Combating Malnutrition in Mazar-e Sharif

A Guidebook for Data Analysis in Monitoring and Evaluating High Diversity Gardening

Nurul Fajriati Setyaningrum
Julia Glaser
Hermine ten Hove

WAGENINGEN UNIVERSITY & RESEARCH
Seeds Afghanistan From Research to Practice
Village of Peace

Colophon:

Nurul Fajriati Setyaningrum, Intern

Wageningen Centre for Development Innovation

Email: nurulfajrs@gmail.com

Julia Glaser, Inclusive food systems advisor

Wageningen Centre for Development Innovation:

Email: julia.glaser@wur.nl

T: +31 (0)317486781

Hermine ten Hove, Advisor

Nutrition in Food Systems | Monitoring, Evaluation, and Learning

Wageningen Centre for Development Innovation:

Email : hermine.tenhove@wur.nl

T: +31 (0)317 485 273

www.wur.nl/wcdi

www.disclaimer-uk.wur.nl

To the trainers: About this Book

This guidebook is written as a complement to the previous guidebook (Auliani et al., 2023). It is part of the High Diversity Gardening project aimed at combating malnutrition and improving food security in Mazar-e Sharif. The guidebook is organized into four chapters: (1) Introduction to Data Analysis, (2) Data Cleaning and Coding, (3) Data Analysis, and (4) Data Visualization.

Disclaimer

Data analysis is a complex task, and mistakes can lead to incorrect conclusions and poor decisions. If you have any doubts about your abilities, it is advisable to consult an expert in data analysis or statistics. Here are situations where seeking expert help is crucial:

- If you are unsure whether you have correctly followed all the necessary steps in the guidebook or have skipped any steps.
- If your data does not meet the required assumptions (see Chapter 3).
- If you have doubts about the realism of your findings. For example, if your analysis suggests that the average household owns 100 cows, it is worth re-evaluating the results.

Table of Contents

1 Introduction to data analysis

In this chapter, we will learn about:

- Basic concepts of statistics
- Microsoft Excel basic

Why do data analysis?

Data are not only numbers, but numbers with context. For example, the number 3 holds no information on its own. However, a newborn with a weight of 3 kg indicates a normal weight. The context makes the number informative and allows us to formulate a viewpoint. Data become more informative or meaningful when we perform data analysis, where we organize, display, and summarize data using certain methods and strategies.

Terminology

Before we plan a statistical study or explore data that has been collected, we need to understand the common terminology used in the field. The following are terminology that we need to learn:

Population and sample

The population refers the entire group of elements (e.g. subjects, animals, plants, etc) about which we want information. A sample is a part of the population from which information is actually collected. For example, the population is all people participated in the home garden intervention, whereas a selection of 150 of them would be the sample. As you may remember, we performed simple random sampling before beginning our data collection.

By doing this, we ensured that the sample was representative of the population, allowing us to make valid generalizations. In other words, the sample is sufficiently large and randomly selected to include a diverse range of individuals, allowing us to draw conclusions about the entire population, not only the people selected as sample.

Variables

Variables are properties or characteristics of some event, object, person or any element that can take on different value or amounts. Variables can be classified as categorical variables and numerical variables. In the following text, you will learn the meaning and examples of categorical and numerical variables.

1. Categorical variables place an element to one of several groups or categories. Categorical variables can be divided into two types, nominal and ordinal.
   - Nominal variables are purely qualitative and unordered, for example, type of livestock.
   - Ordinal variables can be ranked, for example education levels.
2. Numerical variables can be measured and calculated with mathematical operations, such as adding or averaging. The values are usually recorded in a unit of measurement, for

example kilograms or seconds. Numerical variables can be divided into two types, continuous and discrete.

- Continuous variables can take any real numerical value over an interval, for example weight.
- Discrete variables can take limited, finite number of values, for example number of children.

Measure of central tendency

Measure of central tendency, also called as average is mostly used in statistical analysis. It is divided into three types, each providing different information about the data.

1. Mean is the most common type of average. It is calculated by summing all the values in a group, and then dividing by the number of values. For example, in the Excel table below (Table 1), if we want to know the mean time spent going to the market, we could calculate it by summing all the times spent and dividing by 5. You will get a mean of 49.

Table 1: Time spent going to market

Participants	Time (minutes)
A	37,00
B	39,00
C	59,00
D	57,00
E	55,00

2. Median is the middle value of a dataset when the values are ordered from the smallest value to the highest value or from the highest to the smallest. It is calculated by listing the values in order. In the table below (Table 2), you can see the number of months in which participants experienced period of food shortages. The middle-most value is 7, and this is the median.

Table 2: Number of months

Participants	Number of months
A	8
B	6
C	6
D	9
E	7

Participants	Number of months
A	9
B	8
C	7
D	6
E	6

Median is also known as the 50^{th} percentile or Q_2. Percentile points are used to define the percentage of cases equal to and below a certain point in a distribution or set of

scores. For example, if the number of months of food shortage is 7 at the 50th percentile, it means 50% of the households experienced 7 or fewer months of food shortage, and 50% experienced 7 or more months of food shortage.

When to use the median instead of mean? When we have a set of values in which one or more values are extreme. In the table below (Table 3), we have the list of incomes of taxi drivers. The value 150,000 is much higher than the other four in the table. We consider this as an extreme value. The mean of the incomes is 55,800, whereas the median is 32,000. The median is more representative of the group, representing a central position.

Table 3: Median of incomes

Participants	Incomes
A	150,000
B	40,000
C	32,000
D	30,000
E	27,000

Moreover, we would use the median instead of the mean when the data is not normally distributed or is skewed to one side of a graph. We will learn more about distributions later.

3. Mode is the value that occurs most frequently. For example, if you needed to describe the most popular crops, you could list the frequency of the planted crops, and find the one that occurs most frequently. In the table below, we can see that the strawberries are the mode.

Table 4: Frequency of planted crops

Crops	Frequency
Strawberry	5
Onion	2
Tomato	1
Potato	2
Soybean	1
Carrots	1

Measure of dispersion

Dispersion, also called variability is a measure of how different scores are from one another. Variability indicates how data is scattered or clustered together. In the figure below (Figure 1), you can see three set of scores with the same mean (4). However, they have different variability. The set of scores on the left shows more variability compared to the set of scores

on the middle. The set of scores on the right shows no variability at all, the scores are not different from one another.

Figure 1: Variability of scores

The variability can be measured by standard deviation and variance. Standard deviation is the deviation from the standard, which is the mean. It is calculated as the average distance from the mean. The larger the standard deviation, the larger the average distance each data point is from the mean. Variance is simply calculated by standard deviation squared. Understanding variance is important because it's the practical measure of variability in many statistical formulas. We will learn more about this later in the section 'Comparative Analysis'.

Downloading data from Kobo Toolbox Dashboard

In this guidebook and the training, we will use our dataset that is stored online in the Kobo Toolbox dashboard.

Here are the steps:

(1) To download the data that was collected, go to 'Summary' tab and click on 'Downloads' (Figure 2).

(2) Click on 'Select export type' and choose 'XLS'. Click on 'Value and header format' and choose 'XML values and headers' (Figure 3).

(3) Click on 'Export' button, wait for the export process to complete, then click the 'Download' button once the file is ready.

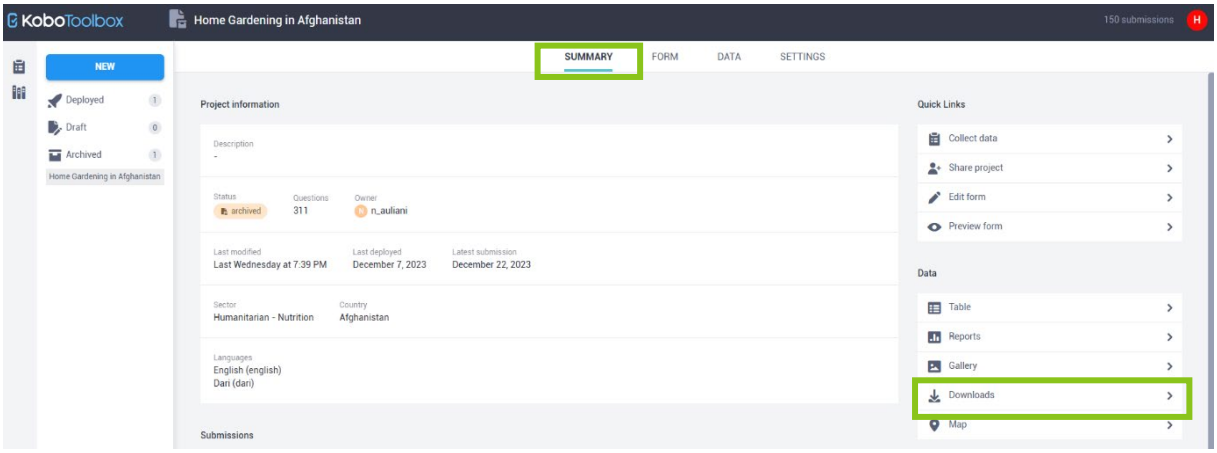


Figure 2: Download dataset (1)

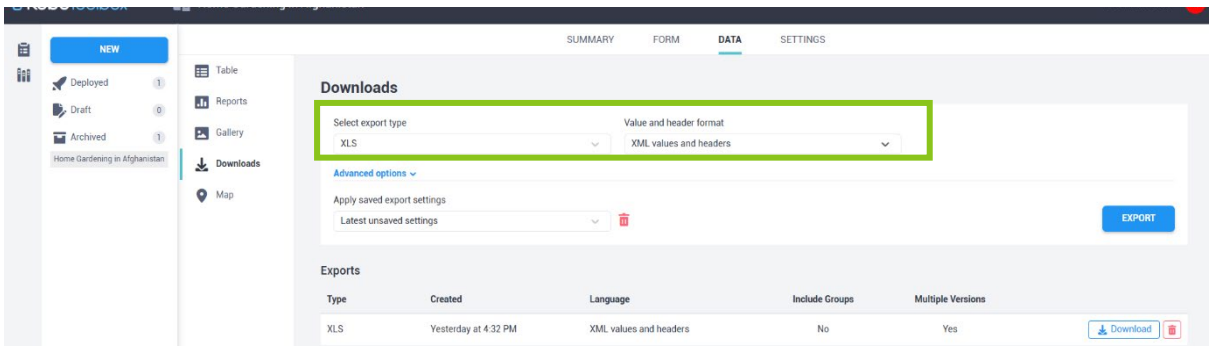


Figure 3: Download dataset (2)

We will use Microsoft Excel to analyse the data. On the next page, you will see a figure of the basic layout and functions of Microsoft Excel (Figure 4). You will also need to familiarize yourself with the structure of our data in the Microsoft Excel before we analyse it (Figure 5).

