
SAMPLING PLAN FOR THE 2022 CENSUS AND DRAWING OF THE MAIN SAMPLE

Steffen Klink, Kai Lorentz

📌 **Keywords:** 2022 Census – 2011 Census – sample – sampling – number of inhabitants

ABSTRACT

The 2022 Census will be conducted with the help of a register-based procedure in Germany, as was the 2011 Census. Population register data are of central importance in the census, especially for determining the number of people living in each municipality. These figures have to be corrected for overcoverage and undercoverage (outdated and missing data entries). Stratified random sampling will be an important tool in this regard. It serves not only to correct the population register data to determine the numbers of inhabitants but also to cover variables not contained in the registers. This article focuses on the sampling design, that is the design of the sampling plan and the sampling methodology.



Steffen Klink

holds a degree in economics from the universities of Heidelberg and Southampton (United Kingdom). He is assistant head of the “Household Sample (Methodology)” section at the Federal Statistical Office, heads the “household generating procedure” subproject and provides support in designing the methodology of the household sample for the 2022 Census.



Dr. Kai Lorentz

studied mathematics at Johannes Gutenberg University Mainz. He is head of the “Mathematical/Statistical Techniques for Population and Social Statistics” section at the Federal Statistical Office. His work focuses in particular on issues relating to the mathematical and statistical methodology for the 2022 Census.

1

Introduction

The basic sampling procedure for the 2022 Census has remained largely unchanged from the procedure used for the 2011 Census. Studies on the sample size and sample design were once again produced with Professor Dr. Ralf Münnich and his team from Trier University (Burgard et al. 2019) for the 2022 Census.

The main new feature of the 2022 sample survey compared with 2011 is that it will now also be used to correct register data in all municipalities irrespective of their size. This means that the sample survey is also conducted in municipalities with fewer than 10,000 inhabitants in order to correct data sent from the population registers.

The objectives of the sample as well as the legal bases for it are examined in Chapter 2. Chapter 3 describes the methodology of the sampling plan for drawing the main sample to achieve Objective 1 (referred to below as the Objective 1 sample) and provides an overview of the sizes of the samples obtained at the level of addresses and individuals. Chapter 4 is dedicated similarly to the sample used to achieve Objective 2 (referred to below as the Objective 2 sample) and Chapter 5 to the follow-up survey that is conducted for quality assurance purposes. The article concludes by looking ahead to the supplementary sample in spring 2022, estimation and error calculation.

2

The role of samples in the 2022 Census

The most important target variable of the census is the number of inhabitants down to the level of the municipalities (Bretsch/Lorentz, 2019). In the 2022 Census, this figure is calculated based on the population register. First, a population count relevant to the census date is created from two submissions from population registers in order to ensure that subsequent registrations and de-registrations are also taken into account. This baseline figure is then adjusted on an automated basis to elimi-

nate impermissible multiple cases¹. This adjustment of the internal statistical population count is followed by a twofold primary statistical adjustment: first, the data are adjusted by conducting a full survey of special facilities, which include collective living quarters and residential establishments. The second adjustment is made for addresses without special facilities, also referred to below as normal addresses, which are corrected to eliminate outdated records and missing records that have been extrapolated from the sample. This fulfils the first key objective of the sample: the identification of over-coverage and undercoverage in the population count is instrumental for establishing the number of inhabitants (Objective 1). In all of the cases mentioned, the adjustments are always internal statistical corrections made to the data stock, i.e. adjusted data are not transferred back to the municipalities for other purposes (so-called “prohibition to transfer the data back”).

The sample also serves to cover variables on which no information is currently available in registers (e.g. data on education) but which need to be provided by the 2022 Census under the requirements of European and national legislation (Objective 2). Both objectives are regulated by Section 11 (1) of the national Census Act (2022 Census Act). The list of variables demanded by the European Union (EU) has remained unchanged from the last census and is regulated in the corresponding implementing regulation (No 543/2017). The additional variables are also surveyed in a sample of residential establishments², but not in collective living quarters (Section 17 of 2022 Census Act). In municipalities with at least 10,000 inhabitants, the survey relating to Objective 2 is conducted at all normal addresses that are also selected to calculate the number of residents/inhabitants, pursuant to Section 11 (3) of the 2022 Census Act. In municipalities with fewer than 10,000 inhabitants, on the other hand, only a subsample of the population is interviewed since there is no requirement for these ‘small’ municipalities to provide data on the additional variables on a municipal-specific basis³. [↪ Overview 1](#)

-
- 1 One example of this is where a personal data set with several main residences or only with a second residence is contained in the data stock.
 - 2 A full survey is carried out at residential establishments in order to determine the number of inhabitants.
 - 3 Their only contribution is towards extensive results at the level of administrative districts.

