

# Il residuo nella predizione

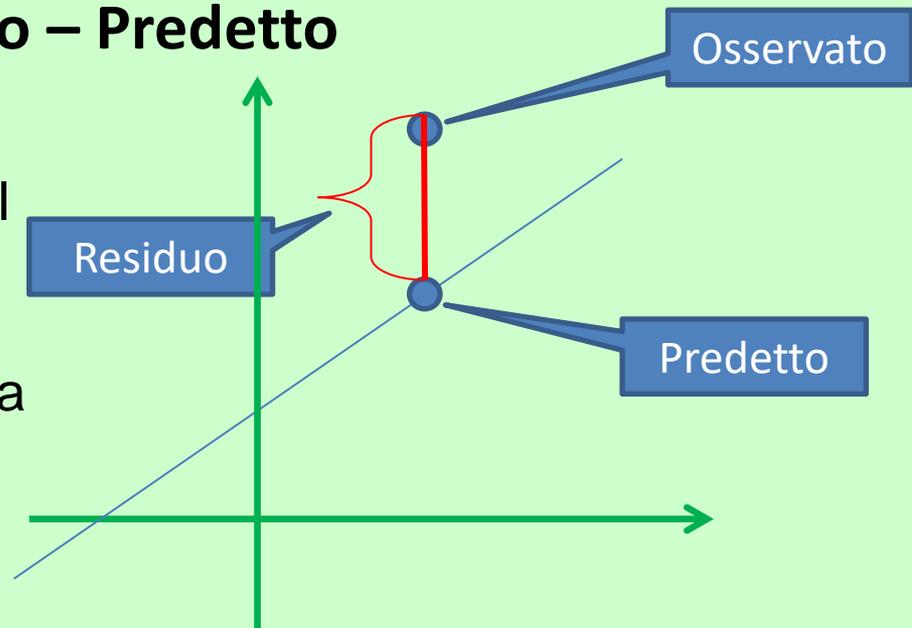
Lezioni di Psicometria  
Giovanni Battista Flebus

# Definizione di residuo

- Il residuo è la differenza fra il punteggio predetto e il punteggio osservato

$$\text{Residuo} = \text{Osservato} - \text{Predetto}$$

Graficamente, è la distanza tra il punto indicante la misurazione realmente effettuata e il suo corrispondente appartenente alla retta di regressione.



# Ricordiamo la formula di predizione

$$Y_i = mX_i + a + \varepsilon_i$$

Variabile  
dipendente,  
spiegata, valore  
osservato

inclinazione

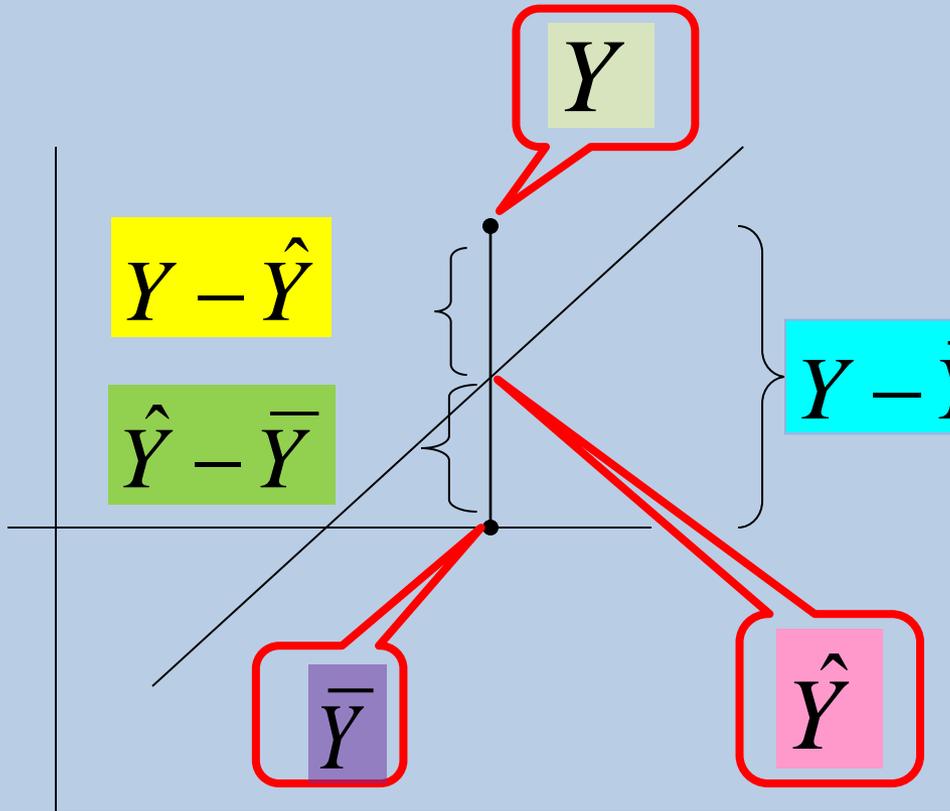
variabile  
indipendente

intercetta

errore

scarti totali,  
somma dei  
quadrati

$$\sum (Y - \bar{Y})^2$$



residui  
spiegati

$$\sum (\hat{Y} - \bar{Y})^2$$

residui non  
spiegati,  
somma  
degli errori

$$\sum (Y - \hat{Y})^2$$

Tre devianze (sum of squares)

