Теоретико-статистическое моделирование использования экспериментальных данных с целью построения математически интерпретированных взаимосвязей динамических элементов мезо- и макроуровней современной экономики

Theoretical and Statistical Modeling of Experimental Data Application for the Development of Mathematically Interpreted Interconnection of the Dynamic Elements of the Meso- and Macro-levels of the Modern Economy

УДК 330.4

Воронов Александр Андреевич
доцент Санкт-Петербургского государственного технологического института (технического университета), кандидат экономических наук
190013, Санкт-Петербург, Московский пр., д. 26

Aleksandr A. Voronov
St. Petersburg State Technological Institute (Technical University)
Moskovskiy Ave 26, St. Petersburg, Russian Federation, 190013

В статье описаны экономико-математические предпосылки организации экспериментов с целью систематизации эмпирических данных, которые могут быть заложены в математическую интерпретацию моделей как внутренних взаимосвязей между динамическими элементами экономических систем различных уровней, так и при моделировании внешних воздействий на всю систему национальной экономики, что в конечном счете позволит построить численную реализацию решения задачи динамической устойчивости механизма национальной экономики под воздействием внешних факторов стохастической природы.

Цель. Представить обоснованный с позиций теоретической статистики аппарат обобщения дискретных экспериментальных данных, изученных в форме одного и нескольких факторов, как независимых, нормально распределенных случайных величин с одинаковыми дисперсиями, который может быть использован при построении экономико-математических моделей внутренних взаимосвязей в сложных динамических экономических системах в сочетании с возможностью проверки устанавливаемых в процессе такого моделирования статистических гипотез.

Задачи. Интерпретировать эти исследования при помощи аппарата современной математики и совокупности математических и инструментальных методов в экономике, что позволит современному экономисту-теоретику относительно быстро перейти к практическим методам и приемам получения всего спектра необходимых математических зависимостей (функций) для численной реализации решения задачи динамической устойчивости и анализа проблемы пограничных состояний механизма национальной экономики под воздействием факторов внешней стохастической природы.

Методология. В методологическом отношении автор апеллирует к проблеме синтеза дискретных экспериментальных данных, теоретико-статистическая интерпретация которых формирует базу, необходимую для относительно точных математических выводов функциональных зависимостей, находящихся в основаниях математической теории устойчивости и переходных процессов в сложных многоуровневых иерархических системах.

Результаты. В процессе исследования автор приходит к тому, что процесс организации и планирования экспериментов в теоретико-статистическом плане является предшествующим этапом, необходимым для получения статистически значимых совокупностей методов и приемов, при помощи которых конкретизируются, уточняются, проверяются статистические гипотезы, являющиеся фундаментом при выводе спектральных функций, моделирующих всю систему внутренних и внешних факторов, обеспечивая численное решение задачи динамической устойчивости сложной и многоуровневой национальной экономической системы.

Выводы. В результате проведенного исследования обобщены задачи, состоящие в проверке гипотезы о равенстве средних значений нескольких нормально распределенных случайных величин с одинаковой дисперсией для случаев учета одного и многих факторов на всю систему в целом; показано, что можно проверять справедливость гипотез о равенстве средних значений многих факторов при помощи закона Беренса–Фишера–Снедекора. Величина несмещенной оценки дисперсии сравнивается с оценкой, найденной по закону Беренса–Фишера–Снедекора, что позволяет принимать или отвергать гипотезу в зависимости от того, существенно или нет она отклоняется от единицы отношения последней оценки к оценке остаточной дисперсии. На основании этой информации можно строить корректные функциональные зависимости, связывающие в определенных допущениях управляющие и управляемые параметры экономических систем, даже при наличии условий неопределенности.

Ключевые слова: экспериментальные данные, дискретные случайные величины, функция распределения, закон Беренса–Фишера–Снедекора, несмещенная оценка, проверка статистических гипотез, экономическая интерпретация статистической информации, динамическая устойчивость сложных многоуровневых экономических систем.

This article provides an economic and mathematical background for experiments aimed at systematizing empirical data that can serve as the basis for the mathematical interpretation of models of internal interconnection between dynamic elements of economic systems of different levels, and for the modeling of external influences on an entire system of a national economy. This would eventually allow the development of a numerical implementation of the solution to the problem of dynamic stability of a national economy mechanism under the influence of external stochastic factors.

Aim. This study aims to propose an apparatus substantiated from the perspective of theoretical statistics for the integration of discrete experimental data examined in the form of one or several factors as independent, normally distributed random values, with possible application of the approach in the development of internal interconnection of complex dynamic economic systems, combined with the possibility of testing the statistical hypotheses determined during such modeling.

Tasks. This study interprets the aforementioned research using the apparatus of modern mathematics and a combination of mathematical and instrumental methods in economics. This allows a modern economic theorist to make a swift transition to practical methods and techniques for obtaining the entire spectrum of the necessary mathematical dependencies (functions) for the numerical implementation of the solution to the problem of dynamic stability and for analysis of the problem of boundary conditions of the entire mechanism of a national economy under the influence of external stochastic factors.

Methods. In terms of methodology, the author appeals to the problem of synthesis of discrete experimental data, the theoretical and statistical interpretation of which forms the basis for relatively accurate mathematical determination of functional relationships that serve as the basis for the mathematical theory of stability and transition processes in complex multilevel hierarchy systems.

Results. During the study, the author concludes that the process of organizing and planning experiments is a predecessor step in terms of theory and statistics. This step is essential to obtain statistically significant combinations of methods and techniques for the concretization, specification, and testing of statistical hypotheses that are fundamental to the determination of spectral functions for modeling the entire system of internal and external factors that provide a numerical solution to the problem of dynamic stability of the complex and multilevel system of national economy.

Conclusion. As a result, the study summarizes the problems of testing the hypothesis of equality of mean values of multiple normally distributed random values with equal variance for the cases, considering the influence of one or multiple factors on the entire system. The study also shows that the hypotheses of equality of mean values of multiple factors can be tested using the Fisher–Snedecor distribution. The unbiased variance estimate is compared to the variance determined using the Fisher–Snedecor distribution, which allows the author to accept or reject the hypothesis depending on the significance of its deviation from the ratio between the last estimate to the residual variance estimate. Using this data, it is possible to plot valid functional dependencies linking, with certain assumptions, the control and controlled parameters of economic systems even under conditions of uncertainty.
Keywords: experimental data, discrete random values, distribution function, Fisher–Snedecor distribution, unbiased estimator, statistical hypothesis testing, economic interpretation of statistical information, dynamic stability of complex multilevel economic systems

Многие современные экономисты используют различные по своей качественной структуре и уровню математической приспособленности модели исследования динамических экономических систем, такие как модель Солоу, модель Кейнса, модель Самуэльсона–Хикса, динамическая модель Леонтьева и др. При этом экономическая система часто представляется как совокупность составляющих ее элементов и взаимосвязей между ними [1, с. 12]. Однако эффективное построение всех существующих экономико-математических моделей не может быть осуществлено без внесения, обработки, анализа и уточнения огромного массива статистических данных, отражающих реальные факторы совокупности национальных хозяйственных единиц (предприятий, организаций), а также отраслевых комплексов, объединенных производственно-технологическими и организационно-хозяйственными связями. В процессе построения названных выше моделей всегда приходится сталкиваться с множеством вероятностных и теоретико-статистических проблем, решение которых находится в области теоретической статистики. Для достижения обозначенной в данной статье цели обратимся к ряду выводов этой науки, учитывая, тем не менее, прикладной характер исследования.

