Program
Evaluation
Toolkit

Program
Evaluation
Toolkit

Module
7
Data
Analysis

October 5, 2021

REL Central at Marzano Research
COLORADO   KANSAS   MISSOURI   NEBRASKA   NORTH DAKOTA   SOUTH DAKOTA   WYOMING

Module 7
Data Analysis
Chapter Progression

1 — Approaches to Data Preparation and Analysis

2 — Examples of Data Analysis

3 — From Findings to Interpretation to Recommendations

Program Evaluation Toolkit

REL Central at Marzano Research

COLORADO KANSAS MISSOURI NEBRASKA NORTH DAKOTA SOUTH DAKOTA WYOMING

Authors
- *Douglas Gagnon, Joshua Stewart, Christopher Rhoads, Mckenzie Haines, and Kyle Luke*

Module 7 Data Analysis
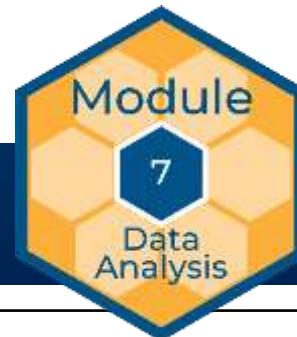
Chapter 1

Approaches to Data Preparation and Analysis
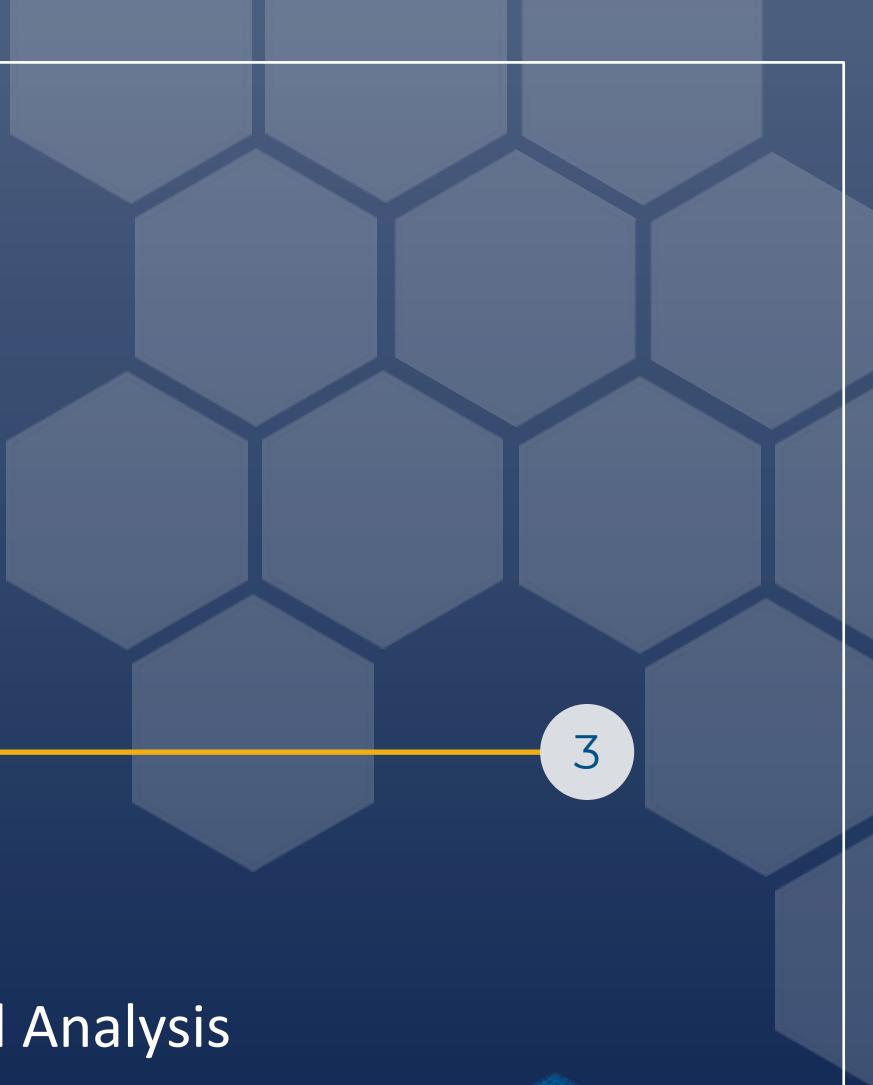
Strategies for Data Preparation

- Develop a system to organize your data:

  - Define a unit of measurement.

  - Assign a unique participant ID (you do not need to create one; you can use existing ID numbers for teachers, schools, students).

  - Develop a codebook with variable names, response options, and numerical codes.

Guidelines fora Codebook

What Is a Codebook?

A codebook is a document describing each variable in your dataset. A codebook includes the following:

- Variable names and labels.
- Values/codes assigned to the variables and their labels.
- Missing data codes.
- Special instructions on how to use data.
- Other relevant information.

Why Is a Codebook Important?

