City University of New York (CUNY)
CUNY Academic Works

Open Educational Resources | Lehman College

2025

MAT 301 - Applied Statistics and Data Analysis

Eric Aragundi
CUNY Lehman College, eric.aragundi@lehman.cuny.edu

How does access to this work benefit you? Let us know!

More information about this work at: https://academicworks.cuny.edu/le_oers/85
Discover additional works at: https://academicworks.cuny.edu

This work is made publicly available by the City University of New York (CUNY).
Contact: AcademicWorks@cuny.edu

MAT 301: Applied Statistics and Data Analysis

Lehman College, CUNY
Spring, 2025

Instructor: Eric Aragundi
Email: eric.aragundi@lehman.cuny.edu
Office location: Gillet Hall 100B
Office Hours: Monday and Wednesday | 2 pm to 3 pm

Official Course Description: *4 hours, 3 credits*. Data analysis using standard statistical methods and relevant computer software. Emphasis on real-world data, interpretation, and misinterpretation of computer output.
Prerequisite: MAT 132 or placement by the department

Textbook and Materials: Free OpenIntro Statistics, Author: David Diez, Mine Cetinkaya-Run, Publisher: openintro.org, Edition: 4, Year Published: 2022, https://www.openintro.org/book/os/. Links for Python tutorials will be in the daily schedule according to the topics.

Technology: All students are required to have access to their Lehman email accounts and Blackboard. Additionally, students must have a Google account, as we will be using Google Colab (Python) and Google Drive to facilitate a more seamless and engaging learning experience with Python for data analysis.

Learning Goals:

- **Classify and Summarize Data**

Collect data using appropriate methods and present key findings with descriptive statistics (measures of central tendency, variation, frequency distributions, and visualizations).

- **Apply Probability and Distributions**

Understand fundamental probability rules, work with common distributions (e.g., normal), and apply the Central Limit Theorem to infer population parameters from sample data.

- **Construct Confidence Intervals**

Determine and interpret confidence intervals for population means under various conditions (known and unknown standard deviations).

- **Conduct Hypothesis Testing**

Formulate and test hypotheses for one or two sample means, analyze differences (independent/dependent samples), and employ chi-square tests for categorical data, using statistical software (e.g., Python).

- **Examine Relationships and Communicate Findings**

Assess correlation and build simple regression models, interpret results, and communicate conclusions effectively and ethically in both written and verbal formats.

These objectives will be assessed on the final project along with other important technique.

Course format. Students are expected to learn both the mathematics covered in the assigned readings and coding examples for all materials covered during the lectures. This course is being taught in person.

Communication. Communication is essential in succeeding in this course – this cannot be stressed enough. After identifying topics that may be giving you trouble, it is your responsibility to share this information with me. Join our Slack channel using this link: https://join.slack.com/t/mat-301-lc-sp25/shared_invite/zt-31o8bg3zq-qKGvVj57Ec6sLK4wW1NdXg

Where to get help:

- **Me.** I will be available during my office hours to answer any questions, also you can send me an email with any questions, and I will get back as soon as time permits.

· **Your classmates.** If you missed a class or are struggling with a topic, it is a good idea to communicate with another student in the course and work together to understand the material

.

Tutoring. The Mathematics and Statistics Student Success Center (MS^3) offers tutoring support for students enrolled in this course. Tutoring is available in person at Gillet Hall, Room 233-A. The following tutors have been designated to provide assistance for this class:

Caleb Adjei
Email : caleb.adjei@lc.cuny.edu
Time: Mon - Thu 11AM - 2PM

For more information visit the website here and you can make an appointment Here

Your final grade. Provided that you attend class regularly and on time, your grade for the course will be made up of the following:

Assignment Type	Grade Percentage
Participation	5%
Homework	15%
Quizzes	20%
Midterm Project	25%
Final Project	35%
Total	100%

Grading policy:

Homework assignments. There will be no additional costs for completing the homework, as we will utilize free tools. Students are expected to follow the homework instructions carefully to avoid losing points. A more detailed explanation of how to complete the homework will be provided when it is assigned.

