Approved by the order of the Chairman of the Committee on Statistics of the Ministry of National Economy of the Republic of Kazakhstan dated August 22, 2016 169

**Methodology for conducting sample surveys of enterprises**

**Chapter 1. General provisions**

1. Methodology for conducting sample surveys of enterprises (hereinafter - Methodology) refers to the statistical methodology, formed in accordance with international standards and approved in accordance with the [Law](http://adilet.zan.kz/rus/docs/Z100000257_#z0) of the Republic of Kazakhstan dated March 19, 2010 "On State Statistics" (hereinafter - Law).

2. The methodology is intended for use by structural units Committee on Statistics of the Ministry of National Economy of the Republic of Kazakhstan (hereinafter – Committee).

3. The methodology defines the main aspects and methods of analysis of the surveyed population, as well as the selection of surveyed sample units; sample quality characteristics; and recommendations for extending sample results to the general population.

4. The purpose of this methodology is to reduce the burden on respondents; some nationwide statistical observations in the Republic of Kazakhstan are carried out by a selective method.

The main advantages of using sampling method in modern statistics are:

1) reduction of the time for carrying out statistical work;

2) reducing the information load on respondents;

3) significant savings in labor costs, material and financial resources for conducting the survey;

4) reducing the number of registration errors, since with a decrease in the volume of the observed population to work attract more professionally trained personnel and more carefully control their activities;

5) significantly accelerated, compared with continuous examination, obtaining the results of the study.

5. This Methodology uses concepts in the meanings defined in the Law , as well as the following definitions:

1 ) **panel observation method** - a method of collecting information in which a certain group of units of analysis is periodically polled for a relatively long time, and the subject of study remains constant;

2) the general population - a complete group of all units of analysis, whose characteristics are subject to evaluation;

3) individual catalog - a list of surveyed units with the corresponding classification attributes, intended for a specific statistical observation;

4) representativeness - the correspondence of the characteristics [of the sample](http://ru.wikipedia.org/wiki/%D0%92%D1%8B%D0%B1%D0%BE%D1%80%D0%BA%D0%B0) to the characteristics of the [population](http://ru.wikipedia.org/wiki/%D0%9F%D0%BE%D0%BF%D1%83%D0%BB%D1%8F%D1%86%D0%B8%D1%8F) or [the general population](http://ru.wikipedia.org/wiki/%D0%93%D0%B5%D0%BD%D0%B5%D1%80%D0%B0%D0%BB%D1%8C%D0%BD%D0%B0%D1%8F_%D1%81%D0%BE%D0%B2%D0%BE%D0%BA%D1%83%D0%BF%D0%BD%D0%BE%D1%81%D1%82%D1%8C) ;

5) mathematical expectation - the average value of a particular characteristic, as well as the weighted average of all results with a weight of probabilities reflecting the possibility of occurrence in each result;

6) parameter - a value calculated from all values in the set of the general population, that is, a descriptive measurement of the general population;

 7) stratum - division into special layers of units (respondents) with the same or similar indicators.

8) sampling plan is a set of specifications that define the general population and sample units, as well as the degree of probability of possible samples;

9) sample population (sample) - a set of cases (subjects, objects, events, samples), using a certain procedure, selected from the general population for participation in the study;

 10) sample size - the total number of units of observation in the sample;

6. The sample set is formed on the basis of an order for sample design submitted by the Committee's structural subdivisions to the Department of Statistical Registers and Classifications of the Committee.

The sample design order specifies:

1) the name of the structural unit;

2) the name of the survey;

3) the purpose of the survey;

4) description of the population;

5) the main feature for calculating descriptive statistics;

6) requirement for representativeness;

7) examination period;

8) the deadline for submitting the sample to the customer.

7. The main stages in the formation of a sample population are:

1) definition of the general population and division of e e on strata;

2) assessment of the main characteristics of the general population;

3) determination of the sample size by stratum;

4) calculation of sampling error;

5) formation of a sample population;

6) weighing, if necessary, re-weighing.

8. A completed sample order is submitted electronically and contains the following information:

1) a list of sampling units;

2) the size of the general population;

3) sample size

4) calculations of descriptive statistics (average value, dispersion and

standard deviation, standard error, relative standard error);

5) distribution coefficient (weight of inclusion) by stratum.

