

Metodi statistici per le ricerche di mercato

Prof.ssa Isabella Mingo
A.A. 2016-2017



SAPIENZA
UNIVERSITÀ DI ROMA

Facoltà di Scienze Politiche, Sociologia, Comunicazione

Corso di laurea Magistrale in «Organizzazione e marketing per la comunicazione d'impresa»

Dalla matrice dei dati ... alle distribuzioni di frequenza

- **La matrice dei dati**
 - è anche detta distribuzione unitaria multipla poiché rappresenta le modalità di ciascuna variabile rilevate in ciascuna unità.
 - è una elencazione puntuale di tutti gli aspetti rilevati, **ma non consente di cogliere in modo sintetico le caratteristiche del fenomeno**
- **La distribuzione di frequenza** di ciascuna variabile
 - È l'elencazione del numero delle volte che una modalità (o un valore) di quella variabile carattere viene osservata nel collettivo.
 - Corrisponde alla suddivisione delle unità statistiche in sottogruppi.
 - Consente dunque di ottenere delle prime informazioni di sintesi

Dalla distribuzione unitaria a quella di frequenza

Cognome	sexo	età	Livello soddisfazione	Tipo albergo (categoria)	Giorni permanenza
Bianchi	F	28	basso	2	4
Alimov	F	52	medio	3	6
More	M	19	alto	4	2
Wieser	F	23	alto	3	10
Lincoln	M	60	medio	5	8

a.a. 2016-2017

Distribuzione di frequenza assoluta

- Elencazione del numero delle volte che una modalità di un carattere viene osservata nel collettivo.
- Corrisponde alla suddivisione delle unità statistiche in **sottogruppi**.

Sexo	frequenza
F	3
M	2
totale	5

Livello soddisfazione	frequenza
basso	1
medio	2
alto	2
Totale	5

Osservazioni:

- Evidenzia le caratteristiche del fenomeno
- Implica perdita di informazione

a.a. 2016-2017

Distribuzione di frequenza relativa e percentuale

- **Relativa** $f_j = n_j/n$
 - rapporto tra la frequenza assoluta e il numero totale delle unità osservate
- **Percentuale** $p_j = f_j * 100$
 - frequenza relativa moltiplicata per 100
 - $p_1 + p_2 + \dots + p_j + \dots + p_K = 100$

Sesso	Freq. assoluta
F	3
M	2
totale	5

Sesso	Freq. rel
F	$(3/5)=0.6$
M	$(2/5)=0.4$
totale	1

Sesso	Freq. %
F	$(3/5*100)=60$
M	$(2/5*100)=40$
totale	100

Es. calcolo

a.a. 2016-2017

Distribuzione di frequenza percentuale: esempio

Sesso	Freq. assol.
F	3
M	2
totale	5

Sesso	Freq. assol.
F	180
M	120
totale	300

Sesso	Freq. %
F	$(3/5*100)=60$
M	$(2/5*100)=40$
totale	100

Sesso	Freq.%
F	$(180/300*100)=60$
M	$(120/300*100)=40$
totale	100

• Attenzione: le distribuzioni percentuali implicano la perdita dell'informazione sulla numerosità dei collettivi

• Possibile il confronto tra due o più collettivi di dimensioni diverse rispetto ad un carattere

a.a. 2016-2017

Esercizio 1

Data la seguente distribuzione unitaria multipla, calcolare le distribuzioni di frequenze assolute relative e percentuali delle variabili : titolo di studio e facilità accesso.

	municipio di residenza	sezzo	eta	titolo di studio	facilità accesso
1	XVII MUNICIPIO	maschi	47	medio superiore	abbastanza gradito
2	XVII MUNICIPIO	femmine	21	medio superiore	poco gradito
3	XVIII MUNICIPIO	maschi	30	medio superiore	molto gradito
4	XVIII MUNICIPIO	maschi	22	laurea triennale	molto gradito
5	XVIII MUNICIPIO	femmine	23	laurea triennale	poco gradito
6	XVIII MUNICIPIO	femmine	37	laurea magistrale	molto gradito
7	XVI MUNICIPIO	maschi	84	medio inferiore	abbastanza gradito
8	XIX MUNICIPIO	maschi	18	medio inferiore	molto gradito
9	XIX MUNICIPIO	maschi	48	medio superiore	molto gradito
10	XIX MUNICIPIO	maschi	56	medio superiore	poco gradito
11	XIX MUNICIPIO	maschi	53	laurea magistrale	molto gradito
12	IV MUNICIPIO	femmine	44	laurea magistrale	molto gradito
13	IV MUNICIPIO	femmine	15	medio inferiore	molto gradito
14	II MUNICIPIO	maschi	22	medio superiore	molto gradito
15	II MUNICIPIO	femmine	53	laurea magistrale	molto gradito
16	XVI MUNICIPIO	femmine	40	laurea magistrale	per niente gradito
17	XVI MUNICIPIO	femmine	49	laurea magistrale	poco gradito
18	VI MUNICIPIO	femmine	38	medio superiore	abbastanza gradito
19	VI MUNICIPIO	maschi	43	laurea magistrale	poco gradito
20	I MUNICIPIO	maschi	26	laurea	abbastanza gradito

