STATISTICA SOCIALE

corso di laurea in
SCIENZE TURISTICHE

A.A. 2007/2008

Docente: *Carla Rampichini*

carla@ds.unifi.it

Programma del corso

- La statistica e i suoi obiettivi.
- Rilevazione dei dati.
- Variabili statistiche univariate.
- Il sistema statistico del turismo.
- Principali indicatori del fenomeno turistico.
- L'indagine ad hoc.
- Indagini campionarie ISTAT in materia di viaggi e vacanze.

Testi consigliati

Pasetti P. (2002) *Statistica del turismo,* Roma: Carocci

Testi di utile consultazione

- Bailey K.D. (1995) *Metodi della ricerca sociale*, Bologna: Il Mulino. Capitoli V-VIII e XIII-XV.
- Fortini M. (2000), *Linee guida metodologiche per rilevazioni statistiche*, ISTAT.
- Lavine D. M. *et al.* (2002) Statistica, Milano: Apogeo.
- Tassinari G. (1997), *Appunti di Statistica del turismo*, Bologna: CLUEB.

Esame e prove intermedie

Esame: prova scritta + prova orale

Per gli studenti frequentanti le prove intermedie costituiscono parte integrante dell'esame di profitto.

Pagina WEB del corso:

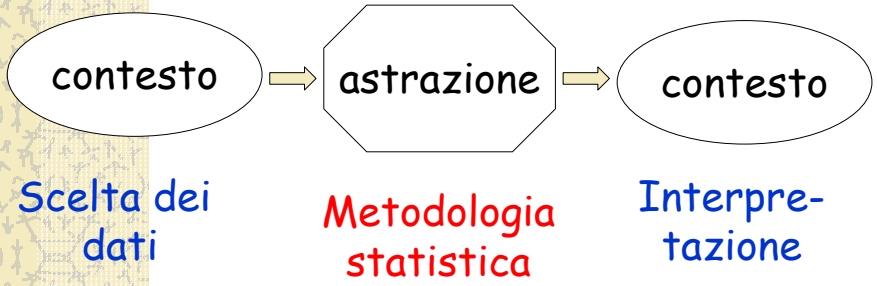
www.ds.unifi.it/rampi/statistica_sociale2006_07.htm

Si accede da:
http://www.ds.unifi.it/didattica/mat_didat.htm

Cos'è la Statistica?

- Trattamento dell'informazione in forma quantitativa.
- Analisi e interpretazione dei dati attraverso concetti e strumenti di tipo analitico.

Cos'è la Statistica?

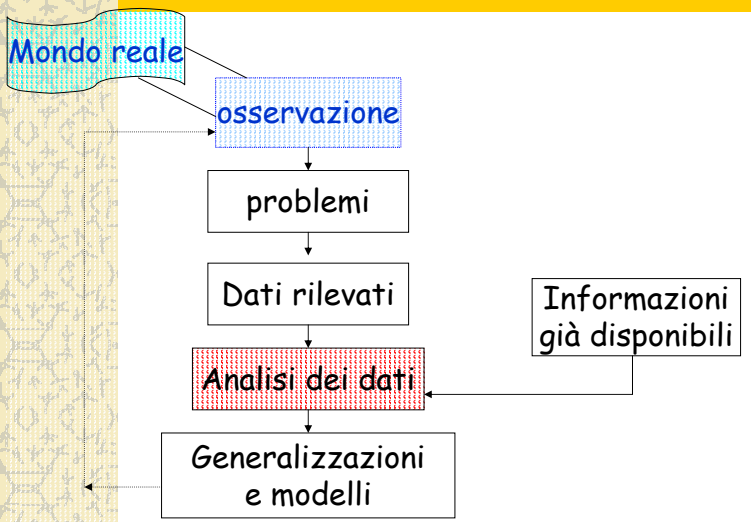


Apprendimento e valutazione del metodo prescindono dal tipo di applicazione

STATISTICA

- sec. XVII e XVIII: misurazione "ricchezza" e "potenza" dello Stato;
- Nasce anche Calcolo delle Probabilità.
- Statistica descrittiva (Censimenti) verso la seconda metà del XIX secolo;
- Inferenza Statistica inizi '900 (Fisher 1920), sviluppi teorici dopo 2a guerra mondiale.

Procedimento di ricerca



Statistica e analisi dei dati

Punto di partenza: matrice dei dati

Controllo dei dati

- Valori anomali o errati;
- Dati mancanti.

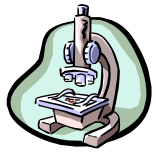
Elaborazione delle informazioni

- Relazioni tra caratteri;
- Somiglianza tra caratteri e/o unità;
- Rappresentazione caratteri e/o unità in uno spazio ridotto.

STATISTICA

Matematica applicata volta allo studio di dati ottenuti per:

(a) sperimentazione;

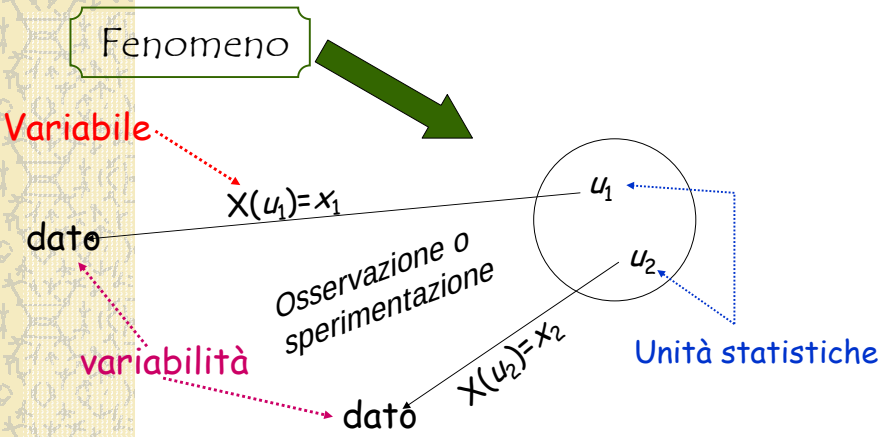


(b) osservazione.

Alcuni termini 'statistici'

- Popolazione di riferimento: insieme di individui che interessa studiare
- Unità statistiche: elementi della popolazione di riferimento
- Dati: risultato della rilevazione-misurazione di caratteristiche delle unità statistiche
- Variabili: caratteristiche rilevate/misurate sulle unità statistiche
- Modalità: valori distinti assunti da una variabile
- Campione: sottoinsieme della popolazione oggetto della rilevazione

Unità statistiche, popolazioni, variabili



Esempio: analisi del comportamento dei turisti stranieri

- Popolazione oggetto di indagine: turisti stranieri presenti nel periodo giugno-agosto 2004
- Variabili di interesse:
  - Giorni di permanenza secondo l'esercizio ricettivo
  - Km percorsi
  - Spesa turistica complessiva

Tipologie di dati

- Dati di stato (stock): si riferiscono ad un preciso istante di tempo (p.e. popolazione censita al 31-12-2004)
- Dati di flusso: riferiti ad un intervallo di tempo (p.e. presenze turistiche a Firenze durante il 2004). Sono i più frequenti nelle statistiche del turismo.

Dati di stock: esercizi ricettivi

Prospetto 1 - Capacità degli esercizi ricettivi (a) - Anni 1999-2002 (valori assoluti)

ANNI	Esercizi alberghieri		Esercizi complementari		Bed and breakfast (b)		Totale esercizi	
	Numero	Letti	Numero	Letti	Numero	Letti	Numero	Letti
1999	33.341	1.807.275	35.856	1.816.616			69.197	3.623.891
2000	33.361	1.854.101	83.858	2.055.897			117.219	3.909.998
2001	33.421	1.891.281	94.860	2.133.049			128.281	4.024.330
2002	33.411	1.929.544	75.966	2.150.643	4.338	19.398	113.715	4.099.585

(a) (b) I dati si riferiscono alla consistenza di fine anno 2002. A partire dalla rilevazione relativa all'anno 2002 è stata inserita una nuova sezione "Alloggi privati in affitto" nell'ambito della quale sono contemplati i bed and breakfast.

I FLUSSI NELLE STRUTTURE RICETTIVE

Nel 2004 gli alberghi e le strutture complementari operanti in Italia hanno registrato 82 milioni e 968 mila arrivi, per un totale di 336 milioni e 843 mila presenze. Rispetto al 2003 gli arrivi sono aumentati dello 0,3% mentre le giornate di presenza sono diminuite del 2,2%.

Tabella 1 – Arrivi e presenze di italiani e stranieri negli esercizi ricettivi. Anni 2003-2004
(valori in migliaia e variazioni percentuali)

		2003		2004	
MOVIMENTO CLIENTI		Numero	Variaz. % su 2002	Numero	Variaz. % su 2003
ITALIANI	Arrivi	47.719	4,5	47.498	-0,5
	Presenze	204.760	2,5	199.618	-2,5
	Perman. media	4,29	-1,8	4,20	-2,1
STRANIERI	Arrivi	35.006	-3,7	35.470	1,3
	Presenze	139.653	-4,1	137.225	-1,7
	Perman. media	3,99	-0,3	3,87	-3,0
TOTALE	Arrivi	82.725	0,8	82.968	0,3
	Presenze	344.413	-0,2	336.843	-2,2
	Perman. media	4,16	-1,2	4,06	-2,4

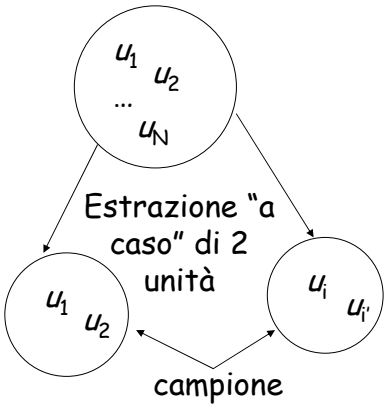
Fonte: Istat, Rilevazione sul movimento dei clienti negli esercizi ricettivi – Dati 2004 provvisori.