A codebook is a reference guide for your evaluation team. It also helps others understand your data.

Codebook Contents

Variable names must be unique and short. If you will analyze data with a statistical program, the variable names cannot contain spaces and special characters. Use underscores instead of spaces.

Variable labels describe the meaning of variables. A good label is concise but descriptive.

Values and labels represent the contents of a variable. For example, the variable *sex* in the codebook below has two possible values, 0 and 1. Each value should have a corresponding label describing the meaning of the value (for example, 0=Male and 1=Female). As a general rule, give numeric codes to categorical variables, and assign a value to each category.

- If missing data have special codes, indicate these codes in the codebook. Typically, missing data for continuous variables are represented with an impossible value such as −88, −99, or −999.

A codebook can also include a description of the measures used, instructions on how to use data, and information about who created the data, when, and how.

Program Evaluation Toolkit

REL Central at Marzano Research

COLORADO KANSAS MISSOURI NEBRASKA NORTH DAKOTA SOUTH DAKOTA WYOMING

Additional Resources

- *Guidelines for a Codebook*

Module 7 Data Analysis

Data Errors

- Data errors come from a variety of sources and can occur at any stage during an evaluation:

  - Data collection

  - Data entry and cleaning

  - Data analysis

Program Evaluation Toolkit

REL Central at Marzano Research

Additional Resources

- *Common Sources of Data Errors and Error-Checking Techniques*

Module 7 Data Analysis

COLORADO   KANSAS   MISSOURI   NEBRASKA   NORTH DAKOTA   SOUTH DAKOTA   WYOMING
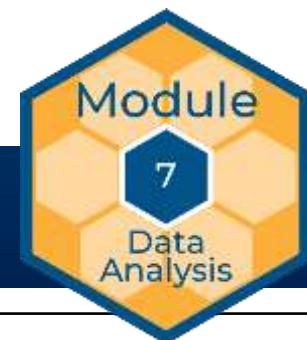
Data Cleaning: Detecting Errors

- Spot-checking

- Logic check

- Double entry

- Descriptive analysis

MicrosoftExcel Functions for Data Cleaning

Function	What It Does	Example Formulas
EXACT	Checks for duplicates. For example, EXACT(A2,A3) checks whether the values in second and third columns match.	=EXACT(A2,A3)
TRIM	Removes spaces from beginning and end of text.	=TRIM(A2:A300)
CLEAN	Removes all nonprintable characters from text.	=CLEAN(A2:A300)
SUBSTITUTE	Substitutes one or more instances of a given character with new character. For example, the formula on the right will substitute all instances of 1 with a 3 in cell C2.	=SUBSTITUTE(C2,"1","3")
REPLACE	Replaces characters in a specified position of a text string. For example, the formula on the right will replace the second character in cell C2 with "new." Replace only one character.	=REPLACE(C2, 2, 1, "new")
LEFT/RIGHT	Extracts a given number of characters from the left/right side of a text string. For example, LEFT("parking", 4) changes "parking" to "park."	=LEFT(A2,4) =RIGHT(A2, 4)
PROPER	Changes the first letter in each word to upper case. For example, PROPER("id") changes "id" to "Id."	=PROPER(A2)
LOWER/UPPER	Changes text to lower/upper case letters. For example, UPPER("id") changes "id" to "ID."	=LOWER(A2) =UPPER(A2)
CONCATENATE	Joins text from different cells into one cell. & " " adds space between the cells. For example, =CONCATENATE("John"& " ", "Smith") changes "John" and "Smith" in separate cells to "John Smith" in a single cell.	=CONCATENATE(A2& " ", B2)
ISBLANK	Identifies whether a cell is blank and returns a "True" or "False."	=ISBLANK(A2)
MIN/MAX	Identifies the lowest or highest value in a range of cells.	=MIN(A2:A300) =MAX(A2:A300)

Note. Adapted from the following sources:

Cunningham, K. (2016, November 7). *10 Excel functions you need to know for data analysis*. Adept. https://blog.adeptmarketing.com/10-excel-functions-you-need-to-know-for-data-analysis

De Groot, L. (2014). *Intro to cleaning data*. Berkeley Graduate School of Journalism, Advanced Media Institute. https://multimedia.journalism.berkeley.edu/tutorials/cleaning-data/

Microsoft. (n.d.). *Excel functions (by category)*. https://support.office.com/en-us/article/Excel-functions-by-category-5f91f4e9-7b42-46d2-9bd1-63f26a86c0eb

