Program Evaluation Toolkit
Module 7, Chapter 1: Data Analysis

Regional Educational Laboratory Central

From the National Center for Education Evaluation at IES

Speaker 1:

Welcome to the seventh module of the *Program Evaluation Toolkit*. It may be helpful to review modules 1 through 6 before beginning module 7, which refers to content from those previous six modules. Module 7 also features the AMMP! example introduced in module 1. AMMP! is a fictitious after-school middle-grades math program started in response to lower-than-expected homework completion rates and a lack of meaningful after-school activities for middle school students. The complete *AMMP! Logic Model*, which includes citations, is available on the resources page of the website.

With an understanding of logic models, the types of evaluation questions, evaluation design categories, a sampling plan, and data collection instruments, you now need to consider how to analyze data and generate implications from the findings. This module introduces common approaches to data analysis and walks you through two basic analyses in the context of the AMMP! evaluation. The module also provides a framework for understanding the implications of the findings.

This module is divided into three chapters, each covering key considerations as you progress from analyzing data to making recommendations based on the findings. Chapter 1 covers common approaches to data preparation and analysis. Chapter 2 focuses on two basic analyses, illustrated in the context of the AMMP! evaluation. Finally, chapter 3 introduces a framework for understanding the implications of the findings from data analyses.

Let’s get started with the first chapter, in which you will review data preparation before discussing data analysis.

Data preparation is essential to any evaluation. Even the best designed and implemented evaluation cannot achieve its goals if data are not collected, organized, and cleaned in a manner that ensures reliable analysis. This process can begin before data are even collected.

Before collecting data, it is important to develop a system to organize your data. Some strategies include the following.

Define a unit of measurement. For instance, are you collecting data from students? Teachers? Schools?

Assign a unique participant ID. This will help ensure that cases are not confused. Note that ID numbers often already exist for teachers, schools, and students.

Next, develop a codebook with variable names, response options, and numerical codes. This will allow you to clearly specify how data elements will be defined and operationalized.

The handout *Guidelines for a Codebook*, available on the resources page of the website, includes a more in-depth definition of what a codebook is and does, a strong rationale and guidance for developing a codebook, and a sample codebook to illustrate its core contents.

When you prepare your data, check for errors or inaccuracies. Data errors are simply differences between actual data values and reported data values. Data errors come from a variety of sources and can occur at any stage during the evaluation process, including data collection, data entry and cleaning, and data analysis. To better understand data errors, review the handout *Common Sources of Data Errors and Error-Checking Techniques*, available on the resources page of the website. In chapter 2 of this module, you will review some examples of data errors and learn how to check for errors and correct them.

There are a number of techniques that you can use to detect errors in order to clean data. One technique is simply spot-checking or eyeballing data—do any values look wrong? You can also perform a logic check—are any data outside the bounds of response options? You can also check for double entry—do any respondents have multiple entries where there should be only one? Finally, you can conduct descriptive analysis—calculating means, looking at distributions, and so on—to help you spot errors.

Luckily, you can use many functions in data analysis software to help you clean data. Microsoft Excel is the primary data cleaning and analyses tool used in this toolkit due to its wide availability. The handout *Microsoft Excel Functions for Data Cleaning*, available on the resources page of the website, lists and describes functions you can use to find duplicate cases and missing values as well as replace values. The handout also includes other functions you can use to expedite data cleaning.

Other programs such as SPSS, SAS (commonly referred to as SAS), Stata, and R are also useful in cleaning, coding, and analyzing data, and they include additional functionality that you can use to perform more complex tasks. These programs are also particularly helpful with larger datasets for which cleaning and coding cannot be done manually.

Creating visualizations of your data can be a good way to spot errors. Let’s say you are looking at a survey question with a 5-point scale and find the following distribution: three 1s, six 2s, five 3s, and one 5. In this case, the 5 is an outlier because all other responses are between 1 and 3. However, there is also a negative 1 and a 12. Whereas the 5 is an outlier and should remain in the dataset, the negative 1 and 12 are impossible values. Therefore, they should be corrected, if possible, to their true or intended values, or else they should be removed. This error might be the result of a software issue, or entry into a spreadsheet if responses were added manually.

