

УДК 519.2

МАТЕМАТИЧЕСКИЕ МОДЕЛИ ПОВЫШЕНИЯ ТОЧНОСТИ РАСЧЕТА ПОКАЗАТЕЛЕЙ УРОВНЯ ЖИЗНИ НАСЕЛЕНИЯ

В. К. Клочко, д.т.н.; профессор кафедры АИТУ РГРТУ, Рязань, Россия;
orcid.org/0000-0003-2550-999X, e-mail: klochkovk@mail.ru

В. П. Кузнецов, к.т.н.; доцент кафедры АИТУ РГРТУ, Рязань, Россия;
orcid.org/0000-0001-6189-9011, e-mail: kuznetsovaitu@yandex.ru

Даны математические модели для расчета показателей уровня жизни населения регионов РФ по данным Росстата, публикуемым в виде гистограмм дохода, полученных путем опроса населения. Известный подход к расчету показателей основан на аппроксимации гистограмм логнормальным распределением. Предложенный подход отличается использованием смеси распределений и дополнительным построением трендов для оценок параметров логнормальных распределений в последовательности нескольких лет. Цель работы – разработка математических моделей дохода населения регионов РФ, отличающихся учетом данных опроса населения за несколько предыдущих лет, позволяющих повысить точность расчета показателей уровня жизни населения регионов РФ. В работе представлены конкретные алгоритмы расчета показателей уровня жизни на основе предложенного подхода. Обсуждается возможность вычисления показателей в регионах РФ с учетом их корреляции с показателями РФ. Приводятся примеры расчетов, подтверждающие работоспособность предложенных алгоритмов. Алгоритмы могут найти применение в практике Росстата.

Ключевые слова: алгоритмы, прогнозирование, доходы населения, математические модели.

DOI: 10.21667/1995-4565-2020-73-105-115

Введение

Одним из важнейших факторов принятия политических и экономических решений в РФ является анализ доходов населения регионов и РФ в целом по данным Росстата. Исходным материалом для анализа служат гистограммы распределения дохода на душу населения в рублях по всем регионам РФ, полученные по результатам опроса населения первичной обработкой алгоритмами, принятыми в Росстате. Эти данные публикуются ежегодно и являются предметом исследований, например [1, 2]. Известно из мировой практики, что доход населения, измеренный в денежных единицах, подчинен логнормальному закону распределения [1, 3]. Этот закон берется за основу для аналитического описания гистограмм, подверженных искажениям, для вычисления числовых характеристик, более точных по сравнению с выборочными характеристиками, найденными непосредственно по гистограммам. Такими характеристиками являются мода, медиана, среднее и СКО, а также коэффициенты фондов и уровень бедности, вычисляемый в процентах по квантилям распределения – минимальным размерам оплаты труда в регионах.

Так как данные опроса подвержены большим искажениям, то актуальна задача повышения точности расчета показателей уровня жизни. Один из путей повышения точности – описание дохода смесью логнормальных распределений [3]. Однако такой путь лишь частично решает проблему, так как любая аппроксимация в определенной степени повторяет ошибки гистограмм. Предлагается подход, который дополняет первый путь анализом временных последовательностей параметров распределения дохода, найденных на основе аналитических описаний гистограмм в разные годы, с последующим их прогнозированием и установлением связи с характеристиками по РФ в целом, как менее подверженными искажениям.

Цель работы – разработка математических моделей дохода населения регионов РФ, отличающихся учетом данных опроса населения за несколько предыдущих лет, позволяющих повысить точность определения показателей уровня жизни населения регионов РФ.

Аппроксимация гистограмм логнормальным распределением

Плотность логнормального распределения $g_i(y)$ дохода y i -го региона ($i = \overline{2, M}$, M – число субъектов или регионов РФ, $i = 1$ соответствует РФ) зависит от двух параметров – m_i и σ_i [1, 3]:

$$g_i(y) = \frac{1}{\sqrt{2\pi}\sigma_i} \frac{1}{y} \exp\left[-\frac{(\ln y - m_i)^2}{2\sigma_i^2}\right], \quad y > 0. \quad (1)$$

Параметры m_i и σ_i , с одной стороны, выражаются через моду $y_{mo}(i)$ распределения дохода y и макроэкономическое среднее $y_{MAK}(i)$ дохода (если оно задано) [3] как

$$\sigma_i = \sqrt{\frac{2}{3}[\ln(y_{MAK}(i)/y_{mo}(i))]}, \quad m_i = \ln y_{MAK}(i) - \sigma_i^2/2. \quad (2)$$

С другой стороны, параметры m_i и σ_i выражаются через моду $y_{mo}(i)$ и средний доход $y_{cp}(i)$:

$$\sigma_i = \sqrt{\frac{2}{3}(\ln[y_{cp}(i)/y_{mo}(i)])}, \quad m_i = \ln y_{cp}(i) - \sigma_i^2/2. \quad (3)$$

Также параметры m_i и σ_i выражаются через средний доход $y_{cp}(i)$ и СКО дохода $\sigma_y(i)$:

$$\sigma_i = \sqrt{\ln[1 + \sigma_y^2(i)/y_{cp}^2(i)]}, \quad m_i = \ln y_{cp}(i) - \sigma_i^2/2. \quad (4)$$

