



UNIVERSITÀ DEGLI STUDI DI ROMA "FORO ITALICO"

Rappresentazione grafica delle distribuzioni

Corso di Laurea in Scienze Motorie e Sportive (L22)

Insegnamento di Informatica

Prof. Federico Mari

Statistica di base

Seconda edizione

David S. Moore



Libro di testo

D.S. Moore. [Statistica di base](#). Idee & strumenti. Apogeo, 2013. ISBN: 978-8838786426.

Parte I

Capitolo 1

Statistica

Descrittiva

- Analisi dei dati

Inferenziale

- Produzione dei dati dalla popolazione
- Induzione e previsione su tutta la popolazione

Dati

- **Unità statistiche**
Oggetti descritti tramite un insieme di dati (persone, animali, cose)
- **Variabili**
Caratteristica associata a un'unità (valore e unità di misura)

Esempio: Registro degli studenti di un'università

- **Unità:** gli studenti
- **Variabili:** data di nascita, sesso (m/f), facoltà, media dei voti

Nome	Facoltà	Corso	Esonero 1	Esonero 2	Esame	Totale	Valutazione
Bianchi, Mario	Economia	Statistica	24	26	29	79	Buono
Blu, Andrea	Scienze	Informatica	28	26	28	82	Buono
Carso, Guido	Scienze	Informatica	27	30	29	86	Ottimo

Variabili

- **Variabili categoriche**
Collocano una unità in un gruppo o categoria

- **Variabili quantitative**
Assumono valori numerici per cui operazioni aritmetiche (somme, media, ...) hanno senso

Esempio: Registro degli studenti di un'università

- **Categoriche:** facoltà, corso, val.
- **Quantitative:** esoneri, esame, totale

Nome	Facoltà	Corso	Esonero 1	Esonero 2	Esame	Totale	Valutazione
Bianchi, Mario	Economia	Statistica	24	26	29	79	Buono
Blu, Andrea	Scienze	Informatica	28	26	28	82	Buono
Carso, Guido	Scienze	Informatica	27	30	29	86	Ottimo

Analisi esplorativa dei dati



Si serve di grafici e indici numerici per descrivere il comportamento delle variabili in un insieme di dati e le relazioni tra esse



Chi? Quanti?

Quale tipo di unità statistiche e quante unità osserviamo



Quali e quante variabili?

Definizioni esatte delle variabili e delle unità di misura



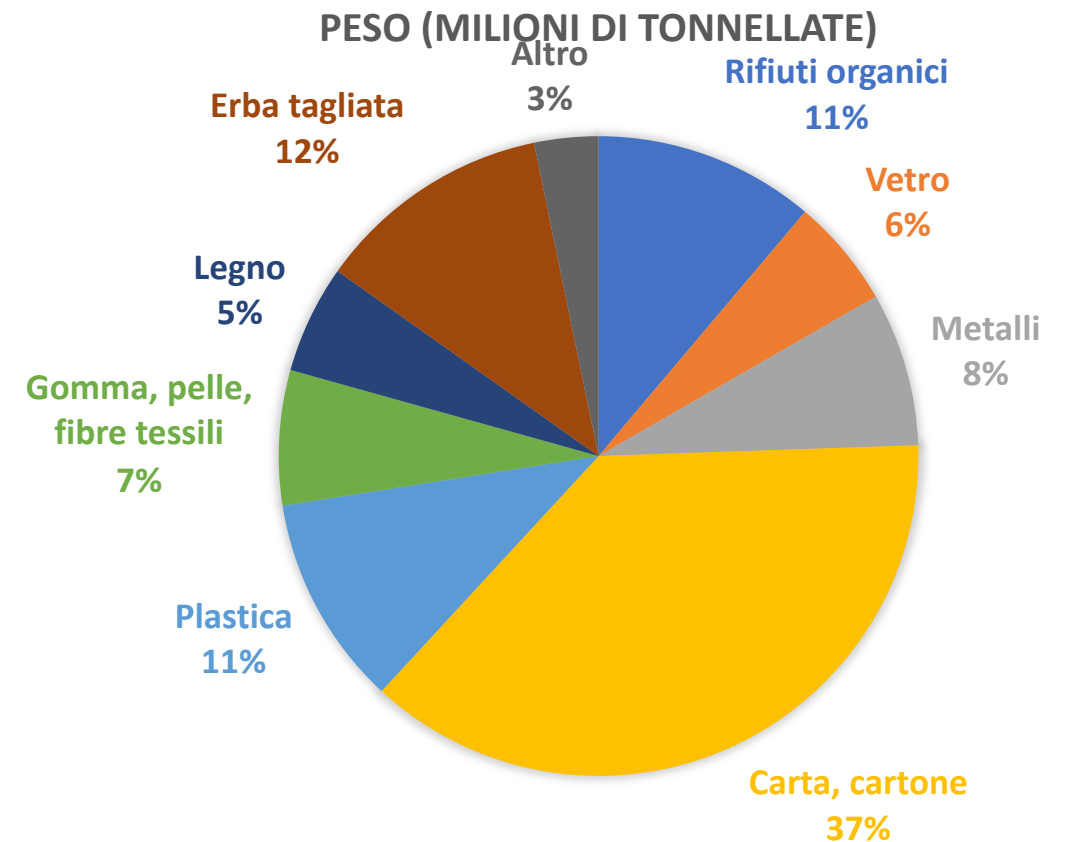
Perché?

