УДК 330.45

Р.С.Каренов

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

Проблемы использования ранговой корреляции и корреляции альтернативных признаков в экономических исследованиях

Выделено практическое значение ранговой корреляции в экономических исследованиях. Отмечено, что в настоящее время более полно разработаны способы однофакторной ранговой корреляции. Подчеркнуто, что принцип нумерации вариантов статистических рядов является основой непараметрических методов изучения связи между явлениями или ранговой статистики. Доказано, что в экономических исследованиях большое применение при изучении связи между количественными и качественными признаками находят ранговые коэффициенты корреляции Спирмена и Кендэла. Предложена методика применения коэффициента конкордации в экономических расчетах и обоснованиях. Значительное внимание уделено анализу корреляционных связей, существующих между различными альтернативными признаками. Рекомендована методика измерения тесноты корреляционной связи между факторными и результативными признаками с помощью таких коэффициентов, как эмпирический коэффициент корреляционной связи, коэффициенты взаимной сопряженности Пирсона и Чупрова, коэффициенты контингенции и ассоциации.

Ключевые слова: ранг, корреляция, альтернативный признак, эндогенные, экзогенные переменные, модель, матрица, гипотеза, Спирмен.

Практическое значение ранговой корреляции

В экономических исследованиях нередко случается, что исходные данные не подчиняются нормальному закону распределения. Поэтому построение производственных функций является затруднительным и к тому же оно может привести к ошибочным результатам. В таких случаях целесообразно использовать так называемую ранговую корреляцию. К этой корреляции прибегают также тогда, когда отдельные показатели признаков не являются строго точными. Проводя предварительные, приближенные изучения, следует также применять способ ранговой корреляции.

Рангом в корреляционном анализе называют номер возрастающих (или убывающих) вариантов факторного или зависимого признака. Определяя ранги одинаковых уровней, им присваивают средние номера. Уровни признаков всей совокупности наблюдений ранжируют, т.е. их нумеруют в порядке возрастания или убывания. Таким образом, каждый уровень всех факторных и зависимого признаков получает свой ранг (номер). Эти ранги по всем признакам идут в следующем порядке: 1, 2,..., n (или наоборот).

Выполняя корреляционный анализ, исследуют корреляцию не между факторными и зависимыми признаками, а между их рангами. Корреляция рангов фактора (или факторов) с рангами результативного показателя называется ранговой корреляцией. Существует ряд способов определения ранговой корреляции. Изначально их разрабатывал К.Пирсон, затем эти способы получили развитие в работах Спирмена, Кендэла и других ученых. Более полно разработаны способы однофакторной ранговой корреляции [1–3].

Для ранжированных признаков показателем тесноты связи служит коэффициент ранговой корреляции. Если связь между изучаемыми признаками прямая, то с увеличением рангов фактора (x) также будут возрастать и ранги зависимого показателя (y). При обратной (отрицательной) зависимости возрастанию рангов факторного признака (x) будет соответствовать уменьшение рангов результативного показателя (y). В случае отсутствия корреляционной связи между указанными признаками нельзя обнаружить какую-либо закономерность в изменениях их рангов.

Априорный статистический анализ

При проведении всякого исследования необходимо учитывать: а) время, в течение которого данное исследование должно быть закончено, и б) имеющиеся материальные, денежные и людские ресурсы.

В процессе спецификации переменных нельзя переоценивать мнение отдельных, даже весьма квалифицированных специалистов о механизме развития отрасли и факторах, определяющих эффективность этого развития. Поэтому обычные консультации часто не дают возможности определить все факторы, отражающие с высокой степенью достоверности взаимосвязи экономических явлений. Из литературных источников также можно получить весьма неполную информацию такого рода.

Чтобы объективно оценить круг эндогенных и экзогенных переменных и конкретный набор их в каждом из уравнений модели, необходимо учесть точки зрения многих специалистов, принадлежащих к различным научным школам. Для этого целесообразно формализовать имеющиеся априорные сведения.

Рассмотрим возможную процедуру формализации априорных сведений и выбора факторов, существенно влияющих на функции (по мнению специалистов) и связанных с ними причинно-следственными отношениями.

Вначале составляется опросная анкета (табл. 1), подлежащее которой содержит набор предполагаемых для изучения эндогенных переменных-функций, а сказуемое — набор эндогенных переменных-аргументов и предопределенных переменных.

Т а б л и ц а 1

Опросная анкета

Эндогенные переменные-функции	Эндогенные переменные-аргументы и предопределенные переменные									
	y_1	...	y_k	...	y_K	z_1	…	z_l	...	z_L
1 y_1										
2 y_2										
… …										
k y_k										
... …										
m y_m										
Примечание. Данные работы [4; 81].										

Данная анкета рассылается соответствующим специалистам. Каждому из экспертов предлагается проранжировать отобранные аргументы по степени их влияния на функции. При желании опрашиваемые специалисты могут расширить список показателей в анкете. Фактору, который, по мнению данного специалиста, оказывает наибольшее влияние на соответствующую функцию, присваивается ранг 1, следующему — ранг 2 и т.д. Значения рангов проставляются в соответствующие графы.

Покажем процедуру статистического анализа априорной информации на примере спецификации одного уравнения модели.

Допустим, $K+L=n$ (i=1,2,…, n). Тогда для этого уравнения можно составить сводную анкету, так называемую «матрицу рангов» (табл. 2).

В таблице 2 x_{ij} — ранг j-го аргумента, данный i-м экспертом; m — число экспертов; n — число факторов-аргументов.

Если специалисту не удается различить по силе влияния некоторые факторы, то он вынужден приписывать им один и тот же ранговый номер. В этом случае вводятся так называемые «связанные ранги». Например, трем факторам в анкете i-го эксперта присвоен ранг 3. Их ранговый номер в сводной анкете равен: (3+4+5): 3=4. Если следующие два фактора в анкете i-го эксперта имели ранг 4 и 5, то в сводной анкете их ранг будет 6 и 7, т.е. происходит переформирование рангов. Иногда в сводной анкете появляются дробные ранги. Например, если в анкете восьмому и девятому фактору приписан ранг 8, то в сводной анкете их ранг будет 8, 5.

