

Correlation Analysis (Part 2)

Munmun Akter

Lecturer, NFE, DIU

2. Partial correlation

- It is used to explore the relationship between two variables while controlling for the effects of one or more additional variables, known as "control variables" or "covariates."
- The purpose of partial correlation is to isolate the unique association between the two main variables of interest, removing the influence of the control variables.

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{1 - r_{13}^2} \sqrt{1 - r_{23}^2}}$$

Partial Correlation

Useful with three variables

predictor variable

predicted variable

control variable

is

relationship between two variables while controlling for a third variable

type of Pearson correlation coefficient

Used in models which assume a linear relationship

data is supposed to be interval in nature

example relationship between height & weight, while controlling for age

assist in understanding regression

range in value from -1 to +1

conducted to understand why two variables are correlated

Data of academic achievement, anxiety and intelligence for 10 subjects

Subject	Academic Achievement	Anxiety	Intelligence
1	15	6	25
2	18	3	29
3	13	8	27
4	14	6	24
5	19	2	30
6	11	3	21
7	17	4	26
8	20	4	31
9	10	5	20
10	16	7	25

A

Academic Achievement

B

Anxiety

C

Intelligence

$$r_{AB} = -0.369 \qquad r_{AC} = 0.918$$

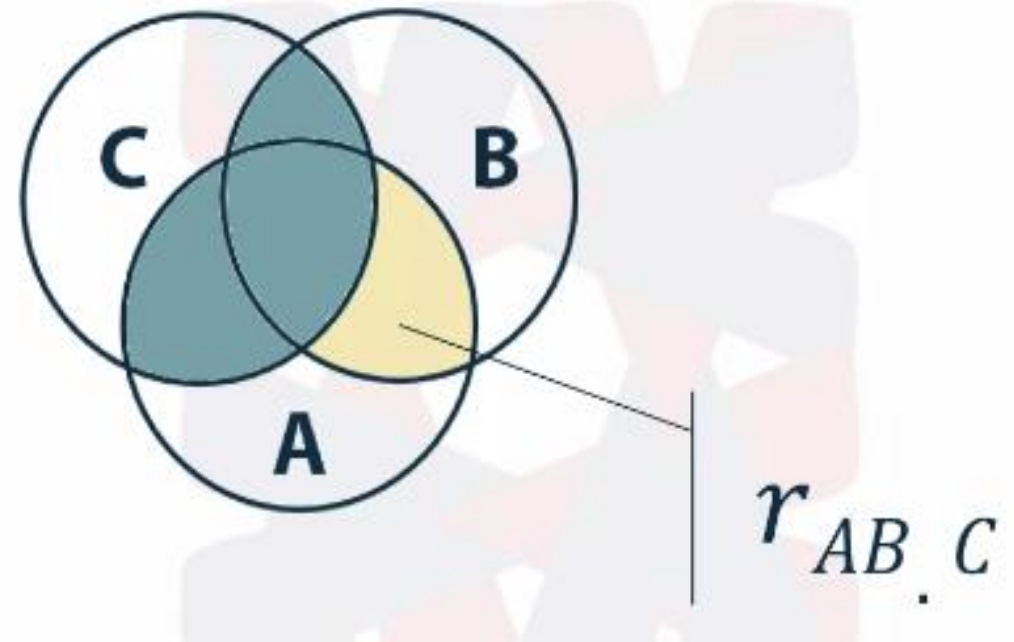
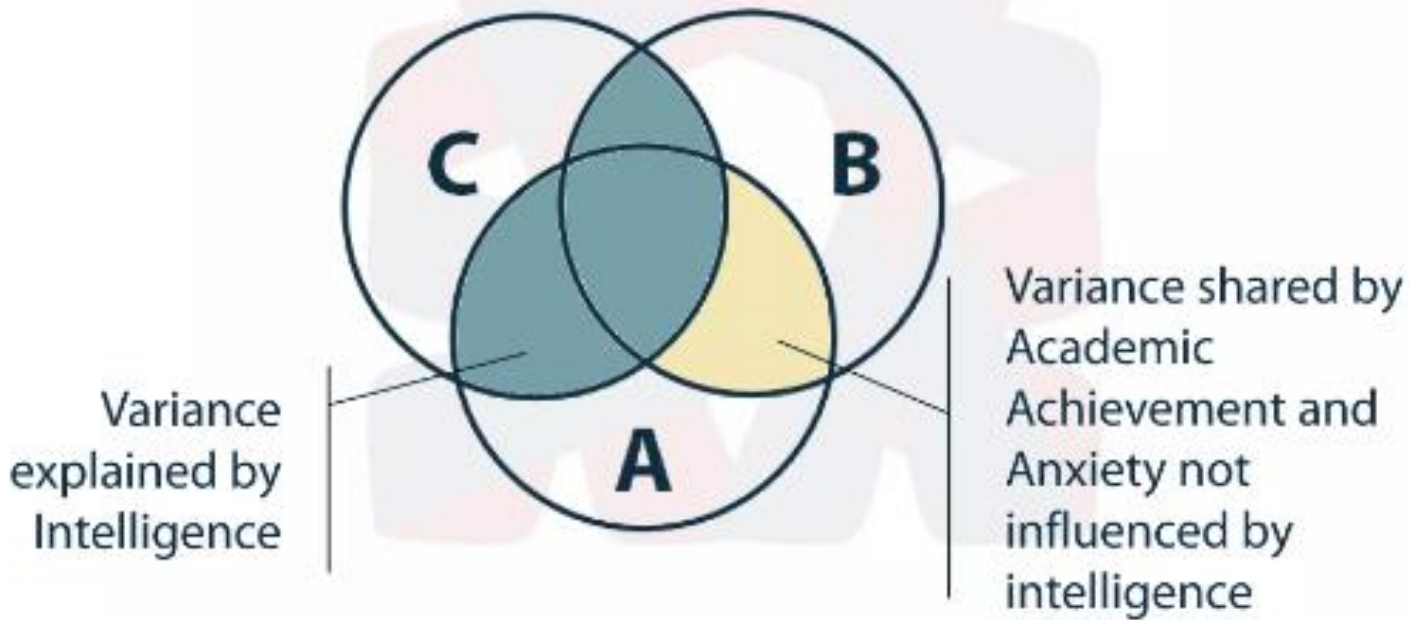
$$r_{BC} = -0.245$$

