УДК 338.45:[622.33:004](574)

Р.С.Каренов

Карагандинский государственный университет им. Е.А.Букетова (E-mail: karenov@mail.ru)

Методика применения способов парной корреляции в экономических исследованиях

Рассмотрены основные предпосылки применения корреляционно-регрессионного анализа в экономических исследованиях. Проанализированы этапы алгоритма расчетов при корреляционном анализе связи парной корреляции. Предложены методические основы корреляционно-регрессионного анализа изменения урожайности зерновых культур в зависимости от качества пахотной земли. Разработан методический подход корреляционно-регрессионного анализа динамики численности населения районного центра Карагандинской области. Рекомендованы методические принципы анализа экономического показателя работы уранового рудника с применением математического аппарата парной корреляции. Составлены методические положения по изучению деятельности крестьянских хозяйств зерносеющих районов Карагандинского региона с использованием прикладного аппарата корреляционно-регрессионного анализа.

Ключевые слова: связь, корреляция, регрессия, анализ, коэффициент, применение, оценка, аппроксимация, зависимость, уравнение.

Корреляционная и регрессионная связи как частные случаи стохастической формы связи

Связь между двумя переменными x и y может быть функциональной (полной) или корреляционной (неполной). Последнее означает, что каждому конкретному значению x соответствует не одно, а несколько значений y; с изменением x меняется распределение значений признака y.

В производстве строгие функциональные зависимости между исследуемыми величинами относительно редки. В абсолютном большинстве случаев изменения одних явлений не вызывают строго обусловленных изменений других — налицо стохастическая форма связи. Частными случаями стохастической формы связи являются корреляционная и регрессионная связи [1–5].

Две случайные величины являются корреляционно связанными, если математическое ожидание одной из них меняется в зависимости от изменения другой. Соответственно метод математической статистики, изучающий корреляционные связи между явлениями, называется корреляционным анализом. Связь между случайной и неслучайной величинами называется регрессионной. Исследование вида зависимости между переменными величинами называется регрессионным анализом.

Регрессионный анализ тесно связан с корреляционным. Основными предпосылками их применения являются:

1) нормальное распределение случайных величин;

2) случайные величины x_1 и x_2 (в многомерном случае — $x_1, x_2 \ \ldots, \ x_m$) можно рассматривать как выборку из двумерной (в многомерном случае — многомерной) генеральной совокупности с нормальным законом распределения;

3) дисперсия случайной величины x_1 остается постоянной при изменении величины x_2 или пропорциональна известной функции 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.

Функция $f(x, \text{a})$ считается линейно зависимой от a, если соблюдается равенство

$$f(x_1, a_1 + a_2) = \varphi_1(x_1, a_1) + \varphi_2(x_1, a_2).$$

Регрессионный анализ предъявляет менее жесткие требования к исходной информации. Его применение возможно даже при некотором отличии распределения случайных величин от нормального, что важно для анализа процессов горного производства, так как распределение изучаемых величин процессов горного производства, включая экономические, часто бывает асимметрично.

В качестве функции (зависимой переменной) в регрессионном анализе принимают случайную переменную, а аргументом (независимой переменной), как правило, является неслучайная переменная.

Необходимые условия применения корреляционного анализа

Корреляция может быть парной и множественной [6–10]. *Парная корреляция* — это связь между двумя показателями, один из которых является фактором, другой — результативным показателем.

Множественная корреляция — связь между несколькими факторами и одним результативным показателем.

Приемы корреляционного анализа на сегодняшний день нашли наибольшее применение. Необходимыми условиями применения корреляционного анализа в экономических исследованиях являются:

1. Наличие достаточно большого количества наблюдений о величине исследуемых факторных и результативных показателей (в динамике или за текущий год по совокупности однородных объектов).

2. Исследуемые факторы должны иметь количественное измерение и отражение в тех или иных источниках информации.

Применение корреляционного анализа позволяет решить следующие основные задачи [11]:

а) определить изменение результативного показателя под воздействием одного или нескольких факторов (в абсолютном измерении), это значит, определить, на сколько единиц изменяется величина результативного показателя при изменении факторного на единицу;

б) установить относительную степень зависимости результативного показателя от каждого фактора.

Исследование корреляционных соотношений имеет огромное значение в АХД (анализе хозяйственной деятельности) промышленных и сельскохозяйственных предприятий. Это проявляется в том, что значительно углубляется факторный анализ, устанавливаются место и роль каждого фактора в формировании уровня исследуемых показателей, углубляются знания об изучаемых явлениях, определяются закономерности их развития и как итог обосновываются планы и управленческие решения, более объективно оцениваются итоги деятельности предприятий и более полно определяются внутрихозяйственные резервы.

Теснота связи между двумя явлениями измеряется парным коэффициентом корреляции (парным корреляционным отношением при криволинейной парной зависимости). Количественная оценка тесноты связи в зависимости от корреляционного отношения (коэффициента корреляции) приведена в таблице 1.

Таблица 1

Качественная оценка тесноты связи при различных значениях корреляционного отношения (коэффициента корреляции)*

Величина корреляционного отношения	0,1–0,3	0,3–0,5	0,5–0,7	0,7–0,9	0,9–0,99
Теснота связи	Слабая	Умеренная	Заметная	Высокая	Весьма высокая

**Примечание*. Данные работы [12].

Общая формула корреляционного отношения:

$$\mathrm{R}_{yx} = \sqrt{\frac{\sigma_y^2 - \sigma_{y_x}^2}{\sigma_y^2}} = \sqrt{1 - \frac{\sigma_{yx}^2}{\sigma_y^2}},$$

где $\sigma_{y_x}^2$ — среднее квадратическое отклонение y от теоретических значений y_x; y_x определяется на основе уравнений регрессии; σ_y^2 — среднее квадратическое отклонение эмпирических (фактических) значений y.

В случае прямолинейной зависимости корреляционное отношение называется коэффициентом корреляции и обозначается буквой r.

Корреляционное отношение (коэффициент корреляции) принимает значения от 0 до 1:

– если $R_{yx}(r) = 0,$ то связь между показателями отсутствует;

– если $R_{yx}(r) = 1,$ то связь функциональная (детерминированная);

– если $R_{yx}(r)$ — отрицательная величина, то связь между показателями обратная;

– если $R_{yx}(r)$ — положительная величина, то связь между показателями прямая.