Come si ottengono le informazioni?

- Consultazione di esperti (es. metodo Delphi)
- Analisi di dati esistenti (fonti statistiche)
- Rilevazioni statistiche *ad hoc*

Esperimenti di campionamento e campioni

Popolazione reale
N "grande"



Esperimento di campionamento casuale:

- senza reinserimento
- semplice (con reinserimento)

Il campione non è una riproduzione fedele della popolazione

⇓

Inferenza statistica

Esempio di esperimento campionario

Quante persone scrivono con la mano sinistra?

N molto grande; M= n. di "mancini"

?? p=M/N ??

Campione di dimensione n=100

$\hat{p}$ =4/100=0.04

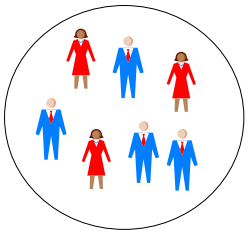
$p \neq \hat{p}$ per errore di campionamento

Quanto $p \neq \hat{p}$? Inferenza statistica

$p \in [0.01;0.08]$

Modelli statistici descrittivi e inferenziali

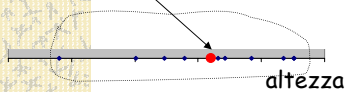
La statistica studia la variabilità attraverso 'modelli' matematici



1 variabile (es. altezza)

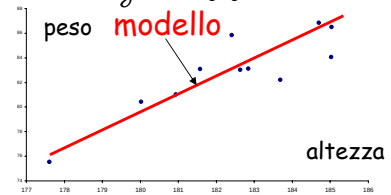
Quanto ci si discosta da un valore 'tipico'

modello



2 variabili (altezza-peso)

Quanto ci si discosta da una relazione sistematica



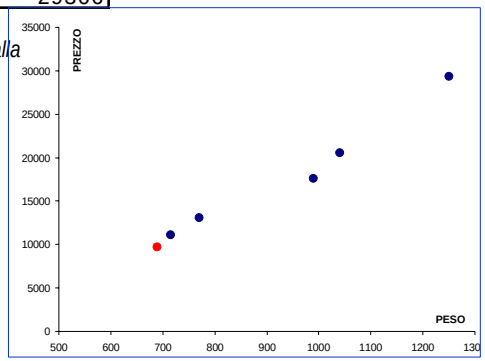
Esempio: le automobili si vendono a peso?

Peso e prezzo dei modelli base della FIAT, marzo 1993

Autovettura	PESO (kg)	PREZZO (mil lire)
Cinquecento 700 ED	690	9697
Panda 1.0 i.e. L.	715	11071
Uno Fire 1.0 i.e 3P	770	13041
Tipo 1.4 i.e.	990	17580
Tempra 1.4 e.e.	1040	20549
Croma 2.0	1250	29366

Fonte: Gente Motori, febbraio 1993

PACE L. e SALVAN A., 1996, *Introduzione alla Statistica - I Statistica Descrittiva*, CEDAM, Padova



Peso e prezzo dei modelli base della FIAT, marzo 1993

Autovettura	PESO (kg)	PREZZO (mil lire)
Cinquecento 700 ED	690	9697
Panda 1.0 i.e. L.	715	11071
Uno Fire 1.0 i.e 3P	770	13041
Tipo 1.4 i.e.	990	17580
Tempra 1.4 e.e.	1040	20549
Croma 2.0	1250	29366

Esempio: le automobili si vendono a peso? (continua)

PREZZO=a+b*PESO+ERRORE

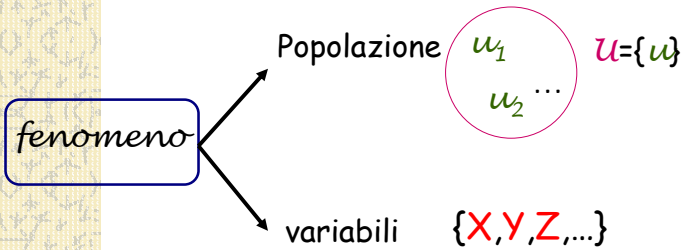
parte sistematica

scostamento dovuto ad altri 'fattori'

Definizione e rilevazione dei dati (1)

Definizione dei dati

- Individuazione del fenomeno
- Individuazione della popolazione di riferimento e delle unità che la compongono
- Individuazione delle variabili rilevanti



Rilevazione dei dati

- Predisposizione lista unità da rilevare
- Predisposizione del questionario
- Istruzione dei rilevatori (intervistatori)

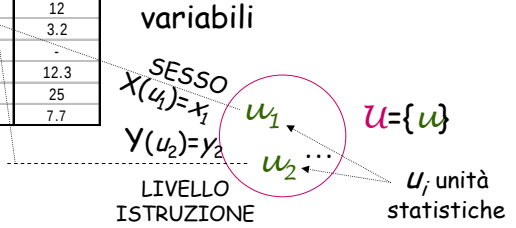
Piano di rilevazione

$U=\{u\}$ nota

Rilevazione materiale

unità u	etichetta	SESSO	ETA' (a.c.)	LIVISTR	DIST(KM)
Alpio	1	M	28	2	5
Caio	2	M	17	4	7.5
Prima	3	F	20	4	12
Velio	4	M	32	2	3.2
Rufa	5	F	16	1	-
Sesto	6	M	34	2	12.3
Beowulf	7	M	18	1	25
Sebaste	8	F	25	2	7.7

SPOGLIO



Esempio matrice dei dati

Etichette di identificazione

VARIABILI

unità u	etichetta	SESSO	ETA' (a.c.)	LIVISTR	DIST(KM)
Alpio	1	M	28	2	5
Caio	2	M	17	4	7.5
Prima	3	F	20	4	12
Velio	4	M	32	2	3.2
Rufa	5	F	16	1	-
Sesto	6	M	34	2	12.3
Beowulf	7	M	18	1	25
Sebaste	8	F	25	2	7.7

Modalità delle variabili

LIVISTR: 1=LIC. Elem., 2=Lic. Media, 3=Diploma, 4=Laurea

Ogni riga della matrice rappresenta una unità statistica

segue esempio matrice dei dati

unità u	etichetta	SESSO	ETA' (a.c.)	LIVISTR	DIST(KM)
Alpio	1	M	28	2	5
Caio	2	M	17	4	7.5
Prima	3	F	20	4	12
Velio	4	M	32	2	3.2
Rufa	5	F	16	1	
Sesto	6	M	34	2	12.3
Beowulf	7	M	18	1	25
Sebaste	8	F	25	2	7.7

Importanza delle etichette

Controllare coerenza dei dati!

Dato mancante (missing)

Attenzione alla qualità dei dati!!

Dal questionario alla matrice dei dati

anno_	gener	scuola	anno_cc	CFU	soddis	vaca	organi	durata	stato	mezzo	tipo	allogg	spesa
82	1	1	2	30	3	1	2	7	1	3	1	12	500
85	2	5	2	33	3	1	2	7	2	1	1	1	650
85	2	5	2	33	3	1	2	7	2	1	1	1	650
69	2	6	.	27	3	1	2	14	2	1	1	1	1500

La matrice dei dati

$$\mathbf{X}=\{x_{ij}\}$$

Raccoglie i risultati codificati di una rilevazione statistica

- p colonne ➔ variabili $\underline{X}_j$, j=1,2,…,p
- n righe ➔ unità statistiche $\underline{x}_i$, i=1,2,…,n

x_{ij} valore della variabile $\underline{X}_j$ misurata sull'unità $\underline{x}_i$

Variabili e modalità

Variabile
caratteristica delle unità statistiche che al variare delle unità può assumere almeno due valori

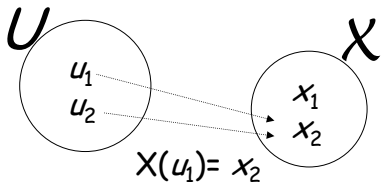
Modalità
Valori assumibili da una variabile (generalmente note a priori). L'insieme di tali valori è detto INSIEME delle MODALITÀ della variabile

Notazione
X, Y, Z …
$X_1, X_2, \ldots, X_p$

x,y,z,… modalità

X, Y, Z insieme delle modalità

$X(u): U \rightarrow X$



CLASSIFICAZIONE DELLE VARIABILI

Variabile qualitativa
Modalità espresse da nomi, aggettivi, attributi

Variabile quantitativa
Modalità espresse da numeri

- QUALITATIVE
  - SCONNESSE o NOMINALI
    Modalità non ordinabili
    SESSO
    RESIDENZA
    PROFESSIONE
  - ORDINALI
    Modalità ordinabili
    LIV. ISTRUZIONE
    VOTAZIONE LIC. MEDIA
- QUANTITATIVE
  - DISCRETE
    Insieme modalità finito o numerabile
    N. FIGLI
    N. DI STANZE
    N. AUTO
  - CONTINUE
    Insieme modalità infinito non numerabile
    ETÀ
    ALTEZZA
    PESO

SCALE DI MISURAZIONE DEI CARATTERI

- *SCALA NOMINALE*

$$x_i = x_j, \quad x_i \neq x_j$$

- *SCALA ORDINALE*

$$x_i = x_j, \quad x_i \prec x_j, \quad x_i \succ x_j$$
$$x_i \succ x_j \text{ e } x_j \succ x_k \Rightarrow x_i \succ x_k, \quad \text{proprietà transitiva}$$

- *SCALA D'INTERVALLI*
  Fissare unità di misura e **origine** del sistema

$$(x_i - x_j) = (x_k - x_h), \quad (x_i - x_j) > (x_k - x_h), \quad (x_i - x_j) < (x_k - x_h)$$

- *SCALA DI RAPPORTI*
  Fissare unità di misura, 0=assenza fenomeno

$$(x_i / x_j) = (x_k / x_h), \quad (x_i / x_j) > (x_k / x_h), \quad (x_i / x_j) < (x_k / x_h)$$

Trasformazione di variabili quantitative

I numeri corrispondenti alle modalità di variabili quantitative possono essere trasformati: x_i equivalenti a x^*_i

- Trasformazione lineare crescente per scala di intervalli

$$x_i^* = \underset{\text{origine}}{a} + \overset{\text{unità di misura}}{b} x_i$$

- Trasformazione lineare omogenea crescente per scala di rapporti

$$x_i^* = \overset{\text{unità di misura}}{b} x_i$$

Proprietà delle trasformazioni

Additività (vale solo per scale di rapporti!)

