

А.А.Нургалиева

*Карагандинский государственный университет им. Е.А.Букетова*

## **Процесс корреляционно-регрессионного анализа экономических явлений в сельском хозяйстве**

В статье подчеркнута, что в социально-экономических явлениях преобладают стохастические взаимосвязи. Отмечено, что эти связи часто носят корреляционный характер, т.е. одному значению изучаемого признака-фактора может соответствовать много значений результативного признака. Показано, что процесс корреляционно-регрессионного анализа экономических явлений на сельскохозяйственных предприятиях состоит из ряда последовательных этапов. Дана характеристика коэффициентам, применяемым для измерения тесноты связи между результатом и признаками-факторами. Обращено внимание на то, что на основе парных коэффициентов корреляции составляется уравнение множественной регрессии между результативным и факторными признаками в стандартизованном масштабе. Рассмотрены возможности использования корреляционно-регрессионных моделей для экономико-статистического анализа явлений и процессов в аграрном секторе экономики. Доказано, что метод наименьших квадратов может быть использован при вычислении параметров уравнений регрессии криволинейной формы связи.

*Ключевые слова:* связь, корреляция, регрессия, параметр, этап, теснота, коэффициент, анализ, величина, прогноз.

### *Функциональная и стохастическая взаимосвязи признаков*

В общественном производстве все процессы находятся в тесной зависимости. Различают функциональную и стохастическую взаимосвязь признаков. Под функциональными понимают такие взаимосвязи, при которых величина результативного признака изменяется на одну и ту же величину, с изменением факториальных признаков. Такого рода зависимости, встречающиеся в экономике, изучаются с помощью индексного приема. В социально-экономических явлениях преобладают стохастические взаимосвязи, когда результативный признак зависит от значений признаков-факторов (например, продуктивность скота зависит от породы, уровня кормления, ухода, содержания и др.). Причем эти связи часто носят корреляционный характер, т.е. одному значению изучаемого признака-фактора может соответствовать много значений результативного признака, варьирующих в различных направлениях.

Корреляционные связи бывают однофакторные и многофакторные (по количеству признаков связи); положительные и отрицательные (по направленности); прямолинейные и криволинейные (по аналитическому выражению) [1–4].

Характер корреляционной связи определяют с помощью корреляционной решетки, построенной в прямоугольных осях координат.

Если частоты распределяются ближе к диагонали, то между признаками будет обнаружена высокая связь. Размещение частот близко к диагонали, пересекающей решетку с левого нижнего в правый верхний угол, свидетельствует о положительной направленности, а с верхнего левого в правый нижний угол — об отрицательной. Дугообразное размещение частот в решетке характеризует криволинейную связь, а беспорядочное — отсутствие связи.

### *Значение и основные этапы процесса корреляционно-регрессионного анализа экономических явлений*

Корреляционный и регрессионный анализ позволяет решать такие задачи, которые пока другими методами выполнить нельзя, например, определение совместного и отдельного влияния многих взаимно связанных и одновременно действующих факторов на какой-то процесс или явление. С помощью корреляционно-регрессионного анализа можно оценить силу связи между отдельными факторами (факторными признаками), между факторами и результативным явлением или процессом (результативным признаком) и подобрать уравнение регрессии, которое определяет форму данной связи [5–8].

Процесс корреляционно-регрессионного анализа экономических явлений состоит из следующих этапов: 1) выбор факторов-аргументов и предварительная обработка статистических данных; 2) оценка тесноты связи между отдельными признаками и выявление формы связи; 3) разработка мо-

дели изучаемого экономического явления и ее анализ; 4) использование результатов анализа для совершенствования планирования и управления данным явлением.

Во избежание ложных корреляций отбор факторов-аргументов, влияющих на данный экономический процесс или явление, должен производиться компетентными работниками — специалистами в данной области экономики.

Проверка статистической однородности данных осуществляется в два приема. Сначала выявляются и исключаются значения признаков, резко отличающихся от всей совокупности. Затем проводится математико-статистическое исследование однородности данных путем проверки независимости выборок и их принадлежности к единой, нормально распределенной генеральной совокупности.

