WOMEN COUNT

# MODULE 7

## MULTILEVEL DISAGGREGATION ANALYSIS TO MONITOR THE SDGS FROM A LEAVE NO ONE BEHIND PERSPECTIVE

## TRAINING SYLLABUS

### Curriculum on Gender Statistics Training

This product was developed under the guidance of the Subgroup on Gender Statistics Training, within the Asia-Pacific Network of Statistical Training Institutes.

UN WOMEN

# Introduction

This syllabus has been designed to guide instructors on how to conduct training on multilevel disaggregation analysis. It can also be used by learners who wish to know more about this topic and other people generally interested in analysing data from a gender perspective.

This syllabus is part of a wider module on this area of gender statistics. Other materials within this module might include exercises, sample data sets, Power Point presentations and example quizzes. Please refer to the additional set of materials for a comprehensive and effective learning experience.

# Who is this module for?

- Statisticians and other experts that wish to learn about analyzing gender data and generating estimates in line with the Leaving No One Behind (LNOB) principle of the 2030 Agenda.
- Academics who wish to focus or inform their research through the use of multiply disaggregated data.
- Data journalists and civil society organizations that wish to use gender data to highlight findings for specific groups of women and girls.
- Anyone who wishes to find out how to analyze gender data from a LNOB perspective.

# What do I need to know before going through this module?

This module contains information about survey data analysis and is suitable for learners who are familiar with basic statistical concepts and have some understanding of coding using statistical software, preferably STATA. Additionally, it is recommended that the learner have some basic understanding of the Sustainable Development Goals[1], including their targets and indicators[2].

[1] For additional information on the SDGs see: https://www.un.org/sustainabledevelopment/sustainable-development-goals/

[2] See: https://unstats.un.org/sdgs/indicators/indicators-list/

# Table of Contents

# 1. Introduction to multilevel disaggregation analysis to monitor the SDGs from a Leave No One behind perspective (LNOB analysis)

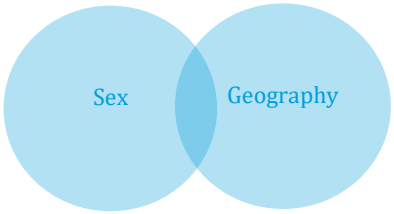
## 1.1. Leaving No One Behind

Fulfilling the 2030 Agenda requires accelerated change. If the aim is to leave no one behind, we must first identify who is furthest behind. Women and girls are often at risk of experiencing disadvantage. Besides sex-based discrimination, they may experience other overlapping forms of discrimination associated with socioeconomic characteristics, such as living in rural areas or poor households. These groups of women facing multiple forms of discrimination will experience severe forms of deprivation[3].

## 1.2. LNOB analysis

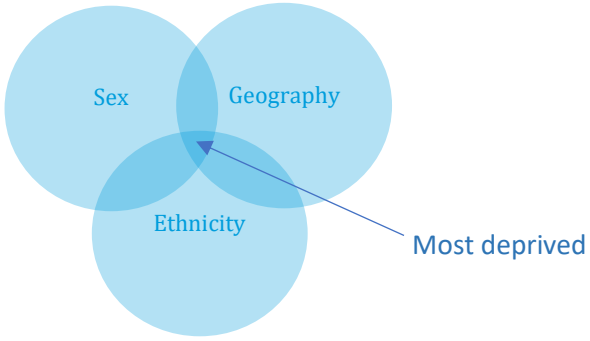
To identify those furthest behind, LNOB analysis focuses on disaggregating data by sex and other variables simultaneously. The first step to conducting this analysis should be identifying which variables should be used for disaggregation. To achieve this, it is important to look at multiple, contextually relevant, socioeconomic characteristics that can be potential causes or drivers of inequality in each country. The rationale is that an individual can be at the intersection of multiple identities that push them behind on the ladder of progress. For example, a woman who lives in a rural location may experience both sex-based discrimination and geography-based discrimination.

*Figure 1: Deprivation due to overlapping discrimination between two variables*




Similarly, a woman of a minority ethnicity who lives in a rural location may be at the intersection of possible discrimination arising from gender identity, geography, and ethnicity.

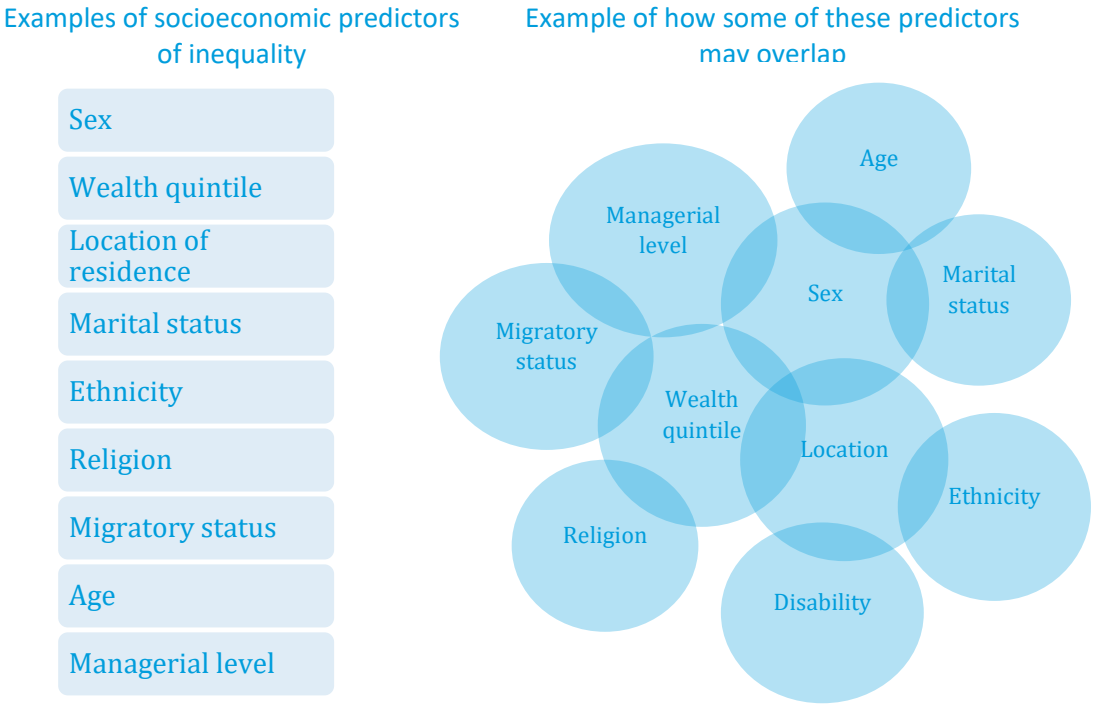
*Figure 2: Deprivation due to overlapping discrimination between three variables*




[3] UN Women. 2018. *Turning Promises into Action* https://www.unwomen.org/-/media/headquarters/attachments/sections/library/publications/2018/sdg-report-gender-equality-in-the-2030-agenda-for-sustainable-development-2018-en.pdf?la=en&vs=4332

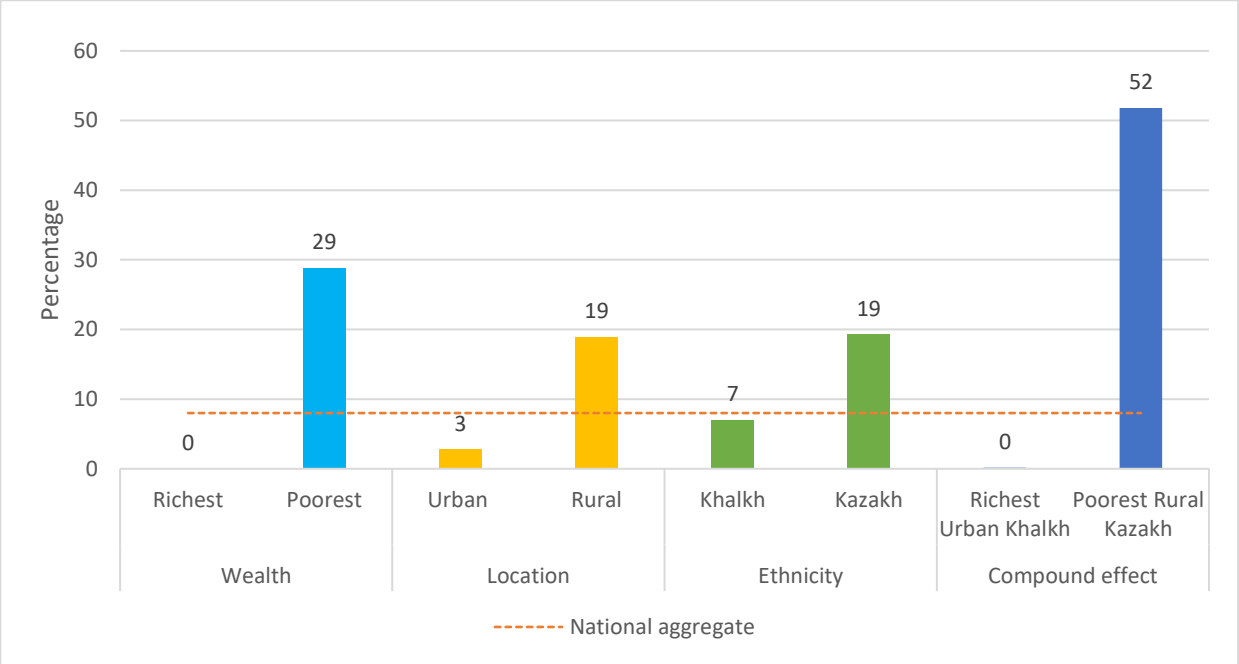
Figure 3 shows examples of socioeconomic predictors of inequality that can produce overlapping disadvantages.

*Figure 3: Overlapping forms of discrimination*



Let's look at data from Mongolia to understand this overlapping effect. In this example, we choose a single indicator, the proportion of people who did not complete more than six years of education (or those who are education-poor):

*Figure 4: Proportion of women and girls who did not complete more than 6 years of education in Mongolia*




*Source: UN Women analysis based on Mongolia's 2014–15 Multiple Indicator Cluster Survey.*

In Mongolia, the likelihood of being education-poor increases if women and girls identify with ethnic minorities, religious minorities and live in a poor household. These factors compound to create substantially deprived groups of women. The most disadvantaged group for this indicator is the poorest rural Kazakh women – who belong to an ethnic minority group. They are nearly 370 times as likely as the richest urban Khalkh women to be education-poor. If the same analytical exercise is repeated for other indicators, such as proportion of women and girls who live in overcrowded households and the proportion of women who lack access to clean cooking fuels, it becomes clear that the poorest rural Kazakh women lag behind across numerous development outcomes, and therefore are multiply deprived.

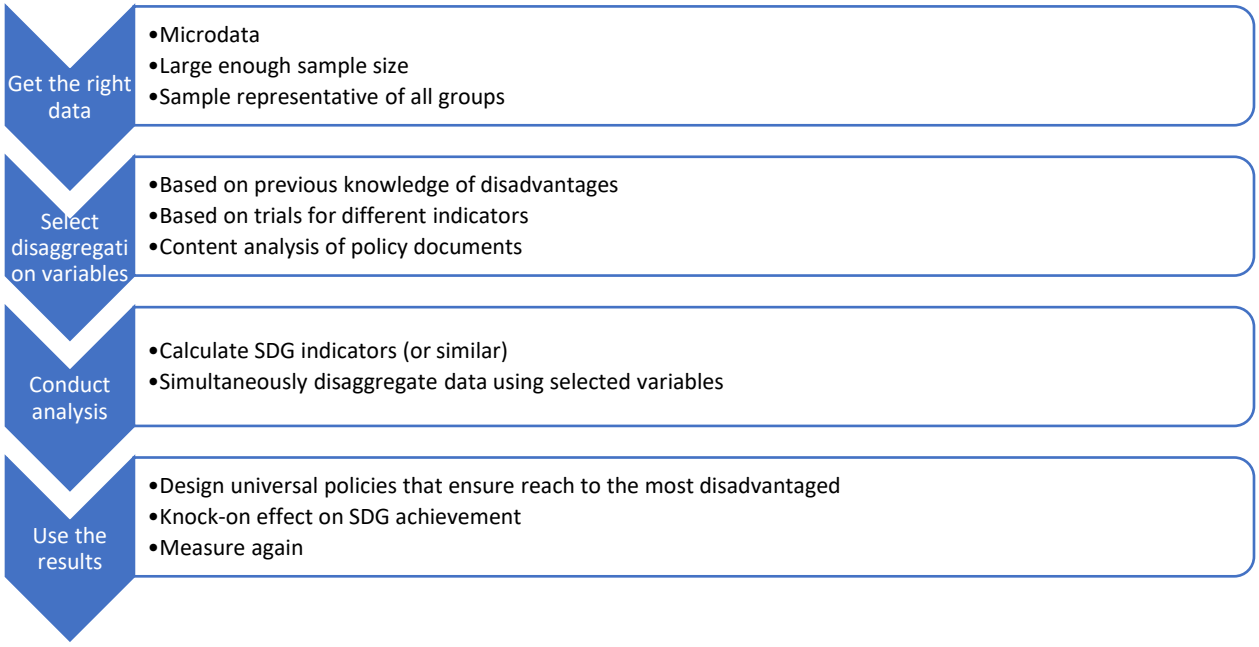
The tendency of deprivations to cluster implies a need to assess deprivations through a multidimensional lens, whereby women and girls facing intersecting forms of discrimination, which are traditionally invisible in aggregated statistics, can be made visible[4].

[4] UN Women, Turning Promises into Action https://www.unwomen.org/-/media/headquarters/attachments/sections/library/publications/2018/sdg-report-gender-equality-in-the-2030-agenda-for-sustainable-development-2018-en.pdf?la=en&vs=4332

# 2. LNOB analysis methodology

Four key steps are necessary to undertake multilevel disaggregation analysis to assess SDG progress from an LNOB perspective. For the purpose of describing the methodology, we will use example indicators for select countries in Asia-Pacific, although the analysis can be replicated for many other indicators and countries.

*Figure 5: Framework for LNOB analysis*




## 2.1. Get the right data

Microdata is necessary to conduct this type of analysis. It is important to utilize survey or census data to be able to generate aggregates for all population groups of interest.

### 2.1.1. Sample

Typically, nationally representative household surveys have sample sizes that are large enough for disaggregating data by sex confidently. However, to undertake the analysis for specific population groups, simultaneous disaggregation by multiple variables – such as wealth and sex or wealth and location – is necessary. Before beginning this analysis, it is important to select a data source that provides large enough samples for all population groups of interest, and samples that are representative of the composition of the population groups of interest. For instance, if the research interest includes disaggregating data to capture phenomena of relatively rare incidence, such as disability, it may be important to choose a survey that oversamples disabled population groups to obtain representative estimates. Looking at the sampling methods of the survey instrument is therefore essential to ensure the reliability of the analysis. If Census data were to be used, no sampling issues would arise.

### 2.1.2. Who is interviewed?

Household surveys that collect information only from the household head (often men) are not suitable to accurately capture the specific realities of women's lives pertaining to areas such as decision-making, violence against women, contraceptive use, perceptions, etc. Hence, data sets with responses from different household members, both women and men, are crucial for LNOB analysis. In short, it is preferable that questions about women are responded by women.

### 2.1.3. Timeliness and periodicity of data

Using data that is collected frequently and released in a timely manner adds to the relevance of the analysis. Household surveys, such as the Demographic and Health Surveys (DHS) or Multiple Indicator Cluster Surveys (MICS), that are rolled out at least once every five to seven years are examples of surveys well placed as data sources for this type of analysis. It is important to chose surveys that have been released relatively recently and surveys with a relatively recent reference period.

### 2.1.4. International comparability

If this analysis is to be carried out for a number of countries, the use of standardized survey instruments is recommended, in order to make the LNOB estimates comparable.

Examples of standardized household surveys suitable for LNOB analysis include DHS and MICS surveys, Labour force Surveys (LFS), Living Standard Measurement Surveys (LSMS), etc. In this module, DHS and MICS surveys will be used as examples, as in these surveys questions are asked to both women and men in the household separately, and their sampling design has taken into consideration the representativeness of various population groups. These surveys are also good sources as they provide information for the calculation of various SDG indicators. Examples of SDG-related indicators (or proxies) used in this module that can be obtained from DHS and MICS survey data include:

*Figure 6: SDG indicators and data sources*

| SDG | Indicator | Data Source |
|---|---|---|
| SDG 2: End hunger, achieve food security and improved nutrition and promote sustainable agriculture | Proportion of women aged 18–49, who are underweight (BMI less than 18.5 kg/$m^2$) | DHS: IR file<br>MICS: wm file |
| SDG 3: Ensure healthy lives and promote well-being for all at all ages | Proportion of births attended by skilled health personnel (births in last five years)[5] | DHS: BR file<br>MICS: wm file |
| SDG 4: Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all | Proportion of women and girls aged 15–49 with primary or less years of education | DHS: IR file<br>MICS: wm file |

[5] Note that the indicator used for this module is 'births not attended by skilled health personnel', as opposed to the SDG indicator, which is 'births attended'. This is because with MDA-LNOB analysis we are aiming at showing different forms of deprivation.

| | | |
|---|---|---|
| SDG 5: Achieve gender equality and empower all women and girls | Proportion of women aged 18–49 who were married before age 18 | DHS: IR file<br>MICS: wm file |
| SDG 6: Ensure availability and sustainable management of water and sanitation for all | Proportion of women and girls aged 15–49 with access to basic drinking water services[6] | DHS: IR file<br>MICS: wm file + hh file |
| | Proportion of women and girls aged 15–49 with access to basic sanitation services[7] | DHS: IR file<br>MICS: wm file + hh file |
| SDG 7: Ensure access to affordable, reliable, sustainable and modern energy for all | Proportion of women and girls aged 15–49 with access to clean cooking fuel[8] | DHS: IR file<br>MICS: wm file |
| SDG 8: Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all | Proportion of women aged 18–49 currently not employed | DHS: IR file<br>MICS: wm file |
| SDG 11: Make cities and human settlements inclusive, safe, resilient and sustainable | Proportion of women and girls aged 15–49 living in overcrowded housing | DHS: PR file + IR file<br>MICS: wm file |

## 2.2. Select disaggregation variables

The first step in conducting LNOB analysis should be to identify research questions. In this case, the questions could be: 1) what are the population groups that are lagging behind? and 2) according to which development indicators do they lag behind? If this analysis is being conducted to inform the monitoring of national policies and priorities, it is important that these questions are answered utilizing national strategies or similar documentation.

UN ESCAP's Every Policy Is Connected (EPIC) tool[9] provides practical guidance for the identification of development indicators in line with national priorities, as reflected in national development strategies or similar documents. It also provides guidance to identify target groups (e.g. groups more likely to be lagging behind) and potential data needs to monitor the development outcomes of those groups. At the very core, the EPIC tool is also a facilitator of dialogue between data producers and data users. However,

[6] Note that the indicator used for this module is 'lack of access to basic water services' as opposed to the SDG indicator, which is 'access to basic water services'. This is because with MDA-LNOB analysis we are aiming at showing different forms of deprivation.
[7] Note that the indicator used for this module is 'lack of access to basic sanitation services', as opposed to the SDG indicator, which is 'access to basic sanitation services'. This is because with MDA-LNOB analysis we are aiming at showing different forms of deprivation.
[8] Note that the indicator used for this module is 'lack of access to clean cooking fuels', as opposed to the SDG indicator, which is 'access to clean cooking fuels'. This is because with MDA-LNOB analysis we are aiming at showing different forms of deprivation.
[9] UN ESCAP. 2019. Every Policy Is Connected Tool. Working Paper No. 9.
https://www.unescap.org/sites/default/files/SD_Working_Paper_no.9_Sep2019_EPIC_tool.pdf

although the scope of the EPIC tool is quite broad, this module will only cover how EPIC can be used to identify priority indicators and disaggregation variables. The following steps are recommended by EPIC[10]:

### 2.2.1. Hold multi-stakeholder consultation

A range of data users and producers must come together to have a dialogue on data needs and priorities[11]. These may include:

- Policymakers at the national, sectoral or local level. They should lead the review of all relevant policy or planning documents (e.g. national strategies) to identify:
    - Priority policy areas for the country, and
    - Specific population groups mentioned in the strategic document as groups of concern or recipients of the policy.
- National planning and budgeting organizations. They should assist in the identification of key priority policies.
- SDG focal points, civil society organizations and private-sector representatives can assist policymakers in identifying priority issues and strategies to address such issues.
- National Statistics Offices (NSOs). Once the priority areas and population groups have been identified, NSOs should guide policymakers in the development of appropriate indicators to monitor progress towards policy goals for the priority areas.
- Other data producers, including representatives from line ministries. They can assist the NSO in the process of developing appropriate indicators (including identification of data sources).

### 2.2.2. Content analysis of gender-relevant policy documents to identify priority issues and target groups

Policymakers need to lead the review of existing policy documents to identify key priority areas, target population groups and data disaggregation needs. This is done by undertaking a content analysis of existing policy documents. Take a look at the following example, an excerpt from India's Niti Ayog three-year Action Agenda (2017). This document sets the plans of action for the years 2017–18 to 2019–2020. Language on one of the key goals indicates:

> *"In order to reduce dropout rates, especially among girls belonging to Scheduled Tribe[12] communities, hostel facilities[13] must be ensured in all areas in which they are currently absent. Moreover, where facilities exist, they need to be upgraded and maintained so that they are usable. State Governments are currently eligible for 100% funding for establishment of all Ashram Schools for girls as well as Ashram Schools for boys in areas that are Naxalite[14] affected."*

[10] For more information and details about the EPIC process, please refer to: UN-ESCAP. "EPIC: A Data Policy Integration Tool." https://www.unescap.org/sites/default/files/EPIC_V1.2_AB15Aug18.pdf

[11] https://www.unescap.org/sites/default/files/EPIC%20Overview%20%28EPIC_V1.1_Final%29.pdf

[12] Scheduled Tribe is a classification of caste

[13] Hostel facility is a residential or accommodation facility where students can stay. In the Naxal affected areas, hostel facilities can especially serve as a safe and secure place for students to stay at as they won't need to travel through conflicted areas to reach their home. Moreover, dropouts due to fear of being attacked on the road from school to home can also be prevented.

