

Об утверждении Методики по формированию прогнозных показателей уровня жизни

Приказ Председателя Комитета по статистике Министерства национальной экономики Республики Казахстан от 24 ноября 2017 года № 185. Зарегистрирован в Министерстве юстиции Республики Казахстан 11 декабря 2017 года № 16074.

В соответствии с подпунктом 5) статьи 12 Закона Республики Казахстан от 19 марта 2010 года "О государственной статистике" и подпунктом 258) пункта 17 Положения о Министерстве национальной экономики Республики Казахстан, утвержденного постановлением Правительства Республики Казахстан от 24 сентября 2014 года № 1011

, ПРИКАЗЫВАЮ:

1. Утвердить прилагаемую Методику по формированию прогнозных показателей уровня жизни.

2. Управлению статистики труда и уровня жизни совместно с Юридическим управлением Комитета по статистике Министерства национальной экономики Республики Казахстан обеспечить в установленном законодательством порядке:

1) государственную регистрацию настоящего приказа в Министерстве юстиции Республики Казахстан;

2) в течение десяти календарных дней со дня государственной регистрации настоящего приказа направление его копии в бумажном и электронном виде на казахском и русском языках в Республиканское государственное предприятие на праве хозяйственного ведения "Республиканский центр правовой информации" для официального опубликования и включения в Эталонный контрольный банк нормативных правовых актов Республики Казахстан;

3) в течение десяти календарных дней после государственной регистрации настоящего приказа направление его копии на официальное опубликование в периодические печатные издания;

4) размещение настоящего приказа на интернет-ресурсе Комитета по статистике Министерства национальной экономики Республики Казахстан.

3. Контроль за исполнением настоящего приказа возложить на заместителя Председателя Комитета по статистике Министерства национальной экономики Республики Казахстан (Орунханов К.К.).

4. Настоящий приказ вводится в действие по истечении десяти календарных дней после дня его первого официального опубликования.

*Председатель Комитета
по статистике Министерства
национальной экономики
Республики Казахстан*

Н. Айдапкелов

Утверждена приказом

Методика по формированию прогнозных показателей уровня жизни

Глава 1. Общие положения

1. Настоящая Методика по формированию прогнозных показателей уровня жизни (далее – Методика) относится к статистической методологии, утверждаемой в соответствии с Законом Республики Казахстан от 19 марта 2010 года "О государственной статистике".

2. Целью Методики является определение основных аспектов формирования прогнозных показателей уровня жизни.

3. Методика предназначена для применения Комитетом по статистике Министерства национальной экономики Республики Казахстан.

4. Для прогнозирования уровня жизни используются следующие показатели: доля населения, имеющего доходы ниже величины прожиточного минимума и доходы населения, использованные на потребление.

5. В Методике используются следующие определения:

- 1) степень свободы – разность числа наблюдений и числа оцененных параметров;
- 2) коэффициент ковариации – средняя величина произведения отклонений переменных от своих средних и является мерой взаимосвязи между двумя переменными;
- 3) доля населения, имеющего доходы ниже величины прожиточного минимума – отношение численности населения, имеющего доходы ниже величины прожиточного минимума к общей численности населения в процентном измерении;
- 4) тренд – изменение, определяющее общее направление развития, основную тенденцию временного ряда;
- 5) временной ряд (динамический ряд) – последовательность значений показателя (признака), упорядоченная в хронологическом порядке, в порядке возрастания временного параметра;
- 6) выборочная дисперсия (вариация) – среднее арифметическое квадратов отклонения случайной величины от среднего значения.

Глава 2. Прогнозирование показателя доли населения, имеющего доходы ниже величины прожиточного минимума

Параграф 1. Использование корреляционного анализа

6. При построении прогноза уровня жизни определяется взаимосвязь между экономическими показателями: уровень безработицы (далее – x) и доля населения, имеющего доходы ниже величины прожиточного минимума (далее – y). Для определения формы связи между экономическими показателями применяется метод корреляционного анализа, рассчитывающий степень зависимости (коэффициент корреляции) двух рассматриваемых показателей.

7. Для расчета коэффициента корреляции используется метод наименьших квадратов, основанный на минимизации суммы квадратов остатков между фактическими и расчетными данными.

Порядок вычисления коэффициента корреляции

(r_{xy})

методом наименьших квадратов:

1) построить вариационные ряды для каждого из сопоставляемых признаков, обозначив первый и второй ряд чисел соответственно x и y .