Разработка регрессионной модели изучаемого экономического процесса или явления осуществляется на основе метода наименьших квадратов, согласно которому обеспечивается наилучшее приближение оценок результативного признака, рассчитанных по уравнению регрессии, к их фактическим значениям.

При разработке регрессионной модели следует избегать автокорреляции и мультиколлинеарности переменных. Автокорреляция может иметь место в тех случаях, когда наблюдения производятся за определенные периоды и существует связь между последующими и предыдущими данными. В этом случае нарушается принцип статистической независимости данных. Наиболее простым способом исключения автокорреляции является включение фактора времени в модель в виде самостоятельной переменной. Мультиколлинеарность имеет место при наличии линейной зависимости между некоторыми переменными. Мультиколлинеарность можно ликвидировать за счет расширения исходной информации, изъятия из модели одного из коррелирующих между собой факторов или путем введения искусственной ортогональности [9–11].

Важнейшими параметрами, характеризующими регрессионную модель, являются [12–15]:

- а) коэффициенты парной корреляции, которые определяют силу связи между двумя признаками;
- б) коэффициент множественной корреляции, который определяет связь результативного признака с совокупностью факторных признаков;
- в) коэффициенты частной детерминации, которые определяют влияние вариации каждого факторного признака в отдельности на вариацию результативного признака;
- г) коэффициент множественной детерминации, который определяет удельный вес совместного влияния всех включенных в модель факторных признаков на вариацию результативного признака;
- д) частные коэффициенты эластичности, которые определяют влияние отдельных факторных признаков на результативный признак в едином масштабе в процентах.

Результаты корреляционно-регрессионного анализа позволяют выявить факторы, оказывающие существенное влияние на исследуемый экономический процесс или явление. Они могут также быть использованы для разработки норм и нормативов, для нахождения передовых коллективов, изучения и распространения их опыта, а также для других целей, направленных на совершенствование планирования и управления экономическими процессами.

#### *Измерение тесноты связи между результатом и признаками-факторами*

Для измерения тесноты связи между результатом и признаками-факторами используются коэффициенты: корреляции линейной ( $r$ ), корреляционного отношения ( $\eta$ ), ассоциации ( $r_a$ ), взаимной сопряженности ( $r_c$ ), корреляции рангов ( $r_p$ ), множественной ( $r_m$ ), индекса корреляций ( $I_r$ ), регрессии ( $R$ ). Коэффициент линейной корреляции — показатель, отображающий направление и меру тесноты связи между признаками при прямолинейных (или близких к ним) взаимозависимостях. Он колеблется в пределах от  $\pm 0$  до  $\pm 1$ . Знак «+» означает прямую, а знак «-» — обратную связь.

На практике плохая связь характеризуется коэффициентом корреляции от  $\pm 0,16$  до  $\pm 0,20$ , слабая — от  $\pm 0,21$  до  $\pm 0,30$ , умеренная — от  $\pm 0,31$  до  $\pm 0,40$ , средняя — от  $\pm 0,41$  до  $\pm 0,60$ , высокая — от  $\pm 0,61$  до  $\pm 0,80$ , очень высокая — от  $\pm 0,81$  до  $\pm 0,90$ , полная связь — от  $\pm 0,91$  до  $\pm 1,0$ .

При малых выборках наиболее часто коэффициент линейной корреляции ( $r$ ) исчисляют по формуле

$$r = \frac{\sum(x - \bar{x}) \cdot (y - \bar{y})}{\sqrt{\sum(x - \bar{x})^2 \cdot \sum(y - \bar{y})^2}},$$

где  $x, y$  — значения изучаемых признаков;  $\bar{x}, \bar{y}$  — средние величины по каждому признаку.

Для больших выборок коэффициент корреляции исчисляются по формуле

$$r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \sigma_y},$$

где  $\overline{xy}, \bar{x}, \bar{y}$  — соответствующие средние изучаемых признаков;  $\sigma_x$  и  $\sigma_y$  — среднеквадратические отклонения.

Обычно предварительная обработка статистических данных при корреляционно-регрессионном анализе заканчивается составлением матрицы парных коэффициентов корреляции (табл. 1).

