



UNIVERSITÀ DEGLI STUDI DI ROMA "FORO ITALICO"

Diagrammi a dispersione e correlazione

Corso di Laurea in Scienze Motorie e Sportive (L22)

Insegnamento di Informatica

Prof. Federico Mari

Statistica di base

Seconda edizione

David S. Moore



Libro di testo

D.S. Moore. [Statistica di base](#). Idee & strumenti. Apogeo, 2013. ISBN: 978-8838786426.

Parte I

Capitolo 4

Analisi statistica con due variabili

- Si vogliono studiare le **relazioni** tra diverse variabili
- Misurare le due variabili sullo **stesso insieme di unità statistiche**
- **Tipo di relazione** che si vuole studiare
 - Rapporto **causa-effetto**?
 - Relazioni ma **non di causa-effetto**?



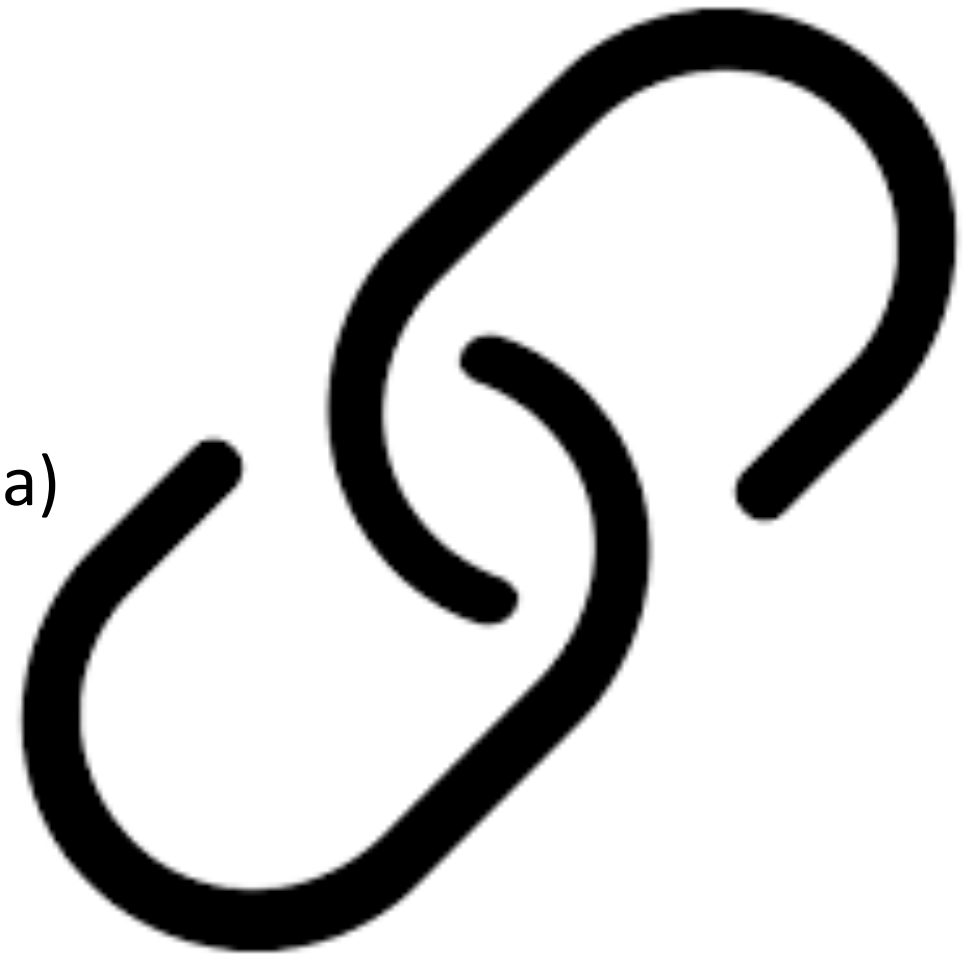
Variabili di risposta, variabili esplicative

- **Variabili di risposta** misura il risultato di una indagine
- **Variabili esplicative** spiega, o influenza, le variazioni di una variabile di risposta
 - Es: gli effetti dell'alcol sulla temperatura corporea
- Possono esserci come no



Indagine statistica con più variabili

- Rappresentare **graficamente** i dati e aggiungere indici numerici (**strumenti già studiati** per l'analisi di una variabile singola)
- Descrivere gli **andamenti generali e le deviazioni** che ne derivano
- Se l'andamento generale è regolare, utilizzare un **modello matematico** per descrivere l'andamento (regressione lineare, etc.)



FEI Dressage World Cup Salzburg, 5-9 Dec. 2019

- Jerich Dressage CDI4*
Grand Prix
- Amadeus Dressage CDI4*
Grand Prix Special
- FEI World Cup™ Dressage
Grand Prix
- FEI World Cup™ Dressage
Freestyle



FEI Dressage World Cup Salzburg, 29-30 Nov. 2019

- Unità statistiche:
 - Cavalieri e cavalli, ipotesi semplificativa: sono diversi in ogni competizione
- Variabili statistiche:
 - Punteggi di ogni Giudice (5 giudici)
- Obiettivo:
 - Valutazioni simili dei giudici?