[14] Areas affected by Naxalites are areas where there is ongoing conflict between Naxals and supporters of the Indian Government.

Content analysis of this excerpt may look as follows. Note that the text highlighted in red represents the key priority area and the text highlighted in blue represents the target groups:

> "In order to reduce dropout rates, especially among girls belonging to Scheduled Tribe communities, hostel facilities must be ensured in all areas in which they are currently absent. Moreover, where facilities exist, they need to be upgraded and maintained so that they are usable. State Governments are currently eligible for 100% funding for establishment of all Ashram Schools for girls as well as Ashram Schools for boys in areas that are Naxalite affected."

The content analysis of the document helps identify the following:

| | |
|---|---|
| **Key issues/Area of action** | Reduce school dropout rates; safety of vulnerable groups |
| **Core concept (see EPIC for complete list of core concepts[15])** | - Prevention<br>- Quality<br>- Services and entitlements |
| **Target groups** | Girls, especially those belonging to the Scheduled Tribe communities. Boys, especially those living in areas affected by Naxalites |
| **Data disaggregation** | By sex<br>By caste<br>By region |

### 2.2.3. Map with existing indicators at national/regional/global levels

Continuing with the same example of school dropout rates, the issues identified from the policy analysis can be mapped with national or global indicators (e.g. SDG indicators). For example, the following indicators from different frameworks may relate to the issue of school dropout rates:

- *National Indicator Framework, India*
    - 1.a.2: Proportion of total government spending on essential services (education, health and social protection)
    - 4.1.7: Out-of-school ratio (primary, upper primary, elementary, secondary and higher secondary)
    - 4.1.1: Net Enrolment Ratio in primary and upper primary education
    - 4.1.2: Adjusted Net Enrolment Ratio in primary, upper primary and secondary education
    - 4.1.3: Gross Enrolment Ratio in higher secondary education
    - 4.3.2: Proportion of women and girls, and men and boys enrolled in higher education, technical and vocational education
    - 4.6.1: Literacy rate of youth in the age group of 15–24 years.

- *SDG Global Indicator Framework*

[15] https://www.unescap.org/sites/default/files/EPIC_V1.2_AB15Aug18.pdf

- 4.1.1: Proportion of children and young people: (a) in grades 2/3; (b) at the end of primary; and (c) at the end of lower secondary achieving at least a minimum proficiency level in (i) reading and (ii) mathematics, by sex
- 4.2.2: Participation rate in organized learning (one year before the official primary entry age), by sex
- 4.3.1: Participation rate of youth and adults in formal and non-formal education and training in the previous 12 months, by sex
- 4.5.1: Parity indices (women/men, rural/urban, bottom/top wealth quintile and others, such as disability status, indigenous peoples and conflict-affected, as data become available) for all education indicators on this list that can be disaggregated.

### 2.2.4. Conceptual framework analysis

Now that the target groups, key policy issues and related development indicators have been identified, it is important to develop a conceptual framework for analysis. This includes identifying variables that may drive or influence some of the indicators. For example, in the above example for India, a range of variables (or socioeconomic characteristics) could be driving dropout, for example, wealth, geography (urban and rural locations), region (living in a Naxalite region), ethnicity, religion, caste, migratory status, etc. In order to identify such variables, it might be necessary to refer to existing research, even if qualitative or anecdotal.

### 2.2.5. Data at multiple levels of disaggregation to LNOB

Often, women and girls face multiple, overlapping forms of discrimination. In this sense, it becomes imperative to disaggregate data not just by each of the individual variables identified in 2.2.4 above but by a combination of these variables. In essence, deprivations are most acutely felt by women and girls who face multiple forms of discrimination[16]. Women and girls who face gender-based discrimination are likely to be less deprived than those who face both gender and race discrimination. This also means that not all members of a social group will face the same types of deprivations. In this case, a focus on race without gender, for example, is inadequate because the specific deprivations experienced by women within a racial group will go unrecognized.

Hence, to leave no one behind and identify the most deprived groups of women and girls, data disaggregation by sex or by location or by age, may not be sufficient in contexts where women are likely to be at the junction of multiple identities and are hence likely to experience discrimination due to a combination of factors.

In the example below, we see how, due to discrimination based on sex, location and region, rural women who reside in certain regions often fare worse than the average woman in a country. Data has to be disaggregated simultaneously by wealth, location and region to identify these differences. This can be illustrated with a wide variety of indicators. For this example, we chose the proportion of the population with access to safely managed sanitation services, as follows:

[16] UN Women 2018.

| Data disaggregation | Rationale | Example of research question (what do we want to find out?) |
|---|---|---|
| by sex<br>(only one variable) | To identify how access to sanitation services varies between women and men or girls and boys. | Difference in access to safely managed sanitation services between men and women in Bangladesh. |
| by location<br>(only one variable) | To identify how access to sanitation services varies between rural and urban populations. | Degree of access to safely managed sanitation services among rural households in Nepal. |
| by sex and location<br>(by a combination of two variables) | To see the combined effect of sex and location-based discrimination on people's likelihood of accessing safely managed sanitation services. This showcases large gaps among population groups (e.g. urban men vs. rural women). | Degree of access to safely managed sanitation services among women and girls who live in rural locations in the Philippines. |
| by sex, location and region<br>(by a combination of three variables) | To see the combined effect of discrimination based on sex, location and region of residence, on people's likelihood of accessing safely managed sanitation services. Typically, the more disaggregation variables are used simultaneously, the higher the likelihood of seeing large gaps among population groups. | Access to safely managed sanitation services among women and girls who live in rural households of Khangai region in Mongolia. |

## 2.3. Conduct analysis

Following the steps listed up to this point will provide the analyst with:

- Priority issues (e.g. relevant topics to include in the analysis)
- Specific development indicators, linked to each of the policy issues, to be used in the analysis
- Relevant variables to utilize for data disaggregation, in line with population groups of interest
- Relevant data sources and microdata to use

Let's imagine that, as a result of the policy analysis, the analyst wishes to calculate estimates for the proportion of women and girls who are currently underweight – a proxy for a measure of undernourishment – to monitor progress towards SDG 2 on eliminating hunger. Before beginning the analysis, it is important to consider the following:

### *Q.1) Who is the target population?*

The selection of the target population (e.g. sex, age group, etc.) should be made based on policy priorities and indicator metadata when using existing indicators. Prior to proceeding with the calculations, however, it is important to take into account the sample size and sampling design of existing data sources. For

instance, the metadata for this SDG indicator refers to women aged 18–24 only. Because this age group yields a very narrow sample in the surveys we will be using for this analysis (e.g. DHS 2016 for Nepal, DHS 2017 for the Philippines and similar surveys), we have broadened the age group to 18–49 years.

*Q.2) Is there any group of women we must exclude?*

The aim is to assess those who are underweight purely as a result of unmet nutritional needs. Because the DHS does not include questions on specific food intake, the body mass index is used as a proxy. Pregnant women and those whose pregnancies terminated less than three months before the time of the survey tend to register a higher Body Mass Index. Including these women in the calculations would skew the results, so pregnant women and those whose pregnancies terminated within three months of the survey must be excluded from the analysis.

*Q.3) What is the definition of underweight?*

The prevalence of undernourishment is an estimate of the proportion of population whose habitual food consumption is insufficient to provide the dietary energy levels that are required to maintain a normal active and healthy life. It is expressed as a percentage.[17] Again, because DHS surveys do not provide this kind of information, Body Mass Index (BMI) is used instead. According to WHO guidelines, an adult woman is considered underweight if her BMI falls below 18.5.

*Q.4) What is required to carry out this analysis?*

To calculate the proportion of women aged 18–49 who are underweight, the following information is required:

a. *Microdata:* Demographic and Health Surveys, Multiple Indicator Cluster Surveys or any other source of microdata that captures BMI at the individual level. In the case of DHS and MICS, different file recodes should be used depending on the indicator. In the case of BMI, the recommended recode is the IR file, or women's recode.

[17] SDG Indicators Metadata Repository. Available from: https://unstats.un.org/sdgs/metadata/

*Box 1: DHS recode files*

Both for DHS and MICS, the data is structured into various recodes, namely:

- *IR File-Individual Recode File (DHS)1 or wm-women's file (MICS)1:* This file has one record for every eligible woman. It contains all the data collected in the women's questionnaire plus some variables about the household. The unit of analysis in this file is individual women.
- *PR File-Personal Recode File (DHS) or hl-list of household member (MICS):* This file has one record for every household member. The unit of analysis in this file are individual household members.
- *BR File-Birth Recode File (DHS) or ch-children's file (MICS):* This file has one record for every child ever born to the interviewed women. It contains the full birth history of all women interviewed. The unit of analysis in this file are births ever born of eligible women.
- *HR File-Household Recode File (DHS) or hh-household file (MICS):* This file has one record for each household. No information from the individual women/men questionnaire is present in this file. The unit of analysis in this file is the household.

Other recodes are available in other rounds of DHS and MICS for different countries, such as men recodes and child recodes. For the purpose of the analysis presented in this module, we will only use the four recodes cited above.

b. *Statistical analysis software such as Stata, SPSS, SAS or R:* In these guidelines, the actual codes provided to calculate each of the estimates are based on STATA coding language.
c. *Indicator metadata:* Metadata is essential to fully understand the definition and method of computation for the indicators. Countries might have metadata guidelines for each of their national indicators. In the case of internationally agreed SDG indicators, it is recommended to use the Global SDG Metadata repository[18]. This metadata repository contains information such as definitions, methods of computation, recommended data sources, potential limitations and other important information that must be fully understood before performing data analysis.
d. *Survey metadata*: The analyst must be familiar with some key aspects of the data before conducting the analysis. These key aspects are as follows[19]:
   - Sample: DHS and MICS data are collected from a sample. Unlike censuses, which interview every individual, survey data is collected from a portion of the total population, i.e. sample. The sample is chosen to be representative of the reference population. Because DHS and MICS include many questions associated with the lives of women, women tend to be over-sampled in these surveys. Therein lies the importance of using the relevant weights for each type of analysis when using this type of data.

[18] UN Statistical Division. "Metadata." https://unstats.un.org/sdgs/metadata/
[19] DHS. "Guide to DHS Statistics". https://dhsprogram.com/pubs/pdf/DHSG1/Guide_to_DHS_Statistics_DHS-7.pdf

- *Sampling frame*: The list of all enumeration areas (EA) with measures of size (population and/or households) covering the entire population from which the sample is drawn. Many surveys use the most recent population census as a sampling frame. The goal of the sampling frame is to identify everyone in the population so that each individual has a probability of being selected for interview.
- *Primary Sampling Units (PSU):* Also known as clusters, PSUs are groups into which the population is divided. For example, for the Bangladesh DHS survey 2014, the primary PSU for the survey is an EA that has an average of about 120 households.
- *Sample design*: There are many forms of sampling design (e.g. simple random sampling, stratified random sampling, etc). DHSs use a two-stage cluster sampling procedure.
- *Stratification:* The process of grouping PSUs into strata, i.e. homogeneous subgroups (such as urban/rural or geographic region). In the case of stratified sampling, strata are selected first, and households within each strata are selected at a second stage.
- *Sampling weights:* A number that is multiplied by each observation (e.g. each response from a woman, child, household, etc. depending on the questionnaire) to weight-up or weight-down that observation if under- or over-sampled. Weights must always be used when conducting survey data analysis. These weights vary, depending on the unit of analysis. In the case of DHS, these are some of the available weight variables:

*Figure 7: Two-stage cluster sampling*

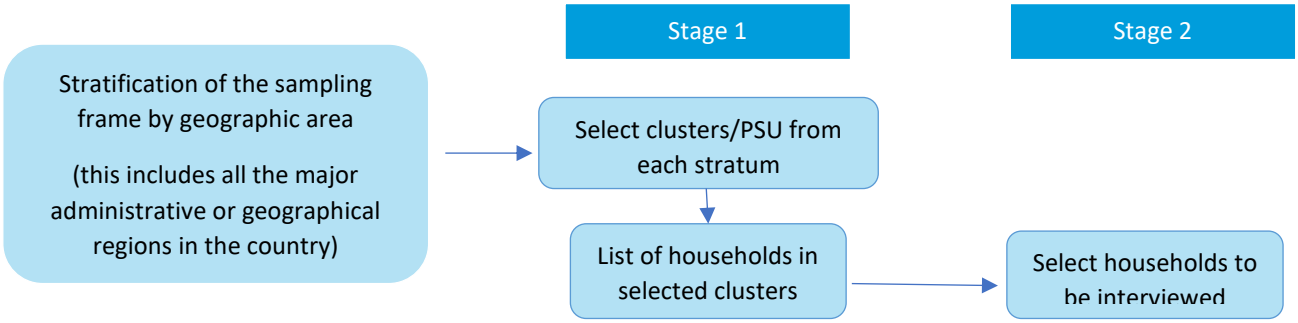



*Figure 8: Weight variables*

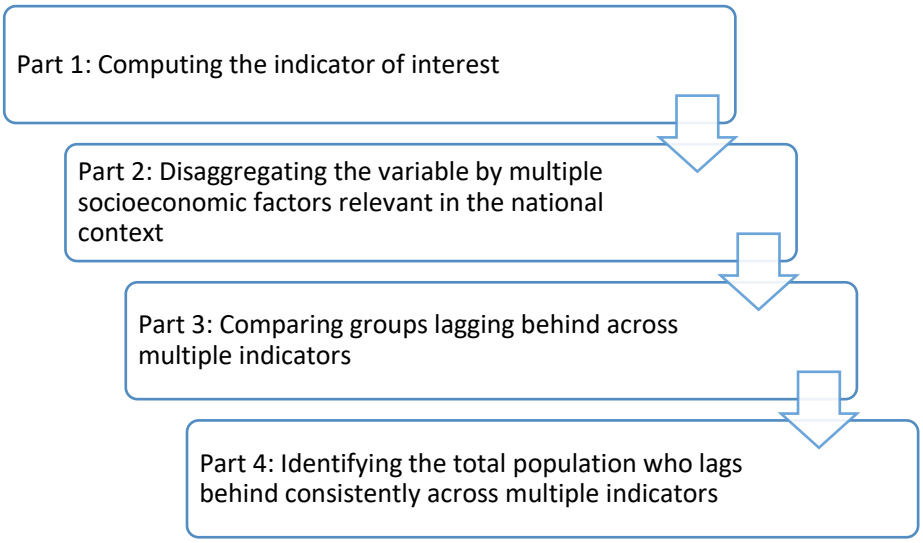
| Unit of analysis | Weight variable |
|---|---|
| Households | hv005 |
| Household members | hv005 |
| Women | v005 |
| Domestic violence | d005 |
| Men | mv005 |

*Box 2: Sampling weights*

- To ensure the survey sample is representative of the population where the survey was conducted, it is important to always apply sampling weights. In DHS and MICS surveys, the sample is often selected with unequal probability. Regions with small populations are over-sampled so they have large enough numbers of observations to produce reliable estimates, and regions with large populations are under-sampled for cost-efficiency purposes. This sample may then look very different from the actual population distribution. Hence, to restore the representativeness of the sample and fix the purposive under- and over-sampling, DHS and MICS apply sampling weights.
- The selection of a weight variable depends on whether you are calculating estimates for households, for women, or for victims of violence, for instance. If you are faced with analysis that uses several variables with different types of weights, always chose the weight associated with the smaller population group. For instance, to assess whether violence against women is larger in urban or in rural areas, you might be using a location variable from a women's recode (typically v005 is the weight variable for women in DHS) and a violence variable (which is associated to weight d005 in DHS). Because the population group that was subjected to the violence questionnaire in DHS is not as large as the total number of women interviewed for the survey, you should use d005.
- In DHS, sample weights are calculated to six decimal places but represented without decimals in the standard recode files. This means, when applying weights during statistical analysis, the following format should be used: weights=v005/1000000.
- For all the analysis demonstrated in this guide, importance weights (iw) are used. These simply denote the importance that each case should be given. Additionally, when declaring to STATA or other statistical analysis software that your data source is a survey, probability weight (pw) should be used. This weight indicates that an observation in a survey represents a certain number of people in a finite population.

After all these elements are well understood, the analyst can proceed with all the stages of LNOB analysis as follows:

*Figure 9: Stages of LNOB analysis*




### 2.3.1. Part 1: Computing the indicator of interest

After selecting the indicators of interest and understanding their metadata, the first step is to calculate each of the indicators for the total population. For instance, in order to calculate the proportion of women aged 18–49 who are underweight, the following steps must be followed:

a) Download the country-specific Standard DHS or MICS data set for the desired year and in a format that is readable by Stata. When working with national governments, National Statistics Offices may be able to provide data files directly. For standardized surveys, such as MICS and DHS, the microdata can be downloaded online. For example, to download the data for the DHS Philippines for the year 2017, the following page (Figure 10) appears upon clicking on this link: https://dhsprogram.com/data/available-datasets.cfm. To analyze the data in STATA, select the option 'ALL STATA' and the 'click here to download the files you selected in one zipped file' link. The data set appropriate to be read in STATA will be downloaded as a zip file.

*Figure 10: DHS data repository comprising multiple file formats*

**Philippines: Standard DHS, 2017**

**>>>Click here to download the files you selected in one zipped file.<<<**

**Select files individually or by file format type then press the button below the list of files to start the download process.**

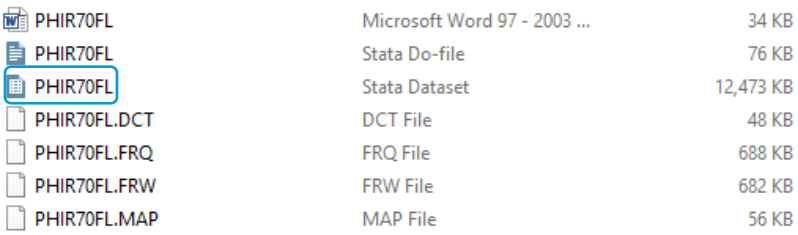
ALL STATA ☑ ALL FLAT ☐ ALL SAS ☐ ALL SPSS ☐ ALL HIERARCHICAL ☐

**Survey Datasets**

| File Name | File Size | File Format |
|---|---|---|
| **Births Recode** | | |
| ☑ PHBR70DT.ZIP | 6.42 MB | Stata dataset (.dta) |
| ☐ PHBR70FL.ZIP | 6.78 MB | Flat ASCII data (.dat) |
| ☐ PHBR70SD.ZIP | 10.9 MB | SAS dataset (.sas7bdat) |
| ☐ PHBR70SV.ZIP | 9.39 MB | SPSS dataset (.sav) |
| **Fieldworker Questionnaire** | | |
| ☐ PHFW71.ZIP | 7.63 KB | Hierarchical ASCII data (.dat) |
| ☑ PHFW71DT.ZIP | 9.45 KB | Stata dataset (.dta) |
| ☐ PHFW71FL.ZIP | 13.9 KB | Flat ASCII data (.dat) |
| ☐ PHFW71SD.ZIP | 9.88 KB | SAS dataset (.sas7bdat) |
| ☐ PHFW71SV.ZIP | 11.9 KB | SPSS dataset (.sav) |
| **Household Recode** | | |
| ☑ PHHR70DT.ZIP | 6.40 MB | Stata dataset (.dta) |
| ☐ PHHR70FL.ZIP | 6.79 MB | Flat ASCII data (.dat) |
| ☐ PHHR70SD.ZIP | 10.3 MB | SAS dataset (.sas7bdat) |
| ☐ PHHR70SV.ZIP | 6.84 MB | SPSS dataset (.sav) |
| **Individual Recode** | | |
| ☐ PHIR70.ZIP | 17.5 MB | Hierarchical ASCII data (.dat) |
| ☑ PHIR70DT.ZIP | 13.7 MB | Stata dataset (.dta) |
| ☐ PHIR70FL.ZIP | 13.4 MB | Flat ASCII data (.dat) |
| ☐ PHIR70SD.ZIP | 21.0 MB | SAS dataset (.sas7bdat) |
| ☐ PHIR70SV.ZIP | 13.9 MB | SPSS dataset (.sav) |
| **Children's Recode** | | |
| ☑ PHKR70DT.ZIP | 3.07 MB | Stata dataset (.dta) |
| ☐ PHKR70FL.ZIP | 3.42 MB | Flat ASCII data (.dat) |
| ☐ PHKR70SD.ZIP | 4.72 MB | SAS dataset (.sas7bdat) |
| ☐ PHKR70SV.ZIP | 3.84 MB | SPSS dataset (.sav) |
| **Household Member Recode** | | |
| ☑ PHPR70DT.ZIP | 6.02 MB | Stata dataset (.dta) |
| ☐ PHPR70FL.ZIP | 6.40 MB | Flat ASCII data (.dat) |
| ☐ PHPR70SD.ZIP | 13.6 MB | SAS dataset (.sas7bdat) |
| ☐ PHPR70SV.ZIP | 9.77 MB | SPSS dataset (.sav) |

b) Open the IR file in Stata by double-clicking on the relevant STATA data set. The reason behind using the IR file is because this information refers to individual characteristics of all women interviewed.

