



ДОНСКОЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
УПРАВЛЕНИЕ ДИСТАНЦИОННОГО ОБУЧЕНИЯ И ПОВЫШЕНИЯ
КВАЛИФИКАЦИИ

Кафедра «Организации перевозок и дорожного движения»

Практикум
по дисциплине

**«Основы научных
исследований»**



Автор
Лазарев Е.Г.

Ростов-на-Дону, 2017

Аннотация

Методические указания предназначены для магистров, обучающихся по направлению подготовки 23.04.01 «Технология транспортных процессов», профессионально-образовательная программа «Транспортная логистика».

В методических указаниях рассматриваются вопросы применения системного подхода к организации научных исследований. Представлены практические методы обоснования объема выборки для проведения статистических исследований. Показаны сферы применения корреляционного анализа в научных исследованиях.

Автор



К.т.н., доцент кафедры «Организации перевозок и дорожного движения»
Лазарев Евгений Георгиевич

Оглавление

1 СИСТЕМНЫЙ ПОДХОД К ОРГАНИЗАЦИИ НАУЧНЫХ ИССЛЕДОВАНИЙ	4
2 ОРГАНИЗАЦИЯ СТАТИСТИЧЕСКИХ НАБЛЮДЕНИЙ В ТРАНСПОРТНЫХ ИССЛЕДОВАНИЯХ	7
3 ПРИМЕНЕНИЕ КОРРЕЛЯЦИОННОГО АНАЛИЗА В НАУЧНЫХ ИССЛЕДОВАНИЯХ	11
3.1 Понятие корреляционной связи.....	11
3.2 Коэффициент корреляции	12
3.3 Основные свойства коэффициентов корреляции	14
3.4 Корреляционные поля	16
3.5 Проведение корреляционного анализа.....	19

1 СИСТЕМНЫЙ ПОДХОД К ОРГАНИЗАЦИИ НАУЧНЫХ ИССЛЕДОВАНИЙ

Характерной особенностью современных транспортных исследований является применение системного подхода к изучаемой проблеме, что позволяет рассматривать процессы не в изолированном виде, а с учетом различных внешних и внутренних факторов. В этом случае предметом исследования становится система объектов. Транспорт обладает всеми признаками сложных вероятностных систем: многомерностью и многофакторностью, взаимосвязанностью различных элементов.

Системный подход как метод научного исследования обеспечивает комплексность, всесторонность и объективность рассмотрения поставленной проблемы. В свою очередь системный анализ – эффективный инструмент исследования сложных систем. Объект исследования (организация, предприятие или процесс) рассматривается как сложная система, состоящая из взаимосвязанных частей подсистем.

Системные исследования широко применяются в различных областях науки, техники и практической деятельности. Системный анализ реализуется в научных исследованиях в виде следующих этапов.

1. Установление границ исследуемой системы, ограничений и показателей функционирования.
2. Формальное (абстрактное) представление содержания, структуры системы, состава элементов и подсистем.
3. Описание внешней среды, изучение связей с системой.
4. Составление алгоритма функционирования системы.
5. Разработка программной модели системы.
6. Реализация модели на ЭВМ, постановка и проведение вычислительных экспериментов.
7. Анализ и оценка результатов экспериментов, выработка решений по совершенствованию и развитию системы.

Наиболее эффективны – имитационные модели исследуемых систем. Такие модели должны отвечать следующим принципам.

1. Быть динамическими, показывать изменение во времени характеристик (входных и выходных переменных) транспортных систем за достаточно большие промежутки времени их работы.
2. Допускать учет влияния случайных факторов на показатели работы транспортных систем.
3. Адекватно воспроизводить технологию работы транс-

портных систем.

4. Оценивать возможные варианты организации и развития технологии работы транспортных систем.

В соответствии с перечисленными принципами технологический процесс работы транспортной системы следует рассматривать как совокупность взаимосвязанных элементов. Требуется формально (абстрактно) представить технологический процесс, установить параметры, входные и выходные переменные, состав элементов, ограничения и показатели функционирования для построения имитационной модели.

Под структурой технологического процесса обычно понимаются физические элементы, связи между ними, и операции, выполняемые в определенной последовательности. Состав элементов транспортной системы может быть однородным (из однотипных элементов), гетерогенным (из разнотипных элементов) и смешанным (большое количество однородных элементов при некотором количестве гетерогенных).

