ЭКОНОМИКА
И МАТЕМАТИЧЕСКИЕ МЕТОДЫ

1967
том III, вып. 3

НАУЧНЫЕ КОНСУЛЬТАЦИИ

КОРРЕЛЯЦИОННЫЙ И РЕГРЕССИОННЫЙ АНАЛИЗ В ЭКОНОМИЧЕСКИХ ИССЛЕДОВАНИЯХ *

Г. В. РОЗАНОВ, А. А. ФРЕНКЕЛЬ

(Москва)

Применение методов математической статистики дает возможность количественно оценивать связи между экономическими явлениями, а также построить и проанализировать экономические модели стохастического характера.

Целью данной консультации является описание методов парной и множественной корреляции и регрессии.

I. ПОНЯТИЕ КОРРЕЛЯЦИОННОГО И РЕГРЕССИОННОГО АНАЛИЗА И ПРЕДПОСЫЛКИ ИХ ПРИМЕНЕНИЯ

Основными условиями объективного анализа экономических взаимосвязей являются корректная постановка задачи, правильный выбор метода исследования и соответствующая данному конкретному объекту интерпретация полученных результатов. Для выполнения этих условий экономист должен иметь четкое представление о видах связи, имеющих место в экономике. Количественно измеримые экономические явления могут быть выражены как случайными, так и неслучайными величинами.

Связь между двумя величинами называется функциональной, если любому определенному значению величины x_2 (из множества ее возможных значений) соответствует одно и только одно определенное значение величины x_1. Связь между двумя величинами называется стохастической, если после определения значения величины x_2 величина x_1 останется случайной, способной принимать различные значения с определенными вероятностями.

Как стохастическая, так и функциональная связь не тождественны с причинной связью между экономическими явлениями. Под причинной будем понимать следующую связь: явление x_2 служит причиной явления x_1, если появлению x_2 всегда сопутствует появление x_1. При изучении связи между явлениями стохастическая зависимость часто указывает на соответствующую причинную зависимость (например, зависимость производительности труда рабочих от стажа работы по данной специальности). Но при наличии стохастической связи между явлениями может и не быть причинной зависимости между ними. Это происходит потому, что оба явления в отдельности зависят от общих факторов. Так, связь между фондоотдачей и себестоимостью (затратами на 1 рубль товарной продукции) является стохастической и непричинной, так как оба эти показателя зависят от фондовооруженности, электровооруженности и т. д.

* В настоящей консультации предполагается, что читатель ознакомился с основными понятиями теории вероятностей и математической статистики, данными в консультации С. А. Айвазяна [1].

Частными случаями стохастической формы связи являются корреляционная и регрессионная связи. Две случайные величины являются корреляционно связанными, если математическое ожидание одной из них меняется в зависимости от изменения другой. Метод математической статистики, изучающий корреляционные связи между явлениями, называется корреляционным анализом. Корреляционный анализ представляет собой инструмент, позволяющий количественно оценивать связи между большим числом взаимодействующих экономических явлений, ряд из которых не известен. Его применение делает возможным проверку различных экономических гипотез о наличии и силе связи между двумя явлениями или между отдельным явлением и группой явлений, а также гипотезы о форме связи.

Применение корреляционного анализа предполагает выполнение следующих предпосылок: 1) случайные величины x_1 и x_2 (в многомерном случае $x_1, x_2, \ldots, x_m$) могут рассматриваться как выборка из двумерной (многомерной) генеральной совокупности с нормальным законом распределения; 2) отдельные наблюдения стохастически независимы, т. е. значение данного наблюдения не должно зависеть от значения предыдущего и последующего наблюдений; 3) дисперсия случайной величины x_1 остается постоянной при изменении величины x_2 или пропорциональна некоторой известной функции от x_2 *; 4) математическое ожидание величины x_1 при x_2, принявшей определенное значение, можно выразить в виде функции $x_1 = f(x_2)$, линейной относительно определяемых параметров **.

Перед началом корреляционного анализа необходимо проверять выполнение этих предпосылок.

Связь между случайной и неслучайной величиной называется регрессионной, а метод анализа таких связей — регрессионным анализом. Применение регрессионного анализа предполагает обязательное выполнение предпосылок 2), 3), 4) корреляционного анализа.

Регрессионный анализ тесно связан с корреляционным. При выполнении предпосылок корреляционного анализа выполняются предпосылки регрессионного анализа. В то же время регрессионный анализ предъявляет менее жесткие требования к исходной информации. Так, например, проведение регрессионного анализа возможно даже в случае некоторого отличия распределения случайной величины от нормального, что существенно для экономики, так как часто распределение экономических величин асимметрично. В качестве функции *** в регрессионном анализе обязательно принимается случайная переменная, а аргументом, как правило, является неслучайная переменная.

Примером возможного применения регрессионного анализа в экономике является исследование производительности труда, себестоимости и других качественных экономических показателей от таких факторов, как размер основных фондов, удельный вес заработной платы в затратах на производстве, уровни специализации, кооперирования и т. д.

---

* Случаи, при которых дисперсия пропорциональна изменению некоторой функции от x_2, в данной консультации не рассматриваются (см. [3, гл. XVIII, § 6]). При статистических исследованиях в экономике обычно принимается, что дисперсия постоянна. О проверке постоянности дисперсии величины x_1 при изменении значений x_2 (см. [3, гл. XI, § 6]).

** Функция $f(x; a)$ называется линейно зависящей от a, если соблюдается равенство $f(x, a_1 + a_2) = \varphi_1(x_1, a_1) + \varphi_2(x_1, a_2)$.

*** В дальнейшем термины «функции» и «аргументы» будут использоваться безотносительно к наличию или отсутствию причинной зависимости между анализируемыми явлениями.

II. ЛИНЕЙНЫЙ КОРРЕЛЯЦИОННЫЙ И РЕГРЕССИОННЫЙ АНАЛИЗ ДВУХ ПЕРЕМЕННЫХ

Корреляционный анализ связи между двумя экономическими явлениями x_1 и x_2 опирается на предпосылки, рассмотренные в разделе I. Для проверки гипотезы о наличии связи между ними и оценки тесноты этой связи вычисляется коэффициент корреляции $r_{x_1x_2}$, если связь линейна, и корреляционное отношение $\eta_{x_1x_2}$, если связь нелинейна. Для нахождения коэффициента корреляции в литературе предлагаются разные формулы. Одной из них является следующая:

$$r_{x_1x_2} = \frac{n\sum_{j=1}^{n} x_{1j}x_{2j} - \sum_{j=1}^{n} x_{1j} \sum_{j=1}^{n} x_{2j}}{\sqrt{\left[n\sum_{j=1}^{n}(x_{1j})^2 - \left(\sum_{j=1}^{n} x_{1j}\right)^2\right]\left[n\sum_{j=1}^{n}(x_{2j})^2 - \left(\sum_{j=1}^{n} x_{2j}\right)^2\right]}} \qquad (1)$$

где n — число наблюдений в выборке.