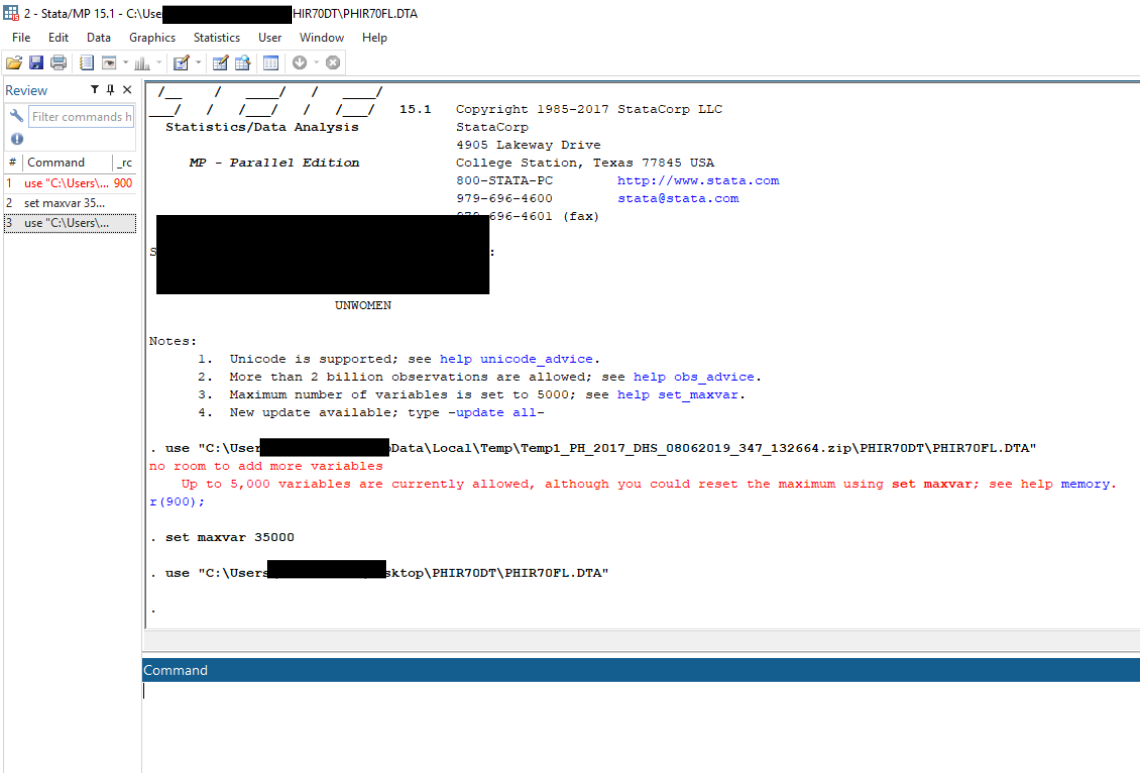
*Figure 11: Individual Recode File for data analysis in STATA*



| | | |
|---|---|---|
| PHIR70FL | Microsoft Word 97 - 2003 ... | 34 KB |
| PHIR70FL | Stata Do-file | 76 KB |
| PHIR70FL | Stata Dataset | 12,473 KB |
| PHIR70FL.DCT | DCT File | 48 KB |
| PHIR70FL.FRQ | FRQ File | 688 KB |
| PHIR70FL.FRW | FRW File | 682 KB |
| PHIR70FL.MAP | MAP File | 56 KB |

If STATA shows an error message upon opening the file, saying 'no room to add more variables', additional steps are necessary to increase the amount of memory that STATA can use in your computer. Memory allocation in STATA is automatic and hence, to resolve this issue, write: 'Set maxvar 30000' in the command window. This simply means STATA is being informed to make room for 30,000 variables. There is no hard-and-fast rule about setting it at 30,000 but it is usually a large enough number to load all the DHS/MICS variables (Should the message reappear, increase the number of variables further[20]).

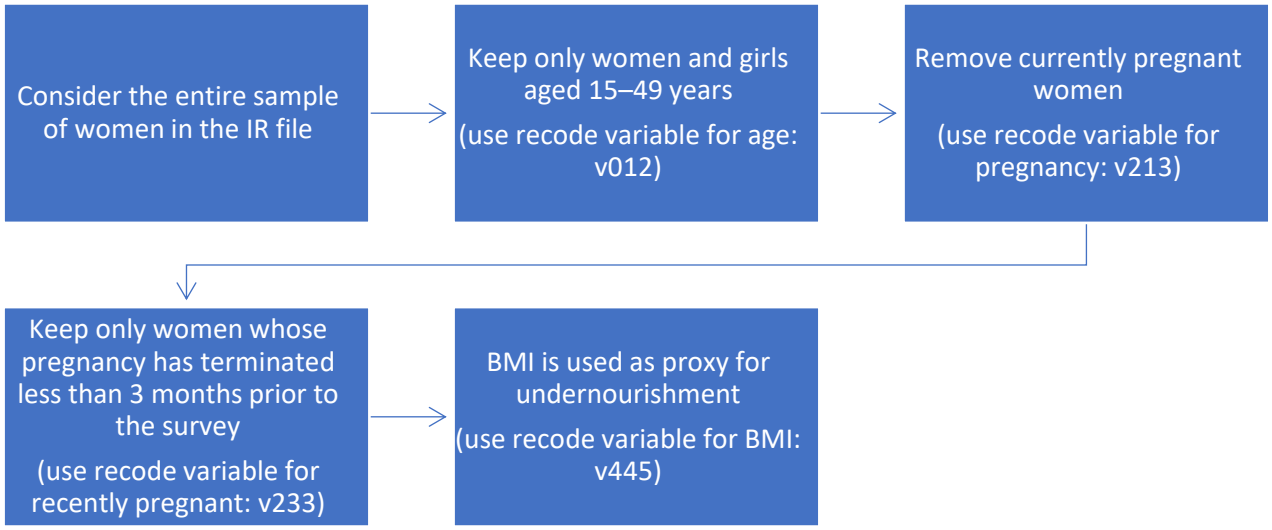
*Figure 12: STATA command window*




c) Run simple checks to make sure you have the right set of observations in your sample. To get an accurate estimate of how many women and girls aged 15–49 are underweight, the sample on which the analysis is conducted must be free of any issues that could skew the results. Women that could skew the estimates (e.g. women outside the reference age group, pregnant women, etc.) must be filtered out, as shown in Figure 13.

[20] STATA's memory capacity depends on the version. For instance, Stata/MP allows a maximum of 120,000 variables and Stata/SE allows a maximum of 32,767. *Stata manual on Managing Memory*: https://www.stata.com/manuals/u6.pdf

*Figure 13: Selecting an adequate sample of women using DHS data for analysis in STATA*




To proceed with this filtering, you must use codes in a format that STATA understands. Thus, to calculate the proportion of women aged 18–49 who are currently underweight, code as follows:

i. You must inform STATA about the nature of the data set (in this case, a survey data set). This can be done by entering information on svyset (PSU) [pweight], strata (). Here, PSU refers to the primary sampling unit and is recoded as v021, 'pweight' refers to the probability weight, which is equal to women's individual sample weight and is recoded as v005, and 'strata' is the grouping of PSUs, which is recoded as v022. Putting these values together, the first line of code will be as follows: svyset v021 [pweight=v005], strata (v022)

ii. To keep only those women who are in the age group of 18–49, type keep if v012>17. Note that most DHS surveys only compile information for women younger than 50. Should the data set in use include information for older women, you would also have to drop women older than 49 years of age for this analysis. This can be done by typing drop if v012 > 49.

iii. For analytical purposes, currently pregnant women and those whose pregnancy terminated less than three months prior to the survey need to be excluded. This can be done by typing drop if v213==1 to drop all women who are currently pregnant, and keep if v233>3 to only keep women whose pregnancy terminated more than three months prior to the survey.

iv. Once the data set contains only the desired set of women (e.g. women 18–49 that are not pregnant nor have they been in the last three months), the computation for share of women with low BMI can be carried out. This begins with creating a new variable, which could be named 'underweight'. To do so, type generate underweight=0. This will create a new variable called underweight' and ascribe a value of 0 to all observations for that variable. For instance, if every woman interviewed would have been asked "Are you underweight?", 0 would mean no and 1 would mean yes. Our new variable with values 0 for all observations would mean that every woman responded no.

v. The next step focuses on replacing 0 with 1 if women responded that they are underweight; that is, their BMI is less than 18.50 (In the Philippines DHS data set, the values for BMI appear without the decimal as 1850 instead of 18.50). Typing replace underweight=1 if v445<1850 will replace 0 with a value of 1 for women whose BMI is less than 18.50 kg/$m^2$.

vi. Missing values are those values that have no observations. It is very important to always take them into consideration in order to avoid over- or under-estimation. For instance, our new "underweight" variable does not have any missing values, as all values are 0 or 1. If the BMI variable (v445) had missing values, in our new variable those have now been converted to 0s. Thus, it is important to transfer the missing values from the original variable v445 to our underweight variable by typing: replace underweight=. If v445==.This will replace all missing values with a dot. If we fail to make this conversion, we will be over-estimating the 0s, or non-underweight people.

vii. Finally, the variable of interest, i.e. underweight, can be tabulated to see the frequency distributions and the proportion of women who are underweight. Appropriate weights must be used to correct our sample. This can be done by typing tab underweight [iw=v005/1000000]

When performing this analysis utilizing data from the Nepal DHS 2016, the results obtained should look as shown in Figure 14.

*Figure 14: Results table showing proportion of underweight women and girls in Nepal, DHS 2016*

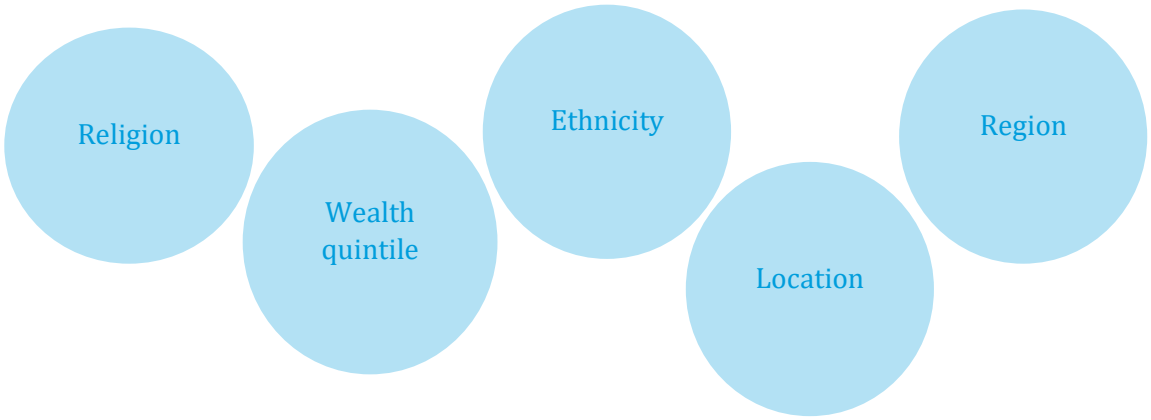
| underweight | Freq. | Percent | Cum. |
|---|---|---|---|
| 0 | 3,821.3473 | 84.41 | 84.41 |
| 1 | 705.680279 | 15.59 | 100.00 |
| Total | 4,527.0276 | 100.00 | |

The proportion of women and girls aged 18–49 who are were underweight at the time of the survey is 15.59 per cent.

### 2.3.2. Part 2: Disaggregating the variable by multiple socioeconomic factors relevant in the national context

As explained in section 2.3 above, the first step to conducting this analysis is to undertake a context-check to see which socioeconomic factors are relevant and could be potential drivers of inequality. These will be used as disaggregation variables.

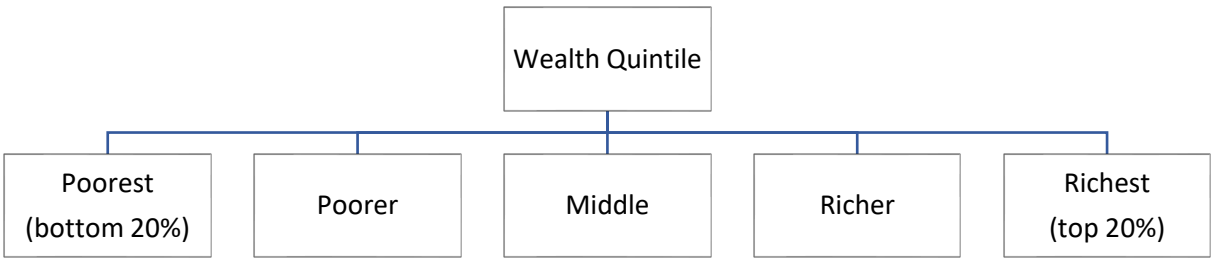
*Figure 15: Examples of context-specific drivers of inequality (sociodemographic variables)*




To show how different population groups perform regarding different development outcomes, the next step is to disaggregate each of the development indicators utilizing some of these sociodemographic variables. Let's take a look at some of the more common disaggregation variables that are often drivers of discrimination and, therefore, deprivation for women:

*Wealth*

*Figure 16: Disaggregation across wealth groups*



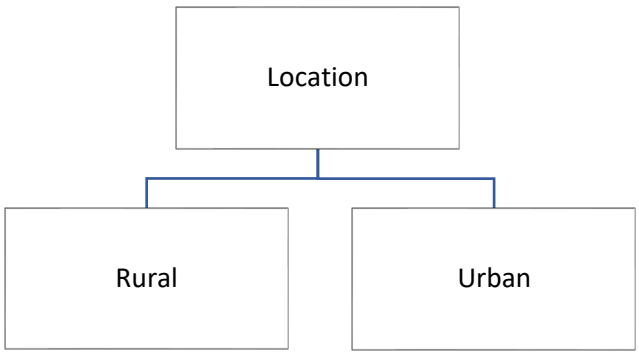

DHS and MICS surveys do not include individual measures of wealth based on income or consumption. However, a wealth index is included that classifies households based on their wealth quintile. That is, the total amount of households has been ranked according to a wealth index, and divided into five groups, each holding 20 per cent of the households. The bottom 20 per cent are the poorest households and top 20 per cent are the richest households. For LNOB analysis, we make an assumption that a woman would be poor if she lives in a household in the poorest quintile. To see the chances for a poor versus a rich woman to be undernourished, use the following code: tab underweight v190 [iw=v005/1000000], cell column row, where v190 is the wealth index variable. Again, for the Nepal DHS 2016, the results would be as follows:

*Figure 17: Results table showing a tabulation of the wealth and undernourishment variables for women*

| underweight | wealth index combined | | | | | |
|---|---|---|---|---|---|---|
| | poorest | poorer | middle | richer | richest | Total |
| 0 | 629.059301 | 702.980134 | 759.211265 | 840.485511 | 889.611132 | 3,821.347 |
| | 16.46 | 18.40 | 19.87 | 21.99 | 23.28 | 100.00 |
| | 82.49 | 81.15 | 80.77 | 83.76 | 93.17 | 84.41 |
| | 13.90 | 15.53 | 16.77 | 18.57 | 19.65 | 84.41 |
| 1 | 133.517904 | 163.242056 | 180.702223 | 163.000407 | 65.217689 | 705.680279 |
| | 18.92 | 23.13 | 25.61 | 23.10 | 9.24 | 100.00 |
| | 17.51 | 18.85 | 19.23 | 16.24 | 6.83 | 15.59 |
| | 2.95 | 3.61 | 3.99 | 3.60 | 1.44 | 15.59 |
| Total | 762.577205 | 866.22219 | 939.913488 | 1,003.486 | 954.828821 | 4,527.028 |
| | 16.84 | 19.13 | 20.76 | 22.17 | 21.09 | 100.00 |
| | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| | 16.84 | 19.13 | 20.76 | 22.17 | 21.09 | 100.00 |

*Location*

*Figure 18: Disaggregation across location of residence*




The location variable in DHS and MICS divides the households into rural and urban households. To see the effect that women's residence location has on being underweight, use the following code: tab underweight v025 [iw=v005/1000000], cell column row, where v025 is the recode variable for location of household. The result for the Nepal DHS 2016 would be:

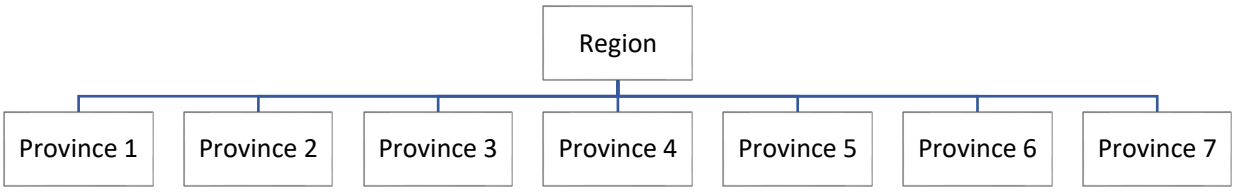
*Figure 19: Results table showing tabulation of location and underweight for women and girls*

| underweight | type of place of residence: urban | rural | Total |
|---|---|---|---|
| 0 | 2,435.424 | 1,385.923 | 3,821.347 |
| | 63.73 | 36.27 | 100.00 |
| | 85.59 | 82.41 | 84.41 |
| | 53.80 | 30.61 | 84.41 |
| 1 | 409.958972 | 295.721307 | 705.680279 |
| | 58.09 | 41.91 | 100.00 |
| | 14.41 | 17.59 | 15.59 |
| | 9.06 | 6.53 | 15.59 |
| Total | 2,845.383 | 1,681.645 | 4,527.028 |
| | 62.85 | 37.15 | 100.00 |
| | 100.00 | 100.00 | 100.00 |
| | 62.85 | 37.15 | 100.00 |

*Region*

The main geographical or administrative regions of a country vary. For example, the regions included in the Nepal DHS 2016 are represented below:

*Figure 20: Disaggregation across region of residence, Nepal DHS 2016*




Note that the DHS asks questions to both '*de jure* members of a household' as well as '*non-de-jure* members of a household'. For this analysis, only women who are *de jure* members of the household are kept. To drop the *non-de-jure* members, use the following code: drop if v139==97, where v139 is the recode variable for region of residence and 97 is the "*non-de-jure* residents" category. Note, however, that different DHS surveys might ascribe different values to this category (e.g. it is not always value 97), so it is important to always check first[21].

[21] This can be achieved by typing codebook v139, tab(100)

To see how women's region of residence affects their chances of low BMI, use the following code: tab underweight v139 [iw=v005/1000000], cell column row. The result obtained for the Nepal DHS 2016 is:

*Figure 21: Results table showing tabulation of province (1,2,3,4) and underweight for women in Nepal*

| underweight | de jure region of residence province | province | province | province | Total |
|---|---|---|---|---|---|
| 0 | 692.498611 | 623.469617 | 856.854652 | 380.493441 | 3,619.026 |
| | 19.13 | 17.23 | 23.68 | 10.51 | 100.00 |
| | 89.47 | 72.09 | 88.86 | 92.76 | 84.28 |
| | 16.13 | 14.52 | 19.95 | 8.86 | 84.28 |
| 1 | 81.498591 | 241.33623 | 107.400897 | 29.712909 | 675.210079 |
| | 12.07 | 35.74 | 15.91 | 4.40 | 100.00 |
| | 10.53 | 27.91 | 11.14 | 7.24 | 15.72 |
| | 1.90 | 5.62 | 2.50 | 0.69 | 15.72 |
| Total | 773.997202 | 864.805847 | 964.255549 | 410.20635 | 4,294.236 |
| | 18.02 | 20.14 | 22.45 | 9.55 | 100.00 |
| | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| | 18.02 | 20.14 | 22.45 | 9.55 | 100.00 |

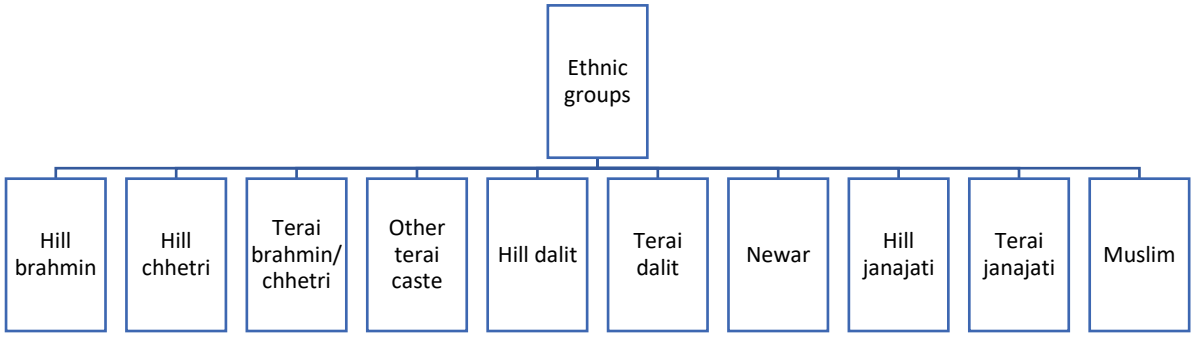
*Figure 22: Results table showing tabulation of province (5, 6, 7) and underweight for women in Nepal*

| underweight | de jure region of residence province | province | province | Total |
|---|---|---|---|---|
| 0 | 560.500749 | 202.284928 | 302.924011 | 3,619.026 |
| | 15.49 | 5.59 | 8.37 | 100.00 |
| | 83.67 | 85.39 | 80.96 | 84.28 |
| | 13.05 | 4.71 | 7.05 | 84.28 |
| 1 | 109.413865 | 34.622823 | 71.224764 | 675.210079 |
| | 16.20 | 5.13 | 10.55 | 100.00 |
| | 16.33 | 14.61 | 19.04 | 15.72 |
| | 2.55 | 0.81 | 1.66 | 15.72 |
| Total | 669.914614 | 236.907751 | 374.148775 | 4,294.236 |
| | 15.60 | 5.52 | 8.71 | 100.00 |
| | 100.00 | 100.00 | 100.00 | 100.00 |
| | 15.60 | 5.52 | 8.71 | 100.00 |

*Ethnicity*

Questions about ethnicity aren't consistently asked in DHS and MICS surveys in all countries. While some DHS and MICS surveys do ask about ethnicity, others use proxy variables such as people's native language and language of the household head. Although not ideal, these proxy variables are sometimes used for LNOB analysis if ethnicity is not available.

