Data Analysis and Use

Michael Long
Andrew MacDonald

2/27/2018

OneStar Evaluation Institute

ICF

Learning Objectives

By the end of this presentation, you will understand:

- Strategies for analyzing qualitative data

- Strategies for analyzing quantitative data

- Tips for displaying results

Analysis of Qualitative Data

- Qualitative data are usually used to answer "how" or "why" research questions.

  - How are AmeriCorps members trained to deliver after school tutoring?

  - Why do program participants sign up for the intervention?

  - Why do some participants drop out of the program?

  - How did AmeriCorps members' career goals change during their service year?

  - Why did program beneficiaries change their behavior after attending seminars run by AmeriCorps members?

Analysis of Qualitative Data: Coding

- Analyzing qualitative typically requires some type of "coding"—that is, taking narrative notes or other information, and turning it into bite-sized chunks that can be more easily compared and understood

  - For example, if you do interviews with people that drop out of your program to find out why, you might read through the notes and identify each specific reason that they give

  - You could then tally up how many people give each specific reason, to see which are most frequent and need to be addressed

ICF

Analysis of Qualitative Data: Coding

- There are two different approaches to coding:

  - *A priori (or predetermined) coding* is when you already have a set list of codes that you are looking for in your data. For example, if you are looking at how beneficiaries heard about your program, your a priori codes might be "Flyer," "Referral," and "Word-of-Mouth."

  - *Emergent coding* is when you develop the list of codes based on what you see in the data.

- It is often useful to combine both approaches. For example, you could use a priori codes as an initial way to make sense of complicated data, but then add emergent codes if you see unexpected things.

Analysis of Qualitative Data: Interpreting Results

- Remember that while trends are important, outliers can also be informative!

  - For example, if almost everyone says that they are satisfied with your program, can you learn anything from the handful of exceptions? Is there anything common between them?

- If possible, look at patterns between different categories of codes. For example, if you identify a list of possible reasons that participants drop out of a program and a separate list of ways that they heard about the program, are there any relationships between those two lists?

  - For example, are people who heard about the program in a certain way more likely to drop out?

ICF

Analysis of Qualitative Data: Using Software

- Qualitative data analysis is often done manually (e.g., with a highlighter and a pen)

  - If this is your approach, save your notes, codes, counts, margin comments, etc. in case you need to return to the data later!

- Software does exist to help with qualitative analysis (e.g., NVivo, Atlas.ti)

  - Allows you to code electronically, store coded text, highlight specific phrases of interest

  - Can be helpful in identifying patterns that are hard to see manually

  - Can be expensive, and takes time to learn

ICF

Analysis of Quantitative Data

- Quantitative data are typically used to answer "counting" questions

  - Does program participants' knowledge improve after the training?

  - How many participants participate in the intervention?

  - How do participants rate the intervention?

  - How long do members typically stay in the field?

  - What percentage of members enroll in higher education after their service?

ICF

Types of Basic Quantitative Analyses

- **Average (or "mean")**

  - A measure of "central tendency"—what is the "middle" of the data?

  - Can be affected by outliers, which can produce misleading results

- **Median**

  - If you line up all the numbers, this is the one in the middle

  - Like average, is a measure of "central tendency"

  - Unlike average, median is not affected by outliers

- **Minimum and maximum**

  - The lowest and highest numbers in a list

  - Gives a sense of how much the numbers vary

Example of Basic Quantitative Analyses

- Imagine that you give 7 trainings, and are analyzing the length of the trainings. In hours, the lengths are:

  - 1, 1, 1, 2, 2, 3, 16

- **Minimum**: 1 hour

- **Maximum**: 16 hours

- **Average**: 4 hours

- **Median**: 2 hours

Frequencies

- Frequencies represent the number and percentage of times that specific numbers or responses occur. For the example on the last slide, the frequencies of lengths for the training sessions are:

  - 1 hour: 3 sessions (43%)

  - 2 hours: 2 sessions (29%)

  - 3 hours: 1 session (14%)

  - 16 hours: 1 session (14%)

- Frequency percentages should total to 100% (or very close to it)

ICF

Cross-Tabulations

- Cross-tabulation analyses (or "cross-tabs") break frequencies down based on a characteristic of the respondent or participant, in order to compare data for different groups

  - For example, if you asked all participants to rate a training…

	Participant Rating				
	Excellent	Good	Fair	Poor	Total
Overall Frequency	119 (40%)	75 (25%)	70 (23%)	36 (12%)	300

Cross-Tabulations

- Cross-tabulation analyses (or "cross-tabs") break frequencies down based on a characteristic of the respondent or participant, in order to compare data for different groups

  - For example, if you asked all participants to rate a training…

	Participant Rating				
Age	Excellent	Good	Fair	Poor	Total
<30	108 (72%)	25 (17%)	13 (9%)	4 (3%)	150
=>30	11 (7%)	50 (33%)	57 (38%)	32 (21%)	150
Overall Frequency	119 (40%)	75 (25%)	70 (23%)	36 (12%)	300

