

Relationships

Computational Social Intelligence - Lecture 08

Prof. Alessandro Vinciarelli
School of Computing Science &
Institute of Neuroscience and Psychology

<http://www.dcs.gla.ac.uk/vincia>
Alessandro.Vinciarelli@glasgow.ac.uk



University
of Glasgow

EPSRC

Engineering and Physical Sciences
Research Council



This lecture is based on the following text (available on Moodle):

- D.C.Howell, "Statistical Methods for Psychology", Chapter 9, pp. 244-257 (included), pp. 273-274 (included), Cengage Learning, 2009.

Outline

- Introduction
- Regression
- Correlation
- Conclusions

Outline

- Introduction
- Regression
- Correlation
- Conclusions

Relationships

“When we are concerned with relationships [...] the experimenter is interested in showing that the dependent variable is some function of the independent variable.”

D.C.Howell, “Statistical Methods for Psychology”, Chapter 9,
Cengage Learning, 2009.

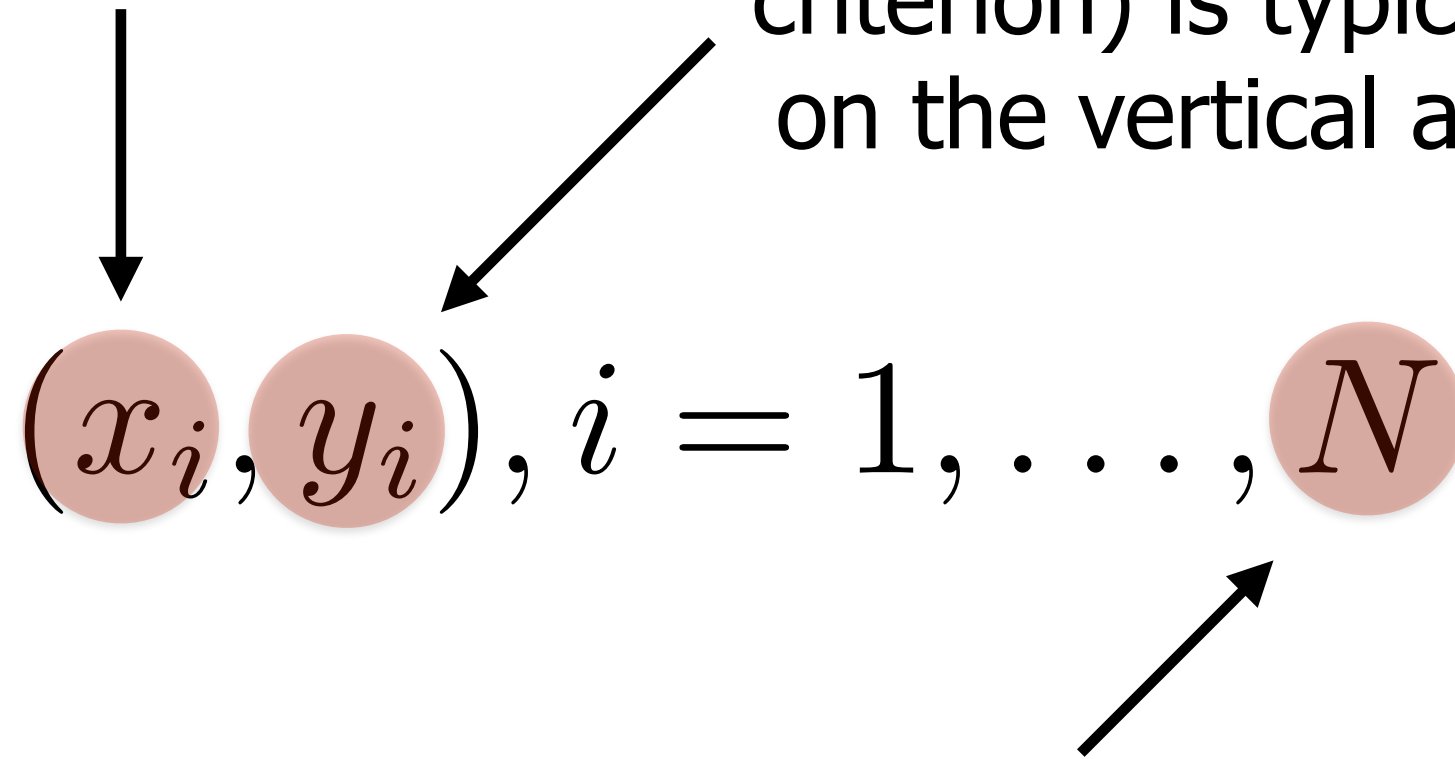
Scatterplot

“One of the most useful techniques for gaining insight into [a] relationship is a scatterplot (also called a scatter diagram or scattergram).”

