

Insegnare statistica: idee e strumenti

Ciclo di incontri per i docenti di scuola secondaria di secondo grado
gennaio-febbraio 2015

Introduzione alla Statistica

13 gennaio 2013

Carla Rampichini
Dipartimento di Statistica, Informatica, Applicazioni 'G. Parenti'
Università di Firenze



Nella vita quotidiana spesso
abbiamo bisogno di informazioni
per prendere una decisione ...

... lo facciamo utilizzando la
Statistica anche se non ce ne
rendiamo conto!

Buongiorno, mi chiamo Statistica.
Ho qualche centinaia di anni,
anche se non li dimostro. La
mia passione sono i dati.....e
le unità statistiche.



Statistica e società



- Nella società in cui viviamo, basata sull'informazione e sulla conoscenza, i cittadini dovrebbero avere la capacità di comprendere e utilizzare correttamente la grande quantità di dati e statistiche (ufficiali e non) diffuse quotidianamente dai mezzi di comunicazione.

The time may not be very remote when it will be understood that for a complete initiation as an efficient citizen of one of the new great complex world wide states that are now developing, it is as necessary to be able to compute, to think in averages and maxima and minima, as it is now to be able to read and write.

H.G. Wells (1903, Mankind in the Making, p. 204)

Perché la Statistica nella scuola ...

- La scuola fornisce ai giovani le **competenze** e gli **strumenti** per affrontare la **vita adulta**
- Il mondo in cui viviamo è sempre più complesso
→ servono strumenti per
 - decidere in condizioni di **incertezza**
 - analizzare e comprendere i **dati**

Cos'è la statistica I

- La statistica è la scienza che studia i metodi per la **raccolta e l'analisi dei dati**.
- La statistica è utilizzata per
 - fare **previsioni** (del tempo, economiche, demografiche)
 - **analizzare** eventi del passato (i prezzi salgono? Il tasso di analfabetismo si è ridotto?)
 - per prendere **decisioni** (un certo vaccino è efficace? Conviene investire in un dato settore?)

Se torturi i numeri abbastanza a lungo, confesseranno qualsiasi cosa.
Gregg Easterbrook
citazione da Darrell Huff (2007) *Mentire con le statistiche*, M&A.

Cos'è la Statistica II

- La statistica è una **disciplina** che non esiste per se stessa ma per offrire ad altri campi di studio un insieme coerente di **idee** e **metodi** per trattare i **dati**.
- La necessità di questa disciplina nasce dall'onnipresenza della **variabilità**
 - gli individui variano, misure ripetute sullo stesso individuo variano, ecc
- La statistica fornisce i metodi per trattare i dati tenendo conto della variabilità
 - per trovare individui **particolari** in una gran mole di dati
 - per trovare gli **effetti sistematici** di un fenomeno scartando il rumore di fondo della variabilità

Il ruolo del contesto

- Il focus sulla variabilità rende la statistica automaticamente differente dalla matematica
- Inoltre, la statistica richiede un modo di pensare diverso, perchè i dati sono numeri in un **contesto**
- Questa differenza ha implicazioni sui metodi di insegnamento
 - per insegnare bene la statistica non basta conoscere la matematica e la teoria statistica
 - è indispensabile disporre di **esempi reali** e sapere come utilizzarli per coinvolgere gli studenti e portarli a sviluppare **capacità di giudizio critico**.

Importanza del giudizio critico

- La statistica prende in prestito dalla matematica un'aria di precisione e certezza, ma necessita del giudizio umano ed è quindi soggetta a distorsioni e interpretazioni errate
- Ecco due semplici **esempi**:
 - **Quanto è pericolosa una malattia?** Dipende! Se la malattia ha un tasso di mortalità dello 0.1% questo può non destare troppa preoccupazione, ma se questa malattia è una malattia comune, può causare migliaia di morti ogni anno!
 - **Come si misura lo stipendio medio di una certa categoria?** Dipende! Per esempio, in una stanza ci sono 10 insegnanti i quali percepiscono uno stipendio tra i 1400 e i 2000 euro, con uno stipendio medio di 1750 euro e una mediana di 1775 euro; entra un'altra persona, il cui stipendio ammonta a 10000 euro: la media cresce di 750 euro mentre la mediana resta invariata.

Un mondo di dati



- Gli studenti di oggi hanno un'esperienza diretta con i dati, anche prima di iniziare lo studio della statistica, e di solito non sono i dati che abbiamo in mente noi!
 - L'ipod tratta le canzoni come dati, le canzoni possono essere "mischiate", sovvertendo l'ordine originale dell'album, inoltre è possibile ottenere statistiche relative alla musica ascoltata (p.e. frequenza di ascolto dei brani)
 - Internet offre un sacco di opportunità di socializzazione (facebook, twitter, ecc)
 - La maggior parte degli studenti crea e porta con sé le proprie banche dati, caricate su supporti piccoli e facili da maneggiare, e capaci di fornire alcune statistiche descrittive (si pensi ai cellulari che contengono indirizzi, foto, musica, ecc)



Un mondo pieno di incertezza ...

- La capacità di prendere decisioni in situazioni di **incertezza** è indispensabile al cittadino consapevole che vuole capire e controllare i fenomeni (naturali, sociali, economici e politici) che lo circondano.
- Alcuni esempi:
 - efficacia di un vaccino
 - gestione del mutuo per l'acquisto di una casa
 - andamento dei prezzi
 - analisi della soddisfazione verso un servizio pubblico
 - giochi d'azzardo (lotterie, gratta e vinci, ecc..)
- **Statistica e Calcolo delle probabilità** sono gli strumenti ideali per prendere decisioni in situazioni di incertezza

Prendere decisioni in situazioni di incertezza

- In molte situazioni della vita quotidiana, è possibile prendere la decisione migliore sul da farsi considerando il VALORE ATTESO

- la Lotteria del Wisconsin ha un gioco tipo "Gratta e vinci", detto "Big Cat Cash" il cui biglietto costa 1\$. Qui a lato leggiamo le probabilità di vincita

winnings, x	Probability, p
\$1	1/10
\$2	1/14
\$3	1/24
\$18	1/200
\$50	1/389
\$150	1/20,000
\$900	1/120,000

Probabilities for Wisconsin scratch-off game.
[Source: www.wilottery.com.]

- Conviene comprare il biglietto e giocare a questo gioco?

Come si fa a sapere se si vince?

- Notiamo prima di tutto che le probabilità non sommano a 1: questo perchè l'evento più probabile è non vincere niente!
- Calcoliamo la probabilità di non vincere niente:

$$P(\text{vincita}=0\$) = 1 - 0.2207 = 0.7793$$



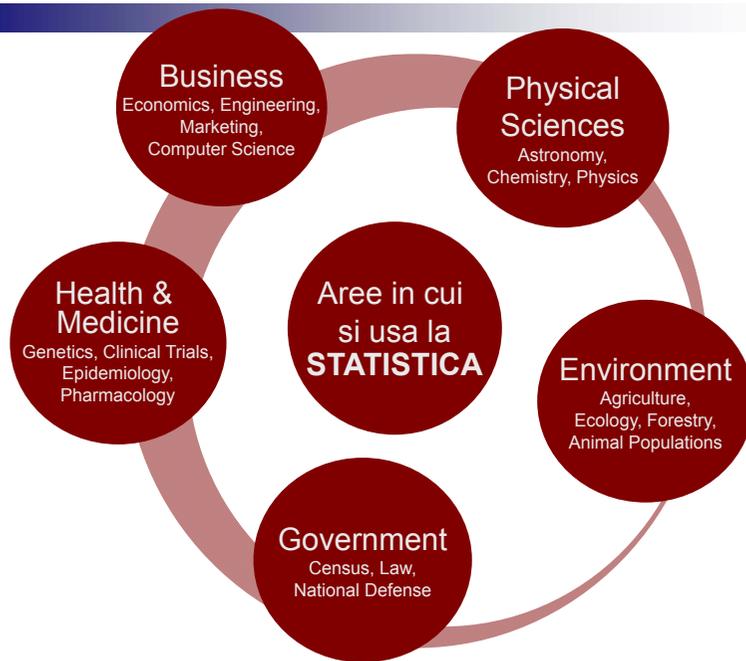
si vince qualcosa circa 2 volte su 10

vincita \$	p(x)
1	0.1000
2	0.0714
3	0.0417
18	0.0050
50	0.0026
150	0.0001
900	0.0000

TOT 0.2207

E a lungo andare, *quanto possiamo sperare di vincere?*

- Calcoliamo la vincita media e otteniamo **0.6014** → se spendiamo 1\$ per giocare, dobbiamo aspettarci di vincere, in media, 60.14¢
- Naturalmente non otterremo indietro 60.14¢ in una giocata, ma questa è la vincita media per molte giocate! Se compriamo **100 biglietti** ci aspettiamo quindi di vincere **60.14\$**, ma ne abbiamo spesi 100\$.
- Il valore medio può non essere molto importante per il singolo giocatore, a meno che non giochi molte volte, ma è di GRANDE importanza per lo Stato del Wisconsin, che paga **60.14\$** per ogni **100\$** scommessi!



