

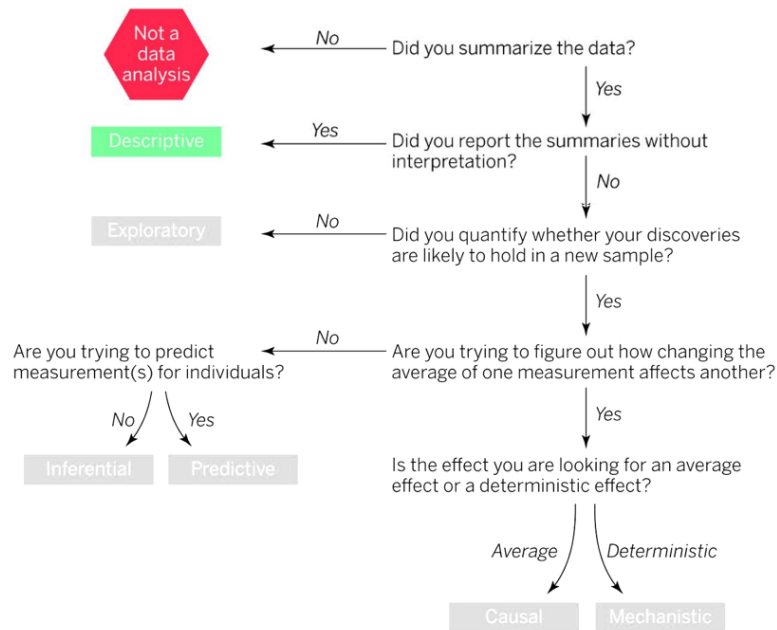
CASA Quantitative Curriculum Outline

Overview: A curriculum for a weeklong series on basic statistics and data visualization for CASA participants.

- Comprised of 10 hours of instructional material and assignments
- Media include digital print and video for each major section
- Designed for mixed proficiency audience

Objectives: Ensure a basic understanding of the following:

1. Operator Precedence, Basic Algebra, Formula Logic, Tabular Data, and Spreadsheet Software
2. Date Types, Data Classes, and Formatting in Spreadsheet Software
3. Basic Summary Statistics, Measures of Central Tendency & Dispersion, and Spreadsheet Macros
4. Calculating & Interpreting Basic Bivariate Models & Coefficients, such as Correlation, Linear Regression
5. Building and Interpreting Basic Univariate, Bivariate, and Multivariate Data Visualization



Caption: *Extent of scope in color.* **Source:** Leek & Peng, 2015.

Tooling: Microsoft Excel (proprietary) and/or Google Sheets (gratis)

Vision & Rationale: This is a 10-hour boot camp to understand a cursory introduction to applied statistics and data visualization. Only a few practical statistical concepts will be visited, while the foundational knowledge of statistics such as theory, formulas, probability, and experimental design will be bypassed for immediate use of spreadsheet software. Ultimately, we want end users to get “over the hump” of spreadsheet software, providing a foundation for a hard skill and guiding them through an otherwise daunting program. Challenges for advanced users will add an additional layer of critical thinking. Visualization is only considered a tool in the data analytic toolkit and is treated accordingly.

Curriculum Outline: The following provides a detailed overview of instructional material and exercises¹.

I. Lecture 1: Review & Material Preparation (2 Hours):

a. Part 1: Operator Precedence

- i. Motivating Example: Why does “ $5 + 5 * 2$ ” not equal “20”?
- ii. What is the “Order of Operations”?
- iii. Acronyms (PEMDAS, or “Please Excuse My Dear Aunt Sally”)
- iv. Further examples
- v. Exercise:
 1. Required: 10 problems, increasing in sophistication
 2. Challenge: 3 advanced order of operation problems

b. Part 2: Basic Algebra

- i. Motivating Example: “Spent \$200 on takeout, 8 orders; how much per order?”
- ii. What is “Algebra”? (Arabic: *al-jabr* or “to make whole”)
- iii. Further examples
- iv. Exercise:
 1. Required: 8 basic algebra questions, 2 easy formulas (e.g. BMI)
 2. Challenge: 2-3 advanced formulas (e.g. amortization)

c. Part 3: Logic

- i. Motivating Example: Philosophical deductive reasoning vs. formulaic logic
- ii. What is “logic”? (Greek: *logos* or “word”)
- iii. Further examples
- iv. Exercise:
 1. Required: 5 logic problems, parentheses-only
 2. Challenge: 5 logic problems with comparators

d. Part 4: Introduction to Spreadsheet Software

- i. Motivating Example: How to apply a simple formula to multiple values.
- ii. What is a spreadsheet? What is spreadsheet software?
- iii. Basic Tour of Microsoft Excel and Google Sheets
- iv. What are rows, columns, values, and tables?
 1. Rows, columns, and values, defined
 2. How series of values coincide as tables
 3. Alternative names of rows (e.g. records, observations, cases)
 4. Alternative names of columns (e.g. fields, variables)
- v. Further examples
- vi. Exercise:
 1. Required: Before-and-after tabular conversions from series
 - a. Includes “tidy” and “untidy data” (Wickham, 2014)
 - b. One column conversion with simple formula
 2. Challenge: 2-3 advanced formula conversions

II. Lecture 2: Data Types (2 Hours)

a. Part 1: Values & Data

- i. What are data?
- ii. What aren’t data?

b. Part 2: Qualitative vs. Quantitative Data

- i. Motivating Example: Which are which?
- ii. What are quantitative values?