Quale obiettivo ha l'analisi

Variabili categoriche: diagrammi a torta

Diagrammi a torta

Materiale	Peso (milioni di tonnellate)	Percentuale sul totale
Rifiuti organici	25,9	11,2%
Vetro	12,8	5,5%
Metalli	18	7,8%
Carta, cartone	86,7	37,4%
Plastica	24,7	10,7%
Gomma, pelle, fibre tessili	15,8	6,8%
Legno	12,7	5,5%
Erba tagliata	27,7	11,9%
Altro	7,5	3,2%
Totale	231,8	100,0%



Distribuzione di una variabile



Indica quali valori la variabile assume e quanto spesso

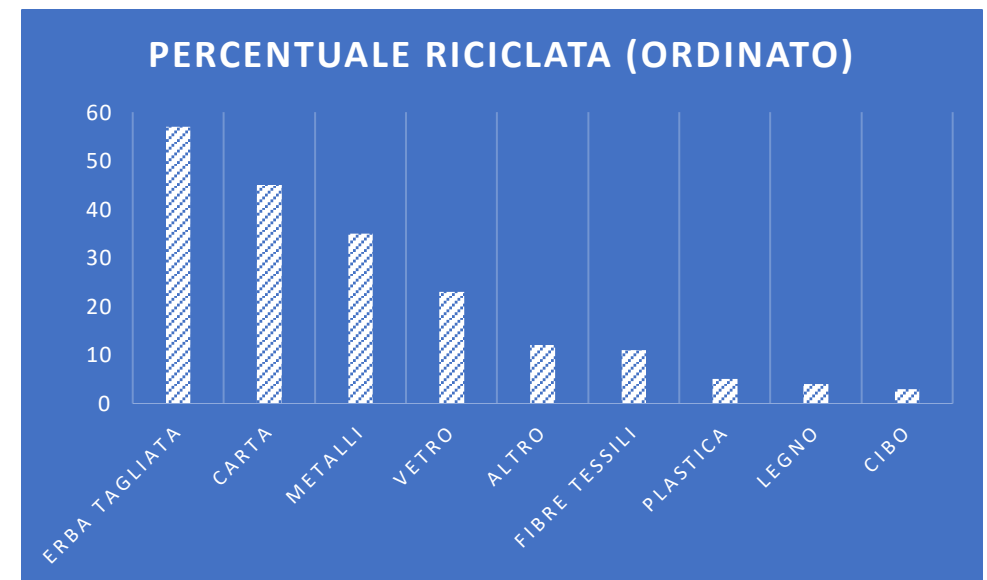
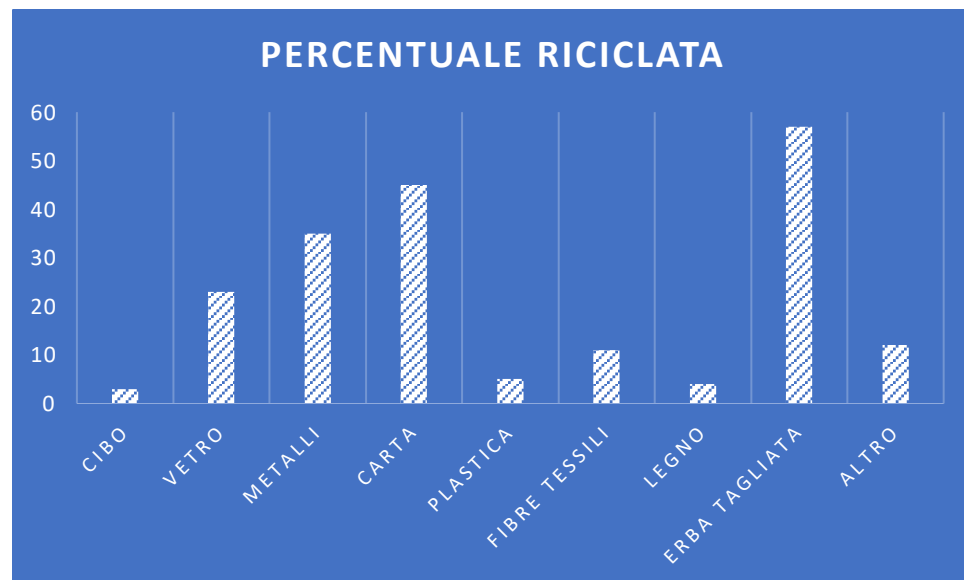


