FS-25-32

Choosing the Right Statistical Test for Your Data: A Road Map for Extension Professionals

Najat Elgeberi, Ph.D.

Assistant Professor for Program Evaluation

A partnership of Nevada counties; University of Nevada, Reno; and the U.S. Department of Agriculture

Introduction

This fact sheet is designed to be a practical step-by-step guide for Extension professionals to help them navigate the process of statistical analysis for any type of data that derives from any type of surveys. The paper will cover some theoretical foundations, while mostly focusing on practical application of the right steps to conduct an analysis. The goal is to empower you to move from raw data to meaningful insights that can inform your programming, demonstrate your impact, and strengthen your accountability to stakeholders. As Majumdar et al. (2020) suggest, a practical evaluation tool and framework is essential for Extension educators to "determine program success and show public value" (Majumdar et al., 2020). This road map will walk you through a systematic process, from preparing your data to interpreting and presenting your results. Each step includes practical tips, checklists and references to the underlying statistical principles. By following this road map, you can ensure that your data analysis is not only statistically sound, but also a powerful tool for telling the story of your program's success.

The Critical Role of Statistical Literacy in Extension Program Evaluation

In the field of Extension, professionals are constantly engaged in a cycle of program development, implementation and evaluation. The data collected through these efforts are the lifeblood of evidence-based practice, providing the necessary insights to assess program effectiveness, demonstrate impact to stakeholders, and make informed decisions about resource allocation. However, the value of this data is entirely dependent on the appropriateness of the statistical methods used to analyze it. The selection of an incorrect statistical test can lead to a cascade of negative consequences, including invalid conclusions, misallocation of resources, and a loss of credibility with funders and the community. As Taylor (2008) notes, statistical analysis of the results yields valuable information that can be used to improve programs and to demonstrate their impact to stakeholders (Taylor, 2008).

This fact sheet serves as a practical guide for Extension professionals to navigate the often-complex landscape of statistical test selection. It provides a comprehensive overview of data types, a decision-making framework for choosing the right statistical test, and best practices for visualizing data to communicate program impact effectively. By strengthening statistical literacy, Extension professionals can ensure their evaluation efforts are not only rigorous and defensible, but also a powerful tool for driving positive changes in the communities they serve.

Step 1: Preparing Your Data for Analysis

Data cleaning, organizing and structuring are the first and most important steps in our road map towards data analysis. This is often the most time-consuming but critical part of the data analysis process to ensure the accuracy and validity of your results.

Think of your data as your child's closet; you have so many items in there that you bought or collected, and you need to organize them to make sense of what items are there and what items are missing. The first step is to carefully review your data for any data entry errors, typos or inconsistencies. Similar to your child's closet, look into all the items there and make sure that things that are out of place be corrected and placed where they do belong. Look for out-of-range values (e.g., an age of 200) or nonsensical responses.

The second step is to decide on a strategy for handling missing data. You might choose to exclude cases with missing data. So, in our closet example you might decide to throw a pajama top away because the bottom part is lost or damaged. In your data, you might use a statistical technique to impute (i.e., fill in) the missing values (fix the damaged pants and keep them for use). The best approach will depend on the amount of missing data and the nature of your research question.

The third step after cleaning is defining and labeling the variables. After organizing your child's t-shirts, make sure to put them in one drawer and put a label on it to show you what items are in the drawer. Ensure that all of your variables are clearly defined and labeled. For categorical variables, make sure that the categories are consistently coded (e.g., 1 = Yes, 2 = No). Your final data table should reflect each respondent in one row and each variable in one column.

Depending on the software you have or the level of statistical skills you have, there are different statistical software that can be used to perform cleaning:

1) If using Excel, use the "Data Validation" feature to prevent data entry errors. The "Find and Replace" tool can be useful for correcting inconsistencies.

2) If using SPSS, the "Data" menu provides a range of tools for data cleaning and preparation, including the ability to define variable properties, identify duplicate cases, and handle missing values.

If this step is difficult for you to perform, reach out to your evaluation specialist or anyone in your institute with statistical skills who can help you with this task.

Step 2: Choosing the Right Statistical Test

With your data cleaned and prepared, the next step is to select the appropriate statistical test. Let's think about the statistical tests as a big wardrobe, and we are on a mission to select an outfit that is appropriate to a specific occasion. It is highly important to find the right outfit. Although picking a T-shirt and a pair of jeans will appear to be comfortable to attend a wedding party, it is not the right choice. The same thing can happen when you pick the wrong statistical test for your data. It might be performed using statistical software, but the given results are wrong and deceiving. Selecting the right test should be guided by your research question and the level of measurement of your variables. The following questions can be used to guide you in determining your test:

1) What is your primary research question?

