



Statistica Descrittiva

Lo studio della relazione tra variabili

Michela Gnaldi

Dipartimento di Scienze Politiche

Università di Perugia

E-mail: michela.gnaldi@unipg.it



Outline

Lo studio della relazione tra due variabili:

1. **Esiste** una relazione tra due variabili? Il CHI-QUADRATO
2. Che segno ha la relazione tra variabili? La COVARIANZA
3. Che forza ha la relazione tra variabili? La CORRELAZIONE
4. Che misura ha la relazione tra variabili? La REGRESSIONE (prossima lezione)



Dati utilizzati

- ❑ Dati Invalsi rilevati a giugno 2009 sugli studenti di terza media
- ❑ Sono considerate solo le scuole/classi con almeno 20 studenti partecipanti alla rilevazione
- ❑ 3774 studenti provenienti da 171 scuole
- ❑ Test di italiano: 30 domande di comprensione del testo e 10 domande di grammatica
- ❑ Test di matematica: 27 domande
- ❑ Tutte le domande sono a scelta multipla con 1 risposta corretta e 3 sbagliate; la codifica usata è 1 per la risposta corretta e 0 per quella sbagliata
- ❑ Informazioni aggiuntive disponibili: sesso dello studente, scuola di provenienza, area geografica sede della scuola



1. Verificare se c'è una relazione tra variabili: il CHI-quadrato

Con i nostri dati, possiamo verificare se esiste una relazione tra

1. **performance scolastiche e genere**; ad es. nelle prove internazionali come PISA o TIMMS, si osserva che le studentesse tendono a fare meglio dei maschi nelle prove di lettura, mentre i maschi tendono ad avere risultati migliori in matematica
2. **performance scolastiche e provenienza geografica**; ad ed. da precedenti studi emerge che gli studenti del Nord-Est sono i più bravi nelle prove INVALSI



Relazione tra performance scolastiche e genere

Tabella: Distribuzione di frequenza secondo il punteggio totale al Test INVALSI di Italiano e il genere degli studenti

Punteggio totale	Sesso		Totale
	M	F	
0-7	6	4	10
8-15	167	123	290
16-23	746	709	1455
24-30	955	1064	2019
Totale	1874	1900	3774

TABELLA OSSERVATA
(dati veri)

Frequenze di associazione: n_{ij}



Relazione tra performance scolastiche e genere

Tabella: Distribuzione di frequenza TEORICA (tra punteggio e genere c'è INDIPENDENZA)

Punteggio totale	Sesso		Totale
	M	F	
0--7	4,97	5,03	
8--15	144,00	146,00	
16--23	722,49	732,51	
24--30	1002,55	1016,45	
Totale			

TABELLA TEORICA

Frequenze “teoriche”, cioè le frequenze che si osserverebbero se tra punteggio e genere ci fosse indipendenza: n_{ij}'



Relazione tra performance scolastiche e genere

Tabella: Tabella per il calcolo del CHI-quadrato

Punteggio totale	Sesso		Totale
	M	F	
0--7	0,22	0,21	
8--15	3,67	3,62	
16--23	0,77	0,75	
24--30	2,25	2,22	
Totale			13,72


$$(n_{ij} - n_{ij}')^2 / n_{ij}'$$


$$\text{CHI-quadrato} = \text{Somma } (n_{ij} - n_{ij}')^2 / n_{ij}'$$



Relazione tra performance scolastiche e genere

L'analisi dell'associazione tra due variabili, attraverso il CHI-quadrato, è il primo necessario step nell'analisi della relazione tra variabili, per verificare se c'è o meno una relazione

Tuttavia, il CHI-quadrato ci consente solo di dire se tra due variabili

- c'è relazione: $\text{CHI-quadrato} > 0$
- non c'è relazione: $\text{CHI-quadrato} = 0$



2. Verificare il verso/direzione della relazione tra variabili: la COVARIANZA

Verificare il verso/direzione della relazione tra variabili significa verificare se tra due variabili c'è una **relazione positiva, negativa o nulla**