Этапы алгоритма расчетов при корреляционном анализе связи парной корреляции

Алгоритм расчетов при корреляционном анализе связи парной корреляции состоит из ряда этапов.

Этап 1. Производится отбор наиболее важных существенных факторов, влияющих на результативный показатель. При отборе факторов учитываются причинно-следственные связи между показателями, причем все факторы должны быть количественно измеримы. Большую помощь при отборе факторов для корреляционной модели оказывают аналитические группировки, способ сравнения параллельных и динамических рядов, линейные графики. Отбор показателей для анализа и придание им статуса фактора или результативного значения осуществляются на основе знания экономических законов. Например, знание закона спроса и предложения помогает изучить влияние ценового фактора на изменение спроса. Отобранные для анализа показатели и результаты наблюдений за их изменением помещаются в таблицу, в которой факторные признаки располагаются в порядке возрастания или убывания, т.е. ранжируются.

Этап 2. Данные из таблицы наносятся на плоскость координат — строится корреляционное поле. Иначе говоря, выявление корреляционной связи между компонентами графическим методом выполняется путем нанесения единичных значений выборки на миллиметровую бумагу в прямоугольной системе координат (рис. 1).

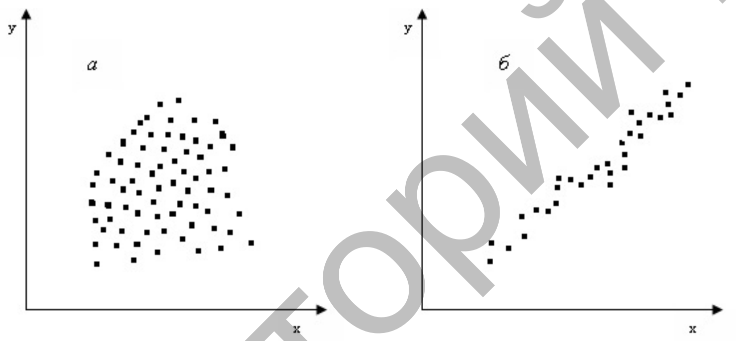

Рисунок 1. Графический способ выявления корреляционной связи между компонентами:
а — отсутствие корреляционной связи; б — наличие корреляционной связи*

**Примечание.* Данные работы [13; 64].

Здесь на осях координат — переменные величины, связь между которыми нужно установить. Если поле точек покажет отсутствие корреляционной связи между компонентами (рис. 1, *а*), то дальнейшее вычисление коэффициентов парной корреляции при анализе двух переменных нецелесообразно. Корреляционная связь или отсутствует, или она завуалирована влиянием прочих величин и необходимо использовать методы множественного корреляционного анализа.

Этап 3. На данном этапе осуществляется выявление характера распределения случайных величин. Если графический метод показал наличие корреляционной взаимосвязи между изучаемыми случайными величинами (рис. 1, *б*), необходимо установить характер распределения их единичных значений в соответствующих статистических выборках.

Поскольку методы корреляционного и регрессионного анализов приемлемы для статистических выборок только с нормальным распределением случайных величин или выборок, которые можно привести к нормальному виду распределений (например, логнормальные распределения), все дальнейшие рассуждения относятся только к этому виду распределений случайных величин.

Этап 4. На данном этапе производится анализ корреляционных и регрессионных связей между изучаемыми явлениями. При этом решаются две основные задачи: устанавливается степень тесноты связи между переменными признаками путем вычисления коэффициента корреляции или корреляционного отношения и с помощью уравнений регрессии определяется аналитическая форма связи между вариациями признаков.

Уравнения регрессии между вариацией признаков x и y определяют форму связи между ними. Влияние прочих посторонних факторов выражено в уравнениях регрессии в виде средних постоянных величин.

Для установления аналитического выражения варьирования признака y_j в зависимости от x_i применяют метод выравнивания, сущность которого заключается в замене эмпирической линии, построенной в прямоугольной системе координат по данным единичных значений выборки (рис. 2), линией, представляющей аналитическое выражение искомой зависимости.

Для отыскания коэффициентов уравнения регрессии применяют метод наименьших квадратов, согласно которому отыскиваются такие значения коэффициентов уравнения регрессии, при которых сумма квадратов отклонений фактических значений варьирующего признака от вычисленных по уравнению была бы наименьшей из всех возможных. Это значит, находят параметры регрессионного уравнения при условии

$$\frac{\Sigma\left(\overline{y}_x - y\right)^2}{n} = \min.$$

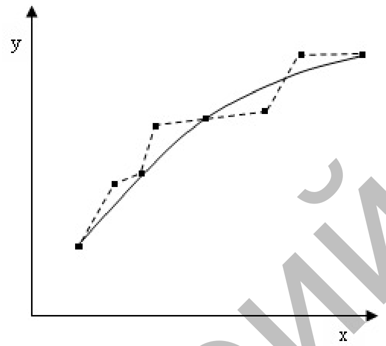

Рисунок 2. Аппроксимация эмпирической линии регрессии расчетной кривой*

Примечание. Данные работы [13; 65].

Исследуемые регрессионные зависимости могут быть прямолинейными или криволинейными. Криволинейная форма связи может быть представлена уравнением гиперболы, параболы, логарифмической функцией и т.д.

Методы расчета параметров нелинейных корреляционных уравнений аналогичны методам расчета прямолинейных, так как и здесь сохраняется требование наименьших квадратов.

Одним из способов нахождения зависимости является метод замены переменной. Этот метод довольно часто используется при решении различных математических задач. Он заключается в том, что независимый фактор заменяется некоторой функцией этого фактора, которая переводит нелинейную зависимость в разряд линейных.

Корреляционно-регрессионный анализ изменения урожайности зерновых культур в зависимости от качества пахотной земли

Наиболее простым уравнением, которое характеризует прямолинейную зависимость между двумя показателями, является уравнение прямой:

$$Y_x = a + bx,$$

где x — факторный показатель; Y — результативный показатель; a и b — параметры уравнения регрессии, которые требуется отыскать.

Это уравнение описывает такую связь между двумя признаками, при которой с изменением факторного показателя на определенную величину наблюдается равномерное возрастание или убывание значений результативного показателя. В качестве примера для иллюстрации корреляционного анализа прямолинейной зависимости могут быть использованы сведения об изменении урожайности зерновых культур (Y) в зависимости от качества пахотной земли (X) по данным 20 крестьянских хозяйств Карагандинской области (табл. 2).