1. Случай одного фактора

Для достижения поставленной цели сделаем попытку обобщить задачу Стьюдента, состоящую в проверке гипотезы о равенстве средних значений нескольких нормально распределенных величин с одинаковой дисперсией. Мы будем рассматривать одномерные независимые нормальные случайные величины с одинаковыми дисперсиями. В планировании эксперимента обычной является следующая схема. Изучается действие некоторого фактора (количественного или качественного), который принимает k различных значений, именуемых уровнями фактора. На i-м уровне производится n_i наблюдений, результаты которых записываются в следующей форме:

$$x_{1,1},\ x_{1,2},\ \ldots,\ x_{1,n_1},$$
$$\ldots\ldots\ldots\ldots\ldots,$$
$$x_{k,1},\ x_{k,2},\ \ldots,\ x_{k,n_k},$$

причем $n_1 + n_2 + \ldots + n_k = N$.

Разумеется, можно не пользоваться терминологией планирования эксперимента, а говорить об n_i наблюдаемых значениях i-й случайной величины, $i = 1, \ldots, k$.

Мы хотим проверить гипотезу о равенстве средних значений в различных группах; таким образом, мы априори выдвигаем предположение

$$\mu_1 = \mu_2 = \ldots = \mu_k = \mu.$$

Результаты наблюдений обозначены $x_{i,j}$, где i обозначает группу наблюдений или уровень фактора, а j — номер наблюдения, относящегося к этому уровню; $i = 1, 2, \ldots, k$, и при данном i

$$j = 1, 2, \ldots, n_i.$$

Элемент вероятности совокупности переменных $x_{i,j}$ имеет вид [2]:

$$A \cdot e^{-\frac{1}{2\cdot\sigma^2}\cdot\sum(x_{i,j}-\mu)^2} \cdot \prod dx_{i,j}.$$

Мы собираемся разложить сумму квадратов, стоящую в показателе, записав

$$\sum_{i,j}\left(x_{i,j}-\mu\right)^2 = \sum_{i,j}\left(x_{i,j}-m_i+m_i-m+m-\mu\right)^2,$$

где $m_i = \sum_{j=1}^{n_i}\frac{x_{i,j}}{n_i}$, $m = \frac{\sum_{i,j}x_{i,j}}{N} = \sum_{i=1}^{k}\frac{n_i \cdot m_i}{N}$.

Можно написать

$$\sum_{i,j}\left(x_{i,j}-\mu\right)^2 =$$
$$=\sum_{i,j}\left(x_{i,j}-m_i\right)^2+\sum_{i,j}\left(m_i-m\right)^2+\sum_{i,j}(m-\mu)^2,$$

что окончательно дает

$$\sum_{i,j}\left(x_{i,j}-\mu\right)^2 =$$
$$=\sum_{i,j}\left(x_{i,j}-m_i\right)^2+\sum_{i=1}^{k}n_i\cdot\left(m_i-m\right)^2+N\cdot(m-\mu)^2,$$

если привести подобные члены в двойных суммах.

Действительно, как легко проверить, все суммы попарных произведений, появляющихся при возведении в квадрат, равны 0. Рассмотрим, например,

$$\sum_{i,j}\left(m_i-m\right)\cdot(m-\mu).$$

Имеем:

$$\sum_{i,j}\left(m_i-m\right)\cdot(m-\mu)=(m-\mu)\cdot\sum_{i,j}\left(m_i-m\right)=$$
$$=(m-\mu)\cdot\sum_{i=1}^{k}n_i\cdot\left(m_i-m\right),$$

а последняя сумма равна 0 в силу определения m.

Аналогично проверяется, что и все остальные попарные произведения равны 0.

Введем

$$S'^2=\sum_{i,j}\left(x_{i,j}-m_i\right)^2,\ s'^2=\sum_{i}n_i\cdot\left(m_i-m\right)^2$$

и найдем закон распределения S' и s'.

С этой целью возьмем в качестве новых переменных m, s', S' и еще $N-3$ переменных, которые мы обозначим символически через U. В дальнейшем мы проинтегрируем по U, что даст нам закон распределения трех величин m, s', S'. В новых переменных элемент вероятности запишется в виде [3; 4]:

$$A\cdot e^{-\frac{1}{2\cdot\sigma^2}\cdot S'^2}\cdot e^{-\frac{1}{2\cdot\sigma^2}\cdot s'^2}\cdot e^{-N\cdot\frac{(m-\mu)^2}{2\cdot\sigma^2}}\times$$
$$\times\frac{D\left(x_{1,1},\ldots,x_{k,n_k}\right)}{D\left(m,s',S',U\right)}dm\,ds'dS'dU.$$

Как мы сейчас увидим, функциональный определитель равен произведению функций от m, s', S' и U.

Итак, имеют место соотношения

$$m_i=\sum_{j=1}^{n_i}\frac{x_{i,j}}{n_i},\ m=\sum_{i=1}^{k}\frac{n_i\cdot m_i}{N},$$
$$s'^2=\sum_{i}n_i\cdot\left(m_i-m\right)^2,$$
$$S'^2=\sum_{i,j}\left(x_{i,j}-m_i\right)^2.$$

Положим:

$$x_{i,j}-m_i=S'\cdot\theta_{i,j},$$

что дает следующие k соотношений для $\theta_{i,j}$:

$$\sum_{j=1}^{n_i}\theta_{i,j}=0$$

при $i=1,\ \ldots,\ k$; кроме того,

$$\sum_{i,j}\theta_{i,j}^2=1,$$

т.е. всего имеется $k+1$ соотношений.

Таким образом, мы имеем всего N переменных $\theta_{i,j}$, связанных $k+1$ соотношениями, т.е. $N-k-1$ свободных переменных.

Аналогично положим:

$$m_i-m=s'\cdot\theta_i.$$

Переменные θ_i (их число равно k) связаны двумя соотношениями

$$\sum_{i}n_i\cdot\theta_i=0,\ \sum_{i}n_i\cdot\theta_i^2=1.$$

Таким образом, из переменных θ_i мы имеем $k-2$ свободных. Следовательно, m, s', S', $N-k-1$ свободных переменных θ_{ij}, $k-2$ свободных переменных θ_i образуют совокупность N новых переменных, заменяющих исходные $x_{i,j}$.

Введение новых величин осуществляется по следующей формуле:

$$x_{i,j}=m+s'\cdot\theta_i+S'\cdot\theta_{i,j}.$$

Сразу видим, что функциональный определитель этого преобразования имеет вид [5]:

$$\frac{D\left(x_{1,1},\ldots,x_{k,n_k}\right)}{D\left(m,s',S',U\right)}=$$
$$=\begin{vmatrix}
1 & \theta_1 & \theta_{1,1} & s' & 0 & \ldots & 0 & S' & 0 & \ldots & 0\\
1 & \theta_1 & \theta_{1,2} & s' & 0 & \ldots & 0 & 0 & S' & \ldots & 0\\
. & . & . & . & . & . & . & . & . & . & .\\
1 & \theta_1 & \theta_{1,n_1} & s' & 0 & \ldots & 0 & \ldots & \ldots & \ldots & 0\\
\ldots & \theta_2 & \theta_{2,1} & 0 & s' & \ldots & 0 & \ldots & \ldots & \ldots & 0\\
. & . & . & . & . & . & . & . & . & . & .\\
1 & \theta_k & \theta_{k,n_k} & s'\cdot\frac{\partial\theta_k}{\partial\theta_1} & s'\cdot\frac{\partial\theta_k}{\partial\theta_2} & \ldots & s'\cdot\frac{\partial\theta_k}{\partial\theta_{k-2}} & \ldots & & & S'\cdot\frac{\partial\theta_{k,n_k}}{\partial\theta_{l,m}}
\end{vmatrix},$$

причем $\theta_{l,m}$ — последняя из свободных переменных $\theta_{i,j}$.