a.a

Esempio Chi sono gli utenti delle Biblioteche di Roma?

- Per rispondere a questa domanda è necessario trarre delle informazioni sintetiche dalla matrice dei dati, riguardanti gli utenti ed in particolare:
 - Il genere
 - La loro nazionalità
 - La loro età
 - Il loro titolo di studio
 - Il luogo di residenza
- Operativamente a partire dalla matrice originaria si dovranno ottenere distribuzioni di frequenze assolute e percentuali di utenti in base ad alcune caratteri di interesse.

2016-2017

Uso del software: distribuzioni di frequenza

I. Mingo 2016-2017

Sesso

	Frequency	Percent
Valid 1 maschi	632	48,9
2 femmine	661	51,1
Total	1293	100,0

Intervistati per sesso

Titolo di studio

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 1 fino licenza ele	10	,8	,8	,8
2 medio inferiore	132	10,2	10,2	11,0
3 medio superiore	650	50,3	50,4	61,4
4 laurea triennale	146	11,3	11,3	72,7
5 laurea magistra	287	22,2	22,2	95,0
6 post laurea	65	5,0	5,0	100,0
Total	1290	99,8	100,0	
Missing 0	3	,2		
Total	1293	100,0		

titolo di studio

I. Mingo 2016-2017

A proposito di ... Frequenze cumulate

- Si calcolano a partire dalle distribuzioni di frequenza (assoluta, relativa o percentuale) ordinate di mutabili ordinali.
- La frequenza per ogni data classe è ottenuta come somma della corrispondente frequenza e di tutte quelle precedenti

$$f_j = f_1 + f_2 + \dots + f_{j-1}$$

- Hanno senso solo se il carattere in esame è almeno un carattere qualitativo ordinale

Durata della vacanza	Freq.	Freq.Cum	%	% cum
Da 1 a 4 notti	60		25	25
Da 5 a 7 notti	70	130	29.17	(25+29.17)= 54.17
Da 8 a 14 notti	20	150	8.33	(54.17+8.33)= 62.5
Da 15 a 21 notti	60	210	25	(62.5+25)= 87.5
Oltre 21 notti	30	240	12.5	(87.5+12.5)= 100
Totali	240		100	

Osservazioni:

Le freq. cumulate consentono di individuare immediatamente il numero di unità che hanno soggiornato meno di un determinato numero di notti.

Es. il 62,5% ha soggiornato meno di 15 notti

a.a. 2016-2017

Esercizio

Nella tabella seguente sono riportati i livelli di soddisfazione di due campioni di consumatori, rilevati in due zone di vendita differenti rispetto a un nuovo prodotto immesso nel mercato.

In quale dei due campioni si registra una percentuale più elevata di clienti che esprimono livelli di soddisfazione inferiori al «medio»?

In quale dei due campioni si registra una percentuale più elevata di clienti che esprimono livelli di soddisfazione superiori al «medio»?

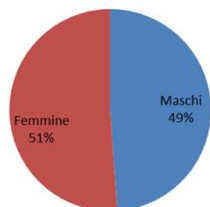
Livello	Campione Zona Nor	Campione Zona Sud
Molto Basso	40	48
Basso	32	28
Medio	53	44
Alto	24	20
Molto Alto	16	13
Totale	165	153

es

a.a. 2016-2017

A proposito di ... Grafici a torta

Intervistati per sesso

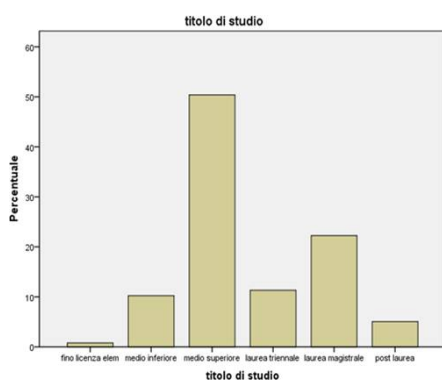


Caratteristiche:

