



TIPS & TOOLS #10: ANALYZING QUANTITATIVE DATA

Statistical analysis can be quite involved. However, there are some common mathematical techniques that can make your evaluation data more understandable. Called **descriptive statistics**¹ because they help describe raw data, these methods include:

- Numerical counts or frequencies
- Percentages
- Measures of central tendency (mean, mode, median)
- Measure of variability (range, standard deviation, variance)

Numerical counts--frequencies

Counts or frequencies tell us how many times something occurred or how many responses fit into a particular category.

- The Youth Tobacco Purchase Survey was conducted in 4662 stores statewide.
- Seven of the twelve participants in the cessation program said that this was their first attempt to quit smoking.
- The city has a total of sixteen pharmacies that sell tobacco products.

In some cases, numerical counts are all that is needed or wanted. In other cases, they serve as a base for other calculations. One such calculation is the percentage.

Percentages

A commonly used statistic, the percentage expresses information as a proportion of a whole, for example:

- Forty-two percent of the illegal sales to minors occurred at convenience stores that sell gasoline.
- Of the 4662 stores, 22 percent were convenience stores with gas, and 19 percent were small markets.
- Sixty percent of the pharmacies in Utopia city said that they would commit to a voluntary policy of not selling tobacco products, and an additional 10 percent mentioned that they would do so next calendar year.

¹ Techniques that allow one to generalize from one group to a larger group are known as tests of statistical significance and fall within the body of knowledge called inferential or inductive statistics.

Percentages tend to be easy to visualize because they show part of a whole. They can easily be transferred into bar graphs, pie charts, and other images and help readers understand the importance of a value.

Percentages are also a good way to show relationships and comparisons- either between categories of respondents or between categories of responses. For example:

- Thirty percent of multi-unit housing complexes in the city have adopted a smoke-free policy. This is up from 10 percent in 1995 (comparing 2010 respondents to 1995 respondents).
- Eighty-five percent of those asked said that they preferred a smoke-free policy, ten percent said that they were against it, and five percent answered that they were not sure (comparing respondents in the same survey).

Percentages are also useful when we want to show a frequency distribution of grouped data. The frequency distribution is a classification of answers or values into categories arranged in some order (size or magnitude is not necessarily the choice of order, it could simply be alphabetical). The following table provides an example:

Table 1. Frequency distribution of participants in an informational meeting
n=120

Place of residence	Frequency	Percentage
Sacramento	69	57.5
Davis	29	24.2
Woodland	22	18.3

When reporting a percentage, common practice is to indicate the number of cases from which the percentage is calculated-either the "N" (the total group) or the "n" (the subsample/subgroup).

Although computing percentages appears to be a simple process, there are a number of possibilities for making errors.

1. Use the correct base

The base (denominator or divisor) is the number from which the percentage is calculated. It is important to use the right base and to indicate which base you've used. Does 75 percent mean 75 percent of all participants, 75 percent of the participants sampled, 75 percent of those who answered the question, or 75 percent of the respondents to whom the question applied?

Sometimes we use the total number of cases or respondents as the base for calculating the percentage. However, erroneous conclusions can result. This is particularly true if the proportion of "no response" is high. For example, we have questionnaires from 100 respondents but not all answered all the questions. For a certain question, 10 people did not respond, 70 answered "yes", and 20 answered "no". If we use 100 as the base or divisor, we show that 70 percent answered "yes". But if we use 90 as the base (those who actually answered the question), we

find that 78 percent of those who responded reported “yes”. We do not know whether the “no response” would have been “yes” or “no”. Consequently, in the analysis, it is essential to say that 10 percent did not answer (table 2 below) or omit the 10 “no answers” in the divisor (table 3).

Table 2. (n=100 participants)

YES	70%
NO	20%
NO RESPONSE	10%

Table 3. (n= 90 respondents)

YES	78%
NO	22%

There are many situations in which a question is not applicable to a respondent. Only the number of persons who actually answer the particular question is used in calculating the percentage. (It is important to make a distinction between not applicable and no response, table 3 would be misleading without the information that 10 out of 100 participants chose not to respond, table 2 is better)

2. Rounding percentages

Round off percentages to the least number of decimal points needed to clearly communicate the findings. To show too many digits (56.529%) may give a false impression of accuracy and make reading difficult. However, showing no decimal points may conceal the fact that differences exist. In rounding percentages, the rule of thumb is that five or greater is rounded off to the next higher number.

