La Probabilità e l'Inferenza statistica

note introduttive

Dipartimento di Statistica "Giuseppe Parenti"

Dipartimento di Statistica "Giuseppe Parenti"

Che cos'è la STATISTICA ?

A che servono gli statistici ?

Che cos'è la statistica ?

per ***Statistica*** **si intendeva in origine la raccolta di dati economico-demografici di vitale interesse per lo Stato**.
Da allora la statistica si è sviluppata in un metodo scientifico d'analisi, oggi applicato a tutte le scienze sociali e naturali.

Definizione formale:

la **Statistica** è ***la disciplina che si occupa dell'elaborazione dei risultati dell'osservazione di uno o più caratteri posseduti dagli elementi di un insieme determinato****, con l'intento di esprimere un giudizio su alcuni aspetti di una realtà di interesse* che, in quanto riferita ad un insieme e non ai singoli elementi che lo compongono, viene chiamata ***fenomeno collettivo***.

Dipartimento di Statistica
"Giuseppe Parenti"

Che cos'è la statistica ?

In altre parole:

La Statistica è Matematica Applicata volta all'analisi dei dati concernenti un particolare fenomeno,
dati che possono derivare da:

- **studi osservazionali**
- **studi di sperimentazione**

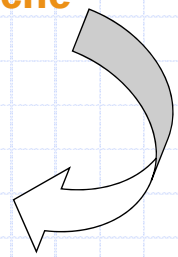
Per ***dato osservazionale*** si intende il dato rilevato su un fenomeno precostituito, esistente in natura e generalmente finito.

Per ***dato sperimentale*** si intende invece un dato creato in circostanze controllate, mediante una replicazione dell'esperimento potenzialmente infinita.

L'essenza della Statistica

La ragion d'essere della statistica è la presenza di un certo livello di variabilità nei dati
(altrimenti, paradossalmente, la comprensione del fenomeno sarebbe possibile ricorrendo ad una sola osservazione)

nello studio dei fenomeni collettivi si è consapevoli che
al variare dell'unità statistica *u* entro una certa popolazione *P* = {*u*} variano certe caratteristiche misurate su *u*



in altre parole,
per lo studio di un fenomeno caratterizzato
da assenza di variabilità **non serve scomodare uno statistico**

Dipartimento di Statistica
"Giuseppe Parenti"

L'essenza della Statistica

Alcuni semplici esempi:

- **altezza e peso degli studenti di una classe**
- **reddito dei parlamentari**
- **votazioni riportate all'esame di Statistica dagli studenti di un certo corso di Laurea**
- **valutazione dell'efficacia dei titoli di studio universitari**
- **soddisfazione nei confronti del trasporto pubblico locale**
- **durata delle lampadine ad alto risparmio energetico**
- **ecc…**

Il problema della misura

Conversione dell'oggetto di studio in dato statistico:

problema del tipo di misura che si intende adottare come espressione delle manifestazioni rilevate

- altezza → metro
- soddisfazione → ????

- **Misura di tutte le unità statistiche**
  (analisi costo/tempo, impossibilità a procedere...)

- **Misura sola di una (piccola) parte di esse**

Dipartimento di Statistica "Giuseppe Parenti"

Le branche della STATISTICA

Statistica DESCRITTIVA

In questo settore rientrano i metodi per **sintetizzare** con opportune grandezze le caratteristiche salienti dei fenomeni collettivi.
La **descrizione** passa attraverso le fasi di formazione del dato statistico e del trattamento matematico dello stesso.

Per formazione del dato statistico si può intendere:

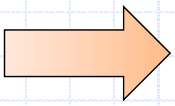
- l'elaborazione di dati preesistenti in natura (dati anagrafici, indici aziendali di bilancio, dati di produzione industriale ecc.)
- la necessità di procedere **all'effettiva rilevazione delle informazioni** necessarie alla comprensione del fenomeno di interesse

Il processo di rilevazione dei dati

Il **processo di rilevazione dei dati** è generalmente distinto nelle fasi di:

- **definizione del piano di rilevazione,**
- **raccolta delle informazioni,**
- **spoglio e classificazione.**

La fase più delicata è senza dubbio la prima.



- **Rilevazioni complete (censuarie)**
- **Rilevazioni campionarie**

Dipartimento di Statistica
"Giuseppe Parenti"

Le rilevazioni campionarie

Fondamentale diviene quindi in statistica il ruolo dell'esperimento campionario.

Si dice **campione** un aggregato di unità statistiche appartenenti ad una certa popolazione e selezionate mediante un esperimento di campionamento (generalmente collegato ad una procedura di randomizzazione).

