



Università degli Studi di Firenze

Statistica per l'Affidabilità e la Qualità

- Parte I – Statistica descrittiva -

Marcantonio Catelani – Facoltà di Ingegneria Firenze
Dipartimento di Elettronica e Telecomunicazioni
Via S. Marta, 3 – 50139 Firenze
marcantonio.catelani@unifi.it

Il concetto di Statistica

La **STATISTICA** è una disciplina che:

- “insegna” a trattare i dati e a prendere decisioni in condizioni di incertezza;
- consente di definire le proprietà di un Universo, cioè di una collezione di elementi.

☯ **STATISTICA DESCRITTIVA:** da una serie di osservazioni sperimentali si costruiscono tabelle, grafici e diagrammi; si sintetizzano tali informazioni mediante parametri (parametri statistici).

La Statistica descrittiva si basa su un ragionamento di tipo deduttivo, dall'Universo al particolare, caratterizzato dalla **certezza delle sue conclusioni**.

☯ **STATISTICA INDUTTIVA:** si basa su un ragionamento di tipo induttivo, dal Particolare all'Universo, caratterizzato da **conclusioni probabili**.

Comporta lo studio di:

☯ Problemi di Stima

☯ Problemi di Verifica delle ipotesi

Termini e definizioni

Popolazione (Universo): collettivo da studiare

Unità statistica (Elemento, Entità): elemento di una popolazione:

- oggetto reale (vite, prodotto intermedio, prodotto finale, ...)
- risultato di un'osservazione ad esso relativa (misura del diametro di un pezzo, ...)

Caratteristica: proprietà che serve per differenziare o qualificare gli elementi di una data popolazione; può essere:

- *Qualitativa*: se esprimibile mediante aggettivo/attributo (leggero, non conforme, ...)
- *Quantitativa*: se esprimibile mediante una misura riferita ad una scala di valori (osservazioni); la caratteristica quantitativa è detta anche **variabile statistica** e può essere di tipo continuo o discreto.

Modalità del carattere: valore associato al carattere di un elemento

(Es. per una caratteristica quantitativa continua: diametro = 3,5 mm)

Frequenza (assoluta), f : numero di elementi che in un dato insieme hanno una certa caratteristica

(Es. $n=150$ pezzi difettosi esaminati, $f=35$ sono difettosi per deformazione)

Frequenza (assoluta) cumulativa, F : somma delle frequenze (assolute) comprese tra l'estremo inferiore della distribuzione (di frequenza) ed un determinato valore

$$F = \sum f = n$$

Frequenza relativa, f' : rapporto tra la frequenza assoluta ed il numero n di elementi dell'insieme

$$f' = f / n \quad (\text{Es. } f'_{\text{deformazione}} = 35/150)$$

Frequenza relativa cumulativa, F' : somma delle frequenze relative comprese tra l'estremo inferiore della distribuzione ed un determinato valore

$$F' = \sum f' = 1$$

Distribuzione di frequenza : espressione della relazione esistente tra le osservazioni e la frequenza con cui si presentano.

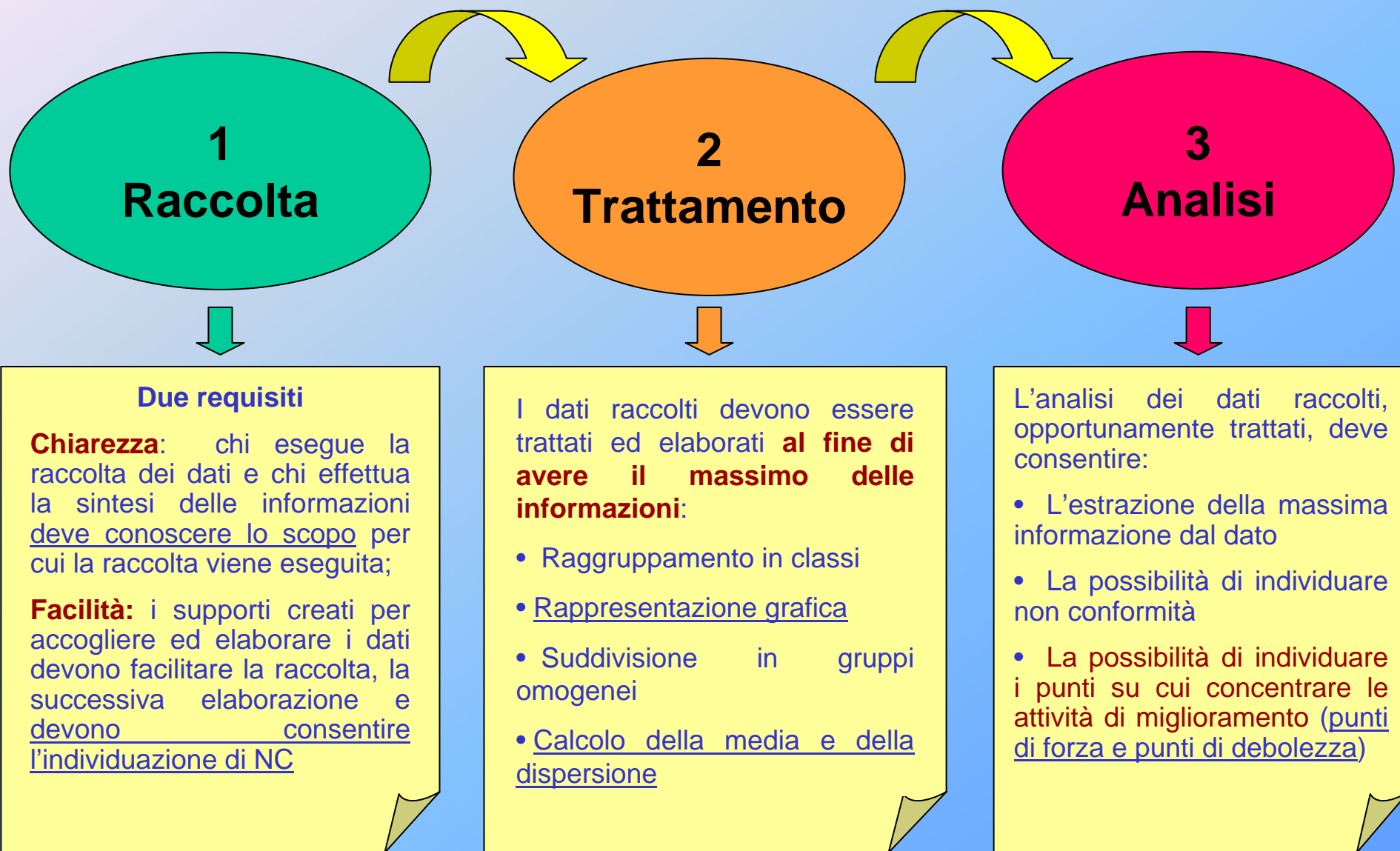
La distribuzione può essere presentata in forma tabellare o grafica avvalendosi di una suddivisione in classi.

