

---

# Probabilità e Statistica Esercitazioni

a.a. 2009/2010

C.d.L.: Ingegneria Elettronica e delle Telecomunicazioni, Ingegneria Informatica

## *Statistica descrittiva I*

Ines Campa

---

—

La STATISTICA DESCRITTIVA si occupa di

- raccogliere e organizzare i dati in tabelle e diagrammi di facile lettura;
- elaborare i dati cioè ricavare indici sintetici che diano informazioni su alcuni aspetti rilevanti. Per esempio misure di tendenza centrale (media, mediana, moda) e di variabilità (varianza, deviazione standard).

La STATISTICA INFERENZIALE si occupa di estendere i risultati dell'analisi descrittiva condotta su un sottogruppo alla popolazione da cui questo proviene.

Per dedurre enunciati formalmente validi dai dati raccolti è necessario che prenda in considerazione l'influenza del caso  $\Rightarrow$  deve fare assunzioni sulla probabilità che i dati rilevati assumano i diversi valori possibili.

# Terminologia

**Carattere:** oggetto della rilevazione statistica.

**Modalità:** manifestazione del carattere.

**Unità statistica:** unità elementare su cui viene osservato il carattere oggetto di studio.

**Popolazione:** insieme delle unità statistiche.

**Campione:** sottogruppo della popolazione.

Il campione è rappresentativo della popolazione quando le sue unità sono scelte in modo completamente casuale.

Distinguiamo tra caratteri:

- qualitativi (es. il colore degli occhi),
- quantitativi (es. l'altezza).

I caratteri quantitativi possono a loro volta essere suddivisi in

- discreti (es. il numero delle stanze in un appartamento),
- continui (es. il peso di una persona).

# Tabelle

Supponiamo che ci sia un solo carattere oggetto di indagine. In una tabella si può riportare o una frequenza o una intensità (una misura).

Esempio di tabella per caratteri qualitativi per una distribuzione di intensità.

Produzione della pesca marittima e lagunare	
tipo di pescato	quantità (in quintali)
pesci	2860590
molluschi	998130
crostacei	310200
totale	4168920

Fonte:Istat

La tabella è già una sintesi dei dati grezzi

Esempio di tabella per caratteri qualitativi per una distribuzione di frequenza.

Impiego dipendenti azienda Beta spa	
Professione	frequenza
operai	17
impiegati	68
dirigenti	15
totale	100

Fonte: Fantasia

Lunghezza di 40 Foglie di alloro in  $mm$  (registrata al  $mm$  più prossimo)

138	164	150	132	144	125	149	157
146	158	140	147	136	148	152	144
168	126	138	176	163	119	154	165
146	175	142	147	135	153	140	135
161	145	135	142	150	156	145	128

Quella sopra riportata è una tabella piuttosto "grezza".

Esempio di tabella per caratteri quantitativi (continui) per una distribuzione di frequenza.

Dividiamo i dati in classi i.e. in intervalli di misura  $5mm$  a partire da  $117mm$

## TITOLO

classi di lunghezza	frequenza
117- 122	1
122- 127	2
127- 132	2
132- 137	4
137- 142	6
142- 147	8
147- 152	5
152- 157	4
157- 162	2
162- 167	3
167- 172	1
172- 177	2
totale	40

Fonte

# Distribuzione di frequenza assoluta

freq. assoluta	
$c_1$	$n_1$
.	.
.	.
.	.
$c_i$	$n_i$
.	.
.	.
.	.
$c_k$	$n_k$
Tot	$n = \sum_{i=1}^k n_i$

Fonte

# Distribuzione di frequenza relativa

Alla classe  $i$ -esima  $c_i$  corrisponde la frequenza relativa  $f_i = \frac{n_i}{n}$ . Si ha

$$0 \leq f_i \leq 1 \text{ e } \sum_{i=1}^k f_i = 1$$

	freq.relativa
$c_1$	$f_1$
.	.
.	.
.	.
$c_i$	$f_i$
.	.
.	.
.	.
$c_k$	$f_k$
Tot	1

Fonte

Qual é il vantaggio di usare la frequenza relativa?

- con la frequenza relativa si percepisce meglio la distribuzione rispetto al totale;
- avendo tolto la dipendenza da  $n$ , si possono confrontare fra loro diverse tabelle.

Per esempio, se gli analfabeti in un paese  $X$  sono 1000 e quelli nel paese  $Y$  sono 500 si può pensare che ci siano più analfabeti in  $X$  che in  $Y$ .