Sampling plan for the 2022 Census and drawing of the main sample

Overview 1

Sampling frames for the 2022 Census

	Municipality with fewer than 10,000 inhabitants	Municipality with 10,000 and more inhabitants
Normal address	Sample to correct misalignments so as to determine the number of inhabitants Objective 2 subsample	Sample to correct misalignments so as to determine the number of inhabitants, congruent to the Objective 2 sample
Residential establishment	Full survey to determine the number of inhabitants Objective 2 sample	Full survey to determine the number of inhabitants Objective 2 sample
Collective living quarter	Full survey to determine the number of inhabitants No Objective 2 coverage	Full survey to determine the number of inhabitants No Objective 2 coverage

In addition to the aforementioned objectives, Section 11 (1) of the 2022 Census Act also defines the spatial units of the sample. In this regard, the sample must take account of the survey areas (referred to below as sampling points) explicitly set out in Section 11 (1) of the 2022 Census Act. These units are primarily municipalities. However, there are also some particular features that are specific to individual Länder; for example, in addition to non-associated municipalities, so-called associations of municipalities or parts thereof can also form a sampling point. Furthermore, provision is made under the Act for cities with a population of more than 400,000 to be disaggregated into separate parts.

As well as regional stratification, the sample is also structured in subject-matter terms. In order to optimise the sample results, it is customary for the population of the sampling units – in this case, addresses with residential space (Section 12 (1) of the 2022 Census Act) – to be subdivided into classes prior to sampling. To achieve the result we are aiming for, these classes need to be as internally homogenous as possible and as externally heterogeneous as possible. These strata are based on address size classes (Section 12 (2) of the 2022 Census Act), since the frequency of register errors generally increases in line with the size of an address. The stratification is examined in closer detail in Chapter 3.

The addresses for the sample are selected on a random basis, which is why the results of the sample survey display a random error (in this case: the simple standard error). The information available at the time of drawing the sample is to be used to determine the size of these random sampling errors. As an ex-ante planning factor, the purpose of the random sample error is therefore to be able to control the accuracy of the sample results to be extrapolated. The rule here is that the greater the

sample size, the more accurate the result. Accordingly, when a sample is designed, one of two approaches can be adopted: either a sample size is determined first, which then results in a certain degree of precision of the extrapolated results; or a level of accuracy that you aim to achieve is determined first, and the necessary sample size can subsequently be derived on this basis. In the 2022 Census Act, the legislator has chosen the second option.

The levels of accuracy for the sample with regard to Objective 1 are stipulated by Section 11 (2) of the 2022 Census Act, for three municipality size classes. In municipalities with at least 10,000 inhabitants, the aim is to achieve a level of accuracy that is no more than 0.5 % of the simple relative standard error. In municipalities with fewer than 10,000 inhabitants but at least 1,000 inhabitants, the required levels of accuracy are determined using a so-called accuracy target function, i.e. the level of accuracy is represented by a gradual transition from a simple relative standard error of 0.5 % to a simple absolute standard error of 15 persons (Bretsch/Lorentz, 2019). In municipalities with fewer than 1,000 inhabitants, a simple absolute standard error of 15 persons is generally given as the desired accuracy target.

$$y(x) = \begin{cases} 15, & \text{für } 0 \leq x < 1\,000 \\ \frac{7}{1\,000}x - \frac{28}{10\sqrt{50}}\sqrt{x-200} + \frac{96}{5}, & \text{für } 1\,000 \leq x < 10\,000 \\ \frac{1}{1\,000}x, & \text{für } x \geq 10\,000 \end{cases}$$

This subdivision is necessary, not only to meet the demand to minimise the burden on those persons required to give information but also to gauge the number of inhabitants as accurately as possible. If, for exam-

ple, the same accuracy targets were to be prescribed in municipalities with fewer than 1,000 inhabitants as in municipalities with more than 10,000 inhabitants, this would lead to full surveys being conducted in many small municipalities. The reason for this is that the absolute sample size (number of persons interviewed) is more important than the relative sample size (persons interviewed as a share of all inhabitants in a municipality) for achieving a particular level of accuracy.

In addition to the household sample survey described above, a further sample survey will be conducted as part of the 2022 Census, the so-called follow-up survey (Klink/Bihler, 2015). As in the 2011 Census, the purpose of the follow-up survey is to subsequently evaluate the quality of the calculated numbers of inhabitants. Both European and national guidelines exclusively provide for subsequent verification of the number of inhabitants calculated in the 2022 Census.

