

ОПРЕДЕЛЕНИЕ, ПРЕДМЕТ И ЗАДАЧИ ЭКОНОМЕТРИКИ. ЭТАПЫ ЭКОНОМЕТРИЧЕСКОГО ИССЛЕДОВАНИЯ

Слово «эконометрика» представляет собой комбинацию двух слов: «экономика» и «метрика» (от греч. «метрон» – правило определения расстояния между двумя точками в пространстве). Формально «эконометрика» означает «измерения в экономике». Однако область исследований данной дисциплины гораздо шире. *Эконометрика — это наука, в которой на базе реальных статистических данных строятся, анализируются и совершенствуются математические модели реальных экономических явлений.* Эконометрика позволяет найти количественное подтверждение либо опровержение того или иного экономического закона либо гипотезы.

Таким образом, *эконометрика – наука, которая дает количественное выражение взаимосвязей экономических явлений и процессов.*

Зарождение эконометрики является следствием междисциплинарного подхода к изучению экономики. Эконометрика представляет собой комбинацию трех областей знания:

- Экономической теории
- Статистики
- Математики

Большинство эконометрических методов и приемов заимствовано из математической статистики. Однако методы математической статистики универсальны и не учитывают специфики экономических данных, которая заключается в следующем:

- 1) данные не являются результатом контролируемого эксперимента;
- 2) невозможность проводить многократные эксперименты (из-за изменения внешних условий);
- 3) экономические данные часто содержат ошибки измерения. В эконометрике разрабатываются специальные методы анализа, позволяющие, если не устранить, то, по крайней мере, снизить влияние этих ошибок на полученные результаты.

Эти особенности рождают ряд специфических проблем, решение которых не входит в математическую статистику.

Таким образом, *эконометрика связывает между собой экономическую теорию и экономическую статистику и с помощью математико-статистических методов придает конкретное количественное выражение общим закономерностям, устанавливаемым экономической теорией.*

Предмет исследования эконометрики как науки – экономические явления. Но в отличие от экономической теории *эконометрика делает упор на количественные, а не на качественные аспекты* этих явлений. Например, экономическая теория утверждает, что спрос на товар с ростом его цены убывает. Но при этом практически неисследованным остается вопрос, как быстро и по какому закону происходит это убывание для определенного товара. Эконометрика отвечает на этот вопрос для каждого конкретного случая.

Основные задачи эконометрики:

1. *Построение эконометрических моделей*, т.е. представление экономических моделей в математической форме, удобной для проведения эмпирического анализа.
2. *Оценка параметров построенной модели*, делающих выбранную модель наиболее адекватной реальным данным.
3. *Проверка качества найденных параметров модели и самой модели в целом.*

4. *Использование построенных моделей* для объяснения поведения исследуемых экономических показателей, прогнозирования и предсказания, а также для осмысленного проведения экономической политики

Таким образом, **стандартная схема анализа** зависимостей состоит в осуществлении ряда последовательных процедур:

1. *Подбор начальной модели (этап спецификации)*. Он осуществляется на основе экономической теории, предыдущих знаний об объекте исследования, опыта исследователя и его интуиции.
2. *Оценка параметров модели* на основе имеющихся статистических данных (*этап параметризации*).
3. Осуществление проверки качества модели (*этап верификации*).
4. При наличии хотя бы одного неудовлетворительного ответа по какому-либо критерию модель совершенствуется с целью устранения выявленного недостатка.
5. При положительных ответах по всем критериям модель считается качественной. Она используется для анализа и прогноза объясняемой переменной.

Однако необходимо предостеречь от абсолютизации полученного результата, поскольку даже качественная модель является подгонкой спецификации модели под имеющийся набор данных. Поэтому вполне реальна картина, когда исследователи, обладающие разными наборами данных, строят разные модели для объяснения одного и того же явления. Проблематичным является и использование модели для прогнозирования значений объясняемой переменной. Иногда хорошие с точки зрения критериев проверки модели обладают весьма низкими прогнозными качествами.

