



Facoltà di Ingegneria
Università degli Studi di Firenze
Dipartimento di Elettronica e Telecomunicazioni

Misure Elettriche

1. Statistica descrittiva

Ing. Andrea Zanobini

Dipartimento di Elettronica e Telecomunicazioni

Distribuzioni di frequenza

Quando si raccolgono dei dati su una popolazione o su un campione, i valori ottenuti si presentano allo statistico come un insieme di dati disordinati; i dati che non sono stati organizzati, sintetizzati o elaborati in qualche modo sono chiamati **dati grezzi**. A meno che il numero delle osservazioni sia piccolo, è improbabile che i dati grezzi forniscano qualche informazione finché non siano stati ordinati in qualche modo.

In ciascuno degli esempi si osserva una variabile, che è rispettivamente
1 – il numero di particelle rilevate in un intervallo di un minuto;
2 – la quantità di gas inquinante emesso in un giorno;
3 – la causa di un guasto verificato.

Esempio 1

Rilevando con uno strumento di misurazione il numero di particelle cosmiche in 40 periodi consecutivi di un minuto si ottengono i seguenti dati

0	2	1	4	3	1	2	3	8	2	5	2	1	3	3	1	3	2	2	5
4	4	4	2	3	5	5	1	1	2	4	4	2	3	3	3	3	3	3	2

Esempio 2

I seguenti dati sono il risultato di 80 determinazioni, in una data unità di misura, dell'emissione giornaliera di un gas inquinante da un impianto industriale

15.8	26.4	17.3	11.2	23.9	24.8	18.7	13.9	9.0	13.2
22.7	9.8	6.2	14.7	17.5	26.1	12.8	28.6	17.6	23.7
26.8	22.7	18.0	20.5	11.0	20.9	15.5	19.4	16.7	10.7
19.1	15.2	22.9	26.6	20.4	21.4	19.2	21.6	16.9	19.0
18.5	23.0	24.6	20.1	16.2	18.0	7.7	13.5	23.5	14.5
14.4	29.6	19.4	17.0	20.8	24.3	22.5	24.6	18.4	18.1
8.3	21.9	12.3	22.3	13.3	11.8	19.3	20.0	25.7	31.8
25.9	10.5	15.9	27.5	18.1	17.9	9.4	24.1	20.1	28.5

Esempio 3

In uno stabilimento vengono registrati i casi di malfunzionamento di una macchina utensile controllata dal computer, e le loro cause. I dati relativi a un certo mese sono i seguenti

fluttuazioni di tensione	6
instabilità del sistema di controllo	22
errore dell'operatore	13
strumento usurato e non sostituito	2
altre cause	5
<i>Totale</i>	48

Le variabili oggetto di rilevazioni statistiche si classificano in più tipi diversi, a seconda del tipo di valori che assumono

variabili	{	numeriche (quantitative)	{ discrete
		non numeriche (qualitative)	continue

Per studiare i dati degli esempi precedenti dividiamo i dati stessi in **classi** e determiniamo il numero di individui appartenenti a ciascuna classe, detto **frequenza della classe**. Costruiamo poi la **tabella di distribuzione di frequenza**, ossia una tabella che raccoglie i dati secondo le classi e le corrispondenti frequenze.

I dati ordinati e riassunti nella tabella di distribuzione di frequenza sono detti **dati raggruppati**.

Variabili numeriche discrete

Nella tabella la prima colonna indica la **classe**; la seconda la **frequenza assoluta**, detta anche semplicemente **frequenza di classe**, ossia il numero di osservazioni che cadono in ciascuna classe; la terza colonna la **frequenza relativa**, ossia il rapporto tra frequenza assoluta e numero totale di osservazioni (in questo caso 40); la quarta è la **frequenza percentuale**, ossia la frequenza relativa moltiplicata per 100.

Variabili numeriche discrete

0	2	1	4	3	1	2	3	8	2	5	2	1	3	3	1	3	2	2	5
4	4	4	2	3	5	5	1	1	2	4	4	2	3	3	3	3	3	3	2

<i>Classe</i>	<i>Freq. assoluta</i>	<i>Freq. relativa</i>	<i>Freq. percentuale</i>
0	1	0.025	2.5%
1	6	0.15	15%
2	10	0.25	25%
3	12	0.3	30%
4	6	0.15	15%
5	4	0.1	10%
6	0	0	0%
7	0	0	0%
8	1	0.025	2.5%
<i>Totale</i>	40	1	100%

