

Project Monitoring Course

Prepared & presented by:
Dr. Dhouha Melliti

2022-2023

جامعة الجميع الذكية

تسعى لتكون الكيان الأول الذي يتبنى مبدأ المجتمع التعليمي
التعاوني ويقدم العلم مقابل العمل لتنمية وبناء المجتمعات.

Lecture 07:

Methodologies for data collection and
analysis for monitoring and evaluation
(Part2)



جامعة الجميع الذكية
EVERYONE'S SMART UNIVERSITY

Lecture7

Session4:Methodologies for data collection and analysis for monitoring and evaluation (Part2)

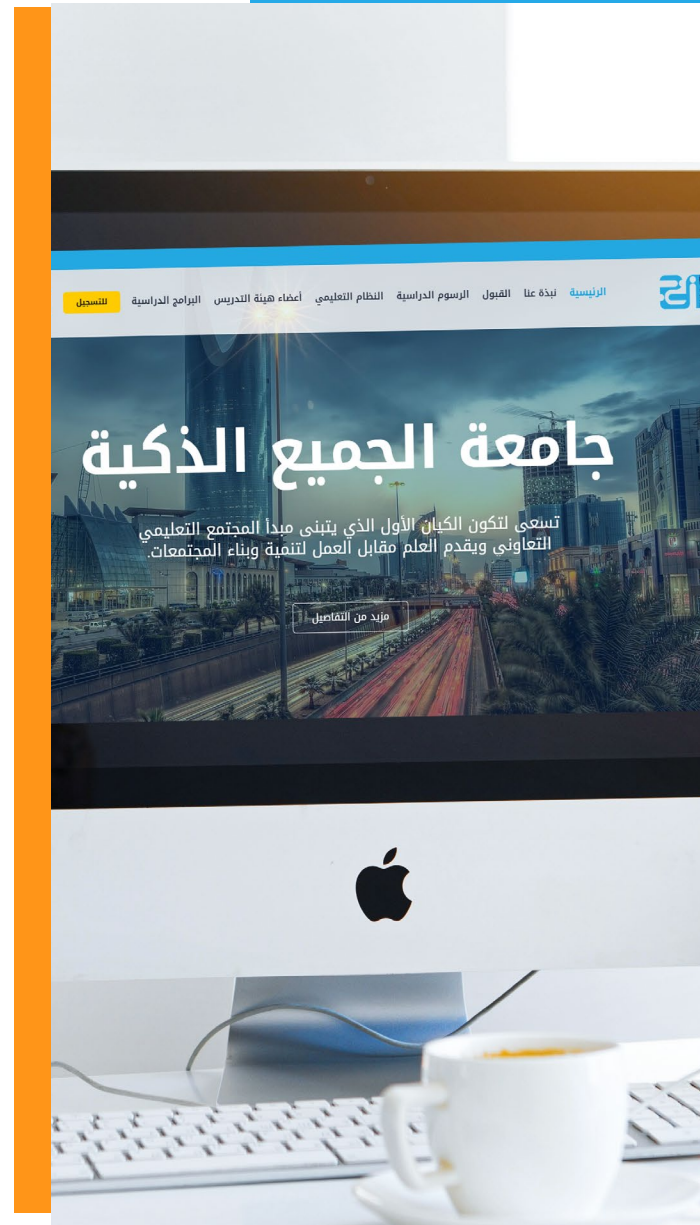
« You can accomplish anything in life if you have the courage to dream it, the intelligence to make it a realistic project, and the will to see this project carried out. » *Sidney A. Friedman*



Lecture 7

Session 4: Methodologies for data collection and analysis for monitoring and evaluation (Part 2)

✓ An overview of chapter 4





4. Collecting and managing data

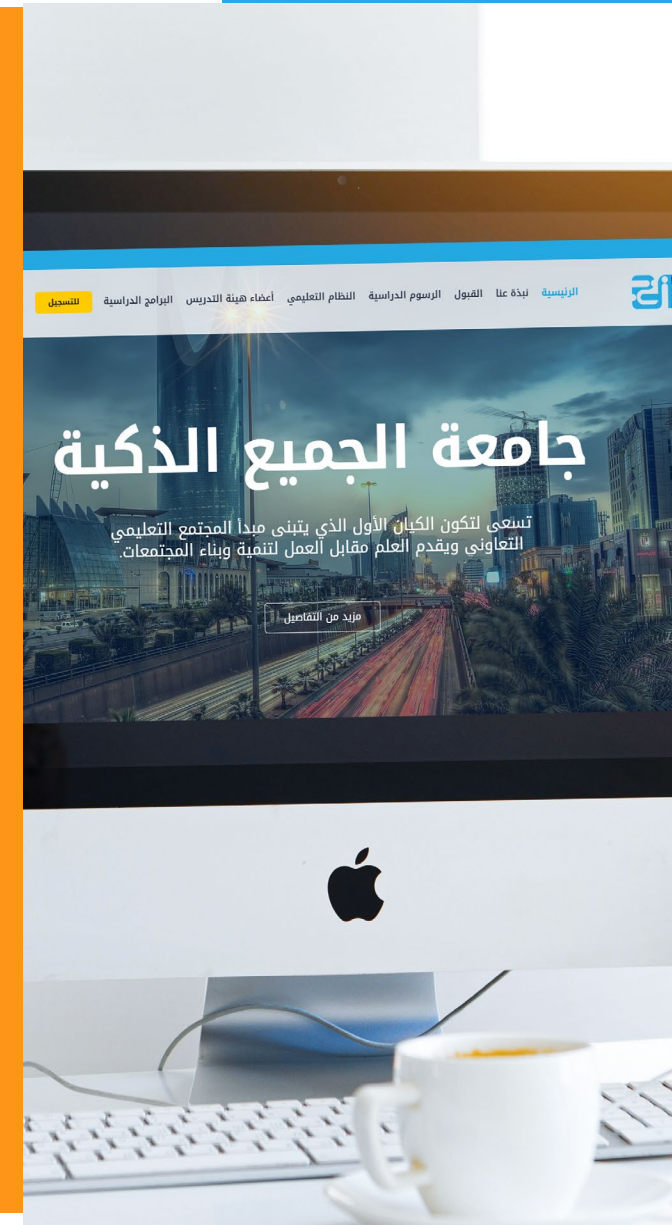
✓ Once the M&E design has been identified and the method(s) and tools have been developed, the data collection can start. It is also recommended to organize a training with the data collection team(s) on the methodology. The training should cover in detail each data collection tool that will be used and include practical exercises of how to implement them.



