Reading and Understanding DHS Tables

Statistical tables can look intimidating at first glance. These examples, from recent DHS reports, highlight ways to read and understand tables.

Example 1: Knowledge of HIV Prevention Methods:
a question asked of all survey respondents

Step 1: Read the title and subtitle. They tell you the topic and the specific population group being described. In this case, the table is about women age 15-49 in Tanzania. This represents the entire female survey population in most DHS surveys.

Step 2: Scan the column headings - the top horizontal row. They describe how the information is categorized. In this case, each column represents one aspect of knowledge of prevention the women report to have. Note that the last column lists the (weighted) number of women in each group. For more on weighting, see example 4.

Step 3: Scan the row headings - the first vertical column. These show the different ways the data are divided up into categories based on population characteristics. In this case, the table presents knowledge of HIV-prevention methods by age, marital status, urban/rural residence, zone of residence, educational level, and wealth. Most of the tables in DHS reports will be divided up in the same categories.

Step 4: Look at very last row at the bottom of the table. These percentages represent the totals of all women age 15-49 who know each method of HIV prevention. In this case, 68.6% of women age 15-49 know that using condoms reduces the risk of getting the AIDS virus.

Step 5: To find out what percentage of women in Eastern zone know that using condoms and limiting sex to one uninfected partner reduces the risk of getting HIV, draw 2 imaginary lines, as shown on the table. 73.4% of women of women age 15-49 in Eastern Zone know that using condoms and limiting sex to one uninfected partner reduces the risk of getting the AIDS virus.

Table 5.2 Knowledge of HIV prevention methods [1]

Percentage of women age 15-49 who, in response to prompted questions, say that people can reduce the risk of getting the AIDS virus by using condoms every time they have sexual intercourse, by having one sex partner who is not infected and has no other partners, and by abstaining from sexual intercourse, by background characteristics, Tanzania HMIS 2007-08

| Background characteristic | Women | | | | |
	Using condoms[1]	Limiting sexual intercourse to one uninfected partner[2]	Using condoms and limiting sexual intercourse to one uninfected partner[1,2]	Abstaining from sexual intercourse	Number of women
Age					
15-24	67.6	79.5	60.0	82.3	3,730
15-19	62.8	76.2	54.8	79.2	1,984
20-24	73.1	83.2	66.0	85.8	1,746
25-29	72.5	84.3	65.7	86.5	1,603
30-39	70.3	84.2	63.3	88.3	2,521
40-49	63.8	80.6	57.1	85.8	1,488
Marital status					
Never married	65.4	80.8	59.2	82.7	2,214
Ever had sex	78.7	85.9	71.9	85.7	983
Never had sex	54.8	76.7	49.2	80.3	1,231
Married/living together	68.7	82.2	61.6	85.7	5,983
Divorced/separated/widowed	74.2	81.4	64.7	87.1	1,147
Residence					
Urban	74.5	89.3	70.5	88.8	2,459
Rural	66.5	79.1	58.2	83.9	6,884
Zone					
Western	77.3	81.8	68.4	87.1	1,682
Northern	66.3	88.6	62.6	88.2	1,449
Central	65.6	82.8	60.7	84.9	532
Southern Highlands	50.1	67.6	40.4	75.8	1,297
Lake	66.9	79.7	58.6	84.1	1,696
Eastern	78.0	89.4	73.4	88.1	1,436
Southern	79.6	83.2	70.0	88.9	942
Zanzibar	46.1	78.8	40.9	81.7	309
Education					
No education	54.7	71.2	46.0	76.8	1,983
Primary incomplete	62.6	75.8	53.4	80.6	1,517
Primary complete	74.4	85.6	67.6	88.6	4,945
Secondary +	77.5	94.2	74.9	92.7	898
Wealth quintile					
Lowest	61.2	74.5	52.2	80.4	1,700
Second	66.5	77.2	56.7	82.1	1,634
Middle	68.9	78.7	60.2	83.3	1,757
Fourth	69.0	84.4	63.2	87.9	1,867
Highest	74.7	90.3	70.8	89.9	2,384
Total	68.6	81.8	61.4	85.2	9,343

