Statistica a.a. 2012/2013

- Orario

  lunedì 14:00-16:00 F

  martedì 14:00-16:00 Leibnitz

  giovedì 11:00-13:00 Leibnitz

  venerdì 11:00-13:00 e 14:00-16:00 F

- Ricevimento

  giovedì 16:00-18:00 in Dipartimento

- Materiale didattico

I lucidi utilizzati a lezione, le esercitazioni e ogni eventuale altro materiale didattico è a disposizione sul sito: http://docenti.unimc.it/docenti/luisa-scaccia

- Organizzazione delle lezioni

Lezioni frontali dal lunedì al giovedì mattina. Esercitazione giovedì pomeriggio.

- Modalità d'esame

L'esame è scritto e orale.

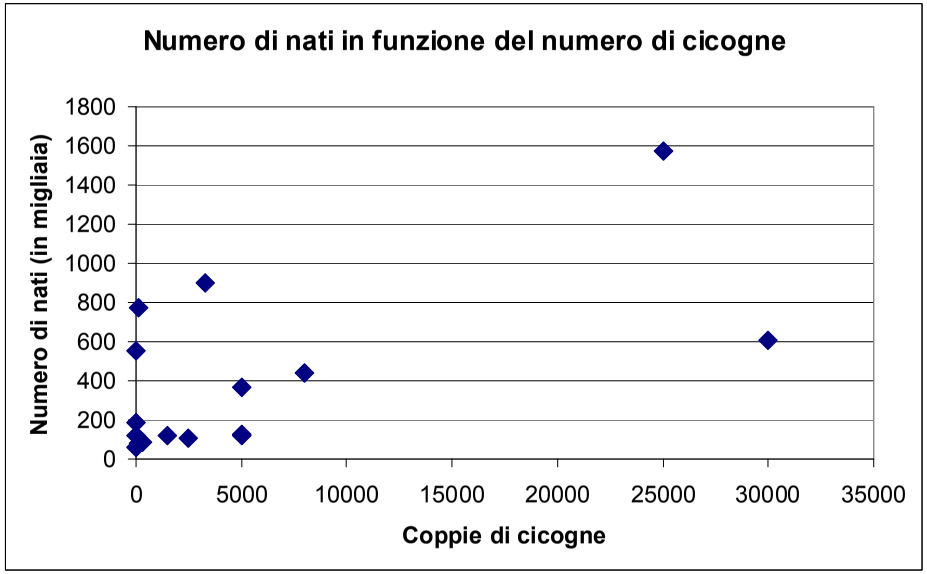
Scritto: 4 esercizi a scelta sui 6 proposti. Ogni esercizio svolto correttamente vale 8 punti (per un totale massimo di 32). Si può utilizzare il libro di testo, gli appunti, la calcolatrice.

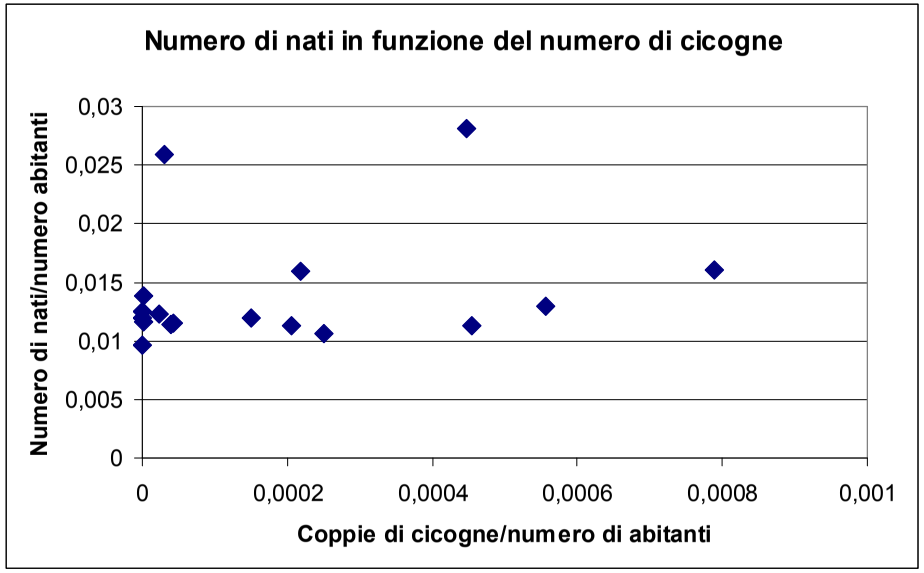
Orale: può essere sostenuto nello stesso appello dello scritto o in uno degli appelli successivi, purchè all'interno della Sessione d'esami. Il risultato dell'orale fa media con quello dello scritto.

Nascite e cicogne

Country	Area (km^2)	Storks (pairs)	Humans (10^6)	Birth rate (10^3/yr)
Albania	28,750	100	3.2	83
Austria	83,860	300	7.6	87
Belgium	30,520	1	9.9	118
Bulgaria	111,000	5000	9.0	117
Denmark	43,100	9	5.1	59
France	544,000	140	56	774
Germany	357,000	3300	78	901
Greece	132,000	2500	10	106
Holland	41,900	4	15	188
Hungary	93,000	5000	11	124
Italy	301,280	5	57	551
Poland	312,680	30,000	38	610
Portugal	92,390	1500	10	120
Romania	237,500	5000	23	367
Spain	504,750	8000	39	439
Switzerland	41,290	150	6.7	82
Turkey	779,450	25,000	56	1576

Table 1. Geographic, human and stork data for 17 European countries





L'inflazione ufficiale e la "vera" corsa del caro vita

Esempio (Corriere della Sera del 16/09/2003)

- "Inflazione ufficiale in Agosto 2,8%" (ISTAT)
- 960 prodotti in 81 città in 29000 negozi
- la media è circa il 7%

- La "vera" corsa del carovita (Altroconsumo)
- 7 prodotti in cinque città

	BA	MI	NA	RM	TO
Caffè	4,5	5,9	14,3	11,9	2,4
Cappuccino	5,0	4,0	20,0	11,8	4,5
Pizza Margherita	12,8	3,1	4,8	8,2	4,8
Biglietto cinema	4,8	3,7	-3,2	3,7	4,8
Piega e taglio	0,0	1,6	6,3	12,4	3,5
Lavanderia	3,0	7,4	11,2	9,0	20,3
CD	nd	8,4	4,6	8,4	5,7

Variazioni percentuali annue

Variazioni annue dei prezzi di alcuni prodotti e relativi pesi sull'indice dell'inflazione media.

