

12.1: Prelude to Linear Regression and Correlation

CHAPTER OBJECTIVES

By the end of this chapter, the student should be able to:

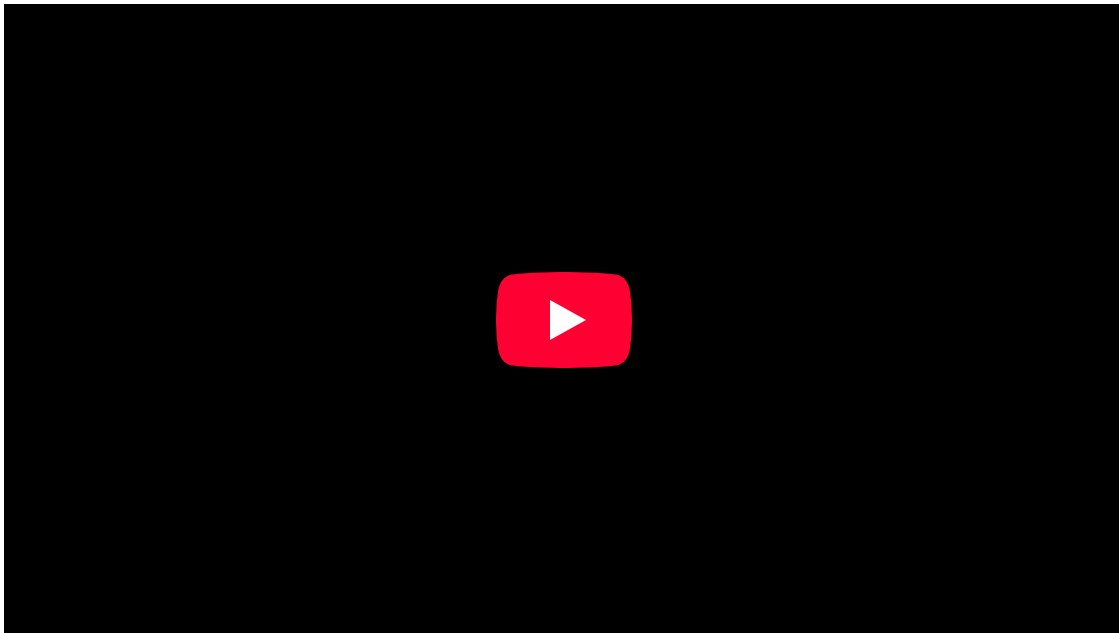
- Discuss basic ideas of linear regression and correlation.
- Create and interpret a line of best fit.
- Calculate and interpret the correlation coefficient.
- Calculate and interpret outliers.

Professionals often want to know how two or more numeric variables are related. For example, is there a relationship between the grade on the second math exam a student takes and the grade on the final exam? If there is a relationship, what is the relationship and how strong is it? In another example, your income may be determined by your education, your profession, your years of experience, and your ability. The amount you pay a repair person for labor is often determined by an initial amount plus an hourly fee.



Figure 12.1.1: Linear regression and correlation can help you determine if an auto mechanic's salary is related to his work experience. (credit: Joshua Rothhaas)

The type of data described in the examples is **bivariate** data — "bi" for two variables. In reality, statisticians use **multivariate** data, meaning many variables. In this chapter, you will be studying the simplest form of regression, "linear regression" with one independent variable (x). This involves data that fits a line in two dimensions. You will also study correlation which measures how strong the relationship is.



Wine Consumption vs. Crime

The computer readout shows the regression analysis for wine consumption per capita in cities and the city's violent crime rate. Interpret r and r^2 and conduct the hypothesis test.

Simple linear regression results:
 Dependent Variable: Violent crime rate
 Independent Variable: Wine consumption per capita
 Violent crime rate = 364.0992 + 99.77388 Wine consumption per capita
 Sample size: 35
 R (correlation coefficient) = 0.2606
 R-sq = 0.06791798
 Estimate of error standard deviation: 104.95475

Parameter estimates:

Parameter	Estimate	Std. Err.	DF	T-Stat	P-Value
Intercept	364.0992	127.843575	33	2.8480055	0.0075
Slope	99.77388	64.342	33	1.5506804	0.1305

Contributors

- Barbara Illowsky and Susan Dean (De Anza College) with many other contributing authors. Content produced by OpenStax College is licensed under a Creative Commons Attribution License 4.0 license. Download for free at <http://cnx.org/contents/30189442-699...b91b9de@18.114>.

This page titled [12.1: Prelude to Linear Regression and Correlation](#) is shared under a [CC BY 4.0](#) license and was authored, remixed, and/or curated by [OpenStax](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.