Важный этап системного анализа – установление свойств и связей выделенных элементов. В первую очередь необходимо выявлять системные свойства и связи элементов, наиболее сильно влияющие на показатели работы транспортной системы.

С использованием моделирования удаётся решать в области планирования и управления работой транспорта следующие оптимизационные задачи:

- планирование грузовых перевозок и работы подвижного состава, в том числе специализированного, а также транспортного оборудования;
- выбор потребителей и поставщиков произведенной продукции, сырья, полуфабрикатов, изделий;
- распределение перевозочной работы между видами транспорта;
- рациональное распределение грузопотоков по параллельным направлениям и участкам перевозок;
- организация грузопотоков, формирование транспортных партий, распределение грузовой работы между терминалами и перевалочными пунктами различных видов транспорта;
- обоснование перерабатывающей способности отдельных устройств транспортной инфраструктуры;
- минимизация простоев подвижного состава в пунктах начала и окончания перевозок;
- размещение и использование вместимости складов;
- развитие транспортной сети с целью освоения пер-

спективных объемов перевозок.

Сфера практического применения методов моделирования ограничивается возможностями и эффективностью формализации рассматриваемых проблем и ситуаций, а также состоянием информационного, математического, технического обеспечения используемых моделей.

2 ОРГАНИЗАЦИЯ СТАТИСТИЧЕСКИХ НАБЛЮДЕНИЙ В ТРАНСПОРТНЫХ ИССЛЕДОВАНИЯХ

В транспортных исследованиях рекомендуется применять специальные методы, позволяющие сокращать объемы статистических наблюдений без потери точности оценок и выводов. Такие методы называются выборочными наблюдениями. Основа выборочных наблюдений – взаимосвязь между единичным и общим, между частью и целым, которая имеет место в транспортных технологических процессах. Сбор статистических данных требуется для анализа и моделирования работы транспортных систем.

Исследуемая часть статистической совокупности называется выборкой, а количество единиц, составляющих ее объем принято обозначать n . Вся совокупность называется генеральной, объем генеральной совокупности обычно обозначают N .

Учитываются некоторые причины, требующие применения выборочных наблюдений. Недостаток временных ресурсов для проведения обследования и анализа большого объема данных. Отсутствие квалифицированных специалистов для проведения наблюдений, а в ряде случаев дефицит материальных ресурсов. Практическая невозможность учета всех единиц совокупности наблюдений в связи с их быстрым изменением. Например, колебания транспортных потоков в течение коротких промежутков времени.

Практическая нецелесообразность наблюдения каждой единицы совокупности (перевозки, выполняемые отдельными транспортными средствами).

Выборочное наблюдение (исследование) заключается в обследовании определенного числа единиц совокупности, отобранного, как правило, случайным образом. При выборочном методе обследованию подлежит сравнительно небольшая часть всей изучаемой совокупности (обычно до 5-10%, реже до 15-20%). Отбор единиц из генеральной совокупности производится таким образом, чтобы выборочная совокупность была представительна (репрезентативна) и характеризовала генеральную совокупность. Степень представительности выборки зависит от способа организации выборки и от ее объема. Полной репрезентативности выборки достичь не удастся. Поэтому необходима оценка надежности результатов выборки и возможности их распространения на генеральную совокупность.

Выборочный метод дает достаточно точные результаты,

поэтому он может применяться для проверки данных полного наблюдения. Минимальная, но достаточная численность обследуемых единиц, позволяет провести исследование в сжатые сроки. Например, при анализе транспортных потоков практикуются выборочные контрольные наблюдения для проверки правильности записей сплошного наблюдения.

В основе теории выборочного наблюдения лежат теоремы законов больших чисел, которые позволяют решить два взаимосвязанных вопроса выборки: рассчитать ее объем при заданной точности исследования и определить ошибку при данном объеме выборки. При использовании выборочного метода обычно используются два вида обобщающих показателей: относительная величина альтернативного признака и средняя величина количественного признака.