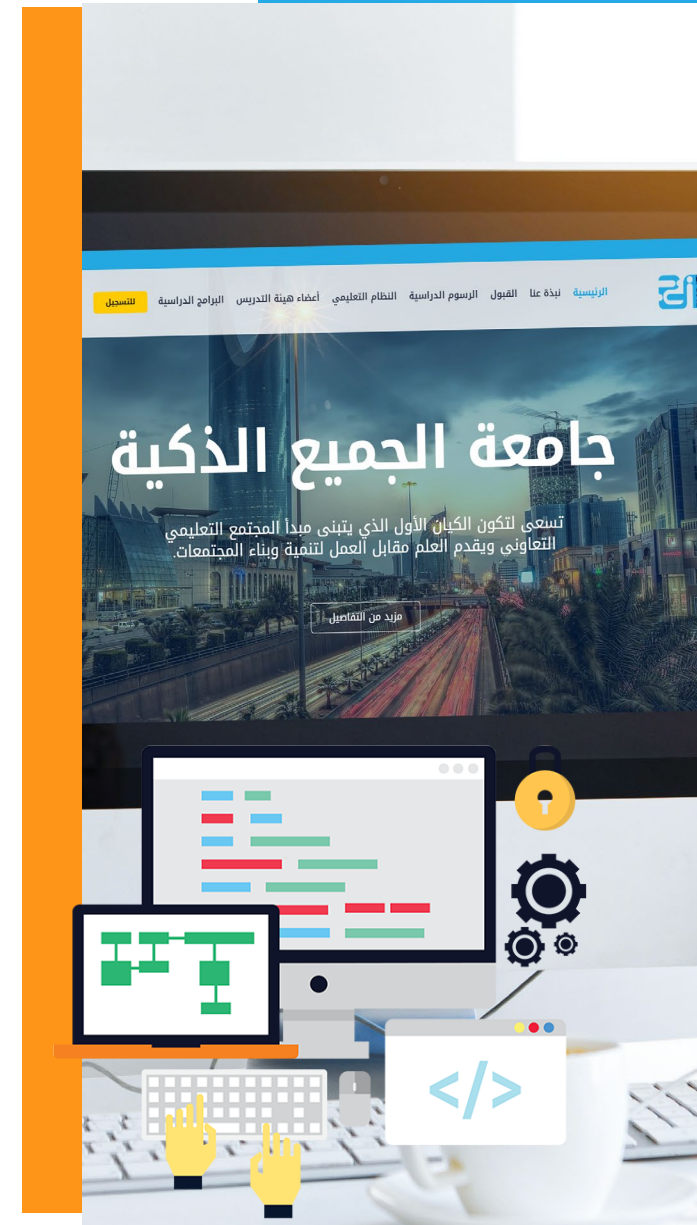


Data collection

✓ Developing a data collection guide with clear instructions for the enumerators is a useful reference tool, both during the training and after, for the actual data collection; see the example provided below for an excerpt from a survey included in a data collection guide.



- **Objective:** To find out how regularly the respondent is receiving financial support from a third party which can be a person or an organization.
- **Instructions:** First read out the question and response options and then circle the respondent's answer (one of the four options listed beside the question). Each data collection team should have a supervisor who can oversee the data collection and check for any errors. During the data collection, it is imperative that the supervisor of the data collection team regularly checks for the following:



Are there any forms missing?

Are there any double forms for a respondent?

Are there any answer boxes or options left blank?

Are there more than one option selected for closed-ended questions with single-option responses?

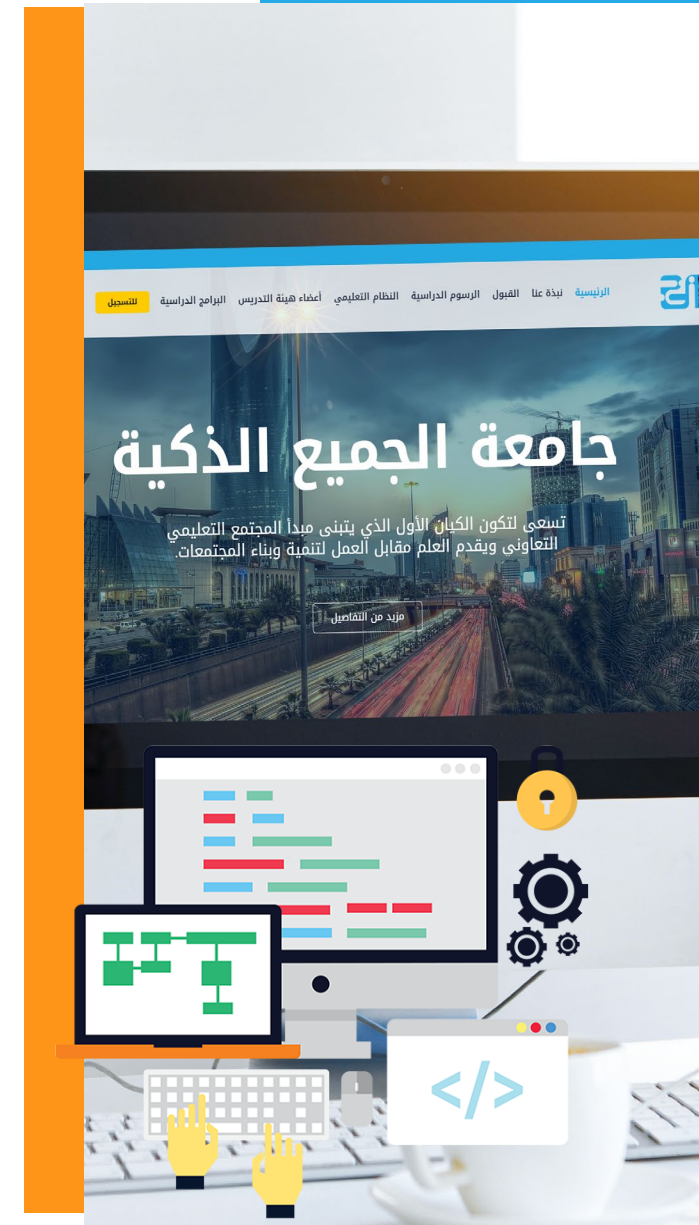
Are correct values filled out in the wrong boxes?

Are the answers readable?

Are there any writing errors?

