

CHAPTER OVERVIEW

17: Linear Regression

Introduction

Regression is a toolkit for developing **models** of **cause and effect** between one **ratio scale** data type **dependent** response variables, and one (**simple linear regression**) or more or more (**multiple linear regression**) ratio scale data type **independent** predictor variables. By convention the dependent variable(s) is denoted by Y , the independent variable(s) represented by X_1, X_2, \dots, X_n for n independent variables. Like ANOVA, linear regression is simply a special case of the **general linear model**, first introduced in [Chapter 12.7](#).

Components of a statistical model

Regression statistical methods return **model** estimates of the intercept and slope **coefficients**, plus statistics of regression fit (e.g., R^2 , aka “R-squared,” the **coefficient of determination**).

Chapter 17.1 – 17.9 cover the simple linear model

$$Y_i = \alpha + \beta X_i + \epsilon_i$$

Chapter 18.1 – 18.5 cover the multiple regression linear model

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \epsilon_i$$

where α or β_0 represent the **Y-intercept** and β or $\beta_1, \beta_2, \dots, \beta_n$ represent the regression **slopes**.

Regression and correlation test linear hypotheses

We state that the relationship between two variables is linear (the **alternate hypothesis**) or it is not (the null hypothesis). The difference? **Correlation** is a test of linear association (are variables correlated, we ask?), imply possible **causation**, but are not sufficient evidence for causation: we do not imply that one variable causes another to vary, even if the correlation between the two variables is large and positive, for example. Correlations are used in statistics on data sets not collected from explicit experimental designs incorporated to test specific hypotheses of cause and effect.

Linear regression, however, is to cause and effect as correlation is to association. With regression and ANOVA, we are indeed making a case for a particular understanding of the cause of variation in a response variable: modeling cause and effect is the goal. Regression, ANOVA, and other general linear models are designed to permit the statistician to control for the effects of **confounding variables** provided the causal variables themselves are uncorrelated.

Assumptions of linear regression

The key assumption in linear regression is that a straight line indeed is the best fit of the relationship between dependent and independent variables. The additional assumptions of parametric tests ([Chapter 13](#)) also hold. In [Chapter 18](#) we conclude with an extension of regression from one to many predictor variables and the special and important topic of correlated predictor variables or **multicollinearity**.

Build a statistical model, make predictions

In our exploration of linear regression we begin with [simple linear regression](#), also called ordinary least squares regression, starting with one predictor variable. Practical aspects of model diagnostics are presented. Regression may be used to describe or to provide a predictive statistical framework. In [Chapter 18](#) we conclude with an extension of regression from one to many predictor variables. We conclude with a discussion of model selection. Throughout, use of `Rcmdr` and R have multiple ways to analyze linear regression models are presented; we will continue to emphasize the general linear model approach, but note that use of linear model in `Rcmdr` provides a number of default features that are conveniently available.

References

Linear regression is a huge topic; references I include are among my favorite on the subject, but are only a small and incomplete sampling. For simplicity, I merged references for Chapter 17 and Chapter 18 into one page at [References and suggested readings](#)

(Ch17 & 18)

- 17.1: Simple linear regression
- 17.2: Relationship between the slope and the correlation
- 17.3: Estimation of linear regression coefficient
- 17.4: OLS, RMA, and smoothing functions
- 17.5: Testing regression coefficients
- 17.6: ANCOVA - analysis of covariance
- 17.7: Regression model fit
- 17.8: Assumptions and model diagnostics for simple linear regression

This page titled [17: Linear Regression](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Michael R Dohm](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.