

МИНИСТЕРСТВО ПУТЕЙ СООБЩЕНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ПУТЕЙ СООБЩЕНИЯ (МИИТ)

Кафедра “Прикладная математика–2”

Л.Ф.Кочнева, А.С.Милевский

ЭКОНОМЕТРИКА.
ЧАСТЬ 1. ПАРНАЯ РЕГРЕССИЯ

Учебное пособие

для экономических специальностей

Москва – 2005

1. Введение

1.1 Предмет эконометрики

Эконометрика – это наука, изучающая конкретные количественные и качественные взаимосвязи экономических объектов и процессов при помощи математических и статистических методов и моделей.

История становления эконометрики как самостоятельной научной дисциплины насчитывает не так уж много лет, но за последние десятилетия эконометрика получила значительное развитие.

Норвежский ученый Р. Фриш определил ее как «...единство экономической теории, статистики и математики». Попробуем в этом разобраться.

Основные результаты *экономической теории* носят *качественный* характер, например, даются ответы на вопросы вида:

- Какие факторы влияют на совокупный объем инвестиций в частном бизнесе?
- Как зависят темпы экономического роста от размеров дефицита государственного бюджета, объема инвестиций, внешнего долга, уровня инфляции?
- Каким образом изменит структуру потребления снижение ставок налогов?

По каждому из этих вопросов, опираясь на теорию, экономисты могут высказывать свои мнения. Однако для принятия обоснованных решений часто требуется отвечать на вопросы, начинающиеся словами «на сколько». Например,

- Какого изменения объема потребления электроэнергии можно ожидать, если предположить, что цены на неё возрастут на 30%?
- На сколько увеличится спрос на данную продукцию, если удвоить текущие расходы на её рекламу?

Это *количественное содержание* и является предметом изучения эконометрики.

Следует отметить, что зачастую не удаётся достичь желаемой степени точности предсказаний. Модели и уравнения, дающие хорошее согласование для одного периода времени, могут оказаться неудовлетворительными для другого. Это связано с большой сложностью экономических процессов. Многообразные внутренние и внешние связи экономики невозможно выделить и детально описать. Здесь

имеется обширное поле деятельности. Но в любом случае анализ моделей в эконометрике производится на основе широкого применения самых разнообразных *математических методов и моделей*. И, конечно, значимые результаты можно получить лишь путем обработки *реальных статистических данных*.

1.2 Модели и статистические данные

Закономерности в экономике проявляются как взаимосвязи между экономическими показателями. В большинстве случаев их стремятся выразить в сравнительно простой математической форме. Всякая математическая модель является лишь упрощённым формализованным представлением реального объекта или процесса. Искусство её построения состоит в совмещении “желаемого с возможным”, лаконичности описания с адекватностью отражения интересующих исследователя закономерностей.

Пример. Изучая, стоимость квадратного метра квартиры в Москве Y , естественно предполагать, что она зависит от количества комнат N , жилой площади квартиры P , нежилой площади Q , расстояния от центра города R , расстояния от ближайшей станции метро S , вида дома (кирпичный, панельный и т.п.) T , этажа U ,.. (а какие ещё факторы вы бы включили в эту модель?): $Y=f(N, P, Q, R, S, T, ..)$. В любом случае влияние неучтённых (или неизвестных) факторов будет представлено как некая случайная добавка ε к Y .

Уже в этом примере видна различная роль рассматриваемых переменных Y, N, P, Q, R, S, T (или, как их ещё называют, *факторов*). В пределах данной модели одни переменные являются *независимыми* (или *объясняющими, экзогенными* – N, P, Q, R, S, T), а другие *зависимыми* (или *эндогенными* – Y).

В общем случае экономическая модель с одной *эндогенной* Y и m *объясняющими* переменными X_1, X_2, \dots, X_m имеет вид:

$$Y=f(x_1, x_2, \dots, x_m, \varepsilon).$$

Пример. Простейшая *модель потребления* может быть представлена в виде

$$\begin{cases} C_t = a + bY_t + \varepsilon_t \\ Y_t = C_t + I_t \end{cases}$$

где C_t – агрегированное потребление, Y_t – национальный доход, I_t – инвестиции в период времени t , ε_t – случайный фактор, a, b – пара-

метры модели. Коэффициент b по понятным причинам носит название “склонность к потреблению”. Это уже пример системы эконометрических уравнений (системы *одновременных уравнений*).

Конкретный вид функциональной зависимости может быть различным. Наиболее часто строятся *линейные эконометрические модели*:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m + \varepsilon, \quad (1.1)$$

разумеется, из-за их большей простоты. Здесь $\beta_0, \beta_1, \dots, \beta_n$ – *параметры модели*, которые оцениваются на основе статистических данных.

Наличие *случайной составляющей* ε в модели обусловлено причинами двоякой природы. Во-первых, она отражает влияние на формирование значения Y *факторов*, не учтённых в перечне *объясняющих переменных* X ; во-вторых, она может включать в себя *погрешность в измерении* Y . В реальности возможны также ошибки в выборе вида функциональной зависимости и т.п. Зависимости другого вида тоже часто сводят к линейным. Так, например, зависимость

$$\ln y = \beta_0 + \beta_1 \ln x_1 + \beta_2 x_2 + \varepsilon$$

фактически является “линейным” представлением зависимости

$$y = A x_1^{\beta_1} \cdot e^{\beta_2 x_2} \cdot u, \quad A = e^{\beta_0}, \quad u = e^{\varepsilon}$$

Используются также квадратичные, кубические, гиперболические, логарифмические, логистические и другие функции.

Выбор вида функциональной зависимости называется *спецификацией модели*, а определение состава объясняющих переменных – *спецификацией переменных*.

Если модель (1.1) содержит только одну объясняющую переменную, она называется *парной регрессией*. При $m > 1$ мы имеем дело с *множественной регрессией*.

По типам модели разделяются на *модели регрессии*, *динамические модели*, *системы одновременных уравнений*...

Основу эконометрического моделирования составляют статистические данные. Их различают по типам.

Пространственные данные собираются по какому-либо экономическому показателю для разных объектов в один момент времени или в разные моменты в случае, когда время несущественно. Например,

данные по курсам покупки/продажи валюты сегодня в различных обменных пунктах Москвы.

Временные ряды – данные для одного объекта в различные моменты времени. Например, ежеквартальные данные об инфляции, данные о денежной эмиссии за последние годы, ежедневный курс доллара.

Промежуточное положение занимают *панельные данные*, которые отражают наблюдения по большому числу объектов за относительно небольшой период времени. Например, данные об объёмах выпуска и капитальных вложениях российских предприятий топливно-энергетического комплекса за 1995–2000 г.

1.3 Основные этапы эконометрического моделирования

В процессе эконометрического исследования можно выделить следующие этапы.

1-й этап (постановочный). Определение конечных целей моделирования, формирование набора участвующих в модели факторов и их ролей.

2-й этап (спецификация модели). Выбор общего вида модели, состава и вида входящих в неё функциональных взаимосвязей.

