ЭКОНОМИКА
И МАТЕМАТИЧЕСКИЕ МЕТОДЫ
1966
том II, вып. 1

МОДЕЛИРОВАНИЕ РАСПРЕДЕЛЕНИЯ ЗАРАБОТНОЙ ПЛАТЫ С ПОМОЩЬЮ ЛОГАРИФМИЧЕСКИ-НОРМАЛЬНОГО ЗАКОНА

И. А. ГЕРАСИМОВА
(Москва)

1. Давно замечено [1—2], что распределение работающих по заработной плате и распределение семей по душевому доходу близки по виду к логарифмически-нормальному распределению*. Этот эмпирический факт вызвал несколько попыток его теоретического обоснования [1—2], результаты которых делают гипотезу о лог-нормальности весьма правдоподобной и привлекательной. Правильность этой гипотезы должна определяться отклонением теоретического (лог-нормального) распределения от эмпирического (фактического), а мера этого отклонения — ошибкой, которую мы получим при использовании в тех или иных расчетах теоретического распределения вместо фактического.

В каких же задачах используются ряды распределения работающих по заработной плате?

Это могут быть, например, расчеты, связанные с изменением налоговой системы. Целью таких расчетов является определение полной суммы средств, необходимых для проведения некоторого мероприятия, или доли этих средств, которая будет распределена между низкооплачиваемыми работниками. Иначе говоря, в этих расчетах необходимо получить по возможности более точное значение величины

$$y = \int_0^\infty y(x)\,dF(x) \tag{1}$$

или

$$y_0 = \int_0^{x_0} y(x)\,dF(x), \tag{1a}$$

где $y(x)$ — сумма налога, взимаемая с работника, получающего заработную плату x, x_0 — уровень заработной платы, ниже которого работники считаются низкооплачиваемыми, $F(x)$ — распределение работающих по заработной плате.

Распределение по заработной плате и доходам может применяться для определения потребительского спроса на товары и услуги. Если $y = y(x)$ есть функция зависимости спроса на некоторый товар от дохода x, то общий спрос на этот товар вычисляется также по формуле (1), где $F(x)$

* Необходимо отметить, что распределение работающих по заработной плате (и распределение семей по душевому доходу) ни в коей мере не рассматривается как распределение одной случайной величины, независимыми реализациями которой являются величины заработной платы отдельных работников, т. е. предположение, что значения заработных плат различных работников есть независимые одинаково распределенные величины, не правомерно. Однако число работников, получающих заработную плату в пределах отдельного интервала, есть некоторая дискретная мера на оси заработной платы. Естественно искать ее аппроксимацию с помощью простых и хорошо изученных непрерывных распределений методами теории вероятностей.

есть распределение по доходу. Доля потребления, приходящаяся на семьи с душевым доходом меньше x_0, определяется по формуле (1а). Очевидно, функции $y(x)$ во всех реальных расчетах являются гладкими (например, вторая производная достаточно мала по абсолютной величине).

Ошибка при определении величины y по теоретическому распределению составит:

$$\Delta y = \int_0^\infty y(x)\,d(F(x) - \bar F(x)) = [F(x) - \bar F(x)]\Big|_0^\infty - \int_0^\infty (F(x) - \bar F(x))\,y'(x)\,dx$$

и

$$\frac{|\Delta y|}{y} \leqslant \max_x |F(x) - \bar F(x)|\,\frac{\int\limits_0^\infty |y'(x)|\,dx}{y}, \qquad (2)$$

где $F(x)$ — теоретическая функция распределения; $\bar F(x)$ — эмпирическая (фактическая) функция распределения; y — среднедушевое потребление (или средний налог).

Распределение по заработной плате используется также при расчете дополнительных средств, необходимых для введения нового минимума заработной платы, для определения средней заработной платы в отрасли или для определения доли низкооплачиваемых рабочих. Гипотеза о лог-нормальности распределения заработной платы оказывалась весьма плодотворной в расчетах такого типа. Ниже будет описан математический аппарат расчетов по заработной плате.

2. Для проверки гипотезы о лог-нормальности распределения заработной платы надо иметь фактический (эмпирический) и теоретический ряды распределения заработной платы. Эмпирическое распределение, как правило, задается как ряд $\bar p_1, \ldots, \bar p_n$ наблюденных частот попадания значения заработной платы x в интервалы (x_0, x_1), $(x_1, x_2), \ldots, (x_{i-1}, x_i), \ldots$ $\ldots, (x_{n-1}, x_n)$ группировки заработной платы. Обычно частоты выражены

в процентах, т. е. $\sum\limits_{i=1} \bar p_i = 100$.

Ряд распределения лог-нормальной случайной величины по тем же интервалам (x_{i-1}, x_i) $(i = 1, \ldots, n)$ можно получить, используя ее функцию распределения.