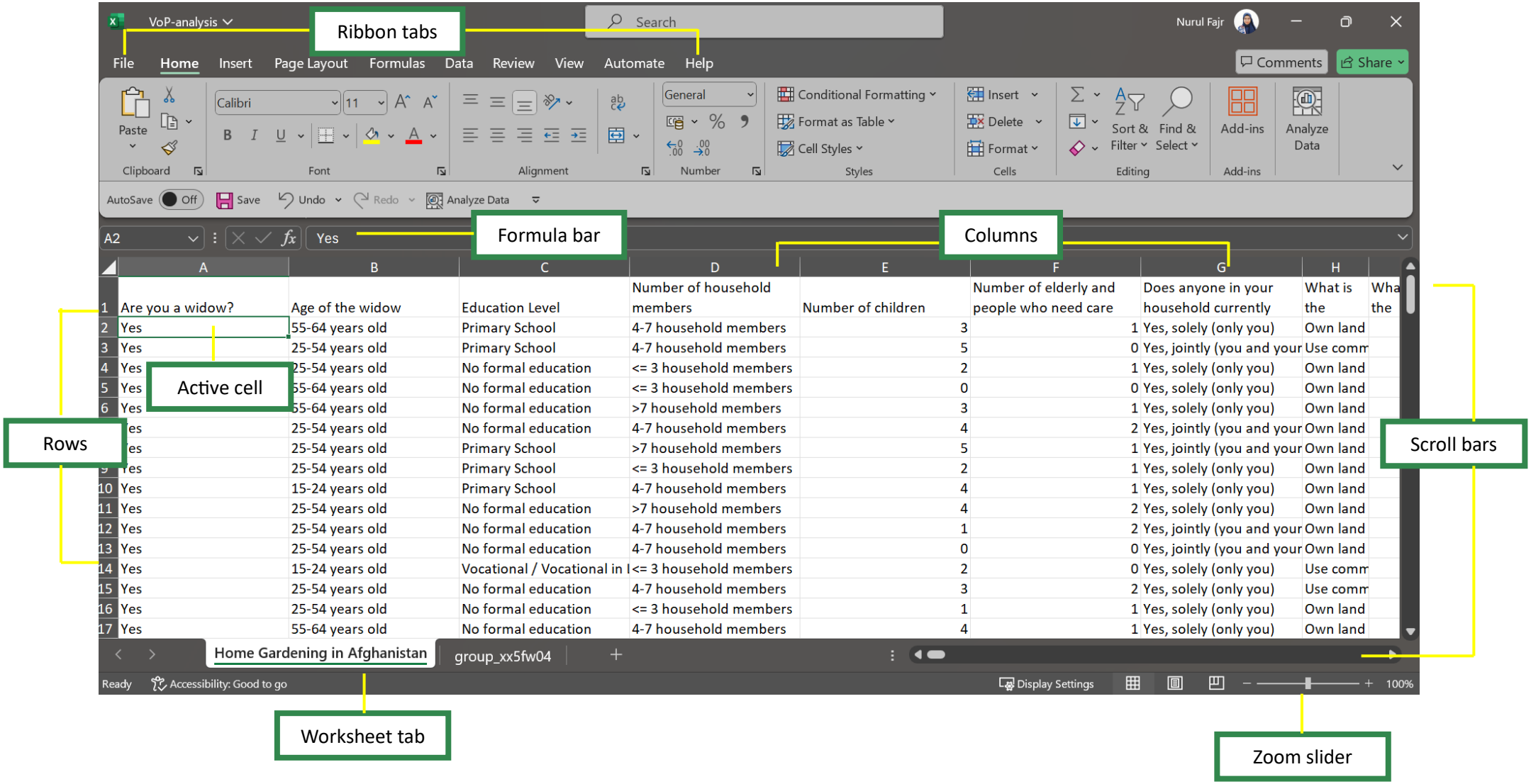


Figure 4: Microsoft Excel basics

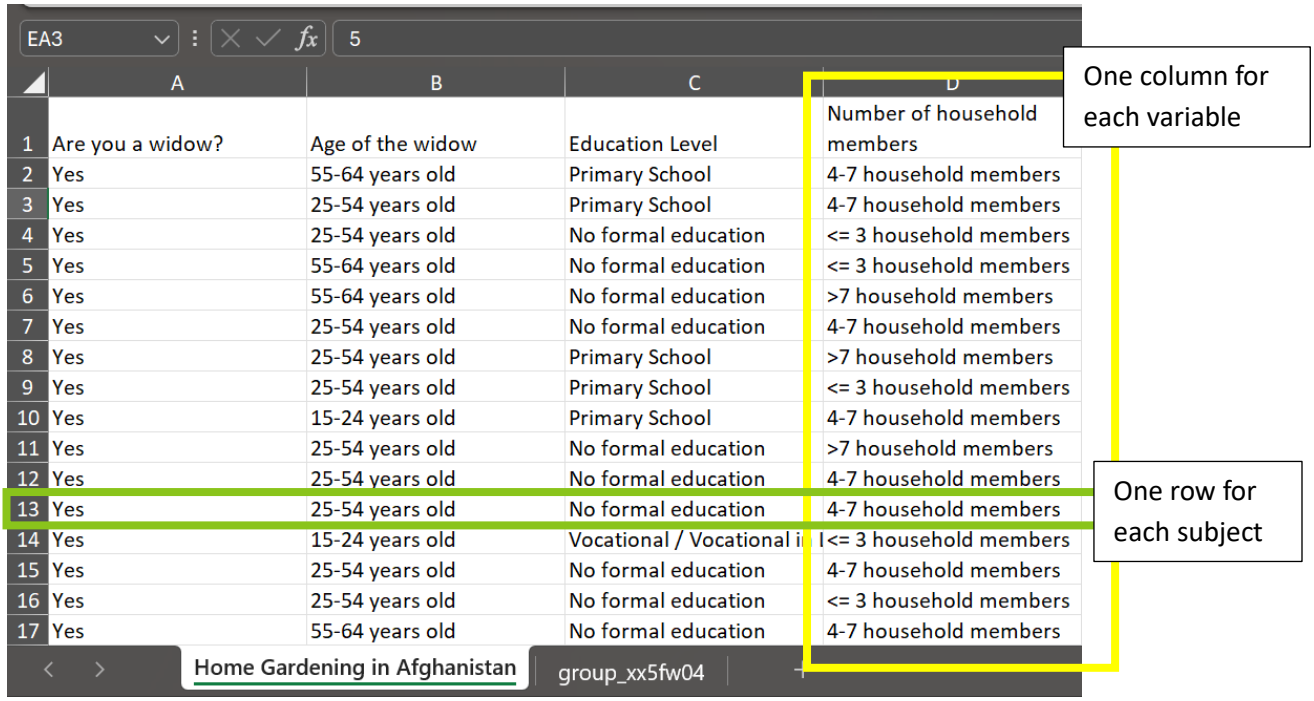


Figure 5: The structure of our data in Microsoft Excel

2 Data cleaning and coding

In In this chapter, we will learn about how to perform data cleaning and data coding

Data Cleaning

In almost all research, errors can occur. The errors occur due to human factors, such as misunderstanding the question, missing answers, filling out the answer in the wrong box, or providing answer that is out of expected range, etc. Through data cleaning, one can screen, diagnose, and edit the data abnormalities. When you modify your data, you should be very careful that changes should only be made when you are certain, such as in instances where an enumerator reported an incorrect response.

Whenever we start data analysis, make a duplicate of the original file. It is important to keep the original file unchanged to ensure that we always have a backup.

Ways to Perform Data Cleaning

1. Editing variables

   This data cleaning involves editing variables in ways that make sense. In the following example, we recode the variable from the original name and hide some columns that are less important for the analysis (Figure 6).

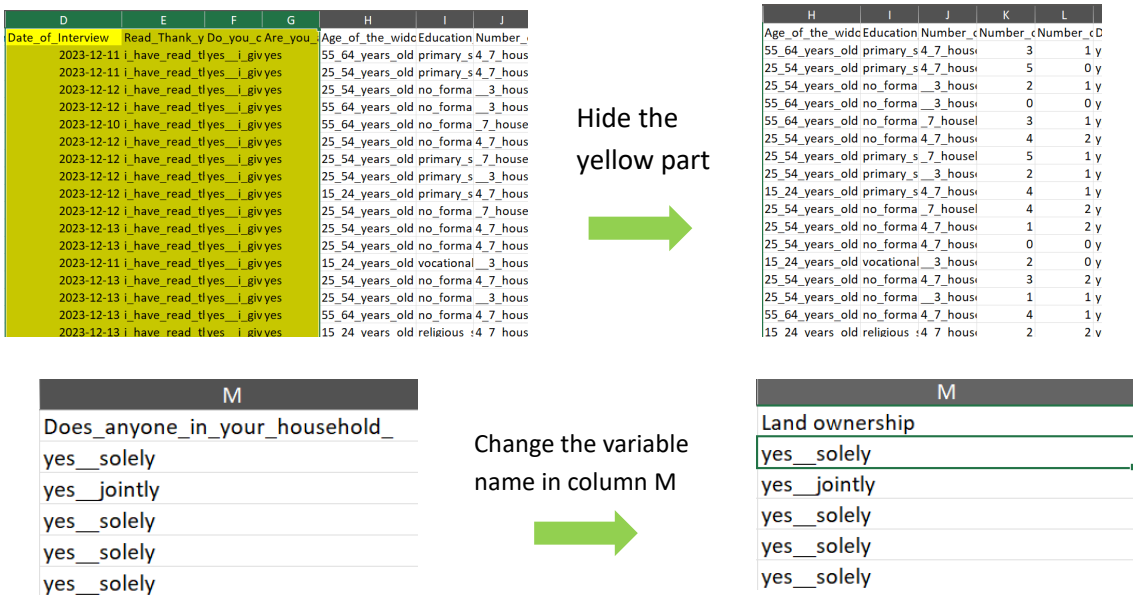


Figure 6: Editing variables

2. Missing data

   Missing value could happen due to error in the questionnaire flow or the unavailability of the information. To check missing data (Figure 7), first, select the whole worksheet or data r 'Styles group'. Choose 'New Rule'.

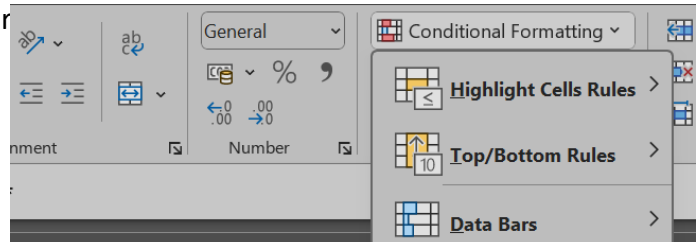


Figure 7: Checking missing data (1)

Click on ‘Use a formula to determine which cells to format’ (Figure 8). Then, enter the formula ‘=ISBLANK(A1)’ (A1 is the first cell in the selected range). Click on ‘Format’, choose a fill color or font color to highlight cells with missing data, then clik ‘OK’.

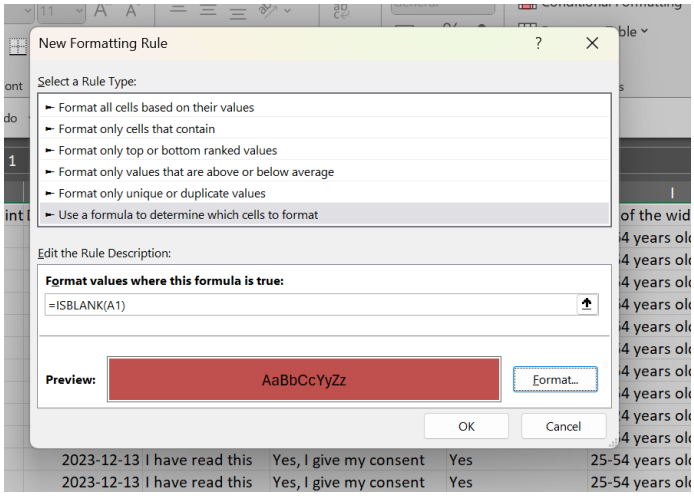


Figure 8: Checking missing data (2)

In the figure below, cells with missing data are highlighted in red (Figure 9). The cell marked with ‘a’ has missing answers and therefore is not useful for analysis. This occurred due to an error in the questionnaire (note: we fixed this problem, but you may need to check whether this type of error occurs in the future)

While the cell marked with ‘b’ has missing answers, but it occurred due to different reason. This happened because in the previous question ("Do you have any livestock at your household?"), respondents answered ‘No’, so there was no answer that needed to be given to the question "What is the livestock you have?"). In this case, it is fine to have missing answers. Meanwhile, respondents who answered ‘Yes’, they gave further answers, such as 'Poultry, Goat'. This condition will appear several times in this project dataset.

H	I
Are_you_a_widow	Age_of_the_widow
1	4
1	3
1	3
1	4
1	4
1	3
1	3

Data

training do you got?/Value training	Do you have any livestock at your household?	Wh live ha					is the ock you /Cattle/Cow	What is the livestock you have?/Other	'other' livestock you have?
	Yes	Po					0	0	
	Yes	Po					0	0	
	No								
	No								
	Yes	Poultry	1	0	0	0	0		
	No								
	Yes	Goat Poultry	1	1	0	0	0		
	Yes	Poultry	1	0	0	0	0		
	Yes	Poultry	1	0	0	0	0		
	Yes	Goat	0	1	0	0	0		

a

b

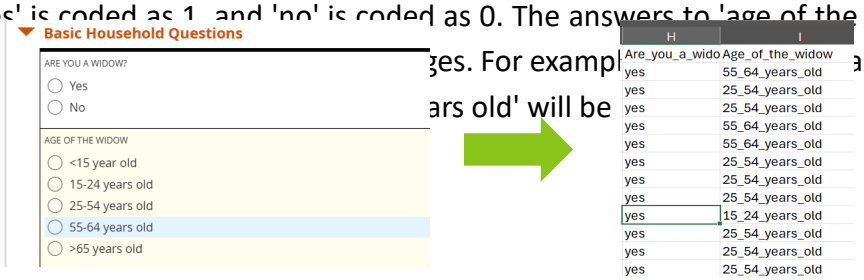
Categ level, howe o code the c nnaire, the a tant to understand the coding used for the different types of questions included in the questionnaire.

1. One answer from various choices (categorical variable)

   In the Basic Household section, you will find these two questions (Figure 10): 'Are you a widow?' with two possible answers, 'Yes' or 'No,' and 'Age of the widow' with five answer choices ranging from '<15 years old' to '> 65 years old.' The current results from Kobo are displayed as shown, but in the updated questionnaire, the results will be in numerical form, as depicted in Figure 11.