I maggiori aumenti

	Var. prezzi %	Peso %
Ortaggi	+ 10,9	1,3
Frutta fresca	+ 7,3	1,0
Patate	+ 15,9	0,2
Serv. bancari	+ 8,1	0,6

Le maggiori diminuzioni

	Var. prezzi %	Peso %
Computer	-15,4	0,2
Medicinali	- 3,1	2,9
App. telef.	- 2,7	0,9
Serv. telef.	- 1,1	2,1

Fonte: Istat, periodo: sett. 2002 - sett. 2003

- "*Gli automobilisti corretti? Solo l'8%*". (Corriere della Sera del 25/08/2003)
- Dal titolo siamo portati a pensare che quasi tutti gli automobilisti sono scorretti.

- Invece, i dati sono i seguenti:
  (Inchiesta Altroconsumo)

	%
guidatori corretti	8
guidatori non completamente corretti	86
guidatori scorretti	6
TOTALE	100

- Un titolo altrettanto parziale, ma un po' più vero di quello scelto:
  "Gli automobilisti scorretti? Solo il 6%".
- E avrebbe trasmesso un'informazione opposta…

Abusivismo, la top ten regionale

(Il Messaggero del 17/09/2003)

	Abitazioni abusive costruite nel 2002
Campania	5925
Sicilia	4260
Puglia	3820
Calabria	2919
Lombardia	1901
Lazio	1697
Veneto	1664
Sardegna	1482
Toscana	1327
Abruzzo	1252

- Nei confronti geografici i numeri assoluti sono spesso fuorvianti.
- Tra le dieci regioni della tabella figurano le sette regioni "più grandi" del paese.
- In testa a graduatorie simili non troveremo mai una "piccola" regione.

Come fare un confronto corretto?

	Abitazioni abusive costruite nel 2002	% sul totale delle abitazioni costruite
Campania	5925	28,3
Sicilia	4260	28,9
Puglia	3820	15,8
Calabria	2919	23,2
Lombardia	1901	4,4
Lazio	1697	8,1
Veneto	1664	4,5
Sardegna	1482	15,5
Toscana	1327	8,4
Abruzzo	1252	19,3

- In questo caso, rapportando in ogni regione il numero di abitazioni abusive a quello delle abitazioni costruite.

Ecco come cambia la top ten dell'abusivismo:
(Fonte: La Repubblica del 17 e 21 settembre 2003)

	Abitazioni abusive costruite nel 2002
Campania	5925
Sicilia	4260
Puglia	3820
Calabria	2919
Lombardia	1901
Lazio	1697
Veneto	1664
Sardegna	1482
Toscana	1327
Abruzzo	1252

	% sul totale delle abitazioni costruite
Molise	30,7
Sicilia	28,9
Campania	28,3
Calabria	23,2
Abruzzo	19,3
Liguria	18,8
Puglia	15,8
Sardegna	15,5
Basilicata	11,6
Toscana	8,4
ITALIA	11,0

Dati OCSE

a) Spesa annuale per studente:

USA:	24370$	Inghilterra:	13506$	Germania:	12446$
Francia:	10995$	Media OCSE:	11512$	**Italia:**	**8026$ (-30,2%)**

b) Spesa pubblica annuale per studente:

USA:	8400$	Inghilterra:	9400$	Germania:	10200$
Francia:	9300$	Media OCSE:	8400$	**Italia:**	**5400$ (-35,7%)**

c) numero studenti / numero docenti:

USA:	15,1	Inghilterra:	16,4	Germania:	12,4
Francia:	17,0	Media OCSE:	15,3	**Italia:**	**20,4 (+33,3%)**

Il confronto con gli altri paesi industrializzati suggerisce che il numero di docenti universitari dovrebbe aumentare invece che diminuire!

Statistica e Economia

- Descrizione dello stato e dell'andamento nel tempo dei fenomeni economici
- Analisi dei comportamenti degli operatori economici
- Previsioni sulla dinamica degli aggregati economici
- Analisi dei processi e dei risultati produttivi e gestionali
- Valutazione delle condizioni del mercato
- Pianificazione delle strategie di marketing
- Scelta tra portafogli alternativi
- Previsioni sulla dinamica delle misure finanziarie