Логарифмически-нормальное распределение $F(x)$ — это распределение некоторой случайной величины ζ, логарифм которой $\ln \zeta$ имеет нормальное (гауссовское) распределение. Поэтому

$$F(x) = P\{\zeta < x\} = P(\ln \zeta < \ln x) = \Phi\left(\frac{\ln x - a}{\sigma}\right),$$

где $\Phi(x) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-\infty}^{x} e^{-y^2/2}\,dy$ — нормальная функция распределения, и a и σ — математическое ожидание и среднеквадратичное отклонение случайной величины $\ln \zeta$. Отсюда получаем функцию плотности $\varphi(x)$ и функцию распределения $F(x)$ в виде:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}\,\sigma x}\,e^{-(\ln x - a)^2/2\sigma^2}, \qquad (3)$$

$$F(x) = \int_0^x \varphi(u)\,du = \frac{1}{\sqrt{2\pi}\,\sigma} \int_0^x \frac{1}{u}\,e^{-(\ln u - a)^2/2\sigma^2}\,du. \qquad (4)$$

Тогда вероятности p_i распределения лог-нормальной случайной величины по интервалам $(x_0, x_1), \ldots, (x_{n-1}, x_n)$ легко выражаются через функцию распределения $F(x)$:

$$p_i = F(x_i) - F(x_{i-1}) = P(x \in (x_{i-1}, x_i)) = \frac{1}{\sqrt{2\pi}\,\sigma} \int\limits_{x_{i-1}}^{x_i} e^{-(\ln u - a)^2/2\sigma^2}\, d\ln u. \quad (5)$$

Чтобы вычислить теоретические вероятности p_i, нам нужны параметры a и σ распределения логарифма заработной платы. Поскольку истинные значения a и σ неизвестны, воспользуемся их статистическими оценками. Для этого от распределения заработной платы перейдем к распределению логарифма заработной платы. При этом границы интервалов $x_0, x_1, \ldots, x_n$ перейдут в точки $\ln x_0, \ln x_1, \ldots, \ln x_n$.

Легко видеть, что логарифм заработной платы $\ln x$ имеет такую же частоту попадания в интервал $(\ln x_{i-1}, \ln x_i)$, с какой значение заработной платы x попадает в интервал (x_{i-1}, x_i), т. е. эмпирический ряд распределения $\ln x$ по интервалам $(\ln x_{i-1}, \ln x_i)$ тот же, что и ряд наблюдения частот для x. Тогда параметры a и σ определятся по известным формулам математической статистики:

$$\begin{aligned} a &= \frac{1}{100} \sum_{i=1}^{n} \ln \bar{x}_i \cdot \bar{p}_i, \\ \sigma^2 &= \frac{1}{100} \sum_{i=1}^{n} (\ln \bar{x}_i - a)^2 \cdot \bar{p}_i, \end{aligned} \quad (6)$$

где $\ln \bar{x}_i = {}^1/_2 (\ln x_i + \ln x_{i-1})$ — середины интервалов группировки логарифма заработной платы, $\bar{p}_1, \ldots, \bar{p}_n$ — эмпирический ряд распределения логарифма заработной платы по логарифмическим интервалам. Подставляя полученные значения a и σ в формулу (5), получим ряд распределения логарифмически-нормальной случайной величины по интервалам $(x_0, x_1), \ldots, (x_{n-1}, x_n)$. Так как нормальная функция распределения табулирована, формула (5) удобна для вычислений. Определенные таким образом значения p_i $(i = 1, \ldots, n)$ выражаются в долях единицы, т. е. $\sum\limits_{i=1}^{n} p_i = 1$. Ряды распределений $\bar{p}_i$ обычно задаются в процентах, поэтому для сравнения p_i с фактическими значениями $\bar{p}_i$ их надо умножить на 100. Не меняя системы обозначений, далее мы всюду будем предполагать, что эмпирический и теоретический ряды распределения представлены в одних и тех же единицах измерения.

Какую же ошибку мы получим, приняв гипотезу о лог-нормальности распределения по заработной плате?

Как было показано выше, относительная ошибка при определении некоторой средней величины (например, среднего потребления или средней заработной платы) есть

$$\frac{|\Delta y|}{y} \leqslant \max_x |F(x) - \bar{F}(x)| \cdot \frac{\int\limits_0^\infty |y'(x)|\, dx}{y},$$

где $y = \int\limits_0^\infty y(x)\, dF(x)$, $\int\limits_0^\infty |y'(x)|\, dx$ — вариация функции $y(x)$. В силу гладкости $y(x)$ ее вариация не может быть слишком большой. Как правило, функции, встречающиеся в реальных плановых расчетах, являются монотонно возрастающими или убывающими (душевое потребление, налог на одного работающего). В этих случаях оценка (2) может быть уточнена:

$$\int_0^\infty |y'(x)|\,dx = |y(x_{\max}) - y(x_{\min})|, \tag{7}$$

где за $x_{\max}$ и $x_{\min}$ принимаются такие уровни заработной платы (или дохода), что сумма заработной платы (дохода) выше уровня $x_{\max}$ и сумма заработной платы ниже уровня $x_{\min}$ составляют пренебрежимо малую долю от всего фонда заработной платы.