Коэффициент корреляции изменяется в интервале $-1 \leqslant r_{x_1x_2} \leqslant +1$. При $r_{x_1x_2} = \pm 1$ между x_1 и x_2 существует прямая или обратная функциональная связь. При коэффициенте корреляции, равном 0, между x_1 и x_2 не существует корреляционной связи. Если коэффициент корреляции находится в интервале $-1 < r_{x_1x_2} < 0$ или $0 < r_{x_1x_2} < 1$, то между x_1 и x_2 существует обратная или прямая корреляционная связь.

Следует еще раз подчеркнуть, что интерпретация коэффициента корреляции в качестве показателя тесноты связи имеет смысл в том и только в том случае, когда полностью выполнены предпосылки корреляционного анализа. В остальных случаях его интерпретация оказывается ненадежной. Часто интерпретация коэффициента корреляции усложняется в связи с тем, что две исследуемые переменные причинно зависят от других факторов, в числе которых оказывается один или несколько факторов, влияющих на оба показателя. В этом случае высокая корреляция между x_1 и x_2 не указывает на причинную зависимость x_1 от x_2 *.

Значение коэффициента парной корреляции является случайной величиной, зависящей от объема выборки. С уменьшением числа наблюдений надежность коэффициента корреляции как измерителя тесноты связи падает. Для оценки этой надежности необходимо изучить вид распределения коэффициента корреляции в выборочной совокупности. Доказано, что с увеличением числа наблюдений (свыше 500) распределение коэффициента $r_{x_1x_2}$, не превышающего 0,9, стремится к нормальному.

Полученный из выборки коэффициент корреляции является оценкой соответствующего «истинного» коэффициента корреляции, существующего в генеральной совокупности, а именно последний и характеризует «истинную» тесноту корреляционной связи между двумя экономическими явле-

---

* Например, при изучении связи между производительностью труда и факторами, влиящими на ее уровень, коэффициент парной корреляции между производительностью труда и объемом произведенной валовой продукции для предприятий цементной промышленности оказался равным 0,71. После исключения влияний других факторов производства парный коэффициент корреляции между этими же факторами оказался равным 0,06, т. е. практически связью можно пренебречь. Завышенная величина коэффициента парной корреляции явилась результатом наслаивающегося влияния на эту связь показателей фондовооруженности, специализации и т. д., находящихся в корреляционной связи с показателем объема произведенной валовой продукции.

ниями. Поэтому вместо того, чтобы дать одно единственное значение оценки коэффициента корреляции генеральной совокупности, лучше построить так называемый доверительный интервал, в котором с определенной вероятностью этот коэффициент находится.

При этом выдвигается и проверяется гипотеза о том, что «истинный» коэффициент корреляции равен нулю, т. е. проверяется гипотеза об отсутствии корреляционной связи между изучаемыми явлениями.

Обозначим через $\rho_{x_1x_2}$ «истинный» коэффициент корреляции в генеральной совокупности. Тогда величина доверительного интервала, в котором он должен находиться, будет равна:

$$r_{x_1x_2} - t_{\text{т}}S_r \leqslant \rho_{x_1x_2} \leqslant r_{x_1x_2} + t_{\text{т}}S_r, \tag{2}$$

где S_r — среднеквадратическая ошибка выборочного парного коэффициента корреляции, вычисляемая по формуле:

$$S_r = (1 - r^2_{x_1x_2})\sqrt{n-1}. \tag{3}$$

Величина $t_{\text{т}}$ (т — табличное значение) имеет распределение Стьюдента с числом степеней свободы * $\nu = n - 2$. Для определения значения $t_{\text{т}}$ необходимо задаться так называемым уровнем значимости α, т. е. вероятностью $P = 1 - \alpha$ того, что $\rho_{x_1x_2}$ находится в доверительном интервале (2). Затем по таблице значений критерия t-Стьюдента (см. [3—5]) при $\nu = n - 2$ находится теоретическое значение и подставляется в формулу (2). В данном случае интервал двусторонний (нас интересуют $\pm t_{\text{т}}S_r$), поэтому для каждой из сторон подставляем $t_{\text{т}}$ для $\alpha / 2$.

При малом числе наблюдений в выборке и высоком коэффициенте корреляции $r_{x_1x_2}$ для проверки гипотезы о наличии корреляционной связи или отличии от предполагаемого «истинного» значения $\rho_{x_1x_2}$, а также для построения доверительного интервала применяется преобразование z-Фишера. Для этого вычисляются величины

$$z_1 = \frac{1}{2}\ln\frac{1 + r_{x_1x_2}}{1 - r_{x_1x_2}}, \quad z_2 = \frac{1}{2}\ln\frac{1 + \rho_{x_1x_2}}{1 - \rho_{x_1x_2}}. \tag{4}$$

Доказано, что даже при небольшом значении n величина $z_1 - z_2$ распределена нормально со средней, равной нулю, и средним квадратическим отклонением:

$$S_{z_1 - z_2} = 1/\sqrt{n-3}. \tag{5}$$

Перейдем теперь к рассмотрению регрессионного анализа при изучении парных линейных связей между экономическими явлениями.

Линейную зависимость средних значений x_1 от изменения значений x_2 можно выразить уравнением регрессии:

$$x_1 = a_0 + ax_2. \tag{6}$$

Угловой коэффициент регрессии a, или коэффициент регрессии x_1 по x_2, показывает, на сколько единиц в среднем изменяется x_1, когда x_2 изменяется на единицу. Коэффициент a_0 показывает значение x_1 при $x_2 = 0$.

Необходимо отметить, что значение коэффициента a зависит от принятых единиц измерения величин x_1 и x_2. Поэтому для интерпретации коэффициента a удобно использовать так называемые коэффициенты эластичности, вычисляемые по формуле:

$$Э = a\bar{x}_2/\bar{x}_1. \tag{7}$$

---

* Число степеней свободы равно числу независимых величин, участвующих в образовании того или иного параметра. Оно равно общему числу величин, по которым вычисляется параметр, минус число условий, связывающих эти величины.

Коэффициент эластичности показывает, на сколько процентов в среднем изменяется величина x_1 с изменением величины x_2 на один процент.

