

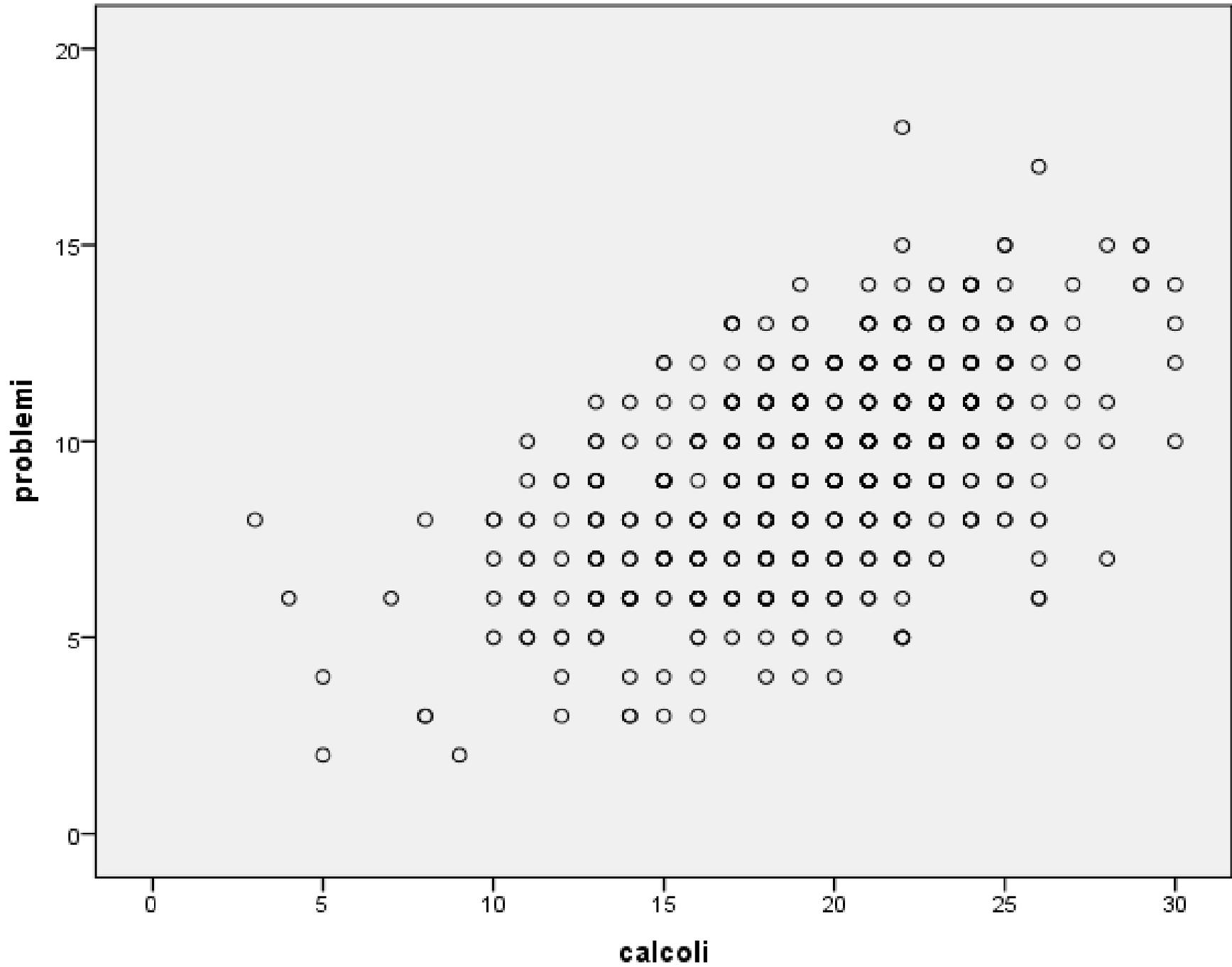
La regressione semplice

- Ricapitolazione, approfondimento e esemplificazione con casi reali

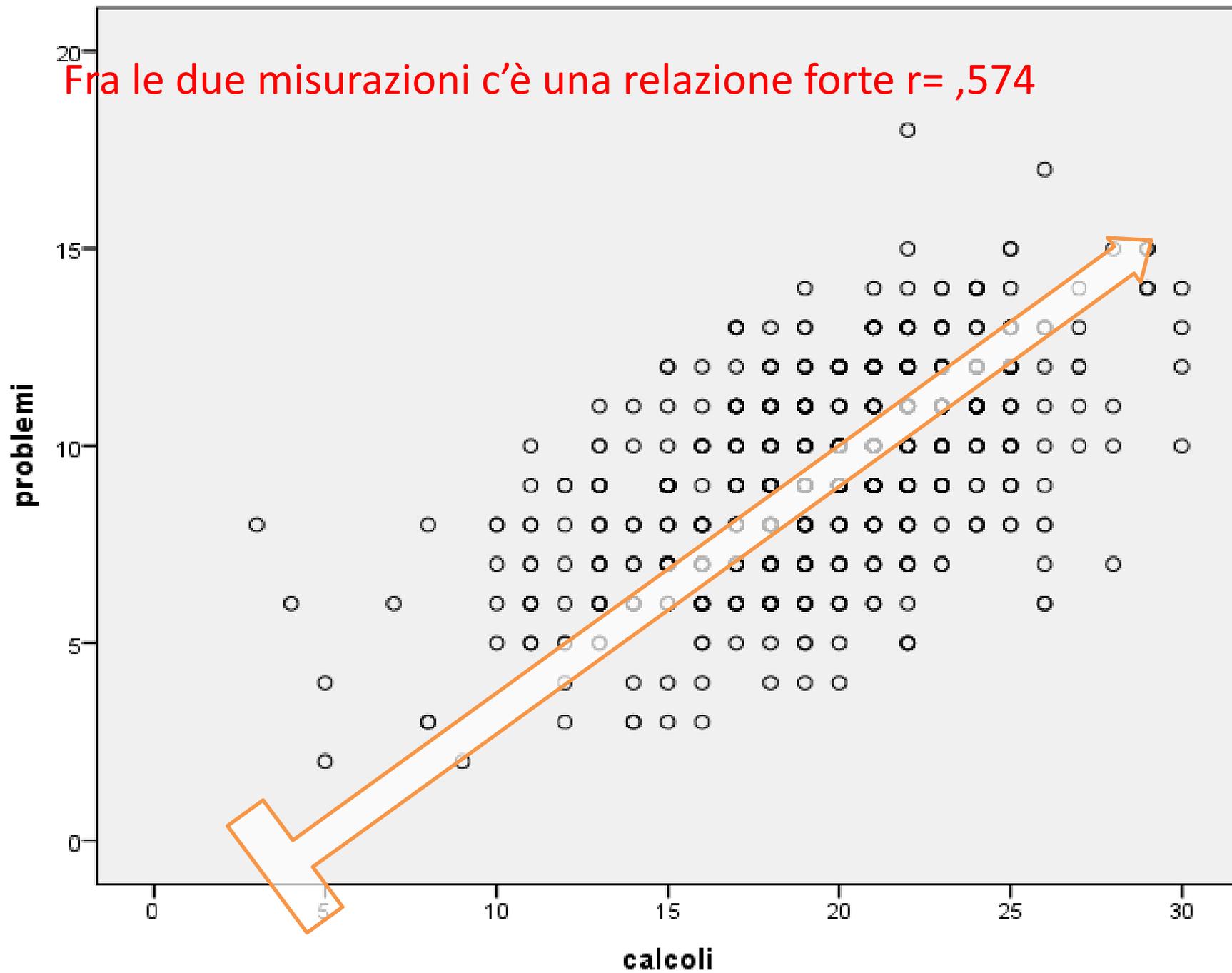
Lezioni di Psicometria
Giovanni Battista Flebus

Usiamo due test del GATB

- Predittore: test 2, risoluzione di calcoli aritmetici
- Predetto: test 6, risoluzione di problemi
- Ipotesi: il numero dei calcoli esatti predice bene il numero di problemi svolti
- Dimostrazione: elaborazione dell'equazione di predizione.
- Preliminare: esame del grafico delle due misurazioni



Fra le due misurazioni c'è una relazione forte $r = ,574$



Il Menu
Analizza
>Regressione
>Lineare
Apri questa
finestra
Scegliamo i
due test,
G6-Problemi
come
dipendente e
G2-Calcoli
come
predittore
Diamo OK

The screenshot shows a software interface for regression analysis. On the left, a list of variables is displayed in a scrollable area:

- scuola
- nprog
- età
- genere
- g1
- g2
- g3
- g4
- g5
- g7
- th1
- th2
- th3
- th4
- th5
- th6
- tipo1

On the right, the configuration for a regression block is shown:

- Dipendente:** g6
- Blocco 1 di 1**
- Indietro** (button) **Avanti** (button)
- Indipendenti:** g2
- Metodo:** Immetti
- Variabile di selezione:** (empty field) **Regola...** (button)
- Etichette casi:** (empty field)
- Peso Minimi quadrati pesati:** (empty field)

At the bottom, there are several buttons: **OK**, **Incolla**, **Reimposta**, **Annulla**, and **Guida**.

