Таким образом, при сохранении общего механизма взаимодействия сил, формируется новый уровень социального обслуживания, характеризуемый иной пропорцией между возмездными и безвозмездными услугами. Эта пропорция может трансформироваться далее, но любое сокращение бюджетной части приводит к еще большему уменьшению объема социальных услуг.

Следовательно, в качестве гипотезы можно предположить, что увеличение бюджетной части само по себе должно приводить к адекватному увеличению объема социальных услуг. Данное предположение следует признать справедливым, потому что финансовая компонента в обеспечении возможности предоставления социальных услуг играет основную роль. Принципиальное увеличение объема социальных услуг за счет фактора финансирования действует до определенного момента, когда дальнейшее увеличение бюджетных расходов на социальные услуги приводит к стабилизации их общего объема за счет сокращения безвозмездных и возмездных небюджетных услуг. Это происходит в силу ограничения показателя R_V теми же причинами, которые лимитируют сферу социального обслуживания в целом[8]. Тем не менее возможность проведения анализа взаимодействия социального обслуживания и форм его финансирования не должна быть проигнорирована. По мнению автора, очередными этапами такого анализа должны стать: разработка целей эволюции системы бюджетного финансирования социальных услуг[9], сопоставление их между собой и выработка алгоритма оптимизации усилий (под которыми, в данном случае, понимается расходование финансовых ресурсов) и результата (удовлетворенность социальной услугой).

Сергей Александрович БАРДАСОВ — старший преподаватель кафедры статистики и информационных технологий финансового факультета, кандидат физико-математических наук

УДК 519.21(075.8)

К ВОПРОСУ О МЕТОДЕ ГРУППИРОВОК В ЭКОНОМИЧЕСКОЙ СТАТИСТИКЕ

АННОТАЦИЯ. Аналитическая группировка и ступенчатая функция, полученная методом наименьших квадратов, дают равные значения индекса корреляции.

Стандартное отклонение, размах вариации, число элементов и отношение средней из групповых дисперсий к общей дисперсии используются для оценки числа групп вариационного ряда. На основе этой оценки предложена формула для расчета числа групп. Эта формула, в отличие от формулы Стерджесса, применима не только для нормального распределения.

Analytical grouping and step-function received by a method of the least squares give equal values of an index correlation.

A standard deviation, amplitude of a variation, number of elements and ratio of average of group dispersions to general dispersions are used for estimate of number of groups of a variational row. This formula, as against the Sturgess' formula, is applicable not only for normal distribution.

---

[8] Эти причины следует искать в области условий и обстоятельств, определяющих пропорциональность развития общества, которое не может быть обеспечено только лишь изменением источника финансирования.

[9] Автор предполагает рассмотреть подходы к решению этой масштабной проблемы в одном из следующих выпусков.

Группировка является одним из основных и наиболее распространенных методов обработки и анализа первичной статистической информации. Для исследования взаимосвязи между экономическими показателями широко используется аналитическая (факторная) группировка. По результатам этой группировки в качестве показателя степени тесноты связи между результативным и факторным признаками рассчитывается эмпирическое корреляционное отношение. Теоретическое корреляционное отношение (индекс корреляции) рассчитывается и на основе уравнения регрессии, найденного методом наименьших квадратов. График уравнения регрессии представляют, как правило, в виде непрерывной линии . Эти методы определения корреляционного отношения рассматривают как различные. Исследуем связь между ними.

При использовании компьютеров для обработки статистических данных часто необходимо определить число групп, на которые будет разбита исследуемая совокупность элементов. Известна только одна аналитическая формула, позволяющая оценить число групп. Для этого используют известную формулу Стерджесса, которая дает хорошие результаты, если совокупность состоит из большого числа элементов и их распределение по признаку, положенному в основание группировки, близко к нормальному. Поставим задачу: оценить число групп для распределений, отличающихся от нормального.

Вначале рассмотрим аналитическую группировку как средство изучения взаимосвязи двух признаков.

Пусть известны значения факторного признака

$$x_i,\ \ i=1,2,\dots,n$$

и соответствующие им значения результативного признака

$$y_i,\ \ i=1,2,\dots,n.$$

Пусть $x_{\min}, x_{\max}$ — минимальное и максимальное значения факторного признака.

Отрезок $\left[x_{\min}, x_{\max}\right]$ разобьем на m равных (для простоты) интервалов длиной $\Delta=\left(x_{\max}-x_{\min}\right)/m$.