После заполнения сводной анкеты следует провести проверку по контрольной сумме. Ее значение находится по формуле

$$\sum_{j=1}^{n} x_{ij} = \frac{(1+n)n}{2}. \qquad (1)$$

Далее вычисляются фактические суммы всех строк. Они должны быть равны друг другу и одновременно контрольной сумме.

Теперь, когда есть уверенность, что матрица рангов составлена правильно, можно перейти к выяснению существенности влияния отобранных факторов на изучаемый показатель с точки зрения опрошенных специалистов. Для этого в таблице 2 подсчитываются суммы всех отдельных столбцов.

Т а б л и ц а 2

Матрица рангов

Эксперты	Аргументы							
	1	2	3	...	j	…	n	$\sum_{j=1}^{n} x_{ij}$
1	x_{11}	x_{12}	x_{13}	…	x_{1j}	…	x_{1n}	
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	
i	x_{i1}	x_{i2}	x_{i3}	…	x_{ij}	…	x_{in}	
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	
m	x_{m1}	x_{m2}	x_{m3}	…	x_{mj}	…	x_{mn}	
$\sum_{i=1}^{m} x_{ij}$								

Примечание. Данные работы [4; 82].

При этом

$$\sum_{i=1}^{m}\sum_{j=1}^{n} x_{ij} = \sum_{j=1}^{n}\sum_{i=1}^{m} x_{ij}. \tag{2}$$

Фактор, который с точки зрения специалистов, оказывает наибольшее влияние на изучаемый показатель, имеет наименьшую сумму рангов, а фактор, оказывающий самое слабое влияние, — наибольшую.

Для того чтобы полнее использовать информацию, содержащуюся в анкетах, все исследование можно разбить на ряд этапов. На каждом этапе ставится и проверяется определенная гипотеза, положительный или отрицательный ответ на которую либо означает переход к следующему этапу исследования, либо показывает, что продолжение его невозможно. Во втором случае в качестве результатов исследования выступают результаты, полученные на уже проведенных этапах. Такой многоэтапный подход помогает максимально использовать возможности статистики, поскольку он позволяет проанализировать различные стороны полученной информации. Сами гипотезы строятся по принципу однообразности, т.е. положительный (отрицательный) ответ на любую из них всегда означает продолжение (окончание) исследования.

Гипотезы располагаются в порядке убывания их мощности, т.е. таким образом, что при окончании исследования на этапе ν проведение этапов ($\nu+1$), ($\nu+2$) ... (или проверка гипотез $H_0^{\nu+1}$, $H_0^{\nu+2}$) не дает никакой новой информации, не содержащейся уже в результатах этапов 1, 2,..., ν (ν – номер этапа, $\nu=1, 2,..., N$). Выполнение этих условий особенно важно при проверке гипотез на ЭВМ.

Для каждого конкретного исследования сами гипотезы и их порядок также конкретны [4; 83, 84]:

1. Гипотеза H_0^1: начальная таблица рангов неадекватна таблице переформированных рангов.

Эта гипотеза проверяется в связи с тем, что дальнейшая работа (кроме гипотез H_0^3 и H_0^4) проводится на переформированных рангах. Поэтому необходима их адекватность. В случае подтверждения данной гипотезы необходимо произвести анкетирование снова (H_0^1— самая мощная гипотеза). Причины неадекватности: а) неоднозначное понимание специалистами каждого фактора и б) недостаточно высокая квалификация специалистов.

Если гипотеза H_0^1 отвергается, переходят к проверке гипотезы H_0^2.

2. Гипотеза H_0^2: нет согласованности во мнениях специалистов. Гипотеза H_0^2 — менее мощная, чем H_0^1, так как используются переформированные ранги. Если ее рассматривать первой, то она будет проверяться даже в условиях, когда таблицы с начальными и переформированными рангами неадекватны.

Если гипотеза H_0^2 подтверждается, то производится новое анкетирование и переходят к проверке гипотезы H_0^1. Причины положительного ответа на гипотезу H_0^2: а) анкетирование проведено неквалифицированно; б) исследуемый процесс недостаточно изучен.

Если гипотеза H_0^2 отвергается, переходят к гипотезе H_0^3.

3. Гипотеза H_0^3: различие в оценках опрошенных специалистов по вопросу о степени влияния отобранных факторов на изучаемый процесс существенно.

Гипотеза H_0^3 проверяется в развитии гипотезы H_0^2. Если гипотеза H_0^3 подтверждается, то необходимо заново произвести анкетирование и перейти к проверке гипотезы H_0^1. Причины положительного ответа на гипотезу H_0^3 те же, что и на гипотезу H_0^2.

Если гипотеза H_0^3 отвергается, переходят к проверке гипотезы H_0^4.

Гипотеза H_0^4: различие во влиянии исследуемых факторов на изучаемый процесс несущественно (по мнению опрошенных специалистов).

Если гипотеза подтверждается, то необходимо расширить список факторов и вторично произвести анкетирование. После этого переходят к гипотезе H_0^1. Причины подтверждения гипотезы H_0^4 те же, что и гипотезы H_0^2.

Если гипотеза H_0^4 отвергается, переходят к гипотезе H_0^5.

5. Гипотеза H_0^5: нет определенной структуры влияния факторов.

Здесь под структурой понимается наличие таких влияний факторов, из которых: а) хотя бы одно не равно нулю; б) хотя бы одно отлично от других при проверке по некоторому критерию.

Если гипотеза отвергается, переходят к гипотезе H_0^6.

6. Гипотеза H_0^6: влияние всех или части факторов подчиняется равномерному распределению.