Practice: Use this table to answer the following questions (answers are upside down, below):

a) What percentage of youth age 15-24 are aware that using condoms AND limiting sex to one partner reduces the risk of getting the AIDS virus?

b) In what zone are women most aware that abstaining from sex reduces the risk of getting the AIDS virus?

c) Are urban or rural residents more likely to know that using condoms reduces the risk of getting the AIDS virus?

a)60; b) Southern-89%; c) urban-75%

Example 2: Payment for Sex and Condom Use
A question asked of a subgroup of survey respondents

Step 1: Read the title and subtitle. In this case, the table is about two separate groups of men a) all men age 15-49 and b) only men in that age group who reported that they paid for sex in the past 12 months.

Step 2: Identify the two panels. First identify the columns that refer to all men (a), and then isolate the columns that refer only to the men who reported having paid for sex (b).

Step 3: Look at the first panel. What percentage of all men age 15-49 have paid for sex in the past year? It's 8.3%.

Now look at the second panel. How many men are included in this group? Only 576, or 8.3% of the 6,975 men who were asked about paying for sex. The second panel is a subgroup of the first.

Step 4: There are very few men who report having paid for sex in the past year. Once these men are further divided into the background characteristics categories, there may be too few cases for the percentages to be reliable.

For example, look to see the percentage of men with secondary+ education who used a condom at last paid intercourse: 79.3%. This percentage is in parentheses because it is based on fewer than 50 men (unweighted) in this category. This means use this number with caution - it may not be accurate. (For more information on weighted and unweighted numbers, see example 4.)

If a number is replaced by an asterisk, it is because there are fewer than 25 people in that category, and results for this group are not reported. The subgroup is too small, and therefore those data are not reliable.

Table 7.4 Payment for sexual intercourse and condom use at last paid sexual intercourse: Men

Percentage of men age 15-49 reporting payment for sexual intercourse in the past 12 months, and among them, the percentage reporting that a condom was used the last time they paid for sexual intercourse, by background characteristics, Tanzania HMIS 2007-08

Background characteristic	Payment for sexual intercourse in the past 12 months		Condom use at last paid sexual intercourse	
	Percentage who paid for sexual intercourse	Number of men	Percentage reporting condom use	Number of men who paid for sexual intercourse in the past 12 months
Age				
15-24	8.1	2,916	57.2	236
15-19	4.6	1,768	55.6	82
20-24	13.4	1,148	58.0	154
25-29	10.8	1,004	60.6	108
30-39	8.9	1,846	62.6	165
40-49	5.5	1,210	61.8	67
Marital status				
Never married	8.4	2,931	58.4	246
Married or living together	6.8	3,701	61.7	253
Divorced/separated/widowed	22.5	343	58.7	77
Residence				
Urban	6.5	1,699	77.9	110
Rural	8.8	5,276	55.7	465
Education				
No education	10.4	829	53.2	86
Primary incomplete	9.0	1,534	60.2	138
Primary complete	8.8	3,597	59.4	316
Secondary +	3.5	1,016	(79.3)	36
Wealth quintile				
Lowest	9.7	1,173	41.8	114
Second	9.1	1,411	53.1	128
Middle	8.8	1,322	57.7	116
Fourth	9.0	1,395	67.6	125
Highest	5.4	1,670	83.7	91
Total	8.3	6,975	59.9	576

Note: Figures in parentheses are based on 25-49 unweighted cases. An asterisk indicates that an estimate is based on fewer than 25 unweighted cases and has been suppressed.

Note: When parentheses or asterisks are used in a table, the explanation will be noted under the table. So if there are no parentheses or asterisks on a table, you can proceed with confidence that enough cases were included in all categories.