Матрица парных коэффициентов — квадратная и симметричная в отношении главной диагонали. Парные коэффициенты корреляции характеризуют тесноту связи между отдельными признаками и их абсолютная величина не может быть больше единицы ( $|r| \leq 1$ ).

Т а б л и ц а 1

**Макет матрицы парных коэффициентов корреляции**

	$y$	$x_1$	$x_2$	$x_3$	...	$x_n$
$y$	1	$r_{yx_1}$	$r_{yx_2}$	$r_{yx_3}$	...	$r_{yx_n}$
$x_1$	$r_{x_1y}$	1	$r_{x_1x_2}$	$r_{x_1x_3}$	...	$r_{x_1x_n}$
$x_2$	$r_{x_2y}$	$r_{x_2x_1}$	1	$r_{x_2x_3}$	...	$r_{x_2x_n}$
$x_3$	$r_{x_3y}$	$r_{x_3x_1}$	$r_{x_3x_2}$	1	...	$r_{x_3x_n}$
...	...	...	...	...	1	...
$x_n$	$r_{x_ny}$	$r_{x_nx_1}$	$r_{x_nx_2}$	$r_{x_nx_3}$	...	1

Расчет парных коэффициентов корреляции — трудоемкий процесс и осуществляется, как правило, на ЭВМ. На основе парных коэффициентов корреляции составляется уравнение множественной регрессии между результативным и факторными признаками в стандартизованном масштабе.

Стандартизованным называется такой масштаб измерения, при котором все фактические значения коррелируемых показателей заменяются расчетными по формуле

$$t_{xi} = \frac{x_i - \bar{x}}{\sigma_{xi}},$$

где  $x_i$  — значение  $i$ -й переменной в натуральном масштабе.

На практике приходится встречаться не только с прямолинейными, но и с криволинейными зависимостями (например, размер фермы и затраты на производство единицы продукции; урожай и осадки, урожай и внесение минеральных и органических удобрений и др.). Криволинейные зависимости принимают формы гиперболы, параболы, логарифмических кривых и т.д.

Для криволинейных зависимостей корреляционное отношение ( $\eta$ ) является наиболее точным измерителем тесноты связи между признаками. Формула его исчисления

$$\eta = \sqrt{\frac{\sigma_{xy}^2}{\sum \sigma_y^2}} = \sqrt{\frac{\sum (\bar{y}_x - \bar{y})^2}{\sum (y - \bar{y})^2}},$$

где  $\sigma_{xy}^2$  — межгрупповая дисперсия результативного признака по фактическому;  $\sigma_y^2$  — общая дисперсия результативного признака.

Тесноту связи между двумя признаками можно определить и с помощью коэффициента корреляции рангов [16; 227, 228]. В этом случае взаимосвязанные признаки размещаются по ранжиру основного признака, их количественные величины расцениваются в рангах (порядковых величинах ранжира). Затем определяется разность рангов ( $d$ ), которая возводится в квадрат. Коэффициент рангов определяется по формуле

$$r_p = 1 - \frac{n \sum d^2}{n(n^2 - 1)},$$

где  $d$  — разность между рангами изучаемых признаков;  $n$  — число наблюдений.

Используя коэффициент рангов, установим зависимость между количеством коров на ферме и себестоимостью молока (табл. 2).

Таблица 2

**Расчет коэффициента корреляции рангов по животноводческим хозяйствам Шетского района Карагандинской области**

Поголовье коров на ферме	Ранг по $x$	Себестоимость 1 ц молока $y$ , тыс. тенге	Ранг по $y$	Разность между рангами	Квадрат разности	Решение
300	1	20,0	6	-5	25	$1 - \frac{6 \cdot 67,5}{6(36-1)} = -0,93$
400	2	19,0	5	-3	9	
500	3	18,0	4	-1	1	
600	4	15,0	3	+1,5	1,25	
700	5	15,0	2	+2,5	6,25	
800	6	12,5	1	+5	25	
—	—	—	—	—	67,5	

Коэффициент свидетельствует о полной связи между концентрацией коров на фермах и себестоимостью молока — связь обратная.

