Google

Reading: Course 5 resources and citations

Module 2: Simple linear regression

Resources

Explore ordinary least squares

- Parameter Estimation - Ordinary Least Squares Method
- SOGA-Py

The four main assumptions of simple linear regression

- qqplot()
- Download the seaborn penguins dataset here
- Introduction to Palmer penguins
- Normal Quantile-Quantile Plots (video from jbstatistics)

Code functions and documentation

- function
- statsmodels.regression.linear_model.OLS
- statsmodels.regression.linear_model.OLSResults
- predict()
- statsmodels documentation
- statsmodels.formula.api interface
- function

Evaluation metrics for simple linear regression

- AIC (Akaike information criterion) and BIC (Bayesian information criterion)

Correlation versus causation: Interpret regression results

- Pearson correlation coefficient
- Spurious Correlations
- randomized controlled experiment

Citations

The four main assumptions of simple linear regression

- pandas 1.4.4 documentation. (2022, August 31). Pandas.

Code functions and documentation

- pandas 1.4.4 documentation. (2022, August 31). Pandas.
- Perktold, J., Seabold, S., & Taylor, J. (n.d.). API Reference — statsmodels. Statsmodels.

Explore linear regression with Python

- Python 3.10.6 documentation. (n.d.). Python.

Correlation versus causation: Interpret regression results

- Numpy and scipy documentation. (n.d.). SciPy.org.
- Spurious correlations. (n.d.). Tylervigen.com.

Module 3: Multiple linear regression

Resources

Multiple linear regression scenarios

- Net Promoter Score
- "Multiple Regression: Definition, Uses, and 5 Examples." Indeed Editorial Team.
- "Multivariate Regression Analysis | STATA Data Analysis Examples." UCLA: Statistical Consulting Group.

Multiple linear regression assumptions and multicollinearity

- pairplot()
- scatterplot()
- Pennsylvania State University's Eberly College of Science
- Practical Econometrics and Data Science
- variance_inflation_factor()
- statsmodels documentation

Underfitting and overfitting

- A detailed description of underfitting and how to mitigate it
- Scikit-learn library documentation for the train_test_split function
- A blog discussing multiple, adjusted, and predicted R-squared values

Citations

Multiple linear regression scenarios

- Net promoter score (NPS): The ultimate guide. (2022, July 21). Qualtrics.

Multiple linear regression assumptions and multicollinearity

- Buteikis, A. (2018, August 7). Practical econometrics and data science. Vilnius University.
- Perktold, J., Seabold, S., & Taylor, J. (n.d.). API Reference — statsmodels. Statsmodels.

Interpret multiple regression results with Python

- Python 3.10.6 documentation. (n.d.). Python.

Module 4: Advanced hypothesis testing

Resources

Chi-squared tests: Goodness of fit versus independence

- chisquare() function
- chi2_contingency() function
- Yates' correction for continuity

Citations

Explore one-way versus two-way ANOVA tests with Python

- Python 3.10.6 documentation. (n.d.). Python.
- statsmodels.stats.anova.anova_lm — statsmodels. (n.d.). Statsmodels.

ANOVA post hoc tests with Python

- Python 3.10.6 documentation. (n.d.). Python.

Module 5: Logistic regression

Resources

Common logistic regression metrics in Python

- precision_score
- recall_score
- accuracy_score
- RocCurveDisplay.from_predictions
- roc_auc_score

Interpret logistic regression models

- LogisticRegression

Prediction with different types of regression

- Different Types of Regression Models
- Hypothesis testing for data scientists

Citations

Construct a logistic regression model with Python

- Python 3.10.6 documentation. (n.d.). Python.
- UCI Machine Learning Repository: Activity recognition with healthy older people using a batteryless wearable sensor Data Set. (n.d.). UCI Machine Learning Repository.

Evaluate a binomial logistic regression model

- Python 3.10.6 documentation. (n.d.). Python.

Key metrics to assess logistic regression results

- Scikit-learn: machine learning in Python. (n.d.). Scikit-learn 2017 Yellow taxi trip data - catalog. (2022, May 12). Data.Gov.