Если гипотеза верна, то по этим факторам необходимы дальнейшие исследования. Если гипотеза отвергается, то анализ априорной информации считается законченным.

Каждая из гипотез проверяется соответствующими статистическими методами.

Коэффициент ранговой корреляции Спирмена

В горном деле при проведении буровых работ нередко возникает необходимость определить степень тесноты связи не только между параметрами, оцениваемыми количественно, но и между явлениями, для которых или вообще нет численных оценок (исходная информация выражается с помощью качественных признаков: влияет — не влияет, брак — годная и т.д.), или такие оценки связаны со значительными трудностями.

Для решения подобных задач может быть использован коэффициент ранговой корреляции Спирмена (в отдельных источниках пишется «Спирмэна») [5–7]. Сущность метода Спирмена состоит в следующем:

– располагают варианты факторного признака (*x*) по возрастанию;

– проставляют ранги для вариантов результативного признака (*y*).

Если связь между признаками прямая, то, наряду с увеличением ранга признака *x*, ранг признака *y* также будет правильно возрастать, и номера рангов признаков *x* и *y* совпадут. При обратной связи возрастанию рангов признака *x* будет соответствовать убывание рангов признака *y*. В случае отсутствия связи ранг признака *y* не будет обнаруживать никакого порядка возрастания или убывания.

Степень тесноты связи между признаками определяется ранговым коэффициентом корреляции (адекватность перехода от рангов таблицы 2 к переформированным рангам проверяется по коэффициенту ранговой корреляции r_s Спирмена):

$$r_s = 1 - \frac{6\sum d^2}{n\left(n^2 - 1\right)}, \tag{3}$$

где d^2 — квадрат разности рангов; n — количество фиксированных наблюдений.

Величина рангового коэффициента корреляции Спирмена находится в границах от -1 до +1. Когда связь отсутствует, ранговый коэффициент корреляции равен нулю (r_s=0). При прямой зависимости коэффициент корреляции (r_s) положительный ($0 \le r_s \le 1$), а в случае обратной связи — отрицательный ($-1 \le r_s \angle 0$).

Для примера вычислим коэффициент ранговой корреляции Спирмена между порядковым номером лактации и надоем молока от коровы в сельскохозяйственном предприятии. Средний уровень годовой продуктивности животных по нечетным лактациям показан в таблице 3.

Таблица 3

Средний надой молока от коровы по нечетным лактациям

Номер лактации	Средняя продуктивность коров, кг	Ранги		Разности рангов	Квадраты разности рангов
		лактаций	продуктивности		
x	y	R_x	R_y	$d = R_x - R_y$	$d^2 = \left(R_x - R_y\right)^2$
1	2541	1	4	-3	9
3	3878	2	7	-5	25
5	3609	3	6	-3	9
7	2874	4	5	-1	1
9	2012	5	3	2	4
11	1737	6	2	4	16
13	1564	7	1	6	36
$\sum 49$	18215	28	28	0	100

В соответствии с формулой (3) рассчитываем коэффициент ранговой корреляции Спирмена

$$r_s = 1 - \frac{6 \cdot 100}{7\left(49 - 1\right)} = -0,7857,$$

т.е. между рангами рассматриваемых признаков имеет место высокая обратная зависимость, что свидетельствует о слабости кормовой базы и несвоевременности выбраковки коров.

Определение значимости коэффициента ранговой корреляции Спирмена

С увеличением объема выборки r_s стремится к p_s (p_s — генеральный коэффициент ранговой корреляции).

Проверка гипотезы H_0: $p_s = 0$ при H: $p_s \ne 0$ осуществляется путем сравнения r_s, вычисленного по формуле (3), с критическим значением, взятым из таблицы 4 для выбранного уровня значимости α и числа пар наблюдений n.

Таблица 4

Критические значения коэффициента ранговой корреляции Спирмена

n	Уровень значимости*		n	Уровень значимости*		n	Уровень значимости*		n	Уровень значимости*	
	$\frac{0,1}{0,05}$	$\frac{0,05}{0,025}$		$\frac{0,1}{0,05}$	$\frac{0,05}{0,025}$		$\frac{0,1}{0,05}$	$\frac{0,05}{0,025}$		$\frac{0,1}{0,05}$	$\frac{0,05}{0,025}$
5	0,800	0,900	12	0,496	0,580	19	0,389	0,458	26	0,330	0,389
6	0,771	0,829	13	0,478	0,555	20	0,379	0,445	27	0,324	0,383
7	0,679	0,745	14	0,459	0,534	21	0,369	0,435	28	0,317	0,375
8	0,595	0,690	15	0,443	0,518	22	0,360	0,424	29	0,311	0,368
9	0,583	0,663	16	0,427	0,500	23	0,352	0,415	30	0,306	0,362
10	0,552	0,636	17	0,412	0,485	24	0,344	0,406			
11	0,527	0,609	18	0,399	0,472	25	0,336	0,398			

* В числителе — для двустороннего критерия, в знаменателе — для одностороннего.

Примечание. Данные работы [8; 95].

Если неравенство

$$\left|r_s\right| \le r_{s(1-\alpha/2)} \tag{4}$$

выполняется, то гипотезу H_0 : p_s =0 не отвергают (в противном случае нет оснований для принятия гипотезы).

При H_1 : p_s >0 или H_1 : $p_s \angle 0$ используют односторонний критерий.

При n>30 гипотезу H_0 проверяют с помощью t-распределения Стъюдента с k=n-2 степенями свободы.

Значение статистики

$$t = \frac{r_s}{\sqrt{1-r_s^2}}\sqrt{n-2} \tag{5}$$

сравнивают с критическим значением $t_{\alpha;k}$, взятым из таблицы 5.

Гипотезу H_0 : p_s =0 принимают, если $\left|t\right| \le t_{\alpha;k}$ (при H_1 : $p_s \ne 0$ используют двусторонний критерий, при H_1 : p_s >0 или H_1 : $p_s \angle 0$ — односторонний).