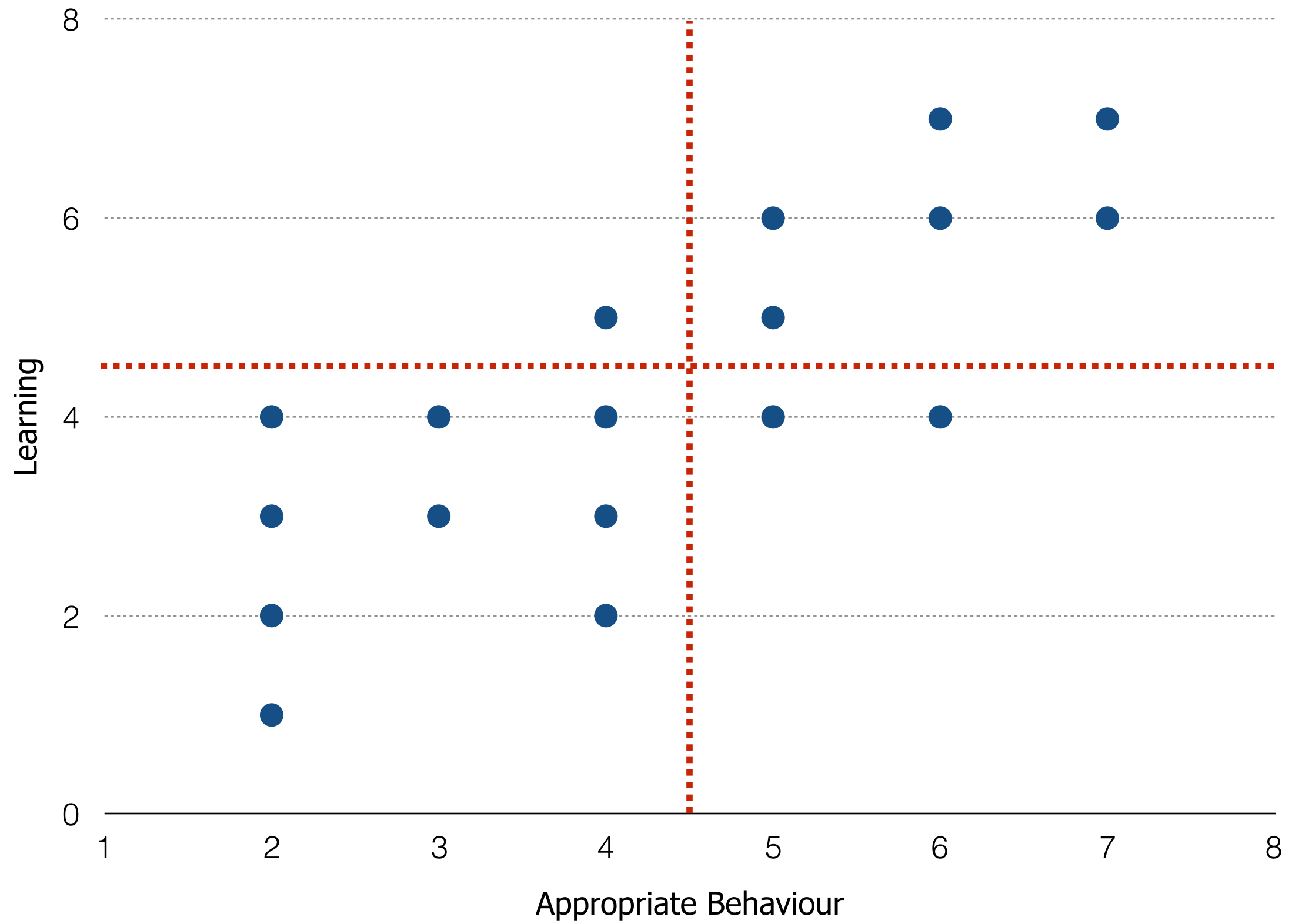
D.C.Howell, “Statistical Methods for Psychology”, Chapter 9,
Cengage Learning, 2009.

The independent
variable (the
predictor) is typically
on the horizontal axis

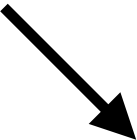
The dependent
variable (the
criterion) is typically
on the vertical axis