Стандартную схему анализа иллюстрирует следующая блок-схема (Рис. 1):



Рис. 1. *Схема анализа*

ЭКОНОМЕТРИЧЕСКИЕ МОДЕЛИ

Многообразие и сложность экономических процессов предопределяет многообразие моделей, используемых для эконометрического анализа. Правильный выбор вида экономической модели является отправной точкой для качественного ее анализа. Безусловно, на практике неизвестно, какая модель является верной, и зачастую подбирают такую модель, которая наиболее точно соответствует реальным данным. При этом необходимо учитывать, что идеальной модели не существует.

Признаки «хорошей» модели:

1. *Скупость (простота).* Модель должна быть максимально простой. Данное свойство определяется тем фактом, что модель не отражает действительность идеально, а является ее упрощением. Поэтому из двух моделей, приблизительно одинаково отражающих реальность, предпочтение отдается модели, содержащей меньшее число объясняющих переменных.
2. *Единственность.* Для любого набора статистических данных определяемые коэффициенты должны вычисляться однозначно.
3. *Максимальное соответствие.* Уравнение тем лучше, чем большую часть разброса зависимой переменной оно может объяснить.
4. *Согласованность с теорией.* Никакое уравнение не может быть признано качественным, если оно не соответствует известным теоретическим предпосылкам. Другими словами, модель обязательно должна опираться на теоретический фундамент, так как в противном случае результат использования регрессионного уравнения может быть весьма плачевным.
5. *Прогнозные качества.* Модель может быть признана качественной, если полученные на ее основе прогнозы подтверждаются реальностью.

Поскольку не существует какого-либо единого правила построения моделей, анализ перечисленных свойств позволяет строить более качественные эконометрические модели.

Одно из главных направлений эконометрического анализа – постоянное совершенствование моделей. Здесь следует отметить, что какого-то глобального подхода, определяющего заранее возможные пути совершенствования, нет и, скорее всего, быть не может. Исследователь должен помнить, что совершенной модели не существует. В силу постоянно изменяющихся условий протекания экономических процессов не может быть и постоянно качественных моделей. Новые условия требуют пересмотра даже весьма устойчивых моделей.

До сих пор достаточно спорным является вопрос, как строить модели: начинать с самой простой и постоянно усложнять ее или начинать с максимально сложной модели и упрощать ее на основе проводимых исследований.

Выделяют *три основных класса моделей.*

I. Регрессионные модели с одним уравнением

- ✓ Линейные
- ✓ Нелинейные

II. Модели временных рядов, полученные с помощью следующих методов

- ✓ Экспоненциального сглаживания
- ✓ Сезонной декомпозиции
- ✓ Авторегрессии
- ✓ ARIMA и др.

III. Системы одновременных уравнений

Пример. Модель спроса и предложения описывается следующей системой уравнений:

$$\begin{cases} Q_t^S = \alpha_1 + \alpha_2 P_t + \alpha_3 P_{t-1} + \varepsilon_t & (\text{спрос}) \\ Q_t^D = \beta_1 + \beta_2 P_t + \beta_3 Y_t + u_t & (\text{предложение}) \\ Q_t^S = Q_t^D & (\text{равновесие}) \end{cases}$$

Здесь: Q_t^S - предложение товара в момент времени t ; Q_t^D - спрос на товар в момент времени t ; P_t - цена товара в момент времени t ; P_{t-1} - цена товара в момент времени $t-1$; Y_t - доход потребителей в момент времени t ; α_i, β_i - параметры ($i=1,2,3$); ε_t, u_t - случайная компонента (отклонение).

Классификация задач, решаемых с помощью эконометрической модели:

I. По конечным прикладным целям

- ✓ прогноз экономических и социально-экономических показателей, характеризующих состояние и развитие анализируемой системы;
- ✓ имитация возможных сценариев социально-экономического развития системы для выявления того, как планируемые изменения тех или иных поддающихся управлению параметров скажутся на выходных характеристиках.

II. По уровню иерархии выделяют задачи, решаемые на:

- ✓ макроуровне (страна в целом);
- ✓ мезоуровне (уровне регионов, отраслей, корпораций);
- ✓ микроуровне (на уровне семьи, предприятия, фирмы).