В целом коэффициент ранговой корреляции Спирмена целесообразно применять в следующих случаях [7; 372]:

Когда необходимо быстро получить приближенную оценку коэффициента корреляции, а точный расчет очень громоздок.

Когда нужно перепроверить согласование решений двух судей относительно рангового упорядочения объектов, например, на конкурсе красоты. С помощью этого критерия можно также проверять измененные способности (сравнение выбранного упорядочения предметов со стандартным упорядочением). Примером может служить расположение детьми кубиков по величине.

Когда имеется подозрение на монотонный тренд: проверяют на значимость коэффициент корреляции между n значениями рангов ряда измерения и рядом натуральных чисел от 1 до n.

Применение в экономических исследованиях рангового коэффициента корреляции Кендэла

В исследованиях находит применение и другой показатель тесноты связи — ранговый коэффициент корреляции Кендэла (τ). Для его расчета предложена следующая формула [9; 280]:

$$\tau = \frac{2S}{n(n-1)} = \frac{2(P+Q)}{n(n-1)}, \tag{6}$$

где P — величина, исчисленная для каждого ранга зависимого признака, как число последующих рангов, больших взятого ранга; Q — отрицательное количество последующих рангов, меньших каждого взятого ранга зависимого признака; n — число наблюдений в исследовании; S — сумма баллов, если баллом +1 оценивается пара рангов, имеющих по обоим признакам одинаковый порядок, а баллом -1 — пара рангов с обратным порядком.

Упрощение расчетов показателя Кендэла достигается следующим образом [10; 228, 229]:

– ряд рангов по признаку x располагаем в возрастающем порядке с указанием соответствующих им рангов по признаку y; получаем ряд рангов по признаку x — 1, 2, 3,..., n и ряд рангов по признаку $y - i_1, i_2, i_3, ..., i_n$;
– подсчитываем баллы для всех рангов по признаку y. Для этого находим, сколько рангов, предшествующих каждому рангу и последующих за ним, превышают его величину. Число предшествующих превышений записываем со знаком минус, а число последующих превышений — со знаком плюс;
– находим сумму положительных и отрицательных баллов по каждому рангу и итоговое число баллов (S);
– далее действуем по указанной формуле (6).

При достаточно больших n между значениями r_s и τ фиксируется определенное соотношение

$$\frac{\tau}{r_s} \approx \frac{2}{3}. \tag{7}$$

Вычислим коэффициент ранговой корреляции Кендэла по данным таблицы 3. Прежде всего определяем величины P и Q. По столбцу R_y находим число предыдущих рангов, которые больше 4. Оно равно 3, это ранги 7, 6, 5. Количество рангов, превышающих 7, равно нулю и т.д. Поэтому величина P составляет

$$P = 3 + 0 + 0 + 0 + 0 + 0 = 3. \tag{8}$$

Таблица 5

Квантили распределения Стьюдента $t_{\alpha;k}$

Число степеней свободы k	Уровень значимости α						
	0,20	0,10	0,05	0,02	0,01	0,005	0,001
1	3,08	6,31	12,71	31,82	63,66	127,32	636,62
2	1,89	2,92	4,30	6,97	9,93	14,09	31,60
3	1,64	2,35	3,18	4,54	5,84	7,45	12,94
4	1,53	2,13	2,78	3,75	4,60	5,60	8,61
5	1,48	2,02	2,57	3,37	4,03	4,77	6,86
6	1,44	1,94	2,45	3,14	3,71	4,32	5,96
7	1,42	1,90	2,37	3,00	3,50	4,03	5,41
8	1,40	1,86	2,31	2,90	3,36	3,83	5,04
9	1,38	1,83	2,26	2,82	3,25	3,69	4,78
10	1,37	1,81	2,23	2,76	3,17	3,58	4,59
11	1,36	1,80	2,20	2,72	3,11	3,50	4,44
12	1,36	1,78	2,18	2,68	3,06	3,43	4,32
13	1,35	1,77	2,16	2,65	3,01	3,37	4,22
14	1,34	1,76	2,15	2,62	2,98	3,33	4,14
15	1,34	1,75	2,13	2,60	2,95	3,29	4,07
16	1,34	1,75	2,12	2,58	2,92	3,25	4,02
17	1,33	1,74	2,11	2,57	2,90	3,22	3,97
18	1,33	1,73	2,10	2,55	2,88	3,20	3,92
19	1,33	1,73	2,09	2,54	2,86	3,17	3,88
20	1,33	1,73	2,09	2,53	2,85	3,15	3,85
21	1,32	1,72	2,08	2,52	2,83	3,14	3,82
22	1,32	1,72	2,07	2,51	2,82	3,12	3,79
23	1,32	1,71	2,07	2,50	2,81	3,10	3,77
24	1,32	1,71	2,06	2,49	2,80	3,09	3,75
25	1,32	1,71	2,06	2,48	2,79	3,08	3,73
26	1,32	1,71	2,06	2,48	2,78	3,07	3,71
27	1,31	1,70	2,05	2,47	2,77	3,06	3,69
28	1,31	1,70	2,05	2,47	2,76	3,05	3,67
29	1,31	1,70	2,04	2,46	2,76	3,04	3,66
30	1,31	1,70	2,04	2,46	2,75	3,03	3,65
40	1,30	1,68	2,02	2,42	2,70	2,97	3,55
60	1,30	1,67	2,00	2,39	2,66	2,91	3,46
120	1,29	1,66	1,98	2,36	2,62	2,86	3,37
∞	1,28	1,64	1,96	2,33	2,58	2,81	3,29

Примечание. Данные работы [8; 205].

Отметим, что количество слагаемых равенства (8) составляет $n-1$, поскольку первый ранг столбца исключается. Это примечание касается и величины Q.