Analisi dei punteggi di ogni giudice (1 var)

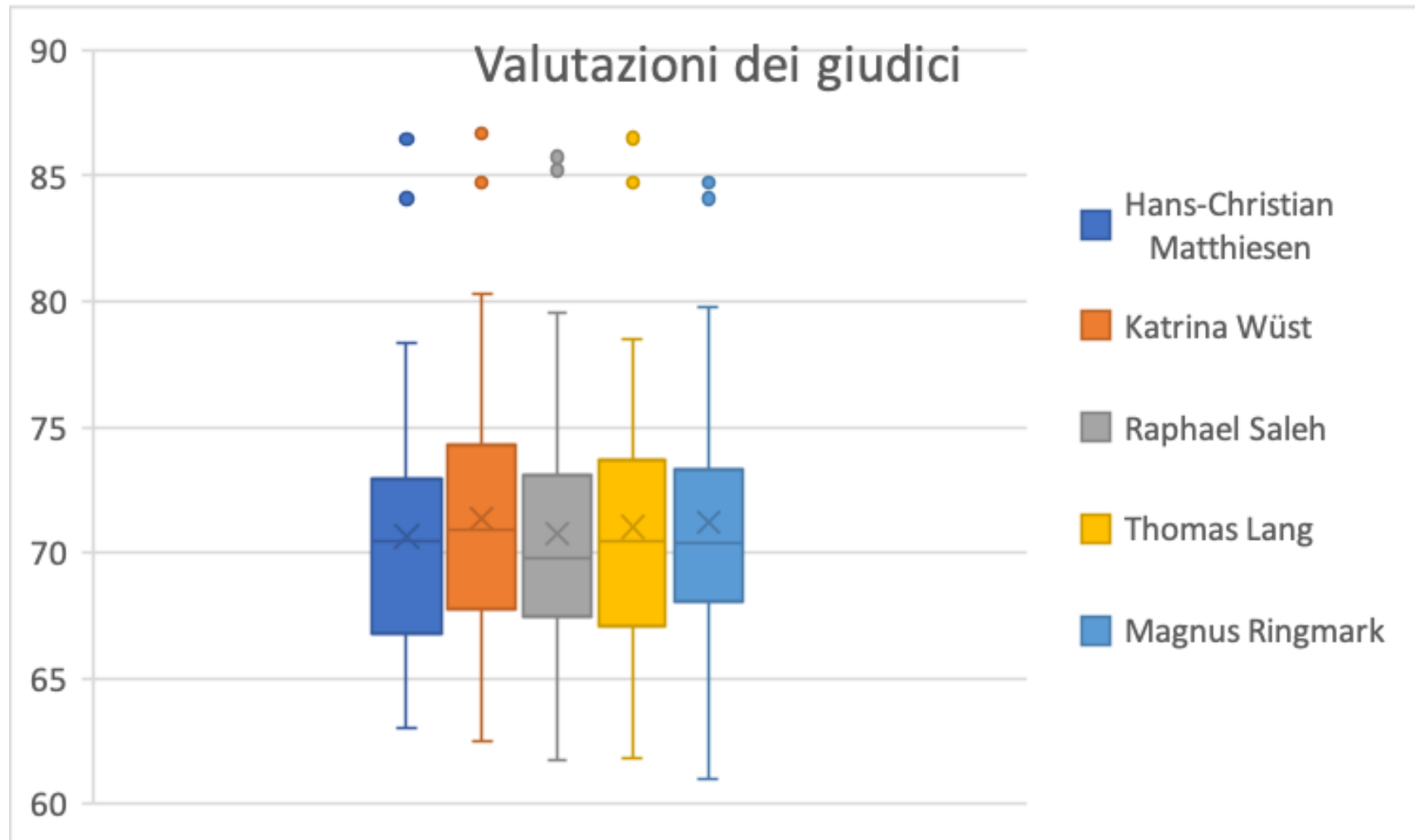
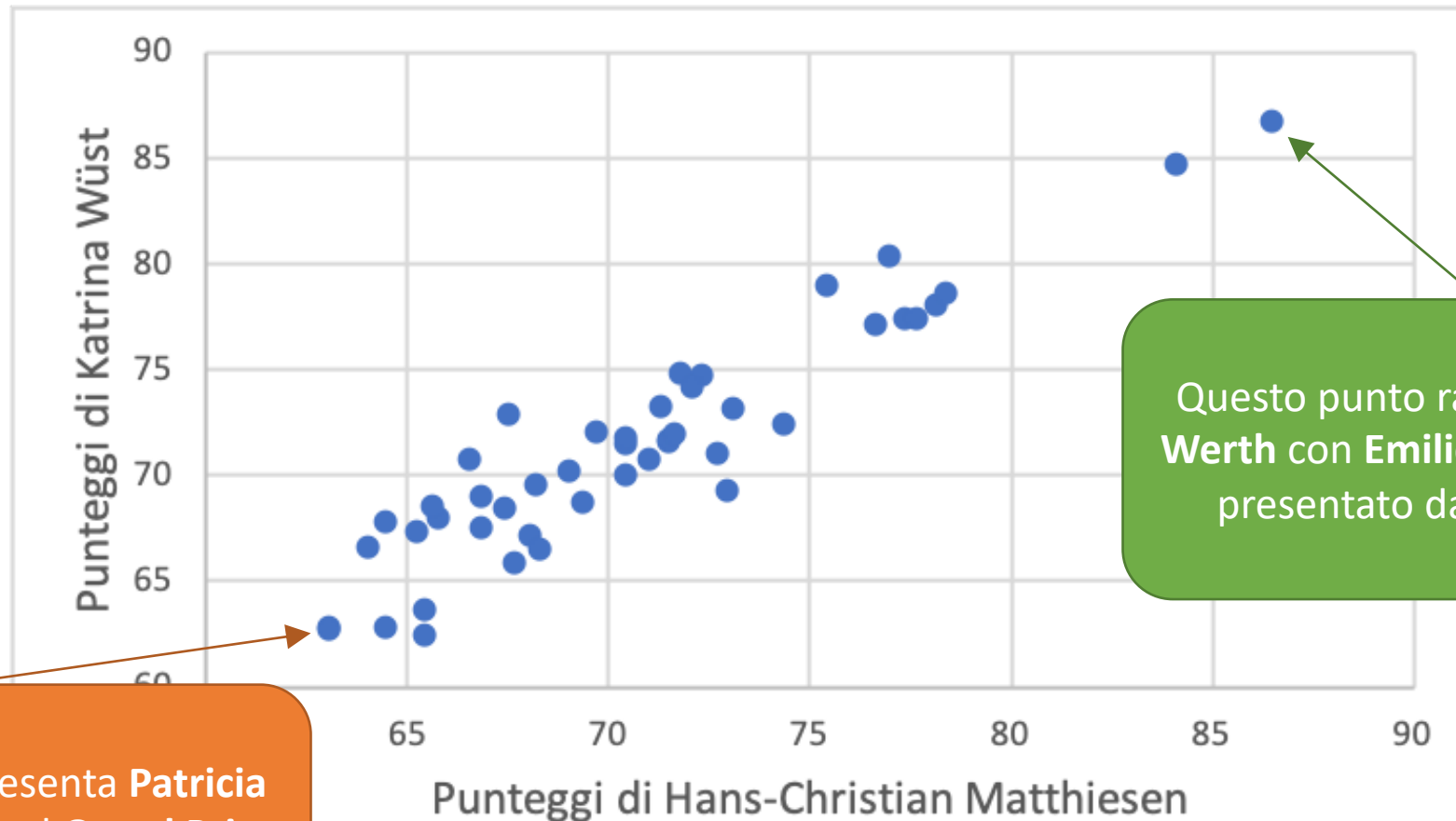