Относительная величина альтернативного признака характеризует долю (удельный вес) единиц в статистической совокупности, обладающих изучаемым признаком. В генеральной совокупности эта доля единиц называется генеральной долей p , а в выборочной совокупности – выборочной долей w . Средняя величина количественного признака в генеральной совокупности называется генеральной средней \bar{x} , а в выборочной совокупности – выборочной средней \tilde{x} .

Доверительная выборка. Для сбора информации исследователь самостоятельно выбирает доступные источники информации. Например, отдельные транспортные службы, подразделения или должностные лица, владеющие необходимыми статистическими данными и материалами.

Метод основного массива. Применяется при исследовании небольших по объему генеральных совокупностей, когда нецелесообразно проводить выборочное исследование. В этом случае достаточно использовать до 60% источников информации. Случайная выборка. Применяется при исследовании генеральной совокупности, границы которой сложно установить.

Пропорциональная или выборка по квоте. Применяется, если известно распределение генеральной совокупности по некоторым существенным признакам. Генеральная совокупность разделяется на взаимоисключающие группы с заданными характеристиками. Каждая группа имеет определенную долю в генеральной совокупности. Выборочная совокупность формируется в соответствии с установленными пропорциями.

Рассмотренные варианты организации выборки позволяют оптимизировать исследования, исходя из специфики генеральной

совокупности, размеров бюджета, целей исследования. Выбирая тот или иной тип выборки важно учитывать множество факторов. Прежде чем приступить к осуществлению выборочного наблюдения необходимо определить количество единиц выборочной совокупности, обеспечивающее репрезентативность, и, следовательно, надежность результатов исследования. Для этого задаются: степень точности исследования (вероятность); предельная допускаемая ошибка, то есть интервал отклонения, зависящий от целей исследования.

Исходя из этих критериев, рассчитывается необходимая численность выборочной совокупности n на основе формулы предельной ошибки выборки. Как указывалось выше, предельная ошибка выборки определяется для средней величины \bar{x} и для доли w , то, соответственно, возможны два варианта определения необходимой численности выборочной совокупности:

а) для повторного отбора:

$$n = \frac{t^2 \sigma_{\bar{x}}^2}{\Delta^2}; \quad n = \frac{t^2 w(1-w)}{\Delta^2} \quad (1)$$

б) для бесповторного отбора:

$$n = \frac{t^2 \sigma_{\bar{x}}^2 N}{\Delta^2 N + t^2 \sigma_{\bar{x}}^2}; \quad n = \frac{t^2 w(1-w)N}{\Delta^2 N + t^2 w(1-w)} \quad (2)$$

где t – коэффициент, зависящий от уровня доверительной вероятности (при уровне доверительной вероятности 0,90; 0,92; 0,95; 0,98 коэффициент соответственно равен 1,64; 1,75; 1,96; 2,32);

σ – среднее квадратичное отклонение наблюдаемой случайной величины;

N – объем генеральной совокупности наблюдаемой случайной величины.

Все величины формул (1, 2), кроме среднего квадратичного отклонения, задаются. Для сокращения объема вычислений значение среднего квадратичного отклонения рекомендуется прогнозировать, учитывая наличие корреляционных связей с другими случайными величинами.

Для обеспечения случайного характера отбора при формировании выборки необходимо выбрать количество дней статисти-

Основы научных исследований

ческих наблюдений (из 365), соответствующих рассчитанному объему репрезентативной выборки. С этой используются случайные числа в интервале $[0, 1]$. Генерируемый диапазон случайных чисел пропорционально распределяется между календарными днями года. Затем датчиком ЭВМ воспроизводятся случайные числа в количестве равном рассчитанному объему выборки. Полученные таким образом календарные дни группируются по месяцам года. Соответственно формируется схема отбора репрезентативной выборки. Так как она неповторная, то при вторичном выпадении какого-либо дня года, соответствующее случайное число не учитывается, а датчиком случайных чисел ЭВМ генерируется новое.

По полученной схеме отбирается требуемый статистический материал (путевые листы, товарно-транспортные накладные, приемосдаточные ведомости и т.п.) за конкретные календарные дни для составления ретроспективной записи об исследуемом транспортном потоке. В таблице 1 дан пример составления схемы отбора статистических данных о выполненных перевозках при требуемом объеме выборки 32 дня.