Таким образом, параметры логнормального распределения определяются одним из трех способов (2), (3), (4). При первом способе (2) макроэкономический доход $y_{MAK}(i)$ каждого i -го региона ($i > 1$) и РФ ($i = 1$) задается, а мода $y_{mo}(i)$ заменяется оценкой $\hat{y}_{mo}(i)$, найденной по гистограмме. Однако нахождение моды по гистограмме сопровождается наибольшими ошибками по сравнению с оценками среднего $\hat{y}_{cp}(i)$ и СКО $\hat{\sigma}_y(i)$. Поэтому при нахождении моды по первому способу разумно идти другим путем. Вначале по гистограмме распределения дохода i -го региона без учета макроэкономического среднего $y_{MAK}(i)$ находятся выборочные среднее $y_{cp}^*(i)$ и СКО $\sigma_y^*(i)$:

$$y_{cp}^*(i) = \frac{1}{n_i} \sum_{j=1}^k n_j(i) y_j, \quad \sigma_y^*(i) = \sqrt{\frac{1}{n_i} \sum_{j=1}^k n_j(i) y_j^2 - (y_{cp}^*)^2}, \quad n_i = \sum_{j=1}^k n_j(i), \quad (5)$$

где k – число разрядов (интервалов) гистограммы; y_j – середина j -го интервала, например: $y_j = (j - 0,5)h_y$, h_y – ширина интервала (шаг); $n_j(i)$ – частота j -го разряда (количество обследованных). На основе выборочных оценок (5) вычисляются оценки параметров логнормального распределения (1) в соответствии с формулами (4):

$$\sigma_i = \sqrt{\ln[1 + (\sigma_y^*(i))^2 / (y_{cp}^*(i))^2]}, \quad m_i = \ln y_{cp}^*(i) - \sigma_i^2/2. \quad (6)$$

Затем уточняются параметры m_i и σ_i методом наименьших квадратов численно минимизацией квадратичного показателя

$$J(m_i, \sigma_i) = \sum_{j=1}^k (\hat{g}_i(y_j) - g_i(y_j, m_i, \sigma_i))^2 \rightarrow \min_{m_i, \sigma_i}, \quad (7)$$

где $\hat{g}_i(y_j)$ – эмпирическое распределение дохода:

$$\hat{g}_i(y_j) = (n_j(i)/n_i)/h_y, \quad j = \overline{1, k}.$$

По найденным оценкам параметров $\hat{m}_i$ и $\hat{\sigma}_i$ находится оценка моды $\hat{y}_{mo}(i)$ по формуле

$$\hat{y}_{mo}(i) = \exp(\hat{m}_i - \hat{\sigma}_i^2)$$

и вычисляются параметры m_i и σ_i с учетом макроэкономического среднего $y_{MAK}(i)$ по формулам (2). Обоснованием такого подхода является то, что мода не зависит от макроэкономического среднего.

Также аналитически по логнормальной аппроксимации вычисляются следующие характеристики дохода каждого i -го региона [3].

Мода: $\hat{y}_{mo}(i) = \exp(\hat{m}_i - \hat{\sigma}_i^2)$.

Медиана: $\hat{y}_{me}(i) = \exp(\hat{m}_i)$.

Среднее: $\hat{y}_{cp}(i) = \exp(\hat{m}_i + \hat{\sigma}_i^2 / 2)$.

Дисперсия: $\hat{\sigma}_y^2(i) = e^{2\hat{m}_i + \hat{\sigma}_i^2} (e^{\hat{\sigma}_i^2} - 1)$ и СКО: $\hat{\sigma}_y(i) = \sqrt{\hat{\sigma}_y^2(i)}$.

Если макроэкономическое среднее не задается, то второй и третий способы объединяются в одном алгоритме нахождения параметров m_i и σ_i в соответствии с (5) – (7).

В итоге по оценкам параметров $\hat{m}_i$ и $\hat{\sigma}_i$ находится логнормальное распределение дохода в i -м регионе:

$$g_i(y) = \frac{1}{\sqrt{2\pi}\hat{\sigma}_i} \frac{1}{y} \exp\left[-\frac{(\ln y - \hat{m}_i)^2}{2\hat{\sigma}_i^2}\right].$$

Аппроксимация гистограмм смесью логнормальных распределений

Смесь распределений применяется для более точной аппроксимации гистограмм. Рассмотрим смесь двух логнормальных распределений:

$$g(y) = \omega_1 g_1(y) + \omega_2 g_2(y) = \frac{\omega_1}{\sqrt{2\pi}\sigma_1} \frac{1}{y} e^{-\frac{(\ln y - m_1)^2}{2\sigma_1^2}} + \frac{\omega_2}{\sqrt{2\pi}\sigma_2} \frac{1}{y} e^{-\frac{(\ln y - m_2)^2}{2\sigma_2^2}}, \quad (8)$$

где $m_1, m_2, \sigma_1, \sigma_2$ – параметры распределения $g(y)$; ω_1, ω_2 – весовые коэффициенты: $0 < \omega_1 < 1, \omega_2 = 1 - \omega_1$.

Найдем числовые характеристики закона распределения (8).

