

Теоретические основы и методика математико-статистического анализа экономических показателей работы предприятий

Theoretical bases and a technique of the mathematical-statistical analysis of economic indicators of work of the enterprises

Каренов Р.С.

Карагандинский государственный университет имени Е.А.Букетова (E-mail: karenov_r@inbox.ru)

Корреляциялық-регрессиялық талдау тәсілімен экономикалық құбылыстар мен үдерістерді зерттеу маңыздылығы негізделген. Оңтайландыру мақсатында экономикалық-математикалық модельдерді құрудың барлық үдерісін бірнеше кезеңдерге бөлу қажеттілігі дәлелденген. Міндет қою кезеңі басты кезең болып табылатындығы көрсетілген; себебі мұнда мәселенің мәні қалыптасып, оны шешу жолдары анықталады. Экономикалық құбылыстар мен үдерістерді корреляциялық тәсілдермен талдап зерттеу барысында екі мәселе шешілуі тиіс екендігі көрсетілген: қандай факторлар бізді қызықтыратын көрсеткішке ықпал етеді және олардың қайсысы зерттеуге және модельге енгізіледі. Ақпараттарды жинау — модель құру кезеңінің еңбекті көп қажет ететін және өте жауапты кезеңі екендігі баса айтылған, себебі тиянақты ақпарат жинауға зерттеудің түпкі нәтижесі тәуелді екендігі пайымдалған. Өндірістік функция құрудағы келесі кезең — оның математикалық теңдеуді таңдау екендігі дәріптелген. Экономикалық-математикалық модель құрудың екінші бөлімін құрайтын қалған келесі кезеңдеріне көңіл бөлінген.

The importance of research of economic events and processes by a method of the correlation-regression analysis is proved. It is proved that with a view of simplification all process of economic-mathematical models construction are expedient for dividing into a number of stages. It is shown that at studying of economic events and processes by correlation methods of the analysis two questions should dare: what factors can influence an indicator interesting us and what of them should be included in research and in model. It is noticed that information gathering is one of the most labor-consuming and responsible stages of construction of model as the research end result depends on carefulness of gathering of the initial data. It is allocated that the subsequent stage in the course of production function construction is an identification of its mathematical model, that is a choice of the corresponding mathematical equation. The attention to other, subsequent stages of process of construction of economic-mathematical model which in aggregate make its second part is focused.

Исследование экономических явлений и процессов методом корреляционно-регрессионного анализа

В условиях рыночной экономики от отечественных предприятий требуется повышение эффективности производства, конкурентоспособности продукции и услуг на основе внедрения достижений НТП, эффективных форм хозяйствования и управления производством, преодоления бесхозяйственности, активизации предпринимательства, инициативы и т.д.

Важная роль в реализации этой задачи отводится анализу хозяйственной деятельности предприятий. С его помощью вырабатываются стратегия и тактика развития предприятия, обосновываются планы и управленческие решения, осуществляется контроль за их выполнением, выявляются резервы повышения эффективности производства, оцениваются результаты деятельности предприятия, его подразделений и работников. Квалифицированный специалист — менеджер, экономист, финансист, маркетолог, аудитор — должен не только хорошо знать общие закономерности и тенденции развития экономики в условиях рыночных отношений, но и тонко понимать проявления общих, специфических и частных экономических законов в практике своего предприятия, своевременно замечать тенденции и возможности повышения эффективности производства. Он должен владеть современными методами экономических исследований, методикой системного, комплексного экономического анализа, мастерством точного, своевременного, всестороннего анализа результатов хозяйственной деятельности.

Вариация экономических явлений — это результат сложного взаимодействия многочисленных факторов, которые действуют непосредственно и опосредованно. Глубокая и довольно тесная взаимосвязь между экономическими показателями приводит к тому, что изменения в одних из них частично определяются изменениями в других. Поэтому возникают большие трудности при определении количественной меры связи экономических явлений. Сложность экономических явлений требует

применения для их анализа математико-статистических методов, которые являются ценным дополнением к обычным методам анализа [1–5].

Основная задача факторного анализа — определить степень влияния каждого фактора на уровень результативного показателя. Для этой цели применяются способы корреляционно-регрессионного, дисперсионного, компонентного, современного многомерного факторного анализа и т.д.

Одним из важнейших для исследования технико-экономических явлений является раздел математической статистики (корреляционно-регрессионный анализ), представляющий совокупность методов и приемов выявления, изучения и количественной оценки взаимосвязей и взаимозависимостей между явлениями и их признаками [6–8]. В экономических исследованиях имеют место две формы связи: корреляционная и регрессионная. Две случайные величины являются корреляционно связанными, если математическое ожидание одной из них меняется в зависимости от изменения другой.

Метод корреляционного анализа является наиболее распространенным при исследовании взаимосвязи производственных факторов с экономическими показателями. Корреляционный анализ позволяет количественно оценить взаимосвязь между величинами в условиях, когда действует большое количество факторов, ряд из них неизвестен. Он дает возможность установить, как в среднем изменяется случайная величина с изменением одной или нескольких других случайных или неслучайных величин при фиксированном значении неучтенных факторов. Это особенно важно в изучении экономических явлений, поскольку при экономическом анализе практически невозможно учесть все действующие факторы.

Однако не все факторы, влияющие на экономические явления, случайные величины. Поэтому в экономических исследованиях часто приходится рассматривать связи между случайными и неслучайными величинами. Такие связи называются регрессионными, а метод их изучения — регрессионным анализом.

Регрессионный анализ тесно связан с корреляционным. В то же время, в отличие от корреляционного, он предъявляет менее жесткие требования к исходной информации. Проведение регрессионного анализа, например, возможно даже в случае некоторого отличия распределения случайной величины от нормального, что особенно важно, так как обычно распределения экономических показателей асимметричны.