The 2022 Census Act requires that the quality of the sample results and of the survey at special-facility addresses be examined with regard to the numbers of inhabitants that have been ascertained (Section 22 (1) of the 2022 Census Act). A key objective of the follow-up survey is to determine (and subsequently extrapolate) survey findings establishing the existence of individuals which either match or deviate from the household sample survey or the survey at special-facility addresses (residential establishments), without these deviations being incorporated into the census result. These comparisons would not be possible in an independent sample survey. For this reason, the follow-up survey is designed as a subsample of the household sample survey pursuant to Section 11 of the 2022 Census Act and as a subsample of the survey at residential establishments pursuant to Section 14 of the 2022 Census Act 2022.

3

The sampling plan for the Objective 1 sample

3.1 Sampling points

The purpose of the Objective 1 sample under the 2022 Census is to determine the numbers of inhabitants in the sampling points. This is done by statistically adjusting population register information, using the data collected by the Objective 1 sample. There are 6,507 sampling points nationwide. [↘ Table 1](#)

Table 1
Sampling points by Land

	Number
Schleswig-Holstein	303
Hamburg	7
Niedersachsen	408
Bremen	4
Nordrhein-Westfalen	407
Hessen	425
Rheinland-Pfalz	170
Baden-Württemberg	1,104
Bayern	2,064
Saarland	52
Berlin	12
Brandenburg	417
Mecklenburg-Vorpommern	206
Sachsen	423
Sachsen-Anhalt	218
Thüringen	287
Germany	6,507

3.2 Statistical units

The address is both the survey unit and sampling unit for the Objective 1 sample. This is necessary in order to be able to identify overcoverage and undercoverage: if the sampling unit were the individual (as exists in the register), only cases of overcoverage would be identifiable since cases of undercoverage would have no chance of being included in the sample in this instance.

3.3 Sampling frame

The sampling frame for the Objective 1 sample is an extract of addresses from the control register, as it is regulated under the Census Preparation Act (Section 3 of the 2022 Census Preparation Act on the status following completion of the update pursuant to Section 5 (2) number 1a). In addition to so-called normal addresses, this register also includes addresses of special facilities, which are broken down into residential establishments and collective living quarters.

Normal addresses without residential space as well as 'special cases'¹⁴ had no chance of being included in the Objective 1 sample. However, an existing population register entry for a person at an address is not necessary in order for the address to have a chance of being included in the sample; if we want to be able to take account of addresses that are purely missing entries, all addresses with no registered occupants must also have the chance of being included in the sample, if they have residential space.

The sampling frame ultimately comprised 20,686,186 addresses, broken down into 20,630,053 normal addresses, 7,651 residential establishments and 49,014 collective living quarters. This sampling frame was determined in advance through a complex process of combining the stocks from a number of different registers, in particular the population registers and land registers, as well as other data stocks, as a universe of all (potential) residential addresses.

3.4 Stratification

The Objective 1 sample is a stratified random sample. Stratification is carried out on a hierarchical basis, with the first level composed of the classification of addresses into the sampling points. At the second hierarchical level, the addresses are stratified on the basis of address size classes at normal addresses as well as at residential establishments and collective living quarters. The number of address size classes per sampling point is flexible, depending on the size of the sampling points. This is different to the methodology used in

the 2011 Census, in which there were precisely eight address size classes for each sampling point. Following an accompanying academic study, the team from Trier University proposed fixing the number of address size classes per sampling point, as shown in [Table 2](#).

Table 2
Number of address size classes per sampling point

Number of registered inhabitants (main and second residences)	Number of address size classes
0	
1 to 49	1
50 to 99	2
100 to 149	3
150 to 199	4
200 to 299	5
300 to 399	6
400 to 599	7
600 to 799	7
800 to 999	9
1,000 to 1,499	10
1,500 to 1,999	11
2,000 to 2,499	12
2,500 to 4,999	13
5,000 to 9,999	14
10,000 to 24,999	15
25,000 and more	16

As a result, there can be a maximum of 16 address size classes per sampling point. In addition, there is a further stratum of so-called void addresses¹⁵ as well as a stratum of residential establishments and a stratum of collective living quarters.

The Objective 1 sample consisted of 97,673 strata in total.

The normal addresses with registered occupants were allocated to the address size classes as follows: the normal addresses with registered occupants were arranged in ascending order according to the number of registered occupants and then allocated to the address size classes in such a way that these show approximately the same overall number of registered occupants.

⁴ This includes seafarers, skippers of inland waterway vessels, foreign armed forces, members of the diplomatic service of other states.

⁵ Void addresses are addresses with no registered occupants whatsoever, i.e. addresses which are not registered as either a main or second residence, but which do have residential space.