3-й этап (информационный). Сбор необходимой статистической информации, т.е. значений участвующих в модели факторов.

4-й этап (идентификация модели). Анализ модели, статистическое оценивание параметров модели.

5-й этап (проверка модели). Сопоставление реальных и оценённых данных, проверка адекватности модели.

В результате проверки модели нередко приходится неоднократно возвращаться к предыдущим этапам и вносить коррективы.

1.4 Метод наименьших квадратов

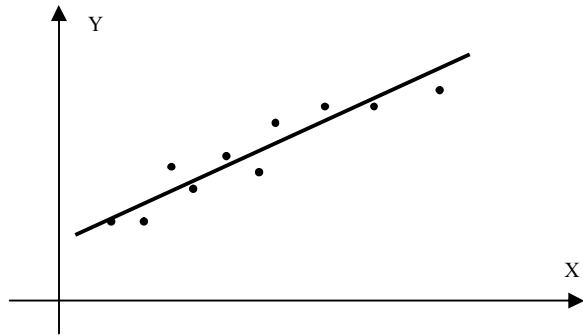
Во многих случаях возникает задача: *найти кривую заданного вида, наиболее точно приближающую экспериментальные данные*.

Математически это формулируется так:

требуется подобрать такие значения параметров β , чтобы график функции

$$f(x) = \beta_1 \varphi_1(x) + \beta_2 \varphi_2(x) + \dots + \beta_m \varphi_m(x).$$

проходил как можно ближе к заданным точкам $(x_i, y_i) \quad i=1, 2..n$.
 Здесь β_i – неизвестные коэффициенты, φ_i – известные функции.



Пример (на рис). Требуется подобрать прямую $y = \beta_0 + \beta_1 x$, наиболее близкую к заданным точкам.

Чтобы такая задача могла считаться корректно сформулированной, нужно как-то конкретизировать понятие близости точек к кривой. Это можно сделать многими способами.

- Можно, например, в качестве критерия близости выбрать сумму расстояний от точек до кривой. Но в таком случае было бы затруднительно выписать явную формулу (чему, например, равно расстояние от точки (1;3) до параболы $y=x^2+x$?).

- Несколько упрощается дело, если расстояния измерять по вертикали, то есть использовать критерий

$$Q(\beta) = \sum |y_i - f(x_i)| \rightarrow \min.$$

В этом случае, однако, есть некоторое неудобство в том, что функция модуль не имеет производной в точке 0, так что мы не сможем использовать производную для поиска минимума.

- Этого недостатка функции $|x|$ лишена функция x^2 , которую и используют в критерии близости метода наименьших квадратов (МНК).

Таким образом, получается следующая постановка задачи для метода наименьших квадратов:

Для заданных точек $(x_i, y_i) \quad i=1, 2..n$, найти такие значения параметров β , чтобы минимизировать остаточную сумму квадратов:

$$Q(\beta) = \sum_{i=1}^n (y_i - f(x_i))^2 \rightarrow \min$$

Оказывается, что в этом случае получаются удобные формулы для подсчёта коэффициентов β .

Дифференцируя $Q(\beta)$, и приравнявая производную нулю, легко получить следующую систему линейных уравнений для нахождения коэффициентов β_i :

$$\begin{cases} A_{11}\beta_1 + A_{12}\beta_2 + \dots + A_{1m}\beta_m = B_1 \\ A_{22}\beta_1 + A_{22}\beta_2 + \dots + A_{2m}\beta_m = B_2 \\ \dots \quad \dots \quad \dots \quad \dots \\ A_{m2}\beta_1 + A_{m2}\beta_2 + \dots + A_{mm}\beta_m = B_m \end{cases}$$

где

$$A_{kl} = \sum_{i=1}^n \varphi_k(x_i)\varphi_l(x_i), \quad B_k = \sum_{i=1}^n y_i\varphi_k(x_i)$$

Пример. Имеются следующие данные за 7 лет о среднем доходе (x) и среднем потреблении (y):

Годы	1	2	3	4	5	6	7
Доход	14.5	15.7	16.3	18.5	20.3	21.7	23
Потребление	12	12.7	13	15.5	16.5	17.3	20

Для этих данных требуется подобрать наилучшую прямую $y = \beta_0 + \beta_1 x$ методом наименьших квадратов.

Здесь $\varphi_1(x)=1$, $\varphi_2(x)=x$. Вычисляем коэффициенты системы уравнений (4.3) по формулам (4.4):

$$A_{11} = \sum \varphi_1\varphi_1 = \sum 1 = 7, \quad A_{12} = A_{21} = \sum \varphi_1\varphi_2 = \sum x = 130,$$

$$A_{22} = \sum \varphi_2\varphi_2 = \sum x^2 = 2476.6,$$

$$B_1 = \sum y\varphi_1 = \sum y = 107, \quad B_2 = \sum (y\varphi_2) = \sum (yx) = 2042.4.$$

Получаем систему уравнений

$$\begin{cases} 7\beta_0 + 130\beta_1 = 107 \\ 130\beta_0 + 2476.6\beta_1 = 2042.4, \end{cases}$$

откуда $\beta_0 = -1.167$, $\beta_1 = 0.8859$. Таким образом, наилучшее линейное приближение для исходных данных имеет вид $y = -1.167 + 0.8859x$. На потребление тратится примерно 88% дохода.

1.5 Контрольные вопросы

1. Дайте определение эконометрики.
2. Какие бывают виды моделей эконометрики?
3. Какие различают типы переменных?
4. Какие функциональные формы могут иметь эконометрические модели?
5. Как проявляется влияние не учтённых в эконометрической модели факторов?
6. Перечислите основные типы статистических данных.
7. Что понимается под спецификацией модели?
8. В чем заключается сущность метода наименьших квадратов?
9. Почему в методе наименьших квадратов используется именно такой критерий близости точек к кривой?
10. Требуется подобрать наилучшую параболу $y = a + bx + cx^2$ для заданных точек. Что здесь играет роль параметров β_i , и каковы функции $\varphi_i(x)$?

2. Некоторые сведения из теории вероятностей и математической статистики

2.1 Случайная величина и ее распределение

Полагаем, что понятие вероятности читателю известно, по крайней мере, на интуитивном уровне. *Вероятность события* служит мерой того, насколько можно быть уверенным, что это событие произойдет. Она принимает значения от 0 до 1, причем 0 соответствует *невозможному* событию, а 1 – *достоверному* событию. *Случайной величиной* называется переменная, которая с определенной вероятностью может принимать значения из заданного множества. Различают *дискретные* и *непрерывные* случайные величины.

Дискретная случайная величина может принимать лишь *конечное или счетное число значений*. Например, число, которое выпадет при бросании игрального кубика, число покупателей, побывавших в магазине в течение дня, количество аварий на электростанциях в течение года – дискретные случайные величины.

