

Laboratorio 2

analisi statistica dei dati sperimentali

aa 2021/22

Laboratorio 2 - analisi statistica dei dati

la mia attività

didattica

ricerca

email

anna.martin@ts.infn.it

pagina web del corso

<https://wwwusers.ts.infn.it/~martin/univ/didattica/lab2/lab2.html>

Laboratorio 2 - analisi statistica dei dati

il corso:

parte (4 CFU) del corso di Laboratorio 2
verrà svolta in parallelo con la parte di laboratorio di elettromagnetismo

l'analisi statistica dei dati fa parte della formazione di base di un fisico,
e' un argomento vastissimo, difficile, e affascinante!
fondamentale verificare e applicare la parte teorica

nel corso,
solo elementi di base, senza pretesa di completezza,
mediando tra tempo disponibile e rigore:
non un "ricettario", molte dimostrazioni
e esercizi e applicazioni

Laboratorio 2 - analisi statistica dei dati

prevede lezioni e laboratori, in presenza

con partecipazione attiva di tutti

lezioni: parte teorica

registrazioni: secondo le modalita' previste
non equivalgono alla presenza

ma anche introduzione alle attività di laboratorio e
discussione dei risultati

lunedì e mercoledì dalle 14 alle 16 (14:30-16:00 ?)
quasi tutte le settimane

Laboratorio 2 - analisi statistica dei dati

laboratori:

- con l'aiuto di A. Kerbizi e A. Moretti (post-doc)
- frequenza obbligatoria
 - segnalare eventuali problemi: troveremo soluzioni
- consistono nello svolgimento di esercizi strettamente legati a quanto fatto a lezione e analisi di dati, reali e simulati
- fondamentale l'uso di un computer (ma non è un "laboratorio di calcolo"...)
- libera scelta sul linguaggio di programmazione (default: fortran+gnuplot)
- impegno: 2.5 ore alla settimana, con settimane di recupero
- turni: lunedì' (16.30 - ...), mercoledì' (16.30-...), venerdì' (14.30-...)

Laboratorio 2 - analisi statistica dei dati

attività di laboratorio

- per ogni laboratorio ci sarà un problema specifico da risolvere, generalmente illustrato e discusso prima a lezione
- durante i laboratori, discussione dei problemi (non di calcolo...) e dei risultati
- dettagli sul software standard utilizzabile prima dell'inizio (prossima settimana)
- ogni volta, verrà compilato l'elenco degli studenti che hanno completato il lavoro, cosa necessaria per poter sostenere l'esame

Laboratorio 2 - analisi statistica dei dati

attività di laboratorio

- per poter seguire tutti bisogna formare tre (? numero massimo di studenti per gruppo: 35 (?)) gruppi, con stesso numero di studenti, che faranno laboratorio in tre pomeriggi diversi
- serve l'elenco degli studenti nei diversi gruppi (evitare sovrapposizione con altri corsi/laboratorio em), che resteranno fissi

Laboratorio 2 - analisi statistica dei dati

materiale didattico

- appunti

e su pagina web del corso:

<https://wwwusers.ts.infn.it/~martin/univ/didattica/lab2/lab2.html>

- dispense sintetiche (disponibile presto una nuova versione, rivista nella parte iniziale: continuazione?)
- testi di consultazione

Laboratorio 2 - analisi statistica dei dati

esame

- necessario aver svolto e discusso tutti gli esercizi proposti durante il corso (verifica durante il corso)
- necessario presentare con una settimana di anticipo una relazione su un lavoro di analisi individuale che includa stima di parametri e test d'ipotesi
 - dati: possibilmente da misure in (altri) laboratori
 - solo per la prima sessione di esami:
 - relazione su uno degli ultimi esercizi proposti (da definire durante il corso)
- all'orale: domande sulla parte teorica, sugli esercizi, sulla relazione
- un unico esame per Laboratorio 2, con un unico voto

Laboratorio 2 - analisi statistica dei dati

sul programma del corso ...

... alcune definizioni

Laboratorio 2 - analisi statistica dei dati

Treccani

“statistica

Scienza che ha per oggetto lo studio dei fenomeni collettivi suscettibili di misura e di descrizione quantitativa: basandosi sulla raccolta di un grande numero di dati inerenti ai fenomeni in esame, e partendo da ipotesi più o meno direttamente suggerite dall'esperienza o da analogie con altri fenomeni già noti

mediante l'applicazione di metodi matematici fondati sul calcolo delle probabilità, si perviene alla formulazione di leggi di media che governano tali fenomeni, dette leggi statistiche

spesso la raccolta dei dati viene limitata a un campione più ristretto, opportunamente predeterminato in modo da rappresentare fedelmente le caratteristiche generali. ...”