Distribuzione cumulativa di frequenza : distribuzione di frequenza in cui ad ogni valore della caratteristica (o ad ogni classe) si fa corrispondere la somma delle frequenze comprese tra l'origine ed il limite superiore della classe considerata.

Classe: per una caratteristica quantitativa, ciascuno degli intervalli consecutivi e mutuamente esclusivi nei quali è diviso l'intervallo totale di variazione.

Istogramma: rappresentazione grafica della distribuzione di frequenza di una caratteristica quantitativa, consistente in un insieme di rettangoli contigui, ciascuno con una base uguale all'ampiezza della classe ed un'area proporzionale alla frequenza della classe.

Misurare e valutare la conformità ai requisiti ed il livello di soddisfazione: l'importanza delle “informazioni (dati)”



Le informazioni (dati): quantitative e qualitative

La **caratteristica** di un servizio (efficienza,...) o di un prodotto (peso,...) è la proprietà che serve per differenziare quel servizio o quel prodotto

La caratteristica deve essere **misurabile**

L'informazione (dato) che si ottiene può essere:

- **quantitativa**: quando può assumere infiniti valori e si esprime con un'unità di misura: *peso, temperatura, lunghezza, ecc.*
- **qualitativa**: quando si esprime mediante aggettivo/attributo: *leggero, pesante, tipo di colore, ecc.*

Sintesi delle informazioni : concetti di media e dispersione

Le informazioni raccolte devono essere riassunte sinteticamente attraverso i parametri statistici

Parametri di posizione: determinano la tendenza centrale della distribuzione dei dati (*media, moda, mediana, ecc.*)

Parametri di dispersione: descrivono il grado di variabilità della distribuzione dei dati (*escursione, s.q.m., ecc.*)

Organizzazione, presentazione e interpretazione dei dati

A) Organizzazione e presentazione dei dati

Dato un insieme di osservazioni (misure), raccolte nell'ordine in cui sono state ottenute (in genere in ordine sparso), occorre organizzarle in maniera opportuna (raggrupparle, ordinarle, ecc.) sotto forma di tabelle e/o grafici.

Foglio raccolta dati

Per raccogliere informazioni quantitative e qualitative, anche successive alla somministrazione di un questionario

Esempio n.2 – Raccolta e presentazione di informazioni ottenute lanciando un dado 50 volte.

Variabile : faccia del dado che può assumere i valori 1,2,...6

Risultato dei 50 lanci: 6, 2, 3, 4, 2, 5, 4, 5,, 5 come riportato in tabella

Numero dei punti	Risultati	Frequenza (f)	Frequenze relative (f ')	---
1		9	0,18	---
2		12	0,24	---
3		6	0,12	---
4		8	0,16	---
5		10	0,20	---
6		5	0,10	---

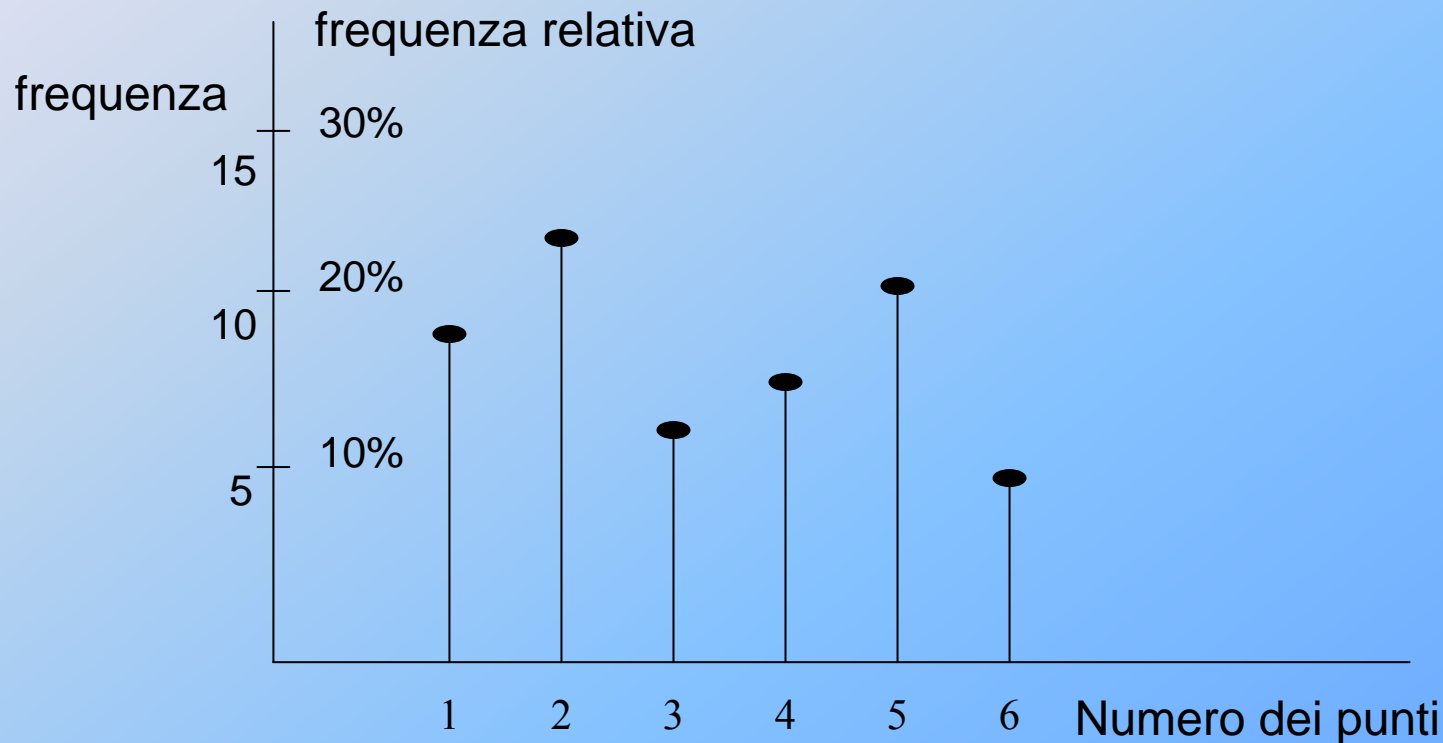
Frequenza (assoluta), f : numero di elementi che, in un dato insieme, hanno la stessa caratteristica (Es. 9 volte si verifica il valore 1)

