**On approval of the Methodology for the formation of predictive indicators of the standard of living**

Order of the Chairman of the Statistics Committee of the Ministry of National Economy of the Republic of Kazakhstan dated November 24, 2017 No. 185. Registered with the Ministry of Justice of the Republic of Kazakhstan on December 11, 2017 No. 16074.

      In accordance with subparagraph 5) of Article 12 of the Law of the Republic of Kazakhstan dated March 19, 2010 "On State Statistics" and subparagraph 258) of paragraph 17 of the Regulation on the Ministry of National Economy of the Republic of Kazakhstan, approved by the Decree of the Government of the Republic of Kazakhstan dated September 24, 2014 No. 1011, I ORDER **:**

      1. Approve the attached Methodology for the formation of predictive indicators of the standard of living.

     2. The Department of Labor Statistics and Living Standards, together with the Legal Department of the Committee on Statistics of the Ministry of National Economy of the Republic of Kazakhstan, shall ensure in the manner prescribed by law:

     1) state registration of this order with the Ministry of Justice of the Republic of Kazakhstan;

     2) within ten calendar days from the date of state registration of this order, sending a copy of it in paper and electronic form in Kazakh and Russian to the Republican State Enterprise on the right of economic management "Republican Center for Legal Information" for official publication and inclusion in the Standard Control Bank of regulatory legal acts of the Republic of Kazakhstan;

     3) within ten calendar days after the state registration of this order, send a copy of it for official publication in print periodicals;

      4) placement of this order on the Internet resource of the Committee on Statistics of the Ministry of National Economy of the Republic of Kazakhstan.

      3. To impose control over the execution of this order on the Deputy Chairman of the Committee on Statistics of the Ministry of National Economy of the Republic of Kazakhstan (Orunkhanov K.K.).

     4. This order shall enter into force ten calendar days after the day of its first official publication.

|  |  |
| --- | --- |
| *Chairman of the Statistics Committee of the Ministry of National Economy of the Republic of Kazakhstan* | *N. Aidapkelov* |
|   | Approved by the order of the Chairman of the Committee on Statistics of the Ministry of National Economy of the Republic of Kazakhstandated November 24, 2017 185 |

**Methodology for the formation of predictive indicators of the standard of living**

 **Chapter 1. General provisions**

      1. This Methodology for the formation of predictive indicators of the standard of living (hereinafter - the Methodology) refers to the statistical methodology approved in accordance with the Law of the Republic of Kazakhstan dated March 19, 2010 "On State Statistics".

     2. The purpose of the Methodology is to determine the main aspects of the formation of predictive indicators of the standard of living.

      3. The methodology is intended for use by the Statistics Committee of the Ministry of National Economy of the Republic of Kazakhstan.

      4. To predict the standard of living, the following indicators are used: the share of the population with incomes below the subsistence level and the income of the population used for consumption.

     5. The following definitions are used in the Methodology:

     1) degree of freedom - the difference between the number of observations and the number of estimated parameters;

     2) coefficient of covariance - the average value of the product of deviations of variables from their means and is a measure of the relationship between two variables;

      3) the proportion of the population with incomes below the subsistence minimum - the ratio of the population with incomes below the subsistence minimum to the total population in percentage terms;

      4) trend - a change that determines the general direction of development, the main trend of the time series;

      5) time series (dynamic series) - a sequence of values of an indicator (attribute), ordered in chronological order, in ascending order of the time parameter;

     6) sample variance (variation) - the arithmetic mean of the squares of the deviation of a random variable from the mean value.

**Chapter 2. Prediction of the share of the population with incomes below the subsistence level**

**Paragraph 1. Use of correlation analysis**

      6. When building a living standard forecast, the relationship between economic indicators is determined: the unemployment rate (hereinafter - x ) and the proportion of the population with incomes below the subsistence level (hereinafter - y ). To determine the form of relationship between economic indicators, the method of correlation analysis is used, which calculates the degree of dependence (correlation coefficient) of the two indicators under consideration.

      7. To calculate the correlation coefficient, the least squares method is used, based on minimizing the sum of squares of the residuals between the actual and calculated data.

       The order of calculation of the correlation coefficient



least squares method:

      1) build variation series for each of the compared features, denoting the first and second series of numbers x and y, respectively.

     2) determine for each variational series x and y the average values of M



3) find deviations



each numerical value from the average value of its variation series;

      4) multiply the resulting deviations



, square each deviation



and sum over each row



     5) substitute the obtained values into the formula for calculating the correlation coefficient:



where:



- correlation coefficient;



– deviation of the value of the series x from its average value;



– deviation of the value of the y series from its average value;



- the sum of squared deviations in x;



- the sum of squared deviations for



.

      8. The value of the correlation coefficient reflects the strength of the connection between the variables (x, y). When assessing the strength of the relationship of the correlation coefficients between the variables (x, y), the Chaddock scale is used (with a correlation coefficient of 0.1-0.3 - the relationship of the variables is weak, 0.3-0.5 - moderate, 0.5-0.7 - noticeable, 0.7-0.9 - high, 0.9-0.99 - very high).