Задачей регрессионного анализа является вычисление неизвестных параметров a_0 и a, а также их статистических характеристик. Значения оценок a_0 и a будут оптимальными в смысле минимизации суммы квадратов отклонений фактически полученных значений функций x_1 от значений, полученных подстановкой в уравнение регрессии (6) значений аргумента x_2, т. е. будут минимизировать функционал:

$$L = \sum_{j=1}^{n} (x_{1j} - \hat{x}_{1j})^2 \to \min, \tag{8}$$

где $\hat{x}_{1j}$ — расчетные значения функции.

Методом, позволяющим выполнить это условие, является метод наименьших квадратов *. В случае наличия регрессионной связи между явлениями x_1 и x_2 параметры a_0 и a_1 можно вычислить по формулам:

$$a_1 = \sum_{j=1}^{n} (x_{1j} - \bar{x}_1)(x_{2j} - \bar{x}_2) \Big/ \sum_{j=1}^{n} (x_{2j} - \bar{x}_2)^2, \tag{9}$$
$$a_0 = \bar{x}_1 - a_1\bar{x}_2.$$

Значимость отличия a от «истинного коэффициента» регрессии $\tilde{a}$ проверяется по критерию t с $\nu = n - 2$ степенями свободы. Величина t_Φ вычисляется по формуле:

$$t_\Phi = (a - \tilde{a}) / S_a, \tag{10}$$

где

$$S_a = \sqrt{\sum_{j=1}^{n} (x_{1j} - \hat{x}_{1j})^2 / (n-2) \sum_{j=1}^{n} (x_{2j} - \bar{x}_2)^2}. \tag{11}$$

Аппарат парного корреляционного анализа тесно связан с парным регрессионным анализом. Анализ корреляционной связи между двумя экономическими явлениями можно дополнить изучением коэффициентов регрессии. Коэффициенты регрессии связаны с коэффициентами корреляции следующим образом:

$$\text{а) } a_{x_1x_2} = r_{x_1x_2}(S_{x_1}/S_{x_2}), \quad \text{б) } a_{x_2x_1} = r_{x_2x_1}(S_{x_2}/S_{x_1}).$$

Тогда: (12)

$$\text{в) } a_{0x_1x_2} = \bar{x}_1 - a_{x_1x_2}\bar{x}_2; \quad \text{г) } a_{0x_2x_1} = \bar{x}_2 - a_{x_2x_1}\bar{x}_1.$$

Из уравнений видно, что

$$r = r_{x_1x_2} = r_{x_2x_1} = \sqrt{a_{x_1x_2} \cdot a_{x_2x_1}}. \tag{13}$$

Найденные через r коэффициенты парной регрессии минимизируют так называемую остаточную дисперсию $S^2_{\text{ост}}$:

$$S^2_{\text{ост}} = S^2_{x_1}(1 - r^2), \tag{14}$$

т. е. ту часть дисперсии величины, принятой в качестве функции, которая не объясняется влиянием аргумента.

Отсюда r^2 (так называемый коэффициент детерминации) показывает долю дисперсии функции, объясняемую аргументом при данном значении коэффициента регрессии a.

* Так как парная корреляция и регрессия являются частным случаем множественной корреляции и регрессии, вычислительная схема метода наименьших квадратов будет дана в разделе III. Подробные сведения о методе наименьших квадратов даны, например, в работе [3].

III. НЕЛИНЕЙНАЯ ПАРНАЯ КОРРЕЛЯЦИЯ И РЕГРЕССИЯ

Связи между экономическими явлениями часто бывают нелинейными. Например, связь между себестоимостью 1 *ц* зерна и урожайностью имеет вид гиперболы, связь между фондоотдачей на 1 тыс. руб. основных фондов и коэффициентом использования производственных мощностей представляет собою параболу, связь между расходами на питание (в пересчете на 1 потребительскую единицу) и числом детей в семье выражается степенным уравнением. При этом нередко эти связи нелинейны относительно параметров. Поэтому применяя регрессионный анализ для изучения связей такого рода необходимо преобразовывать кривую регрессии в функцию, линейную относительно параметров. Например, степенную функцию вида

$$x_1 = a_0 x_2^a \tag{15}$$

можно привести к линейной логарифмированием левой и правой части уравнения (15):

$$\ln x_1 = \ln a_0 + a \ln x_2. \tag{16}$$

Параметры этой функции можно найти вышеизложенными методами. В общем виде принцип преобразования выглядит следующим образом: имея функцию

$$x_1 = f(x_2), \tag{17}$$

нелинейную относительно параметров $\tilde{a}_0$ и $\tilde{a}$, найти функцию

$$\varphi(x) = \tilde{a}_0 + \tilde{a} g(x_2), \tag{18}$$

линейную относительно параметров $\widetilde{a}_0$ и $\widetilde{a}$.

Поскольку в практике экономического анализа класс функций, нелинейных относительно параметров регрессии, достаточно обширен, в рамках настоящей консультации не представляется возможным рассмотреть их. О приведении этих функций к линейным и предпосылках, предшествующих замене переменных и линеаризации, достаточно подробно говорится в работах [2, 3, 5]. Оценка значимости коэффициентов регрессии подобного рода кривых подробно разбираются в этих же работах.

В случае нелинейной зависимости между экономическими явлениями коэффициент корреляции теряет физический смысл. Для измерения тесноты криволинейной корреляционной связи применяется так называемое корреляционное отношение.

Корреляционное отношение $\eta_{x_1x_2}$ вычисляется по формуле:

$$\eta_{x_1x_2} = \sqrt{1 - \left[\sum_{j=1}^{n} (x_{1j} - \hat{x}_{1j})^2 \Big/ \sum_{j=1}^{n} (x_{1j} - \bar{x}_1)^2\right]}. \tag{19}$$

Этот показатель имеет тот же смысл, что и коэффициент корреляции, но только для криволинейной формы связи между переменными.

Величина корреляционного отношения заключена в пределах $0 \leqslant \eta \leqslant \leqslant 1$. В случае функциональной связи $\eta = 1$. Если же корреляционная связь отсутствует, то $\eta = 0$. Однако значение $\eta = 0$ не означает отсутствия причинной зависимости между экономическими явлениями.

