

Welcome to **instats**

The Session Will Begin Shortly

START

Statistics in R with Tidyverse

Session 4: Simple Linear Regression Analysis

Regression Overview

- Regression helps study relationships between an outcome variable and explanatory variables
- Key difference: Explanation (understanding associations) vs Predicting (project outcomes based on inputs)
- Can use for numerical or categorical explanatory variables
- Regression provides a linear model for further analysis

Simple Linear Regression

- Focus on one numerical explanatory variable
- Relationship is represented by a line equation
- Key difference: Simple linear regression involves a single regressor, whereas multiple regression involves more than one regressor
- Used to estimate how one variable influences the outcome

Exploratory Data Analysis (EDA)

- EDA is crucial before regression to understand data patterns
- Key difference: EDA explores data before fitting a model, while regression focuses on the relationship between variables
- Typical EDA includes summary statistics and visualizations (e.g., scatterplots, boxplots)

Interpretation of Coefficients

- The intercept is the outcome when the explanatory variable is 0
- The slope represents how much the outcome changes with one unit of the explanatory variable
- Key difference: Interpretation changes if the variable is categorical vs numerical
- Important for understanding the direction and strength of the relationship

Fit and Residuals

- The fitted value is the predicted outcome from the model
- Residuals are the differences between observed and fitted values
- Key difference: Residuals measure the error in prediction, whereas fitted values represent predictions from the model
- Minimized sum of squared residuals ensures the best-fit line

Demo & Exercises

Q & A

STOP

Welcome to **instats**

The Session Will Begin Shortly

START

Statistics in R with Tidyverse

Session 5: Multiple Linear Regression Analysis (Part 1)

Multiple Linear Regression Overview

- Predicting one variable based on multiple predictors
- Involves two or more explanatory variables to explain the outcome
- Applications in the social sciences, economics, and engineering

Benefits of Multiple Regression

- Explaining Complex Relationships
 - Multiple factors can influence a single outcome
 - Allows for a more complete understanding
- Improved Prediction Accuracy
 - Including additional variables can enhance model accuracy
 - Provides insights into how each variable contributes individually

Key Components

- Explanatory Variables
 - Independent variables used to predict the outcome
 - Can include both numerical and categorical variables
- Outcome Variable
 - Dependent variable
 - The goal/target to be investigated that is numerical

Dummy Variables

- Representing categorical data in regression
- Creating Dummy Variables
 - Binary dummy coding
 - Multiple dummies for multi-level categories
- Interpretation of Dummy Variables
 - Baseline category
 - Interpretation of coefficients relative to baseline

Interactions Between Variables

- Main Effects
 - Impact of a predictor on the outcome when other predictors are held constant
- Interaction Effects
 - Occurs when the effect of one variable depends on the level of another
 - Key in exploring complex, real-world scenarios

Model Interpretation

- Intercept
 - Predicted value of the outcome when all predictors are zero
 - Can be interpreted or ignored based on context
- Coefficients
 - Measure the strength and direction of the relationship between each predictor and the outcome
 - Partial slopes

Model Fit

- Fitted Values
 - Predicted values for the outcome based on the regression model
- Residuals
 - Differences between observed and predicted values
 - Used to evaluate the accuracy and fit of the model

Demo & Exercises

Q & A

STOP

Welcome to **instats**

The Session Will Begin Shortly

START

Statistics in R with Tidyverse

Session 6: Multiple Linear Regression Analysis (Part 2)

Regression Model Without Interaction Effects

- Explanatory variables affect the response independently
- Same slope for all levels of categorical variables
- Intercepts vary by group
- Additive effects of explanatory variables
- Sometimes easier to interpret

Multiple Regression with Two Numerical Regressors

- Models the relationship between one response and two numerical predictors
- Creates a regression plane instead of a line
- Each predictor has a partial slope
 - Slopes represent the effect of each predictor holding the other constant
- Joint influence of both predictors on the response
 - Can reveal complex relationships through the regression plane

Fitted Values

- Predicted values from the regression model
- Visualization
 - Overlaying fitted values on scatter plots
 - Residual plots to assess model fit
- Interpretation
 - How close the fitted values are to actual values
 - Assessing overfitting and underfitting

Partial Slopes

- Definition
 - Coefficients of independent variables in regression
- Interpretation
 - Effect of one predictor while holding others constant
- Calculation
 - Estimated from the regression model

Demo & Exercises

Q & A

STOP