Так как мы рассматриваем не непрерывное распределение по всему множеству значений заработной платы, а распределение по некоторым интервалам группировки $(x_0, x_1), \ldots, (x_{n-1}, x_n)$, то функции распределения $F(x)$ и $\bar{F}(x)$ есть ступенчатые функции со скачками в точках x_i $(i = 1, \ldots, n)$:

$$F(x_i) = \sum_{i=1}^{i} p_j, \quad i = 1, \ldots, n, \tag{8}$$

$$\bar{F}(x_i) = \sum_{j=1}^{i} \bar{p}_j,$$

тогда $\max\limits_x |F(x) - \bar{F}(x)| = \max\limits_x |F(x_i) - \bar{F}(x_i)|$. Обозначив величину $\dfrac{\int_0^\infty |y'(x)|\,dx}{y} = C$, получим:

$$\frac{|\Delta y|}{y} \leqslant C \cdot \max_{x_i} |F(x_i) - \bar{F}(x_i)|. \tag{9}$$

Если $y(x)$ сильно колеблется, то воспользуемся другой оценкой для Δy:

$$|\Delta y| = \left| \int_0^\infty y(x)\,d(F(x) - \bar{F}(x)) \right| \leqslant \sum_{i=1}^{n} \max_{x\in(x_{i-1},\, x_i)} |y(x)| \cdot \left| \int_{x_{i-1}}^{x_i} d(F(x) - \bar{F}(x)) \right| \leqslant \max_{x\in(x_{i-1},\, x_i)} |y(x)| \cdot \sum_{i=1}^{n} |p_i - \bar{p}_i|. \tag{10}$$

В случае $y(x) = x$

$$\max |x| = x_{\max}.$$

Величина $\sum\limits_{i=1}^{n} |p_i - \bar{p}_i|$ есть сумма модулей отклонений теоретических вероятностей от наблюденных частот по всем интервалам группировки.

Итак, пользуясь оценками (9) и (10), мы видим, что величина относительной ошибки при нахождении y определяется величиной

$$\max_{x_i} |F(x_i) - \bar{F}(x_i)| \quad \text{или} \quad \sum_{i=1}^{n} |p_i + \bar{p}_i|.$$

Расчеты, проводившиеся в НИИ труда, показали, что оценка (9) дает лучшие результаты, чем оценка (10). В табл. 1 приводятся результаты этих расчетов (в процентах).

Таблица 1

Отрасль	Год	$\max\limits_{x_i} \lvert F(x_i) - \bar{F}(x_i) \rvert$	$\sum\limits_{i=1}^{n} \lvert p_i - \bar{p}_i \rvert$	$\max\limits_{x_i} \lvert p_i - \bar{p}_i \rvert$	χ
Народное хозяйство	1956	0,7	4	0,5	5
» »	1961	5,3	16	3,5	19
Промышленность	1956	2,2	8	1,2	9
»	1961	2,0	9,8	1,8	12,5
Черная металлургия	1956	2,6	11,2	1,9	15,0
» »	1961	5,8	21,2	3,2	27,0
Машиностроение	1956	2,0	8,2	2,1	11,0
»	1961	3,4	11,4	1,7	15,5
Угольная	1956	2,1	11,2	1,7	18;4
»	1961	1,5	7,2	1,3	29,0
Непроизводственная сфера	1961	6,5	20,8	7,5	31,5
Сельское хозяйство	1956	6,7	19,4	3,7	26,3

Из табл. 1 видно, что максимальное отклонение между эмпирической и теоретической функциями распределения $\max\limits_{x_i} |F(x_i) - \bar{F}(x_i)|$ составляет в среднем 6% и что отклонение меньше для тех отраслей, для которых сами ряды распределения $\{\bar{p}_i\}$ и $\{p_i\}$ близки, т. е. $\max\limits_{x_i} |p_i - \bar{p}_i|$ незначительны (меньше 2%).

Оценки (9) и (10) в большинстве случаев могут оказаться сильно завышены. Как уже говорилось, функции $y(x)$ обычно монотонны, и вид их описывается одним из четырех видов кривых: логарифмической прямой, полулогарифмической прямой, прямой и параболой. В первых двух случаях, когда $\ln y = \alpha + \beta \ln x$ или $y = \alpha + \beta \ln x$, абсолютная ошибка

$$|\Delta y| \text{ или } |\Delta \ln y| = \left| \int_{-\infty}^{+\infty} (\alpha + \beta \ln x)\, d\left[\Phi\left(\frac{\ln x - a}{\sigma}\right) - \right.\right.$$

$$\left.\left. - \bar{\Phi}\left(\frac{\ln x - a}{\sigma}\right)\right]\right| = \left| \alpha \int_{-\infty}^{+\infty} d[\Phi_{a,\sigma}(\ln x) - \bar{\Phi}_{a,\sigma}(\ln x)] + \right.$$

$$\left. + \beta \int_{-\infty}^{+\infty} \ln x d[\Phi_{a,\sigma}(\ln x) - \bar{\Phi}_{a,\sigma}(\ln x)] \right| = 0,$$