Значения коэффициентов a и b находят из системы уравнений, полученных по способу наименьших квадратов. В данном случае система уравнений имеет следующий вид:

$$\begin{cases} na + b\sum x = \sum y; \\ a\sum x + b\sum x^2 = \sum xy, \end{cases}$$

где n — количество наблюдений (в нашем исследовании — 20).

Значения $\sum x, \sum y, \sum xy, \sum x^2$ рассчитываются на основе фактических исходных данных (табл. 3).

Таблица 2

Зависимость урожайности зерновых культур от качества земли по данным 20 крестьянских хозяйств Карагандинской области

Номер хозяйства	Качество земли, балл	Урожайность, ц/га	Номер хозяйства	Качество земли, балл	Урожайность, ц/га
1	32	19,5	11	45	24,2
2	33	19,0	12	46	25,0
3	35	20,5	13	47	27,0
4	37	21,0	14	49	26,8
5	38	20,8	15	50	27,2
6	39	21,4	16	52	28,0
7	40	23,0	17	54	30,0
8	41	23,3	18	55	30,2
9	42	24,0	19	58	32,0
10	44	24,5	20	60	33,0

Таблица 3

Расчет производных величин для определения параметров уравнения связи и коэффициента корреляции

n	x	y	xy	x^2	y^2	Yx
1	32	19,5	624	1024	380,25	19,8
2	33	19,0	627	1089	361,00	20,2
3	35	20,5	717	1225	420,25	21,0
...	...	...	...	...	...	...
20	60	33,0	1980	3600	1089,00	31,0
Итого	900	500,0	22900	41500	12860,00	500,0

Подставив полученные значения в систему уравнений, получим

$$\begin{cases} 20a + 900b = 500; \\ 900a + 41500b = 22900. \end{cases}$$

Отсюда: $a = 7{,}0$; $b = 0{,}4$.

Таким образом, уравнение связи, которое описывает зависимость урожайности от качества почвы, будет иметь вид:

$$Y_x = 7{,}0 + 0{,}4x.$$

Коэффициент a — постоянная величина результативного показателя, которая не связана с изменением данного фактора. Параметр b показывает среднее изменение результативного показателя с повышением или понижением величины фактора на единицу его измерения. Как показало наше исследование, с увеличением качества почвы на один балл урожайность зерновых культур повышается в среднем на 0,4 ц/га.

Подставив в уравнение регрессии соответствующие значения X, можно определить выравненные (теоретические) значения результативного показателя (Y) для каждого хозяйства. Например, чтобы рассчитать урожайность зерновых культур для первого хозяйства, где качество почвы оценивается 32 баллами, необходимо это значение подставить в уравнение связи:

$$Y_x = 7 + 0{,}4 * 32 = 19{,}8 \text{ ц/га}.$$

Полученная величина показывает, какой была бы урожайность при качестве почвы 32 балла, если бы данное хозяйство использовало свои производственные возможности в такой степени, как в среднем все крестьянские хозяйства Карагандинской области.

Наши аналогичные расчеты сделаны для каждого крестьянского хозяйства региона. Данные приведены в последней графе таблицы 3. Сравнение фактического уровня урожайности с расчетным позволяет оценить результаты работы отдельных сельскохозяйственных предприятий (крестьянских хозяйств) Карагандинской области.

Для измерения тесноты связи между факторными и результативными показателями определяется коэффициент корреляции.

В случае прямолинейной формы связи между изучаемыми показателями коэффициент корреляции рассчитывается по следующей формуле:

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left(\sum x^2 - \frac{\left(\sum x\right)^2}{n}\right)\left(\sum y^2 - \frac{\left(\sum y\right)^2}{n}\right)}} =$$

$$= \frac{\overline{xy} - \sum x \sum y}{\sqrt{\left(\sum x^2 n - \left(\sum x\right)^2\right)\left(\sum y^2 n - \left(\sum y\right)^2\right)}}. \qquad (1)$$

Подставляя значения $\sum xy, \sum x, \sum y, \sum x^2$ и $\sum y^2$ в формулу (1), получаем

$$r = \frac{22900\text{-}\frac{900 * 500}{20}}{\sqrt{\left(41500\text{-}\frac{900^2}{20}\right)\left(12860 - \frac{500^2}{20}\right)}} = 0,660.$$

Коэффициент корреляции может принимать значения от 0 до $\pm 1.$ Чем ближе его величина к 1, тем более тесная связь между изучаемыми явлениями, и наоборот. В данном случае величина коэффициента корреляции является существенной $(r = 0{,}660).$ Это позволяет сделать вывод о том, что урожайность почвы — один из основных факторов, от которого в данном районе зависит уровень урожайности зерновых культур.

Если коэффициент корреляции возвести в квадрат, получим коэффициент детерминации $(d = 0{,}435).$ Он показывает, что урожайность зерновых культур на 43,5 % зависит от качества почвы, а на долю других факторов приходится 56,5 % прироста урожайности.

Что касается измерения тесноты связи при криволинейной форме зависимости, то здесь используется не линейный коэффициент корреляции, а корреляционное отношение (индекс корреляции).

Корреляционно-регрессионный анализ динамики численности населения районного центра Карагандинской области

В планово-экономических расчетах широкое использование получила показательная производственная функция

$$y = \mathrm{a}_0 a_i^x.$$

Чтобы это уравнение привести к линейной форме, его логарифмируют и получают

$$\lg y = \lg a_0 + x \lg a_1.$$

При расчете параметров с помощью способа наименьших квадратов записывают

$$S = \sum \left(\lg a_0 + x \lg a_1 - \lg y\right)^2 \to \min.$$

Поэтому система нормальных уравнений имеет вид:

$$n \lg a_0 + \lg a_1 \sum x = \sum \lg y;$$

$$\lg a_0 \sum x + \lg a_1 \sum x^2 = \sum x \lg y.$$

Решая приведенную систему с помощью определителей, имеем

$$\Delta = n\sum x^2 - \left(\sum x\right)^2;$$

$$\Delta_{\lg a_0} = \sum x^2 \sum \lg y - \sum x \sum x \lg y;$$

$$\Delta_{\lg a_1} = n\sum x \lg y - \sum x \sum x \lg y.$$

$$\lg a_0 = \frac{\sum x^2 \sum \lg y - \sum x \sum x \lg y}{n\sum x^2 - \left(\sum x\right)^2}; \quad (2)$$

$$\lg a_1 = \frac{n\sum x \lg y - \sum x \sum x \lg y}{n\sum x^2 - \left(\sum x\right)^2}. \quad (3)$$

После упрощения формул (2) и (3) получают следующие выражения для определения логарифмов параметров:

$$\lg a_1 = \frac{\sigma_{\lg y}}{\sigma_x} r_{x \lg y}; \quad (4)$$

$$\lg a_0 = \overline{\lg y} - \bar{x} \lg a_1. \quad (5)$$

В ходе проведенного исследования сделана попытка рассчитать параметры показательной функции, которая, на наш взгляд, может наиболее точно описать динамику численности населения одного из районных центров Карагандинской области за последние 10 лет.