Program Evaluation Toolkit

REL Central at Marzano Research

COLORADO   KANSAS   MISSOURI   NEBRASKA   NORTH DAKOTA   SOUTH DAKOTA   WYOMING

Additional Resources

- *Microsoft Excel Functions for Data Cleaning*

Module 7 Data Analysis
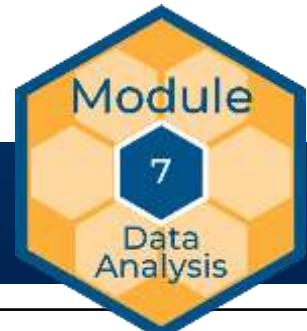
Visualizing Errors[1]
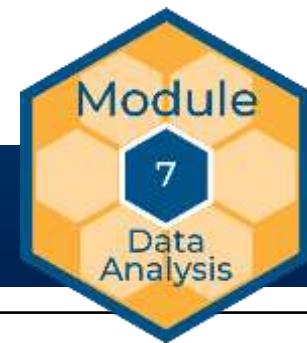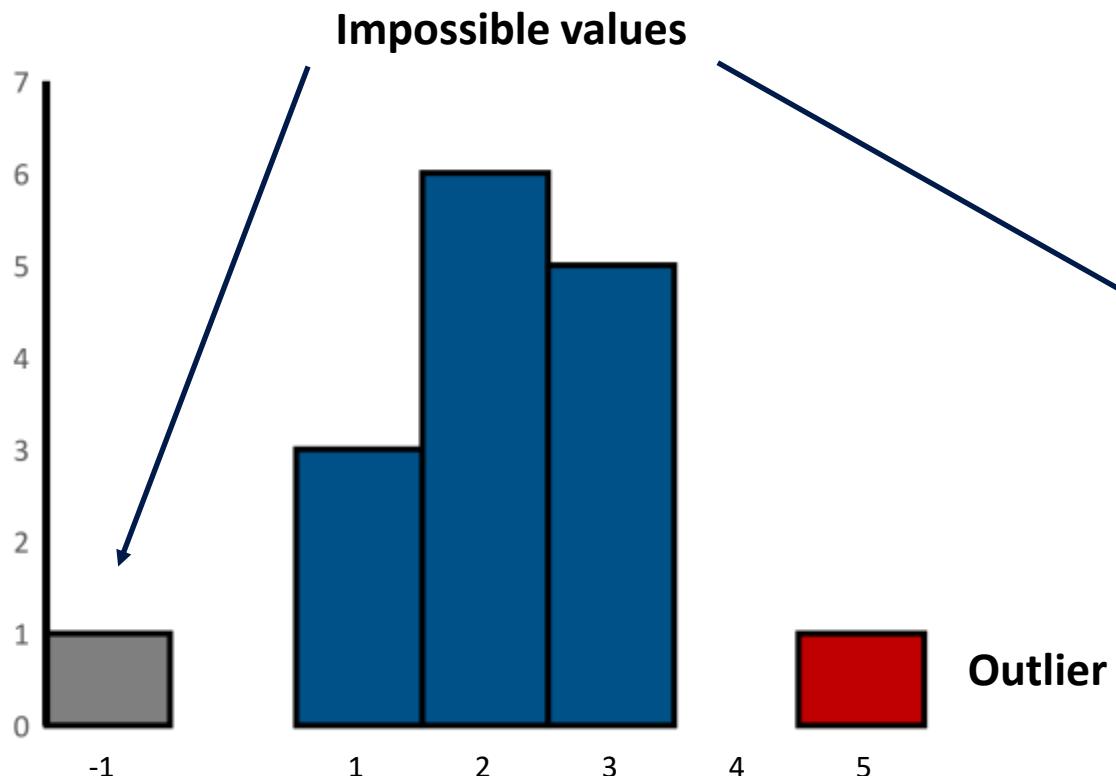
Problematic Data[1]

- What to do with problematic data?

  - Remove duplicates.

  - Check forms or other data sources to correct entry errors or impossible values:

    ◦ If you cannot correct the error, remove the data.

  - Consider outliers:

    ◦ Even though a value is theoretically possible, it may be an error.

    ◦ Outliers influence small datasets more than large datasets.
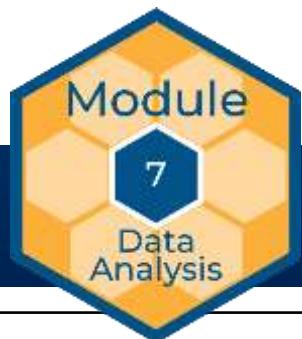
Response Rates

- If a survey has a low response rate, it may be difficult to determine whether the findings reflect the entire population.

- As a general rule, a response rate of 85 percent is high enough to ensure a representative sample.

- If a survey has a lower response rate, examine any possible reasons, such as the background characteristics of respondents compared to the characteristics of the entire population to which you intend to make inferences.[2]

Program Evaluation Toolkit

REL Central at Marzano Research

COLORADO   KANSAS   MISSOURI   NEBRASKA   NORTH DAKOTA   SOUTH DAKOTA   WYOMING

Additional Resources

- *Survey Methods for Educators: Analysis and Reporting of Survey Data*

Module 7 Data Analysis
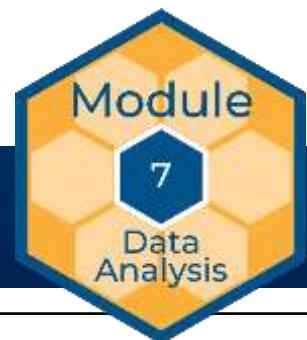
What Is Data Analysis?[3]

- Data analysis is the process of examining and interpreting data in order to answer questions.

