Title of template

**Skill level needed:**Basic/Intermediate/Advanced

**Sample designs supported:**

1, 2, 3, 4, 5, 6

Designing and Implementing Gridded Population Surveys

**gridpopsurvey.com**

**A3. PSU sample – Excel to select PSUs**

Last updated: Aug 2022

**PSU sample selection in Excel**Example: Namibia

**Motivation:** After a PSU sample frame has been constructed, use this tutorial to draw a probability proportional to population size (PPS) sample of PSUs from a list of sample frame units with population totals.

**Example:** In this hypothetical example, the team manually updated the population estimates within Namibia’s EA boundaries using a gridded population dataset in ArcGIS (see Tutorial A2), and now they will sample EAs with PPS in Excel. The team aims to sample 500 EAs nationally; they have approximately equally allocated EAs across regions, and approximately optimally allocated EAs to urban/rural areas within each region, with adjustments to ensure a minimum of 5 EAs per stratum, as follows. During the sampling process, they wish to implicitly stratify by geography using serpentine sampling such that within strata, EAs are ordered from north-to-south and west-to-east.

|  |  |  |
| --- | --- | --- |
| **Region** | **Total estimated population** | **Allocation of PSUs** |
| **Urban** | **Rural** | **Total** | **Urban** | **Rural** | **Total** |
| !Karas | 0 | 88,539 | 88,539 | N/A | 35 | 35 |
| Erongo | 162,749 | 55,444 | 218,193 | 29 | 6 | 35 |
| Hardap | 37,527 | 58,487 | 96,014 | 19 | 16 | 35 |
| Kavango East | 47,020 | 81,311 | 128,331 | 14 | 21 | 35 |
| Kavango West | 31,642 | 91,540 | 123,183 | 10 | 25 | 35 |
| Khomas | 417,743 | 54,363 | 472,106 | 35 | 5 | 40 |
| Kunene | 0 | 109,350 | 109,350 | N/A | 35 | 35 |
| Ohangwena | 17,331 | 254,026 | 271,357 | 5 | 35 | 40 |
| Omaheke | 0 | 78,915 | 78,915 | N/A | 35 | 35 |
| Omusati | 0 | 268,630 | 268,630 | N/A | 35 | 35 |
| Oshana | 65,193 | 135,166 | 200,359 | 11 | 24 | 35 |
| Oshikoto | 10,009 | 204,115 | 214,124 | 5 | 30 | 35 |
| Otjozondjupa | 38,651 | 124,878 | 163,529 | 10 | 25 | 35 |
| Zambezi | 35,507 | 71,387 | 106,893 | 17 | 18 | 35 |
| **Namibia** |  |  | 2,539,524 | **155** | **345** | **500** |

**Steps:**

1. If your sample frame is not readable in Excel, convert it to a file type that Excel can read. For example, use one of these techniques:
* Use OpenOffice (www.openoffice.org) to open the .dbf file that is part of a sample frame shapefile layer, and save the file as .xls or .csv format.
* In ArcGIS, open the sample frame layer attribute table, and export to .txt. Open the .txt file in Excel and use the “Text to Columns” tool to align data in columns.
1. At a minimum, your spreadsheet should include (A) the sample frame unit unique ID (i.e., ea\_id) and (B) the number of people or households per sample frame unit (i.e., pop\_2000). In this example, we show three additional columns because we will explicitly stratify by (C) administrative region and level of urbanicity (i.e., strata), and within stratum we implicitly stratified by geography byway of ordering EAs by (D) latitude and (E) longitude (also called serpentine sampling).



1. Separate the sample frame values for each explicit stratum (i.e., on separate tabs, or in separate sections of a spreadsheet). Then within strata, order the sample frame by any implicit stratification values.
2. Finally, perform the following calculations:
	* + 1. Add column (F) with the cumulative sum of the sorted population.
			2. Document the number of EAs to be sampled in the stratum (n).
			3. Calculate the strata-specific sample interval (SI) by dividing the stratum cumulative population (cell F27) by the number of EAs to be sampled (call J2).
			4. Generate a random start number (RS) between 1 and SI (see cell J4).
			5. Calculate a series of numbers for sample selection: RS, RS+(1\*SI), RS+(2\*SI), ...,
			RS+((n-1)\*SI) (see cells J7:J11).
			6. Add column (G) to look-up and match the series numbers to cumulative population totals (column F). The first time that a unique series number (series\_num) appears in Column G, it is a sampled PSU (highlighted in green).

Figure 1. Example PPS sampling calculations for Oshikoto-Urban stratum (equations shown in red)



 [This work](http://www.gridpopsurvey.com/)© 2022 is licensed under [CC BY 4.0](http://creativecommons.org/licenses/by/4.0/?ref=chooser-v1)