   - Are you trying to compare groups?
   - Are you trying to examine the relationship among variables?
   - Are you trying to predict an outcome?

2) What is the level of measurement of your variables?

   - Are they nominal, ordinal, interval or ratio?

3) If your data is interval or ratio, are they normally distributed?

- Use the normality testing protocol outlined in the fact sheet to make this determination.

By answering these three questions, you can use the decision tree and matrix provided later in this fact sheet to identify the most appropriate statistical test for your analysis.

Step 3: Understanding Your Data and The Four Levels of Measurement

The foundation of sound statistical analysis lies in correctly identifying the type of data you are working with. The level of measurement, including nominal, ordinal, interval and ratio (Scribbr, 2020), not only determines the descriptive statistics that can be used, but also dictates the appropriate inferential statistical tests for hypothesis testing. Each level has unique characteristics and a hierarchical relationship, with each level building upon the properties of the one before it. Descriptive statistics are methods used to summarize, organize and display data so that patterns and key characteristics of a dataset can be clearly understood (Gravetter & Wallnau, 2017). Inferential statistics involve techniques that allow researchers to use data from a sample to make estimates, predictions or generalizations about a larger population, typically through hypothesis testing and probability-based methods (Field, 2018).

3.1. Nominal Data: Categories Without Order

Nominal data consists of categories or labels used to classify observations into mutually exclusive groups. There is no inherent order or ranking among the categories. Examples in an Extension context include, (1) Gender such as male, female, non binary, (2) Farming type such as organic, conventional, mixed, (3) Program participation with yes, no answers, (4) County of residence with answer options, Washoe, Clark, Elko

For nominal data, the only valid measure of central tendency is the **mode**, which represents the most frequently occurring category. Descriptive statistics are limited to frequencies, percentages and proportions. The appropriate statistical tests for nominal data are nonparametric and include:

1) **Chi-square test of independence:** Used to determine if there is a significant association between two nominal variables.
2) **Fisher's exact test:** An alternative to the Chi-Square test for small sample sizes where expected frequencies (or count) are less than five for any of the categories.
3) **McNemar's test:** Used for analyzing paired nominal data, such as in before-and-after studies.

3.2. Ordinal Data: Categories With Meaningful Order

Ordinal data represent categories that have a meaningful order or ranking, but the intervals between the ranks are not necessarily equal or known. This is a common data type in Extension surveys, particularly those using Likert-type scales. Examples include, (1) Satisfaction ratings with answer options: poor, fair, good, excellent, (2) Likert-type scales: strongly disagree, disagree, neutral, agree, strongly agree, and (3) Education level: high school, some college, bachelor's degree, graduate degree.

For ordinal data, the median is the most appropriate measure of central tendency, as it represents the middle value in the ordered set. The interquartile range (IQR) is a useful measure of dispersion. Because the intervals between categories are not equal, nonparametric tests are generally recommended:

1) **Mann-Whitney U test:** Compares two independent groups.
2) **Wilcoxon signed-rank test:** Compares two related groups or paired measurements.
3) **Kruskal-Wallis test:** Compares three or more independent groups.
4) **Spearman's rank correlation:** Measures the strength and direction of the association between two ordinal variables.

3.3. Interval Data: Equal Intervals, Arbitrary Zero

Interval data have all the properties of ordinal data, but with the additional feature of equal intervals between consecutive values. This means that the difference between two values is meaningful. However, interval scales have an arbitrary zero point, which means that ratios are not meaningful. Examples include, (1) Temperature: Celsius or Fahrenheit scales (zero degrees doesn't mean it's not cold), (2) Standardized test scores: Such as pre- and post-workshop knowledge tests (zero test score doesn't mean the participant has zero knowledge), (3) Calendar Dates: the interval between years is consistent.

For interval data, the mean is the most common measure of central tendency, and the standard **deviation** is the primary measure of dispersion. Parametric tests are appropriate for interval data, provided that the assumption of normality is met:

1) **t-tests (independent and paired):** Compare the means of two groups.
2) **Analysis of variance (ANOVA):** Compares the means of three or more groups.
3) **Pearson correlation:** Measures the linear relationship between two interval variables.

3.4. Ratio Data: Equal Intervals, True Zero

Ratio data represent the highest level of measurement. They have all the properties of interval data, but with the addition of a true zero point. A true zero indicates the complete absence of the attribute being measured, which allows for meaningful ratios to be calculated. Examples in Extension include, (1) Age in years, (2) Income in dollars, (3) Farm size in acres, (4) Crop yield in bushels per acre, (5) Work experience.