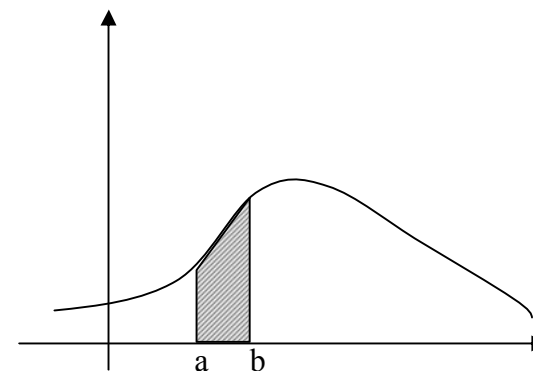
Дискретную случайную величину X можно задать, указав вероятности p_1, p_2, \dots, p_n , с которыми она принимает значения X_1, X_2, \dots, X_n . Это удобно записать в виде таблицы:

X_1	X_2	...	X_n
p_1	p_2	...	p_n

причем все вероятности p_1, p_2, \dots, p_n неотрицательны, а их *сумма равна единице*. Говорят, что такая таблица описывает *закон распределения* дискретной случайной величины.

Случайная величина, которая может принимать любые значения из *некоторого интервала*, является *непрерывной*. Например, размер экспорта за год, курсы валют, объём потребления электроэнергии – непрерывные случайные величины. Для непрерывной случайной величины вероятность того, что она примет любое конкретное значение, равна нулю. Можно лишь говорить о *вероятности попадания в заданный интервал*. Поэтому непрерывную случайную величину невозможно задать таблично. Закон распределения её описывается при помощи *плотности распределения*.

Непрерывная случайная величина задается своей *функцией плотности распределения вероятности* $f(x)$, типичный вид графика которой показан на рисунке. *Вероятность попадания случайной величины в интервал $[a; b]$ равна площади заштрихованной криволинейной трапеции над этим интервалом, ограниченной сверху графиком $f(x)$.*



Плотность вероятности $f(x)$ равна вероятности попадания случайной величины в малый интервал в окрестности точки X в расчете на единицу длины этого интервала. Поэтому вероятность того, что значение случайной величины окажется в малом интервале от X до $X+\Delta X$, приблизительно равна $f(x)\Delta X$. Ясно, что *площадь всей фигуры под графиком плотности вероятности равна 1*.

Пример. Одним из наиболее распространённых законов распределения непрерывной случайной величины является так называемый *нормальный закон распределения*, для которого плотность распределения вероятности имеет вид

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

где m, σ - некоторые параметры. График этой функции симметричен относительно вертикальной прямой $x = m$. Число σ определяет его «растянутость»: чем больше σ , тем более пологий график. Если величина распределена по нормальному закону с параметрами m, σ , будем обозначать это так: $X \sim N(m, \sigma)$.

Нормальное распределение вероятностей, для которого $m=0, \sigma=1$, т.е. $X \sim N(0, 1)$, называется *стандартным нормальным распределением*.

Всякое нормальное распределение можно выразить через стандартное нормальное распределение. Пусть $X \sim N(m, \sigma)$. Рассмотрим новую случайную величину $Z = \frac{X-m}{\sigma}$. Легко проверить непосредственно, что $Z \sim N(0, 1)$. Обозначим через

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$$

вероятность попадания в интервал $(-\infty; X)$ для стандартного нормального закона. Для практических вычислений имеются таблицы значений этой функции. Используя их, можно рассчитать *вероятность попадания в произвольный интервал для нормально распределённой случайной величины X* :

$$P(a \leq X \leq b) = \Phi\left(\frac{b-m}{\sigma}\right) - \Phi\left(\frac{a-m}{\sigma}\right)$$

распространённость нормального распределения объясняет *центральная предельная теорема*. Она утверждает, что если случайную величину можно представить как сумму большого числа не зависящих друг от друга слагаемых, ни одно из которых не является доминирующим, то эта сумма распределена приблизительно по нормальному закону. Например, рост взрослого человека определяется взаимодействием большого числа различных генов и имеет примерно нормальное распределение.

2.2 Математическое ожидание и дисперсия

Для *дискретной* случайной величины X её *математическим ожиданием* (или *средним значением*) называется:

$$MX = x_1 p_1 + x_2 p_2 + \dots$$

Если множество $\{X_1, X_2, \dots\}$ значений X конечно, то в этой сумме будет конечное число слагаемых, если же множество значений счётно, математическое ожидание вычисляется как сумма полученного ряда.

Пример. *Геометрический закон распределения* описывает величину X , равную *количеству испытаний вплоть до первого успеха*, если вероятность успеха в каждом испытании равна p , и испытания независимы. Нетрудно понять, что X описывается таблицей

X	1	2	3	...
P	p	$q p$	$q^2 p$...

где $q = 1 - p$. Математическое ожидание (среднее количество испытаний до первого успеха) будет равно

$$MX = 1 \cdot p + 2q \cdot p + 3q^2 \cdot p + \dots = p(1 + 2q + 3q^2 + \dots) = 1/p$$

В случае *непрерывной* случайной величины суммирование заменяется интегрированием:

$$MX = \int_{-\infty}^{+\infty} x \cdot f(x) dx$$

Для *нормального закона распределения* $MX = m$.

Свойства математического ожидания выводятся из свойств суммирования или интегрирования.

- Математическое ожидание суммы нескольких случайных величин равно сумме их математических ожиданий:
 $M(X+Y)=MX+MY$
- При умножении случайной величины на константу, ее математическое ожидание умножается на эту же константу:
 $M(cX)=cMX$
- Математическое ожидание постоянной величины равно самой этой величине: $Mc=c$.

Дисперсией случайной величины называется математическое ожидание квадрата разности между этой величиной X и ее математическим ожиданием MX т.е.

$$DX=M(X-MX)^2$$

В частности, для дискретной случайной величины

$$DX=(x_1-MX)^2p_1+(x_2-MX)^2p_2+\dots$$

а для непрерывной случайной величины

$$MX = \int_{-\infty}^{+\infty} (x - MX)^2 \cdot f(x) dx$$

Величина $\sigma = \sqrt{DX}$ называется *стандартным отклонением*. Она, также как и дисперсия, характеризует меру «разброса» случайной величины относительно математического ожидания MX .

Пример. Рассмотрим дискретную случайную величину X , принимающую значения $1/N, 2/N, 3/N, \dots, N/N$ с равными вероятностями. Тогда каждая такая вероятность равна $1/N$. Вычислим:

$$MX = \frac{1}{N} \cdot \frac{1}{N} + \frac{2}{N} \cdot \frac{1}{N} + \dots + \frac{N}{N} \cdot \frac{1}{N} = \frac{1}{N^2} (1 + 2 + \dots + N) = \frac{N+1}{2N},$$

$$M(X^2) = \frac{1}{N^2} (1^2 + 2^2 + \dots + N^2) = \frac{(N+1) \cdot (2N+1)}{6N^2},$$

$$DX = M(X^2) - (MX)^2 = \frac{N^2 - 1}{12N^2}$$

В случае нормального распределения дисперсия $DX = \sigma^2$. Отметим следующие два свойства дисперсии.