$$r_p = r_{AB,C} = \frac{(-0.369) - (0.918)(-0.245)}{\sqrt{(1 - 0.918^2)(1 - (-0.245)^2)}}$$

$$r_p = r_{AB,C} = \frac{-0.1441}{0.499}$$

$$r_p = r_{AB,C} = -0.375$$

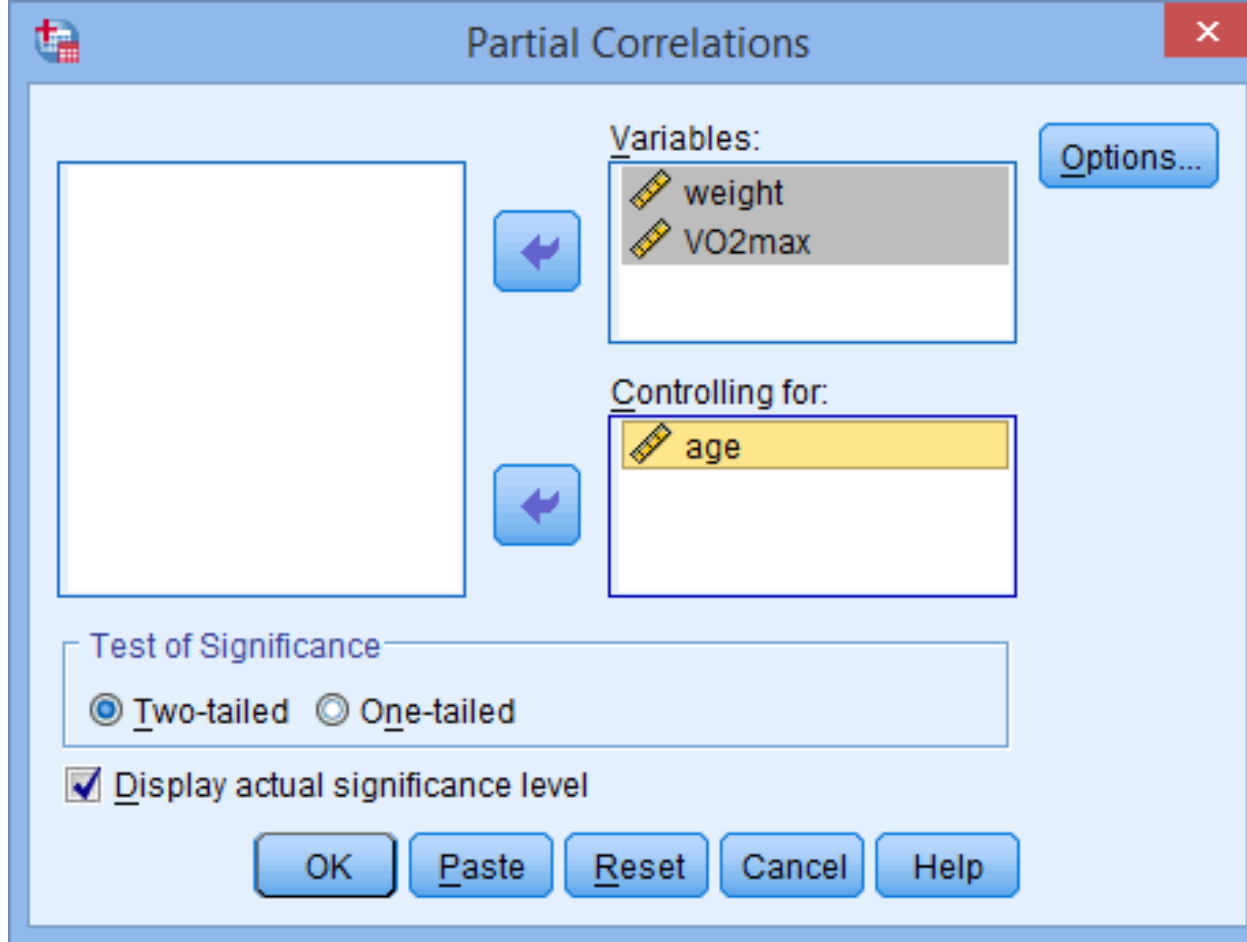
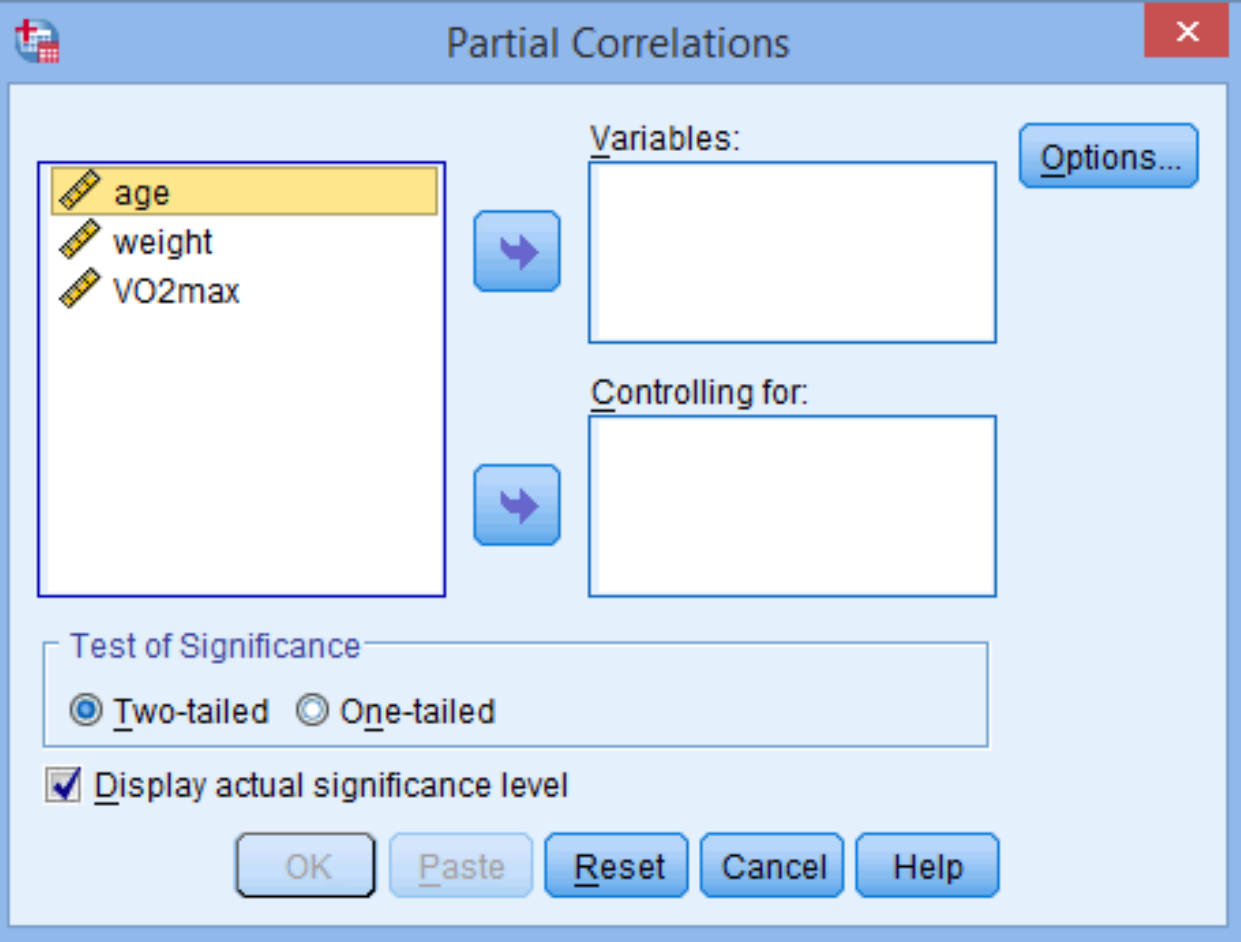
Graphical Explanation of partial correlation between Academic Achievement and Anxiety controlled for Intelligence