Математическое ожидание и дисперсия находятся аналитически на основе свойства линейности интегрирования с учетом

$$y_{cp} = e^{m + \sigma^2 / 2}, \quad \sigma_y^2 = e^{2m + \sigma^2} (e^{\sigma^2} - 1).$$

Имеем:

$$y_{cp} = \omega_1 e^{m_1 + \sigma_1^2 / 2} + \omega_2 e^{m_2 + \sigma_2^2 / 2}, \quad \sigma_y^2 = \omega_1 e^{2m_1 + \sigma_1^2} (e^{\sigma_1^2} - 1) + \omega_2 e^{2m_2 + \sigma_2^2} (e^{\sigma_2^2} - 1). \quad (9)$$

Из двух уравнений (9) с четырьмя неизвестными параметрами $m_1, m_2, \sigma_1, \sigma_2$ не удастся выразить эти параметры через оценки $\hat{y}_{cp}$ и $\hat{\sigma}_y^2$, вычисляемые по гистограмме. Параметры находятся численно путем решения задачи минимизации аналогично (7). В качестве начального приближения берем параметры, найденные для одного логнормального распределения: $m_1(0) = m_2(0) = m$, $\sigma_1(0) = \sigma_2(0) = \sigma$ при начальных весах $\omega_1 = \omega_2 = 1/2$. Простейшим способом пошаговой оптимизации на прямоугольной сетке значений параметров (перебором с помощью четырех циклов) в окрестности начальных значений находим оценки $\hat{m}_1, \hat{m}_2, \hat{\sigma}_1, \hat{\sigma}_2$. Аналогично находятся оценки $\hat{\omega}_1$ и $\hat{\omega}_2 = 1 - \hat{\omega}_1$ пошаговым поиском ω_1 ($\omega_2 = 1 - \omega_1$) в окрестности 0,5 добавлением пятого цикла.

Мода и медиана смеси распределений (8) находятся численно после того, как найдены параметры $\hat{m}_1, \hat{m}_2, \hat{\sigma}_1, \hat{\sigma}_2$: мода – как точка максимума функции $g(y)$ пошаговым поиском, а

медиана – как точка, делящая площадь под кривой распределения пополам. Несложно при аналитическом поиске моды и медианы взять производную и интеграл от суммы функций, но при этом получаются нелинейные уравнения, требующие численного решения более сложным путем.

Функция распределения смеси распределений (8) найдется [3] на основании свойства линейности интеграла и окончательно принимает вид:

$$\begin{aligned} G(y) &= \int_0^y g(t)dt = \omega_1 \int_0^y g_1(t)dt + \omega_2 \int_0^y g_2(t)dt = \\ &= G(y) = \omega_1 m_1 [F(\sigma_1) + F(z_1)] + \omega_2 m_2 [F(\sigma_2) + F(z_2)], \\ z_1 &= \frac{\ln y - m_1}{\sigma_1} - \sigma_1, \quad z_2 = \frac{\ln y - m_2}{\sigma_2} - \sigma_2, \end{aligned} \quad (10)$$

где $F(z)$ – функция Лапласа.

Проверка распределений по критериям согласия

Для проверки гипотез о распределении доходов по одному закону или смеси логнормальных законов воспользуемся двумя критериями согласия: Пирсона и Колмогорова. Логика критерия согласия Пирсона известна [1-4] – вычисляется статистика Z вида

$$Z = \sum_{j=1}^k \xi_j^2, \quad \xi_j = (n_j - np_j) / \sqrt{np_j} \sim N(0, 1),$$

где p_j – вероятность попадания на j -й интервал, вычисленная по аналитическому распределению; $np_j \geq 10$, $j = \overline{1, k}$; ξ_j – случайная величина, близкая к стандартному нормальному распределению. Если проверяемая гипотеза о распределении доходов верна, то случайная величина Z попадает на промежуток $(z_{\min}, z_{\beta})$ с доверительной вероятностью β , где квантили $z_{\min}$ и z_{β} берутся из таблиц с учетом числа степеней свободы. В противном случае $Z \notin (z_{\min}, z_{\beta})$ гипотеза неверна.

Если критерий согласия Пирсона применяется к одному логнормальному распределению $g(y)$, то аналитическое распределение $g(y)$ пересчитывается в нормальное распределение: $f(x) = g(\ln y)$. Соответственно пересчитывается эмпирическое распределение:

$$f^*(x_j) = (n_j / n) / \ln \Delta y_j, \quad j = \overline{1, k}.$$

После чего сравнением $f(x)$ и $f^*(x_j)$ проверяется гипотеза о нормальном распределении случайной величины X .

Если критерий согласия применяется к смеси логнормальных распределений $g(y) = \omega_1 g_1(y) + \omega_2 g_2(y)$, то получается новое аналитическое распределение $g(y)$. Проверяется гипотеза о соответствии выборки или гистограммы распределению $g(y)$. При этом вычисляются вероятности попадания случайной величины Y на промежутки (α, β) численным интегрированием:

$$P\{Y \in (\alpha, \beta)\} = \omega_1 \int_{\alpha}^{\beta} g_1(y)dy + \omega_2 \int_{\alpha}^{\beta} g_2(y)dy$$

или с помощью функции распределения (10).

Второй критерий согласия – критерий Колмогорова основан на вычислении статистики вида:

$$z = \sqrt{n} \max_y |G^*(y) - G(y)|,$$

где $G^*(y)$ и $G(y)$ – соответственно эмпирическая, найденная по гистограмме, и теоретическая, соответствующая проверяемой гипотезе, функции распределения случайной величины Y . Заметим, что в критерии Колмогорова параметры распределения считаются известными.

Построение трендов и прогнозирование характеристик дохода

Зафиксируем номер региона (символ i опускаем) и обозначим символом j номер года формирования гистограммы: $j = \overline{1, N}$, где N – число последних лет, предшествующих текущему году. Задача заключается в построении трендов параметров m_j и σ_j по последовательностям $\{\hat{m}_j\}$ и $\{\hat{\sigma}_j\}$, $j = \overline{1, N}$, с учетом численности населения n_j , обследуемого в j -м году, а также с экспоненциальным забыванием старых данных обследования дохода по сравнению с более актуальными текущими данными. Обозначим символом $x(j)$ – трендовое значение отдельного параметра распределения $p(j)$, найденного в j -м году.