$$(x_i, y_i), i = 1, \dots, N$$

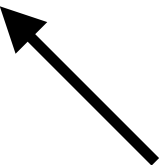
The number of pairs
at disposition



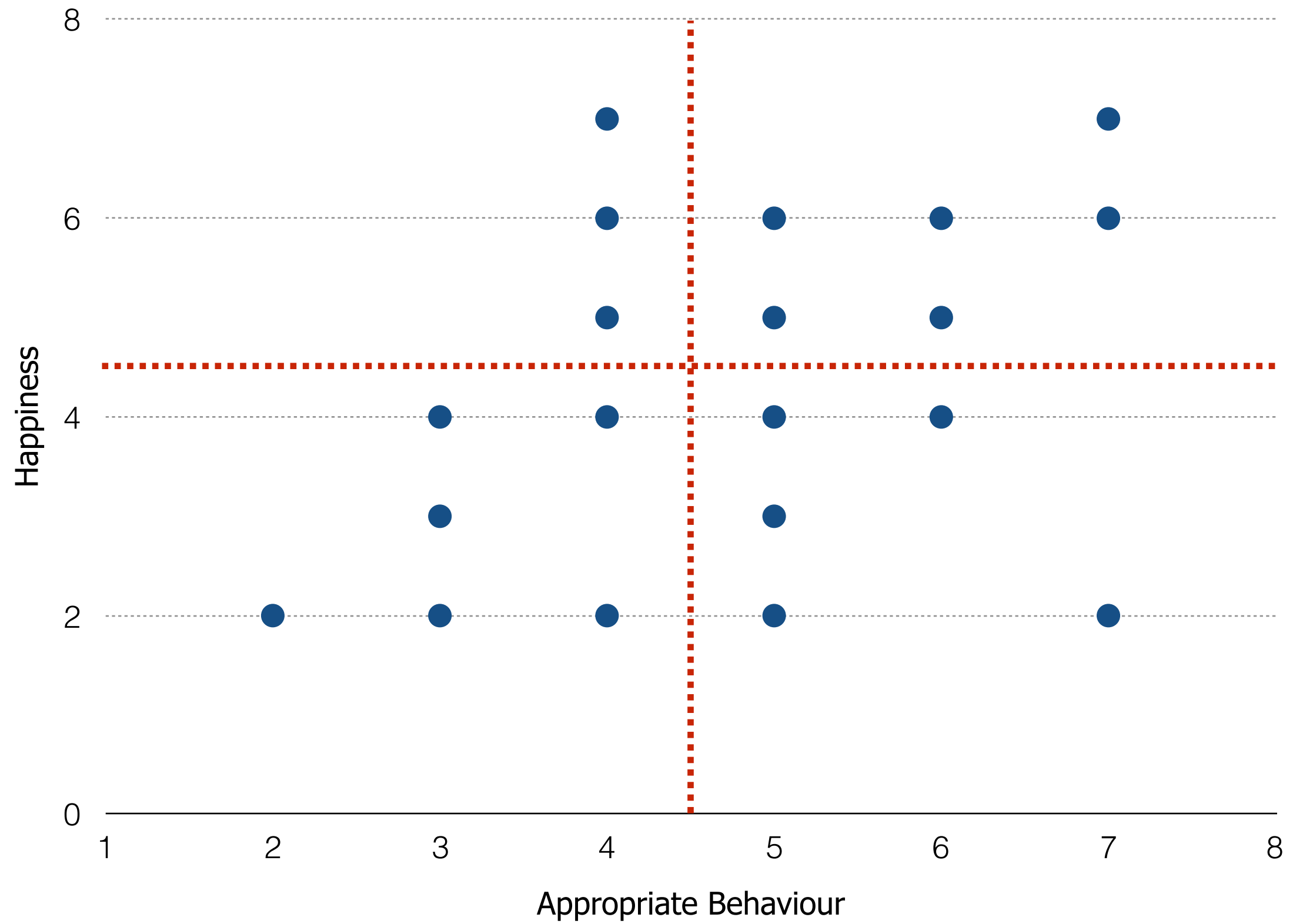
Appropriate
Behaviour



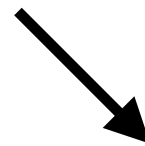
	Above Mean	Below Mean
Above Mean	7	1
Below Mean	2	10



Learning

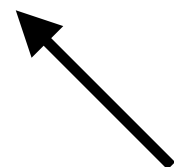


Happiness



	Above Mean	Below Mean
Above Mean	6	3
Below Mean	5	6

Appropriate
Behaviour



The covariance of X
and Y

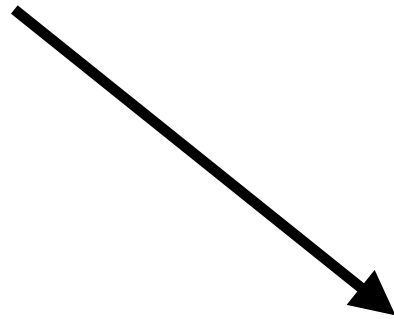
The average
value of X

The average
value of Y

The diagram illustrates the formula for covariance, σ_{xy} , with several explanatory labels and arrows. The label 'The covariance of X and Y' points to the symbol σ_{xy} . The label 'The average value of X' points to the term \bar{x} in the formula. The label 'The average value of Y' points to the term \bar{y} in the formula. The label 'The total number of pairs (x,y)' points to the denominator $N - 1$. The formula itself is
$$\sigma_{xy} = \frac{\sum_{i=1}^N (x_i - \bar{x})(y - \bar{y})}{N - 1}$$

The total number of
pairs (x,y)

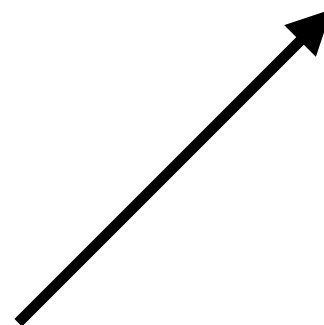
Covariance of
Appropriate
Behaviour and
Learning



$$\sigma_{xy} = 2.34$$

$$\sigma_{xy} = 0.96$$

Covariance of
Appropriate
Behaviour and
Happiness



Covariance

- The expression of the covariance is symmetric with respect to X and Y, its value does not change by switching predictor and criterion;
- It captures the relationship between variables as a systematic tendency to be on the same (or opposite) side of the mean;
- The value of the covariance is difficult to interpret.

Recap

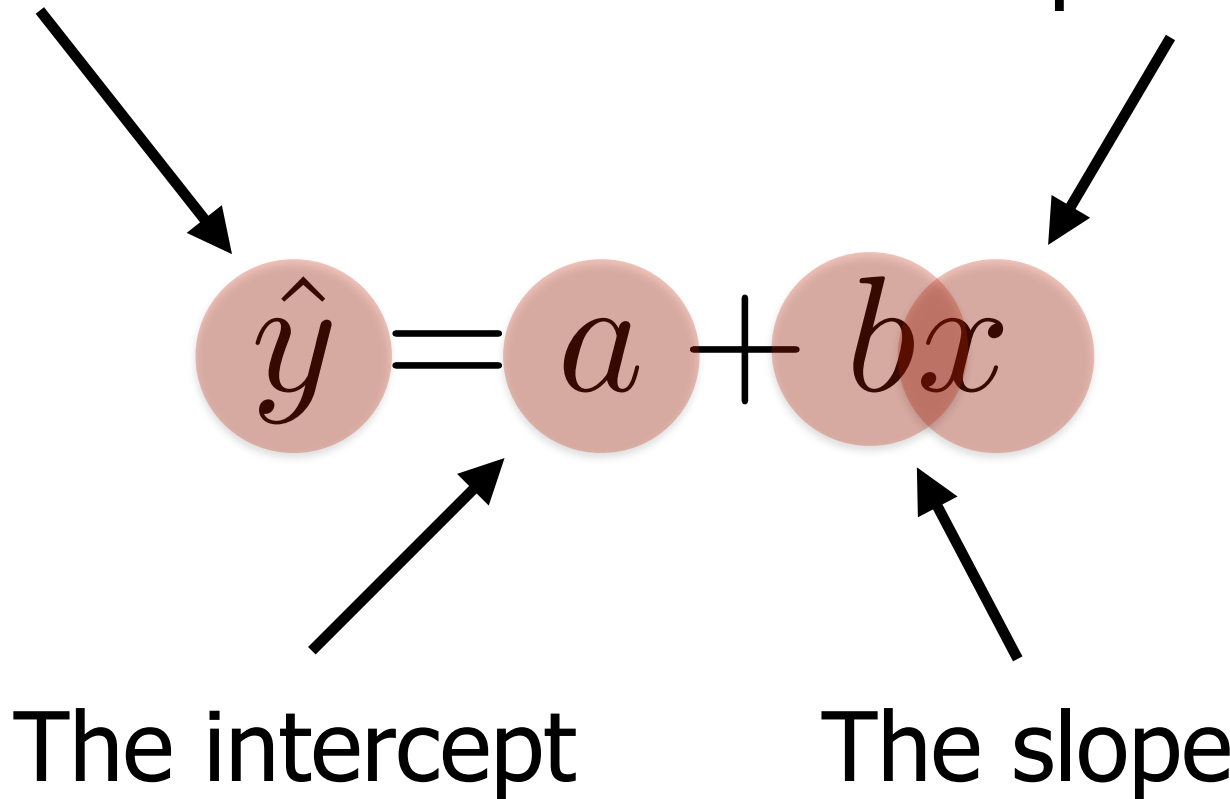
- Some random variables are expected to change according to one another;
- Conventionally, one of the two variables acts as predictor while the other acts as criterion;
- Switching between predictor and criterion does not change the results, the relationship has no direction.