Исходные данные для корреляционно-регрессионного анализа, а также произведенные расчеты по определению параметров показательной функции представлены в таблице 4.

Таблица 4

Информация для определения параметров показательной функции (динамика численности населения в одном из районов Карагандинской области за последние 10 лет)

№ п/п	Год	Население, тыс. чел.	Расчетные величины					
i	x	y	$\lg y$	$x \lg y$	x^2	$\lg^2 y$	$\lg \bar{y}_x$	$\bar{y}_x$
1	1	52	1,7160	1,7160	1	2,9447	1,7356	54,4
2	2	56	1,7482	3,4964	4	3,0562	1,7419	55,2
3	3	57	1,7559	5,2676	9	3,0832	1,7482	56,0
4	4	58	1,7634	7,0537	16	3,1096	1,7543	56,8
5	5	58	1,7634	8,8171	25	3,1096	1,7604	57,6
6	6	59	1,7709	10,6251	36	3,1361	1,7672	58,5
7	7	59	1,7709	12,3960	49	3,1361	1,7731	59,3
8	8	60	1,7782	14,2252	64	3,1620	1,7796	60,2
9	9	61	1,7853	16,0680	81	3,1873	1,7853	61,0
10	10	61	1,7853	17,8533	100	3,1873	1,7917	61,9
$\sum$	55	581	17,6375	97,5184	385	31,1121	17,6375	580,9
$\sum/n$	5,5	85,1	1,76375	×	×	×	1,76375	58,09

Система нормальных уравнений по данным таблицы 4 имеет следующий вид:

$$10\lg a_0 + 55\lg a_1 = 17,6375;$$

$$55\lg a_0 + 385\lg a_1 = 97,5184.$$

Решим эту систему уравнений:

$$\Delta = 10 \cdot 385\text{-}55 \cdot 55 = 825;$$

$$\Delta_{\lg a0} = 17,6375 \cdot 385 - 97,5184 \cdot 55 = 1426,9255;$$

$$\Delta_{\lg a1} = 10 \cdot 97,5184 - 55 \cdot 17,6375 = 5,1215;$$

$$\lg a_0 = 1426,9255 : 825 = 1,729606; \ \lg a_1 = 5,1215 : 825 = 0,006208;$$

$$a_0 = 53,6545; \ a_1 = 1,014397.$$

Искомые корни системы можно получить по формулам (4) и (5). Но для этого необходимо вычислить $\sigma_{\lg y}$, σ_x и $r_{x\lg y}$:

$$\sigma_{\lg y} = \frac{\sqrt{10 \cdot 31{,}1121 - 17{,}6375^2}}{10} = 0{,}01990;$$

$$\sigma_x = \frac{\sqrt{10 \cdot 385 - 55^2}}{10} = 2{,}8723;$$

$$r_{x\lg y} = \frac{10 \cdot 97{,}5184 - 55 \cdot 17{,}6375}{\sqrt{\left(10 \cdot 385 - 55^2\right)\left(10 \cdot 31{,}1121 - 17{,}6375^2\right)}} = 0{,}896030.$$

Поэтому определяемые параметры составляют

$$\lg a_1 = \frac{0{,}01990}{2{,}8723} \cdot 0{,}89603 = 0{,}006208; \; a_1 = 1{,}014397;$$

$$\lg a_0 = 1{,}76375 - 0{,}006208 \cdot 55 = 1{,}729606; \; a_0 = 53{,}6545.$$

Таким образом, показательная производственная функция, моделирующая изменение численности населения, имеет вид

$$y = 53{,}6545 \cdot 1{,}014397^x.$$

В таблице 4 выполнен расчет количества населения на основе полученной функции. Как видно из таблицы, $\sum \lg \bar{y}_x = \sum \lg y$. Но сумма теоретических уровней населения за 10 лет несколько меньше фактической суммы.

В случае применения показательной функции вида $y = a_1^x$ коэффициент регрессии лучше определять не по способу наименьших квадратов, а из уравнения

$$\lg a_1 \sum x = \sum \lg y.$$

Параллельно с показательной функцией находит применение и экспоненциальная зависимость вида

$$y = a_0 e^{a_1 x}.$$

Прологарифмировав (в натуральных логарифмах) уравнение функции, имеем:

$$1ny = 1na_0 + a_1 x.$$

В соответствии с требованием способа наименьших квадратов получаем следующую функцию:

$$S = \sum \left(1na_0 + a_1 x - 1ny\right)^2 \to \min.$$

В конечном счете приходим к такой системе нормальных уравнений:

$$n \ln a_0 + a_1 \sum x = \sum 1ny;$$

$$1na_0 \sum x + a_1 \sum x^2 = \sum x 1ny.$$

Корни этой системы определяются формулами, аналогичными формулам (2) и (3):

$$1na_0 = \frac{\sum x^2 \sum 1ny - \sum x \sum x1ny}{n\sum x^2 - \left(\sum x\right)^2};$$

$$a_1 = \frac{n\sum x1ny - \sum x \sum 1ny}{n\sum x^2 - \left(\sum x\right)^2}.$$

Последние формулы можно представить в следующем виде:

$$a_1 = \frac{\sigma 1ny}{\sigma x} r_{x1ny}; \tag{6}$$

$$1na_0 = \overline{1ny} - a_1 \bar{x}. \tag{7}$$

Нами вычислены в порядке сопоставления параметры экспоненциальной производственной функции по данным таблицы 4, т.е. использована динамика численности населения обследуемого районного центра Карагандинского региона.

На основе таблицы 5 нами получена следующая система нормальных уравнений:

$$10 \ln a_0 + 55a_1 = 40{,}6116;$$

$$55 \ln a_0 + 358a_1 = 224{,}5439.$$

Эта система совершенно аналогична предыдущей системе уравнений, с помощью которой находились параметры функции $y = a_0 a_1^x$. Только в настоящей системе уравнений логарифмы свободных членов увеличились в 2,3026, в модуль перевода десятичных логарифмов в натуральные.