Diagramma a dispersione (scatterplot)

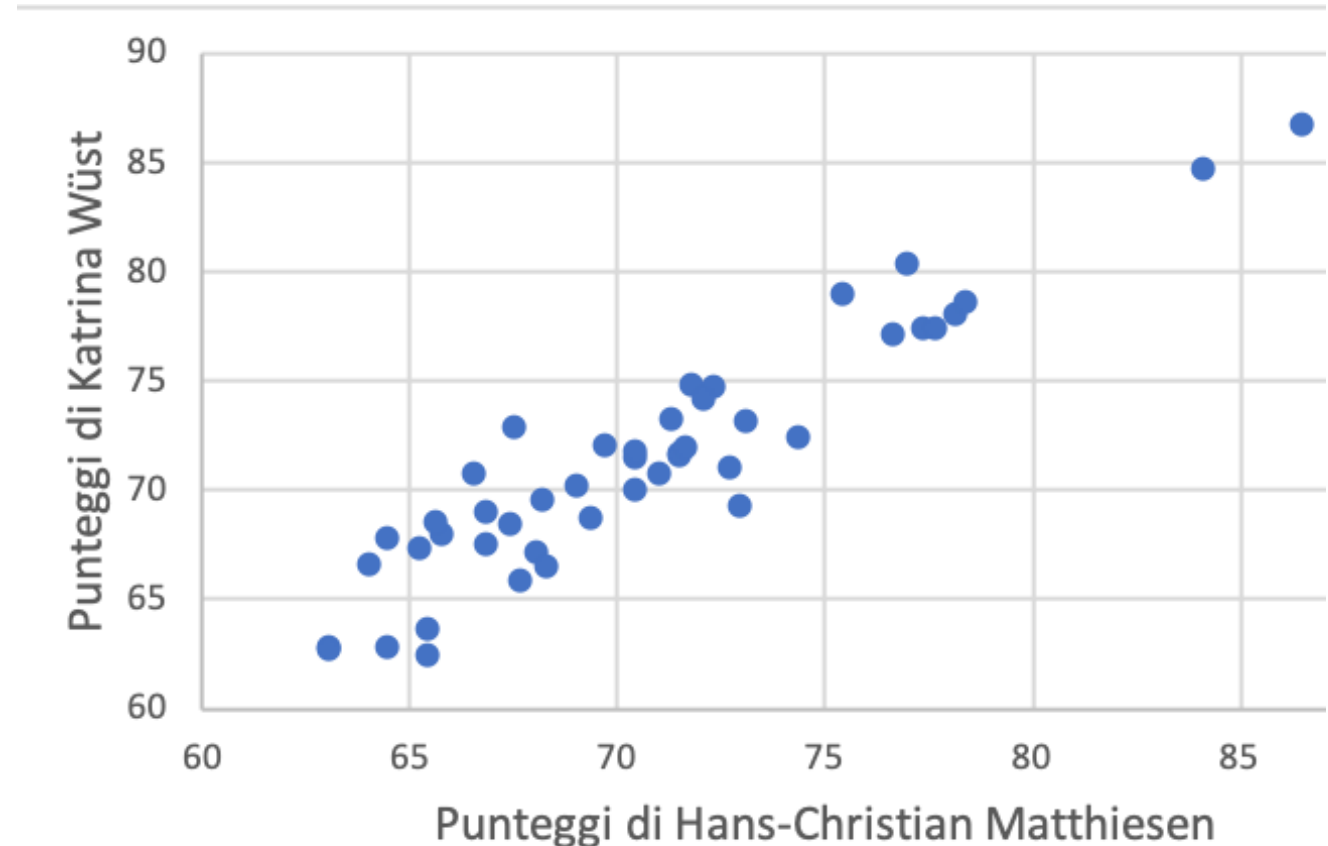


Questo punto rappresenta **Patricia Ferrando** con **Elvis** nel **Grand Prix** presentato da **iWEST**