Example 3: HIV Prevalence by Background Characteristics
Comparing Data and Understanding Patterns

Step 1: Read the title and subtitle. In this case, the table is about HIV prevalence among women age 15-49 in Tanzania.

Step 2: Scan the column headings - the top horizontal row. In this case there is only one variable, the percent of women who are HIV-positive.

The third column- "number tested" indicates how many women in each category were tested for HIV in the survey. In this case, 2,065 women living in urban areas were tested for HIV in the survey. In total, 8,179 women were tested.

Step 3: Scan the row headings - the first vertical column. These show the different ways the data are divided up into categories based on population characteristics. In this table HIV prevalence is presented by urban/rural residence, region of residence, educational level, employment status, and wealth. The data in these categories will help you understand how HIV prevalence varies throughout the country.

Step 4: Answer the following questions to understand how HIV prevalence is spread throughout the population:

   -What is the highest prevalence and the lowest prevalence (range)? Within the regions, prevalence ranges from 0.3% in Pemba to a high of 18.6% among women in Iringa.

   -Look for patterns: do results go up or down? Does prevalence vary within specific populations? For example, is there a clear pattern of HIV prevalence by wealth? by education?

   -Compare different groups: do urban residents have a different HIV prevalence than rural residents?

Step 5: What does all this mean? Once you see that urban residents and residents of Iringa and Dar es Salaam have the highest HIV prevalence, how can this knowledge affect your work? Program planners and policymakers can use this detailed information to inform their activities. Perhaps they should focus prevention campaigns in those areas, and make sure that there are health centers prepared to deal with larger numbers of HIV-positive women in those areas.

What other trends exist? In Tanzania, HIV prevalence is highest among the wealthiest women. What does this mean for the program planner? It means that outreach should not be targeted only at the poorest women in the population. Wealthy women are more likely to carry HIV infection than poorer women. Programs should be designed to address them. There is no clear pattern among educational levels, but it is clear that people in all educational groups are at risk of contracting HIV. Programs should be addressed to the full range of women—those who have received no education through those who have higher levels of education.

Table 9.4 HIV prevalence by socio-economic characteristics

Percentage HIV positive among women 15-49 who were tested, by socio-economic characteristics, Tanzania HMIS 2007-08

Socio-economic characteristic	Women Percentage HIV positive	Number
Residence		
Urban	10.6	2,065
Rural	5.3	6,114
Region		
Arusha	0.8	345
Dar es Salaam	10.4	690
Dodoma	4.1	294
Iringa	18.6	353
Kagera	3.5	451
Kigoma	2.7	361
Kilimanjaro	2.5	328
Lindi	4.9	215
Manyara	2.3	229
Mara	8.5	319
Mbeya	9.3	502
Morogoro	7.1	379
Mtwara	4.4	285
Mwanza	7.1	729
Pwani	8.4	177
Rukwa	5.7	276
Ruvuma	7.4	326
Shinyanga	8.4	660
Singida	3.0	171
Tabora	7.2	447
Tanga	6.7	372
Pemba	0.3	82
Unguja	0.9	187
Education		
No education	6.0	1,727
Primary incomplete	6.0	1,349
Primary complete	7.3	4,332
Secondary +	4.9	771
Employment (past12 months)		
Not employed	3.7	1,569
Employed	7.3	6,596
Wealth quintile		
Lowest	5.0	1,501
Second	6.6	1,455
Middle	5.1	1,582
Fourth	6.0	1,641
Highest	9.5	2,000
Total 15-49	6.6	8,179

Example 4: Understanding Samples and Weighting in DHS Tables

A sample is a group of people that have been selected for a survey. The sample is supposed to represent the entire population that you would like to learn about. Most countries want to collect data and report information for the entire country as well as for a country's regions or provinces.