All descriptive and inferential statistics can be used with ratio data. The choice of statistical test depends on the research question and the distribution of the data. In addition to the parametric tests used for interval data, more advanced statistical methods can be applied:

1) **Regression analysis:** To predict the value of one variable based on the value of another.
2) **Factor analysis:** To identify underlying dimensions in a set of variables.
3) **Time series analysis:** To analyze data collected over time.

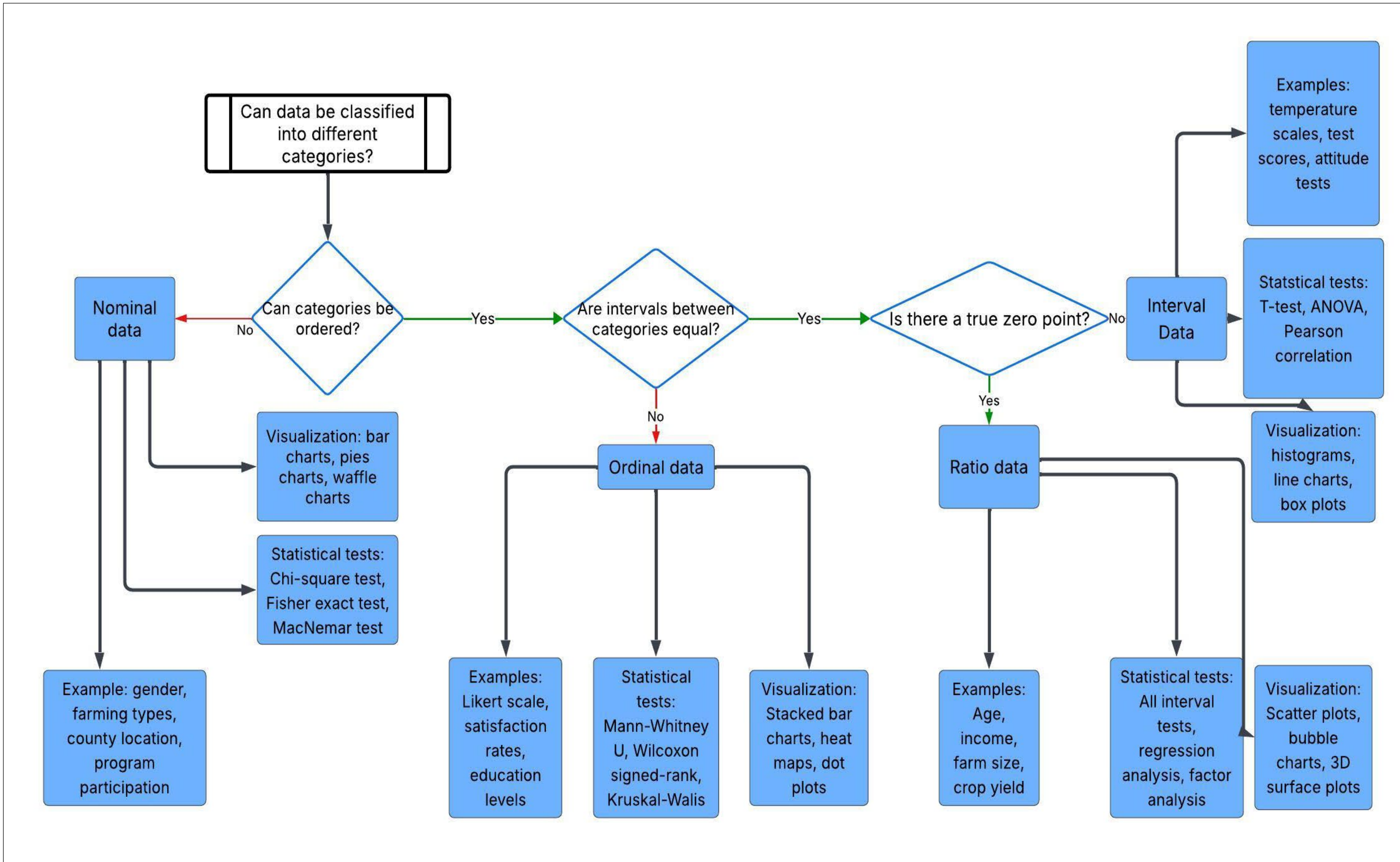

Figure 1. Data Type Decision Tree

Figure 2. Statistical Test Selection Matrix

Step 4: The Importance of Normality Testing
Many of the most powerful and commonly used statistical tests, known as parametric tests, are based on the assumption that the data are normally distributed. The normal distribution, often

referred to as the "bell curve," is a symmetrical distribution where most values cluster around the mean (about 68% of all values fall within 1 standard deviation of the mean, about 95% of all values fall within 2 standard deviations, and about 99.7% of all values fall within 3 standard deviations).

