CHAPTER 13

QUANTITATIVE DATA ANALYSIS

Turning Numerical Data into Meaningful Project Evidence

Presented by

Dr Abenet Yohannes, Ph.D. | MEAL training series

Learning Objectives

By the end of this chapter, participants will be able to:

- Prepare quantitative data for analysis
- Identify variables and levels of measurement
- Calculate frequencies and percentages
- Apply measures of central tendency
- Apply measures of dispersion
- Develop and interpret cross-tabulations
- Conduct disaggregated analysis
- Compare results with targets
- Conduct trend and variance analysis
- Explain basic inferential statistics
- Interpret confidence intervals
- Select tests for relationships and differences
- Distinguish contribution from attribution
- Identify common quantitative-analysis errors
- Select appropriate analytical software
- Prepare clear indicator-performance findings

What is Quantitative Data Analysis?

Quantitative data analysis is the systematic process of organizing, summarizing, comparing and interpreting numerical data.

It helps answer:

- How many?
- How much?
- What percentage?
- How often?
- How different?
- Has performance improved?
- Are groups experiencing different results?
- Is an observed pattern likely to be meaningful?

Analysis, Visualization and Interpretation

Analysis	Visualization	Interpretation
Organizes and summarizes data.	Presents data through charts, graphs, tables and dashboards.	Explains what the findings mean for the project.

Key Message Analysis produces numbers. Interpretation turns numbers into evidence for action.

Descriptive and Inferential Analysis

Two broad forms of quantitative analysis are:

Descriptive Analysis

Summarizes the data collected.

Inferential Analysis

Uses sample data to make conclusions about a wider population.

Descriptive Analysis Questions

Descriptive analysis answers questions such as:

- How many participants were reached?
- What percentage achieved the expected result?
- What was the average score?
- How widely did values vary?
- Which locations performed best?
- What trend occurred over time?

Inferential Analysis Questions

Inferential analysis may ask:

- Are two groups significantly different?
- Is there a relationship between two variables?
- Did the project population improve beyond what might occur by chance?
- How precise is the sample estimate?
- Can findings be generalized to the larger population?

SECTION I

PREPARING DATA FOR ANALYSIS

Begin with the Analysis Plan

Before analysis, review:

- Results Framework
- Indicator definitions
- Performance Indicator Reference Sheets
- Performance Management Plan
- Evaluation questions
- Sampling design
- Data-collection tools
- Data-quality findings
- Intended use of the results

Data-Preparation Process

1. Receive and secure the dataset
2. Verify the file version
3. Review variable structure
4. Check labels and codes
5. Remove or protect identifiers
6. Clean errors and inconsistencies
7. Review missing data
8. Identify duplicates
9. Examine outliers
10. Create derived variables
11. Document all changes
12. Save an analysis-ready file

Raw, Clean and Analysis Files

Maintain separate versions:

Raw Data	Clean Data	Analysis Data
Original data exactly as collected.	Validated and corrected data.	Clean data with calculated variables and analysis fields.

Review Variable Names and Labels

Every variable should have:

- A clear name
- A descriptive label
- Defined response codes
- Correct data type
- Appropriate missing-value codes
- Documented units

Example

Variable Name	Label
hh_size	Number of household members
safe_water	Household uses an improved water source
income_etb	Monthly household income in ETB

Data Types in the Dataset

Variables may be stored as:

- Numeric
- Text or string
- Date
- Time
- Currency
- Binary
- Categorical codes

Coding Categorical Variables

Example:

Sex

- 1 = Female
- 2 = Male
- 3 = Other or context-approved category
- 99 = Prefer not to answer

Project status

- 0 = Not completed
- 1 = Completed

Recoding Variables

Recoding may be required to:

- Create age groups
- Combine rare categories
- Convert scores into performance levels
- Create binary achievement indicators
- Reverse negatively worded items
- Align categories across datasets

Creating Derived Variables

Examples include:

- Percentage scores
- Household dependency ratio
- Total training-attendance days
- Indicator achievement status
- Age categories
- Composite indexes
- Variance from target
- Percentage achievement

Data Cleaning Before Analysis

Check for:

- Invalid codes
- Impossible values
- Missing responses
- Duplicate records
- Inconsistent skip patterns
- Incorrect dates
- Extreme values
- Conflicting answers
- Incorrect units
- Data-entry errors

Handling Missing Data

Before analysis, identify:

- How much data are missing?
- Which variables are affected?
- Which groups have more missing data?
- Why are data missing?
- Should missing records be excluded?
- How will the denominator be affected?

Types of Missingness

Missing data may arise because:

- The respondent refused
- The question was not applicable
- The enumerator skipped the question
- The device failed
- The respondent did not know
- The record was incomplete
- The participant was lost to follow-up

Managing Outliers

An outlier is a value that is unusually far from most other observations.