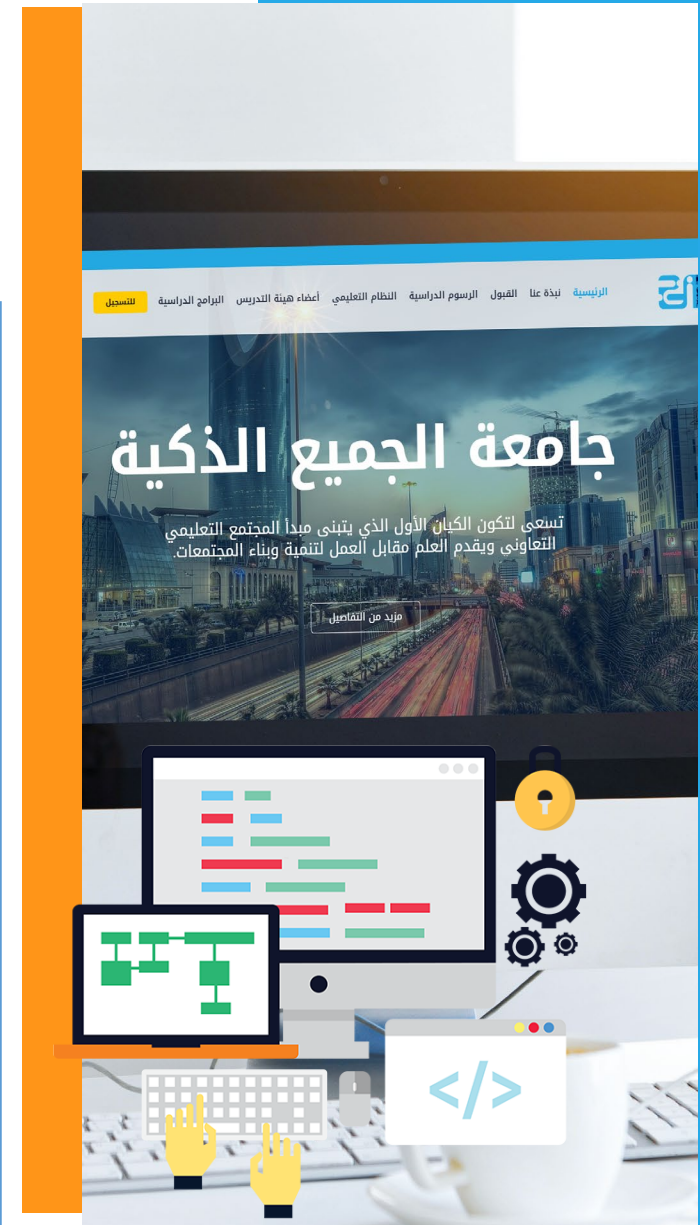
Are there any answers that are out of the expected range (outliers)?

Doing these checks will help reduce the amount of error in the data collected.



Data entry

The data collected needs then to be transferred onto a computer application, such as Microsoft Word or Excel. Having the data in an electronic format will facilitate the data clean-up and data analysis. For quantitative data, the first step in data entry is to create the data file(s) to achieve a smooth transfer between a spreadsheet and a statistical programme package, such as [SPSS](#) and [Stata](#) for conducting statistical analyses.



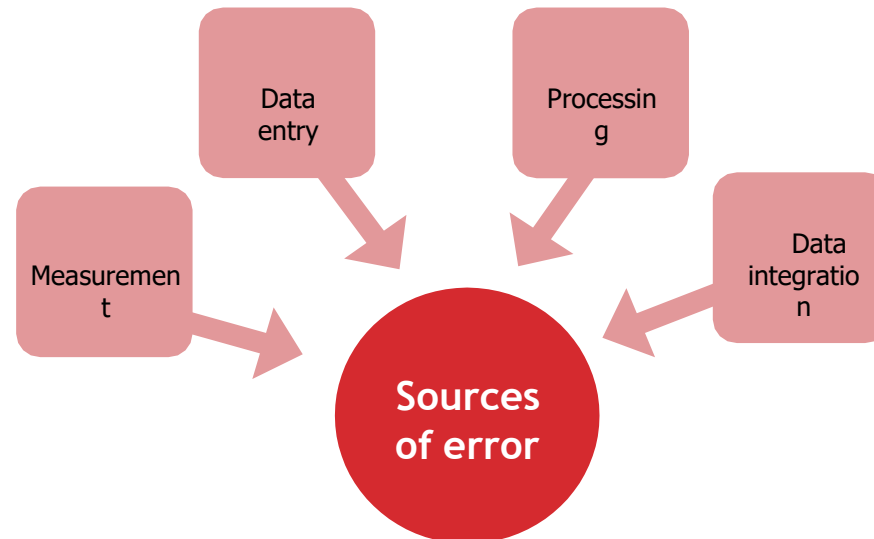
Data structure for cross-sectional data: A table of numbers and text in which each row corresponds to an individual subject (or unit of analysis) and each column corresponds to a different variable or measurement. There is one record (row) per subject.

Data structure for longitudinal data: The data can be structured in a wide data file format or a long data file format. In the wide format, a subject's repeated responses will be in a single row, and each response is in a separate column. In the long format each row is one time point per subject; so each subject (county) will have data in multiple rows. Any variables that don't change across time will have the same value in all the rows.

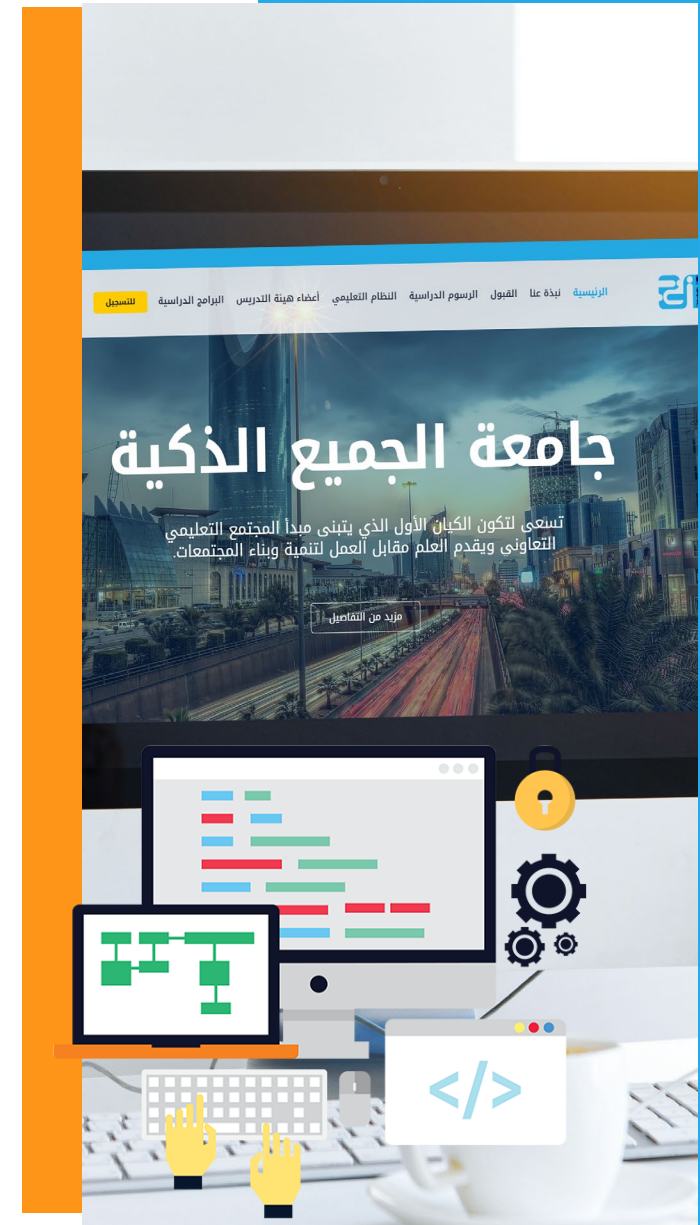


Data clean-up

Once the data has been transferred from the medium used to record the information to a computer application (Word or Excel), it needs to be screened for errors. Following this, any errors need to be diagnosed and treated.

