Lezione 8
Analisi delle relazioni tra due caratteri

PROF. ROBERTO COSTA

SCIENZE DELL'EDUCAZIONE - STATISTICA SOCIALE (305SF)

POLO UNIVERSITARIO FONDAZIONE PORTOGRUARO CAMPUS

Prima di cominciare

Appello anticipato 17/02/2025

Elaborati facoltativi

Contenuti: l'elaborato, in un numero contenuto di facciate (da 4 a 8), analizza un argomento (povertà, stili di vita, salute, istruzione, lavoro, relazioni sociali, ecc.) da un punto di vista statistico. È consigliabile il seguente approccio: ipotesi di partenza, dati a supporto dell'ipotesi, analisi dei dati, conclusioni.

Modalità: gli studenti potranno lavorare in piccoli gruppi (2/3 persone).

Supporto: una volta concluse le lezioni, sono a disposizione al lunedì dalle 14.30 alle 15.30 su appuntamento o via mail.

Consegna: entro venerdì 7/2/2025 via mail a roberto.costa@deams.units.it

Date degli appelli

Appelli d’esame	Data	Iscrizione dal
1° appello anticipato	17/02/2025	03/02/2025
1° appello sessione estiva	09/06/2025	26/05/2025
2° appello sessione estiva	23/06/2025	09/06/2025
3° appello sessione estiva	14/07/2025	30/06/2025
Appello unico sessione autunnale	08/09/2025	25/08/2025
1° appello sessione straordinaria	12/01/2026	29/12/2025

Dove eravamo rimasti

Nella scorsa lezione abbiamo approfondito il tema delle relazioni tra due variabili.

Abbiamo visto come, a seconda del tipo di variabili utilizzate, abbiamo diversi strumenti per misurare la relazione tra due variabili.

Abbiamo imparato a calcolare la covarianza e l'indice di correlazione di Pearson, partendo dai microdati di una variabile quantitativa.

Abbiamo poi visto come possiamo procedere nel caso di variabili qualitative, oppure in presenza di una tabella di contingenza.

Ci sono dei dubbi?

Misure di dipendenza

Come possiamo facilmente immaginare, nelle scienze sociali è pressoché impossibile trovarsi nelle situazioni di indipendenza o dipendenza perfetta.

Ci troveremo in situazioni di connessione intermedia, che dovremo misurare con appositi indici.

Il nostro punto di partenza sarà lo scarto tra il valore osservato e il valore teorico ($n^*_{ij}=\frac{n_{i.}n_{.j}}{N}$).

Chiameremo **contingenze** gli scarti tra valori osservati e teorici:

$$c_{ij}=n_{ij} - n^*_{ij}$$

Misure di dipendenza

Potrebbe funzionare la somma degli scarti tra valori teorici e osservati?

$\sum_{i=1}^{k}\sum_{j=1}^{h} c_{ij}=\sum_{i=1}^{k}\sum_{j=1}^{h} n_{ij} - \sum_{i=1}^{k}\sum_{j=1}^{h} n^{*}_{ij}$

Purtroppo no, perché la somma delle frequenze osservate è sempre pari a N, così come la somma delle frequenze teoriche.

$\sum_{i=1}^{k}\sum_{j=1}^{h} c_{ij}$= N – N = 0

Esercitiamoci

Abbiamo la seguente tabella con le frequenze relative della scuola media di origine e della scuola superiore scelta.

Scuola di orig./scuola scelta	Liceo	Istituto tecnico	Altro	Totale
Scuola A. Manzoni	10	50	5	65
Scuola B. Croce	15	65	10	90
Scuola C. Levi	10	30	5	45
Totale	35	145	20	200

Esercitiamoci

Calcoliamo le contingenze, passo dopo passo.

PASSO 1: Partiamo dal calcolo dei valori teorici $n^*_{ij} = \frac{n_{i.} n_{.j}}{N}$

Scuola di orig./scuola scelta	Liceo	Istituto tecnico	Altro	Totale
Scuola A	11,38	47,13	6,50	65
Scuola B	15,75	65,25	9,00	90
Scuola C	7,88	32,63	4,50	45
Totale	35	145	20	200

Indice di associazione χ^2 di Pearson

Karl Pearson costruisce un indice, noto come **indice di associazione del χ^2 di Pearson**, facendo ricorso ai quadrati delle contingenze, divise per la frequenza teorica.

La somma di questi valori è l'indice di associazione χ^2.

$$\chi^2 = \sum_{i=1}^{k} \sum_{j=1}^{h} \frac{c_{ij}^2}{n_{ij}^*}$$

Le proprietà dell'indice di associazione χ^2 di Pearson

L'indice di associazione del χ^2 di Pearson, gode delle seguenti proprietà:

- L'indice χ^2 è simmetrico. Non tiene conto della dipendenza (causa-effetto) e rimane invariato se scambiamo il ruolo di X e Y.
- È sempre non negativo $\chi^2 \geq 0$
- Assume valore 0 nel caso di indipendenza tra X e Y (associazione nulla)
- Assume valori prossimi allo 0 in caso di bassa associazione
- L'indice è tanto più grande quanto più ci si allontana dal caso di indipendenza
- A parità di associazione l'indice aumenta al crescere di N

L'ultima proprietà di χ^2 ne rappresenta anche il suo limite, poiché l'indice cresce, anche se l'associazione non cambia, se aumenta il collettivo osservato.

Esercitiamoci

n_{ij}

Scuola di orig./scuola scelta	Liceo	Istituto tecnico	Altro	Totale
Scuola A	10	50	5	65
Scuola B	15	65	10	90
Scuola C	10	30	5	45
Totale	35	145	20	200

n^*_{ij}

Scuola di orig./scuola scelta	Liceo	Istituto tecnico	Altro	Totale
Scuola A	11,38	47,13	6,50	65
Scuola B	15,75	65,25	9,00	90
Scuola C	7,88	32,63	4,50	45
Totale	35	145	20	200

PASSO 2: Calcoliamo le contingenze $c_{ij}=n_{ij} - n^*_{ij}$ e vediamo che la somma delle contingenze è uguale a 0.

