

SECTION OVERVIEW

3.3: Correlation and Linear Regression

Often in statistical research, we want to discover if there is a relationship between two variables. The **explanatory variable** is the “cause” and the **response variable** is the “effect”, although a true cause and effect relationship can only be established in a scientific study that controls for all confounding (lurking) variables.

In Chapter 12, we were interested in determining if a person’s gender was a valid explanatory variable of the person’s opinion about legalization of marijuana for recreational use. In this case, both the explanatory and response variables are categorical and the appropriate model was the Chi-square Test of Independence.

In Chapter 13, we explored if tofu pizza sales (the response variable) were affected by location of the restaurant (the explanatory variable). In this case, the explanatory variable was categorical but the response was numeric. The appropriate model for this example is One Factor Analysis of Variance (ANOVA).

What if we want to determine if a relationship exists when both the explanatory and response variables are both numeric? For example, does annual rainfall in a city help explain sales of sunglasses? This chapter explores and defines the appropriate model for this type of problem.

3.3.1: Bivariate Data and Scatterplots Review

3.3.2: The Simple Linear Regression Model

3.3.3: Estimating the Regression Model with the Least-Square Line

3.3.4: Hypothesis Test for Simple Linear Regression

3.3.5: Estimating σ , the standard error of the residuals

3.3.6: r^2 , The Correlation of Determination

3.3.7: Prediction

3.3.8: Extrapolation

3.3.9: Residual Analysis

3.3: Correlation and Linear Regression is shared under a [CC BY-SA](#) license and was authored, remixed, and/or curated by LibreTexts.