- Ogni settore circolare rappresenta una modalità della variabile
- L'angolo al centro di ogni settore circolare corrisponde alla percentuale di ogni modalità
- $|360 * p_j / 100|$
- E' opportuno rappresentare distribuzioni con un numero di modalità non eccessivamente elevato, per garantire la leggibilità del grafico.

2016-2017

A proposito di ... Grafici a barre

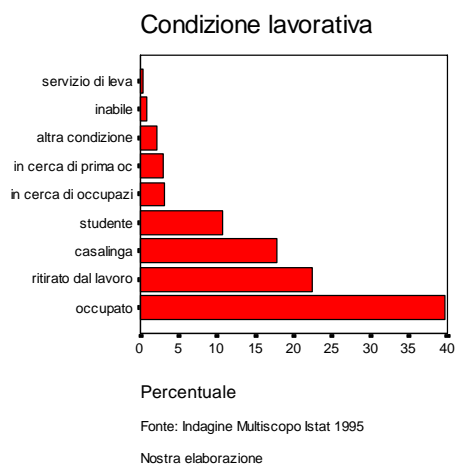


Caratteristiche:

- Ogni barra corrisponde ad una modalità delle variabile
- La base di ogni barra è uguale e arbitraria
- L'altezza è proporzionale alla percentuale
- La successione segue l'ordine delle modalità nel caso di caratteri ordinabili.

2016-2017

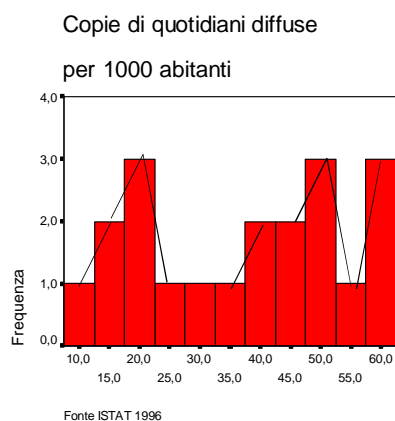
A proposito di ... Grafici a nastro



Caratteristiche:

- Ogni barra orizzontale corrisponde ad una modalità delle variabile
- Le basi delle barre sono uguali e arbitrarie
- La lunghezza è proporzionale alla percentuale
- La successione segue l'ordine delle frequenze per facilitare l'analisi.

A proposito di ... Istogrammi

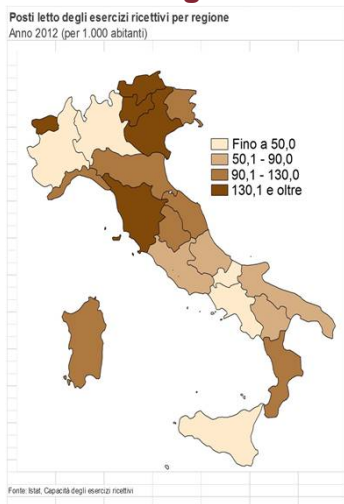


Caratteristiche:

-Adatto a caratteri quantitativi

- Ogni barra corrisponde ad una classe di valori della variabile
- Nell'esempio, le basi delle barre hanno uguale ampiezza come le classi di valori e l'altezza è proporzionale alla frequenza di ogni classe di valori.
- Per classi di ampiezza diversa la base di ciascuna barra dovrebbe essere proporzionale all'ampiezza della classe.
- La somma delle aree di tutti i rettangoli è uguale alla frequenza totale del carattere.
- Anche l'asse delle ascisse è graduato.
- Il poligono di frequenza si ottiene unendo con una linea spezzata i centri dei tetti di ogni rettangolo.

A proposito di ... Cartogrammi



- Si utilizzano per rappresentare serie territoriali
- Hanno come base una mappa in cui sono memorizzati i confini territoriali di ogni unità di analisi di cui si rappresenta la frequenza o l'intensità del carattere.
- Ogni area può essere colorata o retinata in base alla distribuzione di frequenza o all'intensità del carattere.