3. Adding Percentages

Percentages are added only when categories are mutually exclusive (do not overlap). This is not the case in multiple choice questions where the respondent may select several answers. For example, in a question asking respondents to indicate which cessation method they have tried in the past, the respondent might use one or several of the possible answers: cold turkey, acupuncture, cessation class, nicotine patch, etc. These answers are not mutually exclusive and their percentages should not be added.

4. Averaging percentages

Avoid the error of adding percentages and taking an average of the summed percentages. This is done frequently, but is never justified. The following table provides an example.

Table 4. Number of participants completing the cessation workshop (n=192[32 participants enrolled per workshop])

Workshop date	Number of completions	Percent
January 06	10	31.25
March 06	12	19.2
May 06	12	19.2
July 06	25	80
September 06	26	81.25
November 06	29	90.6
Total	148	Average (also called the mean) = 53.58 (incorrect; the correct average or mean is 59)

In this table the evaluator incorrectly reported an average of the participants completing workshops in 2006 by adding the percentages of workshop participants who completed each workshop and dividing that number by six. Rather, the average percentage should have been calculated by dividing the total number of participants completing the workshop (114) by the total of enrolled participants in all workshops (192) for a total of 59 percent. Sometimes the difference is not great. On other instances, the error can be quite large.

Measures of central tendency

Measures of central tendency are used to characterize what is typical for the group. These are measures which allow us to visualize or identify the central characteristic or the representative unit. For our purposes, the most likely measures to be used are the mean, the mode, and the median.

Mean

The mean, or average, is commonly used in reporting data. It is obtained by summing all the answers or scores and dividing by the total number.

For example, to get the average number of tobacco advertisement displays at seventeen rodeos that you are observing, divide the total number of advertisement displays at all the rodeos observed by the total number of rodeos (17). To report the mean number of smoking scenes in a series of observed films, divide the number of observed smoking scenes by the number of films observed.

Mean scores can also be used to summarize findings from rating scales, such as questions about attitudes or opinions. Categories such as "not very important," "somewhat important," and "very important," can be assigned numbers such as 1, 2, and 3. They could be reported in this way:

- On average, 108 apartment residents surveyed rated the importance of having smoke-free units at 4.2, on a scale of one to five, with five being "very important," and 1 being "not important."

A calculation of the mean in a rating of a program's usefulness might look as follows:

Table 5. Program usefulness (n= 100 participants)

	Poor	Fair	Good	Excellent	N
A. Gave me practical information I can use at work	1 (no answers)	2 (10 answers)	3 (60 answers)	4 (30 answers)	100
B. Increased my understanding of the subject	1 (no answers)	2 (20 answers)	3 (70 answers)	4 (10 answers)	100
C. Stimulated me to find out more about the subject	1 (no answers)	2 (20 answers)	3 (30 answers)	4 (50 answers)	100

The mean rating for each item is calculated by multiplying the number of answers in a category by its rating value (1, 2, 3, 4), obtaining a sum and dividing by the total number of answers for that item. To calculate the mean for the first item in the example above, follow these steps:

1. Multiply answers by value.
 Poor = 0 (0 x 1)
 Fair = 20 (10 x 2)
 Good = 180 (60 x 3)
 Excellent = 120 (30 x 4)
2. Sum. 0 + 20 + 180 + 120 = 320
3. Divide by N: 320 ÷ 100 = 3.2 (mean rating)

A summary of the calculations might look like the following:

Table 6. Program usefulness

	Mean rating
Practical information obtained	3.2
Increased understanding of subject	2.9
Stimulated interest in subject	3.3

1-4 scale where 1 = poor to 4 = excellent

A disadvantage of the mean is that it gives undue value to figures at one end or the other of the distribution. For example, if we were to report the average membership for 6 Tobacco control coalitions, with coalition memberships of 5, 9, 9, 11, 13 and 37, the average would be 14. Yet 14 is larger than all but one of the individual club memberships.