Si osservi che la strategia campionaria è la sola possibile quando la popolazione è virtualmente infinita; inoltre la popolazione da cui si estrae il campione, detta popolazione campionata, non sempre coincide con la popolazione obiettivo.

Le rilevazioni campionarie

Tra i vari tipi di campionamento una menzione particolare spetta al **campione casuale semplice CCS** che è una forma di campionamento **equiparabile allo schema dell'urna** in cui le palline - che identificano le diverse unità della popolazione - **hanno tutte la stessa probabilità di essere estratte**.

Si distingue fra:

- **CCS con reimmissione**
- **CCS senza reimmissione.**

Si può in generale affermare che un campione non costituisce quasi mai una riproduzione fedele della popolazione su piccola scala.

Dipartimento di Statistica
"Giuseppe Parenti"

Le branche della STATISTICA

❑ Statistica INFERENZIALE

Tuttavia se l'estrazione è casuale, i dati possono fornire informazioni sulla variabilità della popolazione e sulla fiducia da accordare a tali informazioni. Questi problemi sono oggetto della Statistica Inferenziale o Induttiva.

Il termine inferenza deriva dal latino e letteralmente significa: argomentare, desumere.

Si effettua inferenza quando si generalizza l'esperimento operando una sorta di estensione dal particolare al generale; le generalizzazioni però non sono certe.

Statistica inferenziale e Probabilità

L'inferenza è quindi un processo d'azzardo e l'incertezza viene misurata in termini probabilistici.

La PROBABILITÀ è il fondamento logico per fare inferenza sulla Popolazione oggetto d'indagine.

Ma ...

che cos'è la PROBABILITÀ?

Dipartimento di Statistica "Giuseppe Parenti"

La Probabilità

La PROBABILITÀ è un concetto primitivo.

Per definirla occorre introdurre alcuni ingredienti:

- **esperimento casuale**
- **evento**
- **spazio degli eventi B**

es: lancio del dado o di una moneta

La probabilità è una funzione matematica su B con certe proprietà

La Probabilità

Approcci alla probabilità (*in ordine storico*):

- **impostazione classica;**
  es: moneta
- **impostazione frequentista;**
  es: moneta truccata
- **impostazione soggettiva;**
  es: uomo su Marte

- **impostazione Assiomatica**

Dipartimento di Statistica
"Giuseppe Parenti"

L'impostazione assiomatica delle Probabilità (Kolmogorov)

1) $P(A) \geq 0$

2) $P(\Omega) = 1$

3) $P(A \cup B) = P(A) + P(B)$

$se \quad A \cap B = \varnothing$

4) $P(A / B) = \dfrac{P(A \cap B)}{P(B)}$

Formalizzazione matematica di concetti intuitivi

Un po' di svago

- **Dado dietro la porta**
- **Gioco del lotto**: prob. che esca 23 al secondo estratto
- **Mazzo di 40 carte**: prob. che esca un K alla seconda estraz.
- **Le 3 buste**
- **I 3 prigionieri**
- **35 studenti su un autobus:** prob. che almeno 2 abbiano stessa data di nascita
- Il valore atteso ed il **Paradosso di San Pietroburgo**
- Come misurare **l'area di un lago**

Un po' di svago: le 3 buste

Solo una contiene un grosso premio; le altre due sono vuote. Giochiamo con Gino e proponiamo a Gino di scegliere una busta.

Gino sceglie la busta **A**.

Una volta scelta, facciamo vedere a Gino, aprendola, che una tra le buste **B** e **C** è vuota.

Offriamo a Gino la possibilità di poter cambiare la busta **A** con la busta chiusa rimasta sul tavolo.

Il dubbio di Gino: cosa conviene fare?

Un po' di svago: le 3 buste

Inizialmente: $P(A_{vince}) = P(B_{vince}) = P(C_{vince}) = 1/3$

Supponiamo a Gino venga mostrato che B è vuota;
Gino lo considera un evento e condiziona la sua decisione a questo.

$$P(A_{vince}/B_{vuota}) = \frac{P(A_{vince} \cap B_{vuota})}{P(B_{vuota})}$$

$$= \frac{P(B_{vuota}/A_{vince})P(A_{vince})}{P(B_{vuota})}$$

$$= \frac{P(B_{vuota}/A_{vince})P(A_{vince})}{P(B_{vuota}/A_{vince})P(A_{vince}) + P(B_{vuota}/C_{vince})P(C_{vince})}$$

$$= \frac{1 \cdot 1/3}{1 \cdot 1/3 + 1 \cdot 1/3} = \frac{1}{2}$$

ovvero sapendo che "**B è vuota**"
è indifferente conservare A o operare lo scambio

