

# User information on SAFE

---

## Content

How is confidentiality of the results ensured in the 2011 Census? .....	3
Why has a new procedure been introduced?? .....	3
How does SAFE change the data? .....	3
Does SAFE have an impact on the official numbers of inhabitants?? .....	3
To what extent does SAFE modify the data? .....	4
What should users keep in mind? .....	4
To what extent does SAFE modify the data? – Deviation indicators .....	5

## ***How is confidentiality of the results ensured in the 2011 Census?***

### **How is confidentiality of the results ensured in the 2011 Census?**

The individual data of the 2011 Census are subject to statistical confidentiality according to Article 16 of the Federal Statistics Law. This means it has to be ensured that no conclusions can be drawn from the publication tables which would allow to identify individuals. In the 2011 Census, statistical confidentiality is ensured through the SAFE procedure. "SAFE" stands for safe anonymisation of individual data. The procedure differs from how traditional confidentiality methods work. In the latter, table cells which would allow direct or indirect identification of individuals or their personal or material situation are not published. In SAFE, such identification of individuals is ruled out by slight modification of the individual data.

## ***Why has a new procedure been introduced??***

In the 2011 Census, users are not only provided with a fixed set of tables. Instead, they will be able to compile user-specific tables based on a given set of evaluation variables. This creates an extremely large potential of evaluation in the 2011 Census. In a range of tables as complex as this, the confidentiality activities required by a traditional cell suppression procedure would be highly time-consuming and costly and the information loss caused by table cell suppression would have been unjustifiably high.

## ***How does SAFE change the data?***

Applying SAFE ensures automatically that all tables that can be compiled from the evaluation database will meet the confidentiality requirements.

SAFE changes the data in a way ensuring that, in the protected data, every combination of variables occurring in the original data (e.g. combinations of age, sex, marital status, religion, employment data, etc.) will occur at least three times or not at all. Then it is no longer possible to draw conclusions on individuals or on the information provided by them.

The modifications are done in a controlled way so that they largely offset each other. What is achieved by that approach is that differences in central evaluation tables (including, among other things, all municipality fact sheets) are minimised and all important statistical results are reliable.

## ***Does SAFE have an impact on the official numbers of inhabitants??***

No, it does not. The official numbers of inhabitants (total numbers of inhabitants of the municipalities) are calculated on the basis of the original data and are published without the modifications applied by SAFE. Consequently, at this point, minor differences may occur where users themselves add up, for example, the modified numbers of men and women in a municipality as provided by SAFE.

## ***To what extent does SAFE modify the data?***

The majority of the original frequencies is changed. That modification generally amounts to up to +/-2. The maximum deviations of up to +/-10 in the census of buildings and housing and +/-6 for people occur much less often.

This means that the protected data should be considered as approximate values with deviations of around +/-2, in some cases slightly more, in some slightly less. Especially for smaller frequencies observed (under 10), this produces a distinctly coarsening effect, which is deliberate, so that the protection of individual data can be ensured across the multitude of possible evaluations.

A general rule applying to larger deviations (more than +/-2) is: the larger, the less frequent. Deviations of +/-4, for example, are much less frequent than deviations of +/-3.

For large original values of about 1000 or over, the changes applied are almost irrelevant in terms of percentage.

## ***What should users keep in mind?***

For very small municipalities with low numbers of cases for the values in the results tables – and consequently with many confidentiality cases –, the modifications applied to the data are of course more marked in relative terms. If a value is changed from “2” to “3”, its relative change is as much as 50%.

In rare cases, the confidentiality procedure may produce constellations which at first sight are surprising and may seem illogical. For example, the result for “number of buildings with 2 dwellings” can be compared with the result for “number of dwellings in buildings with 2 dwellings”. In the original results, the number of dwellings in such a case should generally be exactly twice the number of buildings, while in the protected results this is not always the case. Keeping in mind that both figures include a deviation of about +/-2, or slightly more, from the original result, with the deviation possibly having different sizes or even different signs, it is obvious that this can explain even larger deviations and that by no means it indicates errors in data collection or evaluation. Changing the data to 2 fewer buildings and 2 more dwellings would show, in the above relationship between the “number of buildings with 2 dwellings” and the “number of dwellings” in these buildings, 6 too many dwellings.

It should also be noted that such inconsistencies may occur even in the stock of original data. One of the reasons is that residential space used commercially is not shown in the tables. This means that differences are not only due to the SAFE procedure.

The effects described here occur mainly in tables which either contain only data on buildings or refer to small municipalities (under 10,000 inhabitants). In any other tables, only a small part of the data are kept confidential through SAFE, so that the effects described here are practically irrelevant.