Figure 10: One answer from various choices

The answer 'yes' is coded as 1, and 'no' is coded as 0. The answers to 'age of the widow' are coded from 1 t ges. For exampl answers '25-54 years old,' t ars old' will be



H	I
Are_you_a_widow	Age_of_the_widow
yes	55_64_years_old
yes	25_54_years_old
yes	25_54_years_old
yes	55_64_years_old
yes	55_64_years_old
yes	25_54_years_old
yes	25_54_years_old
yes	25_54_years_old
yes	15_24_years_old
yes	25_54_years_old
yes	25_54_years_old
yes	25_54_years_old

2. Selecting multiple answers from various options (categorical variable)

In Figure 12 below, you can see the question 'Which months were there food shortages in the last year?' in the Kobo questionnaire, along with how the results appear in the Excel sheet. For this question, participants could select multiple answers. In column LG, you will find the selected months, and in columns LH to LS, you will see the corresponding number for each month. For example, the first participant selected 'January, February, March.' In the following columns, LH to LJ, January, February, and March are coded as 1, while the remaining months are coded as 0.

Figure 11: Coded answers

Figure 12: Multiple answers from various options

Which_months_were_th_ges_in_the_last_year	Which_months_were_th_ges_in_the_last_year/january	Which_months_were_th_ges_in_the_last_year/february	Which_months_were_th_ges_in_the_last_year/march	Which_months_were_th_ges_in_the_last_year/april	Which_months_were_th_ges_in_the_last_year/may	Which_months_were_th_ges_in_the_last_year/june
january february march	1	1	1	0	0	0
august september october	0	0	0	0	0	0
january november december	1	0	0	0	0	0
august september october	0	0	0	0	0	0
january february march april may june	1	1	1	1	1	1

3. Open question: filling out number (numerical variable)

In Figure 13 below, you can see the questions ‘Number of children’ and ‘Number of elderly and people who need care’ in the Kobo questionnaire, along with how the results appear in the Excel sheet. For these open-ended questions, participants provided numerical answers. In this case, the answers are already in numerical form, so data coding was not necessary.

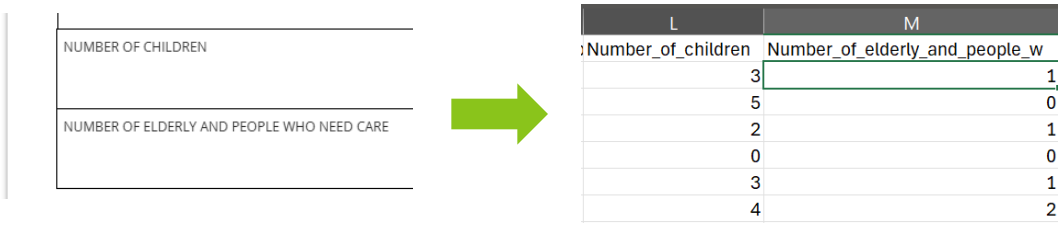
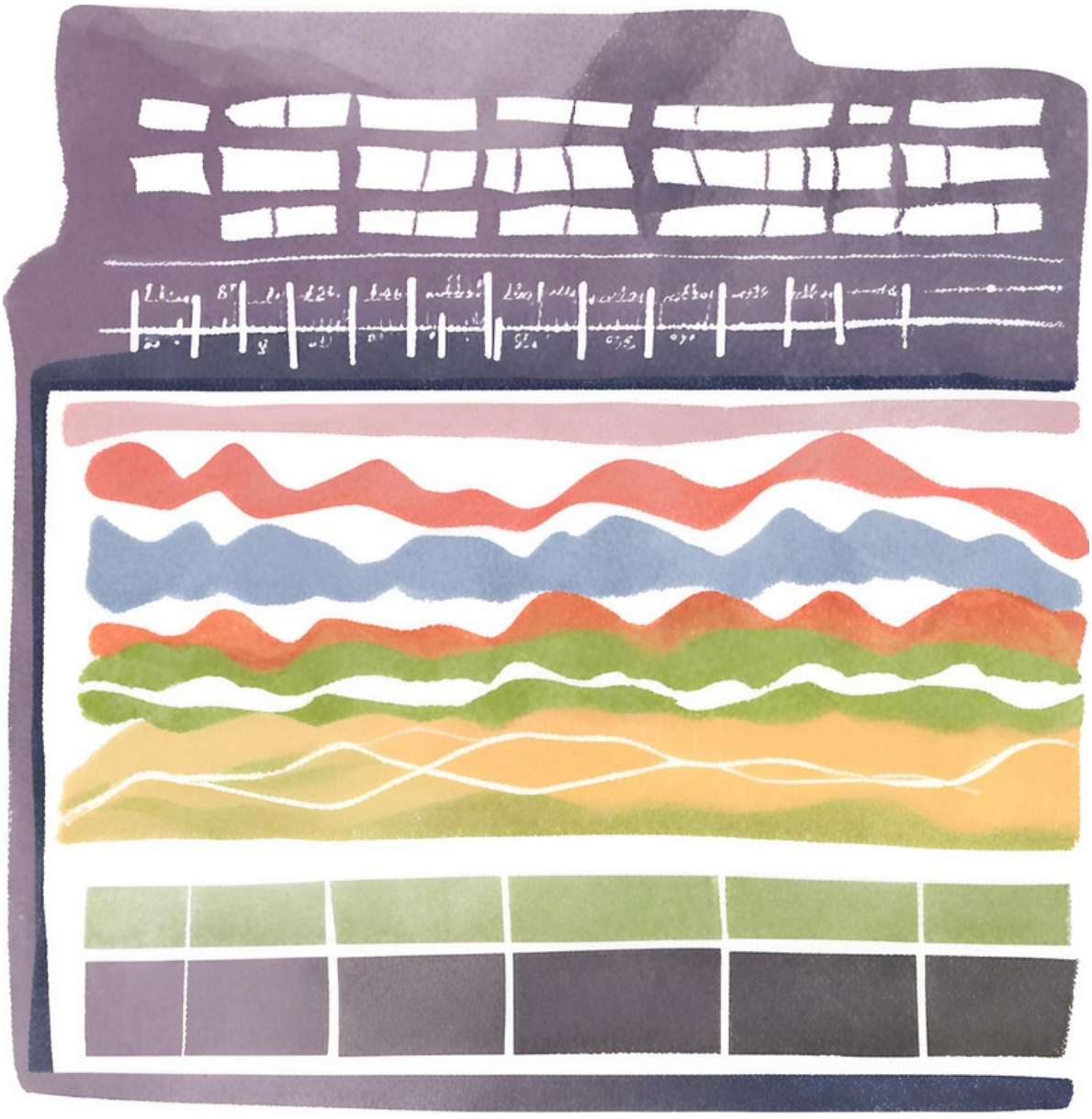


Figure 13: Open question

3 Data analysis



In this chapter, we will learn about:

- Descriptive analysis
- Comparison analysis

Introduction

In this chapter, we will learn about descriptive analysis and comparative analysis. Descriptive statistics are statistical methods that organize and describe the characteristics of collected data in a meaningful way. While it does not involve making comparison or predictions, descriptive analysis is crucial for identifying important characteristics of the data that can inform further analysis and decision-making. In the first chapter, we learned about several terminologies, including measures of central tendency, measures of variability, and measures of distribution shape. When we perform descriptive analysis, we describe our data characteristics using these measures. For example, if we want to know the most frequently planted crops, we can determine the mode. And if we want to know the average score of food diversity score, we can calculate the mean.

After we summarize our data using descriptive analysis, the next steps typically involve making comparison, prediction, or checking relationships between datasets. In our study, we aim to conduct comparative analysis, which is defined as comparing the outcomes or characteristics of two groups from two interventions. For example, we observe that the study population not receiving the home garden intervention has a slightly lower dietary diversity score compared to those who did. We conduct a comparative analysis to determine whether this difference reflects a genuine effect or is merely due to chance, such as accidentally sampling individuals who already consume a wider variety of food groups, even without the intervention.

Meeting the assumption

Before we learn the type of analysis statistic, we need to understand two assumptions that underlie the use of certain analysis statistic. In general, these two assumptions are applicable to parametric statistical test:

1. The Independence Assumption
   This assumption means that the samples are independent, one sample having no influence on the other. This is valid when correct randomization is applied, such as when a properly randomized sample is taken from the population. Independence can be checked by understanding the set-up of the study that was carried out.
2. Data is Normally Distributed
   Both populations are normally distributed, meaning that the data follows a normal distribution curve. The normal distribution curve is symmetrical or bell-shaped, with the mean located at the centre of the curve, which is also the same as the median. Conversely, if the data is not normally distributed, its distribution may be skewed to the left or right, or it may have a uniform distribution (Figure 14).

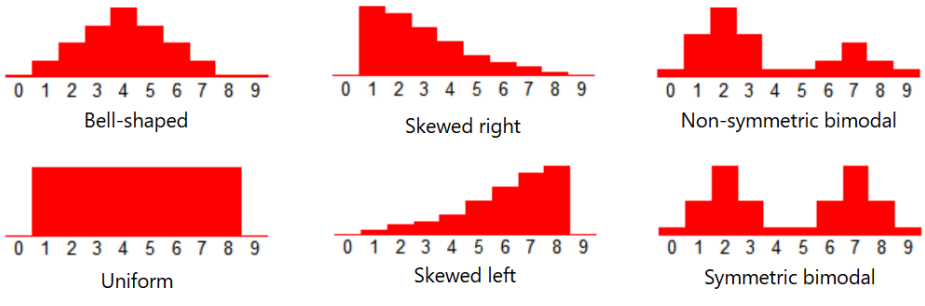


Figure 14: Distribution and shape (Berman et al, 2024)

To check for normal distribution, we can create a histogram, as depicted in Figure 14. However, a more sensitive method is to use a normal quantile plot. If the data set produces a straight line on a normal plot, it indicates a normal distribution. If the data are not normally distributed, the plot will deviate from linearity. We will learn more about how to create this plot later.

Understanding research hypotheses and significance level

When conducting research, we often aim to determine whether there are differences between two or more variables and formulate hypotheses based on these differences. There are two types of hypotheses: null hypothesis (H_0) and alternative hypothesis (H_a). The null hypothesis is a starting claim that assumes that there is no difference between two groups, which we hold until we can prove that there is a difference between two groups due to effect of intervention. In contrast, the alternative hypothesis represents our research hypothesis in which we make the claim about two groups that we are trying to find the evidence for. In many cases, we aim to reject the null hypothesis in favor of the alternative hypothesis.

We can state our hypotheses in two different ways: two-sided and one-sided. In a two-sided hypothesis, we have no specific expectation regarding whether one variable is higher or lower than the other. For example, in this project, the null and alternative hypotheses for a two-sided hypothesis are as follows:

- H_0: *There is no difference* in the mean Food Diversity Group Score (FDGS) between people who participated in the home garden intervention and those who did not.
- H_a: *There is a difference* in the mean FDGS between people who participated in the home garden intervention and those who did not.

For a one-sided hypothesis, the hypotheses can be stated as:

- H_0: *There is no difference* in the mean FDGS between people who participated in the home garden intervention and those who did not.
- H_a: The mean FDGS for people who participated in the home garden intervention *is higher than the* mean score for those who did not.

In the case of one-sided hypothesis, we favor the alternative hypothesis which posits that the mean FDGS is 'significantly higher.' By 'significant,' we mean that any observed difference between the two groups is likely due to a systematic influence rather than random chance. Specifically, we want to attribute this influence to the home garden intervention. However, we

cannot be absolutely certain that the effect is entirely due to the home garden. Statistical tests involve a degree of error or chance that cannot be eliminated, which is quantified by the significance level. For example, a significance level of 0.05 ($p < 0.05$) indicates there is a 5% chance that the observed differences are due to factors other than the hypothesized intervention, such as random variability or unknown influences. Therefore, a lower significance level corresponds to a lower probability of these competing or unknown factors affecting the observed differences.

Deciding Type of Analysis

There are various types of statistical analysis. However, in this chapter, we will focus on descriptive analysis and comparative analysis, both of which can be applied to our dataset. Table 5 below helps determine the appropriate type of analysis based on the data we have. The sections are organized according to the order of questions in Kobo, with the numbering of questions restarting from 1 in each section. Descriptive analysis will be applied to all sections, but comparative analysis will only be necessary if you have hypothesized that two groups would yield different outcomes. For example, if we hypothesize that dietary diversity improves due to the home garden intervention, we should perform a comparative analysis to make a valid claim.