Корреляционный анализ вместе с регрессионным позволяет решать три различные, но связанные между собой задачи: определять коэффициент корреляции, оценивающий силу связи между изучаемыми явлениями; рассчитать и построить уравнение регрессии, определяющее форму связи и, наконец, с привлечением оценок достоверности определять реальность существования связи.

Экономические исследования, использующие аппарат корреляционно-регрессионного анализа, состоят из следующих этапов:

- 1) постановка задачи, экономический анализ объекта исследования;
- 2) выбор зависимого (результативного) показателя и влияющих на него факторов;
- 3) сбор и оценка исходной информации, необходимой для корреляционно-регрессионного анализа;
- 4) изучение характера и моделирование связи между факторами и результативным показателем (подбирается и обосновывается математическое уравнение, которое наиболее точно выражает сущность исследуемой зависимости);
- 5) расчет основных показателей связи корреляционно-регрессионного анализа;
- 6) статистическая оценка результатов корреляционно-регрессионного анализа, экономическая интерпретация полученного уравнения и его практического использования.

Постановка задачи

Этап постановки задачи является главным, так как здесь формулируется сущность проблемы и определяются пути ее решения.

С математической точки зрения всякая задача состоит из условия, в котором указаны известные и неизвестные величины, а также приведены существующие соотношения между ними, и главного вопроса, на который необходимо обоснованно ответить, т.е. задаются какие-то данные и подчеркивается, что следует определить.

Поставить задачу экономического характера — это значит четко указать, что должно быть получено в конечном счете, и привести необходимую информационную базу. Поскольку в условиях про-

изводства различных задач может быть поставлено большое множество, то в самом начале принято формулировать их цель. Само условие задачи приводится, как правило, на втором месте.

Следует отметить также, что сама постановка задачи бывает различной: общей и конкретной. В экономических исследованиях решают в основном конкретные задачи. Но постановка экономических задач в общем виде практикуется довольно часто. Обычно вначале формулируют задачу общего характера, а потом ее конкретизируют по месту и времени, наполняя необходимой информацией.

Правильная постановка экономической задачи имеет большое значение. Ошибка, допущенная в постановке задачи, приводит к тому, что все результаты, полученные при ее решении, являются в лучшем случае бесполезными.

Следует определить, насколько необходимо решать поставленную задачу, в какой мере она актуальна, теоретического или практического она характера.

Ставя задачу получения экономико-математической модели, требуется предварительно выяснить примерную эффективность ее решения и внедрения в производство, а также оценить потенциальные потери от неиспользования изучаемой зависимости.

Необходимо также иметь в виду, что задача должна быть четко сформулирована. Поэтому при ее постановке следует избегать двусмысленных выражений и т.п. Нужно, чтобы все в задаче определялось однозначно. Только при четкой постановке задачи можно рассчитывать на столь же четкие результаты ее решения.

Рассматривая задачу, следует помнить, к какой группе она относится — правильных (обычных) или оригинальных задач, с какой целью ставится — она аналитическая, плановая или исследовательская. Сообразно с характером задачи следует уделять ей соответствующее внимание и усилие. Разумеется, все задачи важны, но в условиях дефицита времени и сил какие-то из них предпочтительнее.

При постановке задачи весьма важно знать ее информационную обеспеченность; имеется ли необходимая информация, в какой мере она достоверна, достаточен ли ее объем; если данные для построения экономико-математической модели отсутствуют, где их можно получить и т.д. Многие поставленные задачи остаются нерешенными из-за отсутствия информации или по причине ее недостоверности.

Иногда отдельные экономические задачи не получают своего решения потому, что не разработаны соответствующие методы. Многие авторы считают, что вначале следует поставить задачу, а затем определять методы и приемы ее решения. Во многих случаях это правомерно, но далеко не всегда. Решая задачу, нередко используют не самые совершенные методы. Все это подтверждает необходимость учета методического обеспечения при постановке задачи. Выяснив все возможные усложняющие моменты, игнорировать которые никогда не следует, формулируют постановку задачи по построению экономико-математической модели.

Выбор зависимого показателя и влияющих факторов

Сущность второго этапа построения экономико-математической модели сводится к выбору результативного зависимого показателя и факторов, оказывающих на него влияние.

Исходя из цели и задачи исследования путем теоретического анализа существующих связей приходят к заключению о том, какой экономический показатель является зависимым. Таким зависимым или результативным признаком обычно выбирается тот, который наиболее полно характеризует исследуемый процесс. Этот результативный признак может выражаться прямым или каким-либо косвенным показателем.

Результативные экономические показатели находятся в зависимости от большого числа различных производственных факторов. Поэтому отбор факторов корреляционно-регрессионного анализа является очень важным моментом в экономическом анализе. От того, насколько правильно сделан отбор факторов, зависит точность выводов по итогам анализа. Главная роль при отборе факторов принадлежит теории, а также практическому опыту анализа. При этом необходимо придерживаться следующих правил.

1. При отборе факторов в первую очередь следует учитывать причинно-следственные связи между показателями, так как только они раскрывают сущность изучаемых явлений. Анализ же таких факторов, которые находятся только в математических соотношениях с результативным показателем, не имеет практического смысла.

2. При создании многофакторной корреляционной модели необходимо отбирать самые значимые факторы, которые оказывают решающее воздействие на результативный показатель, так

как охватить все условия и обстоятельства практически невозможно. Факторы, которые имеют критерий надежности по Стьюденту меньше табличного, не рекомендуется принимать в расчет.