Questo punto rappresenta **Isabell Werth** con **Emilio 107** nel **Freestyle** presentato da **Meggle GmbH**

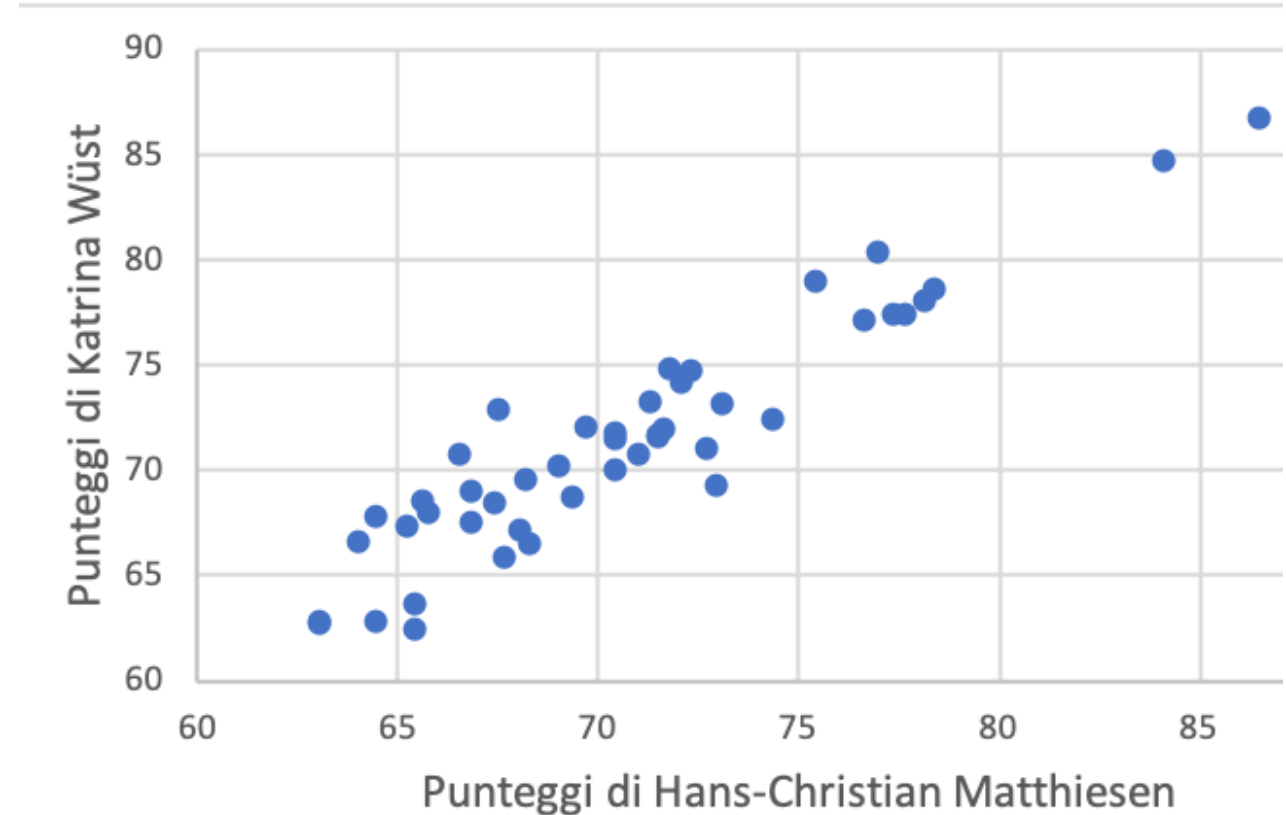
Esaminare un diagramma a dispersione

- Andamento generale
 - Forma (circolare o lineare)
 - Direzione (da basso a sinistra ad alto a destra o viceversa)
 - Forza (compattezza attorno alla retta centrale)
- Deviazioni (outlier)