- Scala di rapporti

$$x_1^* + x_2^* = bx_1 + bx_2 = b(x_1 + x_2) = (x_1 + x_2)^*$$

- Scala di intervalli

$$x_1^* + x_2^* = (a + bx_1) + (a + bx_2) = 2a + b(x_1 + x_2) \neq (x_1 + x_2)^*$$

Invarianza

- Scala di rapporti

$$\frac{x_1^*}{x_2^*} = \frac{bx_1}{bx_2} = \frac{x_1}{x_2}$$

- Scala di intervalli

$$\frac{x_1^* - x_2^*}{x_3^* - x_4^*} = \frac{a + bx_1 - a - bx_2}{a + bx_3 - a - bx_4} = \frac{x_1 - x_2}{x_3 - x_4}$$

Trasformazioni e ricodifiche

Quantitativa continua Y=(100.210, 210) → Quantitativa discreta X=(101, 102, …)

X=101 se Y in (100,101]

Z=basso se Y in (100,150]
Z=medio se Y in (150,175]
Z=alto se Y in (175,210]

Attenzione: perdita di informazione!!!

Qualitativa ordinale Z=(basso,medio,alto) → Qualitativa nominale W=(normale, altro)

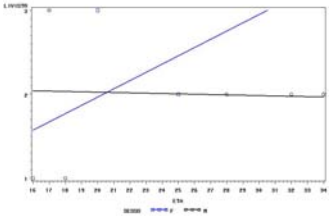
W=normale se Z=medio
W=altro se Z=(alto,basso)

Le analisi statistiche

Univariate studio della variabilità marginale

l'età media è 23.75 anni…

unità u	etichetta	SESSO	ETA' (a.c.)	LIVISTR	DIST(KM)
Alpio	1	M	28	2	5
Caio	2	M	17	4	7.5
Prima	3	F	20	4	12
Velio	4	M	32	2	3.2
Rufa	5	F	16	1	-
Sesto	6	M	34	2	12.3
Beowulf	7	M	18	1	25
Sebaste	8	F	25	2	7.7



Bivariate

l'età media dei maschi è più elevata di quella delle femmine

Multivariate

Al crescere dell'età il livello di istruzione aumenta per le femmine e diminuisce per i maschi

studio della variabilità congiunta

Tabelle di frequenza

Un esempio di matrice dei dati

Unità u	Etichetta	SESSO	ETA'	LIVIST	DIST
Alpio	1	M	28	2	5
Caio	2	M	17	4	7.5
Prima	3	F	20	4	12
Velio	4	M	32	2	3.2
Rufa	5	F	16	1	-
Sesto	6	M	34	2	12.3
Beowful	7	M	18	1	25
Sebaste	8	F	25	2	7.7

Dati in forma grezza

M, M, F, M, F, M, M, F

Tabelle di frequenza

Distribuzione per sesso

SESSO	frequenza
M	5
F	3
Totale	8

Distribuzione per livello di istruzione

LIVISTR	frequenza
1 Lic.elementare	2
2 Lic. Media	4
3 Maturità	0
4 Laurea	2
Totale	8

Definizione Se X è una variabile e x_j è una modalità di X si dice **frequenza assoluta** di x_j il n. di volte che x_j risulta osservata. Si indica con n_j

Definizione L'insieme delle coppie (x_j,n_j) si dice **distribuzione di frequenza assoluta** o **serie statistica**

Indagine sul turismo montano

Alcune delle caratteristiche rilevate

- Età in anni compiuti alla data dell'intervista
- Sesso
- È la prima volta che viene in questa località? (N=no,S=sì)
- Tipo di alloggio
- Numero di pernottamenti
- Pratica dello sci
- Spesa complessiva pro-capite in migliaia di lire

Dal questionario alle tabelle

Matrice dei dati da indagine sul turismo montano (944 turisti)

Obs	ETA	SESSO	PRIMA	ALLOGGIO	PERNOTTO	SCI	SPESA
1	32	F	S	3	4	2	200
2	38	M	S	3	10	4	400
3	30	F	N	2	3	3	250
4	33	F	N	11	2	1	30
5	40	F	N	6	7	4	300
6	35	M	N	7	2	3	100
7	25	M	N	6	2	3	100
8	23	F	N	3	7	3	300
9	51	M	N	2	6	1	600
10	23	F	S	11	3	3	70
…							
943	28	M	N	3	3	4	0
944	29	M	S	3	14	3	900

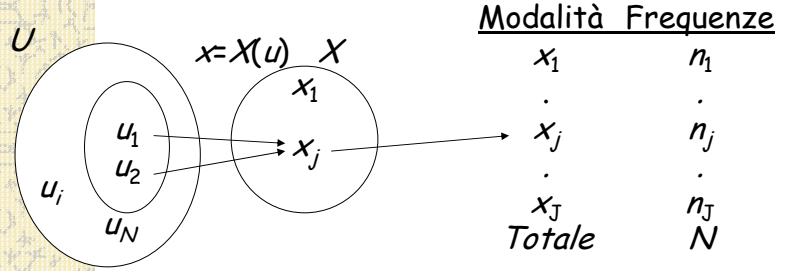
Operazioni di spoglio: dalla matrice dei dati alle tabelle

Turisti per sesso

SESSO	Frequeza	Freq. %
F	470	49.79
M	474	50.21

Distribuzioni di frequenza

n_j numero di unità statistiche che presentano la modalità $X = x_j$



$|U| = N = \Sigma n_j$ *numerosità della popolazione*

Frequenze assolute e relative

Si dice FREQUENZA RELATIVA di una modalità x_j, o di una classe di modalità $(x_{j-1};x_j)$ e si indica con f_j, la frazione o proporzione di u.s. che presentano tale modalità.

$$f_j = \frac{n_j}{\sum_{j=1}^{J} n_j} = \frac{n_j}{N} \quad j = 1,2,\ldots,J$$

Proprietà:

$$0 \le f_j \le 1 \quad j = 1,2,\ldots,J$$

$$\sum_{j=1}^{J} f_j = 1$$

Finalità delle frequenze relative

Facilitare la percezione del PESO delle modalità

Sesso	Freq. Assoluta	Freq. Relativa	Freq. Rel. %
M	1750	0.583	58.3
F	1250	0.417	41.7
Totale	3000	1	100

Facilitare CONFRONTI tra popolazioni

Sesso	Freq. Assoluta		Freq. Rel. %	
	Pop. A	Pop. B	Pop. A	Pop. B
M	1750	850	58.3	85.0
F	1250	150	41.7	15.0
Totale	3000	1000	100	100

Pratica dello sci: decodifica delle modalità

SCI	Frequenza	Freq.%
SOLO ALPINO	540	57.82
SOLO FONDO	88	9.42
ENTRAMBI	134	14.35
NESSUNO	172	18.42

Valori mancanti = 10

Per una migliore lettura: decodifica delle modalità!

Tipo di alloggio

Molte modalità: per una migliore lettura decodifica e raggruppamento delle modalità!

ALLOGGIO	Frequenza	Freq.%
1	46	4.89
2	86	9.15
3	166	17.66
4	80	8.51
5	14	1.49
6	182	19.36
7	200	21.28
9	4	0.43
10	2	0.21
11	56	5.96
12	30	3.19
13	2	0.21
14	18	1.91
15	52	5.53
20	1	0.11
40	1	0.11

Valori mancanti = 4

Tipo di alloggio

ALLOGGIO	Frequenza	Freq.%
ALBERGHIERO	378	40.21
EXTRA - ALBERGHIERI	230	24.47
CASA PROPRIETA'	260	27.66
ALTRO	18	1.91
PENDOLARI	52	5.53
20	1	0.11
40	1	0.11

20 e 40 codici errati!!

Frequency Missing = 4

Decodifica e raggruppamento delle modalità

NB: decodifica consente di individuare codici errati.

Frequenze cumulate e funzione di ripartizione

- Frequenze cumulate: somma delle frequenze sino alla modalità considerata
- Funzione di ripartizione della v.s. quantitativa X

$F(x_j)$=proporzione di u.s. con valori di $X \leq x_j$

$$F(x_j)=\sum_{k=1}^{j} f_k = \Pr\{X \leq x_j\}$$

Età all'intervista

ETA	Frequenza	Freq.%	Freq. Cumulata	Freq.% Cumulata
5	2	0.21	2	0.21
14	2	0.21	4	0.42
15	6	0.64	10	1.06
16	18	1.91	28	2.97
17	16	1.69	44	4.66
18	24	2.54	68	7.20
19	24	2.54	92	9.75
…				
66	2	0.21	938	99.36
72	2	0.21	940	99.58
73	4	0.42	944	100.00

Per una migliore lettura: definizione delle classi di modalità!

Età all'intervista: raggruppamento in classi

ETA	Frequenza	Freq.%	Frequenza Cumulata	Freq.% Cumulata
0-13	2	0.21	2	0.21
14-29	358	37.92	360	38.14
30-59	560	59.32	920	97.46
≥60	24	2.54	944	100.00

Per una migliore lettura: definizione delle classi di modalità!