так как оба слагаемых равны нулю:

$$\int_{-\infty}^{+\infty} d[\Phi_{a,\sigma}(\ln x) - \bar{\Phi}_{a,\sigma}(\ln x)] = 0,$$

$$\int_{-\infty}^{+\infty} \ln x d\,\Phi_{a,\sigma}(\ln x) = a, \qquad \int_{-\infty}^{+\infty} \ln x d\,\bar{\Phi}_{a,\sigma}(\ln x) = a,$$

где $\Phi\left(\frac{\ln x - a}{\sigma}\right)$ сокращенно обозначен $\Phi_{a,\sigma}(\ln x)$, а $\overline{\Phi}_{a,\sigma}(\ln x)$ — эмпирическая функция распределения $\ln x$. Когда $y = \alpha + \beta x$,

$$|\Delta y| = \left| \int_{-\infty}^{+\infty} (\alpha + \beta x)\, d[\Phi_{a,\sigma}(\ln x) - \overline{\Phi}_{a,\sigma}(\ln x)] \right| =$$

$$= \left| \beta \int_{-\infty}^{+\infty} x d[\Phi_{a,\sigma}(\ln x) - \overline{\Phi}_{a,\sigma}(\ln x)] \right| = \beta |\Delta \bar{x}|,$$

т. е. абсолютная ошибка в определении y пропорциональна ошибке в определении средней заработной платы (дохода).

Так как при построении теоретического распределения используется не оценка самой средней заработной платы, а оценка средней и дисперсии от логарифма (a и σ^2), то среднее значение $\bar{x} = \int_0^\infty x d\overline{F}(x)$, рассчитанное по фактическому распределению, может отличаться от среднего, рассчитанного по теоретическому распределению. Как будет показано ниже (13), теоретическое значение средней заработной платы выражается формулой: $m = e^{a+\frac{\sigma^2}{2}}$.

Сравнение этой величины с величиной $\bar{x}$ приведено в табл. 2.

И, наконец, если $y = \alpha + \beta x + \gamma x^2$, то

$$|\Delta y| = \left| \int_{-\infty}^{+\infty} (\alpha + \beta x + \gamma x^2)\, d[\Phi_{a,\sigma}(\ln x) - \overline{\Phi}_{a,\sigma}(\ln x)] \right| =$$

$$= \left| \beta(m - \bar{x}) + \gamma \int_{-\infty}^{+\infty} x^2 d[\Phi_{a,\sigma}(\ln x) - \overline{\Phi}_{a,\sigma}(\ln x)] \right| =$$

$$= |\beta(m - \bar{x}) + \gamma(\mu^2 - \bar{\mu}^2)|.$$

Второй момент, рассчитанный по фактическому распределению $\bar{\mu}^2 = \int_0^\infty x^2 d\overline{F}(x)$, будет отличаться от второго момента μ^2, рассчитанного по теоретическому распределению $F(x)$ по той же причине, по которой различаются средние значения. Сравнение μ^2 и $\bar{\mu}^2$ приведено также в табл. 2.

Можно использовать в качестве меры близости и величину, составленную по типу критерия χ^2 *. Обычно критерий χ^2 применяют для проверки гипотезы о виде распределения случайной величины, если имеется много независимых реализаций этой случайной величины. Конечно, предположение, что заработки различных работников есть независимые, одинаково распределенные величины, ни в какой степени не правдоподобно. Поэтому обычный способ применения критерия χ^2 не правомерен. Можно только проверить близость теоретического и эмпирического распределений по критерию, являющемуся естественным аналогом для χ^2:

$$\chi^2 = \sum_{i=1}^{n} \left(\frac{p_i - \bar{p}_i}{p_i} \right)^2 \cdot p_i, \tag{11}$$

* Проверка гипотезы о лог-нормальности по критерию χ^2 проводится также в ГВЦ Госплана СССР.

где $(p_i - \bar{p}_i) / p_i$ — относительная ошибка в определении p_i, т. е. χ^2 есть сумма взвешенных (с весами p_i) квадратов относительных ошибок.

В табл. 1 приведены расчетные значения величины $\chi = \sqrt{\sum_{i=1}^{n}\left(\frac{p_i - \bar{p}_i}{p_i}\right)^2 \cdot p_i}$.

Как видно из таблицы, значения χ в основном велики. Это происходит за счет крайних интервалов группировки, в которых сами значения p_i и $\bar{p}_i$ невелики, но относительная ошибка $(p_i - \bar{p}_i) / p_i$ принимает большое значение (из таблицы же видно, что абсолютные ошибки в определении p_i составляют в большинстве случаев только 2—3%).