Projects and Exams: This course will focus heavily on projects; there will not be traditional exams. However, you can expect a quiz worth more points during midterm and final weeks. Detailed explanations about the projects will be provided as the course progresses.

Grading Scale

93-100 = A	**83-86.9 = B**	**73-76.9 = C**	**60-66.9 = D**
90-92.9 = A-	**80-82.9 = B-**	**70-72.9 = C-**	**Below 60 = F**
87-89.9 = B+	**77-79.9 = C+**	**67-69.9 = D+**	*Purposedly left blank*

Accommodating Disabilities. Lehman College is committed to provide access to all programs and curricula to all students. Students with disabilities who may need classroom accommodations are encouraged to register with the Office of Disabilities Services, Shuster Hall, Room 238, Phone number 718-960-8841

Tentative Course Schedule (will be updated dynamically):

Day/date	Topic	Reference material
Day 1- Mon 01/27/25	Syllabus, Introduction to Python, and Google Collab	Notebook: Day1_introduction_to_python_notebooks.ipynb Helpful Links: ● Learn Python in Less than 10 Minutes for Beginners (Fast & Easy) ● Python for Beginners - Learn

		Coding with Python in 1 Hour
Wed 01/29/2025	NO CLASSES	
Day 2 - Mon 02/03/25	Managing Data	Notebook: Day2_basics_managing_data Helpful links: ● How to Read CSV files in Google Colab from Drive (from computer) (python pandas) ● How to Import Files from GitHub to Google Colab? How to read csv files in Google Colab? ● Data Cleaning in Pandas | Python Pandas Tutorials
Day 3 - Wed 02/05/25	Wrangling data - Measure of Central Tendency Homework_0 (due 02/10/2025 at 11:59 pm)	Notebook: Day3_wrangling_data_measure_central_tendency Helpful Links: ● Grouping, Sorting, and Shuffling in Python Pandas (2.3) ● Measures of Central Tendency | Statistics with Python | Meritshot Tutorials
Day 4 - Mon 02/10/25	Measures of Variation - Graphing Box Plots using MatPlotLib	Notebook: Day4_Measures_of_Variation

	Homework_1 (due 02/26/2025 at 11:59 pm) Explanation can be found here Mat 301 11am is section 01 Mat 301 6pm is section 81 Quiz 1 link: 11 am class: Here 6 pm class: Here NOTE: The quiz is supposed to be in person, anyone taking the quiz from home will get 0.	_Graphing_Box_Plots_Using_Matplotlib.ipynb Helpful Links: ● Find Variance, Standard Deviation and Interquartile Range Using Python (Measures of Variation) ● Box Plot in Python
Wed 02/12/25	NO CLASSES	
Mon 02/17/25	NO CLASSES	
Day 5 - Tues 02/18/25	More on graphs using MatPlotLib - Normalization	Notebook: Day5_more_about_graphs_standardization.ipynb Helpful links: ● Python Data Visualization: Matplotlib Basic Plotting (Histograms, Scatter Plots and More Graphs) ● Playlist with plotting videos ● Master Min-Max Normalization in Python: Step-by-Step Tutorial!

Day 6 - Wed 02/19/25	Advanced Graphing Techniques and Regression	Notebook: day6_Advanced_Graphing_Techniques_and_Regression.ipynb Helpful links: ● Effortless Data Visualization in Python using Seaborn | Python for Data Science ● Seaborn scatter plot | How to make and style a scatterplot in Python seaborn ● Seaborn Is The Easier Matplotlib ● 02. Linear Regression in Scikit-learn | Machine Learning With Scikit-Learn | Sklearn | Codersarts
Day 7 - Mon 02/24/25	Exploring Distributions	Notebook: Day7_Exploring Distributions.ipynb Helpful links: ● The 6 MUST-KNOW Statistical Distributions MADE EASY [4/13] ● Python Tutorial: Probability distributions ● How to Create Distribution/Histogram Plot using Seaborn ,