Outline

- Introduction
- **Regression**
- Correlation
- Conclusions

The estimated value
of y when using a
linear plot

The predictor (the
independent variable)



The diagram shows the linear regression equation $\hat{y} = a + bx$ centered in the image. Each term (\hat{y} , a , b , and x) is enclosed in a light red circle. Four arrows point from descriptive text labels to these terms: one from the top-left label to \hat{y} , one from the top-right label to x , one from the bottom-left label to a , and one from the bottom-right label to b .

$$\hat{y} = a + bx$$

The intercept

The slope

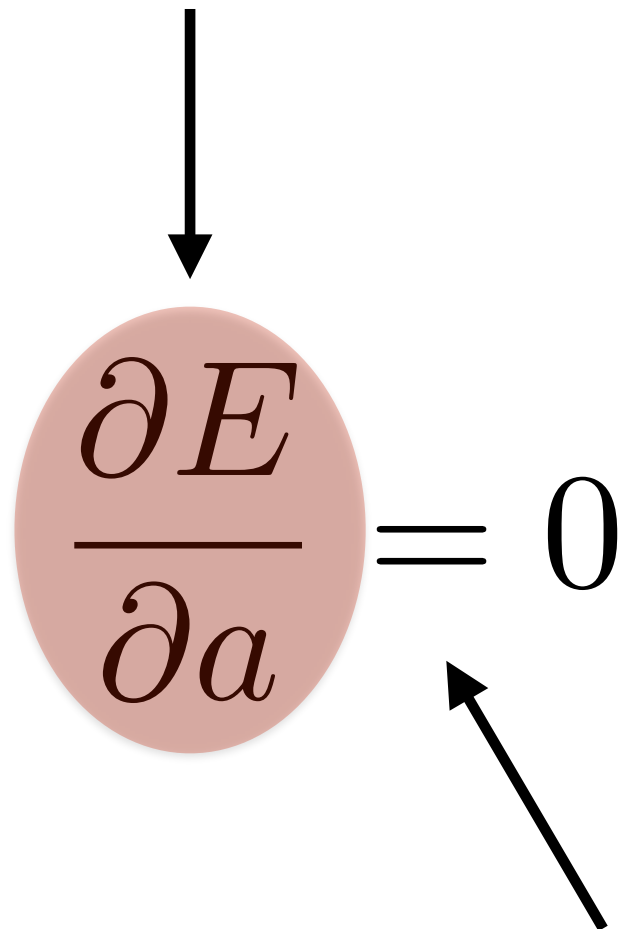
The prediction error

The estimated value
of y when using a
linear plot

$$E = \sum_{i=1}^N (y_i - \hat{y}_i)^2 = \sum_{i=1}^N (y_i - a - bx_i)^2$$

What is the "best" value for
slope and intercept?

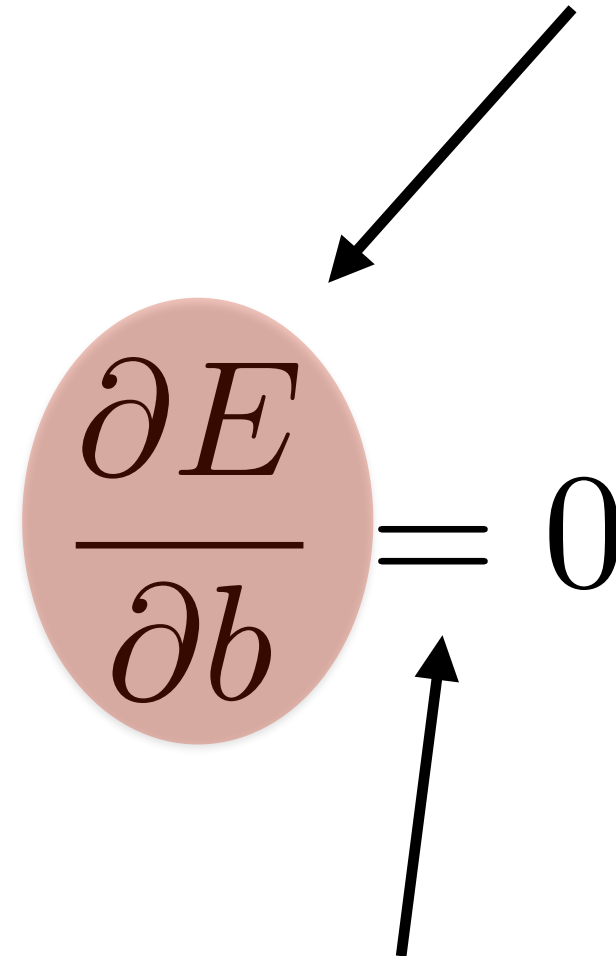
The derivative of the Error with respect to the intercept