3.5 Sample size

In contrast to the 2011 Census, the 2022 Census Act does not fix an upper limit for the overall sample size. Instead there are accuracy targets for estimating the number of inhabitants at the level of the sampling points. The sample size is derived from the accuracy targets and the ex-ante assessment regarding their attainment.

To this end, a number of preliminary decisions were required, which were taken in cooperation with Trier University:

- 1) The estimation method: The numbers of inhabitants in the sampling points are calculated using the so-called GREG estimator (generalised regression estimator; Särndal et al., 1992). It is expected that this will reduce the sample size that is required, compared with the simple expansion estimation, while achieving the same level of accuracy in results. This approach requires the existence of an auxiliary variable correlated with the target variable. In this case, the auxiliary variable is that of the number of main residences recorded with registered occupants.
- 2) In order to quantify the gain in accuracy described under 1), corresponding variance reduction factors were estimated for the sampling points from the 2011 Census sample data. However, these estimates themselves are subject to a degree of uncertainty since they are derived from a sample. This factor therefore needs to be taken into account, in the form of a global mark-up factor on the sample size.
- 3) For the sampling points, the aim is to have a minimal sample size of 100 addresses. Simulations by Trier University had shown (Burgard et al., 2020) that this minimum sample size is required in order to be able to make inferential statistical statements.
- 4) Constraints of the sampling fractions per stratum (so-called box constraints): In terms of achieving a good balance in the sample with regard to the address size and the wish for not too great a spread in the design weights (inverse values of the probabilities of being included in the sample, also referred to as a simple expansion estimation factor), the sampling fractions in the address size classes 1 to 16 were fixed, as far as possible, within an interval between 5 and 50%.

5) For the “0” address size classes (so-called strata of void addresses) of the addresses with no registered occupants whatsoever, the sampling fractions cannot be determined mathematically from the outset. At the recommendation of Trier University, a standard sampling fraction of 20% was established in this case.

- 6) The strata of residential establishments and collective living quarters are considered as total strata for the a-priori estimation of the accuracy level (i.e. fixed sampling fraction of 100%, design weight 1).

In accordance with the accuracy targets, the requisite minimum sample size at address level was calculated for each sampling point under the aforementioned boundary conditions. In cases where it was not possible to achieve the accuracy targets – which may occur due to the boundary condition of the upper box constraint – this was gradually increased for the sampling points in question until the target was finally met. The minimum random sample size target of 100 addresses per sampling point was prioritised over the box constraint requirements.

These calculations produced an overall sample size at federal level as a first intermediate step. Due to the uncertainty in the variance reduction factors, which are a key determinant for the ex-ante accuracy estimation, the overall sample size that was determined for the population of normal addresses with registered occupants was increased by 40%. This mark-up factor was a recommendation by academics from Trier University on the basis of the simulations carried out.

3.6 Allocation of the sample size

The overall sample size determined above was allocated to the individual strata using the BCOpt (‘optimum allocation under box constraints’) method. The method was applied without taking into account the variance reduction factors (in order to prevent any unequal treatment of sampling points) and using the Euclidean norm. As a result, a certain balance is achieved between the quality of the anticipated results at federal level and at the level of the sampling points (Bankier, 1988). Here too, the minimum sample size target of 100 addresses per sampling point was implemented as a matter of priority.

The allocation of the sample sizes for the 2022 Census by Länder is shown in [Table 3](#).

