad una sola modalità

Capitale preferita	n
Parigi	20
Londra	20
Berlino	20
Madrid	20
Praga	20

Capitale preferita	n
Parigi	0
Londra	100
Berlino	0
Madrid	0
Praga	0

Statistica descrittiva Monovariata

Operatori di dispersione: mutabilità

Indice di GINI

Uno meno la sommatoria per k che va da 1 a K del quadrato delle frequenze relative

$$E_1 = 1 - \sum_{k=1}^{K} f_k^2$$

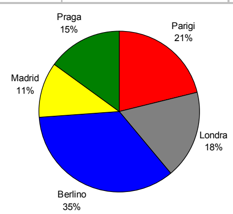
(K-1)/K **Eterogeneità** ⟷ **0** **Omogeneità**

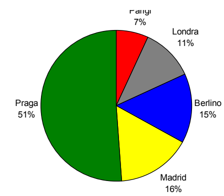
Statistica descrittiva Monovariata

Operatori di dispersione: mutabilità

Indice di GINI: esempio

Capitale preferita	n	f	f^2		Capitale preferita	n	f	f^2
Parigi	21	0,21	0,044		Parigi	7	0,07	0,005
Londra	18	0,18	0,032		Londra	11	0,11	0,012
Berlino	35	0,35	0,123		Berlino	15	0,15	0,023
Madrid	11	0,11	0,012		Madrid	16	0,16	0,026
Praga	15	0,15	0,023		Praga	51	0,51	0,26
	somma=		0,234			somma=		0,325
	E_1		0,766			E_1		0,675





Statistica descrittiva Monovariata

Operatori di dispersione: mutabilità

Indici relativi

Tramite l'operazione di ranging possiamo far variare l'indice di GINI tra 0 ed 1

Le misure relative (e) si ottengono sottraendo a quelle assolute (E) il valore minimo che possono raggiungere e dividendo il risultato per il suo intervallo di variazione

$$e = \frac{E - \min}{\max - \min}$$

Indici relativi

Nel caso dell'indice di GINI

$$e_1 = \frac{E_1 - 0}{\frac{(K-1)}{K} - 0} = \frac{E_1}{\frac{(K-1)}{K}} = \frac{K}{K-1} E_1$$

Altri indici

Nome	Formula	min	max
Gini	$E_1 = 1 - \sum_{k=1}^{K} f_k^2$	0	$\frac{K-1}{K}$
Leti	$E_3 = \frac{1}{\prod_{k=1}^{K} f_k^{f_k}}$	1	K
Entropia	$E_2 = 1 - \sum_{k=1}^{K} f_k \log_a f_k$	0	$\log_a K$

Statistica descrittiva Monovariata

Operatori di dispersione: variabilità non metrica

Per variabili ordinali

In una variabile ordinale è ragionevole considerare come più simili le osservazioni che cadono in modalità contigue, piuttosto che osservazioni che cadono in modalità estreme.

L'operatore D* gode di queste proprietà

$$D^* = 2\sum_{k=1}^{K-1}\left[f_k^{'}\left(1-f_k^{'}\right)\right]$$

Valore minimo =0

Valore massimo = Se N è dispari $\frac{K-1}{2}\left(1-\frac{1}{N^2}\right)$

Se N è pari $\frac{K-1}{2}$

Statistica descrittiva Monovariata

Operatori di dispersione: variabilità non metrica

Per variabili ordinali

Titolo di studio	*f*	*f'*	*1-f'*	*f' (1-f')*
Elementari	0,3	0,3	0,7	0,21
Media inferiore	0,2	0,5	0,5	0,25
Media Superiore	0,2	0,7	0,3	0,21
Laurea	0,3	1	0	

$$D^* = 2\sum_{k=1}^{K-1}\left[f_k^{'}\left(1-f_k^{'}\right)\right]$$

$$D^* = 2\left(0{,}21+0{,}25+0{,}21\right) = 2(0{,}67) = 1{,}34$$

Statistica descrittiva Monovariata

Operatori di dispersione: variabilità metrica

Per variabili cardinali

Prendiamo in considerazione due famiglie di operatori

- intervalli di variazione
- Scarti da un valore centrale

Statistica descrittiva Monovariata

Operatori di dispersione: variabilità metrica

Gli intervalli di variazione sono operatori che quantificano la variabilità misurando la diversità tra due particolari termini della distribuzione