*Figure 23: Disaggregation across ethnicity, Nepal DHS 2016*




To see the effect that women's ethnicity has on BMI, use the following code: tab underweight v131 [iw=v005/1000000], cell column row, where v131 is the ethnicity variable.

*Figure 24: Results table showing tabulation of ethnicity and underweight status for women in Nepal*

| underweight | ethnicity | | | | |
|---|---|---|---|---|---|
| | hill dali | terai dal | newar | hill jana | Total |
| 0 | 293.416228 | 114.455658 | 199.417375 | 863.703528 | 3,619.026 |
| | 8.11 | 3.16 | 5.51 | 23.87 | 100.00 |
| | 87.89 | 59.76 | 91.16 | 91.20 | 84.28 |
| | 6.83 | 2.67 | 4.64 | 20.11 | 84.28 |
| 1 | 40.425902 | 77.07466 | 19.328277 | 83.355975 | 675.210079 |
| | 5.99 | 11.41 | 2.86 | 12.35 | 100.00 |
| | 12.11 | 40.24 | 8.84 | 8.80 | 15.72 |
| | 0.94 | 1.79 | 0.45 | 1.94 | 15.72 |
| Total | 333.84213 | 191.530318 | 218.745652 | 947.059503 | 4,294.236 |
| | 7.77 | 4.46 | 5.09 | 22.05 | 100.00 |
| | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| | 7.77 | 4.46 | 5.09 | 22.05 | 100.00 |

*Figure 25: Results table showing tabulation of ethnicity and underweight status for women in Nepal, continued*

| underweight | ethnicity | | | | |
|---|---|---|---|---|---|
| | hill brah | hill chhe | terai bra | other ter | Total |
| 0 | 456.155249 | 665.375611 | 50.651217 | 448.097797 | 3,619.026 |
| | 12.60 | 18.39 | 1.40 | 12.38 | 100.00 |
| | 87.97 | 87.41 | 76.67 | 75.40 | 84.28 |
| | 10.62 | 15.49 | 1.18 | 10.43 | 84.28 |
| 1 | 62.403599 | 95.803497 | 15.415849 | 146.186339 | 675.210079 |
| | 9.24 | 14.19 | 2.28 | 21.65 | 100.00 |
| | 12.03 | 12.59 | 23.33 | 24.60 | 15.72 |
| | 1.45 | 2.23 | 0.36 | 3.40 | 15.72 |
| Total | 518.558848 | 761.179108 | 66.067066 | 594.284136 | 4,294.236 |
| | 12.08 | 17.73 | 1.54 | 13.84 | 100.00 |
| | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| | 12.08 | 17.73 | 1.54 | 13.84 | 100.00 |

*Figure 26: Results table showing tabulation of ethnicity and underweight status for women in Nepal, continued again*

| underweight | ethnicity | | | |
|---|---|---|---|---|
| | terai jan | muslim | other | Total |
| 0 | 359.909073 | 149.390564 | 18.453709 | 3,619.026 |
| | 9.94 | 4.13 | 0.51 | 100.00 |
| | 81.97 | 73.72 | 86.75 | 84.28 |
| | 8.38 | 3.48 | 0.43 | 84.28 |
| 1 | 79.150517 | 53.248104 | 2.81736 | 675.210079 |
| | 11.72 | 7.89 | 0.42 | 100.00 |
| | 18.03 | 26.28 | 13.25 | 15.72 |
| | 1.84 | 1.24 | 0.07 | 15.72 |
| Total | 439.05959 | 202.638668 | 21.271069 | 4,294.236 |
| | 10.22 | 4.72 | 0.50 | 100.00 |
| | 100.00 | 100.00 | 100.00 | 100.00 |
| | 10.22 | 4.72 | 0.50 | 100.00 |

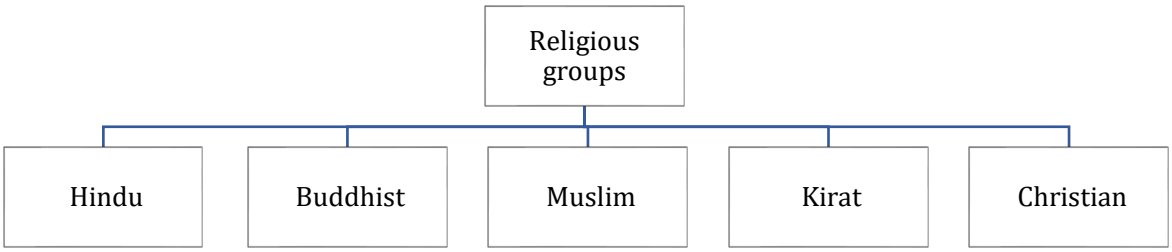
This analysis shows that, while 8.8 per cent of the women belonging to the Hill Janajati ethnic group are undernourished, the likelihood among women in the Terai Dalit ethnic group amounts to nearly 40 per cent. Note that for this type of multilevel disaggregation, it is crucial to consider the sample size of each population group, to ensure that the estimates are reliable. In this example, the sample size is less than 100 for Terai Brahmin and less than 30 for the category “Others”. The estimates for these ethnicities are thus less reliable. For the statistical analysis in this guide, a sample size of more than 45 has been

considered acceptable. Any results from sample groups smaller than 45 have been dropped, and footnotes have been added to any results with samples smaller than 100. In addition, after the analysis is conducted, tests will be necessary to assess the reliability of the estimate. This will be covered in section 3.

### *Religion*

Looking at religion as a driver of inequality may be useful where there are multiple religious groups and enough people belong to each group. For instance, since Islam is the dominant religion in Bangladesh, with approximately 90 per cent of the population being Muslim (based on microdata from DHS Bangladesh 2014), considering religion as a driver of inequality may be difficult because the sample sizes for other religious groups will be very small at the multiple levels of disaggregation. Nevertheless, in other countries, it might be useful to check the effect of this variable on the SDG indicator being measured. The main religious groups in Nepal as per DHS 2016 are:

*Figure 24: Disaggregation across religion in Nepal, according to DHS 2016*

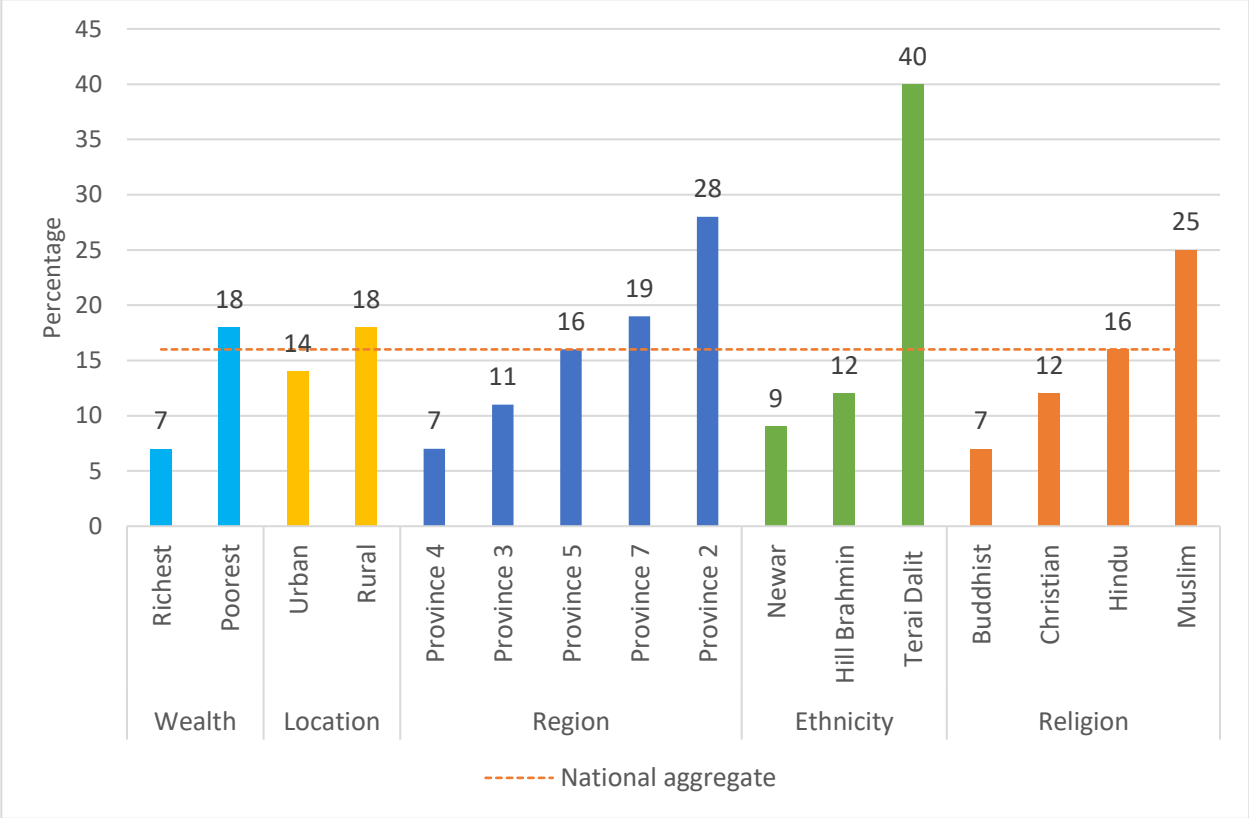



To see the effect of religion on women's likelihood of having a low BMI, use the following code: tab underweight v130 [iw=v005/1000000], cell column row, where v130 represents religion.

*Figure 25: Results table showing tabulation of religion and underweight for women in Nepal*

| underweight | religion | | | | | |
|---|---|---|---|---|---|---|
| | hindu | buddhist | muslim | kirat | christian | Total |
| 0 | 3,105.704 | 212.772173 | 146.088352 | 63.248392 | 91.212732 | 3,619.026 |
| | 85.82 | 5.88 | 4.04 | 1.75 | 2.52 | 100.00 |
| | 84.02 | 92.58 | 73.85 | 95.02 | 88.06 | 84.28 |
| | 72.32 | 4.95 | 3.40 | 1.47 | 2.12 | 84.28 |
| 1 | 590.725316 | 17.061886 | 51.7402 | 3.31762 | 12.365057 | 675.210079 |
| | 87.49 | 2.53 | 7.66 | 0.49 | 1.83 | 100.00 |
| | 15.98 | 7.42 | 26.15 | 4.98 | 11.94 | 15.72 |
| | 13.76 | 0.40 | 1.20 | 0.08 | 0.29 | 15.72 |
| Total | 3,696.43 | 229.834059 | 197.828552 | 66.566012 | 103.577789 | 4,294.236 |
| | 86.08 | 5.35 | 4.61 | 1.55 | 2.41 | 100.00 |
| | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| | 86.08 | 5.35 | 4.61 | 1.55 | 2.41 | 100.00 |

*Figure 26: proportion of women with low BMI in Nepal (DHS 2016)*




As shown in Figure 29, significant inequalities exist across population groups for women in Nepal regarding Body Mass Index. Further, women and girls who are at the intersection of two or more different forms of discrimination may face increased deprivation. For instance, women who are poor and also belong to an ethnic minority are likely to experience greater deprivation or disadvantage than women and girls who are either poor or belong to an ethnic minority. To see the effect of two or more forms of discrimination simultaneously, multiple disaggregation analysis needs to be conducted[22]. In order to achieve that in STATA, enter the commands below as follows:

### *Wealth and Location*

To see the effect of wealth and location simultaneously, write: by v025, sort: tab underweight v190 [iw=v005/1000000], cell column row, where v025 is the location variable and v190 is wealth index. The results for Nepal DHS 2016 would be:

[22] Refer to the methodology developed by Azcona & Duerto in UN Women 2018.

*Figure 27: Results table showing tabulation of underweight, wealth index and location (URBAN) for women in Nepal*

| underweight | wealth index combined | | | | | |
|---|---|---|---|---|---|---|
| | poorest | poorer | middle | richer | richest | Total |
| 0 | 211.834289 | 361.377801 | 436.317845 | 542.496992 | 753.006391 | 2,305.033 |
| | 9.19 | 15.68 | 18.93 | 23.54 | 32.67 | 100.00 |
| | 76.55 | 76.84 | 84.33 | 86.21 | 93.43 | 85.38 |
| | 7.85 | 13.39 | 16.16 | 20.09 | 27.89 | 85.38 |
| 1 | 64.892851 | 108.902453 | 81.079579 | 86.764642 | 52.993236 | 394.632761 |
| | 16.44 | 27.60 | 20.55 | 21.99 | 13.43 | 100.00 |
| | 23.45 | 23.16 | 15.67 | 13.79 | 6.57 | 14.62 |
| | 2.40 | 4.03 | 3.00 | 3.21 | 1.96 | 14.62 |
| Total | 276.72714 | 470.280254 | 517.397424 | 629.261634 | 805.999627 | 2,699.666 |
| | 10.25 | 17.42 | 19.17 | 23.31 | 29.86 | 100.00 |
| | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| | 10.25 | 17.42 | 19.17 | 23.31 | 29.86 | 100.00 |

The proportion of the richest women and girls who live in an urban location and are undernourished is 6.57 per cent. The proportion of the poorest women and girls who live in an urban location and are undernourished is 23.45 per cent. This means they are almost four times as likely as the urban richest women and girls to be undernourished.

*Figure 28: Results table showing tabulation of underweight, wealth index and location (RURAL) for women in Nepal*

| underweight | wealth index combined | | | | | |
|---|---|---|---|---|---|---|
| | poorest | poorer | middle | richer | richest | Total |
| 0 | 390.630914 | 308.260462 | 300.436845 | 236.124752 | 78.539718 | 1,313.993 |
| | 29.73 | 23.46 | 22.86 | 17.97 | 5.98 | 100.00 |
| | 85.17 | 86.17 | 77.56 | 77.37 | 91.68 | 82.40 |
| | 24.50 | 19.33 | 18.84 | 14.81 | 4.93 | 82.40 |
| 1 | 68.024899 | 49.461961 | 86.915166 | 69.05145 | 7.123842 | 280.577318 |
| | 24.24 | 17.63 | 30.98 | 24.61 | 2.54 | 100.00 |
| | 14.83 | 13.83 | 22.44 | 22.63 | 8.32 | 17.60 |
| | 4.27 | 3.10 | 5.45 | 4.33 | 0.45 | 17.60 |
| Total | 458.655813 | 357.722423 | 387.352011 | 305.176202 | 85.66356 | 1,594.57 |
| | 28.76 | 22.43 | 24.29 | 19.14 | 5.37 | 100.00 |
| | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| | 28.76 | 22.43 | 24.29 | 19.14 | 5.37 | 100.00 |

The proportion of the richest women and girls who live in a rural location and are undernourished is 8.32 per cent. The proportion of the poorest women and girls who live in an urban location and are undernourished is 14.83 per cent.

In the case of Nepal, the most nutritionally deprived women and girls belong to the urban-poorest group.

*Region and Location*

To see the effect of region and location simultaneously, type drop if v139==97, to only keep *de jure* members of the household and then type by v025, sort: tab underweight v139 [iw=v005/1000000], cell column row, where v025 is the location variable and v139 is the region variable. The results for the Nepal DHS 2016 are:

*Figure 29: Results table showing tabulation of underweight, region and location (URBAN) for women in Nepal*

| underweight | de jure region of residence | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | province | province | province | province | province | province | province | Total |
| 0 | 440.150254 | 327.973856 | 684.920657 | 247.440987 | 324.345324 | 99.077558 | 181.124682 | 2,305.033 |
| | 19.10 | 14.23 | 29.71 | 10.73 | 14.07 | 4.30 | 7.86 | 100.00 |
| | 90.85 | 74.18 | 87.93 | 92.15 | 83.89 | 86.32 | 80.82 | 85.38 |
| | 16.30 | 12.15 | 25.37 | 9.17 | 12.01 | 3.67 | 6.71 | 85.38 |
| 1 | 44.349783 | 114.165001 | 94.054603 | 21.077013 | 62.305929 | 15.699106 | 42.981326 | 394.632761 |
| | 11.24 | 28.93 | 23.83 | 5.34 | 15.79 | 3.98 | 10.89 | 100.00 |
| | 9.15 | 25.82 | 12.07 | 7.85 | 16.11 | 13.68 | 19.18 | 14.62 |
| | 1.64 | 4.23 | 3.48 | 0.78 | 2.31 | 0.58 | 1.59 | 14.62 |
| Total | 484.500037 | 442.138857 | 778.97526 | 268.518 | 386.651253 | 114.776664 | 224.106008 | 2,699.666 |
| | 17.95 | 16.38 | 28.85 | 9.95 | 14.32 | 4.25 | 8.30 | 100.00 |
| | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| | 17.95 | 16.38 | 28.85 | 9.95 | 14.32 | 4.25 | 8.30 | 100.00 |

Women and girls who live in the urban location of Province 2 are approximately three times as likely as women and girls living in the urban location of Province 4 to be undernourished.

*Figure 30: Results table showing tabulation of underweight, region and location (RURAL) for women in Nepal*

| underweight | de jure region of residence | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | province | province | province | province | province | province | province | Total |
| 0 | 252.348357 | 295.495761 | 171.933995 | 133.052454 | 236.155425 | 103.20737 | 121.799329 | 1,313.993 |
| | 19.20 | 22.49 | 13.08 | 10.13 | 17.97 | 7.85 | 9.27 | 100.00 |
| | 87.17 | 69.91 | 92.80 | 93.91 | 83.37 | 84.51 | 81.18 | 82.40 |
| | 15.83 | 18.53 | 10.78 | 8.34 | 14.81 | 6.47 | 7.64 | 82.40 |
| 1 | 37.148808 | 127.171229 | 13.346294 | 8.635896 | 47.107936 | 18.923717 | 28.243438 | 280.577318 |
| | 13.24 | 45.32 | 4.76 | 3.08 | 16.79 | 6.74 | 10.07 | 100.00 |
| | 12.83 | 30.09 | 7.20 | 6.09 | 16.63 | 15.49 | 18.82 | 17.60 |
| | 2.33 | 7.98 | 0.84 | 0.54 | 2.95 | 1.19 | 1.77 | 17.60 |
| Total | 289.497165 | 422.66699 | 185.280289 | 141.68835 | 283.263361 | 122.131087 | 150.042767 | 1,594.57 |
| | 18.16 | 26.51 | 11.62 | 8.89 | 17.76 | 7.66 | 9.41 | 100.00 |
| | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| | 18.16 | 26.51 | 11.62 | 8.89 | 17.76 | 7.66 | 9.41 | 100.00 |

Women and girls who live in the rural location of Province 2 are approximately five times as likely as women and girls living in the rural location of Province 4 to be undernourished.