Решением приведенной системы являются:

$$\ln a_0 = 3{,}982565;\ a_0 = 53{,}6545;\ a_1 = 0{,}014294.$$

Таблица 5

Данные для построения функции $y = a_0 e^{a_1 x}$

Год	Население, чел.	Расчетные величины					
x	y	$\ln y$	$x \ln y$	$\ln^2 y$	x^2	$\ln \bar{y}_x$	$\bar{y}_x$
1	52	3,9512	3,9512	15,6120	1	3,9964	54,4000
2	56	4,0254	8,0508	16,2038	4	4,0110	55,2000
3	57	4,0431	12,1293	16,3467	9	4,0254	56,0000
4	58	4,0604	16,2416	16,4868	16	4,0395	56,8000
5	58	4,0604	20,3020	16,4868	25	4,0535	57,6000
6	59	4,0775	24,4650	16,6260	36	4,0690	58,5000
7	59	4,0775	28,5425	16,6260	49	4,0826	59,3000
8	60	4,0943	32,7544	16,7633	64	4,0977	60,2000
9	61	4,1109	36,9981	16,8995	81	4,1109	61,0000
10	61	4,1109	41,1090	16,8995	100	4,1255	61,9000
$\sum$ 55	581	40,6116	224,5439	164,9504	385	40,6116	580,9000
$\sum / n$ 5,5	58,1	4,06116	×	×	×	4,06116	58,0900

Это же решение может быть получено по формулам (6) и (7). Следовательно, экспоненциальная производственная функция, описывающая динамику численности населения, имеет следующий вид:

$$y = 53{,}6545 e^{0{,}014294x}.$$

Если учесть, что

$$e^{0{,}014294x} = 1{,}014397,\ \text{то}\ y = 53{,}6545 \cdot 1{,}014397^x,$$

т.е. мы пришли к показательной функции, которая была получена нами раньше.

Корреляционно-регрессионный анализ зависимости суммы прямых затрат на добычу условной единицы металла от продуктивности отрабатываемой жильной площади

В процессе исследования нами сделана попытка выявить зависимость суммы прямых затрат на добычу условной единицы металла (y, долл./усл.ед.) при очистной выемке на урановых месторождениях Республики Казахстан от продуктивности отрабатываемой жильной площади x, выражаемой числом добытых условных единиц металла с 1 м^2 отработанной жильной площади (рис. 3).

Рисунок 3. Зависимость суммы прямых затрат на добычу металла (в условных единицах) от условной продуктивности отрабатываемой жильной площади

Единичные значения условных затрат на добычу условной единицы металла с 1 м^2 жильной площади по очистным блокам представлены в таблице 6.

Нанесенные на график в прямоугольной системе координат, они показали наличие корреляционной связи между независимой переменной x и функцией y, форма которой могла быть аппроксимирована уравнением второй степени.

Это предположение было подтверждено при нанесении единичных значений выборки на график в логарифмическом масштабе (рис. 4), где была получена четкая линейная зависимость

$$\lg y = f\left(\lg x\right).$$

Для аппроксимации было выбрано уравнение вида (гипербола)

$$y = a + \frac{b}{x_i}.$$

Значение коэффициента a_{yx} определено по формуле:

$$a_{yx} = \frac{\sum y_i \sum \frac{1}{x_i^2} - \sum \frac{y_i}{x_i} \sum \frac{1}{x_i}}{n \sum \frac{1}{x_i^2} - \left(\sum \frac{1}{x_i}\right)^2}.$$

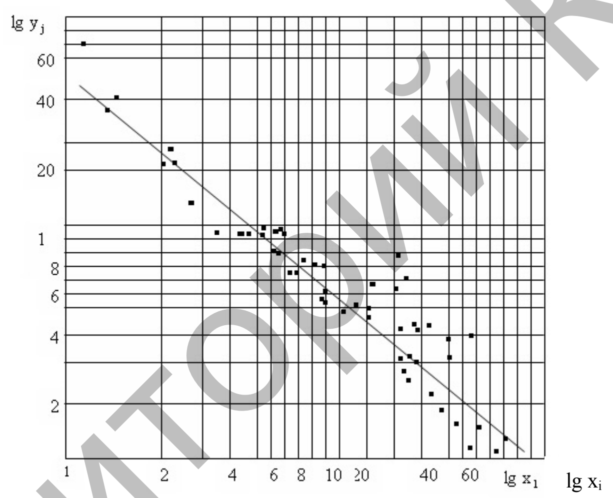

Рисунок 4. Эмпирическая линия регрессии в логарифмическом масштабе $\lg x_i$

Таблица 6

Единичные значения условных затрат на добычу условной единицы металла с 1 м^2 жильной площади по очистным блокам урановых месторождений Казахстана

№ п/п	x_i	y_i	y_i^2	$\frac{1}{x_i^2}$	y_i	y_i^2	$\frac{y_i}{x_i}$
1	2	3	4	5	6	7	8
1	21,0	7,0	49,0000	0,0023	4,2958	18,4539	0,3333
2	2,6	60,55	3666,3025	0,1479	24,3617	593,4924	23,2885
3	10,1	9,25	85,5625	0,0098	7,3558	54,1078	0,9158
4	53,1	4,01	16,0801	0,0004	2,5817	6,6652	0,0755
5	101,1	1,43	2,0449	0,0001	2,0494	4,2000	0,0141
6	59,0	1,79	3,2041	0,0003	2,4696	6,0989	0,0303
7	1,3	32,05	1027,2025	0,5917	47,2630	2233,7912	24,6538
8	6,5	9,25	85,5625	0,0237	10,6209	112,8035	1,4231
9	2,44	22,39	501,3121	0,1736	26,2702	690,1234	9,3292
10	20,3	9,57	91,5849	0,024	4,3936	19,3037	0,4714
11	10,0	8,80	77,4400	0,0100	7,4147	54,9778	0,8800