- При умножении случайной величины на константу дисперсия умножается на квадрат этой константы, т.е. $D(cX) = c^2 DX$

- Справедлива формула $D(X) = M(X^2) - (MX)^2$
- Дисперсия суммы *независимых* (см. п.2.3) случайных величин равна сумме дисперсий слагаемых.

2.3 Совместное распределение случайных величин

В экономике важны связи между экономическими переменными. Они могут быть не детерминированными, а иметь стохастическую природу.

Дискретное распределение вероятностей для двух случайных переменных x и y задается таблицей:

	Y_1	Y_2	...
X_1	p_{11}	p_{12}	...
X_2	p_{21}	p_{22}	...
...

Здесь p_{ij} - вероятность того, что случайная величина X примет значение X_i , а случайная величина Y одновременно примет значение Y_j . По смыслу вероятности должны выполняться условия:

$$0 \leq p_{ij} \leq 1, i = 1, \dots, k, j = 1, 2, \dots, l \quad \sum_i \sum_j p_{ij} = 1$$

Вероятность события $X = X_i$ равна сумме вероятностей p_{ij} по строке i , а вероятность события $Y = Y_j$ равна сумме вероятностей p_{ij} по столбцу j .

Пример. Из колоды 36 карт достали 2 карты. X – количество мастей, Y – количество королей среди них. Нетрудно убедиться, что тогда таблица вероятностей будет иметь вид

X\Y	0	1	2
1	$\frac{112}{630}$	$\frac{32}{630}$	0
2	$\frac{384}{630}$	$\frac{96}{630}$	$\frac{6}{630}$

Вероятность того, что X=1 (только одна масть) равна $\frac{112}{630} + \frac{32}{630} =$

$\frac{144}{630}$, вероятность того, что Y=0 (нет королей) равна $\frac{112}{630} + \frac{384}{630} = \frac{496}{630}$.

Легко подсчитать, что $MX=62/35 \approx 1.77$, $DX=143/811 \approx 0.18$,
 $MY=2/9 \approx 0.22$, $DY=123/641 \approx 0.19$.

Совместное распределение вероятности для двух *непрерывных* случайных величин x и y задается *функцией плотности вероятности* $f(x,y)$. Вероятность попадания в прямоугольник со сторонами X

до ΔX и ΔY с центром в точке (x,y) приблизительно равна

$f(x,y)\Delta X\Delta Y$. При вычислении математических ожиданий MX , MY и дисперсий DX , DY суммы заменяются двойными интегралами.

- Две случайные величины X , Y называются *независимыми*, если

$$P(X < x \text{ и } Y < y) = P(X < x) \cdot P(Y < y)$$

для любых X и Y .

Для *дискретных* случайных величин удобнее проверять равенство

$$P(X=x \text{ и } Y=y) = P(X=x) \cdot P(Y=y)$$

Пример. Для предыдущего примера $P(X=1) = \frac{144}{630}$, $P(Y=2) = \frac{6}{630}$,

$P(X=1 \text{ и } Y=2) = 0$. Значит, эти случайные величины X, Y *зависимы*.

Зависимость между случайными величинами может носить различный характер. Наличие *линейной* связи между X и Y можно обнаружить при помощи *коэффициента корреляции*

$$r_{XY} = \frac{M(XY) - MX \cdot MY}{\sqrt{DX} \sqrt{DY}}$$

Это безразмерная величина, принимающая значения от -1 до 1 , причем 1 соответствует наличию положительной, а -1 – отрицательной линейной связи x и y . Для независимых величин коэффициент корреляции равен 0 . Чем больше модуль коэффициента корреляции, тем больше связь между величинами похожа на линейную.

Пример. Для рассматриваемого примера (X – количество мастей, Y – количество королей) получаем $MX \cdot Y = 124/315 \approx 0.39$ и $r_{XY} = 0$, что говорит об *отсутствии линейной связи* между X и Y , хотя *зависимость*, как отмечалось выше, есть.

2.4 Оценивание характеристик случайной величины

Как правило, закон распределения реальной случайной величины неизвестен (по какому закону, например, распределён среднегодовой спрос на электроэнергию в Москве?). Соответственно, нельзя вычислить математическое ожидание и дисперсию. Их можно лишь *оценить* на основе *выборочных статистических данных*.

Всевозможные реализации случайной величины образуют *генеральную совокупность*. Статистические данные содержат некоторую ее часть, называемую *выборкой*.

Поскольку получение числовых характеристик генеральной совокупности затруднительно или вообще невозможно, их оценивают по *выборочным характеристикам*. Вообще, *точной оценкой параметра закона распределения* называют его *приближённое значение, найденное по выборке*.

Пусть имеется выборка X_1, X_2, \dots значений случайной величины X . Точечная оценка для математического ожидания $MX = m$ называется

«*выборочное среднее*», обозначается \bar{X} и вычисляется как среднее арифметическое чисел - элементов выборки (с учетом кратностей):

$$\bar{X} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

- Оценка числовой характеристики случайной величины называется *несмещенной*, если ее *математическое ожидание равно истинному значению этой характеристики*.

Например, в качестве оценки \hat{m} математического ожидания m можно взять выборочное среднее. По свойствам математического ожидания имеем

$$M(\bar{x}) = M\left(\frac{1}{n} \sum_{i=1}^n x_i\right) = \frac{1}{n} \sum_{i=1}^n M(x_i) = \frac{1}{n} \cdot nm = m$$

Поэтому выборочное среднее \bar{x} является *несмещенной оценкой для математического ожидания*.

Точечная оценка для дисперсии DX называется “*выборочная дисперсия*”, обозначается D_x^* и вычисляется по формуле:

$$D_x^* = \overline{(x - \bar{x})^2} = \overline{x^2} - (\bar{x})^2.$$

Эта оценка – смещенная! Нетрудно проверить, что

$$M(D_x^*) = M\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2\right) = \frac{n-1}{n} DX \neq DX$$

Улучшенная *несмещенная точечная оценка* для DX обозначается S^2 и вычисляется по формуле:

$$S^2 = \frac{n}{n-1} D_x^*.$$

Зачастую можно предложить несколько различных формул (даже несмещенных) для оценки числовой характеристики. Например, в случае симметричного (например, нормального) закона можно попробовать также и следующую оценку для $MX=m$:

$$\hat{m} = \frac{1}{2} (\max_i (x_i) + \min_i (x_i))$$

Какая оценка лучше, эта или выборочное среднее \bar{x} ?

- Оценка называется *эффективной*, если она обладает *наименьшей дисперсией среди всех несмещенных оценок* этой характеристики.

Для примера покажем, что выборочное среднее \bar{x} является эффективной оценкой математического ожидания $MX=m$ в классе *линейных оценок*. Наряду с нею рассмотрим другую линейную несмещенную оценку

$$\tilde{x} = \sum \lambda_i x_i; M\tilde{x} = m$$

Обозначим $\delta_i = \lambda_i - \frac{1}{n}$. Тогда $0 = M(\tilde{x}) - M(\bar{x}) = m \cdot \sum \delta_i$.

