С. М. БОРОДАЧЁВ

ЭКОНОМЕТРИКА

Министерство образования и науки Российской Федерации
Уральский федеральный университет
имени первого Президента России Б. Н. Ельцина

С. М. БОРОДАЧЁВ

ЭКОНОМЕТРИКА

Учебное пособие

Научный редактор – д-р физ.-мат. наук, проф. О. И. Никонов

Екатеринбург
УрФУ
2011

УДК 330.43(075.8)
ББК 65в6я73
Б83

Рецензенты:
кафедра информационных систем в экономике УрГЭУ (зав. кафедрой – д-р физ.-мат. наук, проф. А. Ф. Шориков);
В. И. Зенков, д-р физ.-мат. наук, директор негосударственного образовательного учреждения «Новые технологии и программы».

Бородачёв С. М.
Б83 Эконометрика: учебное пособие / С. М. Бородачёв. Екатеринбург : УрФУ, 2011. 77 с.
ISBN 978–5–321–02037–1

Эконометрика – один из базовых предметов в современном экономическом образовании. Пособие содержит теоретический материал, упражнения, лабораторный практикум и задания для самостоятельной работы. Предназначено для студентов направлений «Экономика», «Менеджмент», «Прикладная информатика».

Библиогр.: 13 назв.

УДК 330.43(075.8)
ББК 65в6я73

ISBN 978–5–321–02037–1

Оглавление

1. Введение

Эконометрика – это совокупность методов, позволяющих придавать конкретное количественное выражение общим (качественным) закономерностям, обусловленным экономической теорией.

Пример 1.1

Будем считать (качественно ясно), что объем продаж магазина определяется его затратами на рекламу и торговыми площадями.

$$Y^n = \beta_0 + \beta_1 x_1^n + \beta_2 x_2^n + E^n, \qquad (1)$$

Y^n – объем продаж произвольного n-го магазина из магазинов одинакового профиля, x_1^n – затраты на рекламу, x_2^n – площадь магазина, E^n – особенность этого магазина (случайная величина), например выгодное расположение в центре, тогда её значение $e^n > 0$, $e^n < 0$ – магазин в подвальном этаже или с плохим менеджментом.

Эконометрика идёт дальше, пытается узнать численные значения коэффициентов $\beta_0, \beta_1, \beta_2$, чтобы, например, для нового магазина с известными x_1, x_2 предсказать объем продаж. Для этого придётся собрать данные по нескольким доступным магазинам ($n = 0, \ldots, N\text{-}1$),

$$\begin{cases} y^0, & x_1^0, & x_2^0 \\ y^1, & x_1^1, & x_2^1, \\ \ldots & \ldots & \ldots \\ y^{N-1}, & x_1^{N-1}, & x_2^{N-1} \end{cases}$$

y^n – фактическое (наблюдённое) значение объёма продаж n-го магазина, и по ним найти оценки $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$ интересующих нас величин.

Пример 1.2

Макроэкономическая модель Кейнса. Объем совокупного потребления в стране в t-ом году C^t связан с совокупным выпуском в t-ом году Y^t (национальным доходом). I^t – инвестиции. E^t – особенность года (неурожай, выплата внешнего долга и т.п.).

$$\begin{cases} C^t = \beta_0 + \beta_1 Y^t + E^t \\ Y^t = C^t + I^t \end{cases} \quad (2)$$

Для оценки коэффициентов модели потребуются данные по T годам для этой страны:

$$\begin{cases} c^0, y^0, i^0 \\ --------- \\ c^{T-1}, y^{T-1}, i^{T-1} \end{cases}$$

Модели типа (1) называют пространственными, а типа (2) – временными (различный смысл индекса наблюдения: n или t), хотя методы оценки коэффициентов в сущности одинаковы.

Конечные цели эконометрического моделирования:

– прогноз социально-экономических показателей (переменных), характеризующих анализируемую систему;

– проигрывание (имитация) различных возможных сценариев развития системы, когда изменяя поддающиеся управлению параметры, можно проследить, как будут меняться «выходные» характеристики.

2. Регрессионный анализ

Примем следующую модель изучаемой системы:

$$Y^n = \beta_0 + \beta_1 x_1^n + ... + \beta_K x_K^n + E^n, \qquad (n = 0, \ldots, N\text{-}1). \tag{3}$$

Y^n – зависимая переменная, отклик, объясняемая переменная;

$\vec{x}^n = \begin{pmatrix} 1 \\ x_1^n \\ ... \\ x_K^n \end{pmatrix}$ – независимые переменные, факторы, регрессоры, объясняющие переменные;

$\vec{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ ... \\ \beta_K \end{pmatrix}$ – вектор коэффициентов, E^n – ошибка (особенность) n-го наблюдения.

(3) можно переписать: $Y^n = \vec{x}^{nT}\vec{\beta} + E^n$, $(n = 0, \ldots, N\text{-}1)$. (4)

Будем считать E^n взаимно независимыми случайными величинами, не зависящими от x^n. $ME^n = 0$, $Cov\left(E^n, E^{n'}\right) = 0 \qquad (n \neq n')$, $DE^n = \sigma^2$ – дисперсия ошибок (одинаковость дисперсий называется гомоскедастичностью).

О названии «регрессионный анализ». Как известно, функцией регрессии Y на X называется зависимость условного математического ожидания величины Y от значений x случайной величины X: $M\left(Y \mid X = x\right)$. Для (4)

$$M(Y^n \mid \vec{X}^n = \vec{x}^n) = M(\vec{x}^{nT}\vec{\beta} \mid \vec{x}^n) + M(E^n \mid \vec{x}^n) = \vec{x}^{nT}\vec{\beta} + ME^n = \vec{x}^{nT}\vec{\beta}.$$

Поэтому в нашей модели $\vec{x}^{nT}\vec{\beta}$ есть функция регрессии. Обозначим $\breve{y} = \vec{x}^T\vec{\beta}$ – уравнение регрессии. Поэтому нахождение оценки $\hat{\vec{\beta}}$ вектора $\vec{\beta}$ называется регрессионным анализом. $\hat{y} = \vec{x}^T\hat{\vec{\beta}}$ – оценка функции регрессии.

Введём

$$\vec{Y} = \begin{pmatrix} Y^0 \\ Y^1 \\ \dots \\ Y^{N-1} \end{pmatrix}, \qquad \vec{E} = \begin{pmatrix} E^0 \\ E^1 \\ \dots \\ E^{N-1} \end{pmatrix},$$

$$z = \begin{pmatrix} \vec{x}^{0T} \\ \vec{x}^{1T} \\ \dots \\ \vec{x}^{N-1T} \end{pmatrix} = \begin{pmatrix} 1 & x_1^0 & x_2^0 & \dots & x_K^0 \\ 1 & x_1^1 & x_2^1 & \dots & x_K^1 \\ - & - & - & - & - \\ 1 & x_1^{N-1} & x_2^{N-1} & \dots & x_K^{N-1} \end{pmatrix}$$ – матрица плана, содержит значения всех факторов во всех наблюдениях. Тогда (4) можно записать:

$$\vec{Y} = z\vec{\beta} + \vec{E}. \tag{5}$$

Классическая линейная модель множественной регрессии (КЛММР):

$$\begin{cases} \vec{Y} = z\vec{\beta} + \vec{E} \\ M\vec{E} = \vec{0} \\ K_{\vec{E}} = \sigma^2 I \\ x_1, x_2, \dots, x_K \text{ - не случайные переменные} \\ \text{ранг матрицы } z = K+1 < N. \end{cases} \tag{6}$$

Если $N < K+1$, то данных в принципе мало для определения $K+1$ параметра в $\vec{\beta}$, если $N = K+1$ то так мало, что не позволит говорить о какой-либо надёжности статистических выводов.

Мы требуем ранг матрицы $z = K + 1$, чтобы матрица $z^T z$ была невырожденной, что будет использовано в (9).

2.1. Метод наименьших квадратов (МНК)

Исходные данные для оценивания – N пар наблюдений $y^n, \vec{x}^n$.