To illustrate how to use the table, let's take the food consumption section as an example. This section contains two types of data: categorical and numerical (see definition of this in Chapter 1). Categorical data are found in questions 1, 2, 4, and 5 of the questionnaire, while numerical data is in question 3. For categorical data, we can perform descriptive analysis, which includes calculating frequency and percentage. Additionally, we can perform chi-square to compare the frequency of two groups. For the numerical data, we can calculate the mean and then perform a t-test to compare the mean of Months of Adequate Household Food Provisioning (MAHFP) score.

Table 5: Table to decide analysis methods

Section	Type of data	Question number	Descriptive analysis	Comparative analysis***
Basic household	Categorical	1, 2, 3, 4, 6	Frequency and percentage	
	Numerical	5, 7	Mean	
Pre-questions – Home garden production	Categorical	All	Frequency and percentage	
Home garden production	Categorical	All	Frequency and percentage	
Crops in the past years	Categorical	1, 2, 4 (each crop)	Frequency and percentage	
	Numerical	3	Mean	

Home garden challenges	Categorical	All (1 - 26)	Frequency and percentage	
Market access	Categorical	1, 2	Frequency and percentage	
	Numerical	5	Mean	
Household Source	Categorical	1, 2	Frequency and percentage	
Food Consumption*	Categorical	1, 2, 4, 5	Frequency and percentage	
	Numerical	3	Mean of MAHFP	T-test for MAHFP
Hygiene	Categorical	All	Frequency and percentage	
In the past 4 weeks…**	Numerical	All	Average HFIAS score, HFIA prevalence	T-test for average HFIA score, Chi-square test for HFIA prevalence
The Diet Quality Questionnaire (DQQ) for Women	Numerical	All	Mean of FDGS, percentage of MDD-W	T-test for FGDS, Chi-square test for MDD-W percentage

Note:

**Food consumption* -> In this section, you will find question number 3, which is used to derive the Months of Adequate Household Food Provisioning (MAHFP) indicator.

**In the past 4 weeks…* -> This section will have this title and contain questions to derive indicators of the Household Food Insecurity Access Scale (HFIAS).

**Comparative analysis* -> In this table, we provide parametric tests to perform with the assumption that conditions are met. Since not all statistical tests are covered in the training, and data analysis often requires consideration of many aspects, we advise consulting experts or statisticians if you encounter difficulties in deciding and performing comparative analysis.

Descriptive Analysis

In general, we can perform three types of descriptive statistics, including measures of central tendency, measures of variability, and measures of distribution shape. In this section, we will provide various example and explore different formulas in Excel that can be used for descriptive statistics.

Percentage of a variable

Example: the age proportion of the widows

When we looked into the questionnaire, we asked for different age groups. The age of widows is categorized into five age groups: (1) <15 years old, (2) 15-24 years old, (3) 25-54 years old, (4) 55-64 years old, (5) >65 years old. This categorical variable was already converted into numerical values (1, 2, 3, 4, 5).

To count how many participants belong to each group. In an empty cell, list the five categories and then, in the next empty cell, use this following formula:

=COUNTIF(C2:C151;1)

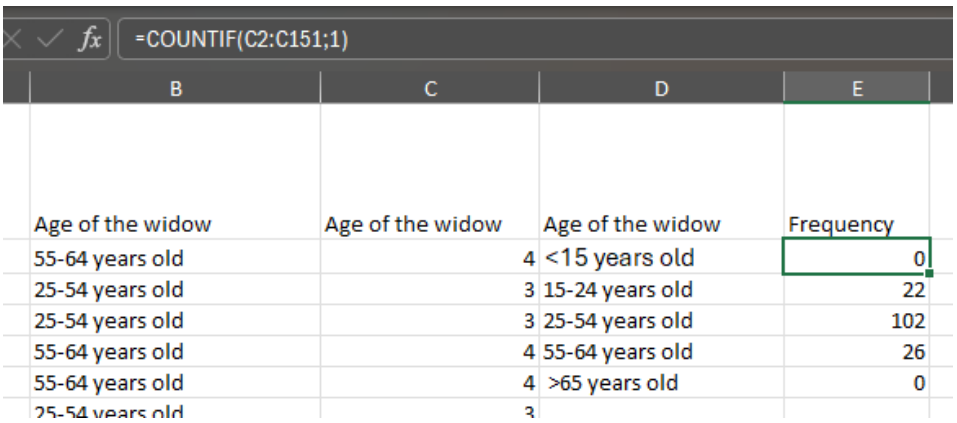


Figure 15: Count frequency for each category

Explanation of the formula:

For each category, enter the formula in a cell and change the number to the corresponding code. For example, for the age 15-24 years old, use 2 instead of 1: =COUNTIF(C2:C151;2)

After we have the frequency of each category, we could calculate the percentage of each category using this following formula (Figure 16):

=E3/(SUM(E2:E6))

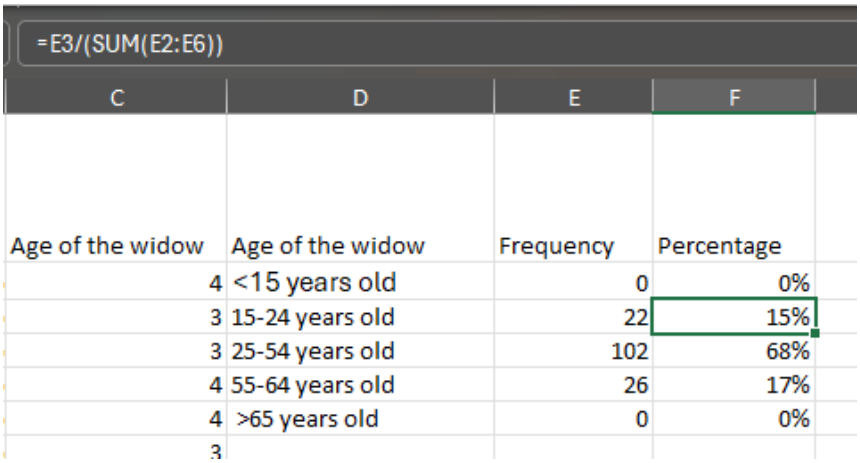
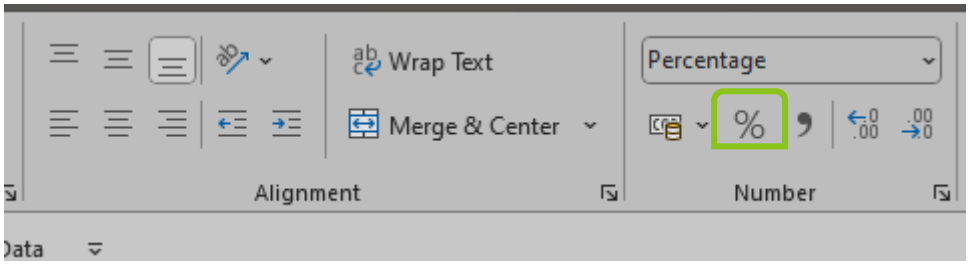


Figure 16: Calculate percentage for each category

Then change the data type into percentage by clicking on this icon:



Or we can use this following formula:
=E3/(SUM(E2:E6)*100)

Explanation of the formula:

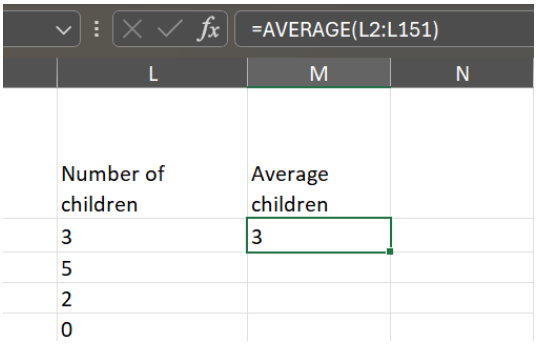
For each category, we need to adjust the formula. For example, for the age 25-54 years old, we change E3 to E4, thus the formula becomes =E4/(SUM(E2:E6)*100)

Mean of a variable

Example: Number of children

When we looked into the questionnaire, we asked for "how many children in each household". Since the number of children is a numerical variable, there is no need to recode it into numbers. To calculate the average (Figure 17), create a new cell next to column 'J' cell. In the new cell, enter the following formula:

=AVERAGE(L2:L151)



Explanation of the formula:

Range 'L2:L151' included the group of cells in column L that contains the data that we want to know the average. This range can be adjusted according to the data location.

Figure 17: Calculate mean

Example: Calculating the score of Months of Adequate Household Food Provisioning (MAHFP)

In this project, we will determine some indicators that can be derived from combination of questions. In this section, we will learn about how to get the Months of Adequate Household Food Provisioning (MAHFP) indicator from series of questions.

The MAHFP is an indicator used to measure household food access over time. It is typically measured by asking households if there were any months in the past year when they did not have enough food to meet the family's needs. To calculate the MAHFP score in Excel, we first need to understand the structure of the data (Figure 18).

	A	B	C	D	E	F	G	H
1	Is there a time of year when there is less food/food shortage compared to other times?	How many months in a year did you face food shortage?	Which months were their food shortages in the last year?	Which months were their food shortages in the last year?/January	Which months were their food shortages in the last year?/February	Which months were their food shortages in the last year?/March	Which months were their food shortages in the last year?/April	Which months were their food shortages in the last year?/May
2	Yes	3	January February March	1	1	1	0	0
3	Yes	3	August September October	0	0	0	0	0
4	Yes	3	January November December	1	0	0	0	0
5	Yes	3	August September October	0	0	0	0	0
6	Yes	6	January February March April May June	1	1	1	1	1
7	No							
8	No							
9	Yes	3	October November December	0	0	0	0	0
10	Yes	1	December	0	0	0	0	0

Figure 18: Structure of MAHFP data in Excel sheet

In column A, each subject answered, 'Yes' or 'No'. Only those who answered 'Yes' continued to the next question, which is why there are empty cells in column B. In column C, each respondent indicated which months they experienced food shortages, resulting in multiple answers. In column D through O, we can see whether there was period of food shortage for each month based on the answers in column B. For example, for the first subject, she

answered ‘January February March’, which is why we see a ‘1’ in column D, E, F while column G and H shows a ‘0’.

Before calculating the MAHFP score, insert a new column next to column B and named it as ‘MAHFP’ (Figure 19). In column C2, calculate MAHFP score (0-12) using the following formula:

=12-B2

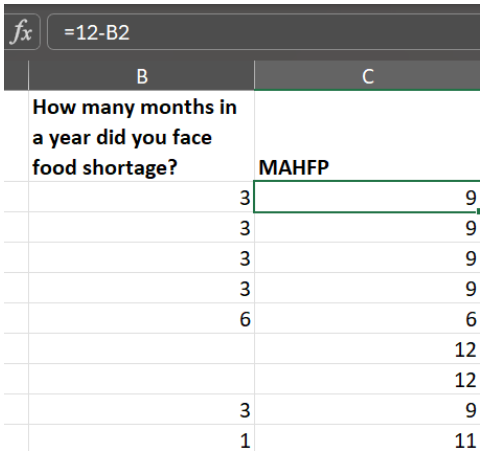


Figure 19: Calculate MAHFP score

Then click on the small square at the bottom-right corner of the cell in C2, this will apply the formula to all the cells in column C. To calculate the average MAHFP score of the households, we can use the previous ‘AVERAGE’ formula:

=AVERAGE(D2:D151)

Explanation of the formula:

In column C, we can see the MAHFP score for each household. For instance, in cell C2, the MAHFP score is 9, indicating that this household experienced some level of inadequate food access in the remaining months of the year (i.e. 3 months).

For further information on MAHFP score interpretation, you can read the MAHFP for Measurement of Household Food Access: Indicator Guide (Bilinsky and Swindale, 2010).

Analysing Dietary Quality Questionnaire (DQQ)

The DQQ is a monitoring tool for diet quality at population level, suitable for characterizing diet patterns and tracking trends over time. This questionnaire is mainly used to calculate Minimum Dietary Diversity for Women (MDD-W) or Dietary Diversity Score (DDS). Additionally, other new indicators such as the non-communicable disease-Protect Score (NCD-Protect), NCD-Risk, and The Global Dietary Recommendations (GDR) Score are also recommended for calculation. For further information on DQQ, you can read the guideline book (Auliani, et al., 2023) and this website Global Diet Quality Project.

Calculating MDD-W/DDS

Before calculating the MDD-W/DDS, we first need to be familiar with the structure of the data. The DQQ questionnaire contains 29 questions. In the following picture, we can see the answers for each question in a column, starting from column A. The number 1 indicates 'Yes', while the number 0 indicates 'No'.

MDD-W Food Groups	Question numbers	Possible points
Grains, white roots, and plantains	1, 2, 3	1
Pulses (beans, peas and lentils)	4	1
Nuts and seeds	21	1
Dairy	14, 15, 25	1
Meat, poultry, and fish	16, 17, 18, 19, 20	1
Eggs	13	1
Dark green leafy vegetables	6*	1
Other vitamin A-rich fruit and vegetables	5, 8	1
Other vegetables	7*	1
Other fruits	9, 10*	1
TOTAL		=SUM (0-10)

*Note that these groups may be asked in two questions (e.g. 6.1 and 6.2).