Età studenti a.a. 2007/08

età a.c.

ETA	Frequency	Percent	Cumulative Frequency	Cumulative Percent
20	54	53.47	54	53.47
21	21	20.79	75	74.26
22	17	16.83	92	91.09
23	3	2.97	95	94.06
25	1	0.99	96	95.05
26	1	0.99	97	96.04
29	1	0.99	98	97.03
39	1	0.99	99	98.02
40	1	0.99	100	99.01
46	1	0.99	101	100.00

Frequency Missing = 1

Soddisfazione studenti a.a. 07/08

sei soddisfatto dei corsi seguiti l'anno scorso?

Soddisfatto?	Frequenza	Freq.%	Freq.Cum.	Freq.cum.%
Decisamente no	1	1.00	1	1.00
Più no che sì	19	19.00	20	20.00
Più sì che no	56	56.00	76	76.00
Decisamente sì	24	24.00	100	100.00

Frequency Missing = 2

Crediti studenti a.a. 07/08

Crediti al 19/9/07

crediti	Freq	Freq.%	Freq.Cum.	Freq.cum.%
0-24	14	15.91	14	15.91
25-40	34	38.64	48	54.55
41-50	19	21.59	67	76.14
>50	21	23.86	88	100.00

Frequency Missing = 14

Altre tabelle

Tabella di frequenza con cumulate

Modalità	Freq.	Fr.rel.	Freq.cum	F(x)
x_1	n_1	f_1	n_1	f_1
x_2	n_2	f_2	n_1+n_2	f_1+f_2
…	…	…	…	…
x_j	n_j	f_j	$n_1+n_2+\ldots+n_j$	$f_1+f_2+\ldots+f_j$
…	…	…	…	…
x_J	n_J	f_J	N	1
Totale	N	1		

Classificazione delle distribuzioni

X

- QUALITATIVE
  - SCONNESSE O NOMINALI → ***Serie sconnessa***
  - ORDINALI → ***Serie ordinata***
- QUANTITATIVE
  - DISCRETE → ***Seriazione***
  - CONTINUE → ***Seriazione***

Serie sconnessa

Modalità	Frequenze
x_1	n_1
x_j	n_j
x_J	n_J
Totale	N

Seriazione

Classi	Frequenze
$x_0 - x_1$	n_1
$x_{j-1} - x_j$	n_j
$x_{J-1} - x_J$	n_J
Totale	N

Struttura generale

Serie univariata

Modalità	Freq.
x_1	n_1
…	…
x_j	n_j
…	…
x_J	n_J
Totale	N

Seriazione univariata

Classi	Freq.
$x_0 - \| x_1$	n_1
…	…
$X_{j-1} - \| x_j$	n_j
…	…
$x_{J-1} - \| x_J$	n_J
Totale	N

$j=1, 2, \ldots, J$ modalità; N=n. unità statistiche

Tabelle di frequenza per analisi bivariate e multivariate

Dati in forma grezza

(M,2),(M,4),(F,4),…,(F,2)

Tabelle di frequenza

Un esempio di matrice dei dati

Unità u	Etichetta	SESSO	ETA'	LIVIST	DIST
Alpio	1	M	28	2	5
Caio	2	M	17	4	7,5
Prima	3	F	20	4	12
Velio	4	M	32	2	3,2
Rufa	5	F	16	1	-
Sesto	6	M	34	2	12,3
Beowful	7	M	18	1	25
Sebaste	8	F	25	2	7,7

Distribuzione per sesso e livello di istruzione

SESSO	*LIVELLO DI ISTRUZIONE*			*TOTALE*
	Lic. Elem.	Lic. Media	Laure a	
M	1	3	1	5
F	1	1	1	3
TOTALE	2	4	2	8

Tabella a doppia entrata o serie bivariata

Tabelle a doppia entrata (bivariate)

Classificazione congiunta delle u.s. in base a coppiedi caratteri

n_{ij} frequenze congiunte: nr u.s. con $X=x_i$ e $Y=y_j$

$X\|Y$	y_1	…	y_j	…	y_c	Tot
x_1	n_{11}	…	n_{1j}	…	n_{1c}	$n_{1.}$
…	…	…	…	…	…	…
x_i	n_{i1}	…	n_{ij}	…	n_{ic}	$n_{i.}$
…	…	…	…	…	…	…
x_r	n_{r1}	…	n_{rj}	…	n_{rc}	$n_{r.}$
Tot	$n_{.1}$	…	$n_{.j}$	…	$n_{.c}$	N

Tabella bivariata: pratica dello sci per titolo di studio

SCI(pratica dello sci) STUDIO(titolo di studio)

Frequenza	nessuno	element.	media	diploma	laurea	Totale
SOLO ALPINO	0	16	104	296	124	540
SOLO FONDO	0	0	18	42	28	88
ENTRAMBI	2	2	14	70	46	134
NESSUNO	0	13	37	88	34	172
Totale	2	31	173	496	232	934

Missing = 10

Tabella bivariata: tipo di alloggio per classe di età

	ETA				
ALLOGGIO	0-13	14-29	30-59	60 E +	Totale
ALBERGHIERO	0	138	238	2	378
EXTRA-ALBERG.	0	96	130	4	230
CASA PROPRIETA'	2	100	144	14	260
ALTRO	0	8	8	2	18
PENDOLARI	0	14	36	2	52
Totale	2	356	556	24	938

Missing = 4

Frequenze relative congiunte pratica dello sci per titolo di studio

SCI(pratica dello sci) STUDIO(titolo di studio)

Freq. Rel.%	nessuno	element.	media	diploma	laurea	Totale
SOLO ALPINO	0.00	1.71	11.13	31.69	13.28	57.82
SOLO FONDO	0.00	0.00	1.93	4.50	3.00	9.42
ENTRAMBI	0.21	0.21	1.50	7.49	4.93	14.35
NESSUNO	0.00	1.39	3.96	9.42	3.64	18.42
Totale	0.21	3.32	18.52	53.10	24.84	100.00

Frequency Missing = 10

Frequenze relative congiunte: $f_{ij}=n_{ij}/N=Pr(X=x_i,Y=y_j)$

Soddisfazione e voto medio esami studenti a.a 07/08

	voto medio esami				
Soddisf	<24	24-<27	27-28	29-30	Total
Decisamente no	0	1	0	0	1
Più no che sì	4	9	2	0	15
Più sì che no	7	33	11	4	55
Decisamente sì	1	7	10	5	23
Total	12	50	23	9	94

Anno di corso e CFU studenti a.a 07/08

Anno	cfu(Crediti al 19/9/07) 0-24	25-40	41-50	>50	Total
2	12	31	18	12	73
	14.63	37.80	21.95	14.63	89.02
3	2	1	0	6	9
	2.44	1.22	0.00	7.32	10.98
Total	14	32	18	18	**82**
	17.07	39.02	21.95	21.95	100.00

Frequency Missing = 20

La tabella contiene le frequenze congiunte in valore assoluto e percentuale, p.e. 14.63=100×12/**82**=100 × f_{ij}

Rappresentazioni grafiche

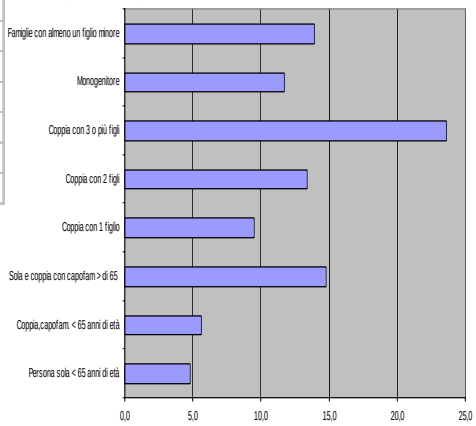
Tabella 1 - Famiglie povere per caratteristiche della famiglia Anno 1998 (ISTAT)

TIPOLOGIE FAMILIARI	1998 Numero (migliaia)	Incidenza *Incidence*
Persona sola < 65 anni di età	98	4,8
Coppia,capofam. < 65 anni di età	127	5,6
Sola e coppia con capofam > di 65	738	14,8
Coppia con 1 figlio	420	9,5
Coppia con 2 figli	521	13,4
Coppia con 3 o più figli	252	23,6
Monogenitore	190	11,7
Famiglie con almeno un figlio minore	851	13,9
TOTALE	**2.557**	**11,8**

Migliore percezione dell'informazione

Dalla tabella alla rappresentazione grafica

Figura 1 - Incidenza della povertà per caratteristiche della famiglia Anno 1998 (ISTAT)



L'Istat nel Sistema statistico nazionale

- Attivo dal 1926, l'Istituto Nazionale di Statistica è il maggior produttore di dati demografici, economici e sociali del paese.

http://www.istat.it

Diagrammi statistici

Dalla tabella alla rappresentazione grafica

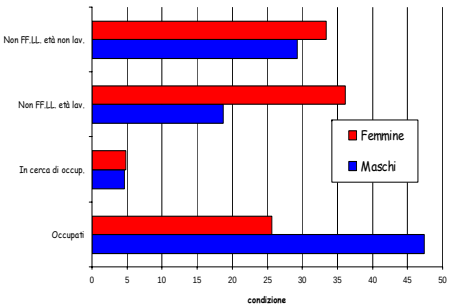
Tab. 1- Popolazione per sesso e condizione anno 1999 (%)

Condizione	*Maschi*	*Femmine*
Occupati	47.4	25.7
In cerca di occup.	4.6	4.8
Non FF.LL. età lav.	18.7	36.1
Non FF.LL. età non lav.	29.3	33.4
TOTALE	100	100

Fonte: Istat, Rapporto sull'Italia edizione 2001, il Mulino

Migliore percezione dell'informazione

Fig. 1- Popolazione per sesso e condizione anno 1999 (%)