Frequenza relativa, f ': rapporto tra la frequenza assoluta f ed il numero n di elementi dell'insieme (Es. 9 volte su 50 = 0,18=18%)

Distribuzione di frequenza

Per rappresentare graficamente una serie di informazioni

La caratteristica in esame (*esempio 1*) è misurabile rispetto ad una scala numerica comprendente solo numeri interi



Esempio – Variabile casuale continua - Analisi e presentazione dei risultati relativi alla misura dei tempi al guasto (in ore) osservati su un campione di 125 lampadine.

X tempo al guasto (variabile casuale continua)

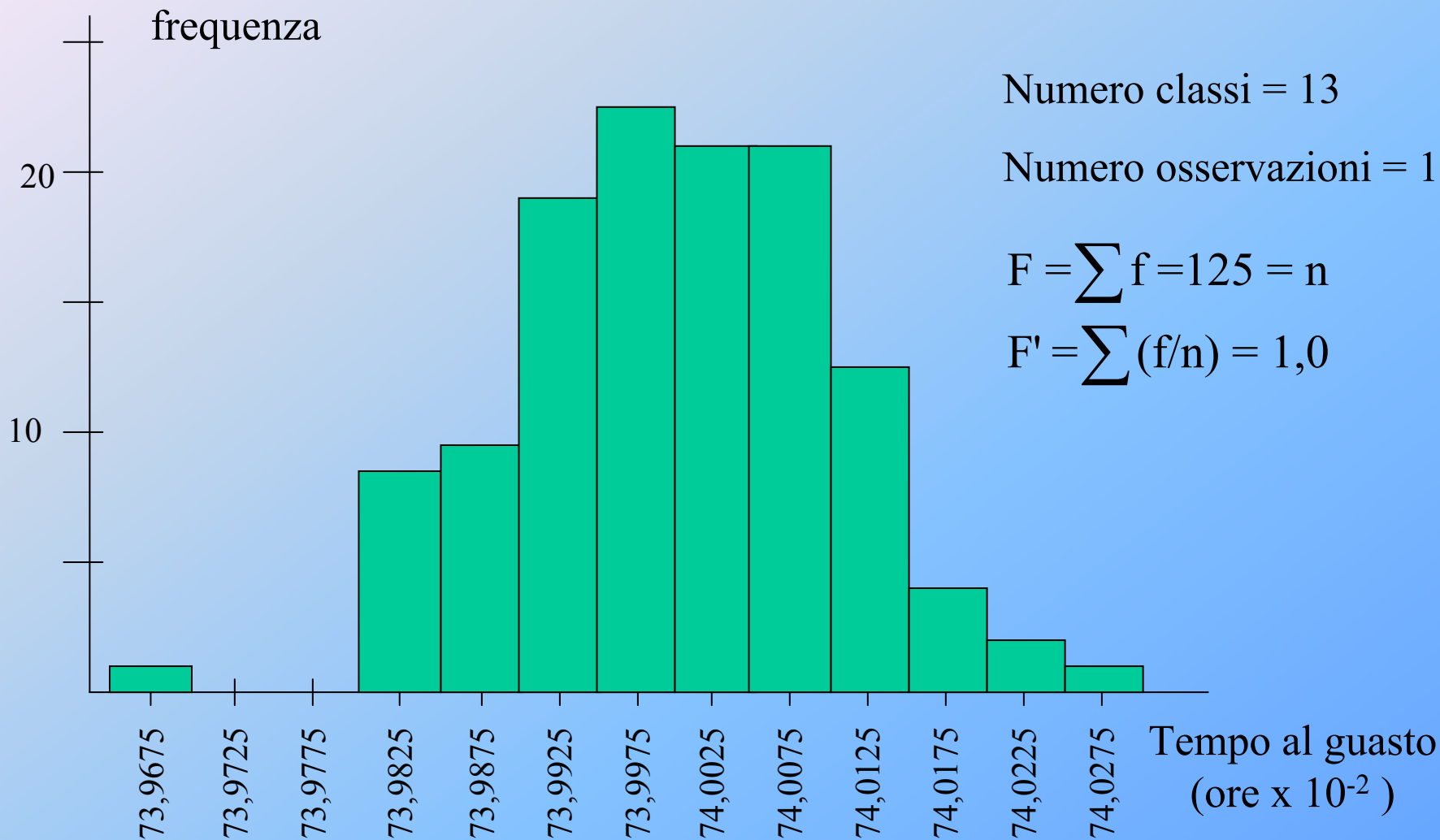
Risultato di 125 osservazioni : valori compresi nell'intervallo [7396,5 7403,0]

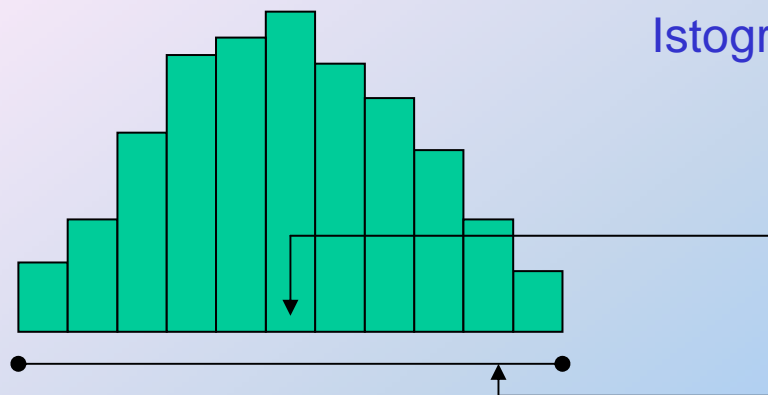


Distribuzione di frequenza del “tempo al guasto”