When this assumption is violated, the results of parametric tests can be misleading, potentially leading to incorrect conclusions about program impact or association between the variables.

4.1. How to Test for Normality

A systematic approach to normality testing involves both visual inspection and formal statistical tests. As Mishra et al. (2019) recommend, this dual approach provides a more complete picture of the data's distribution (Mishra et al., 2019).

4.1.1: Visual Inspection

1) **Histograms:** A histogram provides a graphical representation of the distribution of the data. A normally distributed dataset will have a histogram that is roughly symmetrical and bell-shaped.
2) **Q-Q (Quantile-Quantile) plots:** A Q-Q plot compares the quantiles of your data to the quantiles of a normal distribution. If the data are normally distributed, the points on the plot will fall along a straight diagonal line. (Quantiles split your data into equally sized portions so you can see how the values are spread out. For example, quartiles divide data into four equal parts, percentiles divide data into 100 equal parts, deciles divide data into 10 equal parts).

4.1.2: Formal Statistical Tests

Several statistical tests can be used to formally assess the normality of a dataset. The choice of test often depends on the sample size.

1) **Shapiro-Wilk test:** This test is widely regarded as one of the most powerful normality tests, particularly for small sample sizes ($n < 50$) (Razali & Wah, 2011).
2) **Kolmogorov-Smirnov test:** This test is more suitable for larger sample sizes ($n \geq 50$). However, some researchers caution against its use, as it can be less powerful than other tests (Steinskog et al., 2007).
3) **Anderson-Darling test:** Another powerful test that is sensitive to deviations in the tails of the distribution. To simplify things, imagine you're checking whether students' test scores fit a typical bell curve. A test not sensitive to tails mostly looks at "average" students. The Anderson-Darling test looks carefully at the top scorers and bottom scorers too. If there are too many very high or very low scores, it will flag that as a problem.

For these tests, the null hypothesis is that the data are normally distributed. (The null hypothesis is the default assumption or the thing you test against to see if the data provides strong evidence that something actually changed or is different). Therefore, a p-value greater than 0.05 indicates that the data are likely normally distributed, and the null hypothesis should not be rejected. A p-value that is greater than 0.05 means that an obtained result is probably real, not just luck or random chance. When we run a statistical test, we are basically asking if the effect is strong enough that it's unlikely to have happened just by accident. If the answer is yes, we say the result is statistically significant.

4.2. What to Do if Your Data Are Not Normal

If both visual inspection and formal tests indicate that your data are not normally distributed, you have two primary options:

1) **Use Nonparametric Tests:** A more straightforward approach is to use nonparametric tests. These tests do not assume a normal distribution and are therefore more robust when the assumptions of parametric tests are violated. For every parametric test, there is a corresponding nonparametric alternative. Parametric tests are tests that assume your data follows a certain shape, usually a normal distribution, while nonparametric tests do not assume your data follows a specific distribution and work even when your data is messy or doesn't follow normal rules.
2) **Data transformation:** In some cases, you can apply a mathematical function (e.g., logarithmic, square root) to the data to make it more closely approximate a normal distribution. However, this can make the interpretation of the results more complex. If you plan to go this route, you need to consult with a statistical expert for assistance.