Coefficienti^a

Modello	Coefficients non standardizzati		Coefficienti standardizzati	t	Sign.
	T	Errore std	Beta		
1	(Costante)	2,717		6,989	,000
	g2 calcoli	,340	,574	17,658	,000

a. Variabile dipendente: g6 problemi

-Coefficiente B zero
 -coefficiente additivo
 - Intercetta

Coefficienti^a

Modello	Coefficients non standardizzati		Coefficienti standardizzati	t	Sign.
	T	Errore std	Beta		
1 (Costante)	2,717	,389		6,989	,000
g2 calcoli	,340	,019	,574	17,658	,000

a. Variabile dipendente: g6 problemi

- coefficiente B
- coefficiente moltiplicativo
- Pendenza

Coefficienti^a

Modello	Coefficienti non standardizzati		Coefficienti standardizzati	t	Sign.
	T	Errore std	Beta		
1 (Costante)	2,717	,389		6,989	,000
g2 calcoli	,340	,019	,574	17,658	,000

a. Variabile dipendente: g6 problemi

Coefficiente beta standardizzato,
uguale al coefficiente di
correlazione.
Elevato

Coefficienti^a

Modello	Coefficients non standardizzati		Coefficienti standardizzati	t	Sign.
	T	Errore std	Beta		
1 (Costante)	2,717	,389		6,989	,000
g2 calcoli	,340	,019	,574	17,658	,000

a. Variabile dipendente: g6 problemi

t di Student, entrambi molto elevati, dato il grande numero di casi e della grande forza della relazione

Coefficienti^a

Modello	Coefficients non standardizzati		Coefficienti standardizzati	t	Sign.
	T	Errore std	Beta		
(Costante)	2,717	,389		6,989	,000
g2 calcoli	,340	,019	,574	17,658	,000

a. Variabile dipendente: g6 problemi

Significatività, inferiore a 0,0005

Riepilogo del modello^b

Modello	R	R-quadrato	R-quadrato adattato	Errore standard della stima
1	,574 ^a	,330	,329	2,025

a. Predittori (costante), g2 calcoli

b. Variabile dipendente: g6 problemi

Il coefficiente di correlazione multiplo è molto elevato

Riepilogo del modello^b

Modello	R	R-quadrato	R-quadrato adattato	Errore standard della stima
1	,574 ^a	,330	,329	2,025

a. Predittori: (costante), g2 calcolati

b. Variabile dipendente: g6 problema

Il quadrato di R indica la quota di varianza spiegata dalla variabile indipendente. Moltiplicata per 100 dà la percentuale di varianza spiegata

Riepilogo del modello^b

Modello	R	R-quadrato	R-quadrato adattato	Errore standard della stima
1	,574 ^a	,330	,329	2,025

a. Predittori: (costante), g2 calcoli

Il quadrato di R corretto dà una stima del vero R multiplo su un campione nuovo. Qui il cambiamento è infimo, poiché vi è un solo predittore e i casi sono molto numerosi.

Riepilogo del modello^b

Modello	R	R-quadrato	R-quadrato adattato	Errore standard della stima
1	,574 ^a	,330	,329	2,025