### *Ethnicity and wealth*

To see the effect of ethnicity and wealth simultaneously, type by v131, sort: tab underweight v190 [iw=v005/1000000], cell column row, where v131 is the recode variable for ethnicity and v190 is the recode variable for wealth index. Results for the Nepal DHS 2016 are:

*Figure 31: Results table showing tabulation of underweight, wealth index (POOREST) and ethnicity for women in Nepal*

| underweight | ethnicity | | | | | | |
|---|---|---|---|---|---|---|---|
| | hill brah | hill chhe | terai bra | other ter | hill dali | terai dal | Total |
| 0 | 34.839262 | 188.877683 | 1.806109 | 7.273692 | 100.347408 | 4.227325 | 602.465203 |
| | 5.78 | 31.35 | 0.30 | 1.21 | 16.66 | 0.70 | 100.00 |
| | 81.15 | 82.91 | 100.00 | 81.99 | 82.39 | 31.75 | 81.93 |
| | 4.74 | 25.68 | 0.25 | 0.99 | 13.65 | 0.57 | 81.93 |
| 1 | 8.093583 | 38.941738 | 0 | 1.597869 | 21.455159 | 9.086223 | 132.91775 |
| | 6.09 | 29.30 | 0.00 | 1.20 | 16.14 | 6.84 | 100.00 |
| | 18.85 | 17.09 | 0.00 | 18.01 | 17.61 | 68.25 | 18.07 |
| | 1.10 | 5.30 | 0.00 | 0.22 | 2.92 | 1.24 | 18.07 |
| Total | 42.932845 | 227.819421 | 1.806109 | 8.871561 | 121.802567 | 13.313548 | 735.382953 |
| | 5.84 | 30.98 | 0.25 | 1.21 | 16.56 | 1.81 | 100.00 |
| | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| | 5.84 | 30.98 | 0.25 | 1.21 | 16.56 | 1.81 | 100.00 |

| underweight | ethnicity | | | | | |
|---|---|---|---|---|---|---|
| | newar | hill jana | terai jan | muslim | other | Total |
| 0 | 17.490885 | 228.778804 | 16.375523 | 1.343058 | 1.105454 | 602.465203 |
| | 2.90 | 37.97 | 2.72 | 0.22 | 0.18 | 100.00 |
| | 96.69 | 85.00 | 67.21 | 21.94 | 100.00 | 81.93 |
| | 2.38 | 31.11 | 2.23 | 0.18 | 0.15 | 81.93 |
| 1 | .599179 | 40.377128 | 7.988923 | 4.777948 | 0 | 132.91775 |
| | 0.45 | 30.38 | 6.01 | 3.59 | 0.00 | 100.00 |
| | 3.31 | 15.00 | 32.79 | 78.06 | 0.00 | 18.07 |
| | 0.08 | 5.49 | 1.09 | 0.65 | 0.00 | 18.07 |
| Total | 18.090064 | 269.155932 | 24.364446 | 6.121006 | 1.105454 | 735.382953 |
| | 2.46 | 36.60 | 3.31 | 0.83 | 0.15 | 100.00 |
| | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| | 2.46 | 36.60 | 3.31 | 0.83 | 0.15 | 100.00 |

*Figure 32: Results table showing tabulation of underweight, wealth index (RICHEST) and ethnicity for women in Nepal*

| underweight | ethnicity | | | | | | |
|---|---|---|---|---|---|---|---|
| | hill brah | hill chhe | terai bra | other ter | hill dali | terai dal | Total |
| 0 | 192.433137 | 152.461275 | 19.618124 | 70.78451 | 33.056211 | 2.670797 | 831.546109 |
| | 23.14 | 18.33 | 2.36 | 8.51 | 3.98 | 0.32 | 100.00 |
| | 94.12 | 92.89 | 79.34 | 91.20 | 93.21 | 100.00 | 93.26 |
| | 21.58 | 17.10 | 2.20 | 7.94 | 3.71 | 0.30 | 93.26 |
| 1 | 12.032446 | 11.67535 | 5.106989 | 6.829464 | 2.407085 | 0 | 60.117078 |
| | 20.02 | 19.42 | 8.50 | 11.36 | 4.00 | 0.00 | 100.00 |
| | 5.88 | 7.11 | 20.66 | 8.80 | 6.79 | 0.00 | 6.74 |
| | 1.35 | 1.31 | 0.57 | 0.77 | 0.27 | 0.00 | 6.74 |
| Total | 204.465583 | 164.136625 | 24.725113 | 77.613974 | 35.463296 | 2.670797 | 891.663187 |
| | 22.93 | 18.41 | 2.77 | 8.70 | 3.98 | 0.30 | 100.00 |
| | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| | 22.93 | 18.41 | 2.77 | 8.70 | 3.98 | 0.30 | 100.00 |

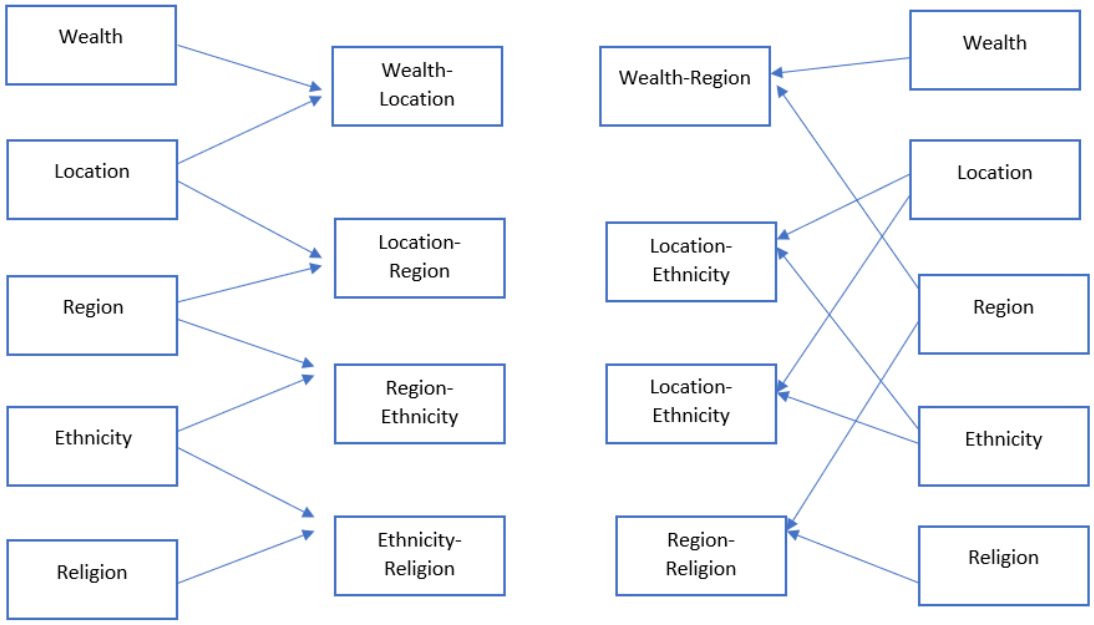
| underweight | ethnicity | | | | | |
|---|---|---|---|---|---|---|
| | newar | hill jana | terai jan | muslim | other | Total |
| 0 | 93.484093 | 183.092902 | 37.86171 | 32.832631 | 13.250719 | 831.546109 |
| | 11.24 | 22.02 | 4.55 | 3.95 | 1.59 | 100.00 |
| | 91.88 | 96.75 | 100.00 | 82.94 | 93.67 | 93.26 |
| | 10.48 | 20.53 | 4.25 | 3.68 | 1.49 | 93.26 |
| 1 | 8.26102 | 6.157839 | 0 | 6.751235 | .89565 | 60.117078 |
| | 13.74 | 10.24 | 0.00 | 11.23 | 1.49 | 100.00 |
| | 8.12 | 3.25 | 0.00 | 17.06 | 6.33 | 6.74 |
| | 0.93 | 0.69 | 0.00 | 0.76 | 0.10 | 6.74 |
| Total | 101.745113 | 189.250741 | 37.86171 | 39.583866 | 14.146369 | 891.663187 |
| | 11.41 | 21.22 | 4.25 | 4.44 | 1.59 | 100.00 |
| | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| | 11.41 | 21.22 | 4.25 | 4.44 | 1.59 | 100.00 |

This analysis shows that the proportion of richest women aged 18–49 who belong to the Other Terai ethnicity are almost three times as likely as the richest Hill Janajati women to be undernourished. Also note that the most disadvantaged group, i.e. the poorest Hill Dalit women are almost six times as likely as the richest Hill Janajati women to be undernourished.

Note that the estimates for various ethnic groups, both in the poorest and richest categories, such as the poorest Terai Janajati and Muslim women, could not be considered because the sample sizes did not exceed 45.

In addition to these three combinations of simultaneous deprivation, users are advised to conduct the analysis for other possible combinations:

*Figure 33: Overlapping discrimination between multiple socioeconomic variables*




The general syntax of the code to perform two-level simultaneous disaggregation for the underweight analysis is: by variable1, sort: tab underweight variable2 [iw=v005/1000000], cell column row

For instance, to disaggregate those who are underweight by region and religion, the code will be: by v139, sort: tab underweight v130 [iw=v005/1000000], cell column row (where v139 is the recode variable for region and v130 is the recode variable for religion).

When samples allow, performing multiple disaggregation by three or more variables simultaneously usually shows even larger inequalities.

### Wealth, location and region

To see the effect of wealth, location and region simultaneously on women's BMI, the code is as follows:

- Drop the non-*de-jure* residents by typing drop if v139==97.
- Calculate the share of underweight women in urban areas, by wealth and region by typing by v190, sort: tab underweight v139 [iw=v005/1000000] if v025==1, cell column row , where, v190 is the wealth variable, v139 is region variable and v025 is the location variable.

The results obtained for the Nepal DHS 2016 are:

*Figure 34: Results table showing tabulation of location (URBAN), wealth index (POOREST) and region for women in Nepal*

| underweight | de jure region of residence | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | province | province | province | province | province | province | province | Total |
| 0 | 57.064921 | 5.39926 | 33.455544 | 14.865044 | 26.525002 | 38.528351 | 35.996167 | 211.834289 |
| | 26.94 | 2.55 | 15.79 | 7.02 | 12.52 | 18.19 | 16.99 | 100.00 |
| | 83.55 | 37.49 | 63.94 | 95.28 | 82.74 | 80.89 | 77.55 | 76.55 |
| | 20.62 | 1.95 | 12.09 | 5.37 | 9.59 | 13.92 | 13.01 | 76.55 |
| 1 | 11.231903 | 9.002608 | 18.864311 | .736343 | 5.532071 | 9.104696 | 10.420919 | 64.892851 |
| | 17.31 | 13.87 | 29.07 | 1.13 | 8.52 | 14.03 | 16.06 | 100.00 |
| | 16.45 | 62.51 | 36.06 | 4.72 | 17.26 | 19.11 | 22.45 | 23.45 |
| | 4.06 | 3.25 | 6.82 | 0.27 | 2.00 | 3.29 | 3.77 | 23.45 |
| Total | 68.296824 | 14.401868 | 52.319855 | 15.601387 | 32.057073 | 47.633047 | 46.417086 | 276.72714 |
| | 24.68 | 5.20 | 18.91 | 5.64 | 11.58 | 17.21 | 16.77 | 100.00 |
| | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| | 24.68 | 5.20 | 18.91 | 5.64 | 11.58 | 17.21 | 16.77 | 100.00 |

The proportion of women and girls who are undernourished varies from 16.45 per cent for the urban-poorest in Province 1 to 36.06 per cent for the urban-poorest in Province 3.

*Figure 35: Results table showing tabulation of location (URBAN), wealth index (RICHEST) and region for women in Nepal*

| underweight | de jure region of residence | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | province | province | province | province | province | province | province | Total |
| 0 | 103.414669 | 70.095276 | 366.141588 | 86.191128 | 88.96947 | 6.777445 | 31.416815 | 753.006391 |
| | 13.73 | 9.31 | 48.62 | 11.45 | 11.82 | 0.90 | 4.17 | 100.00 |
| | 97.10 | 87.68 | 94.58 | 94.60 | 88.65 | 96.08 | 92.62 | 93.43 |
| | 12.83 | 8.70 | 45.43 | 10.69 | 11.04 | 0.84 | 3.90 | 93.43 |
| 1 | 3.084768 | 9.84537 | 20.967154 | 4.921355 | 11.395946 | .276784 | 2.501859 | 52.993236 |
| | 5.82 | 18.58 | 39.57 | 9.29 | 21.50 | 0.52 | 4.72 | 100.00 |
| | 2.90 | 12.32 | 5.42 | 5.40 | 11.35 | 3.92 | 7.38 | 6.57 |
| | 0.38 | 1.22 | 2.60 | 0.61 | 1.41 | 0.03 | 0.31 | 6.57 |
| Total | 106.499437 | 79.940646 | 387.108742 | 91.112483 | 100.365416 | 7.054229 | 33.918674 | 805.999627 |
| | 13.21 | 9.92 | 48.03 | 11.30 | 12.45 | 0.88 | 4.21 | 100.00 |
| | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| | 13.21 | 9.92 | 48.03 | 11.30 | 12.45 | 0.88 | 4.21 | 100.00 |

- Calculate the share of underweight women in rural areas, by wealth and region by v190, sort: tab underweight v139 [iw=v005/1000000] if v025==2, cell column row

The results obtained for the Nepal DHS 2016 are:

*Figure 36: Results table showing tabulation of location (RURAL), wealth index (POOREST) and region for women in Nepal*

| underweight | de jure region of residence | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | province | province | province | province | province | province | province | Total |
| 0 | 64.809015 | 1.343058 | 79.730016 | 52.812312 | 59.183461 | 85.259986 | 47.493066 | 390.630914 |
| | 16.59 | 0.34 | 20.41 | 13.52 | 15.15 | 21.83 | 12.16 | 100.00 |
| | 90.81 | 23.21 | 89.36 | 92.35 | 84.33 | 84.49 | 74.20 | 85.17 |
| | 14.13 | 0.29 | 17.38 | 11.51 | 12.90 | 18.59 | 10.35 | 85.17 |
| 1 | 6.561525 | 4.443804 | 9.489055 | 4.377859 | 10.993345 | 15.648381 | 16.51093 | 68.024899 |
| | 9.65 | 6.53 | 13.95 | 6.44 | 16.16 | 23.00 | 24.27 | 100.00 |
| | 9.19 | 76.79 | 10.64 | 7.65 | 15.67 | 15.51 | 25.80 | 14.83 |
| | 1.43 | 0.97 | 2.07 | 0.95 | 2.40 | 3.41 | 3.60 | 14.83 |
| Total | 71.37054 | 5.786862 | 89.219071 | 57.190171 | 70.176806 | 100.908367 | 64.003996 | 458.655813 |
| | 15.56 | 1.26 | 19.45 | 12.47 | 15.30 | 22.00 | 13.95 | 100.00 |
| | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| | 15.56 | 1.26 | 19.45 | 12.47 | 15.30 | 22.00 | 13.95 | 100.00 |

The proportion of women and girls deprived in nourishment varies from 2.90 per cent for the urban-richest women in Province 1 to 25.8 per cent for the rural-poorest women in Province 1. Disaggregation at multiple levels thus reveals the inequalities and large achievement gaps between groups of women.

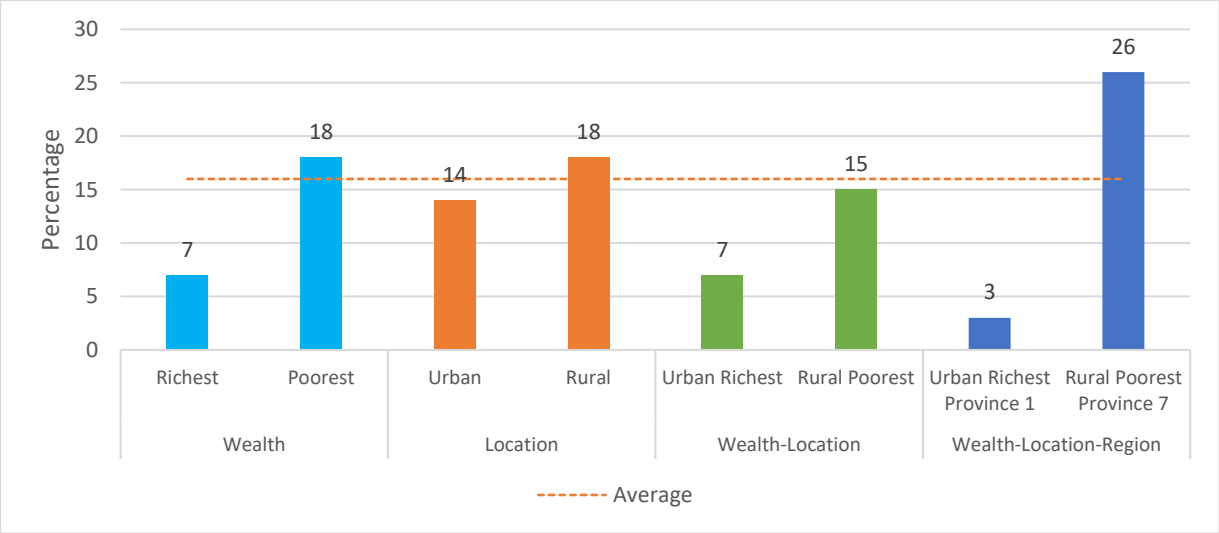
*Figure 37: Results table showing tabulation of location (RURAL), wealth index (RICHEST) and region for underweight women in Nepal*

| underweight | de jure region of residence | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | province | province | province | province | province | province | province | Total |
| 0 | 15.408137 | 16.41769 | 10.442654 | 5.213818 | 21.732854 | 1.551768 | 7.772797 | 78.539718 |
| | 19.62 | 20.90 | 13.30 | 6.64 | 27.67 | 1.98 | 9.90 | 100.00 |
| | 100.00 | 83.50 | 100.00 | 100.00 | 84.85 | 100.00 | 100.00 | 91.68 |
| | 17.99 | 19.17 | 12.19 | 6.09 | 25.37 | 1.81 | 9.07 | 91.68 |
| 1 | 0 | 3.244184 | 0 | 0 | 3.879658 | 0 | 0 | 7.123842 |
| | 0.00 | 45.54 | 0.00 | 0.00 | 54.46 | 0.00 | 0.00 | 100.00 |
| | 0.00 | 16.50 | 0.00 | 0.00 | 15.15 | 0.00 | 0.00 | 8.32 |
| | 0.00 | 3.79 | 0.00 | 0.00 | 4.53 | 0.00 | 0.00 | 8.32 |
| Total | 15.408137 | 19.661874 | 10.442654 | 5.213818 | 25.612512 | 1.551768 | 7.772797 | 85.66356 |
| | 17.99 | 22.95 | 12.19 | 6.09 | 29.90 | 1.81 | 9.07 | 100.00 |
| | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| | 17.99 | 22.95 | 12.19 | 6.09 | 29.90 | 1.81 | 9.07 | 100.00 |

The rural-poorest women living in Province 7 are almost nine times as likely as the urban-richest women living in Province 1 to be deprived – in this case, to have a low BMI. The compound effect of wealth, location and region of residence makes some groups of women lag behind others. To achieve progress for all and to leave no one behind, these women and girls need to be identified and policies must be designed that are inclusive of these groups.

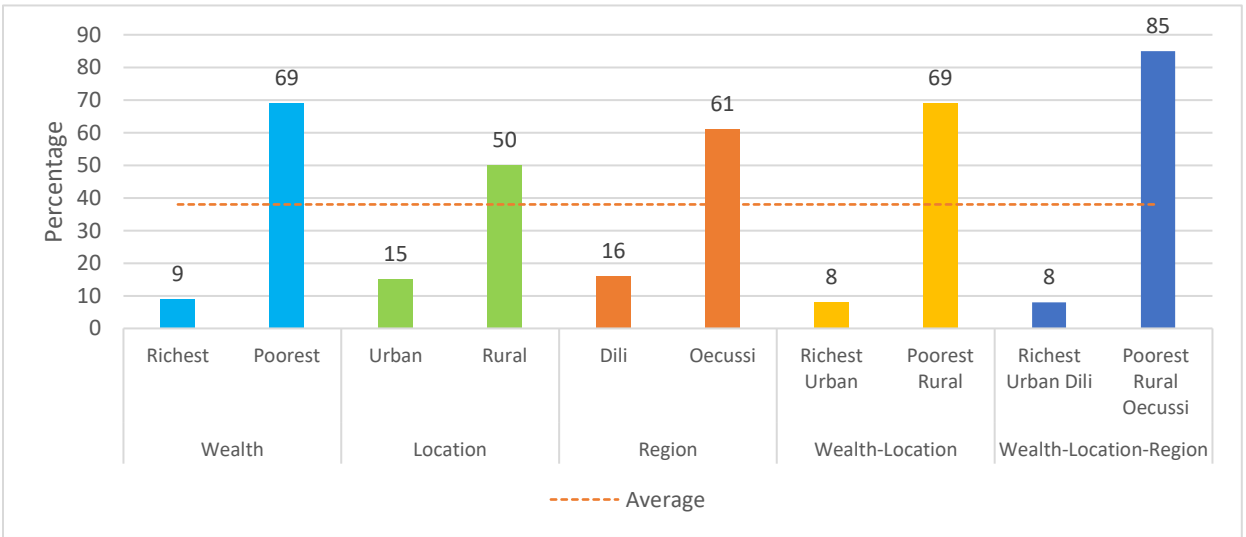
This analysis shows that, in Nepal, the compounding of wealth, location and religion shows the deepest pockets of deprivation.

*Figure 38: Proportion of women undernourished in Nepal, different population groups (DHS 2016)*



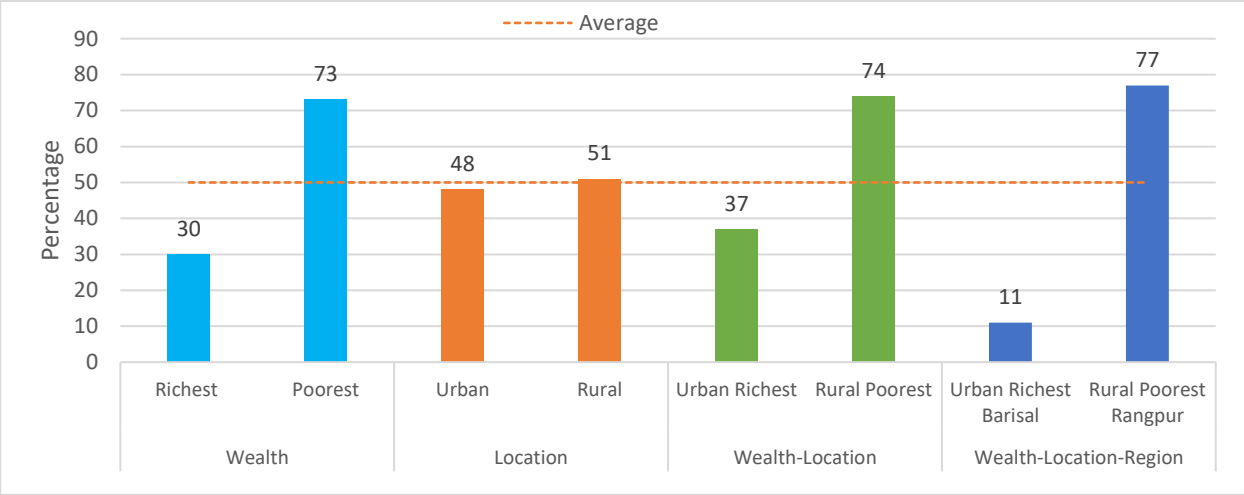

Similar analysis carried out for different countries shows, consistently, large inequalities and groups of women that are victims of multiple forms of discrimination lagging behind on development outcomes. The graph below shows the vast gaps in achievement on education, measured by the proportion of women and girls who are education-poor (i.e. completed six or less years of education) in Timor-Leste. The poorest rural women living in Oecussi are 10 times as likely as the richest urban women living in Dili to be education poor.

