Quantitative data analysis is helpful in evaluation because it provides measurable and easy to understand results. Quantitative data can be analyzed in different ways. Below is the most common way to approach quantitative data analysis in small program evaluation.

Step 1: Identify the level of measurement

Choosing which of the four levels of measurement to use can influence the type of analysis

Nominal data – data has no logical order; data is basic classification data	• Example: Male or Female
	  ○ There is no order associated with male nor female
	  ○ Each category is assigned an arbitrary value (male = 0, female = 1)

Ordinal data – data has a logical order, but the differences between values are not constant	• Example: T-shirt size (small, medium, large) • Example: Military rank (from Private to General)

Interval data – data is continuous and has a logical order; data has standardized differences between values, but no natural zero	• Example: Fahrenheit degrees
	  ○ Ratios are meaningless for interval data. It can't be said, for example, that one day is twice as hot as another day.
	• Example: Items measured on a Likert scale – satisfaction ranked on a scale of 1-5.
	  ○ *1 = Very Dissatisfied*
	  ○ *2 = Dissatisfied*
	  ○ *3 = Neutral*
	  ○ *4 = Satisfied*
	  ○ *5 = Very satisfied*

Ratio data – data is continuous, ordered, has standardized differences between values, and a natural zero	• Example: height, weight, age, length
	• Having an absolute zero means that it can be said that one measure is twice as long as another.
	  ○ For example – 10 inches is twice as long as 5 inches
	  ○ This ratio holds true regardless of which scale the object is being measured in (e.g. meters or yards).

Step 2: Applying Data Analysis Procedures

Once the levels of measurement have been identified, the next step is applying some of the quantitative data analysis procedures outlined below. The procedures available for use are limited by sample size restrictions. However, there are several procedures that can be used to determine what narrative the data is telling.

These procedures include:

- Data tabulation (frequency distributions & percent distributions)
- Descriptives data
- Data disaggregation
- Moderate and advanced analytical methods

Data Tabulation

This procedure involves organizing results for the different variables within the data set. This gives a comprehensive picture of the data and helps to identify patterns. The best way to do this is to construct frequency and percent distributions of those satisfied with the experience.

A frequency distribution is an organized tabulation of the number of individuals or scores located in each category (see the table below).

- This will help determine:
    - If scores are entered correctly
    - If scores are high or low
    - How many are in each category
    - The spread of the scores

As an example, the following table shows that 15 of the clients surveyed who participated in the program reported being satisfied with the experience.

Total sample size=22

Gender	Female=11 Male=11
Ethnicity	Black=7 White=7 Asian=2 South Asian=2 Latino=4

Municipality	Brampton=11 Mississauga=11
Participated in Program	Yes=20 No=2
Satisfied with Program Experience (valid n=20)	Very Dissatisfied=2 Dissatisfied=3 Satisfied=12 Very Satisfied=3 Not Applicable=2

A percent distribution displays the proportion of participants who are represented within each category. In this example, 75% of clients (n = 20) surveyed reported being satisfied with the experience.
Total sample size=22

Gender	Female=11 (50%) Male=11 (50%)
Ethnicity	Black=7 (32%) White=7 (32%) Asian=2 (9%) South Asian=2 (9%) Latino=4 (18%)
Municipality	Brampton=11 (50%) Mississauga=11 (50%)
Participated in Program	Yes=20 (90%) No=2 (9%)
Satisfied with Program Experience (valid n=20)	Very Dissatisfied=2 (10%) Dissatisfied=3 (15%) Satisfied=12 (60%) Very Satisfied=3 (15%)

Descriptives Data

A descriptive refers to calculations that are used to explain the data set. The most common descriptives used are:

- **Mean** – the numerical average of scores for a particular variable
- **Minimum and maximum** values – the highest and lowest value for a particular variable
- **Median** – the numerical middle point or score that cuts the distribution in half for a particular variable
    - Calculate by:
        - Listing the scores in order and counting the number of scores
        - If the number of scores is odd, the median is the number that splits the distribution
        - If the number of scores is even, calculate the mean of the middle two scores
- **Mode** – the most common number score or value for a particular variable

Depending on the level of measurement, descriptives may not be possible for all variables in a dataset.

- A meaningful mean can only be calculated from interval and ratio data.
- Minimum and maximum values can be calculated for all levels of measurement
- A meaningful median can only be calculated from ordinal, interval, and ratio data
- The mode descriptive can be calculated for all levels of measurement

In the table, the average satisfaction level of the clients surveyed who participated in the program (n =20) was 2.8, with a range of 1 = Very Dissatisfied to 4 = Very Satisfied. The mode (most commonly occurring value) is 3, a report of satisfaction.