3. Все факторы должны быть количественно измеримы, т.е. иметь единицу измерения, и информация о них должна содержаться в учете и отчетности.

4. В корреляционную модель линейного типа не рекомендуется включать факторы, связь которых с результативным показателем имеет криволинейный характер.

5. Не рекомендуется включать в корреляционную модель взаимосвязанные факторы. Если парный коэффициент корреляции между двумя факторами больше 0,85, то по правилам корреляционного анализа один из них необходимо исключить, иначе это приведет к искажению результатов анализа.

6. Нельзя включать в корреляционную модель факторы, связь которых с результативным показателем носит функциональный характер.

Большую помощь при отборе факторов для корреляционной модели оказывают аналитические группировки, способ сравнения параллельных и динамических рядов, линейные графики. Благодаря им можно определить наличие, направление и форму зависимости между изучаемыми показателями. Отбор факторов можно производить также в процессе решения задачи корреляционного анализа на основе оценки их значимости по критерию Стьюдента.

Сбор информации и требования к ней

После выбора зависимого и факторных признаков проводится сбор необходимой информации, что составляет основное содержание третьего этапа построения экономико-математической модели. В практике находят применение различные способы получения данных, важнейшие из которых:

- а) постановка эксперимента;
- б) обработка статистических материалов;
- в) комбинированный способ.

Первый способ получения исходной информации наиболее совершенный, поскольку постановка эксперимента обеспечивает высокую достоверность данных. Своевременность получения последних гарантируется самим исследователем. Точность информации и своевременность ее поступления позволяют считать постановку эксперимента наиболее приемлемым способом с научной точки зрения.

Но в экономике постановка эксперимента, как правило, очень дорога и не всегда практически возможна и целесообразна. Поэтому часто используют второй способ получения необходимой информации. Разумеется, он не обеспечивает высокой точности исходных данных и не всегда оперативно они могут быть получены. Результаты построения экономико-математической модели при данном способе получения информации зависят не только от исследователя.

При третьем способе одна часть исходных данных является результатами проведения эксперимента, а другая — получается путем обработки статистических материалов, т.е. данный способ представляет собой комбинацию предыдущих, поэтому называется комбинированным. В нем сочетаются достоинства и недостатки первых двух способов. Исследователь должен стремиться к тому, чтобы это сочетание было оптимальным, что положительно скажется на качестве результатов.

Из описанных трех способов получения информации наилучшим считается первый способ — постановка эксперимента, однако он и более трудоемок. Это закономерно. Более высокое качество достигается дополнительными затруднениями. Наименее трудоемок второй способ — получение данных путем обработки статистических материалов, но он наиболее распространен.

Полученная каким-либо способом информация должна соответствовать определенным требованиям. Прежде всего ее объем должен быть достаточным. Из практики известно, что для построения однофакторной модели необходимо иметь 20–30 наблюдений. Чтобы получить многофакторную зависимость, объем необходимой информации определяют по формуле

$$(n + m) \ll (n - m)^2, \quad (1)$$

где n и m — соответственно число наблюдений и количество факторов в экономико-математической модели.

В дополнение к формуле отметим, что количество производственных факторов должно быть меньше числа наблюдений в 5 раз и более.

Сбор статистических данных и их первичная статистическая обработка — проверка однородности различных групп наблюдений, исключение аномальных (резко выделяющихся) результатов наблюдений. Собранные статистические данные представляются в виде таблицы.

Т а б л и ц а 1

Исходные статистические данные для проведения экономического исследования, использующего аппарат корреляционно-регрессионного анализа

Порядковый номер реализации	Факторы						Функция y
	x_1	x_2	x_j	x_m	
1	x_{11}	x_{12}		x_{1j}		x_{1m}	y_1
2	x_{21}	x_{22}				x_{2m}	y_2
		x_{2j}	
i	x_{i1}	x_{i2}			x_{im}	y_i

n	x_{n1}	x_{n2}		x_{nj}		x_{nm}	y_n

В таблице 1 m — количество исследуемых факторов; n — количество реализаций (объем выборки); x_{ij} — значение i -й реализации j -го фактора. В случае одного фактора ($m=1$) наряду с табличным способом представления данных широко используется графический способ. Для этого в прямоугольной системе координат откладывают по оси абсцисс значения фактора x , по оси ординат — функции y и заполняют квадранты точками с координатами (x_i, y_i) ; i — номер реализации. Подобная точечная диаграмма называется корреляционным полем.

Одно из главных условий корреляционно-регрессионного анализа — однородность исследуемой информации относительно распределения ее около среднего уровня. Если в совокупности имеются группы объектов, которые значительно отличаются от среднего уровня, то это говорит о неоднородности исходной информации. Критерием однородности информации служит среднеквадратическое отклонение (σ) и коэффициент вариации (V), которые рассчитываются по каждому факторному и результативному показателю.

Среднеквадратическое отклонение показывает абсолютное отклонение индивидуальных значений от среднеарифметического. Оно определяется по формуле

$$\sigma = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}. \quad (2)$$

Коэффициент вариации показывает относительную меру отклонения отдельных значений от среднеарифметической. Он рассчитывается по формуле

$$V = \frac{\sigma}{\bar{x}} * 100. \quad (3)$$

Чем больше коэффициент вариации, тем относительно больший разброс и меньшая выравненность изучаемых объектов. Изменчивость вариационного ряда принято считать незначительной, если вариация не превышает 10 %, средней — если составляет 10–12 %, значительной — если она больше 20 %, но не превышает 33 %. Если же вариация выше 33 %, то это говорит о неоднородности информации и ее необходимо исключить или отбросить нетипичные наблюдения, которые обычно бывают в первых и последних ранжированных рядах выборки.