III. По профилю анализируемой экономической системы выделяют задачи, направленные на решение проблем:

- ✓ рынка;
- ✓ инвестиционной, финансовой или социальной политики;
- ✓ ценообразования;
- ✓ распределительных отношений;
- ✓ спроса и потребления;
- ✓ на определенный комплекс проблем. Однако, чем шире комплекс проблем, тем меньше шансов провести эконометрическое исследование достаточно эффективно.

ТИПЫ ДАННЫХ. СУТЬ РЕГРЕССИОННОГО АНАЛИЗА

ИЗМЕРЕНИЯ В ЭКОНОМИКЕ

Для социально-экономических измерений характерны специфические представления о точности. Экономике относят к «неточным» наукам, так как невозможно произвести измерение с произвольно малой погрешностью.

Задача определения погрешности эксперимента далеко не проста. Основная трудность заключается в том, что как *при непосредственном измерении, так и при моделировании* необходим учет разнообразных факторов, в той или иной степени влияющих на результат. В любом конкретном эксперименте проанализировать или хотя бы указать все факторы, воздействующие на результат, невозможно. Поэтому истинное значение погрешности экспериментальной величины на любом этапе развития техники и

методики эксперимента остается неизвестным. Отсюда задачей теории может быть только максимально достоверная оценка погрешности. Степень достоверности этой оценки зависит прежде всего от того, насколько в данном конкретном эксперименте учтены факторы, влияющие на результат измерения или моделирования.

Для облегчения анализа эксперимента его ошибки целесообразно классифицировать в соответствии с причинами, которые их вызывают. В эконометрике выделяют *две категории ошибок*:

I. *Систематические ошибки* – погрешности, которые, практически не изменяясь за время опыта, одинаковым образом входят в каждый конечный результат эксперимента, вызывая смещение его в одну какую-либо сторону. Источниками систематических ошибок могут быть:

- ✓ инструментальные ошибки, возникающие из-за дефектов или неисправностей измерительной аппаратуры;
- ✓ ошибки, связанные с состоянием внешней среды, в которой производятся измерения;
- ✓ ошибки, обусловленные индивидуальными особенностями экспериментатора (субъективные или личные ошибки);
- ✓ ошибки, вносимые самим методом постановки эксперимента из-за приближенности теоретических соотношений, связывающих наблюдаемые на опыте величины с величинами, непосредственно интересующими экспериментатора.

Поскольку систематические ошибки определяются методом постановки эксперимента, то какой-либо общей теории этих ошибок не может существовать. Тем не менее если источник систематической ошибки известен, то в принципе можно учесть ее влияние на результат эксперимента, а в ряде случаев полностью или частично исключить либо устранив источник, ее вызывающий, либо внося поправку, приближенно учитывающую ее влияние.

II. *Случайные ошибки* связывают с факторами, претерпевающими незначительные изменения за время эксперимента.

Поскольку исход каждого отдельного наблюдения зависит от действий большого числа разнообразных факторов, меняющихся за время эконометрического эксперимента, то его результат естественно считать зависимым от случая, т. е. случайной величиной. Следовательно, и ошибку эксперимента, вызванную действием таких факторов, можно рассматривать как случайную величину. Такой подход дает возможность трактовать ошибку как величину, управляемую вероятностными законами, и применять для ее расчета теорию вероятностей. Поэтому случайную ошибку называют иногда также статистической.

Приведенная классификация ошибок на систематические и случайные чисто условна и справедлива только для конкретного эксперимента. Одна и та же ошибка в одном множестве наблюдений выступает как систематическая, а в другом — как случайная. Отсюда следует важный вывод о том, что результат эксперимента всегда можно оценить статистически. Для этого необходимо спланировать эксперимент таким образом, чтобы фактор, дающий систематическую ошибку, действовал случайным образом.

ТИПЫ ДАННЫХ

При моделировании экономических процессов используют следующие типы данных:

- ✓ *пространственные данные*

Пространственными данными является набор сведений по разным объектам, взятым за один и тот же период или момент времени. Например, набор сведений по разным

фирмам (объем производства, численность работников, размер основных производственных фондов и пр.).