1	or Asafe	millets	parsnips	beans	pumpkins	sabar	beans	gourd	mango	oranges	melons	peaches	or nafwa	mimas	Eggs	quroots	doogha
2	1	0	1	1	1	1	0	0	0	0	1	0	0	0	1	1	1
3	0	1	1	0	1	0	1	1	1	0	1	0	0	0	1	1	0
4	1	1	1	1	0	1	0	1	0	0	0	0	0	0	1	1	0
5	0	1	0	1	1	1	0	1	1	1	0	0	0	0	1	0	1

Figure 20: Structure of answers to DQQ

To calculate MDD-W or DDS, we need to sum all the points in the table below (Table 6). The maximum score is 10. In the Excel sheet, next to column AD (the last question of DQQ), make a list of the ten food groups (Figure 21).

	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN
1	KFC, or any place that serves pizza or burgers?	Grains, white roots, and plantains	Pulses	Nuts and seeds	Dairy	Meat, poultry, and fish	Eggs	Dark green leafy vegetables	Other vitamin A-rich fruit and vegetables	Other vegetables	Other fruits
2	0										
5	0										

Figure 21: Summing the points

For some of the food groups, we asked in multiple questions in the questionnaire. For example, in the table below (Table 6) you can see the ten food group with the first food group is grain, white roots, and plantains. In the questionnaire, question number 1, 2, 3 belong to this first food group. Whereas, for pulses (the second food group), only question number 4 belong to this food group.

Table 6: MDDW/DDS calculation (Auliani et al., 2023)

For each food group in the corresponding column enter the following formula (Table 7):

Table 7: Excel formula for each food group

Nr	Food Groups	Formula
1	Grains, white roots, and plantains	=IF(OR(A2=1; B2=1; C2=1); 1; 0)
2	Pulses	=IF(D2=1;1;0)
3	Nuts and seeds	=IF(V2=1;1;0)
4	Dairy	=IF(OR(P2=1; Q2=1; Z2=1); 1; 0)
5	Meat, poultry, and fish	=IF(OR(R2=1; S2=1; T2=1;U2=1); 1; 0)
6	Eggs	=IF(O2=1; 1; 0)
7	Dark green leafy vegetables	=IF(F2=1; 1; 0)
8	Other vitamin A-rich fruit and vegetables	=IF(OR(E2=1; I2=1); 1; 0)
9	Other vegetables	=IF(OR(G2=1; H2=1); 1; 0)
10	Other fruits	=IF(OR(J2=1; K2=1;L2=1); 1; 0)

Explanation of the formula:

The IF function in Excel performs a logical test, resulting one value if the test is TRUE and another value if the test is FALSE.

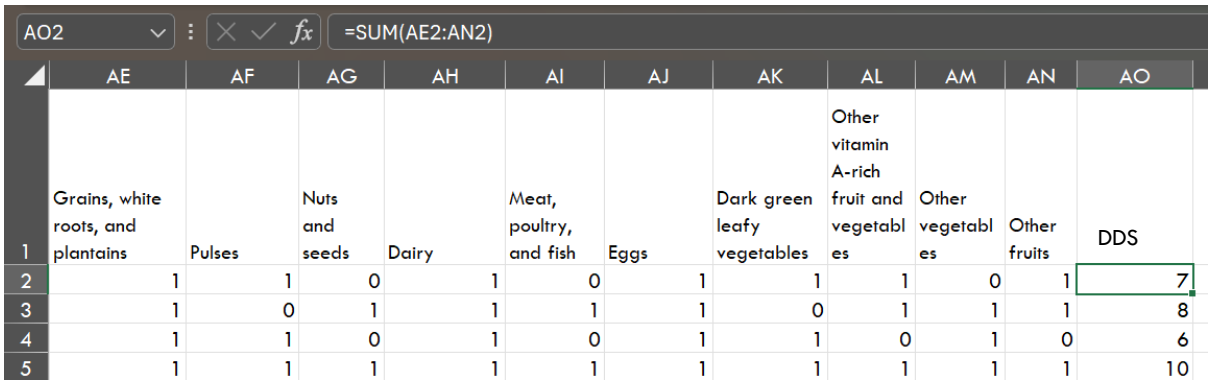
=IF(logical_test, value_if_true, value_if_false)

The OR function is part of the logical test, testing whether any one of the conditions inside the bracket is TRUE. For example, the first food group with formula '(OR(A2=1; B2=1; C2=1)' checks if at least one of the cells in A2, B2, or C2 contains the value 1. It checks whether the subject at least consumed one of the foods that belonged to the first food groups. If cell A2 contains number 1, then the condition is TRUE and the value in column AE will be 1 (i.e., the subject consumed naan or lavash, paratha, rice, macaroni, or aash in the previous day). However, if cell A2 contains the number 0, then the condition is FALSE, and the value in column AE will be 0. This also applies if B2 or C2 contains number 1. This means that respondent did not consume any of the foods asked for in the questionnaire in the previous day.

Another example is the second food group with the formula '=IF(D2=1;1;0)'. 'D2=1' is the logical test that checks whether the value in cell D2 is equal to 1. If the condition is TRUE, then the value in column AF will be 1. In this case, we don't use the OR function because we only need to check one condition.

After calculating the points of each food group, insert a new column. Then, we can calculate the FDGS for each subject by summing the points from all food groups using the following formula (Figure 22):

=SUM(AE2:AN2)



	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN	AO
1	Grains, white roots, and plantains	Pulses	Nuts and seeds	Dairy	Meat, poultry, and fish	Eggs	Dark green leafy vegetables	Other vitamin A-rich fruit and vegetables	Other vegetables	Other fruits	DDS
2	1	1	0	1	0	1	1	1	0	1	7
3	1	0	1	1	1	1	0	1	1	1	8
4	1	1	0	1	0	1	1	0	1	0	6
5	1	1	1	1	1	1	1	1	1	1	10

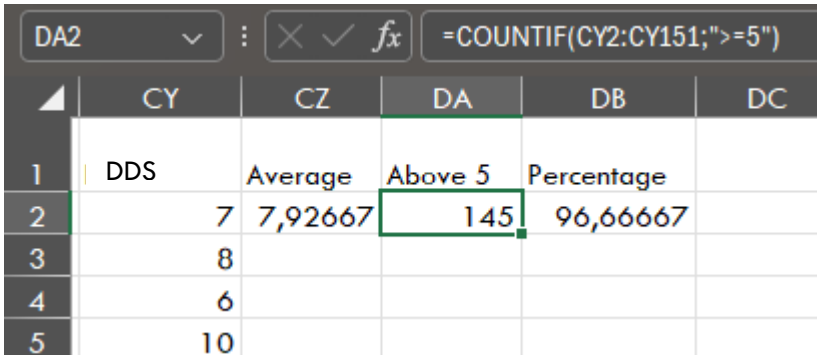
Figure 22: Sum of DDS for each participant

To calculate the average of FGDS score for the population, we can use the following formula:

=AVERAGE(CY2:CY151)

The MDD-W is expressed as the percent of the population consuming at least 5 out of 10 food groups. To calculate this, we can use the COUNTIF function as follows (Figure 23):

=COUNTIF(CY2:CY151;">=5")



	CY	CZ	DA	DB	DC
1	DDS	Average	Above 5	Percentage	
2	7	7,92667	145	96,66667	
3	8				
4	6				
5	10				

Figure 23: Calculate MDD-W

Explanation of the formula:

'CY2:CY151' is the range for the FDGS score for each subject. Then, '>=5' means that we want to count cells (i.e. within the selected range) where the total is at least 5. After we have

the number of subjects with FDGS score equal or above 5, we can calculate the percentage of subjects with FDGS score equal or above 5 using the following formula:

=DA2/150*100

In the above formula, we already know the total number of participants. However, if we don't know the total number of participants, we can calculate it using the COUNT function:

=COUNT(CY2:CY151)

The last step is to calculate the proportion of each food group (Figure 24). In the empty cells, lists the ten food groups. Then, sum the points of each food group using the SUM function: =SUM(CO2:CO151)

The above formula will sum the total point from the Grain, white roots, and plantains group. To get the total point from other food groups, adjust the formula corresponding to the cell range. After that, calculate the proportion with the formula:

=DF4/DC2*100

'DC2 refers to the cell where we counted the total number of participants. The dollar signs '$' make this an absolute reference, meaning that when we copy the formula to the cells below, the reference to DC2 will remain constant.

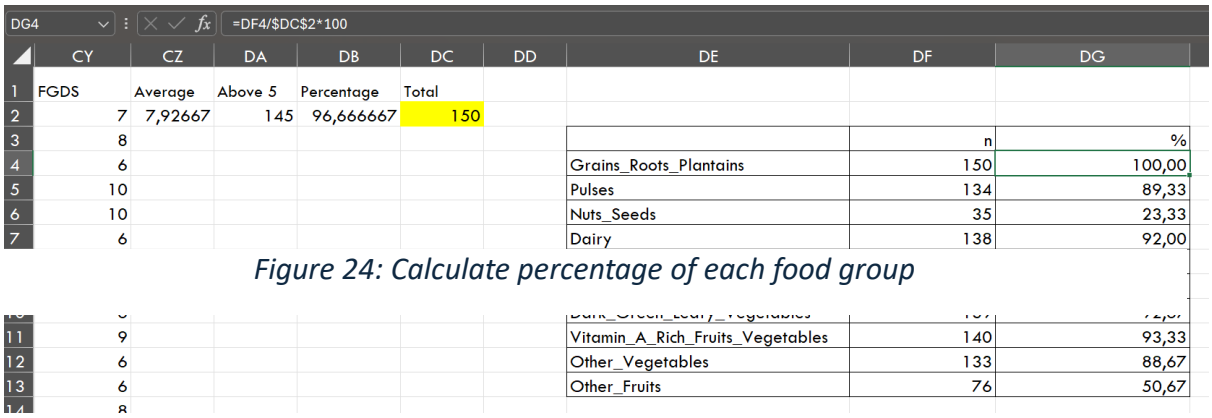


Figure 24: Calculate percentage of each food group

Analysing Household Food Insecurity Access Scale (HFIAS)

The HFIA questionnaire yields information on perceived food insecurity access to food at the household level. Two types of indicators will be calculated to help understand the characteristics of and changes in household food insecurity (access) in the surveyed population. These two indicators provide summary information on Household Food Insecurity Access Scale Score (HFIAS) and Household Food Insecurity Access Prevalence (HFIAP). For detailed explanation about HFIAS, refer to the guideline book (Auliani et al., 2023, Coates et al., 2007).

Household Food Insecurity Access Scale Score

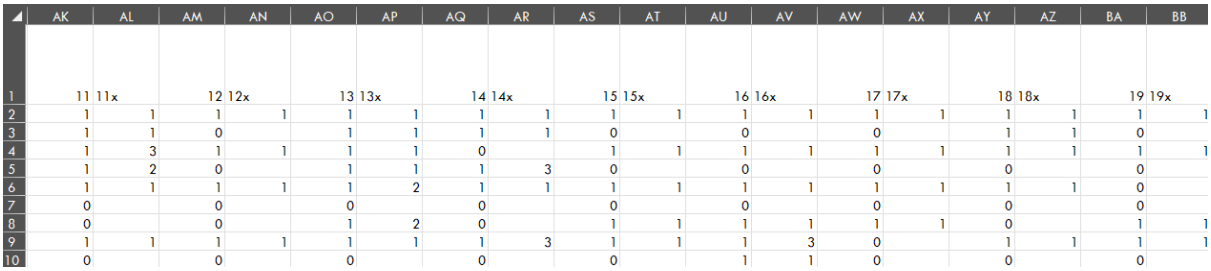
A HFIAS score variable is calculated for each household by summing the codes for each frequency-of-occurrence question. The maximum score for a household is 27 (the household response to all nine frequency-of-occurrence questions was "often", coded with response code of 3); the minimum score is 0 (the household responded "no" to all occurrence questions, frequency-of-occurrence questions were skipped by the interviewer, and subsequently coded as 0).

For example, for Q1 ("Did you worry that your household would not have enough food in the past four weeks?"), the respondent say Yes, then the score is 1. Then for the frequency of occurence (Q1a) the respondent answer "Sometimes", then the score is 2. If the respondent answer Q1 with No, then Q1a will automatically be 0 (Figure 25).