A diagram showing the derivative of the error with respect to the intercept. A red oval contains the mathematical expression $\frac{\partial E}{\partial a}$. To the right of the oval is an equals sign followed by a zero. A black arrow points from the text above down to the oval, and another black arrow points from the text below up to the equals sign.

$$\frac{\partial E}{\partial a} = 0$$

The derivative of the Error with respect to the slope



A diagram showing the derivative of the error with respect to the slope. A red oval contains the mathematical expression $\frac{\partial E}{\partial b}$. To the right of the oval is an equals sign followed by a zero. A black arrow points from the text above down to the oval, and another black arrow points from the text below up to the equals sign.

$$\frac{\partial E}{\partial b} = 0$$

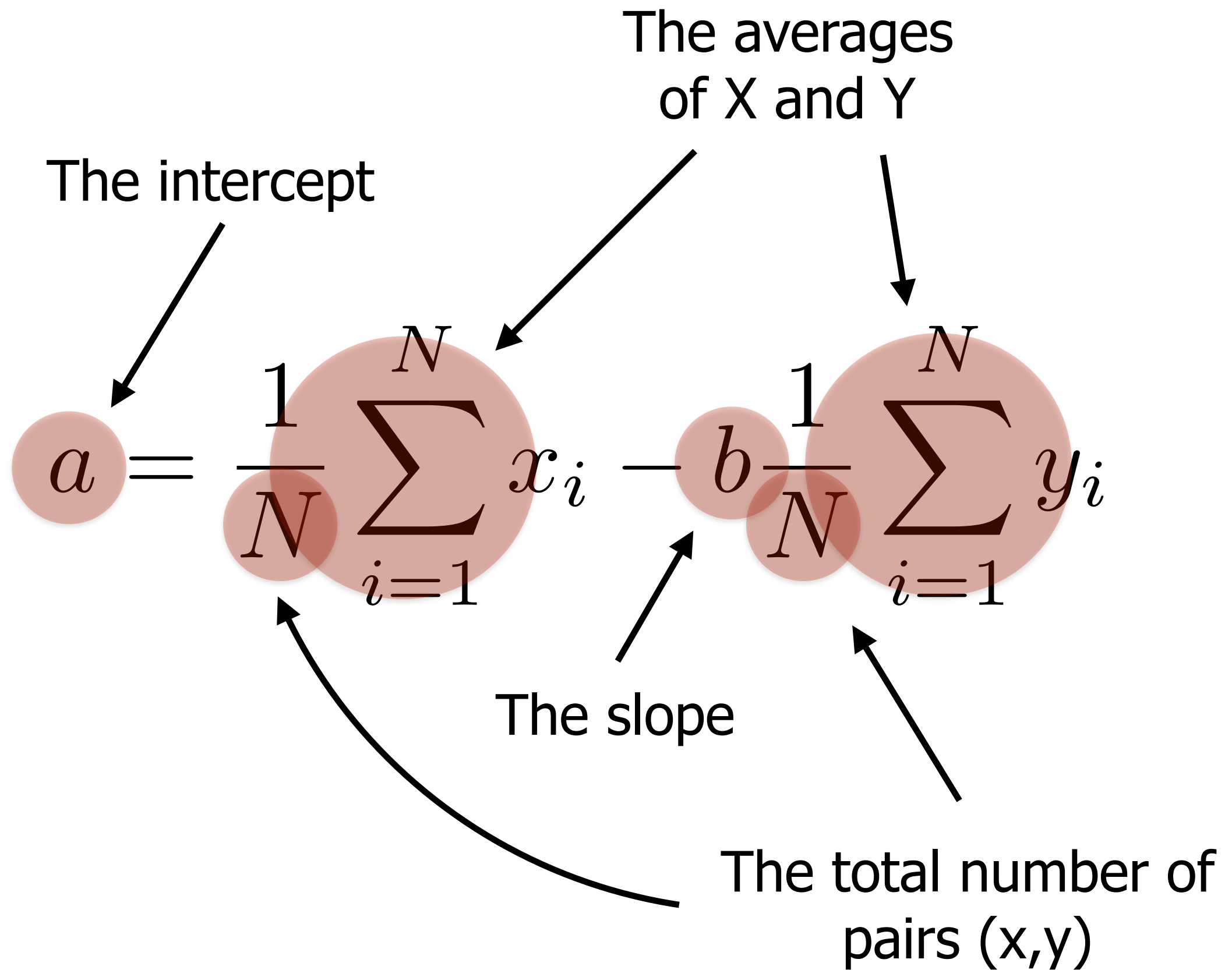
The results of the equations are the values of intercept and slope that correspond to the minimum Error