Таблица 1 – Схема отбора для формирования репрезентативной выборки

День, месяц	День, месяц	День, месяц	День, месяц
14.01	04.04	28.05	22.09
15.01	05.04	22.06	10.10
12.02	29.04	25.06	12.10
13.02	06.05	14.07	19.10
17.02	08.05	21.07	24.10
19.02	14.05	23.07	18.12
06.03	22.05	06.08	21.12
28.03	23.05	16.09	27.12

3 ПРИМЕНЕНИЕ КОРРЕЛЯЦИОННОГО АНАЛИЗА В НАУЧНЫХ ИССЛЕДОВАНИЯХ

3.1 Понятие корреляционной связи

Корреляционный анализ, как и другие статистические методы, основан на использовании вероятностных моделей, описывающих поведение исследуемых признаков в некоторой генеральной совокупности, из которой получены экспериментальные значения x_i и y_i .

Корреляционные связи – это вероятностные изменения, которые можно изучаются на репрезентативных выборках методами математической статистики. Оба термина – корреляционная связь и корреляционная зависимость – часто используются как синонимы. Зависимость подразумевает влияние, связь – любые согласованные изменения, которые могут объясняться различными причинами. Корреляционные связи свидетельствуют о том, что изменениям одного признака, как правило, сопутствуют определенные изменения другого.

Корреляционная зависимость – это изменения, которые вносят значения одного признака в вероятность появления разных значений другого признака. Задача корреляционного анализа сводится к установлению направления (положительное или отрицательное) и формы (линейная, нелинейная) связи между изменяющимися признаками, оценке тесноты связи, и проверке уровня значимости полученных коэффициентов корреляции.

По форме корреляционная связь прямолинейная или криволинейная. Прямолинейная, например, когда имеется связь между количеством рейсов, выполненных автомобилем, и объемом доставленного груза. Криволинейная – наблюдается между себестоимостью и дальностью автомобильных перевозок.

По направлению корреляционная связь положительная (прямая) и отрицательная (обратная). При положительной прямолинейной корреляции более высоким значениям одного признака соответствуют более высокие значения другого, а более низким значениям одного признака – низкие значения другого (рисунок 3). При отрицательной корреляции соотношения обратные (рисунок 4). При положительной – коэффициент корреляции положительный, при отрицательной – отрицательный.

Исходные данные должны быть качественно и количественно однородны. Качественная однородность предполагает приблизительно одинаковые условия и специфику формирования фак-

торных и результативного признаков. Количественная однородность заключается в отсутствии таких наблюдений, которые аномально отличаются от основного массива данных.

3.2 Коэффициент корреляции

Коэффициенты корреляции является общепринятой в математической статистике характеристикой связи между двумя случайными величинами. Коэффициент корреляции – показатель степени взаимозависимости, статистической связи двух переменных; изменяется в пределах от -1 до +1. Значение коэффициента корреляции 0 указывает на возможное отсутствие зависимости, значение +1 свидетельствует о согласованности переменных.

Различают следующие коэффициенты корреляции:

- дихотомический – показатель связи признаков (переменных) измеряемых по дихотомическим шкалам наименований;
- Пирсона (Pearson product-moment correlation) – коэффициент корреляции, используемый для непрерывных переменных;
- ранговой корреляции Спирмена (Spearman's rank-order correlation) – коэффициент корреляции для переменных, измененных в порядковых (ранговых) шкалах;
- точечно-бисериальной корреляции (point-biserial correlation) – коэффициент корреляции, применяемый в случае анализа отношения переменных, одна из которых измерена в непрерывной шкале, а другая – в строго дихотомической шкале наименований;
- j – коэффициент корреляции, используемый в случае, если обе переменные измерены в дихотомической шкале наименований.
- тетракорический (четырепольный) (tetrachoric) – коэффициент корреляции, используемый в случае, если обе переменные измерены в непрерывных шкалах.