Possible causes:

- Data-entry error
- Wrong measurement unit
- Genuine extreme case
- Equipment error
- Fraud or fabrication

Weighting Data

Weights may be needed when:

- Subgroups were sampled disproportionately
- Selection probabilities differed
- Cluster designs were used
- Non-response adjustments are necessary
- Population-level estimates are required

Documenting Data Preparation

Maintain records of:

- Variables recoded
- Records removed
- Values corrected
- Missing-data decisions
- Outlier treatment
- Weighting
- New variables created
- Software syntax or formulas used
- Staff responsible
- Date of changes

SECTION II

VARIABLES AND LEVELS OF MEASUREMENT

II

What is a Variable?

A variable is any characteristic, quantity or condition that can vary between cases.

Examples:

- Age
- Sex
- Income
- Attendance
- Satisfaction
- Training completion
- Water-source type
- Disease incidence

Independent and Dependent Variables

Independent Variable

A factor used to explain or predict another variable.

- Training participation
- Age
- Location
- Intervention type

Dependent Variable

The outcome or result being examined.

- Income
- Test score
- Service utilization
- Employment status

Four Levels of Measurement

The major levels are:

1. Nominal
2. Ordinal
3. Interval
4. Ratio

Nominal Data

Nominal data represent categories with no inherent order.

Examples

- Sex
- District
- Disability status
- Water-source type
- Project participation status
- Marital status

Appropriate analysis

- Frequencies
- Percentages
- Mode
- Cross-tabulation
- Chi-square tests

Ordinal Data

Ordinal data have a meaningful order, but the distance between categories is not necessarily equal.

Examples

- Very dissatisfied to very satisfied
- Low, medium and high
- Poor, fair, good and excellent
- Levels of vulnerability

Appropriate analysis

- Frequencies
- Percentages
- Median
- Percentiles
- Rank-based tests

Interval Data

Interval data have equal distances between values but no absolute zero.

Examples

- Temperature in Celsius
- Some standardized scores

Possible analysis

- Mean
- Standard deviation
- Correlation
- Regression
- Parametric tests, where assumptions hold

Ratio Data

Ratio data have equal intervals and a meaningful absolute zero.

Examples

- Age
- Income
- Weight
- Distance
- Number of visits
- Time spent collecting water

Possible analysis

- All common arithmetic and statistical procedures

Levels of Measurement Summary

Level	Ordered?	Equal Intervals?	True Zero?
Nominal	No	No	No
Ordinal	Yes	No	No
Interval	Yes	Yes	No
Ratio	Yes	Yes	Yes

Choosing Analysis by Measurement Level

Variable Type	Common Analysis
Nominal	Counts, percentages, cross-tabs
Ordinal	Median, ranks, percentages
Interval	Mean, SD, correlation
Ratio	Mean, SD, rates, regression

SECTION III

FREQUENCIES AND PERCENTAGES

Frequency

A frequency is the number of times a value or category occurs.

Example

Training Status	Frequency
Completed	360
Did not complete	90
Total	450

Percentage

A percentage expresses the proportion of cases meeting a condition.

$$\text{Percentage} = \text{Numerator} \div \text{Denominator} \times 100$$

Example

$$360 \div 450 \times 100 = 80\%$$

Percent, Valid Percent and Cumulative Percent

Percent	Valid Percent	Cumulative Percent
Uses all records, including missing cases.	Uses only cases with valid responses.	Adds ordered percentages progressively.

Reporting Counts with Percentages

Good reporting

Of the 450 enrolled participants, 360 completed the training, representing 80.0%.

Weaker alternative

“Eighty percent completed.”

Rate versus Percentage

Percentage	Rate
Part of a defined total.	Events relative to a population and time period.

Examples Attendance percentage • Disease-incidence rate • Complaint rate per 1,000 participants • Staff-turnover rate per year

SECTION IV

MEASURES OF CENTRAL TENDENCY

IV

What is Central Tendency?

Measures of central tendency identify the typical or central value in a dataset.

The main measures are:

- Mean
- Median
- Mode

Mean

The mean is the sum of all values divided by the number of observations.

Mean = Sum of values ÷ Number of observations

Example

Scores: 60, 65, 70, 75, 80

350 ÷ 5 = 70

Median

The median is the middle value after observations are ordered.

Median = Middle value of ordered data

Example

Scores: 40, 50, 55, 60, 95

Median = 55

Mode

The mode is the most frequently occurring value or category.

