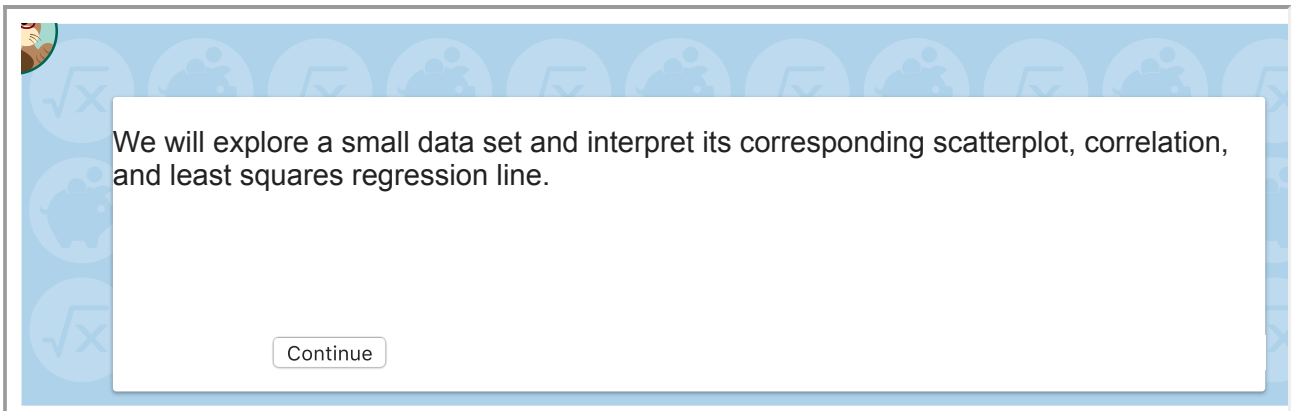
Lagunita is retiring and will shut down at 12 noon Pacific Time on March 31, 2020. A few courses may be open for self-enrollment for a limited time. We will continue to offer courses on other online learning platforms; visit http://online.stanford.edu.

Bookmark this page

Linear Relationships: Summary

Explore the interactive activity below which synthesizes lessons on scatterplots, correlation and linear regression.



Let's Summarize

- A special case of the relationship between two quantitative variables is the **linear** relationship. In this case, a straight line simply and adequately summarizes the relationship.
- When the scatterplot displays a linear relationship, we supplement it with the **correlation coefficient (r)**, which measures the **strength** and direction of a linear relationship between two quantitative variables. The correlation ranges between -1 and 1. Values near -1 indicate a strong

negative linear relationship, values near 0 indicate a weak linear relationship, and values near 1 indicate a strong positive linear relationship.

- The correlation is only an appropriate numerical measure for linear relationships, and is sensitive to outliers. Therefore, the correlation should only be used as a supplement to a scatterplot (after we look at the data).
- The most commonly used criterion for finding a line that summarizes the pattern of a linear relationship is "least squares." The **least squares regression line** has the smallest sum of squared vertical deviations of the data points from the line.
- The slope of the least squares regression line can be interpreted as the average change in the response variable when the explanatory variable increases by 1 unit.
- The least squares regression line predicts the value of the response variable for a given value of the explanatory variable. **Extrapolation** is prediction of values of the explanatory variable that fall outside the range of the data. Since there is no way of knowing whether a relationship holds beyond the range of the explanatory variable in the data, extrapolation is not reliable, and should be avoided.

Open Learning Initiative


Unless otherwise noted this work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

© All Rights Reserved