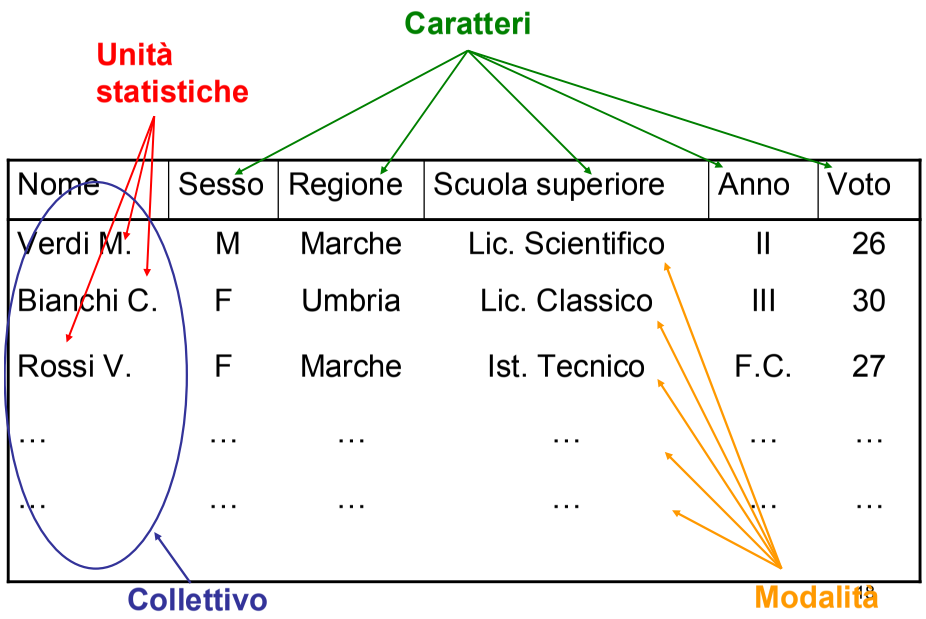
Qualche definizione

- Statistica: insieme dei principi ai quali dovrebbero ispirarsi la raccolta e l'elaborazione dei dati concernenti fenomeni collettivi
- Statistica descrittiva: si occupa dell'analisi di un fenomeno relativo a un certo gruppo di soggetti (popolazione) sulla base di una rilevazione completa delle informazioni (censimento). Tali informazioni vengono sintetizzate tramite opportuni indici statistici.
- Inferenza statistica: basandosi su un campione estratto dalla popolazione di interesse, trae conclusioni sull'intera popolazione

- Popolazione: insieme di riferimento del fenomeno oggetto di studio
- Unità statistica: singolo caso che compone la popolazione
- Carattere: caratteristica oggetto di rilevazione sulle unità statistiche che formano il collettivo
- Modalità di un carattere: diversi modi con cui il carattere si manifesta nelle unità statistiche

Esempio

- Fenomeno collettivo che si intende studiare: rendimento degli studenti della Facoltà di Economia dell'Università di Macerata all'esame di Statistica, nell'A.A. 2009-2010
- Collettivo statistico: insieme degli studenti di Economia dell'Università di Macerata che hanno sostenuto l'esame di Statistica nell'A.A. 2009-2010
- Unità statistica: singolo studente di Economia dell'Università di Macerata che ha sostenuto l'esame di Statistica nell'A.A. 2009-2010
- Caratteri rilevati: sesso, regione di provenienza, tipo di scuola superiore, anno di corso, voto all'esame di statistica
- Matrice di dati: tabella con numero di righe pari al numero di unità statistiche e numero di colonne pari al numero di caratteri rilevati



Nome	Sesso	Regione	Scuola superiore	Anno	Voto
Verdi M.	M	Marche	Lic. Scientifico	II	26
Bianchi C.	F	Umbria	Lic. Classico	III	30
Rossi V.	F	Marche	Ist. Tecnico	F.C.	27
…	…	…	…	…	…
…	…	…	…	…	…

Classificazione dei caratteri

- A seconda di come sono espresse le sue modalità, un carattere viene classificato in
  - Qualitativo: quando le modalità sono espresse tramite espressioni verbali
  - Quantitativo: quando le modalità sono espresse numericamente
- Un **carattere qualitativo** può essere ulteriormente classificato come:
  - Sconnesso: non esiste un ordine naturale delle modalità
  - Ordinato: esiste un ordine naturale delle modalità

- Un **carattere quantitativo** può essere ulteriormente classificato come:
  - Discreto: le modalità possono essere messe in corrispondenza biunivoca con un sottoinsieme dei numeri interi
  - Continuo: si ha una corrispondenza biunivoca con l'insieme dei numeri reali
- Un **carattere quantitativo** si dice **trasferibile** quando la sua intensità può essere trasferita da un'unità all'altra

Rilevazioni statistiche

- Rilevazione statistica: insieme delle operazioni necessarie per il raccoglimento dei dati necessari ad un'indagine statistica
- A seconda del **metodo**, la rilevazione può essere:
  - Sperimentale: il ricercatore controlla le condizioni sotto le quali si svolge l'osservazione
  - Osservazionale: si osserva la realtà senza intervenire su di essa
- A seconda della **complessità**, la rilevazione può essere:
  - Totale: si osservano tutte le unità della popolazione
  - Parziale: si osserva un campione estratto dalla popolazione
- Tipici **strumenti**, di rilevazione sono:
  - Intervista diretta
  - Intervista telefonica
  - Questionario postale

Distribuzioni statistiche

- La rilevazione statistica produce come risultato la matrice dei dati

Nome	Sesso	Regione	Scuola superiore	Anno	Voto
Verdi M.	M	Marche	Lic. Scientifico	II	26
Bianchi C.	F	Umbria	Lic. Classico	III	30
Rossi V.	F	Marche	Ist. Tecnico	F.C.	27
Gialli F.	F	Calabria	Ist. Tecnico	II	30
Neri A.	M	Marche	Lic. Scientifico	III	28

- Ogni colonna della matrice costituisce una ***distribuzione disaggregata*** secondo un singolo carattere. Si tratta dell'elencazione delle modalità osservate per ogni una unità

$$x_1, x_2, \ldots, x_j, \ldots, x_n$$

- Una distribuzione di questo tipo si chiama ***semplice*** (rispetto ad un solo carattere) ***unitaria*** (unità per unità).

$$x_1, x_2, \ldots, x_i, \ldots, x_n$$

- Se si considerassero più caratteri congiuntamente avremmo una distribuzione ***multipla*** (es. ***doppia*** se si considerassero due caratteri)
- Per *sintetizzare* una distribuzione disaggregata si fa uso di una distribuzione di ***frequenza*** che può essere semplice o multipla
- Una ***distribuzione di frequenza semplice*** viene costruita associando a ognuna delle modalità distinte che sono state osservate,

$$x_1, x_2, \ldots, x_j, \ldots, x_k$$

la corrispondente ***frequenza assoluta*** che è pari al numero di unità statistiche che presentano quella modalità. Per la *j*-esima modalità, la frequenza assoluta viene indicata con n_i .

- Una ***distribuzione di frequenza semplice*** viene rappresentata attraverso una tabella di questo tipo

Modalità (x_j)	Frequenze (n_j)
x_1	n_1
x_2	n_2
…	…
x_j	n_j
…	…
x_k	n_k
Totale	n

Esempio

- Dalla matrice di dati precedente, si possono ricavare 5 distribuzioni semplici secondo i caratteri: sesso, regione, scuola di provenienza, anno di corso e voto in Statistica.