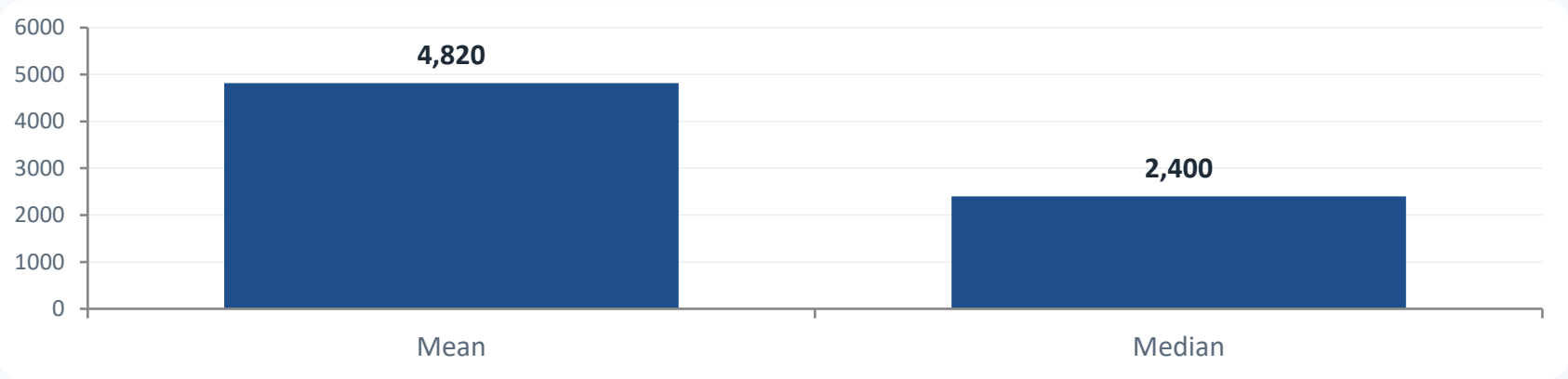
Example — Primary water source:

- Borehole: 210 households
- River: 70 households
- Trucked water: 40 households

Result Mode = Borehole

Mean versus Median

Example monthly income: 2,000; 2,200; 2,400; 2,500; 15,000



Interpretation The mean (4,820) is pulled upward by one unusually high income; the median (2,400) better reflects the typical household.

Selecting the Appropriate Measure

Use:

Mean	Median	Mode
For approximately symmetrical interval or ratio data.	For skewed numerical or ordered data.	For identifying the most common category or value.

SECTION V

MEASURES OF DISPERSION

What is Dispersion?

Measures of dispersion describe how widely values vary around the centre.

Common measures include:

- Range
- Interquartile range
- Variance
- Standard deviation
- Minimum and maximum

Range

The range is the distance between the highest and lowest values.

Range = Maximum value – Minimum value

Example

Maximum distance = 12 km

Minimum distance = 1 km

Range = 11 km

Interquartile Range

The interquartile range, or IQR, covers the middle 50% of observations.

IQR = Third Quartile – First Quartile

Variance

Variance measures the average squared distance of values from the mean.

A higher variance indicates:

- Greater spread
- Less consistency
- More diversity in observations

Standard Deviation

The standard deviation indicates how far observations typically vary from the mean.

Mean test score	Standard deviation
70	5

Interpretation Most scores are relatively close to 70.

Why Dispersion Matters in MEAL

Dispersion can reveal:

- Unequal participant outcomes
- Geographic variation
- Inconsistent service quality
- Different implementation performance
- Hidden subgroups
- Data-quality concerns

Example: Same Mean, Different Results

Two project sites both have an average score of 70.

Site A	Site B
Scores range from 65 to 75.	Scores range from 30 to 100.

Interpretation Site B has much greater inequality and inconsistency. Central tendency alone would hide this difference.

SECTION VI

CROSS-TABULATION

VI

What is Cross-Tabulation?

A cross-tabulation, or crosstab, displays the relationship between two categorical variables.

Example Training completion by sex.

CROSS-TABULATION

Crosstab Example

Sex	Completed	Not Completed	Total
Women	190	60	250
Men	170	30	200
Total	360	90	450

Row Percentages

Using the previous table:

- Women completing: 190 ÷ 250 = 76%
- Men completing: 170 ÷ 200 = 85%

Interpretation Men had a higher completion percentage than women. Use row percentages when comparing outcomes within each group.

Column Percentages

Column percentages show the composition of a category.

Among all participants completing training:

- 52.8% were women
- 47.2% were men

Cross-Tabulation Questions

Cross-tabs can examine:

- Completion by sex
- Satisfaction by location
- Employment by age group
- Service access by disability
- Achievement by intervention type
- Complaint status by category

Cross-Tabs and Small Numbers

Be cautious when:

- Cell counts are very small
- Categories are numerous
- Respondents may be identifiable
- Percentages are based on small denominators
- Statistical tests become unreliable

SECTION VII

DISAGGREGATED ANALYSIS

What is Disaggregated Analysis?