Определим на интервалах следующие функции

$$\theta_j(x)=\begin{cases}1 & \text{при} \begin{cases} x\in\left[x_{\min}+\Delta\cdot(j-1), x_{\min}+\Delta\cdot j\right), & j=1,\dots,m-1,\\ x\in\left[x_{\min}+\Delta\cdot(j-1), x_{\min}+\Delta\cdot j\right], & j=m\end{cases}\\ 0 & \text{при} \begin{cases} x\notin\left[x_{\min}+\Delta\cdot(j-1), x_{\min}+\Delta\cdot j\right), & j=1,\dots,m-1,\\ x\notin\left[x_{\min}+\Delta\cdot(j-1), x_{\min}+\Delta\cdot j\right], & j=m.\end{cases}\end{cases}$$

Очевидно, что

$$\theta_j(x)\cdot\theta_k(x)=\theta_j(x), \quad \text{при} \quad j=k,$$

$$\theta_j(x)\cdot\theta_k(x)=0, \quad \text{при} \quad j\neq k.$$

Уравнение регрессии будем искать в виде

$$\hat{y}(x)=\sum_{j=1}^{m}a_j\cdot\theta_j(x) \ , \tag{1}$$

где $a_j(j=1,\dots,m)$ — неизвестные постоянные параметры. Из определения функции $\hat{y}(x)$ очевидно, что на каждом j-м интервале ее значение остается постоянным и равным a_j.

Для нахождения неизвестных параметров используем метод наименьших квадратов. Критерий этого метода можно записать таким образом:

$$S=\sum_{i=1}^{n}\left(y_i-\hat{y}(x_i)\right)^2 \ \to\min$$

или, поскольку $\hat{y}(x)=\sum_{j=1}^{m} a_j \cdot \theta_j(x)$,

$$S=\sum_{i=1}^{n}\left(y_i-\sum_{j=1}^{m} a_j \cdot \theta_j(x_i)\right)^2 \to \min.$$

Функция m переменных $S(a_1, a_2, \ldots, a_m)$ может достичь экстремума в том случае, когда первые частные производные этой функции равняются нулю.

Найдем производные функции S по параметрам a_k:

$$\frac{\partial S}{\partial a_k}=-2\sum_{i=1}^{n}\left(\left(y_i-\sum_{j=1}^{m} a_j \cdot \theta_j(x_i)\right) \cdot \theta_k(x_i)\right), \quad k=1,\ldots,m.$$

Или , поскольку $\theta_j(x_i) \cdot \theta_k(x_i)=\theta_k(x_i)$, при $j=k$,

$$\frac{\partial S}{\partial a_k}=-2\sum_{i=1}^{n}\left(y_i \cdot \theta_k(x_i)-a_k \cdot \theta_k(x_i)\right), \quad k=1,\ldots,m .$$

Для нахождения неизвестных параметров a_k приравняем производные к нулю (символ k заменим на j):

$$\sum_{i=1}^{n}\left(y_i \cdot \theta_j(x_i)-a_j \cdot \theta_j(x_i)\right)=0, \quad j=1,\ldots,m .$$

Так как $\theta_j(x_i) \neq 0$ только для значений x_i, входящих в j-ю группу (j-й интервал), то $\sum_{i=1}^{n} y_i \cdot \theta_j(x_i)=\bar{y}_j \cdot n_j$, где $\bar{y}_j$ — среднее значение результативного признака по j-й группе, n_j — количество элементов совокупности в j-й группе. Аналогично

$$\sum_{i=1}^{n} a_j \cdot \theta_j(x_i)=a_j\sum_{i=1}^{n}\theta_j(x_i)=a_j \cdot n_j, \quad j=1,\ldots,m .$$

Следовательно,

$$a_j n_j=\bar{y}_j n_j, \quad j=1,2,\ldots,m$$

Таким образом, уравнение регрессии имеет вид:

$$\hat{y}(x)=\sum_{j=1}^{m} \bar{y}_j \cdot \theta_j(x). \qquad (2)$$

Это ступенчатая функция , ее значение на каждом j-м интервале постоянно и равно среднему значению результативного признака $\bar{y}_j$ по j-й группе.

Рассмотрим теоретическое корреляционное отношение (индекс корреляции):

$$\eta=\sqrt{\frac{\delta^2}{\sigma^2}} ,$$

где δ^2 — дисперсия выравненных значений результативного признака, то есть рассчитанных по уравнению регрессии;

σ^2 — дисперсия эмпирических (фактических) значений результативного признака.

