Program Evaluation Toolkit

Chapter 3

1 2 3

Sampling Plan

Module 4 Evaluation Samples

Determining the Right Sample Size[1]

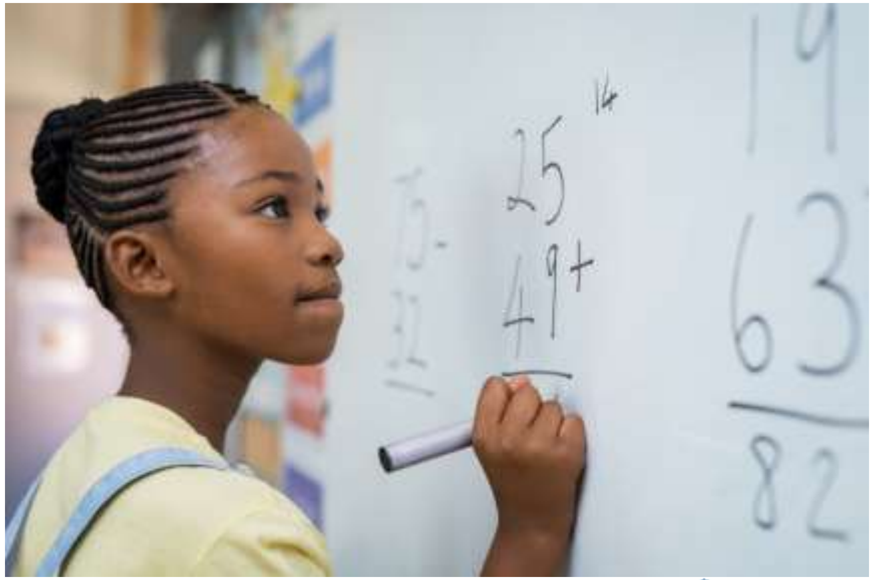
- Before collecting either qualitative or quantitative data, you should determine the adequate sample size.
- Qualitative data collection is less dependent on large sample sizes and tends to involve nonrandom sampling types.
- Quantitative data collection requires larger sample sizes and tends to involve random sampling types.

Determining the Sample Size for Nonrandom Sampling Types

- For qualitative data collection, consider the concept of saturation.
  - *Saturation* is the point at which the data collected begin to yield no new information.
- If the AMMP! evaluation team interviews teachers about student barriers to completing math homework, the team might reach a point at which the teachers identify no new barriers during the interviews.

Determining the Sample Size for Random Sampling Types or Purposive Sampling

- This chapter focuses on quantitative data collection using either random sampling types or purposive sampling.
- You want to determine the sample size needed to detect an effect or group difference related to your program.