Variabili numeriche continue

Nell'esempio 2 la variabile osservata è continua. I valori dei dati sono compresi tra 6.2 e 31.8; il **campo di variazione** o **range** dei dati, cioè la differenza tra il più grande e il più piccolo, vale

$$R = 31.8 - 6.2 = 25.6$$

Troppe classi rendono la tabella poco leggibile; troppo poche classi la rendono poco significativa: il numero delle classi è normalmente compreso fra 5 e 15; se i dati sono molto numerosi si può arrivare a usare un massimo di 20 classi. Una semplice regola pratica che si rivela a volte utile

$$N = 1 + 3.322 \cdot \log_{10} 80 \cong 7$$

$$a = \frac{25.6}{7} \cong 3.7$$

<i>Classe</i>	<i>Freq. assoluta</i>	<i>Freq. relativa</i>	<i>Freq. percentuale</i>
$5 \leq x < 9$	3	0.0375	3.75%
$9 \leq x < 13$	10	0.1250	12.5%
$13 \leq x < 17$	14	0.1750	17.5%
$17 \leq x < 21$	25	0.3125	31.25%
$21 \leq x < 25$	17	0.2125	21.25%
$25 \leq x < 29$	9	0.1125	11.25%
$29 \leq x < 33$	2	0.0250	2.5%
<i>Totale</i>	80	1	100%

Variabili non numeriche

fluttuazioni di tensione	6
instabilità del sistema di controllo	22
errore dell'operatore	13
strumento usurato e non sostituito	2
altre cause	5
<i>Totale</i>	48

<i>Classe</i>	<i>Freq. assoluta</i>	<i>Freq. relativa</i>	<i>Freq. percentuale</i>
<i>fluttuazioni tensione</i>	6	0.125	12.5%
<i>instabilità</i>	22	0.458	45.8%
<i>errore operatore</i>	13	0.271	27.1%
<i>strumento</i>	2	0.042	4.2%
<i>altro</i>	5	0.104	10.4%
<i>Totale</i>	48	1	100%

Distribuzioni cumulative di frequenza

Ci sono altri modi di raggruppare i dati: ad esempio dati “minori di”, “maggiori di”; si ottengono in questo modo le **distribuzioni cumulative**.

La frequenza totale di tutti i valori minori del limite superiore di una data classe è detta **frequenza cumulativa**. Una tabella che presenti frequenze cumulative è detta **tabella di distribuzione cumulativa di frequenza**.

Variabili numeriche discrete

<i>Classe</i>	<i>Freq. assoluta</i>	<i>Freq. relativa</i>	<i>Freq. percentuale</i>
0	1	0.025	2.5%
1	6	0.15	15%
2	10	0.25	25%
3	12	0.3	30%
4	6	0.15	15%
5	4	0.1	10%
6	0	0	0%
7	0	0	0%
8	1	0.025	2.5%
<i>Totale</i>	40	1	100%

<i>Classe</i>	<i>Freq. cumul. assoluta</i>
$x \leq 0$	1
$x \leq 1$	7
$x \leq 2$	17
$x \leq 3$	29
$x \leq 4$	35
$x \leq 5$	39
$x \leq 6$	39
$x \leq 7$	39
$x \leq 8$	40

Distribuzioni cumulative di frequenza

Variabili numeriche continue

<i>Classe</i>	<i>Freq. assoluta</i>	<i>Freq. relativa</i>	<i>Freq. percentuale</i>
$5 \leq x < 9$	3	0.0375	3.75%
$9 \leq x < 13$	10	0.1250	12.5%
$13 \leq x < 17$	14	0.1750	17.5%
$17 \leq x < 21$	25	0.3125	31.25%
$21 \leq x < 25$	17	0.2125	21.25%
$25 \leq x < 29$	9	0.1125	11.25%
$29 \leq x < 33$	2	0.0250	2.5%
<i>Totale</i>	80	1	100%

<i>Classe</i>	<i>Freq. cumul. assoluta</i>
$x < 5$	0
$x < 9$	3
$x < 13$	13
$x < 17$	27
$x < 21$	52
$x < 25$	69
$x < 29$	78
$x < 33$	80

**Fare da soli
esempi 7,8,9 !!**

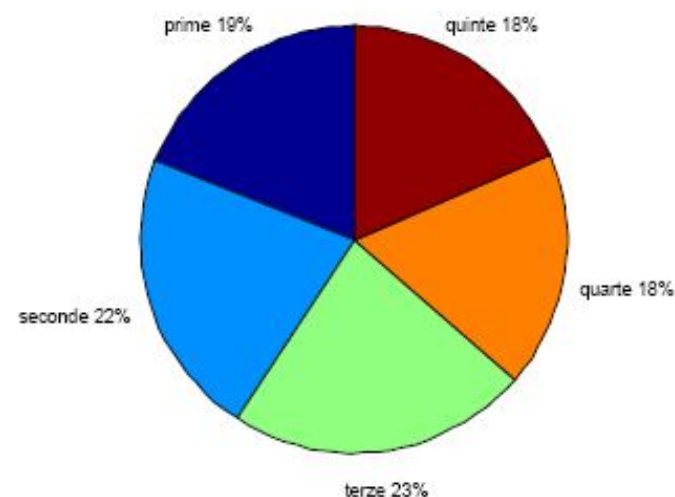
Grafici delle distribuzioni di frequenza