Disaggregated analysis separates results by important population or geographic characteristics.

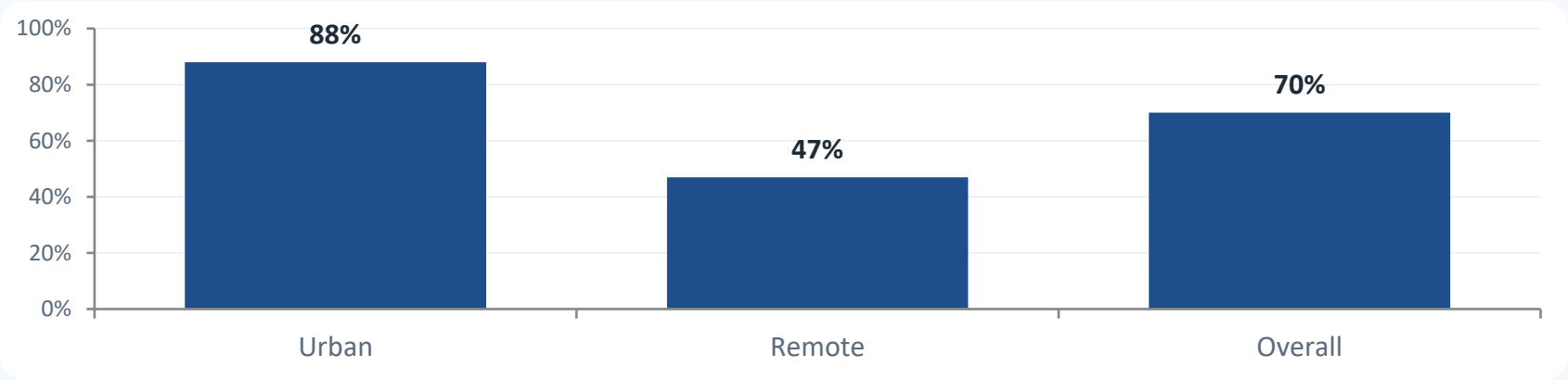
Common categories include:

- Sex
- Age
- Disability
- Location
- Displacement status
- Vulnerability category
- Socioeconomic status
- Project component

DISAGGREGATED ANALYSIS

Why Disaggregate?

Overall service-use rate is 70% — but a total result can hide important differences.



Interpretation The overall figure conceals a major access gap between urban (88%) and remote (47%) communities.

Disaggregation Questions

Ask:

- Which groups achieved the target?
- Which groups did not?
- Are differences increasing or decreasing?
- Are barriers related to location, gender or disability?
- Is the sample sufficient for subgroup analysis?
- Do qualitative findings explain the differences?

Ethical Disaggregation

Protect participants by:

- Avoiding unnecessary sensitive categories
- Suppressing very small groups
- De-identifying data
- Limiting access
- Reporting categories respectfully
- Avoiding stigmatizing conclusions

SECTION VIII

COMPARING ACTUAL RESULTS WITH TARGETS

COMPARING ACTUAL RESULTS WITH TARGETS

Indicator Performance Analysis

For each indicator, compare:

- Baseline
- Milestone
- Current actual
- Period target
- Final target
- Previous period
- Relevant standard or benchmark

Percentage Achievement

Percentage Achievement = Actual ÷ Target × 100

Example

Target = 800 households

Actual = 720 households

720 ÷ 800 × 100 = 90%

Absolute Variance

Absolute Variance = Actual – Target

Example

Actual = 720

Target = 800

Variance = −80 households

Percentage Variance

Percentage Variance = (Actual – Target) ÷ Target × 100

Example

(720 − 800) ÷ 800 × 100 = −10%

Indicators Where Lower Is Better

Examples:

- Disease incidence
- Complaints unresolved
- Dropout rate
- Delivery delays
- Cost per participant
- Stock-out days

Traffic-Light Performance Rating

Example thresholds:

Green	Amber	Red
90% or more of target	75% – 89% of target	Below 75% of target

Variance Explanation

When results differ from targets, examine:

- Implementation delays
- Access or security restrictions
- Budget limitations
- Incorrect assumptions
- Data-quality problems
- Seasonal factors
- Target-setting errors
- Changes in participant demand
- External shocks
- Partner performance

SECTION IX

TREND ANALYSIS

What is Trend Analysis?