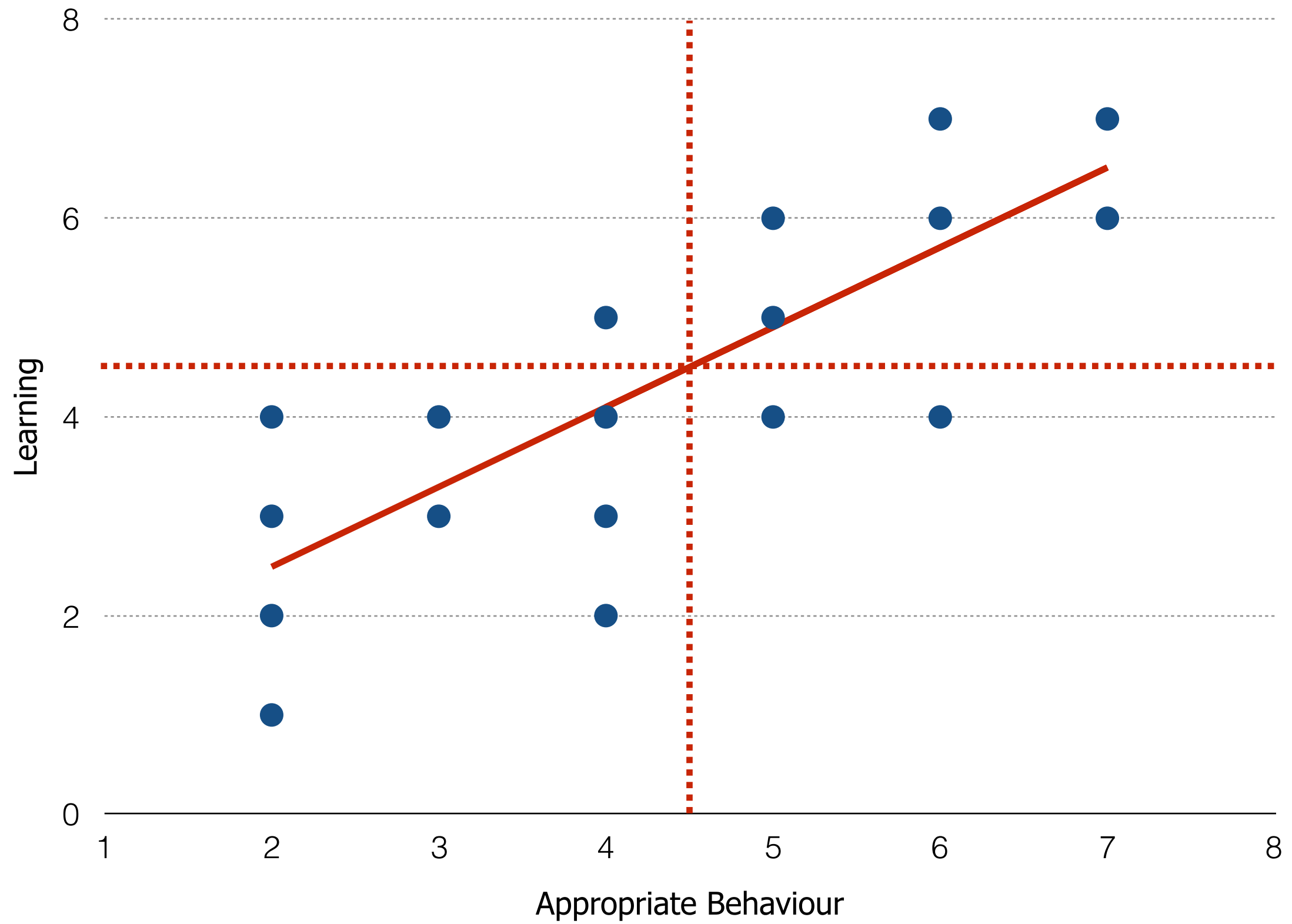
The slope

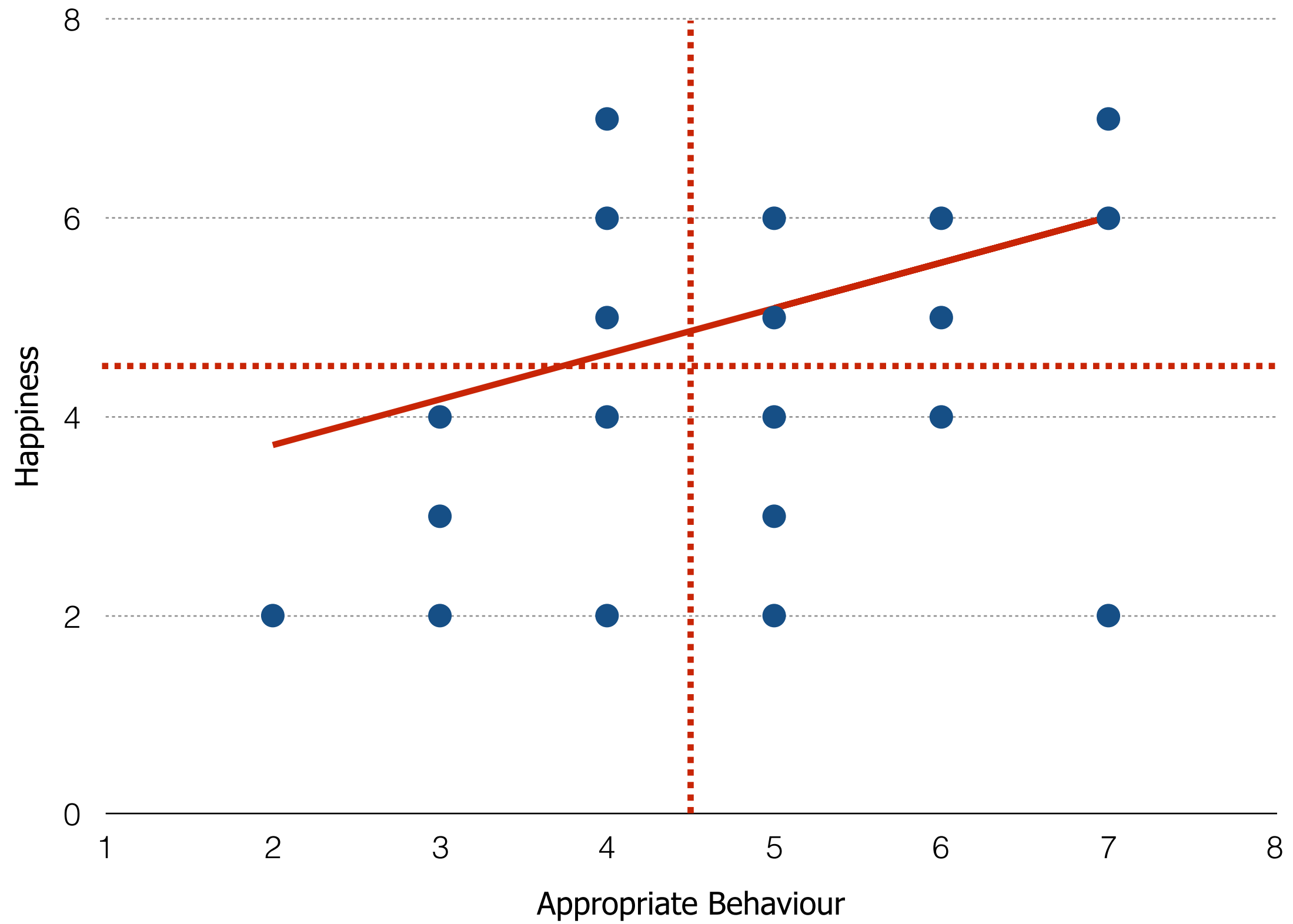
The covariance of X and Y

The variance of X

$$b = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^N (x_i - \bar{x})^2} = \frac{\sigma_{xy}}{\sigma_x^2}$$







Average Prediction
Error (Appropriate
Behaviour vs
Learning)

$$\frac{E}{N} = 1.07$$

$$\frac{E}{N} = 3.00$$

Average Prediction
Error (Appropriate
Behaviour vs
Happiness)

Recap

- The regression allows one to express the criterion as a function of the predictor;
- The use of a line is a constraint that is not necessarily respected by the data, it is an approximation;
- It is difficult to quantify how well the data fits the relationship.

Outline

- Introduction
- Regression
- **Correlation**
- Conclusions

Correlation