Sintesi della distribuzione: gli indici medi

Le distribuzioni delle variabili possono essere sintetizzate mediante un solo valore, calcolando un "indice medio" adeguato al tipo di variabile.

Tipo di variabile	Esempi	Indici medi di sintesi
Qualitativa sconnessa	-Sesso -Nazionalità -Professione	Moda
Qualitativa ordinale	-Titolo studio -Frequenza dell'acquisto	Moda Mediana
quantitativa	-Età -n. prodotti acquistati -spese effettuate	Moda Mediana Media aritmetica

Le medie

- ✓ La **media** deve essere
 - un valore omogeneo con i dati osservati
 - *compreso tra le modalità della distribuzione*, tra le modalità minima e massima se si è in presenza di una mutabile ordinale o di una variabile secondo il principio di Cauchy

- ✓ Si distinguono:
 - *medie di posizione*
 - *medie analitiche*

a.a. 2016-2017

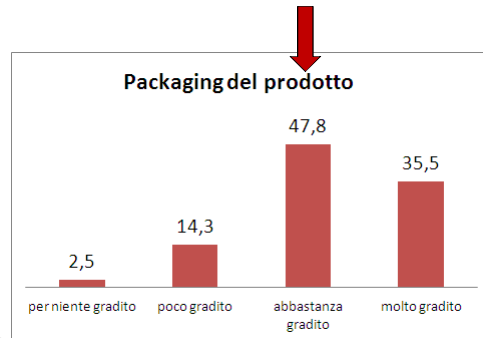
Medie di Posizione

- ✓ *Moda e Mediana*
- ✓ Si possono calcolare sia per caratteri qualitativi che quantitativi
- ✓ Modalità che occupano particolari posizioni all'interno della distribuzione del carattere

a.a. 2016-2017

La moda

- Si può calcolare sia per caratteri qualitativi che quantitativi
- E' la modalità della distribuzione che si presenta con la massima frequenza
- Può non essere unica
- Nelle rappresentazioni grafiche la moda è facilmente individuabile: rappresenta il picco della distribuzione
- Se la distribuzione del carattere è suddiviso in classi si considera la classe a cui corrisponde la massima *densità media di frequenza* (rapporto tra frequenza e ampiezza della classe).



Giudizio dei clienti sul packaging del prodotto

	Frequenza	Percentuale
per niente gradito	30	2,5
poco gradito	175	14,3
abbastanza gradito	584	47,8
molto gradito	434	35,5
Totale	1223	100,0

a.a. 2016-2017

Esempio: distribuzione multimodale

- ✗ Distribuzione dei clienti di un punto vendita per Brand preferito

Brand	n_j	f_j	p_j
Timberland	120	0,300	30,0
Lamberjack	120	0,300	30,0
Geox	70	0,175	17,5
Altro	90	0,225	22,5
Totale	400	1.000	100.0

La distribuzione è **bimodale**
Le mode sono
"Timberland" e "Lamberjack"

a.a. 2016-2017

Esempio: classe modale

- ✓ Distribuzione dei clienti per punteggio assegnato ad un servizio post-vendita.
- ✓ La classe modale è quella a cui corrisponde la max *densità media di frequenza*.

Punteggio	n_j
[87-98]	18
(98-102]	27
(102-105]	25
(105-109]	35
(109-111]	25
Totale	130

Qual è la classe modale?

Voto Laurea	n_j	a_j	$d_j = n_j / a_j$
[87-98]	18	11	1,64
(98-102]	27	4	6,75
(102-105]	25	3	8,33
(105-109]	35	4	8,75
(109-111]	25	2	12,50
Totale	130		

1-Ampiezza della classe:

$$a_j = x_j - x_{j-1}$$

2-Densità media:

$$d_j = n_j / a_j$$

La classe modale è
“(109-111)”

a.a. 2016-2017

Esercizio 2

Qual è il valore modale del titolo di studio e della soddisfazione per la facilità di accesso al punto vendita?