Sampling plan for the 2022 Census and drawing of the main sample

Table 3
Sample sizes (Objective 1) for the 2022 Census by Land

	Addresses	Population (sole residence, main residence, second residence)	Inhabitants (sole residence and main residence)	Size of main sample Objective 1 (Addresses)	Objective 1 sampling – inhabitants	Objective 1 sampling – population
Germany	20,686,718	86,952,257	84,328,028	2,248,923	11,531,900	12,035,171
Normal addresses	20,630,053	85,158,578	82,637,570	2,192,258	9,841,442	10,241,492
Residential establishments and normal addresses	20,637,704	85,502,915	82,936,414	2,199,909	10,140,286	10,585,829
Schleswig-Holstein	926,307	3,052,471	2,955,804	113,301	468,627	494,294
Normal addresses	923,947	2,990,653	2,896,091	110,941	408,914	432,476
Residential establishments and normal addresses	924,058	2,995,791	2,900,851	111,052	413,674	437,614
Hamburg	275,310	1,934,111	1,898,858	18,491	181,636	184,729
Normal addresses	274,231	1,872,590	1,838,479	17,412	121,257	123,208
Residential establishments and normal addresses	274,330	1,883,374	1,848,461	17,511	131,239	133,992
Niedersachsen	2,440,709	8,409,555	8,118,252	194,916	899,822	940,018
Normal addresses	2,434,229	8,243,811	7,962,704	188,436	744,274	774,274
Residential establishments and normal addresses	2,435,235	8,268,886	7,983,599	189,442	765,169	799,349
Bremen	144,249	693,758	681,770	8,620	64,319	65,961
Normal addresses	143,786	677,130	666,086	8,157	48,635	49,333
Residential establishments and normal addresses	143,838	681,269	670,078	8,209	52,627	53,472
Nordrhein-Westfalen	4,193,382	18,627,698	18,296,852	287,348	1,764,704	1,801,831
Normal addresses	4,181,405	18,274,249	17,954,145	275,371	1,421,997	1,448,382
Residential establishments and normal addresses	4,182,766	18,335,754	18,011,792	276,732	1,479,644	1,509,887
Hessen	1,483,275	6,588,320	6,378,120	147,012	874,878	909,578
Normal addresses	1,480,207	6,485,855	6,280,157	143,944	776,915	807,113
Residential establishments and normal addresses	1,480,630	6,505,798	6,298,292	144,367	795,050	827,056
Rheinland-Pfalz	1,328,118	4,272,642	4,136,351	106,734	444,017	460,435
Normal addresses	1,325,676	4,197,438	4,065,777	104,292	373,443	385,231
Residential establishments and normal addresses	1,325,935	4,212,726	4,079,655	104,551	387,321	400,519
Baden-Württemberg	2,635,900	11,528,577	11,140,405	335,334	1,812,991	1,893,508
Normal addresses	2,626,084	11,284,125	10,909,615	325,518	1,582,201	1,649,056
Residential establishments and normal addresses	2,627,937	11,348,637	10,968,411	327,371	1,640,997	1,713,568
Bayern	3,323,074	13,935,218	13,332,694	552,417	2,527,172	2,685,274
Normal addresses	3,313,502	13,622,824	13,049,572	542,845	2,244,050	2,372,880
Residential establishments and normal addresses	3,314,936	13,692,326	13,100,414	544,279	2,294,892	2,442,382
Saarland	335,352	1,041,288	1,004,455	26,751	107,357	111,802
Normal addresses	334,682	1,023,940	988,071	26,081	90,973	94,454
Residential establishments and normal addresses	334,706	1,026,127	990,111	26,105	93,013	96,641
Berlin	335,708	3,866,927	3,768,634	20,577	363,729	373,814
Normal addresses	334,280	3,769,761	3,675,067	19,149	270,162	276,648
Residential establishments and normal addresses	334,480	3,791,863	3,696,080	19,349	291,175	298,750
Brandenburg	741,924	2,634,099	2,558,090	104,188	436,255	455,396
Normal addresses	740,269	2,576,508	2,505,018	102,533	383,183	397,805
Residential establishments and normal addresses	740,462	2,584,040	2,510,910	102,726	389,075	405,337
Mecklenburg-Vorpommern	467,108	1,692,514	1,622,237	73,966	308,845	327,844
Normal addresses	466,014	1,656,072	1,588,334	72,872	274,942	291,402
Residential establishments and normal addresses	466,149	1,660,912	1,592,403	73,007	279,011	296,242
Sachsen	854,818	4,200,372	4,093,406	115,121	631,747	655,860
Normal addresses	852,954	4,111,532	4,009,776	113,257	548,117	567,020
Residential establishments and normal addresses	853,170	4,128,992	4,025,159	113,473	563,500	584,480
Sachsen-Anhalt	611,474	2,271,557	2,212,564	66,721	304,316	317,588
Normal addresses	610,026	2,216,186	2,162,345	65,273	254,097	262,217
Residential establishments and normal addresses	610,116	2,221,348	2,166,407	65,363	258,159	267,379
Thüringen	590,010	2,203,150	2,129,536	77,426	341,485	357,239
Normal addresses	588,761	2,155,904	2,086,333	76,177	298,282	309,993
Residential establishments and normal addresses	588,956	2,165,072	2,093,791	76,372	305,740	319,161

4

Objective 2 sample

The purpose of the Objective 2 sample is to cover additional variables (e.g. labour force participation, education). It is designed as a subsample to the Objective 1 sample described in Chapter 3. Here too, the clear methodological requirements under the 2022 Census Act, as described in Chapter 2, must be fulfilled.