		Matplotlib, and Python
Day 8 - Wed 02/26/25	Catch up day Midterm project instructions are here (due 03/31/2025) Tutorial on Uploading datasets (use .csv files) to GitHub here Quiz 2 link: 11 am class: here 6 pm class: here NOTE: The quiz is supposed to be in person, anyone taking the quiz from home will get 0.	
Day 9 - Mon 03/03/25	More on Filtering Data - Probability Calculations	Notebook: day9_more_on_filtering_data_probability_calculations.ipynb Helpful link: ● Python Pandas Tutorial (Part 4): Filtering - Using Conditionals to Filter Rows and Columns ● How to Filter Data in Python || Using Pandas boolean operators to filter a dataset ● Python Pandas Tutorial 9. Merge Dataframes ● Merging DataFrames in

		Pandas | Python Pandas Tutorials
Day 10 - Wed 03/05/25	Enhancing Data Quality & Conditional Probabilities	Notebook: Day10_Enhancing Data Quality & Conditional Probabilities.ipynb Helpful links: ● How To Use apply() In Pandas (Python) ● 45- Pandas DataFrames: Crosstabs, Cross Tabulation, Generating Contingency Tables ● Conditional Probability (with tables) ● Bayes' Theorem EXPLAINED with Examples
Day 11 - Thurs 03/06/25	Asynchronous class due to conflict. Watch this video: Here	Notebook: Day11_Markdown_Cells_Organazing_Notebooks.ipynb
Day 12 - Mon 03/10/25	Standard Normal Distribution and Assessing Normality	Notebook: Day12_Standard_Normal Distribution_assessing_normality.ipynb Helpful links: ● Quantile-Quantile Plots (QQ plots), Clearly Explained!!! ● Python Tutorial : Skewness and

		Kurtosis • Python Shapiro–Wilk Test for Normality
Day 13 - Wed 03/12/25	Law of large numbers, Sampling Distributions, and Central Limit Theorem.	Notebook: Day13_Sampling_Distributions_applications.ipynb Helpful links: • Python Law of Large Numbers (Numpy) • Simulating Sampling Distributions: A Python Tutorial for Data Science • The Central Limit Theorem – With Examples in Python
Day 14 - Mon 03/17/25	Case Study: Sleep, Health and Lifestyle How to handle issues with big files for GitHub and Dealing with too many Na values.	Notebook: big_files_na_values_issues.ipynb Day14_Case_Study_Sleep_Example.ipynb Helpful links: • Exploratory Data Analysis with Pandas Python • How to Do Data Exploration (step-by-step tutorial on real-life dataset)
Day 15 - Wed 03/19/25	Midterm Exam link: 11 am class: here	

	6 pm class: here NOTE: The Midterm Exam is supposed to be in person; anyone taking the quiz from home will get 0.	
Day 16 - Mon 03/24/25	Normal Distributions Wrap-up, T Distribution, and Intro to Confidence Intervals	Notebook: Day16_Normal_Wrap_up_Confidence_Interval_intro.ipynb Helpful links: ● Normal Distribution (PDF, CDF, PPF) in 3 Minutes ● Python Tutorial - Standard Normal Table - Scipy Tutorial ppf() ● Introduction to the t Distribution (non-technical) ● Confidence Levels & Intervals in Python ● how to build a confidence interval in python
Day 17 - Wed 03/26/25	Day 16 notebook continuation Finishing T Distribution, and Intro to Confidence Intervals	
Mon 03/31/25	NO CLASSES	
Day 18 - Wed 04/02/25	Confidence Intervals for proportions, Estimating sample sizes for confidence intervals,	Notebook: Day18_More_on_Confidence_Intervals.ipynb

	and other types of confidence intervals. Homework 2 (Due 04/23/2025 at 11:59 pm)	Helpful links: ● Confidence Interval for a population proportion | Solved Problems ● How To Calculate The Sample Size Given The Confidence Level & Margin of Error ● Python Tutorial: Creating Bootstrap Confidence Intervals ● Confidence interval for the slope of a regression line | AP Statistics | Khan Academy
Day 19 - Mon 04/07/25	Day 18 notebook continuation Review Estimating sample size for different confidence intervals, learning about other types of confidence intervals. Quiz 3 11 am class: here 6 pm class: here NOTE: The quiz is supposed to be in person; anyone taking the quiz from home will get 0.	
Day 20 - Wed 04/09/25	Class Cancelled	
Mon 04/14/25 Wed 04/16/25	NO CLASSES SPRING RECESS	