# Porzione di varianza spiegata

$$\sum (Y - \bar{Y})^2 = \sum (\hat{Y} - \bar{Y})^2 + \sum (Y - \hat{Y})^2$$

$$\sum (\hat{Y} - \bar{Y})^2 = \sum (Y - \bar{Y})^2 - \sum (Y - \hat{Y})^2$$

$$\frac{\sum (\hat{Y} - \bar{Y})^2}{\sum (Y - \bar{Y})^2} = \frac{\sum (Y - \bar{Y})^2 - \sum (Y - \hat{Y})^2}{\sum (Y - \bar{Y})^2}$$

Facendo il rapporto fra la varianza spiegata e quella totale, la esprimiamo come **proporzione di varianza spiegata**

# Definizione di varianza spiegata

La proporzione di varianza spiegata è anche chiamata “Coefficiente di determinazione”

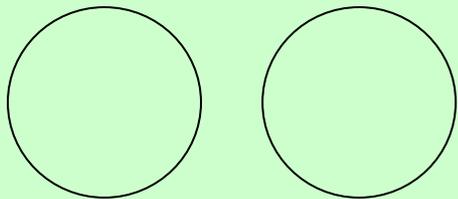
$r^2$

oppure

varianza comune

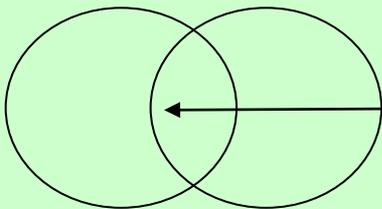
Coefficiente di indeterminazione =  $1 - r^2$

# Rappresentazione con diagrammi di Venn

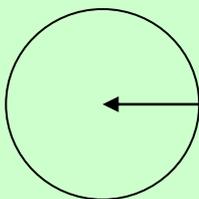


$$r^2 = 0$$

Proporzione di  
varianza comune alle  
due variabili

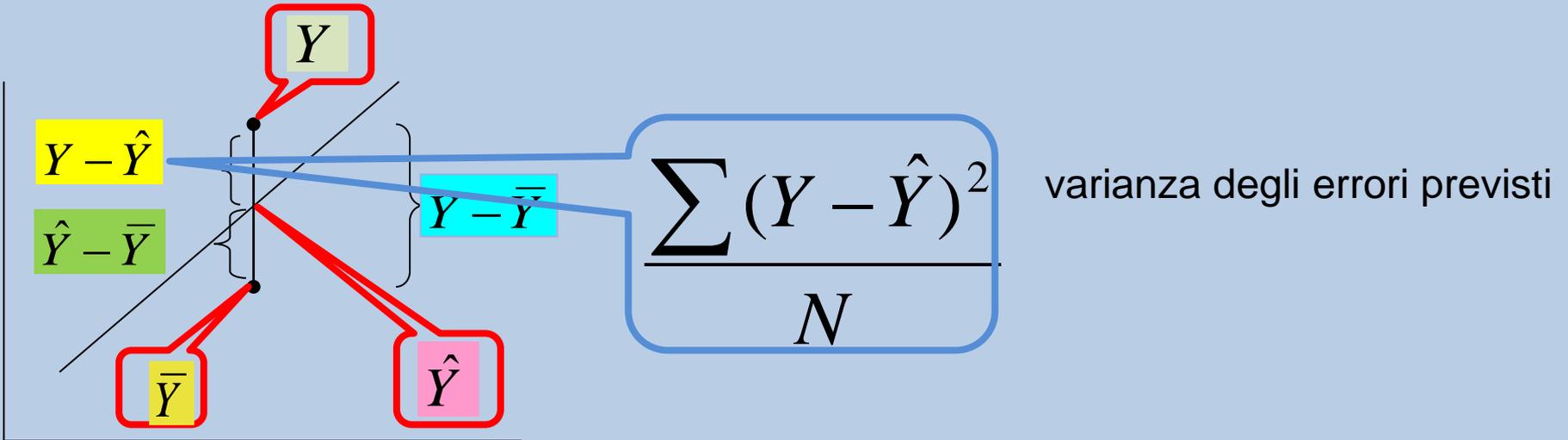


$$r^2 \in (0, 1)$$



$$r^2 = 1$$

# L'errore standard della stima



se  $r=1$ , va 0  
(nessun errore)

$$\sqrt{\frac{\sum (Y - \hat{Y})^2}{N}} = s_{y.x} = s_y \sqrt{1 - r^2}$$

se  $r=0$ , va a  $s_y$   
(massimo errore)