RANGE

$$W = X_{max} – X_{min}$$

SEMIDIFFERENZA INTERQUARTILE

$$W' = Q_3 – Q_1$$

Statistica descrittiva Monovariata

Operatori di dispersione: variabilità metrica

Scarti da un valore centrale

Devianza

$$DEV = \sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2$$

Varianza

$$S^2 = \frac{\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2}{N}$$

Deviazione Standard

$$S = \sqrt{\frac{\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2}{N}}$$

Coefficiente di variazione

$$cv = \frac{s}{\bar{x}} \cdot 100$$

Statistica descrittiva Monovariata

Operatori di dispersione: variabilità metrica

Esempio:range

Campione 1

ID	età
1	27
2	25
3	32
4	27
5	25
6	27
7	31
8	30
9	32
10	27
11	26
12	26
13	33
14	30
15	25
16	33
17	27
18	34
19	30
20	26

W=34-25=9

ID	età
1	45
2	34
3	67
4	34
5	25
6	18
7	17
8	6
9	21
10	8
11	24
12	39
13	41
14	15
15	26
16	45
17	10
18	24
19	63
20	11

W=67-6=61

Campione 2

Statistica descrittiva Monovariata

Operatori di dispersione: variabilità metrica

Esempio.devianza

Campione 1

ID	età	scarti	$scarti^2$
1	27	-1,65	2,723
2	25	-3,65	13,323
3	32	3,35	11,223
4	27	-1,65	2,723
5	25	-3,65	13,323
6	27	-1,65	2,723
7	31	2,35	5,523
8	30	1,35	1,823
9	32	3,35	11,223
10	27	-1,65	2,723
11	26	-2,65	7,022
12	26	-2,65	7,022
13	33	4,35	18,923
14	30	1,35	1,823
15	25	-3,65	13,323
16	33	4,35	18,923
17	27	-1,65	2,723
18	34	5,35	28,623
19	30	1,35	1,823
20	26	-2,65	7,022
media=	28,65	DEV=	174,550

$$DEV = \sum_{i=1}^{N} \left(x_i - \bar{x}\right)^2$$

Statistica descrittiva Monovariata

Operatori di dispersione: variabilità metrica

Esempio.devianza

Campione 2

ID	età	scarti	scarti2
1	45	16,35	267,32
2	34	5,35	28,62
3	67	38,35	1470,72
4	34	5,35	28,62
5	25	-3,65	13,32
6	18	-10,65	113,42
7	17	-11,65	135,72
8	6	-22,65	513,02
9	21	-7,65	58,52
10	8	-20,65	426,42
11	24	-4,65	21,62
12	39	10,35	107,12
13	41	12,35	152,52
14	15	-13,65	186,32
15	26	-2,65	7,02
16	45	16,35	267,32
17	10	-18,65	347,82
18	24	-4,65	21,62
19	63	34,35	1179,92
20	11	-17,65	311,52
media=	28,65	DEV=	5658,55

$$DEV = \sum_{i=1}^{N} \left(x_i - \bar{x}\right)^2$$

Statistica descrittiva Monovariata

Operatori di dispersione: variabilità metrica

Esempio: varianza , deviazione standard, coefficiente var.

		Campione 1	Campione 2
Devianza	$DEV = \sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2$	174,6	5658,6
Varianza	$S^2 = \frac{\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2}{N}$	8,73	282,93
Deviazione standard	$S = \sqrt{\frac{\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2}{N}}$	2,95	16,82
Coefficiente di variazione	$cv = \frac{s}{\bar{x}} \cdot 100$	10,31	58,71

Statistica descrittiva Monovariata

Momenti omogenei ed indici di forma

Momento omogeneo= media dei valori di una variabile presa con esponente positivo

r = ordine del momento

$$M = \frac{\sum_{i=1}^{N}\left(x_i\right)^r}{N} \qquad\qquad M = \frac{\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^r}{N}$$