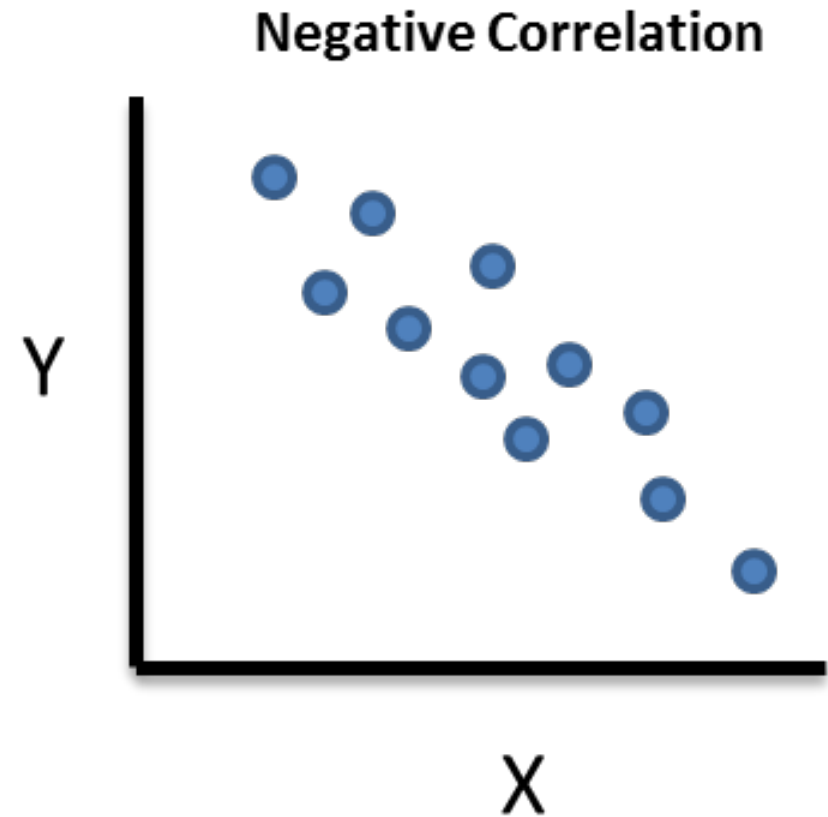
Associazioni positive

- Si ha quando i valori **sopra la media di una variabile tendono ad associarsi con i valori sopra la media dell'altra** e allo stesso modo si comportano i valori sotto la media



Associazioni negative

- Si ha **quando i valori sopra la media di una variabile tendono ad associarsi con i valori sotto la media dell'altra e viceversa.**



Misurare l'associazione lineare: correlazione

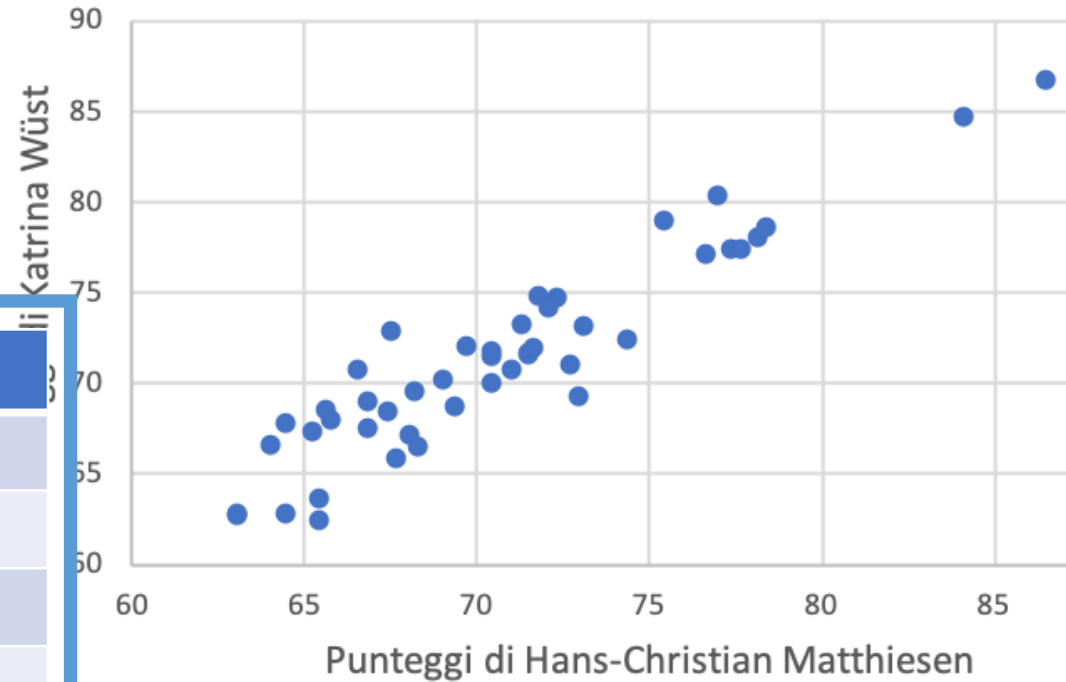
- Misura la **direzione e la forza** della relazione lineare tra due variabili quantitative x e y
- Indicata con la **lettera r**
- In prima istanza, **la correlazione r è la media dei prodotti delle misure standardizzate di x e y**
- x_i siano le n misure della variabile x , \bar{x} la media e s_x la deviazione standard, la **misura standardizzata** è

$$\frac{x_i - \bar{x}}{s_x}$$

Giudice 1	Giudice 2
70,435	71,522
67,5	72,935
69,022	70,217
70,435	70
68,043	67,174
65,217	67,391
65,761	68,043
65,435	62,5
63,043	62,826

Calcolare la correlazione

Giudice 1	Giudice 2	Z_1	Z_2	$Z_1 \cdot Z_2$
70,435	71,522	-0,04	0,03	0,00
67,5	72,935	-0,60	0,29	-0,17
69,022	70,217	-0,31	-0,21	0,07
70,435	70	-0,04	-0,25	0,01
68,043	67,174	-0,50	-0,77	0,38
65,217	67,391	-1,03	-0,73	0,75
65,761	68,043	-0,93	-0,61	0,57
65,435	62,5	-0,99	-1,62	1,61
...