Classe	Limiti (ore)	Valore centrale (ore)	Risultati	f	F	f ’	F ‘
1	$7396,5 \leq x < 7397,0$	7396,75		1	1	0,008	0,008
2	$7397,0 \leq x < 7397,5$	7397,25		0	1	0,000	0,008
3	$7397,5 \leq x < 7398,0$	7397,75		0	1	0,000	0,008
4	$7398,0 \leq x < 7398,5$	7398,25		8	9	0,064	0,072
5	$7398,5 \leq x < 7399,0$	7398,75		10	19	0,080	0,152
6	$7399,0 \leq x < 7399,5$	7399,25		19	38	0,152	0,304
7	$7399,5 \leq x < 7400,0$	7399,75		23	61	0,184	0,488
8	$7400,0 \leq x < 7400,5$	7400,25		22	83	0,176	0,664
9	$7400,5 \leq x < 7401,0$	7400,75		22	105	0,176	0,840
10	$7401,0 \leq x < 7401,5$	7401,25		13	118	0,104	0,944
11	$7401,5 \leq x < 7402,0$	7401,75		4	122	0,032	0,976
12	$7402,0 \leq x < 7402,5$	7402,25		2	124	0,016	0,992
13	$7402,5 \leq x \leq 7403,0$	7402,75		1	125	0,008	1,000

Istogramma - Rappresentazione grafica della distribuzione di frequenza dei tempi al guasto

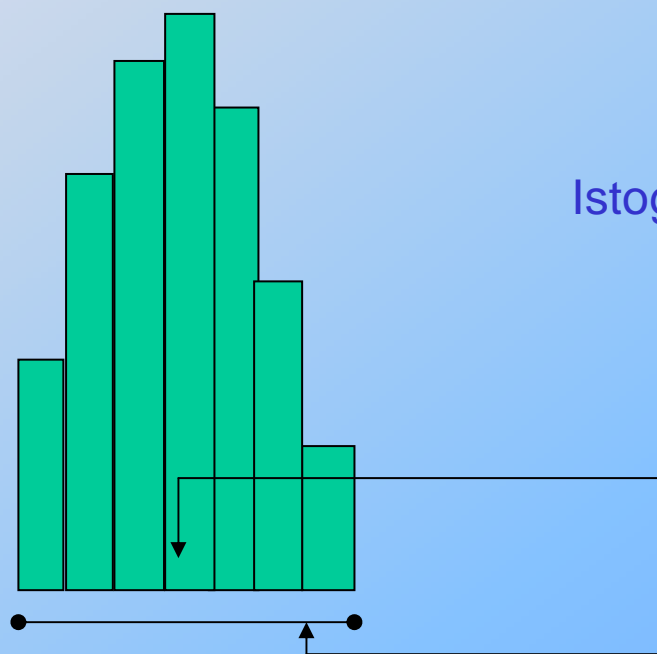




Istogramma simmetrico rispetto al valore medio,
con elevata dispersione (varianza)

Valore medio

Escursione

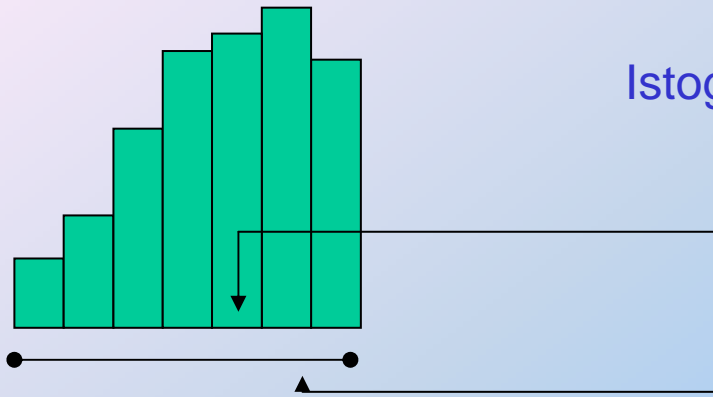


Istogramma simmetrico rispetto al valore medio,
con bassa dispersione (varianza)

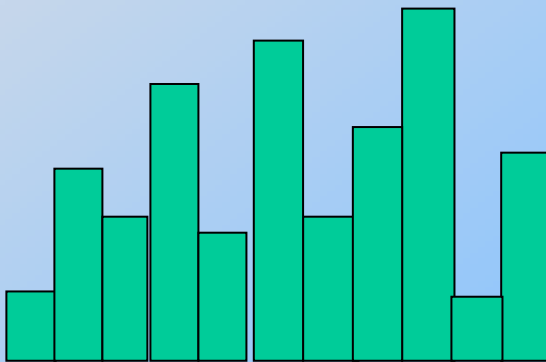
Valore medio

Escursione

Istogramma asimmetrico rispetto al valore medio



Istogramma asimmetrico “a pettine”



Regola empirica per determinare il numero delle classi

$$K \text{ (numero delle classi)} = 3,3 \log (n) + 1$$

n : numero delle osservazioni



n	k
< 50	5 – 7
50 - 100	6 – 10
100 – 250	7 – 12
> 250	10 - 20

B) Sintesi delle osservazioni – Parametri statistici

I parametri statistici consentono di riassumere in maniera sintetica le distribuzioni di frequenza:

Parametri di posizione: determinano la tendenza centrale della distribuzione (moda, media, mediana, ecc.)

Parametri di dispersione: descrivono il grado di variabilità della distribuzione (s.q.m., escursione, ecc.)

Parametri di forma: descrivono la forma della distribuzione (simmetria, curtosi, ecc.)