Таблица 2

Отрасль	$\frac{\bar{x}-m}{\bar{x}}100$	$\frac{\bar{\mu}^2-\mu^2}{\bar{\mu}^2}100$
1956 год		
Народное хозяйство	0,7	5,5
Промышленность	0,1	0,5
Строительство	0,7	4,5
Транспорт и связь	1,6	9,6
Черная металлургия	0,6	1,7
Машиностроение	0,7	1,4
Угольная (все рабочие и служащие)	—	0,4
1961 год		
Народное хозяйство	0,9	5,0
Промышленность	0,3	4,5
Строительство	0,2	1,6
Транспорт и связь	0,7	5,2
Черная металлургия	0,8	10,8
Машиностроение	0,7	1,5
Угольная (все рабочие и служащие)	0,1	3,5
Сельское хозяйство	0,4	6,0
Непроизводственная сфера (с прочими отраслями)	—	6,5
Непроизводственная сфера (без прочих)	2,2	12,0
Угольная (только рабочие)	—	0,7
1930 год		
Промышленность (только рабочие)	0,3	1,3

Подытоживая сказанное выше, можно сделать следующий вывод. Использование логарифмически-нормального распределения в расчетах средней заработной платы дает результаты с достаточной степенью точности, погрешность в определении не превосходит 5—6%. Однако нельзя говорить о близости между теоретическим и фактическим распределением в смысле критерия χ^2 в силу указанных выше причин.

3. При планировании заработной платы в различных отраслях хозяйства важно знать соотношение между ростом фонда заработной платы (или ее средней) и минимальной заработной платой. Такая взаимосвязь, представленная в простом виде, дает возможность принимать обоснованные решения о распределении прироста фонда заработной платы между различными отраслями при заранее заданных условиях. Описанная выше аппроксимация распределения рабочих и служащих по заработной плате логарифмически-нормальным распределением позволяет установить простой вид этой зависимости.

Исходя из гипотезы о логарифмически-нормальном виде исследуемого распределения, установим связь между параметрами распределения логарифма заработной платы $\ln x$ и параметрами распределения заработной платы x.

Пусть a и σ — параметры распределения логарифма заработной платы $\ln x$. Найдем выражения для m и D_x^2 — средней и дисперсии распределения заработной платы. Как было указано выше, логарифм лог-нормальной случайной величины распределен нормально, и его функция плотности есть

$$\bar{\varphi}(\ln x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(\ln x - a)^2/\sigma^2}, \tag{12}$$

тогда функция плотности распределения x есть

$$\varphi(x) = \bar{\varphi}(\ln x) \cdot (\ln x)' = \frac{1}{\sqrt{2\pi}\,\sigma x} e^{-(\ln x - a)^2/2\sigma^2}.$$

Отсюда математическое ожидание распределения заработной платы

$$\begin{array}{l|l}
\ln x = u & \\
du = \frac{1}{x}\,dx & m = \int\limits_0^\infty x\varphi(x)\,dx = \frac{1}{\sqrt{2\pi}\,\sigma}\int\limits_0^\infty e^{-(\ln x - a)^2/2\sigma^2}\,dx = \\
u|_{x=0} = -\infty & \\
u|_{x=\infty} = +\infty & = \frac{1}{\sqrt{2\pi}\,\sigma}\int\limits_{-\infty}^{+\infty} e^u \cdot e^{-(u-a)^2/2\sigma^2}\,du = \\
v = \frac{u-a}{\sigma} & \\
dv = \frac{1}{\sigma}\,du & = \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{+\infty} e^{\sigma v + a - (v^2/2)}\,dv = e^{a+\sigma^2/2}. \\
u = \sigma v + a &
\end{array} \qquad (13)$$

И далее, дисперсия распределения по заработной плате

$$D_x^2 = \int\limits_0^\infty (x-m)^2\varphi(x)\,dx = \frac{1}{\sqrt{2\pi}\,\sigma}\int\limits_0^\infty \frac{(x-m)^2}{x} e^{-(\ln x - a)^2/2\sigma^2}\,dx =$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma}\int\limits_{-\infty}^{+\infty} (e^u - m)^2\, e^{-(u-a)^2/2\sigma^2}\,du =$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma}\int\limits_{-\infty}^{+\infty} (e^{2u} - 2me^u + m^2)\, e^{-(u-a)^2/2\sigma^2}\,du = e^{2(a+\sigma^2)} - m^2. \qquad (14)$$

Этот результат получаем, делая те же замены переменных, что и при выводе формулы (13).

Введение нового минимума заработной платы влечет за собой сдвиг всего распределения рабочих и служащих по заработной плате. Изменение его естественно рассматривать как преобразование шкалы иксов: работник, получавший ранее заработную плату x, будет получать заработную плату $\bar{x} = f(x)$. При этом доля работников $P(x)$ с заработной платой меньше, чем x, будет равна доле работников $\bar{P}(\bar{x})$ с заработной платой меньше, чем x: $P(x) = \bar{P}(\bar{x})$.

Если при этом оба распределения оказываются логарифмически-нормальными, т. е.

$$P(x) = \Phi\left(\frac{\ln x - a}{\sigma}\right), \qquad \bar{P}(\bar{x}) = \Phi\left(\frac{\ln \bar{x} - \bar{a}}{\bar{\sigma}}\right),$$

то в силу однозначности функции нормального распределения аргументы равны, т. е. $\frac{\ln x - a}{\sigma} = \frac{\ln \bar{x} - \bar{a}}{\bar{\sigma}}$, а следовательно, $\ln x$ и $\ln \bar{x}$ связаны линейной зависимостью *.