“The degree to which the points cluster around the regression line (in other words, the degree to which the actual values of Y agree with the predicted values) is related to the correlation between X and Y.”

D.C.Howell, “Statistical Methods for Psychology”, Chapter 9,
Cengage Learning, 2009.

The Pearson
correlation between X
and Y

The covariance of X
and Y

The diagram illustrates the formula for the Pearson correlation coefficient, r . It features a central equation $r = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$ where each term is enclosed in a light red circle. Four arrows point from descriptive text labels to these terms: one from 'The Pearson correlation between X and Y' to r , one from 'The covariance of X and Y' to σ_{xy} , one from 'The standard deviation of X' to σ_x , and one from 'The standard deviation of Y' to σ_y .

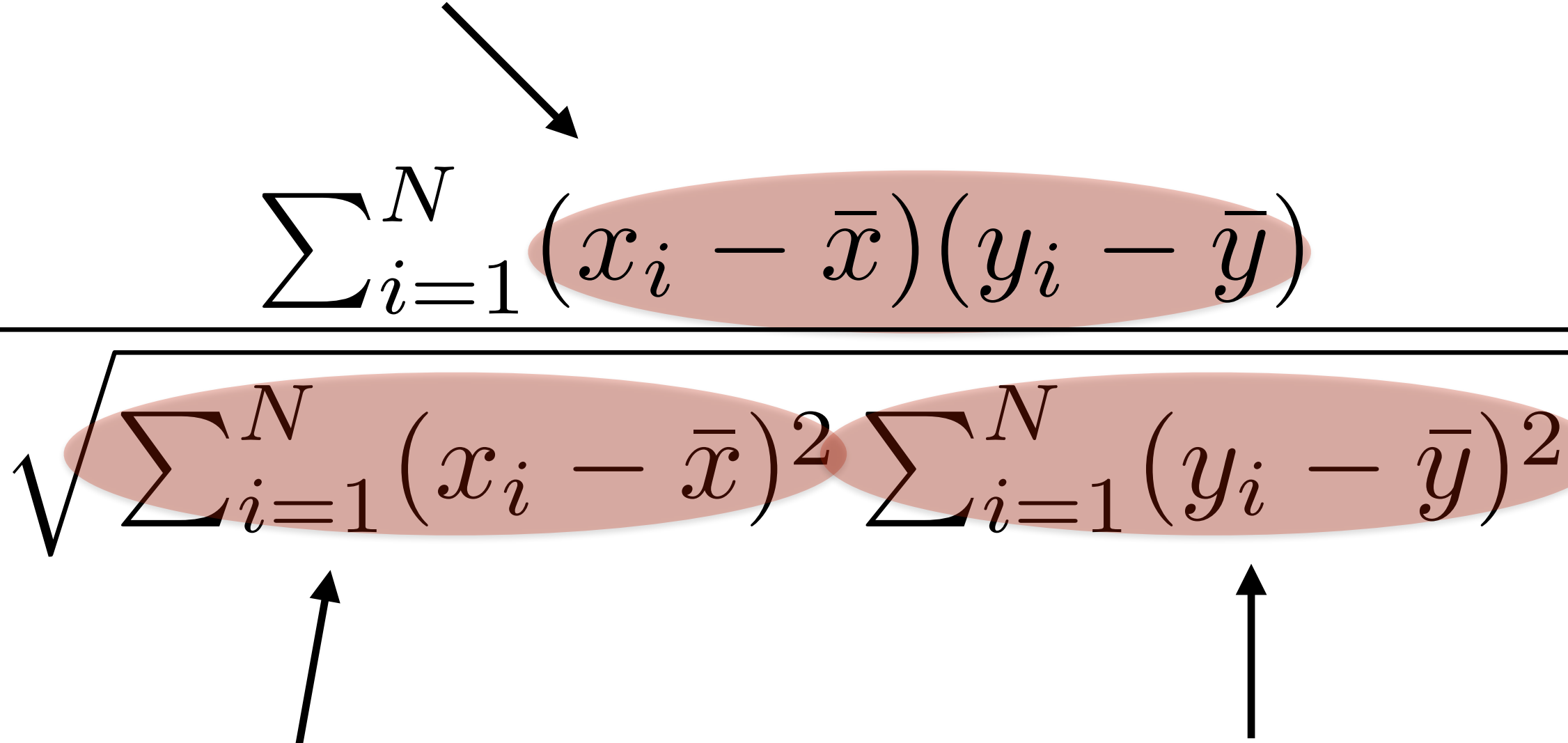
$$r = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