Теоретически показатели $\eta_{x_1x_2}$ и $\eta_{x_2x_1}$ равноправны. Но $\eta_{x_1x_2} = \eta_{x_2x_1}$ лишь в случае линейной зависимости. В этом случае

$$\eta_{x_1x_2} = \eta_{x_2x_1} = r_{x_1x_2} = r_{x_2x_1}.$$

Во всех остальных случаях, т. е. при криволинейной зависимости $\eta_{x_1x_2} \neq \eta_{x_2x_1}$, а коэффициент корреляции по абсолютной величине меньше наименьшего из этих двух корреляционных отношений. Между $\eta_{x_1x_2}$ и $\eta_{x_2x_1}$ нет какой-либо простой зависимости. Так, x_1 может не быть коррелирована с x_2 и $\eta_{x_1x_2}=0$, тогда как $\eta_{x_2x_1}=1$.

В практической работе экономиста обычно интересует только одно из корреляционных отношений, а именно корреляционное отношение между фактором, принятым за функцию, и фактором, принятым за аргумент.

На соотношениях между η и r можно построить критерий криволинейности связи между x_1 и x_2. Упрощенная формула такого рода критерия имеет следующий вид

$$k=\frac{\sqrt{n}}{0{,}67449}\cdot\frac{1}{2}\sqrt{\eta^2-r^2}. \tag{20}$$

Если значение вычисленного $k>2{,}5$, то корреляционную связь можно считать криволинейной *.

В связи с тем, что η вычисляется для выборочной совокупности, корреляционное отношение генеральной совокупности $\eta_{\text{т}}$ находится в интервале:

$$\eta-t_{\text{т}}S_\eta\leqslant\eta_{\text{т}}\leqslant\eta+t_{\text{т}}S_\eta, \tag{21}$$

$$S_\eta=\sqrt{(1-\eta^2)/(n-2)}. \tag{22}$$

Значение $t_{\text{т}}$ берется из таблицы для заданного α и числа степеней свободы $\nu=n-2$. Фактическое значение t_{Φ} вычисляется по формуле:

$$t_{\Phi}=\eta\sqrt{(n-2)/(1-\eta^2)}. \tag{23}$$

Проверка значимости отличия от нуля выборочного значения η по t_{Φ} производится аналогично проверке значимости коэффициента парной корреляции.

Следует обратить особое внимание на то, что использование корреляционного отношения η имеет смысл только для функций криволинейных, но линейных относительно параметров. Для функций, нелинейных относительно параметров, корреляционное отношение не может служить точным измерителем тесноты связи. В случае приведения функции (17) к виду (18) корреляционное отношение η имеет смысл для функции (18), но не (17). Поскольку пока не предложено других измерителей тесноты связи

Вид уравнения регрессии	Формулы для вычисления коэффициентов эластичности	
$x_1=a_0+a_1x_2+a_2\ x_2^2$	$Э=(a_1+2_ax_2)\cdot\frac{x_2}{x_1}$	(24)
$x_1=a_0x_2^{a_1}$	$Э=a_1$	(25)
$x_1=a_0+a_1\ln x_2$	$Э=a_1\cdot\frac{1}{x_1}$	(26)
$x_1=\frac{a_0x_2}{x_2+a_1}$	$Э=\frac{a_0a_1}{(x_2+a_1)^2}\cdot\frac{x_2}{x_1}$	(27)
$x_1=a_0-\frac{a_1}{x_2+a_2}$	$Э=\frac{a_1}{(x_2+a_2)^2}\cdot\frac{x_2}{x_1}$	(28)
$x_1=a_0+\frac{a_1}{x_2-a_2}$	$Э=\frac{a_1}{(x_2-a_2)^2}\cdot\frac{x_2}{x_1}$	(29)
$x_1=\frac{a_0(x_2-a_2)}{(x_2+a_1)}$	$Э=\frac{a_0(a_1+a_2)}{(a_0+a_1)^2}\cdot\frac{x_2}{x_1}$	(30)

* Более точные, но более сложные формулы критерия криволинейности можно найти в работах [2, 3, 6].

для функций вида (17), на практике нередко η считается ее достаточной оценкой.

Для экономической интерпретации нелинейных связей между двумя явлениями удобно пользоваться коэффициентами эластичности. Выше приведены формулы вычисления коэффициентов эластичности для некоторых нелинейных функций, наиболее часто применяемых в экономической практике (см. таблицу).

Экономика характеризуется сложной взаимосвязью между большим числом различных явлений. Поэтому анализ парных связей имеет ограниченное применение. В данной консультации он рассмотрен так подробно из-за того, что ясное понимание задач и проблем множественной корреляции и регрессии невозможно без твердого знания анализа парных связей.

IV. МНОЖЕСТВЕННЫЙ КОРРЕЛЯЦИОННЫЙ И РЕГРЕССИОННЫЙ АНАЛИЗ

Применение множественного корреляционного и регрессионного анализа предполагает выполнение предпосылок, сформулированных выше для парного корреляционного и регрессионного анализа. Многофакторному анализу связей между экономическими явлениями предшествует конкретный качественный анализ объекта исследования.

Лишь после экономического качественного анализа объекта исследования можно переходить к сбору исходных данных применительно к выбранному методу исследования.

Для корреляционного и регрессионного анализа исходные данные удобно представить в виде матрицы:

$$X_{ji} = \begin{Vmatrix} x_{11} & x_{12} \dots x_{1i} \dots x_{1m} \\ x_{21} & x_{22} \dots x_{2i} \dots x_{2m} \\ \dots & \dots \\ x_{j1} & x_{j2} \dots x_{ji} \dots x_{jm} \\ \dots & \dots \\ x_{n1} & x_{n2} \dots x_{ni} \dots x_{nm} \end{Vmatrix}, \tag{31}$$

где n — число наблюдений ($j = 1, 2, 3, \dots, n$), m — число изучаемых переменных ($i = 1, 2, 3, \dots, m$).

Для измерения тесноты связи между двумя из рассматриваемых переменных применяются парные и различного порядка частные коэффициенты корреляции. Частные коэффициенты корреляции характеризуют корреляционную связь между двумя величинами без учета их взаимодействия с другими величинами.