Calculate Partial correlation in SPSS

The screenshot displays the SPSS software interface. The 'Analyze' menu is open, and the 'Correlate' option is selected, which has opened a sub-menu where 'Partial...' is highlighted. The background data table is as follows:

	age	weight					
1	63	50.34					
2	64	73.38					
3	28	103.23					
4	52	88.83					
5	34	50.00					
6	26	62.59					
7	26	90.63					
8	53	78.67					
9	53	87.28					
10	24	72.29					
11	51	93.76					
12	48	97.46					
13	26	53.43					
14	32	64.00					
15	37	56.18					
16	31	112.59					
17	40	58.07	40.90				
18	28	115.42	38.12				
19	21	78.45	38.70				
20	24	60.90	37.09				



The **Correlations** table is split into two main parts:

(a) the Pearson product-moment correlation coefficients for all your variables – that is, your dependent variable, independent variable, and one or more control variables – as highlighted by the blue rectangle; and

(b) the results from the partial correlation where the Pearson product-moment correlation coefficient between the dependent and independent variable has been adjusted to take into account the control variable(s), as highlighted by the red rectangle.

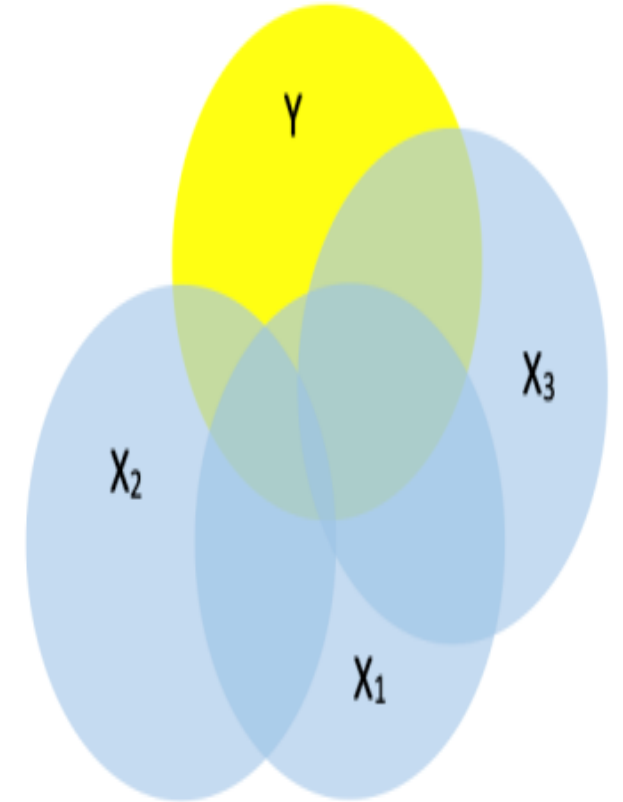
Correlations

Control Variables			Weight	VO2max	Age
-none- ^a	Weight	Correlation	1.000	-.307	-.004
		Significance (2-tailed)	.	.002	.972
		df	0	98	98
	VO2max	Correlation	-.307	1.000	-.191
		Significance (2-tailed)	.002	.	.057
		df	98	0	98
	Age	Correlation	-.004	-.191	1.000
		Significance (2-tailed)	.972	.057	.
		df	98	98	0
Age	Weight	Correlation	1.000	-.314	
		Significance (2-tailed)	.	.002	
		df	0	97	
	VO2max	Correlation	-.314	1.000	
		Significance (2-tailed)	.002	.	
		df	97	0	

a. Cells contain zero-order (Pearson) correlations.

3. Multiple Correlation:

- It examines the relationship between one dependent variable and two or more independent variables, an extension of simple correlation.
- In multiple correlation, the goal is to understand how a single dependent variable is related to a combination of several independent variables.
- Multiple correlation is often used in regression analysis, where the dependent variable is predicted based on the values of multiple independent variables.