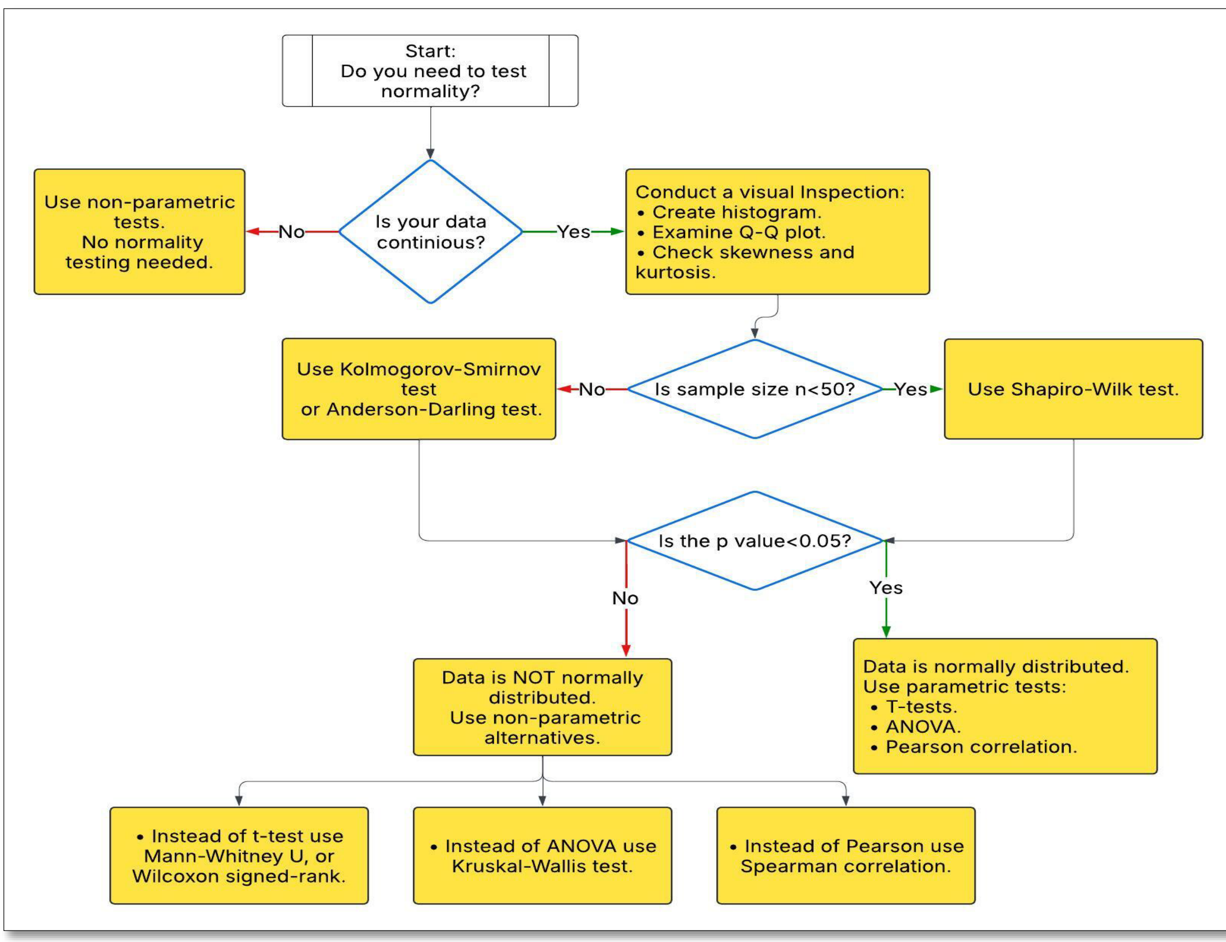

Figure 3. Normality Testing Flowchart

Step 5: Other Scenarios for Statistical Analysis in Extension

5.1. Comparing Groups

A common goal in Extension evaluation is to compare outcomes between two or more groups. For example, you might want to compare the knowledge gain of participants in an online workshop to those in an in-person workshop.

1) **For two independent groups:**
   - **Parametric:** Independent t-test (for interval/ratio data that are normally distributed).
   - **Nonparametric:** Mann-Whitney U test (for ordinal data or interval/ratio data that are not normally distributed).
2) **For three or more independent groups:**
   - **Parametric:** One-way analysis of variance (ANOVA) (for interval/ratio data that are normally distributed).
   - **Nonparametric:** Kruskal-Wallis test (for ordinal data or interval/ratio data that are not normally distributed).

5.2. Comparing Before and After

Another common scenario is to compare the same group of participants before and after a program to assess change over time. For example, you might measure participants' confidence in a particular skill before and after a training program.

1) **Parametric:** Paired t-test (for interval/ratio data that are normally distributed).
2) **Nonparametric:** Wilcoxon signed-rank test (for ordinal data or interval/ratio data that are not normally distributed).

5.3. Examining Relationships

Sometimes, the goal is to understand the relationship or association between two variables. For example, you might want to know if there is a relationship between the number of hours a farmer spends in training and their subsequent crop yield.

1) **For nominal variables:** Chi-square test of independence.
2) **For ordinal variables:** Spearman's rank correlation.
3) **For interval/ratio variables:**
   - **Parametric:** Pearson correlation (if the relationship is linear and the data are normally distributed which means as one variable changes, the other changes in a straight-line pattern)
   - **Non-parametric:** Spearman's rank correlation (if the assumptions for Pearson correlation are not met).

Examples:

The following scenarios are demonstrated.

