# ФЕДЕРАЛЬНОЕ АГЕНТСТВО ПО ОБРАЗОВАНИЮ

Государственное образовательное учреждение высшего профессионального образования

«ТЮМЕНСКИЙ ГОСУДАРСТВЕННЫЙ НЕФТЕГАЗОВЫЙ УНИВЕРСИТЕТ»

# Институт менеджмента и бизнеса

### Кафедра МТЭК

#### КУРСОВАЯ РАБОТА

по дисциплине: «Общая теория статистики»

на тему:

Регрессионный анализ в статистическом изучении взаимосвязи показателей

Выполнил

 Проверил:

## Тюмень, 2010

#### СОДЕРЖАНИЕ

|  |  |
| --- | --- |
| Введение | 3 |
| 1.Статистическое изучение взаимосвязи социально-экономических явлений и процессов | 5 |
| 2.Характеристика регрессионного анализа | 11 |
| 2.1.Оценка взаимосвязи между факторным и результативным признаком на основе регрессионного анализа | 11 |
| 2.2.Отбор факторных признаков для построения множественной регрессионной модели | 13 |
| 2.3.Проверка адекватности моделей, построенных на основе уравнений регрессии | 17 |
| 3.Применение регрессионного анализа для изучения объекта исследования | 26 |
| Заключение | 33 |
| Список литературы  | 34 |

**Введение**

Обработка статистических данных уже давно применяется в самых разнообразных видах человеческой деятельности. Вообще говоря, трудно назвать ту сферу, в которой она бы не использовалась. Но, пожалуй, ни в одной области знаний и практической деятельности обработка статистических данных не играет такой исключительно большой роли, как в экономике, имеющей дело с обработкой и анализом огромных массивов информации о социально-экономических явлениях и процессах. Всесторонний и глубокий анализ этой информации, так называемых статистических данных, предполагает использование различных специальных методов, важное место среди которых занимает корреляционный и регрессионный анализы обработки статистических данных.

В экономических исследованиях часто решают задачу выявления факторов, определяющих уровень и динамику экономического процесса. Такая задача чаще всего решается методами корреляционного и регрессионного анализа. Для достоверного отображения объективно существующих в экономике процессов необходимо выявить существенные взаимосвязи и не только выявить, но и дать им количественную оценку. Этот подход требует вскрытия причинных зависимостей. Под причинной зависимостью понимается такая связь между процессами, когда изменение одного из них является следствием изменения другого.

Основными задачами корреляционного анализа являются оценка силы связи и проверка статистических гипотез о наличии и силе корреляционной связи. Не все факторы, влияющие на экономические процессы, являются случайными величинами, поэтому при анализе экономических явлений обычно рассматриваются связи между случайными и неслучайными величинами. Такие связи называются регрессионными, а метод математической статистики, их изучающий, называется регрессионным анализом.

В своей работе я рассмотрю корреляционно-регрессионный метод выявления взаимосвязи и проиллюстрирую его на примере.

1. **Статистическое изучение взаимосвязи социально-экономических явлений и процессов**

Исследование объективно существующих связей между явлениями – важнейшая задача общей теории статистики. В процессе статистического исследования зависимостей вскрываются причинно-следственные отношения между явлениями, что позволяет выявлять факторы (признаки), оказывающие существенное влияние на вариацию изучаемых явлений и процессов. Причинно-следственные отношения – это связь явлений и процессов, когда изменение одного и них – причины – ведет к изменению другого – следствия.

 Причина – это совокупность условий, обстоятельств, действие которых приводит к появлению следствия. Если между явлениями действительно существуют причинно-следственные отношения, то эти условия должны обязательно реализовываться вместе с действием причин. Причинные связи носят всеобщий и многообразный характер, и для обнаружения причинно-следственных связей необходимо отбирать отдельные явления и изучать их изолированно.

 Особое значение при исследовании причинно-следственных связей имеет выявление временной последовательности: причина всегда должна предшествовать следствию, однако не каждое предшествующее событие следует считать причиной, а последующее следствием.

 В реальной социально-экономической действительности причину и следствие необходимо рассматривать как смежные явления, появление которых обусловлено комплексом сопутствующих более простых причин и следствий. Между сложными группами причин и следствий возможны многозначительные связи, когда за одной причиной будет следовать то одно, то другое действие или одно действие имеет несколько различных причин. Чтобы установить однозначную причинную связь между явлениями или предсказать возможные следствия конкретной причины, необходима полная абстракция от всех прочих явлений в исследуемой временной или пространственной среде. Теоретически такая абстракция воспроизводится. Приемы абстракции часто применяются при изучении взаимосвязей между двумя признаками (парной корреляции). Но чем сложнее изучаемые явления, тем труднее выявить причинно-следственные связи между ними. Взаимное переплетение различных внутренних и внешних факторов неизбежно приводит к некоторым ошибкам в определении причины и следствия.

Социально-экономические явления представляют собой результат одновременного воздействия большого числа причин. Следовательно, при изучении этих явлений необходимо выявлять главные, основные причины, абстрагируясь от второстепенных.

В основе первого этапа статистического изучения связи лежит качественный анализ изучаемого явления, связанный с анализом природы, социального или экономического явления методами экономической теории, социологии, конкретной экономики. Второй этап – построение модели связи. Он базируется на методах статистики: группировках, средних величинах, таблицах и т.д. Третий, последний этап – интерпретация результатов – вновь связан с качественными особенностями изучаемого явления.

Статистика разработала множество методов изучения связей, выбор которых зависит от целей исследования и от поставленных задач. Связи между признаками и явлениями, ввиду их большого разнообразия, классифицируются по ряду оснований. Признаки по их значению для изучения взаимосвязи делятся на 2 класса. Признаки, обуславливающие изменения других, связанных с ними признаков, называются факторными, или просто факторами. Признаки, изменяющиеся под действием факторных признаков, являются результативными. Связи между явлениями и их признаками классифицируются по степени тесноты связи, направлению и аналитическому выражению.

Между различными явлениями и их признаками необходимо прежде всего выделить 2 типа связей: функциональную (жестко детерминированную) и статистическую (стохастически детерминированную).

В соответствии с жестко детерминистическим представлением о функционировании экономических систем необходимость и закономерность однозначно проявляются в каждом отдельном явлении, то есть любое действие вызывает строго определенный результат; случайными (непредвиденными заранее) воздействиями при этом пренебрегают. Поэтому при заданных начальных условиях состояние такой системы может быть определено с вероятностью, равной 1. Разновидностью такой закономерности является функциональная связь. Задача регрессионного анализа состоит в построении модели, позволяющей по значениям независимых показателей получать оценки значений зависимой переменной. Регрессионный анализ является основным средством исследования зависимостей между социально-экономическими переменными. Эту задачу мы рассмотрим в рамках самой распространенной в статистических пакетах классической модели линейной регрессии. Специфика социологических исследований состоит в том, что очень часто необходимо изучать и предсказывать социальные события. Вторая часть данной главы будет посвящена регрессии, целью которой является построение моделей, предсказывающих вероятности событий. Величина называется ошибкой регрессии. Первые математические результаты, связанные с регрессионным анализом, сделаны в предположении, что регрессионная ошибка распределена нормально с параметрами, ошибка для различных объектов считаются независимыми. Кроме того, в данной модели мы рассматриваем переменные как неслучайные значения. Такое, на практике, получается, когда идет активный эксперимент, в котором задают значения (например, назначили зарплату работнику), а затем измеряют (оценили, какой стала производительность труда).