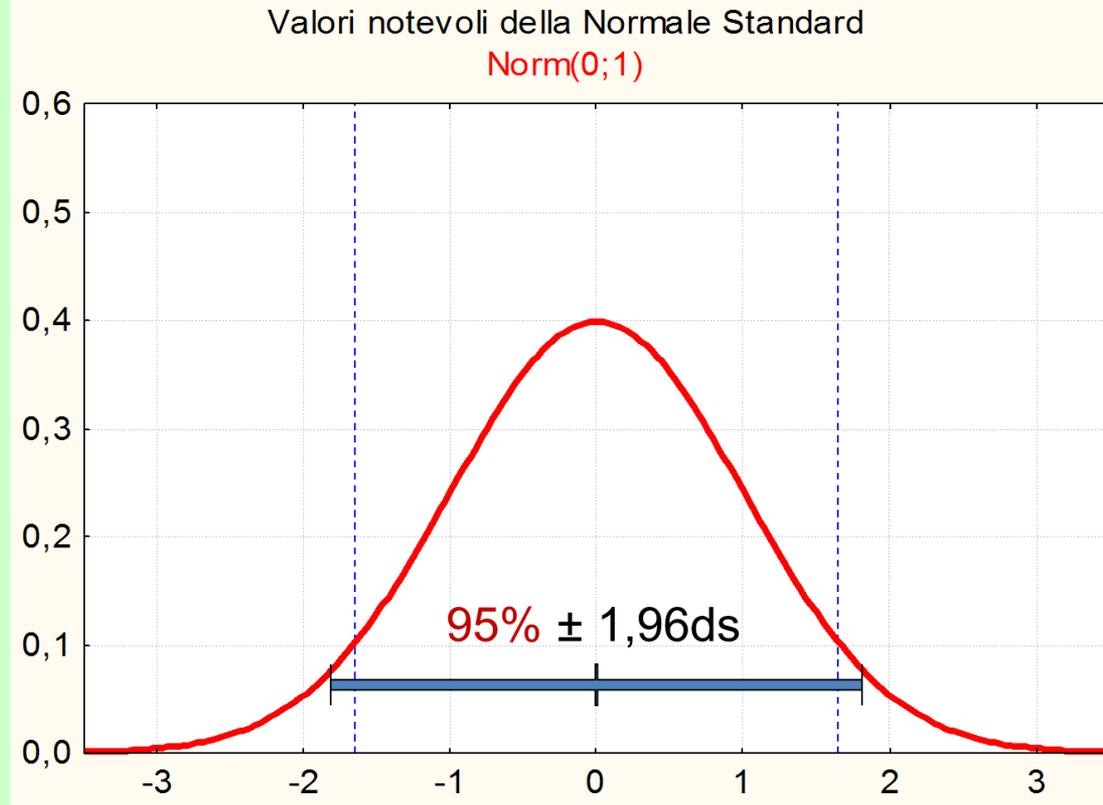
deviazione standard degli errori previsti, errore stimato, errore standard delle stime

A cosa serve l'errore standard delle stime?

$$\hat{Y} - 1.96s_{y.x}$$

$$\hat{Y} + 1.96s_{y.x} \quad e$$

1.96 è il punto z  
corrispondente all'area  
95% attorno alla media



Essendo una deviazione standard, e presumendo che X e Y siano distribuite normalmente, possiamo stimare che il 95% dei valori Y stimati a partire da un certo valore X sarà compreso fra **+ - 1,96 ds**

| studente | ab_verbale | voto |
|----------|------------|------|
| A        | 12         | 8    |
| B        | 10         | 7    |
| C        | 14         | 8    |
| D        | 9          | 5    |
| E        | 9          | 6    |
| F        | 13         | 9    |
| G        | 11         | 7    |
| H        | 8          | 5    |

$$Votopred = 0,135 + 0,626 * abilità$$

| studente | ab_verbale | voto | PRE_1 |
|----------|------------|------|-------|
| A        | 12         | 8    | 7,659 |
| B        | 10         | 7    | 6,405 |
| C        | 14         | 8    | 8,913 |
| D        | 9          | 5    | 5,778 |
| E        | 9          | 6    | 5,778 |
| F        | 13         | 9    | 8,286 |
| G        | 11         | 7    | 7,032 |
| H        | 8          | 5    | 5,151 |

$$Votopred = 0,135 + 0,626 * abilità$$

| studente | ab_verbale | voto | PRE_1 | RES_1  |
|----------|------------|------|-------|--------|
| A        | 12         | 8    | 7,659 | 0,341  |
| B        | 10         | 7    | 6,405 | 0,595  |
| C        | 14         | 8    | 8,913 | -0,913 |
| D        | 9          | 5    | 5,778 | -0,778 |
| E        | 9          | 6    | 5,778 | 0,222  |
| F        | 13         | 9    | 8,286 | 0,714  |
| G        | 11         | 7    | 7,032 | -0,032 |
| H        | 8          | 5    | 5,151 | -0,151 |

$$8 - 7,659 = 0,341$$

# Usiamo SPSS per calcolare i residui

The screenshot shows the 'Residuals' dialog box in SPSS. On the left, a list of variables includes 'a' and 'studente'. The 'Dipendente:' (Dependent) field contains 'voto'. The 'Indipendenti:' (Independent) field contains 'ab\_verbale'. The 'Metodo:' (Method) is set to 'Immetti'. The 'Variabile di selezione:' (Selection variable) field is empty, with a 'Regola...' button next to it. The 'Etichette casi:' (Case labels) and 'Peso Minimi quadrati pesati:' (Weighted least squares) fields are also empty. On the right side, there are buttons for 'Statistiche...', 'Grafici...', 'Salva...', 'Opzioni...', and 'Stile...'. At the bottom, there are buttons for 'OK', 'Incolla', 'Reimposta', 'Annulla', and 'Guida'.

studente

Dipendente:  
voto

Blocco 1 di 1

Indietro Avanti

Indipendenti:  
ab\_verbale

Metodo: Immetti

Variabile di selezione:  
Regola...