B.1) – Parametri di posizione

Media aritmetica (arithmetic mean) \bar{x}, μ :

$$\bar{x} = \frac{\sum_{i=1}^N x_i}{N} ; \quad \text{oppure} \quad \bar{x} = \frac{\sum_{j=1}^k f_j x_j}{N} , \quad N = \sum_j f_j \quad (1)$$

Media pesata (arithmetic weighted mean)

$$\bar{x} = \frac{\sum_j w_j x_j}{\sum_j w_j}$$

- La (1) può essere considerata una media pesata con pesi f_1, \dots, f_k .
- La media ha la stessa unità di misura della variabile

Proprietà della media

- La media è un cattivo indice di posizione quando la distribuzione è asimmetrica; pochi valori fuori scala influiscono pesantemente sulla media.
- La somma algebrica delle differenze tra un insieme di numeri e la loro media aritmetica vale zero:

$$\sum_i (x_i - \bar{x}) = 0$$

- La media è un operatore lineare:

$$\{x_i\} \rightarrow \bar{x}$$

$$\{y_i = a x_i + b\} \rightarrow \bar{y} = a \bar{x} + b$$

- Se x ed y sono due grandezze distinte ma indipendenti:

$$\{x_i\} \rightarrow \bar{x} ; \{y_i\} \rightarrow \bar{y}$$

$$\{z_i = x_i \cdot y_i\} \rightarrow \bar{z} = \bar{x} \cdot \bar{y}$$

Mediana, M_e : valore della caratteristica al di sotto del quale cade la metà dei valori campionari. In un insieme di N numeri ordinati in senso non-decrescente vale:

N dispari : $M_e = (x_{(N+1)/2}) \equiv$ Elemento centrale

N pari : $M_e = \frac{1}{2}(x_{(N/2)} + x_{(N/2+1)}) \equiv$ Media elementi centrali

N = 5: 1 , 3 , 7 , 8 , 9

$$M_e = 7$$

N = 9: 3 , 4 , 4 , 5 , 6 , 8 , 8 , 8 , 10

$$M_e = 6$$

N = 8: 1 , 1 , 3 , 5 , 6 , 9 , 10 , 11

$$M_e = (5+6)/2 = 5,5$$

Quartile, Q_1, Q_3 : valori della caratteristica che dividono l'insieme dei dati in quattro parti uguali

Q_1 primo quartile; Q_2 secondo quartile o Mediana; Q_3 terzo quartile

Decili, D_1, \dots, D_9 : valori della caratteristica che dividono l'insieme dei dati in dieci parti uguali

Percentili, P_1, \dots, P_{99} : valori della caratteristica che dividono l'insieme dei dati in cento parti uguali

Quantile, frattile (di una variabile casuale o di una distribuzione di probabilità): il *p-esimo* quantile è il valore della variabile casuale per il quale la funzione cumulativa, F , è uguale a p .

Nel caso di una variabile continua, il quantile di ordine p è il valore della variabile al di sotto del quale si trova la proporzione p della distribuzione.

Es.: Mediana = quantile 0,5; Primo quartile (Q_1) = quantile 0,25; ecc.

Moda, M_o : valore (valori) della caratteristica a cui corrisponde la massima frequenza.

$N=7$:

2 , 1 , 1 , 3 , 2 , 1 , 5

$M_o=1$ distribuzione unimodale

2 , 1 , 3 , 4 , 11 , 12 , 8

non esiste Moda

2 , 1 , 2 , 1 , 3 , 7 , 5

$M_{o1}=1$; $M_{o2}=2$ distribuzione bimodale

B.2) – Parametri di dispersione

Dispersione è l'attitudine delle osservazioni a distribuirsi intorno al corrispondente valore medio.

Escursione o Range, R: differenza tra il valore massimo e minimo di un insieme di osservazioni quantitative

$$R = x_{\max} - x_{\min}$$

- L'escursione è un cattivo indice di dispersione quando la distribuzione delle osservazioni è dispersa
- L'escursione si esprime con la stessa unità di misura della grandezza a cui si riferisce

Varianza, σ^2 : misura della dispersione, definita come somma dei quadrati degli scarti delle osservazioni rispetto alla loro media aritmetica, divisa per il numero delle osservazioni

Per n osservazioni x_1, x_2, \dots, x_n con media $\bar{x} = \frac{1}{n} \sum_i x_i$, la varianza è:

$$\sigma^2 = \frac{1}{n} \sum_i (x_i - \bar{x})^2$$

L'unità di misura della varianza è il quadrato dell'unità di misura della grandezza a cui si riferisce.

Scarto tipo (scarto quadratico medio), σ : radice quadrata della varianza

L'unità di misura dello scarto tipo corrisponde con quella della grandezza.

Coefficiente di variazione, CV : rapporto tra lo scarto tipo e la media

Proprietà della varianza

- $\{x_i\} \rightarrow \bar{x}, \sigma_x^2$
 $\{y_i = a x_i\} \rightarrow \bar{y} = a \bar{x}, \sigma_y^2 = a^2 \sigma_x^2$
- La varianza non cambia sommando una costante alla distribuzione

$$\{x_i\} \rightarrow \bar{x}, \sigma_x^2$$

$$\{y_i = a x_i + b\} \rightarrow \bar{y} = a \bar{x} + b, \sigma_y^2 = a^2 \sigma_x^2$$

Esempio – Significato di alcuni parametri statistici – Misure del tempo di vita ($\times 10^2$ ore) di due campioni costituiti da cinque lampadine

a) 9 , 13 , 11 , 15 , 7 ; b) 10 , 9 , 12 , 11 , 13

	(a)	(b)	note
Media ($\times 10^2$ ore)	11	11	Parametro non significativo per distinguere i campioni
Escursione ($\times 10^2$ ore)	8	4	Valori apparentemente più concentrati nel campione (b)
Varianza ($\times 10^4$ ore ²)	10	2,5	Valori più concentrati rispetto alla media nel campione (b)

Covarianza, σ_{xy} : somma dei prodotti degli scarti di x e y dalle rispettive medie aritmetiche, divisa per il numero n delle coppie osservate

$$\sigma_{xy} = \frac{1}{n} \sum_i (x_i - \bar{x})(y_i - \bar{y})$$

Coefficiente di correlazione, ρ_{xy} : rapporto tra la covarianza di due caratteristiche ed il prodotto dei loro scarti tipo