Замечаем, что s' входит общим множителем в $k-2$ столбцов, отвечающих переменным $\theta_{i,j}$. Если вынести их за знак определителя, то останется определитель, зависящий только от θ_i и $\theta_{I,\ j}$. После интегрирования элемента вероятности величин m, s', S', U по $N-3$ переменным θ_i и $\theta_{i,\ j}$ получим закон распределения трех случайных величин m, s', S'

$$A\cdot e^{-\frac{1}{2\cdot\sigma^2}\cdot S'^2}\cdot e^{-\frac{1}{2\cdot\sigma^2}\cdot S'^2}\cdot e^{-\frac{N}{2\cdot\sigma^2}\cdot(m-\mu)^2}\times$$
$$\times s'^{k-2}\cdot S'^{N-k-1}\cdot ds'dS'dm.$$

Для проверки равенства средних μ_1, μ_2, …, μ_k сравнивают s' и S'.

После интегрирования по m найдем закон распределения величин s' и S':

$$A' \cdot e^{-\frac{S'^2}{2\sigma^2}} \cdot S'^{N-k-1} \cdot dS' \cdot B \cdot e^{-\frac{s'^2}{2\cdot\sigma^2}} \cdot s'^{k-2} \cdot ds',$$

который показывает, что S' и s' независимы.

Отношение S'/s', очевидно, следует закону Беренса–Фишера–Снедекора [3, с. 424].

Предыдущее рассуждение предполагает, что средние μ_i равны; по известному свойству χ^2-распределения

$$\mathbf{E}\left(\frac{S'^2}{N-k}\right) = \sigma^2,\ \mathbf{E}\left(\frac{s'^2}{k-1}\right) = \sigma^2,$$

т. е. дисперсия σ^2 оценивается двумя независимыми способами величинами $\dfrac{S'^2}{N-k}$ и $\dfrac{s'^2}{k-1}$.

Поэтому определим

$$S^2 = \frac{S'^2}{N-k},\ s^2 = \frac{s'^2}{k-1};$$

S^2 называется остаточной дисперсией выборки, а s^2 — межгрупповой дисперсией.

Величина

$$F = \frac{S'}{s'} \cdot \sqrt{\frac{k-1}{N-k}} = \frac{S}{s}$$

подчиняется закону Беренса–Фишера–Снедекора, и по таблице этого распределения можно видеть, значимо или нет отклонение F от 1 и, следовательно, отвергнуть или принять гипотезу о равенстве средних значений m_i.

Допустим, что найденное значение F привело к отклонению гипотезы о равенстве средних значений. В этих условиях все же можно оценить дисперсию σ^2; мы увидим, что и при неравенстве средних статистика $\dfrac{S'^2}{N-k}$ служит несмещенной оценкой σ^2, чего нельзя сказать о $\dfrac{s'^2}{k-1}$ [2].

Положим:

$$x_{i,j} = \mu_i + y_{i,j},$$

где величины $y_{i,j}$ нормальны с параметрами $(0, \sigma)$. Тогда

$$m_i = \mu_i + m_i',$$

где величины m_i' также нормально распределены с параметрами $\left(0,\ \sigma/\sqrt{n_i}\right)$.

Положим, кроме того,

$$m = \mu + m',\ \mu = \frac{\sum n_i \cdot \mu_i}{N}.$$

Тогда имеем:

$$\sum_{i,j}\left(x_{i,j} - m_i\right)^2 = \sum_{i,j}\left(y_{i,j} - m_i'\right)^2$$

и, следовательно,

$$\mathbf{E}\left[\frac{\sum_{i,j}\left(x_{i,j} - m_i\right)^2}{N-k}\right] = \sigma^2.$$

С другой стороны,

$$\sum_{i=1}^{k} n_i \cdot \left[\left(m_i' + \mu_i\right) - \left(m' + \mu\right)\right]^2 =$$

$$= \sum_{i=1}^{k} n_i \cdot \left(m_i' - m'\right)^2 + \sum_{i=1}^{k} n_i \cdot \left(\mu_i - \mu\right)^2 +$$

$$+ 2 \cdot \sum_{i=1}^{k} n_i \cdot \left(m_i' - m'\right) \cdot \left(\mu_i - \mu\right),$$

откуда

$$\mathbf{E}\left[\sum_{i=1}^{k} n_i \cdot \left[\left(m_i' + \mu_i\right) - \left(m' + \mu\right)\right]^2\right] =$$

$$= \mathbf{E}\left[\sum_{i=1}^{k} n_i \cdot \left(m_i - m\right)^2\right] =$$

$$= \mathbf{E}\left[\sum_{i=1}^{k} n_i \cdot \left(m_i' - m'\right)^2\right] + \mathbf{E}\left[\sum_{i=1}^{k} n_i \cdot \left(\mu_i - \mu\right)^2\right] +$$

$$+ 2 \cdot \mathbf{E}\left[\sum_{i=1}^{k} n_i \cdot \left(m_i' - m'\right) \cdot \left(\mu_i' - \mu\right)\right].$$

Величины m_i' можно считать выборочными средними в различных группах, а m' — суммарным средним (по группам) независимых нормально распределенных случайных величин с одинаковыми дисперсиями и с одинаковыми средними значениями.

В этих условиях применимы полученные выше результаты, что дает

$$\mathbf{E}\left[\sum_{i=1}^{k} n_i \cdot \left(m_i' - m'\right)^2\right] = (k-1) \cdot \sigma^2.$$

К тому же при любом i $\mathbf{E}\left(m_i'\right) = \mathbf{E}\left(m'\right)$, откуда вытекает равенство

$$\mathbf{E}\left[\sum_{i=1}^{k} n_i \cdot \left(m_i' - m'\right) \cdot \left(\mu_i' - \mu\right)\right] = 0.$$

Окончательно находим

$$\mathbf{E}\left(s^2\right) = \mathbf{E}\left[\frac{\sum n_i \cdot \left(m_i - m\right)^2}{k-1}\right] = \sigma^2 + \frac{\sum n_i \cdot \left(\mu_i - \mu\right)^2}{k-1},$$

т. е. $\mathbf{E}\left(s^2\right) \geq \sigma^2$.

2. Случай двух факторов

Пусть при совместном действии двух факторов наблюденное значение случайной величины равно x_{ij}, где i — номер уровня первого фактора, j — номер уровня второго фактора. Смысл этих индексов не совпадает со смыслом пары индексов предыдущего подраздела; чтобы подчеркнуть это, мы опускаем запятую между индексами. Результаты наблюдений можно представить в следующем виде (рис. 1).

				j		q
i				x_{ij}		
p						

Рис. 1. Форма возможного представления случайных величин при совместном действии двух факторов

Другими словами, имеются случайные величины, занумерованные двумя индексами, и x_{ij} — наблюденное значение случайной величины с индексами i и j. В более общем случае имеется не одно, но равное количество наблюдений всех случайных величин.