Q	1			a
Ser.no	Question	In the past 12 months	In the past four weeks	
		1-Yes 0-No 98-Don't Know 99-Refused	1-Yes 0-No 98-Don't Know 99-Refused If ≠ from 'yes' skip to next question	How often did [Question] happen in the past four weeks? 1-Rarely (once or twice in the past four weeks) 2-Sometimes (three to ten times in the past four weeks) 3-Often (more than ten times in the past four weeks)
1	Did you worry that your household would not have enough food?			

Figure 25: Frequency of occurrence for Q1a (Auliani et al., 2023)

Below you can see the data structure in the spreadsheet (Figure 26). We renamed each question to a number for easier calculation later. For example, columns AK represent question numbered 1 (i.e. in the past four weeks…), with responses of 'Yes' or 'No'. Column AL contain the follow-up question, with responses indicating the frequency of occurrence (Rarely, Sometimes, or Often).

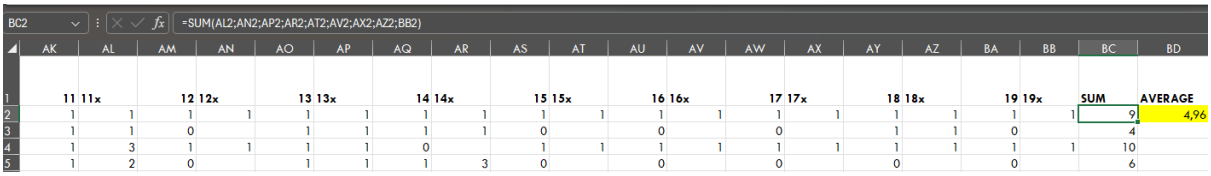


	AK	AL	AM	AN	AO	AP	AQ	AR	AS	AT	AU	AV	AW	AX	AY	AZ	BA	BB
1	11	11x	12	12x	13	13x	14	14x	15	15x	16	16x	17	17x	18	18x	19	19x
2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
3	1	1	0		1	1	1	1	0		0		0		1	1	0	
4	1	3	1	1	1	1	0		1	1	1	1	1	1	1	1	1	1
5	1	2	0		1	1	1	3	0		0		0		0		0	
6	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	0	
7	0		0		0		0		0		0		0		0		0	
8	0		0		1	2	0		1	1	1	1	1	1	0		1	1
9	1	1	1	1	1	1	1	3	1	1	1	3	0		1	1	1	1
10	0		0		0		0		0		1	1	0		0		0	

Figure 26: Structure of HFIAS data in spreadsheet

The HFIAS score is the sum of the frequency-of-occurrence during the past four weeks for the 9-food insecurity-related conditions. Using the SUM function, we can calculate the HFIAS score for each household by adding all columns with 'x':

=SUM(AL2;AN2;AP2;AR2;AT2;AV2;AX2;AZ2;BB2)



BC2 =SUM(AL2;AN2;AP2;AR2;AT2;AV2;AX2;AZ2;BB2)

	AK	AL	AM	AN	AO	AP	AQ	AR	AS	AT	AU	AV	AW	AX	AY	AZ	BA	BB	BC	BD
1	11	11x	12	12x	13	13x	14	14x	15	15x	16	16x	17	17x	18	18x	19	19x	SUM	AVERAGE
2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	9	4,96
3	1	1	0		1	1	1	1	0		0		0		1	1	0		4	
4	1	3	1	1	1	1	0		1	1	1	1	1	1	1	1	1	1	10	
5	1	2	0		1	1	1	3	0		0		0		0		0		6	

Figure 27: Sum of frequency of occurence

Next, the average HFIAS Score of the population is calculated using AVERAGE function:

=AVERAGE(BC2:BC151)

BC2:BC151 is the range where the HFIAS score of each household is located.

Household Food Insecurity Access Prevalence (HFIAP)

The HFIAP indicator categorizes households into four levels of household food insecurity (access): food secure (1), mildly (2), moderately (3) and severely food insecure (4) (Figure 28). To calculate the score for each category, we must meet the following criteria:

HFIAS category	⇒ HFIA category = 1 if [(Q1a=0 or Q1a=1) and Q2=0 and Q3=0 and Q4=0 and Q5=0 and Q6=0 and Q7=0 and Q8=0 and Q9=0]
	⇒ HFIA category = 2 if [(Q1a=2 or Q1a=3 or Q2a=1 or Q2a=2 or Q2a=3 or Q3a=1 or Q4a=1) and Q5=0 and Q6=0 and Q7=0 and Q8=0 and Q9=0]
	⇒ HFIA category = 3 if [(Q3a=2 or Q3a=3 or Q4a=2 or Q4a=3 or Q5a=1 or Q5a=2 or Q6a=1 or Q6a=2) and Q7=0 and Q8=0 and Q9=0]
	⇒ HFIA category = 4 if [Q5a=3 or Q6a=3 or Q7a=1 or Q7a=2 or Q7a=3 or Q8a=1 or Q8a=2 or Q8a=3 or Q9a=1 or Q9a=2 or Q9a=3]

Figure 28: HFIAS category of prevalence (Auliani et al., 2023)

These criteria are then translated into Excel formula as follows:

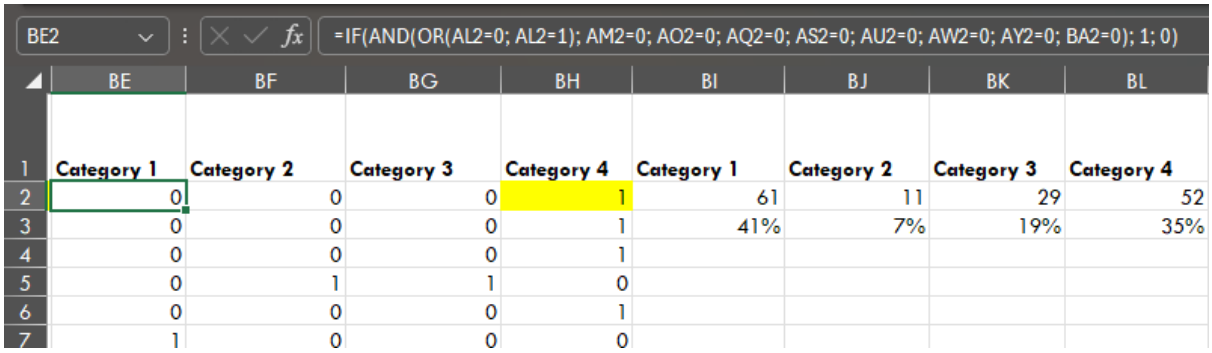
Category 1:
=IF(AND(OR(AL2=0; AL2=1); AM2=0; AO2=0; AQ2=0; AS2=0; AU2=0; AW2=0; AY2=0; BA2=0); 1; 0)
Category 2:
=IF(AND(OR(AL2=2; AL2=3; AN2=1; AN2=2; AN2=3; AP2=1; AR2=1); AS2=0; AU2=0; AW2=0; AY2=0; BA2=0); 1; 0)
Category 3:
=IF(AND(OR(AP2=2; AP2=3; AR2=2; AR2=3; AT2=1; AT2=2; AV2=1;AV2=2); AW2=0; AY2=0; BA2=0); 1; 0)
Category 4:
=IF(OR(AT2=3; AV2=3; AX2=1; AX2=2; AX2=3; AZ2=1; AZ2=2;AZ2=3;BB2=1;BB2=2;BB2=3); 1; 0)

Explanation of the formula:

OR(AL2=0; AL2=1): The OR function checks if the value in cell AL2 (represents Q1a) is either 0 or 1. If the condition met, the value will be TRUE or result in 1.

The AND function checks if the value in AM2 = 0 (represents Q2=0), AO2 = 0 (represents Q3=0), … etc.

In the figure 29, we can see that the first subject (column BH2) is categorized in category 4 (Severely Food Insecure Access), and the other subjects are categorized in different food insecure categories.



	BE	BF	BG	BH	BI	BJ	BK	BL
1	Category 1	Category 2	Category 3	Category 4	Category 1	Category 2	Category 3	Category 4
2	0	0	0	1	61	11	29	52
3	0	0	0	1	41%	7%	19%	35%
4	0	0	0	1				
5	0	1	1	0				
6	0	0	0	1				
7	1	0	0	0				

Figure 29: Category of households prevalence

To find the proportion of each category in the population, we can use the SUM function, then calculate the proportion using the formula:

=SUM(BE2:BE151)

=BI2/150*100

Comparative analysis

In our study, we will collect data from the participants of the second (baseline) and third stages of the study (endline). See the guidebook for detailed information about the stages of the study (Auliani et al., 2023). Using the collected datasets, we aim to compare whether the results for certain variables or indicators differ between participants in the first and second stages, or between those in the second and third stages, after the new home garden intervention is completed. For example, we want to assess whether participants who received home garden intervention at the second stage (baseline) have a different dietary diversity quality compared to those at the third stage (endline). We want to test whether any observed differences are due to the home garden intervention or are the result of random or unknown factors.

Choosing the appropriate statistical test

There are various types of comparative analysis, but we will only discuss those relevant to our study. The variables we want to compare include both numerical and categorical data, and therefore, the appropriate type of comparative analysis will differ. To decide which comparative analysis is applicable to our data, we can follow the flowchart on the next page (Figure 30).

Here is an example of how to use the table:

- What do you compare?
  Mean of FDGS score -> This is numerical data
- How many groups are you dealing with?
  Two
- Is the data meeting the assumption?
  Yes -> we choose the parametric test, t-test