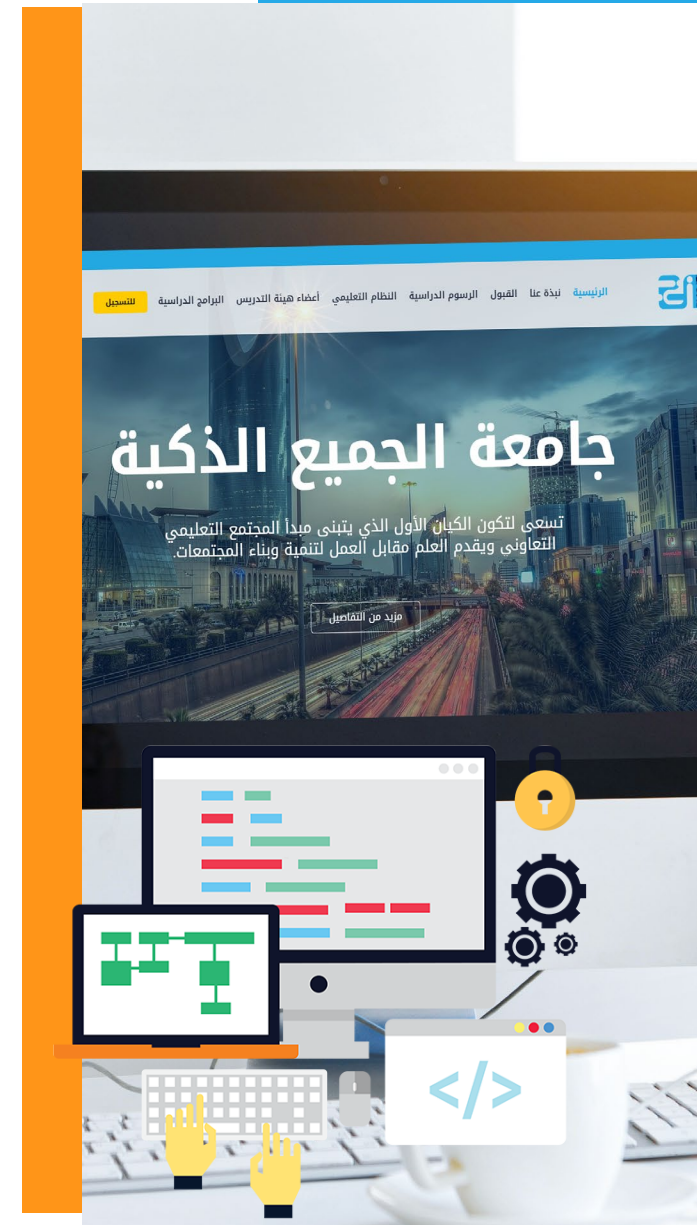


Sources of error



Key errors to look for when screening data

- **Spelling and formatting irregularities:** Are categorical variables written incorrectly? Are date formats consistent?
- **Lack of data:** Do some questions have fewer answers than surrounding questions?
- **Excessive data:** Are there duplicate entries? Are there more answers than originally allowed?
- **Outliers/inconsistencies:** Are there values that are so far beyond the typical distribution that they seem potentially erroneous?
- **Strange patterns:** Are there patterns that suggest cheating rather than honest answers (that is, several questionnaires with the exact same answers)?
- **Suspect analysis results:** Do the answers to some questions seem counter-intuitive or extremely unlikely?



Selected data screening method

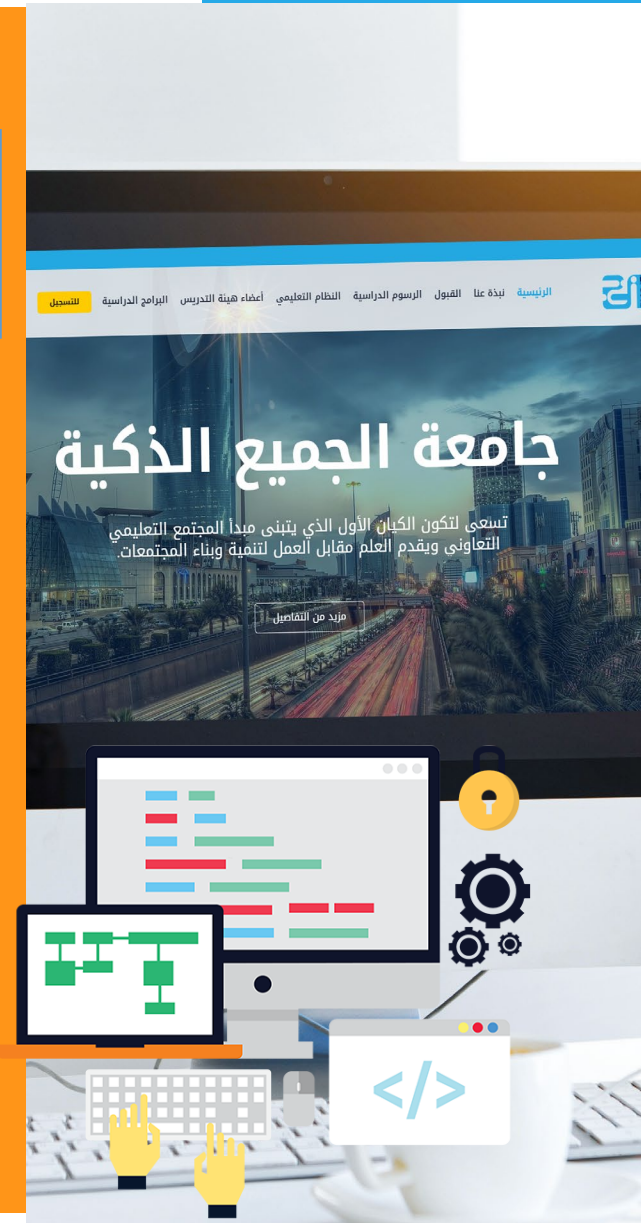
Quantitative data

- Browse data tables after sorting
- Calculate summary statistics
- When time allows, validate data entry
- Create frequency distributions and cross-tabulations
- Graphically explore data distributions using box plots, histograms and scatter plots with the help of visual analysis software.

Qualitative data

- Check for spelling errors
- Compare data with assumptions or criteria
- Take counts of words and phrases
- Create frequency distributions and cross-tabulations.

Depending on the number of data collection tools used and amount of data collected, data entry agents may need to be recruited and trained to do the data entry (and data clean-up).