- Per il carattere sesso, le modalità distinte sono M e F con frequenze pari, rispettivamente, a 2 e 3. La corrispondente distribuzione di frequenza è quindi:

Sesso (x_j)	Frequenze (n_j)
M	2
F	3
Totale	5

- Per gli altri 4 caratteri si ha:

Regione (x_j)	Frequenze (n_j)
Marche	3
Umbria	1
Calabria	1
Totale	5

Anno (x_j)	Frequenze (n_j)
II	2
III	2
F.C.	1
Totale	5

Scuola (x_j)	Frequenze (n_j)
Classico	1
Scientifico	2
Tecnico	2
Totale	5

Voto (x_j)	Frequenze (n_j)
26	1
27	1
28	1
30	2
Totale	5

Esempio

- Distribuzione delle famiglie per numero di componenti - Regione Marche - Censimento 2001.

NUMERO DI COMPONENTI	Numero di famiglie (in migliaia)
1 persona	124,143
2 persone	149,531
3 persone	124,394
4 persone	107,992
5 persone	31,751
6 o più persone	11,663
Totale	**549,474**

Esempio

- Distribuzione doppia delle famiglie per numero di componenti e per ripartizione geografica - Censimento 2001 (dati in migliaia).

RIPARTIZIONI GEOGRAFICHE	Numero di componenti						
	1 persona	2 persone	3 persone	4 persone	5 persone	6 o più persone	Totale
Italia Nord-Occidentale	1.767,208	1.840,037	1.390,009	966,118	207,367	46,461	6.217,200
Italia Nord-Orientale	1.116,042	1.208,212	962,636	701,273	184,009	59,838	4.232,010
Italia Centrale	1.061,905	1.188,248	941,315	780,561	208,574	61,596	4.242,199
Italia Meridionale	940,888	1.100,449	935,550	1.150,759	474,806	145,822	4.748,274
Italia Insulare	541,578	568,465	476,696	537,495	191,070	55,689	2.370,993
Italia	**5.427,621**	**5.905,411**	**4.706,206**	**4.136,206**	**1.265,826**	**369,406**	**2.181,0676**

Suddivisione in classi

- Nel caso di un *carattere quantitativo che assume molte modalità* (tipicamente continuo) è conveniente considerare delle ***classi*** al posto delle singole modalità distinte.

- Ogni classe viene identificata da due ***estremi*** (*di sinistra* e *di destra*) che per la *j*-esima classe sono indicati con $c_{j-1} - c_j$.

- Le classi vanno scelte in modo che:
  - il livello di sintesi sia adeguato
  - siano tra loro disgiunte
  - siano esaustive

- Una ***distribuzione di un carattere in classi*** viene rappresentata attraverso una tabella di questo tipo

Classi ($c_{j-1} - c_j$)	Frequenze (n_j)
$c_0 - c_1$	n_1
$c_1 - c_2$	n_2
…	…
$c_{j-1} - c_j$	n_j
…	…
$c_{k-1} - c_k$	n_k
Totale	n

Esempio

- Per il carattere voto in Statistica abbiamo la seguente distribuzione in classi.

Classi di voto	c_{j-1}	c_j	Frequenze (n_j)
18-21	17,5	21,5	0
22-25	21,5	25,5	0
26-28	25,5	28,5	3
29-30	28,5	30,5	2
Totale	-	-	5

- In questo caso, al fine di avere estremi di classe coincidenti, si effettua la ***correzione per continuità***: si sottrae 1/2 all'estremo di sinistra di ogni classe e lo si aggiunge a quello di destra.

Esempio

- Popolazione residente per classi di età - Regione Marche - Censimento 2001

CLASSI DI ETÀ	Popolazione residente (in migliaia)
Meno di 15	189,811
15-24	154,228
25-34	216,417
35-44	218,079
45-54	193,915
55-64	177,476
65-74	168,371
75 e più	152,284
Totale	**1.470,581**

Frequenze relative e percentuali

Famiglie per numero di componenti - Italia Settentrionale - Censimento 2001.

NUMERO DI COMPONENTI	Numero di famiglie (x 1000)
1 persona	2.883,250
2 persone	3.048,249
3 persone	2.352,645
4 persone	1.667,391
5 persone	391,376
6 o più persone	106,299
Totale	**10.449,210**

Famiglie per numero di componenti - Italia Meridionale e Isole - Censimento 2001.

NUMERO DI COMPONENTI	Numero di famiglie (x 1000)
1 persona	1.482,466
2 persone	1.668,914
3 persone	1.412,246
4 persone	1.688,254
5 persone	665,876
6 o più persone	201,511
Totale	**7.119,267**

	Italia Settentrionale		Italia Meridionale	
NUMERO DI COMPONENTI	Numero di famiglie	Frequenze relative	Numero di famiglie	Frequenze relative
1 persona	2.883,250	0,276	1.482,466	0,208
2 persone	3.048,249	0,292	1.668,914	0,234
3 persone	2.352,645	0,225	1.412,246	0,198
4 persone	1.667,391	0,160	1.688,254	0,237
5 persone	391,376	0,037	665,876	0,094
6 o più persone	106,299	0,010	201,511	0,028
Totale	**10.449,210**	**1**	**7.119,267**	**1**

	Italia Settentrionale		Italia Meridionale	
NUMERO DI COMPONENTI	Numero di famiglie	Frequenze percentuali	Numero di famiglie	Frequenze percentuali
1 persona	2.883,250	27,593	1.482,466	20,823
2 persone	3.048,249	29,172	1.668,914	23,442
3 persone	2.352,645	22,515	1.412,246	19,837
4 persone	1.667,391	15,957	1.688,254	23,714
5 persone	391,376	3,746	665,876	9,353
6 o più persone	106,299	1,017	201,511	2,831
Totale	**10.449,210**	**100**	**7.119,267**	**100**

Esercizio

Una popolazione di 100 individui, di cui 60 donne e 40 uomini, viene intervistata circa le attitudini al fumo ottenendo le seguenti risposte:

Uomini: 0 0 1 0 0 0 1 0 1 1 1 0 0 0 0 1 0 0 1 0 0 1 1 0 0 0 0 0 0 1 0 0 0 1 1 0 0 1 0

Donne: 1 0 0 0 0 1 0 0 1 1 0 0 0 0 1 0 0 1 1 0 0 0 0 0 1 0 0 0 0 0 0 0 1 0 1 0 0 0 0 0 0 1 0 1 0 0 0 0 0 1 0 0 0 0 0 0 1

dove 1=fumatore e 0=non fumatore.

a) Ricavare una tabella di distribuzione doppia
b) Ricavare la distribuzione di frequenza relativa rispetto al carattere “Attitudine al fumo” separatamente per gli uomini e per le donne.
c) Determinare la percentuale di fumatori tra gli intervistati.