More Advanced Analysis Techniques

- **Correlations**: Measure how related two characteristics are (i.e., if you change one, you would expect the other to change as well)

  - E.g., for children, the correlation between age and height would be high

- **Statistical significance tests** (e.g., t-tests, chi-square tests): Measure whether the difference between two groups is "significant"— that is, is "real" and not just a result of coincidence

  - E.g., for adults, the difference in height between men and women is statistically significant; the difference in height between people in Charleston and South Charleston is probably not

More Advanced Analysis Techniques

- **Regression Analyses**: Measure the relationship between different variables while statistically "controlling" for other variables

  - E.g., You might find that girls score higher on a post-test than boys—but that girls also scored higher on the pre-test. A regression allows you to "control" for the pre-test scores, and see whether girls would still have scored higher than the boys if their pre-test scores had been the same.

ICF

Exercise: Brainstorming Types of Data Analysis

- Imagine that there are three trainers in your program who have run a total of 50 training sessions in the past three months, and for each session you have the following data:

  - Date of training session

  - Trainer who facilitated the session

  - Length of training session

  - Location of training session (i.e., school, community center, or church)

- For each of the sessions, you also have surveys completed by the participants, in which they rated the value of the training on a scale of 1 to 5.

- **What questions could you answer with these data, and what analysis would you do?**

ICF

Exercise: Examples of Research Questions/ Possible Analyses

- "On average, how long were the trainings?"

  - Calculate average and median training length

- "How much did the length of the trainings vary?"

  - Calculate the minimum and maximum training length

- "To what extent did the trainings last the expected amount of time?"

  - Calculate the frequency that the training length was between X and Y

- "Did the training length depend on the location of the training or the trainer?"

  - Calculate average and median length for trainings in each type of location and trainer separately

Exercise: Examples of Research Questions/ Possible Analyses (continued)

- "Over time, have the trainings gotten any longer or shorter?"

  - Compare the average/median length of the first 25 trainings to the average/median length of the last 25 trainings

- "To what extent did participants find the training valuable?"

  - Calculate the average rating among participants in all sessions

- "How varied were participants' ratings of the training?"

  - Look at frequencies of participants' ratings (i.e., % giving each rating)

- "Did participants' ratings depend on the location of the training or the trainer?"

  - Calculate cross-tabulations of rating responses vs. training location, and responses vs. trainer
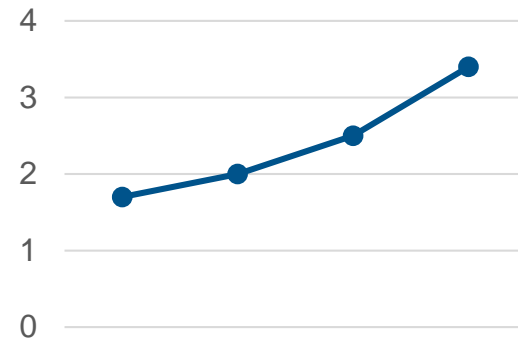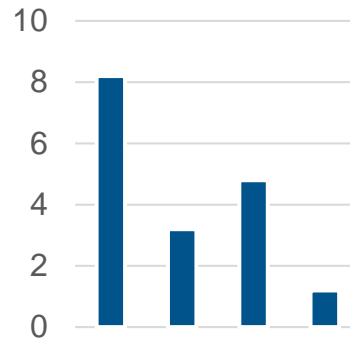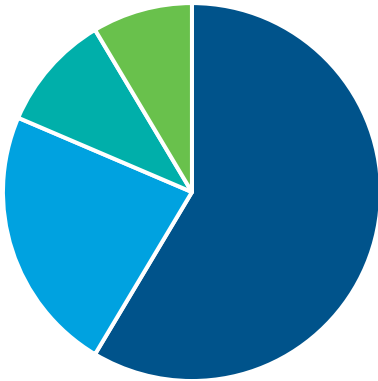
Software for Quantitative Analysis

- Unlike qualitative data, quantitative data are typically not analyzed manually using a paper and pencil (although in theory, you could)

- Most basic analyses can be conducted in Excel

- For more advanced analyses, software like SAS, SPSS, or Stata is typically used