Находим теперь

$$D(\tilde{x}) = D\left(\sum \lambda_i x_i\right) = \sum \left(\frac{1}{n} + \delta_i\right)^2 D(x_i) =$$

$$\left(\sum \frac{1}{n^2} + \frac{2}{n} \sum \delta_i + \sum \delta_i^2\right) \cdot DX = \frac{DX}{n} + DX \cdot \sum \delta_i^2$$

Второе слагаемое в правой части неотрицательно, а первое, как нетрудно убедиться, равно $D(\bar{x})$. Поэтому $D(\bar{x}) \leq D(\tilde{x})$, то есть среднее выборочное является эффективной оценкой по отношению к другим линейным оценкам.

- Оценка характеристики случайной величины называется *состоятельной*, если ее предел по вероятности равен истинному значению этой характеристики. *Сходимость по вероятности* $\{X_n\}$ к A означает, что

$$\lim_{n \rightarrow \infty} P(|X_n - A| > \varepsilon) = 0$$

для любого положительного числа ε .

- Для проверки состоятельности несмещенной оценки достаточно показать, что её дисперсия стремится к нулю с ростом объема выборки.

Так, например, дисперсия $D(\bar{x}) = DX/n$ выборочного среднего \bar{x} стремится к 0. Следовательно, оценка \bar{x} *сходится* к MX по вероятности, т.е. она *состоятельна*.

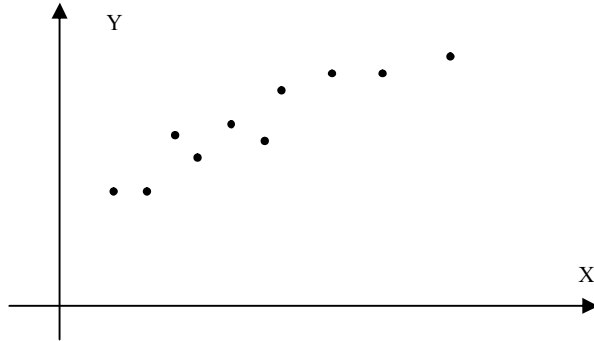
2.5 Выборочные характеристики взаимосвязи между случайными величинами

Как уже отмечалось, на практике приходится иметь дело с выборочными данными, а не со всей генеральной совокупностью. Пусть получена выборка

X_1	X_2	X_3	...
Y_1	Y_2	Y_3	...

где X_i, Y_i – значения, которые принимали случайные величины X и Y в i -м наблюдении.

Первое суждение о взаимосвязи X и Y можно составить по *диаграмме рассеяния*.



На ней результат i -го наблюдения изображается точкой (X_i, Y_i) . Наличие *линейной* связи между факторами X и Y помогает подтвердить или опровергнуть *выборочный коэффициент корреляции*

$$r_{XY}^* = \frac{\overline{XY} - \bar{X} \cdot \bar{Y}}{\sqrt{D_X^*} \sqrt{D_Y^*}} = \frac{\frac{1}{n} \sum_{i=1}^n X_i Y_i - \bar{X} \cdot \bar{Y}}{\sqrt{D_X^*} \sqrt{D_Y^*}}$$

Это – безразмерная величина, принимающая значения от -1 до 1 . Чем ближе модуль этого числа к 1 , тем сильнее линейная зависимость.

2.6 Контрольные вопросы

1. Что такое дискретная случайная величина?
2. Что такое непрерывная случайная величина?
3. Каким образом может быть задана случайная величина?
4. Каковы свойства плотности распределения?
5. Опишите нормальный закон распределения.
6. Как вычисляется вероятность попадания в интервал для нормального закона распределения?
7. Перечислите свойства математического ожидания.

8. Перечислите свойства дисперсии и докажите их.
9. В каком случае случайные величины называются независимыми?
10. Что такое точечная оценка числовой характеристики случайной величины?
11. Какая оценка называется несмещённой? Состоятельной? Эффективной?
12. В случае строгой положительной линейной связи между двумя случайными величинами коэффициент корреляции примет значения, равное: а) 1 ; б) -1 ; в) 0 ?

3. Парная линейная регрессия

3.1 Модель парной линейной регрессии

Если на диаграмме рассеяния точки наблюдений (X_i, Y_i) распределяются случайным образом примерно вблизи некоторой прямой линии, то можно предполагать, что между переменными X, Y существует *линейная* статистическая связь.

Модель парной линейной регрессии имеет вид

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

где Y – *зависимая переменная*, X – *объясняющая переменная* (также называемая *регрессор*), β_0, β_1 – *коэффициенты регрессии*, ε – *случайный член*.

Замечание. На практике трудно ожидать, что значения фактора Y будут определяться значениями только одного фактора X , поэтому модель парной регрессии является лишь простым частным случаем, иллюстрацией методов значительно более важной *модели множественной регрессии*.

Включение случайного члена ε в уравнение связано с возмущениями, которые не учтены в данной модели. Это могут быть невключение других объясняющихся переменных, возможная нелинейность модели, неправильный выбор объясняющей переменной, ошибки измерений и другие факторы.

Допустим, имеется выборка $X_i, Y_i, i = 1, 2, \dots, n$. Для i -го наблюдения

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad i = 1, 2, \dots, n.$$

Значение, ε_i , которое принял случайный член, определяет отклонение зависимой переменной Y_i от теоретического значения

$\beta_0 + \beta_1 X_i$, устанавливаемого моделью.

3.2 Оценивание модели

Параметры β_0, β_1 , которые определяют модель парной линейной регрессии (3.1), на практике не известны, и их можно оценивать по статистическим данным, опираясь на различные принципы.

Пусть $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ – оцененное уравнение регрессии. Обозначим остатки $e_i = y_i - \hat{y}_i, i = 1, 2, \dots, n$.

Хорошими свойствами обладают оценки коэффициентов β_0, β_1 , получаемые по методу наименьших квадратов, рассмотренному выше. Напомним, что в этом случае оценки определяются из условия минимума суммы квадратов отклонений:

$$ESS = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \rightarrow \min$$

Выведем соответствующие формулы ещё раз для этого частного случая. Необходимое условие минимума заключается в обращении в нуль частных производных по β_0 и β_1 :

$$\begin{cases} \frac{\partial(ESS)}{\partial\beta_0} = 2\sum (y_i - \beta_0 - \beta_1 x_i) = 0, \\ \frac{\partial(ESS)}{\partial\beta_1} = 2\sum (y_i - \beta_0 - \beta_1 x_i) \cdot x_i = 0. \end{cases}$$

то есть

$$\begin{cases} \beta_0 \sum 1 + \beta_1 \sum x_i = \sum y_i, \\ \beta_0 \sum x_i + \beta_1 \sum x_i^2 = \sum x_i y_i \end{cases}$$

или, после деления на $\sum 1 = n$,

$$\begin{cases} \beta_0 + \beta_1 \cdot \bar{x} = \bar{y}, \\ \beta_0 \cdot \bar{x} + \beta_1 \cdot \bar{x}^2 = \overline{xy} \end{cases}$$

Решение этой системы линейных уравнений, как нетрудно проверить, находится по формулам:

$$\hat{\beta}_1 = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{x^2 - (\bar{x})^2}; \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad (3.1)$$

Отметим, что уравнение $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ является лишь оценкой для истинной линейной модели (3.1).