Day 21 - Mon 04/21/25	Day 21: Introduction to Hypothesis Testing	Notebook: Day21_Introduction_Hypothesis_Testing.ipynb Helpful links: ● What is a hypothesis test? A beginner's guide to hypothesis testing! ● Statistical Significance, the Null Hypothesis and P-Values Defined & Explained in One Minute ● Type I error vs Type II error ● Hypothesis Testing Problems - Z Test & T Statistics - One & Two Tailed Tests 2
Day 22 - Wed 04/23/25	Day 22: More about Hypothesis Testing Final project instructions are here (due 05/23/2025)	Notebook: Day22_More_on_Hypothesis_testing.ipynb Helpful links: ● Python - One-Sample Proportion Test (Score test and Wald Test) ● Twosample proportions test in Python 3 ● Z-Test For Two Means ● How to Perform a Two Sample Z Test in Python (Step-by-

		Step Tutorial) • https://www.youtube.com/watch?v=YBbfrvggZ24 • Python - Paired Samples t-Test
Day 23 - Mon 04/28/25	Day 23: Chi-Square Distribution - CI and Hypothesis Testing for Variance Homework 2 solution: here	Notebook: Day23_Chi_Square_Distribution_CI_and_Hypothesis_testing_for_Variance.ipynb Helpful links: • An Introduction to the Chi-Square Distribution • Confidence Intervals For Variance And Standard Deviation - Chi Square Distribution • How To Do Hypothesis Testing For Variance And Standard Deviation - Chi-Square Test For Variance
Day 24 - Wed 04/30/25	Day 24: Chi-square Tests Quiz 4 11 am class: here 6 pm class: here NOTE: The quiz is supposed to be in person; anyone taking the quiz from home will get 0.	Notebook: Day24:_Chi-Square_Tests.ipynb Helpful links: • Performing a Chi-Square Test for Independence in Python with Scipy • Chi-Square Goodness of Fit Test in Python: A Step-

		by-Step Guide (Scipy & Numpy) • Introduction to the chi-square test for homogeneity | AP Statistics | Khan Academy • Chi Square Homogeneity Test
Day 25 - Mon 05/05/25	Day 25: A/B testing (introduction) Homework 3 due 05/14/25 (milestone): here	Notebook: Day25_Introduction_AB_testing.ipynb Helpful links: • A/B Testing in Data Science Interviews by a Google Data Scientist | DataInterview • How To A/B Test a Product
Day 26 - Wed 05/07/25	Day 26: Introduction to ANOVA - Understanding Group Differences	Notebook: Day26_Introduction_to_ANOVA_Understanding_Group_Differences.ipynb Helpful links: • Analysis of Variance (ANOVA) • How to Perform One-Way ANOVA in Python (Scipy & Numpy) • TukeyHSD Made Simple: How to Find Significant Differences in Python!

Day 27 - Mon 05/12/25	Project Assistance Day	
Day 28 - Wed 05/14/25	Case Study: Avocados Sales 2015 - 2018 Homework 3 due today, 05/14/25 11:59 pm (milestone)	Notebook: Day27_Case_Study_Avocados_Sale.ipynb
May 16 to May 22	FINAL EXAM WEEK **MAT 301 01 class at 11 am: here** Final Exam Mon 05/19/2025 11:30 am to 1:30 pm, same room **MAT 301 81 class at 6 pm: here** Final Exam Wed 05/21/2025 06:15 pm to 8:15 pm, same room NOTE: The Final quiz is supposed to be in person; anyone taking the quiz from home will get 0. **FINAL PROJECT DUE 05/23/2025**	