✓ *временные данные*

Временными данными является набор сведений, характеризующий один и тот же объект, но за разные периоды или моменты времени. Например, ежеквартальные данные о средней заработной плате, индексе потребительских цен, числе занятых за последние годы, ежедневный курс доллара США. Отличительной особенностью временных данных является то, что они естественным образом упорядочены по времени.

Набор сведений представляет собой множество признаков, характеризующих объект исследования. Признаки являются взаимосвязанными, причем в этой взаимосвязи они могут выступить в одной из двух ролей:

- ✓ в роли *результативного признака* (зависимая переменная, отклик – аналог зависимой переменной y в математике);
- ✓ в роли *факторного признака*, значения которого определяют значение отклика (независимые переменные, факторы – аналог независимой переменной x в математике).

Типы переменных, участвующих в эконометрической модели:

- ✓ *экзогенные* (независимые) – значения которых задаются извне, автономно, в определенной степени они являются управляемыми (планируемыми) (x)
- ✓ *эндогенные* (зависимые) — значения которых определяются внутри модели, или взаимозависимые (y);
- ✓ *лаговые* — экзогенные или эндогенные переменные эконометрической модели, датированные предыдущими моментами времени и находящиеся в уравнении с текущими переменными. Например: y_t — текущая эндогенная переменная, y_{t-1} — лаговая эндогенная переменная, y_{t-2} — тоже лаговая эндогенная переменная;
- ✓ *предопределенные* переменные (объясняющие переменные). К ним относятся лаговые и текущие экзогенные переменные (x_t, x_{t-1}), а также лаговые эндогенные переменные (y_{t-1}).

Любая эконометрическая модель предназначена для объяснения значений текущих эндогенных переменных (одной или нескольких) в зависимости от значений предопределенных переменных.

ВЗАИМОСВЯЗИ ЭКОНОМИЧЕСКИХ ПЕРЕМЕННЫХ

Поведение и значение любого экономического показателя зависят практически от бесконечного числа факторов, и все учесть нереально. Но в этом и нет необходимости. Обычно лишь ограниченное количество факторов действительно существенно воздействует на исследуемый экономический показатель. Доля влияния остальных факторов столь незначительна, что их игнорирование не может привести к существенным отклонениям в поведении исследуемого объекта. Выделение и учет в модели лишь ограниченного числа реально доминирующих факторов и является серьезной предпосылкой для качественного анализа, прогнозирования и управления ситуацией. Экономическая теория выявила и исследовала значительное число устоявшихся и стабильных связей между различными показателями.

Однако в реальных ситуациях даже устоявшиеся зависимости могут проявляться по-разному. Еще более сложной является задача анализа малоизученных и нестабильных

зависимостей. Здесь следует отметить, что такие экономические модели невозможно строить, проверять и совершенствовать без статистического анализа входящих в них переменных с использованием реальных статистических данных. Инструментарием такого анализа являются методы статистики и эконометрики, в частности регрессионного и корреляционного анализа.

В естественных науках большей частью имеют дело со строгими (функциональными) зависимостями, при которых каждому значению одной переменной соответствует единственное значение другой. Однако в подавляющем большинстве случаев между экономическими переменными таких зависимостей нет. Например, нет строгой зависимости между доходом и потреблением, ценой и спросом, производительностью труда и стажем работы и т.д. Это связано с целым рядом причин:

- при анализе влияния одной переменной на другую не учитывается целый ряд других факторов, влияющих на нее;
- это влияние может быть не прямым, а проявляться через цепочку других факторов;
- многие такие воздействия носят случайный характер и т.д.

Поэтому в экономике говорят не о функциональных, а о *корреляционных* зависимостях.

Корреляционная зависимость проявляется в том, что при изменении одной из величин изменяется среднее значение другой.

СУТЬ РЕГРЕССИОННОГО АНАЛИЗА

Рассмотрим некоторый экономический объект (процесс, явление, систему) и выделим только две переменные, характеризующие объект. Обозначим переменные буквами y и x . Будем предполагать, что фактор x оказывает воздействие на отклик y , т.е. имеет место зависимость $y = f(x)$.