2) определить для каждого вариационного ряда x и y средние значения M (M_x и M_y);

3) найти отклонения

$(d_x$ и $d_y)$

каждого числового значения от среднего значения своего вариационного ряда;

4) полученные отклонения перемножить

$(d_x$ и $d_y)$

, каждое отклонение возвести в квадрат

$(d_x^2$ и $d_y^2)$

и суммировать по каждому ряду

$(\sum d_x^2$ и $\sum d_y^2)$;

5) подставить полученные значения в формулу расчета коэффициента корреляции:

$$r_{xy} = \frac{\sum(d_x \times d_y)}{\sqrt{(\sum d_x^2 \times \sum d_y^2)}}, \quad (1)$$

где:

r_{xy}

– коэффициент корреляции;

d_x

– отклонение значения ряда x от своего среднего значения;

d_y

– отклонение значения ряда y от своего среднего значения;

$\sum d_x^2$

– сумма квадратов отклонений по x ;

$\sum d_y^2$

– сумма квадратов отклонений по

y

.

8. Величина коэффициента корреляции отражает силы связи между переменными (x , y). При оценке силы связи коэффициентов корреляции между переменными (x , y) используется шкала Чеддока (при коэффициенте корреляции равном 0,1-0,3 – взаимосвязь переменных слабая, 0,3-0,5 – умеренная, 0,5-0,7 – заметная, 0,7-0,9 – высокая, 0,9-0,99 – весьма высокая).

Коэффициент корреляции принимает значения от -1 до $+1$. Отрицательный коэффициент корреляции указывает на обратную связь показателей (x , y). При нулевой величине коэффициента корреляции переменные не имеют связи между собой.

Параграф 2. Построение регрессионной модели для прогнозирования показателей уровня жизни

9. Уравнение регрессии отражает изменение средней величины одной переменной (y) в зависимости от второй (x).

10. Линейное уравнение регрессии используется для определения зависимости одной (зависимой) переменной y от другой или нескольких других переменных (независимых переменных) x с линейной функцией зависимости:

$$y = bx + a,$$

где:

y

– зависимая переменная;

a и b

– оценки параметров регрессионной модели;

x

– независимая переменная.

11. Оценочное уравнение регрессии (построенное по выборочным значениям экономических показателей):

$$y = bx + a + e, \quad (3)$$

где:

e – случайная ошибка (отклонение);

a и b – оценки параметров регрессионной модели.

12. Для построения уравнения регрессии используются эмпирические коэффициенты регрессии (b):

$$b = \frac{M_{xy} - M_x \times M_y}{\sigma_x^2}, \quad (4)$$

где:

b

– коэффициент регрессии;

M_{xy}

– среднее значение величины xy ;

M_x

– среднее значение величины x ;

M_y

– среднее значение величины y ;

$$\sigma_x^2 = \frac{1}{n} \sum (x - M_x)^2.$$

– выборочная дисперсия переменной x .

Эмпирические коэффициенты регрессии строятся по выборочным значениям экономических показателей (выборочная совокупность) и определяют зависимость средних значений признака-результата (по выборочной совокупности) от средних значений признака-фактора (по генеральной совокупности).

Оценка параметров регрессионной модели рассчитывается по следующей формуле:

$$a = M_y - b \times M_x, \quad (5)$$

где:

a

– оценка параметров регрессионной модели;

M_y

– среднее значение величины y ;

b

– коэффициент регрессии;

M_x

– среднее значение величины x .

13. После определения коэффициентов уравнения регрессии определяется вспомогательный коэффициент

\bar{l}

, показывающий тенденцию роста вариационного ряда переменной x :

$$\bar{l} = \frac{B_n/B_{n-1} + B_{n-1}/B_{n-2}}{2}, \quad (6)$$

где:

\bar{l}

– вспомогательный коэффициент;

B_n

– покупательная способность за n -й год;

B_{n-1}

– покупательная способность за $n-1$ год;

B_{n-2}

– покупательная способность за $n-2$ год;

n

– последний год, от которого производится прогноз на будущий период.

Вспомогательный коэффициент l используется для расчета прогноза переменной x .

Покупательная способность определяется количеством товаров, приобретаемых на сумму среднедушевого денежного дохода обследуемых домашних хозяйств по средним ценам покупки. Покупательная способность также выражается через соотношение доходов с прожиточным минимумом.

Покупательная способность на следующий год после n -го рассчитывается по следующей формуле:

$$B_{n+1} = l \times B_n, \quad (7)$$

где:

B_{n+1}

– покупательная способность за $n+1$ год;

l

– вспомогательный коэффициент;

B_n

– покупательная способность за n -й год.