МНК-оценка вектора коэффициентов регрессии имеет вид

$$\hat{\hat{\beta}}_{МНК} = \arg\min_{\hat{\hat{\beta}}} \sum_{n=0}^{N-1} \left(y^n - \vec{x}^{nT} \hat{\hat{\beta}} \right)^2 . \qquad (7)$$

Это означает, что оценка находится из условия: сумма квадратов отклонений измеренных значений результирующей переменной от получаемых по оценённой функции регрессии должна быть минимальной.

Обозначим сумму в правой части (7) за $Q(\hat{\hat{\beta}})$.

$$Q\left(\hat{\hat{\beta}}\right) = \left(\vec{y} - z\hat{\hat{\beta}}\right)^T \left(\vec{y} - z\hat{\hat{\beta}}\right) = \vec{y}^T\vec{y} - \vec{y}^T z\hat{\hat{\beta}} - \hat{\hat{\beta}}^T z^T\vec{y} + \hat{\hat{\beta}}^T z^T z\hat{\hat{\beta}} =$$
$$\vec{y}^T\vec{y} - \vec{y}^T z\left(z^T z\right)^{-1} z^T\vec{y} + \left(z^T\vec{y} - z^T z\hat{\hat{\beta}}\right)^T \left(z^T z\right)^{-1} \left(z^T\vec{y} - z^T z\hat{\hat{\beta}}\right). \qquad (8)$$

В полученном выражении только последнее слагаемое зависит от $\hat{\hat{\beta}}$ и оно вследствие симметрии (квадратичности) всегда неотрицательно. Поэтому минимум $Q(\hat{\hat{\beta}})$ будет достигаться, когда оно равно нулю, для этого нужно: $z^T\vec{y} - z^T z\hat{\hat{\beta}} = \vec{0}$, отсюда

$$\hat{\hat{\beta}}_{МНК} = \left(z^T z\right)^{-1} z^T\vec{y} \qquad (9)$$

– МНК-оценка вектора коэффициентов регрессии. $P = \left(z^T z\right)^{-1}$ называется матрицей ошибок.

Парная регрессия

Пусть $\breve{y} = \beta_0 + \beta_1 x$ – линейная функция одной переменной. Тогда (9) даёт

$$\hat{\beta}_0 = \frac{\overline{y}\sum (x^n)^2 - \overline{x}\sum x^n y^n}{\sum (x^n)^2 - N\overline{x}^2}, \qquad \hat{\beta}_1 = \frac{\sum x^n y^n - N\overline{xy}}{\sum (x^n)^2 - N\overline{x}^2}, \qquad (10)$$

где $\overline{x} = \frac{1}{N}\sum_{n=0}^{N-1} x^n$, $\overline{y}$ – аналогично – выборочные средние.

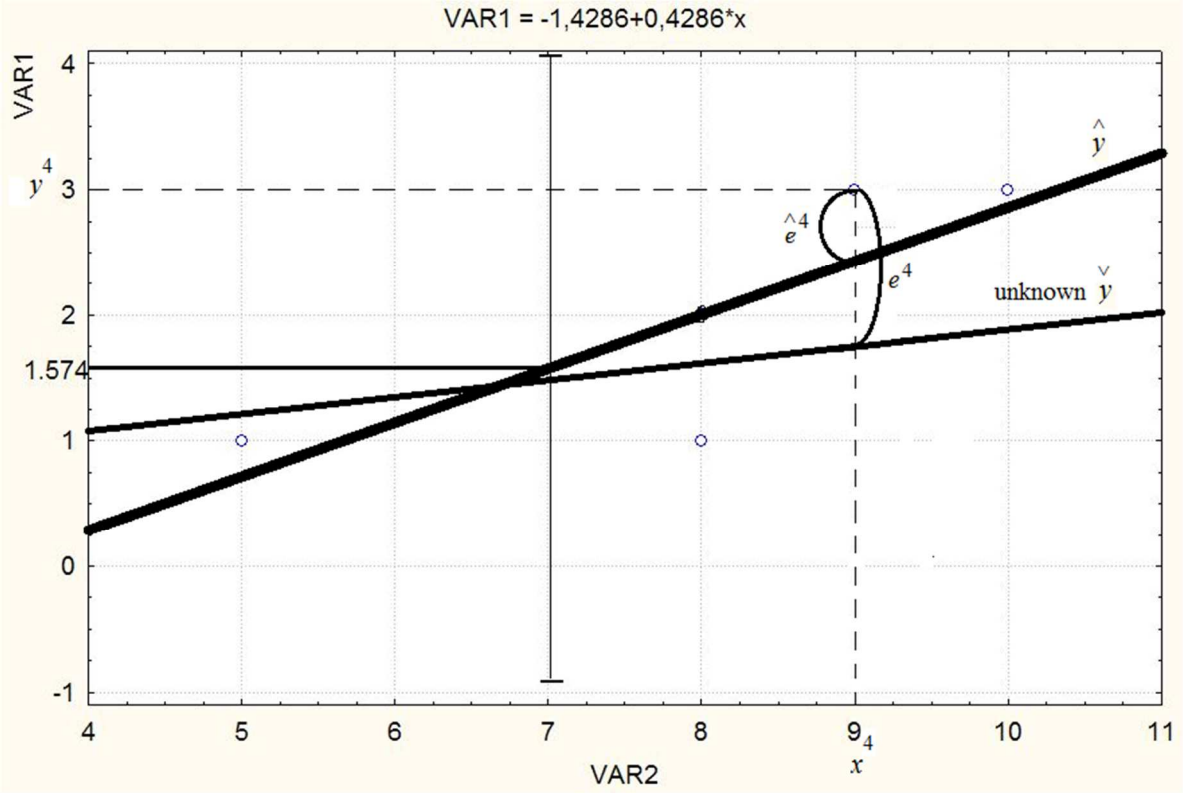
Пример 2.1

По данным 5 магазинов

Затраты на рекламу x^n (тыс. руб.)	8	10	5	8	9
Объем продаж y^n (сотен тыс. руб.)	1	3	1	2	3

найти оценку линейной функции регрессии объёма продаж на затраты на рекламу, построить ее график на фоне диаграммы рассеяния (корреляционного поля).

По (10) получим $\hat{\beta}_0 = -1.429$, $\hat{\beta}_1 = 0.429$ а значит, $\hat{y} = -1.429 + 0.429x$.



Рекуррентный МНК

При появлении дополнительных последовательных наблюдений с номерами N, N+1, …, (NN – 1) нет необходимости пересчитывать оценку вектора коэффициентов регрессии по формуле (9) каждый раз со всеми данными.

Переобозначим все величины, основанные на N данных (с номерами 0, …, N – 1) в формуле (9):

$$\hat{\vec{\beta}}_{МНК}(N) = \left(z(N)^T z(N)\right)^{-1} z(N)^T \vec{y}(N).$$

Аналогично на основе $N + 1$ данных:

$$\hat{\vec{\beta}}_{МНК}(N+1) = \left(z(N+1)^T z(N+1)\right)^{-1} (z(N)^T \vdots \vec{x}^N) \begin{pmatrix} \vec{y}(N) \\ y^N \end{pmatrix} = \tag{11}$$

$$= P(N+1)(z(N)^T \vec{y}(N) + \vec{x}^N y^N);$$

$$P(N+1) = [(z(N)^T \vdots \vec{x}^N) \begin{pmatrix} z(N) \\ \vec{x}^{NT} \end{pmatrix}]^{-1} = [z(N)^T z(N) + \vec{x}^N \vec{x}^{NT}]^{-1}.$$

Применяя тождество $[P^{-1} + M^T Q^{-1} M]^{-1} = P - PM^T [MPM^T + Q]^{-1} MP$, получим

$$P(N+1) = P(N) - \frac{P(N)\vec{x}^N \vec{x}^{NT} P(N)}{1 + \vec{x}^{NT} P(N) \vec{x}^N}. \tag{12}$$

Подставляя это в (11), находим

$$\hat{\vec{\beta}}_{МНК}(N+1) = \hat{\vec{\beta}}_{МНК}(N) + \frac{P(N)\vec{x}^N [y^N - \vec{x}^{NT} \hat{\vec{\beta}}_{МНК}(N)]}{1 + \vec{x}^{NT} P(N) \vec{x}^N}. \tag{13}$$

Эти формулы совпадают с формулами оценивания вектора коэффициентов регрессии как вектора состояния системы фильтром Калмана.