Diagramma circolare

Un primo tipo di diagramma è il **diagramma circolare**; in questo diagramma le frequenze percentuali sono rappresentate da settori circolari aventi ampiezze proporzionali alle frequenze stesse; indicando con f la frequenza percentuale e con g l'ampiezza in gradi, si ha

$$f : 100 = g : 360^\circ$$

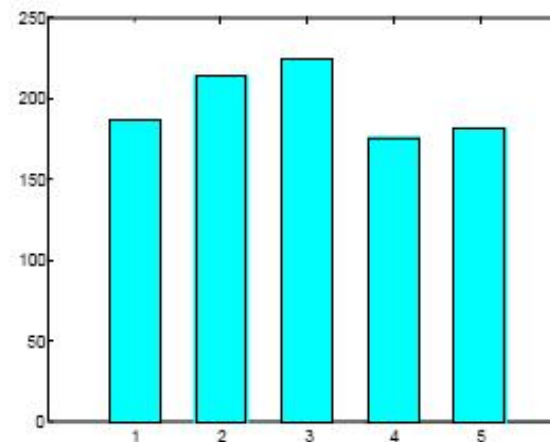
<i>Studenti iscritti ai diversi anni di corso</i>		
	<i>freq. assoluta</i>	<i>freq. percentuale</i>
<i>classi prime</i>	187	19.00%
<i>classi seconde</i>	214	21.75%
<i>classi terze</i>	225	22.87%
<i>classi quarte</i>	176	17.89%
<i>classi quinte</i>	182	18.50%
<i>Totale</i>	984	100.01%



@ minitab

Diagramma a barre

<i>Studenti iscritti ai diversi anni di corso</i>		
	<i>freq. assoluta</i>	<i>freq. percentuale</i>
<i>classi prime</i>	187	19.00%
<i>classi seconde</i>	214	21.75%
<i>classi terze</i>	225	22.87%
<i>classi quarte</i>	176	17.89%
<i>classi quinte</i>	182	18.50%
<i>Totale</i>	984	100.01%



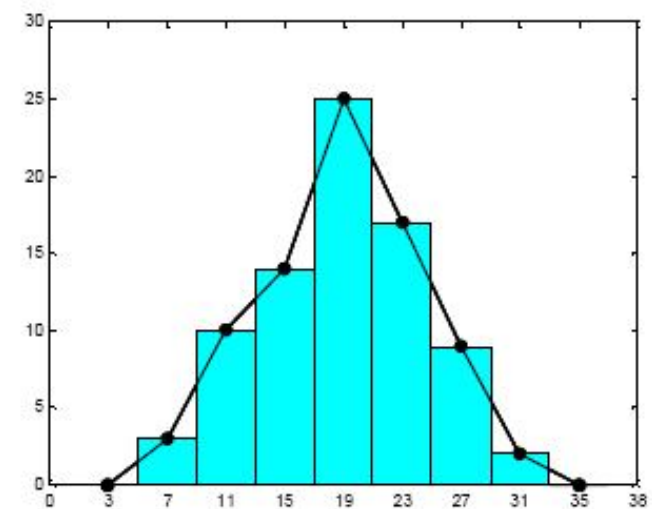
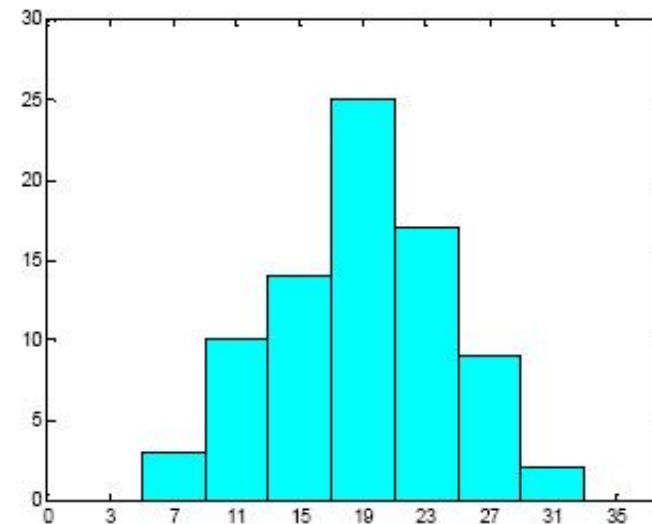
@ minitab

Fare il diagramma a barre nel caso delle particelle anche con minitab!!