	municipio di residenza	sexso	eta	titolo di studio	facilità accesso
1	XVII MUNICIPIO	maschi	47	medio superiore	abbastanza gradito
2	XVII MUNICIPIO	femmine	21	medio superiore	poco gradito
3	XVIII MUNICIPIO	maschi	30	medio superiore	molto gradito
4	XVIII MUNICIPIO	maschi	22	laurea triennale	molto gradito
5	XVIII MUNICIPIO	femmine	23	laurea triennale	poco gradito
6	XVIII MUNICIPIO	femmine	37	laurea magistrale	molto gradito
7	XVI MUNICIPIO	maschi	84	medio inferiore	abbastanza gradito
8	XIX MUNICIPIO	maschi	18	medio inferiore	molto gradito
9	XIX MUNICIPIO	maschi	48	medio superiore	molto gradito
10	XIX MUNICIPIO	maschi	56	medio superiore	poco gradito
11	XIX MUNICIPIO	maschi	53	laurea magistrale	molto gradito
12	IV MUNICIPIO	femmine	44	laurea magistrale	molto gradito
13	IV MUNICIPIO	femmine	15	medio inferiore	molto gradito
14	II MUNICIPIO	maschi	22	medio superiore	molto gradito
15	II MUNICIPIO	femmine	53	laurea magistrale	molto gradito
16	XVI MUNICIPIO	femmine	40	laurea magistrale	per niente gradito
17	XVI MUNICIPIO	femmine	49	laurea magistrale	poco gradito
18	VI MUNICIPIO	femmine	38	medio superiore	abbastanza gradito
19	VI MUNICIPIO	maschi	43	laurea magistrale	poco gradito
20	I MUNICIPIO	maschi	28	laurea	abbastanza gradito

a.a. 2016-2017

La mediana

- Valore della distribuzione che divide il collettivo in due parti uguali: costituiti rispettivamente da modalità inferiori e superiori alla mediana stessa
- Media di posizione calcolabile solo se il carattere è ordinabile .
- La mediana rappresenta quella modalità rispetto alla quale la metà delle unità statistiche osservate presenta una modalità inferiore o uguale e l'altra metà una modalità superiore o uguale.
- Procedura di calcolo :
 - Ordinare in senso crescente la distribuzione unitaria
 - Individuare l'unità che occupa il posto centrale : se il collettivo è formato da un numero dispari di unità la posizione sarà $(n+1)/2$, se è pari si hanno due unità centrali $n/2$ e $n/2 +1$
 - Individuare il valore corrispondente all'unità centrale o alle due unità centrali |nel caso di caratteri quantitativi si considererà la semisomma
- E' una media più robusta della media aritmetica poiché è meno sensibile ai valori estremi

Mediana: come si calcola a partire da una distribuzione di frequenze

- Ordinare in senso crescente le modalità della variabile
 - individuare la modalità a cui è associata una frequenza cumulata almeno pari alla semisomma del collettivo;
- oppure
- Individuare la modalità a cui è associata una frequenza cumulata almeno pari alla metà del collettivo, o una percentuale cumulata almeno pari al 50%.

Giudizio dei clienti sul packaging del prodotto			
	Frequenza	Frequenza cumulata	Percentuale cumulata
per niente gradito	30	30	2,5
poco gradito	175	205	16,8
abbastanza gradito	584	789	64,5
molto gradito	434	1223	100,0
Totale	1223		

Il collettivo è formato da 1223 unità.
L'unità che occupa la posizione centrale è :
 $(1223+1)/2=612$

Le frequenze cumulate superano la semisomma del totale e le percentuali cumulate superano il 50% in corrispondenza della modalità **"abbastanza gradito"** che è pertanto la mediana della distribuzione.

Esercizio 3

- Si supponga di aver rilevato su 114 clienti il numero di prodotti acquistati da ciascuno nell'ultimo fine settimana.
- Qual è la mediana?

Numero di prodotti acquistati	Frequenza assoluta	Frequenza cumulata	% cumulata
7	18	18	15,79
8	15	33	28,95
9	22	55	48,25
10	7	62	54,39
11	41	103	90,35
12	11	114	100,00
Totale	114		

Il collettivo è costituito da $n=114$
i posti centrali sono $114/2=57$ e $(114/2)+1 = 58$

A entrambi è associata la stessa modalità che ha una % cumulata almeno pari al 50%.

La metà dei clienti ha acquistato un numero di prodotti non superiore a 10.

a.a. 2016-2017

I quartili

♦ quartili sono quei valori che dividono la distribuzione in quattro parti.