Diversa a seconda
del tipo di variabile

Variabili
categoriche:
diagramma
a torta

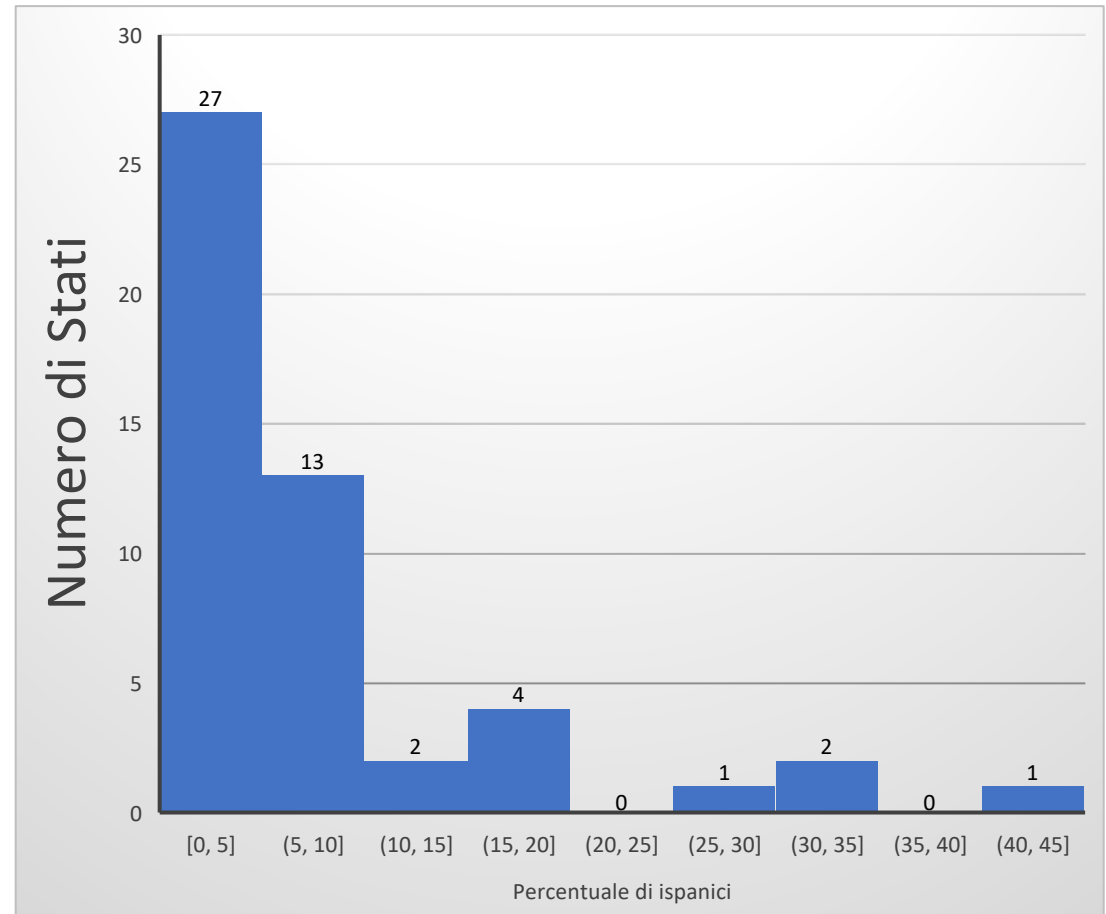
Variabili categoriche: grafici a barre

- Peso di ogni unità è dato dall'altezza delle barre
- Le barre non costituiscono le parti di un insieme
- Spesso ordinare le barre per altezza aumenta l'efficacia comunicativa



Variabili quantitative

- Assumono spesso molti valori
- Grafico della distribuzione è più chiaro se i valori più vicini sono raggruppati
- **Istogramma** è il grafico più comune per questo tipo di distribuzione

