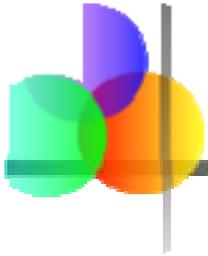
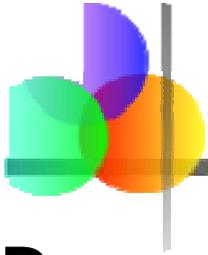


Statistica



Capitolo 2

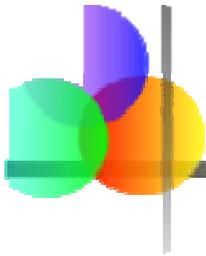
Descrizione grafica dei dati



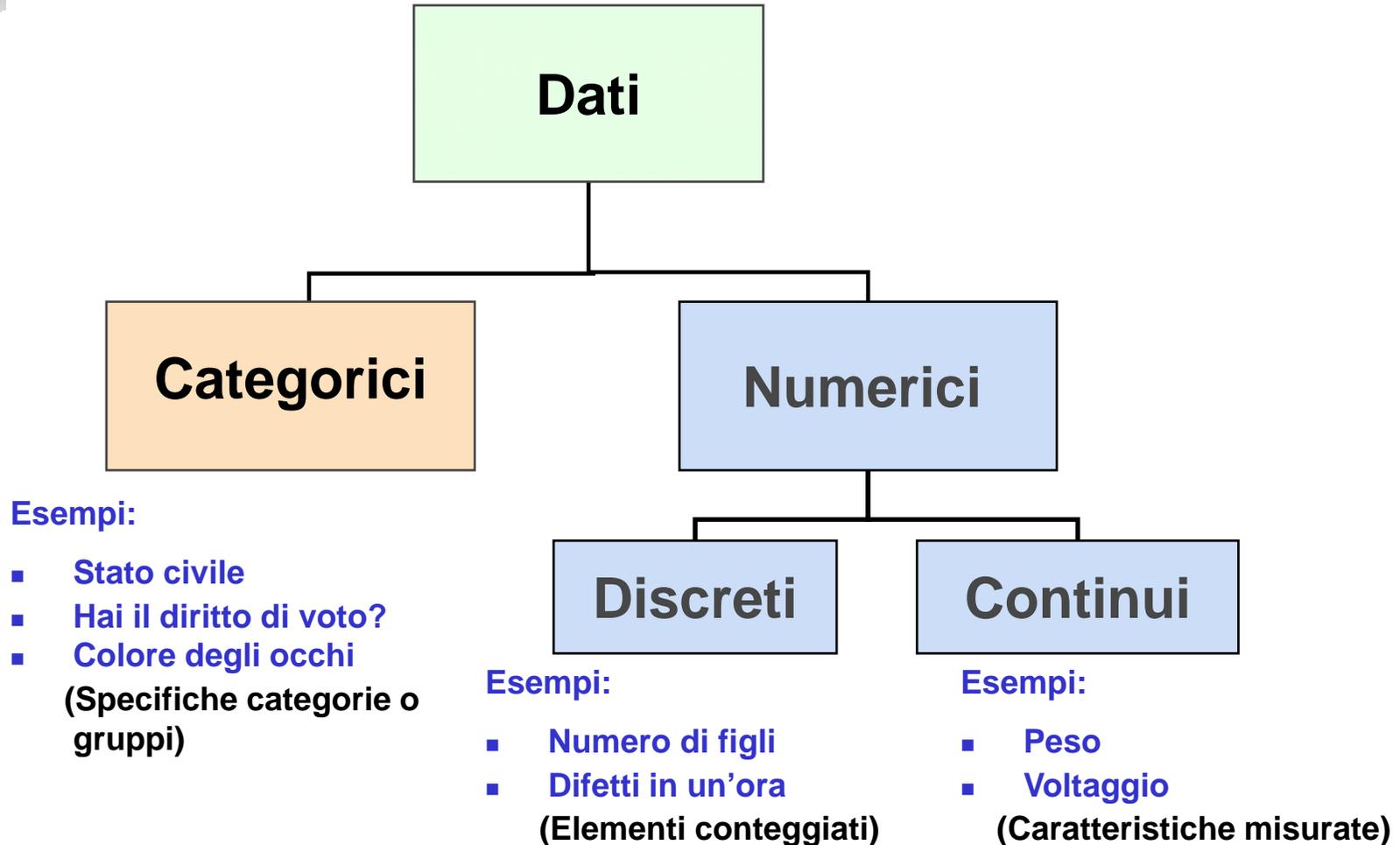
Obiettivi del capitolo

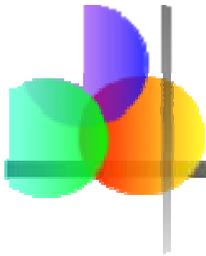
Dopo aver completato il capitolo, sarete in grado di:

- Identificare il tipo di dati e il livello di misurazione
- Creare ed interpretare grafici per descrivere variabili categoriche:
 - distribuzione di frequenze, diagramma a barre, diagramma a torta, diagramma di Pareto
- Creare un grafico per descrivere serie storiche
- Creare ed interpretare grafici per descrivere variabili numeriche:
 - distribuzione di frequenze, istogramma, ogiva, diagramma ramo-foglia
- Costruire ed interpretare grafici per descrivere relazioni tra variabili:
 - diagrammi di dispersione, tabelle a doppia entrata
- Descrivere grafici appropriati e non appropriati per rappresentare i dati

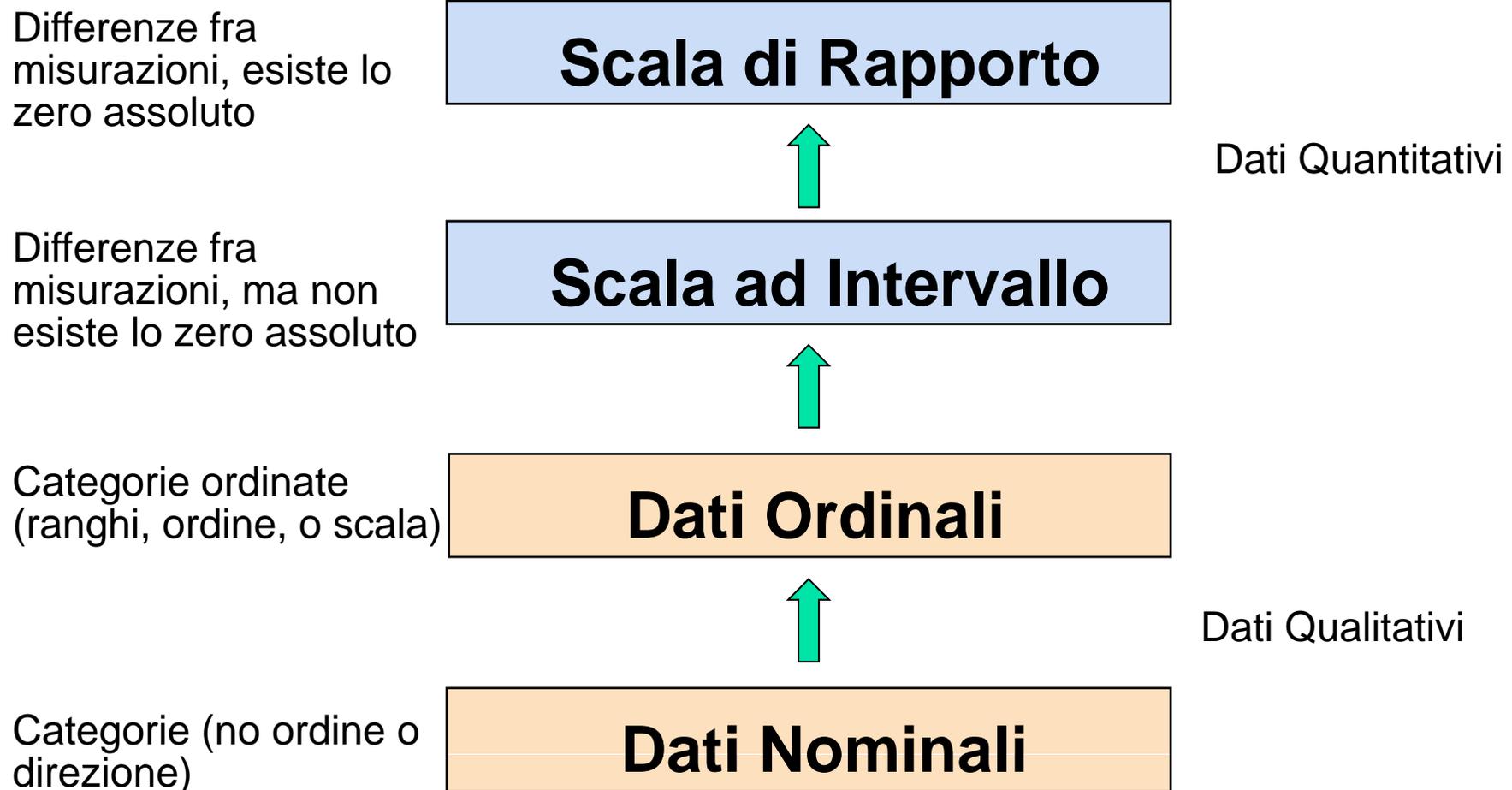


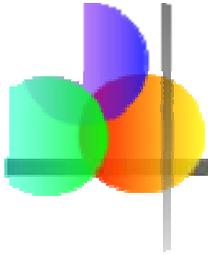
Tipi di Dati





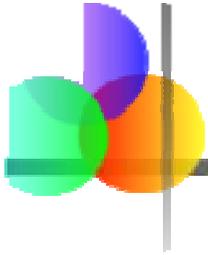
Livelli di Misurazione





Rappresentazione Grafica dei Dati

- I dati in **forma grezza** non sono generalmente facili da usare nel processo decisionale
- Una qualche organizzazione si rende necessaria:
 - Tabelle
 - Grafici
- Il tipo di grafico da usare dipende dalla variabile che vogliamo sintetizzare



Rappresentazione Grafica dei Dati

(continuazione)

- Tecniche presentate in questo capitolo:

Variabili Categorie

- Distribuzione di frequenze
- Diagramma a barre
- Diagramma a torta
- Diagramma di Pareto

Variabili Numeriche

- Grafico per serie storiche
- Distribuzione di frequenze
- Istogramma e ogiva
- Diagramma ramo-foglia
- Diagramma di dispersione