Связь признака у с признаком х называется функциональной, если каждому возможному значению независимого признака х соответствует 1 или несколько строго определенных значений зависимого признака у. Определение функциональной связи может быть легко обобщено для случая многих признаков х1,х2 …хn . Метод включения и исключения переменных состоит в следующем. Из множества факторов, рассматриваемых исследователем как возможные аргументы регрессионного уравнения, отбирается один, который более всего связан корреляционной зависимостью. Далее проводится та же процедура при двух выбранных переменных, при трех и т.д. Процедура повторяется до тех пор, пока в уравнение не будут включены все аргументы, выделенные исследователем, удовлетворяющие критериям значимости включения. Замечание: во избежание зацикливания процесса включения исключения значимость включения устанавливается меньше значимости исключения. Переменные, порождаемые регрессионным уравнением. Сохранение переменных, порождаемых регрессией, производится подкомандой. Благодаря полученным оценкам коэффициентов уравнения регрессии могут быть оценены прогнозные значения зависимой переменной, причем они могут быть вычислены и там, где значения определены, и там где они не определены.

Характерной особенностью функциональных связей является то, что в каждом отдельном случае известен полный перечень факторов, определяющих значение зависимого (результативного) признака, а также точный механизм их влияния, выраженный определенным уравнением.

Функциональную связь можно представить уравнением:

yi= (xi),

где yi - результативный признак ( i = 1, … , n);

 f(xi) - известная функция связи результативного и факторного признаков;

 xi - факторный признак.[11]

В реальной общественной жизни ввиду неполноты информации жестко детерминированной системы, может возникнуть неопределенность, из-за которой эта система по своей природе должна рассматриваться как вероятностная, при этом связь между признаками становится стохастической.

Стохастическая связь – это связь между величинами, при которой одна из них, случайная величина у, реагирует на изменение другой величины х или других величин х1,х2 …хn (случайных или неслучайных) изменением закона распределения. Это обуславливается тем, что зависимая переменная (результативный признак), кроме рассматриваемых независимых, подвержена влиянию ряда неучтенных или неконтролируемых (случайных) факторов, а также некоторых неизбежных ошибок измерения переменных. Поскольку значения зависимой переменной подвержены случайному разбросу, они не могут быть предсказаны с достаточной точностью, а только указаны с определенной вероятностью.

Характерной особенностью стохастических связей является то, что они проявляются во всей совокупности, а не в каждой ее единице. Причём неизвестен ни полный перечень факторов, определяющих значение результативного признака, ни точный механизм их функционирования и взаимодействия с результативным признаком. Всегда имеет место влияние случайного. Появляющиеся различные значения зависимой переменной – реализация случайной величины. Однако при небольшой взаимосвязи между переменными, если стандартизовать переменные и рассчитать уравнение регрессии для стандартизованных переменных, то оценки коэффициентов регрессии позволят по их абсолютной величине судить о том, какой аргумент в большей степени влияет на функцию. Стандартизация переменных. Бета коэффициенты. Коэффициенты в последнем уравнении получены при одинаковых масштабах изменения всех переменных и сравнимы. В случае взаимосвязи между аргументами в правой части уравнения могут происходить странные вещи. Надежность и значимость коэффициента регрессии. Здесь обозначен коэффициент детерминации, получаемый при построении уравнения регрессии, в котором в качестве зависимой переменной взята другая переменная. Из выражения видно, что величина коэффициента тем неустойчивее, чем сильнее переменная связана с остальными переменными. Эта статистика имеет распределение Стьюдента. В выдаче пакета печатается наблюдаемая ее двусторонняя значимость - вероятность случайно при нулевом регрессионном коэффициенте получить значение статистики, большее по абсолютной величине, чем выборочное. Значимость включения переменной в регрессию. При последовательном подборе переменных предусмотрена автоматизация, основанная на значимости включения и исключения переменных.

Модель стохастической связи может быть представлена в общем виде уравнением:

ŷi = (xi) + εi ,

где ŷi - расчётное значение результативного признака;

 f(xi) - часть результативного признака, сформировавшаяся под воздействием учтенных известных факторных признаков (одного или множества), находящихся в стохастической связи с признаком;

 εi - часть результативного признака, возникшая вследствие действия неконтролируемых или неучтенных факторов, а также измерения признаков, неизбежно сопровождающегося некоторыми случайными ошибками. За это иногда зависимую переменную называют откликом. Теория регрессионных уравнений со случайными независимыми переменными сложнее, но известно, что, при большом числе наблюдений, использование метода разработанного корректно. Для получения оценок коэффициентов регрессии минимизируется сумма квадратов ошибок регрессии. В пакете вычисляются статистики, позволяющие решить эти задачи. Существует ли линейная регрессионная зависимость? Для проверки одновременного отличия всех коэффициентов регрессии от нуля проведем анализ квадратичного разброса значений зависимой переменной относительно среднего. Его можно разложить на две суммы следующим образом. Статистика в условиях гипотезы равенства нулю регрессионных коэффициентов имеет распределение Фишера и, естественно, по этой статистике проверяют, являются ли коэффициенты одновременно нулевыми. Коэффициенты детерминации и множественной корреляции. При сравнении качества регрессии, оцененной по различным зависимым переменным, полезно исследовать доли объясненной и необъясненной дисперсии. Корень из коэффициента детерминации называется коэффициентом корреляции. Следует иметь в виду, что является смещенной оценкой. Абсолютные значения коэффициентов не позволяют сделать такой вывод.

Проявление стохастических связей подвержено действию закона больших чисел: лишь в достаточно большом числе единиц индивидуальные особенности сгладятся, случайности взаимопогасятся, и зависимость, если она имеет существенную силу, проявится достаточно отчётливо. [6]

Корреляционная связь существует там, где взаимосвязанные явления характеризуются только случайными величинами. При такой связи среднее значение (математическое ожидание) случайной величины результативного признака у закономерно изменяется в зависимости от изменения другой величины х или других случайных величин х1,х2 …хn. Корреляционная связь проявляется не в каждом отдельном случае, а во всей совокупности в целом. Только при достаточно большом количестве случаев каждому значению случайного признака х будет соответствовать распределение средних значений случайного признака у. Наличие корреляционных связей присуще многим общественным явлениям. Задача регрессионного анализа состоит в построении модели, позволяющей по значениям независимых показателей получать оценки значений зависимой переменной. Регрессионный анализ является основным средством исследования зависимостей между социально-экономическими переменными. Эту задачу мы рассмотрим в рамках самой распространенной в статистических пакетах классической модели линейной регрессии. Специфика социологических исследований состоит в том, что очень часто необходимо изучать и предсказывать социальные события. Вторая часть данной главы будет посвящена регрессии, целью которой является построение моделей, предсказывающих вероятности событий. Величина называется ошибкой регрессии. Первые математические результаты, связанные с регрессионным анализом, сделаны в предположении, что регрессионная ошибка распределена нормально с параметрами, ошибка для различных объектов считаются независимыми. Кроме того, в данной модели мы рассматриваем переменные как неслучайные значения. Такое, на практике, получается, когда идет активный эксперимент, в котором задают значения (например, назначили зарплату работнику), а затем измеряют (оценили, какой стала производительность труда).

Корреляционная связь – понятие более узкое, чем стохастическая связь. Последняя может отражаться не только в изменении средней величины, но и в вариации одного признака в зависимости от другого, то есть любой другой характеристики вариации. Таким образом, корреляционная связь является частным случаем стохастической связи.

Прямые и обратные связи. В зависимости от направления действия, функциональные и стохастические связи могут быть прямые и обратные. При прямой связи направление изменения результативного признака совпадает с направлением изменения признака-фактора, то есть с увеличением факторного признака увеличивается и результативный, и, наоборот, с уменьшением факторного признака уменьшается и результативный признак. В противном случае между рассматриваемыми величинами существуют обратные связи. Например, чем выше квалификация рабочего (разряд), тем выше уровень производительности труда – прямая связь. А чем выше производительность труда, тем ниже себестоимость единицы продукции – обратная связь. Рассмотрим, что представляет собой эта значимость. Обозначим коэффициент детерминации, полученный при исключении из правой части уравнения переменной. При этом мы получим уменьшение объясненной дисперсии, на величину. Для оценки значимости включения переменной используется статистика, имеющая распределение Фишера при нулевом теоретическом приросте. Вообще, если из уравнения регрессии исключаются переменных, статистикой значимости исключения будет. Пошаговая процедура построения модели. Основным критерием отбора аргументов должно быть качественное представление о факторах, влияющих на зависимую переменную, которую мы пытаемся смоделировать. Очень хорошо реализован процесс построения регрессионной модели: на машину переложена значительная доля трудностей в решении этой задачи. Возможно построение последовательное построение модели добавлением и удалением блоков переменных. Но мы рассмотрим только работу с отдельными переменными. По умолчанию программа включает все заданные переменные.