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 (y_i - \bar{y})^2}}$$

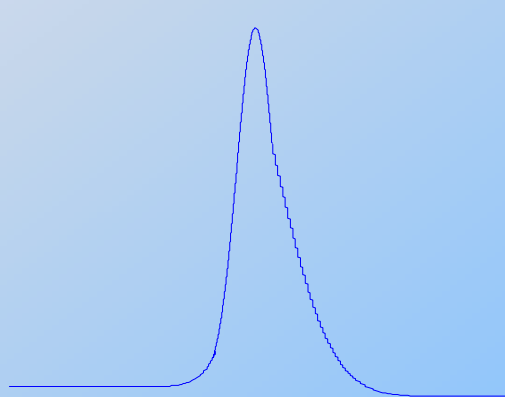
- La definizione data si applica a grandezze misurabili.
- Il valore di ρ_{xy} è sempre compreso tra -1 e $+1$. Quando il coefficiente di correlazione è uguale ad uno di questi limiti si ha una relazione lineare esatta tra X ed Y nell'insieme delle osservazioni accoppiate.
- Se X ed Y sono scorrelate $\rho_{xy} = 0$

B.3) – Parametri di forma

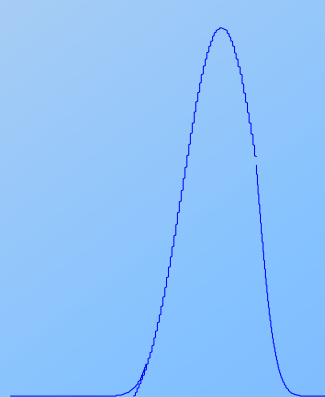
Sono rappresentativi della forma della distribuzione.

Coefficiente di asimmetria o coefficiente di Skewness

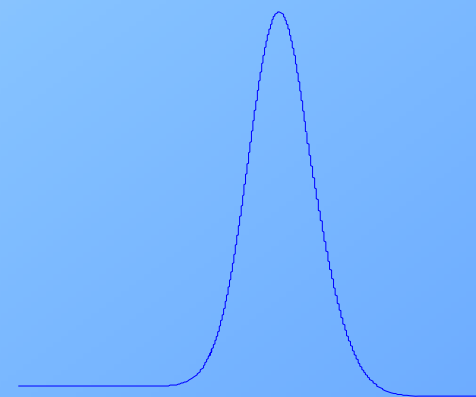
a) valore positivo (distribuzione più spostata a sinistra), b) negativo (distribuzione più spostata a destra), c) nullo (distribuzione simmetrica).



a



b



c

Coefficiente di appiattimento o coefficiente di curtosi

a) valore positivo (distribuzione più allungata rispetto alla normale), b) negativo (distribuzione più appiattita rispetto alla normale), c) nullo (distribuzione normale).