Statistica e matematica

- La statistica è una scienza quantitativa, ma il modo di pensare 'statistico' è diverso da quello 'matematico' per almeno 2 aspetti
 - la statistica non può prescindere dal **contesto** (dati)
 - la logica dell'inferenza statistica non è basata sulla deduzione (come la matematica) ma sull'**induzione**: dal particolare (ciò che si è osservato) al generale

- Per la statistica la matematica ha un ruolo strumentale, cioè consente di costruire gli strumenti che permettono l'analisi statistica (la matematica sta alla statistica come il martello sta al fabbro)

Statistica e matematica

- A livello introduttivo la statistica non necessita di conoscenze di matematica avanzate
- Infatti, gli argomenti trattati dipendono principalmente dal contesto:
 - raccolta dei dati
 - esplorazione dei dati
 - e interpretazione dei risultati
- In tal senso si può insegnare statistica a partire dalla scuola elementare
- L'insegnamento della statistica diventa via via più *matematico* al crescere del livello di approfondimento della disciplina

IL RAGIONAMENTO STATISTICO

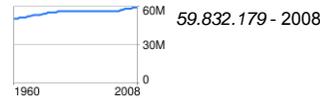


Il ragionamento statistico

- La statistica è un processo orientato alla soluzione di problemi che cerca di rispondere alle domande di interesse attraverso i **dati**.

–La popolazione sta crescendo o diminuendo?
59.832.179 - 2008

Fonte: Banca Mondiale, Indicatori di sviluppo mondiale



- Qual è il modo più sicuro per investire i propri risparmi?
- Se si mangia più frutta e verdura si ha una speranza di vita più lunga?



Il ragionamento statistico

- Obiettivo principale dell'insegnamento della statistica



aiutare gli studenti a sviluppare il
ragionamento statistico

- Diversamente dai teoremi della matematica, i metodi statistici sono efficaci solo se utilizzati con spirito critico e abilità.

Imparare a leggere i dati

- Quando si leggono dei dati bisogna **sempre** ricordare che questi derivano da un processo

- Formulare il problema di interesse
- Raccogliere i dati
- Analizzare i dati
- Interpretare i risultati



- Tutte le fasi del processo devono essere note per poter valutare i risultati

Contestualizzare l'informazione

- Supponiamo che un giornale riporti la notizia che il numero di atti di violenza in una certa città è aumentato negli ultimi 10 anni, passando da 437 casi registrati nel 1999 a 541 nel 2009.

Questi dati sono sufficienti per trarre una conclusione ragionevole in merito al diverso livello di violenza in città?

NO, non è possibile trarre alcuna conclusione in base a questa sola informazione, perchè non si sa se ci sono state modifiche nella popolazione durante il periodo considerato

Cosa dobbiamo fare?

- Dobbiamo conoscere la dimensione della popolazione da cui questi dati derivano, per poter calcolare il tasso di criminalità nei due anni considerati
 - ma il tasso di criminalità da solo, non basta!*
- Per comprendere la rilevanza di un aumento o di una diminuzione è necessario conoscere sia la variazione assoluta che quella relativa
- Se si riporta solo il tasso non si ha idea di quale sia la dimensione del fenomeno
- se la % di crimini violenti aumenta, questo può significare che:
 - a) c'è stato un incremento effettivo di questo tipo di criminalità,
 - b) o il n. di crimini violenti è rimasto invariato, ma si sono ridotti gli altri tipi di crimine
 - Chiaramente (a) e (b) sono due situazioni socialmente molto diverse!

Imparare a leggere i dati: proporzioni e conteggi

- È fondamentale che gli studenti siano in grado di distinguere tra proporzioni e conteggi.
- Come nell'esempio che abbiamo appena visto
 - supponiamo che un giornale riporti la notizia che il numero di atti di violenza in una certa città è aumentato negli ultimi 10 anni, passando da 437 casi registrati nel 1999 a 541 nel 2009.
 - Chiedete ai vostri studenti se questi dati sono sufficienti per trarre una conclusione ragionevole in merito al diverso livello di violenza in città
 - Non è possibile trarre alcuna conclusione in base a questa sola informazione, perchè non si sa se ci sono state modifiche nella popolazione durante il periodo considerato
 - **Viceversa, il confronto tra tasso di criminalità nel 1999 e nel 2009 sarebbe molto informativo!**

Variazioni assolute e relative

- Per comprendere la rilevanza di un aumento o di una diminuzione è necessario conoscere sia la variazione assoluta che quella relativa
- Se si riporta solo una proporzione non si ha idea di quale sia la dimensione del fenomeno
 - p.e., se la % di crimini violenti aumenta, questo può significare che: (a) c'è stato un incremento effettivo di questo tipo di criminalità, o (b) il n. di crimini violenti è rimasto invariato, ma si sono ridotti gli altri tipi di crimine
 - Chiaramente (a) e (b) sono due situazioni socialmente molto diverse!

Prima di valutare, contestualizzare ...

- Prima di reagire davanti ad un valore % è necessario porlo nel suo "contesto".
- Esempio
 - il 99% delle morti nel primo mese di vita avviene nei paesi in via di sviluppo. Questa % sembra terribile!
 - È necessario considerare però che circa il 90% di tutte le nascite avviene nei paesi in via di sviluppo.
 - Pertanto la probabilità che un bambino muoia nel suo primo mese di vita è di **11 volte più alta** nei paesi in via di sviluppo rispetto al resto del mondo.
Christina Pagel (Making sense of statistics, 2010)

Valutare il rischio ...

- Per capire se la variazione di un certo rischio è significativa per un individuo, è necessario sapere qual è il valore del rischio da cui si parte.
 - È stato visto che il rischio di tumore al colon cresce del **20%** se si mangiano carni rosse o salumi.

*Questo **incremento relativo** può risultare allarmante, ma cosa significa?*
 - Dipende da qual è il valore del **rischio** da cui si parte: il rischio di sviluppare il tumore al colon durante la propria vita è circa il **5%** – questo è il **rischio 'assoluto'**.
 - Se si mangiano carni rosse e/o salumi con una certa frequenza, tale rischio cresce del **20%** – che corrisponde all'**incremento relativo** del rischio.
 - il rischio passa dal 5 al 6%, corrispondente ad un **incremento** del rischio del **1%** (pari al 20% del 5%).

Importanza del giudizio critico

- La statistica prende in prestito dalla matematica un'aria di precisione e certezza, ma necessita del giudizio umano ed è quindi soggetta a distorsioni e interpretazioni errate
- Ecco due semplici **esempi**:
 - *Quanto è pericolosa una malattia?* Dipende! Se la malattia ha un tasso di mortalità dello 0.1% questo può non destare troppa preoccupazione, ma se questa malattia è una malattia comune, può causare migliaia di morti ogni anno!
 - *Come si misura lo stipendio medio di una certa categoria?* Dipende! Per esempio, in una stanza ci sono 10 insegnanti i quali percepiscono uno stipendio tra i 1400 e i 2000 euro, con uno stipendio medio di 1750 euro e una mediana di 1775 euro; entra un'altra persona, il cui stipendio ammonta a 10000 euro: la media cresce di 750 euro mentre la mediana resta invariata.

È difficile ragionare 'statisticamente' Un esempio medico

- Una donna di 40 anni fa una mammografia di routine e risulta positiva
- “Che cosa significa?” – chiede al radiologo – “Ho il cancro? Con che probabilità? Il 99%, il 95, il 90? Soltanto il 50%? Che cosa sappiamo?”
- Un terzo dei radiologi della Harvard Medical School ha risposto il 90%
- Ma:
 - La probabilità che una donna di quell'età abbia il cancro al seno è intorno all'1%
 - Se ha il cancro, la probabilità di risultare positiva alla mammografia è del 90%
 - Se non ha il cancro, la probabilità di risultare comunque positiva è del 9%

Un modo diverso di comunicare l'informazione

- Pensate a 100 donne
 - Tutte fanno la mammografia
 - Una ha il cancro (questo è l'1%)
 - Fa la mammografia e risulta positiva
 - Delle altre 99 che fanno la mammografia, 9 risultano positive ma non hanno il cancro
 - Quindi, 10 risultano positive al test
 - Di queste, quante hanno realmente il cancro?
- Non il 90%, né il 50, ma il 10%!
- Vi immaginate come è stata quella signora?

Statistica

DATI E VARIABILITÀ

Statistica, dati, variabilità

- La statistica è una scienza che ragiona partendo dai **dati**
- Il principio fondamentale della Statistica è la **variabilità** dei dati
- Se il mondo fosse perfettamente prevedibile e non ci fosse variabilità, non ci sarebbe bisogno della Statistica



Da cosa deriva l'incertezza?

