INTERACTIVE COURSE

Introduction to Linear Modeling in Python

Continue Course

4 hours

16 Videos

59 Exercises

3,278 Participants

5,050 XP

This course is part of these tracks:

Statistics Fundamentals with Python

Jason Vestuto

Data Scientist, University of Texas at Austin

Jason Vestuto started life as a musician and later studied physics and taught himself to code to survive. Along the way, he has completed a couple of degrees in physics, and another in science education, and discovered that he learns best by trying to teach others. Presently, he works within the Space and Geophysics Lab of the University of Texas at Austin, as a python developer and data scientist focused on GPS satellite navigation and signal processing.

See More

COLLABORATOR(S)

Nick Solomon

Adrián Soto

PREREQUISITES

Statistical Thinking in Python (Part 1)

Intermediate Python for Data Science

DATASETS

Femur length versus body height

Distance hiked versus hike duration

Galaxy distances versus recession velocities

Sea surface height versus year

Mass versus volume of solution

Course Description

One of the primary goals of any scientist is to find patterns in data and build models to describe, predict, and extract insight from those patterns. The most fundamental of these patterns is a linear relationship between two variables. This course provides an introduction to exploring, quantifying, and modeling linear relationships in data, by demonstrating techniques such as least-squares, linear regression, estimatation, and bootstrap resampling. Here you will apply the most powerful modeling tools in the python data science ecosystem, including scipy, statsmodels, and scikit-learn, to build and evaluate linear models. By exploring the concepts and applications of linear models with python, this course serves as both a practical introduction to modeling, and as a foundation for learning more advanced modeling techniques and tools in statistics and machine learning.

1 Exploring Linear Trends FREE

8%

We start the course with an initial exploration of linear relationships, including some motivating examples of how linear models are used, and demonstrations of data visualization methods from matplotlib. We then use descriptive statistics to quantify the

shape of our data and use correlation to quantify the strength of linear relationships between two variables.

Introduction to Modeling Data	50 xp
Reasons for Modeling: Interpolation	100 xp
Reasons for Modeling: Extrapolation	100 xp
Reasons for Modeling: Estimating Relationships	100 xp
Visualizing Linear Relationships	50 xp
Plotting the Data	100 xp
Plotting the Model on the Data	100 xp
Visually Estimating the Slope & Intercept	100 xp
Quantifying Linear Relationships	50 xp
Mean, Deviation, & Standard Deviation	100 xp
Covariance vs Correlation	100 xp
Correlation Strength	100 xp

HIDE CHAPTER DETAILS

Continue Chapter

2 Building Linear Models

0%

Here we look at the parts that go into building a linear model. Using the concept of a Taylor Series, we focus on the parameters slope and intercept, how they define the model, and how to interpret the them in several applied contexts. We apply a variety of python modules to find the model that best fits the data, by computing the optimal values of slope and intercept, using least-squares, numpy, statsmodels, and scikit-learn.

What makes a model linear	50 xp
Terms in a Model	50 xp
Model Components	100 xp
Model Parameters	100 xp
Interpreting Slope and Intercept	50 xp
Linear Proportionality	100 xp
Slope and Rates-of-Change	100 xp
Intercept and Starting Points	100 xp
Model Optimization	50 xp
Residual Sum of the Squares	100 xp
Minimizing the Residuals	100 xp
Visualizing the RSS Minima	100 xp
Least-Squares Optimization	50 xp
Least-Squares with `numpy`	100 xp
Optimization with Scipy	100 xp
Least-Squares with `statsmodels`	100 xp

HIDE CHAPTER DETAILS

Continue Chapter

3 Making Model Predictions

0%

Next we will apply models to real data and make predictions. We will explore some of the most common pit-falls and limitations of predictions, and we evaluate and compare models by quantifying and contrasting several measures of goodness-of-fit, including RMSE and R-squared.

Modeling Real Data	50 xp
Linear Model in Anthropology	100 xp
Linear Model in Oceanography	100 xp
Linear Model in Cosmology	100 xp
The Limits of Prediction	50 xp
Interpolation: Inbetween Times	100 xp
Extrapolation: Going Over the Edge	100 xp
Goodness-of-Fit	50 xp
RMSE Step-by-step	100 xp
R-Squared	100 xp
Standard Error	50 xp
Variation Around the Trend	100 xp
Variation in Two Parts	100 xp

HIDE CHAPTER DETAILS

Continue Chapter

4 Estimating Model Parameters

0%

In our final chapter, we introduce concepts from inferential statistics, and use them to explore how maximum likelihood estimation and bootstrap resampling can be used to

estimate linear model parameters. We then apply these methods to make probabilistic statements about our confidence in the model parameters.

Inferential Statistics Concepts	50 xp
Sample Statistics versus Population	100 xp
Variation in Sample Statistics	100 xp
Visualizing Variation of a Statistic	100 xp
Model Estimation and Likelihood	50 xp
Estimation of Population Parameters	100 xp
Maximizing Likelihood, Part 1	100 xp
Maximizing Likelihood, Part 2	100 xp
Model Uncertainty and Sample Distributions	50 xp
Bootstrap and Standard Error	100 xp
Estimating Speed and Confidence	100 xp
Visualize the Bootstrap	100 xp
Model Errors and Randomness	50 xp
Test Statistics and Effect Size	100 xp
Null Hypothesis	100 xp
Visualizing Test Statistics	100 xp
Visualizing the P-Value	100 xp
Course Conclusion	50 xp

HIDE CHAPTER DETAILS

Continue Chapter