- It summarizes collected data.

- It provides insight that drives decisions.

Program Evaluation Toolkit

Module 7 Data Analysis

Approaches to Data Analysis[4,5]

- Descriptive methods describe a sample. This can involve examining:
  - Counts or percentages.
  - Means, medians, or modes to look at the central tendency of a distribution.
  - Statistics such as standard deviation or interquartile range to look at the spread, or variation, of a distribution.

- Inferential methods involve drawing conclusions about a population from a sample. Inferential approaches may include techniques such as:
  - *t*-tests
  - Analysis of variance (ANOVA)
  - Correlation
  - Regression
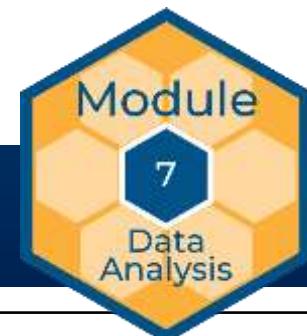
Which Approach Makes Sense?

- What approach to analysis is best suited to answering your evaluation question(s)?

  - Do you want to understand how a program was implemented or perceived?

  - Do you want to explore differences between groups (for example, groups of teachers with different characteristics, teachers who did or did not attend a training)?

  - Are you wondering if a program causes or is associated with changes in student outcomes?

Program Evaluation Toolkit

Module 7 Data Analysis

Descriptive Methods: Counts and Percentages

- Generating counts and percentages are straightforward descriptive methods.

- In the AMMP! example, 25 grade 8 students (a count) participated in the program, which corresponds to 20 percent (a percentage) of the total grade 8 student enrollment.

- Counts and percentages are especially useful for understanding the scope of a program or describing the individuals and sites involved.
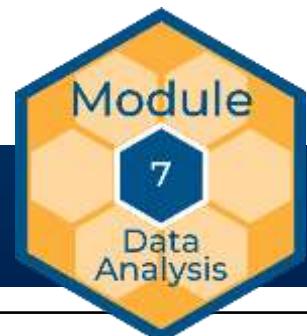
Descriptive Methods: Means, Medians, and Modes[5]

- *Means*, *medians*, and *modes* are statistics that convey unique information about the central tendency of a distribution.

  - **Mean**: The average response across a sample.

  - **Median**: The midpoint of a distribution.

  - **Mode**: The most common response in a distribution.

- In the AMMP! Example, the evaluation team finds that:

  - The mean, or average, student attendance is 95 percent.

  - The median student, or student at the midpoint of the data, attends 97 percent of classes.

 - The mode, or most common attendance rate for students, is 100 percent.

Descriptive Methods: Measures of Variation[5] (Part 1)

- Data may be narrowly or widely spread out, as these two frequency plots, or histograms, illustrate.
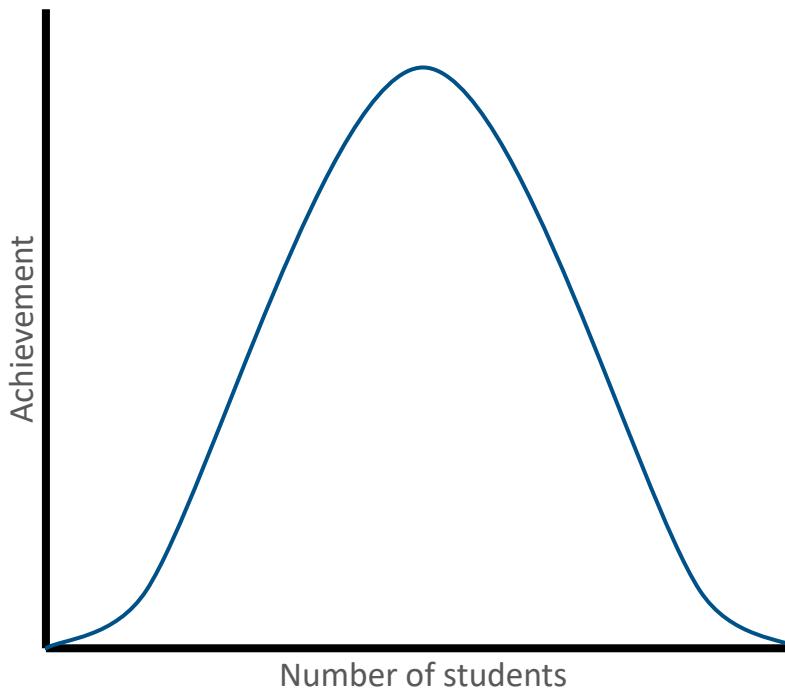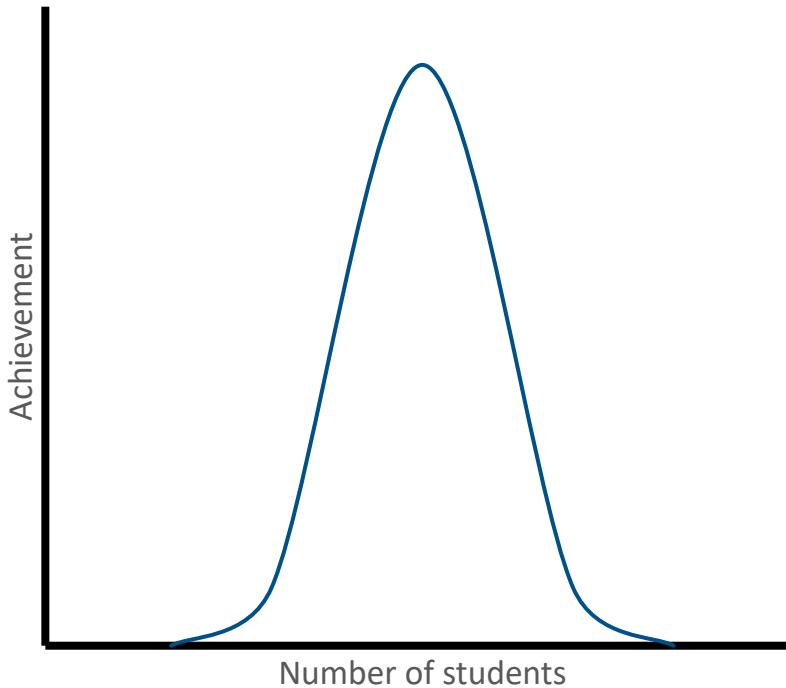
Descriptive Methods: Measures of Variation[5] (Part 2)