Mode

The Mode is the most commonly occurring answer of value in your data. For example, if cities report that they have adopted 3 anti-smoking ordinances more often than they report any other number of anti-smoking ordinances adopted, then 3 is the modal size of adopted ordinances in this study area. The mode is usually what people refer to when they say "the typical." It is the most frequent response or situation found in an evaluation.

The mode is important only when there is a large number of values. It is not as affected by extreme values as the mean.

Median

The median is the middle value. It is the midpoint where half of the cases fall below and half fall above the value. Sometimes we may want to know the midpoint value in our findings, or we may want to divide a group of participants into upper and lower groupings.

To calculate the median, arrange the data from one extreme to the other. Proceed to count halfway through the list of numbers to find the median value. When two numbers tie for the halfway point, take the two middle numbers, add them and divide by 2 to get the median. Like the mode, an advantage of the median is that it is not affected by extreme values or a range in data.

The following example shows the three measures of central tendency. In this example, we are analyzing the number of *tobacco ads below 3 feet in height* in ten stores.

Table 7. The number of *ads below 3 feet in height* in ten store.

Store	Number of ads	
1	3	
2	3	
3	3	Mean = 7.2 (# of ads divided by # of stores)
4	5	
5	6	Mode = 3 (most often occurring response)
6	6	
7	8	Median = 6.5 (mid-point)
8	9	
9	10	
10	19	
Total	72	

Which of the reported calculations make more sense – the mean, the mode, or the median? The answer will depend upon your data and the purpose of your analysis. Often, it is better to calculate all the measures and then decide which provides most meaning. In the example of table 7 the most common answer, the mode, might not be as useful to report as the mean or the median because it does not reflect the reality that most stores have more than 3 signs. The mode or median here give a

better understanding of the situation. If however seven out of the 10 answers had been 3, then it would make sense to report that number because it would be truly "typical," while the remaining 3 values would be out of the ordinary.

Measure of variability

Measures of variability express the spread or variation in responses. As indicated earlier, the mean may mask important differences, or be skewed by extreme values at either end of the distribution. For example, one high value can make the mean artificially high, or one extremely low response will result in overall low mean.

Looking at variability often provides a better understanding of our results. Are all the respondents and responses similar to the mean? Are some very high or very low? Did a few do a lot better than the others? Several measures help describe the variation we might find in our evaluation results.

Range

The range is the simplest measure of variability. It compares the highest and lowest value to indicate the spread of responses or scores. It is often used in conjunction with the mean to show the range of values represented in the single mean score. For example, "Stores displayed an average of 7.5 tobacco advertisement signs below three feet in their stores, ranging from 3 signs to 19 signs."

The range can be expressed in two ways: (1) by the highest and lowest values: "The number of signs ranged between 3 and 19," or (2) with a single number representing the difference between the highest and lowest number: "The range was 16 points."

While the range is a useful descriptor, it is not a full measure of variation. It only considers the highest and lowest scores, meaning that the other scores have no impact.

Standard deviation

The standard deviation measures the degree to which individual values vary from the mean. It is the average distance the scores lie from the mean. A high standard deviation means that the responses vary greatly from the mean. A low standard deviation indicates that the responses are similar to the mean. When all the answers are identical, the standard deviation is zero. The formula for calculating standard deviation is:

$$SD = \sqrt{\frac{\sum (X - \bar{X})^2}{n-1}}$$

If you calculate the standard deviation, chances are that you will have a computer program do this for you without having to apply the formula step-by-step. However, in order to demonstrate how the standard deviation is calculated, the following table shows the steps of the formula in the *number of store ads below three feet* example that was introduced earlier. For the formula we need the total number of ads (N=72) and the mean (7.2):

Table 8. Calculating standard deviation

Number of ads	Mean	Deviation from mean	Squared deviation
3	7.2	- 4.2	17.64
3	7.2	- 4.2	17.64
3	7.2	- 4.2	17.64
5	7.2	- 2.2	4.84
6	7.2	-1.2	1.44
6	7.2	-1.2	1.44
8	7.2	+ 0.8	0.64
9	7.2	+ 1.8	3.24
10	7.2	+ 2.8	7.84
19	7.2	+ 11.8	139.24
N= 72			Sum of squared deviation = 211.6

$$\begin{aligned}
 \text{Standard Deviation} &= \text{Square root}(\text{sum of squared deviations} / (n - 1)) \\
 &= \text{Square root}(211.6 / 71) \\
 &= \text{Square root}(2.98) \\
 &= 1.73
 \end{aligned}$$

Thus, the standard deviation from the mean (7.2) is 1.73, which means that the average difference between the mean and other values is 1.73.