Laboratorio 2 - analisi statistica dei dati

Wikipedia

La **statistica** è una disciplina che ha come fine lo studio quantitativo e qualitativo di un particolare fenomeno collettivo in condizioni di incertezza o non determinismo, cioè di non completa conoscenza di esso o parte di esso.

Strumento del metodo scientifico, si avvale della matematica per studiare i modi in cui un fenomeno collettivo può essere sintetizzato e compreso e ciò avviene attraverso la raccolta e l'analisi delle informazioni relative al fenomeno studiato

Laboratorio 2 - analisi statistica dei dati

Wikipedia

La **statistica** è una disciplina che ha come fine lo studio quantitativo e qualitativo di un particolare fenomeno collettivo in condizioni di incertezza o non determinismo, cioè di non completa conoscenza di esso o parte di esso.

Strumento del metodo scientifico, si avvale della **matematica** per studiare i modi in cui un fenomeno collettivo può essere sintetizzato e compreso e ciò avviene attraverso la raccolta e l'analisi delle informazioni relative al fenomeno studiato

La scienza statistica è comunemente suddivisa in due branche principali:

- statistica **descrittiva**: ha come scopo quello di sintetizzare i dati attraverso ... strumenti grafici (diagrammi a barre, a torta, istogrammi,...) e indici (indicatori statistici, indicatori di posizione, di dispersione, di forma, ...) che descrivono gli aspetti salienti dei dati osservati
- statistica **inferenziale**. ha come obiettivo quello di stabilire delle caratteristiche dei dati e dei comportamenti delle misure rilevate (variabili statistiche) con una possibilità di errore predeterminata. Le inferenze possono riguardare la natura teorica (la legge probabilistica) del fenomeno che si osserva. La conoscenza di questa natura permetterà poi di fare una previsione (...ad esempio, ..."l'inflazione il prossimo anno avrà una certa entità" ... esiste un modello dell'andamento dell'inflazione derivato da tecniche inferenziali). La statistica inferenziale è fortemente legata alla **teoria della probabilità**....

La statistica inferenziale si suddivide poi in altri capitoli, di cui i più importanti sono la teoria della stima (stima puntuale e stima intervallare) e la verifica delle ipotesi.

Laboratorio 2 - analisi statistica dei dati

Wikipedia

La **statistica** è una disciplina che ha come fine lo studio quantitativo e qualitativo di un particolare fenomeno collettivo in condizioni di incertezza o non determinismo, cioè di non completa conoscenza di esso o parte di esso.

Strumento del metodo scientifico, si avvale della matematica per studiare i modi in cui un fenomeno collettivo si manifesta e ciò avviene attraverso la raccolta e l'analisi dei dati. Lo studio di un fenomeno studiato **la useremo, continuamente: "guardare" i dati e' fondamentale**

La scienza statistica è comunemente suddivisa in due branche principali:

- statistica **descrittiva**: ha come scopo quello di sintetizzare i dati attraverso ... strumenti grafici (diagrammi a barre, a torta, istogrammi,...) e indici (indicatori statistici, indicatori di posizione, di dispersione, di forma, ...) che descrivono gli aspetti salienti dei dati osservati
- statistica **inferenziale**. ha come obiettivo quello di stabilire delle caratteristiche dei dati e dei comportamenti delle misure rilevate (variabili statistiche) con una possibilità di errore predeterminata. Le inferenze possono riguardare la natura teorica (la legge probabilistica) del fenomeno che si osserva. La conoscenza di questa natura permetterà poi di fare una previsione (...ad esempio, ..."l'inflazione il prossimo anno avrà una certa entità" ... esiste un modello dell'andamento dell'inflazione derivato da tecniche inferenziali). La statistica inferenziale è fortemente legata alla teoria della probabilità....

La statistica inferenziale si suddivide poi in altri capitoli, di cui i più importanti sono la teoria della stima (stima puntuale e stima intervallare) e la verifica delle ipotesi.

statistica

Wikipedia

La **statistica** è una disciplina che ha come fine lo studio quantitativo e qualitativo di un particolare fenomeno collettivo in condizioni di incertezza o non determinismo, cioè di non completa conoscenza di esso o parte di esso.