- L'incertezza è generata da meccanismi molto diversi ...
- ma fondamentalmente dipende dalla **VARIABILITÀ**
 - se lancio un dado o una moneta, o se estraggo un numero al lotto non so quale sarà il risultato
 - se misuro più volte un oggetto le misure non sono tutte uguali (variabilità delle misure, o errore di misura, p.e. due misurazioni della propria statura)
 - se misuro l'altezza di due studenti ottengo valori diversi (variabilità 'biologica')
 - Non so se domani pioverà, ecc

Il ruolo della variabilità

- La soluzione dei problemi e il prendere decisioni con metodo statistico dipendono dalla comprensione, spiegazione e quantificazione della **variabilità** nei dati.
- La considerazione della variabilità distingue la statistica dalla matematica
- Ci sono più fonti di variabilità nei dati
 - *variabilità nelle misure*
 - *variabilità naturale*
 - *variabilità indotta*
 - *variabilità campionaria*

Una buona introduzione sulla natura della variabilità si trova in *Utts (1999) Journal of Scientific Exploration, Vol. 13, No. 4, pp.615–638.*

Variabilità nelle misure

- Misure ripetute sullo stesso soggetto variano
 - inaffidabilità dello **strumento** di misura (p.e. misurare una distanza con un righello)
 - variazioni nel **sistema** che si sta misurando (p.e. anche utilizzando uno strumento di misura preciso, la pressione sanguigna di uno stesso soggetto può variare da un momento all'altro)

Variabilità naturale

- *La variabilità è insita nei fenomeni naturali*
- Se misuriamo la stessa caratteristica su individui diversi troviamo valori differenti (statura, abilità, opinioni, ecc.)
- in parte queste differenze possono dipendere dallo strumento di misura, ma fondamentalmente dipendono da **differenze tra individui**

p.e. semi diversi della stessa varietà di fagioli cresceranno in modo diverso anche piantati nello stesso terreno, perchè nessun seme è identico ad un altro

Variabilità indotta (sperimentale)

- Se si piantano semi di fagiolo in un campo e altri semi in un altro campo sotto condizioni diverse, allora la differenza osservata nella crescita può essere attribuita a:
 - differenze inerenti ai semi (variabilità naturale)
 - differenze nelle condizioni di crescita. p.e. se il fertilizzante utilizzato nei due campi è diverso.
 - le differenze osservate possono essere dovute a un fattore cui non si è pensato → disegno sperimentale fatto con cura aiuta a determinare l'effetto dei fattori di interesse

Variabilità naturale vs variabilità indotta

- Il confronto tra variabilità naturale e variabilità indotta da altri fattori è il cuore della statistica moderna.
- Questa logica per esempio:
 - ha consentito alla medicina di concludere che alcune medicine sono efficaci e sicure, mentre altre non sono efficaci o hanno effetti collaterali indesiderati
 - è stato utilizzato dagli agronomi per dimostrare che una varietà di grano cresce meglio in un certo clima che in un altro, che un fertilizzante è più efficace di un altro ecc.

Variabilità campionaria

- In un sondaggio elettorale, pare ragionevole utilizzare la proporzione di votanti osservata nel campione (una **statistica campionaria**) come stima della proporzione incognita di votanti a favore di un certo partito.
- Se si estrae un secondo campione di uguale dimensione dalla stessa popolazione quasi certamente NON si osserverà la STESSA proporzione di votanti favorevoli a quel partito
- Il valore della proporzione campionaria varierà da un campione all'altro → **variabilità campionaria**

Il ruolo della variabilità

UN ESEMPIO: MISURE RIPETUTE

Misure ripetute di uno stesso oggetto

- A 30 studenti sono state distribuite altrettante palline da tennis, prodotte da una stessa macchina.
- Agli studenti è stato chiesto di trovare un **metodo** per misurare con un righello o della carta millimetrata il **diametro** della pallina
- Ciascuno studente ha **misurato due** volte il diametro della pallina al millimetro più vicino, e espresso la misura in **millimetri**
- Le misure sono state inserite in un **foglio Excel**
- Osserviamo che non tutte le misure sono uguali: molte sono vicine al diametro programmato dalla ditta che ha prodotto le palline (**valore vero=50**).

Perchè le misure variano?

Variabilità

- Osservate che le misure sono diverse tra loro, anche quelle fatte da uno stesso studente!
- ↓
- La **misurazione** ripetuta del '*diametro della pallina*' può dare luogo a **risultati diversi**

Questo fatto illustra uno dei concetti fondamentali della statistica:
la **variabilità**

Perchè le misure variano?

- Il **metodo di misura** utilizzato può essere diverso
 - qualche studente avrà bloccato la pallina tra due libri e misurato la distanza tra i libri
 - altri avranno avvolto il nastro millimetrato intorno alla pallina, misurato la lunghezza del nastro e quindi diviso per π
- Ogni studente ha misurato la **sua** pallina e le **palline non** sono perfettamente **uguali**

Tuttavia in questo esempio la maggior parte delle differenze sono dovute a **errori di misura**.
È difficile misurare il diametro di una sfera!

Diagramma a barre delle misure

- Contiamo quante volte si presenta ciascun valore → **frequenze**
- Costruiamo il grafico
 - in ascissa scriviamo i valori in millimetri
 - in ordinata riportiamo il valore delle frequenze
- Che forma ha il grafico?
- Come mai ha questa forma?

Il grafico mostra una 'gobba' in corrispondenza del valore **vero**, questo è tipico delle distribuzioni degli errori di misura

Discussione

- Osservando il grafico, scegliete 2 numeri che vi sembrano ragionevoli per completare questa frase:
"La misura tipica del diametro è di circa ____ mm , con uno scarto in più o in meno di circa ____ mm"
- Come potremmo ridurre la variabilità nelle misure?
 - Utilizzando un'unica pallina
 - Utilizzando tutti lo stesso metodo di misura
 - Insegnando a tutti come si misura il diametro
- Potremmo eliminare completamente la variabilità delle misure?

no, non è possibile!

Esempio: se il diametro vero è 49.5 mm. quando misuriamo il diametro con il righello al mm più prossimo, qualcuno riporterà 50 e altri 49 mm, a seconda che pensino che la misura sia un po' sopra o un po' sotto 49.5 mm

IL PROCESSO DI ANALISI DEI DATI

Il processo di analisi dei dati

- La Statistica è un processo orientato alla **soluzione dei problemi** che cerca una risposta alle domande attraverso i **dati**
- Questo processo ha tipicamente 4 componenti:
 - Formulare il problema di interesse
 - Raccogliere i dati
 - Analizzare i dati
 - Interpretare i risultati

Il processo 'statistico'

1. Formulare il problema (quesito)
 2. Raccogliere i dati appropriati
 3. Analizzare i dati
 4. Interpretare i risultati
- Molto spesso si focalizza l'attenzione esclusivamente sul terzo punto (analisi dei dati)
 - Ma per comprendere i dati non basta analizzarli, è necessario considerare anche la natura dei dati e interpretare correttamente i risultati

Quesiti statistici

- Formulare il problema
 - chiarire il problema di interesse
 - formulate una (o più) domande cui si può dare una risposta attraverso i dati
- Non tutti i quesiti sono di tipo statistico!
- La domanda, "*Quanto sono alto?*" non è di tipo statistico perchè ha una risposta deterministica: la mia statura
- La domanda "*Quanto sono alti gli uomini italiani?*" è di tipo statistico perchè ogni italiano ha una statura diversa.
- Un quesito statistico richiede un'anticipazione della variabilità

Raccolta e analisi dei dati

- Raccolta dei dati
 - predisporre disegno e piano di campionamento
 - utilizzare il piano di campionamento per la raccolta dei dati
- Analisi dei dati
 - selezionare uno o più metodi appropriati
 - utilizzo dei metodi per l'analisi
- Obiettivo principale dell'analisi statistica è descrivere la variabilità presente nei dati

Esempio

1. **Quesito:** qual è la proporzione di mancini nella popolazione italiana?
2. **Raccolta dei dati:** campione rappresentativo della popolazione in età da 6 a 80 anni
3. **Analisi dei dati:** calcolo della proporzione campionaria
4. **Risultati:** il 6% degli italiani è mancino, con un margine di errore del +/- 1% al livello di confidenza del 95%

... segue esempio

- Il campione fornisce una **stima** della proporzione di mancini, pari al **6%**.
- Il **margine di errore (+/- 1%)** riflette la variabilità campionaria: indica quanto il risultato campionario può discostarsi dall'effettiva proporzione di mancini nella popolazione
- Il **livello di confidenza (95%)** indica con che frequenza le stime prodotte dal metodo utilizzato producono un risultato corretto.
- Questa analisi si basa sulla distribuzione delle stime nell'universo dei campioni.