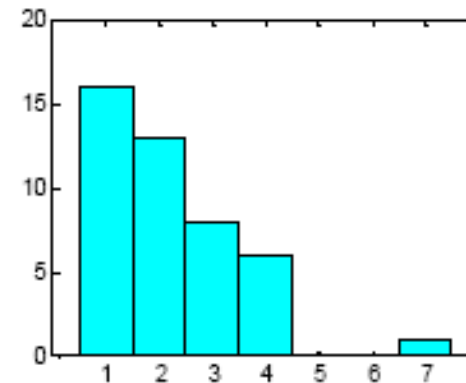
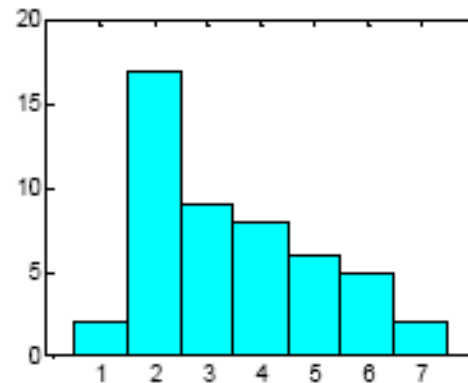
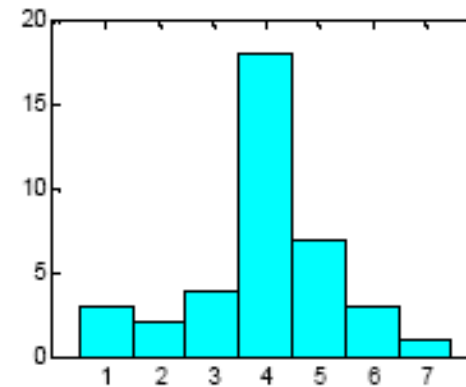
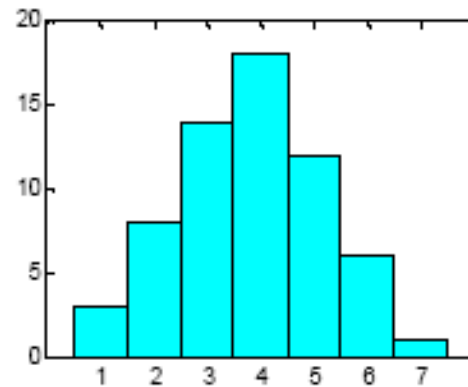
Istogrammi e poligoni di frequenza

<i>Classe</i>	<i>Valori centrali</i>	<i>Freq. assoluta</i>
$5 \leq x < 9$	7	3
$9 \leq x < 13$	11	10
$13 \leq x < 17$	15	14
$17 \leq x < 21$	19	25
$21 \leq x < 25$	23	17
$25 \leq x < 29$	27	9
$29 \leq x < 33$	31	2
<i>Totale</i>		80

**Fare da soli esempi
da 11 a 19 anche
utilizzando minitab!!**



Come misurare quantitativamente le caratteristiche osservate sugli istogrammi seguenti ??



Indici di posizione e di dispersione

Media

Definizione 1

Si definisce **media aritmetica** o **media campionaria** di n dati $x_1, x_2, \dots, x_n$ la quantità

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (1.1)$$

Esempio 20

Media dei dati

$$\bar{x} = \frac{15 + 14 + 2 + 27 + 13}{5} = 14.2$$

Per ogni valore x_i della variabile x si definisce lo **scarto dalla media**

$$s_i = x_i - \bar{x}$$

che indica il grado di scostamento del singolo valore x_i dalla media $\bar{x}$.

Si dimostra facilmente che la somma algebrica S degli scarti dalla media è nulla. Infatti

$$S = \sum_{i=1}^n s_i = \sum_{i=1}^n (x_i - \bar{x}) = \sum_{i=1}^n x_i - \sum_{i=1}^n \bar{x} = n\bar{x} - n\bar{x} = 0$$

Mediana

Definizione 2

La **mediana** M di un insieme di n dati ordinati in ordine di grandezza crescente è il valore centrale dei dati, se il numero di dati è dispari, o la media aritmetica dei due valori centrali, se il numero dei dati è pari.

Esempio 21

a – Mediana dei dati

Dati ordinati in ordine crescente 15 14 2 27 13

Mediana 2 13 14 15 27

$$M = 14$$

b – Mediana dei dati

Dati ordinati in ordine crescente 11 9 17 19 4 15

Mediana 4 9 11 15 17 19

$$M = \frac{11 + 15}{2} = 13$$

La mediana è preferibile alla media quando si vogliono eliminare gli effetti di valori estremi molto diversi dagli altri dati: la ragione è che la mediana non utilizza tutti i dati, ma solo il dato centrale o i due dati centrali.

Moda

Definizione 3

La **moda** $\tilde{x}$ di un insieme di n dati è il valore o la classe a cui corrisponde la massima frequenza assoluta.

Esempio 22

a – Moda dell'insieme di dati

3, 3, 5, 4, 7, 7, 7, 9, 2, 1

L'insieme ha moda $\tilde{x} = 7$.

b – Moda dell'insieme di dati

3, 3, 3, 5, 4, 7, 7, 7, 9, 2, 1

L'insieme ha due mode $\tilde{x} = 3$ e $\tilde{x} = 7$.

c – L'insieme di dati

3, 5, 4, 7, 8, 6, 9, 2, 1

non ha moda, perché ogni dato si presenta una sola volta.

**Fare da soli
esempi 23, 24
25 anche
utilizzando
minitab!!**

Quantili e percentili

Definizione 4

Il **primo quartile** Q_1 è un valore tale che il 25 % dei dati ordinati è minore o uguale a Q_1 . Il primo quartile Q_1 è detto anche **25-esimo percentile** e indicato con $P_{0.25}$.

Il **terzo quartile** Q_3 è un valore tale che il 75 % dei dati ordinati è minore o uguale a Q_3 ed è detto anche **75-esimo percentile** e indicato con $P_{0.75}$.