- Il primo quartile (25esimo percentile) è quel valore che nella graduatoria crescente ha a sinistra il 25% dei casi

$$Q_1 = x_{\left(\frac{n}{4}\right)}$$

- Il secondo quartile è la mediana perché ha a sinistra il 50% dei casi

- Il terzo quartile (75esimo) ha a sinistra il 75% dei casi

$$Q_3 = x_{\left(3 \times \frac{n}{4}\right)}$$

AA 2016-2017

Quartili : esempio

Possiamo affermare che il 25% degli utenti che hanno risposto alla domanda, hanno espresso una valutazione inferiore o pari a 5, il 50% inferiore o pari a 6; il 75%, inferiore o pari a 7

Valutazione ISS: rapidità soluzione problemi

		Frequenza	Percentuale	Percentuale valida	Percentuale cumulata
Validi	1	24	1,6	3,6	3,6
	2	20	1,3	3,0	6,6
	3	10	,7	1,5	8,1
	4	32	2,1	4,8	13,0
	5	96	6,4	14,5	27,4
	6	190	12,7	28,6	56,0
	7	134	8,9	20,2	76,2
	8	124	8,3	18,7	94,9
	9	18	1,2	2,7	97,6
	10	16	1,1	2,4	100,0
	Totale	664	44,3	100,0	
Mancanti	Mancante di sistema	836	55,7		
	Totale	1500	100,0		

1° quartile

mediana

3° quartile

Esercizio 4

Nella tabella seguente è presentata la distribuzione del numero di addetti rilevata sui 115 punti vendita della nostra azienda.

Quanti addetti ha il 25% dei punti vendita?

E il 50%?

E il 75%?

Numero di addetti	Frequenza
10	18
12	15
14	25
16	10
18	29
20	18
Totale	115

Glossario:

La media aritmetica

E' pari alla somma di tutti i valori di un carattere quantitativo divisa per il numero delle unità statistiche su cui tali valori sono stati rilevati

Cognome	Punteggio
Bianchi	98,7
Dandini	99,9
Moreno	80,2
Rossi	90,0
Valeri	96,5

$$M = (98,7 + 99,9 + 90,2 + 90 + 96,5) / 5 = 93,06$$

Osservazioni:

- La media aritmetica realizza l'equipartizione della variabile tra le unità della popolazione
- Risente fortemente dei valori estremi

a.a 2016-2017

Media aritmetica: il calcolo

1) Nel caso di una distribuzione unitaria semplice del carattere X

$$x_1, x_2, \dots, x_j, \dots, x_n$$

$$M = \frac{x_1 + x_2 + x_3 + \dots + x_j + \dots + x_n}{n} = \frac{\sum_{j=1}^n x_j}{n} = \frac{1}{n} \sum_{j=1}^n x_j$$

Cliente	Gradimento
1	98
2	100
3	70
4	72
5	70
6	100
7	85
8	65
9	60
10	88

$$N = 10$$

$$M = \frac{98 + 100 + 70 + 72 + 70 + 100 + 85 + 65 + 60 + 88}{10} = \frac{808}{10} = 80,8$$

$$\sum_{j=1}^{10} x_j = 98 + 100 + 70 + 72 + 70 + 100 + 85 + 65 + 60 + 88 = 808$$

a.a 2016-2017

Media aritmetica: calcolo

2) Nel caso di una distribuzione di frequenza

Consideriamo la distribuzione di frequenza dei clienti di un punto vendita per numero di richieste al servizio di assistenza

$$N = 283$$

$$M = \frac{\sum_{j=1}^7 x_j n_j}{n}$$

$$M = \frac{(1 \times 15) + (2 \times 43) + (3 \times 103) + (4 \times 80) + (5 \times 32) + (6 \times 8) + (7 \times 2)}{283} = \frac{952}{283} = 3.36$$

Num. Richieste (x _j)	Freq (n _j)
1	15
2	43
3	103
4	80
5	32
6	8
7	2
Totale	283

a.a 2016-2017

Esercizio 5

- Data la distribuzione unitaria dell'esercizio 1, calcolare l'età media dei clienti.
- Data la distribuzione di frequenza dell'esercizio 3, calcolare la media aritmetica della variabile numero di prodotti acquistati.

a.a. 2016-2017

Es. Calcolo

Media aritmetica: il caso di variabili quantitative raggruppate in classi

Classi età	frequenza
[20, 25)	416
[25, 35)	792
[35, 45)	783
[45, 55)	670
[55, 65)	537
[65, 80]	392
Totale	3590

Nel caso in cui si disponga della distribuzione semplice di frequenze assolute di una variabile i cui valori sono raggruppate in K classi, come calcolare la media?

