

# Settimana 1

G. M. Marchetti

febbraio 2018

## Organizzazione del corso

- Questo è il corso di Statistica 2018 CdL Economia B034
- lettere A-C B018993 B000319
- Docente: Giovanni Marchetti
- 9 crediti (6 ore a settimana per 12 settimane)

## Orario

---

### Giorni

---

Lunedì 8:30 - 10:00 D4/001

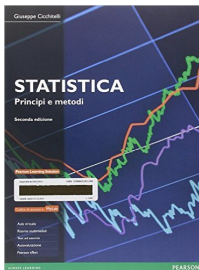
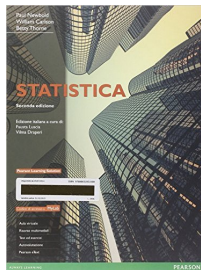
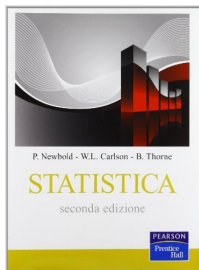
Giovedì 8:30 - 10:00 D5/003

Venerdì 8:30 - 10:00 D4/001

---

## Libro di testo

- **Newbold, Carlson, Thorne** Statistica, Principi e Metodi  
Pearson Prentice Hall
- 



- vecchie edizioni OK
- Useremo MYLAB, spiegazioni su Moodle.

## Materiale didattico e informazioni

- Pagina web con tantissime info su esame: Moodle, e-l.unifi.it (Non c'è password!)
- Pagina web personale con slide e esercizi:  
<http://local.disia.unifi.it/gmm/stateco.html>

## Ricevimento

- **Veloce** dopo la lezione
- **Efficace** per email `giovanni.marchetti@unifi.it`
- **Con calma** per appuntamento email (rispondo sempre)
- **Dipartimento di Statistica e Informatica, viale Morgagni, 59, FI (Chiedete all'ingresso) Stanza 1/25**

## Perché è importante studiare Statistica a Economia?

- Tutto il mondo economico è basato su valutazioni del rischio
- La Probabilità è il linguaggio per esprimere l'incertezza
- La Statistica è un modo essenziale per misurare l'incertezza e valutare il rischio usando i dati

## Intuito?

- Nessuno nasce con l'intuito statistico
- Gli esseri umani hanno già da piccoli un buon intuito per la grammatica.
- Ma l'intuito statistico invece deve essere educato.



## Pertanto ...

- Qualsiasi operatore economico deve imparare a migliorare la conoscenza della logica statistica
- Una buona conoscenza della statistica permette di evitare errori gravi di valutazione del rischio
- Molte informazioni presenti sui media si appoggiano su dati che spesso sono interpretati male

## Iniziamo a imparare le cose fondamentali

- lo spoglio dei dati
- la distribuzione di frequenza
- (questa è una domanda fatta sempre all'orale)

## Popolazione italiana per zona di residenza

Regione	Frequenza	Percentuale
Nord-ovest	15 551	26.5
Nord-est	11 119	18.9
Centro	11 321	19.3
Sud	14 087	24.0
Isole	6 672	11.4
ITALIA	58 751	100.0

## Come si arrotonda

- $15551/58751 = 0.26469337$
- Se vogliamo 3 decimali tronchiamo a 3 decimali dopo la “virgola” e guardiamo la prima cifra a destra

0.264[6]

- se è maggiore di 5 aumenta di 1 l'ultima cifra

0.265

- Se vuoi una percentuale sposta la virgola di due posti a destra

26.5

## Come si arrotonda

- $11119/58751 = 0.18925635$
- Se vogliamo 3 decimali tronchiamo a 3 decimali dopo la “virgola” e guardiamo la prima cifra a destra

0.189[2]

- se è minore di 5 lascia il numero come è

0.189

- Percentuale 18.9

## Come si arrotonda

- $6672/58751 = 0.11356402$
- Se vogliamo 3 decimali tronchiamo a 3 decimali dopo la “virgola” e guardiamo la prima cifra a destra

0.113[5]