Зададим полиномиальную модель m -го порядка для тренда $x(j)$ изменения параметра $p(j)$ в последовательности j -х лет [5-7]:

$$x(j) = \sum_{s=1}^{m+1} a(s)(j-1)^{s-1}, \quad j = \overline{1, N}, \quad (11)$$

где $a(s)$ – s -й параметр модели.

Принимается аддитивная модель временного ряда:

$$p(j) = x(j) + e(j), \quad j = \overline{1, N},$$

где $e(j)$, $j = \overline{1, N}$ – остаточный ряд – дискретный случайный процесс. Если последовательность $\{e(j)\}$ коррелированная с нулевым средним, то адекватность модели (11) можно проверить по известным критериям [5-7].

Оценки $\hat{a}(s)$, $s = \overline{1, m}$, параметров $a(s)$ в (11) находятся методом наименьших квадратов с весами $w(i)$ по критерию

$$J = \sum_{j=1}^N w_j (x(j) - \sum_{s=1}^{m+1} a(s)(j-1)^{s-1})^2 \rightarrow \min_{a(s)}, \quad (12)$$

где весовые коэффициенты w_j определяются численностью населения n_j и параметром α забывания старых данных ($\alpha \geq 0$, например, $\alpha = 0,5$):

$$w_j = n_j \exp(-\alpha(N-j)).$$

В матричной форме критерий (12) принимает вид:

$$J = (X - FA)^T W (X - FA) \rightarrow \min_A, \quad (13)$$

где W – диагональная матрица: $W = \text{diag}[w(1), w(2), \dots, w(N)]$; T – символ транспонирования.

Минимизацией (13) по вектору A параметров стандартным преобразованием находится вектор $\hat{A}$ оценок параметров $a(s)$:

$$dJ/dA = 2(X - FA)^T (-F) = O^T, \quad F^T W (X - FA) = O, \quad \hat{A} = (\hat{a}(1) \dots \hat{a}(m+1))^T = (F^T W F)^{-1} F^T W X, \quad (14)$$

$$X = (x(1), x(2), \dots, x(N))^T, \quad F = (f_1^T, f_2^T, \dots, f_N^T)^T, \quad f_j^T = (1, j-1, (j-1)^2, \dots, (j-1)^m), \quad j = \overline{1, N}.$$

С помощью вектора $\hat{A}$, найденного в (14), вычисляется прогнозное значение параметра $p(j)$ с шагом h на $(N+h)$ -й год в каждом регионе:

$$\hat{p}(N+h) = \sum_{s=1}^{m+1} \hat{a}(s)(N+h-1)^{s-1}.$$

Доверительный интервал прогноза рассчитывается по формуле [5-7]:

$$\hat{p}(N+h) \pm k \hat{\sigma}_e \sqrt{1 + f_{N+h}^T (F^T F)^{-1} f_{N+h}}, \quad (15)$$

где $f_{N+h}^T = (1, N+h-1, \dots, (N+h-1)^m)$; k – параметр, определяемый правилом « k -сигма» (например, $k=2$); $\hat{\sigma}_e$ – выборочное СКО остаточного ряда $\hat{e}(j) = \hat{p}(j) - p(j)$, $j = \overline{1, N}$:

$$\sigma_e = \sqrt{\frac{1}{N} \sum_{j=1}^N \hat{e}^2(j) - (\hat{e}_{cp})^2}, \quad \hat{e}_{cp} = \frac{1}{N} \sum_{j=1}^N \hat{e}(j).$$

Такова классическая процедура математического прогнозирования на основе степенного тренда (возможны и другие модели тренда). Прогнозные значения параметров используются для аналитического описания распределения дохода и вычисления показателей уровня жизни на следующий год ($h=1$) или два ($h=2$).

Корректировка параметров распределения дохода в текущем году

Для снижения ошибок построения гистограмм в текущем году предлагается осуществлять корректировку параметров распределения дохода, найденных в текущем году, с учетом трендов изменения параметров за несколько предыдущих лет следующим образом.

1. С помощью оценок (14) по выборке измерений $p_i(j)$, $j = \overline{1, N}$, для каждого i -го региона РФ ($i = \overline{2, M}$) вычисляется сглаженное значение параметра p_i на текущий N -й год:

$$\tilde{p}_i(N) = \sum_{s=1}^{m+1} \hat{a}_i(s)(N-1)^{s-1}, \quad (16)$$

и устанавливаются границы доверительного интервала в соответствии с (15) при $h=0$:

$$\tilde{p}_i(N) \pm k \hat{\sigma}_e \sqrt{1 + f_N^T (F^T F)^{-1} f_N}. \quad (17)$$

2. Если измеренное значение $p_i(N)$ оказывается в пределах доверительного интервала (17), то в качестве оценки параметра берется само измеренное значение: $\hat{p}_i(N) = p_i(N)$. Если измерение $p_i(N)$ выходит за пределы доверительного интервала, то в качестве оценки параметра берется ближайшая граница доверительного интервала. Тем самым осуществляется корректировка параметра.

3. При наличии хотя бы одной корректировки параметра p_i , $i \in \{2, 3, \dots, M\}$, осуществляется балансировка значений параметра $\hat{p}_i(N)$ по всем i -м регионам ($i = \overline{2, M}$) на соответствие параметру $p_1(N)$, найденному для РФ усреднением с весами по всем регионам:

$$p_1(N) = \sum_{i=2}^M n_i p_i(N) / \sum_{i=2}^M n_i.$$

Сбалансированные значения $\hat{p}_i^*(N)$, $i = \overline{2, M}$, вычисляются умножением на коэффициент γ :

$$\hat{p}_i^*(N) = \gamma \hat{p}_i(N), \quad \gamma = p_1(N) / \sum_{i=2}^M n_i \hat{p}_i(N), \quad i = \overline{2, M},$$

при этом выполняется условие балансировки: $\sum_{i=2}^M n_i \hat{p}_i^*(N) = p_1(N)$.