Scuola di orig./scuola scelta	Liceo	Istituto tecnico	Altro
Scuola A	-1,38	2,88	-1,50
Scuola B	-0,75	-0,25	1,00
Scuola C	2,13	-2,63	0,50

$$c_{ij}=n_{ij} - n^*_{ij}$$

Esercitiamoci

Scuola di orig./scuola scelta	Liceo	Istituto tecnico	Altro
Scuola A	-1,38	2,88	-1,50
Scuola B	-0,75	-0,25	1
Scuola C	2,13	-2,63	0,50

Calcoliamo le contingenze, passo dopo passo.

PASSO 3: Calcoliamo l'indice $\chi^2 = \sum_{i=1}^{k} \sum_{j=1}^{h} \frac{c_{ij}^2}{n_{ij}^*}$

c_{ij}

c_{ij}^2

Scuola di orig./scuola scelta	Liceo	Istituto tecnico	Altro	Totale
Scuola A	1,89	8,27	2,25	12,41
Scuola B	0,56	0,06	1,00	1,63
Scuola C	4,52	6,89	0,25	11,66
Totale	6,97	15,22	3,50	25,69

$\frac{c_{ij}^2}{n_{ij}^*}$

Scuola di orig./scuola scelta	Liceo	Istituto tecnico	Altro	Totale
Scuola A	0,17	0,18	0,35	0,69
Scuola B	0,04	0,00	0,11	0,15
Scuola C	0,57	0,21	0,06	0,84
Totale	0,78	0,39	0,51	1,68

Esercitiamoci

L'indice $\chi^2 = \sum_{i=1}^{k} \sum_{j=1}^{h} \frac{c_{ij}^2}{n_{ij}^*}$ ovvero χ^2 è la sommatoria di tutte le contingenze al quadrato divise per le frequenze relative teoriche.

Scuola di orig./scuola scelta	Liceo	Istituto tecnico	Altro	Totale
Scuola A	0,17	0,18	0,35	0,69
Scuola B	0,04	0,00	0,11	0,15
Scuola C	0,57	0,21	0,06	0,84
Totale	0,78	0,39	0,51	1,68

Contingenza quadratica media

Dal momento che l'indice χ^2 aumenta al crescere del collettivo osservato, è opportuno utilizzare un indice che non dipenda da N.

Sempre Karl Pearson ha proposto un altro indice Φ^2 (Phi quadro), o **contingenza quadratica media**.

L'indice si calcola così: $\Phi^2 = \frac{\chi^2}{N}$

Che può essere calcolato anche nel seguente modo, che non richiede il calcolo delle frequenze teoriche.

$$\Phi^2 = \sum_{i=1}^{k} \sum_{j=1}^{h} \frac{n_{ij}^2}{n_{i.} n_{.j}} - 1$$

Le proprietà dell'indice Φ^2 (Phi quadro)

Le proprietà dell'indice Φ^2 sono le seguenti:

1. È un indice di dipendenza simmetrico
2. È sempre non negativo: $\Phi^2 \geq 0$
3. Assume valore 0 nel caso di indipendenza tra X e Y (associazione nulla)
4. Il valore massimo che può assumere è pari al valore più piccolo tra il numero di righe della tabella -1 (k – 1) e il numero di colonne della tabella – 1 (h – 1).

$$\max \Phi^2 = \min [(k - 1); (h - 1)]$$

5. Di conseguenza assume valore pari a 1 nel caso in cui il numero di righe o il numero di colonne sia pari a 2, altrimenti assumerà valore maggiore di 1.

Esercitiamoci

Calcoliamo l'indice $\Phi^2 = \frac{\chi^2}{N}$.

χ^2 = 1,68 N = 200

$\Phi^2 = \frac{1,68}{200}$ = 0,008

Il valore è prossimo allo 0, quindi siamo in presenza di bassa associazione.

Scuola di orig./scuola scelta	Liceo	Istituto tecnico	Altro	Totale
Scuola A	0,17	0,18	0,35	0,69
Scuola B	0,04	0,00	0,11	0,15
Scuola C	0,57	0,21	0,06	0,84
Totale	0,78	0,39	0,51	1,68

L’indice V di Cramer

Abbiamo visto che la contingenza quadratica media ha come minimo 0 (associazione nulla) e come massimo il valore più piccolo tra il numero di righe – 1 e il numero di colonne – 1.

$$\max \Phi^2 = \min [(k - 1); (h - 1)]$$

Se vogliamo ottenere un indice che vari tra 0 e 1, dovremo rapportare il valore di Φ^2 con il suo massimo.

L’indice normalizzato (che varia tra 0 e 1) più utilizzato è **l’indice V di Cramer** che si ottiene dalla radice quadrata del rapporto tra Φ^2 e il suo valore massimo.

$$V = \sqrt{\frac{\Phi^2}{\min [(k - 1); (h - 1)]}}$$

Esercitiamoci

Calcoliamo l'indice $V = \sqrt{\dfrac{\Phi^2}{\min[(k-1);(h-1)]}}$.

$\Phi^2 = \frac{1,68}{200}$ = 0,008

min [(k - 1); (h - 1)]

k - 1 = 3 – 1 = 2 (numero di colonne)

h - 1 = 3 – 1 = 2 (numero di righe)

