## Regulations on the analysis of microdata

<u>Basic requirements concerning the program code</u>

Hereafter, the formal requirements of the program code are presented. These regulations are to be followed in order for the RDC staff to execute the program code sent for remote execution resp. check the results produced at a safe centre for statistical confidentiality. If these regulations are not followed, the program code will not be executed and results will not be checked for confidentiality and released to the user.

You find further information on the regulations by clicking the following link:
[www.forschungsdatenzentrum/en/confidentiality](www.forschungsdatenzentrum/en/confidentiality).

1. The program code is to have a clear structure (possibly by using a master program code). The sample program code provided by the RDC staff is to be applied.
2. The program code is to be commented broadly and comprehensively.
3. All variables and values are to be labeled unambiguous.
4. The output needs to be identically reproducible by the associated program code.
5. All tabular and analytical results have to be saved in a processable format (e.g. XLSX, SAV). All graphical results have to be saved in a *non*-processable format (e.g. PDF, JPEG).
6. It is to be unambiguously marked in the program code weather a result is supposed to be released to the user or is only serving the RDC check for confidentiality.
7. To enable the check for confidentiality, when analyzing subpopulations, you always have to produce the results for the remaining subpopulation as well. The according references have to be marked (see no. 6).
8. To enable the check for confidentiality, the underlying unweighted number of cases has to be stated for every result that is to be released to the user. The according references have to be marked (see no. 6).
9. To enable the check for confidentiality, the underlying unweighted number of cases has to be stated for every graphical analysis that is to be released to the user. The according references have to be marked (see no. 6).
10. To enable the check for confidentiality, for every value table and every statistical figure based on economic or tax statistics, the highest two values, the number of cases and the sum of all values have to be stated. The according references have to be marked (see no. 6).
11. Identical results can only be marked to be released once. If results have to be released again in duly substantiated exceptional cases, an exact reference to the earlier analysis is to be made.
12. The merging of external variables is only permitted if it was allowed in advance by the supervising RDC location. The external data needs to have a certain format.

List of criteria on the admission of output

Hereafter, you find the most important rules that are used by German official statistics to ensure the confidentiality of statistical results. These rules also apply to the results of scientific analyses. Statistical confidentiality is checked by the supervising RDC location. If these rules are not followed by the scientific user, the RDC have the possibility to reject the output.

1. Rules for numbers of cases and value tables
   1.1 A value is to be kept confidential if only 1 or 2 cases contribute to it. This is also valid for more advanced analyses (so called rule of minimum case number).
   1.2 A table cell is to be kept confidential if the frequency of an inner table field is distinguished by only 1 from the frequency of the corresponding marginal field (so called rule of marginal value).
   1.3 A value is to be kept confidential if the contribution of the largest individual value or the two largest individual values is larger than a set proportion of the overall value (so called rule of dominance).
2. Rules for more advanced analyses
   2.1 Individual values are generally to be kept confidential. This includes, among others, listings of individual values (list-command) or minima and maxima.
   2.2 The result of a quantile calculation is to be kept confidential if less than 3 cases contribute to a quantile segment.
   2.3 Graphic results are to be kept secret if there is at least one confidential value among the underlying case numbers or values (see no. 1.1 to 1.3).
   2.4 Results for subpopulations have to be kept confidential if conclusions that are relevant for confidentiality can be drawn by the comparison of the results from one subpopulation and the whole population.

Not all statistics are being kept confidential by the same rules. For some statistics, not all of the rules described above are relevant. For other statistics, there are additional rules. You find further information on the rules and their validity by clicking the following link: www.forschungsdatenzentrum/en/confidentiality.