What should you do if you identify problematic data? If you find duplicates, simply remove them. Correct all other data errors if possible. For instance, check for possible errors in transferring data from its original form or source to a dataset. Perhaps students completed a paper survey, and responses were then entered into a spreadsheet. If a value in the spreadsheet is supposed to be numeric but is instead text, referring back to the original paper surveys might reveal where the error originated.

If you cannot correct an error, remove the data from the individual cell but try to keep as much of the accurate data as possible. For example, if a student in a dataset has an impossible age, like 1,000, try to preserve the rest of the student's data and remove only the value for age.

Also consider outliers during data cleaning. Even though the value of an outlier is theoretically possible, it may be an error. Explore the outliers in your dataset to be confident that they are, in fact, extreme values but not errors. This is especially important if you have a small dataset, which will be more influenced by outliers than a large dataset will.

When you use survey data, consider examining response rates to ensure the generalizability of your findings. 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 your survey has a lower response rate, you may want to examine whether some groups were more likely not to respond to your survey than others. This is particularly important if the response rate is low among groups you intend to generalize to or if characteristics of nonrespondents might also be associated with key variables of interest on your survey. To learn more, refer to the *Survey Methods for Educators: Analysis and Reporting of Survey Data* report, available on the resources page of the website.

Once you are confident about the accuracy of your data, you can move on to data analysis.

What is data analysis? Simply put, it is the process of examining and interpreting data in order to answer questions. Data analysis summarizes collected data and provides insight that drives decisions.

This chapter briefly introduces common approaches to data analysis. Chapter 2 describes two simplified approaches that you may encounter in your evaluation. The goal is not necessarily to leave you with the tools to perform these common approaches but to help you understand data analysis and recognize how selecting an approach is informed by an evaluation plan and available data. By completing this module, you will have a better understanding of data analysis, even if colleagues or external evaluators conduct the analyses in your evaluation.

You can consult the additional resources referenced throughout module 7 and available on the resources page of the website if you are interested in learning more about data analysis.

There are two broad approaches to data analysis: *descriptive methods* and *inferential methods*.

Descriptive methods describe, or summarize, a sample. Descriptive methods can involve examining counts or percentages; means, medians, or modes to look at the central tendency of a distribution; and 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 methods may include techniques such as *t*-tests, analysis of variance or ANOVA, correlation, and regression.

The following slides discuss these two approaches in more detail.

Whether you use descriptive or inferential methods largely depends on the evaluation questions you are trying to answer.

Do you want to understand how your participants perceived a program or how schools implemented the program? If so, descriptive methods will suffice.

Do you want to explore differences between groups—for instance groups of teachers with different characteristics, or teachers who did or did not attend a training? In this case, descriptive methods will help you answer the question, but inferential methods might give you stronger and more generalizable findings.

Are you wondering if a program causes or is associated with changes in student outcomes? If so, use inferential methods because descriptive methods cannot imply causality.

Generating counts and percentages are straightforward descriptive methods. For instance, in the AMMP! example, the evaluation team determines that 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.

Means, medians, and modes are statistics that convey unique information about the central tendency of a distribution. The *mean* is the average response across a sample. The *median* is the midpoint of a distribution. The *mode* is the most common response in a distribution.

For instance, in the AMMP! example, the evaluation team wants to examine measures of central tendency for math class attendance. The attendance rates for five students are 82, 96, 97, 100, and 100 percent. The 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. And the mode, or most common attendance rate for the students, is 100 percent.

Taken together, the mean, median, and mode also gives the AMMP! evaluation team a sense of the distribution of attendance across participants. More on that soon.

Means, medians, and modes tell you about the central tendency of a distribution but not the spread of the distribution. Data may be narrowly or widely spread out, as these two frequency plots, or histograms, illustrate. The distribution on the right has a wider spread than the distribution on the left. Even though both distributions may have the same mean, there are more observations away from the mean in the distribution on the right. In other words, there is more variability in that distribution.

It is important to understand the spread of a distribution because it describes the entire sample better than averages do. In this example, the students represented in the histogram on the right have more variation in their levels of achievement than the students represented in the histogram on the left do. You would not understand this by looking at only averages.