min [(k - 1); (h - 1)] = 2

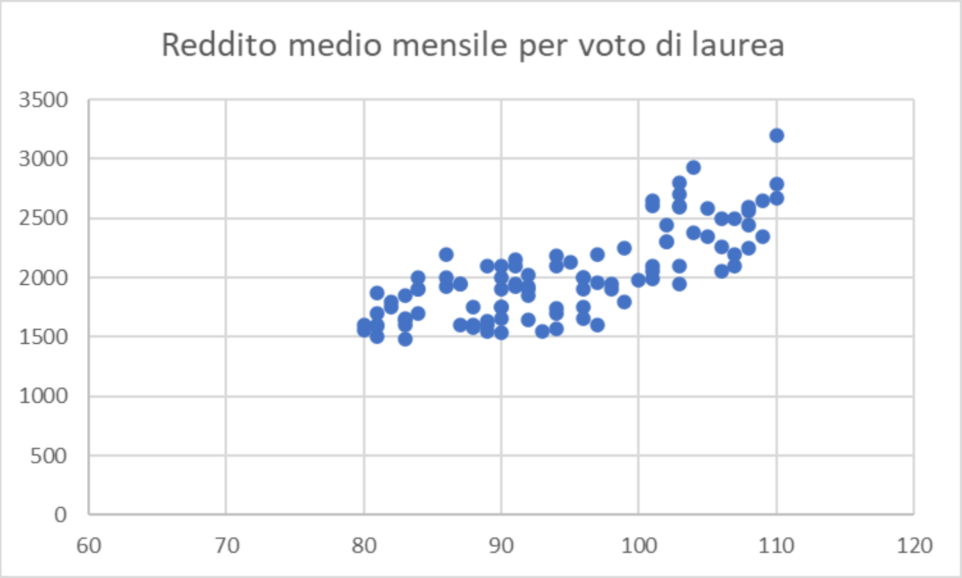
V = $\sqrt{\frac{0,008}{2}}$ = 0,065

Il valore è prossimo allo 0, quindi siamo in presenza di bassa associazione.

La regressione

Partiamo dal comprendere che cosa significa, in statistica, il termine **regressione**.

La regressione è un indicatore statistico che indica l'esistenza o meno di una relazione significativa tra due (analisi bivariata) o più variabili (analisi multivariata) quantitative.



Cosa intendiamo per regressione

In statistica, il termine regressione è stato utilizzato per la prima volta dal biologo inglese Francis Galton nel 1886, quando parlò di «regressione verso la media».

Nell'ambito dei suoi studi sull'ereditarietà dei caratteri, Galton raccolse le stature di 928 figli adulti e dei loro 205 genitori (maschi e femmine). Esaminando le altezze di genitori e figli, notò una relazione tra le due variabili: più alti erano i genitori, più alti erano i figli e viceversa.

Partendo dalla statura media dei genitori ('mid-parent's stature') scoprì che i figli più alti della media avevano genitori ancora più alti di loro e i figli più bassi della media avevano genitori ancora più bassi. A questo fenomeno diede il nome di regressione verso la media.

Ripartiamo da un esempio

Partiamo da un caso concreto con due variabili (dipendente e indipendente) di tipo quantitativo.

Possiamo chiederci che relazione esiste tra il numero di ore dedicate alla preparazione di un esame e il voto conseguito all’esame stesso.

Oppure potremmo chiederci se esiste una relazione tra il voto conseguito all’esame di laurea e il reddito mensile medio.

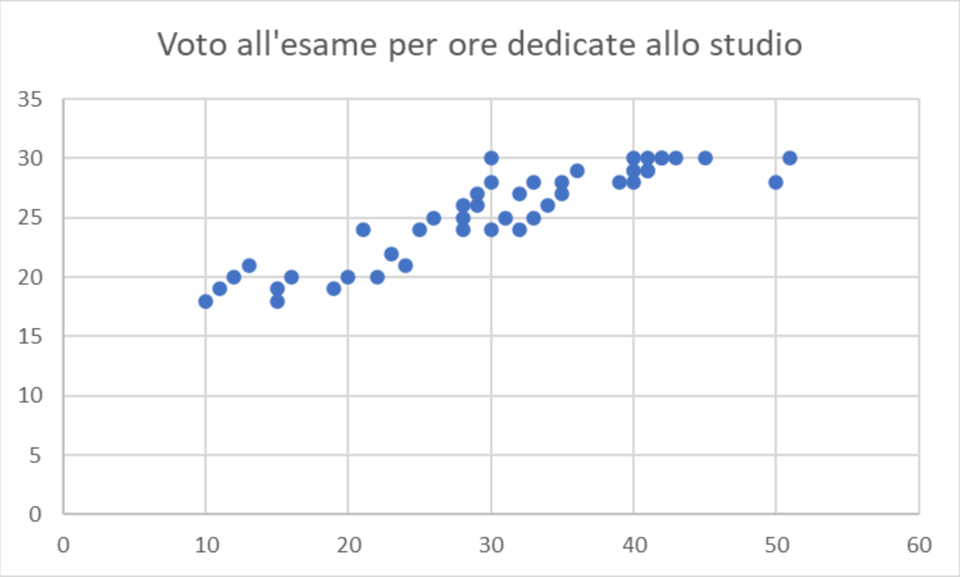
In entrambi i casi abbiamo definito una variabile dipendente (il voto all’esame e il reddito mensile medio) e una variabile indipendente (il numero di ore dedicate alla preparazione di un esame e il voto conseguito all’esame di laurea).

Il diagramma di dispersione

Il **diagramma di dispersione** è la rappresentazione grafica di una possibile relazione tra due variabili.

Sull'asse X troviamo la variabile indipendente e sull'asse Y la variabile dipendente.

L'insieme dei punti che si crea indica come covariano (variano insieme) le due variabili.



