Clinical Biostatistics

Correlation

Martin Bland

Professor of Health Statistics

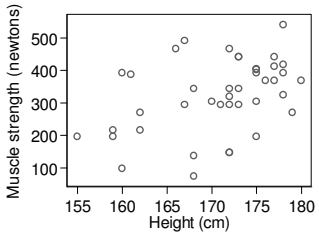
University of York

http://martinbland.co.uk/

Correlation

Example: Muscle strength and height in 42 alcoholics

A scatter diagram:

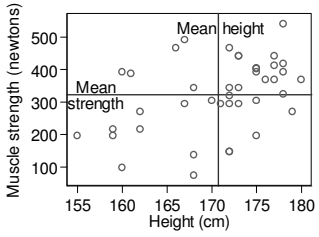


How close is the relationship?

Correlation: measures closeness to a linear relationship.

Correlation coefficient

Subtract means from observations and multiply.

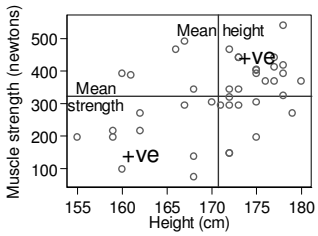


Sum of products about the means.

Like the sum of squares about the means used for measuring variability.

Correlation coefficient

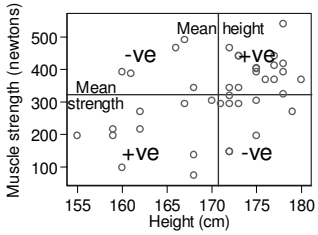
Subtract means from observations and multiply.



Products in top right and bottom left quadrants positive.

Correlation coefficient

Subtract means from observations and multiply.

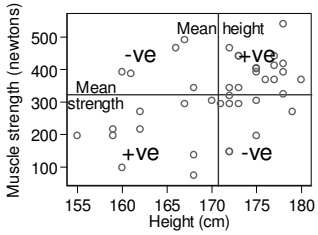


Products in top right and bottom left quadrants positive.

Products in top left and bottom right quadrants negative.

Correlation coefficient

Subtract means from observations and multiply.

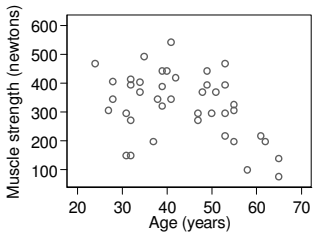


Sum of products positive.

Correlation positive.

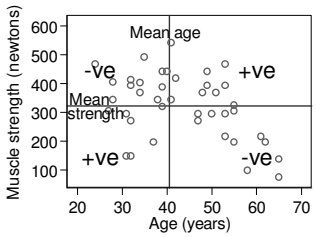
Correlation coefficient

Example: Muscle strength and age in 42 alcoholics



Correlation coefficient

Example: Muscle strength and age in 42 alcoholics



Sum of products negative.

Correlation negative.

Correlation coefficient

Divide sum of products by square roots of sums of squares.

Correlation coefficient, denoted by *r*.

Maximum value = 1.00.

Minimum value = –1.00.

Also known as:

- Pearson's correlation coefficient,
- product moment correlation coefficient.

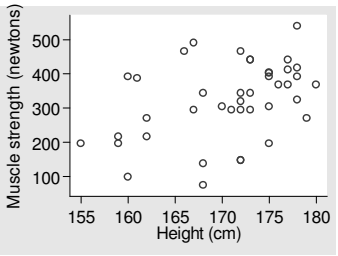
Correlation coefficient

Divide sum of products by square roots of sums of squares.

Correlation coefficient, denoted by *r*.

Maximum value = 1.00.

Minimum value = –1.00.



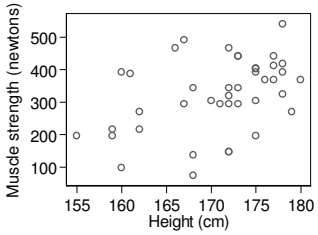
Correlation coefficient

Divide sum of products by square roots of sums of squares.

Correlation coefficient, denoted by *r*.

Maximum value = 1.00.

Minimum value = –1.00.



$r = 0.42$.

Positive correlation of fairly low strength

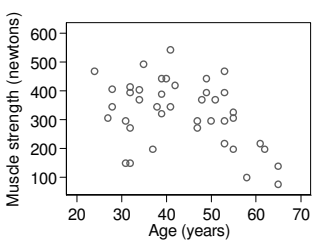
Correlation coefficient

Divide sum of products by square roots of sums of squares.

Correlation coefficient, denoted by *r*.

Maximum value = 1.00.

Minimum value = –1.00.