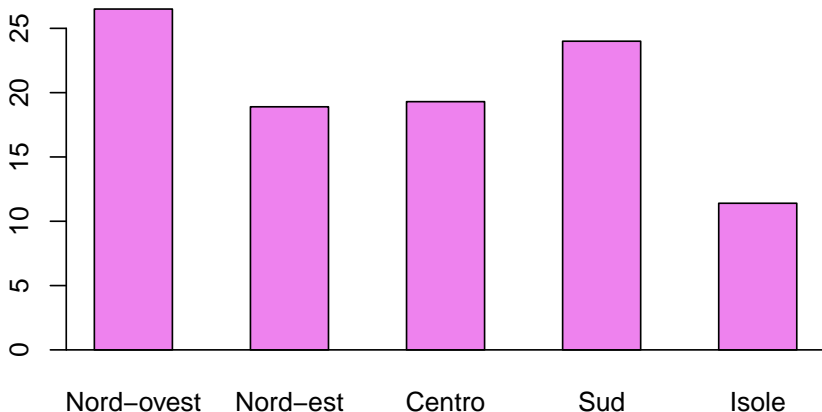
- se è **uguale** a 5 modifica se del caso l'ultima cifra per farla diventare pari.

0.114

- Percentuale 11.4

## Grafico “a barre”

- Si costruisce un grafico in coordinate cartesiane con dei segmenti di lunghezza uguale alla popolazione assoluta della classe o alla percentuale.



## Spoglio

- **Spoglio:** è la classificazione delle unità di osservazione a seconda delle modalità del carattere studiato.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
NE	S	S	C	NO	NO	NO	C	I	NE	S	NO	S	NE	NO

Si costruisce una tabella con le **modalità** della residenza e il totale:

Residenza	Ripetizioni	Frequenza
Nord-Ovest		
Nord-Est		
Centro		
Sud		
Isole		
Totale		



## Spoglio

- **Spoglio:** è la classificazione delle unità di osservazione a seconda delle modalità del carattere studiato.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
NE	S	S	C	NO	NO	NO	C	I	NE	S	NO	S	NE	NO

Quindi si elencano le osservazioni accanto a ciascuna modalità

Residenza	Ripetizioni	Frequenza
Nord-Ovest	NO NO NO NO NO	5
Nord-Est	NE NE NE	3
Centro	C C	2
Sud	S S S	4
Isole	I	1
Totale		

## Spoglio

- **Spoglio:** è la classificazione delle unità di osservazione a seconda delle modalità del carattere studiato.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
NE	S	S	C	NO	NO	NO	C	I	NE	S	NO	S	NE	NO

Infine si fanno il totale e le percentuali

Residenza	Frequenza	Percentuale
Nord-Ovest	5	33.3
Nord-Est	3	20.0
Centro	2	13.3
Sud	4	26.7
Isole	1	6.7
Totale	15	100.0

## Un altro esempio

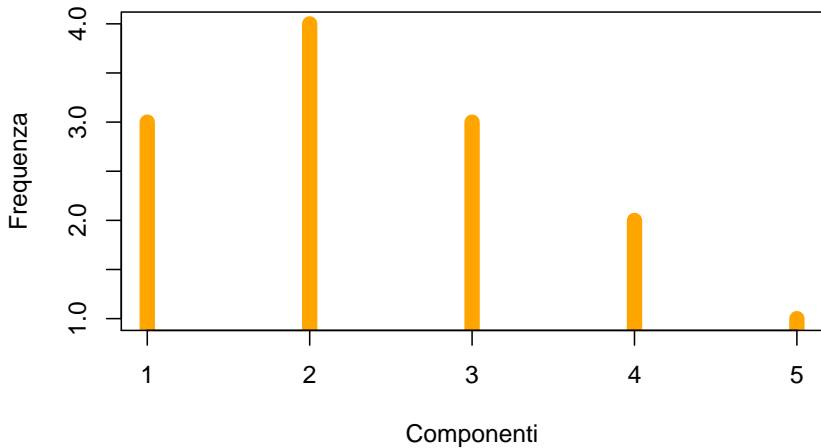
- Abbiamo 13 famiglie e vogliamo studiare il n. di componenti.
- I dati *individuali*:

Famiglie:                    1 2 3 4 5 6 7 8 9 10 11 12 13  
Numero di componenti: 4 2 2 3 2 1 1 3 5 2 4 3 1

- Costruiamo la distribuzione (= classificazione) delle famiglie
- Dati ordinati: 1 1 1 2 2 2 2 3 3 3 4 4 5
- Conto quante sono le famiglie (*dati aggregati*)