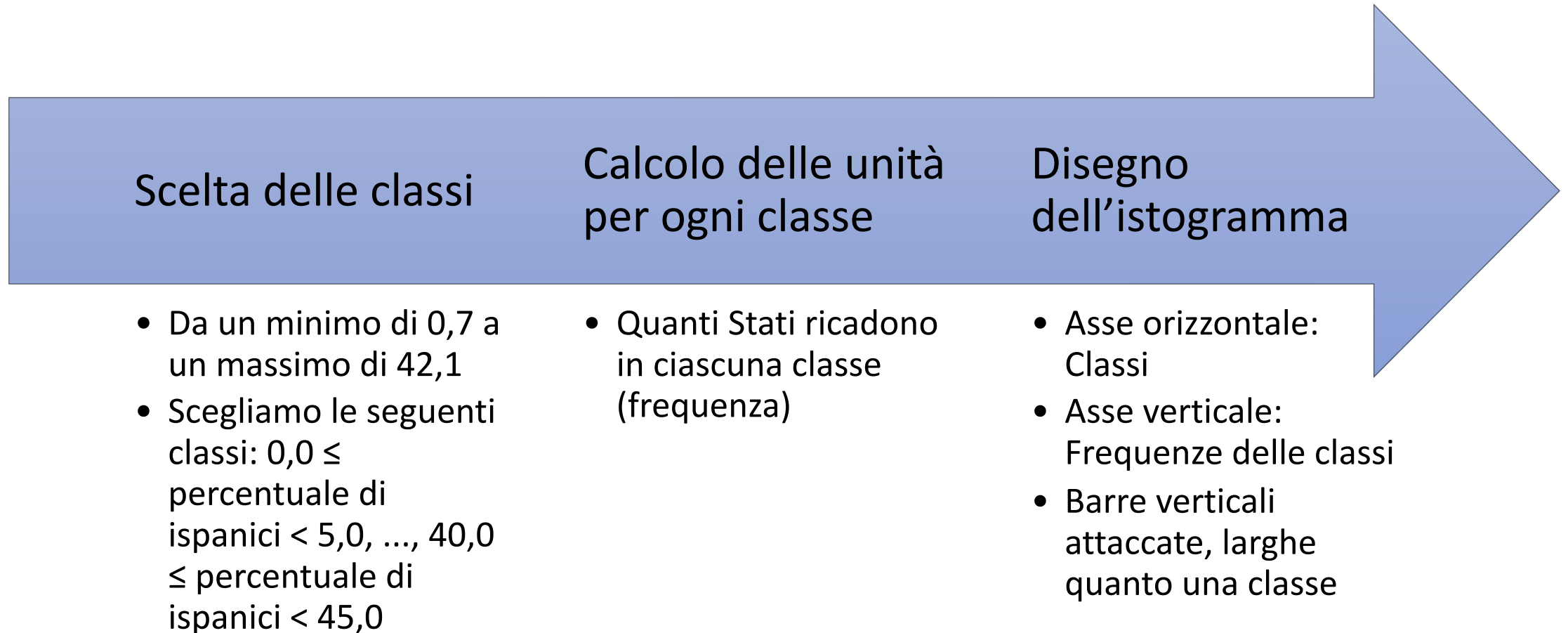


Variabili quantitative: istogrammi

Stato	Percentuale	Stato	Percentuale	Stato	Percentuale
Alabama	1,5	Louisiana	2,4	Ohio	1,9
Alaska	4,1	Maine	0,7	Oklahoma	5,2
Arizona	25,3	Maryland	4,3	Oregon	8
Arkansas	2,8	Massachusetts	6,8	Pennsylvania	3,2
California	32,4	Michigan	3,3	Rhode Island	8,7
Colorado	17,1	Minnesota	2,9	South Carolina	2,4
Connecticut	9,4	Mississippi	1,3	South Dakota	1,4
Delaware	4,8	Missouri	2,1	Tennessee	2