Прямолинейные и криволинейные связи. По аналитическому выражению (форме) связи могут быть прямолинейными и криволинейными. При прямолинейной связи с возрастанием значения факторного признака происходит непрерывное возрастание (или убывание) значений результативного признака. Математически такая связь представляется уравнением прямой, а графически – прямой линией. Отсюда ее более короткое название – линейная связь. При криволинейных связях с возрастанием значения факторного признака возрастание (или убывание) результативного признака происходит неравномерно, или же направление его изменения меняется на обратное. Геометрически такие связи представляются кривыми линиями (гиперболой, параболой и т.д.).

Однофакторные и многофакторные связи. По количеству факторов, действующих на результативный признак, связи различаются: однофакторные (один фактор) и многофакторные (два и более факторов). Однофакторные (простые) связи обычно называются парными (т.к. рассматривается пара признаков). Например, корреляционная связь между прибылью и производительностью труда. В случае многофакторной (множественной) связи имеют в виду, что все факторы действуют комплексно, то есть одновременно и во взаимосвязи. Например, корреляционная связь между производительностью труда и уровнем организации труда, автоматизации производства, квалификации рабочих, производственным стажем, простоями и другими факторными признаками. С помощью множественной корреляции можно охватить весь комплекс факторных признаков и объективно отразить существующие множественные связи.

Метод включения и исключения переменных состоит в следующем. Из множества факторов, рассматриваемых исследователем как возможные аргументы регрессионного уравнения, отбирается один, который более всего связан корреляционной зависимостью. Далее проводится та же процедура при двух выбранных переменных, при трех и т.д. Процедура повторяется до тех пор, пока в уравнение **2.Характеристика регрессионного анализа**

**2.1.Оценка взаимосвязи между факторным и результативным признаком на основе регрессионного анализа**

Для исследования стохастических связей широко используется метод сопоставления двух параллельных рядов, метод аналитических группировок, корреляционный анализ, регрессионный анализ и некоторые непараметрические методы.[1]

адача регрессионного анализа состоит в построении модели, позволяющей по значениям независимых показателей получать оценки значений зависимой переменной. Регрессионный анализ является основным средством исследования зависимостей между социально-экономическими переменными. Эту задачу мы рассмотрим в рамках самой распространенной в статистических пакетах классической модели линейной регрессии. Специфика социологических исследований состоит в том, что очень часто необходимо изучать и предсказывать социальные события. Вторая часть данной главы будет посвящена регрессии, целью которой является построение моделей, предсказывающих вероятности событий. Величина называется ошибкой регрессии. Первые математические результаты, связанные с регрессионным анализом, сделаны в предположении, что регрессионная ошибка распределена нормально с параметрами, ошибка для различных объектов считаются независимыми. Кроме того, в данной модели мы рассматриваем переменные как неслучайные значения. Такое, на практике, получается, когда идет активный эксперимент, в котором задают значения (например, назначили зарплату работнику), а затем измеряют (оценили, какой стала производительность труда).

В общем виде задача статистики в области изучения взаимосвязей состоит не только в количественной оценке их наличия, направления и силы связи, но и в определении формы (аналитического выражения) влияния факторных признаков на результативные.

Задачами регрессионного анализа являются выбор типа модели (формы связи), установление степени влияния независимых переменных на зависимую и определение расчётных значений зависимой переменной (функции регрессии). Задача регрессионного анализа состоит в построении модели, позволяющей по значениям независимых показателей получать оценки значений зависимой переменной. Регрессионный анализ является основным средством исследования зависимостей между социально-экономическими переменными. Эту задачу мы рассмотрим в рамках самой распространенной в статистических пакетах классической модели линейной регрессии. Специфика социологических исследований состоит в том, что очень часто необходимо изучать и предсказывать социальные события. Вторая часть данной главы будет посвящена регрессии, целью которой является построение моделей, предсказывающих вероятности событий. Величина называется ошибкой регрессии. Первые математические результаты, связанные с регрессионным анализом, сделаны в предположении, что регрессионная ошибка распределена нормально с параметрами, ошибка для различных объектов считаются независимыми. Кроме того, в данной модели мы рассматриваем переменные как неслучайные значения. Такое, на практике, получается, когда идет активный эксперимент, в котором задают значения (например, назначили зарплату работнику), а затем измеряют (оценили, какой стала производительность труда).

из правой части уравнения переменной. При этом мы получим уменьшение объясненной дисперсии, на величину. Для оценки значимости включения переменной используется статистика, имеющая распределение Фишера при нулевом теоретическом приросте. Вообще, если из уравнения регрессии исключаются переменных, статистикой значимости исключения будет. Пошаговая процедура построения модели. Основным критерием отбора аргументов должно быть качественное представление о факторах, влияющих на зависимую переменную, которую мы пытаемся смоделировать. Очень хорошо реализован процесс построения регрессионной модели: на машину переложена значительная доля трудностей в решении этой задачи. Возможно построение последовательное построение модели добавлением и удалением блоков переменных. Но мы рассмотрим только работу с отдельными переменными. По умолчанию программа включает все заданные переменные.

Корреляционный и регрессионный анализ. Исследование связей в условиях массового наблюдения и действия случайных факторов осуществляется, как правило, с помощью экономико-статистических моделей. В широком смысле модель – это аналог, условный образ (изображение, описание, схема, чертёж и т.п.) какого-либо объекта, процесса или события, приближенно воссоздающий «оригинал». Модель представляет собой логическое или математическое описание компонентов и функций, отображающих существенные свойства моделируемого объекта или процесса, даёт возможность установить основные закономерности изменения оригинала. В модели оперируют показателями, исчисленными для качественно однородных массовых явлений (совокупностей). Выражение и модели в виде функциональных уравнений используют для расчёта средних значений моделируемого показателя по набору заданных величин и для выявления степени влияния на него отдельных факторов. Метод включения и исключения переменных состоит в следующем. Из множества факторов, рассматриваемых исследователем как возможные аргументы регрессионного уравнения, отбирается один, который более всего связан корреляционной зависимостью. Далее проводится та же процедура при двух выбранных переменных, при трех и т.д. Процедура повторяется до тех пор, пока в уравнение не будут включены все аргументы, выделенные исследователем, удовлетворяющие критериям значимости включения. Замечание: во избежание зацикливания процесса включения исключения значимость включения устанавливается меньше значимости исключения. Переменные, порождаемые регрессионным уравнением. Сохранение переменных, порождаемых регрессией, производится подкомандой. Благодаря полученным оценкам коэффициентов уравнения регрессии могут быть оценены прогнозные значения зависимой переменной, причем они могут быть вычислены и там, где значения определены, и там где они не определены.

По количеству включаемых факторов модели могут быть однофакторными и многофакторными (два и более факторов).

В зависимости от познавательной цели статистические модели подразделяются на структурные, динамические и модели связи.

Двухмерная линейная модель корреляционного и регрессионного анализа (однофакторный линейный корреляционный и регрессионный анализ). Наиболее разработанной в теории статистики является методология так называемой парной корреляции, рассматривающая влияние вариации факторного анализа х на результативный признак у и представляющая собой однофакторный корреляционный и регрессионный анализ. Овладение теорией и практикой построения и анализа двухмерной модели корреляционного и регрессионного анализа представляет собой исходную основу для изучения многофакторных стохастических связей.