Формула частного коэффициента корреляции между переменными x_1 и x_2 при исключении влияния x_3 имеет следующий вид:

$$r_{12.3} = (r_{12} - r_{13}r_{23}) \,/\, \sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}. \tag{32}$$

Коэффициенты частной корреляции, в которых исключено влияние только одного фактора, называются частными коэффициентами корреляции первого порядка. Тогда обычный парный коэффициент корреляции является частным коэффициентом корреляции нулевого порядка. Формула частного коэффициента корреляции любого порядка следующая:

$$r_{12.34,\dots,m} = \frac{r_{12.34,\dots,(m-1)} - (r_{1m.34,\dots,(m-1)} \cdot r_{2m.34,(m-1)})}{\sqrt{(1 - r^2_{1m.34,\dots,(m-1)})(1 - r^2_{2m.34,\dots,(m-1)})}}, \tag{33}$$

С вычислительной точки зрения более экономична следующая схема вычислений. Вначале вычисляются все возможные коэффициенты парной корреляции и сводятся в матрицу:

$$M = \begin{Vmatrix} 1 & r_{12} & r_{13} \ldots r_{1m} \\ r_{21} & 1 & r_{23} \ldots r_{2m} \\ \ldots & \ldots & \ldots \\ r_{m1} & r_{m2} & r_{m3} \ldots 1 \end{Vmatrix}. \tag{34}$$

Матрица (34) симметрична, так как $r_{ik} = r_{ki}$, где i, $k = 1,2,\ldots,m$. Для $i = k$ $r_{ik} = 1$. Тогда:

$$r_{ik\cdot 12,\ldots,i-1,i+1,\ldots,k-1,k+1,\ldots,m} = -D_{ik}/\sqrt{D_{ii}D_{kk}},$$

где D_{ik} — определитель, образованный вычеркиванием i-й строки и k-го столбца; D_{ii}, D_{kk} — определители, полученные вычеркиванием i-й строки и i-го столбца, k-й строки и k-го столбца для каждого определителя соответственно. Знак корня в знаменателе нужно брать положительный.

Парный коэффициент корреляции между факторами обычно не равен соответствующему частному коэффициенту. Если первый измеряет степень тесноты связи между явлениями без учета их взаимодействия с другими явлениями, то второй — при условии, что они связаны с другими явлениями.

Насколько это существенно, покажем на следующем примере. При изучении связей между производительностью труда в цементной промышленности (x_1), электровооруженностью труда (x_2) и часовой производительностью вращающихся печей (x_3) получен следующий коэффициент парной корреляции $r_{13.2} = 0{,}826$, тогда как частный коэффициент корреляции между факторами x_1 и x_3 $r_{13.2} = 0{,}344$. Коэффициент $r_{13.2}$ более чем в два раза меньше соответствующего парного коэффициента корреляции r_{13}. Частные коэффициенты корреляции изменяются в интервале $-1 \leqslant r_{ik.123,\ldots,m} \leqslant 1$.

Значимость и доверительный интервал для частного коэффициента корреляции определяются так же, как для парных коэффициентов корреляции (см. выше). Разница лишь в числе степеней свободы. Для частного коэффициента корреляции число степеней свободы $\nu = n - k$, где k — порядок частного коэффициента корреляции.

Для определения тесноты связи между фактором-функцией и несколькими факторами-аргументами служит множественный коэффициент корреляции. Он равен:

$$R_{i.\,12,\ldots,i-1,i+1,\ldots,m} = \sqrt{1 - D/D_{ii}}, \tag{35'}$$

где D — определитель матрицы парных коэффициентов корреляции.

Множественный коэффициент корреляции изменяется в интервале $0 \leqslant R \leqslant 1$. Если $R = 0$, то не имеется корреляционной связи между явлением i и остальными явлениями. В случае, если $R = 1$, эта связь функциональна. Среднеквадратическая ошибка S_R для коэффициента множественной корреляции определяется по следующей формуле:

$$S_R = (1 - R^2)/\sqrt{n - m - 1}. \tag{36}$$

Существенность R определяется по критерию F (Фишера):

$$F_\Phi = \frac{R^2(n - m - 1)}{(1 - R^2)m}, \tag{37}$$

с числом степеней свободы $\nu_1 = m$; $\nu_2 = n - m - 1$, где F_Φ — расчетное значение критерия F.

Теоретическое значение $F_т$ находится по таблице при заданном уровне значимости α (см. [4—6]). При небольшом числе наблюдений величина выборочного коэффициента множественной корреляции имеет тенденцию завышать долю вариации, характеризуемую остальными явлениями. Поэтому величину R следует корректировать по формуле:

$$\bar{R} = \sqrt{1 - \left[(1 - R^2)\left(\frac{n-1}{n-m-1}\right)\right]}, \tag{38}$$

где $\bar{R}$ — скорректированное значение множественного коэффициента корреляции.

Величина R^2 называется множественным коэффициентом детерминации. Она показывает, какая часть дисперсии функции объясняется за счет вариации линейной комбинации аргументов при данных значениях коэффициентов регрессии a_i.

По аналогии с парным коэффициентом детерминации и множественным коэффициентом детерминации можно ввести частный коэффициент детерминации. Соотношения между коэффициентами детерминации подробно рассматриваются в работе [5].

Одним из наиболее важных и неразработанных вопросов является вопрос о выборе вида уравнения регрессии для описания связи между экономическими явлениями. Иногда указания на соответствующий вид уравнения можно получить на основе теории данного экономического явления, а также из результатов аналогичных исследований других экономистов. Однако предварительные сведения для выбора типа функции часто отсутствуют. В этом случае форму связи можно определять эмпирически путем построения нескольких уравнений регрессии, отличающихся друг от друга как по своей алгебраической форме, так и по набору включенных в них переменных. Сравнение их между собой и выбор наиболее приемлемого для экономического анализа производится статистически с помощью коэффициентов множественной корреляции и F-критерия. Здесь также могут помочь личный практический опыт и знания исследователя.

Рассмотрим два наиболее изученных вида уравнения регрессии, широко используемых в экономической практике:

а) линейное уравнение:

$$x_1 = a_1 + a_2x_2 + \ldots + a_mx_m, \tag{39}$$

б) степенное уравнение:

$$x_1 = a_1x_2^{a_2}x_3^{a_3}, \ldots, x_m^{a_m}, \tag{40}$$

Как уже говорилось выше, исходная информация записывается в виде матрицы (31).

Требуется определить параметры уравнения регрессии $a_1, a_2, \ldots, a_m$ по методу наименьших квадратов.

Для уравнения (39) это равносильно требованию минимизации выражения*:

$$\sum_{j=1}^{n} (x_{1j} - \hat{x}_{1j})^2 = \sum_{j=1}^{n} [x_{1j} - (a_1 + a_2x_{2j} + a_3x_{3j} + \ldots + a_mx_{mj})]^2. \tag{41}$$

Для определения параметров $a_1, a_2, a_3, \ldots, a_m$ необходимо найти первые частные производные (41) по этим параметрам и приравнять их нулю. После соответствующих преобразований получится так называемая система

* В дальнейшем будем опускать индексы при знаке Σ.