Для начала рекуррентного процесса (13) – (12) необходимо рассчитать $P(N)$ и $\hat{\vec{\beta}}_{МНК}(N)$ по (9) по как минимум $N = K + 1$ данным, затем добавлять новые данные по одному. Так будет получаться точная МНК-оценка в любой момент. Если требуется получение оценки вектора коэффициентов регрессии начиная с первых данных, то для $\hat{\vec{\beta}}_{МНК}(0)$ следует использовать разумное начальное приближение, а в качестве $P(0)$ единичную матрицу. С ростом N оценка будет сближаться с точной.

2.2. Приложение: *CAPM* (модель ценообразования активов капитала)

R^t – рентабельность определённого актива в *t*-ом году

$$R^t = \frac{приход^t - расход^t}{расход^t} = \frac{D^t + P^t - p^{t-1}}{p^{t-1}}.$$

D^t – дивиденды в *t*-м году, P^t – цена актива в конце года – случайные величины, p^t – цена покупки актива в начале года (не случайная величина). Таким образом, R^t – случайная объясняемая величина.

От чего она могла бы зависеть? R_m^t – рыночная рентабельность (рентабельность рыночного портфеля, общая конъюнктура, платёжеспособность инвесторов, состояние глобальной экономики) – объясняющая величина. Если все доходности рассматривать относительно без рисковой ставки r_f^t (доходность государственных облигаций), то простейшая модель (*CAPM*):

$$R^t - r_f^t = \beta(R_m^t - r_f^t) + E^t.$$

Слагаемое $\beta(R_m^t - r_f^t)$ называется премией за риск. E^t – особенность года для доходности этого актива. Считаем, что она не зависит от случайного фактора R_m^t, поэтому формула для оценки коэффициента регрессии – как по МНК (подробнее см. [13]) для модели без константы:

$$\hat{\beta} = \frac{\sum z^t z_m^t}{\sum \left(z_m^t\right)^2}, \qquad (13а)$$

$$z^t = r^t - r_f^t,\ z_m^t = r_m^t - r_f^t.$$

Пример 2.2 (J. K. Shim, J. G. Siegel. Managerial finance. McGraw-Hill, Inc, 1986, p. 249-250. Там ошибка)

Для акций некоторой компании собраны данные.

Год	z^t, %	z_m^t, %
19*X*1	-11	4
19*X*2	-2	2
19*X*3	1	6
19*X*4	4	14
19*X*5	6	9

Пусть в 19X6 году $r_f^{19X6} = 4$ %, рынок в целом показал: $r_m^{19X6} = 10.44$ %. Какова оценка ожидаемой рентабельности этого актива? Используя формулу для $\hat{\beta}$, легко получить $\widehat{MR^{19X6}} = 5.3$ %. Поэтому держатели этих акций, проводя такие подсчёты, будут ожидать такой доходности. С этим связано следующее важное понятие.

Стоимость капитала (*cost of capital*) (для фирмы) – это рентабельность, необходимая для поддержания рыночной стоимости фирмы (или цены акций фирмы). Ещё её называют минимально требуемой рентабельностью.

Менеджеры должны знать стоимость капитала своей фирмы, например, при планировании инвестиций она влияет на дисконтную ставку при расчёте *NPV*. Фирма ведь должна выплачивать свою стоимость капитала в виде дивидендов, поэтому будущие её доходы должны дисконтироваться по не меньшей ставке.

2.3. Коэффициент детерминации

Как измерить качество оценённой модели $\hat{y}^n = \vec{x}^{nT}\hat{\vec{\beta}}_{МНК}$, описывающей данные $\vec{y}$? Можно по доле общей изменчивости данных $ss_{tot} = \sum_n \left(y^n - \overline{y}\right)^2$, объяснённой моделью $ss_{mod} = \sum_n \left(\hat{y}^n - \overline{y}\right)^2$.

Коэффициентом детерминации называется отношение

$$R^2 = \frac{ss_{mod}}{ss_{tot}} = 1 - \frac{ss_{res}}{ss_{tot}}. \qquad (14)$$

Последнее равенство следует из тождества

$\sum_n \left(y^n - \overline{y}\right)^2 = \sum_n \left(\hat{y}^n - \overline{y}\right)^2 + \sum_n \left(y^n - \hat{y}^n\right)^2$, или иначе

$$ss_{tot} = ss_{mod} + ss_{res}, \tag{15}$$

которое следует из (9), в предположении модели с константой.

Действительно, введём $\vec{\bar{y}} = \begin{pmatrix} \bar{y} \\ \bar{y} \\ ... \\ \bar{y} \end{pmatrix}$,

$$ss_{tot} = (\vec{y} - \vec{\bar{y}})^T (\vec{y} - \vec{\bar{y}}) = (\hat{\vec{y}} - \vec{\bar{y}} + \hat{\vec{e}})^T (\hat{\vec{y}} - \vec{\bar{y}} + \hat{\vec{e}}) =$$
$$= (\hat{\vec{y}} - \vec{\bar{y}})^T (\hat{\vec{y}} - \vec{\bar{y}}) + \{(\hat{\vec{y}} - \vec{\bar{y}})^T \hat{\vec{e}} + э.с.\} + \hat{\vec{e}}^T \hat{\vec{e}},$$

$\vec{\bar{y}} = z \begin{pmatrix} \bar{y} \\ 0 \\ ... \\ 0 \end{pmatrix} = z\vec{a}$, где z – матрица плана для модели с константой β_0. Тогда

$ss_{tot} = ss_{mod} + \{(\hat{\vec{\beta}}^T - \vec{a}^T) z^T \hat{\vec{e}} + э.с.\} + ss_{res}$. Рассмотрим

$z^T \hat{\vec{e}} = z^T (\vec{y} - z\hat{\vec{\beta}}) = z^T \vec{y} - z^T z (z^T z)^{-1} z^T \vec{y} = \vec{0}$, что и даёт (15).

Из (14) очевидно $0 \le R^2 \le 1$.

Продолжение примера 2.1

$ss_{tot} = (1-2)^2 + ... + (3-2)^2 = 4;$

$ss_{res} = \left[1 - \left(-1.429 + 0.429 * 8\right)\right]^2 + ... + \left[3 - \left(-1.429 + 0.429 * 9\right)\right]^2 = 1.429;$

$R^2 = 0.643$ (64.3 %).

Иногда определяют исправленный коэффициент детерминации $\bar{R}^2 = R^2_{adj}$:

$$\bar{R}^2 = 1 - \frac{N-1}{N-K-1}(1 - R^2). \tag{16}$$

Смысл этого определения будет разъяснён в следующем разделе.

2.4. Оценка дисперсии ошибок

Отметим, что по (5), даже при одинаковых (фиксированных) $\vec{x}^0, \vec{x}^1, ..., \vec{x}^{N-1}$ (т. е. матрице плана z), вектор $\vec{Y}$ случаен, т.к. случаен $\vec{E}$, поэтому и

$$\hat{\vec{\mathrm{B}}}_{МНК} = \left(z^T z\right)^{-1} z^T \vec{Y} \qquad (17)$$

случаен (оценки коэффициентов регрессии), и остаточная сумма квадратов $SS_{res} = \sum_n \left(Y^n - \vec{x}^{nT} \hat{\vec{\mathrm{B}}}_{МНК}\right)^2$ случайна. Рассмотрим вектор остатков (*residuals*):

$$\hat{\vec{E}} = \vec{Y} - \hat{\vec{Y}} = \vec{Y} - z\hat{\vec{\mathrm{B}}}_{МНК} = z\vec{\beta} + \vec{E} - z\left(z^T z\right)^{-1} z^T \left(z\vec{\beta} + \vec{E}\right) =$$

$$= z\vec{\beta} + \vec{E} - z\left(z^T z\right)^{-1} z^T z\vec{\beta} - z\left(z^T z\right)^{-1} z^T \vec{E} = \left[I - z\left(z^T z\right)^{-1} z^T \right]\vec{E} = H\vec{E}.$$

Легко показать, что H симметрична и идемпотентна ($H^2 = H$).