Пример. В таблице приведены индекс цен Y и совокупное производство X цветных телевизоров в США с 1970 по 1979 г.

Год	1970	1971	1972	1973	1974
Y	624	549	520	474	448
X	34,099	41,473	50,218	60,289	68,700
Год	1975	1976	1977	1978	1979
Y	442	430	407	380	360
X	74,919	83,113	92,454	103,674	113,170

Найдём оценки параметров парной регрессии Y на X .

$$\bar{x} = 72.2009, \bar{y} = 463.4, x^2 = 5837.4884; \overline{xy} = 31607.838,$$

$$\hat{\beta}_1 = \frac{31607.838 - 72.2009 \cdot 463.4}{5837.4884 - 72.2009^2} = -2.96,$$

$$\hat{\beta}_0 = 463.4 + (-2.96) \cdot 72.2009 = 677.29,$$

$$ESS = (624 - 576.27)^2 + (549 - 554.72) + \dots = 4268.7108$$

3.3 Теорема Гаусса-Маркова для модели парной регрессии

Рассмотрим модель парной линейной регрессии

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, i = 1, 2, \dots, n. \quad (3.2)$$

Предположим, что выполнены следующие условия:

- (А) X_i – не случайная (детерминированная) величина;
числа X_1, X_2, \dots не все одинаковы.

- (B) $M\varepsilon_i = 0$.
- (C) $D\varepsilon_i = \sigma^2$ не зависит от i .
- (D) $M\varepsilon_i\varepsilon_j = 0$ при всех $i \neq j$.

Как правило, условия (B), (C) заменяют более сильным условием

- (E) $\varepsilon_i \sim N(0, \sigma^2)$ для всех i .

Условие (C) говорит о постоянстве дисперсии случайного фактора. В случае его нарушения говорят о *гетероскедастичности*. Условие (D) означает независимость (точнее, *некоррелированности*) значений случайного фактора в различных наблюдениях. Нарушение этого условия называется *автокорреляцией*.

Теорема. В предположениях (A)–(D) оценки метода наименьших квадратов

$$\hat{\beta}_1 = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{x^2 - (\bar{x})^2}; \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

являются несмещёнными и имеют наименьшую дисперсию среди всех несмещённых линейных оценок параметров β_0, β_1 .

Это утверждение устанавливает обоснованность использования метода наименьших квадратов для оценки коэффициентов линейной регрессии.

Доказательство. Проверим только *несмещённость*. Для этого вычислим

$$\begin{aligned} M\hat{\beta}_1 &= M \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{x^2 - (\bar{x})^2} = M \frac{\overline{x(\beta_0 + \beta_1 x + \varepsilon)} - \bar{x} \cdot \overline{\beta_0 + \beta_1 x + \varepsilon}}{x^2 - (\bar{x})^2} = \\ &= \frac{\overline{x\beta_1 x} - \bar{x} \cdot \overline{\beta_1 x}}{x^2 - (\bar{x})^2} = \beta_1 \frac{\overline{xx} - \bar{x} \cdot \bar{x}}{x^2 - (\bar{x})^2} = \beta_1; \end{aligned}$$

$$M\hat{\beta}_0 = M\bar{y} - M\hat{\beta}_1 \bar{x} = \beta_0 + \beta_1 \bar{x} - \beta_1 \bar{x} = \beta_0.$$

Замечание. В модель парной регрессии, кроме β_0 и β_1 , входит ещё параметр σ^2 . Он характеризует дисперсию случайного фактора (то есть, фактически, отклонение Y от линейной функции $\beta_0 + \beta_1 X$). Можно показать, что несмещённой оценкой для σ^2 является

$$s^2 = \hat{\sigma}^2 = \frac{ESS}{n-2} = \frac{1}{n-2} \sum_{i=1}^n e_i^2$$

Пример. В предыдущем примере находим оценку для дисперсии случайного фактора:

$$S^2 = \frac{1}{8} ESS = \frac{4268.7108}{8} = 533.58.$$

В случае если выполнено условие нормальности (E), можно указать законы, по которым распределены оценки параметров регрессии. Оказывается, что

$$\begin{aligned} \hat{\beta}_1 &\sim N\left(\beta_1, \sqrt{\frac{\sigma^2}{n \cdot D_x^*}}\right); \hat{\beta}_0 \sim N\left(\beta_0, \sqrt{\frac{\sigma^2}{n \cdot D_x^*}} \sqrt{x^2}\right); \\ \frac{(n-2)s^2}{\sigma^2} &\sim \chi^2(n-2) \end{aligned} \quad (3.3)$$

Таким образом, поведение оценок для β_0 и β_1 описывается нормальным законом, а поведение S^2 – законом $\chi^2(n-2)$ (читается “распределение хи-квадрат с $n-2$ степенями свободы”). Это позволяет строить доверительные интервалы и проверять гипотезы о значениях параметров регрессии по экспериментальным данным.

3.4 Доверительные интервалы для параметров регрессии

Будем считать, что условие нормальности (E) выполнено. Тогда, основываясь на (3.2), можно построить *доверительные интервалы* для параметров регрессии. Они имеют следующий вид:

$$\hat{\beta}_0 - t_{1-\alpha/2}(n-2) \cdot S_{\beta_0} < \beta_0 < \hat{\beta}_0 + t_{1-\alpha/2}(n-2) \cdot S_{\beta_0};$$

$$\hat{\beta}_1 - t_{1-\alpha/2}(n-2) \cdot S_{\beta_1} < \beta_1 < \hat{\beta}_1 + t_{1-\alpha/2}(n-2) \cdot S_{\beta_1}; \quad (3.4)$$

$$\frac{ESS}{\chi^2_{1-\alpha/2}(n-2)} < \sigma^2 < \frac{ESS}{\chi^2_{\alpha/2}(n-2)}.$$

Здесь α – заданный “уровень значимости”, т.е. вероятность того, что построенный интервал *не содержит* истинного значения параметра.

Далее, $t_{1-\alpha/2}(n-2)$, $\chi^2_{1-\alpha/2}(n-2)$, $\chi^2_{\alpha/2}(n-2)$ – квантили, соответственно, *законов Стьюдента* и *хи-квадрат*. Это – табличные значения.

Наконец, S_{β} – выборочные среднеквадратичные отклонения оценок параметров (или *стандартные ошибки параметров*), они находятся по формулам (ср. 3.3):

$$S_{\beta_1} = \sqrt{\frac{S^2}{nD_x^*}}; \quad S_{\beta_0} = \sqrt{\frac{S^2}{nD_x^*}} \cdot \sqrt{x^2}; \quad (3.5)$$

Как правило, стандартные ошибки параметров вычисляются в эконометрических программах вместе с оценками самих параметров.