Etichette casi:

Peso Minimi quadrati pesati:

OK Incolla Reimposta Annulla Guida

Statistiche...  
Grafici...  
Salva...  
Opzioni...  
Stile...

Osservati

Residui

### Valori previsti

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

### Residui

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

### Distanze

- Di Mahalanobis
- Di Cook
- Valori di leva

### Statistiche di influenza

- DiffBeta
- DiffBeta standardizzate
- DiffFit
- DiffFit 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

# Eseguiamo un controllo usando la sintassi

```
compute verifica = pre_1+ res_1 .
```

```
DESCRIPTIVES VARIABLES= ab_verbale voto  
PRE_1 RES_1 verifica  
/STATISTICS=MEAN STDDEV VARIANCE.
```

# Caratteristiche dei residui

## Statistiche descrittive

|   | Media | Deviazione<br>std. | Varianza |
|---|-------|--------------------|----------|
| <b>ab_verbale</b>                           | 10,75 | 2,121              | 4,500    |
| <b>voto</b>                                 | 6,88  | 1,458              | 2,125    |
| <b>PRE_1 Unstandardized Predicted Value</b> | 6,88  | 1,330              | 1,769    |
| <b>RES_1 Unstandardized Residual</b>        | ,00   | ,597               | ,356     |
| <b>verifica</b>                             | 6,88  | 1,458              | 2,125    |

Media della variabili predetta  
e osservata uguale

# Caratteristiche dei residui

## Statistiche descrittive

|                                      | Media | Deviazione<br>std. | Varianza |
|--------------------------------------|-------|--------------------|----------|
| ab_verbale                           | 10,75 | 2,121              | 4,500    |
| voto                                 | 6,88  | 1,458              | 2,125    |
| PRE_1 Unstandardized Predicted Value | 6,88  | 1,330              | 1,769    |
| RES_1 Unstandardized Residual        | ,00   | ,597               | ,356     |
| verifica                             | 6,88  | 1,458              | 2,125    |

Media dei residui pari a zero

# Caratteristiche dei residui

## Statistiche descrittive

|                                      | Media | Deviazione<br>std. | Varianza |
|--------------------------------------|-------|--------------------|----------|
| ab_verbale                           | 10,75 | 2,121              | 4,500    |
| voto                                 | 6,88  | 1,458              | 2,125    |
| PRE_1 Unstandardized Predicted Value | 6,88  | 1,330              | 1,769    |
| RES_1 Unstandardized Residual        | ,00   | ,597               | ,356     |
| verifica                             | 6,88  | 1,458              | 2,125    |

Varianza spiegata + varianza dei  
 residui = varianza totale  
 $1,769 + 0,356 = 2,125$

# Caratteristiche dei residui

## Statistiche descrittive

|                                      | Media | Deviazione<br>std. | Varianza |
|--------------------------------------|-------|--------------------|----------|
| ab_verbale                           | 10,75 | 2,121              | 4,500    |
| voto                                 | 6,88  | 1,458              | 2,125    |
| PRE_1 Unstandardized Predicted Value | 6,88  | 1,330              | 1,769    |
| RES_1 Unstandardized Residual        | ,00   | ,597               | ,356     |
| verifica                             | 6,88  | 1,458              | 2,125    |

## Riepilogo del modello

| Modello | R                 | R-quadrato | R-quad<br>ad |
|---------|-------------------|------------|--------------|
| 1       | ,912 <sup>a</sup> | ,832       |              |

Quota di Varianza spiegata = + varianza  
spiegata su varianza totale  
 $1,769 / 2,125 = ,832$

# Predizione usando i punti standardizzati

# Predizione con punti zeta

$$\hat{z}_{yi} = z_{xi} \cdot r_{xy}$$

$\hat{z}_{yi}$  = zeta predetto

$z_{xi}$  = zeta predittore

$r_{xy}$  = coefficiente di correlazione

# Eseguiamo la predizione usando i valori standardizzati

SPSS esegue DESCRIPTIVES con VOTO e  
AB\_VERBALE Salviamo i punteggi standardizzati  
O con i comandi da menu o con la sintassi

# Eseguiamo la predizione usando i valori standardizzati

The screenshot shows a software interface for variable selection and standardization. On the left is an empty box. In the center is a blue button with a left-pointing arrow. On the right is a box titled "Variabili:" containing a list of variables: "ab\_verbale" and "voto". The "voto" variable is highlighted. To the right of the "Variabili:" box are two buttons: "Opzioni..." and "Stile...". At the bottom left, there is a checked checkbox labeled "Salva valori standardizzati come variabili". At the bottom are five buttons: "OK", "Incolla", "Reimposta", "Annulla", and "Guida".

Variabili:

- ab\_verbale
- voto

Salva valori standardizzati come variabili

OK   Incolla   Reimposta   Annulla   Guida

studente  
ab\_verbale  
voto

Dipendente:  
Z\_voto

Blocco 1 di 1

Indietro Avanti

Indipendenti:  
Z\_ab\_verbale

Metodo: Immetti

Variabile di selezione:  
Regola...