нормальных уравнений:

$$na_1 + a_2 \sum x_{2j} + \ldots + a_m \sum x_{mj} = \sum x_{1j}, \tag{42}$$

$$a_1 \sum x_{2j} + a_2 \sum x_{2j}^2 + \ldots + a_m \sum x_{2j}x_{mj} = \sum x_{1j}x_{2j},$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$a_1 \sum x_{mj} + a_2 \sum x_{2j}x_{mj} + \ldots + a_m \sum x_{mj}^2 = \sum x_{1j}x_{mj}$$

или в матричной записи:

$$(\hat{X}^*\hat{X})A = \hat{X}^*X_1, \tag{43}$$

где A и X_1 — векторы-столбцы:

$$A = \begin{Vmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{Vmatrix}; \quad X_1 = \begin{Vmatrix} x_{11} \\ x_{12} \\ \vdots \\ x_{1n} \end{Vmatrix}, \tag{44}$$

$\hat{X}^*$ — транспонированная матрица $\hat{X}$. $\hat{X}$ — матрица значений аргументов, причем

$$\hat{X} = \begin{Vmatrix} x_{01} & x_{21} & x_{31} \ldots x_{i1} \ldots x_{m1} \\ \ldots & \ldots & \ldots\ldots\ldots\ldots \\ x_{0j} & x_{2j} & x_{2j} \ldots x_{ij} \ldots x_{mj} \\ \ldots & \ldots & \ldots\ldots\ldots\ldots \\ x_{0n} & x_{2n} & x_{3n} \ldots x_{in} \ldots x_{mn} \end{Vmatrix}, \tag{45}$$

где $x_{0i} \equiv 1$.

Тогда решение системы (42) запишется в виде:

$$A = (\hat{X}^*\hat{X})^{-1}\hat{X}^* \; X_1. \tag{46}$$

На практике параметры уравнения множественной регрессии находятся иногда также с помощью коэффициентов парной корреляции. Этот способ состоит в нахождении коэффициентов парной корреляции, определении по ним коэффициентов регрессии в стандартизованном масштабе (так называемых β-коэффициентов) и переходе от них к коэффициентам уравнения множественной регрессии в натуральном масштабе.

Если заранее не известен тип функции, описывающей связь между функцией и аргументами, ее можно представить в виде уравнения:

$$X_1 = a_1 + a_2x_2 + \ldots + a_mx_m + a_{22}x_2{}^2 + \ldots + a_{2m}x_m{}^2 + \\ + a_{2m+1}\,x_2x_3 + a_{2m+2}\,x_2x_4 + \ldots + \tag{47}$$

Переменные в (47) могут быть в любой целой степени, но в практической работе обычно достаточно полинома 2—3 порядка. Если теперь заменить

$$x_2{}^2 = x_{m+1};\; x_3{}^2 = x_{m+2}, \ldots, x_m{}^2 = x_{2m};\; x_2x_3 = x_{2m+1};\; x_2x_4 = x_{2m+2}, \ldots,$$

то получим уравнение:

$$x_1 = a_1 + a_2x_2 + \ldots + a_mx_m + a_{m+1}\,x_{m+1} + a_{2m}\,x_{2m} + \\ + a_{2m+1}\,x_{2m+1} + a_{2m+2}\,x_{2m+2} \ldots \tag{48}$$

Уравнение (48) — линейное уравнение, в котором все переменные находятся в первой степени. Следовательно, для определения коэффициентов регрессии к нему можно применить изложенннную выше вычислительную схему.

Для нахождения параметров степенной функции (40) ее необходимо привести к линейному виду. Прологарифмируем левую и правую части и произведем замену переменных:

$$\ln x_1 = z_1;\ \ln a_1 = a_1';\ \ln x_2 = z_2; \ldots;\ \ln x_m = z_m. \tag{49}$$

Тогда уравнение регрессии примет вид:

$$z_1 = a_1' + a_2 z_2 + \ldots + a_m z_m. \tag{50}$$

Параметры его находятся методом наименьших квадратов.

После того, как найдены параметры уравнения регрессии, необходимо проверить его значимость.

В случае анализа корреляционных связей сравнение регрессии считается значимым, если значим коэффициент множественной корреляции R. В случае анализа регрессионных связей мы не имеем права пользоваться для этих целей коэффициентом R, применимыми остаются лишь методы дисперсионного анализа: F-критерий дает общую проверку достоверности коэффициентов регрессии. К тому же с помощью величины F-критерия можно проверить гипотезу о том, что $a_1 = a_2 = \ldots = a_m = 0$.

$$F_\Phi = S_1^2 / S_2^2 \tag{51}$$

с числом степеней свободы $\nu_1 = m$ и $\nu_2 = n - m - 1$, где

$$S_1^2 = \frac{\Sigma(\hat{x}_1 - \bar{x}_1)^2}{m}, \quad S_2^2 = \frac{\Sigma(x_1 - \hat{x}_1)^2}{n - m - 1}. \tag{52}$$

Если величина F_Φ, полученная по уравнению (51), превышает табличное значение $F_т$ при заданном уровне значимости α, то гипотеза о равенстве $a_1 = a_2 = \ldots = a_m = 0$ отвергается.

Можно также проверить значимость отличия отдельных коэффициентов уравнения регрессии a_i от гипотетически предполагаемого значения $\tilde{a}_i$. Для каждого коэффициента регрессии i можно рассчитать величину $t_{i\Phi}$:

$$t_{i\Phi} = \frac{a_i - \tilde{a}_i}{Sa_i} = (a_i - \tilde{a}_i) \Big/ \sqrt{\sum_{j=1}^{n} (x_{1j} - \hat{x}_{1j})^2 C_{ii} / (n - m - 1)}, \tag{53}$$

где C_{ii} — i-й диагональный элемент матрицы, обратной к матрице системы нормальных уравнений (47). Эта функция проверки распределена как t Стьюдента с $n - m - 1$ степенями свободы.

Можно также рассчитать доверительный интервал, в котором с вероятностью $P \geqslant 1 - \alpha$ будет находиться коэффициент регрессии a_i. Он определяется по формуле:

$$a_i \pm t_т S_{a_i}. \tag{54}$$

Коэффициенты линейного уравнения множественной регрессии (39) показывают, на сколько единиц изменяется функция с изменением аргумента на одну единицу при закрепленном положении других аргументов на определенном уровне. Так что, в отличие от коэффициентов парной регрессии, они являются по сути дела коэффициентами частной регрессии. По аналогии можно ввести и частные коэффициенты эластичности. Для уравнения (39) они рассчитываются по формуле:

$$\hat{Э}_i = a_i \bar{x}_i / \bar{x}_1. \tag{55}$$

Частные коэффициенты эластичности показывают, на сколько процентов в среднем изменяется функция с изменением аргумента на 1% при фиксированном положении других аргументов.