Здесь имеем различные средние:

$$x_{i\bullet} = \frac{1}{q} \cdot \sum_{j=1}^{q} x_{ij},\ x_{\bullet j} = \frac{1}{p} \cdot \sum_{i=1}^{p} x_{ij},$$

где $x_{i\bullet}$ — среднее i-й строки, а $x_{\bullet j}$ — среднее j-го столбца. Общее среднее всех наблюдений

$$x_{\bullet\bullet} = \frac{1}{pq} \cdot \sum_{i=1}^{p}\sum_{j=1}^{q} x_{ij} = \frac{1}{p} \cdot \sum_{i=1}^{p} x_{i\bullet} = \frac{1}{q} \cdot \sum_{j=1}^{q} x_{\bullet j}.$$

Мы будем применять тот же метод, что и ранее, предполагая равенство средних μ_{ij} всех случайных величин:

$$\mu_{ij} = \mu.$$

Займемся разложением суммы квадратов, представив ее в виде

$$\sum_{ij}\left(x_{ij} - \mu\right)^2 =$$

$$= \sum_{ij}\left(x_{ij} - x_{i\bullet} - x_{\bullet j} + x_{\bullet\bullet} + x_{i\bullet} - x_{\bullet\bullet} + x_{\bullet j} - x_{\bullet\bullet} + x_{\bullet\bullet} - \mu\right)^2,$$

что дает

$$\sum_{ij}\left(x_{ij} - \mu\right)^2 = \sum_{ij}\left(x_{ij} - x_{i\bullet} - x_{\bullet j} + x_{\bullet\bullet}\right)^2 + q \cdot \sum_{i=1}^{p}\left(x_{i\bullet} - x_{\bullet\bullet}\right)^2 +$$

$$+ p \cdot \sum_{j=1}^{q}\left(x_{\bullet j} - x_{\bullet\bullet}\right)^2 + p \cdot q \cdot \left(x_{\bullet\bullet} - \mu\right)^2.$$

Легко видеть, что все суммы попарных произведений равны 0. Например,

$$\sum_{i}\sum_{j}\left(x_{i\bullet} - x_{\bullet\bullet}\right) \cdot \left(x_{\bullet j} - x_{\bullet\bullet}\right) = \sum_{i}\left[\left(x_{i\bullet} - x_{\bullet\bullet}\right) \cdot \sum_{j}\left(x_{\bullet j} - x_{\bullet\bullet}\right)\right],$$

где сумма

$$\sum_{j}\left(x_{\bullet j} - x_{\bullet\bullet}\right) = 0$$

в силу определения $x_{\bullet\bullet}$. Столь же нетрудно убедиться в том, что и остальные суммы попарных произведений равны 0.

Положим теперь

$$\sum_{ij}\left(x_{ij} - x_{i\bullet} - x_{\bullet j} + x_{\bullet\bullet}\right)^2 = S'^2;$$

$$q \cdot \sum_{i}\left(x_{i\bullet} - x_{\bullet\bullet}\right)^2 = s_l'^2;$$

$$p \cdot \sum_{j}\left(x_{\bullet j} - x_{\bullet\bullet}\right)^2 = s_c'^2.$$

Сделаем замену переменных по следующим формулам:

$$x_{ij} - x_{i\bullet} - x_{\bullet j} + x_{\bullet\bullet} = S' \cdot \theta_{ij},$$

$$x_{i\bullet} - x_{\bullet\bullet} = s_l' \cdot \theta_{i\bullet},$$

$$x_{\bullet j} - x_{\bullet\bullet} = s_c' \cdot \theta_{\bullet j},$$

где новыми переменными служат $x_{\bullet\bullet}$, S', s_c', s_l', θ_{ij}, $\theta_{i\bullet}$, $\theta_{\bullet j}$.

Однако число первоначальных переменных x_{ij} равно $p \cdot q$, а теперь мы имеем четыре переменных $x_{\bullet\bullet}$, S', s_c', s_l'; $p \cdot q$ переменных θ_{ij}; p переменных $\theta_{i\bullet}$; q переменных $\theta_{\bullet j}$, что составляет всего $p \cdot q + p + q + 4$ переменных. Значит, не все последние являются свободными, и нам нужно найти соотношения, связывающие θ_{ij}, $\theta_{i\bullet}$ и $\theta_{\bullet j}$, чтобы количество свободных новых переменных было $p \cdot q$.

Сразу находим одну связь

$$\sum_{ij}\theta_{ij}^2 = 1,$$

p связей

$$\sum_{j=1}^{q}\theta_{ij} = 0,\ i = 1,\ 2,\ \ldots,\ p$$

и q связей

$$\sum_{i=1}^{p}\theta_{ij} = 0,\ j = 1,\ 2,\ \ldots,\ q.$$

Заметим, что если p соотношений

$$\sum_{j=1}^{q}\theta_{ij} = 0,\ i = 1,\ 2,\ \ldots,\ p,$$

выполняются, то

$$\sum_{ij}\theta_{ij} = 0.$$

Следовательно, если выполняются $q-1$ соотношений

$$\sum_{i=1}^{p}\theta_{ij} = 0,\ j = 1,\ \ldots,\ q - 1,$$

то q-е соотношение $\sum_{i=1}^{p}\theta_{iq} = 0$ также выполнено. Оно не является независимым от остальных. Таким образом, переменные θ_{ij} удовлетворяют $1+p+q-1 = p+q$ независимым соотношениям.

С другой стороны, находим

$$\sum_{i=1}^{p}\theta_{i\bullet} = 0;\ q \cdot \sum_{i=1}^{p}\theta_{i\bullet}^2 = 1;$$

$$\sum_{j=1}^{q}\theta_{\bullet j} = 0;\ p \cdot \sum_{j=1}^{q}\theta_{\bullet j}^2 = 1,$$

т. е. имеется две связи между переменными $\theta_{i\bullet}$ и две связи между $\theta_{\bullet j}$. Окончательно получаем:

$p-2$ свободных переменных $\theta_{i\bullet}$,

$q-2$ свободных переменных $\theta_{\bullet j}$,

$p \cdot q - p - q$ свободных переменных θ_{ij},

4 свободных переменных $x_{\bullet\bullet}$, S', s_c', s_l', т. е. всего $p \cdot q$ свободных переменных, как и вначале.

В этих условиях закон распределения совокупности введенных случайных величин запишется так [5]:

$$A \cdot e^{-\frac{S'^2}{2\cdot\sigma^2}} \cdot e^{-\frac{s_l'^2}{2\cdot\sigma^2}} \cdot e^{-\frac{p\cdot q\cdot\left(x_{\bullet\bullet}-\mu\right)^2}{2\cdot\sigma^2}} \times$$

$$\times \frac{D\left(x_{ij}\right)}{D\left(x_{\bullet\bullet}, S', s_l', s_c', \theta\right)} dx_{\bullet\bullet}\, dS' ds_l'\, ds_c'\, d\theta.$$

Найдем вид функционального определителя. Заметим, что

$$x_{ij} = x_{\bullet\bullet} + S' \cdot \theta_{ij} + s_l' \cdot \theta_{i\bullet} + s_c' \cdot \theta_{\bullet j}.$$

Следовательно, функциональный определитель имеет форму

$$D = \begin{vmatrix} 1 & \theta_{11} & \theta_{1\bullet} & \theta_{\bullet 1} & \overbrace{S' \quad 0 \quad 0 \quad 0 \quad \ldots \quad 0}^{p\times q-p-q} & \overbrace{s_l' \quad 0 \quad 0 \quad \ldots \quad 0}^{p-2} & \overbrace{s_c' \quad 0 \quad 0 \quad \ldots \quad 0}^{q-2} \\ 1 & \theta_{12} & \theta_{1\bullet} & \theta_{\bullet 2} & 0 \quad S' \quad 0 \quad 0 \quad \ldots & s' & s_c' \\ & & & & 0 \quad 0 \quad S' & & \\ \ldots & & & & & & \\ . & . & . & . & . \quad . \quad . \quad . \quad \ldots \quad . & . \quad . \quad . \quad \ldots \quad . & . \quad . \quad . \quad \ldots \quad . \\ 1 & \theta_{pq} & \theta_{p\bullet} & \theta_{\bullet q} & & & \end{vmatrix}.$$

Как видим, $x_{\cdot\cdot}$ не входит в определитель, S' является общим множителем в $p \cdot q - p - q$ столбцах, s'_l — в $p-2$ столбцах, а s'_c — в $q-2$. Поэтому

$$D = S'^{\,p\cdot q-p-q} \cdot s_l'^{\,p-2} \cdot s_c'^{\,q-2} \cdot F,$$

где F зависит только от переменных θ_{ij}, $\theta_{i\cdot}$ и $\theta_{\cdot j}$.