- Frequenza relativa: $f_j = \frac{n_j}{n}$
- Frequenza percentuale: $p_j = \frac{n_j}{n} \cdot 100$
- Un'ovvia *proprietà* delle frequenze relative e percentuali è:

$$f_1 + f_2 + \cdots + f_k = \sum_{j=1}^{k} f_j = 1$$

$$p_1 + p_2 + \cdots + p_k = \sum_{j=1}^{k} p_j = 100$$

- Una ***distribuzione di frequenze assolute, relative e percentuali*** viene rappresentata attraverso una tabella di questo tipo

Modalità (x_j)	Frequenze assolute (n_j)	Frequenze relative (f_j)	Frequenze percentuali (p_j)
x_1	n_1	f_1	p_1
x_2	n_2	f_2	p_2
…	…	…	…
x_j	n_j	f_j	p_j
…	…	…	…
x_k	n_k	f_k	p_k
Totale	n	1	100

Esempio

- Per il carattere voto in Statistica abbiamo la seguente distribuzione in classi.

Classi di voto	Frequenze assolute (n_j)	Frequenze relative (f_j)	Frequenze percentuali (f_j)
18-21	0	0	0
22-25	0	0	0
26-28	3	0,6	60
29-30	2	0,4	40
Totale	5	1	100

Frequenze cumulate

Famiglie per numero di componenti - Italia Settentrionale - Censimento 2001.

NUMERO DI COMPONENTI	Numero di famiglie (x 1000)
1 persona	2.883,250
2 persone	3.048,249
3 persone	2.352,645
4 persone	1.667,391
5 persone	391376
6 o più persone	106,299
Totale	**10.449,210**

- Quante sono le famiglie con meno di due componenti?
- Quante sono le famiglie con meno di tre componenti?

NUM. COMP.	Italia Settentrionale Num. famiglie	Freq. ass. cum.	Freq. rel. cum.	Freq. perc. cum.	Italia Meridionale Num. famiglie	Freq. ass. cum.	Freq. rel. cum.	Freq. perc. cum.
1	2.883,250	2.883,250	0,276	27,593	1.482,466	1.482,466	0,208	20,823
2	3.048,249	5.931,499	0,568	56,765	1.668,914	3.151,380	0,443	44,266
3	2.352,645	8.284,144	0,793	79,280	1.412,246	4.563,626	0,641	64,102
4	1.667,391	9.951,535	0,952	95,237	1.688,254	6.251,880	0,878	87,816
5	391,376	10.342,911	0,990	98,983	665,876	6.917,756	0,972	97,169
6 o più	106,299	10.449,210	1	100	201,511	7.119,267	1	100
Totale	**10.449,210**				**7.119,267**			

- Frequenza assoluta cumulata :

$$N_j = n_1 + n_2 + \cdots + n_j = N_{j-1} + n_j$$

- Frequenza relativa cumulata:

$$F_j = \frac{N_j}{n} = f_1 + f_2 + \cdots + f_j = F_{j-1} + f_j$$

- Frequenza percentuale cumulata:

$$P_j = \frac{N_j}{n} \cdot 100 = p_1 + p_2 + \cdots + p_j = P_{j-1} + p_j$$

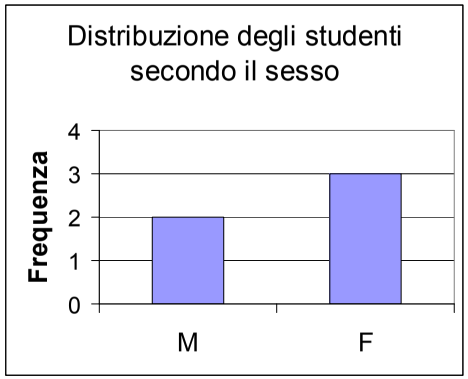
- Una ***distribuzione di frequenze assolute, relative e percentuali cumulate*** viene rappresentata attraverso una tabella di questo tipo

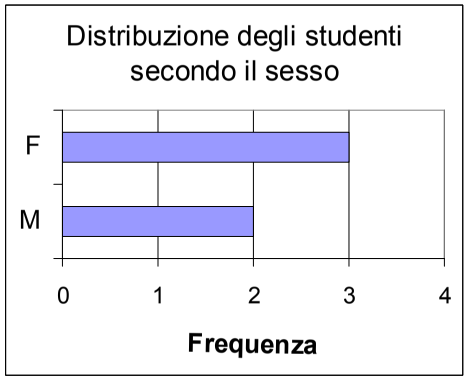
Modalità (x_j)	Frequenze assolute cumulate (N_j)	Frequenze relative cumulate (F_j)	Frequenze percentuali cumulate (P_j)
x_1	N_1	F_1	P_1
x_2	N_2	F_2	P_2
…	…	…	…
x_j	N_j	F_j	P_j
…	…	…	…
x_k	N_k	F_k	P_k