На основании самого высокого показателя вариации можно определить необходимый объем выборки данных для корреляционного анализа по следующей формуле:

$$n = \frac{V^2 * t^2}{q^2}, \quad (4)$$

где n — необходимый объем выборки данных; V — вариация, %; t — показатель надежности связи, который при уровне вероятности $P=0,05$ равен 1,96; q — показатель точности расчетов (для экономических расчетов допускается ошибка 5–8 %).

Следующее требование к исходной информации — соответствие ее закону нормального распределения. Согласно этому закону основная масса исследуемых сведений по каждому показателю должна быть сгруппирована около ее среднего значения, а объекты с очень маленькими или очень большими значениями должны встречаться как можно реже. График нормального распределения имеет следующий вид (рис. 1).

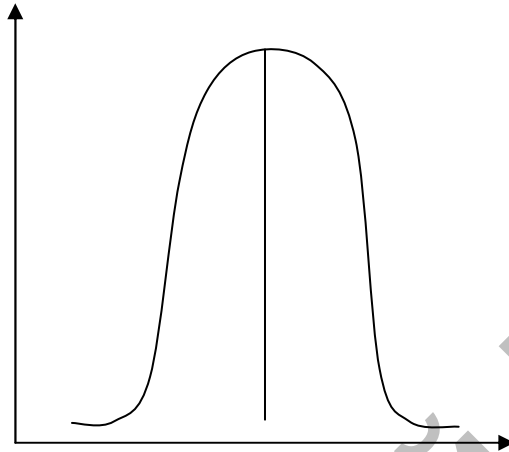


Рис. 1. График нормального распределения информации (данные работы [9; 127])

Для количественной оценки степени отклонения информации от нормального распределения служит отношение показателя асимметрии к его ошибке и отношение показателя эксцесса к его ошибке.

Показатель асимметрии (A) и его ошибка (m_a) рассчитываются по следующим формулам:

$$A = \frac{\sum (x - \bar{x})^3}{n\sigma^3}; m_a = \sqrt{\frac{6}{n}}. \quad (5)$$

Показатель эксцесса (E) и его ошибка (m_e) рассчитываются следующим образом:

$$E = \frac{\sum (x - \bar{x})^4}{n\sigma^4}; m_e = \sqrt{\frac{24}{n}}; \text{ или } m_e = 2m_a. \quad (6)$$

В симметричном распределении $A=0$. Отличие от нуля свидетельствует о наличии асимметрии в распределении данных около средней величины. Отрицательная асимметрия свидетельствует о том, что преобладают данные с большими значениями, а с меньшими значениями встречаются значительно реже. Положительная асимметрия показывает, что чаще встречаются данные с небольшими значениями.

В нормальном распределении показатель эксцесса $E=0$. Если $E > 0$, то данные густо сгруппированы около средней, образуя островершинность. Если $E < 0$, то кривая распределения будет плосковершинной. Однако когда отношения A/m_a и E/m_e меньше 3, то асимметрия и эксцесс не имеют существенного значения и исследуемая информация подчиняется закону нормального распределения.

Выбор вида корреляционной зависимости

После отбора факторов и оценки исходной информации важной задачей в корреляционном анализе является моделирование связи между факторными и результативными показателями, т.е. подбор соответствующего уравнения, которое наилучшим образом описывает изучаемые соотношения.

Для обоснования функции используются те же приемы, что и для установления наличия связи: аналитические группировки, линейные графики и др. Если связь всех факторных показателей с результативным носит прямолинейный характер, то для записи этих зависимостей можно использовать линейную функцию

$$Y_x = a + b_1x_1 + b_2x_2 + b_3x_3 + \dots + b_nx_n. \quad (7)$$

Если связь между функцией и исследуемыми показателями носит криволинейный характер, то может быть использована степенная функция

$$Y_x = b_0 \cdot x_1^{b_1} \cdot x_2^{b_2} \cdot \dots \cdot x_n^{b_n} \quad (8)$$

или логарифмическая

$$\lg Y_x = b_0 + b_1 \lg x_1 + b_2 \lg x_2 + \dots + b_n \lg x_n. \quad (9)$$

Приведенные модели выгодны тем, что их параметрам (b_i) можно дать экономическое объяснение (интерпретацию). В линейной модели коэффициенты b_i показывают, на сколько единиц изменится результативный показатель с изменением факторного на единицу в абсолютном выражении, в степенных и логарифмических — в процентах.

В случаях, когда трудно обосновать форму зависимости, решение задачи можно провести по разным моделям и сравнить полученные результаты.

Выбор вида функции упрощается при наличии лишь одного аргумента, когда возможно графическое представление исходных данных и когда по плотности точек корреляционного поля можно судить о виде корреляционной зависимости $\bar{y} = f(x)$.

Вообще корреляционный анализ, как правило, начинается с изучения связи двух переменных. Принята общая схема такого анализа: выбор функции, аргумента, их математическое обозначение; исчисление средних арифметических, среднеквадратических (σ) значений признака и коэффициента вариации (V); выбор формы связи; расчет эмпирической и теоретической линии регрессии; расчет коэффициента корреляции.

При выборе вида уравнения можно пользоваться формулой

$$K = \frac{\sqrt{N}}{0,67449} \cdot \frac{1}{2} \cdot \sqrt{\eta^2_{yx} - r^2_{yx}}, \quad (10)$$

где N — объем совокупности; η — теоретическое корреляционное отношение; r — парный коэффициент корреляции.

Если $K > 3$, то принимается прямолинейная зависимость. В случае прямолинейной зависимости двух явлений уравнение будет иметь вид: $y = a + bx$.