**Chapter 2** **Population and classification system**

9. When surveying enterprises, the most appropriate source for determining the general population is the Statistical Business Register (hereinafter - SBR), which includes information on individual entrepreneurs registered in the territory of the Republic of Kazakhstan, legal entities and their separate and structural subdivisions.

10. The selection is based on individual catalogs based on the Catalog Order Table (hereinafter - COT). COT contains the conditions for selecting units from the SBR to conduct a particular statistical observation and is formed annually by December 10 for the next reporting year at the republican level.

11. When forming a sample population, for stratification, statistical classifications are used that are available in the structure of an individual catalog, such as GCTEA (by type of economic activity), CSE (enterprise size), CATO (territorial location), CSEC (economic sector) , CPV (by production volume) and other classifications.

12. After stratification, enterprises are selected from the general population by a simple random selection method. This method is called stratified (stratified) simple random selection and its distinguishing feature is that the principle of random selection is always strictly observed.

**Chapter 3. Main characteristics of the general population**

13. To calculate the main characteristics of the general population, the attributes contained in the SBR are mainly used : “number of employees”, “production volume” and “turnover size”. In addition, data obtained from the survey of the past period or data obtained from other sources are used to calculate the main characteristics of the population.

14. Average arithmetic value ( ) (for example : number of employees, volume of production, size of turnover), is calculated for all units of the general population.

= / ,

Where:

- the number of elements of the general population i - stratum,

- indicator of the i -stratum.

15. The measure of variability in a population is called the standard deviation. Its square is called the population variance.

1. Population variance ( ):

,

1. Standard deviation ( *S* ):

= ,

The dispersion shows how close the estimated indicators are to the mathematical expectation of the estimated function, and the standard deviation characterizes the degree of their closeness to the parameter. An evaluation function is a mathematical function that calculates an estimated indicator for a particular parameter.

The main characteristics of the general population are given in Appendix and 1 to this Methodology.

**Chapter 4**

16. Sample size is determined by four factors.

The first factor is the number of groups and subgroups (stratification) whose analysis is recommended.

The second factor is the required accuracy of the results.

The third factor is sampling cost.

The fourth factor is the spread of population values.

17. When organizing sample observation, the sample size primarily depends on the permissible sampling error, it is important to correctly determine **the size of the sample population** , which with a certain probability will provide a given accuracy of the observation results.

18. There are two important criteria for determining the distribution of a sample across different stratums.

The first criterion is convenience, a way of apportioning the sample size:

 N i

n i = ------ \* n,

 N

Where:

n i – sample set i – strata;

i = 1,2,…,h;

N i - the number of enterprises in the i-th stratum, while i = 1,2…., h;

N- the number of enterprises in the general population.

The second criterion is accuracy: the method of optimal distribution is chosen, which gives the smallest mean square error (standard error) of the sample.

The costs of sampling from different strata are the same, the optimal distribution formula is called the Neumann distribution. The sample size is determined by the formula:

 ,

Where:

h - the number of strata in the aggregate;

i = 1,2,…,h;

S i - the standard deviation of the i-th stratum.

19. When organizing sampling, the sample size primarily depends on the size of the sampling error. By increasing the sample size, its error is reduced to small sizes.

20. To obtain the same data quality in each stratum, except for the above proportional and optimal distribution, the sample size is adjusted to the optimal value of the relative standard error (coefficient of variation) as follows:

,

Where:

– sample size;

 - the volume of the general population;

– coefficient of confidence squared;

- the variance of the general population;

- the marginal squared error.

**Chapter 5. Assessing the representativeness of sample survey data on enterprises**

21. In statistical observations carried out by the sampling method, representativeness errors are eliminated by ensuring the principle of random selection.

22. Possible discrepancies between the characteristics of the sample and the general population are measured by the standard error (mean error) of the sample.

In mathematical statistics, the values of the standard error of the sample ( μ) are determined by the formula:

,

The standard error of the sample fraction ( μg ) is given by:

,

Where:

 w - share in the sample population.