Найдём

$$MSS_{res} = M(H\vec{E})^T (H\vec{E}) = M\vec{E}^T H\vec{E} = \sum_{n,n'} h_{nn'} ME^n E^{n'} =$$

$$= \sigma^2 \sum_n h_{nn} = \sigma^2 SpH = \sigma^2 [SpI - Sp(z(z^T z)^{-1} z^T)] = \sigma^2 [N - Sp((z^T z)^{-1} z^T z)] =$$

$$= \sigma^2 [N - (K+1)].$$

Это означает, что

$$\hat{\sigma}^2 = \frac{SS_{res}}{N - K - 1} \qquad (18)$$

– несмещённая оценка дисперсии ошибок. $\hat{\sigma} = \sqrt{\hat{\sigma}^2}$ – оценка стандартного отклонения ошибок, называется ещё *стандартной ошибкой аппроксимации*.

Продолжение примера 2.1

$\hat{\sigma}^2 = \frac{1.429}{5-1-1} = 0.476$, $\hat{\sigma} = \sqrt{0.476} = 0.69$ сотен тыс. руб. – характеризует силу влияния особенности магазина (прочих факторов, кроме затрат на рекламу) на объём продаж.

Если качество оценённой модели измерять присущей ей стандартной ошибкой аппроксимации (куда включены все недочёты модели), то вместо (14) более естественным выглядит определение $\bar{R}^2 = 1 - \frac{\hat{\sigma}^2}{\hat{\sigma}_Y^2}$, которое ведёт к (16).

2.5. Свойства оценок коэффициентов регрессии

Теорема: в КЛММР МНК-оценка (17) есть несмещённая и состоятельная оценка истинного вектора $\vec{\beta}$. Доказательство несмещённости:

$$M\hat{\vec{\mathrm{B}}}_{МНК} = M\left(z^T z\right)^{-1} z^T \left(z\vec{\beta} + \vec{E}\right) = M\left(z^T z\right)^{-1} z^T z\vec{\beta} + M\left(z^T z\right)^{-1} z^T \vec{E} =$$

$$= \vec{\beta} + \left(z^T z\right)^{-1} z^T M\vec{E} = \vec{\beta}.$$

Ковариационная матрица вектора оценок

$$K_{\hat{\vec{\mathrm{B}}}} = M(\hat{\vec{\mathrm{B}}}_{МНК} - \vec{\beta})(\hat{\vec{\mathrm{B}}}_{МНК} - \vec{\beta})^T = M((z^T z)^{-1} z^T \left(z\vec{\beta} + \vec{E}\right) - \vec{\beta})((z^T z)^{-1} z^T \left(z\vec{\beta} + \vec{E}\right) - \vec{\beta})^T =$$

$$= M((z^T z)^{-1} z^T \vec{E})((z^T z)^{-1} z^T \vec{E})^T = \left(z^T z\right)^{-1} z^T M(\vec{E}\vec{E}^T) z\left(z^T z\right)^{-1} = \sigma^2 \left(z^T z\right)^{-1}. \quad (19)$$

Поскольку дисперсия ошибок обычно неизвестна, то её заменяют оценкой по (18): $\hat{K}_{\hat{\vec{\mathrm{B}}}} = \hat{\sigma}^2 \left(z^T z\right)^{-1} = \hat{\sigma}^2 P$.

Диагональные элементы этой матрицы дают оценки дисперсий оценок коэффициентов регрессии $\hat{\mathrm{B}}_0, ..., \hat{\mathrm{B}}_K$: $\hat{\sigma}_k^2 = \widehat{D\hat{\mathrm{B}}_k} = \left(\hat{K}_{\hat{\vec{\mathrm{B}}}}\right)_{kk} = \hat{\sigma}^2 P_{kk}$. Значит, $\hat{\sigma}_k = \sqrt{\hat{\sigma}_k^2} = \hat{\sigma}\sqrt{P_{kk}}$ является стандартной ошибкой (*standard error*) (оценки) коэффициента регрессии β_k.

Продолжение примера 2.1

Было: $P = \begin{pmatrix} 4.771 & -0.571 \\ -0.571 & 0.0714 \end{pmatrix}$. Поэтому $\widehat{\sigma_0} = 0.69\sqrt{4.771} = 1.507$,

$\widehat{\sigma_1} = 0.69\sqrt{0.0714} = 0.184$. С указанием ошибок оценок коэффициентов, результат регрессионного анализа имеет вид:

$$\begin{array}{lll} \hat{y} & = & -1.429 + 0.429x \\ (S.E.) & (1.507) & (0.184). \end{array}$$

Такие величины ошибок в сравнении с величинами самих оценок говорят о низкой точности определения коэффициентов регрессии.

2.6. Нормальная КЛММР

Здесь к требованиям КЛММР (6) добавляется требование нормальной распределённость ошибок: $\vec{E} \sim N\left(\vec{0}, \sigma^2 I\right)$.

Используя обобщённую теорему Фишера, можно показать $\frac{\hat{\mathrm{B}}_k - \beta_k}{\hat{\Sigma}_k} \sim T_{N-K-1}$, что даёт возможность строить доверительные интервалы и проверять гипотезы о коэффициентах регрессии.

Доверительные интервалы для коэффициентов регрессии

Обычным образом получаем

$$P\left\{\widehat{\mathrm{B}_k} - \widehat{\Sigma_k} t_{1-\frac{\alpha}{2},N-K-1} \le \beta_k \le \widehat{\mathrm{B}_k} + \widehat{\Sigma_k} t_{1-\frac{\alpha}{2},N-K-1}\right\} = \gamma, \qquad (20)$$

γ – доверительная вероятность, $\alpha = 1 - \gamma$, $t_{p,n}$ – квантиль порядка p распределения Стьюдента с n степенями свободы.

Продолжение примера 2.1

Построить доверительный интервал с $\gamma = 0.95$ для углового коэффициента (наклона) прямой регрессии.

$N = 5$, $K = 1$, $\left(0.429 - 0.184 \cdot 3.182 \le \beta_1 \le 0.429 + 0.184 \cdot 3.182\right)$;

$\left(-0.09 \le \beta_1 \le 1.08\right)$ – 95 %-ый доверительный интервал.

Проверка гипотез о коэффициентах регрессии

Обычно интерес представляет гипотеза $H_0 : \beta_k = 0$, <=> x_k никак не влияет на y и её можно вообще исключить из модели. Альтернативная гипотеза $H_1 : \beta_k \ne 0$. Уровень значимости данных против H_0:

$$sl = 2\left[1 - \Phi^{Cm}_{N-K-1}\left(\left|\frac{\widehat{\beta_k}}{\widehat{\sigma_k}}\right|\right)\right]$$

Продолжение примера 2.1

$$sl\left(\widehat{\beta_0}\ \textit{против гипотезы}\ \beta_0 = 0\right) = 2\left[1 - \Phi^{Cm}_{3}\left(\left|\frac{-1.429}{1.507}\right|\right)\right] = 0.4132$$ – данные не

значимы против H_0, β_0 может быть равен 0.

Аналогично $sl\left(\widehat{\beta_1}\ \textit{против}\ \beta_1 = 0\right) = 2\left[1 - \Phi^{Cm}_{3}\left(2.33\right)\right] = 0.1027$ – значимость выше, но недостаточна, чтобы опровергнуть H_0.

Проверка гипотезы обо всех коэффициентах сразу:

$H_0 : \beta_1 = \beta_2 = \ldots = \beta_K = 0, \beta_0 \neq 0$, т.е. об отсутствии линейной связи между *y* и x_k (всеми факторами), H_1 – это не так. Уровень значимости

$$sl = P\left\{F_{K,N-K-1} \geq \frac{R^2\left(N-K-1\right)}{\left(1-R^2\right)K}\right\}.$$

Продолжение примера 2.1

$$sl = P\left\{F_{1,3} \geq \frac{0.6428 \cdot 3}{\left(1-0.6428\right)\cdot 1}\right\} = P\left\{F_{1,3} \geq 5.39\right\} = 1 - \Phi^{F}_{1,3}\left(5.39\right) = 0.1027$$ –

ожидаемый результат.