1) Let's assume that you want to know if there is a difference in the average crop yield (ratio data) between farmers who participated in a new irrigation program and those who did not. You will need to test for normality. If the data are normally distributed, use an independent t-test. If not, use the Mann-Whitney U test.

2) You want to know if there is an association between a 4-H member's county of residence (nominal data) and their chosen project area (nominal data). you will use a Chi-square test of independence.

3) You want to assess the change in participants' self-reported confidence (ordinal data) in managing their finances before and after a financial literacy workshop. In this case, the use of a Wilcoxon signed-rank test is the most appropriate.

Step 6: Visualizing Your Data: Telling a Compelling Story

Effective data visualization is not about creating pretty charts; it is about telling a clear and compelling story with your data. The right visualization can make complex information accessible and understandable to a wide range of audiences, from program participants to county commissioners. The choice of visualization should be guided by the type of data you have and the message you want to convey.

6.1. Visualizing Nominal Data

1) **Bar charts:** Ideal for comparing the frequencies of different categories.
2) **Pie charts:** Best for showing the proportional composition of a whole but should be used with caution and limited to a small number of categories (≤5).
3) **Waffle charts and treemaps:** Excellent alternatives to pie charts for showing part-to-whole relationships, especially for public-facing presentations.

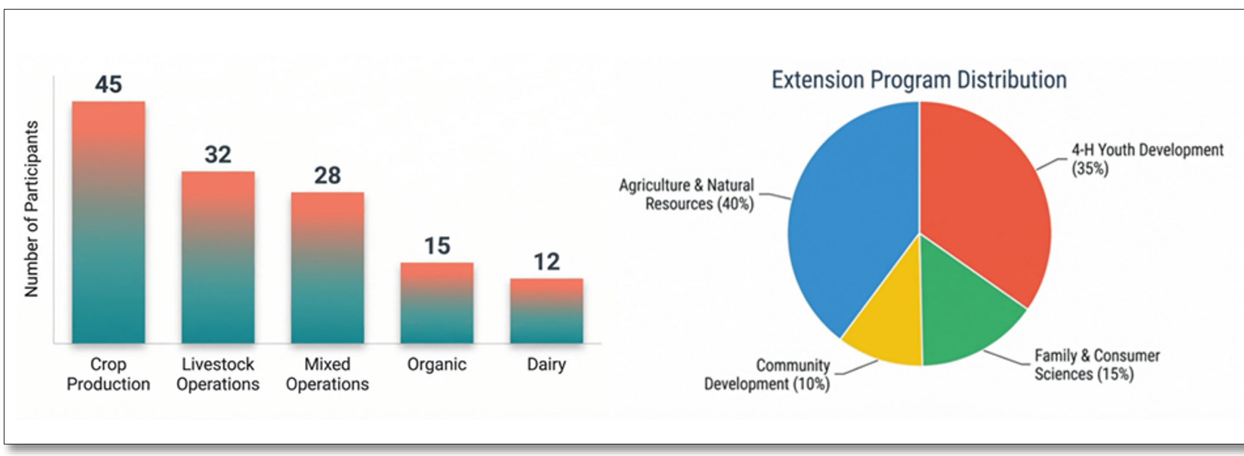

Figure 4. Nominal Data Visualization

6.2. Visualizing Ordinal Data

1) **Stacked bar charts:** Effective for showing the distribution of responses across ordered categories.
2) **Diverging stacked bar charts:** Particularly useful for visualizing Likert scale data with a neutral midpoint.
3) **Heat maps:** Can be used to show the intensity of responses across multiple ordered categories.