 Однако при небольшой взаимосвязи между переменными**2.2.Отбор факторных признаков для построения множественной регрессионной модели**

Важнейшим этапом построения регрессионной модели (уравнения регрессии) является установление в анализе исходной информации математической функции. Сложность заключается в том, что из множества функций необходимо найти такую, которая лучше других выражает реально существующие связи между анализируемыми признаками. Выбор типов функции может опираться на теоретические знания об изучаемом явлении, опыт предыдущих аналогичных исследований, или осуществляться эмпирически – перебором и оценкой функций разных типов и т.п. [10]

При изучении связи экономических показателей производства (деятельности) используют различного вида уравнения прямолинейной и криволинейной связи. Внимание к линейным связям объясняется ограниченной вариацией переменных и тем, что в большинстве случаев нелинейные формы связи для выполнения расчётов преобразуют (путём логарифмирования или замены переменных) в линейную форму. Уравнение однофакторной (парной) линейной корреляционной связи имеет вид:

ŷ = a0 + a1x ,

где ŷ - теоретические значения результативного признака, полученные по уравнению регрессии;

 a0 , a1 - коэффициенты (параметры) уравнения регрессии. Задача регрессионного анализа состоит в построении модели, позволяющей по значениям независимых показателей получать оценки значений зависимой переменной. Регрессионный анализ является основным средством исследования зависимостей между социально-экономическими переменными. Эту задачу мы рассмотрим в рамках самой распространенной в статистических пакетах классической модели линейной регрессии. Специфика социологических исследований состоит в том, что очень часто необходимо изучать и предсказывать социальные события. Вторая часть данной главы будет посвящена регрессии, целью которой является построение моделей, предсказывающих вероятности событий. Величина называется ошибкой регрессии. Первые математические результаты, связанные с регрессионным анализом, сделаны в предположении, что регрессионная ошибка распределена нормально с параметрами, ошибка для различных объектов считаются независимыми. Кроме того, в данной модели мы рассматриваем переменные как неслучайные значения. Такое, на практике, получается, когда идет активный эксперимент, в котором задают значения (например, назначили зарплату работнику), а затем измеряют (оценили, какой стала производительность труда).

Поскольку a0 является средним значением у в точке х=0, экономическая интерпретация часто затруднена или вообще невозможна. За это иногда зависимую переменную называют откликом. Теория регрессионных уравнений со случайными независимыми переменными сложнее, но известно, что, при большом числе наблюдений, использование метода разработанного корректно. Для получения оценок коэффициентов регрессии минимизируется сумма квадратов ошибок регрессии. В пакете вычисляются статистики, позволяющие решить эти задачи. Существует ли линейная регрессионная зависимость? Для проверки одновременного отличия всех коэффициентов регрессии от нуля проведем анализ квадратичного разброса значений зависимой переменной относительно среднего. Его можно разложить на две суммы следующим образом. Статистика в условиях гипотезы равенства нулю регрессионных коэффициентов имеет распределение Фишера и, естественно, по этой статистике проверяют, являются ли коэффициенты одновременно нулевыми. Коэффициенты детерминации и множественной корреляции. При сравнении качества регрессии, оцененной по различным зависимым переменным, полезно исследовать доли объясненной и необъясненной дисперсии. Корень из коэффициента детерминации называется коэффициентом корреляции. Следует иметь в виду, что является смещенной оценкой. Абсолютные значения коэффициентов не позволяют сделать такой вывод.

Коэффициент парной линейной регрессии a1 имеет смысл показателя силы связи между вариацией факторного признака х и вариацией результативного признака у. Вышеприведенное уравнение показывает среднее значение изменения результативного признака y при изменении факторного признака х на одну единицу его измерения, то есть вариацию у, приходящуюся на единицу вариации х. Знак a1 указывает направление этого изменения.

Параметры уравнения a0 , a1 находят методом наименьших квадратов (метод решения систем уравнений, при котором в качестве решения принимается точка минимума суммы квадратов отклонений), то есть в основу этого метода положено требование минимальности сумм квадратов отклонений эмпирических данных yi от выравненных ŷ:

Σ(yi – ŷ)2 = Σ(yi – a0 – a1xi)2 → min [9]

Для нахождения минимума данной функции приравняем к нулю ее частные производные и получим систему двух линейных уравнений, которая называется системой нормальных уравнений:

Решим эту систему в общем виде:

Параметры уравнения парной линейной регрессии иногда удобно исчислять по следующим формулам, дающим тот же результат:



Определив значения a0 , a1 и подставив их в уравнение связи

 ŷ = a0 + a1x , находим значения ŷ, зависящие только от заданного значения х.

Рассмотрим построение однофакторного уравнения регрессии зависимости работающих активов у от капитала х (см. Таблица 1, Приложение 1). Рассмотрим, что представляет собой эта значимость. Обозначим коэффициент детерминации, полученный при исключении из правой части уравнения переменной. При этом мы получим уменьшение объясненной дисперсии, на величину. Для оценки значимости включения переменной используется статистика, имеющая распределение Фишера при нулевом теоретическом приросте. Вообще, если из уравнения регрессии исключаются переменных, статистикой значимости исключения будет. Пошаговая процедура построения модели. Основным критерием отбора аргументов должно быть качественное представление о факторах, влияющих на зависимую переменную, которую мы пытаемся смоделировать. Очень хорошо реализован процесс построения регрессионной модели: на машину переложена значительная доля трудностей в решении этой задачи. Возможно построение последовательное построение модели добавлением и удалением блоков переменных. Но мы рассмотрим только работу с отдельными переменными. По умолчанию программа включает все заданные переменные.

Здесь представлены показатели 32 банков: размер капитала и работающих активов. Передо мной стоит задача определить, есть ли зависимость между этими двумя признаками и, если она существует, определить форму этой зависимости, то есть уравнение регрессии.

За факторный признак я взял размер капитала банка, а за результативный признак – работающие активы. [11]

Сопоставление данных параллельных рядов признаков х и у показывает, что с убыванием признака х (капитал), в большинстве случаев убывает и признак у (работающие активы). Задача регрессионного анализа состоит в построении модели, позволяющей по значениям независимых показателей получать оценки значений зависимой переменной. Регрессионный анализ является основным средством исследования зависимостей между социально-экономическими переменными. Эту задачу мы рассмотрим в рамках самой распространенной в статистических пакетах классической модели линейной регрессии. Специфика социологических исследований состоит в том, что очень часто необходимо изучать и предсказывать социальные события. Вторая часть данной главы будет посвящена регрессии, целью которой является построение моделей, предсказывающих вероятности событий. Величина называется ошибкой регрессии. Первые математические результаты, связанные с регрессионным анализом, сделаны в предположении, что регрессионная ошибка распределена нормально с параметрами, ошибка для различных объектов считаются независимыми. Кроме того, в данной модели мы рассматриваем переменные как неслучайные значения. Такое, на практике, получается, когда идет активный эксперимент, в котором задают значения (например, назначили зарплату работнику), а затем измеряют (оценили, какой стала производительность труда).

Следовательно, можно предположить, что между х и у существует прямая зависимость, пусть неполная, но выраженная достаточно ясно.

Для уточнения формы связи между рассматриваемыми признаками я использовал графический метод. Я нанес на график точки, соответствующие значениям х и у, и получил корреляционное поле (см. График 1, Приложение 2). Метод включения и исключения переменных состоит в следующем. Из множества факторов, рассматриваемых исследователем как возможные аргументы регрессионного уравнения, отбирается один, который более всего связан корреляционной зависимостью. Далее проводится та же процедура при двух выбранных переменных, при трех и т.д. Процедура повторяется до тех пор, пока в уравнение не будут включены все аргументы, выделенные исследователем, удовлетворяющие критериям значимости включения. Замечание: во избежание зацикливания процесса включения исключения значимость включения устанавливается меньше значимости исключения. Переменные, порождаемые регрессионным уравнением. Сохранение переменных, порождаемых регрессией, производится подкомандой. Благодаря полученным оценкам коэффициентов уравнения регрессии могут быть оценены прогнозные значения зависимой переменной, причем они могут быть вычислены и там, где значения определены, и там где они не определены.