Momento non centrale

non è uno scarto dalla media

Momento centrale

è uno scarto dalla media

Statistica descrittiva Monovariata

Momenti omogenei ed indici di forma

Ordine	tipo	Informazione	Esempio
1	non centrale	tendenza centrale della distribuzione	$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$
2	centrale	dispersione della distribuzione	$S^2 = \frac{1}{N}\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2$
3	centrale	Asimmetria	β_1
4	centrale	Curtosi	β_2

Statistica descrittiva Monovariata

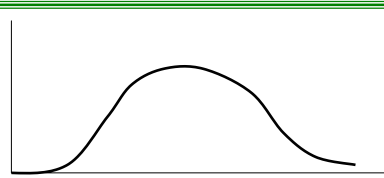
Momenti omogenei ed indici di forma

Asimmetria negativa

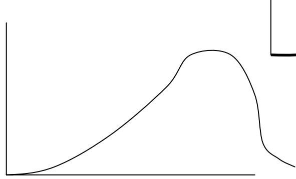
Asimmetria positiva

β_1

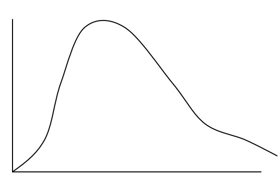
\- 0 +



$\bar{x} = mdn = \text{mod}$



$\bar{x} < mdn < \text{mod}$



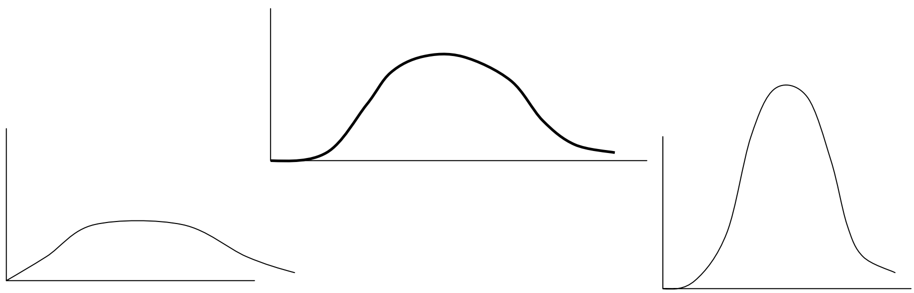
$\bar{x} > mdn > \text{mod}$

Statistica descrittiva Monovariata

Momenti omogenei ed indici di forma

Platicurtica Leptocurtica

β_2 - 0 +

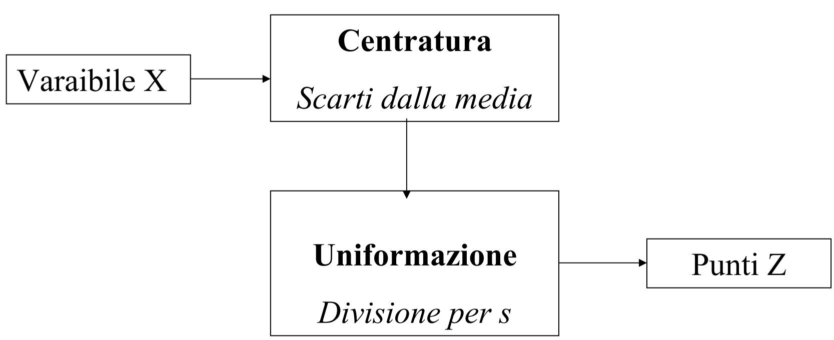


Statistica descrittiva Monovariata

Standardizzazione di una variabile cardianle

Dispositivo per rendere confrontabili distribuzioni diverse.

Le distribuzioni vengono trasformate in distribuzioni con media =0 e deviazione standard=1



Statistica descrittiva Monovariata

Standardizzazione di una variabile cardianle

Dispositivo per rendere confrontabili distribuzioni diverse.

Le distribuzioni vengono trasformate in distribuzioni con media =0 e deviazione standard=1

$$z_i = \frac{(x_i - \bar{x})}{s}$$

Statistica descrittiva Monovariata

Standardizzazione di una variabile cardianle

$$z_i = \frac{(x_i - \bar{x})}{s}$$

ident	altezza	peso	alezza z	peso z
1	1,72	72	-0,48	-0,07
2	1,65	65	-1,61	-0,71
3	1,81	93	0,96	1,83
4	1,75	61	0,00	-1,07
5	1,82	73	1,12	0,02
media	1,75	72,8		
s	0,062	11,034		