4.1 Sampling points with fewer than 10,000 inhabitants

Pursuant to Section 11 (3) number 1 of the 2022 Census Act, the Objective 2 sample at normal addresses is identical to the Objective 1 sample in sampling points with 10,000 or more inhabitants. In the smaller sampling points, however, Section 11 (3) number 2 of the 2022 Census Act prescribes a subsample which must not cover more than 8% of the number of inhabitants nationwide in municipalities of this size. The stratification pattern for the Objective 2 sample at the normal addresses of the small sampling points is identical to the stratification pattern of the Objective 1 sample. This means that for each of the sampling points in question, there are address size classes ranging from “0” (addresses with no registered occupants), “1”, ..., to “14” (maximum). Address size classes “15” and “16” are only possible in sampling points with more than 10,000 inhabitants. In contrast to the Objective 1 sample, the survey aims of the Objective 2 sample are varied. As a result, an approach involving the optimisation of a single variable cannot be used for the design; the aim instead is to achieve proportionate sampling per stratum (relative to the population).

A fixed subsampling fraction of 40% of the Objective 1 sample is set for the strata of the void addresses. That corresponds to a sampling fraction of 8% relative to the population.

For the other address size classes, the aim was also to have a uniform sampling fraction of 8% relative to the population. However, due to the lower box constraint of 5% for the Objective 1 sample, which came into play pri-

marily in the lower address size classes, this could not be achieved in full. The approach was therefore as follows:

The address size classes were considered consecutively, from small to big, and examined to see whether a sampling fraction of 8% relative to the population is achievable, as a subsample to the Objective 1 sample. If yes, the corresponding subsampling fraction was calculated. If no, the subsampling fraction was fixed at 100% and an excess was calculated which was added as an Objective 2 sample size in the next size class up (where possible). The accumulated excesses were implemented as the Objective 2 sample size no later than when the highest size classes were reached, for which the Objective 1 sampling fractions are close to the upper box constraint of 50%.

The addresses for the Objective 2 sample were then drawn. A check was subsequently carried out to establish whether the upper limit of 8% of the overall population nationwide registered in the small sampling points was exceeded. That was not the case. If this had been the case, simple random sampling would have been used to gradually remove addresses from the Objective 2 sample drawn at the normal addresses in the small sampling points, until such time as compliance with the upper limit was achieved.

4.2 Objective 2 sample at the residential establishments

Under the 2022 Census Act, 8% of the national population living in residential establishments is to be surveyed (Section 17 (1) of the 2022 Census Act). Given that the number of residential establishments in many sampling points is low, the regional stratification for the Objective 2 sample at the residential establishments is aggregated.

In order to be able to assess the level of accuracy following the survey, the aim is to have two residential establishments as a minimum sample size for the corresponding Objective 2 stratum. Since the expectation is that 8% of the population also corresponds to 8% of residential establishments, the minimum stratum size for the Objective 2 strata of residential establishments

should be no fewer than 25 residential establishments in order to maintain the minimum sample size of two residential establishments per corresponding stratum.

Within the Länder, the Objective 1 strata of residential establishments in the sampling points are therefore aggregated so that the number of residential establishments in the resulting Objective 2 strata of residential establishments is at least 25.

With regard to implementation:

- › The Objective 1 strata of residential establishments for each Land are first arranged in ascending order by administrative districts/towns not attached to an administrative district, and then by the Objective 1 stratum size.
- › Objective 1 strata of residential establishments are then aggregated to create a new stratum until a stratum size of at least 25 residential establishments is achieved. In the case of the final stratum to be constructed in this manner, the stratum size of 25 residential establishments might not be achieved; this stratum is then aggregated with the penultimate stratum.
- › The Objective 2 sample at residential establishments was drawn from these strata with a sampling fraction of 8%.
- › Finally, a review was carried out to establish whether the upper limit of 8% of the inhabitants at residential establishment addresses nationwide was exceeded in the Objective 2 residential establishment sample addresses.

That was not the case. If the upper limit had been exceeded, random sampling would have been used to once again remove residential establishments from the residential establishment addresses that had been drawn. In this case, the newly created strata of residential establishments nationwide would have been arranged in descending order by stratum size. Based on this new arrangement, random sampling would have been used to remove one unit per stratum until such time as compliance with the upper limit was achieved.

↘ **Table 4** shows the result of the Objective 2 sampling for the normal addresses in small sampling points and at residential establishments.

5

The follow-up survey

The Objective 1 normal addresses that were drawn and the residential establishments are the sampling frame for the follow-up survey.

5.1 Stratification

Strata are combined aggregated in the follow-up survey due to the small sampling fractions and sample sizes compared with the Objective 1 sample at the normal addresses and residential establishments. The procedure is largely the same as the methodology applied in the 2011 Census (Klink/Bihler, 2011).