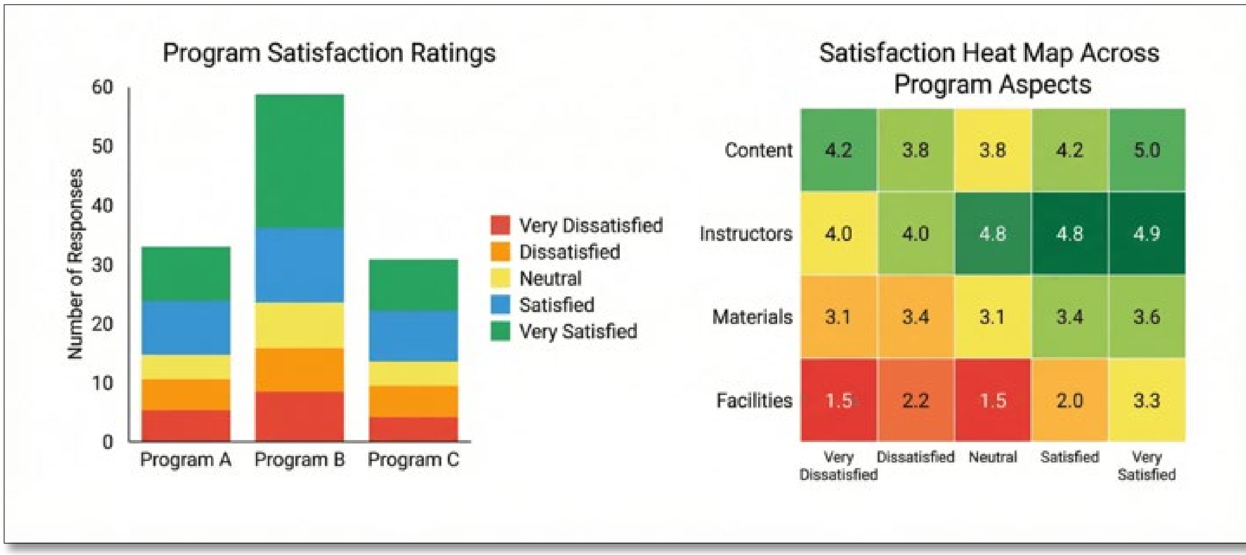

Figure 5. Ordinal Data Visualization

6.3. Visualizing Interval and Ratio Data

1) **Histograms:** The primary tool for visualizing the distribution of continuous data and assessing normality.
2) **Box plots:** Useful for comparing the distributions of two or more groups.
3) **Line charts:** Ideal for showing trends over time.
4) **Scatter plots:** The standard for visualizing the relationship between two continuous variables.
5) **Bubble charts:** An extension of the scatter plot that allows for the visualization of a third variable through the size of the bubbles.

As Franconeri et al. (2021) emphasize, the goal of data visualization is to leverage human perceptual and cognitive processes to help people make sense of data (Franconeri et al., 2021). By choosing the right chart for your data and your story, you can transform your evaluation results from a set of numbers into a powerful tool for communication and impact.

Conclusion: From Data to Decisions

The journey from data collection to data-driven decision-making is paved with critical choices. For Extension professionals, the selection of the right statistical test is a cornerstone of this process. By understanding the different levels of measurement, diligently testing for normality, and thoughtfully selecting the appropriate statistical test, you can ensure that your evaluation findings are not only statistically valid, but also a true reflection of your program's impact. This fact sheet has provided a road map to guide you through this process, empowering you to transform your data into a powerful tool for program improvement, stakeholder engagement and community betterment.