Эту зависимость можно рассматривать с целью установления самого факта наличия или отсутствия значимой связи между y и x , можно преследовать цель прогнозирования неизвестных значений y по известным значениям x , наконец, возможно выявление причинно-следственных связей между x и y .

При изучении взаимосвязи между переменными y и x следует, прежде всего, установить тип зависимости (природу анализируемых переменных y и x). Возможны следующие ситуации:

- y и x являются неслучайными переменными, т.е. значения y строго зависят только от соответствующих значений x и полностью ими определяются. В этом случае говорят о функциональной зависимости, когда y является некоторой функцией от переменной x . Пример: $y = 4 \cdot x$.
- y является случайной переменной, а x – неслучайной. В этом случае считают, что между переменными имеет место регрессионная зависимость. То есть верна модель $y = f(x) + e$, где e - величина случайной ошибки (отклонения).
- y и x зависят от множества неконтролируемых факторов, так что являются случайными по своей сущности. В этом случае к проблемам построения конкретного вида зависимости между переменными присоединяется проблема исследования тесноты связи между ними. Речь в этом случае идет о корреляционно-регрессионной зависимости между y и x .

Предметом дальнейшего рассмотрения будут регрессионные модели вида:

$$y = f(x) + e \text{ или } y = f(x_1, x_2, \dots, x_k) + e.$$

Причины обязательного присутствия в регрессионных моделях случайной ошибки:

1. *Невключение в модель всех объясняющих переменных.*

Любая эконометрическая модель является упрощением реальной ситуации, которая всегда представляет собой сложнейшее переплетение различных факторов, многие из которых в модели не учитываются, что порождает отклонение реальных значений отклика от прогнозов. Проблема еще и в том, что никогда заранее не известно, какие факторы действительно являются определяющими, а какими можно пренебречь. В ряде случаев учесть непосредственно какой-то фактор нельзя в силу невозможности получения по нему статистических данных.

2. *Неправильный выбор функциональной формы модели.*

Из-за слабой изученности исследуемого процесса либо из-за его переменчивости может быть неверно подобрана функция, его моделирующая. Кроме того, неверным может быть подбор факторов.

3. *Агрегирование переменных.*

Во многих моделях рассматриваются зависимости между факторами, которые сами представляют сложную комбинацию других, более простых переменных.

4. *Ошибки измерений.*

Какой бы качественной ни была модель, ошибки измерений отразятся на несоответствии модельных значений эмпирическим данным, что также отразится на величине отклонений.

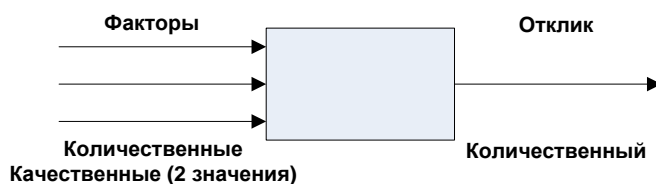
5. *Ограниченность статистических данных.*

Зачастую строятся модели, выражаемые непрерывными функциями. Но для этого используется набор данных, имеющих дискретную структуру. Это несоответствие находит свое выражение в случайном отклонении.

6. *Непредсказуемость человеческого фактора.*

Эта причина может «испортить» самую качественную модель. Действительно, при правильном выборе формы модели, скрупулезном подборе объясняющих переменных все равно невозможно спрогнозировать поведение каждого индивидуума (если речь идет об опросе респондентов).

Регрессионный анализ – статистический метод, с помощью которого можно построить модель с одной зависимой переменной (откликом) и несколькими независимыми переменными (факторами).



Регрессионный анализ позволяет решить **три задачи**:

1. Выявить, какие факторы действуют на отклик.
2. Ранжировать факторы по степени влияния на отклик.
3. Спрогнозировать значение отклика при определенных значениях факторов.

Примеры постановок реальных задач, к которым может быть применен регрессионный анализ:

- Определить, какие факторы влияют на расход электроэнергии на предприятии и построить прогноз расходов электроэнергии на ближайший квартал.