Statistica descrittiva vs inferenziale

Statistica Descrittiva

Metodi di

- raccolta
- presentazione (grafici)
- caratterizzazione (statistiche)

di un insieme di dati allo scopo di descriverne le caratteristiche

Statistica Inferenziale

Metodi di

- stima

di una particolare caratteristica relativa alla popolazione di interesse, sulla base dell'osservazione di un campione, allo scopo di generalizzare il risultato all'intera collettività

Esempio di inferenza statistica

Qual è la proporzione di persone che scrivono con la mano sinistra?

N = numero di persone; M = n. di "mancini"

Quanto vale $p = M/N$?

Campione di 100 persone, di cui 6 sono mancini

$$\hat{p} = 6/100 = 0.06$$

$p \neq \hat{p}$ per errore di campionamento

Inferenza statistica → quantificazione dell'errore

Es. si arriva ad affermare che l'intervallo [0.05, 0.007] contiene p con elevata probabilità

Materiali per attività in classe

- Preparazione personale
 - David S. Moore (2013). Statistica di base 2a edizione. Apogeo.
 - Gaise Reports: <http://www.amstat.org/education/gaise/>
- Per predisporre attività e lezioni
 - Watkins AE, Scheaffer RL, Cobb GW (2008) *Statistics in Action. Understanding a World of Data*. Key Curriculum Press
 - Sitografia e bibliografia indicati
 - Bergamini M., Trifone AM (2009). *Modulo alfa+beta verde - fondamenti probabilità e statistica*, Zanichelli.

Insegnamento della statistica

- L'apprendimento della statistica deve procedere per gradi, affrontando via via con maggiore consapevolezza i 4 punti fondamentali del processo statistico.
- Nel GAISE report e su STatistics Education Web (STEW: <http://www.amstat.org/education/stew/>) si trovano esempi che seguono il processo statistico
- Vediamo un esempio ...

Il primo approccio ...

- Studenti che affrontano per la prima volta lo studio della statistica devono sviluppare il senso dei dati ...
 - ... i dati non sono solo numeri
 - *la statistica trasforma i numeri in informazione*
- Gli studenti dovrebbero
 - riconoscere che i dati sono raccolti, o generati, con riferimento a un **contesto** o a una situazione particolare e che possono essere utilizzati per rispondere a domande che riguardano QUEL contesto o quella situazione
 - saper utilizzare i metodi di base dell'analisi dei dati (statistica descrittiva) e dell'inferenza (almeno informalmente)
 - sviluppare le idee di base della probabilità

Un esempio: la musica preferita

Qual è il tipo di musica preferita dagli studenti della classe?

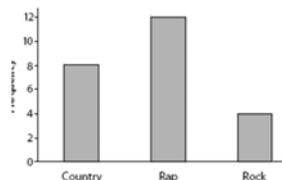
- La caratteristica da misurare, la musica preferita, è una **variabile qualitativa**: ogni studente sarà classificato in una categoria in base alla musica da lui preferita.
- I dati possono essere raccolti in classe (censimento della classe), per esempio una classe di 24 studenti
- I dati possono essere riassunti in una **tabella di frequenza**

Preferenza	Frequenza (conteggio)
Country	8
Rap	12
Rock	4

Rappresentazione e sintesi delle distribuzioni di frequenza

- Per rappresentare la distribuzione di frequenza si può disegnare un diagramma a barre.

Il valore “rappresentativo” o “tipico” di questa distribuzione è la moda, ossia la categoria con il maggior numero di frequenze.



- La lettura e l'interpretazione dei dati precedono sempre l'inferenza.
- È importante cercare di rispondere alla domanda
Perchè i dati si presentano nel modo osservato?

Generalizzazione dei risultati

- È utile invitare gli studenti a pensare se i risultati trovati possono essere “generalizzati” a un gruppo più esteso, per esempio tutti gli studenti della scuola, o tutti i giovani della stessa loro età.
- È importante che gli studenti riflettano su quali caratteristiche individuali possono influenzare i gusti musicali di individui appartenenti al gruppo più esteso cui si vuole generalizzare (p.e. età o luogo di residenza)
 - Gli insegnanti della scuola hanno preferenze musicali simili alle loro?
 - Quali potrebbero essere le preferenze musicali degli studenti di un anno più avanti rispetto a loro?
- Questi ragionamenti dovrebbero aiutare gli studenti a capire che l'inferenza è limitata a una specifica classe.

STATISTICA DESCRITTIVA

Concetti di base
Esempi



Statistica descrittiva

scopo principale

sintetizzare e descrivere i dati attraverso

- grafici (diagrammi a barre, a torta, boxplot, ecc.)
- tabelle
- indici sugli aspetti più importanti
 - posizione, come la media
 - variabilità, come la varianza
 - concentrazione
 - relazione tra variabili

UN ESEMPIO

Conosciamo la classe

Conosciamo la classe (popolazione)

- Non esistono due studenti uguali, gli studenti (unità di osservazione) **variano** da uno all'altro in molti modi interessanti
- Un buon metodo per iniziare a pensare alla variabilità consiste nel raccogliere DATI sugli studenti di questa classe: **come è ciascuno di voi rispetto agli altri?**
- Rispondete al questionario in modo anonimo e saltate le domande cui non volete o non potete rispondere

ATTENZIONE AI PROBLEMI DI MISURA!

Problemi di misura ...

- Come si misura la frequenza del battito cardiaco?

Attenzione

dobbiamo utilizzare tutti lo stesso **STRUMENTO DI MISURA** perché i dati siano confrontabili

- p.e. scegliamo di **contare** i battiti **per 10 secondi** e poi moltiplichiamo **per 6**

Tipi di variabili

- Riprendiamo in esame le domande contenute nel questionario e osserviamo che raccolgono informazioni di vario tipo: quantitative e qualitative
- Le variabili **QUANTITATIVE** misurano caratteristiche numeriche: p.e. il **peso** e l'**altezza** di una persona
- Le variabili **QUALITATIVE** misurano delle qualità: p.e. il **colore degli occhi** e dei capelli
- Le variabili **dicotomiche** sono variabili qualitative con due sole modalità: p.e. la variabile **genere** assume le modalità maschio e femmina

Codici

- Osserviamo che accanto alle modalità delle variabili qualitative compaiono dei **numeri**
- Questi numeri NON sono quantità ma sono dei **CODICI** che facilitano il lavoro di registrazione dei dati

Consideriamo gli studenti presenti come unità di osservazione, e classifichiamo le seguenti variabili

- Tu sei maschio o femmina?
Qualitativa, dicotomica: maschio, femmina
- Quanto tempo hai dedicato allo studio ieri pomeriggio?
Quantitativa: ore, minuti

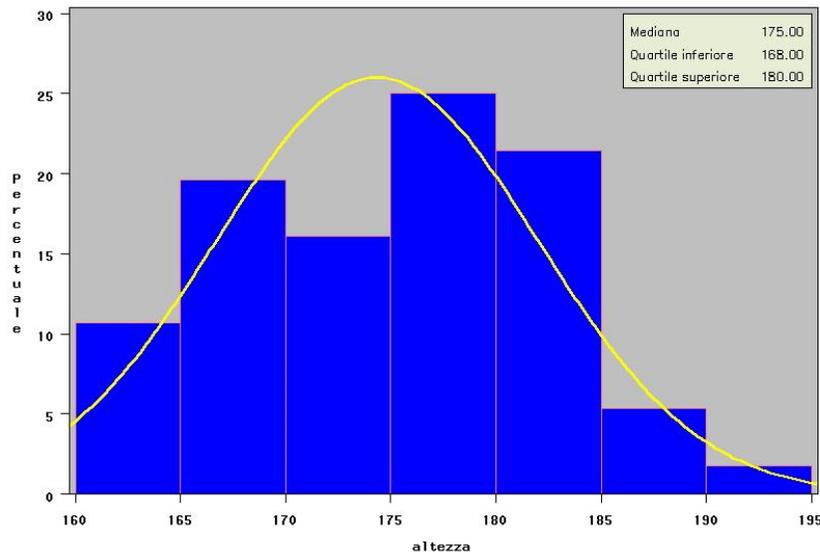
Matrice dei dati

- I dati sono di solito raccolti in forma **RETTANGOLARE**: matrice righe \times colonne
 - ogni **riga** della matrice corrisponde ad una **unità di osservazione**
 - ogni **colonna** della matrice corrisponde ad una **variabile**
 - Esempio dati raccolti in una **classe**: **56** oss \times **15** var
- I dati sono un lungo elenco di valori ed è difficile trovare una regolarità!