*Figure 39: Proportion of women and girls with six or less years of education, Timor Leste (DHS 2013)*



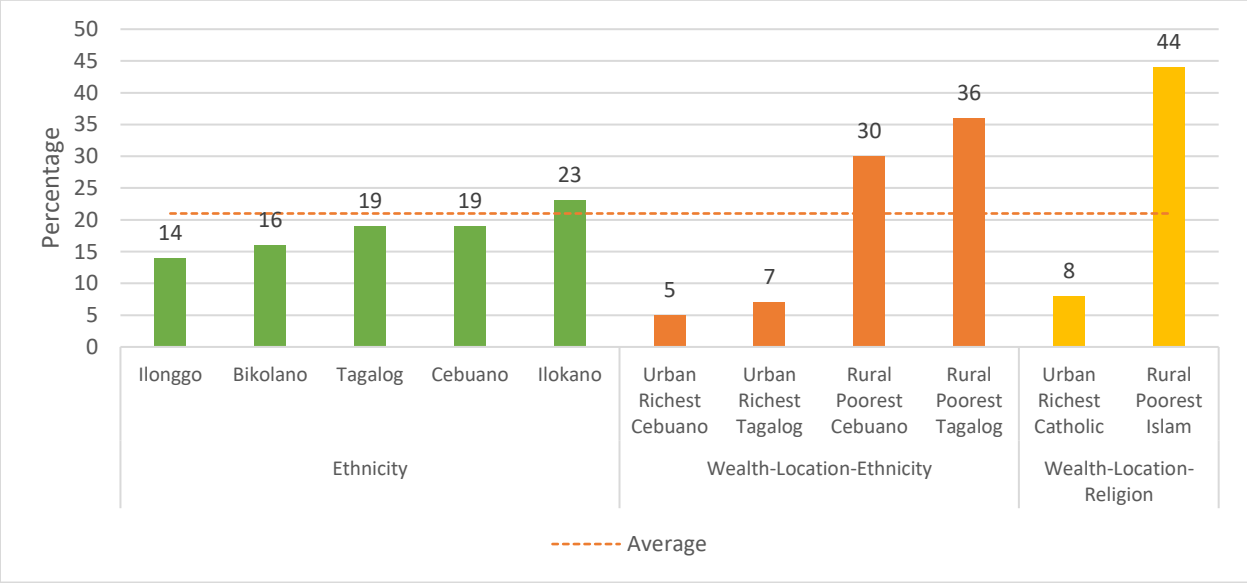
Similarly, the graph below shows the vast gaps in access to basic sanitation facilities by different groups of women and girls in Bangladesh. Basic sanitation facilities are defined as having an improved type of toilet which is not shared by other households (see section 2.3.3. for detailed description and analysis of this indicator). The compounded effect of wealth, location and region significantly deprives women in accessing basic sanitation facilities in Bangladesh. The poorest rural living in Rangpur are 7 times as likely as the urban richest women and girls living in Barisal to be deprived of basic sanitation facilities.

*Figure 40: Proportion of women and girls who lack access to basic sanitation facilities, Bangladesh*




The compounded effect of wealth, location and religion leads to a significant increase in the risk of child marriage in the Philippines. Additionally, the combined effect of wealth, location and ethnicity also increases the risk of child marriage. The poorest rural women belonging to the Tagalog ethnic group are seven times as likely as the richest urban Cebuano women to get married as a child.

*Figure 44: Proportion of women who married as children in the Philippines (DHS 2017)*

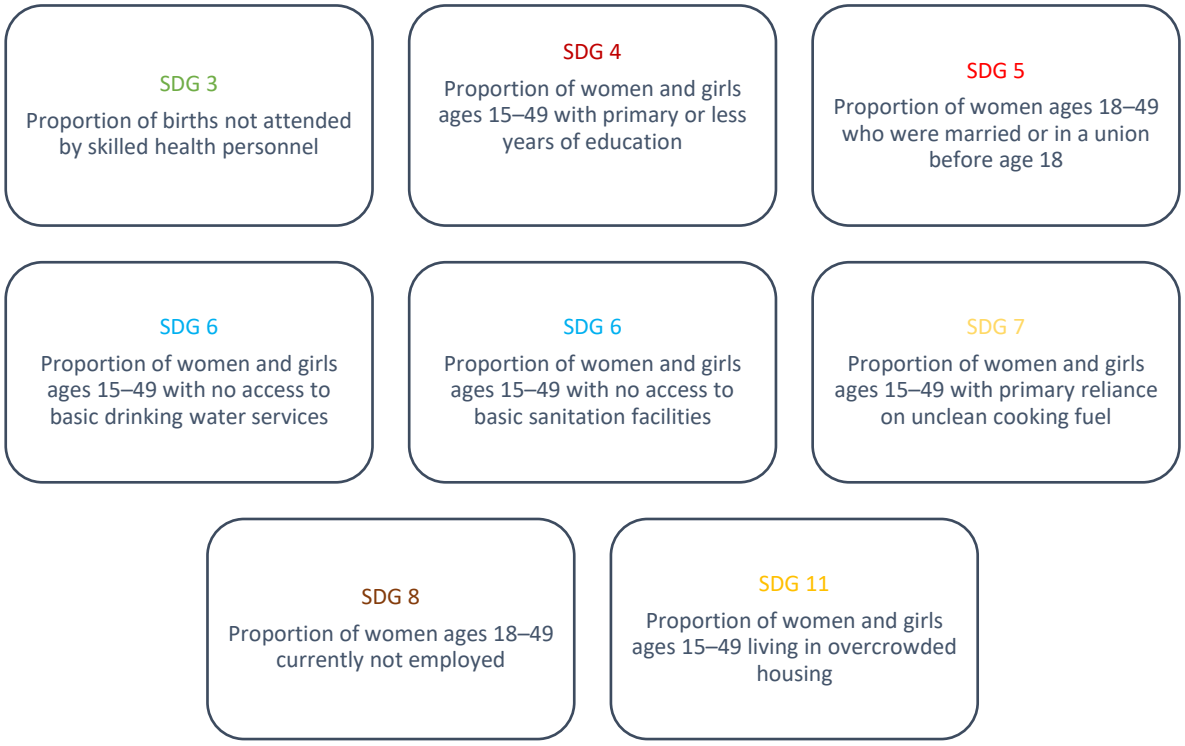


### 2.3.3. Part 3: Identifying the total population who lags behind consistently across multiple indicators

Deprivation in one area is often accompanied by deprivation in another[23]. Analysis often shows that the groups of women who lag furthest behind for one development indicator, are likely to lag behind across many others. Thus, to identify who are these multiply deprived groups of women, it is important to perform an analysis for multiple indicators and population groups simultaneously. To do this, the following section focuses on analysis for eight SDG-related indicators and demonstrates the codes needed to undertake this analysis.

The eight SDG indicators (or proxy SDG indicators in cases where the official indicator could not be calculated with the data source used for this analysis) selected for this analysis are depicted in Figure 45. To showcase deprivations, some SDG indicators have been inverted (e.g. from population with births attended by skilled professionals to population with births not attended):

*Figure 45: Select SDG indicators for upcoming analysis*




[23] UN Women 2018.

The following pages provide step-by-step guidance on how to replicate multilevel disaggregation analysis for different indicators. For each indicator, survey variable information (Bangladesh DHS 2014 and Mongolia MICS 2014) and STATA code lines are provided. As the examples of multilevel disaggregation replicate the logic of the Nepal case study presented above, the description of the code lines and how to interpret results will be a lot more concise in this section.

### *SDG 3: Proportion of births not attended by skilled health personnel (births in last five years)*

To calculate the proportion of births NOT attended by skilled health personnel, refer to the detailed information in Table 1:

*Table 1: Codes to calculate births not attended by skilled health personnel*

| | Demographic and Health Survey (Bangladesh 2014) | Multiple Indicator Cluster Survey (Mongolia 2013–2014) |
|---|---|---|
| Recode File | Birth Recode File (BR) | Women's File (wm) |
| Recode variables | - Births in last 5 years (v208)<br>- Assistance at delivery by skilled personnel:<br>○ m3a: assistance by qualified doctor<br>○ m3b: assistance by midwife/nurse/paramedic<br>○ m3c: assistance by family welfare visitor<br>○ m3d: assistance by community skilled birth attendant<br>- Women's sample weights (v005) | - Births in last 2 years (CM13)[24]<br>- Assistance at delivery by skilled personnel:<br>○ MN17D=assistance by gynecologist,<br>○ MN17E=assistance by physician<br>○ MN17I=assistance by family or soum doctor<br>○ MN17J=assistance by midwife<br>○ MN17C=assistance by auxiliary midwife<br>○ MN17K=assistance by nurse<br>○ MN17F=assistance by traditional practitioner[25]<br>- Women's sample weights (wmweight) |

[24] MICS collects information about births in last two years instead of births in last five years.

[25] In most countries, traditional practitioners are not considered skilled attendants. However, as there are differences across countries, it is important to check the indicator metadata in each case. In Mongolia, the national MICS report indicates that skilled attendant includes all health personnel except relatives and friends.

| Note | Note: This recode variable as well as who is classified as a skilled health personnel is country-specific. User is advised to check the recode file as well as the country final report to confirm. | Note: the recode variables can be in string format and must be converted to numerical format[26]. |
|---|---|---|
| Code | - Drop women who have not given birth in the last 5 years and replace missing values by typing drop if v208==0, followed by drop if v208==.<br>- Generate a new variable named 'skilled' by typing generate skilled=0<br>- Replace the value of 0 with 1 in those cases in which a skilled health personnel attended the delivery. This can be done by typing replace skilled=1 if m3a==1\|m3b==1\|m3c==1\|m3d==1<br>- Replace missing values by typing replace skilled=. if m3a==.\|m3b==.\|m3c==\|m3d==.<br>- Tabulate the results using appropriate weights, by typing tab skilled [iw=v005/1000000] | - Only keep women who have children and replace missing values by typing drop if CM10 == 0 followed by drop if CM10 == .<br>- To only keep births in the last 2 years, type drop if CM13 == 2 followed by drop if CM13 ==.<br>- Convert the recode variables from string format to numerical format using STATA's destring command by typing destring CM10, replace, etc.<br>- Generate a new variable named 'skilled' by typing generate skilled=0<br>- Replace the value of 0 with 1 in those cases in which a skilled health personnel attended the delivery. This can be done by typing replace skilled=1 if gynecologist==1\|physician==1\|familyorsoumdoctor==1\|midwife==1<br>- Tabulate the variable to see the proportion of births attended by skilled personnel by typing tab skilled [iw=wmweight] |

The results for the DHS Bangladesh 2014 will be:

[26] To convert a variable from string into numeric or byte formats, use STATA's command: destring VARIABLE, replace.

*Figure 46: Proportion of unskilled deliveries, Bangladesh (DHS 2014)*

| skilled | Freq. | Percent | Cum. |
|---|---|---|---|
| 0 | 2,831.8727 | 57.86 | 57.86 |
| 1 | 2,062.4799 | 42.14 | 100.00 |
| Total | 4,894.3526 | 100.00 | |

The results for MICS Mongolia 2013–14 will be:

*Figure 417: Proportion of unskilled deliveries, Mongolia (MICS 2013–14)*

| skilled | Freq. | Percent | Cum. |
|---|---|---|---|
| 0 | 52.5685441 | 2.20 | 2.20 |
| 1 | 2,336.4412 | 97.80 | 100.00 |
| Total | 2,389.0097 | 100.00 | |

Since we want to know the proportion of births NOT attended by a skilled health practitioner, we will look at the row alongside 0. As figure 46 shows, 57.86 per cent of the births that took place in the 5 years prior to the survey in Bangladesh (based on DHS 2014) were not attended by a skilled health practitioner. The rate for births not attended by a skilled health personnel (at 2.2 per cent) are much lower in Mongolia[27].

### *SDG 4: Proportion of women and girls ages 15–49 that are education-poor*

Women and girls who have acquired primary or less years of education (taken as six or less years of education) are classified as education-poor. To calculate the proportion of women and girls ages 15–49 years that are education-poor, refer to the information in Table 2.

*Table 2: Codes to calculate the proportion of women and girls who are education-poor*

| | Demographic and Health Survey | Multiple Indicator Cluster Survey |
|---|---|---|
| Recode File | Individual Recode File (IR) | Women's File (wm) |
| Recode variables | - Education in single years (v133) | - Level of education (welevel)[28] |

[27] Births in last two years preceding the survey.

[28] Level of education is recoded as fixed categories (primary, secondary, college, etc.) in MICS Mongolia as opposed to single years of schooling in DHS Bangladesh.

| | | |
|---|---|---|
| | - Women's sample weight (v005) | - Women's sample weight (wmweight) |
| Code | - Generate a new variable named 'primaryorless' by typing generate primaryorless=0<br>- Replace the value of 0 with 1 if single years of education is less than 7 years. This can be done by typing replace primaryorless=1 if v133<7<br>- Replace missing values by typing replace primaryorless=. if v133==.<br>- Tabulate the variable with appropriate sample weights to see the proportion of women and girls with six or less years of education tab primaryorless [iw=v005/1000000] | - Generate a new variable named 'primaryorless' by typing generate primaryorless=0<br>- Replace the value of 0 with 1 if the level of education is primary or less. This can be done by typing replace primaryorless=1 if welevel<=2<br>- Replace missing values by typing replace primaryorless=. if welevel==9\|welevel==.<br>- Tabulate the variable with appropriate sample weights to see the proportion of women and girls with six or less years of education tab primaryorless [iw=wmweight] |

The results for the DHS Bangladesh 2014 will be:

*Figure 48: Proportion of education-poor women and girls in Bangladesh (DHS 2014)*

| primaryorless | Freq. | Percent | Cum. |
|---|---|---|---|
| 0 | 6,968.171 | 39.01 | 39.01 |
| 1 | 10,894.829 | 60.99 | 100.00 |
| Total | 17,862.9997 | 100.00 | |

The results for the MICS Mongolia 2013–14 will be:

*Figure 49: Proportion of education-poor women and girls in Mongolia (MICS 2013–14)*

| primaryorless | Freq. | Percent | Cum. |
|---|---|---|---|
| 0 | 11,777.895 | 91.81 | 91.81 |
| 1 | 1,051.24455 | 8.19 | 100.00 |
| Total | 12,829.139 | 100.00 | |

The proportion of women and girls ages 15–49 years who are education-poor is 60.99 per cent in Bangladesh and 0.01 per cent in Mongolia.

### SDG 5: Proportion of women aged 18–49 who were married before age 18

The international statistical definition for child marriage identifies women and girls married younger than 18 years of age as child brides. In many countries, the age at first cohabitation is used instead of the age at first marriage to calculate child marriage. For detailed information regarding the codes to calculate this indicator, refer to Table 3.

*Table 3: Codes to calculate the estimates for child marriage*

| | Demographic and Health Survey (Bangladesh 2014) | Multiple Indicator Cluster Survey (Mongolia 2013–14) |
|---|---|---|
| Recode File | Individual Recode File (IR) | Women's File (wm) |
| Recode variable | - Age at first marriage/cohabitation (v511)<br>- Women's sample weights (v005)<br>- All-women factor total (awfactt[29]) | - Age at first marriage/cohabitation (wagem)<br>- Current age (WB2)<br>- Women's sample weights (wmweight) |
| Note | - This can be age at first cohabitation or age at first union. User is advised to confirm what this variable denotes in their country context<br>- When applying awfactt, STATA cannot generate means or percentages. Hence, we will compute ratios, where the numerator reflects women aged 18–49 who were married younger than 18 years of age (adjusted with women's sample weights) and the denominator comprises all women aged 18–49. Since, in the case of Bangladesh, this total refers only to ever-married women, it needs to be adjusted by the product of women's sample weight and all-women factor, i.e. (v005/1000000)*(awfactt/100)<br>- Finally, to compute proportion (percentage) of women aged 18–49 | - This can be age at first cohabitation or age at first union. User is advised to confirm what this variable denotes in their country context |

[29] AWFACTT is a 5-digit variable (in IPUMS-DHS) with two implied decimal places that is used to create total population estimates for all women of childbearing age, for those samples interviewing only ever-married women. It is required that you adjust the estimates with an all-woman factor when dealing with an ever-married women sample. Since the DHS Bangladesh data set, an EMW sample, is being used for this example, we need to adjust the estimates with awfactt. To know more about awfactt, see https://www.idhsdata.org/idhs-action/variables/AWFACTT#description_section

| | | |
|---|---|---|
| | married under the age of 18, use the formula: (numerator/denominator)*100 | |
| Code | - Use only the sample of women whose current age is greater or equal to 18 years by typing keep if v012>=18<br>- Generate a new variable named 'childmarriage' by typing generate childmarriage=0<br>- Replace the value of 0 with 1 if age at first marriage is less than 18 years. This can be done by typing replace childmarriage=1 if v511<18<br>- Generate the numerator and adjust with appropriate sample weights by typing generate num= (v005/1000000)*childmarriage<br>- Generate the denominator and adjust with the appropriate all-women factor by typing generate den=(v005/1000000)*(awfactt/100)<br>- Divide numerator and denominator using the ratio command by typing ratio num/den | - Generate a new variable named 'childmarriage' by typing generate childmarriage=0<br>- Replace the value of 0 with 1 if age at first marriage is less than 18, by typing replace childmarriage=1 if WAGEM<18<br>- Replace for missing values and assign missing values to women that were younger than 18 at the time of the survey replace childmarriage =. if WB2<18\| WAGEM==.<br>- Tabulate the new variable using the appropriate sample weights to see the proportion of women who were married as children, tab childmarriage [iw=wmweight] |

The result of the DHS Bangladesh 2014 will be:

*Figure 50: Proportion of women married before turning 18 in Bangladesh (DHS 2014)*

```
Ratio estimation                 Number of obs    =     16,951

      _ratio_1: num/den
```

| | Ratio | Linearized Std. Err. | [95% Conf. | Interval] |
|---|---|---|---|---|
| _ratio_1 | .6927203 | .0040933 | .684697 | .7007437 |

Finally, as per the formula, (numerator/denominator)*100, the proportion of women aged 18–49 who were married before the age of 18 years is 69.27 per cent.

The result of MICS Mongolia 2013–14 will be:

*Figure 51: Proportion of women married before turning 18 in Mongolia (MICS 2013–14)*

| childmarriage | Freq. | Percent | Cum. |
|---|---|---|---|
| 0 | 9,107.7791 | 92.59 | 92.59 |
| 1 | 729.191879 | 7.41 | 100.00 |
| Total | 9,836.971 | 100.00 | |

In total, 69.2 per cent of women in Bangladesh (DHS 2014) and 7.41 per cent of women and girls in Mongolia (MICS, 2013–14) were married under the age of 18 years.

### SDG 6: Proportion of women and girls ages 15–49 with no access to basic drinking water services

Basic water service refers to drinking water from an improved source, provided collection time is not more than 30 minutes for a roundtrip, including queuing. In calculating this indicator, a person is classified as having no access to basic water services if they live in a household with an unimproved source of water or have access to an improved source of water for which the collection time is more than 30 minutes for a roundtrip, including queuing. For detailed information on how to calculate the proportion of women and girls ages 15-49 with no access to basic drinking water services, refer to Table 4.

*Box 3: Classification criteria for improved source of water*

Improved source of drinking water: Piped water into dwelling; piped water into yard; protected spring; bottled water; piped water into plot; public tap/standpipe; Rainwater collection; bore hole/tube well; protected dug well

*Table 4: Codes to calculate proportion of women and girls who lack access to basic water services*

| | Demographic and Health Survey (Bangladesh 2014) | Multiple Indicator Cluster Survey Mongolia (2013-14) |
|---|---|---|
| Recode file | Individual Recode File (IR) | Merge Women's File and Household File (wm + hh files) (see Box 5 for the code to merge two MICS files) |
| Recode variable | - Type of water source (v113)<br>- Time needed to fetch water (v115)<br>- Women's sample weights (v005) | - Type of water source (WS1)<br>- Time needed to fetch water (WS4A)<br>- Women's sample weight (wmweight) |

| | | |
|---|---|---|
| Code | - Drop 'Not a dejure resident' by typing drop if v113==97[30]<br>- For type of source of water, generate a new variable named 'improvedwater' by typing generate improvedwater=0<br>- Replace 0 with a value of 1 if type of water source is improved. This can be done by typing replace improvedwater=1 if v113==11\|v113==12\|v113==13\|v113==21\|v113==32\|v113==41\|v113==61\|v113==62\|v113==71Replace missing values replace improvedwater=. if v113==.\|v113>96<br>- For the time taken to fetch water, generate a new variable named 'nearwater' by typing generate nearwater=0<br>- Replace 0 with 1 if the time taken to fetch water is less than 31 or water source is in the household (on premise). This can be done by typing replace nearwater=1 if v115==996\|v115<31<br>- Replace missing values replace nearwater=. if v115==.\|v115>996<br>- For the combination variable, create a new variable named 'nobasicwater' by typing generate nobasicwater=0<br>- Replace 0 with 1 if the source of water is unimproved or time taken to fetch water is more than 30 minutes. This can be done by typing replace | - For type of source of water, generate a new variable named 'improvedwater' by typing generate improvedwater=0<br>- Replace 0 with 1 if source of water is improved, by typing replace improvedwater=1 if WS1==15\|WS1==16\|WS1==17\|WS1==22\|WS1==31\|WS1==41\|WS1==51\|WS1==62\|WS1==63\|WS1==71\|WS1==91<br>- Replace missing values replace improvedwater=. if WS1==.<br>- For the time taken to fetch water, generate a new variable named 'nearwater' by typing generate nearwater=0<br>- Replace the new variable with a value of 1 if the time taken to fetch water is less than 31 replace nearwater=1 if WS4A==1\|WS4A==2<br>- Replace missing values replace nearwater=. if WS4A==.\|WS4A==8<br>- For the combination variable, create a new variable named 'nobasicwater' by typing generate nobasicwater=0<br>- Replace 0 with 1 if the source of water is unimproved or time taken to fetch water is more than 30 minutes. This can be done by typing replace nobasicwater =1 if improvedwater==0\|nearwater==0 |

[30] We need to drop those respondents who do not reside in the household in which the interview was held. In the case of MICS, since the response from non-de-jure members is not recorded, there is no need to drop those observations.

| | | |
|---|---|---|
| | nobasicwater =1 if improvedwater==0\|nearwater==0<br>- Replace missing values by typing replace nobasicwater=. if improvedwater==.\|nearwater==.<br>- Tabulate the combination variable with appropriate weights tab nobasicwater [iw=v005/1000000] | - Replace missing values by typing replace nobasicwater =. If improvedwater==.\|nearwater==.<br>- Tabulate the combination variable with appropriate weights by typing tab nobasicwater wmweight |

*Box 4: Code to merge wm and hh data sets*

**Multiple Indicator Cluster Survey**

Recode File: Merge wm and hh file using the following code:

```
use "HH FILE PATH"
rename HH1 WM1
rename HH2 WM2
sort WM1 WM2
save temp.dta, replace
use "WM FILE PATH"
sort WM1 WM2
merge WM1 WM2 using temp.dta
tab _merge
keep if _merge==3
```

The result for the DHS Bangladesh 2014 will be:

*Figure 52: Proportion of women and girls with no access to basic water in Bangladesh (DHS 2014)*

| nobasicwater | Freq. | Percent | Cum. |
|---|---|---|---|
| 0 | 16,155.549 | 96.72 | 96.72 |
| 1 | 547.181432 | 3.28 | 100.00 |
| Total | 16,702.73 | 100.00 | |

The result for MICS Mongolia 2013–14 will be:

*Figure 53: Proportion of women and girls with no access to basic water in Mongolia (MICS 2013–14)*

| nobasicwater | Freq. | Percent | Cum. |
|---|---|---|---|
| 0 | 5,808.0609 | 65.22 | 65.22 |
| 1 | 3,097.434 | 34.78 | 100.00 |
| Total | 8,905.4949 | 100.00 | |

## SDG 6: Proportion of women and girls ages 15–49 years who lack access to basic sanitation facilities

Basic sanitation refers to an improved toilet facility which is not shared with another household in a settlement. In calculating this indicator, a person is classified as having no access to basic sanitation if they live in a household with an unimproved toilet facility or in a household with an improved toilet facility which is shared with other households. For detailed information on how to calculate the proportion of women and girls ages 15–49 who lack access to basic sanitation services, refer to Table 5.