¹ Jargon will not necessarily be used in material but is employed here for brevity.

- iii. What are qualitative values?
- iv. Further examples?
- v. Exercise:
 - 1. Required: 5 True or False
 - 2. Challenge: 5 True or False (pitfalls, e.g. ZIP codes)

c. Part 3: Continuous, Discrete, Categorical, & Ordinal

- i. Motivating Example: What's the difference between crime type and crime classification?
- ii. Define continuous, discrete, categorical, and ordinal data
- iii. Alternative names, e.g. "nominal"
- iv. Further examples
- v. Exercise:
 - 1. Required: Multiple choice classification
 - 2. Challenge: Multiple choice classification with pitfalls (e.g. discretized values)

d. Part 4: Data Classes & Formatting

- i. Motivating Example: MS Excel cutting off leading zeroes for ZIP code, e.g.
- ii. Define class numeric, text ("character"), date/datetime, logical
- iii. Demonstrate class coercion/formatting in Excel, Sheets
- iv. Demonstrate fine formatting in Excel, Sheets (e.g. decimal places)
- v. Exercise:
 - 1. Required: Format a simple table, guided
 - 2. Challenge: Format a larger table, unguided

III. Lecture 3: Summary Statistics (2 Hours)

a. Part 1: Summary Statistics Overview

- i. Motivating Example: Average life expectancy in industrialized democracies?

b. Part 2: Measures of Central Tendency

- i. Motivating Example: Average salary (e.g. Amazon), including CEO J. Bezos?
- ii. Defining mean (average), median, and mode
- iii. Further examples
- iv. Exercise:
 - 1. Required: Determine mean, median for 5 sets of values
 - 2. Challenge: Interpret differences in mean, median for 3 sets of values

c. Part 3: Measures of Dispersion

- i. Motivating Example: Vocabulary diversity among hip hop artists ([Source](#))
- ii. Defining minimum, maximum, range, percentile, IQR, variance
- iii. Further examples
- iv. Exercise:
 - 1. Required: Determine min, max, range
 - 2. Challenge: Discretize values into quintiles and deciles

d. Part 4: Excel & Sheets Macros

- i. Motivating Example: What if we want the average of 10,000 groups?
- ii. Defining macros
- iii. Further examples
- iv. Exercise:
 - 1. Required: Mean, median for 5 variables, 50 rows
 - 2. Challenge: Alternate macros and specified decimal precision

IV. Lecture 4: Basic Bivariate Stats (2 Hours)

a. Part 1: Correlation

- i. Motivating Example: Murder ~ Ice cream sales; Drowning ~ Nicholas Cage films ([Source](#))
- ii. What is correlation? What is directionality?

- iii. What is causation? How does it not find?
- iv. What are confounding variables?
- v. Pearson's R Formula
- vi. Further examples
- vii. Exercise:
 - 1. Required: Determine and interpret correlations for 3 bivariate pairs
 - 2. Challenge: Identify spurious relationships or possible confounding variables

b. Part 2: Linear Regression

- i. Motivating Example: Out-of-sample predictions for something cool
- ii. Populations vs. samples (briefly)
- iii. What is linear regression (high-level)?
 - 1. On sum of squares (briefly)
- iv. How to interpret linear regression coefficients (slope, intercept)
- v. Further examples
- vi. Exercise:
 - 1. Required: Simple formulas to predict out-of-sample values
 - 2. Challenge: Interpret direction, strength, and effects of outliers

c. Part 3: Correlation & Regression in Excel, G Sheets

- i. Motivating Example: Linear regression for 200 data points, e.g.
- ii. How to calculate correlation and linear regression in Excel
- iii. Further examples
- iv. Exercise:
 - 1. Required: Linear regression, bivariate pairs, guided
 - 2. Challenge: Determine strongest relationships in multivariate dataset, unguided

V. Lecture 5: Data Visualization (2 Hours)

a. Part 1: What Is Data Visualization?

- i. Motivating Example: Anscombe's Quartet
- ii. What is data visualization?
- iii. Exploratory vs. explanatory visualization
- iv. Advantages of data visualization
- v. Best practices (briefly)
- vi. Recommended data visualizations in Excel: A How-To
- vii. Further examples
- viii. Exercise:
 - 1. Required: Find a data visualization that you like, write briefly on why
 - 2. Challenge: Describe where that same visualization could improve and why

b. Part 2: Univariate Data Visualization

- i. Motivating Example: Average temperature in Atlanta each summer
- ii. On Histograms: One variable for one axis
- iii. Distributions: Normal, skewed, uniform, exponential, etc.
- iv. Further examples
- v. Exercise:
 - 1. Required: Create three histograms from three variables – interpret them
 - 2. Challenge: Use exploratory data visualization to find 3 patterns in a dataset

c. Part 3: Bivariate Data Visualization

- i. Motivating Example: Relationship with Atlanta temperature to CFC emissions
- ii. On scatterplots (continuous) and box/whisker (continuous/categorical)
- iii. Further examples
- iv. Exercise:

1. Required: Create two scatterplots and one box plot with provided variables
2. Challenge: Determine and interpret a systemic discrepancy in given dataset

d. Part 4 Multivariate Data Visualization

- i. Motivating Example: Can we visualize a third variable, too? (Show multiple ways).
- ii. On aesthetic mappings – common aesthetics using color, fill, shape, hatching, etc.
- iii. How-To in Excel
- iv. Further Examples
- v. Exercise:
 1. Required: Map a third variable using an effective aesthetic and given data
 2. Challenge: Recreate the challenge plots