Определение коэффициента корреляции рангов упрощает расчеты.

*Применение регрессионных моделей для экономико-статистического анализа изучаемых явлений*

В процессе корреляционно-регрессионного анализа используют уравнения прямой, параболы, гиперболы, показательной кривой и др. Выбор уравнения регрессии должен быть теоретически и логически обоснован с учетом социально-экономической сущности изучаемого явления, процесса.

В ходе теоретического обоснования необходимо выяснить механизм взаимосвязи результативного признака и признаков-факторов с учетом сущности явления.

В процессе выбора модели уравнения учитывают также характер динамического ряда. Уравнение прямой используют, когда на протяжении изучаемого периода сохраняется более или менее стабильный абсолютный прирост явления. При зигзагообразном возрастающем изменении динамического ряда без стабильного его роста и снижения применяют уравнение параболы. При отрицательной направленности и изменении динамических рядов по зигзагообразной снижающейся, а затем по

повышающейся кривой, без стабильного ее снижения и роста, используют уравнение гиперболы или показательной кривой.

Анализ показывает, что нельзя повсеместно и во всех случаях применять один и тот же вид аналитического уравнения. Обычно при аналитическом выравнивании ряда динамики на базе регрессионных моделей подбирается аналитическая функция (кривая), наиболее точно характеризующая закономерность развития данного явления или процесса во времени. Найденная функция позволяет получить выравненные значения уровней ряда динамики (его теоретические оценки), т.е. те уровни, которые наблюдались бы, если бы динамика явления или процесса полностью совпадала с выбранной кривой (линией регрессии).

Для расчета параметров кривой используют метод наименьших квадратов, который требует, чтобы сумма квадратов отклонений значений, лежащих на линии регрессии (теоретических оценок уровней), от фактических значений уровней была минимальной, т.е. чтобы соблюдалось условие

$$\sum (y_t - \hat{y}_t)^2 \rightarrow \min,$$

где  $y_t$  — фактическое значение уровня ряда динамики;  $\hat{y}_t$  — теоретическая оценка уровня.

Аналитическим выражением прямолинейного тренда является функция

$$y_t = a + bt, \tag{1}$$

где  $a$  и  $b$  — параметры;  $t$  — очередной номер уровня ряда с начала отсчета.

Из курса математической статистики известно, что требование метода наименьших квадратов при прямолинейном тренде приводит к системе двух нормальных уровней относительно параметров  $a$  и  $b$ :

$$\left. \begin{aligned} \sum y_t &= a_n + b \sum t, \\ \sum y_t t &= a \sum t + b \sum t^2, \end{aligned} \right\}$$

где  $n$  — количество уровней в ряду динамики.

Решая данную систему линейных уравнений, получаем формулы для вычисления значений параметров:

$$a = \frac{\sum y_t - b \sum t}{n} = \frac{\sum t^2 \sum y_t - \sum t y_t \sum t}{n \sum t^2 - (\sum t)^2} \tag{2}$$

и

$$b = \frac{n \sum y_t t - \sum t \sum y_t}{n \sum t^2 - (\sum t)^2}. \tag{3}$$

Ниже дан пример расчета параметров линейной регрессии. Имеются следующие данные об урожайности зерновых по годам (ц/га), которые были собраны и обобщены применительно к условиям передовых крестьянских хозяйств Нуринаского района Карагандинской области (2000–2011 гг.):

2000 г. – 19,2,	2006 г. – 25,4,
2001 г. – 21,5,	2007 г. – 21,2,
2002 г. – 20,5,	2008 г. – 29,0,
2003 г. – 18,5,	2009 г. – 27,4,
2004 г. – 23,0,	2010 г. – 20,4,
2005 г. – 23,4,	2011 г. – 27,9.

Эмпирический ряд, несмотря на колебания, отображает тенденцию роста урожайности.