Componenti	Frequenze
1	3
2	4
3	3
4	2
5	1
Totale	13

## Diagramma a barre



## Tipi di dati

- Qualitativi (modalità attributi)
  - nominali (senza ordine: religione, zona di residenza)
  - ordinali (titolo di studio, stato di salute)
- Quantitativi (modalità numeri)
  - discreti (interi: numero di componenti della famiglia)
  - continui (reali: età, valore di un titolo, durata di una chiamata)

## Dati continui

Fatturato: assimilabile a continua

Azienda	Fatturato
NUOVO PIGNONE	2.859.626.881
UNICOOP	2.356.352.604
ARVAL	1.120.295.922
CODIFI	962.226.204
COMPUTER GROSS	939.076.664
ELI LILLY	748.533.015
GUCCI	457.272.325
CARAPELLI	425.278.000
BEYFIN	410.993.356
ARCELORMITTAL	378.700.044
MENARINI	366.501.275
SAMMONTANA	344.735.420

## Suddivisione in classi

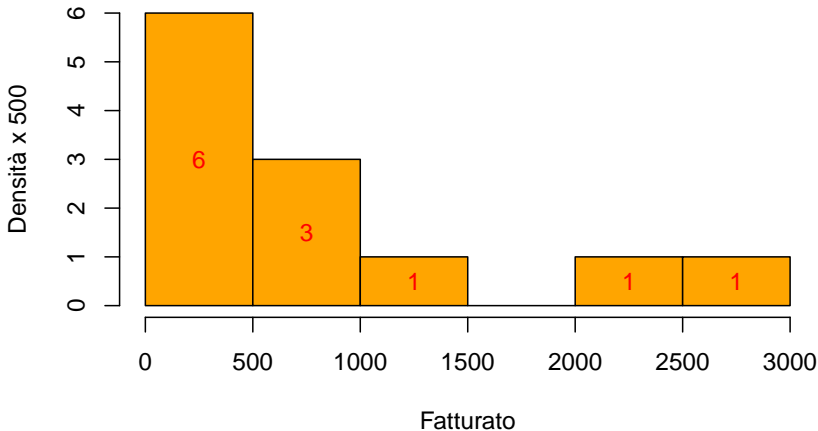
Classi di fatturato (arbitrarie) in milioni di Euro

$[0, 500)$ ,  $[500, 1000)$ ,  $[1000, 1500)$   
 $[1500, 2000)$ ,  $[2000, 2500)$ ,  $[2500, 3000)$

<i>Classi</i>	<i>Frequenza</i>
0 † 500	6
500 † 1000	3
1000 † 1500	1
1500 † 2000	0
2000 † 2500	1
2500 † 3000	1
<i>Totale</i>	12

## Istogramma di dati continui

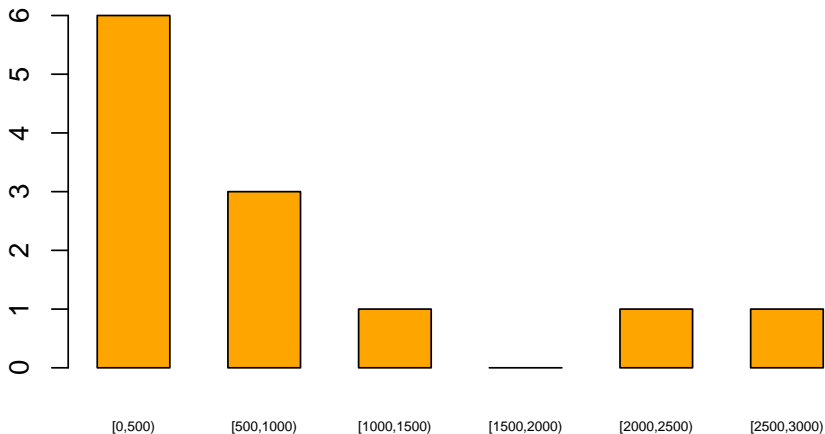
- Si definiscono classi di **ampiezza uguale**
- Si costruiscono dei rettangoli
  - con BASE = classe
  - e AREA = frequenza
- I rettangoli devono essere **affiancati**