      The correlation coefficient takes values from –1 to +1. A negative correlation coefficient indicates the inverse relationship of indicators (x, y). When the correlation coefficient is zero, the variables are not related to each other.

**Section 2. Building a regression model for forecasting**

**living standards indicators**

     9. The regression equation reflects the change in the average value of one variable (y) depending on the second (x).

     10. A linear regression equation is used to determine the dependence of one (dependent) variable y on another or several other variables (independent variables) x with a linear dependence function:

      y=bx+a

where:



- the dependent variable;



– estimates of the parameters of the regression model;



- an independent variable.

      11. Estimated regression equation (built on sample values of economic indicators):

       y=bx+a+e

(3)

where:

e - random error (deviation);

      a and b are estimates of the parameters of the regression model.

     12. Empirical regression coefficients ( b ) are used to build the regression equation :



where:



– regression coefficient;



- the average value of xy;



- the average value of x;



- the average value of y;



- the sample variance of x.

     Empirical regression coefficients are built on sample values of economic indicators (sample population) and determine the dependence of the average values of the result attribute (for the sample population) on the average values of the factor factor (for the general population).

     The estimation of the parameters of the regression model is calculated by the following formula:



where:



– estimation of the parameters of the regression model;



- the average value of y;



– regression coefficient;



- the average value of x.

      13. After determining the coefficients of the regression equation, an auxiliary coefficient is determined



, showing the growth trend of the variational series of the variable x:



where:



- auxiliary coefficient;



– purchasing power for the n -th year;



– purchasing power for n -1 years;



– purchasing power for n -2 years;



- the last year from which the forecast for the future period is made.

     The auxiliary factor l is used to calculate the prediction of the x variable.

     Purchasing power is determined by the quantity of goods purchased for the amount of the average per capita monetary income of the surveyed households at average purchase prices. Purchasing power is also expressed through the ratio of income to the subsistence level.

      Purchasing power for the next year after the nth is calculated by the following formula:



where:



– purchasing power for n +1 years;



- auxiliary coefficient;



– purchasing power for the n -th year.

       By calculating the purchasing power



for the next year after the n -th, a forecast is also made on the variable y (the proportion of the population with incomes below the subsistence level):



where:



- the share of the population with incomes below the subsistence level for n +1 years;



– estimation of the parameters of the regression model;



– regression coefficient;



– purchasing power for n +1 years.

     14. To determine the accuracy of the regression equation, the coefficient of determination is calculated according to the following formula:



where:



– coefficient of determination;



- the sample variance of x values;



- the sample variance of y values.

      The coefficient of determination shows how much variation is due to differences between features and other factors.

      Variation is denoted by the difference in the values of a particular attribute in individual units included in this population.

     The value of R 2 ≥0.5 (Fisher's criterion) shows that the variation is caused by the influence of other factors. Fisher's criterion ( F -statistics) is a parametric criterion and is used to evaluate the quality of the regression model as a whole and by parameters by comparing the obtained criterion value and the table value of the Fisher criterion.

     15. To assess the quality of the regression parameters, the following hypotheses are put forward:

     1) with the null (main) ( H 0 ) hypothesis r xy = 0, there is no linear relationship between the variables;

     2) with an alternative ( H 1 ) hypothesis r xy ≠ 0 – there is a linear relationship between the variables.

     In regression analysis, the null (main) hypothesis H 0 is checked about the absence of a relationship between variables. When the null hypothesis is rejected, the alternative hypothesis about the existence of a relationship between the variables is accepted. The alternative hypothesis is the opposite of the null hypothesis.

      The significance level determines the probability of making an erroneous decision when testing the null hypothesis, which denies the differences between the compared values. The significance level is chosen to be 0.05 or 0.01.

      According to a given level of significance a, the null hypothesis is checked about the equality of the correlation coefficient to zero, with an alternative hypothesis H 1 ≠ 0, the value of the random error is calculated using the following formula:



where:



– the observed value of the criterion (the value of the random error);



– correlation coefficient between variables x and y;



- the number of degrees of freedom;



- the square of the correlation coefficient between x and y.

      According to the table of critical distribution points t -Student's criterion for a given level of significance a and the number of degrees of freedom k = n -2 is the critical point t crit.

      Student's t -test is used to test the significance of each factor in the regression model. The null hypothesis assumes that the means are equal (the negation of this assumption is called the shift hypothesis).

      16. The interval estimate (confidence interval) for the correlation coefficient is determined by the following formula:



where:



– correlation coefficient between variables x and y;



- the critical value at a given level of significance and the number of degrees of freedom;



- the square of the correlation coefficient between variables x and y;



- degree of freedom;



- the number of observations.

     The confidence interval shows in what range the results of sample observations will be located and the permissible deviation of the observed values from the true ones.

      17. Unexplained (residual) variance shows the variation of the result under the influence of factors not taken into account by the regression and is calculated using the following formula:



where:



- unexplained variance (a measure of the spread of the dependent variable around the regression line);



- the sum of the squares of the difference



;



- the number of observations;

      m is the number of independent variables;

      n–m–1 is the degree of freedom;



- the value of the sample value y for each specific observation i;



- the value of the sample quantity y for observing x.