Etichette casi:

Peso Minimi qu

OK Incolla Reimposta

Statistiche...  
Grafici...  
Salva...  
Opzioni...  
Stile...

Salviamo anche i predetti  
e i residui non  
standardizzati

Con la sintassi (o comandi di menu) eseguiamo la verifica:

```
COMPUTE verifica_z = pre_2+ res_2 .
```

Chiediamo anche le descrittive delle nuove variabili:

```
DESCRIPTIVES VARIABLES=Z_ab_verbale  
Z_voto PRE_2 RES_2 verifica_Z  
/STATISTICS=MEAN STDDEV VARIANCE.
```

# Caratteristiche della predizione con variabili standardizzate

## Statistiche descrittive

|                                      | Media | Deviazione std. | Varianza |
|--------------------------------------|-------|-----------------|----------|
| Z_ab_verbale                         | ,00   | 1,000           | 1,000    |
| Z_voto                               | ,00   | 1,000           | 1,000    |
| PRE_2 Unstandardized Predicted Value | ,00   | ,912            | ,832     |
| RES_2 Unstandardized Residual        | ,00   | ,409            | ,168     |
| verifica_Z                           | ,00   | 1,000           | 1,000    |

Le medie sono sempre zero, le ds e varianze delle variabili standardizzate sono uguali a 1

# Caratteristiche della predizione con variabili standardizzate

## Statistiche descrittive

|                                      | Media | Deviazione std. | Varianza |
|--------------------------------------|-------|-----------------|----------|
| Z_ab_verbale                         | ,00   | 1,000           | 1,000    |
| Z_voto                               | ,00   | 1,000           | 1,000    |
| PRE_2 Unstandardized Predicted Value | ,00   | ,912            | ,832     |
| RES_2 Unstandardized Residual        | ,00   | ,409            | ,168     |
| verifica_Z                           | ,00   | 1,000           | 1,000    |

La deviazione standard dei predetti è uguale al coefficiente di correlazione multiplo

La varianza dei predetti è uguale al coefficiente di correlazione quadrato

## Riepilogo del modello

| Modello | R                 | R-quadrato | R-quadrato ad |
|---------|-------------------|------------|---------------|
| 1       | ,912 <sup>a</sup> | ,832       |               |

a. Predittori: (costante), Z\_ab\_verbale

b. Variabile dipendente: Z\_voto

# Caratteristiche della predizione con variabili standardizzate

## Statistiche descrittive

|                                      | Media | Deviazione std. | Varianza |
|--------------------------------------|-------|-----------------|----------|
| Z_ab_verbale                         | ,00   | 1,000           | 1,000    |
| Z_voto                               | ,00   | 1,000           | 1,000    |
| PRE_2 Unstandardized Predicted Value | ,00   | ,912            | ,832     |
| RES_2 Unstandardized Residual        | ,00   | ,409            | ,168     |
| verifica_Z                           | ,00   | 1,000           | 1,000    |

## Riepilogo del mo

| Modello | R                 | R-quadrato | R-q<br>ac |
|---------|-------------------|------------|-----------|
| 1       | ,912 <sup>a</sup> | ,832       |           |

a. Predittori: (costante), Z\_ab\_verbale

b. Variabile dipendente: Z\_voto

La varianza spiegata e quella residua è uguale a 1

# Caratteristiche della predizione con variabili standardizzate

## Statistiche descrittive

|   | Media | Deviazione std. | Varianza |
|---|-------|-----------------|----------|
| ab_verbale                                  | 10,75 | 2,121           | 4,500    |
| voto  | 6,88  | 1,458           | 2,125    |
| <b>PRE_1 Unstandardized Predicted Value</b> | 6,88  | 1,330           | 1,769    |
| RES_1 Unstandardized Residual               | ,00   | ,597            | ,356     |
| verifica                                    | 6,88  | 1,458           | 2,125    |
| Z_ab_verbale                                | ,00   | 1,000           | 1,000    |
| Z_voto                                      | ,00   | 1,000           | 1,000    |
| <b>PRE_2 Unstandardized Predicted Value</b> | ,00   | ,912            | ,832     |
| RES_2 Unstandardized Residual               | ,00   | ,409            | ,168     |
| verifica_Z                                  | ,00   | 1,000           | 1,000    |

La varianza spiegata  
 $1,779 / 2,125 = 0,832$

# Caratteristiche della predizione con variabili standardizzate

## Statistiche descrittive

|   | Media | Deviazione std. | Varianza |
|---|-------|-----------------|----------|
| ab_verbale                                  | 10,75 | 2,121           | 4,500    |
| voto  | 6,88  | 1,458           | 2,125    |
| <b>PRE_1 Unstandardized Predicted Value</b> | 6,88  | 1,330           | 1,769    |
| <b>RES_1 Unstandardized Residual</b>        | ,00   | ,597            | ,356     |
| verifica                                    | 6,88  | 1,458           | 2,125    |
| Z_ab_verbale                                | ,00   | 1,000           | 1,000    |
| Z_voto                                      | ,00   | 1,000           | 1,000    |
| <b>PRE_2 Unstandardized Predicted Value</b> | ,00   | ,912            | ,832     |
| <b>RES_2 Unstandardized Residual</b>        | ,00   | ,409            | ,168     |
| verifica_Z                                  | ,00   | 1,000           | 1,000    |