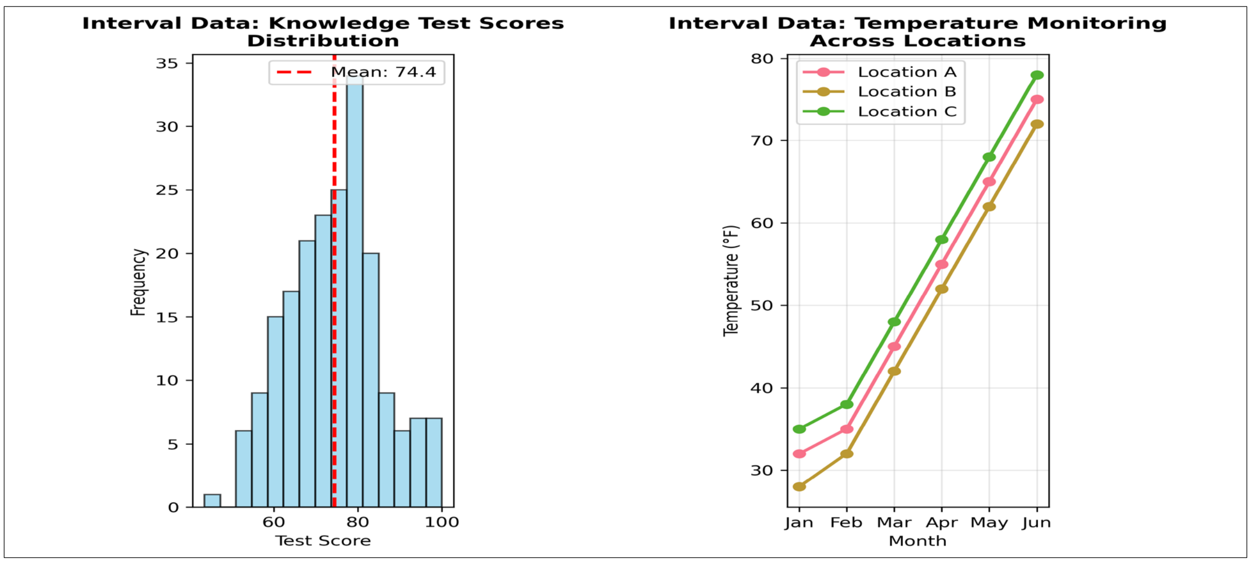

Figure 6. Interval Data Visualization

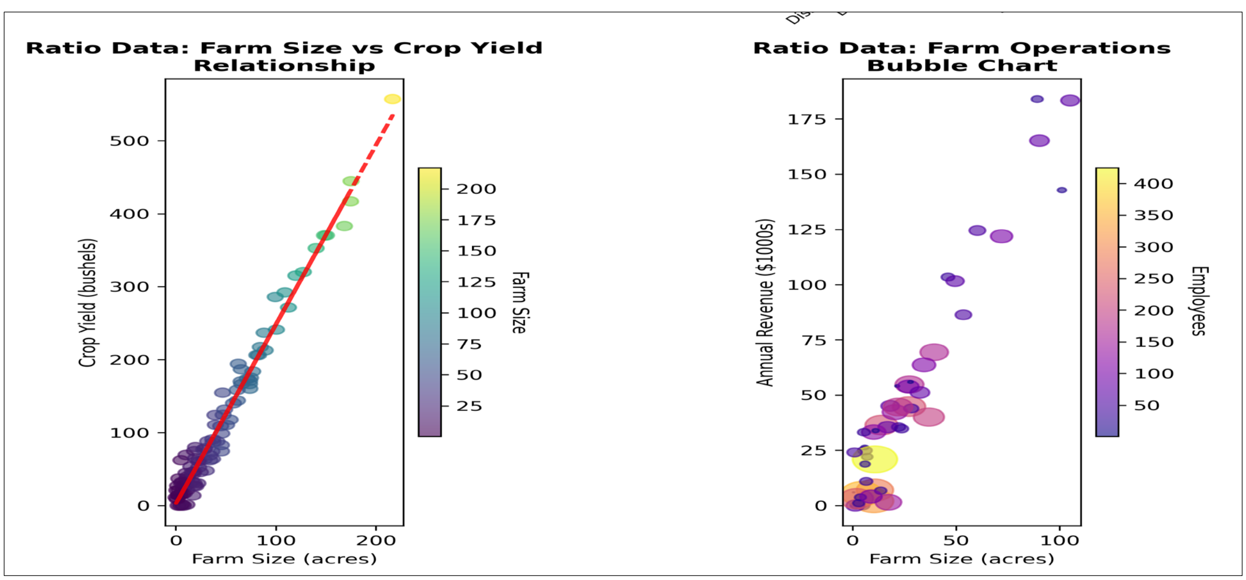

Figure 7. Ratio Data Visualization

References

Field, A. P. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). SAGE Publications.

Franconeri, S. L., Padilla, L. M., Shah, P., Zacks, J. M., & Hullman, J. (2021). The science of visual data communication: What works. *Psychological Science in the Public Interest, 22*(3), 110–161. https://journals.sagepub.com/doi/full/10.1177/15291006211051956

Gravetter, F. J., & Wallnau, L. B. (2017). *Statistics for the behavioral sciences* (10th ed.). Cengage Learning.

Majumdar, A., Chantre, A., Fadamiro, H., & Balusu, R. R. (2020, June). Practical evaluation tool and logic model framework for extension educators and grant writers. *Journal of the NACAA, 13*(1), 1. https://www.nacaa.com/journal/index.php?jid=1096

Mishra, P., Pandey, C. M., Singh, U., Gupta, A., Sahu, C., & Keshri, A. (2019). Descriptive statistics and normality tests for statistical data. *Annals of Cardiac Anaesthesia, 22*(1), 67–72. https://pmc.ncbi.nlm.nih.gov/articles/PMC6350423/

Razali, N. M., & Wah, Y. B. (2011). Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling and Analytics, 2*(1), 21–33.

Scribbr. (2020, July 16). *Levels of measurement | Nominal, ordinal, interval and ratio.* https://www.scribbr.com/statistics/levels-of-measurement/

Steinskog, D. J., Tjøstheim, D. B., & Kvamstø, N. G. (2007). A cautionary note on the use of the Kolmogorov–Smirnov test for normality. *Monthly Weather Review, 135*(3), 1151–1157.

Taylor, G. D. (2008). Better extension programming through statistics: Using statistical analysis for program evaluation. *Journal of Extension, 46*(5), 14. https://open.clemson.edu/joe/vol46/iss5/14/