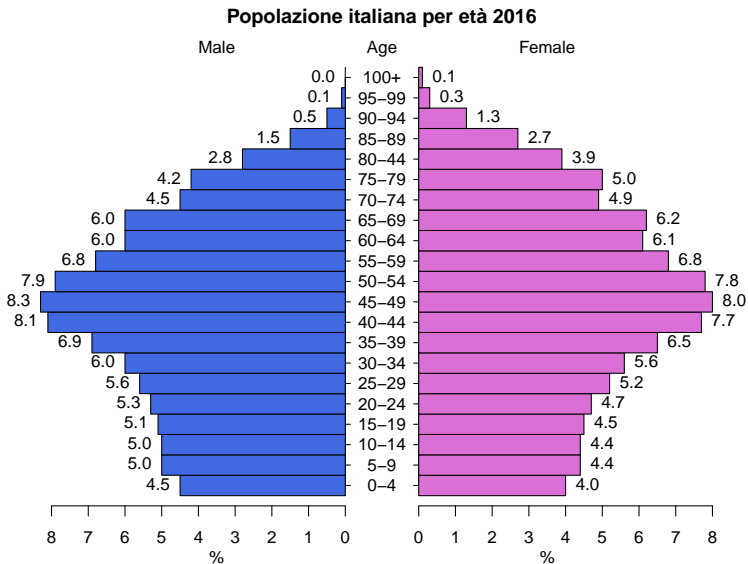


## Istogrammi e diagrammi a barre

- Istogrammi per dati continui
- Diagrammi a barre per dati discreti o qualitativi
- Grafico **sbagliato!** Tipicamente lo fa così Excel!!



## Variabili continue : età



## Vocabolario base della Statistica

- In statistica studiamo **fenomeni collettivi** che si osservano su un gran numero di **casi** o **unità statistiche**
- Il fenomeno studiato si chiama un **carattere**
- Il carattere si manifesta in certe **modalità**
- Esempio: Il carattere *numero di componenti* si rileva su tutte le famiglie italiane in un certo momento
- È un carattere quantitativo discreto con modalità intere  $\geq 1$

## Frequenze

- Per ogni modalità si può calcolare
  - la **frequenza assoluta** = n. di unità con la modalità
  - la **frequenza relativa**:

$$\frac{n. \text{ unità con la modalità}}{n. \text{ totale unità}}$$

- Esempio. Dati sul sesso di 14 studenti:

M M F M F M M M F F M M M M

Termine	Significato
Unità	studente
Carattere	sexso
Modalità	M, F
Frequenza M	10
Freq. rel. M	$10/14 = 0.71$

## Matrici di dati

- I dati statistici sono tipicamente misurazioni di un carattere rilevato su un insieme di unità
- Un campione di voti di 10 studenti partecipanti a un test d'ingresso a Economia a FI. (Totale = 1128)

Mat	Test	Sesso	Scuola
88	13.50	F	T
66	11.50	F	T
76	12.50	F	L
80	7.75	F	P
68	11.75	F	L
74	10.00	M	L
91	21.75	M	L
80	14.75	M	T
74	18.75	M	L
74	12.50	M	T

## Variabili

- Le variabili (le colonne della matrice) sono
  - Mat: Voto alla Maturità
  - Test: Punteggio al test d'ingresso
  - Sesso: Sesso(M,F)
  - Scuola: Tipo di scuola (L = liceo, T = tecnico, P = professionale, A = altro)
- Questa rappresentazione è chiamata *matrice dei dati*.

## Analisi univariate

- Ricorda! Distinguiamo tra
  - analisi di variabili qualitative
  - analisi di variabili quantitative

## Indici di posizione di dati quantitativi

- Si dicono indici di posizione quelli che rappresentano un valore centrale della distribuzione osservata.
- La *media aritmetica*  $\mu$  di una variabile quantitativa  $X$  rilevata su una popolazione di  $N$  unita è definita da

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

- Ha un'interpretazione semplice: è ottenuta equiripartendo il totale  $\sum_i x_i$  della variabile.



## Esempio

- Il voto medio alla maturità:  $\mu_{mat} = 77.2$ .
- Il voto medio al punteggio al test è  $\mu_{test} = 13.3$ .
- Nota: NON si calcolano medie di dati qualitativi

## Alcune proprietà utili:

- È sempre compresa tra il minimo e il massimo

$$60 \leq \text{Voto medio Maturità} \leq 100$$

- La somma degli scarti dalla media è sempre zero. Verifica:

<i>Voti</i> :	70	60	70	100	80	<i>Media</i> = 76
<i>Scarti</i> :	-6	-16	-6	24	4	<i>Somma</i> = 0
$x_i$ :	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	
<i>Scarti</i> :	$x_1 - \mu$	$x_2 - \mu$	$x_3 - \mu$	$x_4 - \mu$	$x_5 - \mu$	<i>Somma</i> = 0