Un po' di svago: le 3 buste

Proviamo a consideriamo invece l'evento "**mostriamo B vuota**"
e condizioniamo la decisione di Gino a questo:

$$P(A_{vince}/mB_{vuota}) = \frac{P(A_{vince} \cap mB_{vuota})}{P(mB_{vuota})}$$

$$= \frac{P(mB_{vuota}/A_{vince})P(A_{vince})}{P(mB_{vuota})}$$

$$= \frac{P(mB_{vuota}/A_{vince})P(A_{vince})}{P(mB_{vuota}/A_{vince})P(A_{vince}) + P(mB_{vuota}/C_{vince})P(C_{vince})}$$

$$= \frac{1/2 \cdot 1/3}{1/2 \cdot 1/3 + 1 \cdot 1/3} = \frac{1/6}{3/6} = \frac{1}{3}$$

si è indifferenti tra mostrare a Gino la busta **B** o la **C,** se **A** è la vincente.

ovvero il vero evento non è quello che Gino vede,
ma l'azione che noi facciamo a seguito della scelta iniziale di Gino.
È certamente conveniente operare lo scambio.

Un po' di svago: i tre prigionieri

Tre prigionieri **A**, **B** e **C** l'indomani saranno condannati a morte.

Il Governatore decide di graziarne uno e comunica la sua scelta al secondino **S**, obbligandolo al silenzio sulla scelta fatta.

A chiede ad **S** di rivelargli il nome di chi si salverà.
S non può parlare pena la sua esecuzione.

In alternativa, **A** chiede ad **S** di comunicargli il nome di chi degli altri due verrà sicuramente condannato.
S accetta ritenendo di non contravvenire agli ordini ricevuti.

A adesso ritiene che la sua probabilità di salvarsi sia pari a ½.

Un po' di svago: i tre prigionieri

Dopo la grazia del Governatore: $P(A) = P(B) = P(C) = 1/3$

$$P(A/S_{diceB}) = \frac{P(A \cap S_{diceB})}{P(S_{diceB})}$$

$$= \frac{P(S_{diceB}/A)P(A)}{P(S_{diceB})}$$

$$= \frac{P(S_{diceB}/A)P(A)}{P(S_{diceB}/A)P(A) + P(S_{diceB}/C)P(C)}$$

$$= ?$$

Occorre fare delle ipotesi.

In primis, assumiamo che S non dica bugie …

Un po' di svago: i tre prigionieri

S è indifferente tra **B** e **C**:

$$P(A/S_{diceB}) = \frac{P(A \cap S_{diceB})}{P(S_{diceB})}$$

$$= \frac{P(S_{diceB}/A)P(A)}{P(S_{diceB})}$$

$$= \frac{P(S_{diceB}/A)P(A)}{P(S_{diceB}/A)P(A) + P(S_{diceB}/C)P(C)}$$

$$= \frac{1/2 \cdot 1/3}{1/2 \cdot 1/3 + 1 \cdot 1/3} = 1/3$$

La probabilità di **A** non cambia
(risultato analogo al gioco delle tre buste)

Un po' di svago: i tre prigionieri

Ad **S** è estremamente antipatico **B, per cui se può fa il suo nome**:

$$P(A/S_{diceB}) = \frac{P(A \cap S_{diceB})}{P(S_{diceB})}$$

$$= \frac{P(S_{diceB}/A)P(A)}{P(S_{diceB})}$$

$$= \frac{P(S_{diceB}/A)P(A)}{P(S_{diceB}/A)P(A) + P(S_{diceB}/C)P(C)}$$

$$= \frac{1 \cdot 1/3}{1 \cdot 1/3 + 1 \cdot 1/3} = 1/2$$

S è costretto a fare il nome di **B** perché è **C** a salvarsi

S se può fa il nome di **B**

La probabilità di **A** sale a ½.

Un po' di svago: i tre prigionieri

Ad **S** è estremamente antipatico **C, per cui se può fa il suo nome**:

$$P(A/S_{diceB}) = \frac{P(A \cap S_{diceB})}{P(S_{diceB})}$$

$$= \frac{P(S_{diceB}/A)P(A)}{P(S_{diceB})}$$

$$= \frac{P(S_{diceB}/A)P(A)}{P(S_{diceB}/A)P(A) + P(S_{diceB}/C)P(C)}$$

$$= \frac{0 \cdot 1/3}{0 \cdot 1/3 + 1 \cdot 1/3} = 0$$

Se è **A** a salvarsi, **S** farebbe il nome di **C per cui, in tal caso, la probabilità che dica B è zero**

A è sicuramente condannato perché se **S** dice **B**, è certo che è stato costretto a dirlo visto che **C** si salverà