- *Standard deviation* is a common measure of variation. It indicates how spread out data points are. A larger standard deviation means that data are more widely spread out.
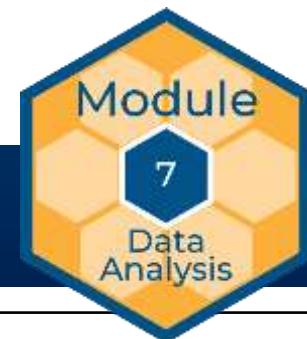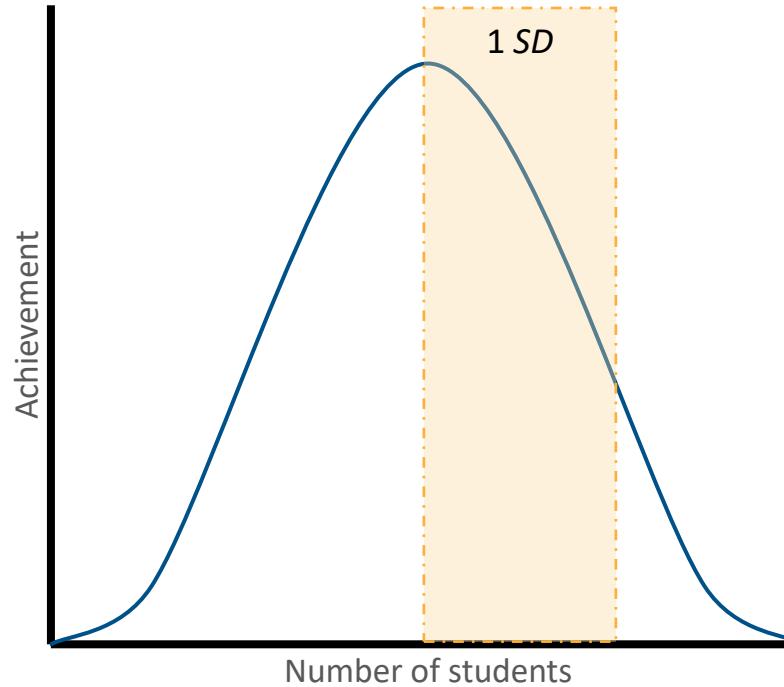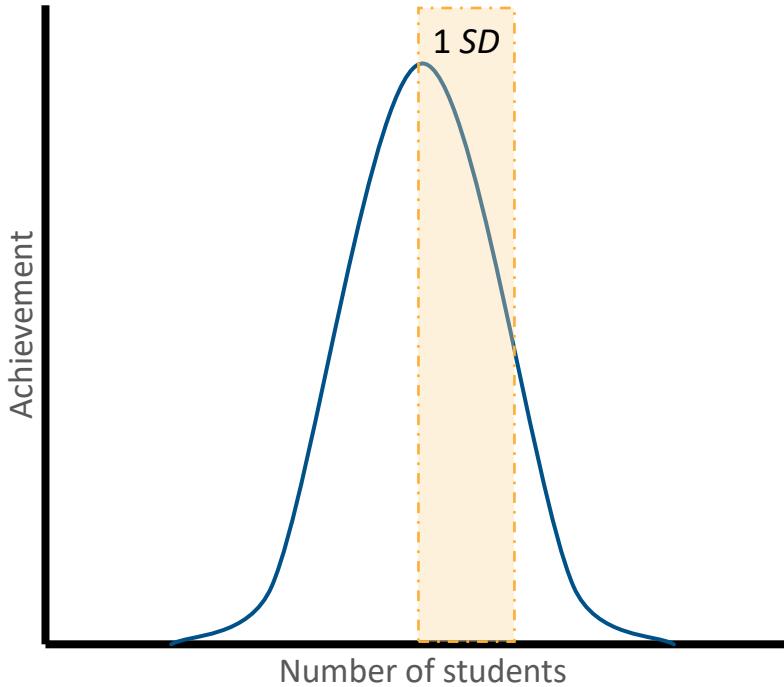
Descriptive Methods: Measures of Variation (Part 3)

- How do you interpret standard deviation?