Зависимость между показателями определяется теоретической линией регрессии. В практическом использовании теоретическая линия регрессии представляет наилучший вариант анализируемого показателя, к которому необходимо стремиться при изменении известного или выбираемого значения функции. Процесс нахождения теоретической линии регрессии называется выравниванием эмпирической линии регрессии. В основе выравнивания лежит метод наименьших квадратов, в котором используется формула

$$\sum (y - \bar{y}_x)^2 = \min, \quad (11)$$

где y — ордината эмпирической линии регрессии; \bar{y}_x — ордината теоретической линии регрессии.

Формула носит название требования наименьших квадратов, а способ отыскания параметров, основанный на этом требовании, называется способом наименьших квадратов. Он является одним из наиболее употребительных методов приближенной оценки неизвестных нам параметров уравнений. По этому способу составляется система нормальных уравнений:

$$\sum x \cdot y = a \cdot \sum x + b \cdot \sum x^2. \quad (12)$$

Однако удобнее и проще при нахождении параметров уравнения пользоваться основными показателями: \bar{x} , \bar{y} , r , σ_x , σ_y , где r — показатель тесноты связи между двумя величинами — парный коэффициент корреляции. Он изменяется от -1 до $+1$; при $r = \pm 1$ наблюдается не корреляционная, а прямая или обратная функциональная связь; при $r = 0$ связь отсутствует.

Разброс точек вокруг эмпирической линии регрессии показывает характер корреляционной связи. Если в результате нанесения точек на корреляционное поле окажется, что ось контура его параллельна одной из осей координат или контур приближается к форме круга, то такое соотношение исследуемых данных показывает на почти полное отсутствие корреляционной связи.

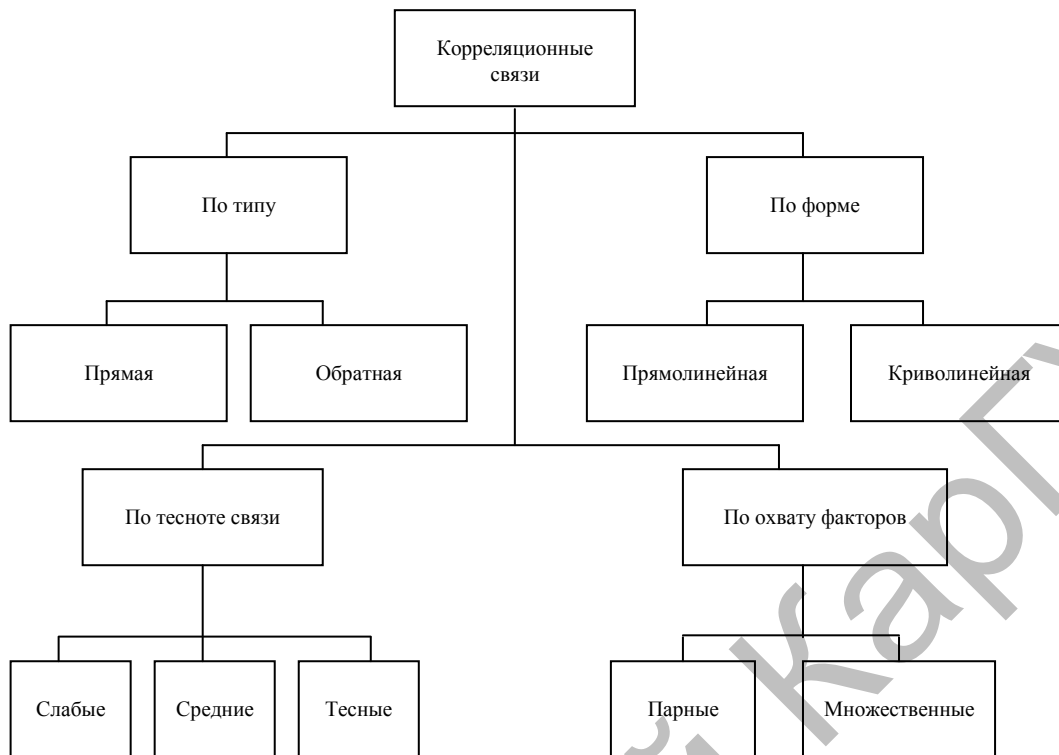


Рис. 2. Схема корреляционных связей (данные работы [10; 11])

Коэффициент корреляции r_{xy} , или коэффициент регрессии в стандартизованной системе единиц, отвечает на вопрос: на сколько сигм изменяется в среднем y , когда x увеличивается на одну сигму. Например, при $r = 0,5$ коэффициент показывает, что если аргумент отклонится от своей средней на свое среднее квадратическое отклонение σ_x , то y — функция отклонится от своей средней на 0,5 своего квадратического отклонения σ_y . Коэффициент корреляции r_{xy} отличается от углового коэффициента уравнения регрессии «в». Численное значение «в» зависит от того, что принято за единицу измерения по x и по y . Если единица измерения x велика, то увеличению последней на единицу соответствует большее изменение среднего y , т.е. «в» велико. Если единица измерения y велика, то соответствующее изменение y выражается меньшим количеством единиц и, следовательно, «в» мало. Так как обычно рассматриваются зависимости между признаками разной природы, имеющими различные единицы измерения (например, можно признак измерять в сантиметрах, метрах, километрах, граммах, килограммах, тоннах и т.д.), то коэффициент «в» является величиной, зависящей от произвола в выборе единиц измерения переменных. Отсюда он не может быть универсальным показателем силы связи.

Данные по различным признакам сравнимы между собой, если они выражены в едином измерении. В статистике есть система, которая использует в качестве единицы измерения признака его среднее квадратическое отклонение (при этом начало отсчета перемещается в точку средней арифметической). Коэффициент корреляции, или регрессии, в этом случае наиболее точно определяет тесноту зависимости или силу связи. Он имеет тот же знак, что параметр «в». На рисунке 2 приведена схема корреляционных связей.