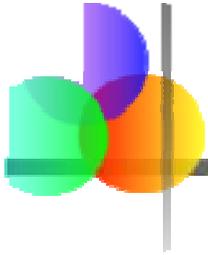
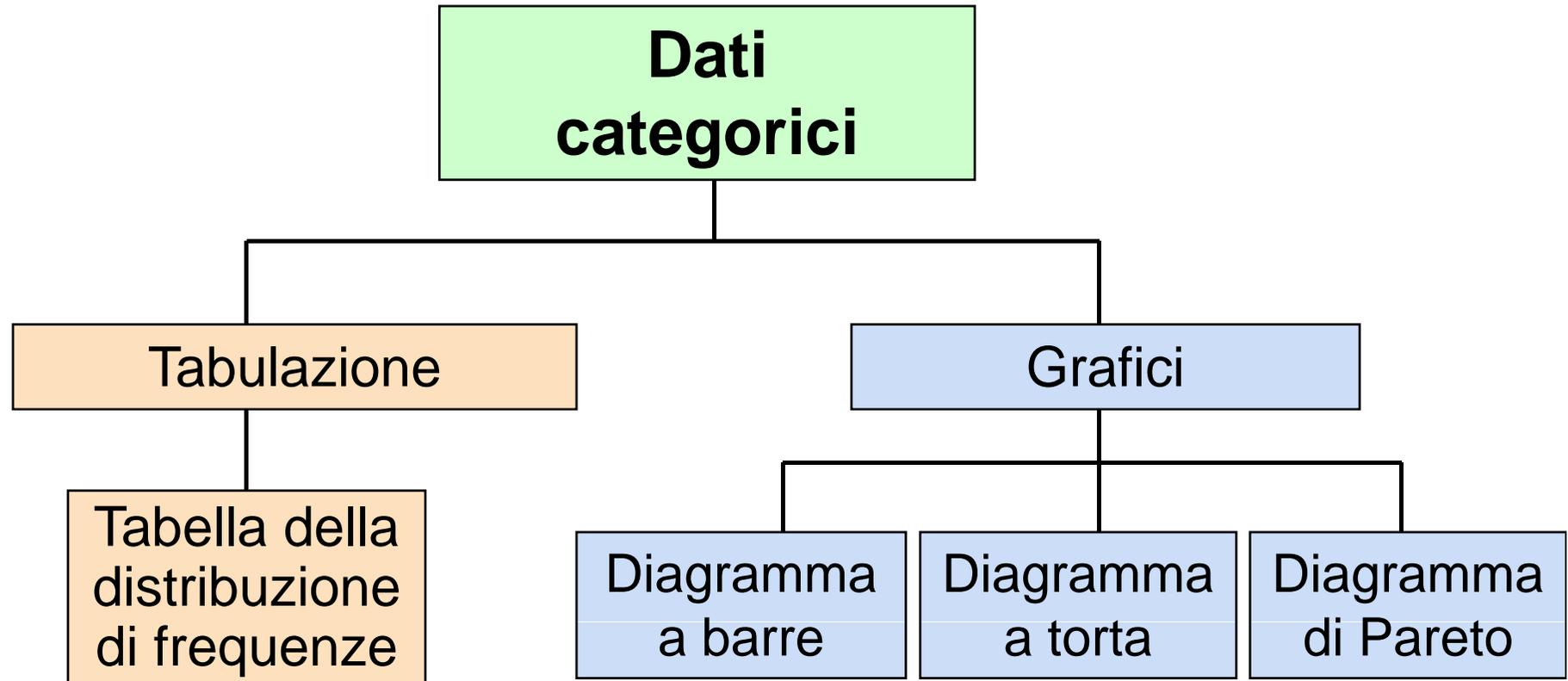
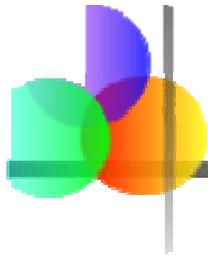


Tabelle e Grafici per Variabili Categorieche





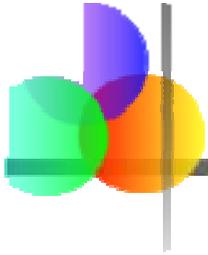
La Tabella della Distribuzione di Frequenze

Riassume i dati per categoria

Esempio: Pazienti per Unità Ospedaliera

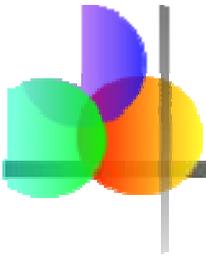
Unità Osped.	Numero di Pazienti
Cardiologia	1,052
Emergenza	2,245
Cura Intensiva	340
Maternità	552
Chirurgia	4,630

(La variabile è
categorica)



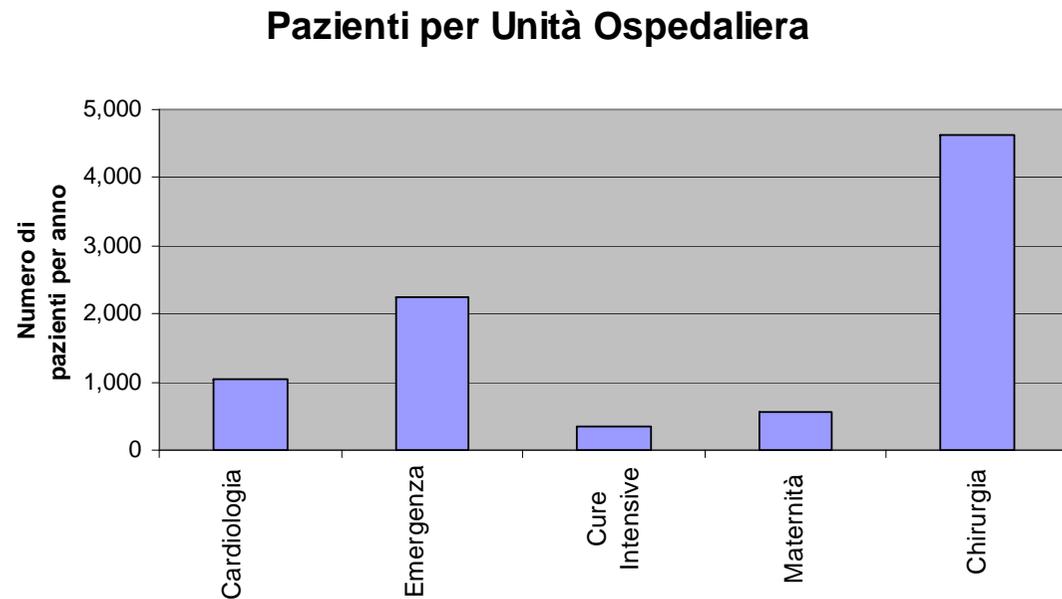
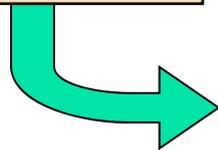
Diagrammi a barre e a torta

- **Diagrammi a barre e Diagrammi a torta** sono spesso usati per dati qualitativi (categorici)
- L'altezza delle barre o l'area dei settori circolari rappresentano la frequenza o percentuale di ciascuna categoria



Esempio diagramma a barre

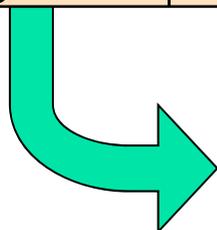
Unità Osped.	Numero di Pazienti
Cardiologia	1,052
Emergenza	2,245
Cure Intensive	340
Maternità	552
Chirurgia	4,630





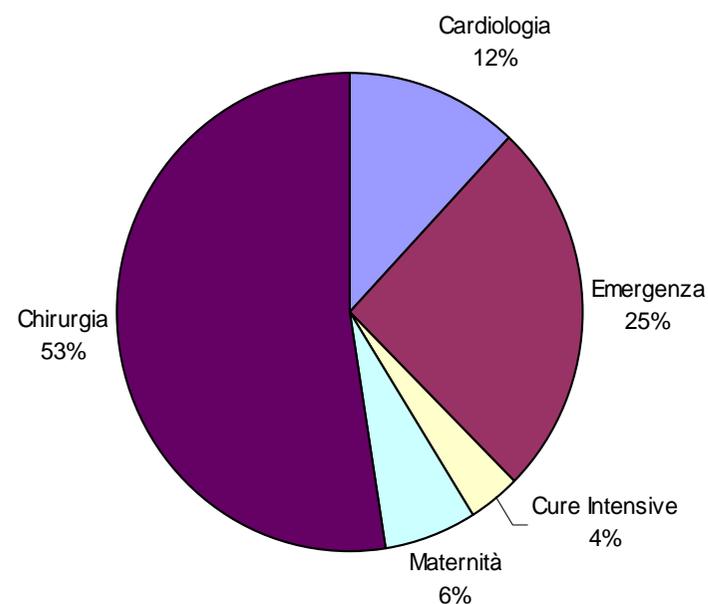
Esempio Diagramma a Torta

Unità Osped.	Numero di Pazienti	% del Totale
Cardiologia	1,052	11.93
Emergenza	2,245	25.46
Cure Intensive	340	3.86
Maternità	552	6.26
Chirurgia	4,630	52.50



(Le percentuali sono arrotondate)

Pazienti per Unità Ospedaliera



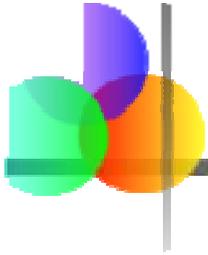
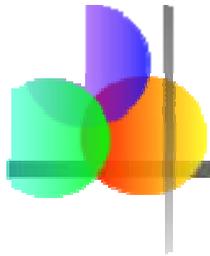


Diagramma di Pareto

E' usato per rappresentare dati categorici. E' formato da:

- Un diagramma a barre, in cui le categorie sono rappresentate in ordine decrescente di frequenza
- Un poligono della frequenza cumulata che viene spesso rappresentato nello stesso grafico

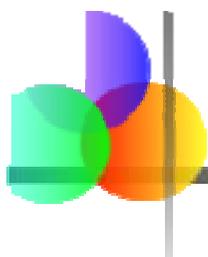
E' usato per separare “**poche cause rilevanti**” dalle “**numerose cause insignificanti**”



Esempio Diagramma di Pareto

Esempio: 400 prodotti difettosi vengono esaminati rispetto al tipo di difetto:

Fonte dell'Errore di Produzione	Numero di difetti
Saldatura Difettosa	34
Cattivo Allineamento	223
Parte Mancante	25
Difetto di Colore	78
Cortocircuito Elettrico	19
Cassa Incrinata	21
Totale	400



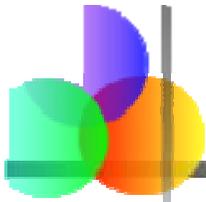
Esempio del Diagramma di Pareto

(continuazione)

Passo 1: Ordinare le categorie rispetto alla causa del difetto, in ordine decrescente

Passo 2: Determinare % per ciascuna categoria

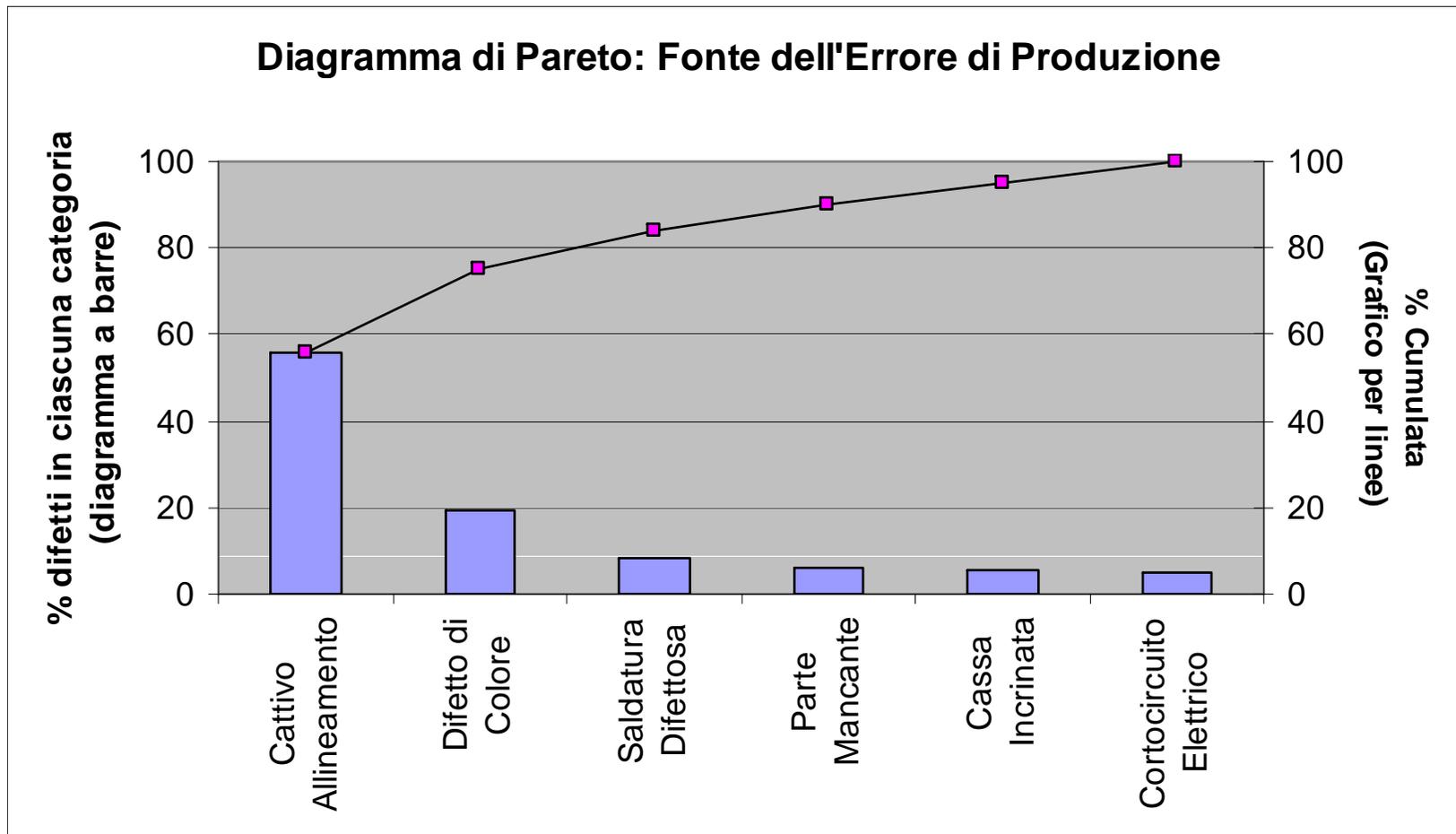
Fonte dell'Errore di Produzione	Numero di difetti	% sul Totale di Difetti
Cattivo Allineamento	223	55.75
Difetto di Colore	78	19.50
Saldatura Difettosa	34	8.50
Parte Mancante	25	6.25
Cassa Incrinata	21	5.25
Cortocircuito Elettrico	19	4.75
Totale	400	100%

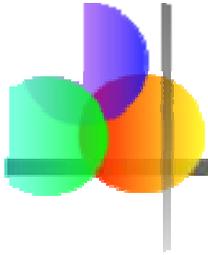


Esempio Diagramma di Pareto

(continuazione)

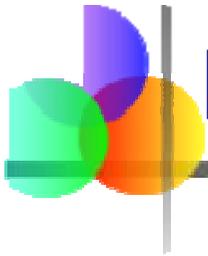
Passo 3: Mostra i risultati graficamente