One common measure of variation, called *standard deviation*, indicates how spread out data points are by describing how far the data are from the mean. A larger standard deviation means that data are further from the mean, or more widely spread out. In the two example histograms here, the standard deviation is wider on the figure to the right due to the greater variation in students' levels of achievement.

Understanding standard deviation can have useful implications. For example, by looking at the standard deviation in these histograms, you can get a better sense of the range of the students' academic ability.

How do you interpret standard deviation? It is helpful to think of standard deviation as indicating how far observations are, on average, from the mean. Standard deviation is in the same units—such as percentage points—as the measure itself.

Let's say 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, whereas class 2 has a standard deviation of 10 percentage points. In this case, you can conclude that the students in class 2 are about twice as spread out from the mean as are students in class 1. This conclusion is true despite the two classes having the same average score on the exam.

This is important for teachers to know. Compared to the teacher of class 1, the teacher of class 2 has more students who are performing very well but also more students who are performing quite poorly. After reviewing the data, the teacher of class 2 may choose to use a scaffolded approach to help learners at different levels.

Range is another common measure of variation. Range indicates the maximum and minimum observed values for a given variable. For instance, in the AMMP! example, the evaluation team finds the range for math achievement is from 35 points to 100 points.

Quartiles divide up the range of values in a dataset into four even segments. *Interquartile range* tells you the spread between the 25th percentile and the 75th percentile. In the AMMP! example, the interquartile range of math achievement goes from 82 points to 92 points. This tells the evaluation team where the middle of the distribution lies. Because the interquartile range is not influenced by extreme values or outliers, it gives a good indication of a range of middle values.

You may look at measures of central tendency in tandem with measures of variation to get a rich, descriptive picture of your data. Here are three figures representing distributions with the same mean and the same spread in data. However, the underlying patterns are quite different. Figure 1 is negatively skewed, figure 2 is symmetrically distributed, and figure 3 is positively skewed. Skewness gives you even more information about your data.

For instance, let's assume these figures represent three distributions of test scores. In the negatively skewed figure, the majority of students had high achievement, but the average was pulled down by relatively few low-scoring students. In the positively skewed figure, the majority of students had relatively low performance, with the average being pulled up by relatively few high-scoring students. The conclusions drawn from these data may be influenced by this skewness.

Now let's discuss inferential methods, which are used to draw conclusions about a population from a sample.

You may already be familiar with some inferential methods. For instance, *t*-tests can be used to determine whether the means of two groups in a population are likely to be different. Analysis of variance, or ANOVA, can be used to test for population differences in three or more groups. Correlation analysis generates correlation coefficients that indicate how differences in one variable correspond to differences in another. Correlation coefficients can be positive, indicating that, as one variable increases, the other variable also increases. Or coefficients can be negative, indicating that, as one variable increases, the other variable decreases. Correlations near positive 1 or negative 1 indicate a very strong relationship between the two variables, whereas correlations near 0 indicate a very weak, or no, relationship.

For any of these methods, it is important to remember that the conclusions you draw will depend, in part, on the evaluation design you select. For example, you cannot evaluate whether a program caused a change in student achievement unless you are using a rigorous design—a quasi-experimental design or a randomized controlled trial. Revisit module 3 for more on evaluation designs.

t-tests, ANOVA, and correlation analysis will not be discussed in depth in this chapter. To learn more about these methods, refer to the glossary of terms in the *Quick Start Guide*, available on the resources page of the website.

Instead, this chapter will focus on 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. For instance, the AMMP! evaluation team graphs math achievement on a state assessment (on the y-axis) and homework completion rates (on the x-axis) of group of individuals. A regression in this case is equivalent to plotting a best-fitting line through the data points. The dashed line on the graph shows a positive relationship—in other words, students with higher homework completion rates tend to have higher achievement. The slope of that line quantifies the relationship between homework completion and achievement—indicating about how much of a difference in achievement the team would expect given a unit change in homework completion. In other words, regression helps the team answer the question "Exactly how do changes in homework completion relate to achievement?"