Вычислив покупательную способность

B_{n+1}

на следующий год после n -го, производится прогноз и на переменной y (доля населения, имеющего доходы ниже прожиточного минимума):

$$A_{n+1} = a + b \times B_{n+1}, \quad (8)$$

где:

A_{n+1}

– доля населения, имеющего доходы ниже величины прожиточного минимума за $n+1$ год;

a

– оценка параметров регрессионной модели;

b

– коэффициент регрессии;

B_{n+1}

– покупательная способность за $n+1$ год.

14. Для определения точности уравнения регрессии производится расчет коэффициента детерминации по следующей формуле:

$$R^2 = 1 - \frac{\sigma_x^2}{\sigma_y^2}, \quad (9)$$

где:

R^2

– коэффициент детерминации;

$$\sigma_x^2 = \frac{1}{n} \sum (x - M_x)^2.$$

– выборочная дисперсия значений x ;

$$\sigma_y^2 = \frac{1}{n} \sum (y - M_y)^2.$$

– выборочная дисперсия значений y .

Коэффициент детерминации показывает насколько вариация обусловлена различиями между признаками и другими факторами.

Вариация обозначается различием в значениях того или иного признака у отдельных единиц, входящих в данную совокупность.

Значение $R^2 \geq 0,5$ (критерий Фишера) показывает об обусловленности вариации влиянием других факторов. Критерий Фишера (F-статистика) является параметрическим критерием и используется для оценивания качества регрессионной модели в целом и по параметрам путем сравнения полученного значения критерия и табличного значения критерия Фишера.

15. Для оценки качества параметров регрессии выдвигаются следующие гипотезы:

1) при нулевой (основной) (H_0) гипотезе $r_{xy} = 0$ – нет линейной взаимосвязи между переменными;

2) при альтернативной (H_1) гипотезе $r_{xy} \neq 0$ – есть линейная взаимосвязь между переменными.

В регрессионном анализе проверяется нулевая (основная) гипотеза H_0 об отсутствии взаимосвязи между переменными. При отклонении нулевой гипотезы, принимается альтернативная гипотеза о существовании взаимосвязи между переменными. Альтернативная гипотеза – противоположность нулевой гипотезе.

Уровень значимости определяет вероятность принятия ошибочного решения при испытании нулевой гипотезы, отрицающей различия сравниваемых величин. Уровень значимости выбирается равным 0,05 или 0,01.

По заданному уровню значимости α проверяется нулевая гипотеза о равенстве коэффициента корреляции нулю, при альтернативной гипотезе $H_1 \neq 0$ вычисляется величина случайной ошибки по следующей формуле:

$$t_{nabi} = r_{xy} \frac{\sqrt{n-2}}{\sqrt{1-r_{xy}^2}}, \quad (10)$$

где:

t_{nabi}

– наблюдаемое значение критерия (величина случайной ошибки);

r_{xy}

– коэффициент корреляции между переменными x и y ;

$n - 2$.

– число степеней свободы;

r_{xy}^2 .

– квадрат коэффициента корреляции между x и y .

Согласно таблице критических точек распределения t -критерий Стьюдента по заданному уровню значимости α и числу степеней свободы $k=n-2$ находится критическая точка $t_{\text{крит}}$.

t -критерий Стьюдента используется для проверки значимости каждого фактора регрессионной модели. При нулевой гипотезе предполагается, что средние равны (отрицание этого предположения называют гипотезой сдвига).

16. Интервальная оценка (доверительный интервал) для коэффициента корреляции определяется по следующей формуле:

$$\left[(r_{xy} - t_{\text{крит}}) \sqrt{\frac{1-r_{xy}^2}{n-2}}; (r_{xy} + t_{\text{крит}}) \sqrt{\frac{1-r_{xy}^2}{n-2}} \right], \quad (11)$$

где:

r_{xy}

– коэффициент корреляции между переменными x и y ;

$t_{\text{крит}}$

– критическое значение при заданном уровне значимости и числе степеней свободы;

r_{xy}^2

– квадрат коэффициента корреляции между переменными x и y ;

$n - 2$

– степень свободы;

n

– число наблюдений.

Доверительный интервал показывает в каком диапазоне расположатся результаты выборочных наблюдений и допустимое отклонение наблюдаемых значений от истинных.