Sometimes, instead of the standard deviation, the **variance** is used. It is simply the square of the standard deviation.

In some cases, variation in response represents a positive outcome. A program designed to help people think independently and to build their individual decision-making skills may reveal a variety of perspectives. In another case, if the goal of the program is to help everyone achieve a certain level of knowledge, skill or production, variation may indicate less than successful outcomes.

Creating ranks

Rankings

The analysis techniques discussed so far involve calculating numbers- using the actual data to provide measures of results. Rankings, on the other hand, are not actual measurements. They are *created* measures to impose sequencing and ordering. Rankings indicate where a value stands in relation to other values or where the value stands in relation to the total. For example:

- Lack of employee training was ranked as the most common reason for non-compliance with the STAKE Act.
- Moving billboards was ranked as the most effective anti-smoking advertisement at rodeo events.
- Utopia City ranked fourth in the number of anti-smoking policies adopted.

While rankings can be meaningful, there is the tendency to interpret rankings as measurements rather than as sequences. Also, only minimal differences may separate items that are ranked. These differences are concealed unless explained. When using rankings, it is best to clearly explain the meaning.

Working with the data

Begin to understand your data by looking at the summary of responses to each item. Are certain answers what you expect? Do some responses look too high or too low? Do the answers to some questions seem to link with responses to other items?

This is the time to *work with your data*. Look at the findings from different angles. Check for patterns. Begin to frame your data into charts, tables, lists and graphs to view the findings more clearly and from different perspectives. A good process is to summarize all your data into tables and charts and write from those summaries. See how the data look in different graphical displays. Think about which displays will most effectively communicate the key findings to others.

Cross-tabulations or subsorting will allow you to explore findings further. For example, suppose you are doing a follow-up evaluation of an annual three-day workshop attended by program participants. You've collected data from 301 participants and one of the items reflects the overall rating of the event. The categories are "Excellent, Good, Fair, Poor." First, you might take a look at the frequency distribution and calculate the average ratings:

Table 9. Overall rating of workshop (n= 301)

Rating	1 st time participants (n=200)	Repeat participants (n= 101)
Excellent (x4)	125	58
Good (x3)	60	35
Fair (x2)	15	6
Poor (x1)	0	2
Average rating*	3.6	3.5

*1-4 scale where 1 = poor and 4 = excellent

These results indicate that the repeat participants are almost equally satisfied with the program. You also might want to check for differences by sorting the respondents into categories (for example "program participants," "program staff," etc.) to see if one group rated the workshop differently than the other. The

possibilities for subsorting will depend upon what data you collected and the purpose of your evaluation.

Cross tabulations may be conveniently presented in contingency tables which display data from two or more variables. As illustrated in table 10, contingency tables are particularly useful when we want to show differences which exist among subgroups in the total population.

Table 10. Changes in participants' satisfaction level with trainings A and B (N=100 participants at each training).

Level of program satisfaction	2005		2006	
	Training A	Training B	Training A	Training B
High	10	20	50	69
Medium	57	62	48	31
Low	33	18	2	0

Summary

The possibilities for analyzing your evaluation data are many. Give priority to those analyses which most clearly help summarize the data relative to the evaluation's purpose and which will make the most sense to your audience.

Acknowledgements: Revised from Ellen Taylor-Powell, Cooperative Extension at the University of Wisconsin-Extension. "Program Development and Evaluation, Analyzing Quantitative Data," Cooperative Extension Publications, Room 170, 630 W. Mifflin Street, Madison, WI 53703, 1996. Retrieved from the web at:
<http://learningstore.uwex.edu/pdf/G3658-6.pdf>

Originally published in 1989 by the Texas Agricultural Extension Service based on material from Oregon State University Extension Service (Sawer, 1984), Virginia Cooperative Extension Service and Kansas State Cooperative Extension Service and with the help of members of the TAEX Program and Staff Development Unit including Howard Ladewig, Mary Marshall and Burl Richardson.

For more resources, visit our website:
<http://programeval.ucdavis.edu>