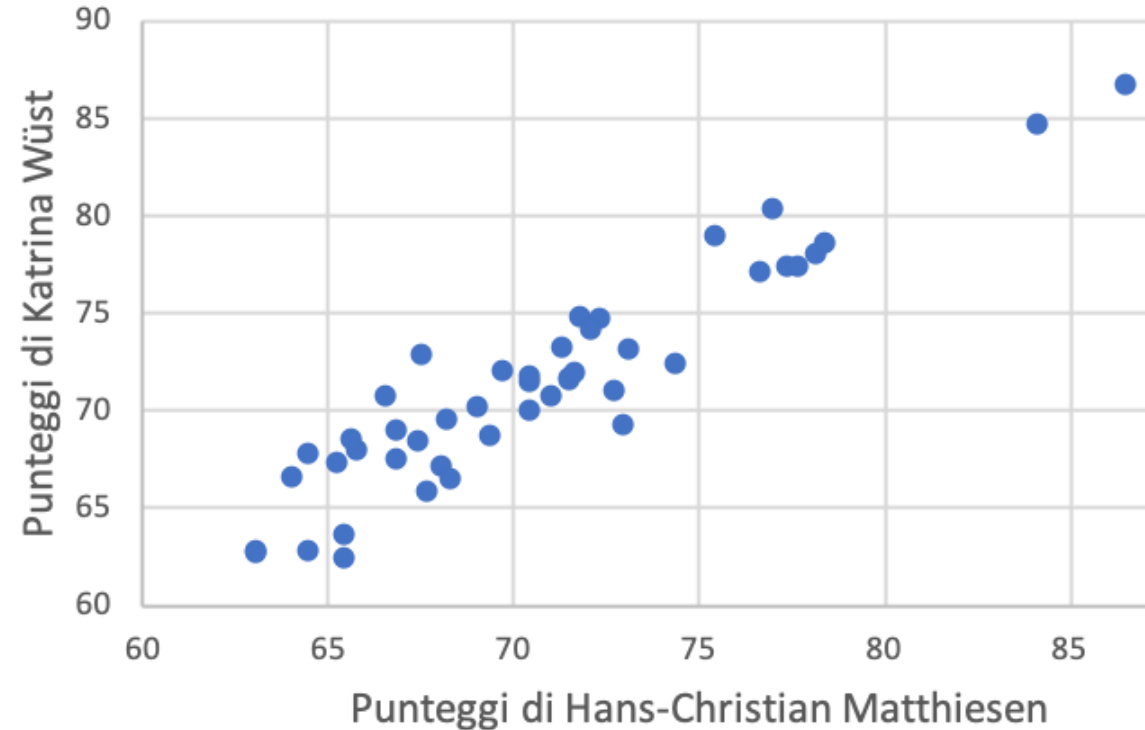
Media di $Z_1 \cdot Z_2$ con $n - 1$ gradi di libertà
= 0,94

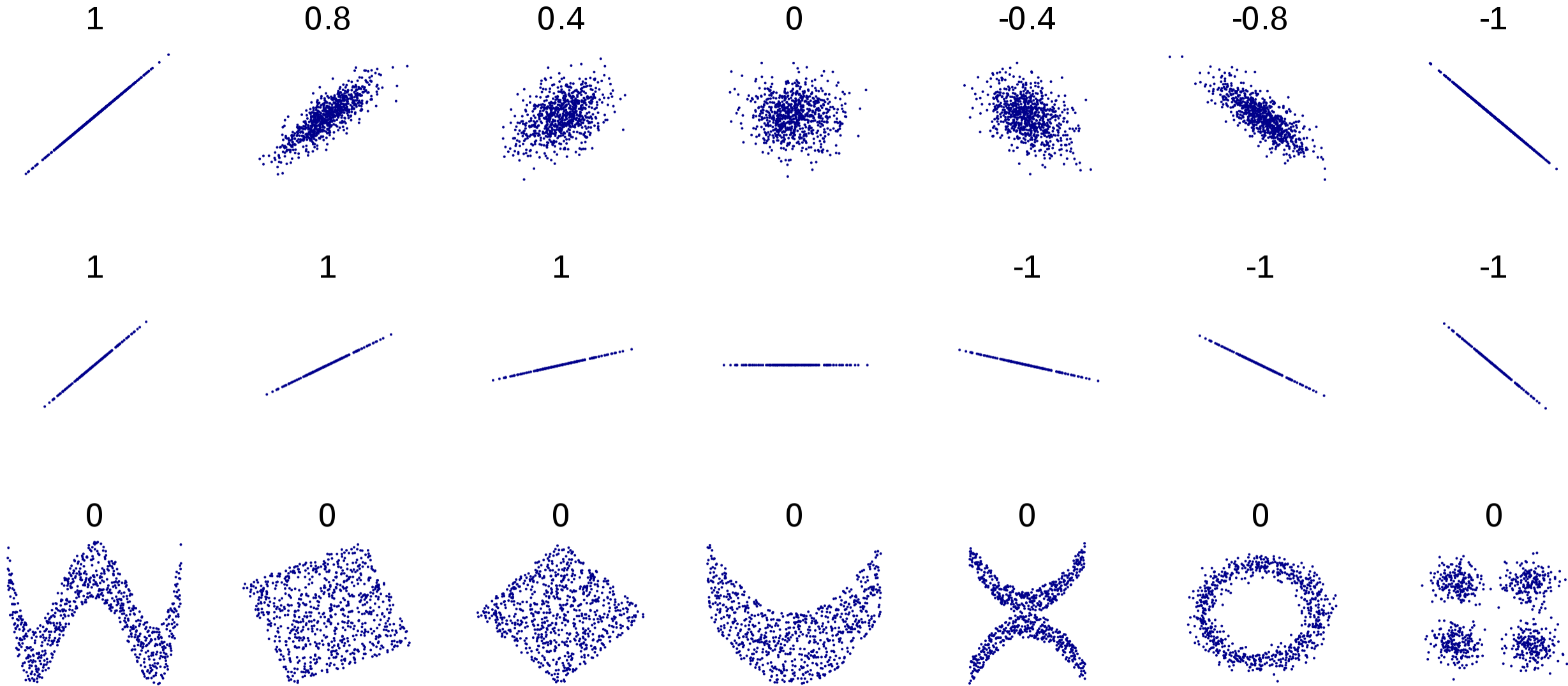
$$\begin{aligned} \overline{G1} &= 70,66 & \overline{G2} &= 71,37 & Z_{1i} &= (G1_i - \overline{G1})/s_1 \\ s_1 &= 5,26 & s_2 &= 5,47 & Z_{2i} &= (G2_i - \overline{G2})/s_2 \end{aligned}$$