Regola per il calcolo dei quartili

- 1 – Si ordinano gli n dati assegnati in ordine crescente;
- 2 – si calcola il prodotto $k = np$, dove $p = 0.25$ per il primo quartile e $p = 0.75$ per il terzo quartile;
- 3 – se k è un intero, il quartile si ottiene facendo la media del k -esimo e del $(k+1)$ -esimo valore dei dati ordinati;
- 4 – se k non è intero, si arrotonda k per eccesso al primo intero successivo e si sceglie come quartile il corrispondente valore dei dati ordinati.

Esempio su quantili e percentili (fare da soli il 27)

Esempio 26

Calcolare il primo e il terzo quartile dell'insieme di dati

32.2 32.0 30.4 31.0 31.2 31.3 30.3 29.6 30.5 30.7

Dati ordinati

29.6 30.3 30.4 30.5 30.7 31.0 31.2 31.3 32.0 32.2

Primo quartile

$$n = 10 \quad p = 0.25 \quad k = np = 2.5$$

k non è intero, perciò si arrotonda per eccesso $k = 3$: il primo quartile è il terzo dei dati ordinati

$$Q_1 = 30.4.$$

Terzo quartile

$$n = 10 \quad p = 0.75 \quad k = np = 7.5$$

k non è intero, perciò si arrotonda per eccesso $k = 8$: il terzo quartile è l'ottavo dei dati ordinati

$$Q_3 = 31.3.$$

Secondo quartile (mediana)

$$n = 10 \quad p = 0.5 \quad k = np = 5$$

k è intero, perciò si fa la media tra il quinto e il sesto dato e si ottiene

$$Q_2 = \frac{30.7 + 31.0}{2} = 30.85$$

(Questo valore coincide con quello che si trova con la regola della mediana).

Varianza e deviazione standard (s.q.m.)

Definizione 5

Si definisce **varianza**, o anche **varianza campionaria**, la quantità

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (1.2)$$

dove $\bar{x}$ indica la media dei dati.

Definizione 6

Si definisce **scarto quadratico medio** o **deviazione standard** la radice quadrata della varianza

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (1.3)$$

$$s^2 = \frac{1}{n-1} \left[\sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2 \right] \quad (1.4)$$

$$s^2 = \frac{1}{n-1} \left[\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right] \quad (1.5)$$

Esempio su varianza e sqm (fare da soli il 29, 30, 31)

Esempio 28

I seguenti dati sono i tempi di esecuzione di una certa operazione misurati in minuti

0.6 1.2 0.9 1.0 0.6 0.8

Calcoliamo la varianza e la deviazione standard.

$$\bar{x} = \frac{0.6 + 1.2 + 0.9 + 1.0 + 0.6 + 0.8}{6} = 0.85 \text{ minuti}$$

Per la varianza, usando la formula (1.2) si dispongono i calcoli nella tabella seguente

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
0.6	-0.25	0.0625
1.2	0.35	0.1225
0.9	0.05	0.0025
1.0	0.15	0.0225
0.6	-0.25	0.0625
0.8	-0.05	0.0025
totale		0.2750

$$s^2 = \frac{0.2750}{5} = 0.055 \text{ minuti}^2$$
$$s = \sqrt{0.055} \cong 0.23 \text{ minuti}$$

Usare la calcolatrice!!

Per confrontare la variazione di molti campioni diversi di dati, ciascuno con media diversa, o misurati in unità di misura diverse, può essere utile usare una misura di variazione relativa, anziché una misura assoluta come lo scarto quadratico medio.

Definizione 7

Il **coefficiente di variazione** CV è definito da

$$CV = \frac{s}{\bar{x}} \cdot 100\% \quad (1.6)$$

Il coefficiente di variazione esprime lo scarto quadratico medio come percentuale della media ed è indipendente dall'unità di misura usata, poiché la media e lo scarto quadratico medio sono espressi nella stessa unità di misura.

Esempio 32

Sia dato un campione di 200 pacchi di cui sono noti il peso e il volume. Calcolando la media e lo scarto quadratico medio delle due misure si ottengono i seguenti valori

Peso medio:	$\bar{x}_P = 9Kg$
Scarto quadratico medio del peso:	$s_P = 1.5Kg$
Volume medio:	$\bar{x}_V = 2.7m^3$
Scarto quadratico medio del volume:	$s_V = 0.6m^3$

Confrontiamo la variabilità del peso e del volume.