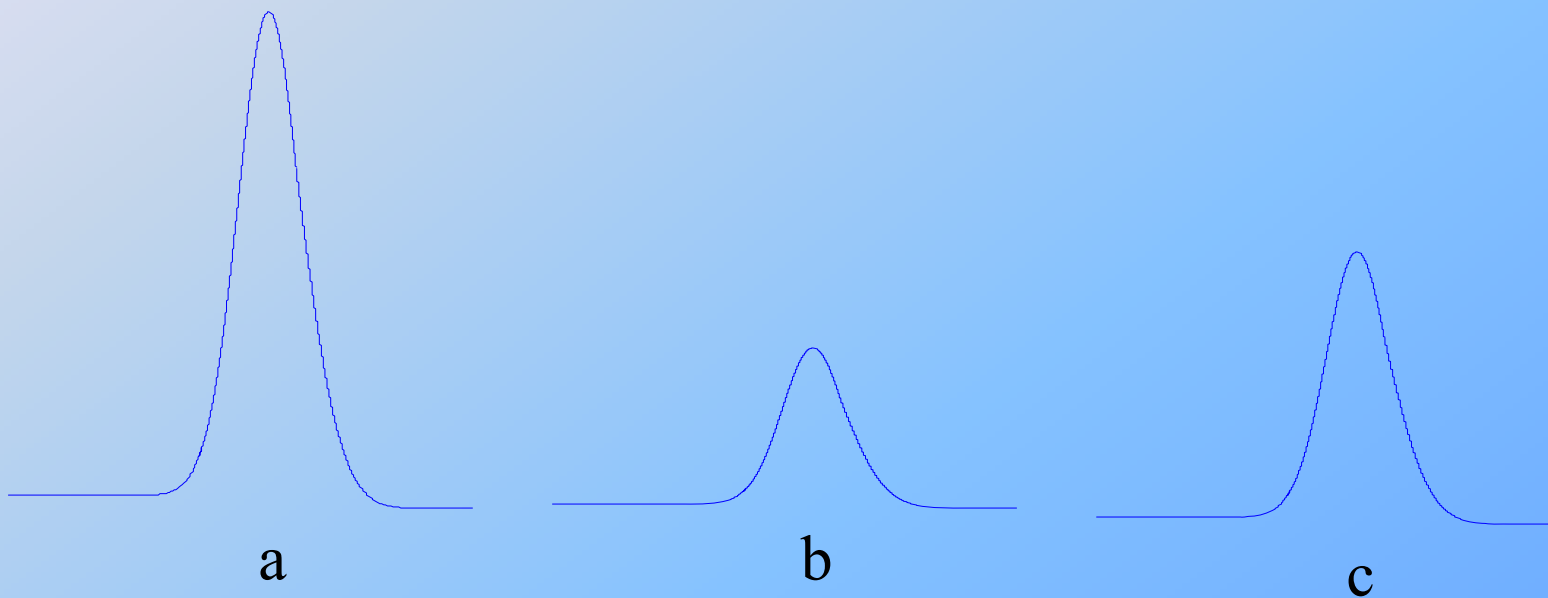


Diagramma di Pareto

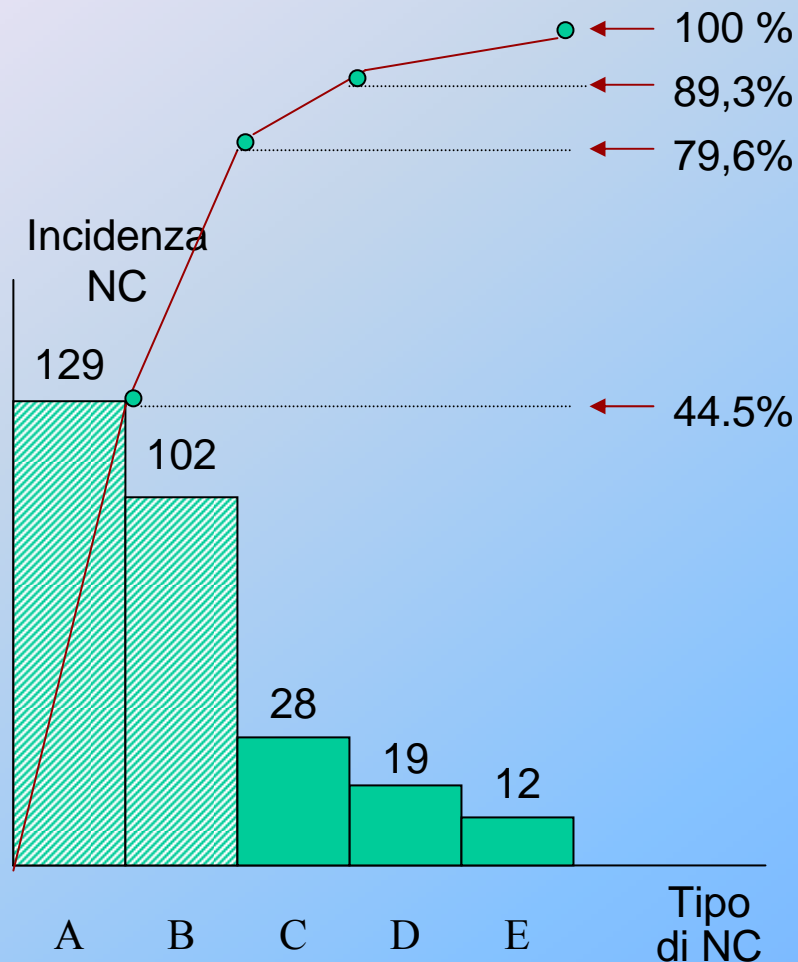
Per visualizzare il contributo delle cause più ricorrenti in un determinato contesto, con particolare riferimento alle non conformità

“Poche cause sono maggiormente responsabili di un determinato effetto”

Da costruire dopo una raccolta dati o somministrazione di un questionario


Consente di individuare le priorità di intervento per le azioni correttive

Esempio n.4 – Elaborazione dei dati relativi alla somministrazione di un questionario per valutare la qualità dell'offerta didattica



NC riscontrate (tot. 290)

- A. Aule e attrezzature non idonee (*)
- B. Cattiva organizzazione dell'orario (*)
- C. Materiale didattico insufficiente
- D. Carico di lavoro inaccettabile
- E. Scarsa chiarezza del docente

 Non conformità su cui occorre intervenire in maniera prioritaria con azioni correttive

(*) Può essere costruito un Pareto delle NC più importanti

Diagramma Causa - Effetto

Rappresentazione grafica per evidenziare le relazioni tra una caratteristica (effetto) e i suoi fattori o cause

Il diagramma, una volta realizzato, costituisce un quadro di insieme ordinato e completo di tutte le cause che concorrono a determinare un certo effetto

La costruzione viene fatta in brainstorming

Per la costruzione e l'applicazione occorre:

1. Identificare chiaramente l'effetto (NC) che deve essere studiato
2. Classificare le cause principali
3. Scomporre le cause principali in cause secondarie, ecc.
4. Individuare le cause su cui concentrare le azioni correttive