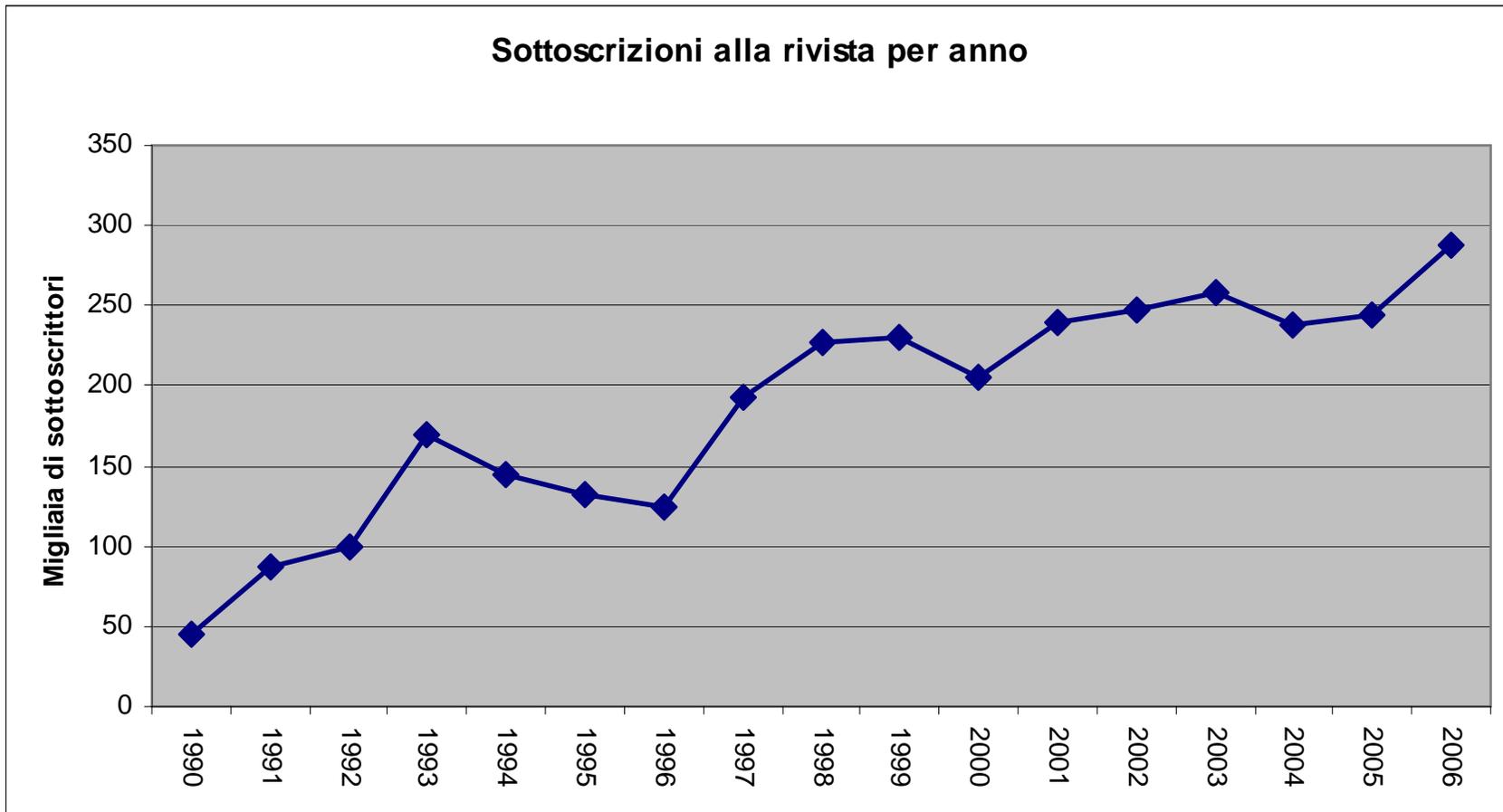


Grafici per Serie Storiche

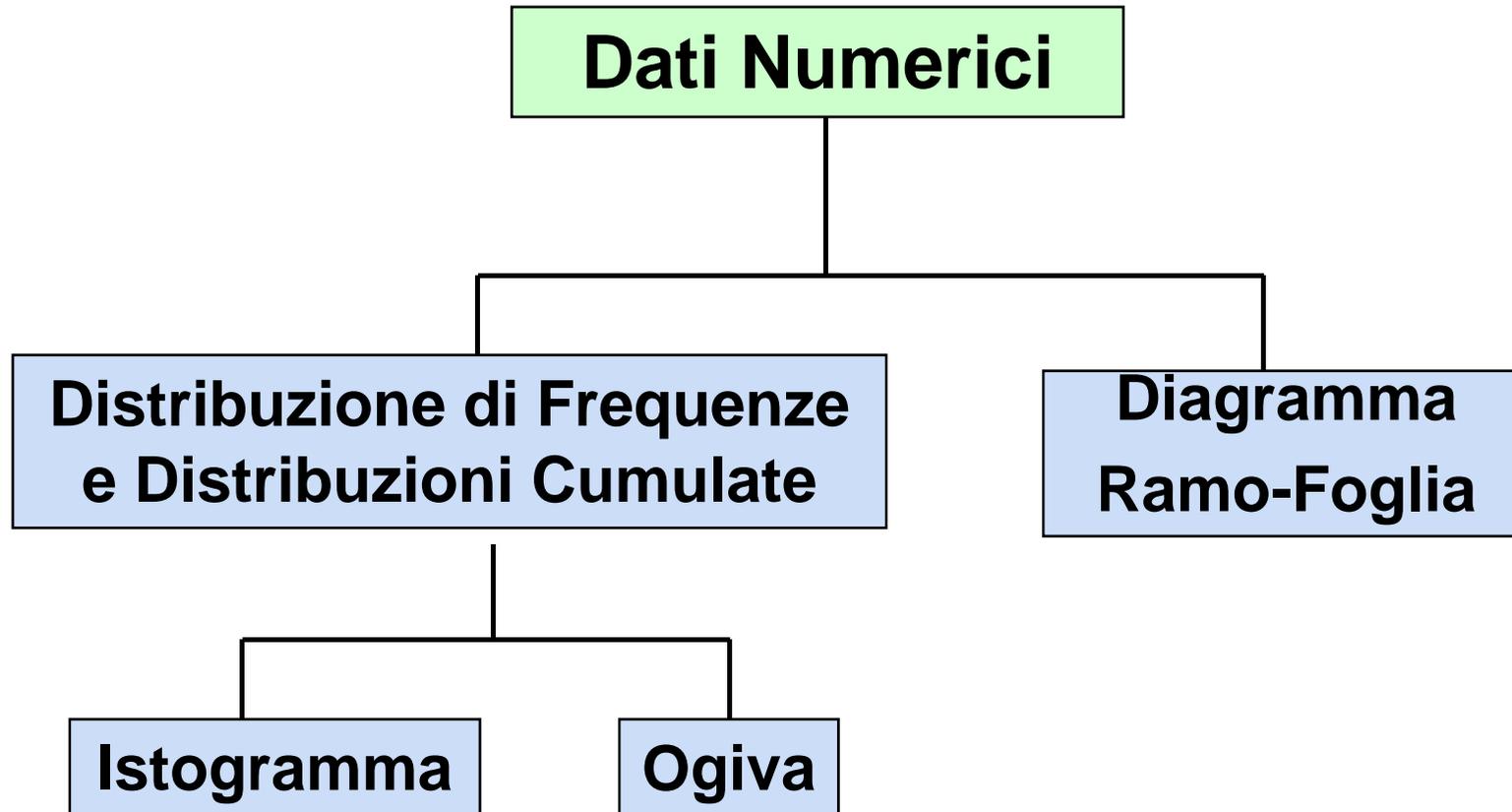
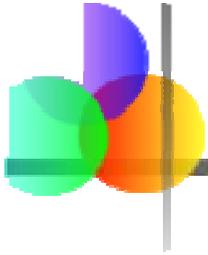
- Un **grafico per serie storiche** viene usato per rappresentare i valori di una variabile nel tempo
- Il tempo viene rappresentato sull'asse orizzontale
- La variabile di interesse viene rappresentata sull'asse verticale

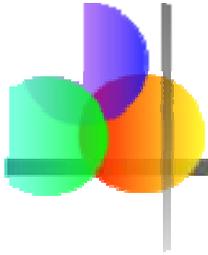


Esempio Grafico per Serie Storiche



Grafici per Descrivere Variabili Numeriche

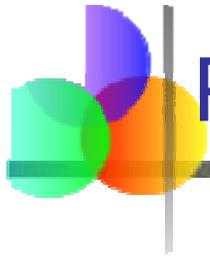




Distribuzione di Frequenze

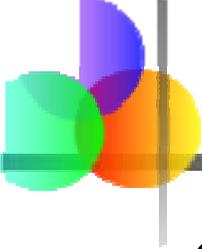
Cosa è una Distribuzione di Frequenze?

- Una distribuzione di frequenze è una **lista o una tabella** ...
- contenente **classi di intervallo** (categorie o intervalli a cui i dati appartengono) ...
- e le **corrispondenti frequenze** con cui i dati appartengono alle classi o categorie



Perchè usare Distribuzioni di Frequenze?

- La distribuzione di frequenze è un modo per riassumere i dati
- La distribuzione condensa i dati grezzi in forma più utile ...
- e consente una veloce interpretazione grafica dei dati

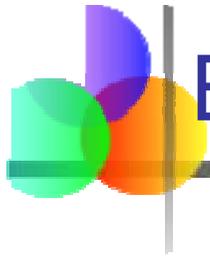


Classi di Intervallo ed Estremi delle Classi

- Ciascuna classe di intervallo dovrebbe avere la stessa ampiezza
- Determinare l'ampiezza di ciascuna classe nel seguente modo:

$$w = \text{Ampiezza dell'intervallo} = \frac{\text{Valore massimo} - \text{Valore minimo}}{\text{Numero di classi}}$$

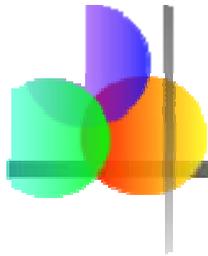
- Usare almeno 5 ma non più di 15-20 intervalli
- E' bene che gli intervalli non si sovrappongono mai
- Arrotondare eventualmente per eccesso (mai per difetto) l'ampiezza dell'intervallo per ottenere i desiderati estremi della classe



Esempio Distribuzione delle Frequenze

Esempio: Un produttore di isolante seleziona a caso 20 osservazioni relative alle temperature registrate dalle componenti elettroniche dei macchinari di un cliente

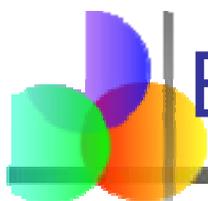
**24, 35, 17, 21, 24, 37, 26, 46, 58, 30,
32, 13, 12, 38, 41, 43, 44, 27, 53, 27**



Esempio Distribuzione delle Frequenze

(continuazione)

- Ordina i dati grezzi in ordine crescente:
12, 13, 17, 21, 24, 24, 26, 27, 27, 30, 32, 35, 37, 38, 41, 43, 44, 46, 53, 58
- Trova il campo di variazione: $58 - 12 = 46$
- Seleziona il numero di classi: **5 (solitamente fra 5 e 15)**
- Calcola l'ampiezza dell'intervallo: **10 (46/5 poi arrotonda per eccesso)**
- Determinare i limiti dell'intervallo: **10 ma meno di 20, 20 ma meno di 30, . . . , 50 ma meno di 60**
- Conta le osservazioni & le assegna alle classi



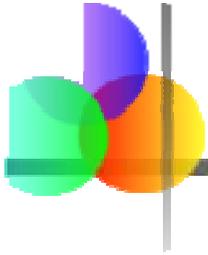
Esempio Distribuzione delle Frequenze

(continuazione)

Dati in sequenza ordinata:

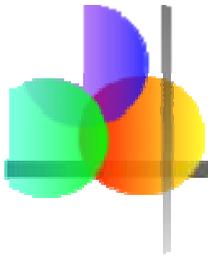
12, 13, 17, 21, 24, 24, 26, 27, 27, 30, 32, 35, 37, 38, 41, 43, 44, 46, 53, 58

Intervallo	Frequenza	Frequenza Relativa	Percentuale
10 ma meno di 20	3	.15	15
20 ma meno di 30	6	.30	30
30 ma meno di 40	5	.25	25
40 ma meno di 50	4	.20	20
50 ma meno di 60	2	.10	10
Totale	20	1.00	100



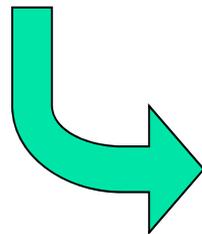
Istogramma

- Un grafico dei dati contenuti in una distribuzione di frequenze è chiamato **istogramma**
- Gli **estremi degli intervalli** sono rappresentati sull'asse orizzontale
- L' **asse verticale** rappresenta le **frequenze solo se** le classi hanno pari ampiezza
- Barre di altezza appropriata sono usate per rappresentare la numerosità di ciascuna classe



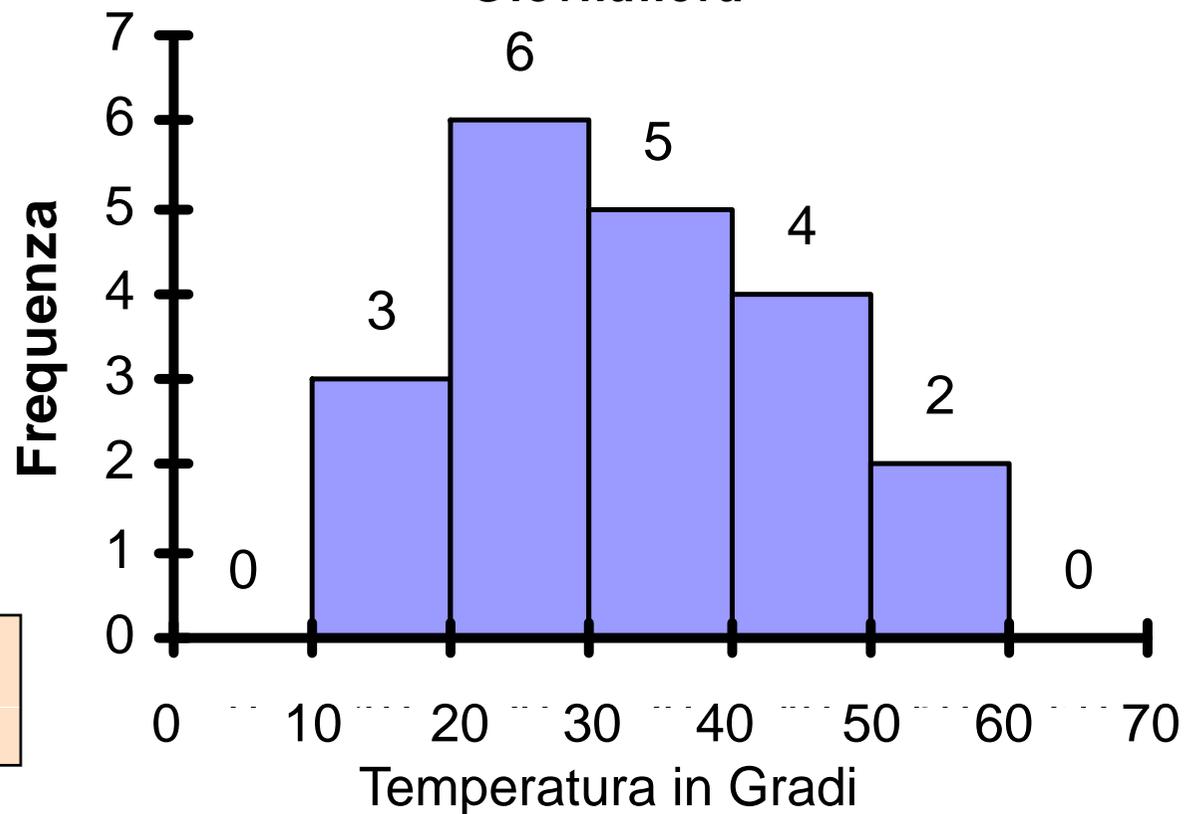
Esempio Istogramma

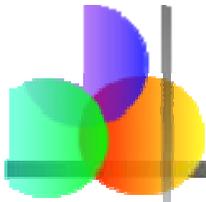
Intervallo	Frequenza
10 ma meno di 20	3
20 ma meno di 30	6
30 ma meno di 40	5
40 ma meno di 50	4
50 ma meno di 60	2