17. Необъясненная (остаточная) дисперсия показывает вариацию результата под влиянием факторов, неучтенных регрессией и рассчитывается по следующей формуле:

$$S^2 = \frac{\sum (y_i - y_x)^2}{n - m - 1}, \quad (12)$$

где:

S^2

– необъясненная дисперсия (мера разброса зависимой переменной вокруг линии регрессии);

$\sum (y_i - y_x)^2$

– сумма квадратов разницы величин

y_i и y_x ;

;

n

– число наблюдений;

m – число независимых переменных;

$n-m-1$ – степень свободы;

y_i

– значение выборочной величины y для каждого конкретного наблюдения i ;

y_x

– значение выборочной величины y для наблюдения x .

18. Стандартная ошибка регрессии рассчитывается по следующей формуле:

$$S = \sqrt{S^2}, \quad (13)$$

где:

S – стандартная ошибка оценки (стандартная ошибка регрессии);

S^2

– необъясненная дисперсия (мера разброса зависимой переменной вокруг линии регрессии).

19. Ошибка прогноза для уравнения

$$y = bx + a$$

вычисляется по следующей формуле:

$$\varepsilon = t_{\text{крит}} \times \sqrt{S^2 \left(\frac{1}{n} + \frac{(x_p - \bar{x})^2}{n \times \sigma_x^2} \right)}, \quad (14)$$

где:

e – случайная ошибка для оценки y ;

n – число наблюдений;

$$\sigma_x^2 = \frac{1}{n} \sum (x - \bar{x})^2 - \text{дисперсия } x.$$

20. Ошибка прогноза для уравнения $y = bx + a + e$ вычисляется по следующей формуле:

$$\varepsilon = t_{\text{крит}} \times \sqrt{S^2 \left(1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{n \times \sigma_x^2}\right)}, \quad (15)$$

где:

ε – независимая случайная ошибка;

n – число наблюдений;

$$\sigma_x^2 = \frac{1}{n} \sum (x - \bar{x})^2 - \text{дисперсия } x.$$

21. Проверка гипотез производится относительно коэффициентов линейного уравнения регрессии. Оценка значимости линейной регрессии производится по следующему алгоритму: выдвигается нулевая гипотеза H_0 о незначимости уравнения $R^2 = 0$ по уровню значимости α .

В регрессионном анализе с помощью F–критерия Фишера оценивается значимость линейных регрессионных моделей. Для оценки значимости линейных регрессионных моделей выполняется сравнение полученного фактического значения $F_{\text{факт}}$ и табличного значения $F_{\text{табл}}$ F–критерия Фишера.

Фактическое значение F–критерия Фишера определяется по следующей формуле:

$$F_{\text{факт}} = \frac{R^2(n-2)}{1-R^2}, \quad (16)$$

где:

$F_{\text{факт}}$

– фактическое значение F – критерия Фишера;

R^2

– коэффициент детерминации;

$n - 2$ – степень свободы;

n – число наблюдений.

Фактическое значение $F_{\text{факт}}$ критерия Фишера сравнивается с табличным значением $F_{\text{табл}}$ по математической таблице критерия Фишера.

$F_{\text{табл}}$ – максимальное значение критерия Фишера под влиянием случайных факторов при текущих степенях свободы и уровне значимости α . При $F_{\text{табл}} > F_{\text{факт}}$ уравнение регрессии считается незначимым.

22. Для анализа наличия автокорреляции остатков в регрессионных моделях используется критерий Дарбина-Уотсона.

Критерий Дарбина-Уотсона используется для сравнения фактической величины критерия Дарбина-Уотсона с теоретическими значениями для заданного числа наблюдений n , и уровня значимости α .

Автокорреляцией остатков модели регрессии (случайных ошибок регрессионной модели) называется корреляционная зависимость между настоящими и прошлыми значениями остатков.

Критерий Дарбина-Уотсона рассчитывается по следующей формуле:

$$DW = \frac{\sum(e_i - e_{i-1})^2}{\sum e_i^2}, \quad (17)$$

где:

DW

– критерий Дарбина-Уотсона;

$\sum(e_i - e_{i-1})^2$

– сумма квадратов разницы величин

e_i и e_{i-1} ;

;

$\sum e_i^2$

– сумма квадратов остатка

e_i

;

e_i

– остаток в i -м наблюдении;

e_i^2

– квадрат остатка в i -м наблюдении;

e_{i-1}

– остаток в $(i-1)$ –м наблюдении.