23. To establish the marginal error ( ) of the sample, the following formula is used:

 = t ⋅ μ

The coefficient t - determined by the probability P (0≤P≤1) given by the researcher. For values of P approaching one, the possibility that the mean of the general population will differ from the mean of the sample population by more than . indicates the accuracy guaranteed by a given confidence level of the probability P. In this case, the higher the confidence level (for example, the values 0.90; 0.95; 0.99, etc. are used), the higher the coefficient t, and, consequently, the value of the limit errors Δ.

*t* =1.28 for 80% confidence level;

*t* =1.64 for 90% confidence level;

*t* =1.96 for 95% confidence level;

*t* =2.58 for 99% confidence level.

24. Confidence interval - the allowable deviation of the observed values from the true ones. The size of this assumption is determined by the expert, taking into account the requirements for the accuracy of information. As the margin of error increases, the sample size decreases and the confidence level remains at 95%.

The confidence interval shows the range of location of the results of sample observations.

Confidence limits of the general mean are determined based on the inequality |-| ≤ , from which we get that x - ≤  ≤x + .

25. The standard error of the sample shows absolute values, and to determine the relative value, the ratio of the standard error to the estimated mean is used. This formula is called relative standard error ( SEi ) (coefficient of variation):

SEi = \*100%

The relative standard error is calculated after sampling surveys. The sample survey database for calculating the relative standard error of the sample after the survey is provided by the structural subdivisions of the Committee to the authorized subdivision of the Committee. For this, a variable is used that is associated with an element of the general population, such as: income, turnover, volume of services rendered, headcount of employees. The value of " relative standard error" estimates the percentage of deviation of the obtained sample results from the true value of the general population. In practice, the larger the value of the relative standard error , the relatively greater the scatter and the lower the evenness of the values under study. If the relative standard error for each stratum is less than 10%, then the variability of the variation series is considered insignificant, from 10% to 20% refers to the average, more than 20% and less than 33% to significant, more than 33% the sample is unrepresentative and a decision is made to increase sample size. To reduce the relative sample standard error by a factor of two, the sample size is quadrupled.

Sampling error calculations given in Appendix and 2 to this Methodology.

**Chapter 6 Formation and rotation of the sample population**

26. The method of panel observation is used in nationwide statistical surveys. The main distinguishing **feature of the panel method** is the assessment of the dynamics of indicators.

In panel observation, in order to achieve an acceptable accuracy of estimates of the studied indicators, changes in the composition of the units of the general and sample population, caused by demographic processes of merger and fragmentation, the emergence of new ones and the elimination of existing ones, are taken into account. Taking into account changes in the composition of the general and sample population , rotation of the composition of the sample.

27. In case of panel observation , the rotation of the sampling population is carried out by the method of simple random selection in a random order, which accordingly requires the recalculation of the distribution coefficients, taking into account the rotational units of the sampling population for a certain date. The rotation of the composition of the sample population ensures the representativeness of the sample until the end of the surveyed period.

28. The rotation of the sample population is carried out in specialized software at the republican level by an authorized structural subdivision of the Committee.

29. For short-term trade statistics, sampling is carried out once in a certain period, after which the sample and the population remain unchanged throughout the survey period and, accordingly, the sampling spread factors also remain unchanged throughout the survey period.

**Chapter 7 Recommendations for weighing and re-weighing**

30. As a result of sample observations, each enterprise is assigned a certain weight (spreading coefficient) to spread the sample data to the general population. This process is called weighing.

The spread factor ( *f i* ) is calculated as follows:



Where:

N *i* - population size in stratum *i;*

*n i* - sample size in stratum *i.*

The probability of inclusion of an enterprise ( *f k* ) in the sample is determined as follows:

*f k = n i* */ N i*

31. If data are not received from enterprises, the replacement of enterprises that did not respond is prohibited, since the replacement of enterprises that did not respond will lead to sampling bias and errors. To compensate for cases of complete non-receipt and if missing data cannot be edited, it is recommended to apply coefficients for incomplete coverage ( *k i* ) for each start , this method is called re-weighting :

*k i = N i / n i -fact*

Where:

N *i* - the size of the general population,

 *n i -fact* - the actual number of surveyed enterprises.