Diagrammi per serie sconnesse o ordinate

Tab. 2- Forze lavoro per condizione anno 1999 (migliaia)

Condizione	*TOTALE*
Occupati	20435
disoccupati	996
in cerca di 1a occup.	1152
altri	596
TOTALE	23179

Fonte: Istat, Rapporto sull'Italia edizione 2001, il Mulino

Fig. 2- Forze lavoro per condizione anno 1999 (migliaia)

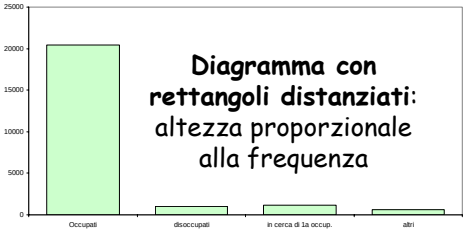


Diagramma con rettangoli distanziati: altezza proporzionale alla frequenza

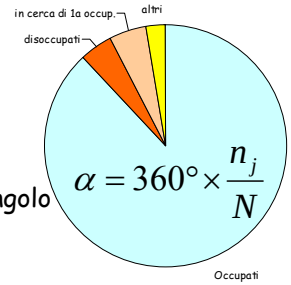
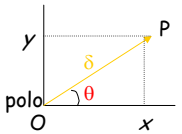


Diagramma circolare (torta): angolo al centro proporzionale alla frequenza

Grafici a stella (in coordinate polari)

Utili per caratteri qualitativi ciclici

Coordinate polari di P (δ,θ)



$x = \delta \cos(\theta)$ raggio vettore

$y = \delta \operatorname{sen}(\theta)$ angolo polare

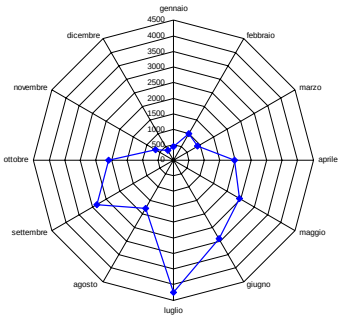
$\delta = (x^2+y^2)^{-0.5}$

$\theta = \operatorname{tang}^{-1}(y/x)$

Bilancia turistica: incassi e pagamenti per mese - Anno 1998 (miliardi di lire)

Fonte:ISTAT, Statistiche del Turismo, anno 1998

mesi	*Saldo*
gennaio	432
febbraio	990
marzo	884
aprile	1958
maggio	2458
giugno	2909
luglio	4235
agosto	1776
settembre	2828
ottobre	2089
novembre	651
dicembre	357
totale	**21567**



Costruzione:

- Suddividere 360° per numero modalità: p.e. 12 raggi distanziati di 30°;
- Lunghezza raggio proporzionale alla frequenza o quantità.

Diagrammi per serie sconnesse o ordinate

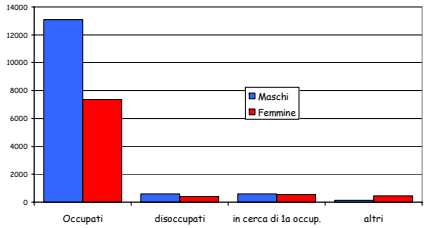
Tab. 2- Forze lavoro per sesso e condizione anno 1999 (migliaia)

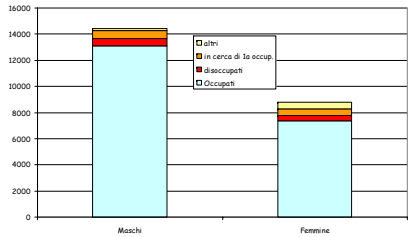
Condizione	*Maschi*	*Femmine*
Occupati	13090	7345
disoccupati	588	408
in cerca di 1a occup.	607	545
altri	118	478
TOTALE	14403	8776

Fonte: Istat, Rapporto sull'Italia edizione 2001, il Mulino

Diagramma a rettangoli distanziati: altezza proporzionale alla frequenza di ciascuna modalità

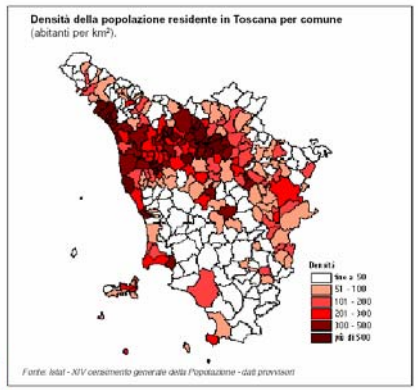
Fig. 2- Forze lavoro per sesso e condizione anno 1999 (migliaia)





Cartogrammi

Rappresentazione di serie territoriali



Aree geografiche: comuni

Carattere: densità della popolazione

Fonte: regione Toscana, Toscana in cifre anno 2002

http://www.regione.toscana.it/statistiche/index.htm

Diagrammi per seriazioni discrete

Tab. 3 – Famiglie per numero di componenti. Italia 1998 (v.a e %)

Componenti	*v.a.*	*%*
1	4594130	21.65
2	5527810	26.05
3	4954870	23.35
4	4466810	21.05
5	1294420	6.1
6 e più	381960	1.8
Totale	21220000	100

Fonte: Istat, Rapporto sull'Italia edizione 2001, il Mulino

Fig. 3 – Famiglie per numero di componenti. Italia 1998 (freq. rel. %)

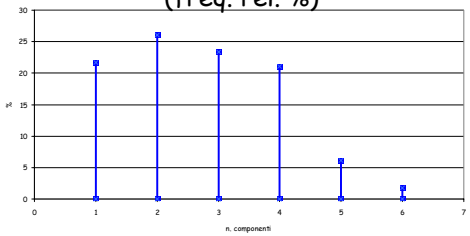


Diagramma a bastoncini: altezza proporzionale alla frequenza

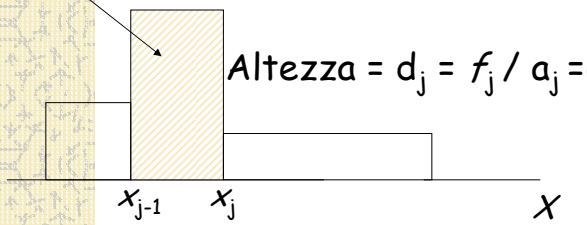
ISTOGRAMMA

Classi	Freq.rel.	Amp.cl.	densità
x_0-| x_1	f_1	a_1	d_1
…	…	…	…
X_{j-1}-| x_j	f_j	a_j	d_j
…	…	…	…
x_{J-1}-| x_J	f_J	a_J	d_J
Totale	1		

Densità di frequenza: $d_j=f_j/a_j$

Ampiezza di classe: $a_j=x_j-x_{j-1}$

Area rettangolo= $a_j * d_j = f_j$ =frequenza classe



Altezza = $d_j = f_j / a_j$ =densità classe

Base rettangolo= $x_j - x_{j-1} = a_j$ =ampiezza classe

ISTOGRAMMA

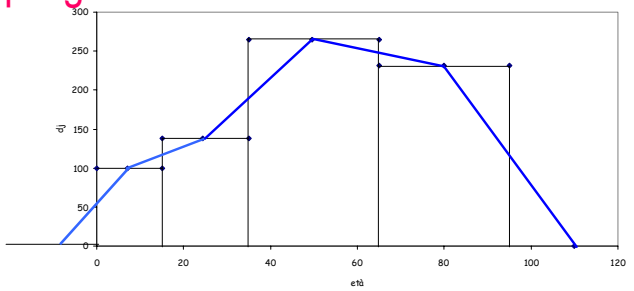
Tab. 4 – Consumatori di farmaci per classe di età, 1998 (migliaia). Istat, Rapporto sull'Italia 2001

età	nj	aj	dj
0-14	1396	14	99.71
15-34	2626	19	138.21
35-64	7687	29	265.07
65 e oltre	6933	30	231.10
TOTALE	18642		

Poligono di frequenza

segmenti che uniscono **i punti centrali** dei lati superiori dei rettangoli che definiscono l'istogramma, comprese due classi terminali con frequenza zero e ampiezza pari all'ampiezza della classe adiacente

Istogramma con poligono di frequenza



Rapporti statistici

R=A/B

- R indica quanta parte dell'intensità di A compete, in media, ad ogni unità di B
- Almeno una delle due grandezze, A o B, deve riferirsi ad un fenomeno collettivo
- Tra A e B deve intercorrere un nesso logico
- A seconda della relazione che intercorre tra A e B si hanno diversi tipi di rapporti statistici

Tipologie di rapporti statistici

- Rapporti di composizione (o di parte al tutto)
- Rapporti di coesistenza
- Rapporti di derivazione
- Rapporti di densità
- Rapporti di incremento
- Numeri indice

Rapporti di composizione

R=100xA/B

- A è una parte di B
- Esempio: A=n. pernottamenti per vacanza
  B=n. totale pernottamenti
  R=Tasso di turismo proprio

Città	*Pernottamenti al 1991*			**TTP**
	A: Vacanza	Altri motivi	B: Totale	**100xA/B**
Lione	222872	2563028	2785900	**8**
Roma	9494633	2523890	12018523	**79**

Fonte:ISTAT

R esprime quanti *turisti in senso proprio* si hanno ogni 100 pernottamenti

Altro esempio: frequenze relative

Rapporti di coesistenza

R=100xA/B

- A + B = totale
- Esempio: A=n. presenze italiani, B=n. presenze stranieri
  R=rapporto composizione italiani/stranieri

CATEGORIE	*Italiani*	*Stranieri*	**Ita/str**
Alberghi di 5 stelle e 5 stelle lusso	1279328	2724885	46.95
Alberghi di 4 stelle	27369943	30772225	88.94
Alberghi di 3 stelle	67517341	43144304	156.49
Alberghi di 2 stelle	19991704	10832685	184.55
Alberghi di 1 stella	8014643	4450227	180.10
Residenze turistico alberghiere	9122097	5912841	154.28
Esercizi alberghieri	**133295056**	**97837167**	136.24