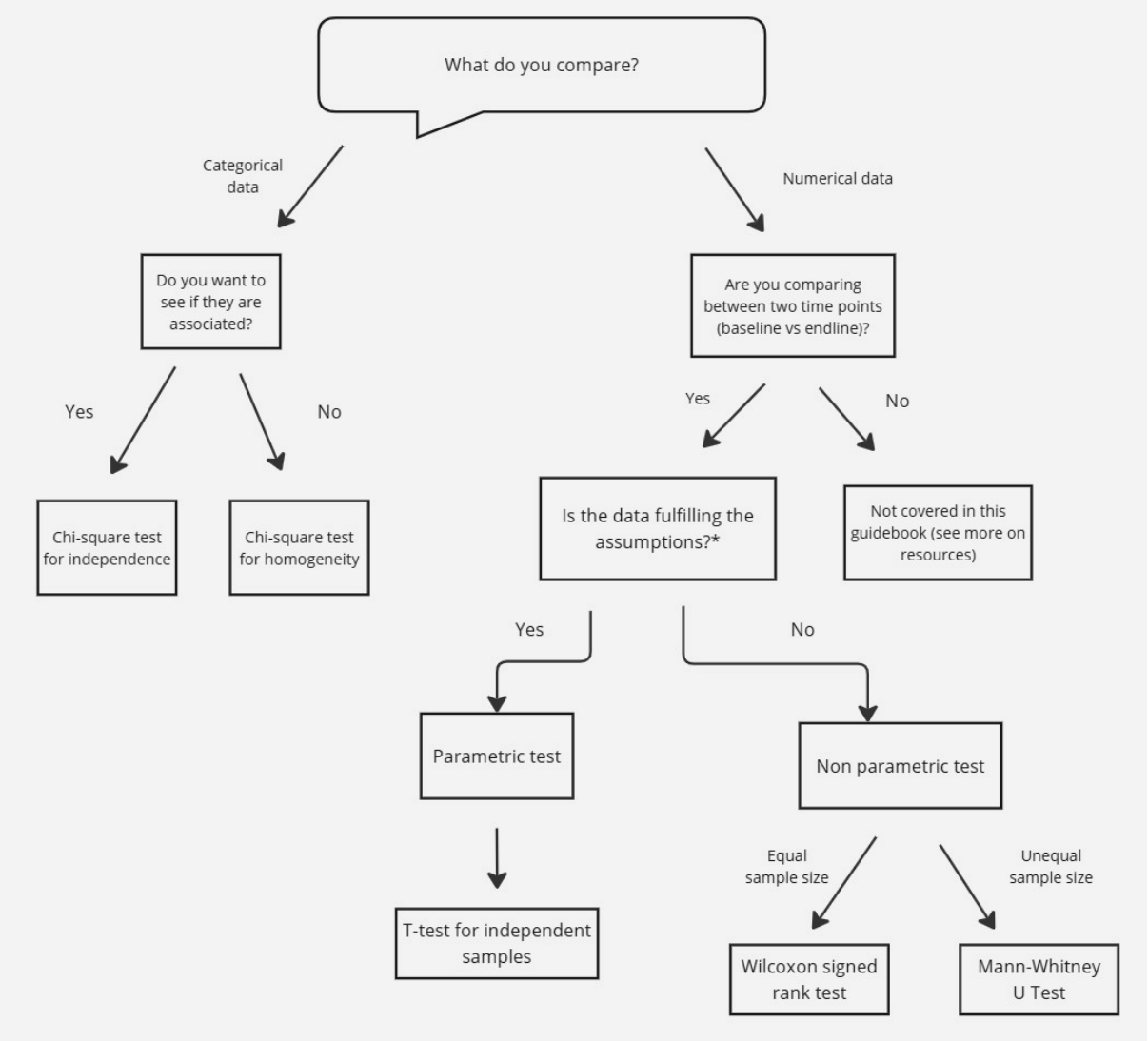


Figure 30: Choosing the appropriate statistical test

Steps of statistical test

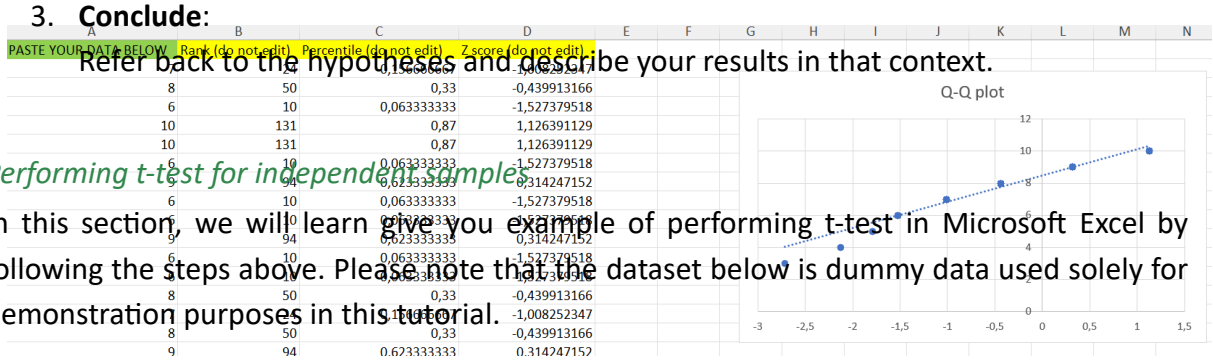
We have learned about statistical test assumptions, hypotheses, and the importance of choosing the right statistical test. In this section, we will compile that knowledge into steps for conducting statistical tests. Following these steps can help make the process more structured:

1. **Plan**:
   a. Define the parameter; in which areas would you expect to see changes as a result of the home garden project?
   b. State the null and alternative hypotheses.
2. **Solve**:
   a. Use the figure from earlier to determine which statistical test to use.
   b. Check whether the dataset meets the assumptions of the chosen test.
   c. Perform the test statistic in Excel.

d. Interpret the p-value.

3. **Conclude**:

Refer back to the hypotheses and describe your results in that context.



Performing t-test for independent samples

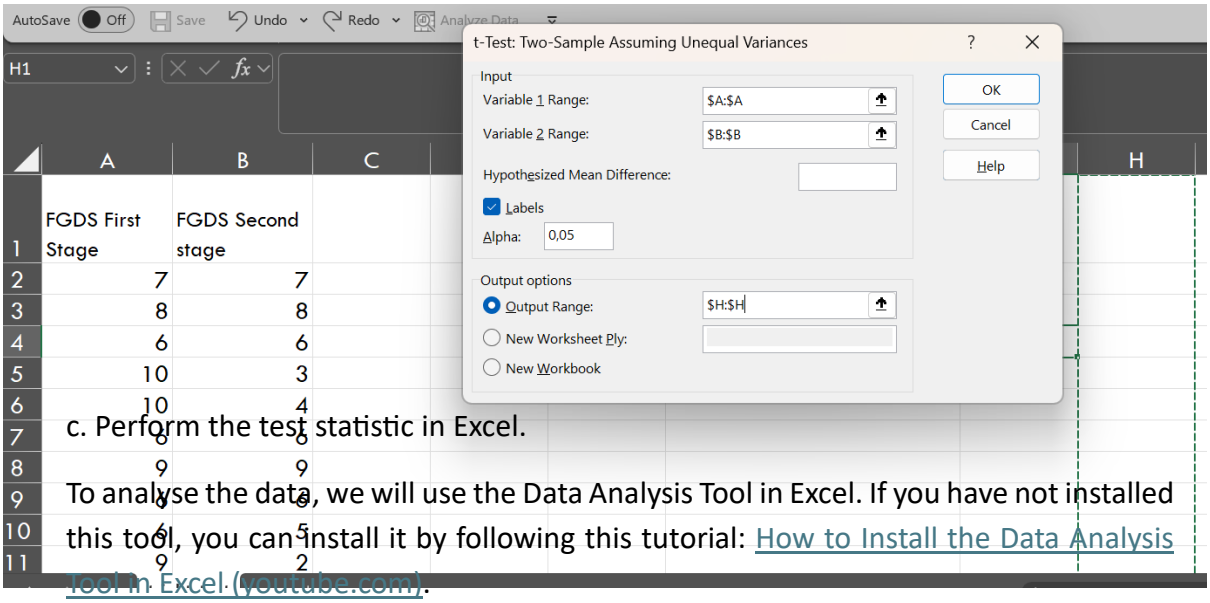
In this section, we will learn give you example of performing t-test in Microsoft Excel by following the steps above. Please note that the dataset below is dummy data used solely for demonstration purposes in this tutorial.

1. **Plan**:
   a. The parameter is mean of DDS.
   b. Null hypothesis (Ho): *There is no difference* in the mean DDS between participants at baseline and endline.
   Alternative hypothesis (Ha): *There is a difference* in the mean DDS between participants in baseline and endline.
2. **Solve**:
   a. We propose to use t-test to compare means of two independent samples.
   b. Check whether the dataset meets the assumptions of the chosen test.
   In Excel template, paste (i.e. paste as values) your data in column A (Figure 31):

A	B	C	D	E
PASTE YOUR DATA BELOW	Rank (do not edit)	Percentile (do not edit)	Z score (do not edit)	
7	24	0,156666667	-1,008252347	
8	50	0,33	-0,439913166	
6	10	0,063333333	-1,527379518	
10	131	0,87	1,126391129	
10	131	0,87	1,126391129	
6	10	0,063333333	-1,527379518	
9	94	0,623333333	0,314247152	
6	10	0,063333333	-1,527379518	
6	10	0,063333333	-1,527379518	

Figure 31: Check data distribution

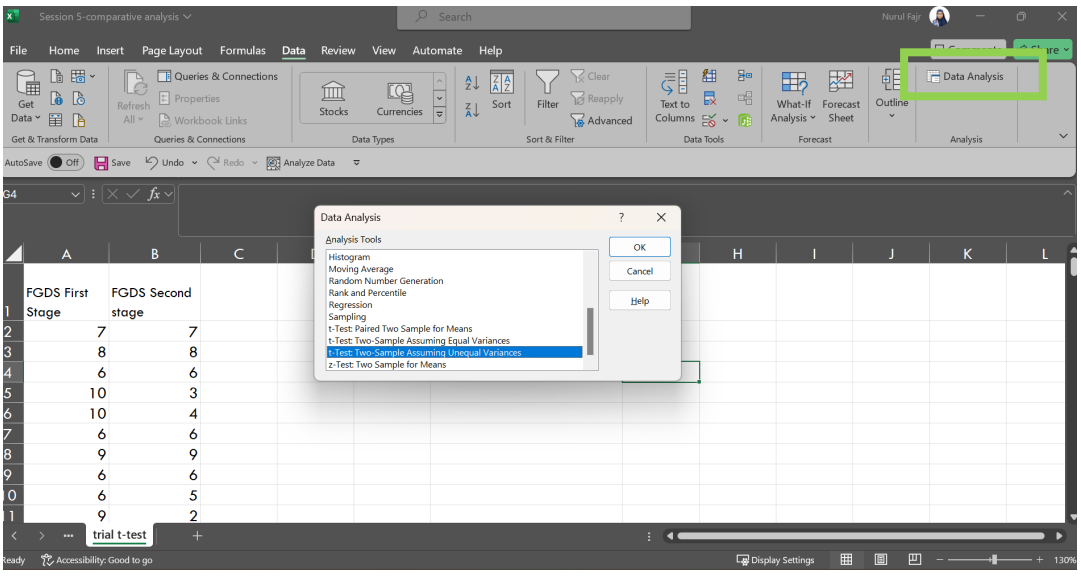
You will notice that the QQ plot on the right has adjusted to reflect the dataset (Figure 32). In this case, we can assume that our data follows a normal distribution, as it aligns with the straight line on the normal plot.



c. Perform the test statistic in Excel.

To analyse the data, we will use the Data Analysis Tool in Excel. If you have not installed this tool, you can install it by following this tutorial: How to Install the Data Analysis Tool in Excel (youtube.com).

Go to 'Data Analysis' then select 't-Test: Two Sample Assuming Unequal Variances (Figure 33).



Then, a dialog box will appear (Figure 34). Fill in the 'Variable 1 Range' by selecting the data in column A, and fill in the 'Variable 2 Range' by selecting the data in column B. Tick the 'Labels' option since we are including the column titles in our selection. Set the alpha level to 0.05. Then, fill in the 'Output Range' by selecting your desired column; in this example, we select column H. Click 'OK'.

DDS baseline DDS endline

d. Interpret the p-value.

After clicking OK, we get the result in column H. Since we used two-sided hypotheses, we look into the P(T<=t) two-tail which results in 3.22955E-28 or p-value < 0.05 (Figure 35). Since p-value is extremely small, this indicates that the difference between the mean FDGS score is not due to random chance.

G	H	I	J	K
	t-Test: Two-Sample Assuming Unequal Variances			
		FGDS First Stage	*FGDS Second stage*	
	Mean	7,926666667	5,38	
	Variance	2,229485459	4,129798658	
	Observations	150	150	
	Hypothesized Mean Difference	0		
	df	274		
	t Stat	12,36840276		
	P(T<=t) one-tail	1,61478E-28		
	t Critical one-tail	1,650433793		

Figure 35: Interpret p-value

3. **Conclude**

We reject the null hypothesis. It is known that there is a difference in the mean DDS between participants in baseline and endline.

Performing chi-square test

In this section, we will learn give you example of performing chi-square test in Microsoft Excel. Please note that the dataset below is dummy data used solely for demonstration purposes in this tutorial.

1. **Plan**
   a. The parameter is frequency.
   b. Null hypothesis (Ho): There is no association between home garden intervention and food insecurity prevalence.
      Alternative hypothesis (Ha): There is an association between home garden intervention and food insecurity prevalence.
2. **Solve**:

   a. We propose to use chi-square test to compare frequency of the two samples (received home garden intervention and do not receive home garden intervention).

   b. Since we have categorical variables, we do not need to check for normal distribution.

   c. Perform the test statistic in Excel.

   In the Excel sheet, fill out the frequency of food insecurity prevalence for each group in the green area, which corresponds to columns B through E (Figure 36). Column I through I contains formula to automatically calculate for p-value.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1		Severe	Moderate	Mild	Secure	Total			Severe	Moderate	Mild	Secure	Total
2	Baseline	51	28	10	61	150		Baseline	51,00	26,50	17,00	55,50	150
3	Endline	51	25	24	50	150		Endline	51,00	26,50	17,00	55,50	150
4	Total	102	53	34	111	300		Total	102	53	34	111	300
5													