Проинтегрировав предыдущий элемент вероятности по $p \cdot q - 4$ переменным θ_{ij}, $\theta_{i\cdot}$ и $\theta_{\cdot j}$, мы найдем закон распределения четырех величин $x_{\cdot\cdot}$, S', s'_l, s'_c в виде

$$A \cdot e^{-\frac{S'^2}{2\cdot\sigma^2}} S'^{\,p\cdot q-p-q} \cdot dS' \cdot e^{-\frac{s_l'^2}{2\cdot\sigma^2}} \cdot s_l'^{\,p-2} \cdot ds'_l \cdot e^{-\frac{s_c'^2}{2\cdot\sigma^2}} \times$$

$$\times\, s_c'^{\,q-2} \cdot ds'_c \cdot e^{-\frac{p\cdot q\cdot(x_{\cdot\cdot}-\mu)^2}{2\cdot\sigma^2}} \cdot dx_{\cdot\cdot}.$$

Таким образом, различные компоненты дисперсии оказываются независимыми случайными величинами, подчиняющимися χ^2-распределению с различным числом степеней свободы. Ниже в таблице мы указываем упомянутые составляющие и соответствующее число степеней свободы:

$$\begin{array}{ll}
\sum\limits_{ij}\left(x_{ij}-\mu\right)^2 & p\cdot q, \\
\sum\limits_{ij}\left(x_{ij}-x_{\cdot\cdot}\right)^2 & p\cdot q-1, \\
\left(x_{\cdot\cdot}-\mu\right)^2 & 1, \\
s_c'^2 & q-1, \\
s_l'^2 & p-1, \\
S'^2 & p\cdot q-p-q+1 = (p-1)\cdot(q-1).
\end{array}$$

Видим, что число степеней свободы у $\sum\left(x_{ij}-\mu\right)^2$ равно сумме количества степеней свободы у $\left(x_{\cdot\cdot}-\mu\right)^2$, $s_l'^2$, $s_c'^2$ и S'^2.

Как мы знаем, в этих условиях

$$\mathrm{E}\left[\frac{S'^2}{(p-1)\cdot(q-1)}\right] = \mathrm{E}\left(S^2\right) = \sigma^2,$$

S^2 — остаточная дисперсия,

$$\mathrm{E}\left[\frac{s_l'^2}{p-1}\right] = \mathrm{E}\left(s_l^2\right) = \sigma^2,$$

s_l^2 — дисперсия по строкам,

$$\mathrm{E}\left[\frac{s_c'^2}{q-1}\right] = \mathrm{E}\left(s_c^2\right) = \sigma^2,$$

s_c^2 — дисперсия по столбцам.

Если теперь образовать отношения s_l/S и s_c/S, то с помощью распределения Беренса–Фишера–Снедекора можно проверить гипотезу о равенстве средних значений μ_{ij}.

Иначе говоря, величины

$$\frac{s'_l}{S'} \cdot \sqrt{\frac{(p-1)\cdot(q-1)}{p-1}},\ \frac{s'_c}{S'} \cdot \sqrt{\frac{(p-1)\cdot(q-1)}{q-1}},$$

т. е.

$$\frac{s'_l}{S'} \cdot \sqrt{q-1},\ \frac{s'_c}{S'} \cdot \sqrt{p-1}$$

подчиняются закону Беренса–Фишера–Снедекора. Если, например, $\frac{s'_l}{S'} \cdot \sqrt{q-1}$ отклоняется значимо от 1 (при данном уровне значимости), а $\frac{s'_c}{S'} \cdot \sqrt{p-1}$ не отклоняется значимо, то следует принять гипотезу о том, что средние μ_{ij} одинаковы в любом столбце, но меняются по строкам.

Если же оба отношения отклоняются значимо от 1, то следует считать, что μ_{ij} изменяется и по строкам и по столбцам.

3. Обобщение на случай *k* факторов

До сих пор наши случайные величины зависели лишь от двух факторов. Теперь мы предположим, что имеется k факторов, и, чтобы облегчить изложение, допустим, что $k = 5$.

Пусть число уровней факторов таково [4]:

Фактор	Число уровней
1	p
2	q
3	r
4	s
5	t

Считаем, что имеется одно и то же количество наблюдений каждой величины x_{ijklm}. Пусть для простоты имеется только по одному наблюдению. Само собой разумеется, все случайные величины считаются нормальными с одинаковой дисперсией, и мы предполагаем, что у них одинаковые средние, т. е. $\mu_{ijklm} = \mu$ при всех i, j, k, l, m.

Как это уже делалось, мы разлагаем полную сумму квадратов, записав ее в следующей форме:

$$\sum\left(x_{ijklm}-\mu\right)^2 = \sum\left(x_{ijklm}-x_{\cdot\cdot\cdot\cdot\cdot}\right)^2 + \\ + p\cdot q\cdot r\cdot s\cdot t\cdot\left(x_{\cdot\cdot\cdot\cdot\cdot}-\mu\right)^2.$$

Затем мы разлагаем сумму квадратов справа

$$\sum\left(x_{ijklm}-x_{\cdot\cdot\cdot\cdot\cdot}\right)^2,$$

представив ее в виде

$$\sum\{[x_{ijklm} - x_{ijkl\cdot} - x_{ij\cdot m} - x_{ij\cdot lm} - x_{i\cdot klm} - x_{\cdot jklm} + x_{ijk\cdot\cdot} +$$
$$+ (9 \text{ аналогичных членов с тремя индексами}) -$$
$$- x_{ij\cdot\cdot\cdot} - (9 \text{ аналогичных членов с двумя индексами}) +$$
$$+ x_{i\cdot\cdot\cdot\cdot} + x_{\cdot j\cdot\cdot\cdot} + x_{\cdot\cdot k\cdot\cdot} + x_{\cdot\cdot\cdot l\cdot} + x_{\cdot\cdot\cdot\cdot m} - x_{\cdot\cdot\cdot\cdot\cdot}] +$$
$$+ [x_{ijkl\cdot} - x_{ijk\cdot\cdot} - x_{ij\cdot l\cdot} - x_{i\cdot kl\cdot} - x_{\cdot jkl\cdot} + x_{ij\cdot\cdot\cdot} + x_{i\cdot k\cdot\cdot} + x_{i\cdot\cdot l\cdot} +$$
$$+ x_{\cdot jk\cdot\cdot} + x_{\cdot j\cdot l\cdot} + x_{\cdot\cdot kl\cdot} - x_{i\cdot\cdot\cdot\cdot} - x_{\cdot j\cdot\cdot\cdot} - x_{\cdot\cdot k\cdot\cdot} - x_{\cdot\cdot\cdot l\cdot} + x_{\cdot\cdot\cdot\cdot\cdot}] +$$
$$+ [4 \text{ квадратные скобки, аналогичные только что выписанному члену, для } x_{ijk\cdot m}, x_{ij\cdot lm}, x_{i\cdot klm}] +$$
$$+ \left[x_{ijk\cdot\cdot} - x_{ij\cdot\cdot\cdot} - x_{i\cdot k\cdot\cdot} - x_{\cdot jk\cdot\cdot} + x_{i\cdot\cdot\cdot\cdot} + x_{\cdot j\cdot\cdot\cdot} + x_{\cdot\cdot k\cdot\cdot} - x_{\cdot\cdot\cdot\cdot\cdot}\right] +$$
$$+ [9 \text{ квадратных скобок, аналогичных только что выписанному члену для величин с тремя индексами}] +$$
$$+ \left[x_{ij\cdot\cdot\cdot} - x_{i\cdot\cdot\cdot\cdot} - x_{\cdot j\cdot\cdot\cdot} + x_{\cdot\cdot\cdot\cdot\cdot}\right] +$$
$$+ [9 \text{ аналогичных квадратных скобок для величин с двумя индексами}] + \left[x_{i\cdot\cdot\cdot\cdot} - x_{\cdot\cdot\cdot\cdot\cdot}\right] +$$
$$+ [4 \text{ аналогичные квадратные скобки для величин с одним индексом}]\}^2.$$

Механизм образования квадратных скобок весьма прост, и лишь длина выкладок производит впечатление сложности.