     18. The standard error of the regression is calculated using the following formula:



     where:

      S - the standard error of the estimate (standard error of the regression);



– unexplained variance (a measure of the spread of the dependent variable around the regression line).

       19. Forecast error for the equation



- calculated using the following formula:



     where:

       e - a random error for estimating y;

      n - the number of observations;



      20. Forecast error for the equation y = bx + a + e is calculated by the following formula:



     where:

       e - an independent random error;

      n - the number of observations;



     21. Hypotheses are tested with respect to the coefficients of the linear regression equation. The assessment of the significance of linear regression is carried out according to the following algorithm: a null hypothesis H0 is put forward about the insignificance of the equation R2 = 0 in terms of the significance level a.

     In regression analysis, the significance of linear regression models is assessed using the Fisher F -test. To assess the significance of linear regression models, a comparison is made between the obtained actual value F fact and the tabular value F table F - Fisher's criterion.

     The actual value of the Fisher F -criterion is determined by the following formula:



where:



- the actual value of F - Fisher's criterion;



– coefficient of determination;

      n – 2 – degree of freedom;

      n = the number of observations.

     The actual value F fact of the Fisher criterion is compared with the tabular value F table according to the mathematical table of the Fisher criterion.

      F table - the maximum value of the Fisher criterion under the influence of random factors at the current degrees of freedom and significance level a. When F table > F fact, the regression equation is considered insignificant.

      22. To analyze the presence of autocorrelation of residuals in regression models, the Durbin-Watson test is used.

      The Durbin-Watson test is used to compare the actual value of the Durbin-Watson test with the theoretical values for a given number of observations n, and a significance level a.

      Autocorrelation of residuals of the regression model (random errors of the regression model) is the correlation between the present and past values of the residuals.

      The Durbin-Watson test is calculated using the following formula:



where:



– Durbin-Watson criterion;



- the sum of the squares of the difference



;



- sum of squares of the remainder



;



- the remainder in the i -th observation;



- the square of the residual in the i -th observation;



– residual in ( i –1) –th observation.

      For a reliable conclusion, the values according to the mathematical table of the Durbin-Watson criterion and the following rules are used:



**Chapter 3 Determination of the predicted value of the indicator of household income used for consumption**

      23. To determine the forecast value of the indicator, household income used for consumption, the exponential smoothing method is used, taking into account the trend and seasonality according to the Holt-Winters method.

      The Holt-Winters method takes into account seasonality and trend in the forecast of indicators for a period of less than a year (monthly, quarterly) and is used to reduce forecasting errors in case of seasonal fluctuations. To account for seasonal variations, an additional equation is applied, and the Holt-Winters method is completely described by four equations:

      1) formula of exponentially smoothed series:



where:



– exponential smoothed series at the current time t;



– constant value of smoothing;



– seasonality in the initial data;



– the duration of the seasonal fluctuation period;



- the exponential smoothed series at the previous moment of time ( t –1);



– trend at the previous time point ( t –1).

     Equation (18) corrects the smoothed series.

      2) trend assessment



,

where:



– trend (main trend of the series);



– constant value of smoothing for the trend;



– exponential smoothed series at the current time t;



- the exponential smoothed series at the previous moment of time ( t –1);



– trend at the previous time point ( t –1).

       3) assessment of seasonality



where:



- seasonal fluctuation of the seasonality series at the current time t;



– constant value of smoothing for seasonality;



– data value at the current time t;



– exponential smoothed series at the current time t;



– seasonal fluctuation of the seasonality series at the moment of time ( t – s ).

       4) forecast for p periods ahead



,

where:



– values of forecast data for p future periods;



– exponential smoothed series at the current time t;

      p – periods;



– trend at the current time t;



– seasonal fluctuation of the seasonality series at the moment of time ( t – s + p ).

     The initial value of the smoothed series is taken equal to the first observation, and the trend is taken with a value equal to zero. Seasonality estimates are set equal to one.

      After determining the initial value of the smoothed series, the value of the smoothed series for the next year is calculated - ( t +1) and the trend value of the corresponding period is determined ( t +1).

      After the trend, a seasonality estimate is calculated. The seasonality estimate of the corresponding period is calculated starting from the period ( t + s ). The first estimates of seasonality s are used equal to one. After calculating the number of estimates n, forecast data is calculated



for future periods p (the number of periods p coincides with the duration of the seasonal fluctuation period s ) according to formula (21).

       In formula (21), the parameters



And



are used as constants and do not change when they are calculated. The variable data are the parameter p (periods), starting from one, and seasonality estimates



, starting from the period ( t - s + p ).

      As a result, periods p of forecast data are formed for a certain indicator of the standard of living (for example, quarterly income indicators of the population used for consumption).

      24. This calculation model for the formation of forecast indicators is similarly applied to other indicators of the standard of living.

© 2012. RSE on REM "Institute of Legislation and Legal Information of the Republic of Kazakhstan" of the Ministry of Justice of the Republic of Kazakhstan