

Intermediate Data Analytics for Technical Professionals

Inference, experiments, regression and time-series evaluation

FA-0590 · 10 days

Data & Analytics · Software Development

Use Python to interpret statistical evidence, analyse experiments and evaluate regression and forecasting examples.

Introduction

This ten-day intermediate course develops analytical skills in inferential statistics, experiment design, regression modelling and time-series analysis. It builds on Python, SQL and descriptive-statistics foundations and ends with a scoped analytical capstone.

Projects use prepared public, synthetic or de-identified datasets. Malaysian business and economic contexts may provide illustrative examples, but results are presented with data and modelling limitations rather than as validated property valuations, investment forecasts or client outcomes. A/B exercises use simulated or previously collected results, not live campaigns.

Learning outcomes

The course teaches participants to:

- Calculate and interpret selected confidence intervals and hypothesis tests within their assumptions.
- Design an illustrative A/B experiment and analyse its results.
- Fit and evaluate regression models using appropriate metrics and held-out data.
- Explore time-series components, decomposition and introductory ARIMA forecasts.
- Use chronological evaluation and communicate uncertainty in forecasting.
- Complete and present a scoped analysis with evidence, limitations and peer feedback.

Prerequisites

- Working Python knowledge and familiarity with pandas and scikit-learn.
- Foundations in descriptive statistics and basic statistical modelling.
- Basic SQL familiarity is helpful for experiment-data exercises.
- A suitable analytical environment and approved learning datasets.
- The capstone assumes participation in the preceding sections.

Training outline

1. Days 1–2 — Inferential Statistics

Day 1 — Sampling and Confidence Intervals

- Sampling Distribution
- Confidence intervals and their repeated-sampling interpretation; a fixed interval is not a probability statement about a fixed parameter.
- Apply a selected interval calculation to a prepared sample dataset with explicit sampling assumptions.

Day 2 — Hypothesis Tests

- Null and Alternative Hypotheses
- p-values and pre-specified significance levels; distinguish statistical evidence from effect size and practical importance.
- Illustrative hypothesis-testing exercise for a fictional e-commerce scenario.

2. Days 3–4 — A/B Testing

Day 3 — Experimental Design

- Importance of A/B Testing
- Define hypotheses, metrics, randomisation, sample-size considerations and analysis assumptions.
- Design an A/B experiment for a fictional website.

Day 4 — Results and Interpretation

- Statistical Methods for Analysis

- Interpret estimated effects, intervals, assumptions and limits rather than declare success from a threshold alone.
- Analyse synthetic or previously collected results from an illustrative advertising experiment; no live campaign is required.

3. Days 5–6 — Regression and Model Evaluation

Day 5 — Regression Models

- Types of Regression Models
- Fitting Models using Python
- Fit a property-style regression example using approved historical or synthetic data.

Day 6 — Evaluation

- R-squared and Adjusted R-squared
- RMSE (Root Mean Square Error)
- Evaluate a model using a public or synthetic economic dataset; do not imply validated GDP forecasting.
- Check residuals, assumptions and out-of-sample performance; predictive association is not proof of causality.

4. Days 7–8 — Time-Series Analysis and Forecasting

Day 7 — Components and Exploration

- Components of Time Series
- Exploratory Time Series Data Analysis
- Explore historical or synthetic market-style time series without trading recommendations.

Day 8 — Models and Forecast Review

- ARIMA Models
- Seasonal Decomposition
- Forecast an illustrative tourism-style series and compare against observed hold-out periods; replace undated 'post-COVID' assumptions with explicit dataset dates.
- Compare a baseline and a selected forecast using time-respecting evaluation; explain uncertainty and historical data limits.

5. Days 9–10 — Capstone

Day 9 — Scoped Project Work

- Choose a feasible question, approved dataset and evaluation plan.
- Use a prepared or approved dataset and document cleaning decisions.
- Apply relevant analysis/model techniques rather than force every method into one project.

Day 10 — Presentation and Feedback

- Finalizing Analysis
- Presentation of Findings
- Peer and Instructor Feedback