La varianza residua o degli errori  
 $0,356 / 2,125 = 0,168$

# Caratteristiche della predizione con variabili standardizzate

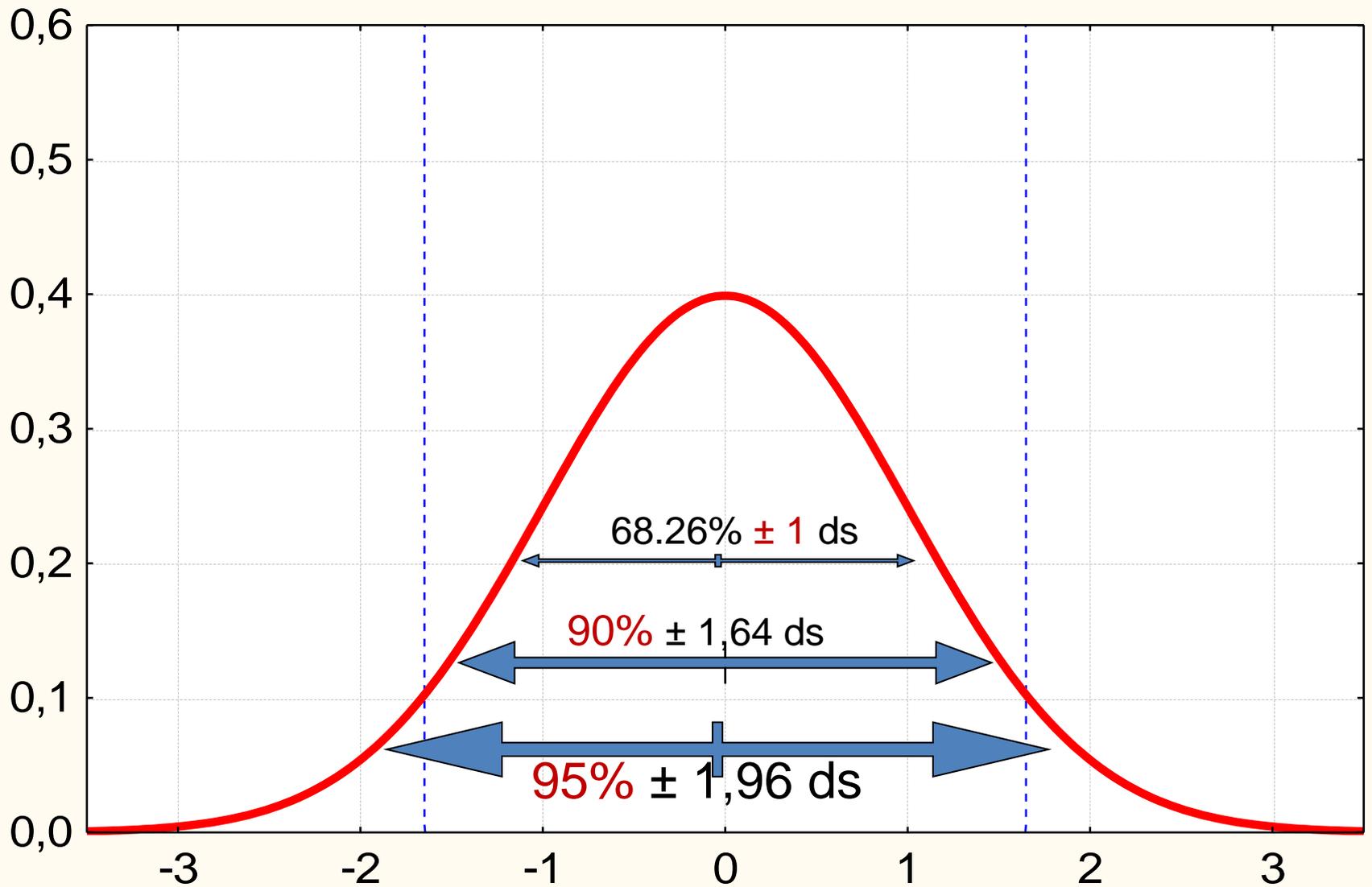
## Statistiche descrittive

|   | Media | Deviazione std. | Varianza |
|---|-------|-----------------|----------|
| ab_verbale                                  | 10,75 | 2,121           | 4,500    |
| voto  | 6,88  | 1,458           | 2,125    |
| <b>PRE_1 Unstandardized Predicted Value</b> | 6,88  | 1,330           | 1,769    |
| <b>RES_1 Unstandardized Residual</b>        | ,00   | ,597            | ,356     |
| verifica                                    | 6,88  | 1,458           | 2,125    |
| Z_ab_verbale                                | ,00   | 1,000           | 1,000    |
| Z_voto                                      | ,00   | 1,000           | 1,000    |
| <b>PRE_2 Unstandardized Predicted Value</b> | ,00   | ,832            | ,832     |
| <b>RES_2 Unstandardized Residual</b>        | ,00   | ,168            | ,168     |
| verifica_Z                                  | ,00   | 1,000           | 1,000    |

La varianza residua o degli errori è uguale alla varianza totale moltiplicata per  $(1-r^2)$