Анализируя поле корреляции, можно предположить, что возрастание признака у идет пропорционально признаку х. В основе этой зависимости лежит прямолинейная связь, которая может быть выражена простым линейным уравнением регрессии:

ŷ = a0 + a1x,

где ŷ - теоретические расчётные значения результативного признака (работающие активы), полученные по уравнению регрессии;

 a0 , a1 - коэффициенты (параметры) уравнения регрессии;

 х – капитал исследуемых банков.

Пользуясь вышеуказанными формулами для вычисления параметров линейного уравнения регрессии и расчётными значениями из Таблицы 1 (Приложение 1), получаем:

Следовательно, регрессионная модель зависимости работающих активов от капитала банков может быть записана в виде конкретного простого уравнения регрессии:

.[4]

Это уравнение характеризует зависимость работающих активов от капитала банка. Расчётные значения ŷ , найденные по этому уравнению, приведены в Таблице 1 (Приложение 1). Правильность расчёта параметров уравнения регрессии может быть проверена сравниванием сумм ∑у = ∑ŷ . В моем случае эти суммы равны. Однако при небольшой взаимосвязи между переменными, если стандартизовать переменные и рассчитать уравнение регрессии для стандартизованных переменных, то оценки коэффициентов регрессии позволят по их абсолютной величине судить о том, какой аргумент в большей степени влияет на функцию. Стандартизация переменных. Бета коэффициенты. Коэффициенты в последнем уравнении получены при одинаковых масштабах изменения всех переменных и сравнимы. В случае взаимосвязи между аргументами в правой части уравнения могут происходить странные вещи. Надежность и значимость коэффициента регрессии. Здесь обозначен коэффициент детерминации, получаемый при построении уравнения регрессии, в котором в качестве зависимой переменной взята другая переменная. Из выражения видно, что величина коэффициента тем неустойчивее, чем сильнее переменная связана с остальными переменными. Эта статистика имеет распределение Стьюдента. В выдаче пакета печатается наблюдаемая ее двусторонняя значимость - вероятность случайно при нулевом регрессионном коэффициенте получить значение статистики, большее по абсолютной величине, чем выборочное. Значимость включения переменной в регрессию. При последовательном подборе переменных предусмотрена автоматизация, основанная на значимости включения и исключения переменных.

Но для того, чтобы применить мою формулу, надо рассчитать, насколько она приближенна к реальности, то есть проверить ее адекватность.

**2.3. Проверка адекватности моделей, построенных на основе уравнений регрессии**

Для практического использования моделей регрессии большое значение имеет их адекватность, т.е. соответствие фактическим статистическим данным.

Корреляционный и регрессионный анализ обычно (особенно в условиях так называемого малого и среднего бизнеса) проводится для ограниченной по объёму совокупности. Поэтому показатели регрессии и корреляции – параметры уравнения регрессии, коэффициенты корреляции и детерминации могут быть искажены действием случайных факторов. Чтобы проверить, насколько эти показатели характерны для всей генеральной совокупности, не являются ли они результатом стечения случайных обстоятельств, необходимо проверить адекватность построенных статистических моделей.

При численности объектов анализа до 30 единиц возникает необходимость проверки значимости (существенности) каждого коэффициента регрессии. При этом выясняют насколько вычисленные параметры, характерны для отображения комплекса условий: не являются ли полученные значения параметров результатами действия случайных причин. Значимость коэффициентов простой линейной регрессии (применительно к совокупностям, у которых n<30) осуществляют с помощью t-критерия Стьюдента. При этом вычисляют расчетные (фактические) значения t-критерия

для параметра a0:

для параметра a1:

 где n - объём выборки;

- среднее квадратическое отклонение результативного признака от выравненных значений ŷ ;

 или

- среднее квадратическое отклонение факторного признака x от общей средней . [8]

Вычисленные по вышеприведенным формулам значения сравнивают с критическими t, которые определяют по таблице Стьюдента с учетом принятого уровня значимости α и числом степеней свободы вариации . В социально-экономических исследованиях уровень значимости α обычно принимают равным 0,05. Параметр признаётся значимым (существенным) при условии, если tрасч> tтабл . В таком случае практически невероятно, что найденные значения параметров обусловлены только случайными совпадениями.

Теперь я рассчитаю t-критерий Стьюдента для моей модели регрессии.

- это средние квадратические отклонения.

Расчетные значения t-критерия Стьюдента:

По таблице распределения Стьюдента я нахожу критическое значение t-критерия для ν= 32-2 = 30 . Вероятность α я принимаю 0,05. tтабл равно 2,042. Так как, оба значения ta0 и ta1 больше tтабл , то оба параметра а0 и а1 признаются значимыми и отклоняется гипотеза о том, что каждый из этих параметров в действительности равен 0 , и лишь в силу случайных обстоятельств оказался равным проверяемой величине.

Проверка адекватности регрессионной модели может быть дополнена корреляционным анализом. Для этого необходимо определить тесноту корреляционной связи между переменными х и у. Теснота корреляционной связи, как и любой другой, может быть измерена эмпирическим корреляционным отношением ηэ, когда δ2 (межгрупповая дисперсия) характеризует отклонения групповых средних результативного признака от общей средней: .

Говоря о корреляционном отношении как о показателе измерения тесноты зависимости, следует отличать от эмпирического корреляционного отношения – теоретическое.

Теоретическое корреляционное отношение η представляет собой относительную величину, получающуюся в результате сравнения среднего квадратического отклонения выравненных значений результативного признака δ, то есть рассчитанных по уравнению регрессии, со средним квадратическим отношением эмпирических (фактических) значений результативности признака σ:

 ,

где ; .

Тогда . [2]

Изменение значения η объясняется влиянием факторного признака.

В основе расчёта корреляционного отношения лежит правило сложения дисперсий, то есть , где - отражает вариацию у за счёт всех остальных факторов, кроме х, то есть являются остаточной дисперсией:

.

Тогда формула теоретического корреляционного отношения примет вид:

,

или .

Подкоренное выражение корреляционного выражения представляет собой коэффициент детерминации (мера определенности, причинности).

Коэффициент детерминации показывает долю вариации результативного признака под влиянием вариации признака-фактора. Задача

Теоретическое корреляционное выражение применяется для измерения тесноты связи при линейной и криволинейной зависимостях между результативным и факторным признаком.

Как видно из вышеприведенных формул корреляционное отношение может находиться от 0 до 1. Чем ближе корреляционное отношение к 1, тем связь между признаками теснее.

Теоретическое корреляционное отношение применительно к моему анализу я рассчитаю двумя способами:

[5]

Полученное значение теоретического корреляционного отношения свидетельствует о возможном наличии среднестатистической связи между рассматриваемыми признаками. Коэффициент детерминации равен 0,62. Отсюда я заключаю, что 62% общей вариации работающих активов изучаемых банков обусловлено вариацией фактора – капитала банков (а 38% общей вариации нельзя объяснить изменением размера капитала).

Кроме того, при линейной форме уравнения применяется другой показатель тесноты связи – линейный коэффициент корреляции:

,

где n – число наблюдений.

Для практических вычислений при малом числе наблюдений (n≤20÷30) линейный коэффициент корреляции удобнее исчислять по следующей формуле:

.

Значение линейного коэффициента корреляции важно для исследования социально-экономических явлений и процессов, распределение которых близко к нормальному. Он принимает значения в интервале:

-1≤ r ≤ 1.

Отрицательные значения указывают на обратную связь, положительные – на прямую. При r = 0 линейная связь отсутствует. Чем ближе коэффициент корреляции по абсолютной величине к единице, тем теснее связь между признаками. И, наконец, при r = ±1 – связь функциональная.