R : *presenze di turisti italiani* ogni 100 presenze di stranieri

Fonte:ISTAT 2003

Rapporti di derivazione

R=100xA/B

- A deriva logicamente da B (B *produce* A)
- Esempio: A=n. vacanze 1-3 gg B=popolazione residente

RIPARTIZIONE GEOGRAFICA	Popolazione residente (migliaia)	viaggi 1-3 notti (migliaia)	viaggi*100 residenti
Nord	25,910	20,399	78.7
Centro	11,046	7,540	68.3
Sud	20,581	10,131	49.2
Italia	**57,537**	**38,069**	**66.2**

R : *vacanze brevi ogni 1000 residenti*

(Fonte: ISTAT 2004)

Rapporti di densità

- Particolari rapporti di derivazione
- relazione logica tra A e B tipo 'affollamento'
- Esempio: A=superficie B=popolazione residente

Ripartiz	A pop	B superf (kmq)	R=A/B	I=B/A
NO	14984766	57950.05	259	3.87
NE	10694115	61981.40	173	5.80
Centro	10946174	58379.55	188	5.33
Sud isole	20532353	123024.98	167	5.99
TOT	57157408	301335.98	190	5.27

R=A/B
abitanti per kmq

I=1000*A/B
Kmq per 1000 abitanti

(Fonte:ISTAT 2003)

Rapporti di incremento

- X_t dato al tempo t, X_{t-1} dato al tempo t-1
- $R=(X_t - X_{t-1})/ X_{t-1}$ ➔ variazione nell'intervallo di tempo
- Esempio : arrivi in Europa per anno (Fonte:WTO)

t Anno	X Arrivi (milioni)	R Tasso incremento	tasso medio annuo
1975	153.86		
1980	186.11	20.96	4.19
1985	212.11	13.97	2.79
1990	282.88	33.36	6.67
1995	335.60	18.64	3.73

$R= (X_t - X_{t-5})/ X_{t-5}$
Variazione nel quinquennio

$T= (X_t - X_{t-5})/ (5*X_{t-5})$
Variazione media annua

Rapporti statistici utili per lo studio del fenomeno turistico

- Indicatori della propensione turistica: misurano quanto una collettività si dedica al turismo
- Indicatori dell'offerta turistica: misurano la capacità ricettiva di una località
- Indicatori dei flussi turistici: misurano la produzione (domanda) turistica di una certa località

Indicatori della propensione turistica

- Tasso di propensione turistica lordo:

  TP_{lordo}=100 x (n. viaggi/pop.media)

- Tasso di propensione turistica netto:

  TP_{netto}=100x(n. persone che hanno viaggiato)/pop.media

  Pop = popolazione residente in un'area (regione, paese)

Si possono calcolare: per durata del viaggio, tipo di alloggio, località di destinazione

Indicatori della propensione turistica (2)

Anno 2003 dati in migliaia
(fonte:ISTAT)

ripartiz	pop_res	viaggi	persone	TP_lordo	TP_netto
Nord	25,674	51,348	13,747	200.0	53.5
Centro	11,104	21,098	5,282	190.0	47.6
Sud	20,217	22,239	6,462	110.0	32.0
Italia	55,697	94,685	25,491	170.0	45.8

- TP_{lordo} nel 2003 sono stati effettuati 170 viaggi ogni 100 italiani con rilevanti differenze tra ripartizioni: al nord 200, al sud 110
- TP_{netto} nel 2003 hanno effettuato almeno un viaggio 46 italiani su 100 (53 su 100 al nord, 32 su 100 al sud)

Indicatori della propensione turistica (3)

Anno 2003 dati in migliaia
(fonte:ISTAT)

ripartiz	pop_res	viaggi	persone	n. medio viaggi x residente	n. medio viaggi x persona
Nord	25,674	51,348	13,747	2	3.7
Centro	11,104	21,098	5,282	1.9	4.0
Sud	20,217	22,239	6,462	1.1	3.4
Italia	55,697	94,685	25,491	1.7	3.7

- nel 2003 il numero medio di viaggi per residente (compresi coloro che non hanno viaggiato) è pari a 1.7 con rilevanti differenze tra ripartizioni: 2 al nord, 1.1 al sud
- mentre il numero medio di viaggi per persona (considerando solo coloro che hanno viaggiato) è pari a 3.7 con poca differenza tra ripartizioni.

Indicatori dell'offerta turistica

Consistenza delle strutture ricettive

- Tasso di funzione ricettiva semplice

  TP_{SEMP}=100 x posti letto/popolazione

- Tasso di funzione ricettiva composto

  TP_{COMP}=10000 x posti letto/(pop x superficie)

TP_{SEMP} e TP_{COMP} si possono calcolare per tipologia di esercizio

- Si possono costruire indicatori analoghi considerando altri servizi turistici (p.e. bar, ristoranti, ecc)

Indicatori dei flussi turistici

- Arrivi — n. di ingressi nel luogo considerato (p.e. n. clienti registrati in un albergo)
- Presenze — n. di notti trascorse dal soggetto nel luogo considerato
- Permanenza media — presenze/arrivi

Movimento clienti negli esercizi alberghieri e nelle strutture complementari: indagine censuaria sulle strutture ricettive

- *Unità/Ente di rilevazione* Enti periferici del turismo
- *Unità d'analisi* Esercizi alberghieri, Esercizi complementari,Alloggi privati in affitto
- *Periodicità:* mensile
- Tecnica d'indagine Indagine per autocompilazione con acquisizione telematica

Indicatori della pressione turistica

- Misurano l'impatto dei turisti sul territorio
  - Sfruttamento del territorio
  - Produttività turistica
- Tasso di funzione turistica
  - Quanti turisti sono presenti (media annua) ogni 100 residenti

TT=100 x presenze/(pop x 365)

Esempio di calcolo: tasso di funzione turistica

- Italia 2002 (fonte ISTAT)
- TT=100*presenze/(pop. Residente*365)
- 100*345.247.050/(57.157.408*365)=1.65 turisti presenti (media annua) ogni 100 residenti in Italia
- 100*38.016.690/(3.506669*365)=2.97 Toscana

Indici di utilizzazione

- Misurano quanto sono stati effettivamente utilizzati i servizi e le strutture turistiche di un territorio
- Utili per valutare il dimensionamento delle strutture e dei servizi in relazione ai flussi turistici che si sono realizzati in un certo periodo
- Pubblicati dall'ISTAT insieme ai dati sui flussi

esempio 2002

Indici di utilizzazione *segue*

- Indice di utilizzazione lordo

$$IU_L = 100 \times \frac{Presenze}{Letti \times 365}$$

- Indice di utilizzazione lordo nel periodo $[t_0,t_1]$

$$IU_L = 100 \times \frac{Presenze_{t_0,t_1}}{Letti \times (t_1 - t_0)}$$

- Indice di utilizzazione netto

g: n. giorni di apertura nell'anno

$$IU_N = 100 \times \frac{Presenze}{Letti \times g}$$

Calcolo IU lordo 1998 alberghi

(fonte ISTAT)

- Lombardia 146.192 letti 18.018.393 presenze
- Toscana 155.583 letti 20.207.540 presenze

- % di utilizzo dei letti media annua
  - IU_lordo=100*18.018.393/(146.192*365)=33.8 Lombardia
  - IU_lordo=100* 20.207.540 /(155.583 *365)=35.6 Toscana

Calcolo IU netto 1998 alberghi

- Lombardia 146.192 letti 18.018.393 presenze
  n. medio gg di apertura 293
- Toscana 155.583 letti 20.207.540 presenze
  n. medio gg di apertura 311

- 100*18.018.393/(146.192*293)=42.1 Lombardia
- 100* 20.207.540 /(155.583 *311)=41.8 Toscana

- % di utilizzo dei letti media annua nel periodo di apertura

Numeri indice

- Misurano la relazione tra due misure di un certo fenomeno nel caso di serie temporali o territoriali

- N.I. a base fissa: variazione rispetto ad un valore scelto come base (denominatore= base)

$$I_0^t = 100 \times \frac{y_t}{y_0}$$

- N.I. a base mobile: variazione rispetto al valore precedente

$$I_{t-1}^t = 100 \times \frac{y_t}{y_{t-1}}$$

esempio

Numeri indice e variazioni percentuali

- È possibile passare dai N.I. alle variazioni percentuali $\Delta\%$ e viceversa
- Le $\Delta\%$ misurano le variazioni relative tra due periodi

$$\Delta\%_{t,0} = 100 \times \frac{y_t - y_0}{y_0} = 100 \times \left(\frac{y_t}{y_0} - \frac{y_0}{y_0} \right) = I_0^t - 100$$

$$\Delta\%_{t,t-1} = 100 \times \frac{y_t - y_{t-1}}{y_{t-1}} = 100 \times \left(\frac{y_t}{y_{t-1}} - \frac{y_{t-1}}{y_{t-1}} \right) = I_{t-1}^t - 100$$

Relazione tra N.I. a base fissa e N.I a base mobile

- È possibile trasformare la serie dei numeri indice a base fissa nella serie dei numeri indice a base mobile

$$I_{t-1}^t = 100 \times \frac{y_t}{y_{t-1}} = 100 \times \frac{y_t / y_0}{y_{t-1} / y_0}$$

$$= 100 \times \frac{I_0^t / 100}{I_0^{t-1} / 100} = 100 \times \frac{I_0^t}{I_0^{t-1}}$$

Relazione tra N.I. a base mobile e N.I a base fissa

- È possibile trasformare la serie dei numeri indice a base mobile nella serie dei numeri indice a base fissa

$$I_0^2 = 100 \times \frac{y_2}{y_0} = 100 \times \frac{y_2}{y_1} \times \frac{y_1}{y_0} = 100 \times \frac{I_1^2}{100} \times \frac{I_0^1}{100}$$

$$I_0^t = 100 \times \frac{I_{t-1}^t}{100} \times \ldots \times \frac{I_1^2}{100} \times \frac{I_0^1}{100}$$