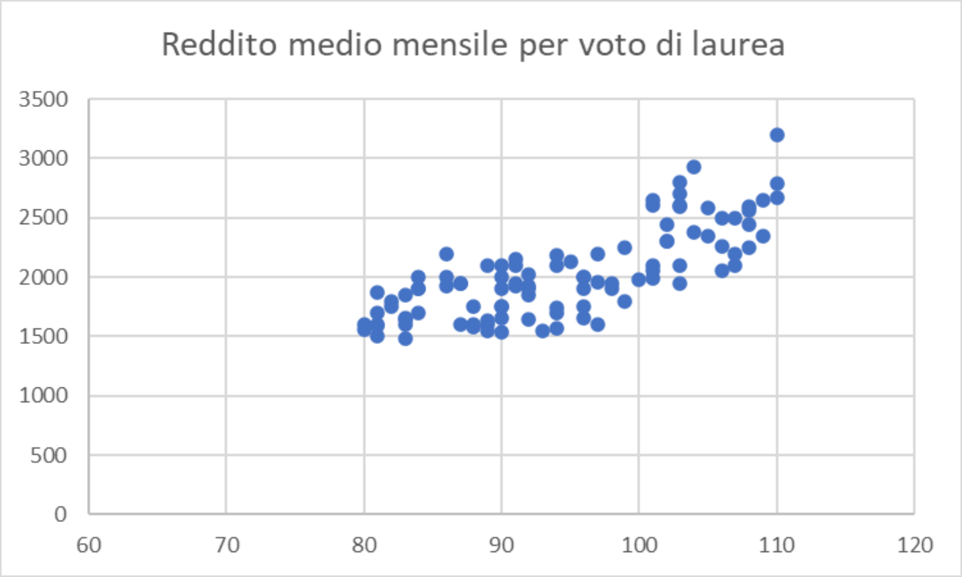
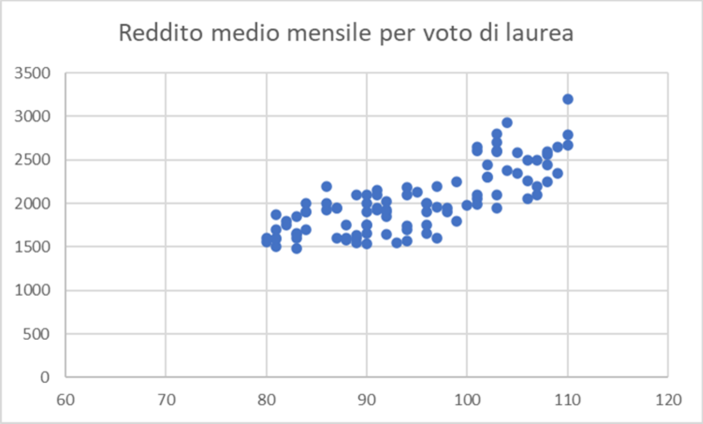


Diagramma di dispersione

L'osservazione del diagramma di dispersione ci consente di trarre alcune conclusioni:

- Le due variabili covariano in modo sistematico, ovvero sono legate da una relazione,
- La distribuzione della «nuvola di punti» dalla sinistra in basso alla destra in alto ci indica che la relazione è positiva, ovvero al crescere della variabile indipendente cresce anche la variabile dipendente.
- La disposizione dei punti ci suggerisce che siamo in presenza di una relazione lineare, ovvero la variabile Y tende a variare sempre nella stessa direzione e nella stessa misura al variare di X.



Non ci aiuta a misurare l'effetto causale, ovvero a quantificare la variazione della variabile dipendente al variare di quella indipendente.

Equazione lineare

Se vogliamo quantificare l’intensità della relazione tra le due variabili, la dobbiamo esprimere attraverso un’equazione matematica.

Ogni equazione è definita dalla sua **forma funzionale** e dai valori che assumono i suoi **parametri**.

Parlando di forma funzionale prenderemo in considerazione solo quella lineare, che possiamo esprime semplicemente così:

$$Y = \alpha + \beta X$$

Il valore di Y è dato dal parametro α, che è costante, più X moltiplicato per il parametro β.

La seguente è invece la funzione di una parabola:

$$Y = \alpha X^2 + \beta X + \gamma$$

Equazione lineare

Facciamo un esempio concreto:

Ho i due parametri α e β

$\alpha = 3$

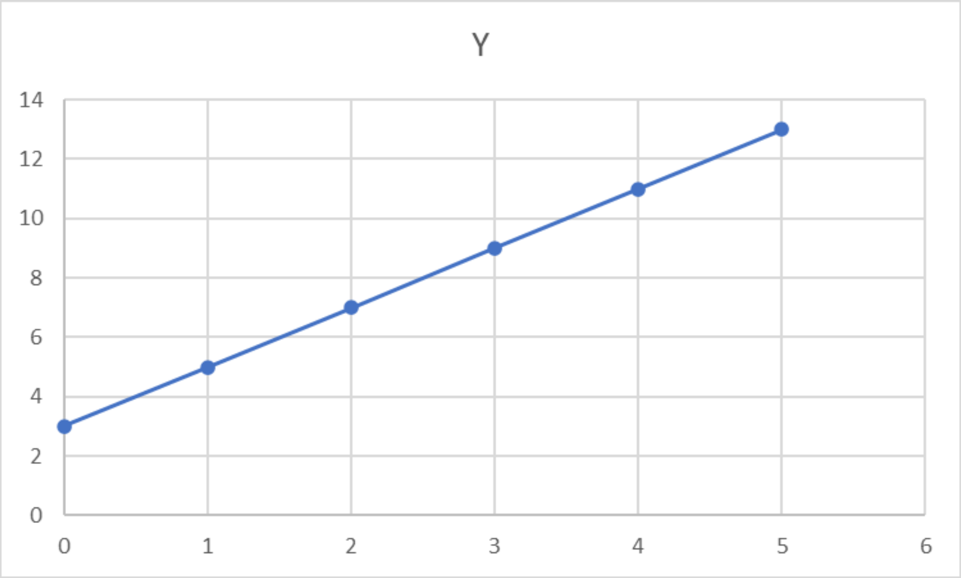
$\beta = 2$

La nostra equazione lineare $Y = \alpha + \beta X$

diventa $Y = 3 + 2X$

X	Y
0	3
1	5
2	7
3	9
4	11
5	13

Equazione lineare



X	Y
0	3
1	5
2	7
3	9
4	11
5	13

Equazione lineare

Facciamo un esempio concreto

$\alpha = 5$

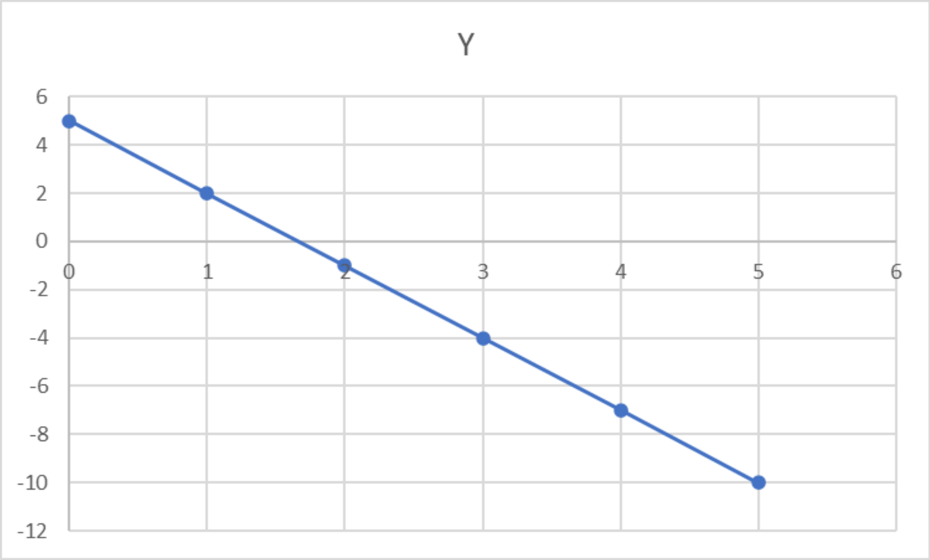
$\beta = -3$

La nostra equazione lineare $Y = \alpha + \beta X$

diventa $Y = 5 - 3X$

X	Y
0	5
1	2
2	-1
3	-4
4	-7
5	-10

Equazione lineare



X	Y
0	5
1	2
2	-1
3	-4
4	-7
5	-10

Equazione lineare

Abbiamo espresso la relazione tra X e Y con una linea retta.

Come influiscono i due parametri α e β?

α stabilisce la distanza dall'asse orizzontale, ovvero il valore di Y in corrispondenza dello 0 della X. Questo parametro viene definito anche come **intercetta o costante**.

β determina **l'inclinazione o coefficiente angolare** e ci dice di quanto varia la Y al variare di X. Questo valore ci spiega l'intensità dell'effetto della variabile indipendente sulla variabile dipendente.

Se β è positivo ci troviamo di fronte a una relazione diretta, mentre se è negativo la relazione è inversa.

Equazione lineare

Abbiamo espresso la relazione tra X e Y con una linea retta $Y = \alpha + \beta X$.

Come abbiamo ripetuto più volte, nelle scienze sociali non è possibile rappresentare esattamente con un'equazione lineare una relazione complessa, come, nel nostro esempio, quella tra reddito e voto di laurea.

Dalle immagini che abbiamo visto in precedenza, al medesimo voto di laurea possono corrispondere diversi livelli di reddito medio.

La relazione tra due variabili non può essere rappresentata esattamente da un'equazione lineare, tuttavia non possiamo negare che la nuvola di punti ci suggerisca una tendenza precisa, ovvero al crescere delle ore di studio aumenta il voto all'esame, oppure al crescere del voto di laure corrisponda un reddito medio più elevato.

Un'equazione lineare ci aiuta a stimare i due parametri α e β, e ad approssimare la covariazione tra le due variabili.

Modello di regressione lineare semplice

La formula che approssima la relazione tra X e Y con una linea retta è la seguente:

$\hat{Y}_i = \alpha + \beta X_i$

$\hat{Y}_i$ rappresenta il **valore atteso** sulla base dei parametri stimati α e β e non quello osservato.

Se vogliamo esprimere i valori osservati di Y dobbiamo aggiungere all'equazione lineare un ulteriore elemento ε_i, che rappresenta la componente erratica (gli errori di predizione).

$Y_i = \alpha + \beta X_i + \varepsilon_i$

La componente ε_i rappresenta la differenza tra il valore osservato e il valore atteso, derivante dal modello di regressione lineare, ovvero:

$\varepsilon_i = Y_i - \hat{Y}_i$

se sostituiamo $\hat{Y}_i$ con $\alpha + \beta X_i$ (visto che $\hat{Y}_i = \alpha + \beta X_i$) otteniamo che:

$\varepsilon_i = Y_i - \alpha - \beta X_i$

I residui

Gli errori di predizione vengono chiamati **residui**, dal momento che corrispondono a quella parte di Y che non viene spiegata dall'effetto lineare di X.

ε_i rappresenta:

- l'influenza su Y di tutti i fattori casuali che non sono stati introdotti nel modello di regressione lineare utilizzato.
- Il fatto che la relazione tra X e Y non è detto sia perfettamente lineare.
- Nelle scienze sociali i comportamenti umani sono di consueto caratterizzati da una componente di casualità che nessun modello di regressione, anche il più sofisticato, sarebbe in grado di stimare con precisione il valore di Y.

Scelta della retta di regressione

Abbiamo chiarito che l’obiettivo della regressione lineare è stimare i valori dei parametri α e β che consentono di approssimare nel modo migliore la covariazione tra X e Y.

Questo significa che la retta di regressione migliore è quella che minimizza i valori osservati di Y e quelli predetti attraverso il modello.

Dal momento che la differenza tra valori attesi e valori osservati:

$\varepsilon_i = Y_i - \hat{Y}_i$

La migliore retta di regressione è quella che minimizza gli errori di predizione.

Scelta della retta di regressione

Se proviamo a sommare gli errori da una determinata retta scopriremo che si annullano, ovvero

$\sum_{i=1}^{N} \varepsilon_i = 0$

Come abbiamo già visto in altre situazioni, possiamo elevare al quadrato gli scarti e la migliore retta di regressione è quella che minimizza il quadrato della somma dei residui.

Questo significa che rende minima la seguente quantità:

$\sum_{i=1}^{N} (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{N} \varepsilon_i^2$

I valori dei parametri α e β che soddisfano questo criterio sono detti **stime dei minimi quadrati**.