Коэффициент корреляции вычисляют по формуле

$$r = \frac{n \sum xy - \sum x \sum y}{\sqrt{n \sum x^2 - (\sum x)^2} \sqrt{n \sum y^2 - (\sum y)^2}}. \quad (13)$$

Существует два приема производства корреляции: когда данные сгруппированы и не сгруппированы.

На первом этапе корреляционного анализа парных уравнений и коэффициентов предварительно отбираются главные факторы, устанавливаются характер и теснота связи между функцией и аргументами.

Профессором Колумбийского университета США Чэддоном предложены следующие показатели качественной оценки тесноты связи (см. табл. 2).

Т а б л и ц а 2

Показатели качественной оценки тесноты связи, предложенные профессором Чэддоном (США)

Величина (r)	0,1–0,3	0,5 — 0,7	0,3 — 0,5	0,7 — 0,9	свыше 0,9
Теснота связи	Слабая	Заметная	Умеренная	Высокая	Весьма высокая

Примечание. Используются данные работы [10; 13].

В литературе стран СНГ [11–14] принято считать: $r < 0,3$ — слабая связь; $r = 0,3 \div 0,7$ — средняя связь; $r > 0,7$ — сильная связь.

При изучении тесноты связи надо иметь в виду, что величина коэффициентов корреляции является случайной, зависящей от объема выборки. Известно, что с уменьшением количества наблюдений надежность коэффициентов корреляции падает, и наоборот, при увеличении количества наблюдений — возрастает.

Значимость коэффициентов корреляции проверяется по критерию Стьюдента:

$$t = \frac{r}{\sigma_r}, \quad (14)$$

где σ_r — среднеквадратическая ошибка коэффициента корреляции, которая определяется по формуле

$$\sigma_r = \frac{1 - r^2}{\sqrt{n - 1}}. \quad (15)$$

Если расчетное значение t выше табличного, то можно сделать заключение о том, что величина коэффициента корреляции является значимой. Табличные значения t находят по таблице значений критериев Стьюдента. При этом учитываются количество степеней свободы ($W = n - 1$) и уровень доверительной вероятности (в экономических расчетах обычно 0,05 или 0,01).

Расчет уравнения связи (регрессии)

Следующий этап корреляционно-регрессионного анализа — расчет уравнения связи (регрессии).

При расчете парных зависимостей определяется влияние выбранного фактора. Однако при этом остается неопределенное скрытое влияние других, не учтенных, но связанных с ним факторов. Для определения влияния большого числа факторов применяются методы множественной корреляции. Путем составления многофакторных экономико-статистических моделей можно ответить на вопрос, как каждый фактор численно влияет на изучаемый показатель и как изменится этот показатель с изменением каждого фактора на 1 % или на одну единицу своего измерения.

Коэффициенты уравнения множественной регрессии определяются двумя способами: способом парных коэффициентов корреляции и методом наименьших квадратов. Для расчета уравнения множественной регрессии все переменные переводятся в стандартизованный масштаб через средние квадратические отклонения, и уравнение в стандартизованном масштабе переводит « w », которые зависят от единиц измерения, в « β » — сравнимые коэффициенты.

Уравнение в стандартизованном масштабе имеет вид

$$\bar{t}_{1 \dots p} = \beta_1 t_1 + \beta_2 t_2 + \dots + \beta_p t_p, \quad (16)$$

где t — стандартизованные значения переменных; β — стандартизованные коэффициенты регрессии.

Для нахождения β -коэффициентов решаем систему уравнений:

$$\begin{aligned} r_{12} &= \beta_2 + \beta_3 r_{23} + \dots + \beta_p r_{2p}, \\ r_{1p} &= \beta_2 r_{2p} + \beta_3 r_{3p} + \dots + \beta_p. \end{aligned} \quad (17)$$

Коэффициенты β показывают, на какую часть сигмы изменилось бы среднее значение функции, если бы соответствующий аргумент увеличился на сигму, а прочие аргументы остались без изменения. Они означают скорость изменения среднего значения функции по каждому из аргументов. Так как все переменные выражены в сравнимых единицах измерения, то коэффициенты показывают силу влияния каждого аргумента. В этом заключается ценность для экономического анализа уравнения множественной регрессии в стандартизованном масштабе. Они дают возможность определить силу и долю влияния каждого фактора в отдельности, одновременное их влияние.

Бетта-коэффициенты и коэффициенты регрессии связаны следующими отношениями:

$$\beta_i = b_i \frac{\sigma_{xi}}{\sigma_y}. \quad (18)$$

Для экономической интерпретации нелинейных связей между функцией и одним из влияющих на нее факторов удобно пользоваться коэффициентом эластичности. Формулы вычисления эластичности для некоторых нелинейных функций сведены в таблицу 3.

Т а б л и ц а 3

Оценка коэффициентов эластичности при нелинейной форме связи

Вид уравнения регрессии	Формулы для вычисления коэффициентов эластичности
$Y = a_0 + a_1x + a_2x^2;$	$\mathcal{E} = (a_1 + 2a_2x) \frac{x}{y};$
$Y = a_0x^{a_1};$	$\mathcal{E} = a_1;$
$Y = \frac{a_0x}{x + a_1};$	$\mathcal{E} = \frac{a_0a_1}{(x + a_1)^2} \cdot \frac{x}{y};$
$Y = a_0 + a_1 \ln x;$	$\mathcal{E} = a_1 \frac{1}{y}$
$Y = a_0 - \frac{a_1}{x - a_2};$	$\mathcal{E} = \frac{a_1}{(x + a_2)^2} \cdot \frac{x}{y};$
$Y = a_0 + \frac{a_1}{x - a_2};$	$\mathcal{E} = -\frac{a_1}{(x - a_2)^2} \cdot \frac{x}{y};$
$Y = a_0 \frac{(x - a_2)}{(x + a_1)};$	$\mathcal{E} = \frac{a_0(a_1 + a_2)}{(a_0 + a_1)^2} \cdot \frac{x}{y}.$

Примечание. Использованы данные работы [15; 70].