# Valori notevoli della Normale Standard

Norm(0;1)



# Caratteristiche della predizione con variabili standardizzate

## Statistiche descrittive

|   | Media | Deviazione std. | Varianza |
|---|-------|-----------------|----------|
| ab_verbale                                  | 10,75 | 2,121           | 4,500    |
| voto  | 6,88  | 1,458           | 2,125    |
| <b>PRE_1 Unstandardized Predicted Value</b> | 6,88  | 1,330           | 1,769    |
| <b>RES_1 Unstandardized Residual</b>        | ,00   | ,597            | ,356     |
| verifica                                    | 6,88  | 1,458           | 2,125    |
| Z_ab_verbale                                | ,00   | 1,000           | 1,000    |
| Z_voto                                      | ,00   | 1,000           | 1,000    |
| <b>PRE_2 Unstandardized Predicted Value</b> | ,00   | ,912            | ,832     |
| <b>RES_2 Unstandardized Residual</b>        |       | ,409            | ,168     |
| verifica_Z                                  |       | ,00             | 1,000    |

La deviazione standard dei residui o errore standard della stima permette di individuare gli intervalli di fiducia al 95% entro cui si situano gli errori della stima (o predizione)  $1,96 * 0,597 = 1,17$

# Ricapitolando...

- La **varianza spiegata** è la varianza dei predetti, cioè la varianza **spiegata dalla regressione**.
- La **varianza residua** (o varianza dei residui) indica quella parte di varianza non spiegata dalla regressione, (attribuibile all'errore).
- La somma della varianza spiegata e della varianza residua è pari alla **varianza totale del punteggio predetto**.

## Caratteristiche dei residui con valori grezzi

- I residui hanno media  $M = 0$
- La varianza dei residui è pari al quadrato del **coefficiente di alienazione moltiplicata per la varianza originale**:  $(1-r^2) \times s^2$
- La correlazione fra i residui e i predetti è **nulla**

## Caratteristiche dei residui con **punti standardizzati**

- La varianza dei predetti è uguale al **coefficiente di determinazione**:  $r^2$
- La deviazione standard dei predetti è uguale al **coefficiente di correlazione** (in quanto radice quadrata della varianza)
- La varianza dei residui è pari al quadrato del **coefficiente di alienazione**:  $(1-r^2)$

# Uso dei residui 1

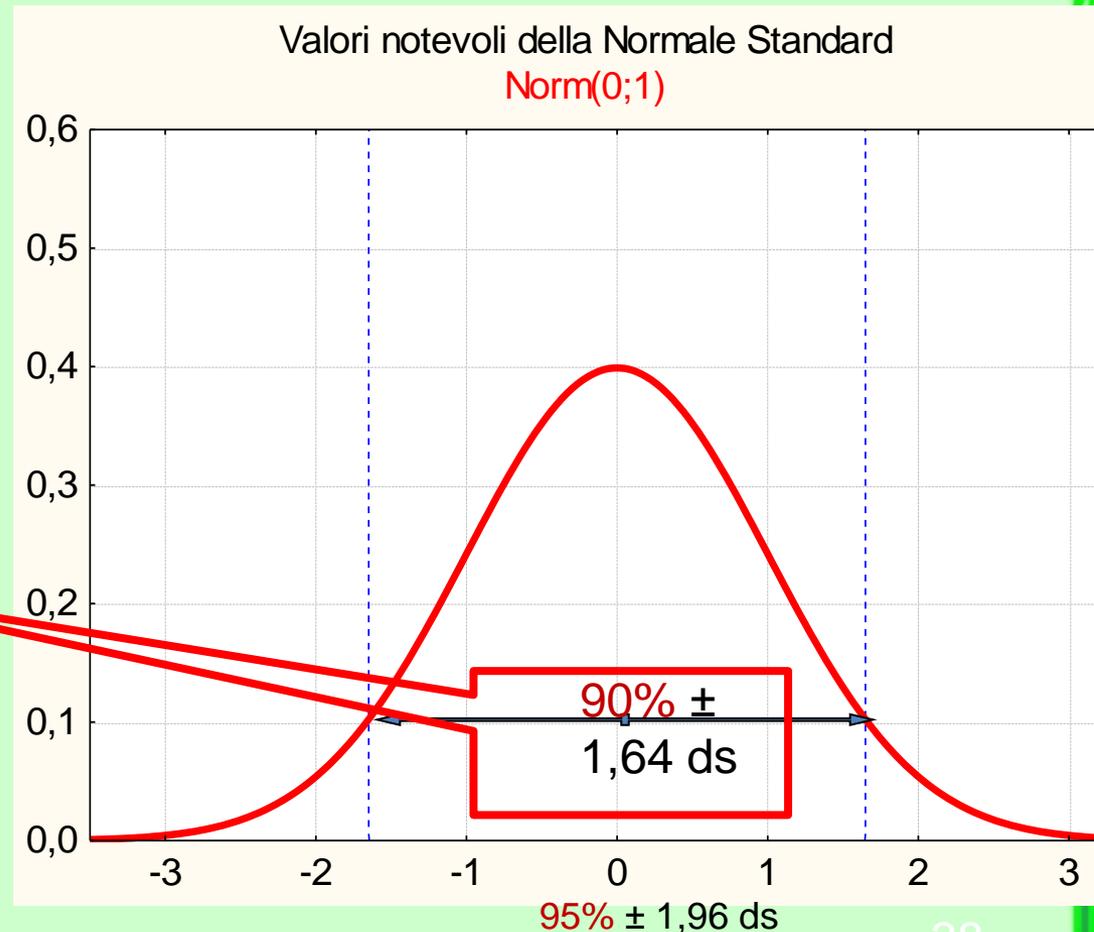
L'esame dei residui permette di valutare distribuzioni anomale, sbilanciate in una direzione o nell'altra, in alcune zone della distribuzione dei punteggi osservati piuttosto che in altre.