## ***To what extent does SAFE modify the data? – Deviation indicators***

The indicators compiled here refer to the total of all statistical results in the cells of the evaluation tables for which the SAFE confidentiality procedure was applied. This is a total of about 6.8 million table cells showing population data (the remaining population data are extrapolated from the household sample; these table cells do not have to be kept confidential by means of SAFE) and about 25.1 million cells with data on buildings and dwellings.

On the whole, the average of the changes occurring in the cells is close to zero because the modifications are done in a specific way so that they largely offset each other. For example, increasing the frequency of a variable's value from two to three is generally accompanied by decreasing the frequency of another value, for example, from one to zero.

The impact of the SAFE confidentiality procedure on the buildings and dwellings data is somewhat larger on the whole. Here, the average *absolute* change in original frequencies caused by SAFE is 1.9. For population data, the absolute changes are an average 1.1.

Table 1 shows that – as mentioned above – the change in original frequencies caused by SAFE is up to +/-2 for most of the cells, that is, for 72% of the buildings and dwellings data shown in the tables and for almost 93% of the population data shown. Deviations of +/-3 occur for 0.4 million out of 6.8 million (6%) of the population data represented and for 13% (3.3 of 25.1 million) of the table cells containing data on buildings and dwellings. Deviations of +/-4 occur much less often, that is, in only 1% of the table cells with population data and in less than 8% of the cells with buildings and dwellings data.

In population data, deviations of up to +/-6 occur in the results presented. This maximum deviation is very rare, i.e. it occurs only in about 1 of 100,000 results (in exactly 64 of the 6.8 million results). Somewhat larger deviations occur for the data shown on buildings and dwellings. About one of 10,000 (exactly 2,204 out of 25.2 million) results involve a deviation of +/-9 or +/-10. The probability of coming across a result with the maximum deviation of +/-10 is less than one in a million because it occurs in only 14 of the 25 million results.

**Table 1: Deviations of the results shown in the evaluation tables, before and after application of confidentiality procedure**

(Absolute) deviation	Buildings and dwellings		Population	
	Number of table cells	Cumulated percentage of table cells with deviation of up to ... (in %)	Number of table cells	Cumulated percentage of table cells with deviation of up to ... (in %)
0	3,431,878	13.6	1,719,409	25.2
1	8,591,297	47.8	3,103,036	70.7
2	6,081,200	72.0	1,503,944	92.7
3	3,306,790	85.2	412,711	98.7
4	1,944,141	92.9	74,752	99.8
5	1,068,964	97.1	10,829	100.0
6	505,685	99.1	64	100.0
7	163,918	99.8	-	100.0
8	47,953	100.0	-	100.0
9	2,190	100.0	-	100.0
10	14	100.0	-	100.0
Total	25,144,030		6,824,745	

As SAFE performs the modifications in a way so that they largely offset each other, the deviations observed for the central key figures are markedly smaller. Table 2 shows the frequency distribution of the SAFE modifications specifically for these key figures. Key figures in this context are, for example, data on the total numbers of buildings and dwellings in the individual municipalities, administrative districts, etc. and federal results broken down by just one variable such as the population's age structure. Also, the discrepancies observed by users when, for example, adding up the number (modified by SAFE) of men and women in a municipality with less than 10,000 inhabitants and comparing it with the number of inhabitants shown for that municipality are included in table 2. For municipalities with more than 10,000 inhabitants, however, most of the results shown are extrapolated from the household sample and are not kept confidential by SAFE. As those results are rounded before shown in the tables, further deviations may occur, which are not included in table 2.

On average, the SAFE-related absolute changes in such key figures are 1.2 for buildings data, 0.7 for dwellings data and just 0.25 for population data. As can be seen in table 2, the key figures on buildings and dwellings were changed by a maximum of just +/-4, the key figures on the population structure by a maximum of 2. For about three quarters (75.3%) of the key figures on the population structure, there is no deviation from the original frequencies counted.

**Table 2: Deviations in central key figures, before and after application of confidentiality procedure**

(Absolute) deviation	Buildings and dwellings		Population	
	Number of table cells	Cumulated percentage of table cells (with deviation of up to ...) (in %)	Number of table cells	Cumulated percentage of table cells (with deviation of up to ...) (in %)
0	8,737	26.7	12,515	75.3
1	13,742	68.8	4,092	99.9
2	7,236	90.9	12	100.0
3	2,886	99.7	-	100.0
4	85	100.0	-	100.0
Total	32,686		16,619	