*Box 5: Classification criteria for type of toilet facility*

Improved sanitation facility: flush - to piped sewer system; flush - to septic tank; flush - to pit latrine; flush - don't know where; flush – unspecified; pit latrine - ventilated improved pit (vip); pit latrine - with slab; composting toilet

Unimproved sanitation facility: flush - to somewhere else; pit latrine - without slab / open pit; no facility/bush/field; bucket toilet; hanging toilet/latrine; other

*Table 5: Codes to calculate the proportion of women and girls who lack access to basic sanitation facilities*

| | Demographic and Health Survey (Bangladesh 2014) | Multiple Indicator Cluster Survey (Mongolia 2013–14) |
|---|---|---|
| Recode File | Individual Recode File (IR) | Women's File (wm) |
| Recode variable | Type of toilet facility (v116)<br>Shared facility (v160) | Type of toilet facility (WS8)<br>Shared facility (WS9) |
| Code | - Drop non-*de-jure* members by typing<br>drop if v160==7 | - For type of sanitation facility, generate a new variable named 'improvedsanitation' by |

| | | |
|---|---|---|
| | - For type of sanitation facility, generate a new variable named 'improvedsanitation' by typing generate improvedsanitation=0<br>- Replace 0 with a value of 1 if the type of sanitation facility is improved, by typing replace improvedsanitation=1 if v116==11\|v116==12\|v116==13\|v116==15\|v116==16\|v116==21\|v116==22\|v116==41<br>- Replace for missing values by typing replace improvedsanitation =. if v116==.\|v116>96<br>- Generate a new variable named 'sharedfacility' by typing generate sharedfacility=0<br>- Replace 0 with 1 if sanitation facility is shared. This can be done by typing replace sharedfacility=1 if v160==1<br>- Replace for missing values by typing replace sharedfacility=. if v160==.<br>- Generate a new combination variable named 'nobasicsanitation' by typing generate nobasicsanitation=0<br>- Replace 0 with 1 if type of sanitation facility is either unimproved or shared, by typing replace nobasicsanitation =1 if | typing generate improvedsanitation=0<br>- Replace with a value of 1 if the type of sanitation facility is improved. This can be done by typing replace improvedsanitation =1 if WS8==11\|WS8==12\|WS8==13\|WS8==21\|WS8==22\|WS8==31<br>- Replace for missing values replace improvedsanitation =. if WS8==.\|WS8>96<br>- For shared facility, create a new variable and ascribe a value of 0 to all of its observations generate sharedfacility=0<br>- Replace the new variable with 1 if sanitation facility is shared replace sharedfacility=1 if WS9==1<br>- Replace for missing values by typing replace sharedfacility=. If WS9==.<br>- Generate new variable named 'nobasicsanitation', which is a combination of inadequate sanitation and shared sanitation facility. This can be done by typing generate nobasicsanitation=0<br>- Replace 0 with a value of 1 by typing replace nobasicsanitation =1 if improvedsanitation ==0\|sharedfacility==1 |

| | | |
|---|---|---|
| | improvedsanitation ==0\|sharedfacility==1<br>- Replace for missing values by typing replace nobasicsanitation =. if improvedsanitation ==.\|sharedfacility==.<br>- Tabulate the combination variable with appropriate sample weights tab nobasicsanitation [iw=v005/1000000] | - Replace for missing values replace nobasicsanitation =1 if improvedsanitation ==.\|sharedfacility==.<br>- Tabulate the combination variable with appropriate sample weights tab nobasicsanitation [iw=wmweight] |

The result for the DHS Bangladesh 2014 will be:

*Figure 54: Proportion of women and girls who lack access to basic sanitation facilities in Bangladesh (DHS 2014)*

| nobasicsanitation | Freq. | Percent | Cum. |
|---|---|---|---|
| 0 | 8,033.4311 | 49.76 | 49.76 |
| 1 | 8,110.55776 | 50.24 | 100.00 |
| Total | 16,143.989 | 100.00 | |

The result for MICS Mongolia 2013–14 will be:

*Figure 55: Proportion of women and girls who lack access to basic sanitation facilities in Mongolia (MICS 2013–14)*

| nobasicsanitation | Freq. | Percent | Cum. |
|---|---|---|---|
| 0 | 7,580.07685 | 59.08 | 59.08 |
| 1 | 5,249.92315 | 40.92 | 100.00 |
| Total | 12,830 | 100.00 | |

As shown in the figures above, over half of all women and girls in Bangladesh and over 40 per cent of women and girls in Mongolia lack access to basic sanitation facilities.

### SDG 7: Proportion of women and girls aged 15–49 with no access to clean cooking fuel

A fuel is defined as "Clean" based on the emission rate targets and specific fuel recommendations included in the *WHO guidelines for indoor air quality: household fuel combustion*. According to these guidelines, unprocessed coal and kerosene (among other types of fuels) are not considered clean[31]. For the classification of type of fuel, see Box 6 and for details on how to calculate the proportion of women and girls aged 15–49 years who lack access to clean cooking fuels, refer to Table 6.

*Box 6: Classification criteria for type of cooking fuel*

Clean cooking fuel: Electricity; Liquified Petroleum Gas (LPG); natural gas; biogas; solar cooker

Unclean cooking fuel: Solid biomass fuel (e.g. wood, animal dung, crop wastes, charcoal); kerosene; open fires

*Table 6: Codes to calculate the proportion of women and girls who lack access to clean cooking fuel*

| | Demographic and Health Survey (Bangladesh 2014) | Multiple Indicator Cluster Survey (Mongolia 2013–14) |
|---|---|---|
| Recode File | Individual Recode File (IR) | Merge Women's File and Household File (wm + hh files) |
| Recode variable | - Type of cooking fuel (v161)<br>- Women's sample weights (v005) | - Type of cooking fuel (HC6)<br>- Women's sample weights (wmweight) |
| Code | - Drop non-*de-jure* residents by typing drop if v161==97<br>- Generate a new variable named 'cleanfuel' by typing generate cleanfuel=0<br>- Replace 0 with a value of 1 type of cooking fuel is clean. This can be done by typing replace cleanfuel=1 if v161==1\|v161==2\|v161==3\|v161==4 | - Generate a new variable named 'cleanfuel' by typing generate cleanfuel=0<br>- Replace 0 with a value of 1 type of cooking fuel is clean. This can be done by typing replace cleanfuel=1 if HC6==1\|HC6==2<br>- Replace missing values replace cleanfuel=. if HC6==.\|HC6==95 |

[31] See WHO https://www.who.int/indoorair/publications/fuelforlife/en/

| | | |
|---|---|---|
| | - Replace missing values and if no food is cooked at home. This can be done by typing replace cleanfuel=. if v161==.\|v161>90<br>- Tabulate the new variable with appropriate sample weights to see the proportion of women and girls aged 15–49 relying on unclean fuels for cooking tab cleanfuel [iw=v005/1000000] | - Tabulate the new variable with appropriate sample weights to see the proportion of women and girls aged 15–49 relying on unclean fuels for cooking tab cleanfuel [iw=wmweight] |

The results for DHS Bangladesh 2014 will be:

*Figure 56: Proportion of women and girls who lack access to clean cooking fuels in Bangladesh (DHS 2014)*

| cleanfuel | Freq. | Percent | Cum. |
|---|---|---|---|
| 0 | 13,736.767 | 82.29 | 82.29 |
| 1 | 2,957.0962 | 17.71 | 100.00 |
| Total | 16,693.863 | 100.00 | |

The results for the MICS Mongolia 2013–14 will be:

*Figure 57: Proportion of women and girls who lack access to clean cooking fuels in Mongolia (MICS 2013–14)*

| cleanfuel | Freq. | Percent | Cum. |
|---|---|---|---|
| 0 | 6,686.6066 | 52.12 | 52.12 |
| 1 | 6,142.2253 | 47.88 | 100.00 |
| Total | 12,828.832 | 100.00 | |

A total of 82.29 per cent of women and girls in Bangladesh and 52.12 per cent of women and girls in Mongolia lack access to clean cooking fuels.

### SDG 8: Proportion of women aged 18–49 who are currently not employed

Currently employed is defined as having done work in the past seven days. Therefore, this indicator refers to the proportion of people who were not employed at the time of the survey. It includes people who: 1) did not work in the seven days prior to the survey date but who are regularly employed and were absent from work for leave; 2) Were away from work at the time of the survey due to illness, vacation, or any other such reason. For information on how to calculate the proportion of women aged 18–49 who are currently NOT employed, refer to Table 7.

*Table 7: Code to calculate the proportion of currently not employed women aged 18–49 years*

| | Demographic and Health Survey | Multiple Indicator Cluster Survey |
|---|---|---|
| Recode File | Individual Recode File (IR) | Women's File (wm) |
| Recode variables | - Respondent worked in last 12 months (v731)<br>- Current age (v012)<br>- Women's sample weight (v005) | - Employed in last 7 days (WB8)<br>- Current age (WB2)<br>- Women's sample weight (wmweight) |
| Code | - Drop if younger than 18 years by typing drop if v012<18<br>- Generate a new variable named 'currentlyemployed' by typing generate currentlyemployed=0<br>- Replace 0 with a value of 1 if the respondent is currently employed. This can be done by typing replace currentlyemployed=1 if v731==2\|v731==3<br>- Replace missing values by typing replace currentlyemployed=. if v731==.<br>- Tabulate the new variable with appropriate sample weights to see the proportion of women aged 18–49 who are | - Drop if younger than 18 years by typing drop if WB2<18<br>- Generate a new variable named 'currentlyemployed' by typing generate currentlyemployed=0<br>- Replace 0 with a value of 1 if the respondent is currently employed. This can be done by typing replace currentlyemployed=1 if WB8==1<br>- Replace missing values by typing replace currentlyemployed=1 if WB8==.<br>- Tabulate the new variable with appropriate sample weights to see the proportion of women aged 18–49 who are |

| | | |
|---|---|---|
| | currently employed tab currentlyemployed [iw=v005/1000000] | currently employed tab currentlyemployed [iw=wmweight] |

The results for the DHS Bangladesh 2014 will be:

*Figure 58: Proportion of women who are currently not employed in Bangladesh (DHS 2014)*

| currentlyemployed | Freq. | Percent | Cum. |
|---|---|---|---|
| 0 | 11,183.334 | 65.85 | 65.85 |
| 1 | 5,799.7755 | 34.15 | 100.00 |
| Total | 16,983.109 | 100.00 | |

The results for the MICS Mongolia 2013–2014 will be:

*Figure 59: Proportion of women who are currently not employed in Mongolia (MICS 2013–14)*

| currentlyemployed | Freq. | Percent | Cum. |
|---|---|---|---|
| 0 | 4,965.91267 | 42.38 | 42.38 |
| 1 | 6,751.3598 | 57.62 | 100.00 |
| Total | 11,717.272 | 100.00 | |

A total of 65.85 per cent of women in Bangladesh and 42.38 per cent of women in Mongolia were not employed at the time of the DHS 2014 and MICS 2013–14 surveys, respectively.

### SDG 11: Proportion of women and girls aged 15–49 living in overcrowded housing

A household is classified as overcrowded if more than three people are sleeping in the same room[32]. For more information on how to calculate the proportion of women and girls aged 15–49 years who live in overcrowded households, refer to Table 8.

[32] Per the Global SDG database metadata, households are considered overcrowded if not more than three people share the same habitable room. This differs from the definition of overcrowded dwelling when calculating the slum indicator, where three or more than three is considered overcrowded. For more details, please see: https://unstats.un.org/sdgs/metadata/files/Metadata-11-01-01.pdf

*Table 8: Codes to calculate the proportion of women and girls who live in overcrowded households*

| | Demographic and Health Survey | Multiple Indicator Cluster Survey |
|---|---|---|
| Recode File | Merge Personal Recode File (PR file) and Individual Recode File (IR file) | Merge Women's file (wm file) and Household File (hh file) |
| Recode Variable | - No. of household members (hv136)<br>- No. of rooms used to sleep (hv216) | - No. of household members (HH11)<br>- No. of rooms used to sleep (HC2) |
| Code | - Create a new variable and ascribe a value of 0 to all of its observations generate overcrowd = 0<br>- Replace the new variable with 1 if the household is overcrowded, replace overcrowd = 1 if hv136/hv216 > 3<br>- Replace missing values replace overcrowd = . if hv136 == . \| hv216 == .<br>- Tabulate the new variable to see the proportion of women that live in overcrowded households tab overcrowd [iw = v005/1000000] | - Create a new variable and ascribe a value of 0 to all of its observations generate overcrowd = 0<br>- Replace the new variable with 1 if the household is overcrowded, replace overcrowd = 1 if HH11/HC2>3<br>- Replace missing values replace overcrowd = . if HH11 == . \| HC2 == .<br>- Tabulate the new variable to see the proportion of women that live in overcrowded households tab overcrowd [iw = wmweight] |

The results for the DHS Bangladesh 2014 will be:

*Figure 60: Proportion of women and girls living in overcrowded households in Bangladesh (DHS 2014)*

| overcrowded housing | Freq. | Percent | Cum. |
|---|---|---|---|
| 0 | 13,400.9353 | 75.02 | 75.02 |
| 1 | 4,462.0644 | 24.98 | 100.00 |
| Total | 17,862.9997 | 100.00 | |

The results for the MICS Mongolia 2013–14 will be:

*Figure 61: Proportion of women and girls living in overcrowded household in Mongolia (MICS Mongolia 2013–14)*

| overcrowded housing | Freq. | Percent | Cum. |
|---|---|---|---|
| 0 | 5,472.8314 | 69.80 | 69.80 |
| 1 | 2,368.0744 | 30.20 | 100.00 |
| Total | 7,840.9058 | 100.00 | |

Nearly 25 per cent of women and girls in Bangladesh and 30 per cent of women and girls in Mongolia live in overcrowded households.

After calculating all indicators of interest, the analyst should perform disaggregation at various levels (e.g. by location, wealth, ethnicity, etc.) to identify the population groups furthest behind for each indicator. Scanning these results often reveals that the same population groups in each country tend to lag behind in many of the development indicators. These multiply deprived population groups are those being left furthest behind. The following example showcases how to identify the furthest behind by conducting this type of analysis.

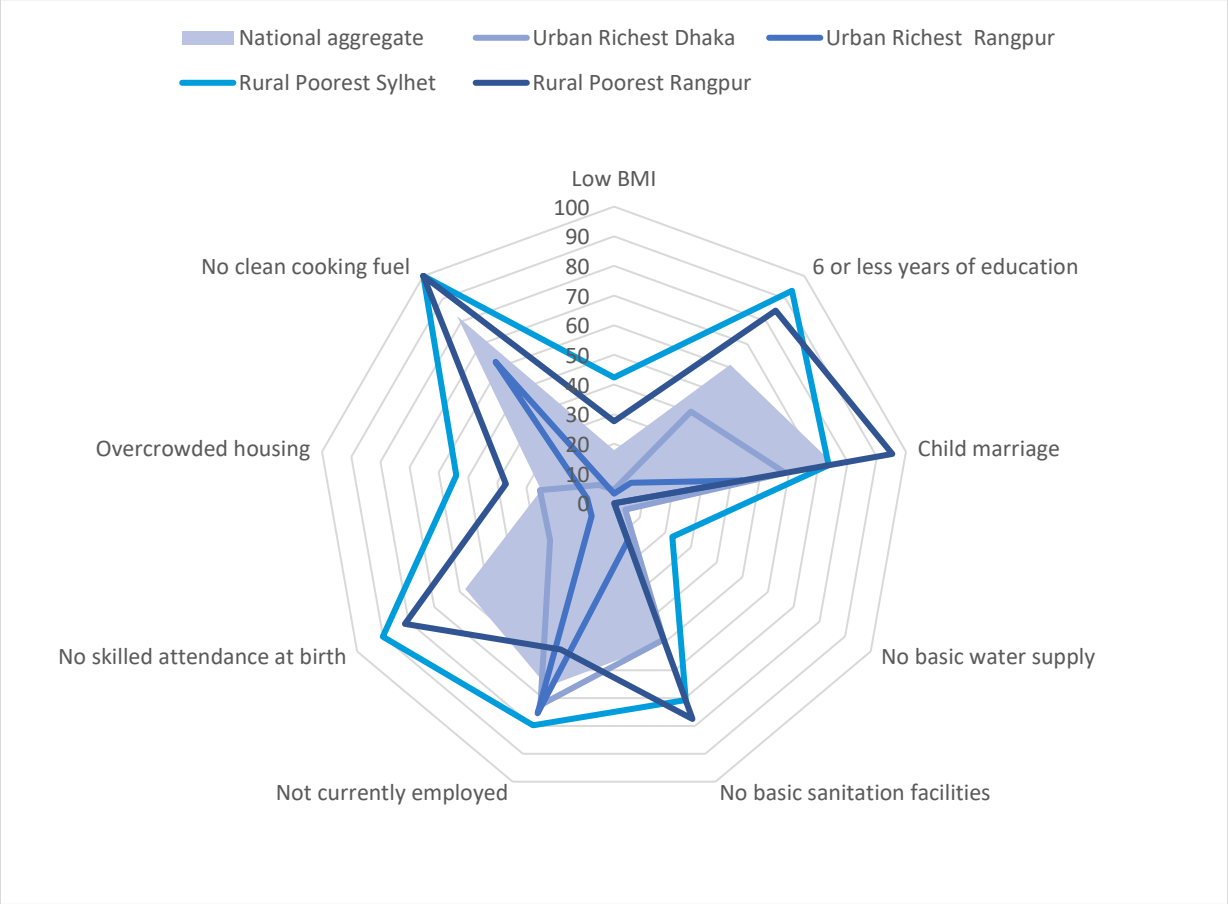
## 2.4. Country example: Bangladesh

### Identifying the furthest behind

Analysis of the 2014 Bangladesh DHS[33] shows how the intersection of poverty, location and region of residence leaves certain groups of women and girls behind. The findings point to large disparities between rich and poor and urban and rural residents. The groups lagging the furthest behind are those facing intersecting forms of discrimination based on wealth, geography and region. Poorest rural women and girls who live in the Sylhet region of Bangladesh are the most deprived in seven out of the nine dimensions tested: nutrition (proxied by low BMI), education, access to basic drinking water services, employment, access to skilled birth attendance, overcrowded housing and access to clean cooking fuel. Wealth and location appear to have a significant effect on progress towards achieving the SDGs. Compounded with region of residence, key population groups emerge as those furthest behind. Figure 62 shows the compound effect of wealth, location and region on four groups of women and girls in Bangladesh, as compared to the national aggregate.

[33] https://dhsprogram.com/data/dataset/Bangladesh_Standard-DHS_2014.cfm?flag=0

*Figure 62. Compound effect of wealth, location and region of residence on SDG-related outcomes for women (Bangladesh, DHS 2014)*




## Multidimensional clustered deprivation

Often, deprivation in one area reinforces deprivations in other areas, that is, deprivations cluster together.[34] To identify the total share of the multiply deprived population, it is important to see who experiences deprivations for more than one indicator. In our example, we look at the proportion of the population that experiences deprivations in three areas (using three indicators), namely: only completed primary or less years of education, they were not employed at the time of the survey, and rely primarily on unclean cooking fuels. To ensure comparability across indicators, we must ensure that all indicators considered utilize the same age group. For this analysis, we have selected women in the age group of 18–49. In other words, we drop from our analysis all women younger than 18.