- Планируется строительство нового торгового центра. Требуется спрогнозировать «проходимость» секций будущего торгового центра с целью обоснования ставки арендной платы и оптимальной площади помещений.
- На основе риэлтерской базы данных по реализованным объектам недвижимости построить прогноз стоимости квартиры с учетом площади, удобств, типа дома и других факторов.
- Выявить факторы, определяющие долю рынка торговой марки определенных товаров.
- При покупке автомобиля требуется выбрать такую модель, которая по истечении трех лет службы на вторичном рынке незначительно потеряет в цене.

Регрессионная модель может быть:

- Линейной
- Нелинейной (экспоненциальной, логарифмической, квадратической, кубической и т.п.)

ПАРНАЯ ЛИНЕЙНАЯ РЕГРЕССИЯ

Парная линейная регрессия – регрессионная зависимость между двумя переменными y и x , т. е. модель вида $y = a + bx + e$, где y – отклик, x – фактор, e – случайная «остаточная» компонента.

Далее рассмотрим алгоритм вычисления для случая парной линейной регрессии.

1. Выбор формы модели

На этом этапе по исходным статистическим данным выбирается наиболее подходящая модель, т. е. уравнение регрессии. Этот выбор может быть осуществлен тремя способами:

- Графическим
На координатной плоскости строят точки с координатами (x, y) , по расположению которых предполагают наличие зависимости определенного вида. Такое изображение статистической зависимости называется *диаграммой рассеяния*.
- Аналитическим, т. е. исходя из теории изучаемой взаимосвязи
- Экспериментальным

2. Вычисление коэффициентов (параметров) регрессионной модели

Для оценки параметров используется *метод наименьших квадратов* (МНК), согласно которому неизвестные параметры a и b выбираются таким образом, чтобы сумма квадратов отклонений фактических значений отклика y от прогнозных (полученных по уравнению регрессии) \hat{y} была минимальна, т. е.

$$\sum (y - \hat{y})^2 \rightarrow \min .$$

Чтобы найти минимум функции, надо вычислить частные производные по каждому из параметров a и b и приравнять их к нулю. Обозначим:

$$\sum (y - \hat{y})^2 = \sum (y - a - bx)^2 = S(a, b) . \text{ Тогда,}$$

$$\begin{cases} \frac{\partial S}{\partial a} = \sum (2(y - a - bx)(-1)) = 0 \\ \frac{\partial S}{\partial b} = \sum (2(y - a - bx)(-x)) = 0 \end{cases} \Rightarrow \begin{cases} \sum (-2y + 2a + 2bx) = 0 \\ \sum (-2yx + 2ax + 2bx^2) = 0 \end{cases} \Rightarrow$$

$$\Rightarrow \begin{cases} -2 \sum y + 2na + 2b \sum x = 0 \\ -2 \sum yx + 2a \sum x + 2b \sum x^2 = 0 \end{cases} \Rightarrow \begin{cases} Na + b \sum x = \sum y \\ a \sum x + b \sum x^2 = \sum yx \end{cases}$$

Делим обе части уравнений системы на N (объем выборки). Получим:

$$\begin{cases} a + b\bar{x} = \bar{y} \\ a\bar{x} + b\bar{x}^2 = \overline{xy} \end{cases} \Rightarrow \begin{cases} a = \bar{y} - b\bar{x} \\ b = \frac{\overline{yx} - \bar{y} \cdot \bar{x}}{\bar{x}^2 - (\bar{x})^2} \end{cases} .$$

Формально a – значение y при $x = 0$. Если фактор x не имеет и не может иметь нулевого значения, то такая трактовка параметра a не имеет смысла. Этот параметр может и не иметь экономического содержания. Попытки экономически интерпретировать параметр a могут привести к абсурду, особенно при $a < 0$.

Интерпретировать можно лишь знак при параметре a . Если $a > 0$, то относительное

изменение отклика происходит медленнее, чем изменение фактора.

Коэффициент b называется **коэффициентом регрессии** и показывает среднее изменение отклика при изменении фактора на одну единицу.