Проверка нормальности распределения ошибок

Вектор остатков (*residuals*): $\widehat{\vec{E}} = H\vec{E}$ пропорционален вектору ошибок, а значит, имеет нормальное распределение, если ошибки нормальны. Поэтому, если отвергнем нормальность остатков, то нет и нормальности ошибок.

Можно применить простые визуальные методы проверки нормальности выборки остатков:

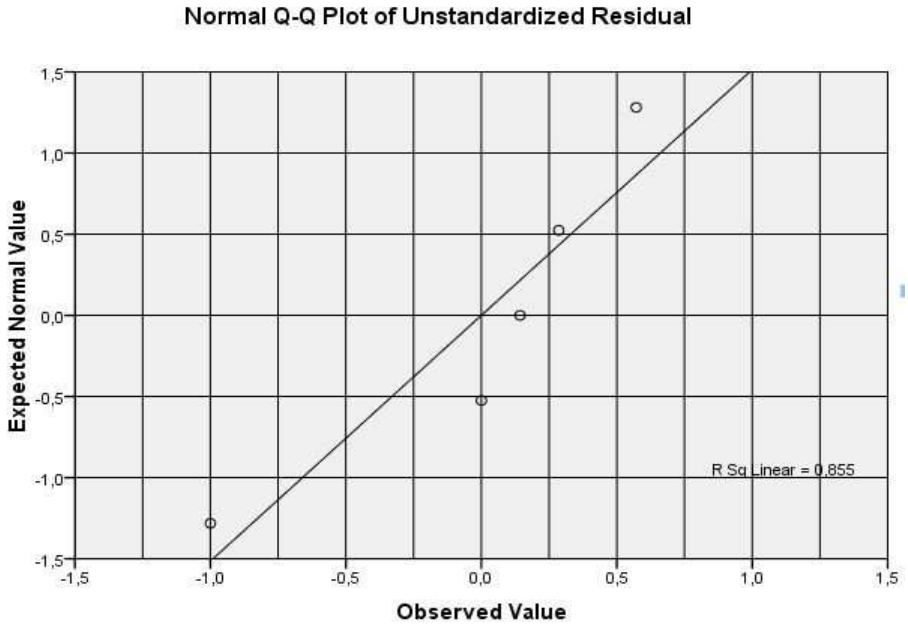
1. Построение гистограммы (похожесть на гауссовский колокол);

2. Если распределение нормально, то квантильное (нормальное стандартное) преобразование по оси ординат эмпирической функции распределения должно быть близко к прямой, что легко распознать визуально.

Продолжение примера 2.1

Остатки $\hat{e}^0 = -1, \hat{e}^1 = 0.143, \hat{e}^2 = 0.286, \hat{e}^3 = 0, \hat{e}^4 = 0.571.$

$u^0 = u_{0.1} = -1.28, u^1 = u_{0.5} = 0, u^2 = u_{0.7} = 0.524, u^3 = u_{0.3} = -0.524, u^4 = u_{0.9} = 1.282.$



Следует считать точки не лежащими на прямой, остатки не нормальными, значит и ошибки тоже.

2.7. Точечный и интервальный прогноз, основанный на КЛММР

Пусть наряду с данными $\begin{matrix} \vec{x}^0 & y^0 \\ \vdots & \vdots \\ \vec{x}^{N-1} & y^{N-1} \end{matrix}$ имеем $\vec{x}^N$ – новое значение вектора факторов, а соответствующее y^N нам неизвестно. Требуется построить наилучший (в смысле среднего квадрата ошибки), линейный по $\vec{y}$ и несмещённый прогноз для y^N.

Результат будет: наилучший прогноз даётся по оценённой по МНК функции регрессии на аргументе $\vec{x}^N$, т.е. $y_*^N = \vec{x}^{N^T}\hat{\vec{\beta}}_{МНК}$.

В нормальной модели можно построить доверительный интервал для y^N:

$$\left(y_*^N - t_{1-\frac{\alpha}{2},N-K-1}\hat{\sigma}\sqrt{\vec{x}^{N^T}P\vec{x}^N+1} \le y^N \le y_*^N + t_{1-\frac{\alpha}{2},N-K-1}\hat{\sigma}\sqrt{\vec{x}^{N^T}P\vec{x}^N+1} \right),$$

обозначения как в (20).

Продолжение примера 2.1

Найти точечный и интервальный (95%) прогноз объёма продаж для нового магазина, который затратил на рекламу 7 тыс. руб.:

$\vec{x}^5 = \begin{pmatrix}1\\7\end{pmatrix}$, $y_*^5 = \vec{x}^{5^T}\hat{\vec{\beta}} = \begin{pmatrix}1 & 7\end{pmatrix}\begin{pmatrix}-1.429\\0.429\end{pmatrix} = 1.574$ сотен тыс. руб. – точечный прогноз.

$$\vec{x}^{N^T}P\vec{x}^N = \frac{1}{N} + \frac{\left(x^N - \bar{x}\right)^2}{\sum_n \left(x^n\right)^2 - N\bar{x}^2} = \frac{1}{5} + \frac{(7-8)^2}{14} = 0.271, \qquad (21)$$

поэтому $\left(1.574 - 3.182\cdot 0.69\cdot\sqrt{1.271} \le y^5 \le 1.574 + 3.182\cdot 0.69\cdot\sqrt{1.271}\right)$, или $\left(-0.90 \le y^5 \le 4.05\right)$ сотен тыс. руб. – интервальный прогноз. Смотри рисунок в разделе 2.1.

2.8. Коэффициент эластичности

Истинная функция регрессии: $\breve{y} = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + \ldots + \beta_K x_K$. Пусть k-й фактор получил приращение $\widetilde{x_k} = x_k + \Delta x_k$, тогда $\tilde{\breve{y}} = \beta_0 + \beta_1 x_1 + \ldots + \beta_k (x_k + \Delta x_k) + \ldots + \beta_K x_K = \breve{y} + \beta_k \Delta x_k$.

Преобразуем $\Delta\tilde{\breve{y}}_{x_k} = \tilde{\breve{y}} - \breve{y} = \beta_k \Delta x_k$, $\frac{\Delta\tilde{\breve{y}}_{x_k}}{\breve{y}} 100\,\% = \frac{\beta_k x_k}{\breve{y}}\left(\frac{\Delta x_k}{x_k} 100\,\%\right)$. Если выбрать проценты приращения k-о фактора (выражение в скобках) = 1 %, то

левая часть (проценты приращения результирующей переменной) окажется равной

$$Э_{yx_k} = \frac{\beta_k x_k}{\breve{y}} \qquad (22)$$

– коэффициент эластичности *y* по x_k.

Значит, коэффициент эластичности показывает, на сколько % изменится величина результирующего признака *y* при изменении независимой переменной x_k на 1 % (при неизменных остальных переменных).

Если заменить: $\beta_k \to \hat{\beta}_k$, $x_k \to \overline{x_k}$, $\breve{y} \to \overline{y}$, то получим средний коэффициент эластичности $\overline{Э}_{yx_k} = \hat{\beta}_k \frac{\overline{x_k}}{\overline{y}}$.

Продолжение примера 2.1

Рассчитать средний коэффициент эластичности продаж по затратам на рекламу.

$\overline{Э}_{yx} = 0.429\frac{8}{2} = 1.72\ \%$, т.е. при увеличении затрат на рекламу на 1 % объём продаж вырастет на 1.72 % (нужно вкладываться в рекламу).

2.9. Упражнения

Упражнение 2.1

Проверить формулы (5), (8), (9).

Упражнение 2.2

Вывести из общей формулы МНК-оценки вектора коэффициентов регрессии формулы оценок коэффициентов парной регрессии. Получить результаты примера 2.1. Вывести формулу (13а).

Упражнение 2.3

Проверить связь для парной регрессии: $\hat{\beta}_1 = \hat{\rho}\frac{\hat{\sigma}_Y}{\hat{\sigma}_X}$.

Упражнение 2.4

Показать, что оценённая прямая парной регрессии всегда проходит через точку с координатами, равными выборочным средним.

Упражнение 2.5

Получить результаты примера 2.2.