Distribuzione e sintesi dei dati

- Se volete confrontare la vostra **altezza** con quella di questi ragazzi come fate?
- La lista di valori è lunga e ci vorrebbe molto tempo per confrontare tutti i numeri!
- Possiamo fare un grafico (p.e. istogramma)
- e fare una **sintesi** dei dati raccolti:
 - Quartili: il 50% centrale dei valori è compreso tra 168 e 180 cm
 - Mediana: la metà delle altezze è superiore a 175 cm e l'altra metà è inferiore a questo valore

Istogramma altezze (cm)



Distribuzione e sintesi dei dati

- Se siete alti p.e. 178 cm, in base a questa sintesi sapete subito che la vostra altezza si trova nella parte centrale della distribuzione, non lontana dal valore centrale
- Notate che questa sintesi vi fornisce due informazioni: il valore **centrale** è 175 cm e le altezze si **distribuiscono** intorno a questo valore, variando tra 170 e 180 cm nella parte centrale (50% delle altezze) della distribuzione
- Spesso questo tipo di sintesi fornisce tutte le informazioni necessarie per capire l'andamento del fenomeno, soprattutto quando la **forma** della distribuzione è una di quelle tipiche

Distribuzione dei dati

L'insieme dei valori assunti da una variabile e la frequenza con cui ogni valore si presenta tra le unità di osservazione produce la **DISTRIBUZIONE** dei dati

Tabella modalità e frequenze
Mezzo di trasporto utilizzato

mezzo	cod	freq	%
a piedi	1	4	7.14
bici	2	1	1.79
motorino	3	23	41.07
auto	4	16	28.57
bus	5	6	10.71
bus+altro	6	6	10.71
Totale		56	100.00

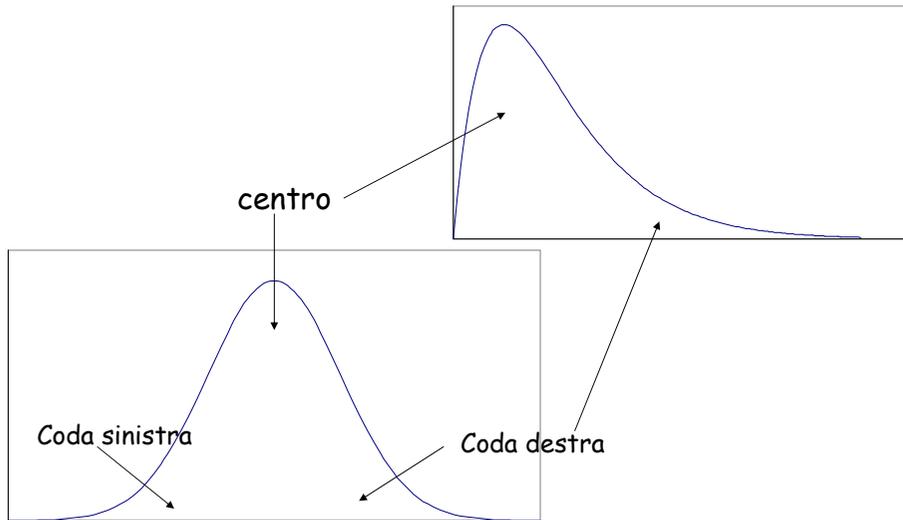
Stem Leaf
 192 0 altezza
 190
 188 00
 186 Steam-and-leaf-plot
 184 00
 182 000
 180 00000000
 178 00000000
 176 0000
 174 000
 172 0000
 170 00000
 168 00000000
 166 0
 164 00
 162 000
 160 000
 -----+-----+-----+

Come si esplorano i dati?

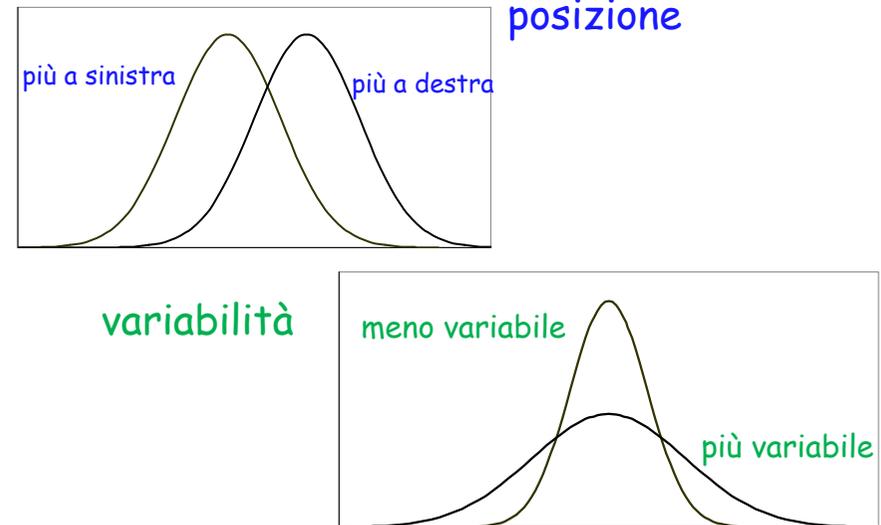
- Ogni analisi esplorativa dovrebbe seguire questi passi
grafico → **forma** → **centro** → **dispersione**
1. Tracciare il **grafico** più appropriato;
 2. Descrivere la **forma** della distribuzione in base al grafico e agli indici di forma;
 3. Calcolare una misura del **centro** della distribuzione, appropriata in base alla forma della distribuzione;
 4. Calcolare un indice di **dispersione** appropriato in base alla forma della distribuzione e coerente con la misura di centro utilizzata;

UTILIZZARE GRAFICI E INDICI APPROPRIATI IN BASE AL TIPO DI VARIABILE

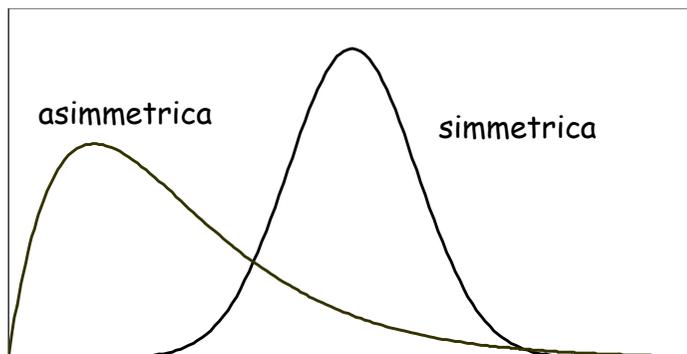
Aspetti notevoli delle distribuzioni



Aspetti caratterizzanti le distribuzioni: posizione e variabilità



Aspetti caratterizzanti le distribuzioni: forma



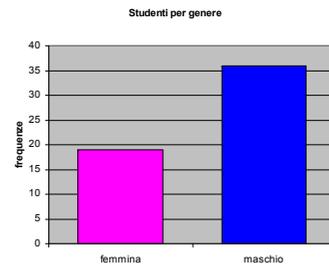
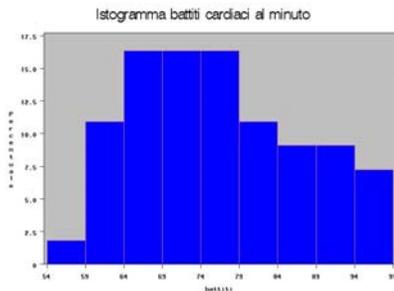
Grafici per variabili quantitative

- Per capire come sintetizzare una distribuzione è utile conoscere la sua **forma**
- La forma di una distribuzione può essere vista attraverso un **grafico**
- Grafici più utilizzati (variabili quantitative)
 - Dotplot
 - Istogramma
 - Boxplot
 - Steam and leaf plot
 - Plot frequenze cumulate

Esempio dati raccolti in [classe](#)

Distribuzioni congiunte

- Osserviamo i grafici del battito cardiaco e del genere: possiamo dall'esame di questi grafici dire che le ragazze hanno in generale il battito più rallentato dei ragazzi?
- No, perchè non abbiamo associato i dati sul battito al genere e quindi non siamo in grado di distinguere quali battiti sono dei maschi e quali delle femmine.

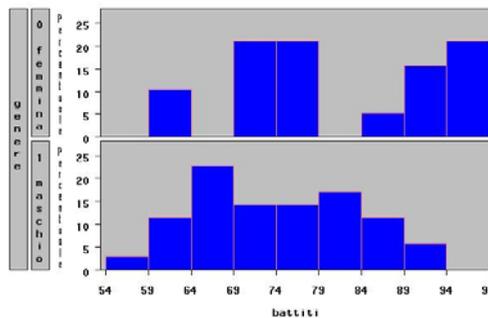


Grafici per distribuzioni congiunte

- La tabella contiene la distribuzione congiunta del battito per genere: che grafico fareste per vedere se ci sono differenze tra maschi e femmine?

