

## Regressione e correlazione lineare

### Indicazioni dei programmi ministeriali

#### PNI per il biennio:

L'argomento in questione si trova trattato nel **punto b (Elementi di statistica descrittiva: rilevazioni di dati, valori di sintesi, indici di variabilità, regressione e correlazione)** del **tema 4 (Elementi di probabilità e statistica)**.

#### PNI per il triennio

Il **tema 4** tratta di probabilità e statistica; **punto a:** statistica descrittiva bivariata, distribuzioni statistiche (congiunte, condizionate, marginali), coefficienti di correlazione; **punto c:** correlazione, formula di Bayes, variabili aleatorie in una e due dimensioni.

### Destinatari

Classe quarta liceo scientifico PNI; si svolge nel secondo quadrimestre e le ore settimanali di matematica previste sono 5.

### Prerequisiti

- Familiarità con formule algebriche ed espressioni contenenti radicali
- Coordinate cartesiane nel piano
- Equazioni della retta e della parabola
- Traslazione
- Elementi di calcolo delle probabilità
- Statistica descrittiva
- Utilizzo del foglio elettronico per l'elaborazione di insiemi di dati più numerosi

### Obiettivi specifici

#### Conoscenze

- Conoscere il concetto di distribuzione congiunta
- Conoscere il concetto di regressione
- Conoscere il metodo dei minimi quadrati

#### Abilità

- Saper valutare l'eventuale dipendenza statistica di due caratteri attraverso diversi metodi
- Saper calcolare indici di contingenza
- Saper calcolare la retta di regressione utilizzando il metodo dei minimi quadrati
- Saper esprimere l'equazione della retta di regressione nella sua forma generale, anche utilizzando varianza e covarianza.

### Metodologia didattica

- Lezioni frontali e dialogiche che si basano sull'insegnamento per problemi: si cerca di stimolare gli allievi a formulare ipotesi ricorrendo a conoscenze possedute ma anche a intuizione e fantasia.
- Metodologie didattiche attive: discussioni guidate e lavoro di gruppo per cercare di coinvolgere maggiormente la classe e per abituare allo scambio di idee.
- Lezioni affrontate utilizzando il metodo induttivo: partire da esempi concreti della vita quotidiana per passare, gradualmente, alla formalizzazione dei concetti.
- Numerosi esempi per agevolare la comprensione di quanto svolto teoricamente.
- Esercizi svolti alla lavagna dall'insegnante o dagli alunni ed assegnazione di esercizi per casa.

- Utilizzo del laboratorio di informatica: in particolare si pensa all'aiuto del foglio elettronico (Excel). L'attività di laboratorio integra gli elementi di contenuto dei vari temi e costituisce essa stessa, un momento di riflessione teorica. Essa consisterà in:
  - analisi dei problemi e loro risoluzione informatica attraverso l'utilizzo di programmi disponibili e di software di utilità;
  - esplorazione e verifiche di proprietà matematiche, rappresentazioni grafiche e calcoli, simulazioni, come momenti costitutivi del processo di apprendimento della matematica e delle sue successive sistematizzazioni.

### **Strumenti**

- Lavagna, gessi
- Calcolatrice scientifica
- Libro di testo, fotocopie
- Carta millimetrata
- Matita, gomma, righello
- Personal computer con foglio di calcolo

### **Tempi di intervento didattico**

Per svolgere questa unità didattica si prevede un totale di 5 ore (comprese 2 ore per verifica e correzione della stessa). La previsione è da intendersi elastica, perché occorre tener conto delle necessità degli studenti.

### **Contenuti**

- Distribuzioni congiunte:
  - Distribuzione condizionata
  - Distribuzione marginale
- Indici di dipendenza:
  - Chi quadrato di Pizzetti Pearson
  - Indice di contingenza di Cramer
- Regressione e correlazione lineare:
  - Retta di regressione
  - Metodo dei minimi quadrati
  - Baricentro

### **Sviluppo dei contenuti**

#### **1° passo: distribuzione statistica congiunta**

Fino a questo momento abbiamo preso in considerazione e confrontato unità statistiche sulla base di un solo carattere.

E' possibile confrontare più unità statistiche sulla base simultanea di più caratteri? Come facciamo, ad esempio, a vedere quale sia la distribuzione dei voti di matematica tra gli studenti che hanno la media del 6 in matematica?

Siamo di fronte a un problema di distribuzione statistica congiunta.

Cominciamo, come al solito, presentando una tabella (a doppia entrata) di valori. Osservando la colonna parliamo di **distribuzione condizionata e marginale** e presentiamo qualche esempio di calcolo di frequenze.

Osserviamo che in una distribuzione congiunta si può verificare la dipendenza o l'assoluta indipendenza di un carattere dall'altro. Mostriamo tutto attraverso esempi!

