# Course 5 overview

WELCOME

to Course 5

Hello, and welcome to **Regression Analysis: Simplify Complex Data Relationships**, the fifth course in the Google Advanced Data Analytics Certificate. You're making great progress on this exciting journey!

In previous courses, you learned how data professionals contribute to the success of an organization, the basic syntax and functions of the Python programming language, the main stages of exploratory data analysis (EDA), and how to use statistics to analyze and interpret data. In this course, you'll explore how to simplify complex data relationships using linear and logistic regression models.

Course descriptions

The Google Advanced Data Analytics Certificate has seven courses. **Regression Analysis: Simplify Complex Data Relationships** is the fifth course.
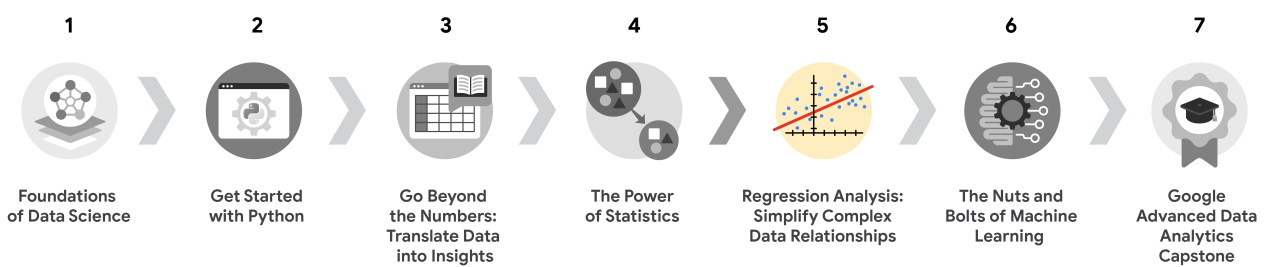
1. **Foundations of Data Science** — Learn how data professionals operate in the workplace and how different roles in the field of data science contribute to an organization's vision of the future. Then, explore data science roles, communication skills, and data ethics.

2. **Get Started with Python** — Discover how the programming language Python can power your data analysis. Learn core Python concepts, such as data types, functions, conditional statements, loops, and data structures.

3. **Go Beyond the Numbers: Translate Data into Insights** — Learn the fundamentals of data cleaning and visualizations and how to reveal the important stories that live within data.

4. **The Power of Statistics** — Explore descriptive and inferential statistics, basic probability and probability distributions, sampling, confidence intervals, and hypothesis testing.

5. **Regression Analysis: Simplify Complex Data Relationships** — *(current course)* Learn to model

variable relationships, focusing on linear and logistic regression.

6. **The Nuts and Bolts of Machine Learning** — Learn unsupervised machine learning techniques and how to apply them to organizational data.

7. **Google Advanced Data Analytics Capstone** — Complete a hands-on project designed to demonstrate the skills and competencies you acquire in the program.

Course 5 content

Each course of this certificate program is broken into modules. You can complete courses at your own pace, but the module breakdowns are designed to help you finish the entire Google Advanced Data Analytics Certificate in about six months.

What's to come? Here's a quick overview of the skills you'll learn in each module of this course.

Regression Analysis: Simplify Complex Data Relationships — Introduction to complex data relationships — Simple linear regression — Multiple linear regression — Advanced hypothesis testing — Logistic regression — Course 5 end-of-course project

Module 1: Introduction to complex data relationships

In this part of the course, you'll examine how PACE can serve as a solid foundation for regression analysis. Then, you'll learn about linear and logistic regression, how they're similar and different, and when to apply each approach. Finally, you'll use real data to learn how to apply estimation techniques in Python and study more use-cases of regression.

Module 2: Simple linear regression

You will explore examples of ordinary least squares estimation in Python and learn the four main model assumptions of simple linear regression: linearity, normality of residuals, independent observations, and homoscedasticity. You'll build a model in Python and learn how to evaluate model fit using R squared, holdout samples, and measures of uncertainty, like confidence intervals and p-values. Then, you'll begin to develop a story using numbers and statistics and learn how to communicate the results of your simple linear regression model using Python visualizations.

Module 3: Multiple linear regression

In this section, you will build on your knowledge of simple linear regression by extending the concepts into multiple

regression. You'll learn how to test assumptions using math, built-in Python functions, and visualizations during the EDA process. Next, you'll interpret and evaluate multiple regression results, examine the idea of overfitting, and discuss ways to overcome the effects of overfitting using regularization techniques.

Module 4: Advanced hypothesis testing

In this part of the course, you'll examine a variety of hypothesis tests: Chi-squared, ANOVA, ANCOVA, MANOVA, and MANCOVA. You will also investigate null and alternative hypotheses and how they help articulate testing results.

Module 5: Logistic regression

You will explore logistic regression models, with a heavy emphasis on binomial logistic regression. You'll build binomial, multinomial, and ordinal logistic regression models, and log-linear Poisson regression models. Next, you'll discover important classification model metrics such as ROC and AUC, precision, recall, and type I and type II errors.

Module 6: Course 5 end-of-course project

As you conclude this part of the Advanced Data Analytics Certificate, you will put everything you've learned into one end-of-course project. You'll be tasked with solving a business problem using the provided data. The concepts

and skills you will learn in this part of the course will be critical to your success as a data professional.

What to expect

Each course offers many types of learning opportunities:

- **Videos** led by Google instructors teach new concepts, introduce the use of relevant tools, offer career support, and provide inspirational personal stories.

- **Readings** build on the topics discussed in the videos, introduce related concepts, share useful resources, and describe case studies.

- **Discussion prompts** explore course topics for better understanding and allow you to chat and exchange ideas with other learners in the **discussion forums**.

- **Self-review activities** and **labs** give you hands-on practice in applying the skills you are learning and allow you to assess your own work by comparing it to a completed example.

- **Interactive plug-ins** encourage you to practice specific tasks and help you integrate knowledge you have gained in the course.

- **In-video quizzes** help you check your comprehension

as you progress through each video.

- **Practice quizzes** allow you to check your understanding of key concepts and provide valuable feedback.

- **Graded quizzes** demonstrate your understanding of the main concepts of a course. You must score 80% or higher on each graded quiz to obtain a certificate, and you can take a graded quiz multiple times to achieve a passing score.

Tips for success

- It is strongly recommended that you go through the items in each lesson in the order they appear because new information and concepts build on previous knowledge.

- Participate in all learning opportunities to gain as much knowledge and experience as possible.

- If something is confusing, don't hesitate to replay a video, review a reading, or repeat a self-review activity.

- Use the additional resources that are referenced in this course. They are designed to support your learning. You can find all of these resources in the **Resources** tab.

- When you encounter useful links in this course, bookmark them so you can refer to the information later for study or review.

- Understand and follow the [Coursera Code of Conduct](#) to ensure that the learning community remains a welcoming, friendly, and supportive place for all members.