**Fare da soli
esempio 33 !!**

$$CV = \frac{1.5}{9} \cdot 100\% = 16.67\% .$$

$$CV = \frac{0.6}{2.7} \cdot 100\% = 22.22\% .$$

Calcolo di media e varianza per dati raggruppati

Definizioni 8

La media per dati raggruppati è definita da

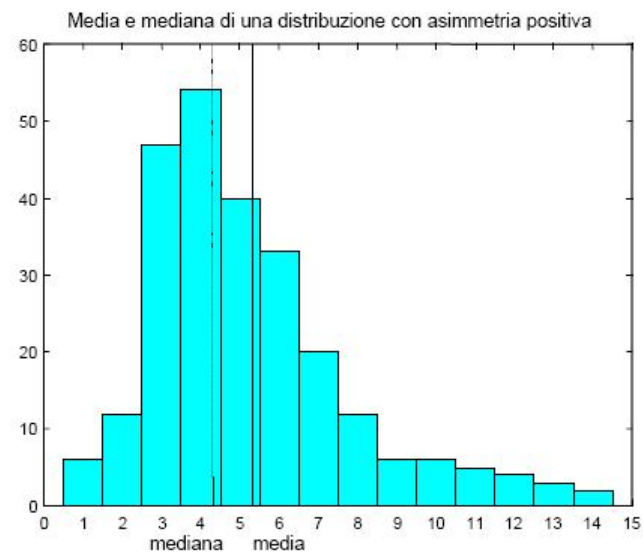
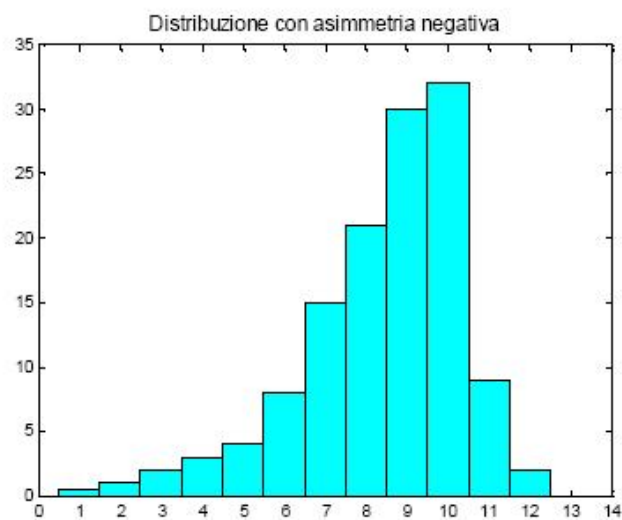
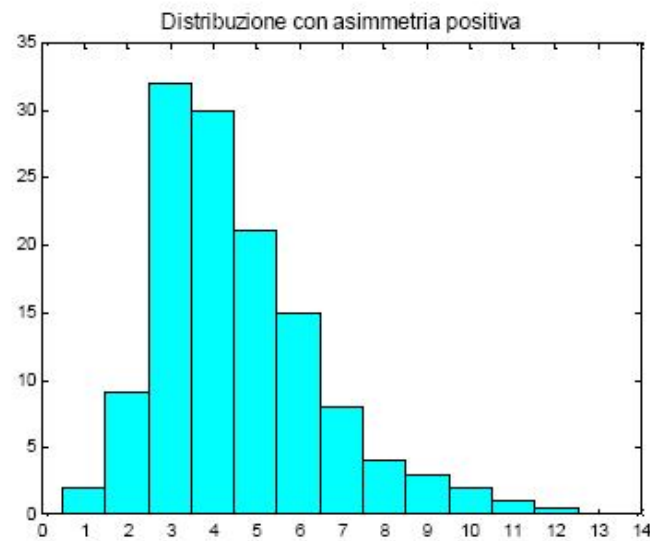
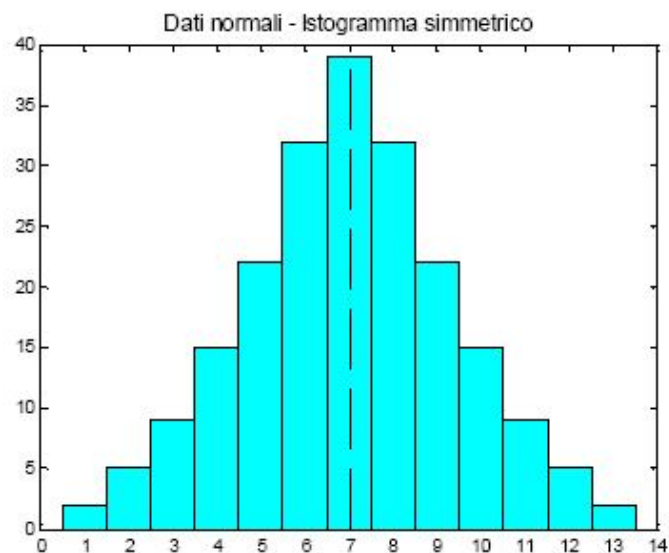
$$\bar{x} = \frac{1}{n} \sum_{i=1}^k m_i f_i \quad (1.7)$$