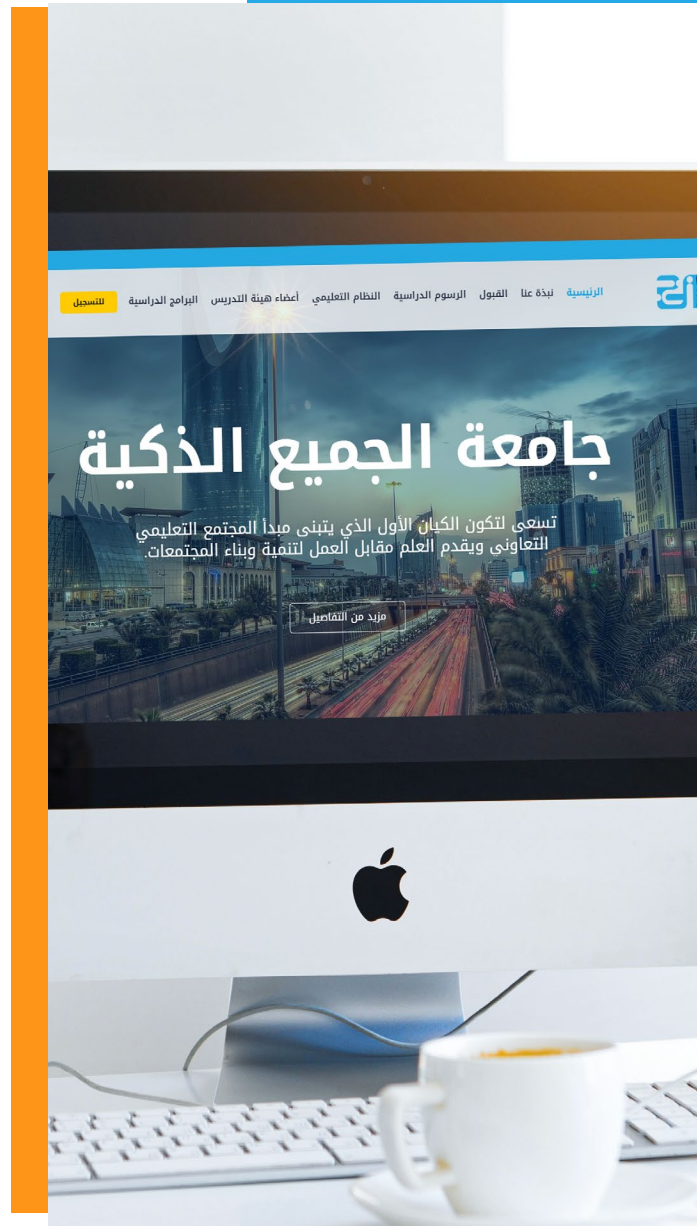
Diagnosis

- ✓ Once the suspect data has been identified, the next step is to review all the respondent's answers to determine if the data makes sense given the context in which it was collected. Following this review, there are several possible diagnoses for each suspect data point identified:
 - ❖ The data point is missing. Missing data can be a result of omitted answers by the respondents (no response), questions that are skipped over by the enumerator (erroneous entry or skip pattern) or the data entry agents, or there are dropouts (for longitudinal research).
 - ❖ The data point is a true extreme value. True extreme values are answers that seem high but can be justified by other answers.
 - ❖ The data point is a true normal value. True normal values are valid answers.



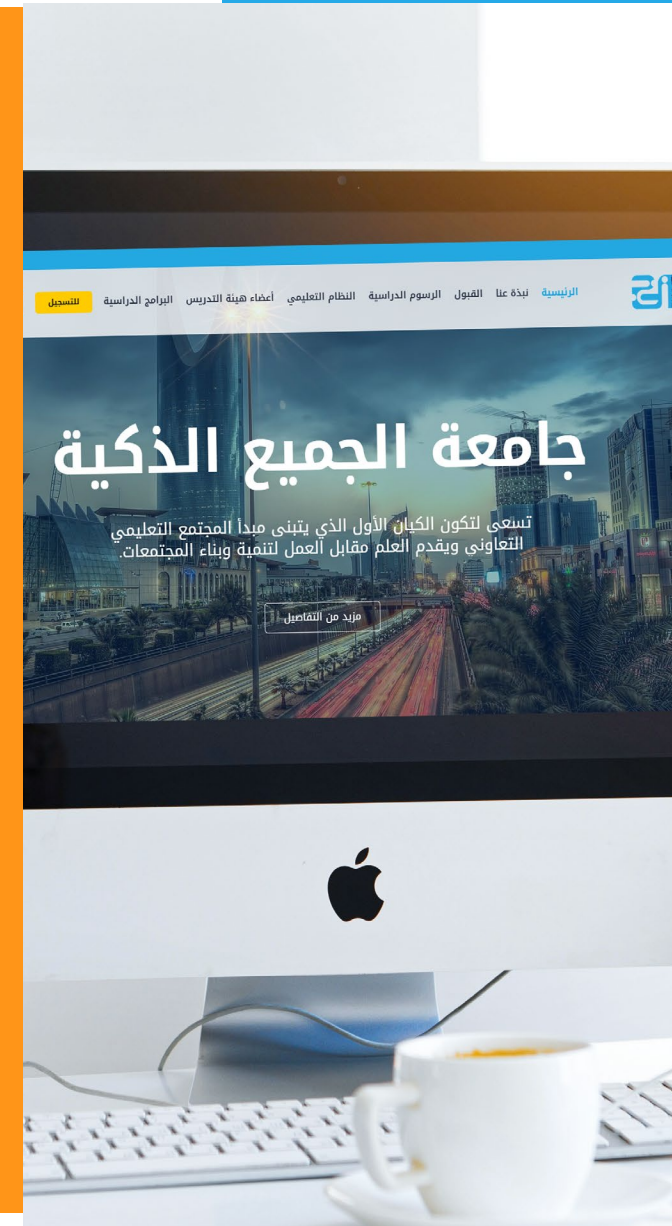
Treatment

- ✓ Once the problematic observations have been identified, these need to be treated before the data can be analysed. The following are some of the key approaches to dealing with data errors:
 - ❖ Leave the data unchanged. This approach is the most conservative as it entails accepting the erroneous data as valid response(s) and making no changes. For large-n studies, leaving one erroneous response may not affect the analysis. However, for small-n studies, the decision of leaving the data unchanged may be more problematic.
 - ❖ Correct the data, however without modification of the intention of or meaning given by the respondent.
 - ❖ Delete the data. It is important to remember that leaving out data can make it seem as if the data is being “cherry-picked” to obtain the desired results