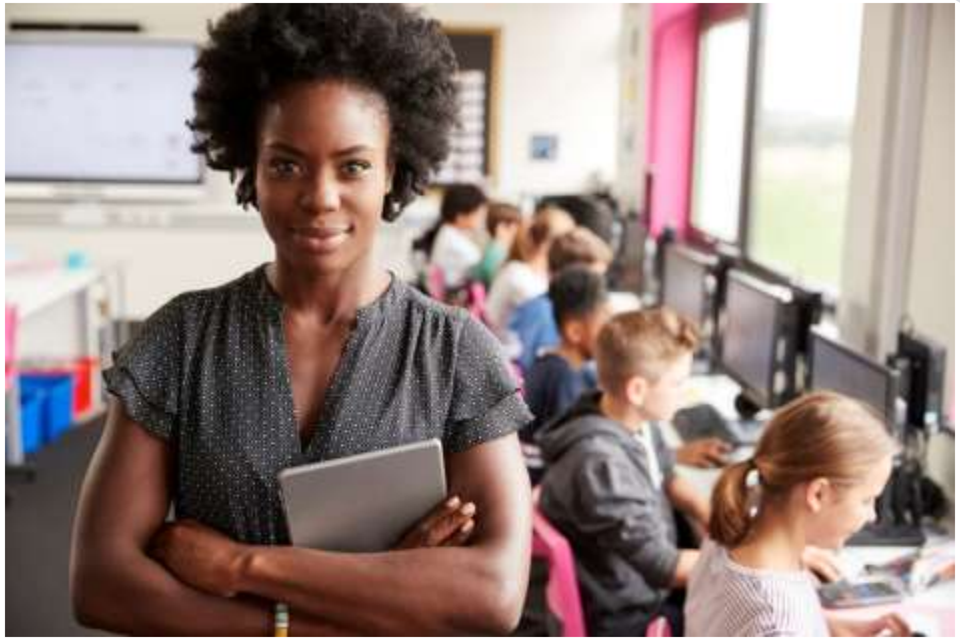


Identifying the Unit of Measurement

- The unit of measurement is the level at which you plan to collect data.
- You may need to protect individual identities by focusing on a higher level.
  - For example, obtaining average measures of classroom or school achievement rather than test scores of individual students.
    - To work with large populations.
    - To make inferences about classrooms or schools rather than individuals.
    - To protect individual personally identifiable information.

Identifying the Sampling Unit

- The sampling unit is the level at which you identify individuals for the sample.
  - Can be the same as unit of measurement (for example, teachers, students, parents, administrators).
  - Can be larger unit (for example, classrooms, grades, schools, districts).
  - Sample the highest level at which measurements are occurring (for example, sample schools if students, teachers, and principals at each school will be measured).



Limitations of the *Sample Size Workbook*[2]

- The *Sample Size Workbook* is useful for simple random sampling.
- You should not use the tool for stratified or clustered random sampling.
    - Choosing a sample size in these cases is more complicated.
    - More advanced tools are available to help in these cases. For instance, the PowerUp! tool is available for free:
        - https://www.causalevaluation.org/power-analysis.html

Determining the Sample Size Based on Inferential Statistics

- Methods for determining the appropriate sample size for quantitative studies are based on *inferential statistics*.
- There are two main procedures that make up inferential statistics:
  - Confidence intervals
  - Hypothesis testing

Confidence Intervals

- *Confidence intervals* are calculated for statistics generated from a sample of a larger population. These intervals depict a range of values in which the true value for the population would fall a certain proportion of times.
- Standard practice is to calculate 95 percent confidence intervals.
- When that proportion needs to be high, the range of values is typically wider and a larger sample size is typically required.

Hypothesis Testing

- *Hypothesis testing* involves writing a **null hypothesis of no effect**:
  - A statement that suggests that there will be no difference between a control and treatment group involved in an evaluation.
- Purpose of hypothesis testing:
  - To determine whether the information in the sample is sufficient to reject the null hypothesis of no effect.
- All conclusions in inferential statistics may be in error.
- Hypothesis testing is most concerned with the error of rejecting the null hypothesis when it is true.
  - Standard practice is to limit the probability of making an error of this sort to 5 percent.

Statistical Power

- Statistical power is the probability of rejecting the null hypothesis when a particular alternative hypothesis is true.
- Statistical power will increase if the sample size is larger.
- Standard practice is to ensure that statistical power is at least 80 percent.

Choosing a Procedure to Plan Sample Size

- You can use either confidence intervals or statistical power for planning the sample size in any type of evaluation.
- Confidence intervals may be more appropriate for descriptive and correlational designs in which there is no obvious comparison group or hypothesis of interest.
- However, statistical power is particularly appropriate for QEDs and RCTs because these designs compare two groups and have a natural null hypothesis (that the groups are the same on average).

Continuous and Binary Data

- The next step is to identify the type of quantitative data you will collect to inform your evaluation questions.
- Although there are many types of quantitative data, this chapter focuses on the two most common types: continuous and binary.
- Continuous data can take on a wide range of possible values.
    - Examples: student test scores, years of teaching experience, schoolwide percentage of students eligible for the National School Lunch Program.
- Binary data can take on only two values.
    - Examples: an exam with a pass/fail score, course completion, graduation, college acceptance.

Considering How Spread Out the Data Are

- When planning the sample size for an evaluation with continuous data, the required sample size will usually depend on how spread out the data are.
  - Described in more detail in Module 7, a standard deviation tells you how spread out your data are within a given sample.
- Over the next few slides, when the data are continuous, we will discuss distance in terms of standard deviation units. This will simplify sample size planning.
  - Note: When there are two groups of interest, calculations will be based on the assumption that the standard deviation in the two groups is the same.
- We will also use the *Sample Size Workbook* to calculate sample sizes for different scenarios.