Funzioni che descrivono la v.s. Y

- Y: $\mathcal{U} \rightarrow S_Y$
- S_Y supporto di Y (insieme delle modalità osservate)
- Se le etichette non sono rilevanti, sintetizziamo la v.s. Y in una tabella di frequenza
- Frequenza assoluta n_j: $S_y \rightarrow \mathcal{N}$

n. unità statistiche che possiedono la j-ma modalità

Proprietà

$n_j > 0$ per y_j in S_y ;

$\Sigma\, y_j\, n_j = N$

Funzioni che descrivono la v.s. Y
frequenza relativa

$f_j: S_y \rightarrow (0;1]$

$f_j = pr\{Y=y\} = n_j / N$

Utile per:

– confrontare popolazioni;

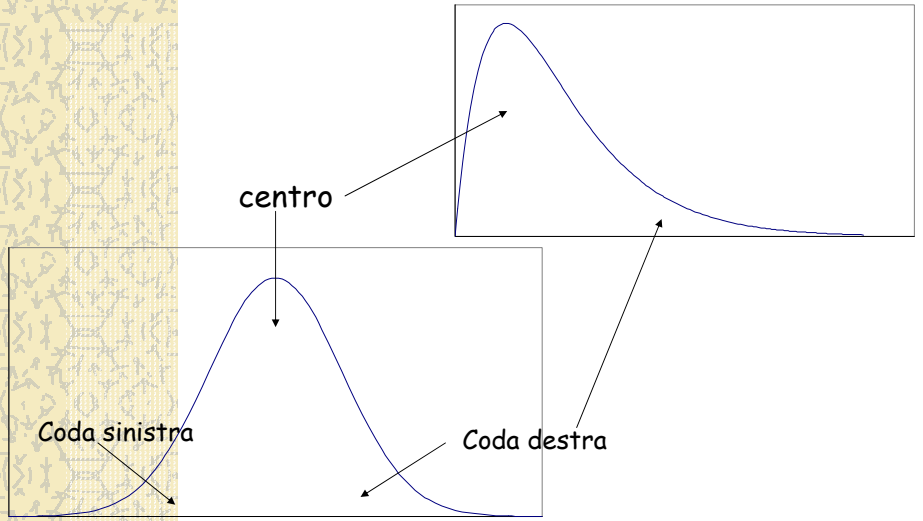
– attribuire un peso alle modalità;

Proprietà

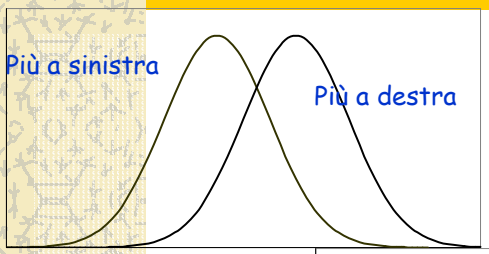
$f_j > 0$ per y_j in S_y ;

$\Sigma\, y_j\, f_j = 1$

Aspetti notevoli delle distribuzioni

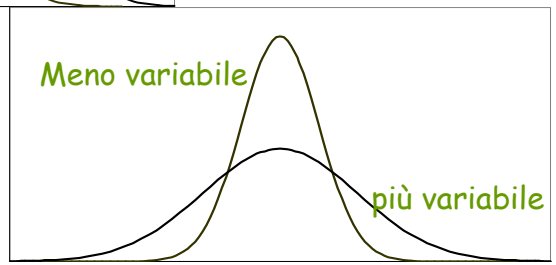


Aspetti caratterizzanti le distribuzioni: posizione e variabilità

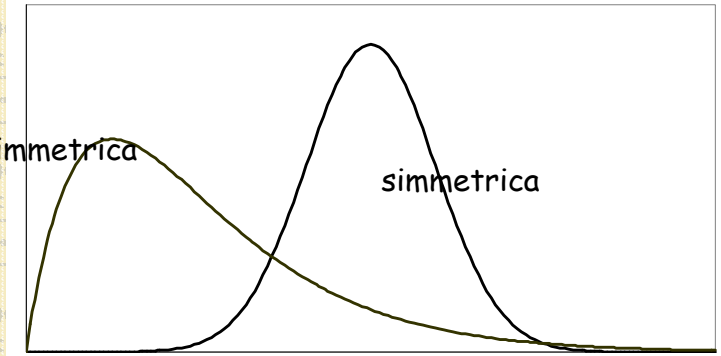


Posizione

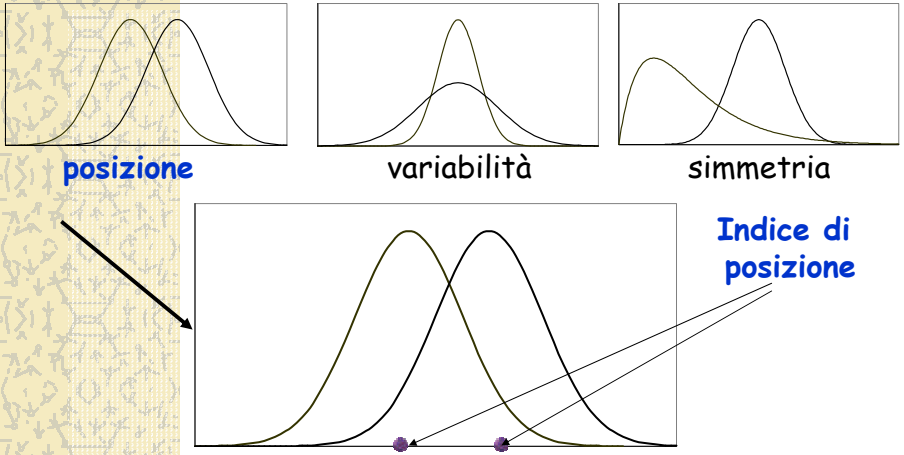
Variabilità



Aspetti caratterizzanti le distribuzioni: forma



Indici caratteristici delle distribuzioni



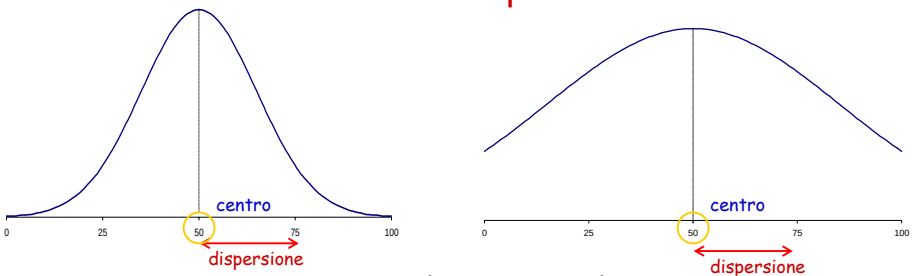
Gli indici di posizione: medie

Sintesi della distribuzione attraverso un valore rappresentativo

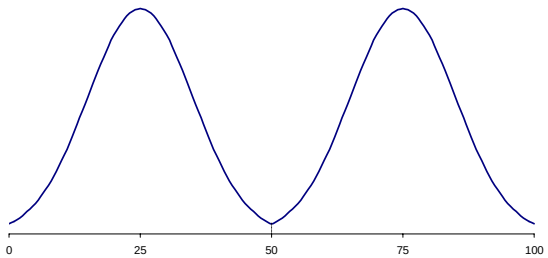
VARIABILI

- sconnesse: moda
- ordinali: mediana
- quantitative: media aritmetica, media geometrica

Il centro e la dispersione



stesso centro, dispersione diversa



Il centro non sintetizza la distribuzione!

Media aritmetica

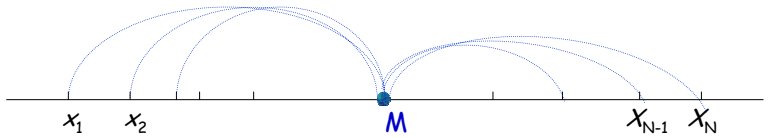
X v.s. quantitativa

$\{x_1, x_2, \ldots, x_N\}$ successione

$$\mu = M = \frac{1}{N}\sum_{i=1}^{N} x_i$$

esempio X: 2 3 2 1 ➔ M=(2+3+2+1)/4=2

MEDIA: CENTRO DELL'INSIEME DEGLI *N* PUNTI



Distanza tra x_i e x_j in $\mathbb{R}$: $d_{ij} = |x_i - x_j|$

Media aritmetica di una serie

Tabella di frequenza

Mod.tà	Freq.	Fr.rel.
x_1	n_1	f_1
x_2	n_2	f_2
...	...	...
x_j	n_j	f_j
...	...	...
x_J	n_J	f_J
Totale	N	1

$$M = \frac{1}{N}\sum_{j=1}^{J} x_j n_j = \sum_{j=1}^{J} x_j f_j$$

Serie: $\{(x_j, n_j)\}$

j=1,2,...,J

Esempio media aritmetica di una serie

ETA	nj	fj*100	Fj	Fj*100
20	54	53.47	54	53.47
21	21	20.79	75	74.26
22	17	16.83	92	91.09
23	3	2.97	95	94.06
25	1	0.99	96	95.05
26	1	0.99	97	96.04
29	1	0.99	98	97.03
39	1	0.99	99	98.02
40	1	0.99	100	99.01
46	1	0.99	101	100.00

Età media=21

Media aritmetica per dati in classi

Ipotesi istogramma:

equidistribuzione frequenze all'interno delle classi

Tabella di frequenza

Mod.tà	Freq.	Fr.rel.
x_0-x_1	n_1	f_1
x_1-x_2	n_2	f_2
...	...	...
x_{j-1}-x_j	n_j	f_j
...	...	...
x_{J-1}-x_J	n_J	f_J
Totale	N	1