Линейная связь между переменными X_i и X_j оценивается коэффициентом корреляции по формуле (3).

$$p = \frac{M(X_i - mX_i)(X_j - mX_j)}{\sqrt{\sigma_{xi}\sigma_{xj}}} \quad (3)$$

где X_i и X_j – исследуемые переменные;
 mX_i и mX_j – математические ожидания переменных;
 σ_{xi} и σ_{xj} – дисперсии переменных.
 Выборочный коэффициент корреляции определяют по фор-

муле (4)

$$r_{X_j X_i} = \frac{\frac{1}{N} \sum_{u=1}^N (x_{iu} - x_i)(x_{ju} - x_j)}{S_{X_i} S_{X_j}} \quad (4)$$

или по преобразованной формуле (5)

$$r_{X_i X_j} = \frac{\frac{1}{N} \sum_{u=1}^N x_{iu} x_{ju} - x_i x_j}{S_{X_i} S_{X_j}} \quad (5)$$

где $i = 1, 2, \dots, n, j = 1, 2, \dots, m, u = 1, 2, \dots, N$;

N – число опытов(объем выборки);

x_i, x_j – оценки математических ожиданий;

S_{X_i}, S_{X_j} – оценки среднеквадратических отклонений.

Только при совместном нормальном распределении исследуемых случайных величин X_i и X_j коэффициент корреляции имеет определенный смысл связи между переменными. В противном случае коэффициент корреляции может только косвенно характеризовать эту связь.

Используя понятие коэффициента корреляции, матрице экспериментальных данных можно поставить в соответствие квадратную матрицу оценок коэффициентов корреляции (корреляционную матрицу). Матрица записывается следующим образом.

$$Q = \begin{vmatrix} Q_{11} & Q_{12} & \dots & Q_{1m} \\ Q_{21} & Q_{22} & \dots & Q_{2m} \\ \dots & \dots & \dots & \dots \\ Q_{m1} & Q_{m2} & \dots & Q_{mm} \end{vmatrix}$$

Корреляционная матрица обладает следующими свойствами:

- элементы матрицы симметричны относительно главной диагонали, то есть $Q_{jk} = Q_{kj}, j, k = \overline{1, m}$;
- элементы главной диагонали равны единице, $Q_{jj} = 1, j = \overline{1, m}$.

3.3 Основные свойства коэффициентов корреляции

К основным свойствам коэффициента корреляции необходимо отнести следующие:

- коэффициенты корреляции способны характеризовать только линейные связи, т.е. такие, которые выражаются уравнением линейной функции. При наличии нелинейной зависимости между варьирующими признаками следует использовать другие показатели связи;
- значения коэффициентов корреляции – это отвлеченные числа, лежащее в пределах от -1 до $+1$, т.е. $-1 < r < 1$;
- при независимом варьировании признаков, когда связь между ними отсутствует, $r = 0$;
- при положительной, или прямой, связи, когда с увеличением значений одного признака возрастают значения другого, коэффициент корреляции приобретает положительный знак и находится в пределах от 0 до $+1$, т.е. $0 < r < 1$;
- при отрицательной, или обратной, связи, когда с увеличением значений одного признака соответственно уменьшаются значения другого, коэффициент корреляции сопровождается отрицательным знаком и находится в пределах от 0 до -1 , т.е. $-1 < r < 0$;
- чем сильнее связь между признаками, тем ближе величина коэффициента корреляции к 1 . Если $r = \pm 1$, то корреляционная связь переходит в функциональную, т.е. каждому значению признака X будет соответствовать одно или несколько строго определенных значений признака Y ;
- только по величине коэффициентов корреляции нельзя судить о достоверности корреляционной связи между признаками. Этот параметр зависит от числа степеней свободы $f = n - 2$, где n – число коррелируемых пар показателей X и Y . Чем больше n , тем выше достоверность связи при одном и том же значении коэффициента корреляции.