Occorre innanzitutto individuare il **valore centrale della classe** che si ottiene come semisomma degli estremi della classe

Clienti per classe di età

Classi di età	Frequenze assolute n_i	Semisomma estremi $(c_i + c_{i-1})/2$	Valori centrali c_i
[20, 25)	416	$(25+20)/2$	22,5
[25, 35)	792	$(35+25)/2$	30
[35, 45)	783	$(45+35)/2$	40
[45, 55)	670	$(55+45)/2$	50
[55, 65)	537	$(65+55)/2$	60
[65, 80]	392	$(80+65)/2$	72,5
Totale	3590		



$$M = \frac{\sum_{i=1}^K c_i n_i}{n}$$

a.a 2016-2017

Esercizio 6

Nella tabella seguente è riportato il fatturato annuo di alcuni punti vendita.

Quale è stato il fatturato medio ?

Fatturato per punti vendita (migliaia di euro)	
	n_i
[400-600)	10
[600-1000)	16
[1000-1500)	8
[1500-2500]	6
	40

a.a. 2016-2017

Stesso valore medio per distribuzioni diverse

u_i	X_1	X_2	X_3
A	22	1	21
B	22	8	21
C	20	10	21
D	23	32	21
E	19	34	21
F	20	41	21

- Un uguale valore medio può sintetizzare distribuzioni molto diverse tra loro
- Le tre distribuzioni hanno la medesima media aritmetica, ma la tendenza di ogni unità ad assumere valori diversi dalla media è differente in ciascuna distribuzione

M	21	21	21
---	----	----	----

a.a. 2016-2017

Caratteri quantitativi La variabilità

La variabilità o la dispersione di una distribuzione esprime la tendenza dei caratteri o dei fenomeni ad assumere differenti valori o determinazioni.

Requisiti di un indice di variabilità:

- assume valore minimo se tutte le unità presentano uguale modalità del carattere
- aumenta all'aumentare della diversità tra modalità

Attenzione: ogni indice di variabilità esprime un concetto diverso pertanto non è corretto confrontare la variabilità ottenuta con indici diversi

a.a. 2016-2017

Caratteri quantitativi Scostamento semplice medio

Si calcola sommando le differenze in **valore assoluto**, cioè non tenendo conto del segno positivo o negativo, tra ogni valore della distribuzione e il valore medio e dividendo la somma per la numerosità della distribuzione.

$$s = \frac{\sum_{j=1}^n |x_j - M|}{n}$$

A seconda della media scelta si può ottenere uno specifico scostamento medio per esempio dalla media aritmetica o dalla mediana.

Per qualsiasi distribuzione di frequenze lo scostamento semplice dalla mediana è sempre minore o uguale allo scostamento dalla media aritmetica

a.a. 2016-2017

Caratteri quantitativi Scostamento semplice medio: calcolo

v_i	X_{11}	X_{12}	X_{13}
A	22	1	21
B	22	8	21
C	20	10	21
D	23	32	21
E	19	34	21
F	20	41	21
M	21	21	21

- Per la prima distribuzione si avrà:

$$S = [|22-21| + |22-21| + |20-21| + |23-21| + |19-21| + |20-21|] / 6 = 1,33$$

- Per la seconda distribuzione si avrà:

$$S = [|1-21| + |8-21| + |10-21| + |32-21| + |34-21| + |41-21|] / 6 = 14,77$$

- Per la terza distribuzione si avrà

$$S = [|21-21| + |21-21| + |21-21| + |21-21| + |21-21| + |21-21|] / 6 = 0$$

a.a. 2016-2017

Caratteri quantitativi Varianza

E' la media dei quadrati degli scarti dalla media aritmetica
si calcola sommando gli scarti elevati al quadrato e dividendoli per la
numerosità della distribuzione.

$$\sigma^2 = \frac{\sum_{j=1}^n (x_j - M)^2}{n}$$

Osservazioni: L'elevazione a quadrato trasforma tutte le differenze negative in positive e mette in maggiore risalto le differenze grandi rispetto a quelle piccole.

La varianza non possiede la stessa unità di misura dei valori della distribuzione

- La radice quadrata della varianza è la deviazione standard o scarto quadratico medio
- Il numeratore della varianza è detto devianza

Esercizio

Nell'ambito di una indagine di mercato si sono rilevati i costi giornalieri per una camera doppia standard in alcuni Hotel della zona (tab.seguente).

Calcolare il costo medio e lo scarto quadratico medio.