После подстановки дисперсий получим:

$$\eta=\sqrt{\frac{\sum_{i=1}^{n}\left(\hat{y}(x_i)-\bar{y}\right)^2}{\sum_{i=1}^{n}\left(y_i-\bar{y}\right)^2}}=\sqrt{\frac{\sum_{i=1}^{n}\left(\bar{y}_j \cdot \theta_j(x_i)-\bar{y}\right)^2}{\sum_{i=1}^{n}\left(y_i-\bar{y}\right)^2}}.$$

В силу свойств функции $\theta_j(x_i)$ в числителе подкоренного выражения из среднего значения $\bar{y}_j$ результативного признака по каждой j-й группе вычитается общее среднее значение $\bar{y}$ результативного признака столько раз, сколько эмпирических значений факторного и, соответственно, результативного признака содержится в j-й группе. Следовательно,

$$\eta=\sqrt{\frac{\sum_{j=1}^{m}\left(\bar{y}_j-\bar{y}\right)^2\cdot n_j}{\sum_{i=1}^{n}\left(y_i-\bar{y}\right)^2}}, \tag{3}$$

где $\sum_{j=1}^{m} n_j=n$, n_j — количество эмпирических значений результативного признака, попавших в j-ю группу. Равенство (3) есть эмпирическое корреляционное отношение, которое находят методом аналитической группировки.

Таким образом, для уравнения регрессии (2) теоретическое корреляционное отношение (индекс корреляции) совпадает с эмпирическим. Ступенчатая функция, найденная методом наименьших квадратов, дает те же результаты, что и аналитическая группировка. Очевидно, что этот вывод справедлив для аналитической группировки в случае, когда имеется один факторный признак и несколько результативных, так как при этом зависимости между факторным и результативными признаками рассматриваются независимо. Легко показать, что рассмотренный метод можно распространить на случай, когда имеется несколько количественных факторных признаков.

Рассмотрим теперь оценку числа групп вариационного ряда.

Часто используют следующую формулу Стерджесса для определения оптимального числа групп:

$$n=1+\log_2 N\approx 1+3{,}322\cdot\lg N\ , \tag{4}$$

где N — количество элементов совокупности, n — число групп.

Результат округляют до целого числа, так как число групп не может быть дробным числом. Формула (4) получена для биномиального распределения, которое при большом N близко к нормальному. Следовательно, она дает хорошие результаты, если совокупность состоит из большого числа единиц и их распределение близко к нормальному закону.

Группировка по количественному признаку имеет задачу отразить распределение единиц совокупности по этому признаку. Чем больше групп, тем точнее будет воспроизведен характер исследуемого объекта. Однако слишком большое число групп затрудняет выявление закономерностей при исследовании социально-экономических явлений и процессов. Часто полагают, что внутригрупповая дисперсия должна составлять не более чем 0,25 или 0,3 от общей дисперсии. Если задать отношение средней из внутригрупповых дисперсий к общей дисперсии, то можно оценить число групп, на которые разбиваем исследуемую совокупность.

Положим правило сложения дисперсий в основу определения количества групп. Согласно этому правилу, общая дисперсия σ_0^2 равна сумме средней из внутригрупповых $\bar{\sigma}_j^2$ и межгрупповой δ^2 дисперсий:

$$\sigma_0^2=\bar{\sigma}_j^2+\delta^2 \tag{5}$$

Межгрупповая дисперсия δ^2 характеризует систематическую вариацию, то есть часть вариации, обусловленную влиянием группировочного признака. Средняя из внутригрупповых дисперсий $\bar{\sigma}_j^2$ отражает случайную вариацию, то есть часть вариации, происходящую под влиянием неучтенных факторов и не зависящую от группировочного признака. Общая дисперсия σ_0^2 отражает вариацию признака за счет всех условий и причин, действующих в совокупности.

Разобьем исследуемую совокупность на группы с равной длиной интервала так, чтобы средняя из внутригрупповых дисперсий не превышала некоторую долю α общей дисперсии. Тогда доля межгрупповой дисперсии, согласно равенству (5), будет равна $1-\alpha$. Очевидно, что для отражения влияния группировочного признака необходимо соблюдать условие $(1-\alpha)>\alpha$. Итак, имеем:

$$\frac{\bar{\sigma}_j^2}{\sigma_0^2}=\alpha \quad . \tag{6}$$

Общую дисперсию σ_0^2 легко вычислить по имеющимся данным, но средняя из внутригрупповых дисперсий $\bar{\sigma}_j^2$ зависит от того, на сколько групп разбита изучаемая совокупность.