- You have two classes, each with an average of 80 percent on a standardized exam. However:
  - Class 1 has a standard deviation of 5 percentage points.
  - Class 2 has a standard deviation of 10 percentage points.

- Thus, class 2 has a much larger spread in performance than class 1 does, despite having the same average score on the exam.
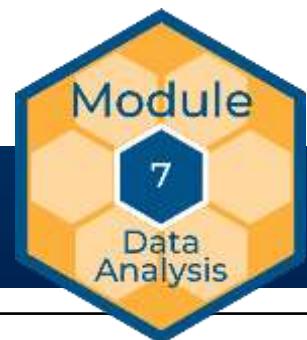
Descriptive Methods: Measures of Variation (Part 4)

- *Range* indicates the maximum and minimum observed values for a given variable.

- Interquartile range indicates the spread between the 25th percentile and the 75th percentile.
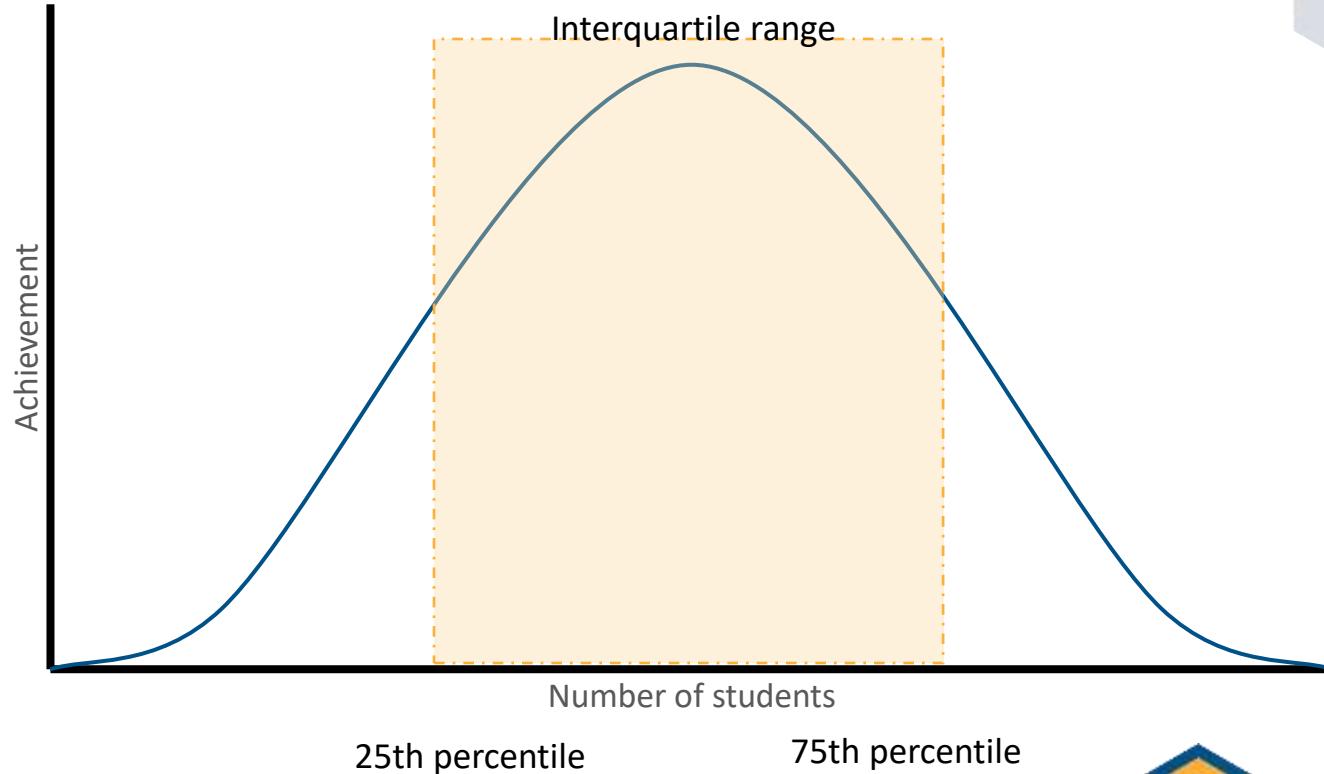
Descriptive Methods: Combining Measures
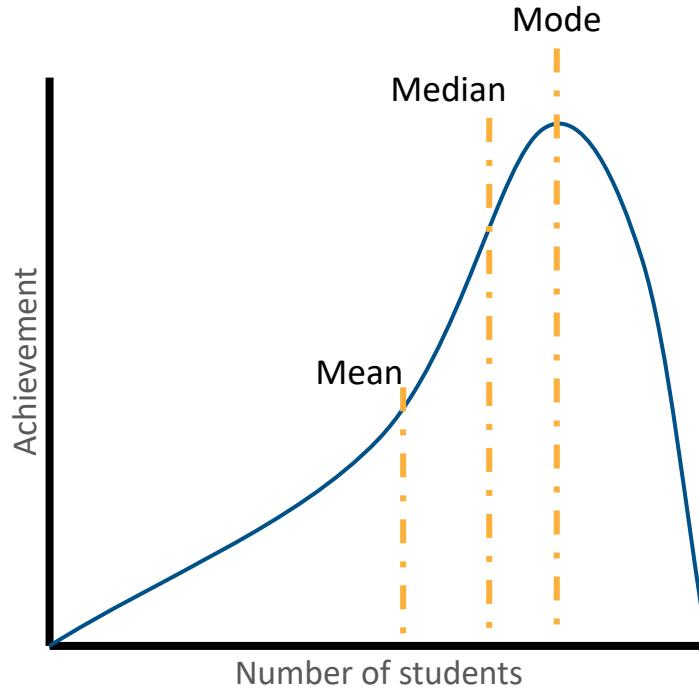
Figure 1:
Negatively skewed
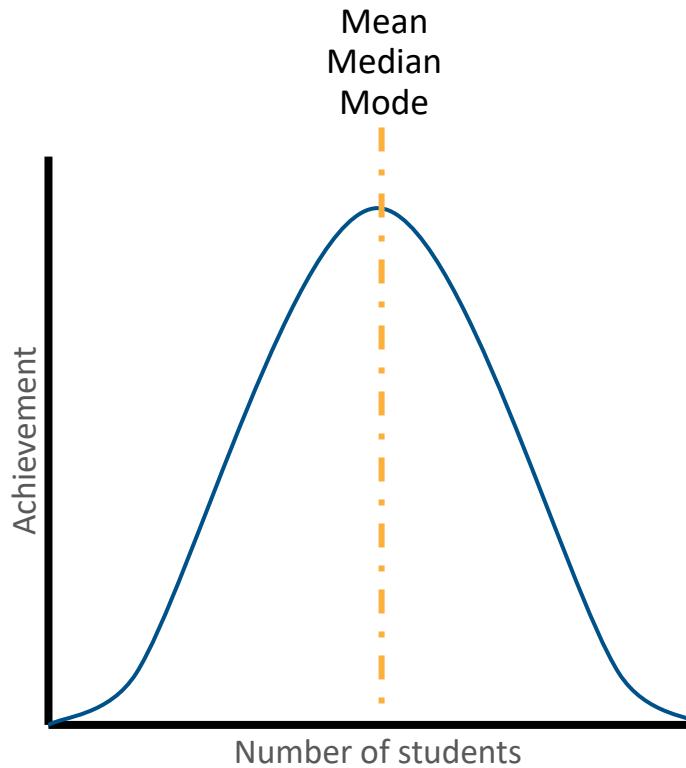
Figure 2:
Normal (no skew)
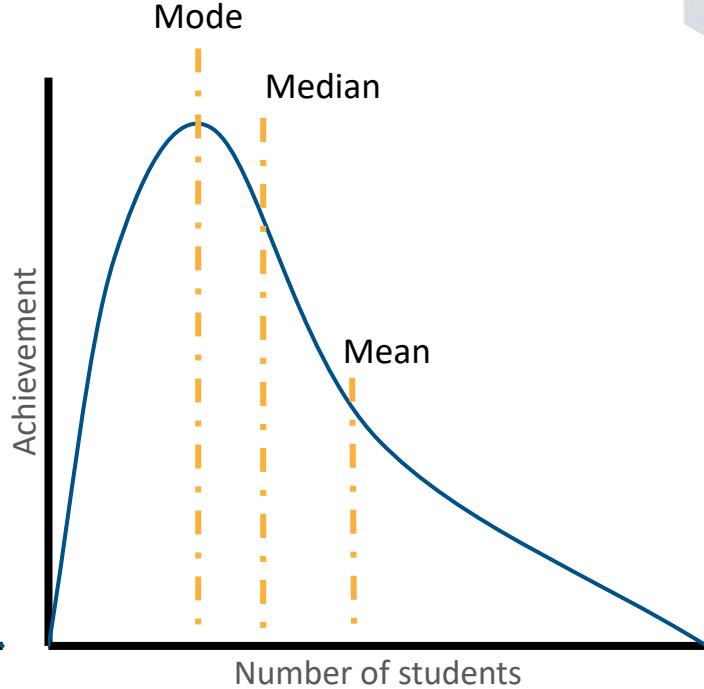
Figure 3:
Positively skewed

# Inferential Methods[5]

- Used to draw conclusions about a population from a sample.