Для надежного вывода используются значения по математической таблице критерия Дарбина-Уотсона и следующие правила:

- 1) $0 < DW < d_L$ – есть положительная автокорреляция;
- 2) $d_L < DW < d_U$ – зона неопределенности;
- 3) $d_U < DW < 4 - d_U$ – автокорреляции нет;
- 4) $4 - d_U < DW < 4 - d_L$ – зона неопределенности;
- 5) $4 - d_L < DW < 4$ – есть отрицательная автокорреляция.

Глава 3. Определение прогнозного значения показателя доходы населения, использованные на потребление

23. Для определения прогнозного значения показателя доходы населения, использованные на потребление, используется метод экспоненциального сглаживания с учетом тренда и сезонности по методу Хольта-Винтерса.

Метод Хольта-Винтерса учитывает сезонность и тренд в прогнозе индикаторов за период меньше года (месячной, квартальной периодичностью) и используется для уменьшения ошибок прогнозирования при сезонных колебаниях. Для учета сезонных вариаций применяется дополнительное уравнение, и полностью метод Хольта-Винтерса описывается четырьмя уравнениями:

- 1) формула экспоненциально сглаженного ряда:

$$L_t = k \times \frac{Y_t}{S_{t-s}} + (1 - k) \times (L_{t-1} + T_{t-1}), \quad (18)$$

где:

L_t

– экспоненциальный сглаженный ряд в текущем моменте времени t ;

k

– постоянная величина сглаживания;

$\frac{Y_t}{S_{t-s}}$

– сезонность в исходных данных;

s

– длительность периода сезонного колебания;

L_{t-1}

– экспоненциальный сглаженный ряд в предыдущем моменте времени (t-1);

T_{t-1}

– тренд в предыдущем моменте времени (t-1).

Уравнение (18) корректирует сглаженные ряды.

2) оценка тренда

$$T_t = b \times (L_t - L_{t-1}) \times (1 - b) \times T_{t-1}, \quad (19)$$

,

где:

T_t

– тренд (основная тенденция ряда);

b

– постоянная величина сглаживания для тренда;

L_t

– экспоненциальный сглаженный ряд в текущем моменте времени t;

L_{t-1}

– экспоненциальный сглаженный ряд в предыдущем моменте времени (t-1);

T_{t-1}

– тренд в предыдущем моменте времени (t-1).

3) оценка сезонности

$$S_t = q \times \frac{Y_t}{L_t} + (1 - q) \times S_{t-s}, \quad (20)$$

где:

S_t

– сезонное колебание ряда сезонности в текущем моменте времени t;

q

– постоянная величина сглаживания для сезонности;

Y_t

– значение данных в текущем моменте времени t ;

L_t

– экспоненциальный сглаженный ряд в текущем моменте времени t ;

S_{t-s}

– сезонное колебание ряда сезонности в моменте времени $(t-s)$.

4) прогноз на p периодов вперед

$$\hat{Y}_{t+p} = (L_t + p \times T_t) \times S_{t-s+p}, \quad (21)$$

,

где:

\hat{Y}_{t+p}

– значения прогнозных данных на p будущих периодов;

L_t

– экспоненциальный сглаженный ряд в текущем моменте времени t ;

p – периоды;

T_t

– тренд в текущем моменте времени t ;

S_{t-s+p} .

– сезонное колебание ряда сезонности в моменте времени $(t-s+p)$.

Начальное значение сглаженного ряда принимается равным первому наблюдению, а тренд берется со значением равным нулю. Оценки сезонности устанавливаются равными единице.

После определения начального значения сглаженного ряда, рассчитывается значение сглаженного ряда на следующий год – $(t+1)$ и определяется значение тренда соответствующего периода $(t+1)$.

После тренда вычисляется оценка сезонности. Оценка сезонности соответствующего периода рассчитывается начиная с периода $(t+s)$. Первые оценки сезонности s используют равными единице. После расчета количества оценок n , рассчитываются прогнозные данные

\hat{Y}_{t+p}

на будущие периоды p (количество периодов p совпадает с длительностью периода сезонного колебания s) по формуле (21).

В формуле (21) параметры

L_t

и

T_t

используются как константы и не изменяются при их подсчете. Изменяемыми данными выступают параметр p (периоды), начиная с единицы, и оценки сезонности

S_{t-s+p} .

, начиная с периода $(t-s+p)$.

В результате формируются периоды p прогнозных данных по определенному показателю уровня жизни (например, квартальные показатели доходов населения, использованные на потребление).

24. Данная модель расчетов по формированию прогнозных показателей аналогично применяется относительно других показателей уровня жизни.