Пример. В рассматриваемом примере

$$S_{\beta_1} = \sqrt{\frac{463.4}{8 \cdot 624.52}} = 0.29;$$

$$S_{\beta_0} = S_{\beta_1} \cdot \sqrt{5837.4884} = 22.33$$

Таким образом, оценка зависимости индекса цен телевизоров Y от объема их выпуска X в рассматриваемом примере может быть записана так:

$$y = 677.29 - 2.96x$$

(22.33) (0.29)

где в скобках подписаны стандартные ошибки параметров.

3.5 Проверка гипотез о параметрах регрессии

В этом параграфе также будем считать, что условие нормальности (E) выполнено.

При исследовании модели приходится проверять гипотезы о равенстве истинных значений параметров регрессии заданным числам. Для этого достаточно построить доверительный интервал и посмотреть, попадает ли в него это заданное число.

Рассмотрим, например, гипотезу $H_0: \beta_1 = A$ при альтернативной гипотезе $H_1: \beta_1 \neq A$. Учитывая формулы (3.4) получаем, что гипотеза H_0 принимается в случае, если число A лежит в интервале

$(\hat{\beta}_1 - t_{1-\alpha/2}(n-2) \cdot S_{\beta_1}; \hat{\beta}_1 + t_{1-\alpha/2}(n-2) \cdot S_{\beta_1})$, в противном случае принимается гипотеза H_1 .

Особенно часто проверяют гипотезу $H_0: \beta_1 = 0$. Дело в том, что если принимается альтернативная гипотеза

$$H_1: \beta_1 \neq 0,$$

то говорят, что коэффициент β_1 *значим*. Таким образом, получаем следующий критерий значимости параметра β_1 :

$$\frac{\hat{\beta}_1}{S_{\beta_1}} \geq t_{1-\alpha/2}(n-2) \quad (3.5)$$

Пример. В рассматриваемом примере доверительный интервал для коэффициента β_1 на уровне значимости $\alpha = 0.1$ имеет вид

$$-2.96 - t_{0.95}(8) \cdot 0.29 < \beta_1 < 2.96 + t_{0.95}(8) \cdot 0.29;$$

$$t_{0.95}(8) = 1.86 \Rightarrow -3.51 < \beta_1 < -2.52$$

Мы видим, что доверительный интервал не содержит 0, так что коэффициент β_1 значим.

3.6 Коэффициент детерминации

Дисперсия зависимой переменной (или *полная сумма квадратов*) обозначается TSS:

$$TSS = \sum (y_i - \bar{y})^2 = nD_y^*$$

Сумма

$$RSS = \sum (\hat{y}_i - \bar{y})^2 = nD_{\hat{y}}^*$$

определяет разброс оценки \hat{y} относительно среднего, или “объяснённую часть дисперсии”. И, наконец, сумма

$$ESS = \sum (y_i - \hat{y}_i)^2 = \sum e_i^2 = nD_e^*$$

представляет собой “необъясненную часть дисперсии” зависимой переменной. Нетрудно проверить, что

$$RSS+ESS=TSS,$$

Замечание. Это соотношение проще всего объяснить геометрически. Будем рассматривать $Y=(y_1, y_2, \dots, y_n)$ и $X=(x_1, x_2, \dots, x_n)$ как n -мерные векторы (или точки в n -мерном пространстве). Обозначим через H вектор $(1, 1, \dots, 1)$. Тогда при всевозможных значениях параметров β_0, β_1 вектор $\hat{Y} = \beta_0 H + \beta_1 X$ пробегает плоскость в n -мерном пространстве, натянутую на векторы H и X . Величина ESS равна квадрату длины вектора $A = Y - \hat{Y}$. Поскольку в методе наименьших квадратов минимизируется ESS , вектор \hat{Y} является ортогональной проекцией вектора Y на эту плоскость. Таким образом, вектор A ортогонален векторам H и X . Далее, ортогональной проекцией вектора Y на вектор H является $B = \bar{y}H = (\bar{y}, \bar{y}, \dots, \bar{y})$. Величины TSS и RSS равны, соответственно, квадратам длин векторов $Y-B$ и $\hat{Y}-B$. Соотношение $RSS+ESS=TSS$ поэтому есть просто теорема Пифагора, применённая к треугольнику $Y, Y-B, \hat{Y}-B$.

Величина

$$R^2 = \frac{RSS}{TSS} = 1 - \frac{ESS}{TSS}$$

называется коэффициентом детерминации и представляет собой долю дисперсии зависимой переменной Y , объяснённую при помощи оцененного уравнения регрессии. Эта величина принимает зна-

чения в пределах от 0 до 1. Крайний случай $R^2=1$ означает, что $ESS=0$, т.е. все точки наблюдений лежат на регрессионной прямой. Для случая парной регрессии R^2 равен квадрату выборочного коэффициента корреляции X и Y : $R^2 = r_{XY}^2$.

Пример. В рассматриваемом примере

$$TSS=59074.4, \quad RSS=54805.6892, \quad ESS=4268.7108,$$

так что $TSS=RSS+ESS$ и $R^2=0.9277$. Это – очень большое значение коэффициента детерминации.

3.7 Проверка значимости уравнения регрессии

При построении эконометрической модели обязательно возникает вопрос: в какой мере можно ей доверять, насколько уравнение значимо?

Для оценки значимости уравнения используется F -статистика Фишера. Такое распределение вероятности имеет случайная величина

$$F = \frac{(U_1^2 + U_2^2 + \dots + U_m^2)/m}{(W_1^2 + W_2^2 + \dots + W_l^2)/l},$$

где U_i, W_j – независимые случайные величины, распределенные по стандартному нормальному закону. Распределение Фишера обозначается $F(m, l)$, а числа m, l называются степенями свободы.

Оказывается, для парной регрессии величина

$$F = \frac{RSS}{ESS/(n-2)}$$

где n – число наблюдений в выборке, распределена по закону $F(1, n-2)$.

Разделив в этих двух последних выражениях числитель и знаменатель на TSS , получим ещё одну формулу для F

$$F = \frac{R^2}{(1-R^2)/(n-2)}.$$

Для распределения Фишера имеются *таблицы квантилей*

$F_{1-\alpha}(m, k)$, соответствующих уровню значимости α .

Пусть выбрано некоторое значение α .

- Если вычисленное значение *F*-статистики оказывается больше $F_{1-\alpha}(1, n-2)$, то оцененное уравнение регрессии считается значимым при выбранном α .

Пример. В рассматриваемом примере $F = \frac{54805.6892}{4268.7108/8} =$

102.7115, а квантиль $F_{0,9}(1;8)$ по таблице равен 3.46. Таким образом, оцененное уравнение *значимо на уровне 0.1* (и с большим запасом).