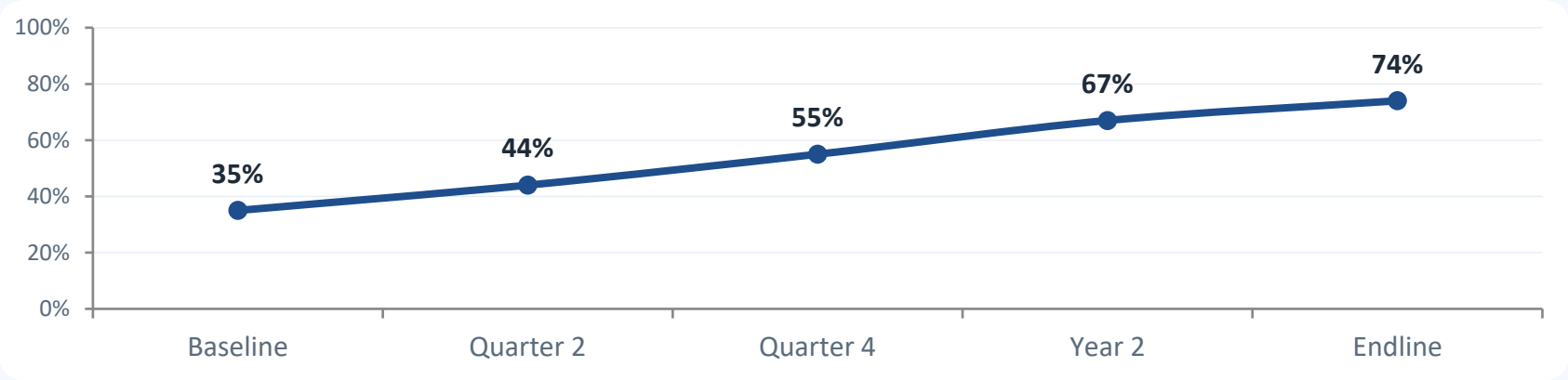
Trend analysis examines how an indicator changes over time.

Possible periods:

- Daily
- Monthly
- Quarterly
- Annually
- Baseline, midline and endline

Trend Example

Percentage of households using safe water:



Interpretation The indicator shows steady improvement.

Trend Patterns

Possible patterns include:

- Increasing
- Decreasing
- Stable
- Seasonal
- Irregular
- Sudden increase
- Sudden decline
- Plateau

Seasonal Effects

Indicators may vary because of:

- Rainfall
- Harvest seasons
- School calendars
- Disease cycles
- Migration
- Conflict
- Religious events
- Market fluctuations

Trend Visualization

Appropriate visualizations include:

- Line charts
- Column charts
- Target-versus-actual charts
- Control charts
- Dashboards
- Geographic maps over time

SECTION X

BASIC INFERENTIAL STATISTICS

What is Inferential Statistics?

Inferential statistics use sample data to estimate or test characteristics of a larger population.

They help answer:

- Is an observed difference likely to be real?
- Is there evidence of a relationship?
- How precise is the estimate?
- Can findings be generalized?

Population Parameter and Sample Statistic

Population Parameter

The true value in the full population.

Sample Statistic

The value calculated from the sample.

Example Population percentage using safe water: unknown • Sample estimate: 68%

Hypothesis Testing

A statistical test commonly compares:

Null Hypothesis	Alternative Hypothesis
No relationship or difference exists.	A relationship or difference exists.

Significance Level

A common significance level is:

$$\alpha = 0.05$$

If p < .05 The observed result is considered statistically significant under the assumptions of the test.

Statistical and Practical Significance

Statistical Significance

The pattern is unlikely to be explained only by sampling variation.

Practical Significance

The size of the difference is meaningful for participants or decisions.

Effect Size

Effect size measures the magnitude of a difference or relationship.

Examples:

- Difference in percentages
- Mean difference
- Correlation coefficient
- Odds ratio
- Cohen’s d
- R-squared

SECTION XI

CONFIDENCE INTERVALS

What is a Confidence Interval?

A confidence interval provides a range of plausible values for a population estimate.

Example:

- Estimated safe-water use = 68%
- 95% confidence interval: 63% to 73%

Interpreting Confidence Intervals

Narrow interval

Suggests greater precision.

Wide interval

Suggests lower precision.

Interval width is influenced by Sample size • Variability • Sampling design • Confidence level

Confidence Interval Example

District A

- 70% service use
- 95% CI: 66% – 74%

District B

- 64% service use
- 95% CI: 51% – 77%

Interpretation District B’s estimate is much less precise.

Confidence Intervals and Targets

Suppose:

- Project target = 70%
- Sample estimate = 72%
- 95% CI = 65% – 79%

Interpretation The point estimate exceeds the target, but the confidence interval includes values below 70%.