- The multiple correlation coefficient is represented by "R" and is the square root of the coefficient of determination (R-squared) in regression analysis. R-squared represents the proportion of the variance in the dependent variable that is explained by the independent variables. It ranges from 0 to 1.



Multiple Correlation Coefficient

The formula for R is

$$R = \sqrt{\frac{r_{yx_1}^2 + r_{yx_2}^2 - 2r_{yx_1} \cdot r_{yx_2} \cdot r_{x_1x_2}}{1 - r_{x_1x_2}^2}}$$

where

r_{yx_1} = correlation coefficient for y and x_1

r_{yx_2} = correlation coefficient for y and x_2

$r_{x_1x_2}$ = correlation coefficient for x_1 and x_2

Problem 3

If the simple correlation coefficients have the values $r_{12} = 0.6$, $r_{13} = 0.65$, $r_{23} = 0.8$, find the multiple correlation coefficient $R_{1,23}$.

Solution:

We have

$$\begin{aligned} R_{1,23} &= \frac{\sqrt{r_{12}^2 + r_{13}^2 - 2 r_{12} r_{13} r_{23}}}{\sqrt{1 - r_{23}^2}} \\ &= \frac{\sqrt{(0.6)^2 + (0.65)^2 - 2 \times 0.6 \times 0.65 \times 0.8}}{\sqrt{1 - (0.8)^2}} \end{aligned}$$

Data of academic achievement, anxiety and intelligence for 10 subjects

Subject	Academic Achievement	Anxiety	Intelligence
1	15	6	25
2	18	3	29
3	13	8	27
4	14	6	24
5	19	2	30
6	11	3	21
7	17	4	26
8	20	4	31
9	10	5	20
10	16	7	25

A= Academic achievement

B= Anxiety

C= Intelligence

$$r_{AB} = -0.369$$

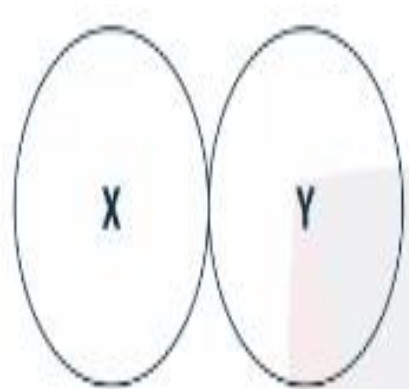
$$r_{AC} = 0.918$$

$$r_{BC} = -0.245$$

$$R_{A.BC} = 0.929$$

Interpretation of multiple correlation coefficient

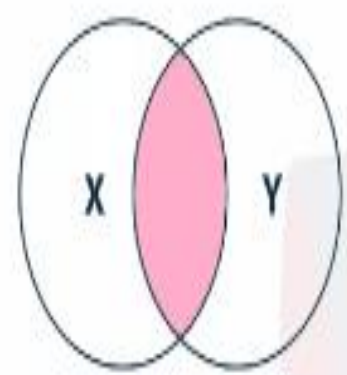
- What is the meaning of $R_{A.BC} = 0.929$?
- This means that the multiple correlation between academic achievement and the linear combination of intelligence and anxiety is 0.929 or 0.93.
- What is R^2 ?
- The R^2 is the percentage of variance in academic achievement explained by the linear combination of intelligence and anxiety.
- In this sample $R=0.929$ or 0.93; so, the value of R^2 is about 0.865. It means that the linear combination of intelligence and anxiety explain 86.5 percent variance in the academic achievement.