Florida	16,8	Montana	2	Texas	32
Georgia	5,3	Nebraska	5,5	Utah	9
Hawaii	7,2	Nevada	19,7	Vermont	0,9
Idaho	7,9	New Hampshire	1,7	Virginia	4,7
Illinois	10,7	New Jersey	13,3	Washington	7,2
Indiana	3,5	New Mexico	42,1	West Virginia	0,7
Iowa	2,8	New York	15,1	Wisconsin	3,6
Kansas	7	North Carolina	4,7	Wyoming	6,4
Kentucky	1,5	North Dakota	1,2		

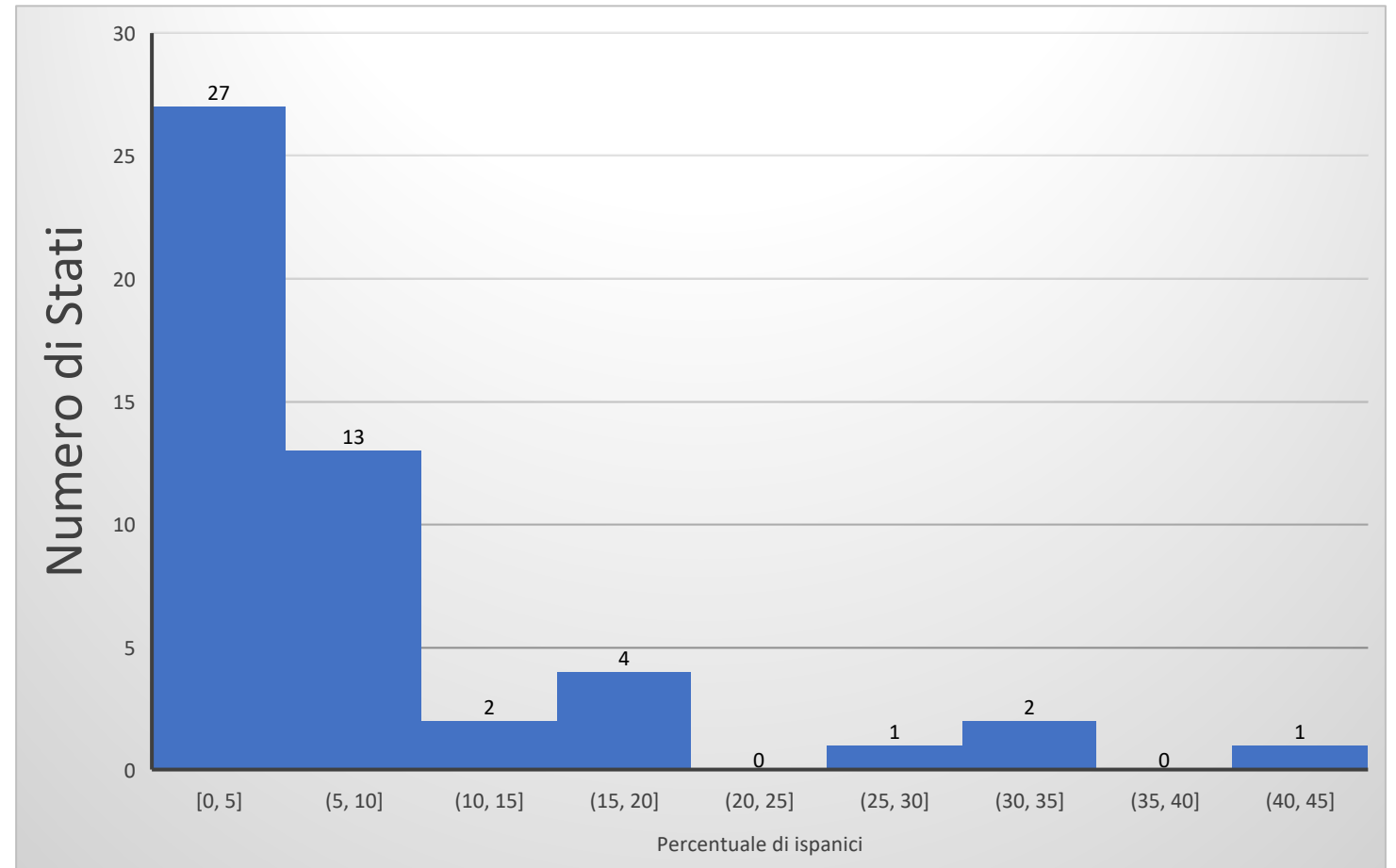
Popolazione di origine ispanica per Stato americano (2000)