* Теоретическое обоснование того факта, что при изменении минимума заработной платы общий вид распределения не изменяется, но логарифмы значений заработной платы претерпевают линейное изменение, разработанное в лаборатории математических методов НИИ труда и описано в [3].

Как показали исследования НИИ труда, предположение о линейной зависимости между логарифмами аргумента (заработной платы) при прогнозах распределения оказывается приемлемым (причем даже в тех случаях, когда гипотеза лог-нормальности распределения не дает требуемой точности).

Определим вид линейной зависимости между $\ln x$ и $\ln \bar{x}$. Производится такое преобразование шкалы заработной платы, что $x_{\min} \to \bar{x}_{\min}$, $x_{\max} \to \bar{x}_{\max}$. Тогда преобразование шкалы логарифмов ($\ln x$) может быть записано в виде:

$$\frac{\ln x - \ln x_{\min}}{\ln x_{\max} - \ln x_{\min}} = \frac{\ln \bar{x} - \ln \bar{x}_{\min}}{\ln \bar{x}_{\max} - \ln \bar{x}_{\min}} \tag{15}$$

или $\ln \bar{x} = A \ln x + B$, где $A = \dfrac{\ln \bar{x}_{\max} - \ln \bar{x}_{\min}}{\ln x_{\max} - \ln x_{\min}}$,

$$B = \ln \bar{x}_{\min} - A \ln x_{\min}.$$

Как при этом изменятся параметры a и σ распределения $\ln x$? По известным законам теории вероятностей

$$\begin{gathered} \bar{a} = M \ln \bar{x} = M(A \ln x + B) = AM \ln x + B = Aa + B, \\ \bar{\sigma}^2 = D^2 \ln \bar{x} = D^2(A \ln x + B) = A^2 D^2(\ln x) = A^2\sigma^2, \end{gathered} \tag{16}$$

где $\bar{a}$ и $\bar{\sigma}$ — параметры распределения $\ln \bar{x}$. Соответственно изменяются параметры распределения заработной платы: $\bar{m} = e^{\bar{a} + \bar{\sigma}^2/2}$

$$\bar{D}^2_{\bar{x}} = e^{2(\bar{a}+\bar{\sigma}^2)} - \bar{m}^2, \tag{17}$$

где $\bar{m}$ и $\bar{D}_x{}^2$ — новые параметры распределения $\bar{x}$.

Пусть надо определить, как изменится средняя заработная плата (а следовательно, и фонд заработной платы) при введении нового минимума в некоторой отрасли или в целом по народному хозяйству, т. е. выразить значение средней заработной платы через $\bar{x}_{\min}$. Подставим выражения для $\bar{a}$ и $\bar{\sigma}^2$ через $\ln \bar{x}_{\max}$, $\ln \bar{x}_{\min}$ в формулу для $\bar{m}$

$$\bar{m} = \exp\left\{\bar{a} + \frac{\bar{\sigma}^2}{2}\right\} = \exp\left\{Aa + B + \frac{A^2\sigma^2}{2}\right\} =$$

$$= \exp\left\{a\frac{\ln \bar{x}_{\max} - \ln \bar{x}_{\min}}{\ln x_{\max} - \ln x_{\min}} + \ln \bar{x}_{\min} - \ln x_{\min}\frac{\ln \bar{x}_{\max} - \ln \bar{x}_{\min}}{\ln x_{\max} - \ln x_{\min}} + \right.$$

$$\left. + \frac{\sigma^2}{2}\left(\frac{\ln \bar{x}_{\max} - \ln \bar{x}_{\min}}{\ln x_{\max} - \ln x_{\min}}\right)^2\right\}.$$

Сделав необходимые упрощения и приведя подобные члены, получим:

$$\bar{m} = \exp\{p_1 \ln^2 \bar{x}_{\min} + p_2 \ln \bar{x}_{\min} + p_3\}, \tag{18}$$

где

$$p_1 = \frac{\sigma^2}{2(\ln x_{\max} - \ln x_{\min})^2},$$

$$p_2 = \frac{\ln x_{\max} - a}{\ln x_{\max} - \ln x_{\min}} - \frac{\sigma^2 \ln \bar{x}_{\max}}{(\ln x_{\max} - \ln x_{\min})^2},$$

$$p_3 = \frac{\sigma^2 \ln^2 \bar{x}_{\max}}{2(\ln x_{\max} - \ln x_{\min})^2} - \frac{\ln \bar{x}_{\max}(a - \ln x_{\min})}{\ln x_{\max} - \ln x_{\min}}$$

Аналогично можно получить выражение для новой дисперсии распределения $\bar{x}$ через новый $\bar{x}_{\min}$

$$\overline{D}^2_{\bar{x}} = \exp 2\cdot \{p_1' \ln^2 \bar{x}_{\min} + p_2' \ln \bar{x}_{\min} + p_3'\}, \tag{19}$$

$$p_1' = 2p_1,$$

$$p_2' = \frac{\ln x_{\max} - a}{\ln x_{\max} - \ln x_{\min}} - \frac{2\sigma^2 \ln \bar{x}_{\max}}{(\ln x_{\max} - \ln x_{\min})^2},$$

$$p_3' = \frac{\sigma^2 \ln^2 \bar{x}_{\max}}{(\ln x_{\max} - \ln x_{\min})^2} + \frac{\ln \bar{x}_{\max}(a - \ln x_{\min})}{\ln x_{\max} - \ln x_{\min}}.$$