Коэффициенты эластичности показывают, на сколько процентов в среднем изменяется функция с изменением аргумента на 1 %.

Для линейной зависимости коэффициент эластичности рассчитывается по формуле

$$\mathcal{E} = b_i \frac{\bar{x}_i}{y}. \quad (19)$$

Корреляционная модель может быть решена методом последовательного включения факторов.

Метод последовательного включения факторов позволяет вычислить для каждой зависимости показатель R — множественный коэффициент корреляции, характеризующий полноту учета факторов, и показатель $\sigma_{ост}$ — остаточное среднее квадратическое отклонение, которое показывает колеблемость признака вокруг ее расчетных значений за счет неучтенных факторов в данном исследовании. При этом методе в окончательную модель можно включать наиболее значимые факторы и не загромождать ее лишними, не влияющими на окончательный результат. Из уравнений множественной регрессии могут быть получены уравнения частных и чистых регрессий, широко используемые в экономическом анализе.

Математико-статистический метод частной корреляции помогает устанавливать взаимосвязь между двумя явлениями при условии постоянства остальных. Частный коэффициент корреляции имеет ту же природу, что и парный, и изменяется в пределах от -1 до $+1$. Анализ коэффициентов частной корреляции делается для того, чтобы уточнить наиболее существенные факторы, влияющие на функцию, ранжировать их по степени связи с функцией и окончательно отобрать главные.

Если в уравнении множественной регрессии один из аргументов, влияние которого определяется, оставить без изменения, а на место остальных поставить их средние значения, можно определить уравнение чистой регрессии. Эта операция отражается на свободном члене, а уравнения примут вид:

$$\begin{aligned} y_{12(345\dots p)} &= a + bx_2 \\ &\dots\dots\dots \\ y_{1p(2345\dots n)} &= a + bx_p. \end{aligned} \quad (20)$$

В этом случае изменяется один из аргументов и оказывает влияние на изменение функции, остальные закреплены (показаны в скобках), и влияние их на функцию исключается. Создается возможность определения влияния аргументов в их чистом виде. Уравнения чистой регрессии имеет смысл вычислять тогда, когда исключаемые переменные не сильно связаны с переменными, остающимися в уравнении. Если зависимость велика, целесообразно рассчитывать уравнения частных регрессий, где исключаемые переменные закрепляются не на средних, а на других выбранных или интересующих исследователя уровнях.

Тесноту связи переменной y с совокупностью переменных $x_1, x_2 \dots x_p$ характеризует коэффициент множественной корреляции, который в случае прямолинейной корреляции выражается формулой

$$R_{12\dots p} = \sqrt{\beta_2 r_{12} + \beta_3 r_{13} + \dots + \beta_p r_{1p}}. \quad (21)$$

По величине R определяется теснота связи между функцией и введенными в уравнения аргументами, удельный вес отобранных факторов в общем их количестве, оказывающих влияние на функцию, и данных, не учтенных уравнением.

Когда R значительно меньше единицы, необходимо расширить набор факторов.

Если при небольшом числе наблюдений n имеется много переменных p , то величина R корректируется на систематическую погрешность по формуле

$$R = \sqrt{1 - (1 - R^2) \cdot \frac{n-1}{n-p}}. \quad (22)$$

Квадрат коэффициента корреляции — коэффициент детерминации — показывает удельный вес анализируемых факторов на вариацию изучаемого признака: $\Phi = R^2$, а $1 - R^2$ — долю неучтенных факторов.

Коэффициент множественной корреляции характеризует ту часть первоначальной колеблемости зависимого показателя вокруг его средней, которую удалось устранить благодаря применению корреляционной формулы.

Если исходные данные выборочные, то корреляционная формула считается выборочной характеристикой, подлежащей распространению на всю совокупность. Отсюда возникает проблема оценки совместимости величины множественного коэффициента корреляции, полученного при выборочном обследовании, с гипотезой о равенстве его нулю в генеральной совокупности.

Если вероятность такой совместимости мала, гипотеза отвергается, и связь между зависимой переменной и исследуемыми факторами-аргументами в генеральной совокупности можно считать реально существующей. В математической статистике известен ряд методов оценки справедливости нулевой гипотезы при корреляционном анализе.

Простейшая схема оценки сводится к построению доверительных интервалов для R и выяснению вопроса о том, находится ли нуль внутри построенного интервала. Доверительный интервал для R :

$$R - t\sigma_R \leq R \leq R + t\sigma_R, \quad (23)$$

где σ_R — средняя квадратическая ошибка:

$$\sigma_R = \frac{1 - R^2}{\sqrt{n - p - 1}};$$

$$\frac{R}{\sigma_R} \geq t = 2,58. \quad (24)$$

Здесь t — это критерий Стьюдента — число, показывающее, во сколько раз R больше его средней квадратической ошибки. Доверительная вероятность принимается 0,99, т.е. выводы не подтверждаются одним случаем из ста. Если значение t превосходит величину, соответствующую принятой доверительной вероятности, то нулевая гипотеза о равенстве R нулю считается опровергнутой, так как нуль не содержится в его доверительных интервалах; R признается существенным.