В этом разложении, например, член x_{ijk} обозначает среднее всех величин, у которых первые три индекса равны соответственно i, j, k. Число таких величин равно $s \cdot t$. С помощью этого разложения можно непосредственно проверить, что все суммы попарных произведений квадратных скобок равны 0 (выражения в квадратных скобках, как говорят, попарно ортогональны).

Как следствие, введенная выше полная (суммарная) дисперсия

$$\sum_{i,j,k,l,m} \left(x_{ijklm} - \mu\right)^2$$

разлагается в сумму 32 сумм квадратов (выражений, стоящих в квадратных скобках). Легко показать (единственная трудность здесь состоит в чрезвычайной сложности записи) с помощью замены переменных, аналогичной замене, сделанной в случае двух факторов, что все эти различные суммы квадратов оказываются независимыми случайными величинами, подчиняющимися закону χ^2. Можно определить число степеней свободы всех законов. Их можно найти непосредственно (как мы сделали в двухфакторном анализе). Для этого достаточно определить число связей между переменными θ, которые появляются при замене переменных. Можно также исходить из результата, когда сумма двух независимых случайных величин, имеющих χ^2-распределения соответственно с n_1 и n_2 степенями свободы, имеет χ^2-распределение с n_1+n_2 степенями свободы.

Рассмотрим, например,

$$\sum_{i,j,k} \left(x_{ijk\bullet\bullet} - x_{\bullet\bullet\bullet\bullet\bullet}\right)^2.$$

Эта случайная величина подчиняется закону χ^2 с $p \cdot q \cdot r - 1$ степенями свободы.

Имеем:

$$\sum_{i,j,k} \left(x_{ijk\bullet\bullet} - x_{\bullet\bullet\bullet\bullet\bullet}\right)^2 =$$

$$= \sum_{i,j,k} \begin{pmatrix} x_{ijk\bullet\bullet} - x_{ij\bullet\bullet\bullet} - x_{i\bullet k\bullet\bullet} - x_{\bullet jk\bullet\bullet} + \\ + x_{i\bullet\bullet\bullet\bullet} + x_{\bullet j\bullet\bullet\bullet} + x_{\bullet\bullet k\bullet\bullet} - x_{\bullet\bullet\bullet\bullet\bullet} \end{pmatrix}^2 +$$

$$+ r \cdot \sum_{i,j} \left(x_{ij\bullet\bullet\bullet} - x_{i\bullet\bullet\bullet\bullet} - x_{\bullet j\bullet\bullet\bullet} + x_{\bullet\bullet\bullet\bullet\bullet}\right)^2 +$$

$$+ p \cdot \sum_{j,k} \left(x_{\bullet jk\bullet\bullet} - x_{\bullet j\bullet\bullet\bullet} - x_{\bullet\bullet k\bullet\bullet} - x_{\bullet\bullet\bullet\bullet\bullet}\right)^2 + q \cdot r \cdot \sum \left(x_{i\bullet\bullet\bullet\bullet} - x_{\bullet\bullet\bullet\bullet\bullet}\right)^2 +$$

$$+ p \cdot r \cdot \sum \left(x_{\bullet j\bullet\bullet\bullet} - x_{\bullet\bullet\bullet\bullet\bullet}\right)^2 + p \cdot q \cdot \sum \left(x_{\bullet\bullet k\bullet\bullet} - x_{\bullet\bullet\bullet\bullet\bullet}\right)^2.$$

Обозначим через $N_{i,j,k}$ число степеней свободы суммы

$$\sum_{i,j,k} \left(x_{ijk\bullet\bullet} - x_{ij\bullet\bullet\bullet} - x_{i\bullet k\bullet\bullet} - x_{\bullet jk\bullet\bullet} + x_{i\bullet\bullet\bullet\bullet} + x_{\bullet j\bullet\bullet\bullet} + x_{\bullet\bullet k\bullet\bullet} - x_{\bullet\bullet\bullet\bullet\bullet}\right)^2.$$

В силу только что упомянутого результата, поскольку число степеней свободы суммы

$$\sum \left(x_{ij\bullet\bullet\bullet} - x_{i\bullet\bullet\bullet\bullet} - x_{\bullet j\bullet\bullet\bullet} + x_{\bullet\bullet\bullet\bullet\bullet}\right)^2,$$

как мы видели в п. 2 (в случае двух индексов), равно $(p-1)\cdot(q-1)$, число степеней свободы двух других сумм с двумя индексами соответственно равно $(p-1)\cdot(r-1)$ и $(r-1)\cdot(q-1)$, количество степеней свободы для сумм с одним индексом соответственно равно $p-1$, $q-1$, $r-1$, имеем

$$p \cdot q \cdot r - 1 = N_{i,j,k} + (p-1)\cdot(q-1) + (p-1)\cdot(r-1) +$$

$$+ (r-1)\cdot(q-1) + (p-1) + (q-1) + (r-1).$$

Отсюда следует равенство

$$N_{i,j,k} = (p-1)\cdot(q-1)\cdot(r-1),$$

что видно из разложения

$$p \cdot q \cdot r - 1 = [(p-1)+1]\cdot[(q-1)+1]\cdot[(r-1)+1] - 1 =$$

$$= (p-1)\cdot(q-1)\cdot(r-1) + (p-1)\cdot(q-1) + (q-1)\cdot(r-1) +$$

$$+ (r-1)\cdot(p-1) + (p-1) + (q-1) + (r-1).$$

Приведенное рассуждение, очевидно, является общим и, следовательно, число степеней свободы суммы квадратов с λ индексами, которые принимают p_1, p_2, …, p_λ различных значений, равно $(p_1-1)\cdot(p_2-1)\times$ $\times\ldots\cdot(p_\lambda-1)$.

Недостаток изложенной модели состоит в том, что она требует слишком большого количества наблюдений. В случае пяти факторов с p, q, r, s, t уровнями соответственно нужно произвести $p \cdot q \cdot r \cdot s \cdot t$ наблюдений. Еще одним неудобством являются слишком громоздкие вычисления.

4. Дисперсия взаимодействия

Вернемся к случаю, когда имеется всего два фактора. Представим разложение суммы квадратов в следующей форме:

$$\sum \left(x_{ij} - x_{\bullet\bullet}\right)^2 = \sum_{j,k} \left(x_{ij} - x_{i\bullet} - x_{\bullet j} + x_{\bullet\bullet}\right)^2 +$$

$$+ \sum_{i=1}^{p} q \cdot \left(x_{i\bullet} - x_{\bullet\bullet}\right)^2 + \sum_{j=1}^{q} p \cdot \left(x_{\bullet j} - x_{\bullet\bullet}\right)^2.$$

Если средние для различных уровней не равны, то две последние суммы квадратов уже не являются оценками для дисперсии σ^2.