Пусть $x_{\max}, x_{\min}$ — максимальное и минимальное значения элементов совокупности. Тогда, при разбиении совокупности на n групп, длина интервалов будет равна

$$\frac{x_{\max}-x_{\min}}{n} .$$

Дисперсия σ_j^2 каждой j-й группы не превосходит значения:

$$\left(\frac{x_{\max}-x_{\min}}{2n}\right)^2 .$$

Следовательно и средняя из внутригрупповых дисперсий $\bar{\sigma}_j^2$ не превосходит этого значения:

$$\bar{\sigma}_j^2 \le \left(\frac{x_{\max}-x_{\min}}{2n}\right)^2 \tag{7}$$

Из соотношений (6) и (7) получим:

$$\left(\frac{x_{\max}-x_{\min}}{2n}\right)^2 \ge \alpha\sigma_0^2 \cdot$$

Тогда для количества групп получим следующую оценку:

$$n \le \frac{x_{\max}-x_{\min}}{2\sqrt{\alpha}\,\sigma_0} \quad . \tag{8}$$

Формула Стерджесса (4) дает хорошие результаты для нормального распределения, поэтому используем ее для оценки параметра α. Для нормального распределения $x_{\max}-x_{\min}\approx 6\sigma_0$, тогда из (4) и (8) получим:

$$1/\sqrt{\alpha}=(1+3{,}322\bullet\lg N)/3 \cdot$$

Тогда для оценки числа групп получим формулу:

$$n=\frac{x_{\max}-x_{\min}}{6\sigma_0}(1+3{,}322\bullet\lg N) \quad \cdot \tag{9}$$

Оценим число групп, применяя формулу (8). Как уже отмечалось выше, часто полагают, что отношение групповой дисперсии к общей не превосходит значения $\alpha=0{,}25$. Подставляя это значения в (8), получим:

$$n \le \frac{x_{\max}-x_{\min}}{\sigma_0} \tag{10}$$

Применим (10) для нормального распределения. Тогда получим $n\approx 6$.

Отличительной особенностью формул (8) и (9) является то, что число групп оценивается исходя из отношения двух показателей вариации — размаха вариации и среднего квадратического отклонения. Эти формулы, в отличие от формулы Стерджесса, учитывают внутреннюю структуру распределения. В учебниках часто утверждают: чем больше вариация признака, тем на большее число групп необходимо делить изучаемую совокупность. При этом совершенно не учитывают вид распределения. Согласно соотношениям (8), (9), при одном и том же размахе вариации $(x_{max}-x_{\min})$ распределение с большей дисперсией необходимо делить на меньшее число групп. Рассмотрим, например, равномерное распределение. Если его разбить на группы, то в каждой группе будет одинаковое число элементов. То есть различия между группами не будет. Поэтому нет необходимости делить такое распределение на большое количество групп. Подтвердим эти рассуждения. Для этого применим формулу (10) для равномерного распределения. Как известно, в этом случае $\sigma_0 = (x_{max} - x_{min})/\sqrt{12}$. Тогда $n \approx \sqrt{12} \approx 3$. Это в два раза меньше результата, который дает эта же формула для нормального распределения.

Формула (9) дает тот же результат, что и формула Стерджесса, для распределений, близких к нормальному. В остальных случаях она дает результат больший или меньший в $(x_{max} - x_{min}) / (6\sigma_0)$ раз по сравнению с формулой (4). Все рассуждения, как и формула Стерджесса, справедливы для одномодальных распределений.

ЛИТЕРАТУРА

1. Ефимова М. Р., Петрова Е. В., Румянцев В. Н. Общая теория статистики. М.: ИНФРА-М, 1996. 416 с.

2. Елисеева И. И., Юзбашев М. М. Общая теория статистики. М.: Финансы и статистика, 1996. 368 с.

3. Теория статистики / Под ред. проф. Р. А. Шмойловой М.: Финансы и статистика. 1996. 464 с.

4. Статистика / Под ред. В. Г. Ионина. Новосибирск: Изд-во НГАЭиУ, 1996. 310 с.

5. Венецкий И. Г., Венецкая В. И. Основные математико-статистические понятия и формулы в экономическом анализе. М.: Статистика, 1979. 447 с.

Анна Николаевна ЕРМАКОВА —
старший преподаватель кафедры
статистики и информационных технологий
финансового факультета,
кандидат социологических наук

УДК 550.812

ОБЩАЯ ПОСТАНОВКА ЗАДАЧ ПОСТРОЕНИЯ КРИТЕРИЕВ УПРАВЛЕНИЯ НЕДРОПОЛЬЗОВАНИЕМ И ИХ МАТЕМАТИЧЕСКОЕ ОБОСНОВАНИЕ

АННТОТАЦИЯ. В работе приведено математическое обоснование критериев управления недропользованием. Дана классификация критериев в зависимости от шкал их измерения. Рассмотрены задачи, решаемые для разных классов критериев. Проведен анализ современного состояния постановки задач каждого типа. Разработана минимальная теоретическая база для постановки и решения задач построения критериев на основе отношения стратегической эквивалентности.

In work the mathematical substantation of criteria of a management the using of bowels of the earth is indicated. Classification of criteria depending on scales of their measurement is given. Problems, soluble for different classes of criteria are considered. The analysis of a modern condition of statement of problems of each type is conducted. Minimum theoretical base for statement and decision of problems of construction of criteria on the basis of relation of a strategic equivalence is developed.