Используя данные Таблицы 1 (Приложение 1), я рассчитал линейный коэффициент корреляции r. Но чтобы использовать формулу для линейного коэффициента корреляции рассчитаем дисперсию результативного признака σy:

Квадрат линейного коэффициента корреляции r2 называется линейным коэффициентом детерминации. Из определения коэффициента детерминации очевидно, что его числовое значение всегда заключено в пределах от 0 до 1, то есть 0 ≤ r2 ≤ 1. Степень тесноты связи полностью соответствует теоретическому корреляционному отношению, которое является более универсальным показателем тесноты связи по сравнению с линейным коэффициентом корреляции.

Факт совпадений и несовпадений значений теоретического корреляционного отношения η и линейного коэффициента корреляции r используется для оценки формы связи. [4]

Выше отмечалось, что посредством теоретического корреляционного отношения измеряется теснота связи любой формы, а с помощью линейного коэффициента корреляции – только прямолинейной. Следовательно, значения η и r совпадают только при наличии прямолинейной связи. Несовпадение этих величин свидетельствует, что связь между изучаемыми признаками не прямолинейная, а криволинейная. Установлено, что если разность квадратов η и r не превышает 0,1 , то гипотезу о прямолинейной форме связи можно считать подтвержденной. В моем случае наблюдается примерное совпадение линейного коэффициента детерминации и теоретического корреляционного отношения, что дает мне основание считать связь между капиталом банков и их работающими активами прямолинейной.

При линейной однофакторной связи t-критерий можно рассчитать по формуле:

,

где (n - 2) – число степеней свободы при заданном уровне значимости α и объеме выборки n.

Так, для коэффициента корреляции между капиталом и работающими активами получается:

Если сравнить полученное tрасч с критическим значением из таблицы Стьюдента, где ν=30, а α=0,01 (tтабл=2,750), то полученное значение t-критерия будет больше табличного, что свидетельствует о значимости коэффициента корреляции и существенной связи между капиталом и работающими активами.

Таким образом, построенная регрессионная модель ŷ=245,75+1,42x в целом адекватна, и выводы, полученные по результатам малой выборки можно с достаточной вероятностью распространить на всю гипотетическую генеральную совокупность.

**Экономическая интерпретация параметров регрессии**

После проверки адекватности, установления точности и надежности построенной модели (уравнения регрессии), ее необходимо проанализировать. Прежде всего, нужно проверить, согласуются ли знаки параметров с теоретическими представлениями и соображениями о направлении влияния признака-фактора на результативный признак (показатель).

В рассмотренном уравнении ŷ=245,75+1,42х, характеризующем зависимость размера *работающих активов* (у) от *капиталов* банков (х), параметр а1>0. Следовательно, с возрастанием размера капитала банка размер работающих активов увеличивается.

Из уравнения следует, что возрастание капитала банка на 1 млн рублей приводит к увеличению работающих активов в среднем на 1,4 млн рублей (величину параметра а1).

Для удобства интерпретации параметра **a1** используют коэффициент эластичности. Он показывает средние изменения результативного признака при изменении факторного признака на 1% и вычисляется по формуле, %:

.

В представленном анализе деятельности банков эта величина равна:



Это означает, что с увеличением размера капитала на 1% следует ожидать повышения размера работающих активов банков в среднем на 0,78%.

Этот вывод справедлив только для данной совокупности банков при конкретных условиях их деятельности.

Если же эти банки и условия считать типичными, то коэффициент регрессии может быть применен для расчета размера работающих активов по их капиталу и для других банков.

Имеет смысл вычислить остатки *εi = y – ŷ*, характеризующие отклонение i-х наблюдений от значений, которые следует ожидать в среднем.

Анализируя остатки, можно сделать ряд выводов о деятельности банков. Значения остатков (Таблица 1, графа 8, Приложение 1) имеют как положительные, так и отрицательные отклонения от ожидаемого. Таким образом, выявляются банки, которые вкладывают больше денежных средств в оборот (положительные значения), и банки, предпочитающие пускать в оборот небольшую часть своих денежных средств (отрицательные значения остатков).

В итоге положительные отклонения размеров работающих активов уравновешиваются отрицательными значениями, то есть получается ***∑εi=0***.

Таким образом, в данной работе я установил корреляционную зависимость показателей 32 российских банков, провел регрессионный анализ и нашёл регрессионную модель данной взаимосвязи показателей.

Полученное уравнение ŷ=245,75+1,42х позволяет проиллюстрировать зависимость размера работающих активов банков от размера их капитала.

А также я проверил мою модель на адекватность по критерию Стьюдента, результат оказался положительным (модель адекватна, т.е. ее можно применять), а затем дал экономическую оценку этой модели.

И так, с помощью корреляционно-регрессионного анализа, я исследовал показатели банков.

**3.Применение регрессионного анализа для изучения объекта исследования**

На основе ранжированных данных о производительности труда и стаже работы двадцати рабочих бригады ЗАО «Роспан Интернешнл» (Таблица 2, Приложение 3) необходимо:

1.Установить результативный и факторный признаки.

2.Определить наличие и форму корреляционной связи между производительностью труда рабочих бригады и стажем работы.

3.Построить на графике поле корреляции и эмпирическую линию корреляционной связи.

4.Построить регрессионную модель парной корреляционной зависимости и определить её параметры.

5.Построить на графике теоретическую кривую корреляционной зависимости.

6.Рассчитать показатели тесноты связи между выработкой рабочего и стажем работы. Дать качественную оценку степени тесноты связи.

7.Оценить существенность параметров регрессивной модели и показателей тесноты связи. Дать оценку надёжности уравнения регрессии.

8.Дать экспериментальную интерпретацию параметров построенной регрессионной модели.

9.На основании регрессионной модели парной зависимости указать доверительные границы, в которых будет находиться прогнозное значение уровня производительности труда рабочего бригады, если стаж его работы составит 10,5 лет при уровне доверительной вероятности 95%.

**Решение:**

Установим результативный и факторный признаки: результативный признак (y) - выработка, факторный (x) - стаж работы, лет.

Определим наличие и форму корреляционной связи между производительностью труда рабочих бригады и стажем работы. Так как увеличение значений признака-фактора влечёт за собой увеличение величины результативного признака. То можно предположить наличие прямой корреляционной связи между выработкой и стажем работы. Проведём группировку работников бригады по признаку-фактору - стажу работы. Результаты оформим в Таблицу 2 (Приложение 3). Сравнив средние значения результативного признака по группам, можно сделать вывод о наличии связи между выработкой и стажем работы. Причём она будет являться прямой, так как рост значений признака фактора влечёт рост средних значений признака результата.

Построим поле корреляции.

Рисунок 1. Поле корреляции

х

8 9 10 11 12

1200

1100

1000

900

800

700

Теоретическая кривая корреляционной зависимости

Построим регрессионную модель парной корреляционной зависимости и определим её параметры:  - уравнение парной линейной корреляционной зависимости (регрессионная модель).

→,

→

Найдём среднее произведение факторного и результативного признака по формуле:

; .

Рассчитаем средние значение факторного и результативного признака:

факторного по формуле:

; .

результативного, по формуле:

; .

Подставим значения результативного и факторного признака в уравнение парной линейной корреляционной зависимости получим регрессионную модель парной корреляционной зависимости: - регрессионная модель зависимости выработки от стажа работы.



; .



5. Построим на графике теоретическую кривую корреляционной зависимости.

6. Рассчитаем показатели тесноты связи между выработкой рабочего и стажем работы. Для прямолинейных зависимостей измерителем тесноты связи между признаками является коэффициент парной корреляции, который рассчитывается по формуле: .

Для расчёта коэффициента парной корреляции рассчитаем среднее квадратическое отклонение факторного и результативного признака:

результативного признака, по формуле:

;  (штук)

факторного признака, по формуле:

;  (лет)

Подставим полученные значения в формулу: , рассчитаем показатель тесноты связи:



Дадим качественную оценку степени тесноты связи. Для этого рассчитаем коэффициент детерминации, который показывает какая часть общей вариации результативного признака (y) объясняется влиянием изучаемого фактора (x).