Strumento del metodo scientifico, si avvale della matematica per studiare i modi in cui un fenomeno collettivo può essere sintetizzato e compreso e ciò avviene attraverso la raccolta e l'analisi delle informazioni relative al fenomeno studiato

La scienza statistica è comunemente suddivisa in due branche principali:

- statistica **descrittiva**: ha come scopo quello di sintetizzare i dati attraverso ... strumenti grafici (torta, istogrammi,...) e indici (indicatori statistici, indicatori di posizione, di forma, ...) che descrivono gli aspetti salienti dei dati osservati
- statistica **inferenziale**. ha come obiettivo quello di stabilire delle caratteristiche dei dati e dei comportamenti delle misure rilevate (variabili statistiche) con una possibilità di errore predeterminata. Le inferenze possono riguardare la natura teorica (la legge probabilistica) del fenomeno che si osserva. La conoscenza di questa natura permetterà poi di fare una previsione (...ad esempio, ..."l'inflazione il prossimo anno avrà una certa entità" ... "l'andamento dell'inflazione derivante da questa previsione").
L'inferenza statistica è fortemente legata alla teoria della stima e al test d'ipotesi.

La statistica inferenziale si suddivide poi in altri capitoli, di cui i più importanti sono la **teoria della stima** (stima puntuale e stima intervallare) e la **verifica delle ipotesi**.

statistica

inferenza statistica.

- **stima dei parametri** (puntuali o di intervalli)
 - estrarre da un campione rappresentativo informazioni sull'intera popolazione può essere la stima dei parametri della distribuzione di una variabile casuale (grandezza fisica, misurabile)
oppure dei parametri che compaiono in una relazione tra variabili casuali (grandezze fisiche)
a partire da un campione (insieme di misure)
- **test d'ipotesi**
 - estrarre informazioni quantitative sulla validità di un'ipotesi statistica (funzione di distribuzione di variabili casuali, relazioni tra variabili casuali) a partire da un campione rappresentativo

statistica

inferenza statistica.

- **stima dei parametri** (puntuali o di intervalli)
 - estrarre da un campione rappresentativo informazioni sull'intera popolazione può essere la stima dei parametri della distribuzione di una variabile casuale (grandezza fisica, misurabile)
oppure dei parametri che compaiono in una relazione tra variabili casuali (grandezze fisiche)
a partire da un campione (insieme di misure)
- **test d'ipotesi**
 - estrarre informazioni quantitative sulla validità di un'ipotesi statistica (funzione di distribuzione di variabili casuali, relazioni tra variabili casuali) a partire da un campione rappresentativo

argomenti molto vasti, di cui vedremo solo alcuni metodi e loro applicazioni
saranno trattati nella seconda parte del corso: per affrontarli servono più nozioni di **teoria della probabilità**

darò per scontati i concetti fondamentali di fenomeni statistici, probabilità, variabili casuali, distribuzioni di probabilità e funzioni di distribuzione,....
anche se molte cose le rivedremo

Laboratorio 2 - analisi statistica dei dati

programma del corso

- fenomeni statistici caratterizzati da una variabile casuale
 - rapido ripasso
 - indicatori statistici, momenti, funzioni generatrici dei momenti
 - alcune funzioni di distribuzione
- più variabili casuali (campioni)
 - funzioni di distribuzione congiunte
 - funzione di distribuzione multinormale, distribuzione multinomiale
- funzioni di una o più variabili casuali
 - considerazioni generali
 - alcune funzioni importanti: media, somma di quadrati,...
- stima dei parametri
 - metodo del maximum likelihood
 - metodo dei minimi quadrati
 - intervalli di confidenza
- test d'ipotesi
 - test parametrici
 - test di χ^2

Laboratorio 2 - analisi statistica dei dati

prima di iniziare il corso vero e proprio

formule per la **stima dei parametri di una retta**
utili da subito per le misure

anticipando quello che faremo nella seconda metà del corso,
formule per la stima dei parametri di una retta

ipotesi: $Y = mX + q$

dati: n coppie di valori misurati (x_i, y_i)

con incertezze statistiche trascurabili su x_i , e σ_i su y_i

anticipando quello che faremo nella seconda metà del corso,
formule per la stima dei parametri di una retta

ipotesi: $Y = mX + q$

dati: n coppie di valori misurati (x_i, y_i)

con incertezze statistiche trascurabili su x_i , e σ_i su y_i

metodo dei minimi quadrati: stima migliore di m e di q valori che minimizzano

$$X^2 = \sum_{i=1}^n \frac{(y_i - y_i^t)^2}{\sigma_i^2}, \quad y_i^t = mx_i + q$$