Additional Resources

- *Sample Size Workbook*
- *Sample Size Workbook User's Guide*

Scenario 1: Single Mean

- Focus first on estimating a single mean for a continuous variable.
- Suppose AMMP! parents rate their satisfaction with their students' math homework completion on a scale of 0 percent to 100 percent satisfied.
  - The team assumes that the standard deviation of parent ratings is 4 percentage points.
  - The evaluation team wants to estimate the true average rating for all AMMP! parents to within plus or minus 1 percentage point.
- The evaluation team finds that they need 62 parents.

Confidence Intervals

Confidence interval for a single mean	
Desired interval width (in standard deviation units)	0.5
Required sample size	**62**

Confidence interval for comparing two means	
Desired interval width (in standard deviation units)	0.5
Proportion of sample in group 1	0.75
Required sample size	**328**

Confidence interval for a single proportion	
Desired interval width (in percentage points)	8
Estimate of true proportion	0.6
Required sample size	**577**

Confidence interval for comparing two proportions	
Desired interval width (in percentage points)	16
Estimate of true proportion in group 1	0.7
Estimate of true proportion in group 2	0.5
Proportion of sample in group 1	0.25
Required sample size	**705**

Confidence interval for correlation	
Desired interval width	0.3
Expected correlation	0.1
Required sample size	**168**

Confidence Intervals | Power-Mean Difference | Power-Proportion Diff

Enter desired interval width divided by the standard deviation here.

Program Evaluation Toolkit

Module 4 Evaluation Samples

Scenario 2: Two Means

- Suppose the evaluation team wants to compare parent satisfaction of homework completion between parents with students in AMMP! and parents with students not in AMMP!.
  - They want to estimate the difference between the true average rating for all AMMP! parents and the true average rating for all non-AMMP! Parents to within plus or minus 2 percentage points.
- The team needs to specify what proportion of the sample will be in each group.
  - Because there are fewer students in AMMP! the team samples 25 percent of AMMP! parents and 75 percent of non-AMMP parents.
  - They need to sample 328 parents.

Confidence Intervals

Confidence interval for a single mean	
Desired interval width (in standard deviation units)	0.5
Required sample size	**62**

Confidence interval for comparing two means	
Desired interval width (in standard deviation units)	0.5
Proportion of sample in group 1	0.75
Required sample size	**328**

Confidence interval for a single proportion	
Desired interval width (in percentage points)	8
Estimate of true proportion	0.6
Required sample size	**577**

Confidence interval for comparing two proportions	
Desired interval width (in percentage points)	16
Estimate of true proportion in group 1	0.7
Estimate of true proportion in group 2	0.5
Proportion of sample in group 1	0.25
Required sample size	**705**

Confidence interval for correlation	
Desired interval width	0.3
Expected correlation	0.1
Required sample size	**168**

Confidence Intervals | Power-Mean Difference | Power-Proportion Diff

Enter desired interval width divided by the standard deviation and the proportion of the sample in the first group here.

Scenario 3: Single Proportion

- Now suppose the team will ask parents to answer "yes" or "no" to the question "Are you satisfied with your student's homework completion?"
  - The team wants to estimate the true proportion of all parents in the district who are satisfied to within plus or minus 4 percentage points.
  - With binary data, instead of assuming a standard deviation, the team needs a ballpark estimate of the proportion who will answer "yes."
    - The team assumes that this estimate is 60 percent.
- The team finds that they need to sample 577 parents.
- Binary data require a larger sample size.

Confidence Intervals

Confidence interval for a single mean	
Desired interval width (in standard deviation units)	0.5
Required sample size	**62**

Confidence interval for comparing two means	
Desired interval width (in standard deviation units)	0.5
Proportion of sample in group 1	0.75
Required sample size	**328**

Confidence interval for a single proportion	
Desired interval width (in percentage points)	8
Estimate of true proportion	0.6
Required sample size	**577**

Confidence interval for comparing two proportions	
Desired interval width (in percentage points)	16
Estimate of true proportion in group 1	0.7
Estimate of true proportion in group 2	0.5
Proportion of sample in group 1	0.25
Required sample size	**705**

Confidence interval for correlation	
Desired interval width	0.3
Expected correlation	0.1
Required sample size	**168**

Confidence Intervals | Power-Mean Difference | Power-Proportion Diff

Enter desired interval width and estimate of true proportion here.