Вычисление показателей уровня жизни населения в регионах с учетом их корреляции с показателями по РФ

При расчете показателей уровня жизни в текущем N -м году на основе аналитического распределения дохода, разумно параметры этого распределения устанавливать с учетом корреляции с параметрами распределения дохода по РФ на протяжении N лет, так как распределение РФ менее подвержено искажениям за счет усреднения по регионам. В плане обсуждения данного подхода предлагается следующий алгоритм учета корреляции.

1. Вычисляются выборочные ковариация $K_p(i)$ и коэффициенты корреляции $r_p(i)$ между параметром РФ p_1 и параметром i -го субъекта РФ p_i :

$$K_p(i) = \frac{1}{N} \sum_{j=1}^N \hat{p}_1(j) \hat{p}_i(j) - \bar{p}_1 \bar{p}_i, \quad k_p(i) = K_p(i) / (\sigma_1 \sigma_i), \quad i = \overline{2, M}.$$

где $\hat{p}_1(j)$ и $\hat{p}_i(j)$ – оценки параметров, найденные в j -м году; $\bar{p}_1$ и $\bar{p}_i$ – выборочные средние; σ_1 и σ_i – выборочные СКО:

$$\bar{p}_i = \frac{1}{N} \sum_{j=1}^N \hat{p}_i(j), \quad \sigma_i = \sqrt{\frac{1}{N} \sum_{j=1}^N \hat{p}_i^2(j) - (\bar{p}_i)^2}, \quad i = \overline{1, M}.$$

2. Устанавливается тренд $x_1(j)$, $j = \overline{1, N}$ параметра РФ p_1 , например (11):

$$x_1(j) = \sum_{s=1}^{m+1} \hat{a}_1(s)(j-1)^{s-1}, \quad j = \overline{1, N}.$$

3. Корректируется тренд параметра i -го региона по формуле

$$x_i^*(j) = \sum_{s=1}^{m+1} \hat{a}_i^*(s)(j-1)^{s-1}, \quad j = \overline{1, N}.$$

Коррекция заключается в том, что параметры модели $\hat{a}_i(s)$, $s = \overline{2, m+1}$, отвечающие за производные изменения параметра (скорость, ускорение и др.), меняются с учетом корреляции с РФ:

$$\hat{a}_i^*(s) = \hat{a}_i(s) + \lambda(\hat{a}_1(s) - \hat{a}_i(s))k_p(i), \quad s = \overline{2, m+1}, \quad (18)$$

где множитель λ выбирается эмпирически ($0 < \lambda \leq 1$), а параметр $\hat{a}_i(1)$, отвечающий за параллельный сдвиг тренда (по вертикали – уровню дохода) при коррекции (18), находится легко методом МНК:

$$J = \sum_{j=1}^N w_j [p_i(j) - \hat{a}_i(1) - \sum_{s=2}^{m+1} \hat{a}_i^*(s)(j-1)^{s-1}]^2 \rightarrow \min_{\hat{a}_i(1)}.$$

Примеры расчетов

На рисунке 1 показаны: графики гистограмм (ломаные линии), аппроксимация гистограмм одним логнормальным распределением и аппроксимация смесью двух логнормальных распределений по доходам на душу населения 2018 года в рублях. При этом нормированные показатели близости аппроксимаций к гистограммам составили для одного логнормального распределения: РФ – 0,38; Ряз. обл. – 0,72; Москва – 0,53. Соответственно для смеси двух логнормальных распределений: РФ – 0,22; Ряз. обл. – 0,59; Москва – 0,48. Видно, что показатели близости смеси логнормальных распределений к гистограмме лучше (меньше), чем аналогичные показатели для одного логнормального распределения. Проверка распределений по критериям согласия подтверждает гипотезу о логнормальном распределении.

В таблице 1 сравниваются числовые характеристики, найденные по гистограмме и по смеси двух логнормальных распределений. Видно, что такая характеристика, как мода, наиболее сильно подвержена искажениям при гистограммной обработке.

Таблица 1 – Числовые характеристики распределения дохода

Table 1 – Numerical characteristics of income distribution

	По гистограмме			По смеси распределений		
	РФ	Ряз. обл.	Москва	РФ	Ряз. обл.	Москва
Мода	13750	13750	36250	14259	13515	30378
Медиана	17500	15000	35000	18351	15375	36702
Среднее	21917	17487	40745	22300	17474	40454