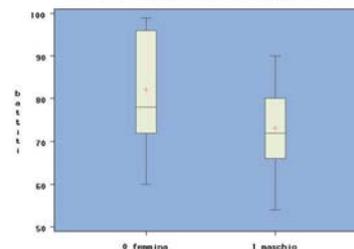
battiti(battiti)	genere(genere)		
	0 femmin	1 maschi	Totale
Frequenza			
54-59	0	2	2
60-64	2	3	5
65-69	0	9	9
70-74	4	4	8
75-79	4	5	9
80-84	0	8	8
85-89	1	2	3
90-94	3	2	5
95-99	5	0	5
Totale	19	35	54
Frequenza dei valori mancanti = 2			

Istogramma battiti cardiaci al minuto



Grafici del battito cardiaco per genere

Box Plot battiti al min. per genere



Sintesi della distribuzione attraverso un valore rappresentativo

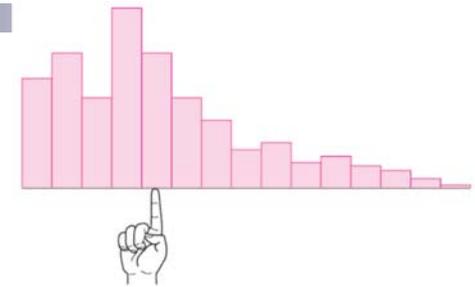
Indici di posizione o medie
attenzione... di media NON ce n'è una sola!

- Distribuzioni sconnesse: moda
- Distribuzioni ordinate: mediana
- Seriazioni: media aritmetica
- Altre medie: media armonica, media geometrica

Media e mediana

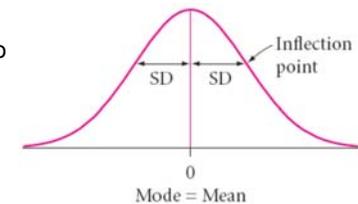
- Sono entrambi indici di posizione
 - indicano il **centro** della distribuzione
- La mediana divide la distribuzione in due parti uguali
- La media è il punto di equilibrio dell'istogramma, come una bilancia, si ottiene sommando i valori e dividendo per il numero di valori
- Negli istogrammi visti è più grande la media o la mediana?

La media è il punto di equilibrio di una distribuzione



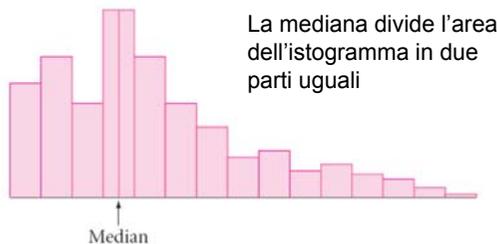
Per trovare la media osservando un istogramma, trovate il punto in cui dovrete mettere un dito sotto l'asse orizzontale per tenere in equilibrio la distribuzione immaginando che i rettangoli abbiano un peso proporzionale alla loro area.

Se una distribuzione è approssimativamente normale, il suo punto di equilibrio corrisponde all'asse di simmetria, quindi la media si trova sull'asse orizzontale direttamente sotto il punto di massimo.



Mediana

La **mediana** è il valore che divide i dati in due metà

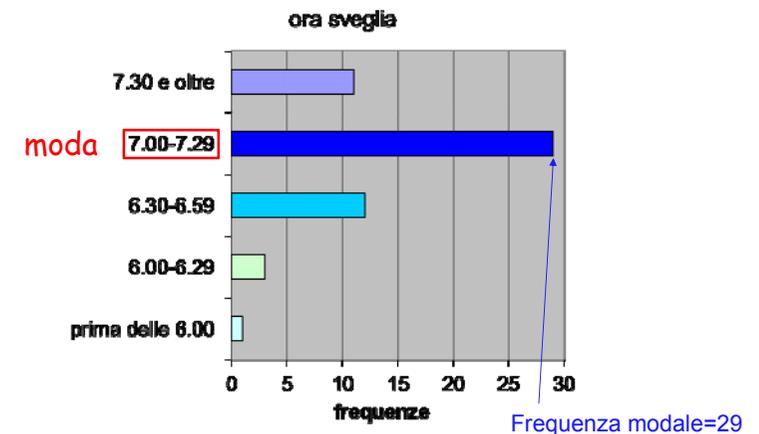


Per trovare la mediana:

- ordinare i valori dal più piccolo al più grande
- e prendere quello che sta nel mezzo (N dispari) o la media dei due che stanno nel mezzo (N pari)

La moda

modalità cui corrisponde la frequenza più alta.



Un esempio

Un'agenzia che effettua indagini di mercato ha rilevato per una rete televisiva i seguenti dati medi giornalieri di ascolto, nel periodo invernale e nella fascia oraria dalle 20 alle 21.

Giorno della settimana	Lun	Mar	Mer	Gio	Ven	Sab	Dom
N medio spettatori (migliaia)	1200	1800	2000	1600	1200	800	900

In quale giorno l'agenzia potrebbe consigliare a un proprio cliente di inserire uno spot pubblicitario di un prodotto per la neve, volendo usare la fascia oraria 20 – 21?



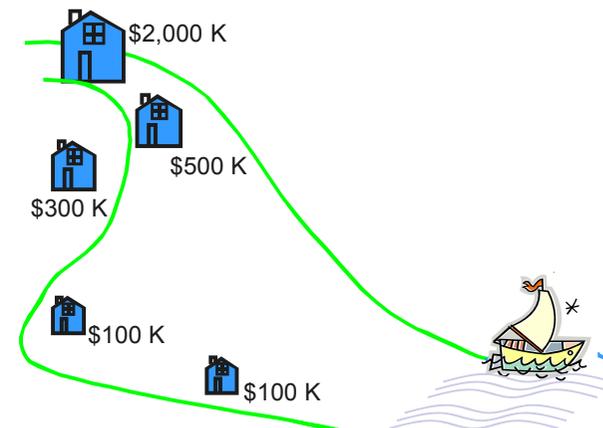
la moda è **il mercoledì**, il giorno della settimana con la massima frequenza di ascolto.

Moda, media o mediana?

- Cinque case su una collina presso una spiaggia

Prezzi delle case:

\$2,000,000
500,000
300,000
100,000
100,000



Moda, media e mediana danno informazioni diverse!

Prezzi delle case:

\$2,000,000
500,000
300,000
100,000
100,000

Somma 3,000,000

- **Moda:** valore più frequente
= **\$100,000**
- **Mediana:** valore centrale dati ordinati
= **\$300,000**
- **Media:** $(\$3,000,000/5)$
= **\$600,000**

Quando è utile calcolare la media aritmetica?

- Uno studente ha riportato i seguenti voti nella pagella del primo quadrimestre
- Il padre gli ha promesso un regalo se la media dei suoi voti è superiore al 7.
- Lo studente otterrà il regalo?

materia	Voto
italiano	8
storia	5
geografia	6
inglese	7
scienze	5
matematica	8
educ. Fisica	8
TOTALE	47

Calcoliamo la media aritmetica dei voti
 $46/7=6.57$ Sembra di no però ...

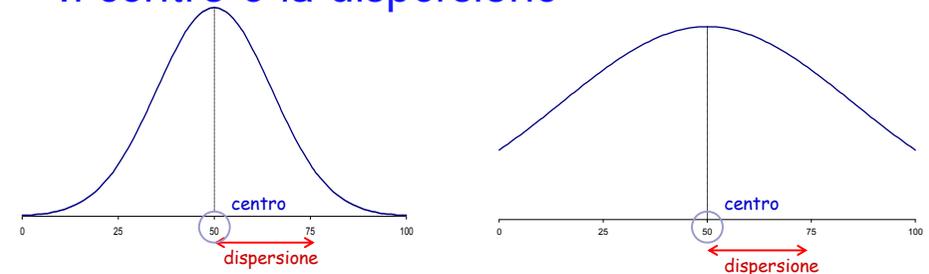
Come convincere il padre a fare il regalo?

- Non tutte le materie sono 'difficili' allo stesso modo
- Possiamo ricalcolare la media assegnando un grado di difficoltà diverso a ciascuna materia

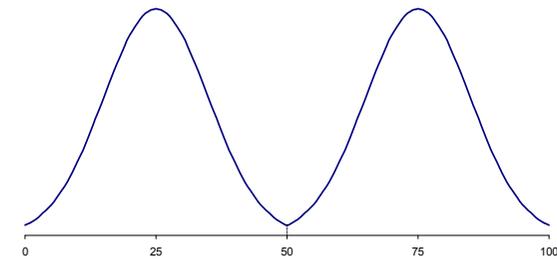
- Calcoliamo la media
 $70/10=7$

materia	Voto	peso	voto*peso
italiano	8	2	16
storia	5	1	5
geografia	6	1	6
inglese	7	2	14
scienze	5	1	5
matematica	8	2	16
educ. Fisica	8	1	8
TOTALE	47	10	70

Il centro e la dispersione



stesso centro, dispersione diversa



Il centro non sintetizza la distribuzione!

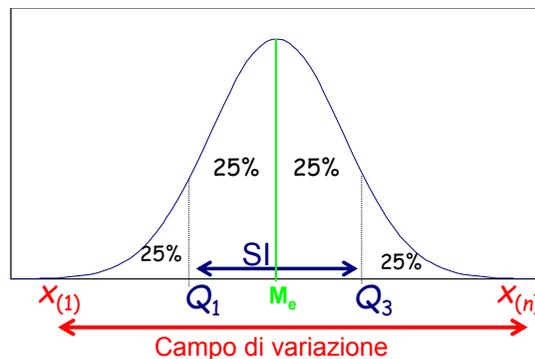
Indici di dispersione

- Se utilizziamo la mediana come indice del centro della distribuzione, dividendo la distribuzione in due parti, possiamo usare la stessa idea per misurare la dispersione
 - Troviamo i valori che dividono queste due metà ancora in due metà!
 - Q_1 è il primo quartile e Q_3 è il terzo quartile
 - Adesso la distribuzione è divisa in 4 parti uguali!