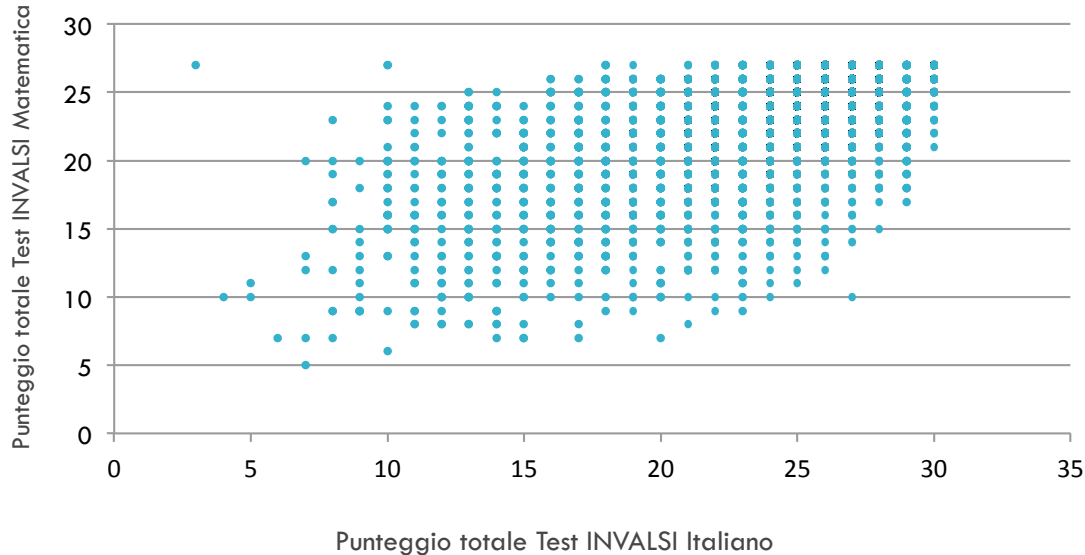
Con i dati INVALSI, possiamo verificare se tra i risultati conseguiti dagli studenti nel test di italiano e risultati nel test di matematica c'è una relazione positiva, negativa o nulla

- ◆ Se la relazione è positiva, gli studenti che ottengono i risultati migliori nel test di italiano, tendono anche ad avere i punteggi più alti nel test di matematica
- ◆ Invece, se la relazione è negativa, migliori sono i risultati in italiano, peggiori i risultati in matematica
- ◆ Infine, se non c'è relazione, i risultati conseguiti dagli studenti in uno dei due test non sono legati ai risultati conseguiti nell'altro test



Relazione tra punteggio al test di italiano e punteggio al test di matematica

Distribuzione dei punteggi ai test INVALSI di Italiano e Matematica

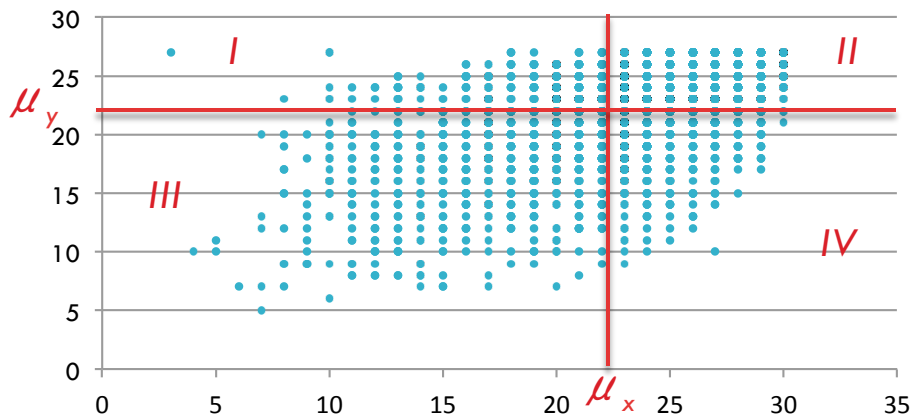


Relazione positiva



Relazione tra punteggio al test di italiano e punteggio nel test di matematica

Distribuzione dei punteggi ai test INVALSI di Italiano e Matematica



Quadrante I:

somma di $(x_i - \mu_x) < 0$

somma di $(y_i - \mu_y) > 0$

Quadrante III:

somma di $(x_i - \mu_x) < 0$

somma di $(y_i - \mu_y) < 0$

Quadrante II:

la somma di $(x_i - \mu_x) > 0$

la somma di $(y_i - \mu_y) > 0$

Quadrante IV:

la somma di $(x_i - \mu_x) > 0$

la somma di $(y_i - \mu_y) < 0$



Relazione tra punteggio al test di italiano e punteggio nel test di matematica

Riassumendo:

1. Se la maggior parte dei punti della nuvola dei punti si posiziona sul quadrante II e III 
somma di $(x_i - \mu_x) * (y_i - \mu_y) > 0$ e  **relazione positiva**
2. Se la maggior parte dei punti della nuvola dei punti si posiziona sul quadrante I e IV 
somma di $(x_i - \mu_x) * (y_i - \mu_y) < 0$ e  **relazione negativa**

La COVARIANZA esprime la tendenza a covariare di due variabili:

$$\text{COV}(X;Y) = \text{somma } (x_i - \mu_x) * (y_i - \mu_y) / N$$

Se:

- $\text{COV}(X;Y) > 0$ allora relazione positiva
- $\text{COV}(X;Y) < 0$ allora relazione negativa
- $\text{COV}(X;Y) = 0$ allora relazione nulla



Relazione tra punteggio al test di italiano e punteggio nel test di matematica

- Quello che ci interessa della COVARIANZA non è tanto il valore numerico che assume, quanto il **segno**
- La COVARIANZA ci da indicazioni sul SEGNO o DIREZIONE della relazione tra variabili
- Nell'esempio che utilizza i dati INVALSI: $COV(X;Y) = 37551,87 / 3774 = + 9,95$
- Tra punteggio al test di italiano e punteggio al test di matematica c'è una **relazione positiva**. Quindi, gli studenti che ottengono i risultati migliori nel test di italiano, tendono anche ad avere i punteggi più alti nel test di matematica
- Ancora però non sappiamo se questa relazione positiva è forte, perché **la covarianza non è un indice relativo**

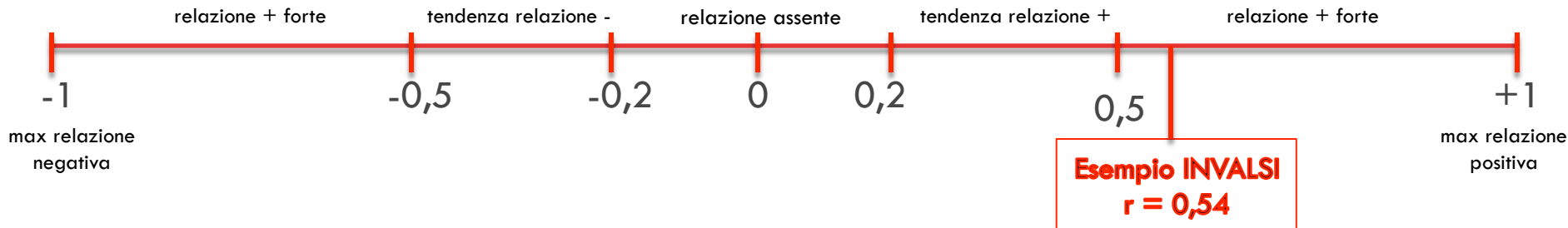


Relazione tra punteggio al test di italiano e punteggio nel test di matematica

La **forza/intensità** della relazione tra variabili si verifica attraverso il **Coefficiente di correlazione di Bravais-Pearson**

$$r = \text{COD}(X;Y) / \sqrt{\text{DEV}(X) \text{ DEV}(Y)}$$

L'indice è relativo e assume valori compresi tra $-$ e $+$ 1. Convenzionalmente:





Relazione tra punteggio al test di italiano e punteggio nel test di matematica

- Nell'esempio che utilizza i dati INVALSI: $r = 0,54$ quindi tra risultati al test di italiano e al test di matematica c'è una **forte relazione positiva**. La relazione osservata è il 54% di quella massima positiva osservabile
- Il coefficiente di correlazione non ci dice nulla sulla **misura della relazione**, ad es. non ci dice **di quanto** aumenta il punteggio al test di matematica al variare del punteggio di italiano. Ci dice che solo che se il punteggio in italiano aumenta, allora aumenta anche quello in matematica e che questa relazione è forte
- La misura della relazione tra variabili è fornita dal **coefficiente di regressione**... prossima lezione!



Grazie!

michela.gnaldi@unipg.it