В самом деле, пусть

$$x_{ij} = \mu_{ij} + y_{ij},$$

где y_{ij} имеют нулевые средние. Отсюда

$$x_{i\bullet} = \mu_{i\bullet} + y_{i\bullet}, \text{ где } \mu_{i\bullet} = \frac{\sum_{j=1}^{q} \mu_{ij}}{q},$$

$$x_{\bullet j} = \mu_{\bullet j} + y_{\bullet j}, \text{ где } \mu_{\bullet j} = \frac{\sum_{i=1}^{p} \mu_{ij}}{p}$$

$$\text{и } x_{\bullet\bullet} = \mu_{\bullet\bullet} + y_{\bullet\bullet}, \text{ где } \mu_{\bullet\bullet} = \frac{\sum_{i,j} \mu_{ij}}{p \cdot q}.$$

Остаточную сумму квадратов можно записать в виде

$$\sum \left(y_{ij} - y_{i\bullet} - y_{\bullet j} + y_{\bullet\bullet} + \mu_{ij} - \mu_{i\bullet} - \mu_{\bullet j} + \mu_{\bullet\bullet}\right)^2 =$$

$$= \sum \left(y_{ij} - y_{i\bullet} - y_{\bullet j} + y_{\bullet\bullet}\right)^2 + \sum \left(\mu_{ij} - \mu_{i\bullet} - \mu_{\bullet j} + \mu_{\bullet\bullet}\right)^2 +$$

$$+ 2 \cdot \sum \left(y_{ij} - y_{i\bullet} - y_{\bullet j} + y_{\bullet\bullet}\right) \cdot \left(\mu_{ij} - \mu_{i\bullet} - \mu_{\bullet j} + \mu_{\bullet\bullet}\right).$$

Математическое ожидание суммы попарных произведений равно 0, поскольку все величины y_{ij} имеют по определению нулевое математическое ожидание, а, следовательно, и все $y_{i\bullet}$, $y_{\bullet j}$ и $y_{\bullet\bullet}$ имеют нулевые ожидания.

Переходя в последнем равенстве к математическим ожиданиям, получим

$$\mathrm{E}\left[\sum\left(y_{ij}-y_{i\bullet}-y_{\bullet j}+y_{\bullet\bullet}+\mu_{ij}-\mu_{i\bullet}-\mu_{\bullet j}+\mu_{\bullet\bullet}\right)^2\right]=$$

$$=\mathrm{E}\left[\sum\left(y_{ij}-y_{i\bullet}-y_{\bullet j}+y_{\bullet\bullet}\right)^2\right]+\mathrm{E}\left[\sum\left(\mu_{ij}-\mu_{i\bullet}-\mu_{\bullet j}+\mu_{\bullet\bullet}\right)^2\right]=$$

$$=\mathrm{E}\left[\sum\left(y_{ij}-y_{i\bullet}-y_{\bullet j}+y_{\bullet\bullet}\right)^2\right]+\sum\left(\mu_{ij}-\mu_{i\bullet}-\mu_{\bullet j}+\mu_{\bullet\bullet}\right)^2.$$

Отсюда видно, что в данном случае остаточная дисперсия не является более несмещенной оценкой σ^2, поскольку

$$\frac{1}{(p-1)\cdot(q-1)}\cdot\mathrm{E}\left[\sum\left(x_{ij}-x_{i\bullet}-x_{\bullet j}+x_{\bullet\bullet}\right)^2\right]=$$

$$=\sigma^2+\frac{1}{(p-1)\cdot(q-1)}\cdot\sum_{ij}\left(\mu_{ij}-\mu_{i\bullet}-\mu_{\bullet j}+\mu_{\bullet\bullet}\right)^2.$$

В общем случае, если второй член справа не равен 0, эта оценка систематически завышает значение σ^2.

Однако имеется случай, когда второй член равен 0, если μ_{ij} представляется в виде суммы двух слагаемых, одно из которых зависит только от i, а другое — только от j:

$$\mu_{ij}=\alpha_i+\beta_j.$$

Вклад в μ_{ij}, отвечающий одному из индексов, не зависит от другого. Другими словами, отсутствует взаимодействие между i и j. Действительно, в этом случае

$$\mu_{i\bullet}=\alpha_i+\beta,$$

$$\mu_{\bullet j}=\alpha_{\bullet}+\beta_j,$$

$$\mu_{\bullet\bullet}=\alpha_{\bullet}+\beta_{\bullet},$$

и ясно, что сумма

$$\sum\left(\mu_{ij}-\mu_{i\bullet}-\mu_{\bullet j}+\mu_{\bullet\bullet}\right)^2$$

равна 0, поскольку каждое слагаемое равно 0.

Но даже в этом случае суммы

$$\sum_i q\cdot\left(x_{i\bullet}-x_{\bullet\bullet}\right)^2 \text{ и } \sum_j p\cdot\left(x_{\bullet j}-x_{\bullet\bullet}\right)^2$$

не являются уже несмещенными оценками σ^2, как видно из разложения

$$\sum_i q\cdot\left(x_{i\bullet}-x_{\bullet\bullet}\right)^2=\sum q\cdot\left(y_{i\bullet}-y_{\bullet\bullet}+\mu_{i\bullet}-\mu_{\bullet\bullet}\right)^2=$$

$$=\sum q\cdot\left(y_{i\cdot}-y_{\cdot\cdot}\right)^2+\sum q\cdot\left(\mu_{i\bullet}-\mu_{\bullet\bullet}\right)^2+$$

$$+2\cdot\sum q\cdot\left(y_{i\bullet}-y_{\bullet\bullet}\right)\cdot\left(\mu_{i\bullet}-\mu_{\bullet\bullet}\right).$$

При переходе к математическим ожиданиям попарные произведения исчезают, но остается член

$$\sum q\cdot\left(\mu_{i\bullet}-\mu_{\bullet\bullet}\right)^2,$$

который, впрочем, сводится к

$$\sum q\cdot\left(\alpha_{i\bullet}-\alpha_{\bullet}\right)^2.$$

Аналогичный результат получим для суммы

$$\sum p\cdot\left(x_{\bullet j}-x_{\bullet\bullet}\right)^2,$$

от которой появляется дополнительный член

$$\sum p\cdot\left(\beta_j-\beta_{\bullet}\right)^2.$$

Рассмотрим теперь случай трех факторов. Мы получим следующее разложение суммы квадратов отклонений от общего среднего (пусть i меняется от 1 до p, j — от 1 до q и k — от 1 до r):

$$\sum\left(x_{ijk}-x_{\bullet\bullet\bullet}\right)^2=$$

$$=\sum\begin{pmatrix}x_{ijk}-x_{ij\bullet}-x_{i\bullet k}-x_{\bullet jk}+\\+x_{i\bullet\bullet}+x_{\bullet j\bullet}+x_{\bullet\bullet k}-x_{\bullet\bullet\bullet}\end{pmatrix}^2+$$

$$+r\cdot\sum\left(x_{ij\bullet}-x_{i\bullet\bullet}-x_{\bullet j\bullet}+x_{\bullet\bullet\bullet}\right)^2+$$

$$+q\cdot\sum\left(x_{i\bullet k}-x_{i\bullet\bullet}-x_{\bullet\bullet k}+x_{\bullet\bullet\bullet}\right)^2+$$

$$+p\cdot\sum\left(x_{\bullet jk}-x_{\bullet j\bullet}-x_{\bullet\bullet k}+x_{\bullet\bullet\bullet}\right)^2+$$

$$+q\cdot r\cdot\sum\left(x_{i\bullet\bullet}-x_{\bullet\bullet\bullet}\right)^2+p\cdot r\cdot\sum\left(x_{\bullet j\bullet}-x_{\bullet\bullet\bullet}\right)^2+$$

$$+p\cdot q\cdot\sum\left(x_{\bullet\bullet k}-x_{\bullet\bullet\bullet}\right)^2.$$

Первая сумма справа может служить мерой взаимодействия второго порядка, а три следующие суммы — мерой взаимодействия первого порядка.

Пусть средние μ_{ijk} неодинаковы, но можно ввести величины y_{ijk}, имеющие нулевые средние, причем

$$y_{ijk}=\mu_{ijk}+y_{ijk}$$

и

$$\mu_{ijk}=\alpha_i+\beta_j+\gamma_k.$$

В этих условиях можно показать совершенно аналогично предыдущему, что три дисперсии взаимодействия первого порядка и дисперсия взаимодействия второго порядка доставляют несмещенные оценки σ^2, причем эти статистики независимы. Следовательно, образовав отношение двух этих оценок, можно проверять гипотезу о независимости действия факторов, т. е. проверять возможность представления μ_{ijk} в виде $\alpha_i+\beta_j+\gamma_k$.