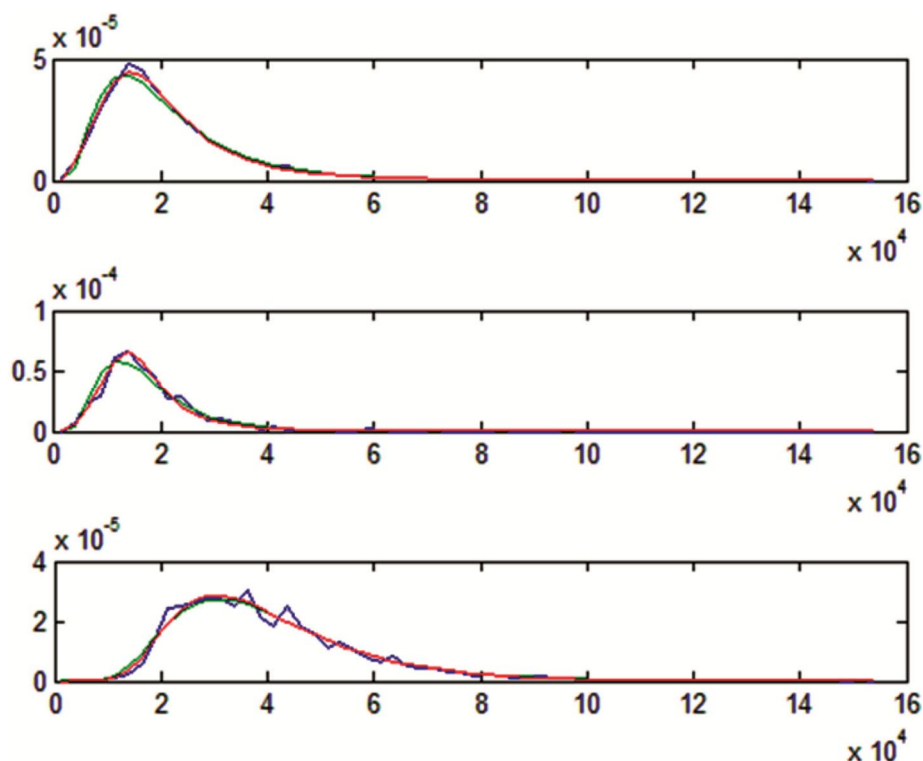


Рисунок 1 – Распределения дохода (сверху-вниз: РФ, Рязанская область, Москва)
Figure 1 – Income distributions (top-down: Russian Federation, Ryazan region, Moscow)

На рисунке 2, *а* показаны линейные тренды зависимостей среднего (выше) и СКО (ниже) по 6 годам (2013 – 2018) с прогнозом и доверительным интервалом (вероятности 0,95) на 7-й год (2019-й) для РФ. На рисунке 2, *б* – зависимости для степенной модели 2-го порядка.

На рисунках 3, *а-5, а* показаны гистограммы распределения дохода по РФ, Ряз. обл. и Москве за 2018 год. На рисунках 3, *б-5, б* – логнормальная аппроксимация распределения дохода по РФ, Ряз. обл. и Москве при линейном прогнозе распределения на 2019 год. Указаны числовые характеристики, найденные по результатам прогноза распределения дохода.

В таблице 2 показан уровень бедности, рассчитанный без учета корреляции региона (Ряз. обл. и Москвы) с РФ и с учетом корреляции по данным 2018 года с использованием макроэкономического среднего. Приводится коэффициент корреляции k_p . Заметим, что после учета корреляции по всем регионам требуется балансировка измененных расчетных значений на их соответствие значению по РФ аналогично рассмотренной выше.

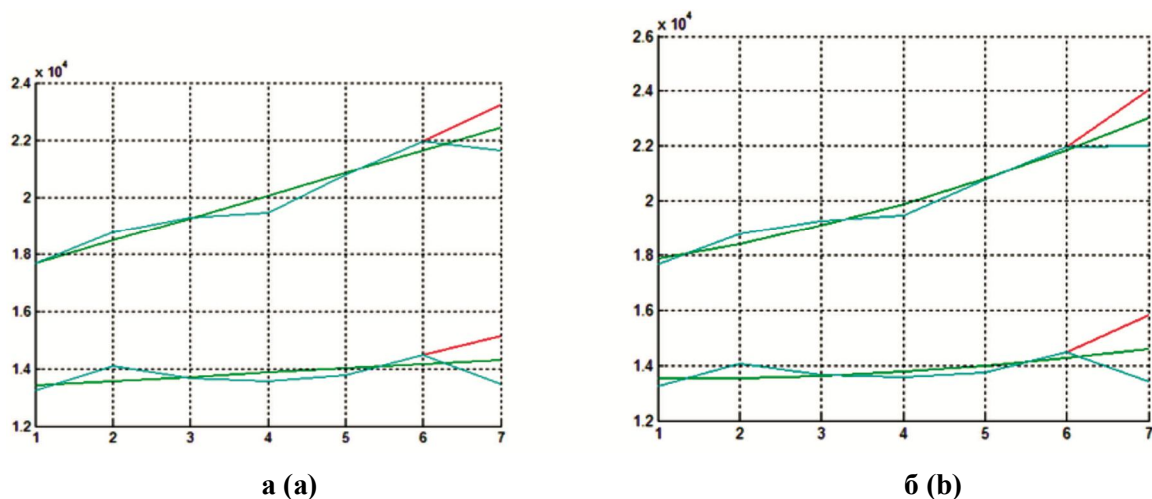
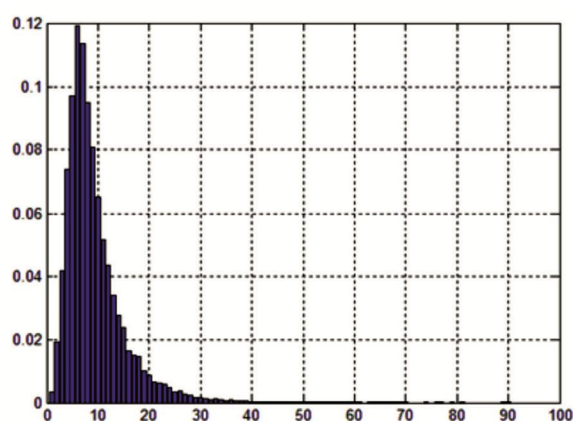
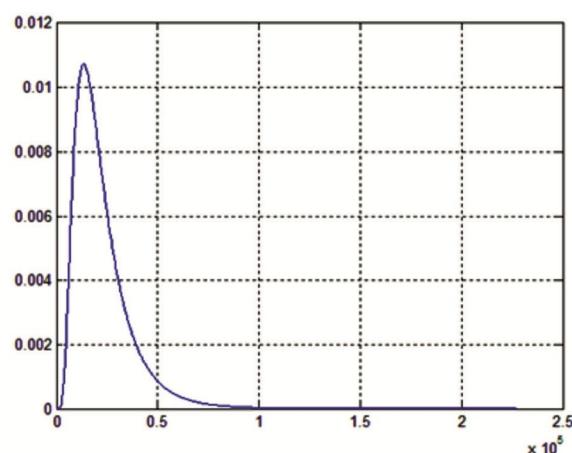