La trattazione termina con la spiegazione di due **indici di dipendenza: indice chi quadrato di Pizzetti – Pearson e indice di contingenza di Cramer.**

NOTA: meglio riassumere attraverso uno schema i passi da fare per ottenere tali indici

## 2° passo: introduzione alla regressione

Di fronte alla necessità di stabilire la dipendenza di un carattere da un altro, la ricerca scientifica utilizza la statistica per vedere se esiste una funzione matematica che sia in grado di esprimere tale relazione.

Si introduce il concetto di **regressione** come quel processo di ricerca di un legame di dipendenza tra due caratteri quantitativi.

Spieghiamo che esistono diversi tipi di regressione ma noi ci occuperemo solo di quella **lineare**.

## 3° passo: retta di regressione lineare

Cominciamo presentando un esempio: allungamento di una molla in relazione alla massa ad essa applicata. Forniamo una tabella (a doppia entrata) e chiediamo di riportare i dati in un grafico, utilizzando la carta millimetrata (si può usare anche il computer).

Come sono distribuiti i punti? Non si distribuiscono tutti attorno a una retta?

Questa retta è detta **retta di regressione**.

OSSERVAZIONE

Facciamo dunque osservare l'importanza degli argomenti studiati l'anno precedente

A questo punto diciamo che un modo per determinare una legge di dipendenza che approssimi i nostri dati è **il metodo dei minimi quadrati**

OSSERVAZIONI

- vogliamo cercare di rendere minima la somma dei quadrati delle differenze tra dati reali e dati teorici
- Poiché il nostro scopo è quello di scrivere l'equazione di una retta dobbiamo determinare  $m$  e  $q$  della retta, come facciamo?

NOTA

Forse è il caso di ripetere velocemente le cose fondamentali sulla retta perché probabilmente sono state rimosse!

- Il problema si riduce a trovare quella funzione lineare che renda minima la misura di distanza data da  $\sum_{i=1}^n (y_i - \bar{y}_i)^2$ . DOBBIAMO SPIEGARE BENE CHE COSA SIANO  $y_i$  e  $\bar{y}_i$ .
- Cerchiamo di capire se il concetto è stato afferrato chiedendo: Quanto vale questa somma nel caso in cui i punti stiano esattamente sulla retta?

## 4° passo: calcolo della retta di regressione

Ma come si fa a trovare la retta?

- Definiamo il **baricentro** di una distribuzione doppia e affermiamo (ma non dimostriamo) che la retta di regressione (ottenuta con il metodo dei minimi quadrati) deve passare per il baricentro dei dati.

- Calcoliamo quindi il baricentro dei nostri dati (prima di X poi di Y)

NOTA

gli allievi devono ricordarsi come poter calcolare la media aritmetica

- Scriviamo l'equazione del fascio di rette passanti per il baricentro: la retta di regressione **deve** appartenere a quel fascio.
- Siamo ora di fronte al problema di far coincidere il baricentro con l'origine degli assi.

→RIPASSARE LA TRASLAZIONE

- E' proprio a questo punto che dobbiamo inserire la somma detta prima: il problema è quello di minimizzarla
- Facciamo notare che, inserendo i dati e svolgendo correttamente i conti, si ottiene una funzione polinomiale di secondo grado nella variabile  $m$ . Analiticamente, che cosa rappresenta una equazione di secondo grado?

→PARABOLA

### 5° passo: generalizziamo

E' arrivato il momento di generalizzare quanto visto con il nostro esempio; presentiamo dunque l'equazione della retta di regressione ottenuta con il metodo dei minimi quadrati.

#### OSSERVAZIONI

- Cerchiamo di far osservare che non abbiamo scritto niente di diverso da quello che abbiamo svolto pochi momenti prima
- Se si ritiene opportuno si possono ricordare le espressioni della **varianza** e della **covarianza** ed esprimere la retta di regressione utilizzando questi strumenti

Infine si presentano altri esercizi di applicazione.

### Verifica sommativa

Per la verifica sommativa di questo modulo si pensa di lasciare un'ora di tempo. Gli esercizi saranno simili a quelli svolti in classe.

Ad esempio:

- Si riporta una tabella di dati e si chiede di calcolare l'indice di contingenza quadratica media e discutere la dipendenza tra i caratteri.
- Riportiamo una tabella in cui compaiono 2 caratteri. Si chiede agli allievi di verificare la dipendenza lineare di una dall'altra e di determinare l'equazione della retta di regressione.

This document was created with Win2PDF available at <http://www.win2pdf.com>.  
The unregistered version of Win2PDF is for evaluation or non-commercial use only.  
This page will not be added after purchasing Win2PDF.