- Inferential methods include:

  - *t*-tests

  - Analysis of variance (ANOVA)

  - Correlation analysis

  - Regression analysis
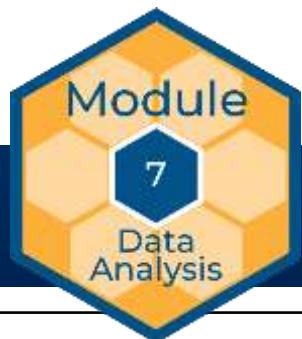
What Is Regression Analysis?

- Regression analysis is a family of statistical procedures that estimate relationships between variables.

- In its simplest form, regression analysis can show the relationship between two variables.

Types of Variables[4,5]

Dependent variable: A variable that may be predicted or caused by one or more other variables.

Independent variable: A variable that can have an influence on or association with the **dependent variable**.

Covariate: A variable that has a relationship to the **dependent variable** that should be considered but that is not directly related to the program.

Confound: A variable that creates a spurious relationship between the **independent** and **dependent variable**.

Program Evaluation Toolkit

Module 7 Data Analysis

Qualitative Methods[6]

- Qualitative methods are descriptive, allowing for a deeper exploration of "how" and "why" questions.

- Focus groups and interviews are common qualitative methods.

- To build validity, conduct focus groups or interviews in pairs so that one person takes notes while the other facilitates.

- Examine focus group or interview notes and transcripts systematically to identify themes or major topics. These themes may be predetermined by a codebook or may be identified ad hoc (open coding).

- To build reliability, have two individuals code and compare findings.

- For more information, visit the online *Qualitative Research* guide at https://guides.library.duke.edu/qualitative-research.

Program Evaluation Toolkit

REL Central at Marzano Research

COLORADO   KANSAS   MISSOURI   NEBRASKA   NORTH DAKOTA   SOUTH DAKOTA   WYOMING

Additional Resources
- *Qualitative Research Methods: A Data Collector's Field Guide*
- *Qualitative Research*

Module 7 Data Analysis

Cost Analysis[7]

- Federal and state agencies are increasingly considering the financial costs associated with programs and the relative benefits of implementing those programs.

- Cost analysis is the process of estimating what resources are needed to implement a program and what those resources will cost.

Program Evaluation Toolkit

REL Central at Marzano Research

COLORADO KANSAS MISSOURI NEBRASKA NORTH DAKOTA SOUTH DAKOTA WYOMING

Additional Resources

- *Cost Analysis: A Starter Kit*

Module 7 Data Analysis
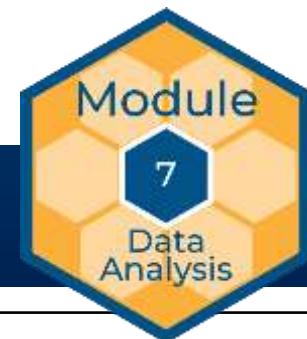
Chapter 1 Complete

Recommended next: Chapter 2 – Data Analysis Examples

Thank You

Please visit our website and follow us on Twitter
for information about our events, priorities, and research alliances,
and for access to our many free resources.

ies.ed.gov/ncee/edlabs/regions/central/index.asp
@RELCentral

or contact us at

RELCentral@marzanoresearch.com

REL Central at Marzano Research

COLORADO KANSAS MISSOURI NEBRASKA NORTH DAKOTA SOUTH DAKOTA WYOMING

References

1. Van den Broeck, J., Argeseanu Cunningham, S., Eeckels, R., & Herbst, K. (2005). Data cleaning: detecting, diagnosing, and editing data abnormalities. *PLoS Medicine, 2*(10), Article e267. https://doi.org/10.1371/journal.pmed.0020267

2. Pazzaglia, A. M., Stafford, E. T., & Rodriguez, S. M. (2016). *Survey methods for educators: Analysis and reporting of survey data (Part 3 of 3)* (REL 2016–164). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. Regional Educational Laboratory Northeast & Islands. https://eric.ed.gov/?id=ED567753

3. Rossi, P. H., Lipsey, M. W., & Henry, G. T. (2018). *Evaluation: A systematic approach* (8th ed.). Sage.

References

4. Loeb, S., Dynarski, S., McFarland, D., Morris, P., Reardon, S., & Reber, S. (2017). *Descriptive analysis in education: A guide for researchers* (NCEE 2017-4023). U. S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. https://ies.ed.gov/ncee/pubs/20174023/

5. Glass, G. V., & Hopkins, K. D. (1995). Statistical methods in education and psychology (3rd ed.). Allyn & Bacon.

6. Creswell, J. W., & Poth, C. N. (2016). *Qualitative inquiry and research design: Choosing among five approaches* (3rd ed.). Sage.

7. Institute of Education Sciences. (2020). *Cost analysis: A starter kit* (Version 1.1; IES 2020-001). U.S. Department of Education. https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=IES2020001rev

Program Evaluation Toolkit

Module 7 Data Analysis