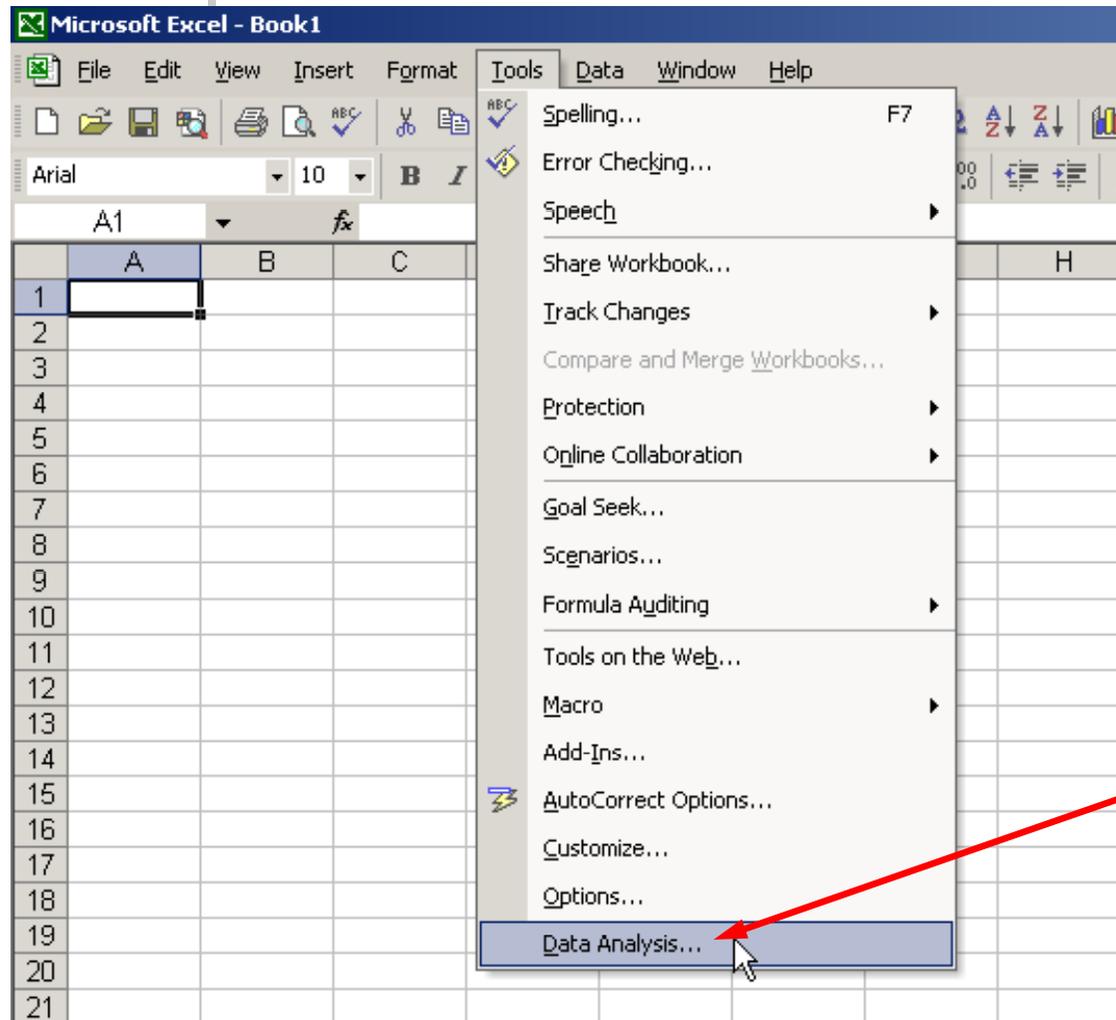
(No spazio
fra le barre)

Istogramma: Temperatura Massima
Giornaliera



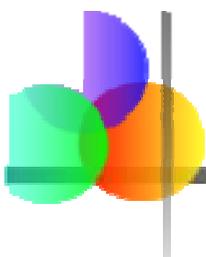


Istogramma in Excel



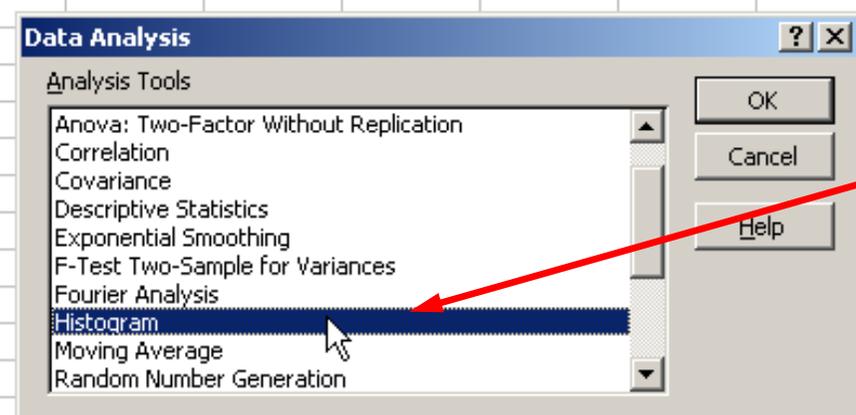
1

Selezionare
Strumenti/Analisi Dati



Istogramma in Excel

(continuazine)



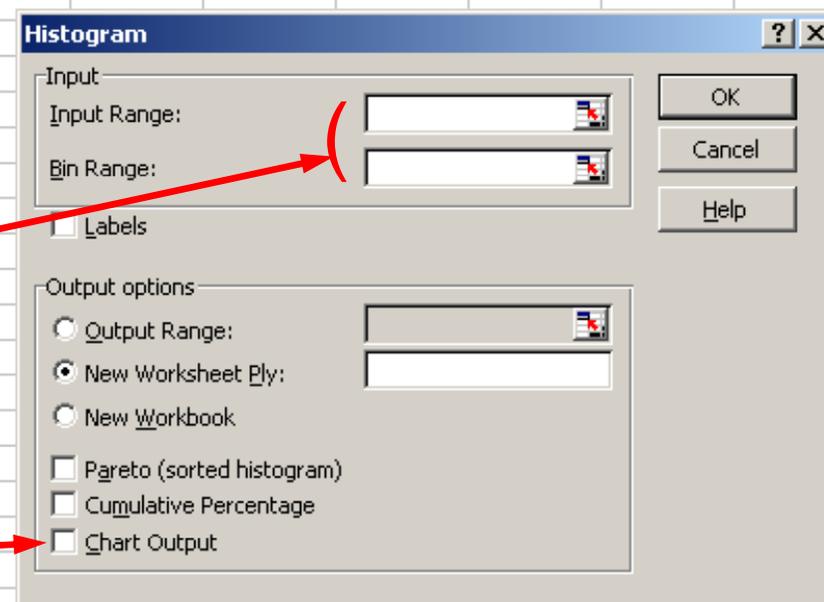
2

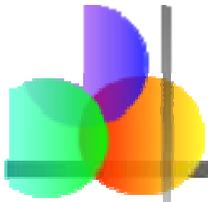
Scegliere Istogramma

Inserire le celle contenenti i dati e i bins (celle contenenti i limiti superiori delle classi)

3

Selezionare Grafico in Output e cliccare su "OK"

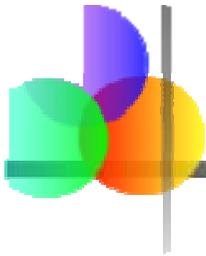




Domande per Raggruppare i Dati in Intervalli

- 1. Quale dovrebbe essere l'ampiezza di ciascun intervallo?
(Quante classi dovrebbero essere usate?)

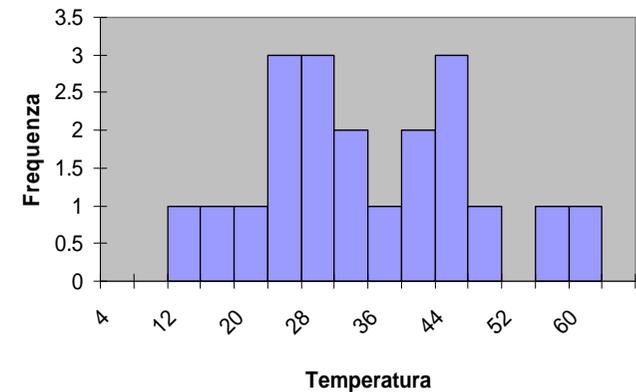
- 2. Come dovrebbero essere determinati gli estremi degli intervalli?
 - Spesso la risposta si trova per tentativi, ed è soggettiva
 - L'obiettivo è di creare una distribuzione che non è troppo "frastagliata" ma nemmeno troppo "a blocchi"
 - L'obiettivo è di mostrare appropriatamente l'andamento della variazione nei dati



Quante classi di intervallo?