SECTION XII

TESTING RELATIONSHIPS AND DIFFERENCES

Selecting a Statistical Test

The choice depends on:

- Analytical question
- Variable type
- Number of groups
- Independent or related groups
- Sample size
- Distribution
- Study design
- Statistical assumptions

Testing Differences Between Two Groups

Examples:

Continuous Outcome	Categorical Outcome
• Independent-samples t-test • Mann–Whitney U test	• Chi-square test • Fisher’s exact test

Testing Differences Across Three or More Groups

Examples:

Continuous Outcome

- One-way ANOVA
- Kruskal–Wallis test

Categorical Outcome

- Chi-square test

Testing Before-and-After Change

Examples:

Continuous Outcome

- Paired-samples t-test
- Wilcoxon signed-rank test

Binary Outcome

- McNemar test

Testing Relationships

Examples:

Two Continuous Variables

- Pearson correlation
- Spearman correlation

Prediction

- Linear regression
- Logistic regression

Two Categorical Variables

- Chi-square test

Correlation

Correlation measures the direction and strength of association between variables.

Possible values:

- Positive relationship
- Negative relationship
- No relationship

Regression

Regression examines how one or more predictors are associated with an outcome while other variables may be controlled.

Examples:

- Predicting school attendance
- Examining factors associated with income
- Predicting service utilization
- Estimating the probability of program completion

Statistical Assumptions

Depending on the test, assumptions may include:

- Independence
- Normality
- Equal variances
- Linearity
- Adequate cell sizes
- No severe outliers
- No excessive multicollinearity

SECTION XIII

CONTRIBUTION VERSUS ATTRIBUTION

What is Attribution?

Attribution means establishing that the observed change was caused by the project.

Strong attribution usually requires:

- Counterfactual evidence
- Comparison or control groups
- Rigorous design
- Control of alternative explanations
- High-quality baseline and endline data

What is Contribution?

Contribution examines whether the project plausibly helped produce the observed change.

It asks:

- Did the expected pathway occur?
- Were project outputs used?
- Are alternative explanations understood?
- Does evidence support the Theory of Change?
- What role did other actors play?

Attribution versus Contribution

Attribution	Contribution
Seeks causal proof	Seeks credible influence
Often uses counterfactuals	Uses Theory of Change and multiple evidence sources
More demanding design	More feasible in complex settings
Strong causal claim	Cautious, evidence-based claim

Contribution Analysis Steps

1. Define the key question
2. Review the Theory of Change
3. Identify assumptions and risks
4. Collect existing evidence
5. Develop the contribution story
6. Examine alternative explanations
7. Seek additional evidence
8. Revise and conclude

Example of Responsible Interpretation

Overstated

The project caused household income to increase by 30%.

More Defensible

Household income increased by 30% among surveyed participants. Monitoring and qualitative evidence indicate that project-supported training, grants and market linkages contributed to the improvement, alongside favourable market conditions.

SECTION XIV

COMMON QUANTITATIVE-ANALYSIS ERRORS

Type I Error

A Type I error is concluding that an effect or difference exists when it does not.

Also called False positive.

Type II Error

A Type II error is failing to detect a real effect or difference (a false negative).

Possible causes:

- Small sample
- High variability
- Weak measurement
- Low statistical power

Common Descriptive Errors

Avoid:

- Incorrect denominators
- Mixing counts and percentages
- Reporting percentages without counts
- Using the mean for highly skewed data
- Ignoring missing data
- Combining incompatible categories
- Double counting participants
- Comparing different time periods without context

Common Inferential Errors

Avoid:

- Using statistical tests with non-random samples
- Ignoring assumptions
- Selecting tests based only on desired results
- Running many tests without adjustment
- Confusing correlation with causation
- Reporting only p-values
- Ignoring effect size
- Ignoring confidence intervals
- Generalizing beyond the sampled population

Cherry-Picking Results

Cherry-picking means selecting only favourable findings.

Examples:

- Reporting only successful indicators
- Ignoring groups with weak outcomes
- Selecting one time period
- Hiding non-significant findings
- Excluding unexplained negative results

Denominator Manipulation

Example — completion rate:

Reported as	Instead of
Completed participants ÷ participants attending the final session	Completed participants ÷ all enrolled participants

Over-Aggregation

Overall results may hide:

- Gender gaps
- Disability exclusion
- Geographic inequalities
- Partner differences
- Vulnerability patterns
- Implementation weaknesses

Misleading Visualization

Avoid:

- Truncated axes exaggerating differences
- Three-dimensional charts
- Inconsistent scales
- Too many categories
- Decorative graphics obscuring values
- Unlabelled denominators
- Colours without explanation

SECTION XV

SOFTWARE TOOLS

Selecting an Analysis Tool

Consider:

- Dataset size
- Analysis complexity
- Staff skills
- Cost
- Collaboration needs
- Security
- Automation
- Visualization requirements
- Donor expectations

Microsoft Excel

Best suited for

- Data cleaning
- Frequencies
- Percentages
- Pivot tables
- Target-versus-actual analysis
- Trend charts
- Small and medium datasets
- Routine project dashboards

Advantages

- Widely available
- Familiar
- Flexible

SOFTWARE TOOLS

SPSS

Best suited for

- Descriptive statistics
- Cross-tabulation
- Hypothesis testing
- Correlation
- Regression
- Survey analysis
- Academic and evaluation datasets

Advantages

- User-friendly menus
- Strong statistical capabilities
- Clear output tables

SOFTWARE TOOLS

Power BI

Best suited for

- Interactive dashboards
- Large linked datasets
- Automated refresh
- Performance visualization
- Geographic analysis
- Management reporting

Advantages

- Dynamic filtering
- Data-model integration
- Strong visualization

Other Suitable Tools

Other tools may include:

- R
- Python
- Stata
- Tableau
- Kobo exports
- DHIS2
- SQL databases
- Google Looker Studio

Tool Comparison

Tool	Best Use
Excel	Routine analysis and tracking
SPSS	Statistical analysis
Power BI	Dashboards and visualization
R / Python	Advanced and automated analysis
Stata	Survey and econometric analysis
Tableau	Interactive visual reporting

Reproducible Analysis

Good practice includes:

- Saving formulas
- Saving syntax or scripts
- Recording filters
- Documenting weights
- Preserving codebooks
- Version controlling files
- Keeping analysis logs
- Assigning approval responsibilities

SECTION XVI

INTERPRETING AND REPORTING RESULTS

From Analysis to Interpretation

For every result, ask:

- What does the number show?
- Is the change meaningful?
- Who benefited?
- Who did not?
- How does it compare with the target?
- What factors may explain it?
- How reliable is the data?
- What should management do?

Triangulating Quantitative Findings

Compare quantitative findings with:

- Interviews
- Focus groups
- Observations
- Community feedback
- Monitoring reports
- Context information
- Partner data
- Evaluation findings

Writing an Indicator Finding

A strong finding includes:

1. Indicator name
2. Actual result
3. Target or comparison value
4. Disaggregation
5. Trend or variance
6. Explanation
7. Limitation
8. Management implication

Example Indicator Finding

Example indicator finding

At the end of Year 2, 68% of supported households used an improved water source, compared with a target of 75% and a baseline of 41%. Achievement was higher in urban locations (82%) than remote rural communities (51%). Access constraints, delayed construction and seasonal shortages contributed to the shortfall. The finding is based on a probability sample of 520 households.

Variance Narrative

A good variance narrative explains:

- Size of the variance
- Direction
- Root causes
- Affected locations or groups
- Data-quality concerns
- Corrective measures
- Revised expectations

Quantitative Findings Table

Indicator	Baseline	Target	Actual	Achievement	Key Interpretation
Safe-water use	41%	75%	68%	90.7%	Progress strong but rural gap remains
Training completion	0	500	540	108%	Target exceeded; verify quality and duplication
Complaint resolution	54%	90%	84%	93.3%	Improved but below service standard

Recommendations from Quantitative Findings

Recommendations should be:

- Evidence-based
- Specific
- Linked to performance gaps
- Assigned to responsible teams
- Time-bound
- Feasible
- Monitored through follow-up indicators

SECTION XVII

PRACTICAL EXERCISE

PRACTICAL EXERCISE

Practical Exercise Overview

Participants will analyse a sample project dataset and prepare indicator-performance findings.

1. Review and clean the data
2. Classify variables
3. Produce descriptive statistics
4. Develop cross-tabs
5. Disaggregate findings
6. Compare actual results with targets
7. Conduct trend and variance analysis
8. Interpret findings
9. Prepare a results summary

PRACTICAL EXERCISE

Sample Project Scenario

A livelihoods project supports 600 households.