Затем определяют число последующих рангов столбца R_y, меньших данного ранга. Для ранга 4 количество последующих рангов, которые меньше 4, равно 3. Это ранги 3, 2, 1. Ранг 7 имеет 5 последующих меньших рангов (6, 5, 3, 2, 1). Таким образом подсчет ведется до конца. В результате имеем

$$Q = -3 - 5 - 4 - 3 - 2 - 1 = -18. \tag{9}$$

Поэтому ранговый коэффициент корреляции Кендэла в рассматриваемом примере составляет

$$\tau = \frac{2(3-18)}{7(7-1)} = -0{,}7143.$$

Выводы, вытекающие из полученных результатов, аналогичны тем, что приведены выше относительно рангового коэффициента Спирмена.

В итоге подчеркнем, что особенностью ранговых коэффициентов корреляции является то, что их расчет может выполняться в условиях отсутствия достоверной исходной информации, но при известных рангах изучаемых признаков. Разумеется, ранговые коэффициенты корреляции не могут совпадать с коэффициентами корреляции действительных уровней изучаемых признаков. Так, в рассматриваемом примере коэффициент парной линейной корреляции между порядковым номером лактации и продуктивностью коров составляет

$$r = \frac{7 \cdot 109885 - 49 \cdot 18215}{\sqrt{\left(7 \cdot 455 - 49^2\right)\left(7 \cdot 52291731 - 18215^2\right)}} = -0,7526.$$

Как видно, полученный коэффициент парной линейной корреляции несколько больше рангового коэффициента корреляции Спирмена и меньше такого же коэффициента Кендэла. Но, несмотря на это, ранговые коэффициенты корреляции имеют большое практическое значение, поскольку они позволяют в сложных условиях выполнять корреляционный анализ, значительно сокращая объем вычислительной работы и время получения необходимых результатов.

Метод формализации априорной информации

Этот метод имеет особо важное значение при исследовании влияния различных факторов на качественный показатель технологического процесса бурения. Сущность метода — сбор и анализ данных об объекте исследования до проведения экспериментальных работ, составление примерного перечня факторов, которые, по мнению исследователя, могут влиять на качественный показатель технологического процесса. Затем предлагается нескольким специалистам данной области провести ранжирование этих факторов, т.е. расстановку в ряд в порядке убывания степени их важности.

Вклад каждого фактора оценивается по величине ранга — места, которое отведено данному фактору исследователем (экспертом) при опросе. Наиболее важному, по мнению каждого эксперта, факторному признаку присваивается ранг 1, следующему по важности ранг 2 и т.д. Если эксперт не может при ранжировании отдать предпочтение какому-нибудь одному фактору (например 2 и 3), то каждому из факторов присваивается связанный ранг, представляющий среднюю из соответствующих рангов (например (2+3)/2=2,5). Каждый специалист может по своему усмотрению к имеющимся факторам добавлять новые с учетом их при ранжировке.

Полученные результаты мнений m экспертов о рангах факторных признаков сводятся в таблицу (табл. 6).

Таблица 6

Форма заполнения результатов мнений экспертов о рангах факторных признаков

Факторный признак	Эксперт					Сумма
	1	2	3	…	m	
x_1	a_{11}	a_{12}	a_{13}	…	a_{1m}	$\sum a_{1j}$
x_2	a_{21}	a_{22}	a_{23}	…	a_{2m}	$\sum a_{2j}$
…	…	…	…	…	…	…
x_k	a_{k1}	a_{k2}	a_{km}	…	a_{km}	$\sum a_{kj}$

Примечание. Данные работы [8; 96].

Результаты опроса обрабатываются следующим образом. Для каждого фактора x_i определяется сумма рангов, выставленных всеми специалистами ($\overset{\wedge}{\sum} a_{1j}$ — сумма рангов первого факторного признака и т.д.). Находят общую сумму оценок для всех факторов, т.е. $\sum_{i=1}^{k}\sum_{j=1}^{m} a_{ij}$, общая сумма оценок делится на число факторных признаков

$$T = 1/k \sum_{i=1}^{k}\sum_{j=1}^{m} a_{ij}, \tag{10}$$

где T — средняя сумма рангов.

Затем находят сумму квадратов отклонений сумм рангов от их средней

$$s = \sum \left(a_{ij} - T\right)^2. \tag{11}$$

Степень согласованности мнений специалистов о влиянии различных факторов на величину результативного признака оценивается коэффициентом конкордации

$$W_k = s \Big/ \frac{1}{12}\left[m^2\left(k^3 - k\right) - m\sum_{j=1}^{m} T_j\right], \tag{12}$$

где T_j — показатель связанности рангов.

$$T_j = \sum_{\nu=1}^{\nu_i}\left(t^3 V_j - tV_j\right), \tag{13}$$

tV_j — число одинаковых рангов в j-ом ранжировании.