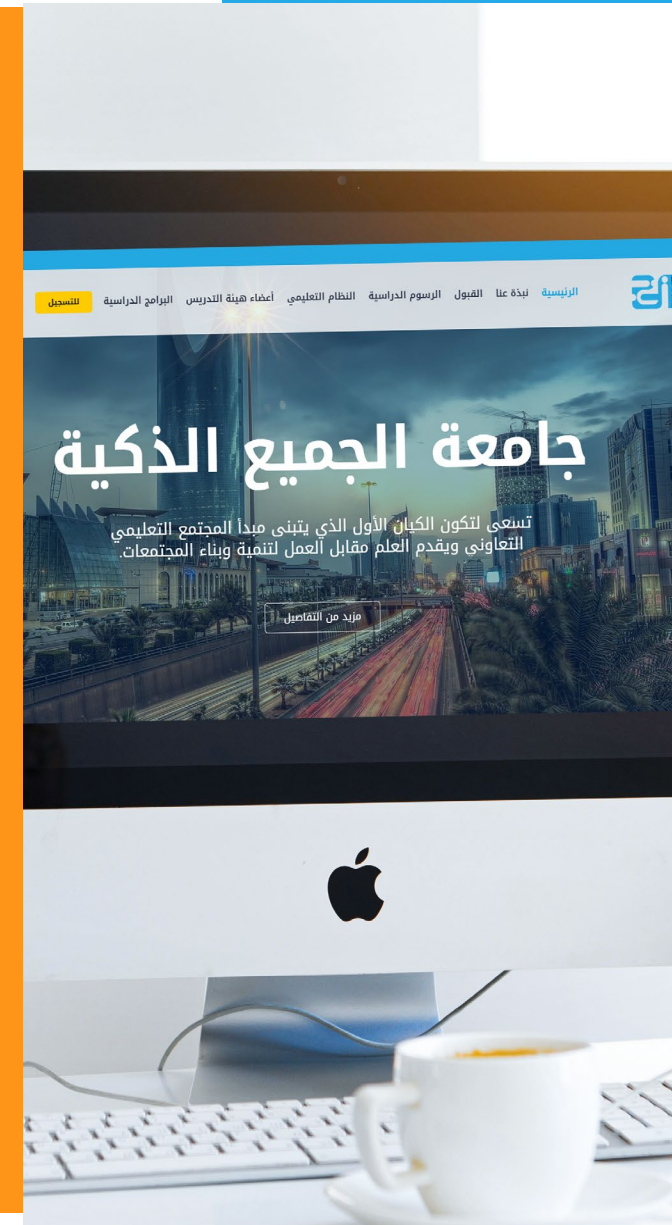
General decision-making rules

- If the person doing the data entry has entered values different from the ones in the survey, the value should be changed to what was recorded in the survey form.
- When variable values do not make sense and there is no data entry error nor notes to determine where the error comes from, the value should be left as it is. Any changes will bias the data.
- When blank cases are present for questions that required an answer, or if erroneous values cannot be corrected, these may be deleted from the data file.
- When there are still suspect and true extreme values after the diagnostic phase, it is necessary to next examine the influence of these data points, both individually and as a group, on the results before deciding whether or not to leave the data unchanged.
- Any data points taken out of the data set should be reported as “excluded from analysis” in the methodology chapter of the final report.



Missing data

- Missing values require attention because they cannot be simply ignored. The first step is to decide which blank cells need to be filled with zeros (because they represent negative observation; for example “no”, “not present” and “option not taken”) and which to leave blank (if using blanks to indicate missing or not applicable). Blank cells can also be replaced with missing value codes; for example, 96 (I don’t know), 97 (refused to answer), 98 (skip question/not applicable) and 99 (blank/missing).
- If the proportion of missing or incomplete cases is substantial for a category of cases, this will be a major M&E concern. Once a set of data is known to be missing, it is important to determine whether the missing data are random or whether they vary in a systematic fashion, and also the extent to which the problem exists. Random missing values may occur because the subject inadvertently did not answer some questions.is missing?



Recoding and creating new variables

Guidelines for collapsing data

- Ordinal variables need to be collapsed in a way that preserves the order of the categories.
- Combine only those categories that go together.
- The way in which categories are collapsed can easily affect the significance level of statistical tests. Categories should be collapsed a priori to avoid the criticism that the data were manipulated to obtain a certain result.
- Do not oversimplify the data. The unnecessary reduction in the number of categories can reduce statistical power and make relationships in the data ambiguous. Generally, any categories that include 10 per cent or more of the data (or 5 cases for very small samples) should be kept intact.



Documenting change

Two good data management practices are transparency and the proper documentation of all the procedures followed, including the data cleaning process.

Documenting errors, changes and additions is essential to the following:

- ✓ Determining and maintaining data quality;
- ✓ Avoiding duplication of error checking by different data entry staff;
- ✓ Knowing what and by whom data quality checks have been carried out;
- ✓ Recovering data cleaning errors;
- ✓ Informing data users of the changes made to the last version of the data accessed.



Analyzing data

- Once the data has been collected and cleaned, these are ready to be analysed. Data analysis makes it possible to assess whether, how and why the intervention being monitored and/or evaluated is on track towards achieving, or has achieved, the established objectives. This part of the chapter will discuss and provide examples of how to analyse qualitative and quantitative M&E data, as well as the triangulation of data sources.



Qualitative data analysis

- ❖ **Qualitative data analysis** is a process aimed at reducing and making sense of vast amounts of qualitative information – very often from multiple sources, such as focus group discussion notes, individual interview notes and observations – in order to deduce relevant themes and patterns that address the M&E questions posed. When analysing qualitative data, the focus is on the words spoken by the respondents, the context in which the data was collected, the consistency and contradictions of respondents' views, the frequency and intensity of participants' comments, their specificity and emerging themes and patterns. For example, as part of monitoring an ongoing project, it is decided to conduct 10 focus groups with a select number of beneficiaries. What should be done once all the discussion notes are collected? Should the data be analysed in an ad hoc or systematic fashion, that is, highlight the relevant information or code it?