Продолжение таблицы 6

1	2	3	4	5	6	7	8
12	41,1	2,43	5,9049	0,0006	2,9091	8,4629	0,0591
13	3,4	15,01	225,3001	0,0865	18,9732	359,9823	4,4147
14	36,3	2,08	4,3264	0,0008	3,1007	9,6143	0,0573
15	32,7	4,48	20,0704	0,0009	3,2813	10,7669	0,1370
16	4,2	11,89	141,3721	0,0567	15,637	244,5283	2,8310
17	16,7	5,27	27,7729	0,0036	5,0259	25,2597	0,3156
18	14,6	6,46	41,7316	0,0047	5,5387	30,6772	0,4425
19	23,2	7,51	56,4001	0,0019	4,0269	16,2159	0,3237
20	1,2	69,43	4820,5249	0,6944	51,0799	2609,1562	57,8583
21	28,0	4,61	21,2521	0,0013	3,5870	12,8666	0,1646
22	9,8	6,91	47,7481	0,0104	7,5363	56,7958	0,7051
23	21,2	4,51	20,3401	0,0022	4,2691	18,2252	0,2127
24	41,0	3,34	11,1556	0,0006	2,9127	8,4838	0,0815
25	45,4	1,78	3,1684	0,0005	2,7719	7,6834	0,0392
26	7,6	8,39	70,3921	0,0173	9,2951	86,3989	1,1039
27	27,5	3,48	12,1104	0,0013	3,6256	13,1450	0,1265
28	8,5	9,22	85,0084	0,0138	8,4655	71,6647	1,0847
29	16,9	7,51	56,4001	0,0035	4,9837	24,8378	0,444
30	5,0	7,82	61,1524	0,0400	13,3691	78,7328	1,5640
31	8,3	5,37	28,8369	0,0156	8,6343	74,5511	0,6470
32	71,2	1,36	1,8496	0,0002	2,2907	5,2748	0,0191
33	29,4	4,17	17,3849	0,0012	3,4857	12,1501	0,1418
34	10,0	7,43	55,2049	0,0100	7,4147	54,9778	0,7430
35	2,1	22,63	512,1169	0,2268	29,8144	888,8984	10,7762
36	2,2	29,04	843,3216	0,2066	28,5256	813,7099	13,2000
37	10,0	7,26	52,7076	0,0100	7,4147	54,9778	0,7260
38	39,2	3,97	15,7609	0,0007	2,9794	8,8768	0,1013
39	10,0	6,71	45,0241	0,0100	7,4147	54,9778	0,6710
40	22,7	3,50	12,2500	0,0019	4,0835	16,6750	0,1542
41	107,9	1,20	1,4400	0,0001	2,0122	4,0489	0,0111
42	20,8	3,75	14,0625	0,0023	4,3231	18,6892	0,1803
43	6,9	12,16	147,8656	0,0210	10,0899	101,8061	1,7623
44	11,4	6,12	37,4544	0,0077	6,6835	44,6692	0,5368
45	66,4	1,12	1,2544	0,0002	2,3571	5,5559	0,0169
46	11,2	6,69	44,7561	0,0080	6,7768	45,9250	0,5973
47	1,4	36,34	1320,5956	0,5102	43,9914	1935,2433	25,9571
48	2,3	24,70	6010,0900	0,1890	27,3488	747,9569	10,7391
49	84,4	1,03	1,0609	0,0001	2,1659	4,6911	0,0122
50	15,9	6,06	36,7236	0,0040	5,2053	27,0951	0,3811
51	7,0	11,25	126,5625	0,0204	9,9666	99,3331	1,6071
52	9,4	8,80	77,4400	0,0113	7,7948	60,7589	0,9362
53	5,5	11,11	123,4321	0,0331	12,2865	150,9581	2,0200
54	1,6	38,74	1500,7876	0,3906	38,6750	1495,7556	24,2125
55	1,5	42,84	1835,2656	0,4444	41,1560	1693,8163	28,5600
56	24,5	2,89	8,3521	0,0017	3,8907	15,1375	0,1180
57	3,7	12,66	160,2556	0,0730	17,5532	308,1148	3,4216
58	107,6	1,5625	0,0001	2,0138	4,0554	4,0554	0,0116
59	8,0	8,29	68,7241	0,0156	8,9033	79,2688	1,0369
60	6,0	9,46	89,4916	0,0278	11,3843	129,6023	1,5867
61	5,5	12,77	163,0729	0,0331	12,2865	150,9581	2,3218
62	6,7	13,64	186,0496	0,0223	10,3475	107,0708	2,0358
63	4,5	11,66	135,9556	0,0494	14,6923	215,8637	2,5911
64	2,7	16,96	287,6416	0,1372	34,5135	552,8847	6,2815
65	98,9	1,12	1,2544	0,0001	2,0625	4,2539	0,0113
66	25,3	2,55	6,5025	0,0016	3,8139	14,5458	0,1008
67	27,3	3,51	12,3201	0,0013	3,6415	13,2605	0,1286
68	5,9	9,69	93,8961	0,0287	11,5525	133,4603	1,6424
$\sum$ Σ^*	1557,0	758,02	20015,7862	4,4205	11,147353**	17737,3638	279,3679

* $(1/x_i) = 11,0627.$

** Среднее значение $\overline{y}$.

После подстановки численных значений

$$a_{yx} = \frac{758,02 \cdot 4,4205 - 279,3679 \cdot 11,0627}{68 \cdot 4,4205 - 11,0627^2} = 1,46.$$

Значение коэффициента b_{yx} определено по формуле:

$$b_{yx} = \frac{n\sum \frac{y_i}{x_i} - \sum \frac{1}{x_i} \sum y_i}{n\sum \frac{1}{x_i^2} - \left(\sum \frac{1}{x_i} \right)^2}.$$

После подстановки численных значений

$$b_{yx} = \frac{68 \cdot 279,3679 - 11,0627 \cdot 758,02}{68 \cdot 4,4205 - 11,0627^2} = 59,54.$$

Отсюда

$$y_i = 1,46 + \frac{59,54}{x_i}. \qquad (8)$$

Для оценки тесноты корреляционной связи был рассчитан индекс корреляции (парное корреляционное отношение):

$$R_{yx} = \pm \sqrt{\frac{n\sum_i y_i^2 - \left(\sum_i y_i \right)^2}{n\sum_i y_i^2 - \left(\sum_i y_i \right)^2}} =$$

$$= \pm \sqrt{\frac{68 \cdot 17737,36 \cdot 38 - 574594,3204}{68 \cdot 20015,7862 - 574594,3204}} = \pm\sqrt{0,8030} = \pm 0,896,$$

что свидетельствует о высокой тесноте связи. Значит, выбранное нами аппроксимирующее уравнение хорошо соответствует экспериментальным данным выборки.