Основы научных исследований

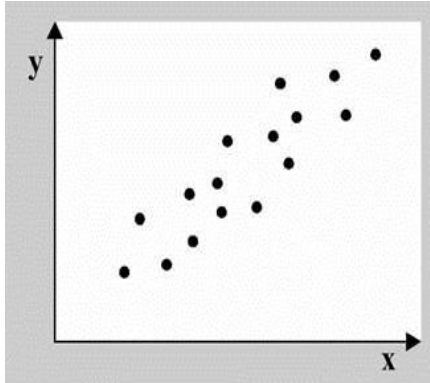


Рисунок 2 – Прямая корреляция

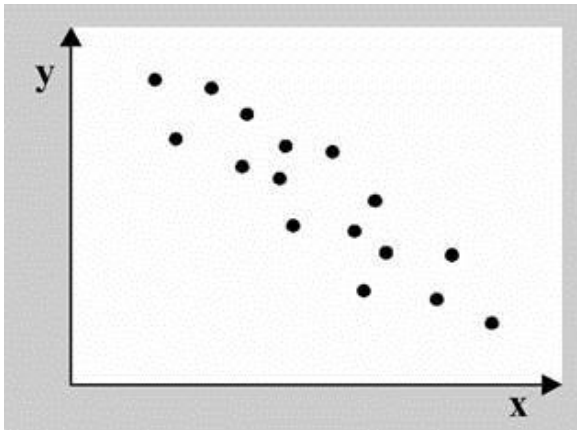


Рисунок 3 – Обратная корреляция

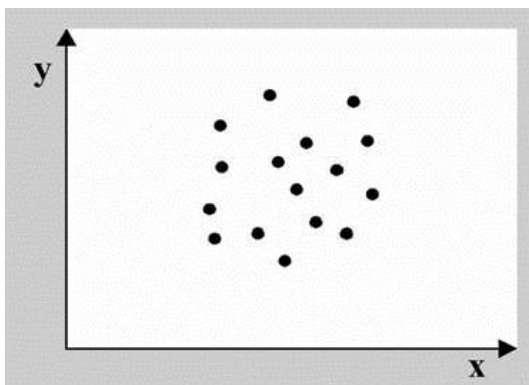


Рисунок 4 – Отсутствие корреляции

Степень, сила или теснота корреляционной связи определяется по величине коэффициента корреляции. Сила связи не зависит от ее направленности и определяется по абсолютному значению коэффициента корреляции.

В зависимости от коэффициента корреляции различают следующие корреляционные связи:

- сильная, или тесная при коэффициенте корреляции $r > 0,70$;
- средняя (при $0,50 < r < 0,69$);
- умеренная (при $0,30 < r < 0,49$);
- слабая (при $0,20 < r < 0,29$);
- очень слабая (при $r < 0,19$).

3.4 Корреляционные поля

Корреляция изучается на основании экспериментальных данных, представляющих собой измеренные значения (x_i , y_i) двух признаков. Если экспериментальных данных немного, то двумерное эмпирическое распределение представляется в виде двойного ряда значений x_i и y_i . При этом корреляционную зависимость между признаками можно описывать разными способами. Соответствие между аргументом и функцией может быть задано таблицей, формулой, графиком и т. д.

Корреляционный анализ, как и другие статистические методы, основан на использовании вероятностных моделей, описывающих поведение исследуемых признаков в некоторой генеральной совокупности, из которой получены экспериментальные значения x_i и y_i . Когда исследуется корреляция между количествен-

ными признаками, значения которых можно точно измерить в единицах метрических шкал (метры, секунды, килограммы и т.д.), то очень часто принимается модель двумерной нормально распределенной генеральной совокупности. Такая модель отображает зависимость между переменными величинами x_i и y_i графически в виде геометрического места точек в системе прямоугольных координат. Эту графическую зависимость называют также диаграммой рассеивания или корреляционным полем.

Данная модель двумерного нормального распределения (корреляционное поле) позволяет дать наглядную графическую интерпретацию коэффициента корреляции, т.к. распределение в совокупности зависит от пяти параметров: математических ожиданий $E[x]$, $E[y]$ величин x , y ; стандартных отклонений σ_x , σ_y случайных величин x , y ; Q – коэффициент корреляции, который является мерой связи между случайными величинами X и Y .

Примеры корреляционных полей.

Если $Q = 0$, то значения x_i , y_i , полученные из двумерной нормальной совокупности, располагаются на графике в пределах области, ограниченной окружностью. В этом случае между случайными величинами x и y отсутствует корреляция, и они называются некоррелированными. Для двумерного нормального распределения некоррелированность означает одновременно и независимость случайных величин x и y .

Если $Q = 1$ или $Q = -1$, то говорят о полной корреляции, то есть между случайными величинами x и y существует линейная функциональная зависимость.

При $Q = 1$ значения x_i , y_i определяют точки, лежащие на прямой линии, имеющей положительный наклон (с увеличением x_i значения y_i также увеличиваются).