3. Оценка значимости коэффициента регрессии

Используется *t*-критерий Стьюдента. Выдвигается гипотеза $H_0: b = 0$ об отсутствии влияния фактора на отклик. Вычисляется фактическое значения *t*-критерия:

$$t_b = \frac{|b|}{SE_b},$$

где SE_b - стандартная ошибка, вычисляемая по формуле:

$$SE_b = \sqrt{\frac{\sum (y - \hat{y})^2}{(N-1) \cdot (N-2) \cdot \sigma_x^2}} = \left| \text{остатки } e_i = y_i - \hat{y}_i \right| = \sqrt{\frac{\sum e_i^2}{(N-1) \cdot (N-2) \cdot \sigma_x^2}}.$$

Стандартные отклонения для фактора и отклика вычисляются по формулам:

$$\sigma_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{N-1}} = \sqrt{\frac{N}{N-1} \cdot (\overline{x^2} - (\bar{x})^2)} \quad \text{и} \quad \sigma_y = \sqrt{\frac{\sum (y_i - \bar{y})^2}{N-1}} = \sqrt{\frac{N}{N-1} \cdot (\overline{y^2} - (\bar{y})^2)}.$$

Сравнивая фактическое и табличное $t_{табл}$ значения на уровне значимости $\alpha = 0,05$ и числе степеней свободы $N - 2$ принимается решение:

- Если $t_{табл} < t_b$, то H_0 отклоняется, влияние фактора на отклик обнаружено.
- Если $t_{табл} > t_b$, то гипотеза H_0 принимается, фактор на отклик не оказывает существенного влияния.

Далее необходимо рассчитать **границы доверительного интервала для коэффициента регрессии**. Вычисленное выше значение коэффициента регрессии – это среднее значение. Доверительный интервал показывает, каким может быть коэффициент регрессии в 95% случаев. Формулы для расчета имеют следующий вид:

$$н.гр. = b - t_{табл} \cdot SE_b, \quad в.гр. = b + t_{табл} \cdot SE_b.$$

Рассчитывать границы доверительного интервала следует лишь тогда, когда влияние фактора на отклик обнаружено.

4. Оценка качества всей модели

Тесноту связи отклика и фактора оценивает **линейный коэффициент корреляции**

Пирсона r , который можно вычислить, например, по следующей формуле: $r = b \frac{\sigma_x}{\sigma_y}$.

Важным свойством коэффициента корреляции является следующее: $-1 \leq r \leq 1$. Если $r > 0$, то корреляционная связь между переменными называется *прямой*, если $r < 0$ – *обратной*. Качественная оценка тесноты связи может быть выявлена на основе *шкалы Чеддока*.

Теснота связи	Значение $ r $
слабая	0,1 – 0,3
умеренная	0,3 – 0,5
заметная	0,5 – 0,7
высокая	0,7 – 0,9
весьма высокая	0,9 – 0,99

Коэффициент детерминации – для парной линейной регрессии – это квадрат линейного коэффициента корреляции. Величина R^2 показывает, сколько процентов отклика объясняется с помощью включенного в модель фактора. Чем больше R^2 , тем лучше построенная модель. Если $R^2 < 30\%$, то прогнозировать по такой модели нецелесообразно.

Для оценивания качества уравнения регрессии используется *F-критерий Фишера*, который состоит в проверке гипотезы H_0 о статистической незначимости уравнения регрессии и коэффициента детерминации ($R^2 = 0$). Для этого необходимо вычислить фактическое значение *F-критерия*:

$$F_{\text{факт}} = \frac{R^2}{1 - R^2} (N - 2).$$

Сравнивая фактическое значение $F_{\text{факт}}$ и табличное $F_{\text{табл}}$ на уровне значимости $\alpha = 0,05$ и числе степеней свободы 1 и $N - 2$ принимается решение:

- Если $F_{\text{табл}} < F_{\text{факт}}$, то гипотеза H_0 отклоняется, полученное уравнение значимо, т.е. построенная модель «лучше» прогноза по среднему.
- Если $F_{\text{табл}} > F_{\text{факт}}$, то гипотеза H_0 принимается, уравнение незначимо, качество построенной модели сравнимо с точностью прогноза по среднему.