Уравнение (8) для определенных условий может служить математической моделью прямых затрат на добычу условной единицы продукции очистными работами. При этом оно в скрытой форме учитывает взаимосвязь комплекса определенных технических параметров, находящих отражение в затратах на добычу, с геологическим — продуктивностью отрабатываемой жильной площади.

При невозможности определения характера связи между изучаемыми признаками при качественном анализе их значений в поле корреляции определяют и коэффициент корреляции, и корреляционное отношение. Если они оказываются равными или близкими друг другу, то зависимость линейна, если же корреляционное отношение существенно больше значения коэффициента корреляции, то связь нелинейна. В этих случаях характеристикой тесноты связи служит корреляционное отношение, интерпретация которого не зависит от вида исследуемой корреляционной зависимости.

Корреляционно-регрессионный анализ зависимости урожайности зерновых культур от количества выпавших осадков

Нами по данным 15 крестьянских хозяйств ряда зерносеющих районов Карагандинской области исследована зависимость урожайности зерновых культур *Y* (ц/га) от количества осадков *X* (см), выпавших в вегетационный период в период роста, развития растений. Исходные данные для корреляционно-регрессионного анализа приведены в таблице 7.

Таблица 7

Исходные данные для расчета

№ п/п	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Количество осадков x_i, см	25	27	30	35	36	38	39	41	42	45	46	47	50	52	53
Урожайность y_i, ц/га	23	24	27	27	32	31	33	35	34	32	29	28	25	24	25

Из качественных соображений можно предположить, что увеличение количества выпавших осадков приводит к увеличению урожайности до некоторого предела, после чего урожайность будет снижаться. Кроме того, учитывая расположение точек корреляционного поля (см. рис. 5), можно предположить, что наиболее подходящим уравнением регрессии будет уравнение параболы

$$y_x = b_0 + b_1 x + b_2 x^2 .$$

Его параметры b_0, b_1, b_2 находим, применяя метод наименьших квадратов:

$$S = \sum_{i=1}^{n}\left(y_{x_i} - y_i\right)^2 = \sum_{i=1}^{n}\left(b_0 + b_1 x + b_2 x^2 - y_i\right)^2 \to \min .$$

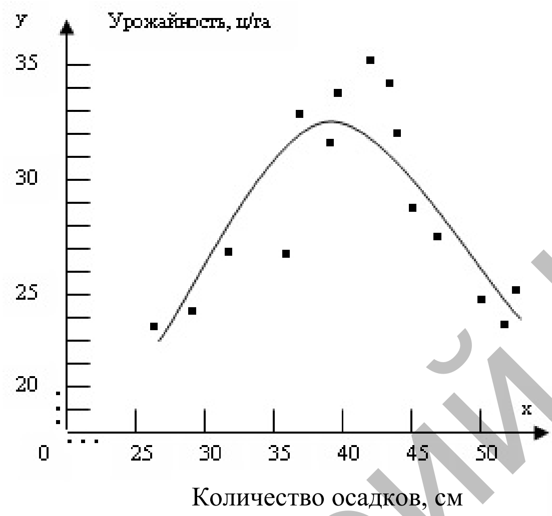

Рисунок 5. Графическая зависимость урожайности от количества выпавших осадков

Система нормальных уравнений имеет вид:

$$\begin{cases} b_0 n + b_i \sum_{i=1}^{n} x_i + b_2 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i ; \\ b_0 \sum_{i=1}^{n} x_i + b_i \sum_{i=1}^{n} x_i^2 + b_2 \sum_{i=1}^{n} x_i^3 = \sum_{i=1}^{n} y_i x_i ; \\ b_0 \sum_{i=1}^{n} x_i^2 + b_1 \sum_{i=1}^{n} x_i^3 + b_2 \sum_{i=1}^{n} x_i^4 = \sum_{i=1}^{n} y_i x_i^2 . \end{cases} \qquad (9)$$

Для расчета необходимых сумм составим вспомогательную таблицу (табл. 8).

Теперь система (9) примет вид

$$\begin{cases} 15 b_0 + 606 b_1 + 25548 b_2 = 429; \\ 606 b_0 + 25548 b_1 + 1115808 b_2 = 17371; \\ 25548 b_0 + 1115808 b_1 + 50158200 b_2 = 730123. \end{cases}$$

Решая эту систему, например, методом Гаусса, получим $b_0 = -43,93$; $b_1 = 3,8342$; $b_2 = -0,048361$, т.е. уравнение регрессии имеет вид

$$y_x = -43,93 + 3,8342x - 0,048361x^2 .$$

Т а б л и ц а 8

Вспомогательная расчетная таблица

i	x_i	y_i	x_i^2	x_i^3	x_i^4	$y_i x_i$	$y_i x_i^2$	y_i^2	$\overline{y}_{x_i}$	$\left(\overline{y}_{x_i} - y_i\right)^2$
1	25	23	625	15625	390625	575	14375	529	21,7	1,69
2	27	24	729	19683	531441	648	17496	576	24,3	0,11
3	30	27	900	27000	810000	810	24300	729	27,6	0,33
4	35	27	1225	42875	1500625	945	33075	729	31,0	16,20
5	36	32	1296	46656	1679616	1152	41472	1024	31,4	0,33
6	38	31	1444	54872	2085136	1178	44764	961	31,9	0,88
7	39	33	1521	59319	2313441	1287	50193	1089	32,0	0,91
8	41	35	1681	68921	2825761	1435	58835	1225	32,0	9,14
9	42	34	1764	74088	3111696	1428	59976	1156	31,8	4,85
10	45	32	2025	91125	4100625	1440	64800	1024	30,7	1,75
11	46	29	2116	97336	4477456	1334	61364	841	30,1	1,23
12	47	28	2209	103823	4879681	1316	61852	784	29,4	2,10
13	50	25	2500	125000	6250000	1250	62500	625	26,9	3,52
14	52	24	2704	140608	7311616	1248	64896	576	24,7	0,46
15	53	25	2809	148877	7890481	1325	70225	625	23,4	2,44
$\sum$	606	429	25548	1115808	50158200	17371	730123	12493	-	45,94

Для оценки тесноты связи вычислим индекс корреляции.