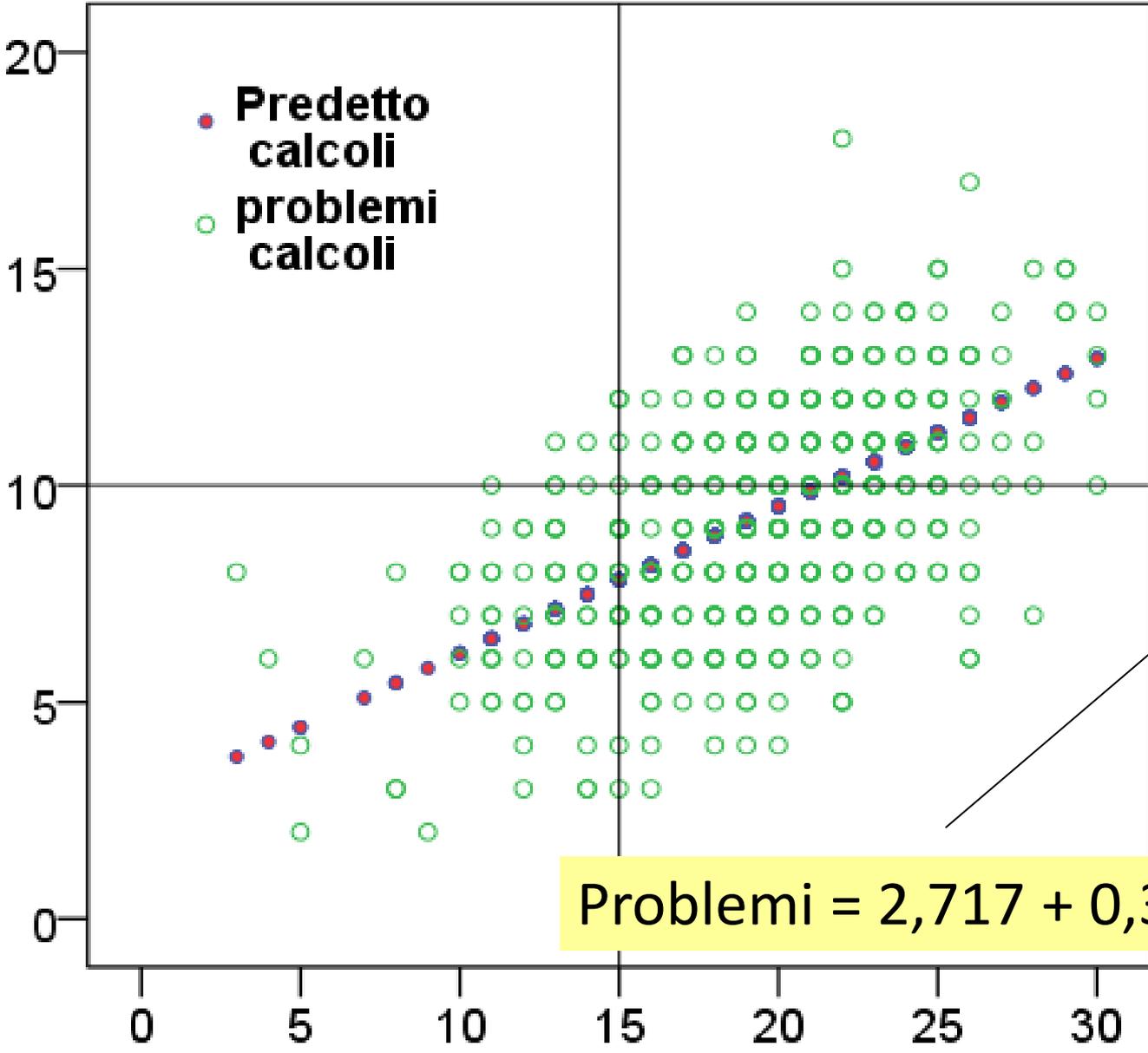
a. Predittori: (costante), g2 calcoli

b. Variabile dipendente: g6 problemi

L'errore standard della stima è la deviazione standard dei residui, con media zero e distribuzione normale. Il 68 % degli errori è inferiore a 2,025 in valore assoluto