Для установления прогноза урожайности используем уравнение прямой  $y_t = a + bt$ . Параметры прямой определяются по формулам (2) и (3). Для исчисления параметров находим необходимые величины в таблице 3.

$$a = \frac{(650 \cdot 277,4) - (1894,7 \cdot 78)}{(12 \cdot 650) - (78 \cdot 78)} = \frac{180310,0 - 147786,6}{7800 - 6084} = \frac{32523,6}{1716} = 18,95 \text{ ц/га};$$

## Исчисление необходимых величин для отыскания параметров уравнения прямой

Годы	Порядковый номер года, $t$	Фактическая урожайность зерновых $y_t$ , ц/га	Произведение порядкового номера года на урожайность, $y_t t$	Квадрат порядкового номера $t^2$
2000	1	19,2	19,2	1
2001	2	21,5	43,0	4
2002	3	20,5	61,5	9
2003	4	18,5	74,0	16
2004	5	23,0	115,0	25
2005	6	23,4	140,4	36
2006	7	25,4	177,8	49
2007	8	21,2	169,6	64
2008	9	29,0	261,0	81
2009	10	27,4	274,0	100
2010	11	20,4	224,4	121
2011	12	27,9	334,8	144
$\Sigma$	78	277,4	1894,7	650

$$b = \frac{(12 \cdot 1894,7) - (78 \cdot 277,4)}{(12 \cdot 650) - (78 \cdot 78)} = \frac{22736,4 - 21637,2}{7800 - 6084} = \frac{1099,2}{1716} = 0,64 \text{ ц/га.}$$

Параметр  $a$  показывает постоянную исходную урожайность зерновых культур для данного ряда.

Параметр  $b$  характеризует среднюю прибавку урожайности зерновых за год. Следовательно, расчетный уровень урожайности в 2011 г. должен равняться

$$\bar{y}_t = a + bt = 18,95 + 0,64 \cdot 12 = 26,63 \text{ ц/га.}$$

Расхождение с фактической урожайностью составляет 4,6 %, что характеризует обоснованность расчетных параметров.

Метод наименьших квадратов может быть использован при вычислении параметров уравнений регрессии криволинейной формы. В этом случае приходится сначала функцию «линеаризовать». В практике криволинейного выравнивания широко распространены два вида линеаризации: с помощью натуральных логарифмов ( $\ln$ ) и обратных преобразований  $\left(\frac{1}{t}\right)$ . Так, экспоненциальную кривую

$y_t = ab^t$  можно превратить в линейную логарифмированием, в результате которого получаем уравнение  $\ln y_t = \ln a + \ln b t$ . Заменяя  $\ln y_t$  выражением  $y'$ ,  $\ln a$  — выражением  $a'$  и  $\ln b$  — выражением  $b'$ , получаем линеаризованную функцию  $y' = a' + b't$ . В этом уравнении значения  $a'$  и  $b'$  могут быть найдены на основе приведенных выше формул (2) и (3). Затем вычисляем значение  $y'$  по формуле (1) и с помощью антилогарифмов находим значение  $y$ .

## Список литературы

- 1 Гришин А.Ф., Кочерова Е.В. Статистические модели: построение, оценка, анализ: Учеб. пособие. — М.: Финансы и статистика, 2005. — 416 с.
- 2 Рыжова В.В., Кузнецова Л.А. Математические методы в анализе хозяйственной деятельности предприятий. — М.: Финансы, 1970. — 88 с.
- 3 Федосеев В.В., Гармаш А.Н., Дайитбегов Д.М., Орлова И.В., Половников В.А. Экономико-математические методы и прикладные модели: Учеб. пособие. — М.: ЮНИТИ, 1999. — 391 с.
- 4 Ричард Томас. Количественные методы анализа хозяйственной деятельности: Пер. с англ. — М.: Изд-во «Дело и Сервис», 1999. — 432 с.
- 5 Чавкин А.М. Методы и модели рационального управления в рыночной экономике: разработка управленческих решений: Учеб. пособие. — М.: Финансы и статистика, 2001. — 320 с.