Multiple regression analysis allows you to control for other factors by including additional variables. For instance, the evaluation team might want to look at how homework completion relates to math achievement while also accounting for such variables as student socioeconomic status and AMMP! participation.

Before conducting a regression analysis, it is important to identify how the variables will be used. A *dependent variable* is a variable that may be predicted or caused by one or more other variables. In the AMMP! example on the previous slide, math achievement is the dependent variable of interest, because the team is interested in examining how math achievement is influenced by homework completion. Dependent variables are sometimes called outcome variables.

An *independent variable* is a variable that can be introduced or adjusted to determine its influence on or association with a dependent variable. In the previous AMMP! example, homework completion is the independent variable because it may influence math achievement and strategies can be introduced to help students complete more homework.

A *covariate* is a variable that could explain differences in the dependent variable but that is not directly related to the program. Examples of covariates are student race/ethnicity, gender, socioeconomic status, and prior achievement. In the previous AMMP! example, prior math achievement is a covariate because it generally has a strong relationship to future math achievement. Therefore, differences in students' prior math achievement could explain differences in current math achievement, the dependent variable. As a result, the evaluation team would want to include prior math achievement as a covariate in their analyses. Approaches to include covariates in analyses will be described in chapter 2 of this module.

Finally, a *confound* is a variable that could be introduced that is not part of the program that could influence the dependent variable. One example of a confound is a related intervention being implemented at the same time as the intervention of interest. Another example is when there are characteristics in the group receiving the intervention that are not found in the group that does not receive the intervention, such as all teachers in the intervention group having advanced degrees. In the AMMP! example, a newly implemented supplemental math curriculum used in only the intervention schools would be a confound. The evaluation team must consider that this additional supplemental math curriculum may lead to higher math achievement, rather than the AMMP!. As a result, during the planning stages of the evaluation, the AMMP! team might consider excluding schools that are implementing the new supplemental curriculum. The team could also ensure that both intervention and control schools are both using the supplemental curriculum so that the effect is present in both conditions. In this way, the confound does not align completely with one condition and can therefore be taken into account during interpretation of the results.

In the next chapter, you will complete an activity to conduct a regression analysis and interpret the results, using a mock dataset.

Next, let's briefly discuss qualitative methods. Qualitative methods are descriptive, allowing 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. The pair can then help each other clarify what they heard and saw. Alternatively, a single data collector can audio-record the session, with the participants' permission, and later listen to or transcribe the conversation.

Afterward, examine the 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 (called open coding). Try to reduce the total number of themes so that it includes only the most essential. To build reliability, have two individuals code and compare findings. Do similar themes emerge? Why or why not?

Qualitative methods can provide rich contextual information that is essential for answering certain evaluation questions. *Qualitative Research Methods: A Data Collector's Field Guide*, available on the resource page of the website, provides more in-depth information on qualitative methods.

You can visit the online *Qualitative Research* guide by Duke University Libraries. This resource includes information from textbooks and experts as well as recommendations for software and tools for managing and analyzing qualitative data.

Federal and state agencies are increasingly considering the financial costs associated with programs and the relative benefits of implementing those programs. You might also conduct a cost analysis to help you consider which program you want to use to address your problem. Cost analysis is the process of identifying what resources are needed to implement a program and estimating what those resources will cost. If the estimated cost of a program means it is not feasible for you to implement it, then the cost analysis can prompt you to consider alternative programs or ways to procure other funding to implement the program.

Cost Analysis: A Starter Kit, published by the Institute of Education Sciences, describes a three-phase process that you might use. Phase 1 focuses on identifying the resources needed to implement a program, phase 2 involves understanding the costs of the resources, and phase 3 entails estimating the costs, thinking about assumptions, and using the findings. *Cost Analysis: A Starter Kit* is available on the resources page of the website.

This concludes chapter 1. In the next chapter, you will complete an activity to conduct a qualitative analysis and interpret results, using mock interview data, and you will complete an activity to conduct a regression analysis and interpret the results, using a mock dataset.

This handout was prepared under Contract ED-IES-17-C-0005 by Regional Educational Laboratory Central, administered by Marzano Research. The content does not necessarily reflect the views or policies of IES or the U.S. Department of Education, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.