$R = 0$

$R^2 = 0$

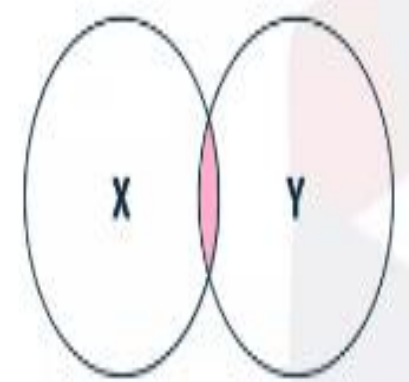
Variance = 0%



$R = \pm 0.6$

$R^2 = 0.36$

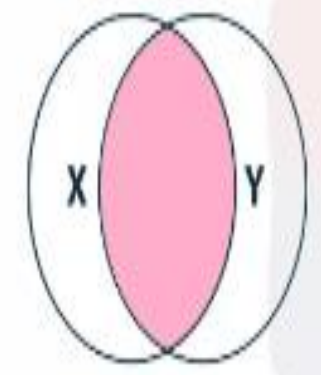
Variance = 36%



$R = \pm 0.2$

$R^2 = 0.04$

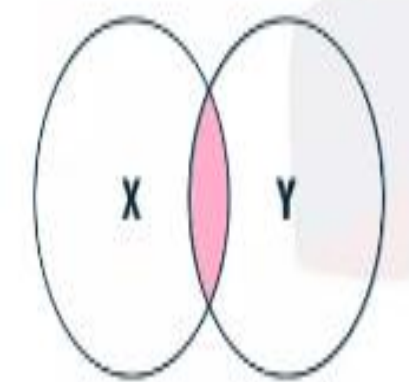
Variance = 4%



$R = \pm 0.8$

$R^2 = 0.64$

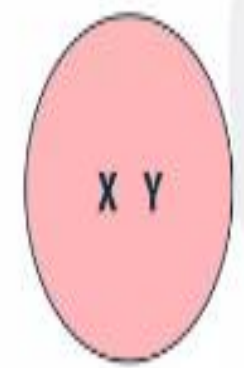
Variance = 64%



$R = \pm 0.4$

$R^2 = 0.16$

Variance = 16%



$R = \pm 1$

$R^2 = 1$

Variance = 100%

Multiple Correlation Coefficient

Population Value

Sample Value

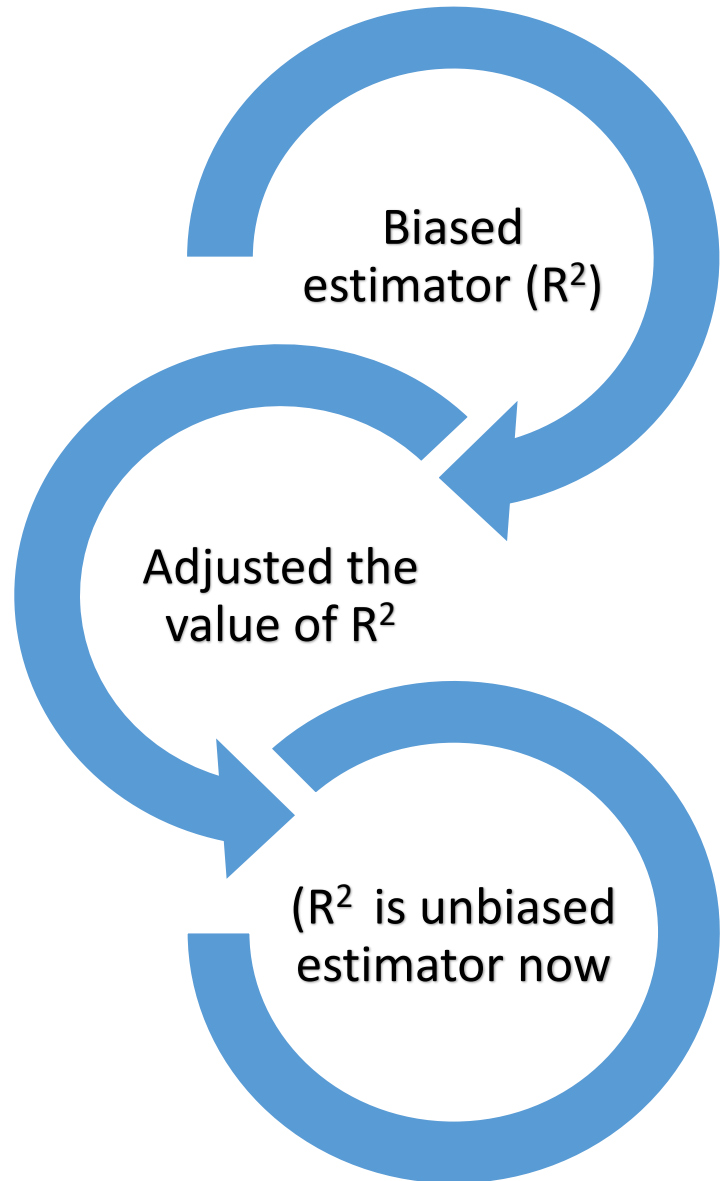
ρ^2

R^2

Estimated from



Reduced biasness



A biased estimator is one that deviates from the true population value.

An unbiased estimator is one that does not deviate from the true population parameter.

$$\text{Formula: } R^{*2} = 1 - (1 - R^2)(n-1)/(n-k-1)$$

Where,

n = number of participants / sample size

k = number of predicted variables

R^{*2} = is the adjusted value of R

$$\tilde{R}^2 = 1 - \frac{(1 - R^2)(n - 1)}{n - k - 1}$$

$$= 1 - \frac{(1 - 0.865)(10 - 1)}{10 - 2 - 1}$$

$$= 1 - \frac{1.217}{7}$$

$$= 0.826$$

Multiple Correlation Coefficient by SPSS

The screenshot shows the IBM SPSS Statistics main window. The 'Analyze' menu is open, and the 'Regression' option is selected, which has opened a sub-menu where 'Linear...' is highlighted. The background shows a data editor with columns for 'caseno', 'age', 'gender', 'VO2max', and 'var'. The 'Analyze' menu items include: Reports, Descriptive Statistics, Compare Means, General Linear Model, Generalized Linear Models, Mixed Models, Correlate, Regression, Loglinear, Classify, Dimension Reduction, Scale, Nonparametric Tests, Forecasting, Survival, Multiple Response, Simulation..., Quality Control, and ROC Curve... The 'Regression' sub-menu items are: Automatic Linear Modeling..., Linear..., Curve Estimation..., Partial Least Squares..., Binary Logistic..., Multinomial Logistic..., Ordinal..., Probit..., Nonlinear..., Weight Estimation..., and 2-Stage Least Squares... The data editor shows the following data:

	caseno	age	gender	VO2max	var
1	1	28	Male	55.79	
2	2	6	Female	35.00	
3	3	3	Male	42.93	
4	4				
5	5				
6	6				
7	7				
8	8				
9	9				
10	10				
11	11				
12	12				
13	13				
14	14				
15	15				
16	16				
17	17	37	Male	56.18	163
18	18	30	Male	86.13	156

The screenshot shows the 'Linear Regression' dialog box. The 'Dependent' variable is 'VO2max'. The 'Independent(s)' list contains 'age', 'weight', 'heart_rate', and 'gender'. The 'Method' is set to 'Enter'. The 'Selection Variable' field is empty. The 'Case Labels' field is empty. The 'WLS Weight' field is empty. The 'Method' dropdown is set to 'Enter'. The 'Previous' and 'Next' buttons are visible. The 'Statistics...', 'Plots...', 'Save...', and 'Options...' buttons are also visible. The 'OK', 'Paste', 'Reset', 'Cancel', and 'Help' buttons are at the bottom.

Scatter Plot or Scatter Diagram

- The scatter plot and correlation complement each other in understanding the relationship between two variables.
- It is a type of graphical representation used to visualize the relationship between two continuous variables.
- Each data point is plotted as a dot on the graph, with one variable represented on the x-axis and the other variable on the y-axis.
- When the points in the scatter plot are tightly clustered along a straight line, the correlation coefficient tends to be close to +1 or -1, indicating a strong linear relationship.
- When the points in the scatter plot are scattered randomly with no apparent pattern, the correlation coefficient tends to be close to 0, indicating a weak or no linear relationship.

Scatter Plots & Correlation Examples