Т а б л и ц а 7

χ^2 – распределение

Степень свободы ν	Уровень значимости α					
	0,20	0,10	0,05	0,02	0,01	0,001
1	1,642	2,706	3,841	5,412	6,635	10,827
2	3,219	4,605	5,991	7,824	9,210	13,815
3	4,642	6,251	7,815	9,837	11,345	16,266
4	5,989	7,779	9,488	11,668	13,277	18,467
5	7,289	9,236	11,070	13,388	15,086	20,515
6	8,558	10,645	12,592	15,033	16,812	22,457
7	9,803	12,017	14,067	16,622	18,475	24,322
8	11,030	13,362	15,507	18,168	20,090	26,125
9	12,242	14,684	16,919	19,679	21,666	27,877
10	13,442	15,987	18,307	21,161	23,209	29,588
11	14,631	17,275	19,675	22,618	24,725	31,264
12	15,812	18,549	21,026	24,054	26,217	32,909
13	16,985	19,812	22,362	25,472	27,688	34,528
14	18,151	21,064	23,685	26,783	29,141	36,123
15	19,311	22,307	24,996	28,259	30,578	37,697
16	20,465	23,542	26,296	29,633	32,000	39,252
17	21,615	24,769	27,587	30,995	33,409	40,790
18	22,760	25,989	28,869	32,346	34,805	42,312
19	23,900	27,204	30,144	33,687	36,191	43,820
20	25,038	28,412	31,410	35,020	37,566	45,315
21	26,171	29,615	32,671	36,343	38,932	46,797
22	27,301	30,813	33,924	37,659	40,289	48,268
23	28,429	32,007	35,172	38,968	41,638	49,728
24	29,553	33,196	36,415	40,270	42,980	51,179
25	30,675	34,382	37,652	41,566	44,314	52,620
26	31,795	35,563	38,885	42,856	45,642	54,052
27	32,912	36,741	40,113	44,140	46,963	55,476
28	34,027	37,916	41,337	45,419	48,278	56,893
29	35,139	39,087	42,557	46,693	49,588	58,302
30	36,250	40,256	43,773	47,962	50,892	59,703

Примечание. Данные работы [8; 207, 208].

Величина W_k колеблется в пределах от нуля, соответствующего полной несогласованности мнений специалистов, до единицы, указывающей на полную согласованность мнений, т.е. принимает значения в интервале $0 \le W_k \le 1$.

Существенность коэффициента конкордации оценивается критерием χ^2

$$\chi^2 = s \Big/ \frac{1}{12}\left[mk\left(k+1\right) - \frac{1}{k-1}\sum T_j\right]. \tag{14}$$

Фактическое значение сравнивается с табличным (см. табл. 7).

При $\chi^2 \rangle \chi^2_{табл}$ коэффициент W_k существен (т.е. значим) и согласованность мнений специалистов высокая.

Воспользовавшись методикой применения коэффициента конкордации, можно также отобрать наиболее значимые факторы и отбросить наименее влияющие, что уменьшит трудоемкость исследований и создаст условия для получения достоверных результатов.

Это можно проследить на следующем примере, взятого из практики горной промышленности.

Для выяснения причин обрыва секторов матриц алмазных коронок был проведен опрос трех специалистов (m = 3). Опрос проводился с помощью анкеты, содержащей три фактора (k = 3): работа коронкой с матрицей, не соответствующей твердости пород (x_1); работа при сильных вибрациях (x_2); некачественное спекание матрицы с корпусом коронки (x_3). Результаты опроса сведены в таблицу 8. Сравним мнения специалистов о важности трех факторных признаков.

Таблица 8

Расчетная таблица

Факторный признак	Ранг, установленный специалистами			Сумма рангов по каждому факторному признаку	Отклонение суммы рангов от средней суммы	Квадрат отклонения
	первым	вторым	третьим			
x_1	1	1	2	4	-2	4
x_2	2,5	2	3	7,5	1,5	2,25
x_3	2,5	3	1	6,5	0,5	0,25
Итого	6	6	6	18	0	6,5
В среднем	2	2	2	6	-	-

По данным этой таблицы, определим T_j.

Число связанных рангов первого специалиста составляет y = 2, второго и третьего y = 0. Тогда

$$T_j = 2^3 - 2 = 6.$$

По формуле (12) находим

$$W_k = \frac{6,5}{\frac{1}{12}\left[3^2\left(3^2-3\right)-6\right]} = 0,3.$$

Найденное значение отличается от нуля, поэтому можно считать, что имеется достаточная степень согласованности.

Оцениваем существенность W_k согласно выражению (14)

$$\chi^2 = \frac{6,5}{\frac{1}{2}3\cdot 3\left(3+1\right)-\frac{6}{2}} = 2,36.$$

Из таблицы 7 находим, что для 5 %-ного уровня значимости при ν =3–1=2 (две степени свободы) $\chi^2_{табл} = 5,991$. Поскольку $\chi^2 \langle \chi^2_{табл}$, необходимо для уточнения последующих выводов привлечь большее число специалистов и некоторые дополнительные факторные признаки.

Альтернативные признаки в экономике

В экономических науках признаки подразделяют на качественные и количественные. Качественные — это признаки, отдельные значения которых отличаются друг от друга существенными свойствами. Количественными именуют признаки, отдельные значения которых отличаются друг от друга только величиной. Количественные признаки выражаются числами.

Качественные признаки также именуются атрибутивными (лат. *attribuo* — придаю, наделяю). Атрибут — неотъемлемое качественное свойство предмета, без которого он не существует. В экономических явлениях атрибутивные признаки широко распространены и поэтому имеют большую практическую значимость. В этой связи их необходимо, как и количественные признаки, тщательно анализировать, прогнозировать, планировать и исследовать [9; 282].

В большой совокупности атрибутивных (не имеющих количественного выражения) признаков особое место принадлежит альтернативным признакам. Альтернативным называют такой признак, который может иметь данная единица совокупности, а может и не иметь, т.е. альтернативный признак может характеризоваться только двумя следующими значениями: или да, или нет. Альтернатива (лат. *alter*, фр. *alternative* — один из двух) переводится как попеременный. Примерами альтернативных признаков могут быть рентабельные и убыточные горнодобывающие предприятия, высококачественная и бракованная продукция и т.д.

Учитывая большую значимость альтернативных признаков в экономических процессах, им следует уделять надлежащее внимание как в анализе, так и в планировании горного производства. Для этого целесообразно составлять ряды динамики и вариации различных альтернативных признаков и определять их основные аналитические характеристики.

Если обозначить наличие альтернативного признака у единицы совокупности через 1, а его отсутствие — через 0, то вариация этого признака p в совокупности составляет

$$p = M / N, \tag{15}$$

где M — число единиц совокупности, обладающих данным признаком; N — общее количество учтенных единиц совокупности, т.е. вариация альтернативного признака (p) представляет собой долю единиц, обладающих определенным свойством, во всей их совокупности.

