Moneyball The Art of Winning an Unfair Game. That's the title of a book by Michael Lewis about the Oakland Athletics baseball team and the person tasked with building the team, their general manager, Billy Beane.

Alderson began using statistical approaches to find inefficiencies in the market. Alderson was a mentor to Billy Beane, who succeeded him in 1998and fully embraced data science as opposed to scouts as a method for finding low-cost players that data predicted would help the team win. Today, this strategy has been adopted by most baseball teams.As we will see, regression plays a large role in this approach.

1. Introduction to Regression

In this section, you'll learn the basics of linear regression through this course's motivating example, the data-driven approach used to construct baseball teams. You'll also learn about correlation, the correlation coefficient, stratification, and the variance explained.

1.1 introduction to Regression Overview

In the Introduction to Regression section, you will learn the basics of linear regression.

After completing this section, you will be able to:

- Understand how Galton developed linear regression.
- Calculate and interpret the sample correlation.
- Stratify a dataset when appropriate.
- Understand what a bivariate normal distribution is.
- Explain what the term variance explained means.
- Interpret the two regression lines.

This section has three parts: Baseball as a Motivating Example, Correlation, and Stratification and Variance Explained. There are comprehension checks that follow most videos.

1.2 Baseball as a Motivating Examples

1.2.1 Motivating Example: MoneyBall

This all changed with Bill James. In the late 1970s, this aspiring writer and baseball fan started publishing articles describing more in-depth analysis of baseball data. He named the approach of using data to predict what outcomes best predict if a team wins sabermetrics. Until Billy Beane made sabermetrics the center of his baseball operations, Bill James' work was mostly ignored by the baseball world. Today, pretty much every team uses the approach,

1.2.2 Baseball Basics

1.2.3 Bases on Balls or Stolen Bases?

1.3 Correlation

1.3.1 Correlation

1.3.2 Correlation Coefficient

1.3.3 Sample Correlation is a Random Variable

1.4 Stratification and Variance Explained

1.4.1 Anscombe's Quartet/Stratification

1.4.2 Bivariate Normal Distribution

1.4.3 Variance Explained

1.4.4 There are two Regression Lines

2. Linear Models

In this section, you'll learn about linear models. You'll learn about least squares estimates, multivariate regression, and several useful features of R, such as tibbles, lm, do, and broom. You'll learn how to apply regression to baseball to build a better offensive metric.

2.1 Linear Models Overview

2.2 Introduction to Linear models

2.2.1 Confounding: Are BBs More Predictive?

2.2.2 Stratification and Multivariate Regression

2.2.3 Linear Models

2.3 Least Squares Estimates

2.3.1 Least Squares Estimates (LSE)

2.3.2 The lm function

2.3.3 LSE are Random Variables

2.3.4 Advanced notes on LSE

2.3.5 Predcted Variables are Random Variables

2.4 Tibles, do and broom

2.4.1 Advanced dplyr: Tibbles

2.4.2 Tibbles: Differences from Data Frames

2.4.3 do

2.4.4 broom

2.5 Regression and Baseball

2.5.1 Building a better Offensive Metric for Baseball

2.5.2 Building a better Offensive Metric for Baseball: Linear Regression

2.5.3 On Base Plus Slugging (OPS)

2.5.4 Regression Fallacy

2.5.5 Measurement Error Models

3. Confounding

In the final section of the course, you'll learn about confounding and several reasons that correlation is not the same as causation, such as spurious correlation, outliers, reversing cause and effect, and confounders. You'll also learn about Simpson's Paradox.

3.1 Confounding Overview

3.2 Correlation is not Causation

3.2.1 Correlation is not Causation: Spurious Correlation

3.2.2 Correlation is not Causation: Outliers

3.2.3 Correlation is not Causation: Reversing Cause Effect

3.2.4 Correlation is not Causation: Confounders

3.2.5 Simpson's Paradox