Scenario 4: Two Proportions

- Now suppose the team wants to compare the proportion of AMMP! parents who answer “yes” with the proportion of non-AMMP! parents who answer “yes.”
  - The team wants to estimate the difference between the true proportion of AMMP! parents who are satisfied and the true proportion of non-AMMP! parents who are satisfied.
    - The team assumes the proportion of satisfied AMMP! parents is about 70 percent, the proportion of satisfied non-AMMP! parents is about 50 percent, and 25 percent of the sample will be AMMP! parents.
    - The team wants to estimate the difference between the proportions to within plus or minus 8 percentage points.
- The team finds that they need to sample 673 parents.

Confidence Intervals

Confidence interval for a single mean	
Desired interval width (in standard deviation units)	0.5
Required sample size	**62**

Confidence interval for comparing two means	
Desired interval width (in standard deviation units)	0.5
Proportion of sample in group 1	0.75
Required sample size	**328**

Confidence interval for a single proportion	
Desired interval width (in percentage points)	8
Estimate of true proportion	0.6
Required sample size	**577**

Confidence interval for comparing two proportions	
Desired interval width (in percentage points)	16
Estimate of true proportion in group 1	0.7
Estimate of true proportion in group 2	0.5
Proportion of sample in group 1	0.25
Required sample size	**705**

Confidence interval for correlation	
Desired interval width	0.3
Expected correlation	0.1
Required sample size	**168**

Confidence Intervals | Power-Mean Difference | Power-Proportion Diff

Enter desired width, estimates of the true proportions in the two groups, and the proportion of the sample that will come from group one here.

Scenario 5: Causal Designs (Part 1)

- Now suppose that the evaluation team wants to make a causal comparison that applies to the whole district, and so the team will use power analysis.
- For instance, the team wants to know whether participation in AMMP! causes an increase in scores on high school math placement tests.
  - The team will compare the average scores of students who participate in AMMP! with the average scores of those who do not participate in AMMP!.
  - The team must first specify the anticipated average difference between AMMP! students and non-AMMP! students. The specification should be in standard deviation units. This specification is known as the *minimum relevant effect size* (MRES).[3]

Scenario 5: Causal Designs (Part 2)

- The evaluation team also needs to specify the proportion of the sample that will have participated in AMMP!.
  - The number of students who will have participated in AMMP! divided by the total number of students eligible to participate.

Scenario 5: Causal Designs (Part 3)

- The evaluation team wants to determine the causal impact of AMMP! participation on high school math placement test scores.
- Students will obtain different scores on the test. The extent of these differences is called the variation.
- AMMP! may explain some variation in scores but other variables may also explain variation in scores.
- To determine the required sample size, the evaluation team needs to estimate the percentage of variation in math scores that can be explained by other variables.

Scenario 5: Causal Designs (Part 4)

- In education research, it is common to assume MRES about 0.25.
- Again, assume 25 percent of sample will have participated in AMMP!.
- Finally, assume 60 percent of variance in reading scores can be explained by other variables.

Power-Mean Difference

Power analysis for comparing two means	
Effect size (MRES)	0.25
Proportion of variance explained by covariates (R^2)	0.6
proportion of sample in group 1	0.75
Required sample size	**268**

Note:

Effect size is the difference between the average scores in the two groups, expressed in standard deviation units.

Enter the effect size, proportion of variance explained by the covariates, and the proportion of the sample in group 1.

Now It's Your Turn²

- Choose an evaluation question that you drafted in module 2. Then fill in the sampling plan.

Additional Resources

- *Sampling Plan for Evaluation Questions*
- *Sample Size Workbook*

Program Evaluation Toolkit

Module 4 Evaluation Samples

Chapter 3 Complete

1 — 2 — 3

Recommended next: Module 5 – Data Quality

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

This presentation 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. of Department of Education, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

References

1. Anderson, S. F., Kelley, K., & Maxwell, S. E. (2017). Sample-size planning for more accurate statistical power: A method adjusting sample effect sizes for publication bias and uncertainty. *Psychological Science*, *28*(11), 1547–1562. https://doi.org/10.1177%2F0956797617723724
2. Dong, N., Kelcey, B., Maynard, R., & Spybrook, J. (2015). PowerUp! Tool for power analysis. https://www.causalevaluation.org/power-analysis.html
3. Schochet, P. Z. (2008). Statistical power for random assignment evaluations of education programs. *Journal of Educational and Behavioral Statistics, 33*(1), 62–87. https://doi.org/10.3102%2F1076998607302714