Rappresentazioni grafiche

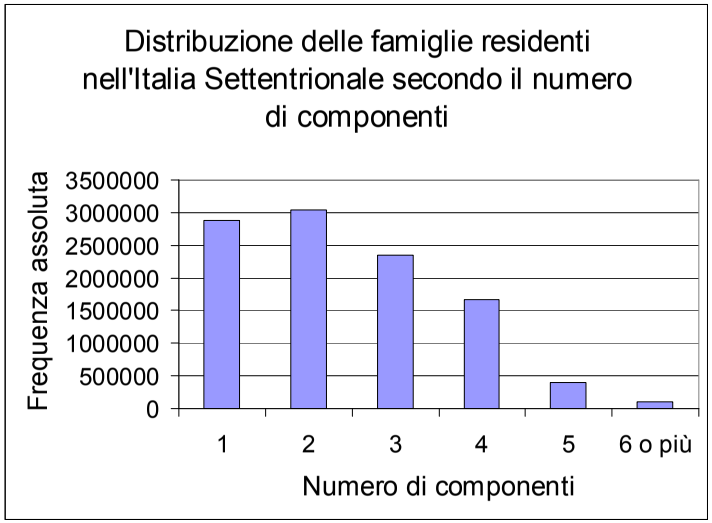
- Per una *distribuzione di frequenza* di un carattere **qualitativo** o **quantitativo discreto**, si utilizza un **grafico a barre** che consiste nel rappresentare, su un piano cartesiano, *k* barre di altezza $n_1, \ldots, n_k$ in corrispondenza delle ascisse $x_1, \ldots, x_k$.

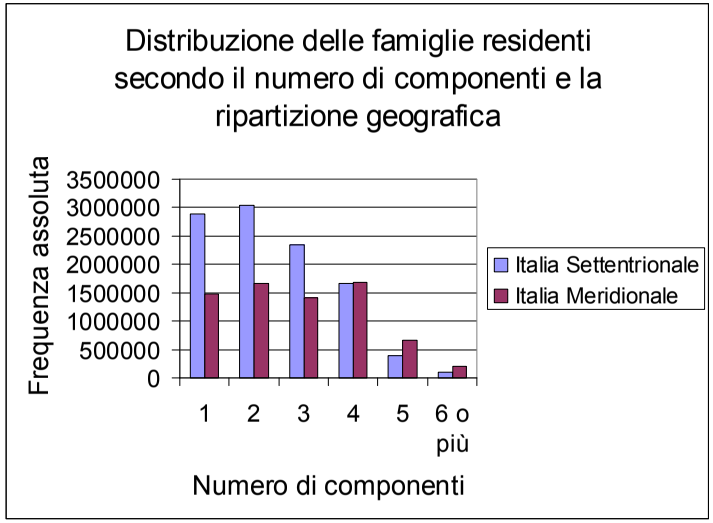
Esempio

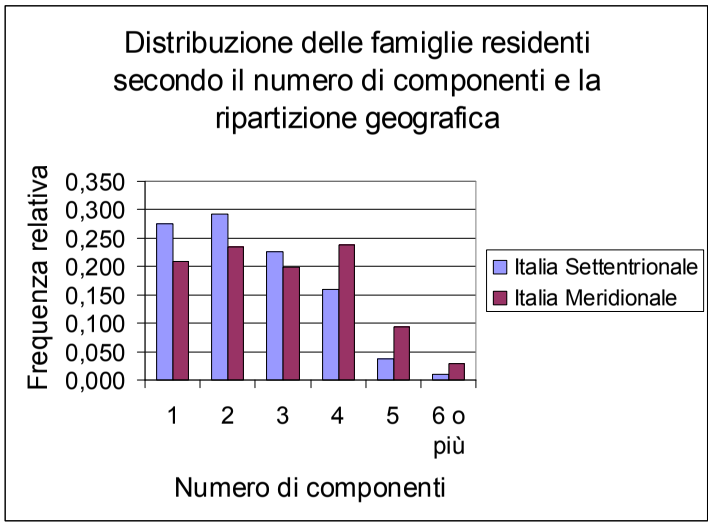




Esempi







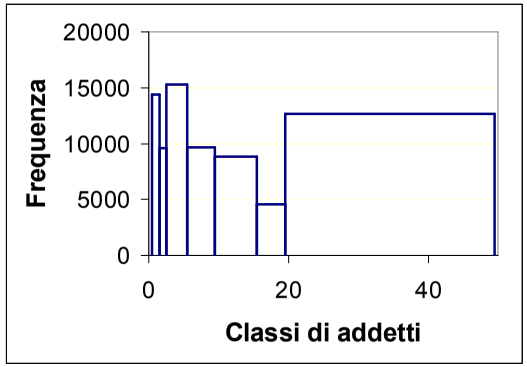
- Nel caso di un **carattere quantitativo in classi**, la distribuzione viene rappresentata tramite un **istogramma di frequenza** costruito tramite una serie di rettangoli corrispondenti alle varie classi. Il rettangolo corrispondente alla *j*-esima classe ha:
  - ***base*** pari all'ampiezza della classe: $a_j = c_j - c_{j-1}$
  - ***altezza*** pari alla densità di frequenza: $h_j = \dfrac{f_j}{a_j}$
- L'altezza della classe (densità) corrisponde alla frequenza che compete a un sottointervallo di ampiezza unitaria nel caso di uniforme distribuzione delle unità nelle classi. L'istogramma permette quindi di confrontare tra loro classi di diversa ampiezza.
- La caratteristica fondamentale dell'istogramma è che l'area di ogni rettangolo corrisponde alla frequenza della classe a cui si riferisce:

$$a_j \cdot h_j = a_j \cdot \frac{f_j}{a_j} = f_j$$

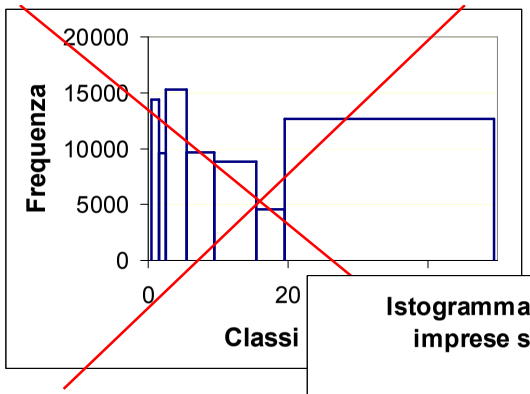
Esempio

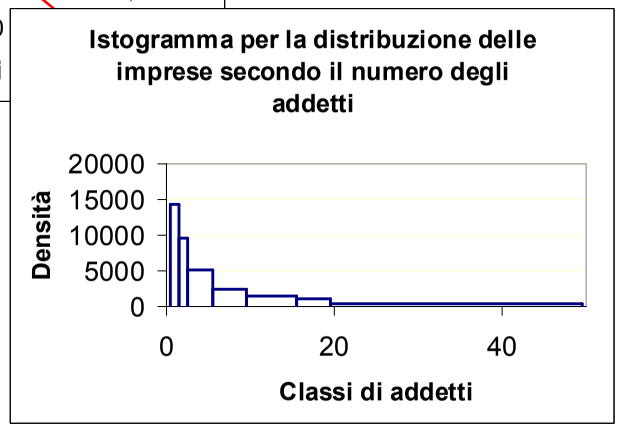
•Distribuzione in classi delle imprese della provincia di Macerata con meno di 50 addetti, secondo il numero di addetti – Censimento 2001

Classi di addetti	Frequenze assolute
1	14.349
2	9.588
3--5	15.263
6--9	9.651
10--15	8.837
16--19	4.570
20--49	12.653
Totale:	**74.911**

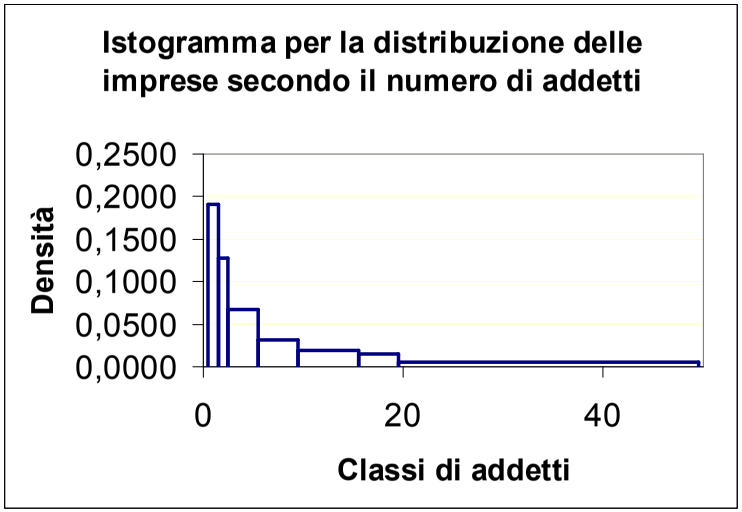


Classi di addetti	Frequenze assolute (n_j)	Ampiezza classi (a_j)	Densità di frequenza (h_j)
1	14.349	1	14.349,0
2	9.588	1	9.588,0
3--5	15.263	3	5.087,7
6--9	9.651	4	2.412,8
10--15	8.837	6	1.472,8
16--19	4.570	4	1.142,5
20--49	12.653	30	421,8
Totale:	**74.911**		





Classi di addetti	Frequenze relative (f_j)	Ampiezza classi (a_j)	Densità di frequenza (h_j)
0,5--1,5	0,1915	1	0,19155
1,5--2,5	0,1280	1	0,12799
2,5--5,5	0,2037	3	0,06792
5,5--9,5	0,1288	4	0,03221
9,5--15,5	0,1180	6	0,01966
15,5--19,5	0,0610	4	0,01525
19,5--49,5	0,1689	30	0,00563
Totale:	**1**		



Funzione di ripartizione

•La **funzione di ripartizione**, *F(x)* , fornisce la frequenza relativa delle osservazioni che presentano una modalità del carattere non superiore a *x*. Quindi si ha sempre:

$$F(-\infty) = \lim_{x\to-\infty} F(x) = 0 \qquad F(\infty) = \lim_{x\to\infty} F(x) = 1$$

•Per un carattere **qualitativo ordinato** o **quantitativo** non in classi, la funzione di ripartizione è pari a:

$$F(x) = \begin{cases} 0 & \forall\ x < x_1 \\ F_j & x_j < x < x_{j+1} \\ 1 & \forall\ x > x_k \end{cases}$$

quindi, se *x* è compreso tra la modalità più piccola (x_1) e quella più grande (x_k), *F(x)* è uguale alla frequenza cumulata (F_j) corrispondente alla più grande modalità (x_j) minore o uguale a *x*. Altrimenti, *F(x) = 0* o *F(x) = 1*.

Esempio

NUMERO DI COMPONENTI (x_j)	Numero di famiglie (n_j)	Frequenze relative (f_j)	Frequenze relative cumulate (F_j)
1 persona	2.883,250	0,276	0,276
2 persone	3.048,249	0,292	0,568
3 persone	2.352,645	0,225	0,793
4 persone	1.667,391	0,160	0,952
5 persone	391,376	0,037	0,990
6 persone	106,299	0,010	1
Totale	**10.449,210**	**1**	

•F(2) = 0,568; F(4) = 0,952; F(4,5) = 0,952.

•Nel caso di un **carattere in classi** (tipicamente continuo) ci si basa sull'ipotesi che in ogni classe ci sia uniforme distribuzione: si ha sempre la stessa frequenza in ogni sottointervallo della classe di ampiezza unitaria

•In questo caso la funzione di ripartizione è pari a:

$$F(x) = \begin{cases} 0 & \forall\ x < c_0 \\ F_{j-1} + h_j(x - c_{j-1}) & c_{j-1} < x < c_j \\ 1 & \forall\ x > c_k \end{cases}$$

e quindi, se *x* è compreso tra l'estremo sinistro della prima classe (c_0) e l'estremo destro dell'ultima classe (c_k), per il calcolo occorre innanzitutto individuare la classe che contiene *x* ($c_{j-1} – c_j$) e poi $F(x) = F_{j-1} + h_j(x - c_{j-1})$; altrimenti, *F(x) = 0* oppure *F(x) = 1*. Si noti che se *x* è uguale a un estremo di classe (c_j), si ha $F(x)= F_j$

