Stratified data sampling with XLSTAT

demo_data_sampling.xls

Dataset for stratified sampling

The dataset used in this tutorial is the list of the employee of a company with some details about their gender (male/female) and their type of employment (full-time/part-time). The HR team wants to conduct a survey on the working condition that will be representative of the general opinion without interviewing every employee. They decide to conduct a stratified sampling.

There are 46% of female employee and 54% of male employee. 66% of the employee work full-time.
The part-time employees are more often female than male; female part-time employees represent 25% of the employees against 9% for the male part-time employee.

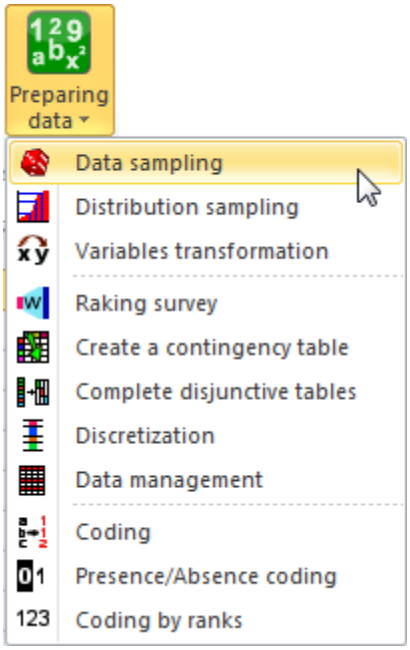
Descriptive statistics (Qualitative data):

Sample	No. of observations	No. of missing values	Sum of weights	No. of categories	Mode	Mode frequency	Category	Frequency per category	Rel. frequency per category (%)
Gender	97	0	97	2	Male	52	Female	45,000	46,392
							Male	52,000	53,608
Time	97	0	97	2	Full-time	64	Full-time	64,000	65,979
							Part-time	33,000	34,021
Strata	97	0	97	4	MF	43	FF	21,000	21,649
							FP	24,000	24,742
							MF	43,000	44,330
							MP	9,000	9,278

An Excel sheet with both the data and the results can be downloaded by clicking here.

Setting-up a stratified sampling of the data

Open the **Data sampling dialog boxPreparing data – Data sampling**.



Select the data including all the available columns (employee, gender, time, strata).

Choose the sampling option **Random stratified (2)**. This option takes into account the proportions of each strata.

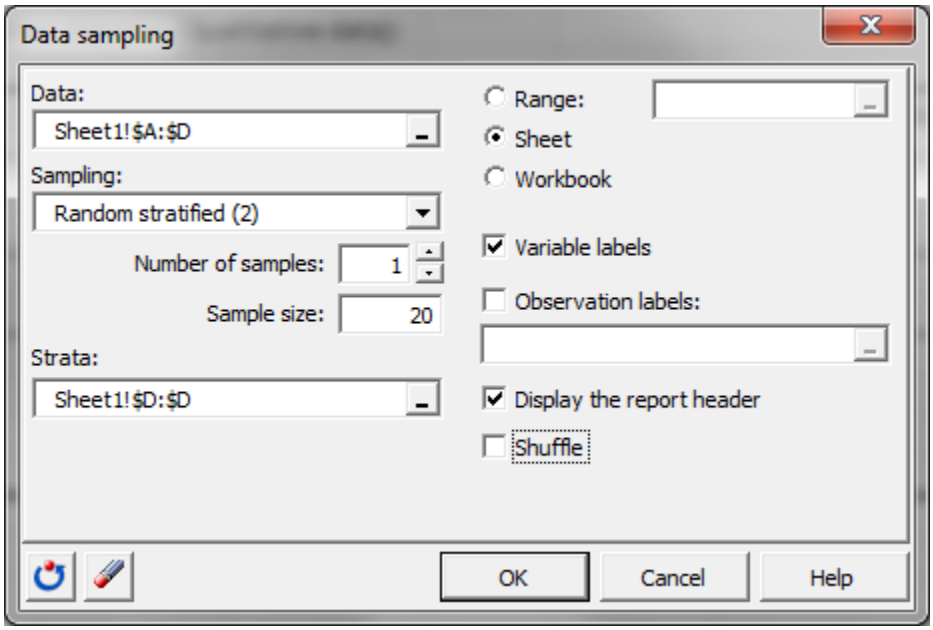
We want to generate a sample of 20 employees for the interviews. So enter “20” in the field **Sample size**.

Select the “strata” as the last column of the dataset.

The names of the variables are included in the selection so the option **variable labels** must be ticked.

We do not need to shuffle the individuals so we do not activate the option **shuffle**.

When everything is set, press **OK**.



Results of stratified data sampling

The results of the stratified data sampling appear in a new sheet. You find a table of 20 samples. As the sampling is random you may not have the exact same results. However you may have the same proportions for each category. This results in having the same amount of sample in each strata:

- 4 full-time female employees,
- 5 part-time female employees,
- 9 full-time male employees,
- 2 part-time male employees.

Sampled data:

Employee	Gender	Time	Strata
Em003	Male	Full-time	MF
Em023	Male	Full-time	MF
Em027	Male	Full-time	MF
Em049	Male	Full-time	MF
Em055	Male	Full-time	MF
Em067	Male	Full-time	MF
Em085	Male	Full-time	MF
Em088	Male	Full-time	MF
Em094	Male	Full-time	MF
Em025	Male	Part-time	MP
Em039	Male	Part-time	MP
Em017	Female	Part-time	FP
Em022	Female	Part-time	FP
Em058	Female	Part-time	FP
Em064	Female	Part-time	FP
Em096	Female	Part-time	FP
Em031	Female	Full-time	FF
Em045	Female	Full-time	FF
Em047	Female	Full-time	FF
Em050	Female	Full-time	FF

Below you have the descriptive statistics that are computed on the stratified sample. You can compare these statistics with the one obtained on the population:

- 20% for the 22% of full-time female employees,
- 25% for the 25% of part-time female employees,
- 45% for the 44% of full-time male employees,
- 10% for the 9% of part-time male employees.

Descriptive statistics (Qualitative data):

Sample	No. of observations	No. of missing values	Sum of weights	No. of categories	Mode	Mode frequency	Category	Frequency per category	Rel. frequency per category (%)
Gender	20	0	20	2	Male	11	Female	9,000	45,000
							Male	11,000	55,000
Time	20	0	20	2	Full-time	13	Full-time	13,000	65,000
							Part-time	7,000	35,000
Strata	20	0	20	4	MF	9	FF	4,000	20,000
							FP	5,000	25,000
							MF	9,000	45,000
							MP	2,000	10,000