The regional stratification level is raised from the sampling point level to the level of the administrative regions.⁶

Furthermore (with the exception of the stratum of void addresses), the Objective 1 address size classes within the administrative regions are, where necessary, aggregated in an appropriate manner so that these new size classes contain at least 200 inhabitants of the Objective 1 sample for each new (real) address size class. This is done so as to ensure that there are at least eight sample units available per stratum.

The Objective 1 strata of the void addresses and the residential establishments are aggregated at the level of the administrative regions.

5.2 Sampling of the follow-up survey

In each case, 4% of the addresses are drawn from the strata of the follow-up survey that were created beforehand (Section 22 (1) of the 2022 Census Act).

A check was subsequently carried out to ensure that the upper limits set for the follow-up survey – namely 4% of inhabitants at the residential establishments and at the normal addresses included in the Objective 1 sample – were not exceeded nationwide. This was once again

⁶ In Länder without administrative regions, the regional stratification level is the Länder.

Table 4
Sample sizes (Objective 2) for the 2022 Census by Land

	Addresses	Size of main sample Objective 2 (addresses)	Objective 2 sampling – inhabitants	Objective 2 sampling – population
Germany	20,686,718	1,525,344	7,195,786	7,431,395
Normal addresses	20,630,053	1,524,735	7,173,653	7,405,932
Residential establishments and normal addresses	20,637,704	1,525,344	7,195,786	7,431,395
Schleswig-Holstein	926,307	73,712	268,557	278,880
Normal addresses	923,947	73,703	268,356	278,665
Residential establishments and normal addresses	924,058	73,712	268,557	278,880
Hamburg	275,310	17,419	121,856	123,846
Normal addresses	274,231	17,412	121,257	123,208
Residential establishments and normal addresses	274,330	17,419	121,856	123,846
Niedersachsen	2,440,709	173,507	677,899	702,781
Normal addresses	2,434,229	173,426	675,658	700,067
Residential establishments and normal addresses	2,435,235	173,507	677,899	702,781
Bremen	144,249	8,161	49,078	49,791
Normal addresses	143,786	8,157	48,635	49,333
Residential establishments and normal addresses	143,838	8,161	49,078	49,791
Nordrhein-Westfalen	4,193,382	270,848	1,400,179	1,425,917
Normal addresses	4,181,405	270,740	1,396,030	1,421,580
Residential establishments and normal addresses	4,182,766	270,848	1,400,179	1,425,917
Hessen	1,483,275	112,165	627,881	648,817
Normal addresses	1,480,207	112,131	626,794	647,622
Residential establishments and normal addresses	1,480,630	112,165	627,881	648,817
Rheinland-Pfalz	1,328,118	103,161	370,661	382,342
Normal addresses	1,325,676	103,140	369,206	380,788
Residential establishments and normal addresses	1,325,935	103,161	370,661	382,342
Baden-Württemberg	2,635,900	203,513	1,030,209	1,065,492
Normal addresses	2,626,084	203,366	1,025,731	1,060,568
Residential establishments and normal addresses	2,627,937	203,513	1,030,209	1,065,492
Bayern	3,323,074	267,331	1,203,907	1,263,139
Normal addresses	3,313,502	267,215	1,200,523	1,258,345
Residential establishments and normal addresses	3,314,936	267,331	1,203,907	1,263,139
Saarland	335,352	24,852	85,976	89,243
Normal addresses	334,682	24,850	85,796	89,062
Residential establishments and normal addresses	334,706	24,852	85,976	89,243
Berlin	335,708	19,164	271,506	278,101
Normal addresses	334,280	19,149	270,162	276,648
Residential establishments and normal addresses	334,480	19,164	271,506	278,101
Brandenburg	741,924	57,575	236,822	244,104
Normal addresses	740,269	57,560	236,400	243,610
Residential establishments and normal addresses	740,462	57,575	236,822	244,104
Mecklenburg-Vorpommern	467,108	37,759	142,200	148,711
Normal addresses	466,014	37,748	141,967	148,429
Residential establishments and normal addresses	466,149	37,759	142,200	148,711
Sachsen	854,818	64,654	340,718	350,254
Normal addresses	852,954	64,637	339,488	348,889
Residential establishments and normal addresses	853,170	64,654	340,718	350,254
Sachsen-Anhalt	611,474	44,515	182,323	187,389
Normal addresses	610,026	44,508	182,167	187,147
Residential establishments and normal addresses	610,116	44,515	182,323	187,389
Thüringen	590,010	47,008	186,014	192,588
Normal addresses	588,761	46,993	185,483	191,971
Residential establishments and normal addresses	588,956	47,008	186,014	192,588

Table 5

Sample sizes of the follow-up survey for the 2022 Census

	Addresses	Sample size of the follow-up survey	Sampling of the follow-up survey	
		addresses	inhabitants	population
Germany	20,686,718	84,401	392,880	409,419
Normal addresses	20,630,053	84,114	381,799	396,731
Residential establishments and normal addresses	20,637,704	84,401	392,880	409,419


not the case. If the upper limits had been exceeded, the sampling for the follow-up survey would have been repeated until such time as compliance with the upper limits was achieved.