  - Can be expensive and difficult to learn

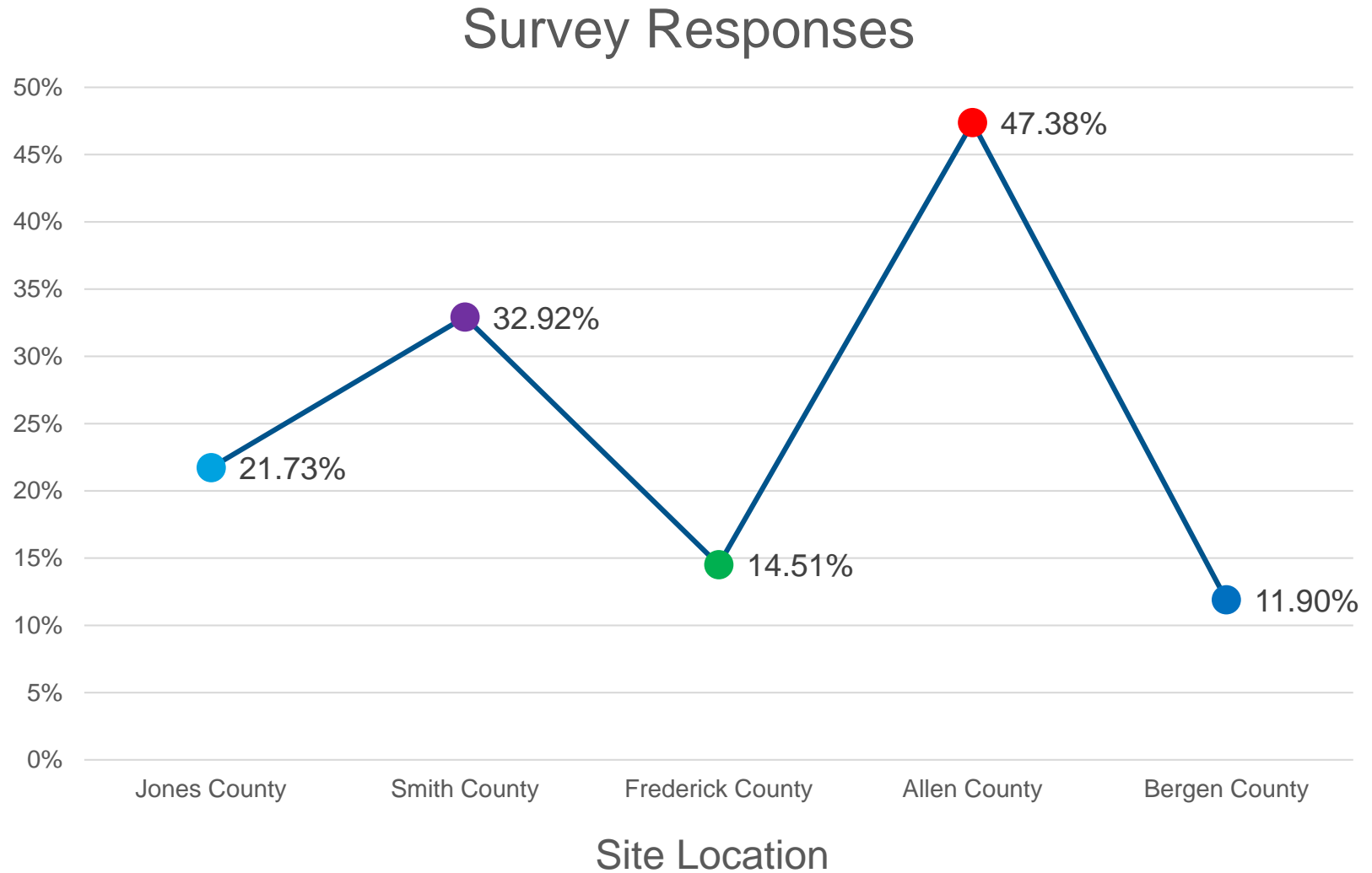**ICF**

Tips on Showing/Reporting Data

- Keep the display as simple as possible

  – There is no reason to say 56.12% when you could say 56%

  – Consider combining categories to simplify the "number of numbers" on the page (e.g., combine "strongly agree" and "agree")

- Think about what you want the reader to focus on

  – Do not add potentially distracting colors, bolding, etc. unless you have a reason

- Unless you have a reason not to, show the data in an intuitive way— that is, in a way that the reader would expect

  – List categories in a logical order

Tips on Showing/Reporting Data

- If you are charting data, choose the right kind of chart
  - A **pie chart** is a good way of showing frequencies—that is, parts of a whole
  - A **bar chart** is a good way of comparing different categories of data
  - A **line chart** is a good way of showing a trend over time

Example of Flawed Chart

Survey Responses

- Jones County: 21.73%
- Smith County: 32.92%
- Frederick County: 14.51%
- Allen County: 47.38%
- Bergen County: 11.90%

Site Location

ICF

Example of Improved Chart

Percentage of Members Who Are "Very Satisfied" With Their Experience

Bar chart showing percentage of members very satisfied by site location:
- Allen County: 47%
- Smith County: 33%
- Jones County: 22%
- Frederick County: 15%
- Bergen County: 12%

Y-axis: 0% to 50%. X-axis label: Site Location

ICF

Reporting Qualitative Data

- Use pull quotes or call-out boxes to highlight important points

- Consider alternatives to blocks of text

  - Bullet lists

  - Timelines

  - Word clouds

- Infographics are increasingly being used by organizations to highlight key evaluation findings in an engaging way

- For more ideas, see http://annkemery.com/qual-dataviz/

ICF