Calcolare la correlazione

- x_i sono le n misure della variabile x , \bar{x} la media e s_x la deviazione standard
- y_i sono le n misure della variabile y , \bar{y} la media e s_y la deviazione standard

$$r = \frac{1}{n-1} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s_x} \right) \left(\frac{y_i - \bar{y}}{s_y} \right)$$





Correlazione

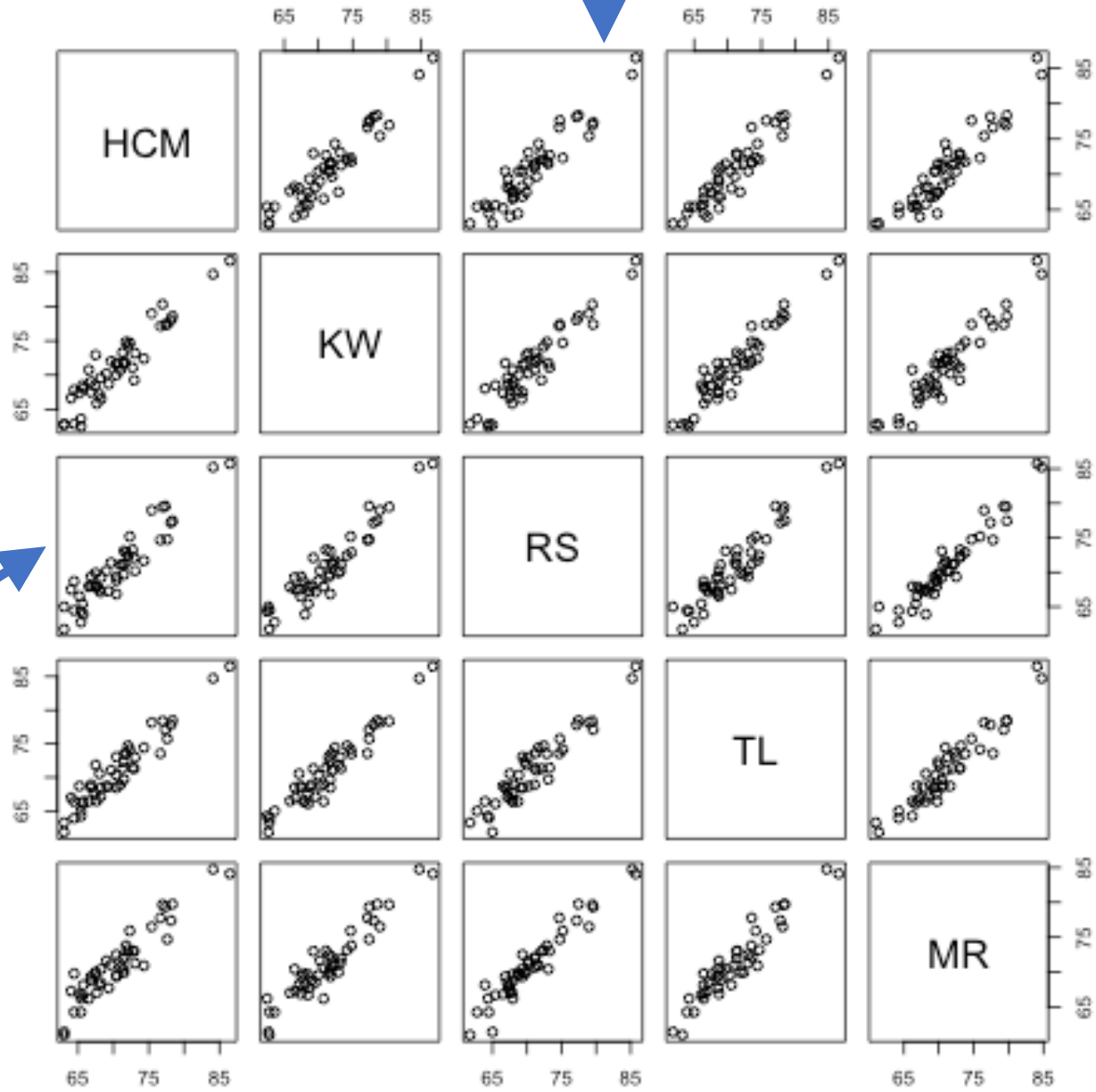
- Assume valori tra -1 e 1 . Correlazione 0 indica relazione lineare molto debole
- Non distingue tra variabili esplicative e di risposta
- Entrambe le variabili devono essere quantitative
- Cambiare le unità di misura delle variabili non cambia la loro correlazione
- Se r è positivo l'associazione è positiva (basso-sinistra verso alto-destra)
- Se r è negativo l'associazione è negativa (alto-sinistra verso basso-destra)
- Come la media e la deviazione standard, anche la correlazione non è una misura molto robusta (occhio agli outlier)

FEI Dressage World Cup Salzburg, 29-30 Nov. 2019

- Unità statistiche:
 - Cavalieri e cavalli, ipotesi semplificativa: sono diversi in ogni competizione
- Variabili statistiche:
 - Punteggi di ogni Giudice (5 giudici)
- Obiettivo:
 - Valutazioni simili dei giudici?