To calculate the proportion of women ages 18–49 who are simultaneously deprived in three SDG-related dimensions, follow these steps:

[34] UN Women 2018.

a) Open the DHS IR file in STATA for Bangladesh DHS 2014.
b) Drop the observations for women younger than 18 by typing drop if v012<18
c) Calculate the proportion of women and girls who are deprived in education (i.e. primary or less), deprived in employment (i.e. currently not employed) as well as rely primarily on unclean cooking fuels.
    - To calculate the first indicator, *primary or less years of education*, type: generate primaryorless=0. Replace the value of 0 with 1 if woman obtained only primary or less years of education, by typing replace primaryorless=1 if v133<=6. Replace missing values by typing replace primaryorless=. if v133==.
    - To calculate the second indicator, *currently employed*, type: generate currentlyemployed=0. Replace the value of 0 with 1 if woman is currently working or has a job but has been on leave in the last 7 days, by typing replace currentlyemployed=1 if v731==2|v731==3. Replace missing values by typing replace currentlyemployed=. if v731==.
    - To calculate the third indicator, *lack of access to clean fuels*, type: generate nocleanfuel=0. Replace the value of 0 with 1 if woman lacks access to a clean cooking fuel, by typing replace nocleanfuel=1 if v161>4. If observations are missing, or no food is cooked at home, or the resident is non dejure, replace as missing values. This can be done by tying replace nocleanfuel=. if v161==.|v161>=95
    - Finally, to calculate the composite indicator, generate a new variable "deprivedin3". This can be done by typing generate deprivedin3=0. Replace the value of 0 with 1 if woman is deprived in all the three indicators calculated above, by typing: replace deprivedin3=1 if primaryorless==1 & currentlyemployed==0 & nocleanfuel==1. Replace missing values by typing replace deprivedin3=. if primaryorless==. | currentlyemployed==. | nocleanfuel==.

    - Tabulate the results using appropriate sample weights by typing tab deprivedin3 [iw=v005/1000000]

The results for the Bangladesh DHS 2014 will be:

*Figure 63: Proportion of women aged 18–49 in Bangladesh simultaneously deprived in three SDG-related dimensions (DHS 2014)*

| deprivedin3 | Freq. | Percent | Cum. |
|---|---|---|---|
| 0 | 10,615.594 | 66.47 | 66.47 |
| 1 | 5,354.205 | 33.53 | 100.00 |
| Total | 15,969.799 | 100.00 | |

d) To further disaggregate these results by wealth, use the following code:
   - tab deprivedin3 v190 [iw=v005/1000000], cell column row

The results will look as follows:

*Figure 64: Poorest women are overrepresented in the simultaneously deprived group in Bangladesh (DHS 2014)*

| deprivedin3 | wealth index | | | | | |
|---|---|---|---|---|---|---|
| | poorest | poorer | middle | richer | richest | Total |
| 0 | 1,515.5029 | 1,578.557 | 1,894.015 | 2,444.321 | 3,183.199 | 10,615.59 |
| | 14.28 | 14.87 | 17.84 | 23.03 | 29.99 | 100.00 |
| | 50.22 | 52.04 | 60.45 | 73.20 | 92.36 | 66.47 |
| | 9.49 | 9.88 | 11.86 | 15.31 | 19.93 | 66.47 |
| 1 | 1,502.217 | 1,454.633 | 1,239.235 | 894.954106 | 263.166455 | 5,354.205 |
| | 28.06 | 27.17 | 23.15 | 16.71 | 4.92 | 100.00 |
| | 49.78 | 47.96 | 39.55 | 26.80 | 7.64 | 33.53 |
| | 9.41 | 9.11 | 7.76 | 5.60 | 1.65 | 33.53 |
| Total | 3,017.72 | 3,033.19 | 3,133.249 | 3,339.275 | 3,446.3653 | 15,969.8 |
| | 18.90 | 18.99 | 19.62 | 20.91 | 21.58 | 100.00 |
| | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| | 18.90 | 18.99 | 19.62 | 20.91 | 21.58 | 100.00 |

e) To disaggregate the estimates by location, use the following codes:
   - tab deprivedin3 v025 [iw=v005/1000000], cell column row

The results will look as follows:

*Figure 65: Rural women are overrepresented in the simultaneously deprived group in Bangladesh (DHS 2014)*

| deprivedin3 | type of place of residence | | |
|---|---|---|---|
| | urban | rural | Total |
| 0 | 3,716.478 | 6,899.117 | 10,615.59 |
| | 35.01 | 64.99 | 100.00 |
| | 81.39 | 60.50 | 66.47 |
| | 23.27 | 43.20 | 66.47 |
| 1 | 849.959524 | 4,504.245 | 5,354.205 |

f) To disaggregate the estimates by region, use the following codes:
   - Drop non-*de-jure* residents by typing drop if v139==97
   - tab deprivedin3 v139 [iw=v005/1000000], cell column row

The results will look as follows:

*Figure 66: Women living in the Sylhet region are overrepresented in the simultaneously deprived group in Bangladesh 2014*

| deprivedin3 | de jure region of residence | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | barisal | chittagon | dhaka | khulna | rajshahi | rangpur | sylhet | Total |
| 0 | 576.255533 | 1,934.663 | 4,097.041 | 1,044.884 | 1,264.532 | 1,187.886 | 510.332029 | 10,615.59 |
| | 5.43 | 18.22 | 38.59 | 9.84 | 11.91 | 11.19 | 4.81 | 100.00 |
| | 59.16 | 66.12 | 73.23 | 63.16 | 66.96 | 65.72 | 45.37 | 66.47 |
| | 3.61 | 12.11 | 25.65 | 6.54 | 7.92 | 7.44 | 3.20 | 66.47 |
| 1 | 397.864755 | 991.311791 | 1,497.403 | 609.401378 | 624.051691 | 619.732069 | 614.440394 | 5,354.205 |
| | 7.43 | 18.51 | 27.97 | 11.38 | 11.66 | 11.57 | 11.48 | 100.00 |
| | 40.84 | 33.88 | 26.77 | 36.84 | 33.04 | 34.28 | 54.63 | 33.53 |
| | 2.49 | 6.21 | 9.38 | 3.82 | 3.91 | 3.88 | 3.85 | 33.53 |
| Total | 974.120288 | 2,925.975 | 5,594.4444 | 1,654.285 | 1,888.584 | 1,807.618 | 1,124.772 | 15,969.8 |
| | 6.10 | 18.32 | 35.03 | 10.36 | 11.83 | 11.32 | 7.04 | 100.00 |
| | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| | 6.10 | 18.32 | 35.03 | 10.36 | 11.83 | 11.32 | 7.04 | 100.00 |

Single-level disaggregation analysis fails to adequately reflect the level of deprivation of those who are disadvantaged across multiple dimensions. Identifying these population groups will require multidimensional data disaggregation and targeted analysis. Descriptive statistics for each of the subgroups should be accompanied by qualitative work to understand the root causes and the drivers of the inequalities observed. Only after assessing the full effects of multiple discrimination and clustered deprivation can policies be tailored to meet the needs of all population groups[35].

## 3. Testing the significance of results

Typically, statistical analysis involves a research question or an exploratory hypothesis, which is tested using the best-suited methodological techniques, to finally answer the problem question or accept/reject the hypothesis. Broadly, the research question for the LNOB analysis pertains to identifying the most deprived groups of women and girls who are at the greatest risk of being left behind. For instance, the problem question can be phrased as follows:

*Who are the most deprived groups of women and girls regarding access to education?*

Statistically, the LNOB analysis aims to do the following:

[35] UN Women 2018.

- *It makes an inference about a population group (e.g. women and girls), based on a sample of the population (DHS or MICS survey data of women and girls), with a certain degree of probability.*
- *It hypothesizes to test the effect of a particular socioeconomic characteristic (e.g. wealth, location, ethnicity, etc.) or a combination of more than one socioeconomic characteristic (e.g. wealth and location simultaneously, or wealth, location and religion simultaneously), on the outcome of women and girls for a given development indicator (e.g. proportion of women and girls who are education-poor).*

To answer the LNOB research question, we have disaggregated DHS or MICS survey data by multiple socioeconomic variables, at multiple levels, to simply see the change in the proportion of women and girls who are deprived on the given indicator. For example, Figure 67 shows that 86.80 per cent of the poorest women and girls are education-poor and 31.16 per cent of the richest women and girls are education-poor. Clearly, there is a large difference in the proportions of the educated among the richest and poorest groups.

*Figure 427: Proportion of women and girls with primary or less years of education in Bangladesh, by wealth*

| primaryorless | wealth index | | | | | |
|---|---|---|---|---|---|---|
| | poorest | poorer | middle | richer | richest | Total |
| 0 | 443.3788 | 783.190081 | 1,395.17 | 1,745.724 | 2,600.708 | 6,968.171 |
| | 6.36 | 11.24 | 20.02 | 25.05 | 37.32 | 100.00 |
| | 13.20 | 22.98 | 39.19 | 46.45 | 68.84 | 39.01 |
| | 2.48 | 4.38 | 7.81 | 9.77 | 14.56 | 39.01 |
| 1 | 2,916.061 | 2,624.329 | 2,165.072 | 2,012.391 | 1,176.976 | 10,894.83 |
| | 26.77 | 24.09 | 19.87 | 18.47 | 10.80 | 100.00 |
| | 86.80 | 77.02 | 60.81 | 53.55 | 31.16 | 60.99 |
| | 16.32 | 14.69 | 12.12 | 11.27 | 6.59 | 60.99 |
| Total | 3,359.44 | 3,407.519 | 3,560.242 | 3,758.115 | 3,777.6834 | 17,863 |
| | 18.81 | 19.08 | 19.93 | 21.04 | 21.15 | 100.00 |
| | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| | 18.81 | 19.08 | 19.93 | 21.04 | 21.15 | 100.00 |

This difference suggests that wealth can affect the likelihood of women and girls to be education-poor. From a statistical standpoint, it is important to check if this difference is statistically significant. That is: with how much confidence can we say this difference is a true difference and not merely caused by a chance factor or error?

## 3.1. Testing the statistical significance of LNOB estimates

To understand this, let's continue with the same example as above. Suppose we want to test whether the difference in the proportion of women and girls who are education-poor is significant between the poorest and the richest wealth groups. To do this, first set the hypothesis:

Null hypothesis: There is no difference between the proportion of poorest and richest women and girls with primary or less years of education, or

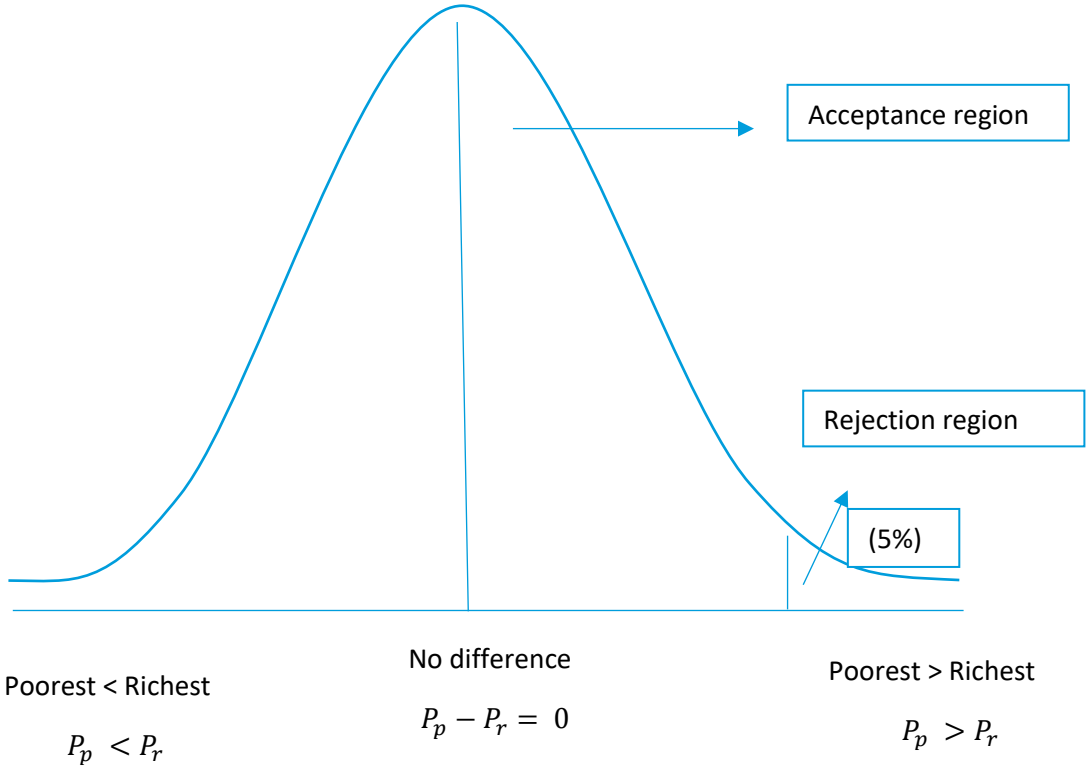
$H_0: P_p - P_r = 0$

where $P_p$ is the proportion of poorest women and girls having completed only primary or less years of education and $P_r$ is the proportion of richest women and girls with primary or less education.

We want to test this null hypothesis against the alternative hypothesis:

Alternative hypothesis: The proportion of women and girls who are education-poor is higher in the poorest households, or

Ha: $P_p - P_r \neq 0$




From the LNOB analysis, we found that the proportion of poorest women and girls having completed primary or less years of education is 86.80 per cent and that of the richest women and girls is 31.16 per cent. This estimation about the population is based on a sample (the sample of the survey used), and therefore it concerns probability theory (e.g. our results indicate that it is likely that 86.80 per cent of poorest women and girls in Bangladesh are education poor, but we cannot know if this is the figure with certainty), these proportion can be understood as probabilities:

$P_p$ = 0.86 (probability of the proportion of poorest women and girls having completed primary or less years of education)

$P_r$ = 0.31 (probability of the proportion of richest women and girls having completed primary or less years of education)

The difference between the proportions of the two groups (poorest and richest) = 0.55 $(i.e. 0.86 - 0.31)$

The question we have to ask at this stage is: is this difference large enough to warrant our acceptance of the alternative hypothesis or is it in the range within which we will accept the null hypothesis? In other words, how likely it is that we obtained this result if there is no difference in educational attainment between the rich and poorest groups in the actual population?

## 3.2. Testing Significance using Stata

To find the statistical significance of the difference between the two groups of women, we can run a test of two-sample proportions (prtest) in Stata. Below are the steps needed to run this test for two select groups: richest women who have completed primary or less years of education vs. poorest women who have completed primary or less years of education. For this example, the unit of analysis are women and girls in Bangladesh (DHS 2014).

a) Step 1: Primary or less years of education
   - Generate a new variable for primary or less years of education. This can be done by typing generate primaryorless=0
   - Replace 0 with 1 if single years of education are six or less than six. This can be done by typing replace primaryorless=1 if v133<7
   - Replace missing values by typing replace primaryorless=. if v133==.
b) Step 2: Create a new group of poorest women
   - Generate a new variable named 'poorest' by typing generate poorest=0
   - Replace 0 with 1 if the observation belongs to the poorest wealth quintile. This can be done by typing replace poorest=1 if v190==1
   - Replace missing values by typing replace poorest=. if v190==.
c) Step 3: Create a new group of poorest women who have only completed primary or less years of education
   - Generate a new group named 'poorestprimaryorless' by typing generate poorestprimaryorless=0
   - Replace 0 with 1 if a woman belongs to the poorest wealth quintile and has only completed primary or less years of education. This can be done by typing replace poorestprimaryorless=1 if poorest==1&primaryorless==1
   - Replace missing values if poorest is not equal (!=) to 1. This command treats all other wealth groups (poorer, middle, richer and richest as missing) and prevents underestimation of results for the poorestprimaryorless group. This can be done by typing replace poorestprimaryorless=. if poorest!=1
   - Replace missing values if observations for primaryorless are missing. This can be done by typing replace poorestprimaryorless=. if primaryorless==.
d) Step 4: Create a new group of richest women
   - Generate a new variable named 'richest' by typing generate richest=0

- Replace 0 with 1 if the case belongs to the richest wealth quintile. This can be done by typing replace richest=1 if v190==5
- Replace for missing values by typing replace richest=. if v190==.

e) Step 5: Create a new group of richest women who have primary or less years of education
- Generate a new group named 'richestprimaryorless' by typing generate richestprimaryorless=0
- Replace 0 with 1 if woman belongs to the richest wealth quintile and has only completed primary or less years of education. This can be done by typing replace richestprimaryorless=1 if richest==1&primaryorless==1
- Replace missing values if richest is not equal (!=) to 1. This command treats all other wealth groups (poorer, middle, richer and richest) as missing and prevents underestimation of results for the richestprimaryorless group. This can be done by typing replace richestprimaryorless=. if richest!=1
- Replace missing values if observations for primaryorless are missing. This can be done by typing replace richestprimaryorless=. if primaryorless==.

f) Step 6: Test whether the difference between the proportion of poorest women who have completed primary or less years of education (poorestprimaryorless) and that of richest women who have completed primary or less years of education (richestprimaryorless) is significant and interpret the results.
- This can be done by typing: prtest poorestprimaryorless=richestprimaryorless

The results table will look as follows:

*Figure 68: Result for two-sample test of proportions*

```
Two-sample test of proportions                poorestprima: Number of obs =     3251
                                              richestprima: Number of obs =     3862
------------------------------------------------------------------------------
       Group |      Mean   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
poorestprima |   .8637342   .0060169                      .8519413    .8755272
richestprima |    .290782   .0073075                      .2764596    .3051044
-------------+----------------------------------------------------------------
        diff |   .5729523   .0094659                      .5543995     .591505
             |  under Ho:   .0118348    48.41   0.000
------------------------------------------------------------------------------
    diff = prop(poorestprima) - prop(richestprima)                z =  48.4125
Ho: diff = 0

 Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
Pr(Z < z) = 1.0000    Pr(|Z| > |z|) = 0.0000     Pr(Z > z) = 0.0000
```

To interpret the results, make note of the values shown in Table 9:

*Table 9: Interpretation guide for test of significance*

| Parameters | Meaning | Value | Interpretation |
|---|---|---|---|
| z-score | A measure of how many standard deviations below or above the population mean a raw score is. | 48 | The obtained difference in the proportions between the richest and poorest is 48 standard deviations away from the hypothesized difference of 0. A z-score this high indicates that a difference of .57 between the poorest and richest groups is very unlikely to have happened by chance. |
| P value | The p-value or probability value is the probability of obtaining test results at least as extreme as the results actually observed during the test under the assumption that $H_0$ is true. | 0.00 | Usually, a p value of 0.05 is set as the cutoff for significance. Since the z score of 48 has been obtained at a p value of 0.00, which is less than 0.05, we can reject the null hypothesis that our two groups were not different. |
| Confidence interval | A confidence interval refers to the probability that a population parameter will fall between two set values for a certain proportion of times. | .5543 to .5915 | The 95 per cent confidence interval indicates that we are 95 per cent confident that the difference in the proportions between the poorest and richest groups of women with primary or less years of education in Bangladesh lies between this range. |

## 4. Reporting the estimates

Once we have obtained the estimates for SDG indicators disaggregated by multiple socioeconomic characteristics, what is next? In order to properly monitor the SDGs and meet the promise of Leaving No One Behind, data disaggregation at multiple levels must become common practice.

Ideally, LNOB-disaggregated gender data (or multilevel disaggregated gender data) must be included in routine monitoring of progress. For this to happen, the following recommendations could be considered:

- Consistently review policy priorities/national strategies and identify priority indicators and population groups of interest to prioritize routine data collection and analysis.
- Consistently produce multilevel disaggregated data for such indicators, utilizing the population groups selected.
- Ensure that all forms of data collection are gathering sufficient information on sex, location, ethnicity or any other relevant socioeconomic characteristics for each individual.
- Train national statisticians for the production of LNOB estimates and include the consistent production of such estimates in their Terms of Reference.

- Once the estimates are available, ensure the estimates are published and widely distributed. For instance, ensure that national SDG databases, gender databases, or any other relevant data repositories showcase information at various levels of disaggregation.
- Include LNOB estimates in survey reports, monitoring reports and strategy/policy progress reports.
- Ensure LNOB estimates are also reported to the international statistical system. Although some global repositories do not allow for disaggregation for all indicators, others (such as the SENDAI Monitor) do offer the possibility of uploading disaggregated data, if desired.
- Overall, ensure LNOB data is widely used for policymaking and efficiently communicated (refer to Modules 10 and 11 of this Training Curriculum for more about this).

# 5. KEY TAKEAWAYS

- *Disaggregating data by sex is important to reveal the inequalities between men and women.*
- *Disaggregating data by variables other than sex – such as location, region, ethnicity and religion – is key to identify the multiple effects of discrimination on women's well-being.*
- *The compound effect of more than one form of discrimination often pushes women and girls furthest behind.*
- *Individual-level household surveys are a good source of microdata for this type of analysis because they interview women in the household. Ideally, choose surveys that are nationally representative and produced in a timely manner.*
- *To disaggregate data at multiple levels, the sample size has to be large enough and representative of the reference population to make claims about very specific population groups.*
- *When calculating estimates for an indicator in STATA or any other software for data analysis, always use sample weights to interpret results accurately.*
- *Hold multi-stakeholder consultations to identify the key priority areas, target groups and data disaggregation needs in your country.*
- *Read the metadata to understand international definitions, computation methods and recommended disaggregation variables before starting the LNOB analysis.*
- *Always test for significance of results at a 95 or 99 per cent significance level.*
- *Use the 'prtest' to test the significance of differences between two-sample proportions.*
- *Use the z statistic to test the significance of differences between the proportion of two samples.*
- *Consistently publish data disaggregated at multiple levels to ensure that the promise of leaving no one behind is being met.*