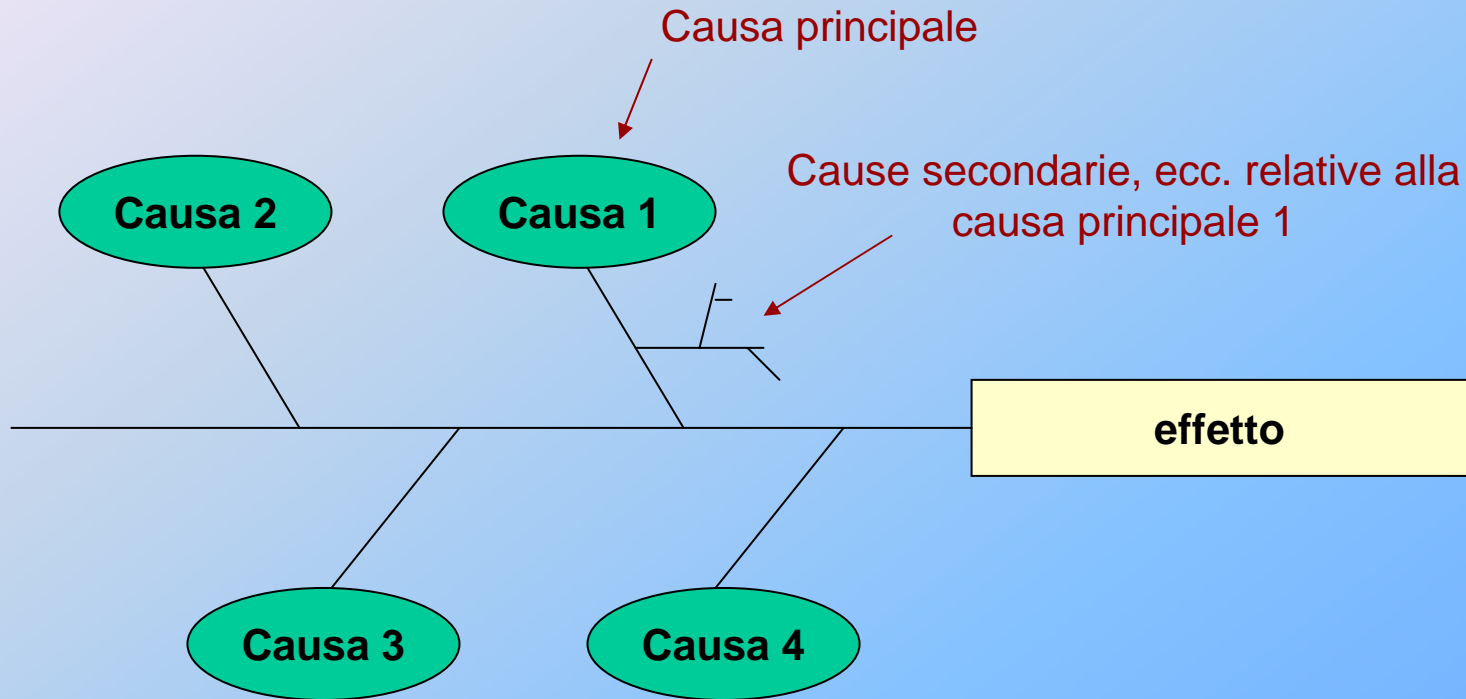
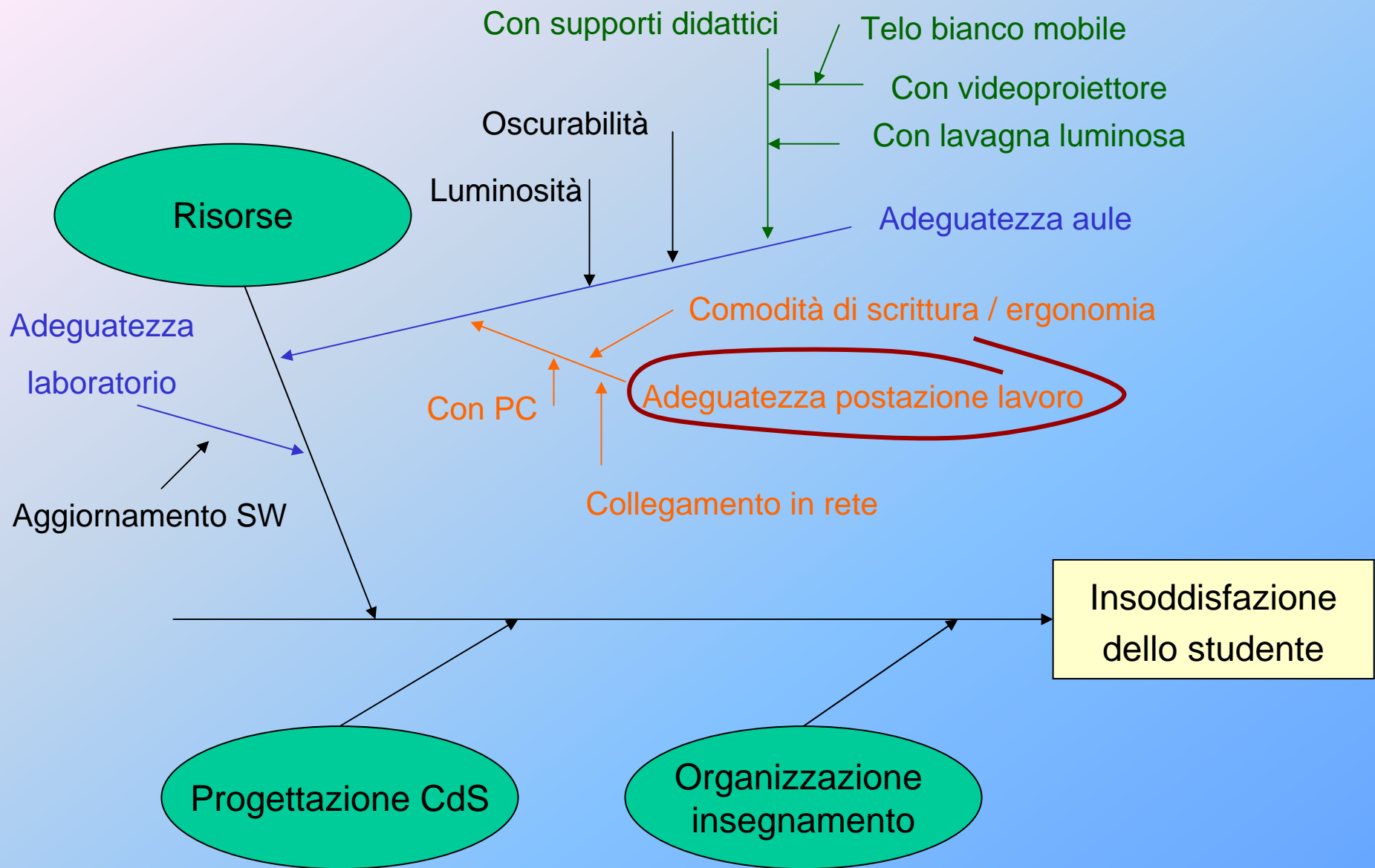


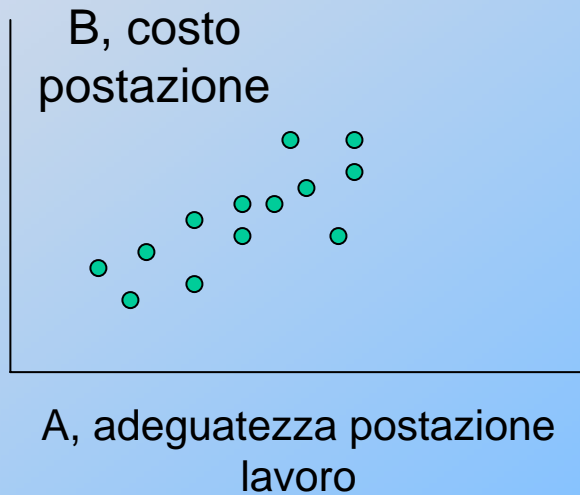
Diagramma Causa - Effetto



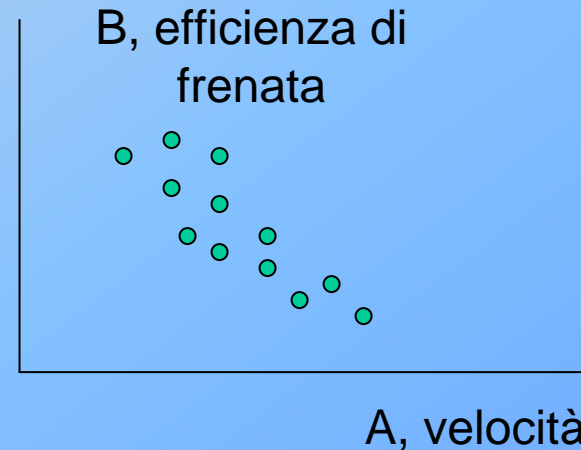
Correlazione

Serve per comprendere se e quanto due grandezze A e B siano legate (correlate) tra loro

Può essere applicata dopo il diagramma causa-effetto per studiare il legame che eventualmente esiste tra cause diverse



A ↑ ; B ↑



A ↑ ; B ↓