Metodo dei minimi quadrati

Tralasciamo la dimostrazione e vediamo qual è la formula che ci consente di stimare i due parametri.

$$\beta = \frac{\sum_{i=1}^{N}(x_i - M(X))(y_i - M(Y))}{\sum_{i=1}^{N}(x_i - M(X))^2} = \frac{Codev\ (X,Y)}{Dev\ (X)}$$

$\alpha = M(Y) - \beta\ M(X)$

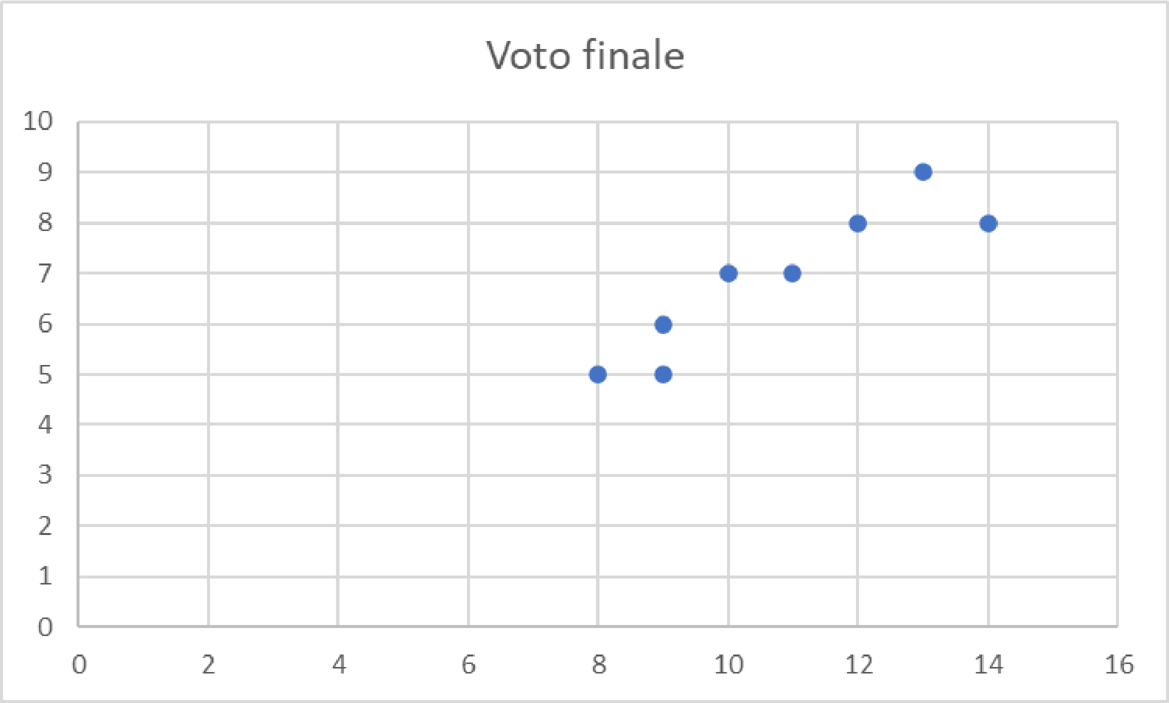
Esercitiamoci

(Esempio di pag. 150 del libro) Abbiamo due variabili misurate su 8 studenti: voto del test di ingresso a inizio anno scolastico e voto finale di matematica.

Studente	Test ingresso	Voto finale
1	12	8
2	10	7
3	14	8
4	9	5
5	9	6
6	13	9
7	11	7
8	8	5

Esercitiamoci

Come prima cosa costruiamoci uno scatterplot (diagramma di dispersione), per vedere come si dispongono sul piano cartesiano i punti individuati dalle coppie di valori.



Esercitiamoci

Dalla rappresentazione grafica possiamo individuare una andamento lineare positivo.

Procediamo quindi con il calcolo dei vari indicatori che ci servono:

M(X) = 10,75 M(Y) = 6,88

Var(X) = 3,94 Var(Y) = 1,86

Dev(X) = 31,50 Dev(Y) = 14,88

Codev(X;Y) = 19,75

Cov(X;Y) = Codev(X;Y)/N = 2,47

Esercitiamoci

Possiamo calcolare il valore di ρ coefficiente di correlazione lineare di Pearson.

$$\rho = \frac{Codev(X,Y)}{\sqrt{Dev(X)Dev(Y)}}$$

$$\rho = \frac{19,75}{\sqrt{31,50*14,88}} = \frac{19,75}{21,65} = 0,91$$

Il coefficiente ρ pari a 0,91 indica che c'è correlazione positiva tra le due variabili.

Esercitiamoci

Possiamo calcolare il parametri della retta di regressione.

β = $\frac{Codev(X,Y)}{Dev(X)} = \frac{19,75}{31,50}$ = 0,63

α = M(Y) - β M(X) = 6,88 – (0,63 * 10,75) = 0,13

Esercitiamoci

La nostra retta di regressione è la seguente:

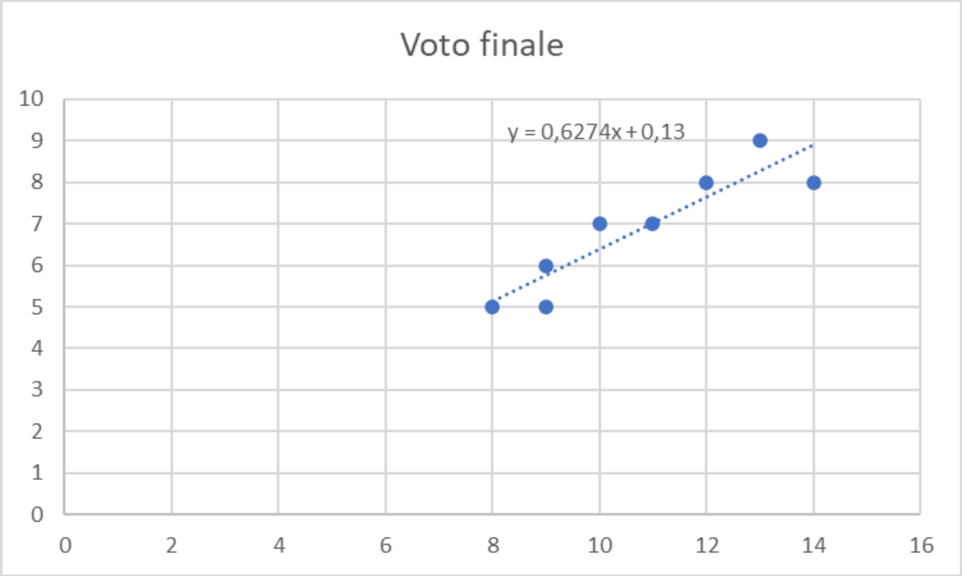
$\hat{Y}_i = \alpha + \beta X_i$

$\hat{Y}_i = 0{,}13 + 0{,}63X_i$

Possiamo costruire la retta di regressione con due punti:

l'intercetta (ovvero il valore di $\hat{Y}_i$ con $\beta = 0$) che corrisponde a (0 e 0,13) e

il valore individuato da M(X) e M(Y), ovvero (10,75 e 6,88).



La bontà del modello

Come abbiamo visto, il metodo dei minimi quadrati ci garantisce che la retta individuata sia la migliore possibile.

Questo però non ci assicura che la retta sia il miglior modello per rappresentare i dati.

Come possiamo capire se la retta di regressione è adatta a rappresentare i dati:

- Calcoliamo un apposito indice;
- Analizziamo graficamente i residui.

L’indice di determinazione R^2

Dal punto di vista teorico possiamo dire che un modello di regressione è tanto migliore quanto i valori della Y e quelli ottenuti con la retta di regressione hanno una correlazione vicina a 1.

L’indice di determinazione, detto anche coefficiente **R^2** è il quadrato del coefficiente di correlazione lineare fra Y e $\hat{Y}$.

$R^2 = [\rho(Y;\hat{Y})]^2$

L’indice di determinazione varia tra 0 e 1.

È pari a 1 quando la variabilità totale di Y è totalmente spiegata dalla retta di regressione.

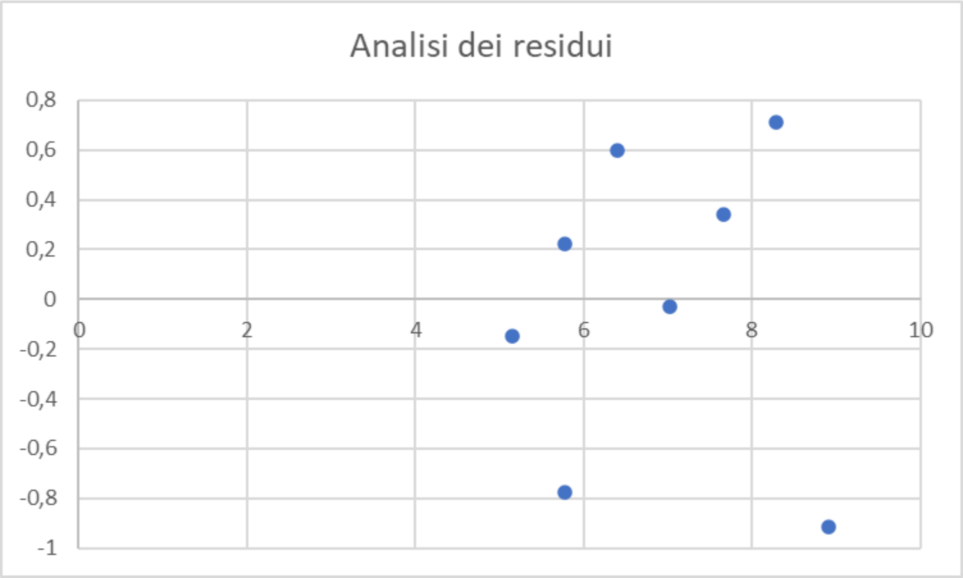
È pari a 0 quando la variabilità totale di Y non è per nulla spiegata dalla retta di regressione.

Analizzare graficamente i residui

Affinché la retta di regressione possa essere considerata una buona approssimazione della relazione tra X e Y, i residui devono avere un andamento casuale rispetto ai valori della X.

Ad esempio se all'aumentare dei valori di X crescessero sistematicamente anche i residui, ci potremmo trovare in presenza di una relazione non lineare e quindi la retta di regressione non è il modello più adeguato.

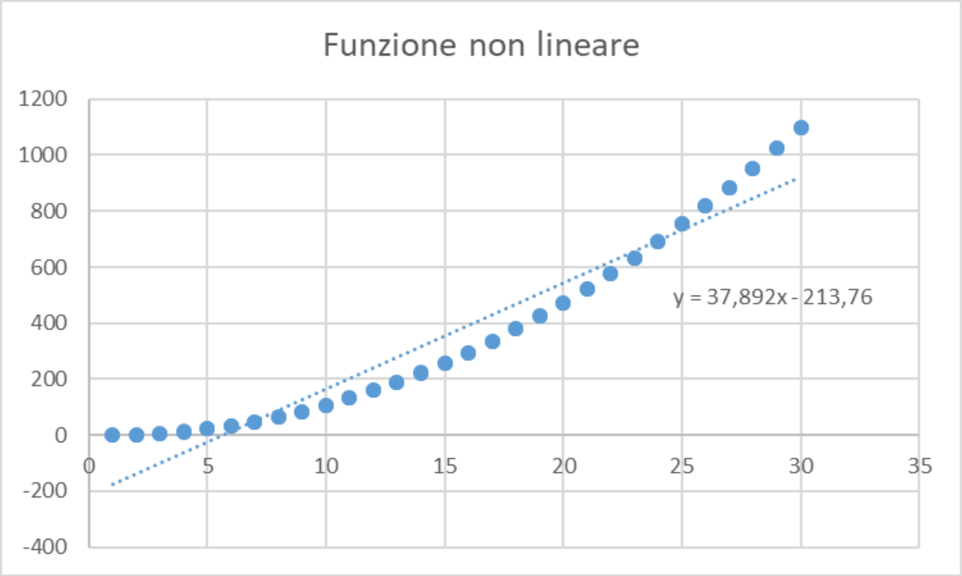
Dovremo quindi costruire un diagramma di dispersione con i valori di x_i e i relativi residui.