; .

На основе шкалы Чеддока можно сделать вывод о том, что между выработкой т стажем работы существует прямая высокая связь.64% изменения выработки обусловлено изменением стажа работы рабочих.

7. Оценим существенность параметров регрессионной модели и показателей тесноты связи и дадим оценку надёжности уравнения регрессии.

Значимость параметров простой линейной регрессии осуществляется с помощью t-критерия Стьюдента. Рассчитаем значения t-критерия Стьюдента для параметра a0 и a1: для параметра а0, по формуле: . Для этого рассчитаем средне квадратическое отклонение результативного признака у от выровненных значений уx по формуле:

, , 

для параметра a1 по формуле:

, 

Для оценки значимости линейного коэффициента корреляции r применяется t-критерий Стьюдента. При этом определяется фактическое (расчетное) значение критерия (trф). Рассчитаем это значение по формуле:

, 

Для всей совокупности наблюдаемых значений рассчитаем среднюю квадратическую ошибку уравнения регрессии по формуле:

,  (штук).

Так как <, то уравнение регрессии целесообразно и может быть использовано в дальнейшем статистическом анализе.

81,98 < 133,8423.

Так как  (фактическое) >  (критическое), то значение параметра  признаётся существенным, то есть оно не является результатом стечения случайных обстоятельств.

Так как > , то  также признаётся существенным.

Так как > , то связь между произвольностью труда и стажем работы признаётся существенной.

8. Дадим экспериментальную интерпретацию параметров построенной регрессионной модели. Так как коэффициент регрессии  > 0, то это подтверждает теоретические представления о прямой зависимости между выработкой и стажем работы. Значение = 83,84 шт. можно интерпретировать так: при увеличении стажа на 1 год выработка увеличивается на 83,84 шт.

Рассчитаем коэффициент эластичности, который показывает среднее изменение результативного признака при изменении факторного признака на 1%:

, %.

То есть при увеличении стажа на 1% их выработка увеличивается на 0,88%.

9. Укажем доверительные границы, в которых будет находиться прогнозное значение уровня производительности труда рабочего бригады, если стаж его работы составит 10,5 лет при уровне доверительной вероятности 95% по формуле:

****

 штук





Таким образом, с вероятностью 95% можно ожидать, что при стаже работы работника 10,5 лет составит не менее 956 штук и не более 1040 штук.

**ЗАКЛЮЧЕНИЕ**

В ходе написания курсовой работы мной были раскрыты поставленные задачи.

В теоретической части работы были изучены статистические взаимосвязи социально-экономических явлений и процессов. Описаны характеристики регрессионного анализа, выполнена оценка взаимосвязи между факторным и результативным признаком на основе регрессионного анализа, отмечены факторные признаки для построения множественной регрессионной модели, произведена проверка адекватности модели, построенной на основе уравнений регрессии.

В расчетной части было продемонстрировано применение регрессионного анализа на конкретном примере.

**СПИСОК ЛИТЕРАТУРЫ**

1. Аверкин А.Н., Батыршин И.З., Блишун А.Ф. и др. Нечеткие множества в моделях управления и искусственного интеллекта // Под ред. Д.А. Поспелова. – М.: Наука, 1986. – 312 с.
2. Аветисян Д.О. Проблемы информационного поиска: (Эффективность, автоматическое кодирование, поисковые стратегии) - М.: Финансы и статистика, 1981. - 207 с.
3. Айвазян С.А., Бежаева З.И., Староверов О.В. Классификация многомерных наблюдений. – М.: Статистика, 1974. – 240 с.
4. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика. Основы моделирования и первичная обработка данных. Справочное издание. – М.: Финансы и статистика, 1983. – 472 с.
5. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика: Исследование зависимостей: Справочник. – М.: Финансы и статистика, 1985. – 182с.
6. Айвазян С.А. , Мхитарян В.С. Прикладная статистика и основы эконометрики. – М. Юнити, 1998. – 1024 с.
7. Ван дер Варден Б.Л. Математическая статистика. – М.: Изд-во иностр. лит., 1960. – 302 с.
8. Гайдышев И.П. Анализ и обработка данных: специальный справочник. - СПб.: Питер, 2001. - 752 с.
9. Гмурман В.С. Теория вероятностей и математическая статистика. – М.: Высш. шк., 1972. – 368 с.
10. Калинина В.Н., Панкин В.Ф. Математическая статистика. – М.: Высш. шк., 2001. – 336 с.
11. Кендалл М., Стьюарт А. Теория распределений. – М.: Наука, 1966. – 566 с.
12. Кендалл М., Стьюарт А. Статистические выводы и связи. – М .: Наука, 1973. – 899 с.