Total Valid Sample Size=20

Satisfied with Program Experience *(valid n=20)*	Mean=2.8 Maximum=4 Minimum=1 Mode=3 Median=3

Disaggregating the Data

After tabulating the data, it can be explored further by disaggregating it across different variables and subcategories of variables. Crosstabs allow data to be disaggregated across multiple categories.

The table below explores participant demographics (gender and ethnicity) within each program city. Presented in this way, it can be seen that the demographic makeup of each program city is different.

Total sample size, n=22

Crosstabulation - Gender and Ethnicity by Program City

		Municipality		
		Brampton	Mississauga	Total
Gender	Female	36% (n=4)	64% (n=7)	50% (n=11)
	Male	64% (n=7)	36% (n=4)	50% (n=11)
TOTAL		50% (n=11)	50% (n=11)	100% (n=22)
Ethnicity	White	71% (n=5)	29% (n=2)	32% (n=7)
	Black	14% (n=1)	86% (n=6)	32% (n=7)
	Asian	50% (n=1)	50% (n=1)	9% (n=2)
	South Asian	100% (n=2)	0% (n=0)	9% (n=2)
	Latino	50% (n=2)	50% (n=2)	18% (n=4)
TOTAL		50% (n=11)	50% (n=11)	100% (n=22)

From this table, it can be seen that:

- Females are overrepresented in the Mississauga program, and males are overrepresented in the Brampton program
- Over 70% of the White sample is in the Brampton program while only 14% of the Black sample is represented in that program
- Asian and Latino participants are evenly distributed across both program cities
- The entire South Asian sample (n=2) is the Brampton program

This data can also be disaggregated by subcategories within a variable. This enables a deeper look at the units that make up that category. From the sample data, 25% of clients reported being dissatisfied with the program experience.

In the table below, this subcategory of participants is explored in more in-depth.

Total Valid Sample Size, n=5
Dissatisfaction by Gender, Ethnicity and Program City

Dissatisfied with Program Experience	Gender Female=40% (n=2) Male=60%(n=3)
	Ethnicity White=0% (n=0) Black=20% (n=1) Asian=20% (n=1) South Asian=20% (n=1) Latino=40% (n=2)
	Program City Brampton=100% (n=5) Mississauga=0% (n=0)

This table shows that:

- All of the clients who were dissatisfied with the program were clients of color.
- All but one of the clients of color in the Brampton program were dissatisfied with their experience since there were 6 clients of color in the Brampton program.

From these results it may be inferred that the Brampton program is not meeting the needs of its clients of color. This result is masked when the average satisfaction level of **all participants** is reported as 2.8 (on a 4-point scale) and that 75% of the clients were satisfied with their experience.

Moderate and Advanced Analytical Methods

In addition to the basic methods described above there are a variety of more complicated analytical procedures that can be performed with data. These include:

- Correlation
- Regression
- Analysis of variance

These types of analyses generally require computer software (e.g., SPSS, SAS, STATA, MINITAB) and a solid understanding of statistics to interpret the results. Here, each method is briefly described.

Correlation

A correlation is a statistical calculation which describes the nature of the relationship between two variables (i.e., strong and negative, weak and positive, statistically significant).

An important thing to remember when using correlations is that a correlation does not explain causation. A correlation merely indicates that a relationship or pattern exists, but it does not mean that one variable is the cause of the other.

For example, there might be a strong positive correlation between participation in a summer program and students' grades the following school year; however, the correlation doesn't identify if the summer program is the reason why students' grades were higher.

Regression

Regression is an extension of correlation and is used to determine whether one variable is a predictor of another variable.

A regression can be used to determine how strong the relationship is between an intervention and the outcome variables. More importantly, a regression will tell whether a variable (e.g., program participation) is a statistically significant predictor of the outcome variable (e.g., GPA, SAT, etc.). A variable can have a positive or negative influence, and the strength of the effect can be weak or strong.

For example, a regression would help determine if the length of participation (number of weeks) in the summer program is actually predictor of students' high school grades the following year. Like correlations, causation can't be inferred from regression.

Analysis of Variance

An analysis of variance (ANOVA) is used to determine whether the difference in means (averages) for two groups is statistically significant.

For example, an analysis of variance will help determine if the high school grades of those students who participated in the summer program are significantly different from the grades of students who did not participate in the program.

Links to additional resources on advanced methods of data analysis:

StatSoft Electronic Statistics Textbook
Hyperstat Online Statistics Textbook