Упражнение 2.6

Проверить формулы (16), (19).

Упражнение 2.7

Показать, что в однофакторном регрессионном анализе $R^2 = (\hat{\rho})^2$.

Упражнение 2.8

По $N = 20$ оценена функция регрессии $\hat{y} = 8 - 7x$. Известно также $\hat{\rho} = -0.5$.

- Найти выражение стандартной ошибки аппроксимации через R^2 и $\hat{\sigma}_Y$.
- Найти стандартную ошибку коэффициента регрессии.
- Проверить гипотезу о коэффициенте регрессии.
- Построить 90 процентный доверительный интервал для коэффициента регрессии.

Упражнение 2.9

Изучается зависимость потребления материалов y от объёма производства продукции x. По 20 наблюдениям было получено: $\hat{y} = 3 + 2x$. Фактическое значение t-критерия 6.48. Найти коэффициент детерминации.

Упражнение 2.10

По совокупности 6 предприятий торговли изучается зависимость между признаками: x – цена на товар, тыс. руб.; y – прибыль торгового предприятия, млн. руб. При оценке регрессионной модели были получены следующие промежуточные результаты:

$$\Sigma(y^n - \hat{y}^n)^2 = 39000; \Sigma(y^n - \overline{y})^2 = 120000.$$

- Найти коэффициент детерминации.
- Найти оценку коэффициента корреляции.
- Проверить значимость всего уравнения регрессии.

Упражнение 2.11

Показать для парной регрессии формулу (21).

Упражнение 2.12

По 20 фермам области получена информация, представленная в таблице:

Показатель	Среднее значение	Коэффициент вариации (%)
Урожайность (ц/га)	27	20
Внесено удобрений (кг/га)	5	15

Фактическое значение F-критерия Фишера составило 45.

- Найти коэффициент детерминации.
- Записать оценённое уравнение линейной регрессии.
- Проверить его значимость.
- Найти средний коэффициент эластичности.
- Найти 95 процентный доверительный интервал прогнозируемого значения урожайности в предположении внесения количества удобрений на 10 % больше среднего уровня по 20 фермам области.

Упражнение 2.13

По 19 предприятиям оптовой торговли изучается зависимость объёма реализации y от размера торговой площади x_1 и товарных запасов x_2 . Были получены следующие варианты уравнений регрессии:

$$\hat{y} = 25 + 15x_1 \qquad \hat{\rho}^2 = 0.9$$

$$\hat{y} = 42 + 27x_2 \qquad \hat{\rho}^2 = 0.84$$

$$\hat{y} = 30 + 10x_1 + 8x_2 \quad R^2 = 0.92$$

Проанализируйте тесноту связи результата с каждым из факторов. Выберите наилучшее уравнение регрессии.

Упражнение 2.14

Для изучения рынка жилья в городе по данным о 16 коттеджах было построено уравнение множественной регрессии:

$\hat{y} = 21.1 - 6.2x_1 + 0.95x_2 + 3.57x_3 \qquad R^2 = 0.7$

(*S.E.*) (1.8) (0.54) (0.83)

где y – цена объекта, тыс. \$;

x_1 – расстояние до центра города;

x_2 – полезная площадь объекта, кв. м.;

x_3 – число этажей в доме.

- Проверьте гипотезу о независимости цены от расстояния до центра города.
- Проверьте гипотезу о независимости цены от полезной площади объекта.
- Проверьте гипотезу о всем уравнении.

Упражнение 2.15

В результате исследования факторов, определяющих экономический рост, по 73 странам оценено следующее уравнение регрессии:

$\hat{G} = 1.4 - 0.52P + 0.17S + 11.16I - 0.38D - 4.75In \qquad R^2 = 0.6$

(*t*) (-5.9) (4.34) (3.91) (-0.79) (-2.7)

где G – темпы роста среднедушевого ВВП в % к базисному периоду;

P – реальный среднедушевой ВВП;

S – бюджетный дефицит, % к ВВП;

I – объем инвестиций, % к ВВП;

D – внешний долг, % к ВВП;

In – уровень инфляции, %.

До получения результатов этого исследования ваш однокурсник заключил с вами пари, что эмпирические результаты докажут наличие обратной связи между темпами роста среднедушевого ВВП и объёмом внешнего долга. Выиграл ли пари ваш однокурсник?

3. Мультиколлинеарность

Мультиколлинеарность (строгая) – это линейная зависимость между столбцами матрицы плана

$$z = \begin{pmatrix} 1 & x_1^0 & x_2^0 & \cdots & x_K^0 \\ 1 & x_1^1 & x_2^1 & \cdots & x_K^1 \\ \ldots & - & - & - & - \\ 1 & x_1^{N-1} & x_2^{N-1} & \cdots & x_K^{N-1} \end{pmatrix}.$$

Это означает, что одна из объясняющих переменных есть линейная комбинация остальных объясняющих переменных.

Пример 3.1

Перепись населения, собираются сведения: x_1 – число мальчиков в семье, x_2 – число девочек, x_3 – число детей в семье. Ясно, $\forall n$ $x_3^n = x_1^n + x_2^n \Rightarrow z^{<3>} = z^{<1>} + z^{<2>}$, третий столбец равен сумме первого и второго.

Линейная зависимость ведёт к тому, что $r_z(\textit{ранг матрицы } z) < K+1$, $\det\left(z^T z\right) = 0$, $\nexists \left(z^T z\right)^{-1}$ – нет обратной матрицы, и по формуле (9) оценку вектора коэффициентов регрессии не найти.

Реальная (или частичная) мультиколлинеарность возникает в случае существования тесных статистических связей между объясняющими переменными. Скажем, обследуются *N* фермерских хозяйств и x_1 – число тракторов, x_2 – число борон. При этом $\hat{\rho}(X_1, X_2) >\approx 0.75$ уже плохо, так как ведёт к плохой обусловленности матрицы $z^T z$, $\det\left(z^T z\right) \approx 0$, и сравним с ошибками округления, в результате обращение неверно и оценка вектора коэффициентов регрессии непредсказуема.

Как распознать отягощённость мультиколлинеарностью?

- Предварительно полезно рассчитать $\hat{\rho}(X_i, X_j)$ (нет ли больших значений).

- Странные знаки $\widehat{\beta_k}$ и неоправданно большие абсолютные значения.
- Большинство или даже все $\widehat{\beta_k}$ оказываются незначимыми, а уравнение в целом значимо.

Методы устранения мультиколлинеарности

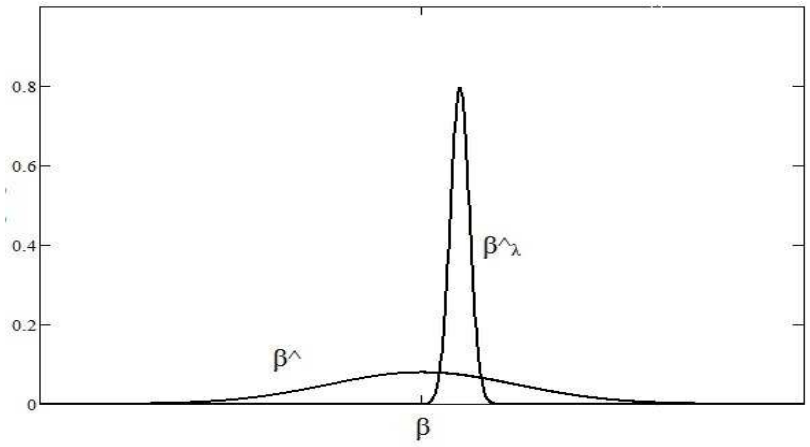
1. Ридж-регрессия. Вместо (9) используется формула

$$\overrightarrow{\hat{\beta}_\lambda} = (z^T z + \lambda I)^{-1} z^T \vec{y}, \qquad (23)$$

где $0.1 < \lambda < 0.4$.

Добавление гребня λI (по диагонали): $\begin{pmatrix} \lambda & .... & 0 \\ ... & ... & ... \\ 0 & ... & \lambda \end{pmatrix}$ делает матрицу в скобках уже

хорошо обусловленной и вычисления пройдут устойчиво, но оценка (23) будет смещённой. Однако смещённая оценка с малой дисперсией будет лучше (с меньшей среднеквадратической погрешностью), чем несмещённая, но с большой дисперсией (помним, что $\sigma_{\beta_k} \sim \left(z^T z\right)^{-1}_{kk}$).