Calcoliamo le frequenze relative in  $X$ :

$$\frac{1000}{\text{numero abitanti}} = 0,02$$

in  $Y$ :

$$\frac{500}{\text{numero abitanti}} = 0,10$$

Pertanto l'asserzione è corretta in senso assoluto, ma gli analfabeti sono di meno, in percentuale, in  $X$  che in  $Y$ .

## Diagrammi statistici o grafici

Per i caratteri qualitativi vi sono vari tipi di diagrammi statistici.

Qualifica	operai impiegati nella		
	indus. meccanica	indus. elettrica	indus. chimica
operai spec.	10	10	18
operai qualif.	20	15	8
manovali	30	15	5
totale	60	40	31
		Tot op.	131

Fonte

La tabella utilizzata è a doppia entrata

# Grafico a rettangoli distanziati

## Grafico a torta

Le superfici degli oceani	milioni di $km^2$
Pacifico	183.4
Atlantico	106.7
Indiano	73.8

N.B. Gli angoli al centro dei settori circolari sono proporzionali all'intensità.

# Diagramma a figura

Censimento della popolazione italiana

anno	frequenza in migliaia
1881	29791
1901	33778
.	.
.	.
.	.

Ogni omينو rappresenta  $x$  milioni.

# Serie temporale

---

La serie temporale è un aggregato di osservazioni fatte in tempi diversi.

La serie temporale può svolgere la funzione di

- descrizione,
- interpolazione,
- previsione.

---

Secondo i medici contrari alla vaccinazione, facendo una previsione tramite la serie temporale, il valore di mortalità doveva essere quello in rosso; quindi se non fossero stati usati i vaccini i morti sarebbero stati meno. Questo è assurdo, perchè secondo questo ragionamento fra poco si "risorgerebbe" tutti (valore in blu). L'errore è usare l'interpolazione lineare per fare la previsione al posto di quella esponenziale.

---

# Grafico a bastoncini

---

Per i caratteri quantitativi discreti.

Le altezze dei bastoncini sono proporzionali alla frequenza.

# Istogramma

Per i caratteri quantitativi continui.

Le aree dei rettangoli sono proporzionali alle frequenze osservate.

# Frequenza cumulata

**Definizione 1.** Diciamo frequenza cumulata corrispondente al valore  $x_i$ , la somma delle frequenze fino a comprendere quella di  $x_i$ .

$$N_i = \sum_{j=1}^i n_j \quad \text{dove } N_i = \text{freq. cumulate, } n_j = \text{freq. di } x_j.$$

Osservazione

Per le frequenze cumulate assolute

$$0 \leq N_i \leq n \quad N_i \leq N_k \text{ se } i \leq k$$

Per le frequenze cumulate relative

$$F_i = \sum_{j=1}^i f_j \quad F_i \leq F_k \text{ se } i \leq k$$

# Poligono delle frequenze cumulate

Riprendiamo l'esempio delle 40 foglie di alloro

$C_i$	$n_i$	$N_i$
117- 122	1	1
122- 127	2	3
127- 132	2	5
132- 137	4	9
137- 142	6	15
142- 147	8	23
147- 152	5	28
152- 157	4	32
157- 162	2	34
162- 167	3	37
167- 172	1	38
172- 177	2	40
	totale	40



# Diagramma stem and leaf

Una maniera efficiente di organizzare un numero non troppo elevato di dati è il diagramma **stem and leaf**, in italiano **ramo-foglia**. Per costruirlo, occorre dividere le cifre di ogni dato numerico in due parti, una più significativa (stem) e una meno significativa (leaf).

Esempio: in uno studio scientifico (D. G. Hoel "A representation of mortality data by competing risks" Biometrics) un gruppo di topi di cinque settimane fu sottoposto a una dose di radiazioni di 300 rad. I topi furono divisi in due gruppi. Il primo fu tenuto in ambiente sterile, il secondo in normali condizioni di laboratorio. La seguente matrice di dati riporta il numero di giorni di vita dei topi, tenuti in normali condizioni di laboratorio, che in seguito morirono di linfoma al timo.

159	189	191	198	235
245	250	256	261	265
266	280	343	356	383
403	414	428	432	

Realizziamo il diagramma stem and leaf.

1	59	89	91	98				
2	35	45	50	56	61	65	66	80
3	43	56	83					
4	03	14	28	32				