Bookmark this page

Linear Relationships: Correlation

Learning Objective: Interpret the value of the correlation coefficient, and be aware of its limitations as a numerical measure of the association between two quantitative variables.

The Correlation Coefficient—r

The numerical measure that assesses the strength of a linear relationship is called the **correlation coefficient**, and is denoted by **r**. We will:

- give a definition of the correlation r,
- discuss the calculation of r,
- explain how to interpret the value of r, and
- talk about some of the properties of r.

Definition: The **correlation coefficient (r)** is a numerical measure that measures the **strength** and **direction** of a linear relationship between two quantitative variables.

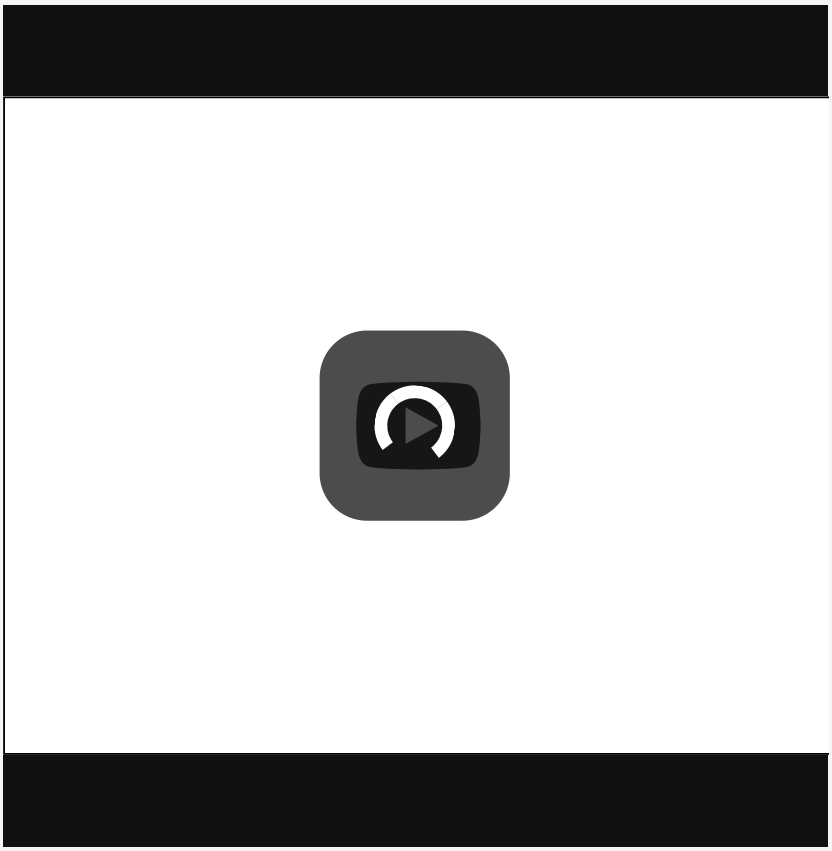
Calculation: r is calculated using the following formula: $r = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{S_x}\right)\left(\frac{y_i-\bar{y}}{S_y}\right)$

However, the calculation of the correlation (r) is not the focus of this course. We will use a statistics package to calculate r for us, and the emphasis of this course will be on the *interpretation* of its value.

Interpretation

Once we obtain the value of **r**, its interpretation with respect to the strength of linear relationships is quite simple, as this video will illustrate:

Interpreting the value of r



Video

Download video file

Transcripts

Download SubRip (.srt) file

Download Text (.txt) file

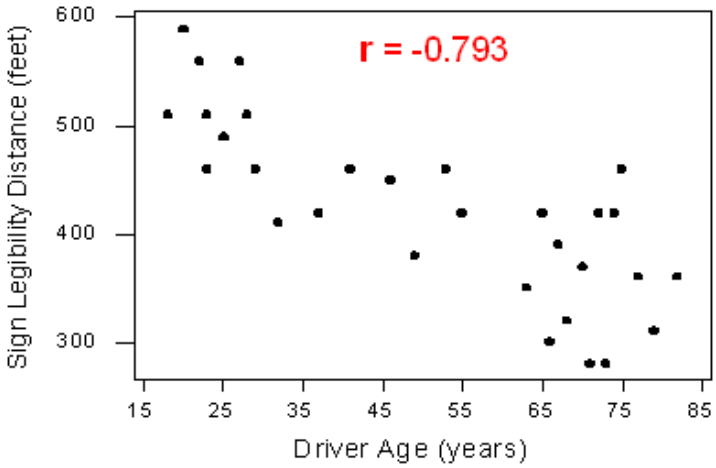
In order to get a better sense for how the value of r relates to the strength of the linear relationship, take a look at the interactive simulation below.

The slider bar at the top of the simulation allows us to vary the value of the correlation coefficient (**r**) between -1 and 1 in order to observe the effect on a scatterplot. Click the **Switch Sign** button to change the sign of the correlation (positive or negative) while keeping the value the same.