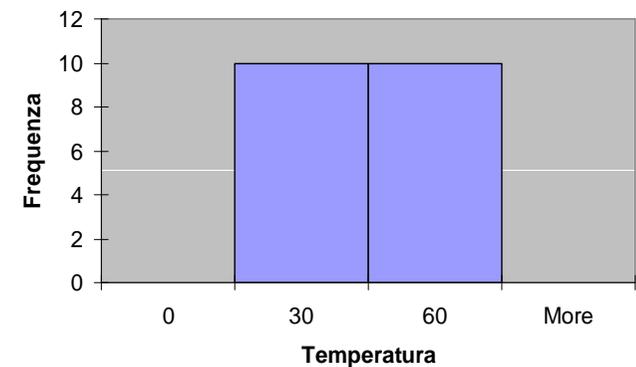
■ Troppe (classi di intervallo piccole)

- può produrre una distribuzione molto frastagliata con spazi dovuti a classi vuote
- può dare una cattiva indicazione di come la frequenza cambia nelle classi

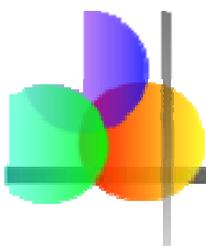


■ Poche (classi di intervallo ampie)

- può comprimere troppo la variazione e produrre una distribuzione a blocchi
- può oscurare importanti andamenti nella variazione.



(etichette sull'asse delle X sono i limiti superiori delle classi)

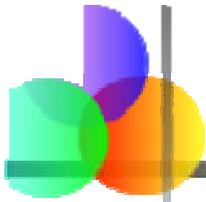


La Distribuzione di Frequenze Cumulate

Dati in sequenza ordinata:

12, 13, 17, 21, 24, 24, 26, 27, 27, 30, 32, 35, 37, 38, 41, 43, 44, 46, 53, 58

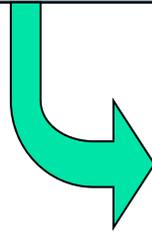
Classi	Frequenza	Percentuale	Frequenza Cumulata	Percentuale Cumulata
10 ma meno di 20	3	15	3	15
20 ma meno di 30	6	30	9	45
30 ma meno di 40	5	25	14	70
40 ma meno di 50	4	20	18	90
50 ma meno di 60	2	10	20	100
Totale	20	100		



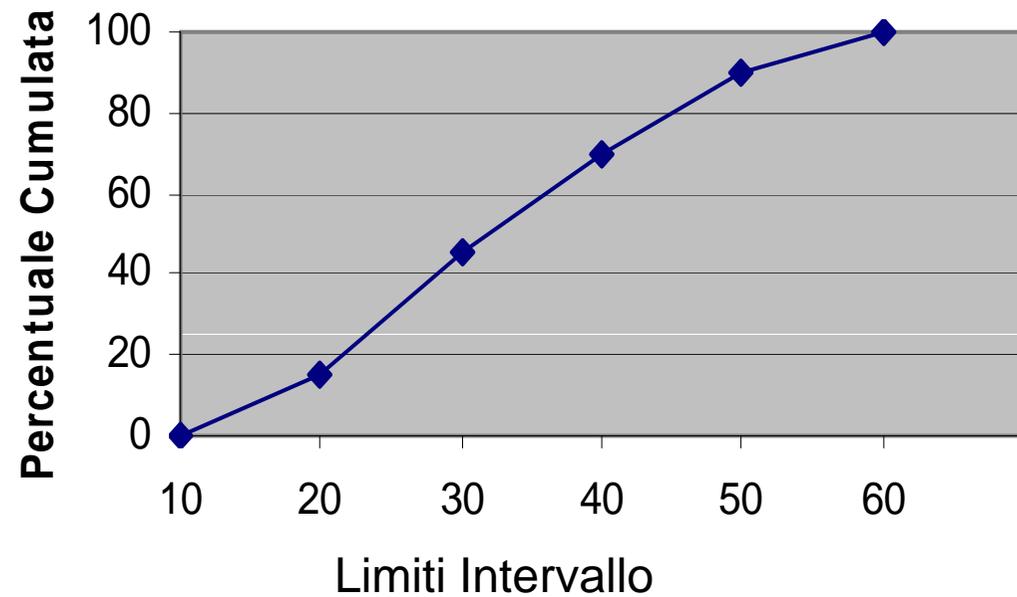
Le Ogive

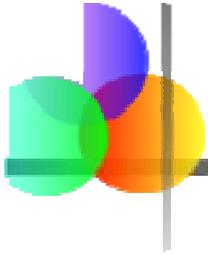
Grafici per Frequenze Cumulate

Intervallo	Limite Superiore Intervallo	Percentuale Cumulata
Meno di 10	10	0
10 ma meno di 20	20	15
20 ma meno di 30	30	45
30 ma meno di 40	40	70
40 ma meno di 50	50	90
50 ma meno di 60	60	100



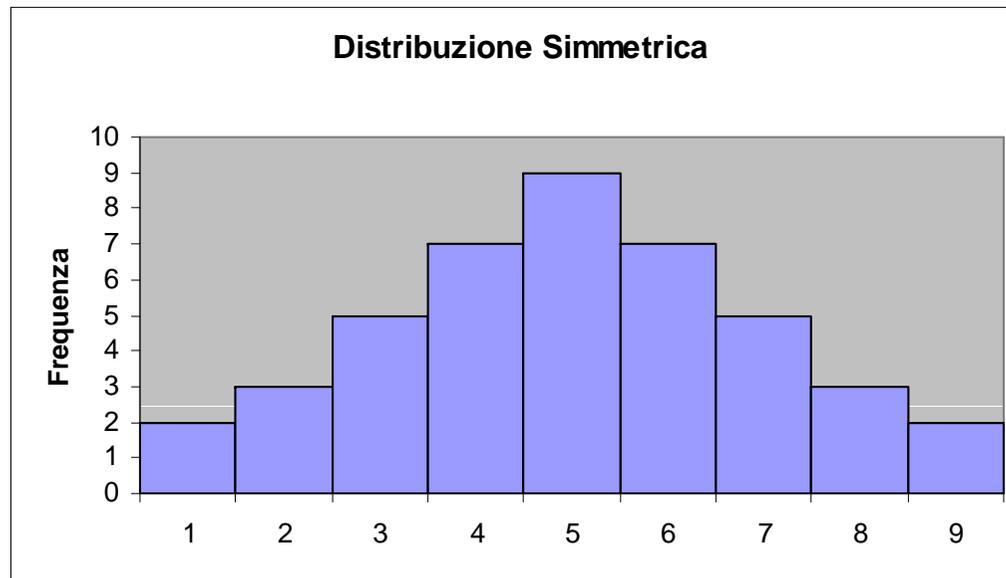
Ogiva: Temperatura Massima Giornaliera

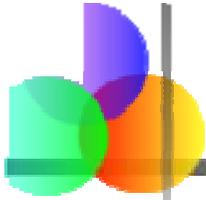




Forma della Distribuzione

- La forma della distribuzione si dice **simmetrica** se le osservazioni sono bilanciate, o distribuite in modo approssimativamente regolare attorno al centro.



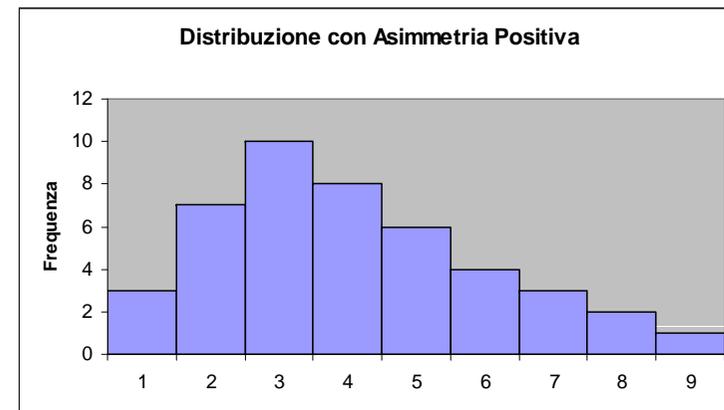


Forma della Distribuzione

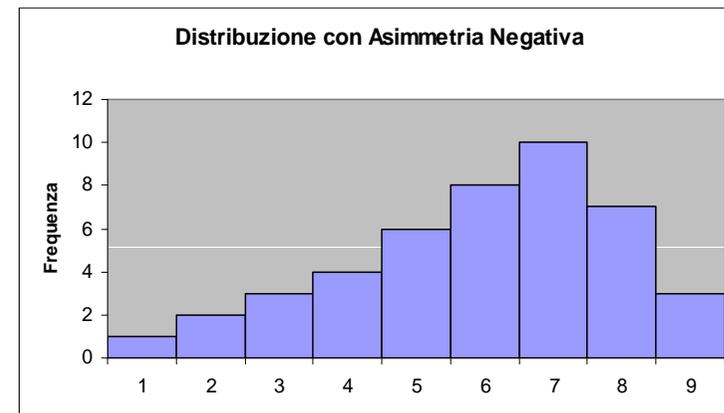
(continuazione)

- La forma della distribuzione è detta **asimmetrica** se le osservazioni non sono distribuite in modo simmetrico rispetto al centro.

Una distribuzione con **asimmetria positiva** (obliqua a destra) ha una coda che si estende a destra, nella direzione dei valori superiori.



Una distribuzione con **asimmetria negativa** (obliqua a sinistra) ha una coda che si estende a sinistra, nella direzione dei valori inferiori.