Key indicators

- Percentage completing business training
- Percentage starting an income-generating activity
- Average monthly income
- Percentage reporting improved market access
- Percentage satisfied with project support

Required disaggregation

- Sex
- District
- Disability status
- Displacement status

PRACTICAL EXERCISE

Exercise Part 1 — Data Preparation

Participants should:

- Identify variable types
- Review codes
- Check missing values
- Identify duplicates
- Examine outliers
- Create age groups
- Create achievement variables
- Document cleaning decisions

PRACTICAL EXERCISE

Exercise Part 2 — Descriptive Analysis

Produce:

- Frequency tables
- Percentages
- Mean, median and mode
- Minimum and maximum
- Standard deviation
- Appropriate charts

PRACTICAL EXERCISE

Exercise Part 3 — Cross-Tabs and Disaggregation

Analyse:

- Training completion by sex
- Business start-up by district
- Market access by displacement status
- Satisfaction by disability status
- Income by location

PRACTICAL EXERCISE

Exercise Part 4 — Target Analysis

For each indicator, calculate:

- Actual value
- Target value
- Percentage achievement
- Absolute variance
- Percentage variance
- Performance status

PRACTICAL EXERCISE

Exercise Part 5 — Trend Analysis

Using quarterly data:

- Plot indicator trends
- Identify increases and decreases
- Compare actuals with milestones
- Note unusual values
- Suggest possible explanations

PRACTICAL EXERCISE

Exercise Part 6 — Indicator Findings

Prepare three written findings that include:

- Result
- Comparison with target
- Disaggregation
- Trend
- Explanation
- Limitation
- Recommended action

PRACTICAL EXERCISE

Practical Analysis Checklist

Confirm that:

- ✓ Dataset was cleaned
- ✓ Variables were correctly classified
- ✓ Denominators were verified
- ✓ Missing values were addressed
- ✓ Counts accompany percentages
- ✓ Appropriate central tendency was used
- ✓ Dispersion was examined
- ✓ Results were disaggregated
- ✓ Targets were compared consistently
- ✓ Limitations were reported
- ✓ Findings lead to decisions

PRACTICAL EXERCISE

Peer Review Questions

Review another group's analysis:

- Are calculations accurate?
- Are percentages based on correct denominators?
- Were appropriate statistics selected?
- Are important differences visible?
- Are interpretations supported by evidence?
- Are causal claims overstated?
- Are data limitations disclosed?
- Are recommendations practical?

SECTION XVIII

COMMON MEAL ANALYSIS MISTAKES

Common Preparation Mistakes

Avoid:

- Analysing the wrong file version
- Overwriting raw data
- Undocumented recoding
- Ignoring missing data
- Deleting outliers without investigation
- Mixing units
- Failing to remove identifiers
- Creating calculations without documentation

Common Interpretation Mistakes

Avoid:

- Treating association as causation
- Ignoring external factors
- Focusing only on average results
- Ignoring negative findings
- Overgeneralizing from non-random samples
- Interpreting statistically significant results as automatically important
- Hiding uncertainty
- Ignoring qualitative evidence

Common Reporting Mistakes

Avoid:

- Excessive decimal places
- Percentages without denominators
- Tables without titles
- Charts without labels
- Different definitions across reports
- Unexplained target variance
- No disaggregation
- No limitations
- No management implications

SECTION XIX

CHAPTER REVIEW

CHAPTER REVIEW

Chapter Summary

In this chapter, we learned that:

- Quantitative analysis organizes and interprets numerical data.
- Data preparation is essential before analysis.
- Measurement levels determine appropriate statistical methods.
- Frequencies and percentages describe categories.
- Mean, median and mode describe central tendency.
- Range, IQR, variance and standard deviation describe dispersion.
- Cross-tabs reveal relationships between categorical variables.
- Disaggregation identifies equity and performance differences.
- Targets, trends and variance support management decisions.
- Inferential statistics quantify uncertainty and test patterns.
- Confidence intervals communicate estimate precision.
- Statistical relationships do not automatically establish causation.
- Contribution is often a more appropriate claim than attribution.
- Analysis software should match organizational needs and skills.
- Findings must be interpreted, triangulated and converted into action.

CHAPTER REVIEW

Knowledge Check

1. What is the difference between descriptive and inferential analysis?
2. Why must raw data be preserved?
3. What are the four levels of measurement?
4. When is the median preferable to the mean?
5. What does standard deviation show?
6. What is the purpose of a cross-tabulation?
7. Why is disaggregation important?
8. How is percentage achievement calculated?
9. What does a confidence interval communicate?
10. What is the difference between contribution and attribution?
11. What is a Type I error?
12. Why should effect size be considered?
13. Which software is suitable for routine target analysis?
14. What should a strong indicator finding contain?

CHAPTER REVIEW

Action Planning

Each participant should identify:

- One dataset requiring cleaning
- One indicator with a denominator problem
- One missing disaggregation
- One report using an inappropriate average
- One unexplained performance variance
- One overstated causal conclusion
- One required analytical improvement
- Responsible person
- Completion date

THANK YOU

Questions and Discussion

Dr Abenet Yohannes, Ph.D
abenetyohannes@gmail.com

Quantitative analysis is not simply about producing numbers. It is about using reliable evidence to understand performance, identify inequalities, explain change and make better project decisions.