Step Three — Summarize and Analyze Your Data — *Quantitative Methods*

The first step in analyzing quantitative data is to summarize the responses using *descriptive* statistics. When you collect and summarize quantitative data, your result is a *distribution of scores* for each item on your survey (except open-ended items). A distribution is simply the collection of all ratings or scores for a particular item, ordered from the lowest to the highest value. Table 2 presents some of the most common descriptive statistics: frequency counts, percentages, and measures of central tendency (mean, median, and mode).

Table 2: Examples of Descriptive Statistics

Question: *Please indicate your level of agreement with this statement.* I am more confident about finding prescription drug information on the Web after taking this training session.								
Response		Strongly agree	Somewhat agree	Uncertain	Somewhat disagree	Strongly disagree	Total	Missing
Response value		(5)	(4)	(3)	(2)	(1)		
N	100							
Frequencies		54	36	5	2	0	97	3
Percent		54.0%	36.0%	5.0%	2.0%	0.0%	97.0%	3.0%
Valid Percent		55.7%	37.1%	5.2%	2.1%	0.0%		
Mean	4.41							
Median	5							
Mode	5							

Definitions

N	Number of people responding to the survey. (Note: 100 people returned a survey, but only 97 responded to this particular question.)
Frequencies	The number of respondents choosing each response.
Percent	The number of those choosing that response divided by the number of people who completed the *survey*.
Valid Percent	The number of respondents choosing that response divided by the number of respondents who answered the *question*. In this example, we had 100 people complete the survey, but only 97 actually responded to this particular question.
Mean	The mean is the "average" response in your distribution. It is computed by adding all responses and dividing by the number of respondents who answered the *question*.
Median	The median is the score that is in the middle of the distribution, with half of the scores above and half below. To find it, sort your distribution from highest to lowest ratings, then find the number that equally divides the distribution in half. For the 97 people who completed this distribution, the 49th score divides the distribution in half. The 49th (median) score is a "5." When the majority of ratings fall either at the high or low end of a rating scale, as they do here, the median is usually the preferable measure of central tendency because it is not affected by a few extremely low or high ratings.
Mode	The mode is the most frequent response. For many demographic and two-option questions, the mode is the only measure of central tendency that can be reported. This is also true for questions that ask respondent to provide more than one response, such as "check all that apply" questions.

Table 3: Participants' Self-Report of Confidence in Using Databases N=50

	Strongly Agree	Agree	Neither Agree or Disagree	Disagree	Strongly Disagree
The training session helped me develop more confidence in using MedlinePlus.	23 46%	16 32%	9 18%	2 4%	0 0%
The training session helped me develop more confidence in using PubMed.	10 20%	22 44%	13 26%	3 6%	2 4%
Analysis: The majority of respondents agreed or strongly agreed that the training sessions helped them gain confidence in using the NLM online resources. Ratings seemed to be slightly more positive for MedlinePlus. This indicates that we achieved our objective of increasing confidence in use of online resources with the majority of our participants.					

Tables are very helpful for understanding your data. Tables 3-7 show formats that will help you analyze your descriptive data. After you compile a table, write a few notes interpreting the numbers.

You may simplify your data to make the positive and negative trends more obvious. For instance, in Table 4, the "Strongly Agree" and "Agree" responses were combined into a "Positive" category and the "Disagree/Strongly Disagree" responses were put into a "Negative" category.

Table 4: Participants' Self-Report of Confidence in Using Databases N=50

	Positive (Strongly Agree/ Agree)	Neutral (Neither Agree or Disagree)	Negative (Disagree/Strongly Disagree)
The training session helped me develop more confidence in using MedlinePlus.	39 78%	9 18%	2 4%
The training session helped me develop more confidence in using PubMed.	32 64%	13 26%	5 10%
Analysis: This table makes the pattern of positive ratings more obvious for the items introduced in Table 3. It also confirms that ratings were more positive for the MedlinePlus session compared to the PubMed session. One explanation might be that PubMed is more difficult to use and requires a longer training session or more training sessions compared to MedlinePlus.			

Table 5: Average Number of NLM Resources Used Before and One Month After Training N=80

	Average # of Websites Before Training	Average # of Websites One Month After Training	Difference
How many of the following websites have you used in the past month. (Check all that apply of 6 resources.)	1.85	3.37	1.52
Analysis: Of the six websites we demonstrated in the training session, participants on average had used less than two of them before training. One month after training, they had, on average, visited more than three of the websites. This finding suggests that we chose websites that our participants found to be useful.			

Sometimes, you may want to see how participants' attitudes, feelings, or behaviors have changed over the course of the project. Table 5 also shows you how to organize pre-project and post-project data into a chart that will help you assess change. Table 5 also presents means rather than percentages. Data that represent a wide range of scores, such as attendance rates for a large number of training sessions, sometimes are easier to analyze using averages. You could also use means or medians in place of percentages if you have rating scales such as those presented above in Step 1 (see Example 4).

You may wonder if the findings vary for the different groups you surveyed. For instance, you may wonder if nurses, social workers, or members of the general public found your resources as useful as the health librarians who had your training. To explore this question, you would create cross-tabulation tables, as in Table 6.

Table 6: Average Number of NLM Resources Used Before and One Month After Training Broken Down by Profession N=80

	N	Average # of Websites Before Training	Average # of Websites One Month After Training	Increase in Use
Health Science Librarians	20	3.7	4.3	.6
Social Workers	20	1.3	3.0	1.7
Nurses	20	2.2	3.6	1.4
General public	20	.2	2.6	2.4
Analysis: We did not seem to increase the variety of websites used by the health science librarians, probably because, on average, they already had used more than half of the websites we demonstrated. Our training seemed to have the greatest impact on the general public, who had used very few of the websites. For planning future sessions, we may want to conduct a preliminary survey to find out what websites are popular with health science librarians so we can adjust our training content to cover websites they do not know.				

Table 7: Comparison of Those Who Used Resources After Training Compared to Targets in Objectives

	Actual	Goal	Difference
Numbers of participants using MedlinePlus after training	62%	50%	+12%
Number of participants using PubMed after training.	45%	50%	-5%
Analysis: We exceeded our criterion for the number of participants who used MedlinePlus after they took our training sessions. However, we were slightly under our goal for PubMed. On the other hand, because PubMed is more academic and MedlinePlus is more consumer-oriented, it is possible our users simply had more occasion to use MedlinePlus the month following the session. We may want to explore this in a follow-up interview with a few users who took both sessions to see if there are ways to improve the PubMed training.			

Finally, you also may want to compare your findings against the criteria you identified in your objectives. Table 7 gives an example of how to present a comparison of objectives to actual results.