Приложение 1

Таблица 1

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Банк | Капитал (млн.руб.) **x** | Раб./риск. активы (млн.руб.) **y** | **x2** | **y2** | **xy** | **ŷ** | **y-ŷ** | **(y-ŷ)2** | **y-y** | **(y-y)2** | **ŷ-у** | **(ŷ-у)2** |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
| Славянский банк | 936 | 1545 | 876096 | 2387025 | 1446120 | 1571,40 | 26,40 | 697,04 | 454,03 | 206143,24 | 480,43 | 230814,48 |
| Локо-Банк | 877 | 1758 | 769129 | 3090564 | 1541766 | 1487,84 | -270,16 | 72986,25 | 667,03 | 444929,02 | 396,87 | 157506,06 |
| Союзобщемаш-банк | 833 | 1075 | 693889 | 1155625 | 895475 | 1425,52 | 350,52 | 122866,71 | -15,97 | 255,04 | 334,55 | 111926,03 |
| БВТ | 823 | 1369 | 677329 | 1874161 | 1126687 | 1411,36 | 42,36 | 1794,42 | 278,03 | 77300,68 | 320,39 | 102650,11 |
| Финпромбанк | 805 | 966 | 648025 | 933156 | 777630 | 1385,87 | 419,87 | 176288,55 | -124,97 | 15617,50 | 294,90 | 86964,42 |
| Московско-Парижский | 750 | 1005 | 562500 | 1010025 | 753750 | 1307,97 | 302,97 | 91791,57 | -85,97 | 7390,84 | 217,00 | 47089,54 |
| Оптбанк | 748 | 1590 | 559504 | 2528100 | 1189320 | 1305,14 | -284,86 | 81145,98 | 499,03 | 249030,94 | 214,17 | 45868,21 |
| Ми-Банк | 730 | 1620 | 532900 | 2624400 | 1182600 | 1279,65 | -340,35 | 115841,25 | 529,03 | 279872,74 | 188,68 | 35598,41 |
| Интурбанк | 703 | 1423 | 494209 | 2024929 | 1000369 | 1241,41 | -181,59 | 32976,56 | 332,03 | 110243,92 | 150,44 | 22630,84 |
| БРП | 615 | 906 | 378225 | 820836 | 557190 | 1116,77 | 210,77 | 44424,76 | -184,97 | 34213,90 | 25,80 | 665,73 |
| Алеф-Банк | 613 | 817 | 375769 | 667489 | 500821 | 1113,94 | 296,94 | 88172,91 | -273,97 | 75059,56 | 22,97 | 527,59 |
| "Аверс" | 607 | 780 | 368449 | 608400 | 473460 | 1105,44 | 325,44 | 105912,16 | -310,97 | 96702,34 | 14,47 | 209,42 |
| "Первомайский" | 603 | 1277 | 363609 | 1630729 | 770031 | 1099,78 | -177,22 | 31408,23 | 186,03 | 34607,16 | 8,81 | 77,55 |
| Русский Банкирский Дом | 586 | 1426 | 343396 | 2033476 | 835636 | 1075,70 | -350,30 | 122710,54 | 335,03 | 112245,10 | -15,27 | 233,19 |
| "Электроника" | 570 | 1410 | 324900 | 1988100 | 803700 | 1053,04 | -356,96 | 127421,38 | 319,03 | 101780,14 | -37,93 | 1438,79 |
| Первый Республиканский | 551 | 1161 | 303601 | 1347921 | 639711 | 1026,13 | -134,87 | 18190,15 | 70,03 | 4904,20 | -64,84 | 4204,34 |
| "Снежинский" | 546 | 1208 | 298116 | 1459264 | 659568 | 1019,05 | -188,95 | 35702,98 | 117,03 | 13696,02 | -71,92 | 5172,82 |
| Национальный Банк Развития | 543 | 1355 | 294849 | 1836025 | 735765 | 1014,80 | -340,20 | 115736,86 | 264,03 | 69711,84 | -76,17 | 5802,05 |
| Меритбанк | 526 | 872 | 276676 | 760384 | 458672 | 990,72 | 118,72 | 14094,87 | -218,97 | 47947,86 | -100,25 | 10049,70 |
| ВКАБанк | 518 | 736 | 268324 | 541696 | 381248 | 979,39 | 243,39 | 59239,42 | -354,97 | 126003,70 | -111,58 | 12449,76 |
| Ланта-Банк | 511 | 1293 | 261121 | 1671849 | 660723 | 969,48 | -323,52 | 104666,84 | 202,03 | 40816,12 | -121,49 | 14760,44 |
| "Транснациональ-ный" | 510 | 722 | 260100 | 521284 | 368220 | 968,06 | 246,06 | 60546,09 | -368,97 | 136138,86 | -122,91 | 15106,58 |
| "Адмиралтейский" | 510 | 678 | 260100 | 459684 | 345780 | 968,06 | 290,06 | 84135,48 | -412,97 | 170544,22 | -122,91 | 15106,58 |
| Центральное ОВК | 506 | 1072 | 256036 | 1149184 | 542432 | 962,40 | -109,60 | 12013,04 | -18,97 | 359,86 | -128,57 | 16531,28 |
| Российский Промышленный | 504 | 1209 | 254016 | 1461681 | 609336 | 959,56 | -249,44 | 62218,61 | 118,03 | 13931,08 | -131,41 | 17267,69 |
| ПРОДОЛЖЕНИЕ ТАБЛИЦЫ 1  |
| "Смоленский" | 490 | 1001 | 240100 | 1002001 | 490490 | 939,74 | -61,26 | 3753,36 | -89,97 | 8094,60 | -151,23 | 22871,93 |
| АПР-Банк | 459 | 1268 | 210681 | 1607824 | 582012 | 895,83 | -372,17 | 138510,31 | 177,03 | 31339,62 | -195,14 | 38079,52 |
| СудКомБанк | 448 | 817 | 200704 | 667489 | 366016 | 880,25 | 63,25 | 4000,70 | -273,97 | 75059,56 | -210,72 | 44402,47 |
| "Военный" | 440 | 665 | 193600 | 442225 | 292600 | 868,92 | 203,92 | 41583,66 | -425,97 | 181450,44 | -222,05 | 49305,88 |
| "Золото-Платина" | 425 | 743 | 180625 | 552049 | 315775 | 847,68 | 104,68 | 10957,14 | -347,97 | 121083,12 | -243,29 | 59191,81 |
| "Андреевский" | 410 | 618 | 168100 | 381924 | 253380 | 826,43 | 208,43 | 43443,88 | -472,97 | 223700,62 | -264,54 | 69980,38 |
| Народный Банк Сбережений | 401 | 526 | 160801 | 276676 | 210926 | 813,69 | 287,69 | 82762,85 | -564,97 | 319191,10 | -277,28 | 76886,79 |
| **Сумма** | **19097** | **34911** | **12055479** | **41516175** | **21767209** | **34911** | **0** | **2107994,57** | **9,96** | **3429375,97** | **11,96** | **1321383,40** |
| **Средние***(сумма/**кол-во банков)* | **596,78** | **1090,97** | **376733,72** | **1297380,47** | **680225,28** |  |  |  |  |  |  |  |

Приложение 2

График 1

ПРИЛОЖЕНИЕ 3

Таблица 2

Расчётная таблица

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |  |
| 8 | 800 | 6400 | 640000 | 64 | 789,02 | -1,95 | 3,8025 | -152,5 | 23256,25 | 10,98 | 120,56 |
| 8 | 850 | 6800 | 722500 | 64 | 789,02 | -1,95 | 3,8025 | -102,5 | 10506,25 | 60,98 | 3718,56 |
| 8 | 720 | 5760 | 518400 | 64 | 789,02 | -1,95 | 3,8025 | -232,5 | 54056,25 | -69,02 | 4763,76 |
| 9 | 850 | 7650 | 722500 | 81 | 872,86 | -0,95 | 0,9025 | -102,5 | 10506,25 | -22,86 | 622,57 |
| 9 | 800 | 7200 | 640000 | 81 | 872,86 | -0,95 | 0,9025 | -152,5 | 23256,3 | -72,86 | 5308,57 |
| 9 | 880 | 7920 | 774400 | 81 | 872,86 | -0,95 | 0,9025 | -72,5 | 5256,25 | 7,14 | 50,98 |
| 9 | 950 | 8550 | 902500 | 81 | 872,86 | -0,95 | 0,9025 | -2,5 | 6,25 | 77,14 | 5950,57 |
| 9 | 820 | 7380 | 672400 | 81 | 872,86 | -0,95 | 0,9025 | -132,5 | 17556,25 | -52,86 | 2794,17 |
| 10 | 900 | 9000 | 810000 | 100 | 956,7 | 0,05 | 0,0025 | -52,5 | 2756,25 | -56,7 | 3114,89 |
| 10 | 1000 | 10000 | 1000000 | 100 | 956,7 | 0,05 | 0,0025 | 47,5 | 2256,25 | 43,3 | 1874,89 |
| ПРОДОЛЖЕНИЕ ТАБЛИЦЫ 2 |
| 10 | 920 | 9200 | 846400 | 100 | 956,7 | 0,05 | 0,0025 | -32,5 | 1056,25 | -36,7 | 1346,89 |
| 10 | 1060 | 10600 | 1123600 | 100 | 956,7 | 0,05 | 0,0025 | 107,5 | 11556,25 | 103,3 | 10670,89 |
| 10 | 950 | 9500 | 902500 | 100 | 956,7 | 0,05 | 0,0025 | -2,5 | 6,25 | -6,7 | 44,89 |
| 11 | 900 | 9900 | 810000 | 121 | 1040,54 | 1,05 | 1,1025 | -52,5 | 2756,25 | -140,54 | 975,15 |
| 11 | 1200 | 13200 | 1440000 | 121 | 1040,54 | 1,05 | 1,1025 | 247,5 | 61256,25 | 159,46 | 25421, 19 |
| 11 | 1150 | 12650 | 1322500 | 121 | 1040,54 | 1,05 | 1,1025 | 197,5 | 39006,5 | 109,46 | 11981,49 |
| 11 | 1000 | 11000 | 1000000 | 121 | 1040,54 | 1,05 | 1,1025 | 47,5 | 2256,25 | -40,54 | 1643,49 |
| 12 | 1200 | 14400 | 1440000 | 144 | 1124,38 | 2,05 | 4, 2025 | 247,5 | 6156,25 | 75,62 | 5718,38 |
| 12 | 1100 | 13200 | 1210000 | 144 | 1124,38 | 2,05 | 4, 2025 | 147,5 | 21756,25 | -24,38 | 594,38 |
| 12 | 1000 | 12000 | 1000000 | 144 | 1124,38 | 2,05 | 4, 2025 | 47,5 | 2256,25 | -124,38 | 5470,38 |
| **199** | **19050** | **192310** | **18497700** | **2013** | **19050,16** | **0** | **32,95** | **0** | **358275** | **0** | **120969,33** |