In ascissa il test dei calcoli

In ordinata i problemi, osservati e predetti



Equazione di predizione

$$\text{Problemi} = 2,717 + 0,340 \times \text{Calcoli}$$

Predizione e residui

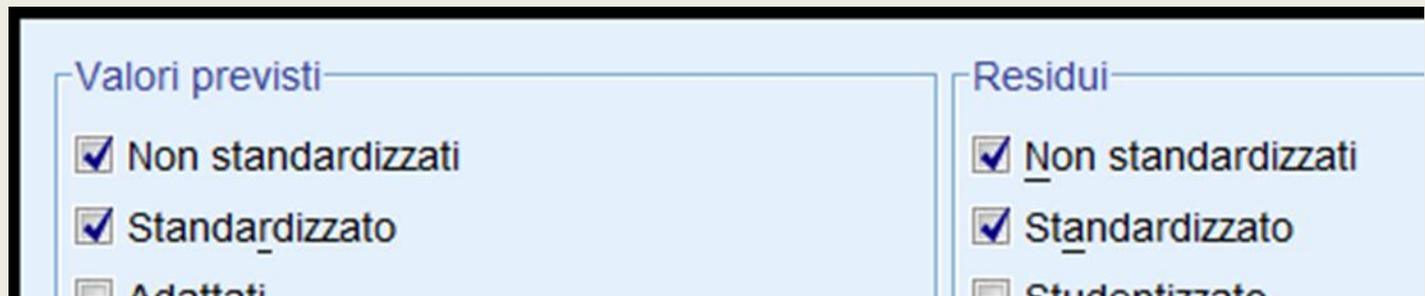
- I valori predetti possono essere confrontati con i valori realmente osservati.
- La loro differenza è detta RESIDUO
- Hanno
 - media uguale a zero
 - d.s. $s_y \sqrt{1 - R^2}$
 - una distribuzione normale

Che (cor)relazioni fra predittori, predetti e residui?

- Usiamo ancora SPSS per esaminare le correlazioni fra
- Test dei calcoli
- Test dei problemi
- Test dei problemi predetti
- Residuo dei problemi predetti

Per analizzare meglio i valori predetti e i residui, è più opportuno standardizzarli. Si possono ottenere con questo comando:

>Analizza >Regressione >Lineare >Salva



Statistiche residui^a

	Minimo	Massimo	Media	Deviazione std.	N
Valore previsto	3,74	12,92	9,43	1,420	635
Residuo	-5,561	7,799	,000	2,023	635
Valore previsto std.	-4,012	2,456	,000	1,000	635
Residuo std.	-2,746	3,852	,000	,999	635

a. Variabile dipendente: g6 problemi

I valori previsti e i residui servono a tenere sotto controllo il processo di predizione, anche se per queste due righe è improbabile individuare gli elementi da approfondire

Statistiche residui^a

	Minimo	Massimo	Media	Deviazione std.	N
Valore previsto	3,74	12,92	9,43	1,420	635
Residuo	-5,561	7,799	,000	2,023	635
Valore previsto std.	-4,012	2,456	,000	1,000	635
Residuo std.	2,746	3,852	,000	,999	635

a. Variabile dipendente: g6 problemi

I valori previsti standardizzati invece possono aiutare a individuare i valori anomali o estremi: per esempio, un punto zeta come questo (-4) è estremamente raro e su un campione di 635 rappresenta un'anomalia da sorvegliare

Statistiche residui^a

	Minimo	Massimo	Media	Deviazione std.	N
Valore previsto	3,74	12,92	9,43	1,420	635
Residuo	-5,561	7,799	,000	2,023	635
Valore previsto std.	-4,012	2,456	,000	1,000	635
Residuo std.	-2,746	3,852	,000	,999	635

a. Variabile dipendente: g6 problemi

Parimenti, un residuo standardizzato pari a 3,85 è molto raro in un campione di 635. Merita approfondimento

Matrice di correlazione

	g2 calcoli	g6 problemi	PRE_1 Predetto	RES_1 Residuo
Correlazione	g2 calcoli	,574	1,000	,000
	g6 problemi	,574	1,000	,819
	PRE_1 Predetto	1,000	1,000	,000
	RES_1 Residuo	,000	,000	1,000