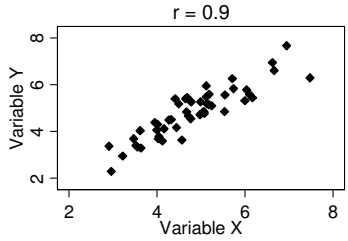


$r = -0.42$.

Negative correlation of fairly low strength.

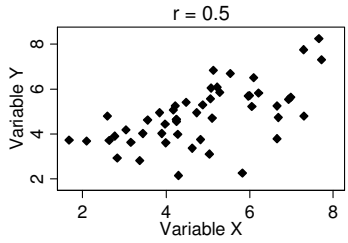
Correlation coefficient

Positive when large values of one variable are associated with large values of the other.



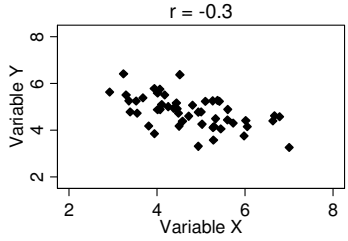
Correlation coefficient

Positive when large values of one variable are associated with large values of the other.



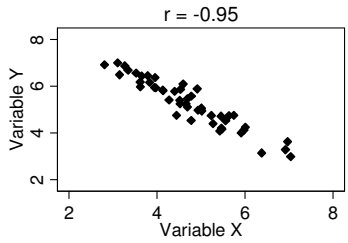
Correlation coefficient

Negative when large values of one variable are associated with small values of the other.



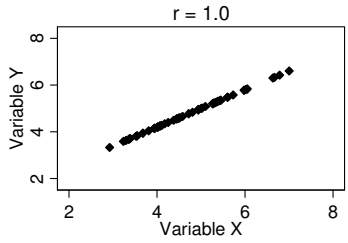
Correlation coefficient

Negative when large values of one variable are associated with small values of the other.



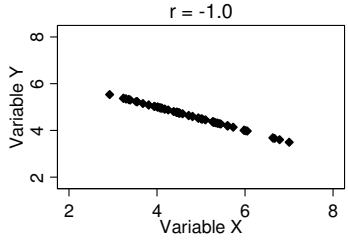
Correlation coefficient

r = +1.00 when large values of one variable are associated with large values of the other and the points lie on a straight line.



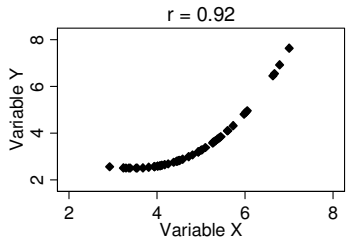
Correlation coefficient

r = –1.00 when large values of one variable are associated with small values of the other and the points lie on a straight line.



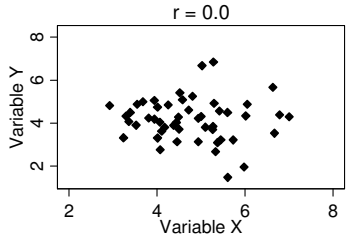
Correlation coefficient

r will not equal –1.00 or +1.00 when there is a perfect relationship unless the points lie on a straight line.



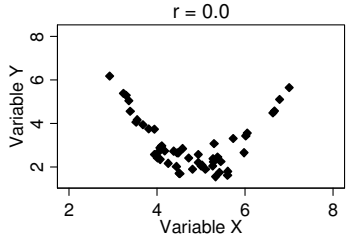
Correlation coefficient

r = 0.00 when there is no linear relationship.



Correlation coefficient

It is possible for *r* to be equal to 0.00 when there is a relationship which is not linear.



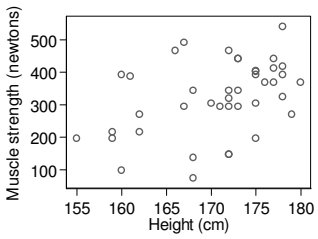
Correlation coefficient

We can test the null hypothesis that the correlation coefficient in the population is zero.

Simple t test, tabulated.

Assume: one of the variables is from a Normal distribution.

Large deviations from assumption ➔ P very unreliable.



$r = 0.42$, P = 0.006.

Easy to do, simple tables.

Computer programs almost always print this.

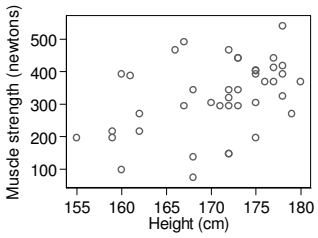
Correlation coefficient

We can find a confidence interval for the correlation coefficient in the population.

Fisher's z transformation.

Assume: both of the variables are from a Normal distribution.

Large deviations from assumption ➔ CI very unreliable.



$r = 0.42$, approximate 95% confidence interval: 0.13 to 0.64

Tricky, approximate.

Computer programs rarely print this.