2. Отбор наиболее существенных объясняющих переменных. Так как мультиколлинеарность вызывается дублированием информации от сильно связанных переменных, то нужно исключить «повторяющиеся» переменные, что

упростит модель и исключит мультиколлинеарность. Одну из таких процедур рассмотрим в следующем примере.

3. Если явление мультиколлинеарности возникает при увеличении числа наблюдений, то средством борьбы с ним может стать рекуррентный МНК. При этом получение исходной *P(N)* идёт без проблем, а в дальнейшем не приходится обращать плохо обусловленные матрицы. Если и получение исходной *P(N)* затруднительно, то можно попробовать поменять местами наблюдения или добавить при получении *P(N)* «гребень».

Пример 3.2 (продолжение примера 2.1)

Вы решили уточнить прогноз продаж, включив ещё 3 регрессора: x_2 – закуп товара магазином за тот же период (тыс. руб.), x_3 – торговые площади магазина (м^2), x_4 – затраты на рекламу, по сведениям рекламных агентств. Попутно добавили данные ещё одного магазина. Расчёт ведём в *Statistica*:

	1 Var1	2 Var2	3 Var3	4 Var4	5 Var5
1	1	8	200,1	20	8,01
2	3	10	298	30	9,8
3	1	5	300	30	5
4	2	8	400	40	8
5	3	9	500	50	9
6	2	9	450	45	9

При попытке оценить функцию регрессии со всеми четырьмя регрессорами *Statistica* выдаёт сообщение, что матрица плохо обусловлена. Ясно, что имеет место мультиколлинеарность, переменные x_2 и x_3, x_1 и x_4 почти пропорциональны.

Если идти «напролом» (параметр *Tolerance* уменьшить до 0), расчёт производится с результатом

$$\hat{y} = -1.28 + 6.59x_1 + 0.047x_2 - 0.42x_3 - 6.43x_4, \quad R^2 = 0.928, SL = 0.39, \hat{\sigma} = 0.54.$$
$$SL \quad 0.53 \quad 0.99 \quad 1.0 \quad 1.0 \quad 0.99$$

Значимость всего уравнения выше значимости отдельных коэффициентов регрессии, странные знаки у коэффициентов при двух последних регрессорах. Если оценить корреляционную матрицу регрессоров:

Variable	**Var2**	Var3	Var4	Var5
Var2	1,000000	0,299602	0,303934	0,999117
Var3	0,299602	1,000000	0,999974	0,318071
Var4	0,303934	0,999974	1,000000	0,322164
Var5	0,999117	0,318071	0,322164	1,000000

то видно, что $\hat{\rho}(X_2, X_3)$ и $\hat{\rho}(X_1, X_4)$ очень велики. Есть все признаки отягощённости мультиколлинеарностью.

Расчёт в *MathCAD* по (9), если отменить строгую проверку не вырожденности матриц, даёт оценки коэффициентов регрессии: – 1.284 4.183 – 0.204 2.027 – 3.896. Сравнивая с расчётом в *Statistica*, убеждаемся, что оба результата зависят от непредсказуемых ошибок округления.

Перейдём на ридж-регрессию с $\lambda = 0.1$ (*Tolerance* вернём к значению по умолчанию 0.0001). Результат:

$\hat{y} = -1.78 + 0.20x_1 + 0.0015x_2 + 0.017x_3 + 0.124x_4$, $R^2 = 0.72, SL = 0.72, \hat{\sigma} = 1.07$.

SL 0.59 0.8 0.91 0.89 0.88

Ситуация со значимостями выровнялась, знаки сменились на ожидаемые.

Вместо ридж-регрессии применим отбор наиболее существенных объясняющих переменных – пошаговая вперёд (*stepwise forward*) регрессия в *Statistica*. Суть её в следующем.

На первом шаге в модель вводится одна переменная, дающая максимальное значение *F* (фактически R^2) среди всех предикторов (а не первая по порядку).

Далее алгоритм: если уже включено *K* переменных, то при добавлении ещё одной вычисляется значение статистики $F_{to\ enter} = \dfrac{ss_{res}(K) - ss_{res}(K+1)}{ss_{res}(K+1)/(N-K-2)}$, которая есть статистика критерия в проверке гипотезы $H_0 : \beta_{K+1} = 0$, против альтернативы $H_1 : \beta_{K+1} \neq 0$. Оно сравнивается с граничным F_{enter} (фактически с

необходимым уровнем значимости). Если $F_{to\ enter}$ больше F_{enter}, то дополнительная переменная вводится в модель.

В нашем примере на первом шаге введена x_1, $ss_{res}(1) = 1.573$. Если включить ещё x_3, то легко увидеть, что $ss_{res}(2) = 0.985$ (см. *ANOVA*). $F_{to\ enter} = 1.791$, поэтому если установить $F_{enter} < 1.791$, то x_3 будет включена в модель, и наоборот. По умолчанию в *Statistica* $F_{enter} = 1$, т.е. добавляются все новые (лучшие из пока не вошедших), если $F_{to\ enter} > 1$. x_2 и x_4 проиграли x_3 по включению после x_1. Итоговый результат пошаговой регрессии (2 шага, т. к. далее опять возникает мультиколлинеарность):

$$\begin{array}{l}\hat{y} = -1.94 + 0.34x_1 + 0.032x_3 \\ SL \quad 0.24 \quad 0.12 \qquad 0.27\end{array},\ R^2 = 0.75, SL = 0.122, \hat{\sigma} = 0.57.$$

Упражнение 3.1

Применить рекуррентный МНК к данным примера 3.2. (*MathCAD*).

Попробовать начать с $P(0) = I$, $\hat{\beta}_{МНК}(0) = 0$.

4. Введение фиктивных переменных в линейную модель регрессии

Пример 4.1

Исследуется зависимость y (руб.) – потребление пива человеком за неделю от x (руб.) – доход его за тот же период. Но ясно, что если оценить модель $Y^n = \beta_0 + \beta_1 x^n + E^n$, то её прогноз для мужчины и женщины будет одинаков, что, очевидно, неточно. Поэтому в модель следует ввести зависимость от пола путём добавления слагаемого с фиктивной переменной $z^n = \begin{cases} 1, n\text{-й респондент – мужчина} \\ 0, n\text{-й респондент – женщина} \end{cases}$, то есть $Y^n = \beta_0 + \beta_1 x^n + \beta_2 z^n + E^n$. Матрица плана, например, будет иметь вид:

$$z = \begin{pmatrix} 1 & x^0 & z^0 \\ 1 & x^1 & z^1 \\ - & - & - \\ 1 & x^N & z^{N-1} \end{pmatrix} = \begin{pmatrix} 1 & 500 & 1 \\ 1 & 300 & 0 \\ - & - & - \\ - & - & - \end{pmatrix}.$$

Далее находим по МНК оценки коэффициентов и получаем оценённую модель $\hat{y} = \widehat{\beta_0} + \widehat{\beta_1} x + \widehat{\beta_2} z$. Или сразу модель для мужчин $(z = 1)$ $\hat{y} = (\widehat{\beta_0} + \widehat{\beta_2}) + \widehat{\beta_1} x$, а для женщин $(z = 0)$ $\hat{y} = \widehat{\beta_0} + \widehat{\beta_1} x$.

Можно показать, что если оценивать модели для мужчин и женщин отдельно, то при том же объёме выборочной информации, оценки коэффициентов будут хуже (их значимость меньше).

Анализируя значимость коэффициента β_2 (при фиктивной переменной), можно установить наличие либо отсутствие влияния качественной переменной (пол) на результирующий количественный признак.