La distanza tra Q_1 e Q_3 è una misura di dispersione detta **scarto interquartile**

$$SI = Q_3 - Q_1$$

La distanza tra il valore più piccolo e il valore più grande è il **campo di variazione** (range)



Quale indice?

- Quale indice di posizione e dispersione utilizzare dipende anche dall'obiettivo con cui si calcolano questi indici.
- Se l'obiettivo è **descrivere** il fenomeno, e la variabile è **quantitativa**, gli indici più informativi sono:
 - la **media aritmetica** e la **deviazione standard** se la distribuzione è approssimativamente normale
 - la **mediana** e lo **scarto interquartile** se la distribuzione è asimmetrica

Un esempio

DATI SUI MAMMIFERI

Esempio: dati sui mammiferi

- dati mammiferi: 39 osservazioni × 6 variabili

Gli scoiattoli comunemente detti **hipmunk** (*Tamias*, Illiger 1811) costituiscono un genere dell'ordine dei Roditori.



Curiosità: i personaggi Disney *Cip & Ciop*, sono due Chipmunk.

Proviamo a fare un esercizio di analisi descrittiva

- Costruite l'istogramma per la durata di vita media e massima per alcuni mammiferi
- Convieni usare classi di 10 anni per la durata massima e di 5 anni per la durata media
- Vediamo insieme cosa possiamo leggere da questi grafici

Istogramma durata vita

- la vita media varia tra 1 e 41, mentre la vita massima varia da 4 a 70 anni
→ Meglio utilizzare **classi** di 5 anni per la vita media e di 10 per la vita massima

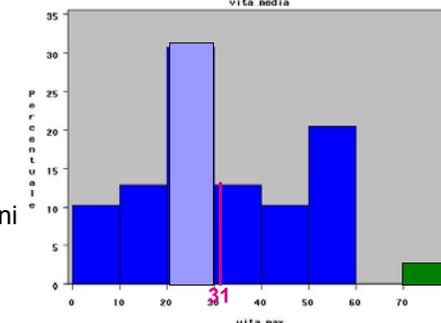
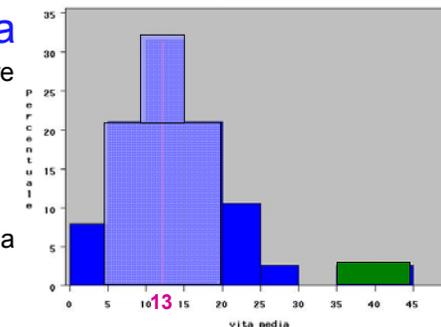
La **forma** dell'istogramma della vita media è molto diversa da quella della vita massima

La distribuzione della **vita media**

- è **centrata sul valore 13** (media)
- la maggior parte dei valori sono **compresi tra 5 e 20**
- ci sono valori anomali (**outliers**) a destra, nelle classi [35,40) e [40,45)

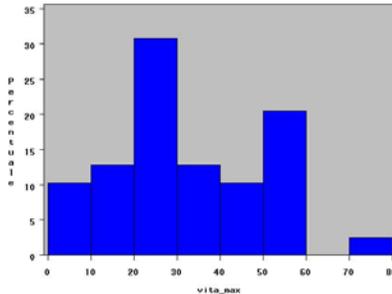
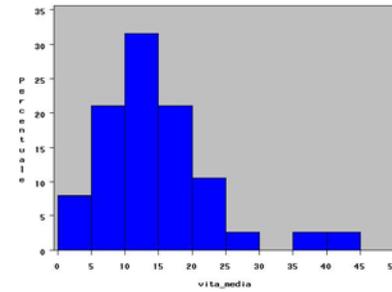
La distribuzione della **vita massima** è più regolare

- c'è un **picco** nella classe [20,30) anni
- e un **valore anomalo** nella classe [70,80) anni.
- il **valore centrale** è molto più alto!



Media e mediana: speranza di vita dei mammiferi

- Il valore in cui l'istogramma sta in equilibrio (media) è più grande del valore che divide l'area in due parti uguali (la mediana) perché la distribuzione non è simmetrica
- Se la distribuzione fosse simmetrica media e mediana sarebbero uguali
- I valori anomali a destra tendono a far crescere il valore medio ma non hanno effetto sulla mediana
- Per esempio, se i valori della classe [35, 40) fossero spostati nella classe [45, 50) la mediana resterebbe uguale mentre la media sarebbe più grande!



Esempio mammiferi

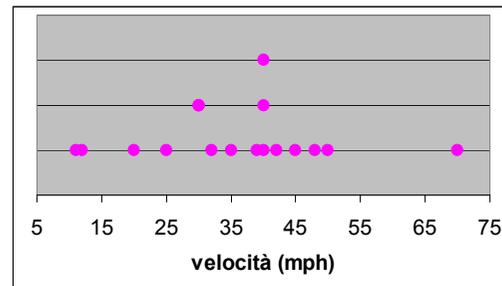


- I gatti domestici vivono circa 12 anni, ma possono vivere anche fino a 28 anni.
Questa durata è tipica di altri carnivori domestici?
- Cosa si può dire per i mammiferi domestici che non sono predatori, come le mucche e i maiali?
E per i mammiferi che vivono allo stato selvaggio?
- I rinoceronti, mammiferi non predatori, vivono in media 15 anni, con una durata massima di 45 anni. Mentre l'orso grizzly, che è un predatore, vive in media 25 anni, con una durata di vita massima di circa 50 anni.
Possiamo affermare che tra i mammiferi selvaggi i carnivori vivono più a lungo dei vegetariani?

Possiamo rispondere ad alcune di queste domande osservando i grafici dei dati raccolti

DOTPLOT

- mostra i singoli casi osservati come punti
- dal dotplot possiamo vedere la forma, il centro e la dispersione dei dati



Il dot plot è **utile** quando:

- si hanno **pochi casi**
- si vogliono vedere i **singoli valori**
- si vuole vedere qual è la **forma** della distribuzione
- si vogliono confrontare **pochi gruppi**

Quando leggete un dot plot, tenete presente che software diversi fanno dot plot diversi: a volte 1 punto rappresenta 1 singolo caso, a volte 2 o più casi, a volte i valori vengono arrotondati

Dotplot per durata della gestazione

- Costruiamo il dotplot (vedi foglio di [lavoro](#))
- Osservando il dotplot proviamo a descrivere la forma, il centro e la dispersione



- Quali mammiferi presentano la gestazione più lunga?

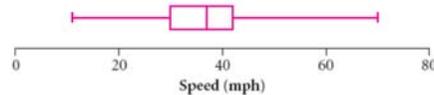
Quelli più grandi!

Box-plot

È possibile sintetizzare la distribuzione attraverso 5 numeri:

- **minimo:** il più piccolo valore osservato
- **Q1:** la mediana della prima metà dei valori
- **Mediana:** il valore che divide i dati in due parti
- **Q3:** la mediana della metà superiore dei valori
- **massimo:** il valore più grande osservato

1	1 2	min	11
2	0 5	Q1	30
3	0 0 0 2 5 9	median	37
4	0 0 0 2 5 8	Q3	42
5	0	max	70
6			
7	0		



Quando sono utili i box-plot

- per rappresentare una variabile **quantitativa**
- se si vogliono **confrontare** la forma, il centro e la dispersione di due o più gruppi di osservazioni
- Quando si hanno molti valori e non è possibile fare un dotplot o uno stemplot
- Quando non si è interessati a rappresentare i valori individuali
- Quando si vogliono individuare chiaramente gli outlier

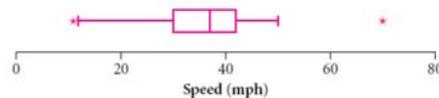
Valori anomali (outliers)

Gli outliers sono valori '*distanti*' dagli altri

Per trovare gli outliers: si cercano i valori che sono più grandi o più piccoli di 1.5 volte SI rispetto a $Q3$ o $Q1$, cioè:

$$> Q3 + 1.5 \cdot SI \text{ o } < Q1 - 1.5 \cdot SI$$

- **Esempio** velocità dei mammiferi
 $Q1 = 30$ e $Q3 = 42 \rightarrow SI = 42 - 30 = 12$, $1.5 \cdot SI = 18$
- **All'estremo inferiore** $Q1 - 1.5 \cdot SI = 30 - 18 = 12$
 - Il **maiale** con una velocità di 11 mph è un outlier.
 - Lo scoiattolo è il non-outlier con la velocità più piccola: 12 mph
- **All'estremo superiore** $42 + 1.5 \cdot SI = 42 + 18 = 60$
 - Il **ghepardo** con una velocità di 70 mph, è un outlier.
 - il leone con 50 mph è il non-outlier con la velocità più alta.