La correlazione fra il test dei calcoli e il test dei problemi è elevata

Matrice di correlazione

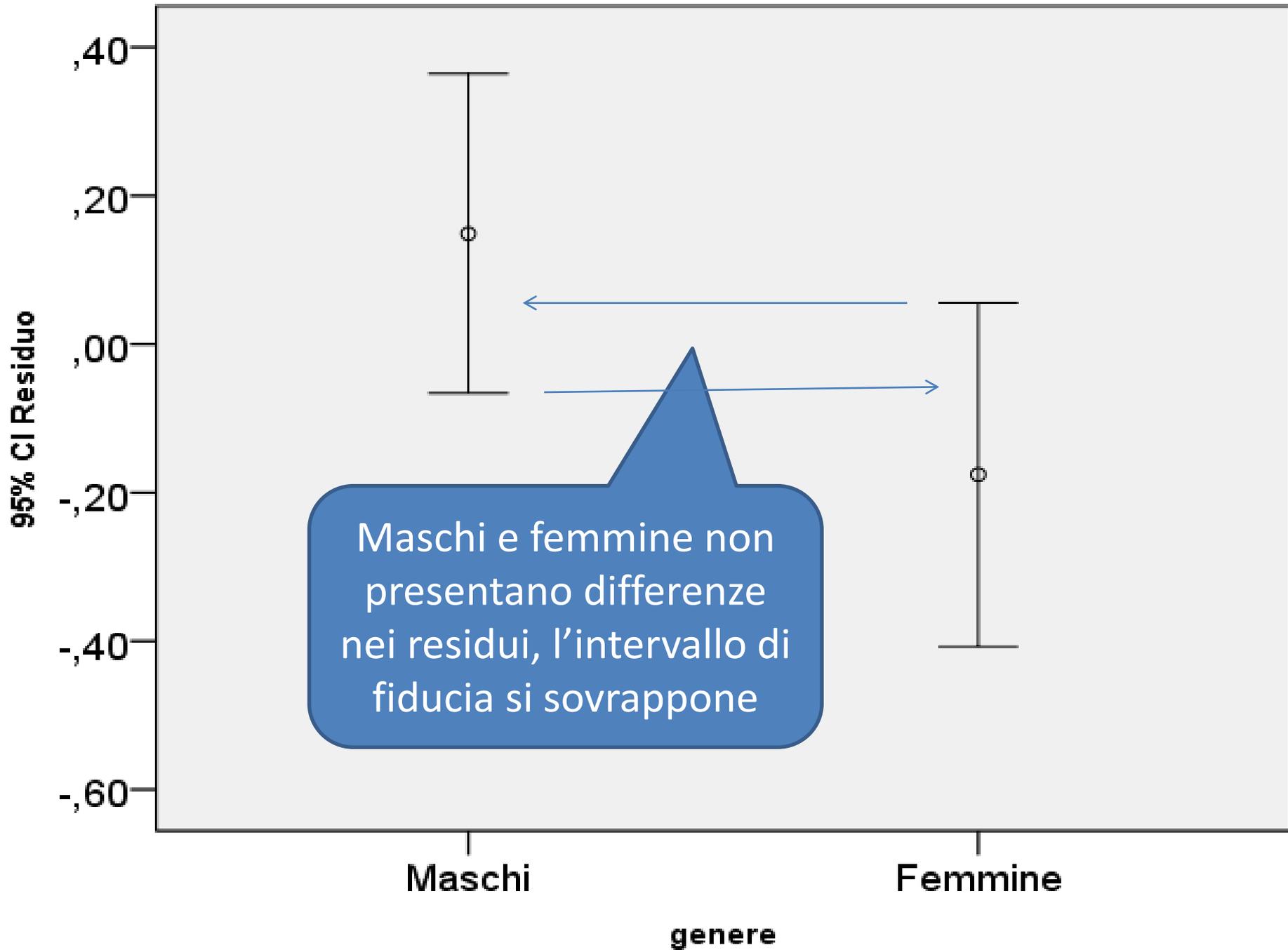
	g2 calcoli	g6 problemi	PRE_1 Predetto	RES_1 Residuo	
Correlazione	g2 calcoli	1,000	,574	1,000	,000
	g6 problemi	,574	1,000	,574	,819
	PRE_1 Predetto	1,000	,574	1,000	,000
	RES_1 Residuo	,000	,819	,000	1,000