Для уравнения вида (40) коэффициенты регрессии a_i равны соответствующим частным коэффициентам эластичности.

V. ПРОБЛЕМЫ И ТРУДНОСТИ, ВОЗНИКАЮЩИЕ ПРИ ПРИМЕНЕНИИ МЕТОДОВ МНОЖЕСТВЕННОГО КОРРЕЛЯЦИОННОГО И РЕГРЕССИОННОГО АНАЛИЗА В ЭКОНОМИЧЕСКИХ ИССЛЕДОВАНИЯХ

Автокорреляция. В практике экономико-статистических исследований для увеличения числа исходных данных часто пользуются так называемым методом заводо-лет. Сущность его заключается в следующем: показатели по каждому объекту исследования (например заводу) берутся за ряд лет и рассматриваются как отдельные наблюдения. Между показателями одного и того же объекта, взятыми за ряд лет, будет наблюдаться автокорреляция.

Под автокорреляцией понимается корреляция между последовательными отдельными наблюдениями временно́го ряда. Например, если показатели какого-либо объекта взяты за ряд лет: $x_{11}, x_{12}, \ldots, x_{1m}$, то под автокорреляцией для этого временно́го ряда понимают корреляцию между рядами: $x_{11}, x_{12}, \ldots, x_{1m-1}$ и $x_{12}, x_{13}, \ldots, x_{1m}$.

Наличие автокорреляции приводит к искажению величин среднеквадратических ошибок коэффициентов регрессии, что затрудняет построение их доверительных интервалов, а также проверку их значимости. Кроме того, автокорреляция приводит к сокращению числа эффективных наблюдений, так как показатели одного и того же объекта, скажем, за пять лет — это качественно нечто совершенно иное, чем показатели пяти объектов за один год. В первом случае мы фактически имеем дело лишь с одним независимым наблюдением, ввиду того, что состояние этого объекта в k-м году определяется состоянием этого объекта в предыдущие годы. Во втором же случае рассматривается пять независимых наблюдений.

Поэтому следует с особенно большой осторожностью относиться к исходным данным: необходимо проверять, имеется ли автокорреляция, и в случае значимости последней применять меры по ее устранению. Вопросы вычисления и исключения автокорреляции рассмотрены в работе [5].

Мультиколлинеарность. Нередко в практике экономико-статистического анализа в исследование включаются два или несколько линейно связанных аргументов. В таких случаях наряду с уравнением регрессии имеются и другие линейные соотношения. Наличие такой линейной связи между аргументами называется мультиколлинеарностью. Теоретически это может быть лишь в случае, если между некоторыми аргументами имеется коэффициент корреляции, равный единице. Но известно, что ошибки наблюдений снижают этот коэффициент, т. е. на практике он не равен единице только из-за ошибок, содержащихся в наблюдениях. Поэтому принимается, что два аргумента коллинеарны (для нескольких аргументов — мультиколлинеарны), если парный коэффициент корреляции (все парные коэффициенты корреляции) по абсолютной величине больше 0,8.

Полученные при наличии мультиколлинеарно связанных аргументов коэффициенты регрессии и другие оценки являются неустойчивыми, т. е. имеющими большую среднеквадратическую ошибку, что часто приводит к невозможности их экономической интерпретации. В этом случае необходимо один или несколько из линейно связанных аргументов из рассмотрения исключить. Вопрос о том, какую переменную исключить и какую оставить, следует решать, исходя из теоретических или практических соображений.

Соотношение между числом аргументов и числом наблюдений. При проведении множественного корреляционного и регрессионного анализа экономистам часто приходится иметь дело с малыми выборками. Причем число рассматриваемых аргументов бывает очень велико. Это приводит к тому, что сильно увеличиваются доверительные интервалы коэффициентов регрессии и уменьшается их достоверность. В таких случаях экономическая интерпретация коэффициентов регрессии бывает крайне ненадежна.

Чтобы избежать этого, предлагают следующий грубый критерий соотношения между числом аргументов и числом наблюдений: при построении уравнения множественной регрессии число наблюдений в выборке должно превышать число аргументов примерно в восемь раз.

В литературе также приводятся некоторые критерии, связывающие коэффициент множественной корреляции с числом аргументов и числом наблюдений. Рекомендуется, например, соотношение: $R^2=(2m+1)\ /\ (2n-3)$.

Отсев несущественных аргументов. Одним из важнейших требований, предъявляемых к экономико-статистическим исследованиям, является отбор наиболее существенных аргументов, влияющих на изучаемое явление. Однако сложность и многообразие экономических явлений, наличие в них перекрещивающихся тенденций часто приводят к тому, что теоретический анализ не позволяет однозначно ответить на вопрос о влиянии данного круга отобранных аргументов на изучаемое явление. Поэтому в таких случаях целесообразно применять двухстадийный отбор: на первой стадии отбираются все аргументы, связанные с исследуемым явлением, численное значение которых можно определить, а затем, уже во второй стадии, на основе математико-статистического анализа отбираются существенно влияющие аргументы.

В таком случае встает вопрос об отсеве несущественных аргументов. Мы несколько коснулись этого вопроса, когда рассматривали проблему мультиколлинеарности. Проблема отсева несущественных аргументов полностью не решена, да и вряд ли ее можно окончательно решить на все случаи жизни. В литературе приводится большое число способов решения этой задачи (см., например, [5]). Все они обладают как достоинствами, так и недостатками. Во всяком случае возможность применения каждого из них нужно рассматривать в конкретном исследовании.

Не вызывает никакого сомнения только один способ отсева несущественных аргументов, предложенный Андерсоном: аргумент можно исключать из уравнения регрессии тогда, когда средняя квадратическая ошибка коэффициента регрессии превышает абсолютное значение вычисленного коэффициента регрессии.

Некоторые проблемы экономической интерпретации результатов регрессионного анализа. Каждый экономист, берущий на вооружение методы корреляционного и регрессионного анализа, в определенный момент задает себе вопросы :1) можно ли вообще применять методы математической статистики в экономике и если можно, то какие?; 2) как интерпретировать полученные результаты?