$$\left\{ \begin{array}{l} \frac{\partial X^2}{\partial m} = 0 \\ \frac{\partial X^2}{\partial q} = 0 \end{array} \right. \Rightarrow \begin{array}{l} \hat{m} = \frac{1}{D} (S_{00}S_{11} - S_{10}S_{01}) \\ \hat{q} = \frac{1}{D} (S_{01}S_{20} - S_{11}S_{10}) \end{array} \quad \begin{array}{l} D = S_{00}S_{20} - S_{10}^2 \\ S_{jk} = \sum_{i=1}^n \frac{x_i^j y_i^k}{\sigma_i^2} \end{array}$$

e, con la legge di propagazione della varianza $\sigma_{\hat{m}}^2 = \sum_i \left(\frac{\partial \hat{m}}{\partial y_i} \right)^2 \sigma_i^2, \dots$

$$\sigma_{\hat{m}}^2 = \frac{S_{00}}{D} \quad \sigma_{\hat{q}}^2 = \frac{S_{20}}{D}$$

prima di iniziare 2

anticipando quello che faremo nella seconda metà del corso,
formule per la stima dei parametri di una retta

ipotesi: $Y = mX + q$

dati: n coppie di valori misurati (x_i, y_i)

con incertezze statistiche trascurabili su x_i , e σ_i su y_i

metodo dei minimi quadrati: stima migliore di m e di q valori che minimizzano

$$X^2 = \sum_{i=1}^n \frac{(y_i - y_i^t)^2}{\sigma_i^2}, \quad y_i^t = mx_i + q$$

$$\left\{ \begin{array}{l} \frac{\partial X^2}{\partial m} = 0 \\ \frac{\partial X^2}{\partial q} = 0 \end{array} \right. \Rightarrow \begin{array}{l} \hat{m} = \frac{1}{D} (S_{00}S_{11} - S_{10}S_{01}) \\ \hat{q} = \frac{1}{D} (S_{01}S_{20} - S_{11}S_{10}) \end{array} \quad \begin{array}{l} D = S_{00}S_{20} - S_{10}^2 \\ S_{jk} = \sum_{i=1}^n \frac{x_i^j y_i^k}{\sigma_i^2} \end{array}$$

e, con la legge di propagazione della varianza $\sigma_{\hat{m}}^2 = \sum_i \left(\frac{\partial \hat{m}}{\partial y_i} \right)^2 \sigma_i^2, \dots$

$$\sigma_{\hat{m}}^2 = \frac{S_{00}}{D} \quad \sigma_{\hat{q}}^2 = \frac{S_{20}}{D}$$

verificare

note

- a) 1. se $Y = mX$ **formule simili, più semplici, facili da ricavare**
- b) se è l'incertezza su y_i trascurabile, **basta scambiare X e Y**

note

a) 1. se $Y = mX$ **formule simili, più semplici, facili da ricavare**

b) se è l'incertezza su y_i trascurabile, **basta scambiare X e Y**

c) se **entrambe le incertezze sono non trascurabili**

al primo ordine

si stimano una prima volta i parametri come se le incertezze su x_i fossero trascurabili $\rightarrow \hat{m}_0, \hat{q}_0$

si propaga l'errore $\sigma_{y_i, x_1}^2 = \hat{m}_0^2 \sigma_{x_i}^2$, $\sigma_i'^2 = \sigma_i^2 + \sigma_{y_i, x_i}^2$

si stimano di nuovo i parametri

....fino ad ottenere risultati stabili

note

- a) 1. se $Y = mX$ **formule simili, più semplici, facili da ricavare**
- b) se è l'incertezza su y_i trascurabile, **basta scambiare X e Y**
- c) se **entrambe le incertezze sono non trascurabili**
- d) se **la relazione tra X e Y non è lineare...**

spesso si può “linearizzare”

ad es $V(t) = V_0 e^{-t/c}$ “campione” (t_i, V_i)

però $\ln V = \ln V_0 - t/c$

e quindi, introducendo $Y = \ln V$ e $X = t$, ci si riconduce al caso precedente

NB: $\sigma_i^2 = \dots \sigma_{V_i}^2$

note

- a) 1. se $Y = mX$ **formule simili, più semplici, facili da ricavare**
- b) se è l'incertezza su y_i trascurabile, **basta scambiare X e Y**
- c) se **entrambe le incertezze sono non trascurabili**
- d) se **la relazione tra X e Y non è lineare...**
- e) incertezze sui parametri e correlazione

sara' il primo esercizio