Esempio

Imprese attive al 31.12.2010 per classi di fatturato (migliaia di euro) - dati relativi al bilanci anno 2009 - Comune di Padova

Classi di fatturato ($c_{j-1} - c_j$)	Frequenze assolute (n_j)	Frequenze relative (f_j)	Frequenze relative cumulate (F_j)	Densità (h_j)
0--250	2156	0,482219	0,482219	0,001929
250--500	666	0,14896	0,631179	0,000596
500--1.000	517	0,115634	0,746813	0,000231
1.000--2.500	539	0,120555	0,867367	8,04E-05
2.500--5.000	260	0,058153	0,92552	2,33E-05
5.000--10.000	171	0,038246	0,963766	7,65E-06
10.000--25.000	95	0,021248	0,985015	1,42E-06
25.000--50.000	32	0,007157	0,992172	2,86E-07
50.000--100.000	35	0,007828	1	1,57E-07
Totale	**4471**	**1**		

•F(500) = 0,631179; F(450) = 0,482219 + 0,000596(450-250) = 0,601387

Rappresentazione grafica della funzione di ripartizione

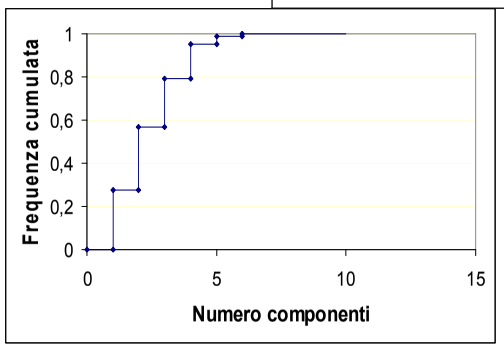
•Rappresentando la funzione di ripartizione si mostra l'andamento delle frequenze cumulate al variare della modalità del carattere.

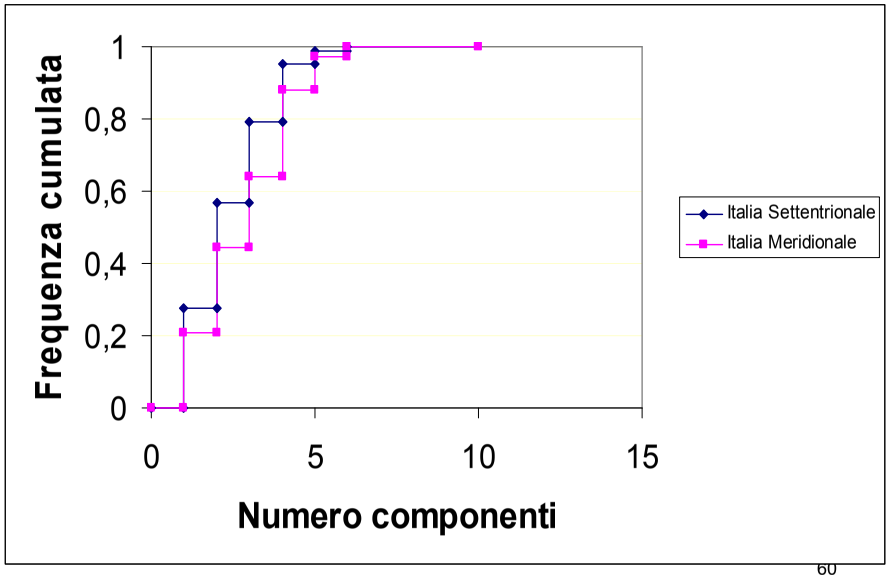
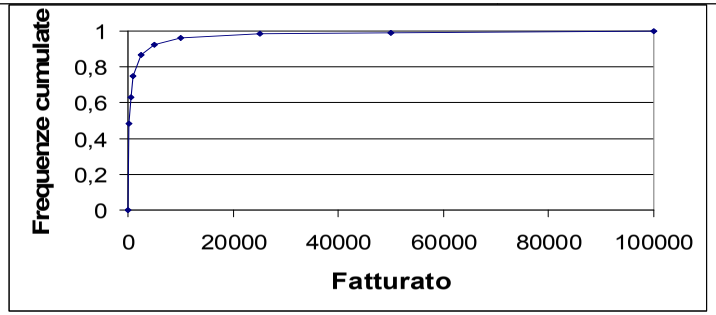
•Per un carattere **qualitativo ordinato** o **quantitativo** la funzione di ripartizione ha una forma "a gradini" ottenuta congiungendo i punti di coordinate (x_j , F_j), per $j = 1,\ldots,k$.

•Nel caso di un carattere in classi si congiungono i punti i coordinate (c_j ,F_j) , $j = 1,\ldots,k$, e il punto (c_{j-1},0) dando origine a una "spezzata".

Esempi

NUMERO DI COMPONENTI (x_j)	Numero di famiglie (n_j)	Frequenze relative (f_j)	Frequenze relative cumulate (F_j)
1 persona	2.883,250	0,276	0,276
2 persone	3.048,249	0,292	0,568
3 persone	2.352,645	0,225	0,793
4 persone	1.667,391	0,160	0,952
5 persone	391,376	0,037	0,990
6 persone	106,299	0,010	1
Totale	**10.449,210**	**1**	





Classi di fatturato ($c_{j-1} - c_j$)	Frequenze assolute (n_j)	Frequenze relative (f_j)	Frequenze relative cumulate (F_j)	Densità (h_j)
0--250	2156	0,482219	0,482219	0,001929
250--500	666	0,14896	0,631179	0,000596
500--1.000	517	0,115634	0,746813	0,000231
1.000--2.500	539	0,120555	0,867367	8,04E-05
2.500--5.000	260	0,058153	0,92552	2,33E-05
5.000--10.000	171	0,038246	0,963766	7,65E-06
10.000--25.000	95	0,021248	0,985015	1,42E-06
25.000--50.000	32	0,007157	0,992172	2,86E-07
50.000--100.000	35	0,007828	1	1,57E-07
Totale	**4471**	**1**