The standard
deviation of X

The standard
deviation of Y

The covariance
of X and Y (multiplied
by N-1)

The Pearson
correlation coefficient
between X and Y


$$\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})$$

$$\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \sum_{i=1}^N (y_i - \bar{y})^2}$$

The variance of X
(multiplied by N-1)

The variance of Y
(multiplied by N-1)

The correlation can
be adjusted when N
(the number of pairs)
is small

The non-adjusted
value of the
correlation

$$r_{adj} = \sqrt{1 - \frac{(1 - r^2)(N - 1)}{N - 2}}$$

The total number of
pairs (x,y)

The Student's t variable can show whether r is statistically significant

The total number of pairs (x,y)

The diagram illustrates the formula for the Student's t test for a correlation coefficient r . The formula is presented as $t = \frac{r \sqrt{N-1}}{\sqrt{1-r^2}}$. Each variable in the formula (t , r , N , and r^2) is enclosed in a light red circle. Three arrows provide context: one points from the text 'The Student's t variable...' to the t circle; another points from 'The total number of pairs (x,y) ' to the N circle; and a third points from 'The value of the correlation' to the r^2 circle. A curved arrow also points from the r circle to the r^2 circle, indicating its derivation.

$$t = \frac{r \sqrt{N-1}}{\sqrt{1-r^2}}$$

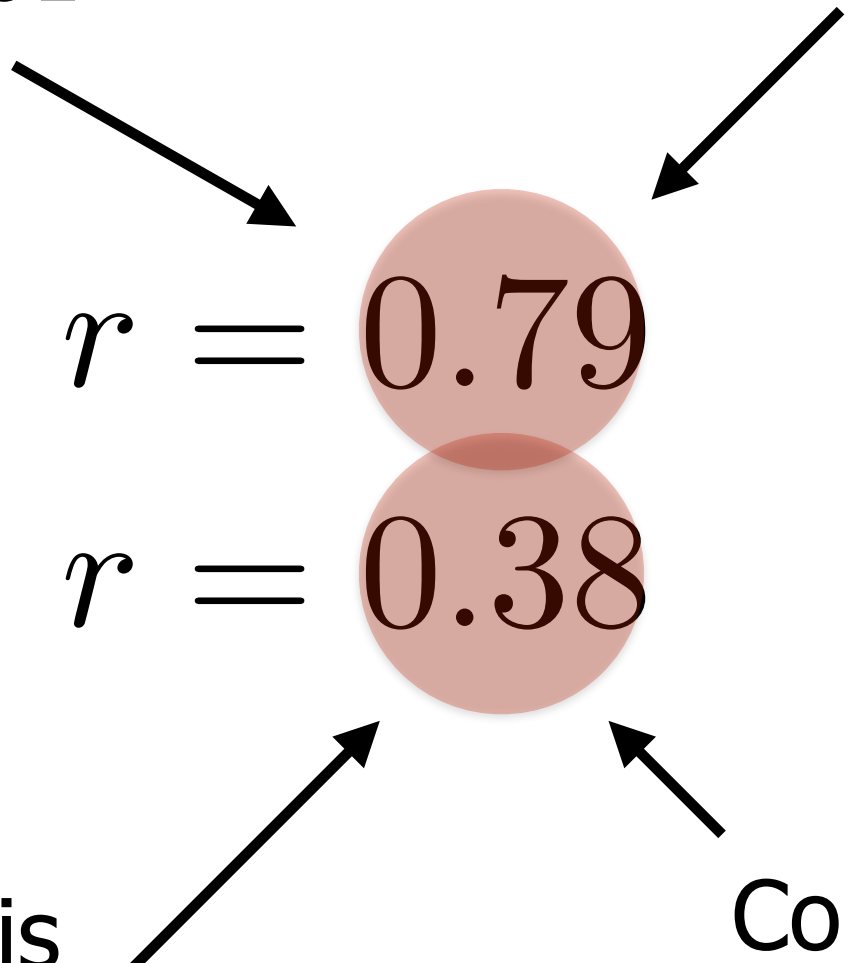
The value of the correlation

Research Hypothesis

- Research Hypothesis: The correlation is higher (or lower for the negative values) than what is expected by chance;
- Null Hypothesis: The correlation is lower (or higher for the negative values) than what is expected by chance.

The Null Hypothesis
can be rejected with
confidence level 0.01
(two-tailed)

Correlation between
Appropriate
Behaviour and
Learning


$$r = 0.79$$

$$r = 0.38$$

The Null Hypothesis
fails to be rejected

Correlation between
Appropriate
Behaviour and
Happiness

Recap

- The correlation measures the fraction of common variance with respect to the variance of the individual variables;
- The values of the correlation are comparable and easy to interpret;
- It is possible to test whether the correlation is statistically significant, i.e., higher than what is expected by chance.

Outline

- Introduction
- Regression
- Correlation
- **Conclusions**

Conclusions

- It is possible to measure the relationship between two variables, i.e., their tendency to change according to each other;
- The relationship has no direction (it is not possible to say whether one variable influences the other or vice versa);
- However, the analysis of the relationships can provide insight about the phenomena under exam.

Thank You!