Valore centrale di classe

$c_j = (x_j + x_{j-1})/2$

$$M = \frac{1}{N}\sum_{j=1}^{J} c_j n_j = \sum_{j=1}^{J} c_j f_j$$

Seriazione: $\{(x_{j-1}; x_j), n_j)\}$

j=1,2,...,J

Esempio: media di una seriazione

CFU	nj	fj*100	Fj	Fj*100
0-24	14	15.91	14	15.91
25-40	34	38.64	48	54.55
41-50	19	21.59	67	76.14
>50	21	23.86	88	100.00

Frequency Missing = 14
MIN=9, MAX=142, P90%=69

foglio di calcolo

Proprietà della media aritmetica

- Centro di ordine 2 (minimi quadrati)
- Lascia invariato l'ammontare complessivo: $NM = \sum_{i=1}^{N} x_i$
- Baricentro $\sum_{i=1}^{N} (x_i - M) = 0$
- Internalità (propr. di Cauchy) $x_{[1]} \le M \le x_{[n]}$
- Invarianza per trasformazioni lineari

$$Y = a + bX \Rightarrow M(Y) = a + bM(X)$$

Media aritmetica ponderata

$$M_p = \frac{\sum_{j=1}^{J} x_j p_j}{\sum_{j=1}^{J} p_j}$$

- Esempio: indici sintetici dei prezzi con pesi pari al valore delle merci scambiate

  (A) Indice di LASPEYRES: pesi= $p_{iA}q_{iA}$
  (B) Indice di PAASCHE: pesi= $p_{iA}q_{iB}$

  TEMPI: A<B

Indici sintetici dei prezzi

- Paniere di beni: i=1,…,k
- NI elementari $I_i=p_{iB}/p_{iA}$ A<B
- Indice di Laspeyres

$$I_L = \frac{\sum_{i=1}^{k} \frac{p_{iB}}{p_{iA}} \times p_{iA}q_{iA}}{\sum_{i=1}^{k} p_{iA}q_{iA}} = \frac{\sum_{i=1}^{k} p_{iB}q_{iA}}{\sum_{i=1}^{k} p_{iA}q_{iA}}$$

- Indice di Paasche

$$I_L = \frac{\sum_{i=1}^{k} \frac{p_{iB}}{p_{iA}} \times p_{iA}q_{iB}}{\sum_{i=1}^{k} p_{iA}q_{iB}} = \frac{\sum_{i=1}^{k} p_{iB}q_{iB}}{\sum_{i=1}^{k} p_{iA}q_{iB}}$$

Media geometrica

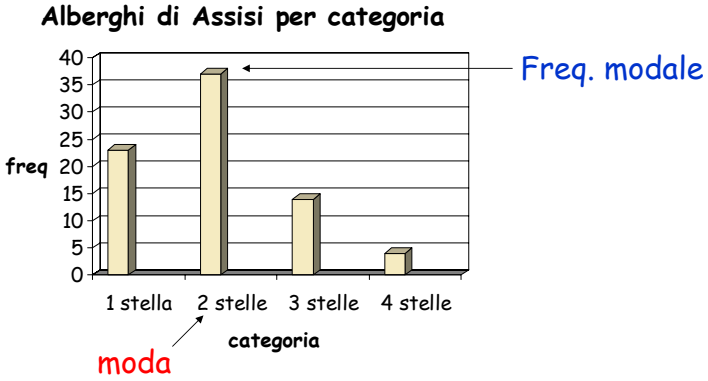
$$M_g = \sqrt[N]{\prod_{i=1}^{N} x_i} = \left(\prod_{i=1}^{N} x_i \right)^{\frac{1}{N}}$$

$$\log M_g = \frac{1}{N} \sum_{i=1}^{N} \log x_i$$

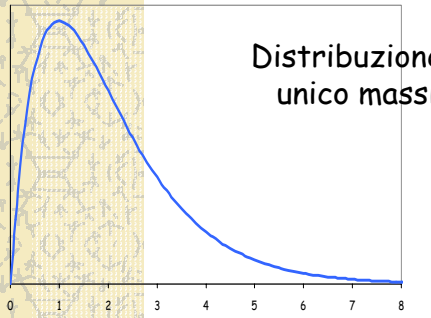
- M_g valore che sostituito agli N termini della successione ne lascia invariato il prodotto.
- M_g applicata ad una progressione geometrica, con N dispari, fornisce termine centrale della progressione.

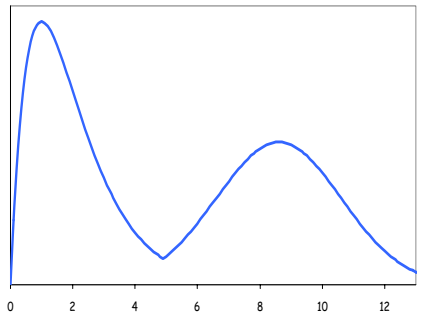
La moda

Modalità cui corrisponde la frequenza più alta.



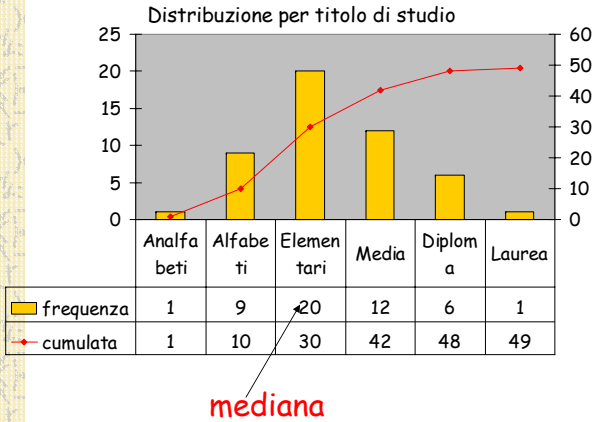
Moda
utile per distribuzioni unimodali





La mediana

Modalità che è più grande del 50% delle osservazioni e più piccola del 50% delle osservazioni



Mediana di una successione

La mediana M_e di n numeri ordinati in senso non decrescente $\{x_{(1)}, \ldots, x_{(n)}\}$ è:

- per n dispari $M_e = x_{(n+1)/2}$
- per n pari M_e in $[x_{n/2}; x_{(n/2)+1}]$

se X è quantitativa, si considera

$$M_e = [x_{n/2} + x_{(n/2)+1}]/2$$

Calcolo della mediana attraverso le frequenze cumulate

(1) Calcolare **0.5*N**=N/2

(2) guardare il valore ottenuto:

- se *N*0.5* è un numero frazionario (N dispari) e *M* numero intero tale che *M<N*0.5<M+1:*

$$M_e = x_{(M+1)}$$

- se *N*0.5 è* numero intero (N pari):

$$M_e \in [x_{(N*0.5)}; x_{(N*0.5+1)}]$$

(3) Individuare il valore mediano in base al valore delle frequenze cumulate

Calcolo della mediana attraverso la cumulata relativa F(X)

$$M_e = x_m : \begin{cases} F(x_m) \geq 0.5 \\ F(x_{m-1}) \leq 0.5 \end{cases}$$

Se x è una v.s. ordinale o discreta si ha:

$$F(x_{m-1}) = F(x_m) - f_m$$

Proprietà della mediana

- Centro di ordine 1 $\quad M_e : \sum_{i=1}^{n} |x_i - M_e| = \min$

•Si può utilizzare anche per v.s. ordinali

•M_e resta invariata se si sostituiscono i termini x< M_e o x> M_e : non risente di valori anomali

Calcolo della mediana per v.s. ordinali

$$M_e = x_{0.5} : \begin{cases} pr(X \geq x_{0.5}) \geq 0.5 \\ pr(X \leq x_{0.5}) = F(x_{0.5}) \geq 0.5 \end{cases}$$

TITSTUD	Freq	Percent	Cumulative Frequency	Cumulative Percent
lic. elementare	2	0.34	2	0.34
lic. media	105	17.89	107	18.23
diploma	413	70.36	520	**88.59**
diploma univ.	7	1.19	527	89.78
laurea	61	10.22	588	100.00

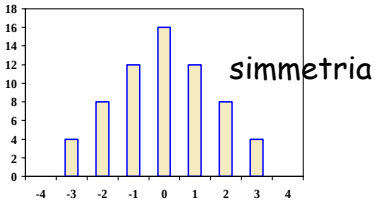
N*0.5=588*0.5=294➔ $M_e = x_{0.5} \in [x_{294}; x_{295}] = diploma$

Calcolo della mediana per v.s. ordinali

Titolo di studio	*Frequenze*	*Frequenze cumulate*
Nessun titolo	1862	1862
Licenza elementare	9903	11765
Licenza media	4491	16256
Diploma	3093	19349
Laurea	1160	20509
Totale	20509	

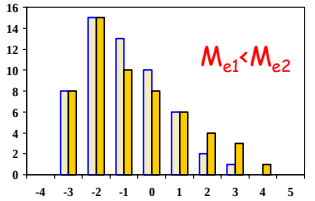
Mediana di un istogramma:

$$M_e : pr(X \leq M_e) = F(M_e) = 0.5$$

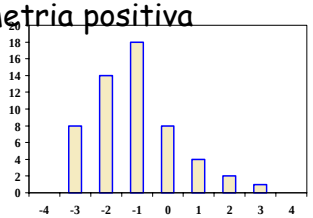


Me=M

Non risente delle code

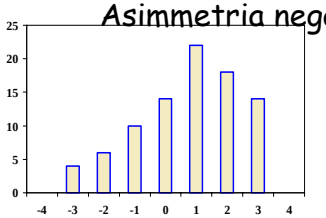


Asimmetria positiva



Me<M

Asimmetria negativa



M< Me