La varianza per dati raggruppati è definita da

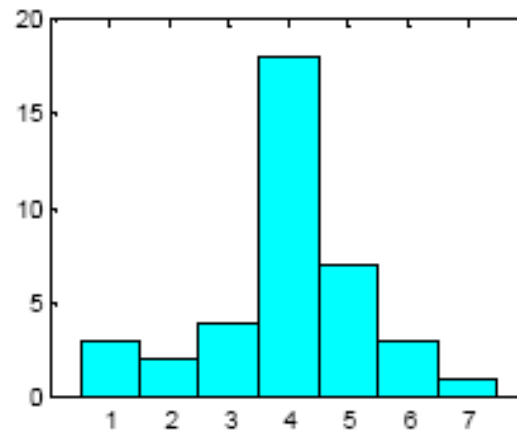
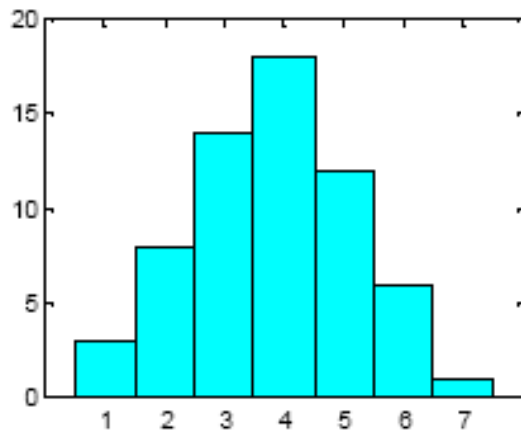
$$s^2 = \frac{1}{n-1} \sum_{i=1}^k (m_i - \bar{x})^2 f_i \quad (1.8)$$

Fare da soli esempi da 34 a 38 anche utilizzando minitab!!

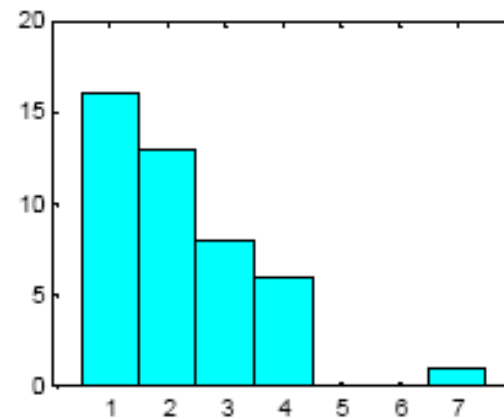
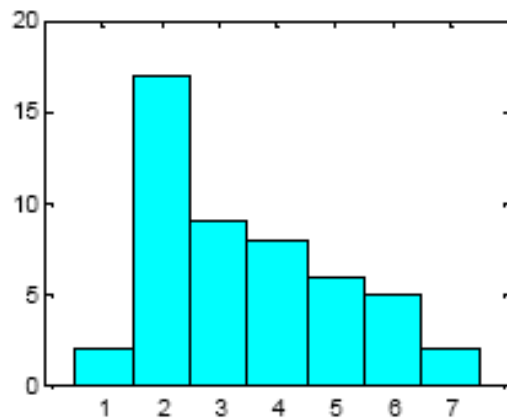
Forma di una distribuzione



Caso di varianza differente



Caso decrescente e presenza di *outlier*



Coefficiente SK (fare da soli gli esempi 39, 40)

Definizione 9

Siano $\bar{x}$, M e s rispettivamente la media, la mediana e lo scarto quadratico medio di un insieme di dati; il **coefficiente di asimmetria di Pearson** è definito da

$$SK = \frac{3(\bar{x} - M)}{s} \quad (1.11)$$

Per una distribuzione perfettamente simmetrica il coefficiente SK vale 0; per una distribuzione asimmetrica positivamente il valore di SK è positivo, mentre è negativo per una distribuzione asimmetrica negativamente.

In generale i valori di SK cadono fra -3 e 3 . La divisione per lo scarto quadratico medio rende il valore di SK indipendente dall'unità di misura dei dati.

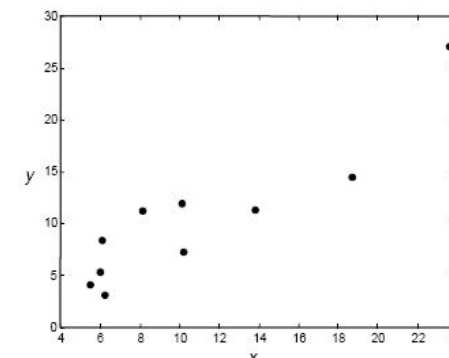
Regola empirica

Se un insieme di dati è approssimativamente normale, con media $\bar{x}$ e scarto quadratico medio s , allora:

- 1 – circa il 68% dei dati è compreso fra $\bar{x} - s$ e $\bar{x} + s$;
- 2 – circa il 95% dei dati è compreso fra $\bar{x} - 2s$ e $\bar{x} + 2s$;
- 3 – circa il 99.7% dei dati è compreso fra $\bar{x} - 3s$ e $\bar{x} + 3s$;

Correlazione fra variabili

diagramma di dispersione o scatterplot.