Figure 36: Chi square test in Excel

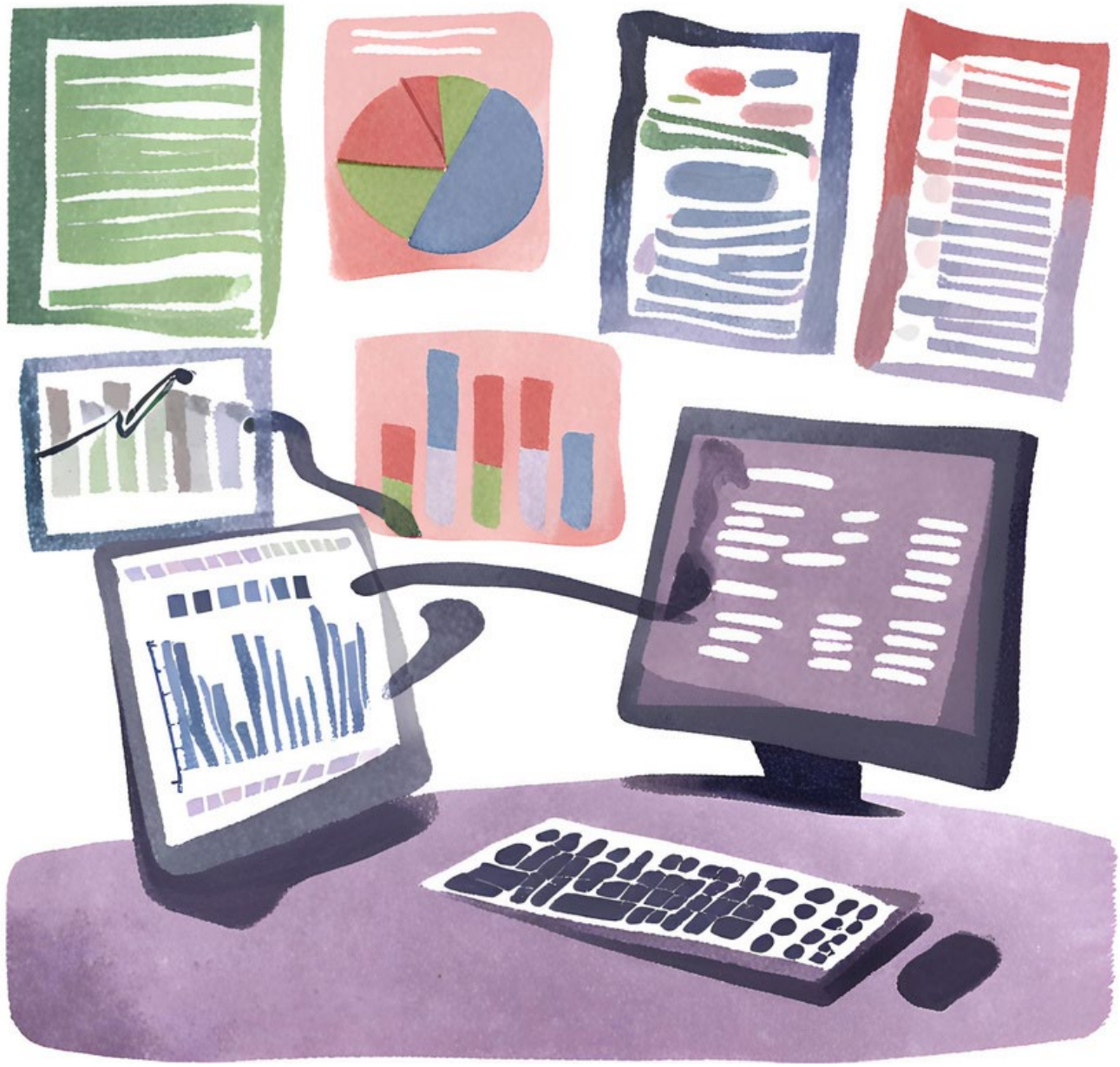
   d. Interpret the p-value.

   In the figure above, you can see that p-value is 0.07 ($p > 0.05$).

3. **Conclude**:

   We accept the null hypothesis. It is shown that there is no association between home garden intervention and food insecurity prevalence

4 Data visualization



In this chapter, we will learn the fundamentals of data visualization and how to make charts in Microsoft Excel.

Data visualization

In this chapter, we will learn how to visualize data by making different types of statistical charts. Data visualization allows audience to easily understand the information, patterns, or relationships within a data set. There are various tools available for data visualization, such as Microsoft Excel, Tableau, Google Data Studio, and others. However, we will focus on using Microsoft Excel because our datasets are relatively small datasets, and our visualizations are simple.

Choosing the appropriate chart

Knowing the right chart for the data is important to deliver the key information and avoid misinterpretation. The data can be visualized through different type of charts, such as bar chart, line chart, histogram, and others. The type of chart that we will use is depending on the purposes and type of variables that we have. In this section, we will only explain the charts that are relevant to our dataset, if you want to learn more about data visualization you can go to this page: https://extremepresentation.com/ (Abela, 2021).

Comparison

If you are interested to make comparison by displaying few categories, you can use either column chart or bar chart. In our dataset, we will use bar chart or column chart for several times, for example to display the different type of toilet facilities and livestock type (Figure 37).

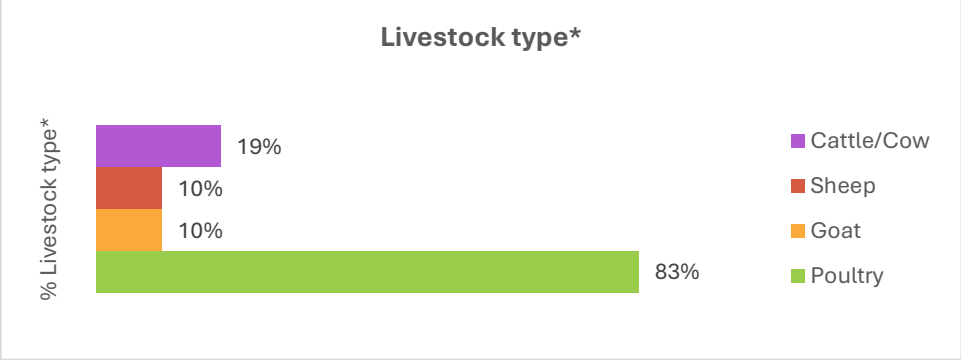


*Figure 37: Livestock type (*participants could give multiple answers)*

Composition

If you want to see the composition of static data (i.e. do not change over different time points), you can use pie chart. For example, we use pie chart to display the proportion of widow from our dataset (Figure 38).

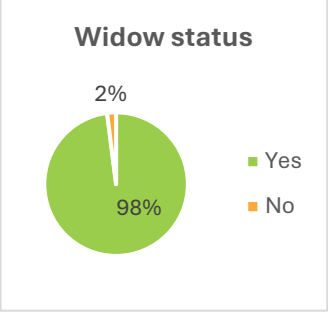


Figure 38: Widow status

Making a chart

Bar chart

First, open the Excel file containing the results of analysis. Then, make a new sheet containing the data that you want to display. Select the relevant table (Figure 39), in this example we are using the variable 'age of children'. Click on the 'Insert' tab, then choose 'Recommended Charts'.

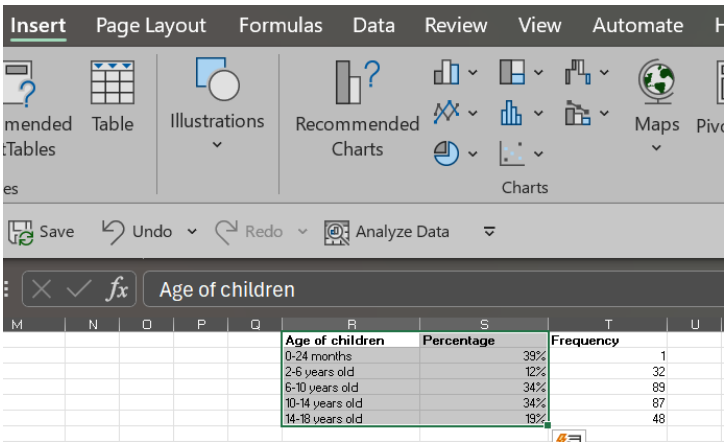


Figure 39: Age of children

Sometimes it can be helpful to see the list of recommended charts, however sometimes Excel can give misleading recommendation. Instead of selecting chart from 'Recommend chart', we need to click on 'All charts' tab (Figure 40). Since we want to show the proportion of different categories (in this case, the category is children age), we click on 'Bar', then click 'OK'.

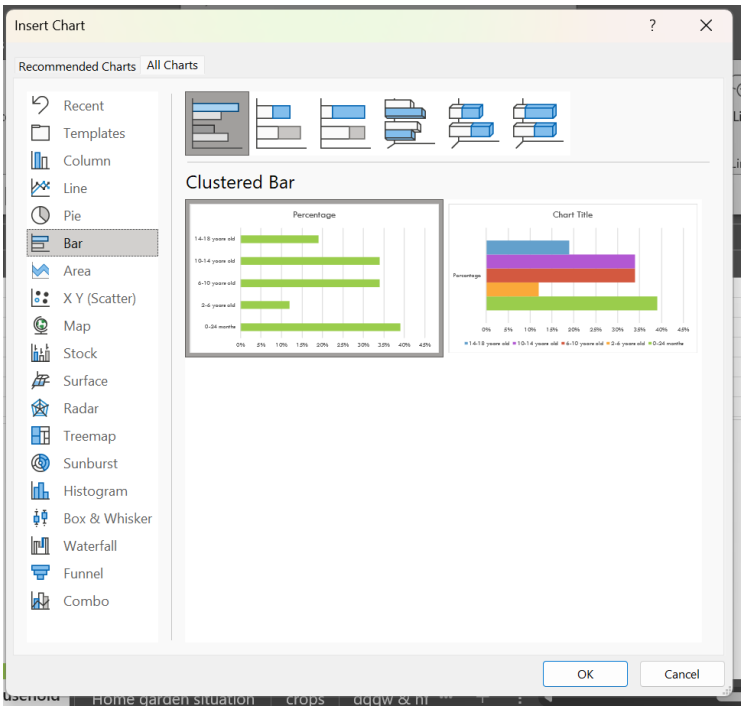


Figure 40: Insert chart

The bar chart will appear as follows. Click on '+' button to customize the chart elements, for example the chart title, data labels, and legend (Figure 41).

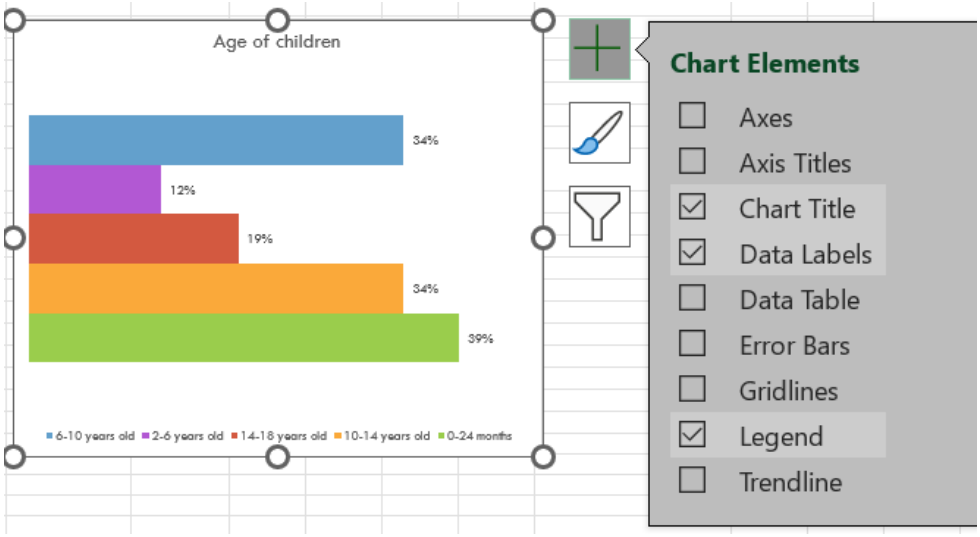


Figure 41: Bar chart of age of children

References

Abela, A. (2021). *The extreme presentation method* [Internet]. Available from: https://extremepresentation.com/

Auliani, N.A., Glaser, J., Pittore, K. 2023. *A guideline to conduct: Combating malnutrition in Mazar-e Sharif Study*. Wageningen Centre for Development Innovation.

Berman H.B. (n.d). "*AP Statistics Tutorial*". https://stattrek.com/tutorials/ap-statistics-tutorial

Bakker, E.J., & de Kovel, C.G.F. (2022). *Advanced Statistics MAT20306.* Biometris Wageningen University and Research.

Baldi, B., & Moore, D. S. (2017). *The practice of statistics in the life sciences*. Macmillan Higher Education.

Bilinsky, P., & Swindale, A. Months of Adequate Household Food Provisioning (MAHFP) for Measurement of Household Food Access: Indicator Guide: Version 4 [Internet]. Available from: https://www.fantaproject.org/monitoring-and-evaluation/mahfp

Coates J, Swindale A, Bilinsky P. Household Food Insecurity Access Scale (HFIAS) for Measurement of Food Access: Indicator Guide: Version 3 [Internet]. 2007. Available from: https://www.fantaproject.org/monitoring-and-evaluation/household-food-insecurity-access-scale-hfias

Healy, Y. H. a. C. (2018). *From data to Viz | Find the graphic you need*. Available from: https://www.data-to-viz.com/#story

Martin-Prevel Y, Arimond M, Allemand P, Wiesmann D, Ballard TJ, Deitchler M, Dop MC, Kennedy G, Lartey A, Lee WTK, and Moursi M on behalf of the Women's Dietary Diversity Project (WDDP) Study Group. (2017).Development of a Dichotomous Indicator for Population-Level Assessment of Dietary Diversity in Women of Reproductive Age. Current Developments in Nutrition, 1(12) cdn-117. https://doi.org/10.3945/cdn.117.001701

Salkind, N. J. (2007). *Statistics for people who (Think they) hate statistics: The Excel Edition*. SAGE.

Yau, N. (2011). *Visualize this: The Flowing Data Guide to Design, Visualization, and Statistics*. John Wiley & Sons.

2024

WAGENINGEN
UNIVERSITY & RESEARCH

Village of Peace