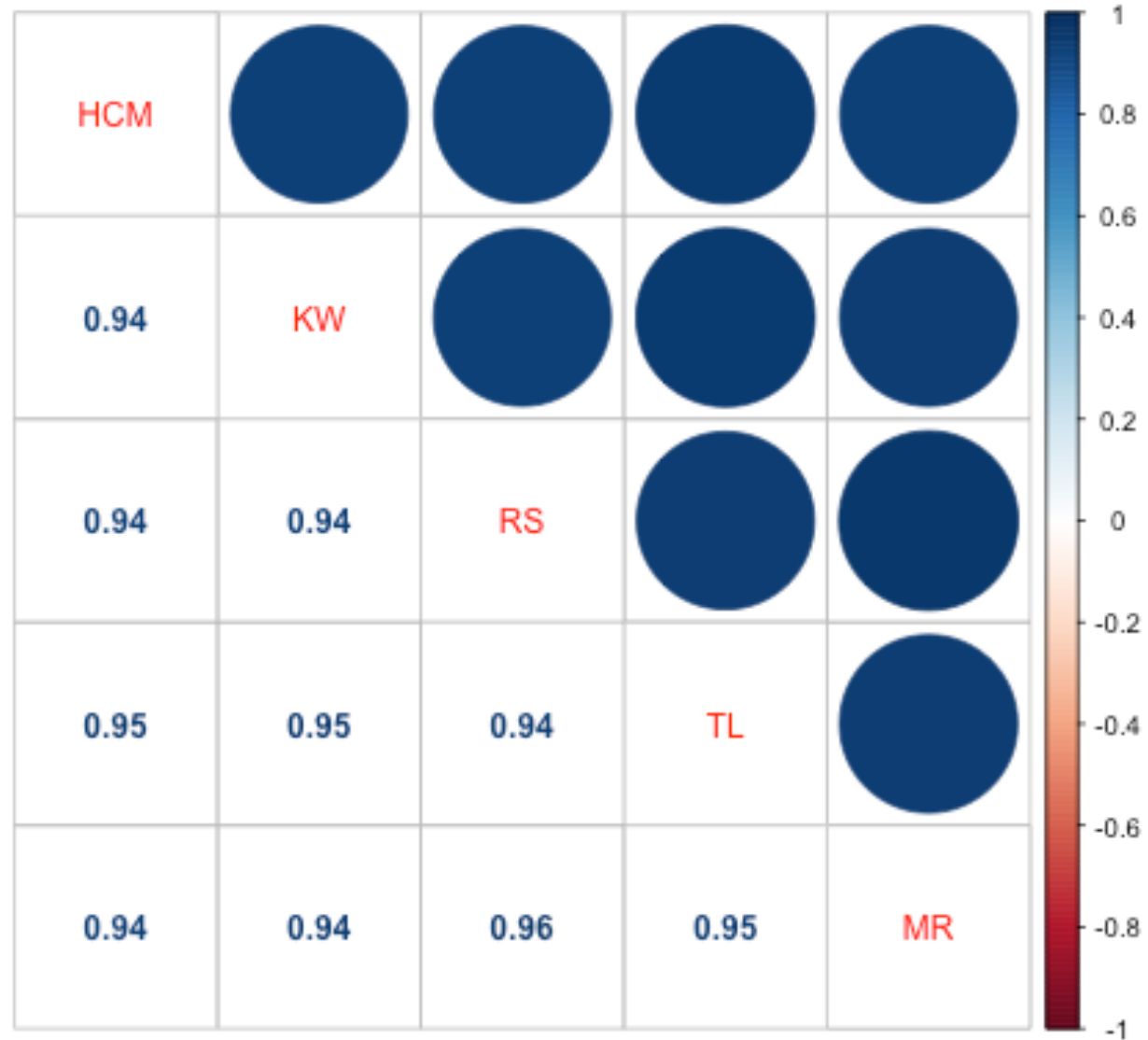
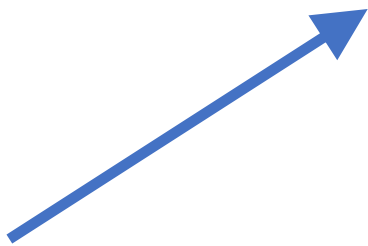
Correlazione tra HCM (sulle x) e RS (sulle y)



Correlazione tra RS (sulle x) e HCM (sulle y)

$$r(HCM, RS) = 0.94$$

$$r(RS, HCM) = 0.94$$



I giudici hanno seguito lo stesso schema di valutazione

FEI Dressage World Cup Salzburg

Esercizi

- Svolgere gli esercizi 4.9, 4.10, 4.11 a pag. 108 del libro di testo e caricare i relativi file Excel sulla propria cartella del canale Esercizi