Definizione 10

Date n osservazioni congiunte di due variabili x e y

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$$

si dice **covarianza** delle due variabili x, y il numero

$$S_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \quad (1.12)$$

Definizione 11

Si dice **coefficiente di correlazione** delle due variabili x, y il numero

$$r = \frac{S_{xy}}{\sqrt{s_x^2 \cdot s_y^2}} \quad (1.13)$$

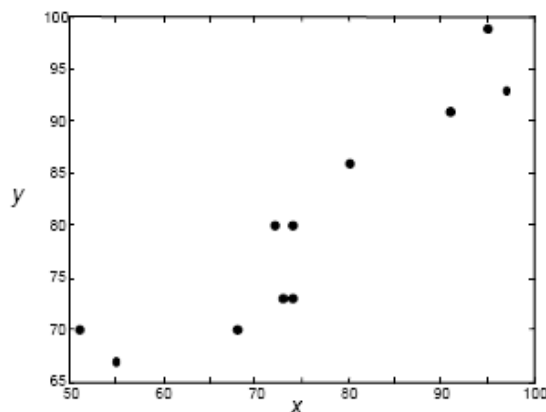
dove s_x^2 e s_y^2 sono le varianze delle variabili x e y .

Correlazione fra variabili

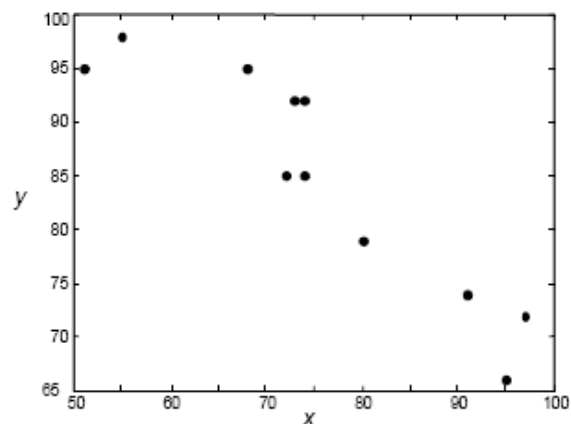
Definizione 12

Si dice che fra le variabili x, y c'è una **correlazione diretta o positiva** se $S_{xy} > 0$; si dice che c'è una **correlazione inversa o negativa** se $S_{xy} < 0$; si dice infine che le variabili sono **non correlate** se $S_{xy} = 0$

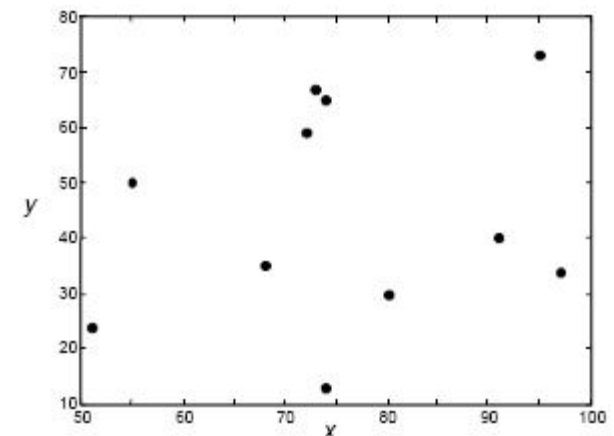
dati correlati positivamente
(covarianza positiva)



dati correlati negativamente
(covarianza negativa)



dati non correlati



Esempio 41

I seguenti dati sono i punteggi che 10 studenti hanno conseguito in due esami di Analisi Matematica (punteggio massimo = 100). Calcolare la covarianza e il coefficiente di correlazione.

Analisi I	Analisi II
51	74
68	70
97	93
55	67
95	99
74	73
20	33
91	91
74	80
80	86

Tabella 51

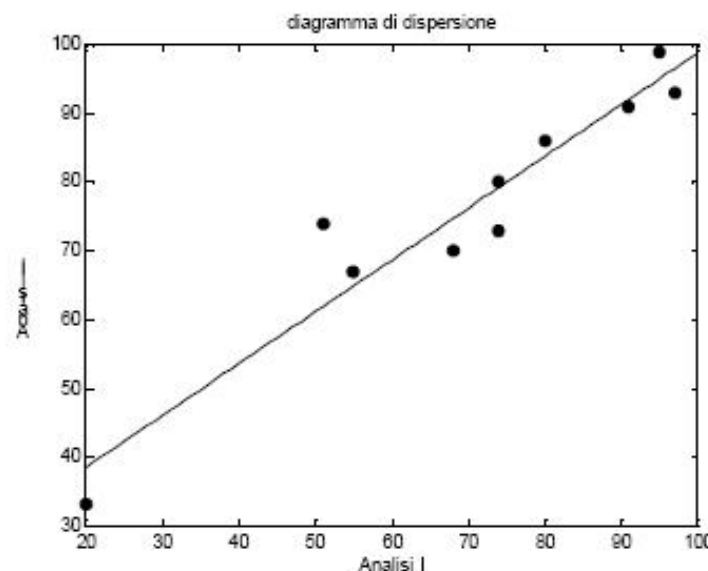


Figura 37

Si ottengono i seguenti risultati

$$\bar{x} = 70.5$$

$$\bar{y} = 76.6$$

$$S_{xy} = 420.55$$

$$s_x^2 = 557.17$$

$$s_y^2 = 348.27$$

$$r = 0.955$$

I dati sono positivamente correlati; il diagramma di dispersione e il valore di r prossimo al valore 1 indicano una relazione lineare fra i dati.