Для работающих, получающих заработную плату ниже некоторого интересующего нас уровня T, определяется по формуле:

$$F(T) = \frac{1}{\sqrt{2\pi}\,\sigma}\int_0^T \frac{1}{x} e^{-(\ln x - a)^2/2\sigma^2}\,dx = \frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\ln T} e^{-(u-a)^2/2\sigma^2}\,du = \Phi\left(\frac{\ln T - a}{\sigma}\right)$$

Определим, как меняется доля работающих, заработная плата которых ниже уровня T, при введении нового минимума.

При новом минимуме $a \to \bar{a}$, $\sigma \to \bar{\sigma}$. Подставим вместо $\bar{a}$ и $\bar{\sigma}$ их выражения через $\ln \bar{x}_{\min}$. Тогда:

$$\frac{\ln T - \bar{a}}{\bar{\sigma}} = \frac{\ln T - Aa - B}{A\sigma} = \frac{\ln T - Aa - \ln \bar{x}_{\min} + A \ln x_{\min}}{A\sigma} =$$

$$= \frac{\ln T - \ln \bar{x}_{\min}}{A\sigma} - \frac{a - \ln x_{\min}}{\sigma} = \frac{\ln T - \ln \bar{x}_{\min}}{\ln \bar{x}_{\max} - \ln \bar{x}_{\min}} k - r,$$

где

$$k = \frac{\ln x_{\max} - \ln x_{\min}}{\sigma}, \qquad r = \frac{a - \ln x_{\min}}{\sigma},$$

тогда

$$P(T) = \Phi\left(\frac{\ln T - \ln \bar{x}_{\min}}{\ln \bar{x}_{\max} - \ln \bar{x}_{\min}} k - r\right). \tag{20}$$

С помощью формулы (20) можно решать обратную задачу. Пусть нас интересует верхний предел интервала заработной платы t, в котором находится определенный процент низкооплачиваемых рабочих и служащих. Естественно, что t будет меняться вместе с изменением $\bar{x}_{\min}$.

Обозначим $\dfrac{\ln t - \ln \bar{x}_{\min}}{\ln \bar{x}_{\max} - \ln \bar{x}_{\min}} k - r = u$. По таблицам функции нормального распределения найдем то значение $\bar{u}_N$, при котором $\Phi(\bar{u}_N) = \dfrac{N}{100}$.

Решая уравнение $\dfrac{\ln t - \ln \bar{x}_{\min}}{\ln \bar{x}_{\max} - \ln \bar{x}_{\min}} \cdot k - r = \bar{u}_N$ относительно $\ln t$, получим:

$$\ln t = \left(1 - \frac{\bar{u}_N + r}{k}\right) \ln \bar{x}_{\min} + \frac{\bar{u}_N + r}{k} \ln \bar{x}_{\max} = c_1 \ln \bar{x}_{\min} + c_2,$$

$$c_1 = -\frac{\bar{u}_N + r}{k}, \qquad c_2 = \frac{\bar{u}_N + r}{k} \ln \bar{x}_{\max},$$

$$t = \exp\{c_1 \ln \bar{x}_{\min} + c_2\}. \tag{21}$$

Итак, средняя заработная плата m, доля работающих p, получающих заработную плату ниже некоторого уровня T, и уровень заработной платы t, ниже которого получает N процентов работающих, представляются в виде известных функций полиномов первого и второго порядка от $\ln \bar{x}_{\min}$ (17), (20), (21). В небольших областях изменения аргумента эти функции обычно хорошо приближаются линейными функциями.

Проведенные в НИИ труда по данным формулам расчеты по нескольким отраслям для ряда значений $\bar{x}_{\min}$ подтверждают это предположение. Для каждой отрасли точки с координатами $\bar{x}_{\min}$, m легли на одну прямую и, кроме того, значения m, рассчитанные по формуле (17), дали хорошее совпадение со значениями средней заработной платы, определенными по эмпирическим данным за предыдущие годы (см. табл. 2) *. Точно так же и функция

$$t = \exp\{c_1 \ln \bar{x}_{\min} + c_2\}$$

очень близка к линейной функции от $\bar{x}_{\min}$.