Unità statistica	Costo x_j	$(x_j - M)$	$(x_j - M)^2$
Albergo A	90		
Albergo B	75		
Albergo C	135		
Albergo D	83		
Albergo E	105		
Albergo F	90		
Albergo G	115		

Calcolo

Caratteri quantitativi Coefficiente di variazione

- Indice di variabilità percentuale, non influenzato dall'unità di misura e dall'ordine di grandezza dei dati.
- Rapporto tra lo scarto quadratico medio e la media aritmetica x100.

$$Cv = \sigma / M * 100$$

- Consente di confrontare fenomeni che presentano diverse unità di misura e/o sono rilevati su collettivi di diversa numerosità
- Ci indica se una distribuzione è più variabile di un'altra

Caratteri quantitativi Coefficiente di variazione esercizio

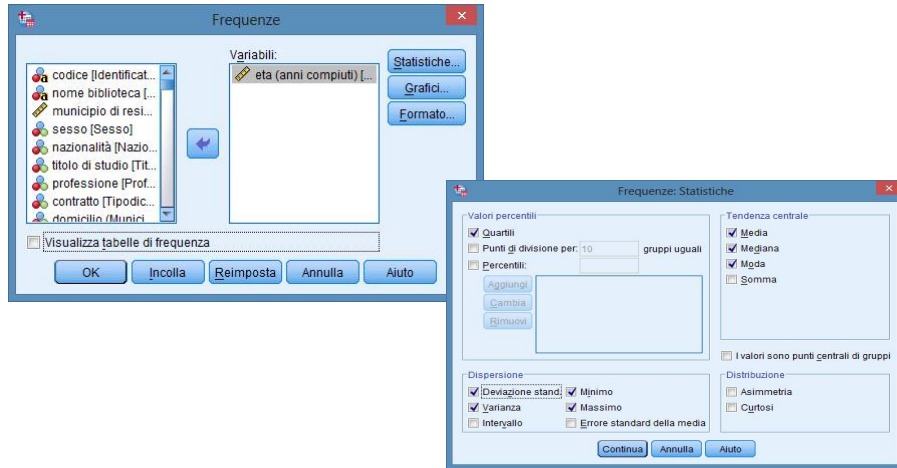
- Confrontiamo i volumi di vendita di due prodotti.
 - Il prodotto A nei diversi punti vendita rilevati ha avuto un volume di vendite medio di 750 euro e una deviazione standard (σ) pari a 150 euro.
 - Il prodotto B ha avuto un volume di vendite medio di 487 euro e una deviazione standard pari a 115 euro.
- Quale prodotto registra una maggiore variabilità del volume di vendite?

$$Cv = \sigma / M * 100$$

	σ	M	CV*100
Prodotto A	150,00	750,00	20,00
Prodotto B	115,00	487,00	23,61

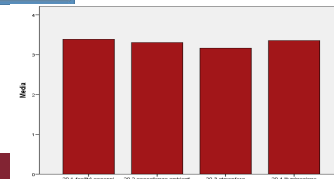
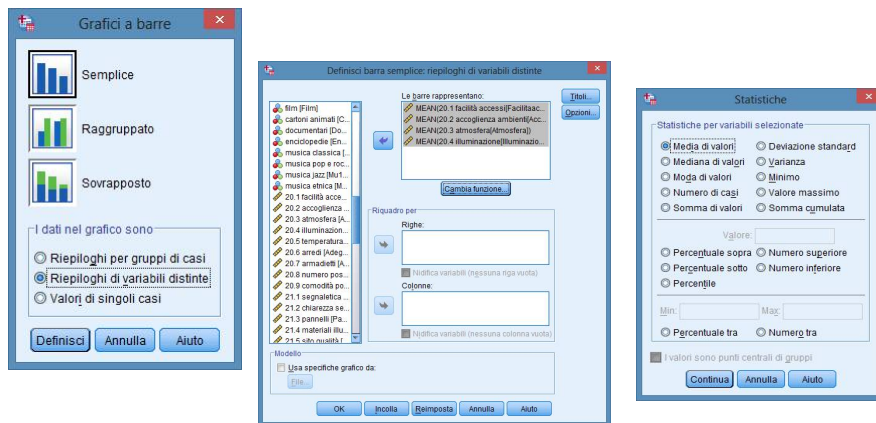


Uso del software :
indici medi e di variabilità



I. Mingo 2016-2017

Uso del software :
confrontare variabili



I. Mingo 2016-2017