Now that we understand the use of **r** as a numerical measure for assessing the direction and strength of linear relationships between quantitative variables, we will look at a few examples.

Example: Highway Sign Visibility

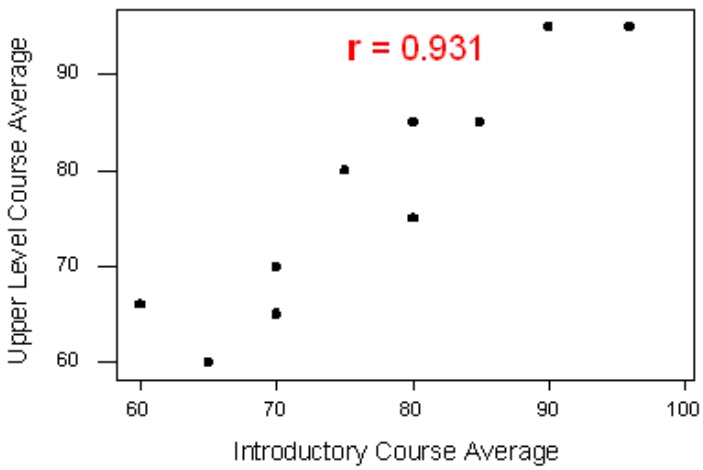
Earlier, we used the scatterplot below to find a **negative linear** relationship between the age of a driver and the maximum distance at which a highway sign was legible. What about the strength of the relationship? It turns out that the correlation between the two variables is r = -0.793.



Since r < 0, it confirms that the direction of the relationship is negative (although we really didn't need r to tell us that). Since **r** is relatively close to -1, it suggests that the relationship is moderately strong. In context, the negative correlation confirms that the maximum distance at which a sign is legible generally decreases with age. Since the value of r indicates that the linear relationship is moderately strong, but not perfect, we can expect the maximum distance to vary somewhat, even among drivers of the same age.

Example: Statistics Courses

A statistics department is interested in tracking the progress of its students from entry until graduation. As part of the study, the department tabulates the performance of 10 students in an introductory course and in an upper-level course required for graduation. What is the relationship between the students' course averages in the two courses? Here is the scatterplot for the data:



The scatterplot suggests a relationship that is **positive** in direction, **linear** in form, and seems quite strong. The value of the correlation that we find between the two variables is **r** = 0.931, which is very close to 1, and thus confirms that indeed the linear relationship is very strong.

Comment

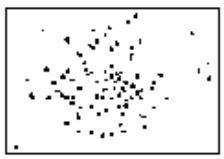
Note that in both examples we supplemented the scatterplot with the correlation (r). Now that we have the correlation (r), why do we still need to look at a scatterplot when examining the relationship between two quantitative variables?

The **correlation** coefficient can **only** be interpreted as the **measure of the strength of a linear relationship**, so we need the scatterplot to verify that the relationship indeed looks linear. This point and its importance will be clearer after we examine a few properties of r.

Did I Get This

1/1 point (graded)

Choose the most likely correlation value for this scatterplot:



- r = 0.436
- r = 0.100 ✔
- r = −0.897
- r = 0.995
- r = −0.575

Answer

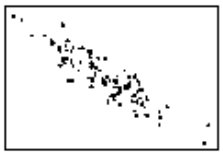
Correct: The direction of the relationship is positive and the linear relationship is weak (close to zero).

Submit

Did I Get This

1/1 point (graded)

likely correlation value for this scatterplot:



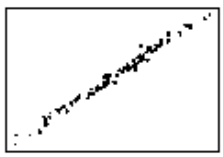
- r = 0.436
- r = 0.100
- r = −0.897 ✔
- r = 0.995
- r = −0.575

Answer
Correct: The direction of the relationship is negative, and the linear relationship is quite strong.

Submit

Did I Get This

1/1 point (graded)
Choose the most likely correlation value for this scatterplot:



- r = 0.436
- r = 0.100
- r = −0.897

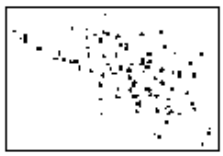
- (●) r = 0.995 ✔
- ( ) r = −0.575

Answer
Correct: The direction of the relationship is positive, and the linear relationship is quite strong.

Submit

Did I Get This

1/1 point (graded)
Choose the most likely correlation value for this scatterplot:



- ( ) r = 0.436
- ( ) r = 0.100
- ( ) r = −0.897
- ( ) r = 0.995
- (●) r = −0.575 ✔

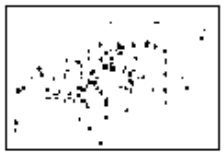
Answer
Correct:
The direction of the relationship is negative, and the linear relationship is only moderately strong.

Submit

Did I Get This

1/1 point (graded)

Choose the most likely correlation value for this scatterplot:



- (●) r = 0.436 ✔
- ( ) r = 0.100
- ( ) r = −0.897
- ( ) r = 0.995
- ( ) r = −0.575

Answer
Correct:
The direction of the relationship is positive and the linear relationship is only moderately strong.

Submit