5. Анализ остатков

Для проверки целесообразности использования линейной регрессионной модели используется процедура графического анализа остатков. **Остатки должны быть нормально распределены и не зависеть от предсказанных по уравнению регрессии значений отклика.** Для такой проверки по столбцу остатков строится гистограмма и ящик с усами – для проверки на нормальность и диаграмма рассеяния для проверки независимости остатков от прогноза.

Алгоритм оценки остатков:

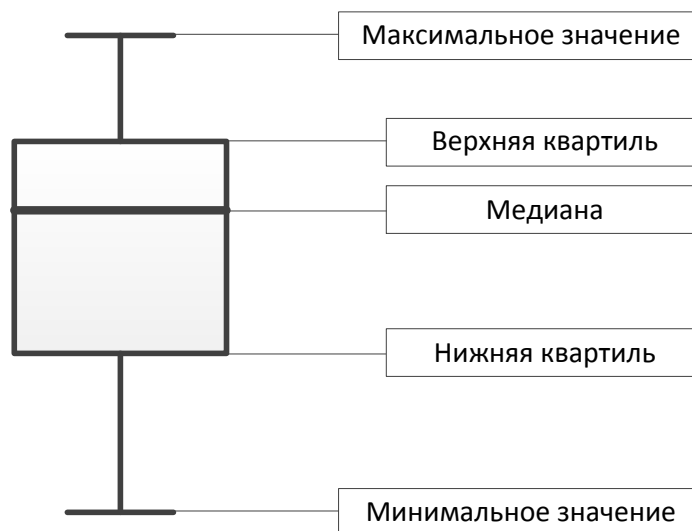
- 1) по формуле Стерджесса определить количество интервалов, на которые следует разбить остатки: $n = 1 + 3,322 \cdot \lg N$;
- 2) определить длину интервала: $h = \frac{e_{\text{max}} - e_{\text{min}}}{n - 1}$;
- 3) найти границы первого интервала: $\left[e_{\text{min}} - \frac{h}{2}; e_{\text{min}} + \frac{h}{2} \right]$;
- 4) рассчитать границы остальных интервалов и частоты (количество остатков, попавших в тот или иной интервал). Получим интервальный вариационный ряд;
- 5) построить гистограмму. В случае нормального распределения остатков гистограмма будет иметь (почти) симметричный вид;
- 6) упорядочить столбик остатков по возрастанию;
- 7) по полученным данным найти медиану, верхнюю и нижнюю квартили, минимальное и максимальное значения:

Медиана – значение в упорядоченной выборке, которое делит ее на две равные части.

Нижняя квартиль – значение в упорядоченной выборке, которое отделяет 25%, начиная с минимального.

Верхняя квартиль – значение в упорядоченной выборке, которое отделяет 25%, начиная с максимального.

- 8) построить ящик с усами по следующей схеме:



В случае нормального распределения ящик будет (почти) симметричным.

- 9) построить диаграмму рассеяния, на которой изобразить точки с координатами (y_i, e_i) . В случае «хорошей» модели никакой закономерности в расположении точек не должно прослеживаться.

6. Оценка точности прогнозов и прогнозирование

Средняя абсолютная ошибка модели (Mean Absolute Percent Error – MAPE) показывает, на сколько процентов в среднем прогноз отклоняется от факта. Вычисляется по формуле:

$$MAPE = \frac{\sum \left| \frac{e_i}{y_i} \right|}{N} \cdot 100\% .$$

Показатель MAPE полезен при использовании разных методов построения модели на основе одних и тех же данных. Чем меньше MAPE, тем лучше модель.

Прогнозное значение y_f определяется путем подстановки в уравнение линейной регрессии соответствующего прогнозного значения x_f . Далее вычисляется стандартная ошибка прогноза:

$$SE_f = \sqrt{\frac{\sum e^2}{N-2} \cdot \left(1 + \frac{1}{N} + \frac{(x_f - \bar{x})^2}{\sigma_x^2 \cdot (N-1)} \right)} .$$

95-процентный доверительный интервал прогноза строится по формуле:

$$н.зп. = y_f - t_{табл} \cdot SE_f, \quad в.зп. = y_f + t_{табл} \cdot SE_f .$$