Рисунок 2 – Тренды РФ зависимостей среднего и СКО по 6 годам с прогнозом
Figure 2 – Trends of the Russian Federation of middle and MSD for 6 years with forecast



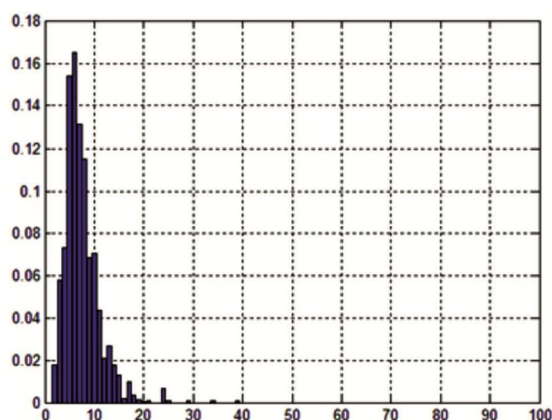
а (a)



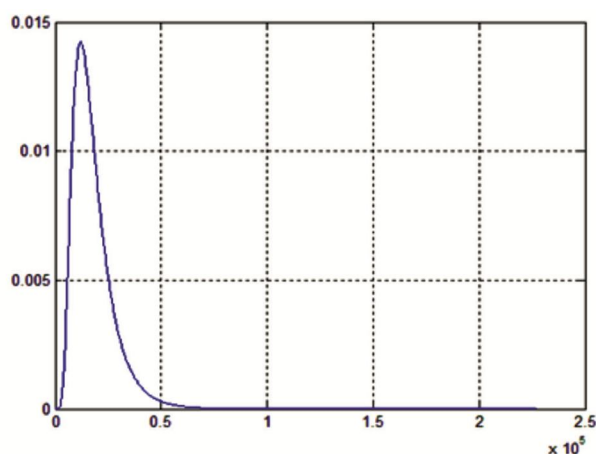
б (b)

Рисунок 3 – Линейный прогноз распределения дохода по РФ
(Мода – 13446, Медиана – 18915, Среднее – 22434)

Figure 3 – Linear forecast of income distribution in the Russian Federation
(Mode – 13446, Median – 18915, Middle – 22434)



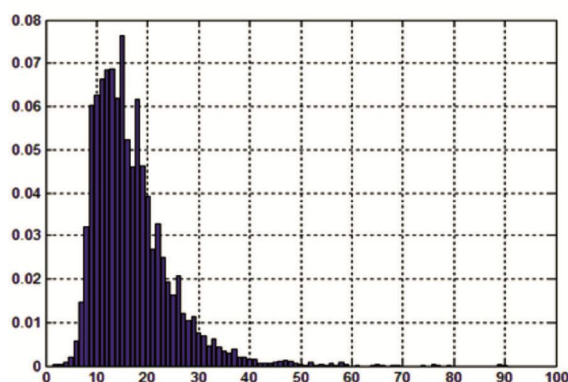
а (a)



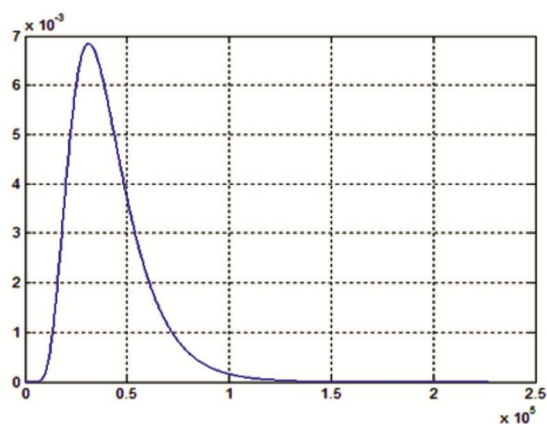
б (b)

Рисунок 4 – Линейный прогноз распределения дохода по Рязанской области
(Мода – 12003, Медиана – 15608, Среднее – 17798)

Figure 4 – Linear forecast of income distribution in Ryazan region
(Mode – 12003, Median – 15608, Middle – 17798)



а (a)



б (b)

Рисунок 5 – Линейный прогноз распределения дохода по Москве
(Мода – 31149, Медиана – 37369, Среднее – 40930)

Figure 5 – Linear forecast of income distribution in Moscow
(Mode – 31149, Median – 37369, Medium – 40930)

Таблица 2 – Уровень бедности без учета корреляции и с учетом
Table 2 – Poverty rate excluding and including correlation

РФ	Ряз. обл.	Москва
12,57 %	12,74 %	3,39 %
С учетом корреляции	12,72 %, $k_p = -0,14$	3,51 %, $k_p = 0,85$

Заключение

Изложены математические модели расчета показателей уровня жизни населения регионов РФ, основанные на известных и предложенных подходах. Известные подходы используют логнормальную аппроксимацию гистограмм распределений дохода одним распределением. Использование смеси логнормальных распределений дает более точное аналитическое описание гистограмм. Однако при этом по-прежнему остаются ошибки, присутствующие в гистограммах. Предложенный подход отличается построением трендов для оценок параметров логнормальных распределений в последовательности нескольких лет. Это позволяет сглаживать ошибки гистограмм при расчете показателей дохода на текущий год, а также прогнозировать распределения доходов населения. При построении трендов учитывается численность населения в регионах по годам и актуальность текущих данных.