Это дает возможность представить m и t в виде линейных функций от $\bar{x}_{\min}$:

$$\begin{aligned} m &= \beta_1 + \alpha_1 x_{\min}, \\ t &= \beta_2 + \alpha_2 x_{\min}. \end{aligned} \tag{22}$$

Коэффициенты β_1, β_2 и α_1, α_2, можно определить по двум точкам (скажем, крайним) рассматриваемого интервала изменения $x_{\min}$. Если рассматривались значения $x_{1\,\min}, x_{2\,\min}, \ldots, x_{k\,\min}$, то

$$\begin{aligned} \alpha_1 &= \frac{m_k - m_1}{x_{k\,\min} - x_{1\,\min}}, & \beta_1 &= \frac{m_k x_{1\,\min} - m_1 x_{k\,\min}}{x_{1\,\min} - x_{k\,\min}}, \\ \alpha_2 &= \frac{p_k - p_1}{x_{k\,\min} - x_{1\,\min}}, & \beta_2 &= \frac{p_k x_{1\,\min} - p_1 x_{k\,\min}}{x_{1\,\min} - x_{k\,\min}}, \end{aligned} \tag{23}$$

где m_1, m_k, p_1, p_k, t_1, t_k — значения m, p и t, соответствующие значениям $x_{1\,\min}$, $x_{k\,\min}$.

Представление средней заработной платы m и уровня t в виде линейных функций от значения $x_{\min}$ дает удобный аппарат для прогноза и исследования поведения этих величин.

4. Аппроксимация распределения заработной платы логарифмически-нормальной функцией открывает возможность новых подходов к изучению вопроса о формировании распределения семей по душевому доходу.

Полученные простые связи между основными параметрами распределения: средней заработной платой m и минимумом заработной платы $x_{\min}$, между $x_{\min}$ и долей работающих $p(T)$, получающих ниже некоторого интересующего нас уровня заработной платы T, побуждают нас искать пути, которые позволили бы непосредственно оценить влияние изменения параметров распределения заработной платы на распределение по душевому доходу.

Как показали расчеты, проведенные в последнее время в НИИ труда (по данным для различных областей СССР), фактическое распределение по доходу хорошо приближается логарифмически-нормальной функцией распределения, отклонения накопленных частот эмпирического и теоретического распределений составляют не более 2—3% (точность получается лучшая, чем при аппроксимации распределения по заработной плате).

Поэтому один из возможных подходов к решению задачи о непосредственном переходе от распределения по заработной плате к распределению по душевому доходу состоит в непосредственном переходе от распределе-

* Значения минимума заработной платы $x_{\min}$ и средней заработной платы m могут быть использованы как параметры в общей модели народного хозяйства

ния по заработной плате к распределению семей по душевому доходу. Для этого составляется специальная таблица. Работа по составлению такой таблицы и изучению возможности ее использования ведется совместно с Лабораторией математических и статистических методов в экономике труда в НИИ труда.

Условно предположим, что имеется десять интервалов группировки по душевому доходу и десять интервалов группировки по заработной плате. Номер столбца означает соответствующий интервал для размера заработной платы, номер строки — интервал группировки по душевому доходу. Числа $n_{0,j}$ $(j = 1, \ldots, m)$, стоящие в последней строке таблицы, есть полные численности работающих, размер заработной платы которых находится в j-м интервале. Каждый из столбцов $1, \ldots, X$ представляет собой распределение работающих с указанной заработной платой, т. е. коэффициенты k_{ij} $(i = 1, \ldots, 10;\ j = 1, \ldots, 10)$ показывают долю работающих, получающих заработную плату в j-м интервале и находящихся в i-м интервале по душевому доходу. Следовательно, зная $n_{0,j}$ и коэффициенты k_{ij}, можно определить $n_{ij} = n_{0j} \cdot k_{ij}$ — численности работающих, получающих заработную плату в j-м интервале и находящихся в i-й группе по душевому доходу. Отсюда легко получить и $n_{i0} = \sum_{j=1}^{10} n_{ij}$ $(i = 1, \ldots, 10)$ — полное число работающих для каждой группы по душевому доходу. Далее, коэффициенты k_i, стоящие в первом столбце таблицы, представляют для каждой группы по душевому доходу отношение числа работающих к числу семей N_i, находящихся в этой группе душевого дохода. Поэтому с помощью этих коэффициентов можно перейти от числа работающих n_{i0} к числу семей N_i, имеющих данный душевой доход. Таким образом, если мы знаем, как изменится на плановый период распределение работающих по заработной плате, т. е. можем определить новые численности n_{0j} работающих в каждой группе по душевому доходу, то, зная матрицу описанных выше коэффициентов, можем определить новые численности n_{i0} работающих в каждой группе по душевому доходу, а затем по коэффициентам k_i определить численности семей N_i, находящиеся в каждой группе по доходу, или, другими словами, перейти к распределению семей по душевому доходу.

В заключение я хочу поблагодарить В. А. Волконского, предложившего тему данной работы, за постоянное внимание при ее выполнении, а также Н. М. Римашевскую за помощь при проведении всех необходимых расчетов.

ЛИТЕРАТУРА

1. I. Aitchison, G. A. C. Broun. The Lognormal Distribution. Cambridge, 1957.
2. Г. Крамер. Математические методы статистики. М., Изд-во иностр. лит., 1948.
3. Н. Е. Рабкина, Н. М. Римашевская. Дифференциация заработной платы и ее прогнозирование. Экономика и матем. методы, 1965, № 6.

Поступила в редакцию
12 VI 1965