↘ Table 5 shows the sample sizes of the follow-up survey for the 2022 Census.

6

Outlook

The 2022 Census Act regulates the data contained in the population registers and based on which the main sample is drawn. Following the drawing of the main sample in September 2021, new addresses that are added to the sampling frame will also be taken into account in a supplementary sample, to be conducted in March/April 2022. Residential establishments are not included in this supplementary sample. Addresses which are added between the supplementary sample and the census date, are included in the number of inhabitants without being corrected by the sample.

An error calculation is performed following the estimation procedure. The quality of the estimated results (number of inhabitants at sampling point level as well as demographic Objective 2 results) can then be assessed. 

BIBLIOGRAPHY

Bankier, Michael D. *Power allocations: Determining sample sizes for subnational areas*. In: The American Statistician. Volume 42. Issue 3/1988, p. 174 ff. Published online in 2012. DOI: www.tandfonline.com

Burgard, Jan Pablo/Münnich, Ralf/Rupp, Martin. *Die Entwicklung des Stichprobenkonzepts für den Zensus 2021*. In: WISTA Wirtschaft und Statistik. Special edition 2021 Census. 2019. p. 23 ff.

Burgard, Jan Pablo/Münnich, Ralf/Rupp, Martin. *Qualitätszielfunktionen für stark variierende Gemeindegrößen im Zensus 2021*. In: AStA Wirtsch Sozialstat Arch 14, p. 5 ff. 2020. DOI: doi.org/10.1007/s11943-019-00256-6.

Bretsch, Corinna/Lorentz, Kai. *Präzisionsziele für die Ermittlung der Einwohnerzahl bei der Haushaltsstichprobe im Zensus 2021*. In: WISTA Wirtschaft und Statistik. Special edition 2021 Census. 2019. p. 12 ff.

Klink, Steffen/Bihler, Wolf. *Die Wiederholungsbefragung beim Zensus 2011*. In: WISTA Wirtschaft und Statistik. Issue 2/2015, p. 42 ff.

Särndal, Carl-Erik/Swensson, Bengt/Wretman, Jan. *Model Assisted Survey Sampling*. In: Springer Series in Statistics. 1992.

LEGAL BASIS

Act on Conducting the Census in 2022 (2022 Census Act – ZensG 2022) of 26 November 2019 (Federal Law Gazette I, p. 1851), amended by Section 1 of the Act of 3 December 2020 (Federal Law Gazette I, p. 2675).

Regulation (EC) No 763/2008 of the European Parliament and of the Council of 9 July 2008 on population and housing censuses (Official Journal of the EU No L 218, p. 14).

Commission Implementing Regulation (EU) 2017/543 of 22 March 2017 laying down rules for the application of Regulation (EC) No 763/2008 of the European Parliament and of the Council on population and housing censuses as regards the technical specifications of the topics and of their breakdowns (Official Journal of the EU No L 78, p. 13).

Commission Implementing Regulation (EU) 2017/881 of 23 May 2017 implementing Regulation (EC) No 763/2008 of the European Parliament and of the Council on population and housing censuses, as regards the modalities and structure of the quality reports and the technical format for data transmission, and amending Regulation (EU) No 1151/2010 (Official Journal of the EU No L 135, p. 6).

Extract from the journal WISTA Wirtschaft und Statistik

Published by:

Statistisches Bundesamt (Federal Statistical Office)

www.destatis.de

You may contact us at

www.destatis.de/kontakt

Abbreviations

WISTA	=	Wirtschaft und Statistik
JD	=	annual average
D	=	average (for values which cannot be added up)
Vj	=	quarter of a year
Hj	=	half-year
a. n. g.	=	not elsewhere classified
o. a. S.	=	no main economic activity
St	=	piece
Mill.	=	million
Mrd.	=	billion

Explanation of symbols

–	=	no figures or magnitude zero
0	=	less than half of 1 in the last digit occupied, but more than zero
.	=	numerical value unknown or not to be disclosed
...	=	data will be available later
X	=	cell blocked for logical reasons
I or —	=	fundamental change within a series affecting comparisons over time
/	=	no data because the numerical value is not sufficiently reliable
()	=	limited informational value because numerical value is of limited statistical reliability