В плане обсуждения предложен подход к вычислению показателей уровня жизни населения в регионах с учетом их корреляции с показателями по РФ, менее подверженными искажениям. Приводятся примеры расчетов, подтверждающие работоспособность предложенных алгоритмов. Результаты работы могут найти применение при расчете показателей уровня жизни населения РФ после соответствующей апробации в Росстате.

Библиографический список

1. Айвазян С. А., Мхитарян В. С. Теория вероятностей и прикладная статистика. М.: ЮНИТИ-ДАНА, 2001. 656 с.
2. Елисеева И. И., Юзбашев М. М. Общая теория статистики: учебник / Под ред. И. И. Елисеевой. М.: Финансы и статистика, 2004. 656 с.
3. Теория вероятностей и математическая статистика в задачах Росстата: методич. указания к курсовой работе / В. К. Ключко; Рязан. гос. радиотехн. ун-т. – Рязань, 2020. – 16 с.
4. Гмурман В. Е. Теория вероятностей и математическая статистика: учеб. пособие. М.: Высшая школа, 2003. 479 с.
5. Чураков Е. П. Математические методы обработки экспериментальных данных в экономике: учеб. пособие. М.: Финансы и статистика, 2004. 240 с.
6. Чураков Е. П. Прогнозирование эконометрических временных рядов: учеб. пособие. М.: Финансы и статистика, 2008. 208 с.
7. Математические методы прогнозирования: учеб. пособие / В. К. Ключко; Рязан. гос. радиотехн. ун-т. Рязань, 2019. – 68 с.

UDC 519.2

MATHEMATICAL MODELS FOR IMPROVING ACCURACY CALCULATION OF POPULATION LIVING STANDARDS

V. K. Klochko, Dr. Sc. (Tech.) Professor, Department of Automation and Information Technologies in Control, RSREU, Ryazan, Russia;

orcid.org/0000-0003-2550-999X, e-mail: klochkovk@mail.ru

V. P. Kuznetsov, associate professor, Department of Automation and Information Technologies in Control, RSREU, Ryazan, Russia;

orcid.org/0000-0001-6189-9011, e-mail: kuznetsovaitu@yandex.ru

*Mathematical models to calculate indicators of population living standards in Russian Federation regions according to the data of Federal State Statistics Service (FSSS) published in the form of histograms of income obtained by interviewing the population are given. The known approach to calculating the indicators is based on histograms approximation with lognormal distribution. The approach proposed differs from others by the use of distributions mixture and estimation of lognormal distributions parameters in several years sequence. **The aim of the work** is to develop different mathematical models of population income of the Russian Federation regions taking into account the data of population survey for several previous years, which make it possible to increase the accuracy of calculating indicators of population living standard in the regions of Russian Federation. Specific algorithms for calculating living standards based on the proposed approach have been presented. The possibility of calculating indicators in the regions of Russian Federation taking into account their correlation with indicators of Russian Federation is being discussed. Examples of the calculations confirming the operability of the algorithms offered are given. Algorithms can be used in FSSS practice.*

Key words: Algorithms, forecasting, population incomes, mathematical models.

DOI: 10.21667/1995-4565-2020-73-105-115

References

1. **Ajvazyan S. A., Mhitaryan V. S.** *Teoriya veroyatnostej i prikladnaya statistika* (Probability theory and applied statistics). Moscow.: YUNITI-DANA, 2001. 656 p. (in Russian)
2. **Eliseeva I.I., YUzbashev M.M.** *Obshchaya teoriya statistiki: uchebnik* (General theory of statistics: textbook) / Pod red. I. I. Eliseevoj. Moscow.: Fi-nansy i statistika, 2004. 656 p. (in Russian)
3. *Teoriya veroyatnostej i matematicheskaya statistika v zadachah Rosstata: metodich. ukazaniya k kursovoj rabote* (Probability theory and mathematical statistics in Rosstat problems: methodic. instructions for coursework) / V. K. Klochko; Ryazan' gos. radiotekhn. un-t. Ryazan, 2020. 16 p. (in Russian)
4. **Gmurman V. E.** *Teoriya veroyatnostej i matematicheskaya statistika: ucheb. posobie* (Probability theory and mathematical statistics: manual). Moscow.: Vysshaya shkola, 2003. 479 p. (in Russian)
5. **Churakov E. P.** *Matematicheskie metody obrabotki eksperimental'nyh dannyh v ekonomike: ucheb. posobie* (Methods of processing experimental data in economics: manual). Moscow.: Finansy i statistika, 2004. 240 p. (in Russian)
6. **Churakov E. P.** *Prognozirovanie ekonometricheskikh vremennyh ryadov: ucheb. posobie* (Forecasting econometric time series: manual). Moscow.: Finansy i statistika, 2008. 208 p. (in Russian)
7. *Matematicheskie metody prognozirovaniya: ucheb. posobie* (Mathematical forecasting methods: manual) / **V. K. Klochko**; Ryazan. gos. radiotekhn. un-t. Ryazan', 2019. 68 p. (in Russian).