Такой тест можно построить только для случая, когда имеется не менее трех факторов, так как при двух факторах мы имеем лишь одну дисперсию взаимодействия, которую не с чем сравнивать [Ibid.].

Понятие взаимодействия без труда обобщается на случай, когда количество факторов превышает три.

5. Случай различного числа наблюдений в ячейках

Вернемся к случаю двух факторов. Пару (i, j) удобно интерпретировать как ячейку (см. рис. 1). До сих пор мы предполагали, что в ячейке (i, j) имелось лишь одно наблюдение, т. е. лишь одна величина x_{ij} с индексами i, j. Пусть теперь имеется n_{ij} наблюдений в ячейке (i,j); мы будем обозначать наблюдения через $x_{ij,k}$, где i и j обозначают уровни факторов, а k — номер наблюдения для уровней i и j.

Мы вновь можем разложить сумму квадратов отклонений, но более сложным способом [2]. Имеем:

$$\sum_{i,j,k}\left(x_{ij,k}-m\right)^2=\sum_{i,j,k}\left(x_{ij,k}-m_{ij}\right)^2+\sum_{i,j}n_{ij}\cdot\left(m_{ij}-m\right)^2$$

и

$$\underbrace{\sum_{i,j,k}\left(x_{ij,k}-\mu\right)^2}_{N}=\underbrace{\sum_{i,j,k}\left(x_{ij,k}-m_{ij}\right)^2}_{N-p\cdot q}+$$

$$\underbrace{\sum_{i,j}n_{ij}\cdot\left(m_{ij}-m\right)^2}_{p\cdot q-1}+\underbrace{N\cdot(m-\mu)^2}_{1}.$$

Под каждой суммой мы указали отвечающее ей число степеней свободы.

Мы видим, что сумма (остаточная дисперсия)

$$\frac{1}{N-p\cdot q}\cdot\sum\left(x_{ij,k}-m_{ij}\right)^2$$

является оценкой σ^2 и любая функция от m_{ij} не зависит от этой оценки. Данное свойство можно будет использовать для проверки гипотезы о равенстве средних μ_{ij}, если мы найдем еще одну оценку σ^2.

Рассмотрим теперь таблицу, аналогичную той, которую мы имели ранее. Пусть в клетке i, j этой таблицы записано значение m_{ij} нормально распределенной со средним μ и дисперсией σ^2/n_{ij} случайной величины. Мы приходим к задаче, аналогичной рассмотренной выше, в которой, однако, дисперсии не одинаковы.

Введем следующие величины:

$$M_j=\frac{1}{p}\cdot\sum_{i=1}^{p}m_{ij},\ j=1,\ldots,q.$$

Величина M_j нормально распределена со средним μ и дисперсией

$$\frac{\sigma^2}{p^2}\cdot\sum_{i=1}^{p}\frac{1}{n_{ij}}.$$

Положим:

$$\frac{1}{N_j}=\frac{1}{p^2}\cdot\sum_{i=1}^{p}\frac{1}{n_{ij}}.$$

Случайные величины M_j и M_k независимы при $j\neq k$.

В этих условиях совокупность величин M_j имеет плотность вероятности

$$K\cdot e^{-\frac{1}{2}\cdot\sum_j\frac{N_j\left(M_j-\mu\right)^2}{\sigma^2}}.$$

Разложим входящую сюда сумму квадратов, положив

$$M=\frac{\sum M_j\cdot N_j}{\sum N_j}.$$

Можем записать:

$$\sum N_j\cdot\left(M_j-\mu\right)^2=\sum N_j\cdot\left(M_j-M+M-\mu\right)^2=$$

$$=\sum N_j\cdot\left(M_j-M\right)^2+\left(\sum N_j\right)\cdot(M-\mu)^2.$$

Как легко проверить, попарные произведения исчезают.

Далее определим

$$S'^2=\sum N_j\cdot\left(M_j-M\right)^2.$$

Затем сделаем следующую замену переменных:

$$M_j-M=S'\cdot\theta_j,$$

причем старыми переменными считаем M_j, а новыми — M, S' и q величин θ_j. Сразу же устанавливаем, что q переменных θ_j связаны двумя следующими соотношениями:

$$\sum N_j\cdot\theta_j=0,\ N_j\cdot\theta_j^2=1.$$

Следовательно, мы имеем $q-2$ свободных переменных θ_j, которые вместе с M и S' образуют совокупность q новых свободных переменных, столько же, как и вначале.

Легко проверить способом, неоднократно уже примененным, что якобиан преобразования может быть записан в следующей форме [5]:

$$\frac{D\left(M_1,\ M_2,\ldots,M_q\right)}{D\left(M,\ S',\ \theta_1,\theta_2,\ldots,\theta_{q-2}\right)}=S'^{q-2}\cdot f\left(\theta_1,\theta_2,\ldots,\theta_{q-2}\right).$$

Обращаясь к виду совместной плотности, сразу находим, что M подчиняется нормальному закону, а S' — закону χ^2 с $q-1$ степенями свободы, причем M и S' независимы.

Следовательно, величина

$$\frac{\sum N_j\cdot\left(M_j-M\right)^2}{q-1}$$

является несмещенной оценкой дисперсии σ^2 и можно сравнивать ее с оценкой, найденной выше. Отношение этих двух независимых оценок подчиняется закону Беренса–Фишера–Снедекора, и, таким образом, мы можем проверять справедливость гипотезы о равенстве средних μ_{ij} с разными j. Эта гипотеза отвергается или принимается в зависимости от того, значимо или нет отклоняется от 1 отношение последней оценки к оценке с помощью остаточной дисперсии. С помощью указанного теста проверяется влияние уровней j второго фактора. Аналогично исследуется влияние уровней i первого фактора. На основании этой информации можно непосредственно строить корректные функциональные зависимости, связывающие в определенных допущениях управляющие и управляемые параметры экономических систем, даже при наличии условий неопределенности (рыночной волатильности).

Литература

1. *Колемаев В. А.* Экономико-математическое моделирование. Моделирование макроэкономических процессов и систем: Учебник. М.: Юнити-Дана, 2005. 295 с.
2. *Fisher R. A.* The design of experiments. London: Oliver and Boyd, 1935. 260 p.
3. *Кобзарь А. И.* Прикладная математическая статистика. Для инженеров и научных работников. 2-е изд., М.: ФИЗМАТЛИТ, 2012. 816 с.
4. *Mann H. B.* Analysis and design of experiments: Analysis of variance and analysis of variance designs. N.Y.: Dover publications, 1949. 195 p.
5. *Fréchet M.* Généralités sur les probabilités. Eléments aléatoires. 2nd ed. Paris: Gauthier-Villars, 1950. 355 p.

References

1. Kolemaev V. A. *Ekonomiko-matematicheskoe modelirovanie. Modelirovanie makroekonomicheskikh protsessov i sistem* [Economic-mathematical modeling. Modeling of macroeconomic processes and systems]. Moscow, YuNITI-DANA Publ., 2005. 295 p.
2. Fisher R. A. *The design of experiments.* London, Oliver and Boyd Publ., 1935. 260 p.
3. Kobzar' A. I. *Prikladnaya matematicheskaya statistika. Dlya inzhenerov i nauchnykh rabotnikov* [Applied mathematical statistics. For engineers and scientists]. Moscow, FIZMATLIT Publ., 2012. 816 p.
4. Mann H. B. *Analysis and design of experiments: Analysis of variance and analysis of variance designs.* New York, Dover Publ., 1949. 195 p.
5. Fréchet M. *Généralités sur les probabilités. Eléments aléatoires* [Overview of probabilities. Random elements]. Paris. Gauthier-Villars Publ., 1950. 355 p.