DHS surveys are designed to provide these national and regional statistics. We want the sample surveyed in each region to resemble the actual population of that region, just as we want the national sample to resemble the actual population of the country. If the regions in a particular country vary in size and especially if some regions have very small populations, then a randomly-drawn sample may not include enough people from each region for analysis.

For example, let's say that you have enough money to interview about 9,000 women for a survey that should be representative of both the regions and the entire country (as in the Tanzania table below). In Tanzania, the regions are not evenly distributed - some regions are very heavily populated, while others have very small populations.

A sampling statistician can determine how many women should be interviewed in each region in order to get reliable statistics for the specific indicators the country is interested in. In the case of Tanzania, the blue column (1) shows the actual (unweighted) number of women selected and interviewed in each region, ranging from 233 in Pwani to 1,486 in Pemba. Now there are enough interviews to get reliable results in each region.

But now there is a new challenge. With this distribution of interviews throughout the regions, some regions are overrepresented and some regions are underrepresented. For example, Pemba's population in 2002 was about 362,000 people, which only represents 1% of the entire Tanzanian population. On the other hand, Mwanza's population in 2002 was almost 3,000,000, or approximately 9% of the Tanzanian population. But based on the blue column, our survey has interviewed more women in Pemba than women in Mwanza! Does this represent the population of the country?

In order to get statistics that are representative of the entire country, the distribution of the women in our sample needs to resemble the distribution of the women in the country. Women from a very small region, like Pemba, should only contribute a very small amount to the national total. Likewise, women from a largely populated region, like Mwanza, should contribute more. So DHS statisticians adjust or "weight" the numbers of women from each region so that each region's contribution to the total is proportionate to the actual population of the country. The numbers in the purple column (2) represent the "weighted" values. The total sample size of 9,343 women has not changed, but the distribution of the women in the regions has been changed to represent their contribution to the total population size. The weighted numbers suggest the true relationship between the populations of Mwanza and Pemba: the population of Mwanza is about nine times as large as Pemba.

Table 3.1 Background characteristics of respondents		
Percent distribution of women age 15-49 by background characteristics, Tanzania HMIS 2007-08		

Background characteristics	Women		
	Weighted percentage	Weighted number	Unweighted number
Region/Island			
Mainland	96.7	9,034	6,673
Arusha	4.1	383	306
Dar es Salaam	8.5	797	448
Dodoma	3.6	338	263
Iringa	4.3	403	280
Kagera	5.3	495	312
Kigoma	4.4	414	362
Kilimanjaro	4.1	379	321
Lindi	2.6	246	287
Manyara	2.8	263	326
Mara	3.9	368	393
Mbeya	6.2	581	292
Morogoro	4.7	436	247
Mtwara	3.5	324	267
Mwanza	8.9	833	373
Pwani	2.2	203	233
Rukwa	3.4	314	312
Ruvuma	4.0	372	335
Shinyanga	8.0	750	390
Singida	2.1	194	261
Tabora	5.5	518	357
Tanga	4.5	424	308
Zanzibar	3.3	309	2,670
Pemba	1.0	94	1,486
Unguja	2.3	214	1,184
Total 15-49	100.0	9,343	9,343

How do statisticians weight each category? They recalculate the categories to reflect the real population of the country. If you were to compare the light red column (3) to the actual population distribution of Tanzania, you would see that women in each region are contributing to the total sample with the same weight that they contribute to the population of the country. The weighted number of women in the survey now accurately represents how many women live in Mwanza- 9% of the Tanzanian population - and how few women live in Pemba - about 1% of the population.

With sampling and weighting, it is possible to interview enough women to provide reliable statistics at both the national and regional level, without distorting the overall population distribution within the country. In general, only the weighted numbers are shown in each of the DHS tables, so don't be distressed if these numbers seem low- they may actually represent a larger number of women interviewed. And remember, the table will use parentheses and asterisks to warn you if there are too few unweighted cases in any category.