- 6 Терехов Л.Л., Куценко В.А., Сиднев С.П. Экономико-математические методы и модели в планировании и управлении. — Киев: Вища шк., 1984. — 231 с.
- 7 Шикин Е.В., Чхартишвили А.Г. Математические методы и модели в управлении: Учеб. пособие. — М.: Дело, 2000. — 440 с.
- 8 Эренберг А. Анализ и интерпретация статистических данных: Пер. с англ. — М.: Финансы и статистика, 1981. — 406 с.
- 9 Замков О.О., Толстопятенко А.В., Черемных Ю.Н. Математические методы в экономике: Учебник. — М.: МГУ им. М.В.Ломоносова, Изд-во «ДИС», 1997. — 368 с.
- 10 Кейн Э. Экономическая статистика и эконометрия. Введение в количественный экономический анализ. Вып. 2: Пер. с англ. — М.: Статистика, 1977. — 232 с.
- 11 Демиденко Е.З. Линейная и нелинейная регрессии. — М.: Финансы и статистика, 1981. — 302 с.
- 12 Елисеева И.И., Курьшева С.В., Костеева Т.В., Бабаева И.В., Михайлов Б.А. Эконометрика: Учебник. — М.: Финансы и статистика, 2001. — 344 с.
- 13 Ларионов А.И., Юрченко Т.И., Новоселов А.Л. Экономико-математические методы в планировании: Учебник. — М.: Высш. шк., 1991. — 240 с.
- 14 Шелобаев С.И. Математические методы и модели в экономике, финансах, бизнесе: Учеб. пособие. — М.: ЮНИТИ-ДАНА, 2000. — 367 с.
- 15 Карасев А.И., Кремер Н.Ш., Савельева Т.И. Математические методы и модели в планировании: Учеб. пособие. — М.: Экономика, 1987. — 240 с.
- 16 Венецкий И.Г., Венецкая В.И. Основные математико-статистические понятия и формулы в экономическом анализе. — М.: Статистика, 1974. — 280 с.

А.А.Нурғалиева

### **Ауыл шаруашылығындағы экономикалық құбылыстарды корреляциялық-регрессиялық талдау**

Әлеуметтік-экономикалық құбылыстарда стохастикалық өзара байланыстардың басым болатындығы дәріптелген. Бұл байланыстардың көбінесе корреляциялық сипатқа ие екендігі, яғни зерттелетін белгі-фактордың бір мәніне нәтижелік белгінің көптеген мәндері сәйкес келетіндігі айтылған. Ауыл шаруашылығындағы экономикалық құбылыстарды корреляциялық-регрессиялық талдау үдерісі бірқатар тізбекті кезендерден тұратындығын көрсетті. Нәтиже және белгі-факторлар арасындағы байланыстың қаншалықты тығыз екендігін анықтауға қолданылатын коэффициенттерге сипаттама берілген. Корреляцияның жұп коэффициенттері негізінде нәтижелік және факторлық белгілер арасындағы стандартты деңгейде көптік регрессия тендеуі құрылатындығына көңіл бөлінген. Экономиканың аграрлық секторындағы құбылыстар мен үдерістерді экономикалық-статистикалық талдау үшін корреляциялық-регрессиялық модельдерді қолдану мүмкіндіктері қарастырылған. Байланыстың қисық сызықты түрдегі регрессиясы тендеулерінің параметрлерін есептеуде ең төмен квадраттар тәсілін қолдануға болатындығы дәлелденген.

A.A.Nurgaliyeva

### **Process of the correlation and regression analysis of economic events in agriculture**

It is emphasized that the social and economic phenomena stochastic interrelations prevail. It is noted that these communications often have correlation character, i.e. to one value of a studied sign factor there can correspond many values of a productive sign. It is shown that process of the correlation and regression analysis of economic events at the selkoxhozyaychvenny enterprises consists of a number of consecutive stages. The characteristic is given to the factors applied to measurement of narrowness of communication between result and signs factors. That on the basis of pair factors of correlation the equation of multiple regression between productive and factorial signs in the standardized scale is worked out is paid attention. Possibilities of use of correlation and regression models for the economical and statistical analysis of the phenomena and processes in agrarian sector of economy are considered. It is proved that the method of the smallest squares can be used at calculation of parameters of the equations of regression of a curvilinear form of communication.