Variabili quantitative: istogrammi



Variabili quantitative: istogrammi

- Istogramma della distribuzione della percentuale di ispanici tra i residenti dei 50 Stati americani
- Distribuzione Asimmetrica a destra



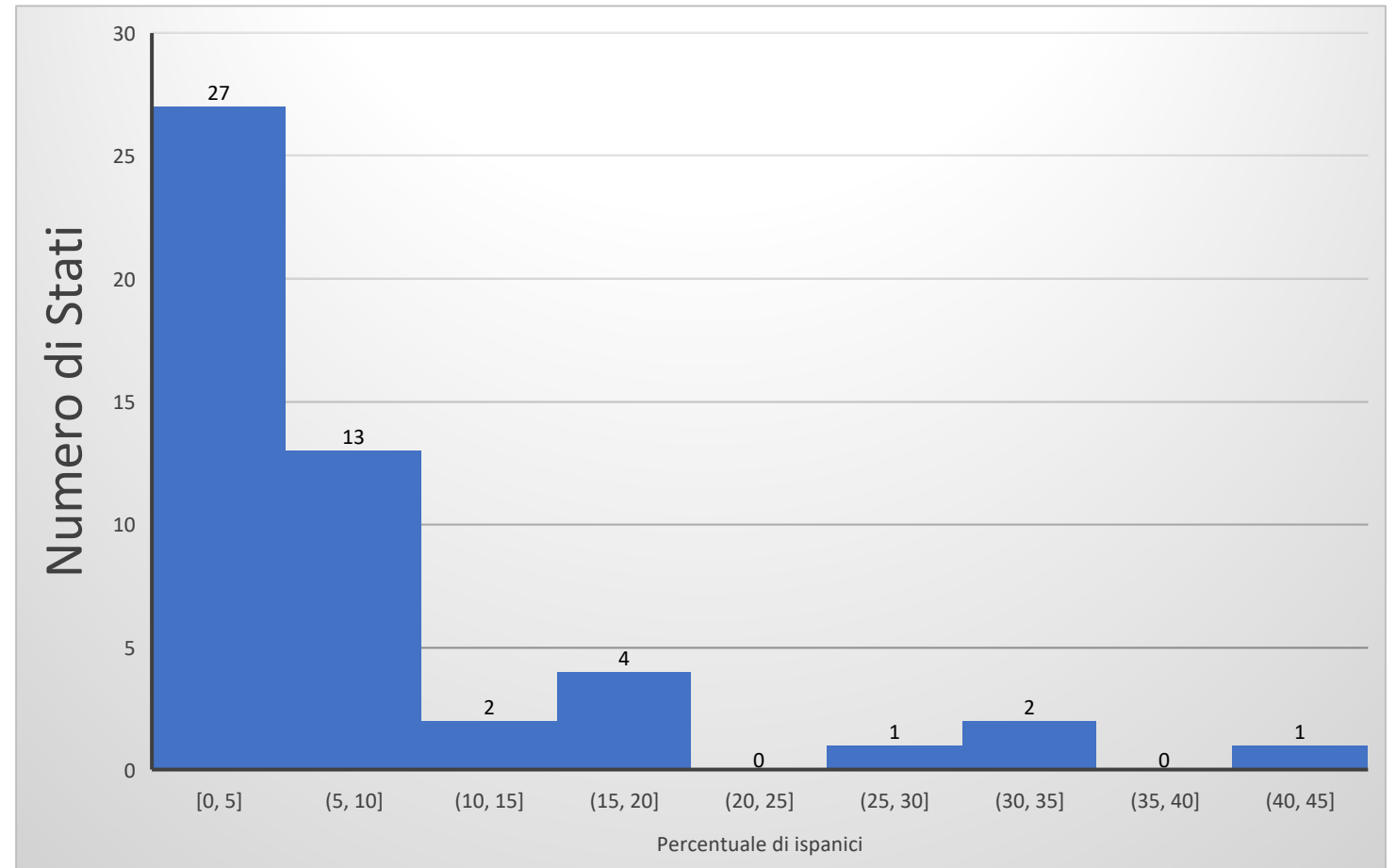
Analizzare una distribuzione

- Cosa si può imparare da una distribuzione?
- È importante individuare il **modello generico** e le **deviazioni** evidenti rispetto a tale modello
- Si individua il modello generico attraverso
 - **Forma**: picchi, simmetria
 - **Punto centrale**: metà osservazioni sono più piccole e metà sono più grandi
 - **Dispersione**: minimo, massimo, *range* o campo di variazione (max–min)
- Deviazione importante: **Outlier** cioè un singolo valore che si discosta molto dallo schema generale osservato

Variabili quantitative: istogrammi

Analisi della distribuzione:

- Forma
 - Picco in [0,5]
 - Asimmetrica a destra
- Centro
 - Attorno al 5% (cfr. Tabella)
- Dispersione
 - Min 0,7% - Max 42%
 - Outlier 4 in (25,45]



Variabili quantitative: diagrammi ramo-foglia

Stato	Percentuale	Stato	Percentuale	Stato	Percentuale
Alabama	1,5	Louisiana	2,4	Ohio	1,9
Alaska	4,1	Maine	0,7	Oklahoma	5,2
Arizona	25,3	Maryland	4,3	Oregon	8
Arkansas	2,8	Massachussetts	6,8	Pennsylvania	3,2
California	32,4	Michigan	3,3	Rhode Island	8,7
Colorado	17,1	Minnesota	2,9	South Carolina	2,4
Connecticut	9,4	Mississippi	1,3	South Dakota	1,4
Delaware	4,8	Missouri	2,1	Tennessee	2
