

City University of New York (CUNY)

CUNY Academic Works

Open Educational Resources

City College of New York

2018

Test: Intro to Data Science - Midterm Review

Grant Long

CUNY City College

NYC Tech-in-Residence Corps

[How does access to this work benefit you? Let us know!](#)

More information about this work at: https://academicworks.cuny.edu/cc_oers/259

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

Review

Week 2: Where to Find Data

- The data science lifecycle.
- Structured vs unstructured data.
- Common sources of data.
- Common ways to access data.

Week 3: Processing and Cleaning Data

- Elements of the ETL Process
- Processing Tools
- Data Cleaning Considerations for Data Scientists
 - Missing Value
 - Handling Outliers
 - Normalizing Data

Week 4: Statistics and the Stories We Tell Ourselves

- Types of Data
- Useful Statistical Distribution
- Important Summary Statistics
- Independence
- Key Theorems

Week 5: Intro to Linear Models

- What Makes Linear Regression Linear
- Assumptions for Linear Models
- Measuring Performance for Linear Models

Week 6: Regression vs Classification

- Regression vs. Classification
- Logistic Regression
- Measuring Performance for Classification Models