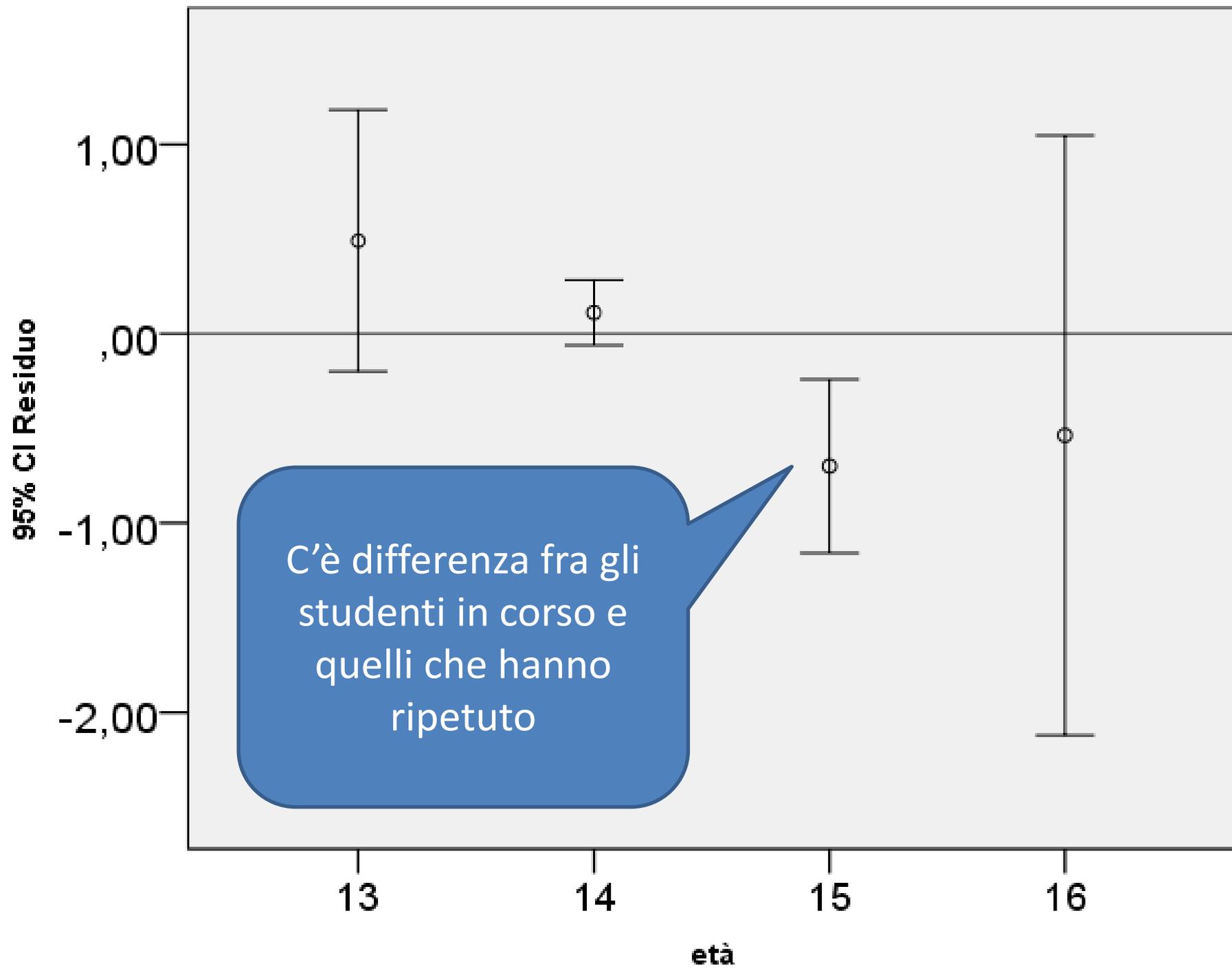
La correlazione fra il test dei calcoli e i valori che predicono i problemi è perfetta, poiché si tratta della stessa variabile, trasformata in modo lineare

Matrice di correlazione

	g2 calcoli	g6 problemi	PRE_1 Predetto	RES_1 Residuo	
Correlazione	g2 calcoli	1,000	,574	1,000	,000
	g6 problemi	,574	1,000	,574	,819
	PRE_1 Predetto	1,000	,574	1,000	,000
	RES_1 Residuo	,000	,819	,000	1,000

La correlazione fra il test dei calcoli e i residui invece è perfettamente uguale a zero





Per analizzare meglio i valori predetti e i residui, è più opportuno standardizzarli. Si possono ottenere con questo comando:
>Analizza
>Regressione
>Lineare
>Salva

Valori previsti

- Non standardizzati
- Standardizzato
- Adattati
- Errore standard delle previsioni delle medie

Residui

- Non standardizzati
- Standardizzato
- Studentizzato
- Eliminato
- Per cancellazione studentizzata

Distanze

- Di Mahalanobis
- Di Cook
- Valori di leva

Statistiche di influenza

- DiffBeta
- DiffBeta standardizzate
- DifFit
- DifFit standardizzata
- Rapporto di covarianza

Intervalli di previsione

- Media
- Singolo

Intervallo di confidenza: %

Statistiche dei coefficienti

- Crea statistiche coefficienti
 - Crea nuovo dataset
Nome dataset:
 - Scrivi nuovo file di dati