*Методика оценки и практического применения результатов
корреляционно-регрессионного анализа*

Уравнения должны быть подвержены статистическому качественному и количественному анализу. Качественный анализ результатов решения корреляционных моделей при экономических исследованиях состоит в основном из двух частей: проверки соответствия знаков и относительной величины полученных параметров при отдельных факторах-аргументах общеэкономическим представлениям об их влиянии на исследуемый показатель. Если такое соответствие имеется, формула может быть признана пригодной.

Определенный интерес представляет исследование типа распределения отклонений фактических значений от расчетных.

Нормальное распределение этих отклонений может указать на случайный их характер.

Для того чтобы убедиться в точности (надежности) уравнения связи и правомерности его использования для практической цели, необходимо дать статистическую оценку надежности показателей связи. Для этого используются критерий Фишера (F -отношение); средняя ошибка аппроксимации ($\bar{\varepsilon}$).

Критерий Фишера рассчитывается следующим образом:

$$F = \frac{\sigma_{воспр}^2}{\sigma_{ост}^2}, \quad (25)$$

где $\sigma_{воспр}^2 = \frac{\sum (Y_{xi} - \bar{Y}_x)^2}{m-1}$; $\sigma_{ост}^2 = \frac{\sum (Y_i - Y_{xi})^2}{n-m}$; Y_{xi} — индивидуальные значения результативного показателя, рассчитанные по уравнению; \bar{Y}_x — среднее значение результативного показателя, рассчитанное по уравнению; Y_i — фактические индивидуальные значения результативного показателя; m — количество параметров в уравнении, учитывая свободный член уравнения; n — количество наблюдений (объем выборки).

Фактическая величина F -отношения сопоставляется с табличной и делается заключение о надежности связи. Если $F_{факт} > F_{табл}$, то гипотеза об отсутствии связи между результативным показателем и исследуемыми факторами отклоняется.

Для статистической оценки точности уравнения связи используется также средняя ошибка аппроксимации:

$$\bar{\varepsilon} = \frac{1}{n} \sum_{i=1}^n \left[\frac{Y_{xi} - \bar{Y}}{Y_i} \right]^2. \quad (26)$$

Чем меньше теоретическая линия регрессии (рассчитанная по уравнению) отклоняется от фактической (эмпиричной), тем меньше средняя ошибка аппроксимации.

Влияние каждого фактора на прирост (отклонение от плана) результативного показателя рассчитывается следующим образом:

$$\Delta Y_{xi} = b_i * \Delta x_i. \quad (27)$$

Результаты многофакторного регрессионного анализа могут быть также использованы для планирования и прогнозирования уровня результативного показателя. С этой целью необходимо в полученное уравнение связи подставить плановый (прогнозный) уровень факторных показателей.

Таким образом, многофакторный корреляционно-регрессионный анализ имеет важную научную и практическую значимость. Он позволяет изучить закономерности изменения результативного пока-

зателя в зависимости от поведения разных факторов, определить их влияние на величину результирующего показателя, установить, какие из них являются основными, а какие второстепенными. Этим достигается более объективная оценка деятельности предприятия, более точное и полное определение внутрихозяйственных резервов и плановых (прогнозных) показателей.

References

1. *Ezekiel M., Fox K.A.* Methods of the analysis correlation and regression linear and curvilinear: Trans. from engl. — M.: Statistics, 1966. — 558 p.
2. *Ionescu K., Iordake V. et al.* Statistical methods of the study correlation in economy: Trans. from rum. — M.: Statistics, 1972. — 160 p.
3. *Richard Thomas.* Quantitative methods of the analysis to economic activity: Trans. from engl. — M.: Publishing house «Deal and Service», 1999. — 432 p.
4. *Gambarov G.M., Zhuraveli N.M. et al.* Statistical modeling and forecasting: Tutorial. — M.: Finance and statistics, 1990. — 383 p.
5. *Karasev A.I., Karasev N.Sh., Savelieva T.I.* Mathematical methods and models in planning and abutted against-pouring: Tutorial. — M.: Economy, 1987. — 240 p.
6. *Eliseeva I.I., Kurysheva S.V. et al.* Econometrics. — M.: Finance and statistics, 2001. — 344 p.
7. *Dzhonston Dzh.* Econometric methods: Trans. from engl. — M.: Statistics, 1980. — 444 p.
8. *Grishin A.F., Kocherova E.V.* Statistical models: building, estimation, analysis: Tutorial. — M.: Finance and statistics, 2005. — 416 p.
9. *Savickaya G.V.* Analysis of economic activity of enterprises. — Minsk: IP «Ecoperspective», 1997. — 498 p.
10. *Ryzhova V.V., Kuznecova L.A.* Mathematical methods in analysis of economic activity of enterprises. — M.: Finance, 1970. — 88 p.
11. *Frenkeli A.A.* Mathematical methods of the analysis speakers and forecastings of productivity of labour. — M.: Economy, 1972. — 190 p.
12. *Bro G.G., Shnaydman L.M.* Methods of the analysis and forecastings to productivity of labour. — M.: Economy, 1972. — 190 p.
13. *Bolgozhin Sh.A., Vorob'ev B.M., Karenov R.S.* Forecasting of the work coal enterprise (on base of economic and mathematical models). — Alma-Ata: Science, 1984. — 200 p.
14. *Karenov R.S.* Modeling and forecasting to efficiency mountain production in market conditions. — Karaganda: IPC «Profeducation», 2006. — 280 p.
15. *Pereverzev M.P.* The mathematical methods and models of the economic analysis of production reserves. — Harkov: Publishing association «High school»; Harkov State University Publ., 1974. — 188 p.