3.8 Контрольные вопросы

1. Опишите модель парной линейной регрессии.
2. Перечислите основные причины включения случайного члена в модель парной линейной регрессии.
3. Какие параметры оцениваются в этой модели?
4. Какой общий вид оцененного уравнения регрессии?
5. В чём состоит различие между теоретическим и оценённым уравнением регрессии?
6. По каким формулам рассчитываются параметры оцененного уравнения регрессии, полученные при использовании МНК?
7. Как соотносятся истинные и оценённые значения параметров регрессии?
8. Что показывает коэффициент детерминации?
9. Как проверяется значимость уравнения?
10. Насколько большую практическую значимость имеет модель парной линейной регрессии? Почему?

Литература

1. Магнус Я.Р., Катышев П.К., Пересецкий А.А. Эконометрика. Начальный курс: Учеб. пособие. – М.: Дело, 2004.
2. Бородич С.А. Эконометрика: Учеб. пособие. – Мн.: Новое знание, 2001.
3. Доугерти К. Введение в эконометрику – М.: ИНФРА-М, 2001.
4. Берндт Э.Р. Практика эконометрики: классика и современность. – М.: ЮНИТИ–ДАНА, 2005.

Приложение. Таблицы квантилей

Квантили u_p стандартного нормального закона распределения

$N(1,0)$.

p	0,9	0,95	0,975	0,99	0,995	0,999	0,9995
u_p	1,282	1,645	1,960	2,325	2,576	3,090	3,291

Квантили $\chi_p^2(k)$ закона распределения $\chi^2(k)$,

$k \backslash p$	0,01	0,025	0,05	0,1	0,9	0,95	0,975	0,99
1	0,0002	0,001	0,004	0,0158	2,71	3,84	5,02	6,63
2	0,02	0,05	0,103	0,211	4,61	5,99	7,38	9,21
3	0,115	0,216	0,352	0,584	6,25	7,81	9,35	11,3
4	0,297	0,484	0,711	1,06	7,78	9,49	11,1	13,3
5	0,554	0,831	1,15	1,61	9,24	11,1	12,8	15,1
10	2,56	3,25	3,94	4,87	16	18,3	20,5	23,2
20	7,63	9,59	10,9	12,4	28,4	31,4	34,2	37,6
30	14,3	16,8	18,5	20,6	40,3	43,8	47,0	50,9
40	22,2	24,4	26,5	29,1	51,8	55,8	59,3	63,7
50	29,7	32,4	34,8	37,7	63,2	67,5	71,4	76,2
75	49,5	53	56,1	59,8	91,1	96,2	100,8	106,4
100	70,1	74,2	77,9	82,4	118,5	124,3	129,6	135,6

Квантили $t_p(k)$ закона распределения Стьюдента $T(k)$,

$k \backslash p$	0,9	0,95	0,975	0,99	0,995
1	3,078	6,314	12,706	31,821	63,657
2	1,886	2,92	4,303	6,965	9,925
3	1,638	2,353	3,182	4,541	5,841
4	1,533	2,132	2,776	3,747	4,604
5	1,476	2,015	2,571	3,365	4,032
10	1,372	1,812	2,228	2,764	3,169
20	1,325	1,725	2,086	2,528	2,845
30	1,312	1,697	2,042	2,457	2,750
40	1,303	1,684	2,021	2,423	2,704
120	1,289	1,658	2,98	2,358	2,617

Квантили $F_{0,9}(n_1, n_2)$ закона распределения Фишера

$n_1 \backslash n_2$	1	2	3	4	5	10	15	20	30	120
1	40	8,53	5,54	4,54	4,06	3,29	3,07	2,97	2,88	2,75
2	49,5	9	5,46	4,32	3,78	2,92	2,7	2,59	2,49	2,35
3	53,6	9,16	5,39	4,19	3,62	2,73	2,49	2,38	2,28	2,13
4	55,8	9,24	5,34	4,11	3,52	2,61	2,36	2,25	2,14	1,99
30	62,2	9,46	5,17	3,82	3,17	2,16	1,87	1,74	1,61	1,41

Квантили $F_{0,95}(n_1, n_2)$ закона распределения Фишера

$n_1 \backslash n_2$	1	2	3	4	5	10	15	20	30	120
1	161	18,5	10,13	7,71	6,61	4,96	4,54	4,35	4,24	3,92
2	199	19	9,55	6,94	5,79	4,1	3,68	3,49	3,39	3,07
3	216	19,16	9,28	6,59	5,41	3,71	3,29	3,1	2,99	2,68
4	225	19,25	9,12	6,39	5,19	3,48	3,05	2,87	2,76	2,45
30	250	19,46	8,62	5,75	4,5	2,7	2,25	2,04	1,84	1,55

Критические точки d_H, d_B распределения Дарбина-Уотсона при $\alpha=0,05$

$k \backslash n$	6	7	8	9	10	20	30	40	
1	d_H	0,61	0,7	0,763	0,824	0,879	1,2	1,352	1,44
	d_B	1,4	1,356	1,332	1,32	1,32	1,41	1,49	1,544
2	d_H		0,467	0,359	0,629	0,697	1,1	1,284	1,39
	d_B		1,896	1,777	1,7	1,641	1,537	1,567	1,6
3	d_H			0,368	0,435	0,525	0,998	1,214	1,338
	d_B			2,287	2,128	2,016	1,676	1,65	1,659
4	d_H				0,296	0,376	0,894	1,143	1,285
	d_B				2,39	2,414	1,828	1,74	1,721

Критические точки $K_H - K_B$ для проверки автокорреляции методом рядов при $\alpha=0,05$

$n_- \backslash n_+$	3	4	5	6	7	8	9	10	15	20
3	-	-	-	2-	2-	2-	2-	2-	3-	3-
4		-	2-9	2-9	2-	3-	3-	3-	3-	4-
5			2-10	3-10	3-11	3-11	3-	3-	4-	5-
6				3-11	3-12	3-12	4-13	4-13	5	6-
7					3-13	4-13	4-14	5-14	6-15	6-
8						4-14	5-14	5-15	6-16	7-17
9							5-15	5-16	7-18	8-18
10								6-16	7-18	9-20
15									10-22	12-25
20										14-28

Оглавление

1. Введение	3
1.1. Предмет эконометрики	3
1.2. Модели и статистические данные	4
1.3. Метод наименьших квадратов	6
1.4. Основные этапы эконометрического моделирования	6
1.5. Контрольные вопросы	8
2. Некоторые сведения из теории вероятностей и математической статистики	9
2.1. Случайная величина и ее распределение	9
2.2. Математическое ожидание и дисперсия	11
2.3. Совместное распределение случайных величин	13
2.4. Оценка характеристик случайной величины	15
2.5. Выборочные характеристики взаимосвязи между случайными величинами	18
2.6. Контрольные вопросы	19
3. Парная линейная регрессия	19
3.1. Модель парной линейной регрессии	19
3.2. Оценивание модели	20
3.3. Теорема Гаусса-Маркова для модели парной регрессии	22
3.4. Доверительные интервалы для параметров регрессии	24
3.5. Проверка гипотез о параметрах регрессии	25
3.6. Коэффициент детерминации	26
3.7. Проверка значимости уравнения регрессии	27
3.8. Контрольные вопросы	28
Литература	29
Приложение. Таблицы квантилей	30