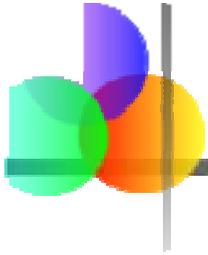
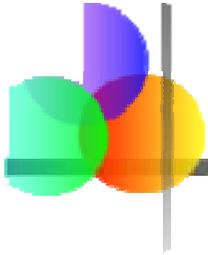


Diagramma Ramo-Foglia

- Si tratta di un modo semplice per vedere i dettagli della distribuzione di un set di dati, abbinando il dettaglio della distribuzione di frequenze e la sintesi grafica propria dell'istogramma

METODO: Separare la serie di dati ordinata in cifre più significative (i **rami**) e cifre meno significative (le **foglie**)



Esempio

Dati ordinati:

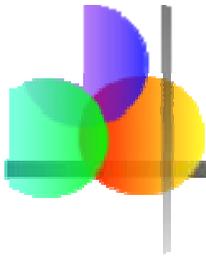
21, 24, 24, 26, 27, 27, 30, 32, 38, 41

- Qui usiamo le decine come unità per i rami:

- 21 è mostrato come

- 38 è mostrato come

Ramo	Foglia
2	1
3	8



Esempio

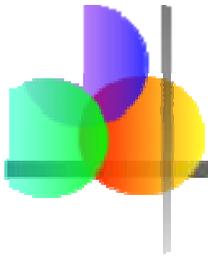
(continuazione)

Dati ordinati:

21, 24, 24, 26, 27, 27, 30, 32, 38, 41

- Diagramma ramo-foglia completato:

Ramo	Foglia
2	1 4 4 6 7 7
3	0 2 8
4	1

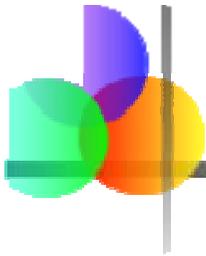


Usando altre unità per i rami

- Usando le centinaia come unità per i rami:
 - Arrotondiamo alla decina per formare le foglie

- 613 diventerebbe
- 776 diventerebbe
- . . .
- 1224 diventa

	Ramo	Foglia
→	6	1
→	7	8
→	12	2



Usando altre unità per i rami

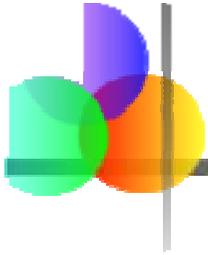
(continuazione)

- Usando le centinaia come unità per i rami:
 - Il diagramma ramo-foglia completato:

Dati:
613, 632, 658, 717, 722, 750, 776, 827, 841, 859, 863, 891, 894, 906, 928, 933, 955, 982, 1034, 1047, 1056, 1140, 1169, 1224

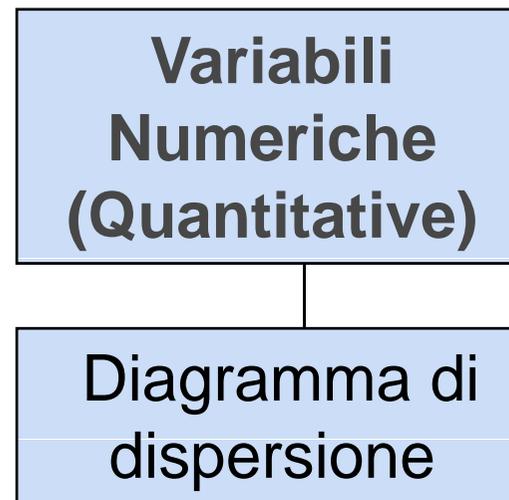
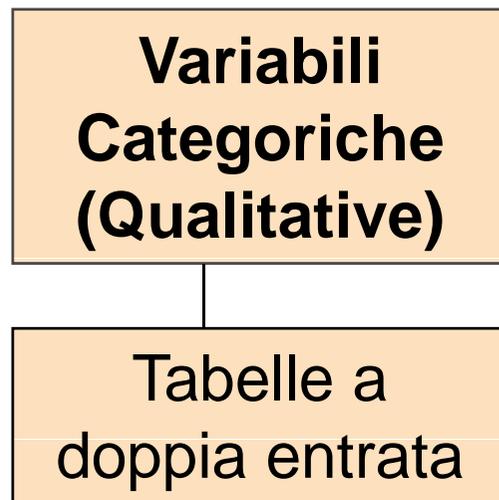


Ramo	Foglia
6	1 3 6
7	2 2 5 8
8	3 4 6 6 9 9
9	1 3 3 6 8
10	3 5 6
11	4 7
12	2



Relazioni fra Variabili

- I grafici illustrati finora si riferiscono ad una sola variabile
- Quando consideriamo **due variabili** vengono usate altre tecniche:



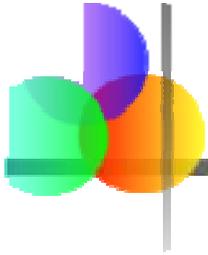
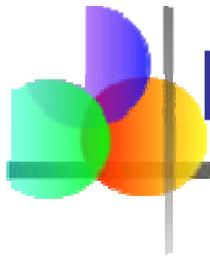


Diagramma di Dispersione

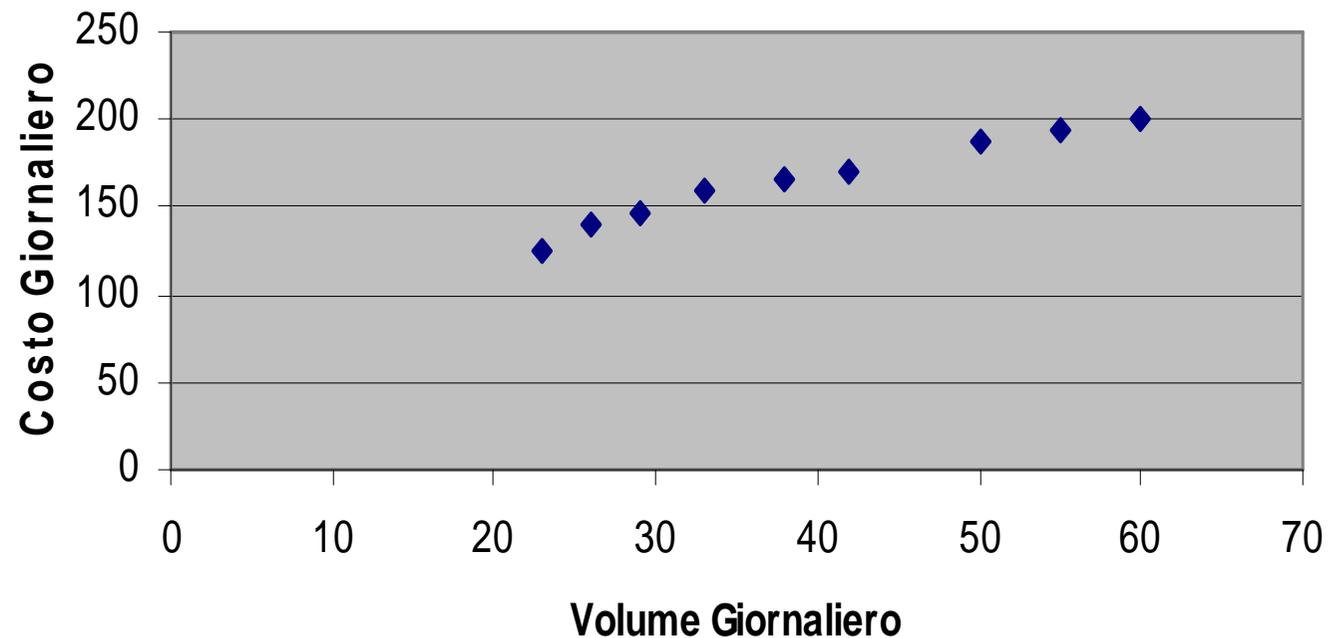
- **Diagrammi di dispersione** sono usati per osservazioni accoppiate relative a due variabili numeriche
- Il diagramma di dispersione:
 - Una variabile viene rappresentata sull'asse verticale e l'altra variabile viene rappresentata sull'asse orizzontale



Esempio Diagramma di Dispersione

Volume Produzione	Costo Giornaliero
23	125
26	140
29	146
33	160
38	167
42	170
50	188
55	195
60	200

Costo Giornaliero vs. Volume di Produzione



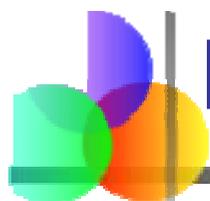


Diagramma di Dispersione in Excel

1

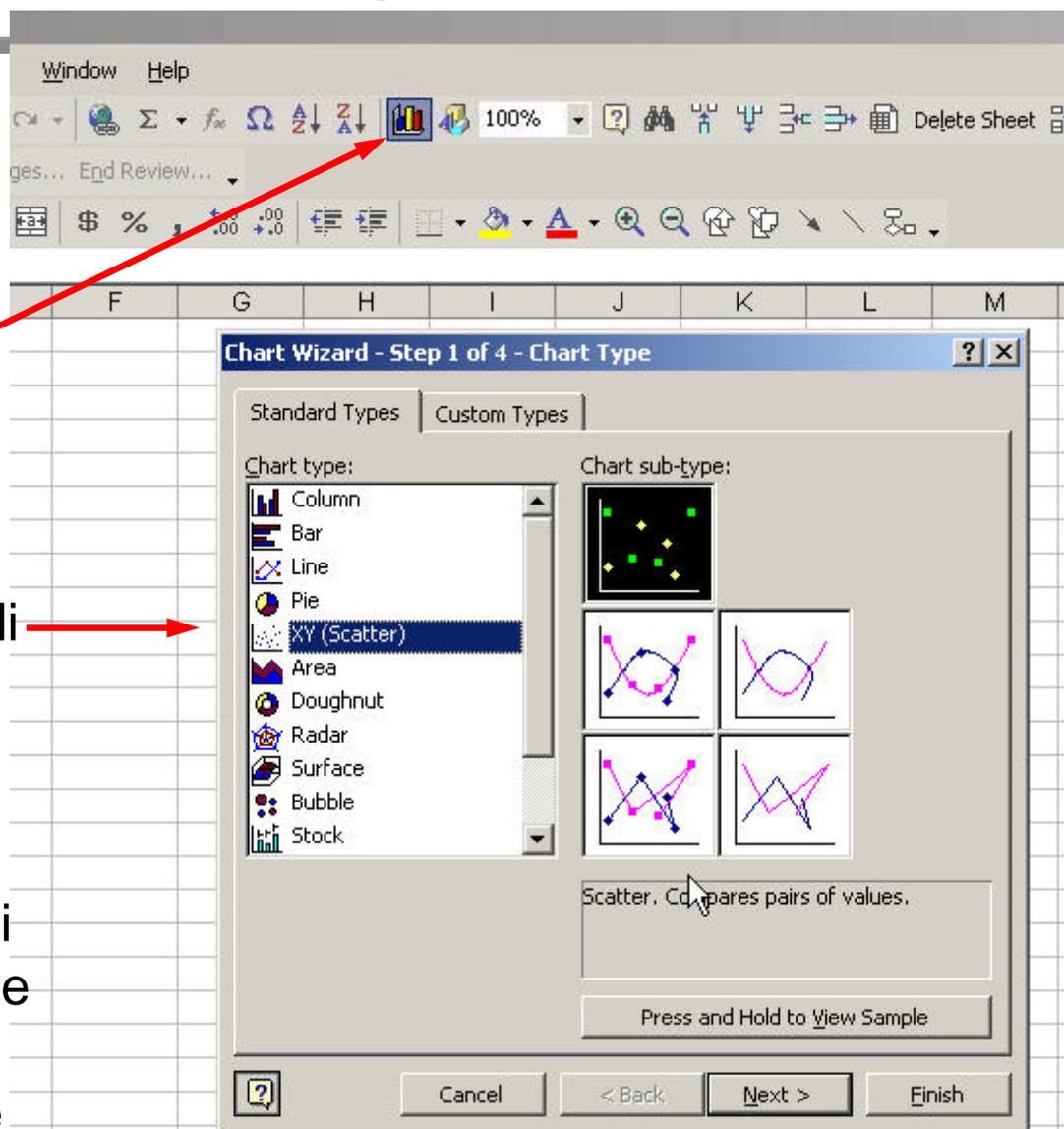
Selezionare Creazione Guida Grafica

2

Selezionare l'opzione Dispers. (XY), e quindi cliccare su "Avanti"