- Infatti

$$\text{Somma} = (x_1 + \dots + x_5) - 5\mu = 5\mu - 5\mu = 0.$$

## La media è il valore più vicino ai dati

- Soddisfa il **criterio dei minimi quadrati**
- Consideriamo i dati precedenti

Voti : 70 60 70 100 80    *Media* = 76

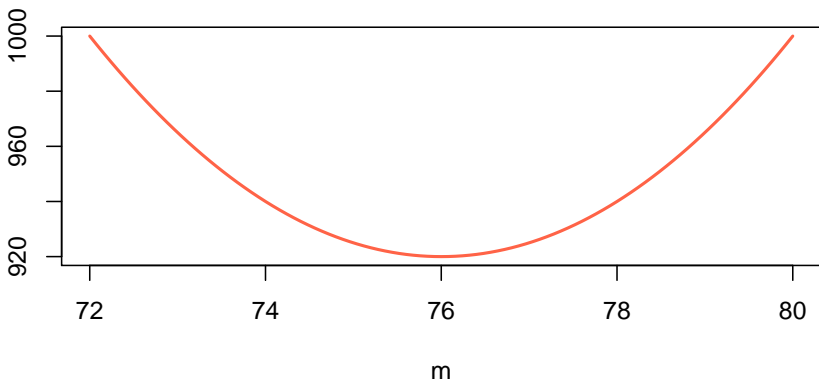
- Il valore  $m$  che minimizza

$$d = (70 - m)^2 + (60 - m)^2 + (70 - m)^2 + (100 - m)^2 + (80 - m)^2$$

è proprio la media aritmetica  $m = 76$ !

## Dimostrazione

- Infatti  $d = 29800 - 760m + 5m^2$
- che è una parabola col vertice in  $m = 76$



## Media calcolata da dati aggregati

- Qual è il n. medio di componenti della famiglia?

Componenti	Frequenza	Totale Parziale
1	20	20
2	30	60
3	20	60
4	10	40
Totale	80	180

- Il numero medio è il totale dei componenti diviso per il numero di famiglie

$$\mu = \frac{1 \cdot 20 + 2 \cdot 30 + 3 \cdot 20 + 4 \cdot 10}{80} = 180/80 = 2.25$$

## Media calcolata dalla distribuzione di frequenza relativa

- Nota che

$$\bar{x} = 1 \cdot \frac{20}{80} + 2 \cdot \frac{30}{80} + 3 \cdot \frac{20}{80} + 4 \cdot \frac{10}{80} = 180/80 = 2.25$$

Componenti	Freq. rel.	Totale Parziale
1	0.250	0.25
2	0.375	0.75
3	0.250	0.75
4	0.125	0.50
Totale	1.000	2.25

- Quindi

$$\mu = x_1 f_1 + \dots + x_4 f_4$$

## Media di tassi di interesse

- La *media geometrica* è un indice di posizione per dati positivi  $x_i > 0$  che hanno una natura moltiplicativa.
- Supponiamo di avere un capitale di 100 Euro che è investito e rende

Anno	Tasso	Capitale
0	-	100
1	0.25	$100 \times 1.25 = 125$
2	0.20	$125 \times 1.20 = 150$
3	0.02	$150 \times 1.02 = 153$

- Qual è il tasso medio?
- È ragionevole calcolare  $(0.25 + 0.20 + 0.02)/3 = 0.16$ ?
- NO. Perché? Perché i tassi non sono additivi.

- Si ragiona così: qual è il tasso costante che applicato al capitale di 100 lo fa diventare 153 alla fine dei tre anni?
- Se  $g$  è questo tasso, dopo 3 anni a interesse composto il capitale è

$$100 \cdot (1 + g)^3$$

- Quindi per trovare  $g$  bisogna risolvere l'equazione

$$100 \cdot (1 + g)^3 = 153$$

- Soluzione

$$1 + g = \sqrt[3]{153/100} = 1.152 \quad \therefore g = 0.152$$

- Questo tipo di media è la **media geometrica**.



## Media calcolata in una popolazione e in un campione

- La media calcolata per una popolazione di dimensione  $N$  è

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

- La media calcolata su un campione di  $n$  elementi con  $n < N$  con l'intenzione di stimare  $\mu$  è

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