Speranza di vita media (anni) per 38 mammiferi

- Ordiniamo i **dati**
- Troviamo i 5 numeri che sintetizzano la distribuzione
- Troviamo $SI = Q3 - Q1$
 - $SI = 15 - 8 = 7$
- Calcoliamo $Q1 - 1.5 \cdot SI$
 - $Q1 - 1.5 \cdot SI = 8 - 1.5 \cdot 7 = -2.5 < \min = 1$
 - no outliers nella metà inferiore
- Calcoliamo $Q3 + 1.5 \cdot SI$
 - $Q3 + 1.5 \cdot SI = 15 + 1.5 \cdot 7 = 25.5 < \max = 41$
 - outliers: elefante 35 anni e ippopotamo 41 anni
 - Il valore più grande che non è un outlier è 25 anni (grizzly)
- Disegniamo il boxplot.
- È possibile che un boxplot non abbia baffi?

Alcune esperienze a livello nazionale e internazionale

COME TRASMETTERE LE COMPETENZE?



Quale metodo usare?

- È importante utilizzare **dati reali**
 - Analizzare il contesto durante la discussione degli esempi proposti
- Se possibile, fare in modo che gli studenti imparino a utilizzare un **software statistico** per:
 - costruire grafici
 - costruire tabelle di frequenza
 - calcolare medie e altri indici
- Essere in grado di scegliere i grafici e gli indici statistici appropriati, saper descrivere l'evidenza presente nei dati è molto più importante che conoscere i dettagli computazionali.

Lavorare insieme

- È più semplice 'catturare' l'attenzione dei ragazzi affrontando temi che sono vicini alle **esperienze quotidiane**
- L'approccio statistico alla conoscenza della realtà si presta ad attività di '**laboratorio**'
- Varie attività proposte per la scuola primaria e secondaria di primo grado dal documento UMI-SIS **La matematica per il cittadino**
<http://www.sis-statistica.it/files/pdf/2008/datieprev.pdf>

Interazione con altre discipline

- Il modo di pensare statistico e i metodi quantitativi possono essere veicolati anche attraverso altre discipline:
 - metodi per le scienze **sperimentali**: fisica, chimica, scienze in generale
 - metodi per l'**osservazione** dei fenomeni economico/sociali: geografia, storia, lettere
 - utilizzo di tecnologie informatiche per l'**elaborazione** dei dati

Imparare facendo

- Le recenti teorie di apprendimento suggeriscono che una buona pratica di insegnamento consiste nel disegnare ambienti di apprendimento che stimolino gli studenti a costruire la conoscenza.
 - p.e attività che diano agli studenti più opportunità di pensare e discutere
- Esempi di attività sul sito WEB AIMS (Adapting and Implementing Innovative Materials in Statistics <http://www.tc.umn.edu/~aims>).

Esperienze nell'insegnamento della Statistica

Iniziano a essere molte le esperienze proposte, sia in ambito nazionale che internazionale

- Sito [DiSIA](http://local.disia.unifi.it/gmm/scuola/) dedicato all'insegnamento della Statistica e della Probabilità nella Scuola (a cura di G. Marchetti e A. Moro)
<http://local.disia.unifi.it/gmm/scuola/>
- Dalla sezione dedicata alla **didattica della Statistica** nel sito della **SIS** è possibile scaricare **materiali e strumenti didattici** e raggiungere altri siti interessanti
<http://new.sis-statistica.org/formazione/didattica-per-la-statistica/>

Materiale per lezioni



- Sito dell'American Statistical Association dedicato ai docenti di matematica che devono insegnare concetti di statistica.
- Contiene progetti di lezione, corredati di dati e indicazioni pratiche, proposte e riviste dai docenti stessi
- I concetti statistici proposti seguono le raccomandazioni delle [*Guidelines for Assessment and Instruction in Statistics Education \(GAISE\) Report: A Pre-K-12 Curriculum Framework, Common Core State Standards for Mathematics*](#), e [*NCTM Principles and Standards for School Mathematics*](#).
- Le risorse sono organizzate per livello di apprendimento e argomento seguendo i 4 elementi delle linee guida GAISE :
 - Formulazione della domanda di ricerca
 - Disegno e raccolta dei dati
 - Analisi dei dati attraverso misure e grafici
 - Interpretazione dei dati nel contesto

<http://www.amstat.org/education/stew/>

Esperienze in ambito internazionale

- Elenco completo sul sito della SIS e dell'ISTAT
- Da segnalare
 - [IASE International Statistical Literacy Project](#) materiale per gli insegnanti da usare in classe o per migliorare la conoscenza della Statistica , per valutare le conoscenze acquisite dagli studenti e molto altro!)
 - [Statistics Canada](http://www.statcan.gc.ca/edu/index-eng.htm) materiale di supporto per insegnanti e studenti
 - [Royal Statistical Society](http://www.rss.org.uk) due riviste Teaching Statistics e Significance e varie risorse

Un invito a esplorare e divertirsi ...

CONSORTIUM FOR THE ADVANCEMENT OF UNDERGRADUATE STATISTICS EDUCATION

- CAUSE è un'organizzazione USA il cui scopo è quello di supportare e migliorare l'istruzione statistica, promuovendo lo sviluppo professionale e gli scambi tra docenti e ricerca
- Disponibili molti materiali per l'insegnamento della statistica:
 - Lezioni e dati
 - Fumetti, canzoni e rompicapo

<http://www.causeweb.org/>



Swivel Una miniera di dati

- Swivel è nato in California il 6 dicembre 2006
- Swivel permette di esplorare i dati e condividere con altri le proprie conclusioni.
- Swivel contiene dati sulla politica, l'economia, il tempo, gli sport, gli affari e altri ancora

<http://www.swivel.com/>

V edizione Olimpiadi della Statistica

- Le Olimpiadi Nazionali di Statistica, bandite dalla Società Italiana di Statistica (SIS), in collaborazione con l'Istituto Nazionale di Statistica (ISTAT), hanno come obiettivo principale quello di avvicinare gli studenti al ragionamento statistico, incoraggiandoli ad un maggiore interesse verso l'analisi dei dati, la probabilità e le sue applicazioni.
- Le Olimpiadi di Statistica sono rivolte agli studenti frequentanti le classi IV e V di tutte le scuole secondarie di secondo grado presenti sul territorio italiano.
- Iscrizioni: **Scadenza 26 febbraio 2015**

<http://new.sis-statistica.org/7970-2/>

La Società Italiana di Statistica



- La SIS è una **società scientifica** senza fine di lucro, costituita nel 1939 con lo scopo fondamentale di promuovere lo **sviluppo delle Scienze Statistiche** e delle loro applicazioni in campo economico, sociale, sanitario, demografico, tecnologico, produttivo e in molti altri ambiti di ricerca.
- ha una sezione è dedicata a materiali e informazioni utili per la formazione in statistica e la divulgazione della cultura statistica nella società
- I materiali sono scaricabili dal sito WEB della SIS

<http://www.sis-statistica.it>

Riassumendo: strumenti per la trasmissione delle competenze

- Consultare materiale in rete (DiSIA, SIS e ISTAT buon punto di partenza)
- Attenzione: i libri di testo scolastici contengono ancora diversi errori!
- UMI-SIS-MIUR (2003, 2004, 2005) *Raccolta di attività di supporto curricolare per la scuola secondaria di secondo grado* Pubblicazioni MIUR, nell'ambito del protocollo d'intesa UMISIS-MIUR (a cura di G. Anichini *et al.*)

■ INDUZIONI

Demografia, probabilità, statistica a scuola
Fabrizio Serra editore, Pisa - Roma

Rivista semestrale pubblicata con il patrocinio della Società Italiana di Statistica



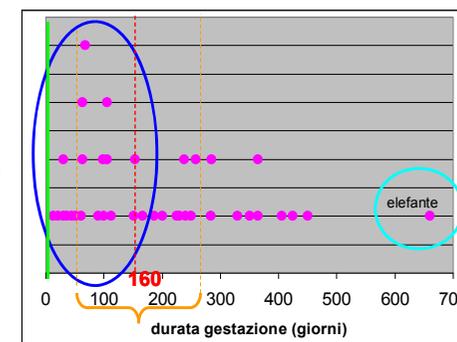
Grazie per la vostra attenzione ...



risposte

- La distribuzione è centrata verso i **valori più bassi**, senza gruppi o buchi particolari

- C'è una sorta di **'muro'** a 0 giorni, perché nessun mammifero può avere un periodo di gestazione più piccolo!



- L'elefante è l'unico mammifero fuori norma (**outlier**)
- Circa la **metà** dei mammiferi hanno un periodo di gestazione superiore a 160 giorni e la metà hanno un periodo più breve
- La **metà centrale** ha un periodo di gestazione che varia tra i 63 e i 284 giorni.