Re-weighing is always done after sampling.

**Chapter 8 Recommendations for extending sampling results to the general population**

32. Ultimate purpose selective observations is characteristic general aggregates on basis data received By sample. The distribution of the data of sample observation to the general population is carried out according to the following formula.

|  |  |
| --- | --- |
|  | Z = ,or:Z = .Where:- the sum of the indicator of the sample i -stratum;- the average value of the indicator of the sample population i -stratum. |

# Appendix 1

# to the Methodology for conducting sample surveys of enterprises

Main characteristics of the general population

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Cities of republican significance, regions** | **Population** | **Average value** | **Dispersion** | **Standard deviation** |
| = **/** |  | = |
| Astana city | 12 829 | 5.91 | 77.86 | 8.82 |
| Almaty city | 33 607 | 5.14 | 65.18 | 8.07 |
| Akmola | 3 201 | 10.19 | 150.24 | 12.26 |
| Aktobe | 4 941 | 6.98 | 91.34 | 9.56 |
| Almaty | 6 959 | 8.29 | 122.68 | 11.08 |
| Atyrau | 3 860 | 8.2 | 129.27 | 11.37 |
| Batys Kazakhstan | 3 350 | 9.02 | 141.55 | 11.9 |
| Zhambyl | 2860 | 8.03 | 99 | 9.95 |
| Karaganda | 7991 | 7.49 | 106.83 | 10.34 |
| Kostanai | 4 164 | 9.41 | 135.71 | 11.65 |
| Kyzylorda | 2875 | 7.4 | 100.15 | 10.01 |
| Mangystau | 5 885 | 6.69 | 82.13 | 9.06 |
| Ontustik Kazakhstan | 8 219 | 8.38 | 103.35 | 10.17 |
| Pavlodar | 5 216 | 7.85 | 113.07 | 10.63 |
| Soltustik Kazakhstan | 2978 | 8.78 | 129.08 | 11.36 |
| Shygys Kazakhstan | 7 358 | 8.59 | 109.92 | 10.48 |
| **Republic of Kazakhstan** | **116 293** | **7.02** | **97.31** | **9.86** |

# Appendix 2

# to the Methodology for conducting sample surveys of enterprises

Sampling error calculations

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Cities of republican significance, regions** | **Population** | **Sample population** | **Standard sampling error** | **Marginal sampling error** | **Relative standard error** |
| **N** | **n** |  |  = t ⋅ μ | SEi = μ/ |
| Astana city | 12 829 | 2475 | 0.1287 | 0.2523 | 2.18% |
| Almaty city | 33 607 | 4244 | 0.1198 | 0.2348 | 2.33% |
| Akmola | 3 201 | 1037 | 0.322 | 0.6312 | 3.16% |
| Aktobe | 4 941 | 1438 | 0.217 | 0.4253 | 3.11% |
| Almaty | 6 959 | 2180 | 0.2014 | 0.3947 | 2.43% |
| Atyrau | 3 860 | 1078 | 0.2956 | 0.5793 | 3.60% |
| Batys Kazakhstan | 3 350 | 975 | 0.3289 | 0.6447 | 3.64% |
| Zhambyl | 2860 | 877 | 0.2902 | 0.5688 | 3.61% |
| Karaganda | 7991 | 2347 | 0.1831 | 0.3589 | 2.44% |
| Kostanai | 4 164 | 1286 | 0.2792 | 0.5472 | 2.97% |
| Kyzylorda | 2875 | 879 | 0.2924 | 0.5731 | 3.95% |
| Mangystau | 5 885 | 1657 | 0.1895 | 0.3714 | 2.83% |
| Ontustik Kazakhstan | 8 219 | 2638 | 0.1677 | 0.3286 | 2.00% |
| Pavlodar | 5 216 | 1558 | 0.2288 | 0.4484 | 2.91% |
| Soltustik Kazakhstan | 2978 | 959 | 0.313 | 0.6135 | 3.57% |
| Shygys Kazakhstan | 7 358 | 2196 | 0.1902 | 0.3728 | 2.21% |
| **Republic of Kazakhstan** | **116 293** | **27824** | **0.0445** | **0.0872** | **0.63%** |