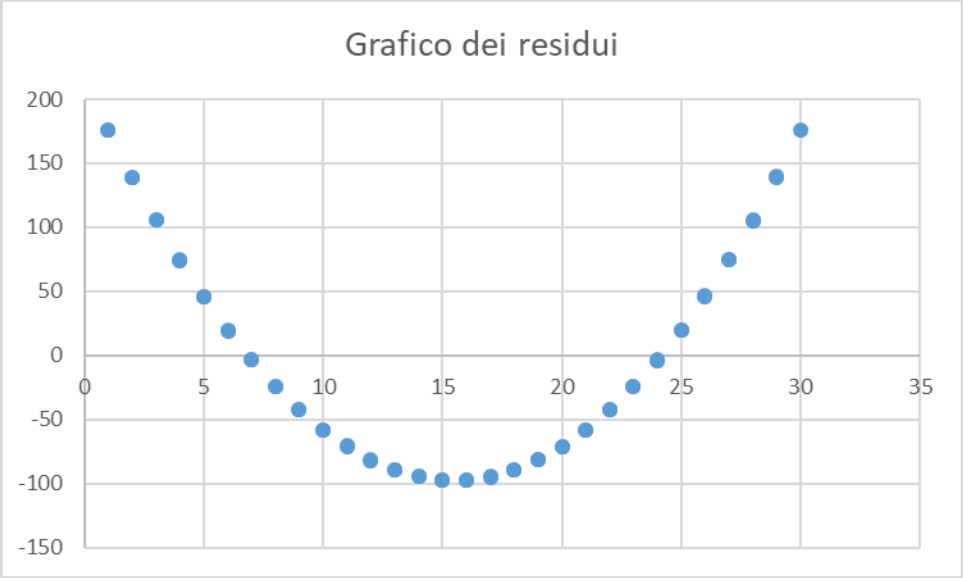


Analizzare graficamente i residui

Partiamo da un funzione non lineare e calcoliamo la nostra retta di regressione.

Dal punto di vista grafico la retta di regressione non sembra così inadeguata, ma guardando il grafico dei residui possiamo vedere che questi non si distribuiscono in modo causale.





Come sarà l’esame

Domande legate alla conoscenza teorica:

1 - I **microdati** sono:

A. l’informazione trattata ad un livello minimo, dati grezzi
B. Una sintesi di dati aggregati
C. Informazioni aggiuntive che aiutano a contestualizzare i dati prodotti

2 - Una **popolazione o collettivo di stato** è:

A. Una popolazione definibile precisando un unico istante di tempo
B. Una popolazione definibile in un intervallo di tempo

Come sarà l’esame

Domande legate all’applicazione delle conoscenze teoriche:

3 - Il numero di **abitanti di un comune al 1.1.2022** è un esempio di:

A. Popolazione o collettivo di stato
B. Popolazione o collettivo di movimento

4 - Il numero di **laureati in Scienze dell’educazione nell’anno accademico 2021/22** è un esempio di:

A. Popolazione o collettivo di stato
B. Popolazione o collettivo di movimento

Come sarà l’esame

Domande legate alla conoscenza teorica:

5 - Si definisce **popolazione empirica**:

A. Se tutte le unità che la compongono possono entrare a far parte di un campione
B. Se alcune delle sue unità non possono essere effettivamente osservate

Domande legate all’applicazione delle conoscenze teoriche

6 - Gli **abbonati ad una rivista online** ad una certa data è un esempio di:

A. Popolazione empirica
B. Popolazione teorica

Come sarà l’esame

Domande legate alla conoscenza teorica:

7 - Si definisce **rapporto di composizione**:

A. Il rapporto tra l’ammontare di una modalità e l’ammontare complessivo
B. Il rapporto tra l’ammontare di una modalità e quello di un’altra modalità della stessa variabile
C. Il rapporto tra l’ammontare di un fenomeno e quella di un altro che può essere considerato il suo presupposto logico

Domande legate all’applicazione delle conoscenze teoriche

8 - Il **tasso di attività** (rapporto percentuale tra le forze di lavoro e la corrispondente popolazione di riferimento) è un esempio di:

A. Rapporto di derivazione
B. Rapporto di coesistenza
C. Rapporto di composizione

Come sarà l’esame - soluzioni

1 - I **microdati** sono:

l’informazione trattata ad un livello minimo, dati grezzi

2 - Una **popolazione o collettivo di stato** è:

Una popolazione definibile precisando un unico istante di tempo

3 - Il numero di **abitanti di un comune al 1.1.2022** è un esempio di:

Popolazione o collettivo di stato

4 - Il numero di **laureati in Scienze dell’educazione nell’anno accademico 2021/22** è un esempio di:

Popolazione o collettivo di movimento

Come sarà l’esame - soluzioni

5 - Si definisce **popolazione empirica**:

Se tutte le unità che la compongono possono entrare a far parte di un campione

6 - Gli **abbonati ad una rivista online** ad una certa data un esempio di:

Popolazione empirica

7 - Si definisce **rapporto di composizione**:

Il rapporto tra l’ammontare di una modalità e l’ammontare complessivo

8 - Il **tasso di attività** (rapporto percentuale tra le forze di lavoro e la corrispondente popolazione di riferimento) è un esempio di:

Rapporto di composizione

Esercitiamoci

Studente	Inglese	Sociologia	Storia soc.	Geografia	Pedagogia
Voti agli esami	27	28	30	25	30

Calcoliamo i seguenti valori centrali:

media ____

mediana ____

media aritmetica ____

Esercitiamoci

Studente	Inglese	Sociologia	Storia soc.	Geografia	Pedagogia
Voti agli esami	27	28	30	25	30

Calcoliamo i seguenti valori di disuguaglianza:

range ____

scarto quadratico medio ____

varianza ____