Поскольку альтернативный признак принимает всего два значения (1 и 0) с весами соответственно p и q, то его дисперсия

$$\sigma^2 = \frac{(1-p)^2 p + (0-p)^2 q}{p+q} = \frac{q^2 p + p^2 q}{p+q} = pq, \tag{16}$$

где q — доля единиц, которые не имеют данного признака, причем $q + p = 1$.

Но особенно важно, изучая альтернативные признаки в экономических процессах, использовать корреляционный анализ. На основе результатов последнего вскрываются корреляционные связи и зависимости между различными альтернативными признаками.

Тетрахорические показатели альтернативных зависимостей

Корреляционные связи, существующие между различными альтернативными признаками, именуются альтернативными зависимостями (корреляцией альтернативных признаков). При изучении корреляции между альтернативными признаками обычно используют тетрахорические показатели. Для их расчета подготавливают специальную четырехпольную (четырехклеточную) таблицу. Последняя представляет собой группировочную таблицу, которая позволяет вычислить тесноту связи между двумя альтернативными признаками [9; 283].

Для примера приведем таблицу распределения (в общем виде) рабочих горно-обогатительной фабрики (ГОФ) по полу и состоянию здоровья (табл. 9).

Расчетная таблица содержит четыре поля, в которых показывают следующую исходную информацию: a, b, c, d. Каждое из полей соответствует определенной альтернативе первого и второго признаков.

На основе таким образом подготовленных таблиц ведется расчет тетрахорических показателей тесноты связи. Из этих показателей получили следующие коэффициенты [10]:

– эмпирический коэффициент корреляционной связи (коэффициент Фехнера);

– коэффициенты взаимной сопряженности Пирсона и Чупрова;

– коэффициент контингенции;

– коэффициент ассоциации.

Таблица 9

Распределение (в общем виде) рабочих предприятий по полу и состоянию здоровья

Признак *A* / Признак *B*		Число рабочих по полу		Всего
		женщины	мужчины	
Число рабочих по состоянию здоровья	Здоровье	a	b	$a + b$
	Нуждающиеся в лечении	c	d	$c + d$
Всего		$a + c$	$b + d$	$a + b + c + d$

Примечание. Данные работы [9; 284].

Эмпирический коэффициент корреляционной связи (коэффициент Фехнера, $K_э$) измеряет тесноту связи по следующей формуле:

$$K_э = \frac{\sum C - \sum H}{\sum C + \sum H}, \tag{17}$$

где C и H — число совпадений и несовпадений знаков отклонений значений признаков факторного и результативного от своих средних, т.е. $\bar{x}$ и $\bar{y}$.

При этом фиксируется совпадение и несовпадение знаков в отклонениях от средней y различных пар значений признаков.

$K_э$ изменяется в пределах от - 1 до + 1. Если связь между признаками обратная, то $K_э$ отрицателен, в случае прямой связи — положителен. Чем ближе $K_к$ ± 1, тем связь более тесная.

Коэффициенты взаимной сопряженности Пирсона (C) и Чупрова (K) используются в качестве показателя тесноты связи качественных признаков x_1 и x_2 и имеют вид:
коэффициент Пирсона

$$C = \sqrt{\frac{\varphi^2}{1+\varphi^2}}; \tag{18}$$

коэффициент Чупрова

$$K = \frac{\varphi^2}{\sqrt{(k_1 - 1)(k_2 - 1)}}, \tag{19}$$

где

$$\varphi^2 = \sum_{j_1=1}^{k_1} \sum_{j_2=1}^{k_2} \frac{n_{j_1/j_2}}{n_{j_1} \cdot n_{j_2}},$$

k_1 и k_2 — число групп по каждому из признаков.

Коэффициент контингенции (показатель сходства, $K_к$) — показатель, используемый для изучения зависимости между альтернативными признаками на основе таблицы четырех полей:

$$K_k = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}, \tag{20}$$

где a, b, c, d — исходные данные четырехпольной таблицы.

Контингенция, контингент (*лат.* contingens — достигающийся на долю) представляет собой совокупность людей, предметов или явлений, которые составляют однородную в каком-то отношении группу или категорию [9; 285].

Коэффициент контингенции находится в пределах от –1 до +1. В случае обратной функциональной связи этот коэффициент равен –1. Судя по формуле (20), $K_к$ = –1 тогда, когда $a = d = 0$. Если коэффициент контингенции составляет +1, то изучаемая связь является прямой функциональной. Такая зависимость имеет место тогда, когда в четырехпольной таблице $b = c = 0$.

Сравнительно простым является способ вычисления коэффициента ассоциации, который обосновал Юл. Ассоциация — это связь между отдельными представлениями, при которой одно из представлений вызывает другое. Например, ассоциация по сходству, по расположению. Коэффициент ассоциации (K_a), вскрывающий наличие ассоциативной связи между двумя признаками, определяется формулой

$$K_a = \frac{ad - bc}{ad + bc}. \tag{21}$$

Величина этого коэффициента варьируется от –1 до +1. $K_a = -1$, если $ad = 0$; $K_a = +1$, когда $bc = 0$. Вся последующая характеристика коэффициента ассоциации является такой же, как и коэффициента контингенции.

Для корреляционного анализа альтернативных признаков Юл также предложил другой показатель, который получил название «коэффициент коллигации» (K_{KO}). Его уровень определяется следующим равенством:

$$K_{KO} = \frac{\sqrt{ad} - \sqrt{bc}}{\sqrt{ad} + \sqrt{bc}}. \tag{22}$$

Отметив сходство формул Юла (21) и (22), подчеркнем, что абсолютные значения коэффициента коллигации намного меньше по сравнению с коэффициентом ассоциации. Подчеркнем также, что при использовании в расчетах формулы (22) находят только арифметические корни.