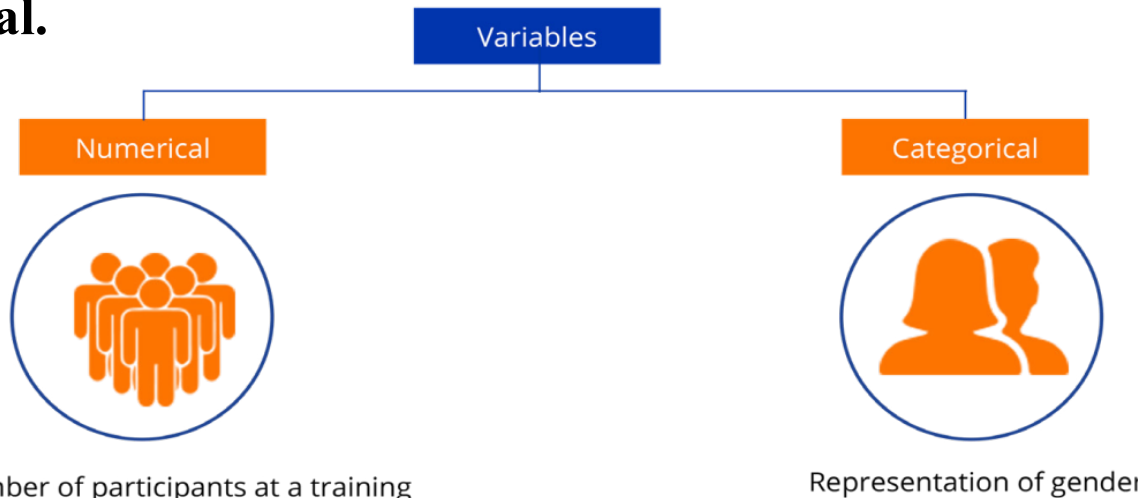


Quantitative data analysis

- ❖ **Once the Quantitative data** has been entered in a spreadsheet, it is ready to be used for creating information to answer the monitoring or evaluation questions posed. Statistics help transform quantitative data into useful information to help with decision-making, such as summarizing data and describing patterns, relationships and connections. Statistics can be descriptive or inferential. As its name already reveals, descriptive statistics give information that help describe data and help to summarize it. Other methods of descriptive statistics are graphical representations in form of histograms, pie charts and bar charts, to name a few. This provides a quick method of making comparisons between different sets of data and spotting smallest and largest values, trends or changes over a period of time. Inferential statistics use data drawn from a sample of the population to make generalizations about populations.



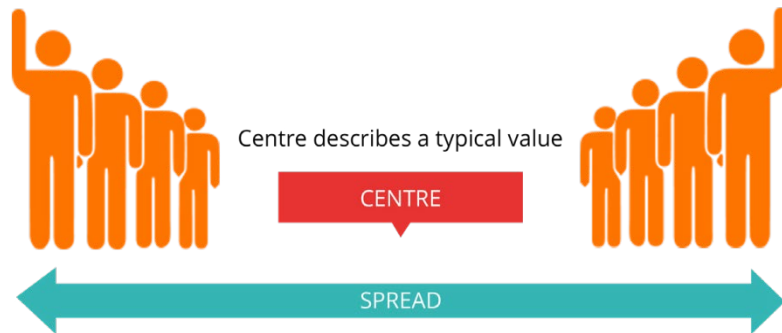
- ❖ **Categorical data groups** are all units in distinct categories, which can be summarized by determining how many times a category occurs. For example, the number of females in a community, which can be described as the frequency of females in the community. This information is presented using a frequency table. The frequency table shows how many individuals in the community fall into each category (male/ female).
- ❖ This can also then be represented as a percentage or proportion of the total.



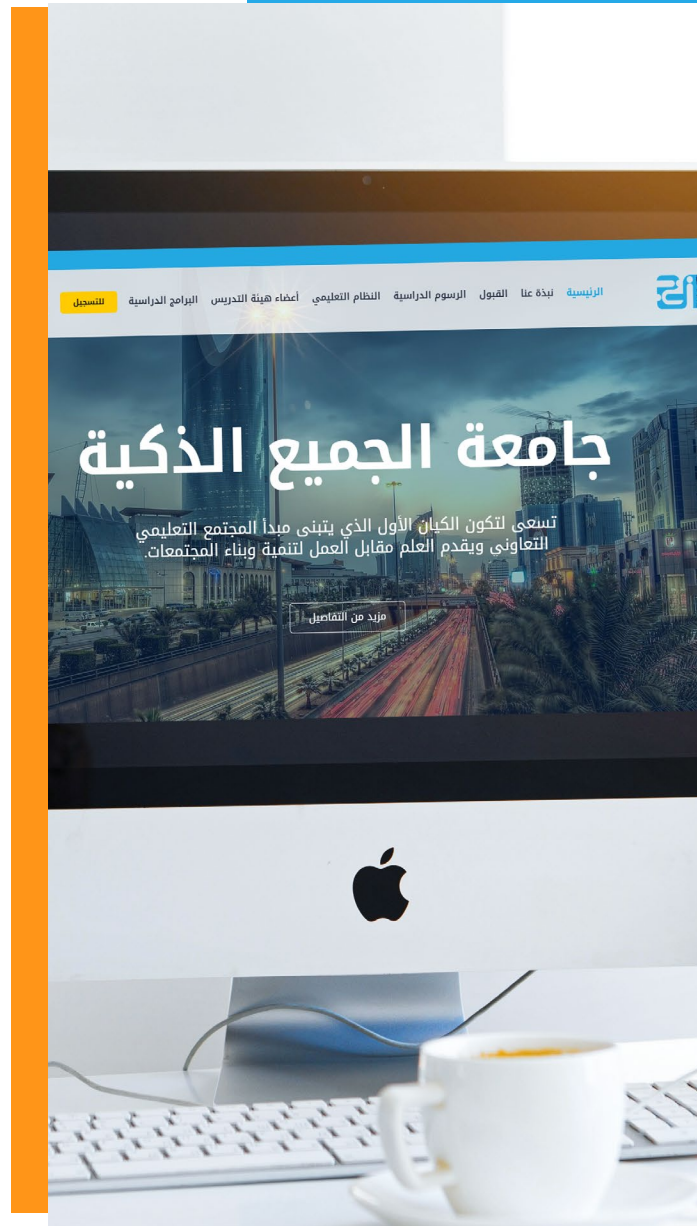
Analysis of numerical variables

- ❖ The center and the spread of the data are two commonly used descriptive statistics. Whereas the centre describes a typical value, the spread describes the distance of a data point from the centre of the data.
- ❖ The most common statistics used to describe the center are the mean (that is, the average) and the median. The median is the middle value in a data set; half the data are greater than the median and half are less. The mean is calculated by adding up all the values and then dividing by the total number of values.

Two Commonly Used Descriptive Statistics



Spread describes the distance of a data point from the centre of the data



Triangulation of data sources

- **Triangulation** is the process of comparing several different data sources and methods to corroborate findings and compensate for any weaknesses in the data by the strengths of other data.
- Triangulation can enhance the validity and reliability of existing observations about a given issue. The ability to compare and contrast different findings and perspectives on the same situation and/or phenomenon is an effective way to find inconsistencies in data and identify areas for further investigation. When findings converge, this can lead to new, credible findings about an issue and create new ways of looking at it.



Presenting findings

- **M&E efforts aim to generate, and make available, relevant information for decision-making and the management of the intervention being monitored or evaluated. All data visualizations should summarize the collected data and communicate the findings obtained in a simple and intuitive way for the reader. This final part of this chapter will discuss and provide examples of the different types of techniques and tools available for visualizing data, depending on what it demonstrates and what is the aim of conveying it to the reader.**



How to visualize findings

- **Step 1: Identify the data visualization goal**
 - ✓ Before M&E practitioners start designing any data visualization, the following questions should be asked:
 - ✓ What is the data trying to communicate?
 - ✓ How will it engage or persuade the audience to act upon the information being presented?
 - ✓ What is the takeaway message the audience should be left with?
- **Step 2: Know the audience**
 - ✓ Knowing the audience means asking what the audience already knows, what additional information they wish to have to learn and how much detail they require to understand the message being conveyed.
- **Step 3: Think about how to visualize the story**
 - ✓ Once the data collected is cleaned and has been analyzed, a more precise idea about what findings to present should emerge.