При $Q = -1$ прямая имеет отрицательный наклон.

Основы научных исследований

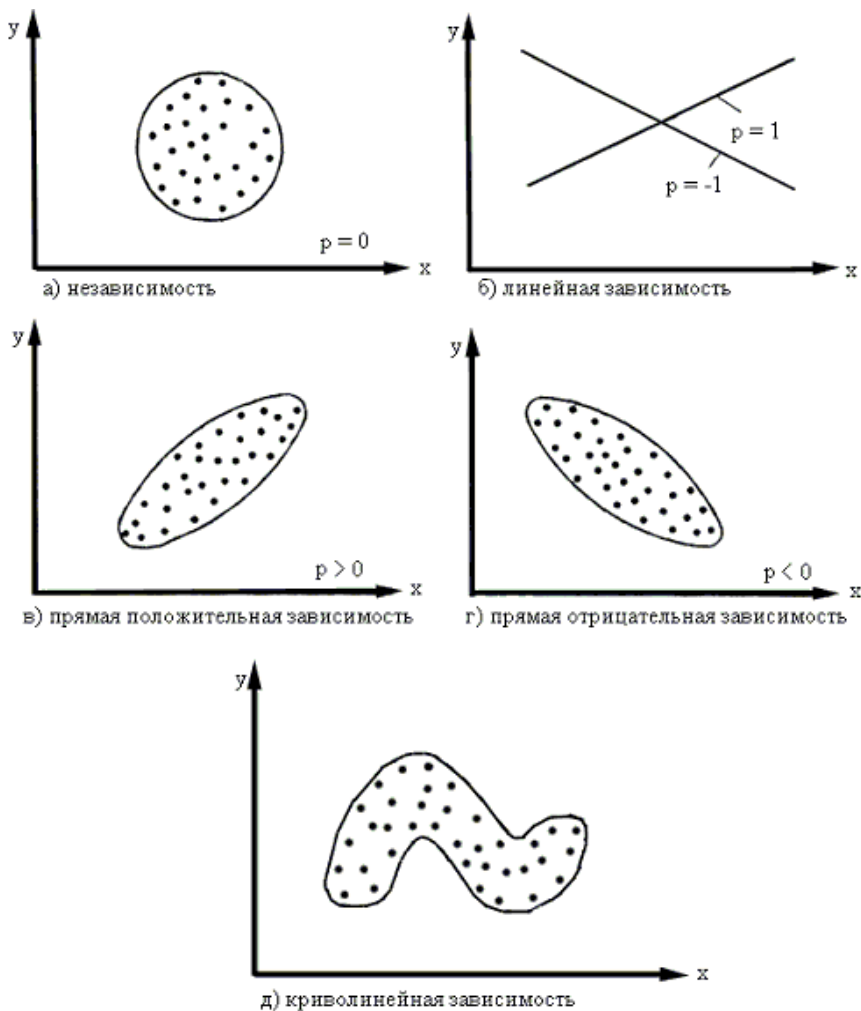


Рисунок 5 – Графическая интерпретация взаимосвязи между показателями

В промежуточных случаях, когда $-1 < Q < 1$, определяемые значениями x_i , y_i точки попадают в область, ограниченную некоторым эллипсом, причём при $Q > 0$ имеет место положительная корреляция (с увеличением x значения y в целом имеют тенденцию к возрастанию), при $Q < 0$ корреляция отрицательная. Чем ближе Q к ± 1 , тем уже эллипс и тем теснее точки, определяемые экспериментальными значениями, группируются около прямой

линии.

Здесь же следует обратить внимание на то, что линия, вдоль которой группируются точки, может быть не только прямой, а иметь любую другую форму: парабола, гипербола и т. д. В этих случаях рассматривают нелинейную корреляцию.

Корреляционную зависимость между признаками можно описывать разными способами, в частности, любая форма связи может быть выражена уравнением общего вида $y = f(x)$, где признак y – зависимая переменная, или функция от независимой переменной x , называемой аргументом.

Таким образом, визуальный анализ корреляционного поля помогает определить не только наличие статистической связи (линейной или нелинейной) между исследуемыми признаками, но и ее тесноту и форму.