Рассмотрим примеры вычисления показателей тесноты связи альтернативных признаков. Для этого приведем данные, характеризующие надежность эксплуатации горно-шахтного оборудования на угольном предприятии в зависимости от качества их ремонта (табл. 10).

Коэффициент контингенции в соответствии с формулой (20) и данными таблицы 10 составляет

$$K_k = \frac{40 \cdot 10 - 7 \cdot 3}{\sqrt{(40+7)(3+10)(40+3)(7+10)}} = 0,5671.$$

Таблица 10

Влияние качества ремонта горно-шахтного оборудования на угледобывающем предприятии на надежность его эксплуатации

Качество ремонта	Эксплуатация		Всего
	надежная	ненадежная	
Высокое	40	7	47
Низкое	3	10	13
Всего	43	17	60

По формуле (21) коэффициент ассоциации Юла

$$K_a = \frac{40 \cdot 10 - 7 \cdot 3}{40 \cdot 10 + 7 \cdot 3} = 0,9002.$$

Обращает на себя внимание значительное расхождение в уровнях исчисленных коэффициентов. Если базовым считать коэффициент контингенции, то коэффициент ассоциации Юла больше первого на 58,7 %.

Коэффициент коллигации Юла, вычисляемый по формуле (22), составляет

$$K_{KO} = \frac{\sqrt{40 \cdot 10} - \sqrt{7 \cdot 3}}{\sqrt{40 \cdot 10} + \sqrt{7 \cdot 3}} = 0,6272.$$

Как видно, этот коэффициент намного меньше коэффициента ассоциации (на 30,3 %).

На основе полученных результатов можно сделать заключение о том, что коэффициент контингенции и коэффициент коллигации Юла характеризуются определенным сходством своих уровней. Кстати, это сходство подтверждает и опыт корреляционного анализа альтернативных признаков. Уровень же коэффициента ассоциации Юла значительно превосходит другие показатели. Поэтому с какой-то вероятностью можно считать, что между надежностью эксплуатации горно-шахтного оборудования и качеством его ремонта существует корреляционная связь, теснота которой является средней.

В целом же проблема ранговой корреляции и корреляции альтернативных признаков не является до конца решенной. В идеальном варианте должны быть установлены единые и объективные показатели тесноты связи, проблематичным остается расчет параметров функций, их ошибок, критериев согласия и т.д.

References

1 *Kendel M.* Rank correlations / Transl. from English. — M.: Statistics, 1974. — P. 214.

2 *Hettmansperger T.* The statistical conclusions based on ranks / Transl. from English. — M.: Finance and statistics, 1987. — P. 334.

3 *Ul J.A., Kendel M. Dzh.* The statistics theory / Transl. from English. — M.: Statestatpubl, 1960. — P. 780.

4 *Rosanov G.V.* Statistical modeling of branch development. — M.: Statistics, 1976.

5 *Richard Thomas*. Quantitative methods of the analysis of economic activities / Transl. from English. — M.: Publishing house Business and Service, 1999. — P. 432.

6 *Pollard J.A.* directory on computing methods of statistics / the Lane with English. — M.: Finance and statistics, 1982. — P. 344.

7 *Sachs L.* Statistical estimation / transl. from German. — M.: Statistics, 1976.

8 *Gandzhumjan R.A.* The mathematical statistics in prospecting drilling: the Handbook. — M.: Bowels, 1990.

9 *Grishin A.F., Kocherova E.V.* Statistical models: construction, estimation, analysis: Tutorial. — M.: Finance and statistics, 2005.

10 *Venetskij I.G., Venetskaia V.I.* The basic mathematic-statistical concepts and formulas of the economic analysis. — M.: Statistics, 1974.

Р.С.Каренов

Экономикалық зерттеулерде рангтік корреляцияны және баламалы белгілер корреляциясын қолдану мәселелері

Экономикалық зерттеулердегі рангтік корреляцияны қолданудың практикалық маңызы көрсетілген. Қазіргі кезде дара факторлар арасындағы рангтік корреляция тәсілдері толығырақ жасалғандығы туралы айтылған. Статистикалық қатарлар нұсқаларын нөмірлеу принципі құбылыстар арасындағы немесе рангтік статистикадағы байланыстарды зерттеудің параметрлік емес тәсілдердің негізі болып табылатынына ерекше көңіл бөлінген. Экономикалық зерттеулерде сандық және сапалық белгілер арасындағы байланыстарды айқындағанда Спирменнің және Кендэлдың рангтік корреляция коэффициенттері кеңінен қолданылатындығы дәлелденген. Экономикалық есептеулерде және негіздеуге конкордация коэффициентін қолдану әдістемесі ұсынылған. Әр түрлі баламалы белгілер арасындағы корреляциялық байланыстарды талдауға едәуір көңіл бөлінген. Корреляциялық байланыстың эмпирикалық коэффициенті, Пирсон және Чупровтың өзара үндесу коэффициенттері, контингенция және ассоциация коэффициенттері арқылы факторлар және нәтижелі белгілер арасындағы корреляциялық байланыс күшін өлшеу әдістемесі ұсынылған.

R.S.Karenov

Use problems ранговой correlations and correlations of alternative signs in economic researches

Practical value ranks correlations in economic researches is allocated. It is noticed that ways one-factorial ranks correlations are now more full developed. It is underlined that the principle of numbering of variants of statistical numbers is a basis of nonparametric methods of studying of communication between the phenomena or ranks statisticans. It is proved that in economic researches the big application at communication studying between quantitative and qualitative signs find ranks factors of correlation of Spirmena and Kendela. The technique of application of factor konkordacii in economic calculations and substantiations Is offered. The considerable attention is given to the analysis of the correlation communications existing between various alternative signs. The technique measurements narrownesses of correlation communication between factorial and productive signs by means of such factors, as empirical factor of correlation communication, factors of a mutual associativity of Pirsona and Tchuprov, factors contingent and associations is recommended.