**Program Development
and Evaluation**

# Analyzing Quantitative Data

**Ellen Taylor-Powell**

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. For example:

- Thirty-two of the participants were over 55 years of age.
- Twenty-seven of the 30 participants rated the content of the Extension newsletter as very useful in helping deal with family communication problems.
- A total of 330 producers took soil samples in 1992.

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. Calculating percentages for the examples above we find:

- Eighty-five percent of the participants were over 55 years of age.
- Ninety percent of the participants rated the content very useful.
- Seventy-four percent of the producers in the county took soil samples in 1992.

Percentages tend to be easy to interpret. For example, it is more understandable to say that 40 percent of the respondents use chlorination to correct water quality problems than to say that 96 of 240 people chlorinate their water.

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

- Fifty-nine percent of producers in 1983 were vaccinating their calves for brucellosis as compared to 1987 when 76 percent reported brucellosis vaccination. (Comparing 1983 respondents to 1987 respondents).
- While 76 percent of the producers vaccinated their calves for brucellosis, only 27 percent used a vet to plan their herd health management. (Comparing responses from the same respondents.)

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 order of size or magnitude. The following table provides an example.

---

[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**.*

**Table 1.** Frequency distribution of Extension participants by place of residence (n = 860)

| Place of residence | Frequency | Percentage |
|---|---|---|
| Rural farm | 269 | 31.3 |
| Rural nonfarm | 288 | 33.5 |
| Small town | 303 | 35.2 |

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/subgrouping).

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 to 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.

## 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 sources of information they use in corn production, the respondents might use one or several of the possible answers: fertilizer dealer, consultant, printed materials, county agent, 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.** Projects completed by 4-H members by district (n = 165,000 4-H members)

| District | Projects completed | Percent |
|---|---|---|
| A | 31,000 | 74 |
| B | 8,000 | 60 |
| C | 12,000 | 65 |
| D | 26,000 | 75 |
| E | 11,000 | 50 |
| F | 28,000 | 72 |
| TOTAL | 116,000 | Mean 66 (incorrect) |

In this table, the evaluator incorrectly reported an average of the six district percentages. Rather, the percentage should have been calculated by dividing the total number of completed projects (116,000) by the total state membership (165,000) for a total mean percentage of 70 percent. Sometimes, as in this example, the differences are not great. In 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 average acreage for a population of farmers, divide the total number of acres reported by the total number of respondents. To report mean income of program participants, divide the total income reported by the total number of responding participants.

The mean is also useful for summarizing findings from rating scales. Even with narrative scales, we can assign a number value to each category and thus calculate the mean. For example, "not important, slightly important, fairly important, very important" can be assigned 1, 2, 3, 4. "Almost never, sometimes, frequently" could become 1, 2, 3. Consider the information in table 5 from a question asking respondents to rate the usefulness of an Extension program.

**Table 5.** Program usefulness (n = 100 participants)

| | Poor | Fair | Good | Excellent | N |
|---|---|---|---|---|---|
| A Gave me practical information I can use at work | 1 | 2 (10 answers) | 3 (60 answers) | 4 (30 answers) | 100 |
| B. Increased my understanding of the subject | 1 | 2 (20 answers) | 3 (70 answers) | 4 (10 answers) | 100 |
| C. Stimulated me to find out more about the subject | 1 | 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 clubs in the district, with club 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 or value. For example, if farmers report the size of their farm as 120 acres more often than any other size, then 120 is the modal size of a farm in the 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 the evaluation.

The mode is important only when a large number of values is available. 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 income data from 10 families.

**Table 7.** Income data

| Family | Family income (1991) |          |
| ------ | -------------------- | -------- |
| 1      | $ 25,000             |          |
| 2      | 30,000               | **Mean** |
| 3      | 30,000               | $55,000  |
| 4      | 30,000               | **Mode** |
| 5      | 30,000               | $30,000  |
| 6      | 40,000               | **Median** |
| 7      | 40,000               | $35,000  |
| 8      | 45,000               |          |
| 9      | 50,000               |          |
| 10     | 230,000              |          |
| TOTAL  | $550,000             |          |

Which of the reported calculations makes 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 the most meaning. Is it more useful and important to know that the average income for these ten families is $55,000? Or, is the most common income more meaningful? Sometimes, it will be important to look at variability, not just the central tendencies.

# Measures 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 an 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, "Soil testing for phosphorus saved producers an average $15/acre, ranging from $12 to $20/acre."

The range can be expressed in two ways: (1) by the highest and lowest values: "The scores ranged from 5 to 20"; or (2) with a single number representing the difference between the highest and lowest score: "The range was 15 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 average score lies 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.

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

## Variance

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

In some cases, variation in responses 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:

■ Producers ranked the county Extension agent as the most important source of information.

■ Homemakers in Oneida County ranked second in the state in adoption of energy saving practices.

■ The Crawford County 4-H club ranked 5th in the overall state competition.

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 graphic displays. Think about which displays will most effectively communicate the key findings to others.

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

**Table 8.** Overall rating of workshop (n = 301)

| Rating | 1st 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 occupational categories to see if one group rated the program 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 9, contingency tables are particularly useful when we want to show differences which exist among subgroups in the total population.

**Table 9.** Changes in young people's self-confidence levels (N = 100 girls, 100 boys)

| Level of confidence | 1990 | | 1991 | |
|---|---|---|---|---|
| | Girls % | Boys % | Girls % | Boys % |
| 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.

**Author:** Ellen Taylor-Powell is a program development and evaluation specialist for Cooperative Extension, University of Wisconsin–Extension.