3.5 Проведение корреляционного анализа

Многофакторный корреляционный анализ позволяет установить наличие, тесноту и форму связи между факторами и изучаемым показателем. Он состоит из нескольких этапов, деление на которые условно, так как отдельные стадии тесно связаны между собой.

Этапы проведения корреляционного анализа:

1. На первом этапе определяются цели и задачи исследования и на основе качественного анализа подбираются факторы, которые предположительно влияют на изучаемый показатель.

При их подборе необходимо учитывать: наличие причинно-следственных связей между показателями; значимость факторов, то есть степень их влияния на результирующий показатель; возможность количественного измерения фактора.

2. На втором этапе осуществляется сбор и первичная обработка исходной информации.

Совокупность данных должна быть достаточно большой. Информация должна соответствовать закону нормального распределения, согласно которому основная масса наблюдений по каждому показателю должна быть сгруппирована около его среднего значения.

Исходные данные должны быть качественно и количественно однородны. Качественная однородность предполагает приблизительно одинаковые условия и специфику формирования факторных и результирующих признаков. Количественная однородность заключается в отсутствии таких наблюдений, которые значительно (аномально) отличаются от основной массы данных.

Критерием однородности информации служит среднеквадратическое отклонение и коэффициент вариации, которые рассчитываются по каждому факторному и результативному показателю. Среднеквадратическое отклонение показывает абсолютное отклонение индивидуальных значений от среднеарифметической, а коэффициент вариации характеризует относительную меру отклонения отдельных значений от среднеарифметической. Причем, чем больше коэффициент вариации, тем относительно больший разброс данных в совокупности.

Изменчивость вариационного ряда принято считать:

- незначительной, если вариация не превышает 10%;
- средней, если вариация составляет 10-20%;
- значительной, если она больше 20%, но не превышает

33%. Если вариация больше 33 %, то следует исключить из выборки нетипичные наблюдения.

3. На третьем этапе осуществляется моделирование связей между факторами и результативным признаком, т.е. решается вопрос о выборе формы связи.

На основе экономического и логического анализа природы и сущности изучаемого явления подбирается тип математического уравнения, которое наилучшим образом отражает характер изучаемых зависимостей.

Обоснование уравнения связи проводится с помощью группировки данных, построения графика и т.д.

Размещение точек на графике показывает, какая зависимость образовалась: прямолинейная или криволинейная.

Для парной корреляции прямолинейную зависимость описывает уравнение прямой:

$$y = a_0 + a_1x_1 \quad (6)$$

Для множественной корреляции – уравнение линейной функции:

$$y = a_0 + a_1x_1 + a_2x_2 + \dots + a_nx_n, \quad (7)$$

где y – результативный показатель;

a_0 – постоянная величина, которая не связана с изменением факторов;

a_i – коэффициенты при переменных в уравнении регрессии;

x_i – факторные показатели (переменные).

Криволинейные зависимости в экономических исследованиях

ях встречаются реже. Они могут быть описаны уравнением

- параболы;
- гиперболы.

В случаях, когда трудно обосновать форму зависимости, решение задачи можно провести по разным моделям и сравнить полученные результаты. Наилучший вариант выбирается после оценки их по специальным критериям.

4. На четвертом этапе проводится расчет численных значений параметров регрессионного уравнения. Он осуществляется методом наименьших квадратов, при котором в случае прямолинейной зависимости прямая на графике пройдет наиболее близко к точкам фактических наблюдений. Для того решается система нормальных уравнений.

Построение уравнения регрессии проходит шаговым способом. При этом существуют два варианта:

а) сначала в расчет принимается один фактор, который является наиболее значимым, потом добавляется второй, третий и т.д. На каждом шаге рассчитываются уравнение связи, множественный коэффициент корреляции, коэффициент детерминации и статистические показатели, характеризующие надежность уравнения связи. Чем выше величины коэффициентов корреляции и лучше статистические характеристики, тем точнее уравнение связи описывает изучаемые зависимости. Если добавление новых факторов не улучшает оценочные показатели, то их надо отбросить и остановиться на том уравнении, где эти показатели наиболее оптимальны.

б) при втором варианте решение модели происходит в обратном порядке путем последовательного исключения наименее значимых факторов.

5. На пятом этапе дается статистическая оценка уравнения связи и экономическая интерпретация результатов корреляционного анализа.