# Uso dei residui 2

Se la distribuzione dei residui segue la curva normale, si può stabilire l'intervallo di fiducia al 90 (95) % entro cui ricadono i gli errori di predizione. Per esempio, se l'errore standard della stima è pari a 3, possiamo dire che il 90 % degli errori è compreso fra

$$3 \times 1,64 = + - 4,92$$

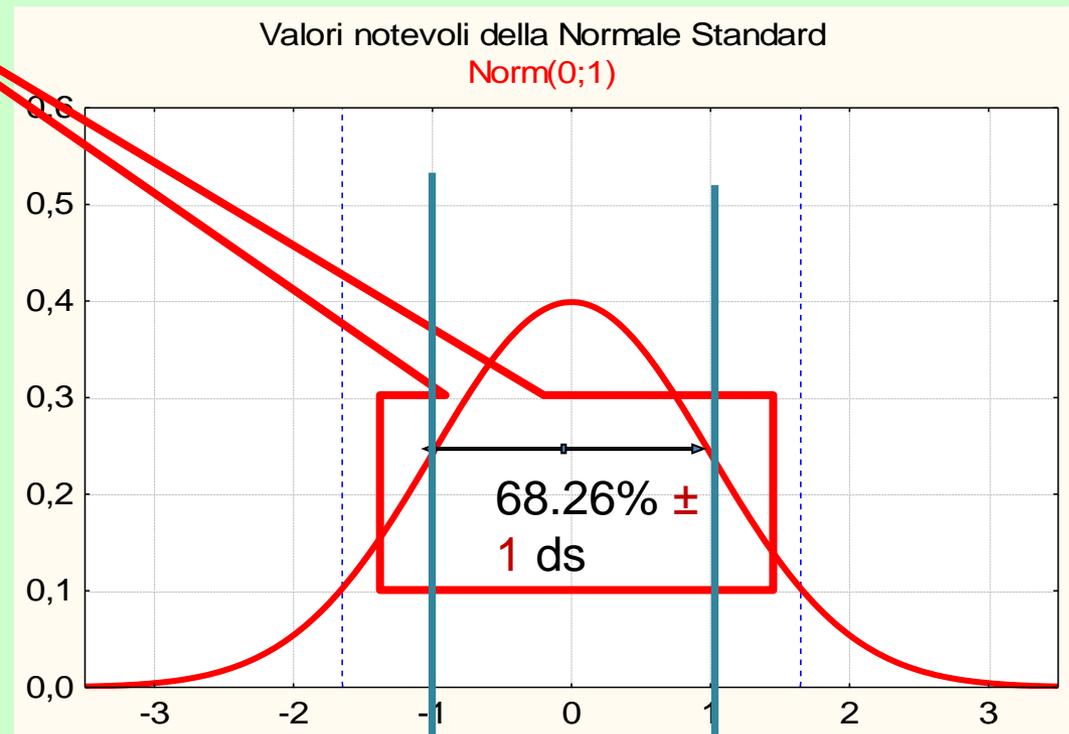
Quindi, usando quell'equazione di regressione, 90 volte su 100 l'errore è superiore a -4,92 e inferiore a 4,92. Nel 10 % dei casi l'errore sarà più grande.



# Uso dei residui 3

Oppure anche (ricordando i valori notevoli), due terzi degli errori sono compresi fra  $-3$  e  $+3$ .

- L'esame dei residui è veramente proficuo nella **regressione multipla**



# Meccanismo della predizione o della stima

1. Per ogni individuo, l'equazione della regressione predice un valore di  $Y$ , indicato con  $\hat{Y}$ , simile ma non uguale al valore osservato  $Y$
2. La differenza fra  $Y$  e  $\hat{Y}$  è chiamata **residuo**, o errore
3.  $Y$  sta vicino a  $\hat{Y}$ , con alta probabilità è molto vicino, con bassa probabilità è molto lontano dal valore vero
4. Questa vicinanza al risultato reale è definibile con la curva gaussiana, con  $m = 0$  e  $\sigma = \text{err. stand. della stima}$
5. Perciò, se non si può calcolare il punteggio reale, si può affermare che esso deve trovarsi con il 90 % (o altri livelli) di probabilità entro un certo intervallo calcolabile.

# Riassumendo

- Nella predizione del **singolo caso** non è mai possibile sapere se la predizione è precisa o no
- Si può quantificare la precisione **totale**, fatta su tutti i casi (presenti e futuri): la quota di varianza spiegata ( $r^2$ ) è un utile indice per definire la precisione della predizione.