Florida	16,8	Montana	2	Texas	32
Georgia	5,3	Nebraska	5,5	Utah	9
Hawaii	7,2	Nevada	19,7	Vermont	0,9
Idaho	7,9	New Hampshire	1,7	Virginia	4,7
Illinois	10,7	New Jersey	13,3	Washington	7,2
Indiana	3,5	New Mexico	42,1	West Virginia	0,7
Iowa	2,8	New York	15,1	Wisconsin	3,6
Kansas	7	North Carolina	4,7	Wyoming	6,4
Kentucky	1,5	North Dakota	1,2		

Popolazione di origine ispanica per Stato americano (2000)

Variabili quantitative: diagrammi ramo-foglia

- Utili per piccoli insiemi di dati

1. Osservazioni in diversi rami

Ogni ramo è formato da tutte le cifre di un'osservazione tranne l'ultima che è la foglia

2. Rami in verticale partendo dalle osservazioni più piccole

3. Foglie alla destra dei rami partendo dalle più piccole

0		779	
1		2345579	
2		00144889	
			⋮
3		2356	
4		13778	14
5		235	15 1
6		48	16 8
7		0229	17 1
8		07	18
9		04	19 7
10		7	20
11			21
12			22
13		3	23
			24
			25 3
			⋮

Diagrammi temporali

- Variabili misurate in diversi momenti del **tempo**
- Interesse nel **cambiamento della variabile** rispetto al tempo
- **Diagramma temporale** riporta le diverse osservazioni insieme all'indicazione del periodo in cui sono state misurate
- L'**asse orizzontale** misura il tempo
- L'**asse verticale** misura la variabile che stiamo studiando
- Unire i punti del **grafico con le linee** enfatizza l'andamento nel tempo

Diagrammi temporali

Tasse pagate in università pubbliche e private in USA dal 1971 al 2001