Lecture7

Session4:Methodologies for data collection and analysis for monitoring and evaluation (Part2)

What the data shows	Appropriate visualization technique use
Summary	Summary table
Facts and figures	Icons and images draw attention to the data values
Comparison, rank and distribution	Bar charts and heat maps using shapes and colours represent numerical values
Proportion or part-to-whole	Pie charts, donuts, stacked bar charts, tree maps show distribution within a group
Change over time	Line graphs for time-series analyses with optional trend lines
Relationships and trends	Scatter plot and bubble graphs can help show correlation
Text analysis	Word clouds to visually display the most common words in a qualitative data set



- **How to conduct a desk review?**

- Steps to conduct a desk review*

- Step 1: Identify all possible sources.

- Step 2: Categorize documents.

Because some documents will be more pertinent than others, not all documents should be given equal weight or attention.

- Step 3: Decide on the approach (Structured/Unstructured).

Often times, due to time constraints, monitoring and evaluation activities limit themselves to first-tier and partially second-tier documents for which an unstructured approach is suitable. To include a full second- and third-tiers documents in the desk review, a structured approach will be required.



Reviewing, translating, pretesting and piloting the survey(1/2)

- ✓ The translator(s) should be fluent in both languages and, to the extent possible, be familiar with the issues being covered in the survey. Once the survey is translated into the local language, a second translator should translate it back into the original language.
- ✓ This process ensures an accurate translation. Any gaps or misunderstanding need to be addressed before the survey can be pretested



Reviewing, translating, pretesting and piloting the survey(2/2)

- ✓ When conducting a pretest, it is important to test the survey among a group of people from diverse backgrounds relevant to the issues being monitored/evaluated. The goal of a pretest is to ensure that the survey is collecting the information it is aimed to collect. A good pretest should look at the survey at three levels:
 - ❖ **As a whole:** Are all sections of the survey consistent? Are there any sections that ask the same question?
 - ❖ **Each section:** If the survey has more than one section, does each section collect the intended information? Are all major activities/issues being monitored/evaluated accounted for? Are there any questions that are not relevant?



Preparing, conducting and moderating a focus group

Preparing for a focus group discussion

- An average focus group discussion involves 6 to 8 people and a maximum of 15 and lasts between one to two hours. When selecting the participants for a focus group discussion, it is important that the group is homogeneous so that participants feel more comfortable expressing their opinions. To select homogenous groups of participants, consider, among other things, the following:



Preparing, conducting and moderating a focus group

- **Gender:** Will men and women feel comfortable discussing this topic in a mixed-gender group? For example, women might feel uncomfortable discussing issue X if men are in the group.
- **Age:** Will age affect the way that people react to this topic? For example, a young person might feel uncomfortable talking about issue X if older people from his community are in the group.
- **Hierarchy:** Will people of different hierarchical positions be able to discuss this topic equally? For example, a resident from village A might feel uncomfortable discussing issue X if the local administrator is in the group.



How to conduct a focus group discussion

- At the start of the focus group, the moderator should present the aim of the discussion and an overview of the topics that will be covered, assure participants that their responses will remain confidential and anonymous, and lay down the ground rules. The following is an example of an introduction to read at the outset of the focus group to all the participants:
- "Thank you for agreeing to participate in this focus group discussion. We will discuss ... Let me remind you that the information given will be treated confidentially and that you may refuse to answer any of the questions and end your participation in the discussion at any time.
- For the group discussion to proceed smoothly and respectfully for all participants, the following ground rules should be respected at all times by everyone:
 - ✓ One person speaks at a time.
 - ✓ What is shared in the room stays in the room.
 - ✓ Everyone has a chance to share their ideas, experience and opinions.
 - ✓ There are no right or wrong answers.
 - ✓ Everyone actively listens to and respects each other.



Lecture7

Session4:Methodologies for data collection and analysis for monitoring and evaluation (Part2)



Icebreaker example

One-words: This icebreaker allows the group to be familiar with one another by sharing their thoughts on a common topic. First, divide the participants into subgroups of four or five people by having them number off. This allows participants to get acclimated to the others in the group. Mention to the groups that their assignment is to think of one word that describes X; give the groups a minute to generate a word. After, the group shares the one word that describes X with the entire group of participants. For example, in a session about mobile usability testing, request the group to think about their smartphone and come up with one word to describe it.



Steps for analysing qualitative data

Step 1: Get to know the data

Before beginning to analyse the data, M&E practitioners need to familiarize themselves with them. This process requires reading and rereading the notes taken (the data) in their entirety. As they go through the data, it is important to take notes of any thoughts that come to mind and summarize each transcript and piece of data that will be analysed. The goal at this stage is to absorb and think about the data that has been collected, jotting down reflections and hunches but reserving judgement. Some open-ended questions that can be asked include: “What is happening here?” or “What strikes you?”



Step 2: Initial round of coding

Once the content of the data is known, practitioners can begin coding the material to condense the information into key themes and topics that can help answer the M&E questions posed.

There are two main approaches to coding qualitative data. The first approach consists of creating a framework that reflects the monitoring or evaluation aims and objectives, which is then used to assess the data gathered. This is a deductive approach, as the concepts and issues of interest are first identified, which allows one to focus on particular answers of respondents and disregard the rest.

The initial round of coding begins with an examination of the data and assignment of words or phrases that capture their essence. Those who use a manual approach could write the code in the margin, and those who use a word-processing software could type the code in the Comment function or in another column.



Step 3: Second round of coding (1/2)

The second round of coding is about beginning to critically examine the codes identified in the initial round of coding and organize, synthesize and categorize them into interpretive concepts. This second round aims to explain, theorize and synthesize the codes from the first round by identifying patterns, rules or cause–effect progressions and making interpretations.

For instance, if codes that continually reappear in the data are identified, M&E practitioners may decide to link them together in a specific way that responds to the monitoring/evaluation question posed.



Step 3: Second round of coding (2/2)

Accordingly, at this point, a better understanding of which data will be most important for the analysis will emerge. Certain data, even if they are already collected, may only tangentially relate to the questions being explored, and therefore, they will not be included in the analysis at hand. It is also at this point that M&E practitioners will see whether additional data needs to be collected to flesh out an emerging code or explanation of what is being observed in the data collected.

One way to identify whether additional data is required is to ask this question: “Does the emerging analysis address the monitoring/evaluation question posed in an interesting and significant way?” If not, this may suggest the need for more data. It might also suggest the need for additional synthesizing activities.



Lecture 2



جامعة الجميع الذكية
EVERYONE'S SMART UNIVERSITY