3

Quando richiesto, per completare il diagramma di dispersione, inserire le celle contenenti i dati, le etichette, e la destinazione



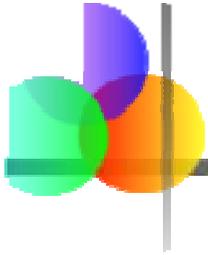
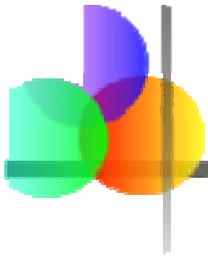


Tabelle a Doppia Entrata

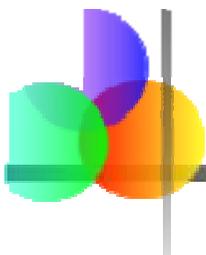
- **Tabelle a doppia entrata** (o tabelle di contingenza) elencano il numero di osservazioni per ogni combinazione di valori per le due variabili categoriche o ordinali.
- Se ci sono r categorie per la prima variabile (righe) e c categorie per la seconda variabile (colonne), la tabella viene chiamata tabella a doppia entrata $r \times c$.



Esempio Tabella a Doppia Entrata

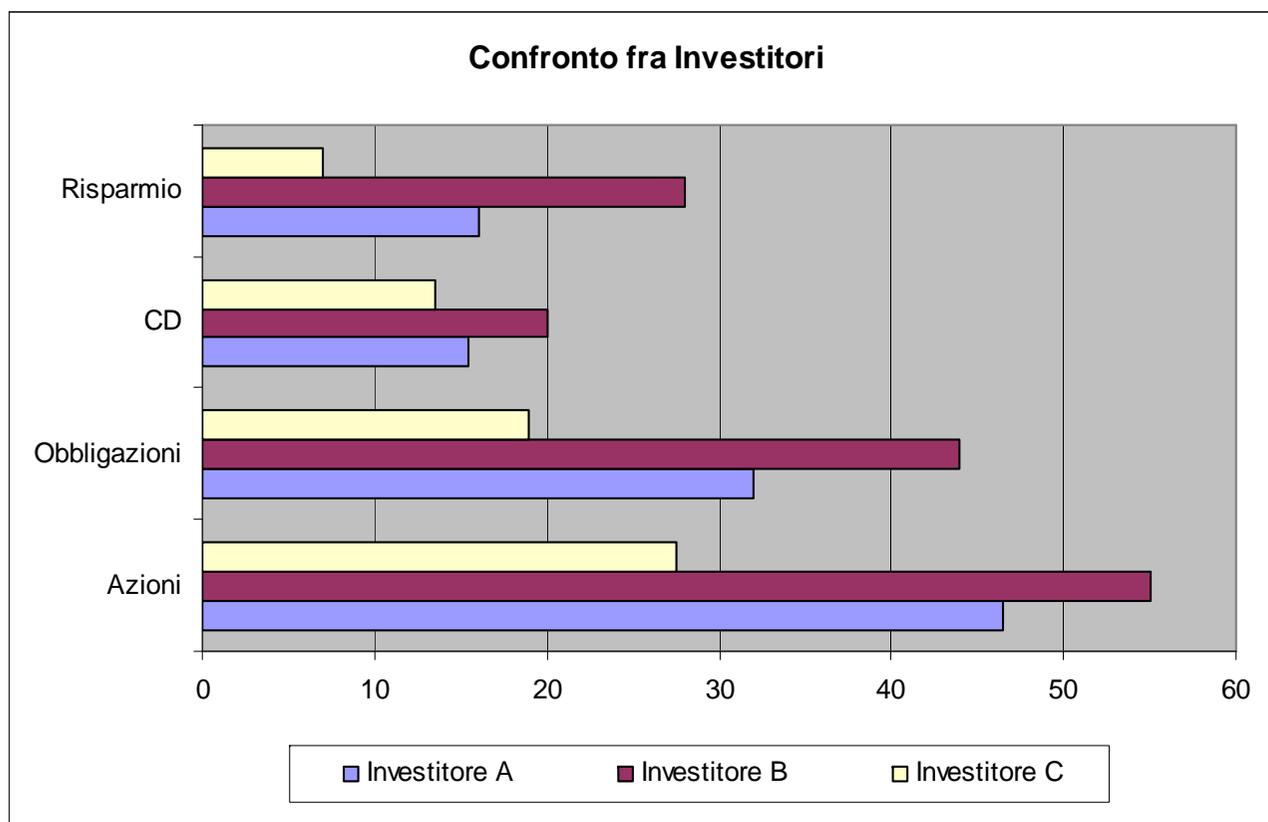
- Tabella a doppia entrata 4 x 3 per Scelta di Investimento per Investitore (valori in migliaia di dollari)

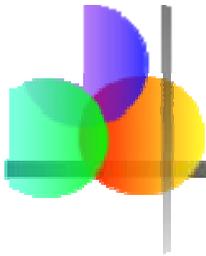
Categoria Investimento	Investitore A	Investitore B	Investitore C	Totale
Azioni	46.5	55	27.5	129
Obbligazioni	32.0	44	19.0	95
CD	15.5	20	13.5	49
Risparmio	16.0	28	7.0	51
Totale	110.0	147	67.0	324



Rappresentazione Grafica di Dati Categorici Multivariati

- Diagramma a barre accostate

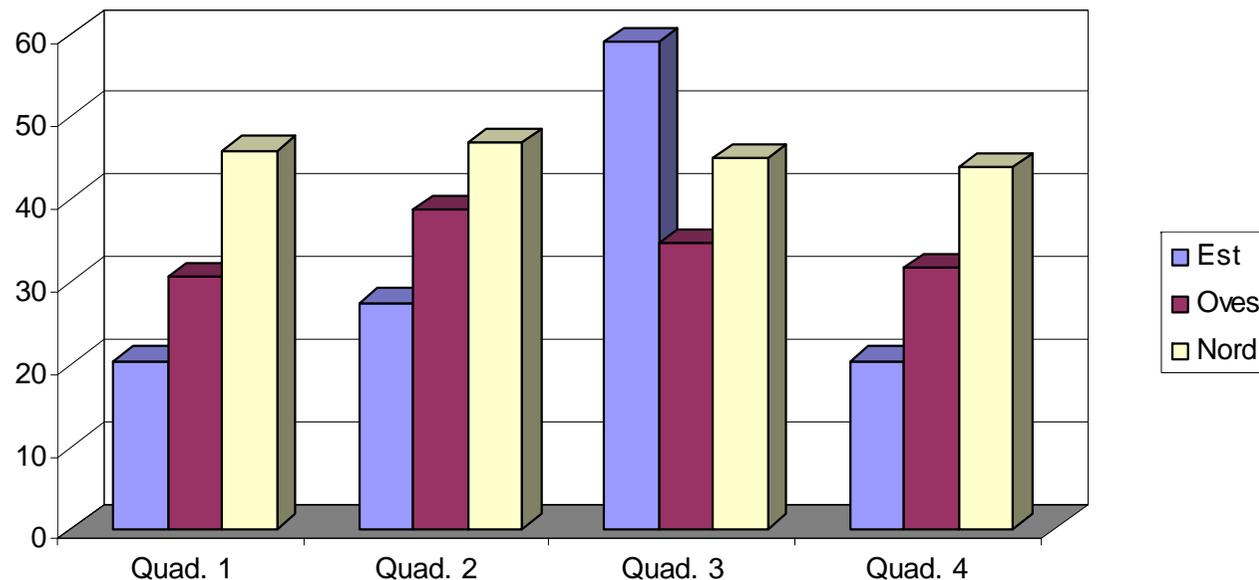


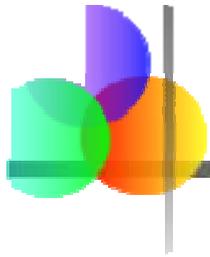


Esempio Grafico Accostato

- Vendite quadrimestrali per tre zone di vendita:

	Quad. 1	Quad. 2	Quad. 3	Quad. 4
Est	20.4	27.4	59	20.4
Ovest	30.6	38.6	34.6	31.6
Nord	45.9	46.9	45	43.9

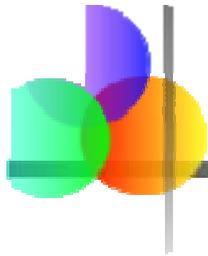




Errori nella Presentazione dei Dati

Obiettivi per una presentazione efficace dei dati:

- Presentare i dati in modo da mostrare le informazioni essenziali
- Comunicare idee complesse chiaramente ed in modo accurato
- Evitare distorsioni che possono comunicare il messaggio sbagliato

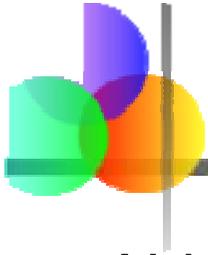


Errori nella Presentazione dei Dati

(continuazione)

- Diversa ampiezza delle classi di intervallo in un istogramma
- Compressione o distorsione dell'asse verticale
- Omissione dello zero sull'asse verticale
- Non fornire una base di riferimento per il confronto di dati di diversi gruppi





Riepilogo del Capitolo

- Abbiamo discusso tipi di dati e livelli di misurazione
- Dati in forma grezza di solito non sono facili da usare nel processo decisionale – Qualche tipo di organizzazione è necessaria:

◆ Tabelle

◆ Grafici

- Tecniche viste in questo capitolo:

- Distribuzione di frequenze
- Diagramma a barre
- Diagramma a torta
- Diagramma di Pareto
- Plot per serie storiche
- Distribuzione di frequenze
- Istogramma and ogiva
- Diagramma ramo-foglia
- Diagramma di dispersione
- Tabelle a doppia entrata e diagrammi a barre accostate