Welcome to

instats

The Session Will Begin Shortly

START

Statistics in R with Tidyverse

Session 4: Simple Linear Regression Analysis

instats

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

instats

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

instats

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)

instats

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

instats

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

instats

Demo & Exercises

Q & A

STOP