В последние годы наступил новый этап переосмысливания результатов применения корреляционного и регрессионного анализа к экономическим задачам. Вновь были подняты вопросы соотношений между статистическими и причинными связями. Следствием этого явилось создание так называемого структурного анализа, который пытается модифицировать метод наименьших квадратов, считая оценки последнего «смещенными». Вместо одного уравнения множественной регрессии, включающего как причинно, так и непричинно связанные факторы, для целей описания объекта и управления им вводятся системы рекурсивных и нерекурсивных уравнений, включающих в себя только причинно связанные переменные. В советской литературе эти вопросы, к сожалению, пока не рассматриваются.

При проведении корреляционного и регрессионного анализа важно выделить круг экономических задач, для которых целиком выполняются все предпосылки математической статистики: наличие гипотетической генеральной совокупности, равновероятность попадания отдельных наблюдений в выборку и т. п.

Такой же конкретности требует и интерпретация результатов применения методов множественной корреляции и регрессии к анализу связей между экономическими явлениями.

Здесь легко прийти к ложным выводам. Корреляционный и регрессионный анализ предназначен лишь для измерения степени взаимосвязанности тех или иных явлений. Поэтому результаты нельзя интерпретировать в понятиях причинно-следственной терминологии без каких-либо предварительных предположений. После того, как такие предположения выдвинуты, множественный корреляционный и регрессионный анализ может подтвердить или опровергнуть их и дать количественные оценки влияния различных факторов. Часто ошибочное истолкование результатов связано с существенными недостатками исходного материала, слишком упрощенным подходом к задаче. Следовательно, применение регрессии и корреляции к решению экономических задач должно приводиться с крайней осторожностью.

Надо всегда иметь в виду различие в задачах, которые могут возникнуть в практической работе при рассмотрении стохастически связанных переменных, для того, чтобы сознательно выбирать и рационально строить применяемые методы, а также правильно истолковывать результаты, получаемые с их помощью. Это невозможно без четкого и сознательного понимания цели исследования. Здесь наиболее важной предпосылкой успеха является критическое отношение к предмету и методу исследования.

ЛИТЕРАТУРА

1. С. А. Айвазян. Основные понятия теории вероятностей и математической статистики. Экономика и матем. методы, 1966, т. 2, вып. 5.
2. С. А. Айвазян. Применение методов корреляционного и регрессионного анализов к обработке результатов эксперимента. Заводск. лаборатория, 1964, № 7—8.
3. А. Хальд. Математическая статистика с техническими приложениями. М., Изд-во иностр. лит., 1956.
4. Ж. Мот. Статистические предвидения и решения на предприятии. М., «Прогресс», 1966.
5. М. Езекиел, К. Фокс. Методы анализа корреляции и регрессии. М., «Статистика», 1966.
6. Дж. Э. Юл, М. Дж. Кендал. Теория статистики. М., Госстатиздат, 1960.
7. Ф. Миллс. Статистические методы. М., Госстатиздат, 1958.
8. Р. А. Фишер. Статистические методы для исследователей. М., Госстатиздат, 1958.
9. А. А. Чупров. Основные проблемы теории корреляции. М., Госстатиздат, 1960.

Поступила в редакцию
6 X 1966

ЭКОНОМИКА
И МАТЕМАТИЧЕСКИЕ МЕТОДЫ

1967
том III, вып. 3

ПРАКТИЧЕСКИЙ ОПЫТ

ОПРЕДЕЛЕНИЕ ВНУТРИРАЙОННОЙ СПЕЦИАЛИЗАЦИИ СЕЛЬСКОХОЗЯЙСТВЕННОГО ПРОИЗВОДСТВА НА ПЕРСПЕКТИВУ

И. Д. БЛАЖ

(Москва)

В [1] приведена методика составления экономико-математической задачи по отчетным данным колхозов Оргеевского района Молдавской ССР. Решение этой задачи показывает, какими производственными резервами располагают колхозы данного района и как они могут быть использованы за счет более обоснованной специализации производства.

Однако при планировании специализации сельскохозяйственного производства на перспективу возникает необходимость использовать не только отчетные данные, но и большое количество разнообразных нормативных материалов, а также учесть более широкий круг производственных и экономических условий.

Поэтому нами на примере колхозов Оргеевского района Молдавской ССР составлена другая экономико-математическая задача по определению оптимального варианта специализации сельского хозяйства на ближайшую перспективу (1970 г.), которая формулируется также на основе общей задачи линейного программирования. При составлении этой задачи использованы основные положения разработанной в ВНИЭСХ методики [2], которые конкретизировались и уточнялись применительно к местным условиям колхозов района.

В составленной задаче специализация сельскохозяйственного производства на перспективу определяется в разрезе выделенных в [1] трех групп колхозов. Поэтому система неравенств и уравнений, характеризующих нормы производственных затрат и наличие производственных ресурсов, строится по каждой группе колхозов в виде отдельного матричного блока. Каждый блок матрицы содержит необходимую информацию, относящуюся к данной группе колхозов. Эта информация включает: перечень переменных, систему условий и ограничений производства с их технико-экономическими коэффициентами по переменным данного блока и оценки переменных в целевой функции. Матрица первого блока с некоторыми сокращениями приводится в табл. 1.

Переменными величинами в задаче являются все культуры, выращиваемые в колхозах Оргеевского района, и все виды разводимого скота. Условия и ограничения производства представлены в матрице задачи тремя группами: 1) ограничения по использованию производственных ресурсов; 2) ограничения, вызванные биологическими особенностями сельскохозяйственных культур и отраслей; 3) ограничения по обязательному производству основных видов сельскохозяйственной продукции.

В составленной задаче единицей измерений переменных по сельскохозяйственным культурам является гектар посева, по свиноводству — центнер привеса, по крупному рогатому скоту, овцеводству и птицеводству — структурная маточная голова.

Оптимальный план специализации сельскохозяйственного производства по группам колхозов (на перспективу) находится одновременно с определением урожайности культур и продуктивности скота, рационов кормления скота, распределением удобрений, техники и капиталовложений.

По условиям задачи урожайность сельскохозяйственных культур находится в функциональной зависимости от количества основных элементов питания (азота, фосфора и калия), вносимых в почву с органическими и минеральными удобрениями. Количество питательных элементов, требуемых каждой культурой для получения соответствующего уровня урожайности, рассчитывается как разность между общим выносом этих элементов вместе с урожаем и запасом питательных элементов, содержащихся в почве на 1 *га*. В связи с этим каждая культура представлена в задаче двумя переменными — с минимально и максимально возможными уровнями урожайности. Нижняя граница — это фактически достигнутый уровень урожайности в колхозах каждой группы в среднем за 1961, 1962 и 1964 гг. Верхняя граница — это уро-