Из таблицы 8 найдем суммы:

$$\delta_{yx}^2 = \sum_{i=1}^{n}\left(y_{x_i} - y_i\right)^2 = 45,94;$$

$$\delta_y^2 = \sum_{i=1}^{n}\left(y_i - y\right)^2 = \sum_{i=1}^{n} y_i^2 - \frac{\left(\sum\limits_{i=1}^{n} y_i\right)^2}{n} = 12493 - \frac{429^2}{15} = 223,6.$$

Подставляя эти данные, находим

$$R_{yx} = \sqrt{1 - \frac{\delta_{yx}^2}{\delta_y^2}} = \sqrt{1 - \frac{45,94}{223,6}} = \sqrt{0,795} = 0,891,$$

т.е. полученная зависимость весьма тесная. Коэффициент детерминации $R_{yx}^2 = 0,795$ показывает, что вариация урожайности зерновых культур на 79,5 % обусловлена регрессией, или изменчивостью количества выпавших в вегетационный период осадков.

В некоторых случаях нелинейность связей является следствием качественной неоднородности совокупности, к которой применяют регрессионный анализ. Например, объединение в одной совокупности предприятий различной специализации или предприятий, существенно различающихся по природным условиям, и т.д. В этих случаях нелинейность может являться следствием механического объединения разнородных единиц. Регрессионный анализ таких совокупностей не может быть эффективным. Поэтому любая нелинейность связей должна критически анализироваться.

Следует отметить, что по расположению точек корреляционного поля далеко не всегда можно принять окончательное решение о виде уравнения регрессии. Если теоретические соображения или опыт предыдущих исследований не могут подсказать точного решения, то необходимо сделать расчеты по двум или нескольким уравнениям. Предпочтение отдается уравнению, для которого меньше величина остаточной дисперсии. Однако при незначительных расхождениях в остаточных дисперсиях следует всегда останавливаться на более простом уравнении, интерпретация показателей которого не представляется сложной.

Весьма соблазнительным представляется увеличение порядка выравнивающей параболической кривой, ибо известно, что всякую функцию на любом интервале можно как угодно точно приблизить полиномом

$$y = b_0 + b_1 x + b_2 x^2 + \ldots + b_k x^k.$$

Так можно подобрать такой показатель k, что соответствующий полином пройдет через все вершины эмпирической линии регрессии. Однако повышение порядка гипотетичной параболической кривой может привести к неоправданному усложнению вида искомой функции регрессии, когда случайные отклонения осредненных точек неправильно истолковываются как определенные закономерности в поведении кривой регрессии. Кроме того, за счет увеличения числа параметров снижается точность кривой регрессии (особенно в случае малой по объему выборки) и увеличивается объем вычислительных работ. В связи с этим в практике регрессионного анализа для выравнивания крайне редко используются полиномы выше третьей степени.

References

1 *Ezekiyel M., Fox K.A.* Methods of the analysis of correlations and regressions linear and curvilinear: The lane with English. — Moscow: Statistics, 1966. — 558 p.

2 *Ionesku K., To Iordak V., Moynyagu K., Postelnika V., Shatteles T.* Statistical methods of research of correlations in economy: The lane with рум. — Moscow: Statistics, 1972. — 160 p.

3 *Yeliseyev I.I., Kurysheva S.V., Kosteev T.V., Babayeva I.V., Mikhaylov B.A.* Ekonometrika. — Moscow: Finance and statistics, 2001. — 344 p.

4 *Politova I.D.* The dispersive and correlation analysis in economy: Manual. — Moscow: Economy, 1972. — 224 p.

5 *Siskov V.I.* The correlation analysis in economic researches. — Moscow: Statistics, 1975. — 168 p.

6 *Chavkin A.M.* Methods and models of rational management in market economy: development of administrative decisions: Manual. — Moscow: Finance and statistics, 2001. — 320 p.

7 *Thomas R*. Quantitative methods of the analysis of economic activity: The lane with English. — Moscow: Business and Service, 1999. — 432 p.

8 *Grishin A.F., Kocherova E.V.* Statistical models: construction, mark, analysis: Manual. — Moscow: Finance and statistics, 2005. — 416 p.

9 *Shikin E.V., Chkhartishvili A.G.* Mathematical methods and models in management: Manual. — Moscow: Business, 2000. — 440 p.

10 *Karenov R.S.* Modeling and forecasting of efficiency of mining in market conditions. — Karaganda: IPTs «Profobrazovaniye», 2006. — 280 p.

11 *Savitsky G.V.* Analiz of economic activity of the enterprise. — Minsk: SP «Ekoperspektiva», 1997. — P. 115.

12 *Lyubushin N.P., Leshcheva V.B., Dyakova V.G.* Analysis of financial and economic activity of the enterprise: Manual. — Moscow: UNITI — it is GIVEN, 1999. — P. 49.

13 *Karasev A.I., Kremer N.Sh., Savelyev T.I.* Mathematical methods and models in planning: Manual. — Moscow: Economy, 1987. — 240 p.

Р.С.Каренов

Экономикалық зерттеулерде жұп корреляция тәсілдерін қолдану әдістемесі

Экономикалық зерттеулерде корреляциялық-регрессиялық талдауды қолданудың негізгі алғышарттары қарастырылған. Жұп корреляция байланыстарын корреляциялық талдаудағы есептеулер алгоритмінің кезеңдері талданған. Астық дақылдары түсімінің жыртпалы жер алқаптарының сапасына байланысты өзгерісін корреляциялық-регрессиялық талдаудың әдістемелік негіздері берілген. Қарағанды облысы аудан орталығының халқы санының динамикасын корреляциялық-математикалық аппаратын қолдану арқылы уран кеніші жұмысының экономикалық көрсеткішін талдаудың әдістемелік принциптері зерттелген. Корреляциялық-регрессиялық талдаудың қолданбалы аппаратын пайдаланып, Қарағанды аймағының астық егуші аудандарының шаруа қожалықтарының қызметін зерттеу бойынша әдістемелік қағидалар ұсынылған.

R.S.Karenov

A technique of application of ways of pair correlation in economic researches

The main preconditions of application of the correlation and regression analysis in economic researches are considered. Stages of algorithm of calculations are analyzed in the correlation analysis of communication of pair correlation. Methodical bases of the correlation and regression analysis of change of productivity of grain crops depending on quality of an arable land are offered. The methodical approach of the correlation and regression analysis of dynamics of population of the regional center Karaganda облатси is developed. Methodical principles of the analysis of an economic indicator of work of uranium mine with use of mathematical apparatus of pair correlation are recommended. Methodical provisions on studying of activity of country farms of zernoseyushchy areas of the Karaganda region with use of the applied device of the correlation and regression analysis are made