Если качественная переменная имеет l градации (уровней, категорий), то для отражения её влияния на результирующую переменную нужно ввести l – 1 фиктивных дихотомических переменных $z_1, z_2, ..., z_{l-1}$. Например, мужчина, женщина, юноша, девушка:

	z_1	z_2	z_3	
$l=1$	0	0	0	*М*
$l=2$	1	0	0	*Ж*
$l=3$	0	1	0	*Ю*
$l=4$	0	0	1	*Д*

$$\hat{y} = \widehat{\beta_0} + \widehat{\beta_1} x + \widehat{\beta_2} z_1 + \widehat{\beta_3} z_2 + \widehat{\beta_4} z_3 .$$

Пример 4.2 (продолжение примера 2.1)

Исследуем влияние качественной переменной *расположение* (уровни: центр, средний пояс, окраина города) на объем продаж, данные представлены ниже.

1	8	0	1	средний пояс
3	10	1	0	окраина
1	5	1	0	окраина
2	8	0	1	средний пояс
3	9	0	0	центр

По МНК получаем оценённую модель

$\hat{y} = -0.6 + 0.4x - 0.4z_1 - 1.1z_2$. $R^2 = 0.875, SL = 0.441, \hat{\sigma} = 0.707$.

Или для центра ($z_1 = 0$, $z_2 = 0$) $\hat{y} = -0.6 + 0.4x$, среднего пояса (0, 1) $\hat{y} = -1.7 + 0.4x$, окраины (1, 0) $\hat{y} = -1 + 0.4x$. И прогнозы объёмов продаж для магазинов, тратящих на рекламу по 7 тыс. руб. будут в зависимости от расположения 2.2, 1.1, 1.8 сотен тыс. руб.

При попытке оценить уравнение для окраины только по соответствующим данным анализ вообще не удаётся провести (тем более для центра).

Упражнение 4.1

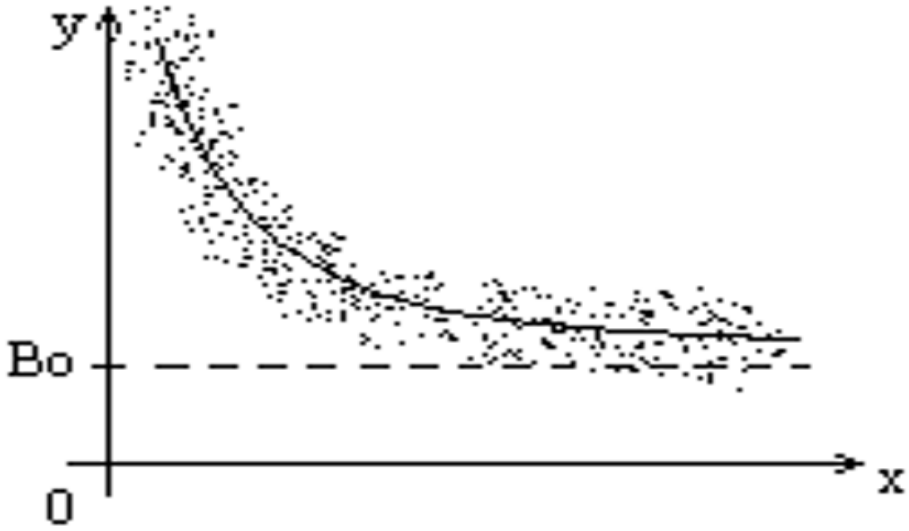
По данным примера 4.2 оценить в *MathCAD* и *Statistica* уравнение для окраины только по соответствующим данным.

5. Нелинейные модели регрессии и линеаризация

До сих пор мы считали, что неизвестная функция регрессии хорошо описывалась линейным законом $M\left(Y \mid \overrightarrow{X} = \vec{x}\right) = \vec{x}^T \vec{\beta} = \beta_0 + \beta_1 x_1 + \ldots + \beta_K x_K$. Но это не всегда так.

Пример 5.1

Y – спрос на необходимый товар (хлеб), *x* – его цена. Собрав наблюдения, строим корреляционное поле и предполагаем зависимость гиперболического типа.



$Y = \beta_0 + \beta_1 \frac{1}{x} + E$, т.е. $\breve{y} = M(Y \mid x) = \beta_0 + \beta_1 \frac{1}{x}$. Переходя к новой объясняющей переменной $\frac{1}{x} = \tilde{x}$, эта зависимость сводится к линейному виду $\breve{y} = \beta_0 + \beta_1 \tilde{x}$. Применяем МНК, $\hat{\vec{\beta}}_{МНК} = \left(\tilde{z}^T \tilde{z}\right)^{-1} \tilde{z}^T \vec{y}$, где матрица плана имеет вид

$$\tilde{z} = \begin{pmatrix} 1 & \tilde{x}^0 \\ 1 & \tilde{x}^1 \\ - & - \\ 1 & \tilde{x}^{N-1} \end{pmatrix} = \begin{pmatrix} 1 & \frac{1}{x^0} \\ 1 & \frac{1}{x^1} \\ - & - \\ 1 & \frac{1}{x^{N-1}} \end{pmatrix},$$

здесь $x^0, x^1, ..., x^{N-1}$ – наблюдённые данные для цены товара.

Пример 5.2

Производственная функция Кобба–Дугласа описывает зависимость объёма произведённой предприятием продукции *Y* от основных факторов производства $x_1, x_2, ..., x_K$ (x_1 – труд, x_2 – капитал, x_3 – сырье, и т.д.).

$$Y = e^{\beta_0} x_1^{\beta_1} x_2^{\beta_2} \dots x_K^{\beta_K} e^E .$$

Если взять логарифм от обеих частей: $\ln Y = \beta_0 + \beta_1 \ln x_1 + \dots + \beta_K \ln x_K + E$, или

$$\tilde{Y} = \beta_0 + \beta_1 \tilde{x}_1 + \dots + \beta_K \tilde{x}_K + E - \text{линейная функция.} \qquad (24)$$

Оценка параметров $\tilde{\vec{\beta}}_{МНК} = \left(\tilde{z}^T \tilde{z}\right)^{-1} \tilde{z}^T \tilde{\vec{y}}$, где

$$\tilde{z} = \begin{pmatrix} 1 & \ln x_1^0 & \cdots & \ln x_K^0 \\ 1 & \ln x_1^1 & \cdots & \ln x_K^1 \\ - & - & - & - \\ 1 & \ln x_1^{N-1} & \cdots & \ln x_K^{N-1} \end{pmatrix}, \quad \tilde{\vec{y}} = \begin{pmatrix} \ln y^0 \\ \ln y^1 \\ \cdots \\ \ln y^{N-1} \end{pmatrix}.$$

Подбор преобразований исходных переменных $Y, x_1, \dots, x_K$: $\tilde{Y} = \theta(Y)$, $\tilde{x}_k = \varphi_k(x_k)$, которые делают нелинейную зависимость *Y* от $\vec{x}$ линейной (24), называется *линеаризацией*.

Пример 5.3 (продолжение примера 2.1)

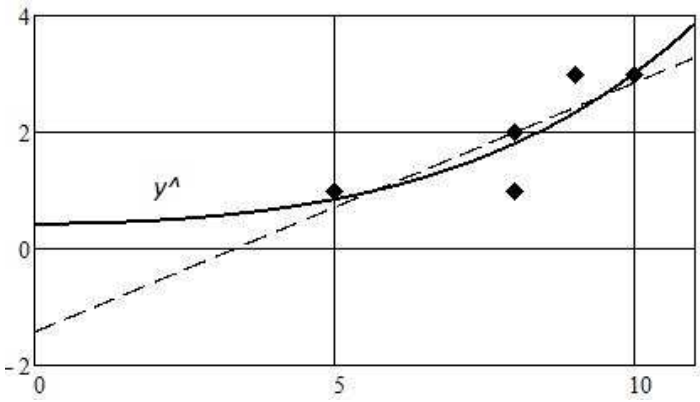
Попробуем подобрать преобразования $\tilde{y} = \theta(y)$ $\tilde{x} = \varphi(x)$ так, чтобы точки $\left(\tilde{x}^n, \tilde{y}^n\right)$ лучше ложились на прямую чем исходные $\left(x^n, y^n\right)$. Это означает, что должно быть $\hat{\rho}\left(\tilde{x}, \tilde{y}\right) > \hat{\rho}(x, y)$. В *Statistica* последовательно выбираем: *Statistics*, *Advanced Linear/Nonlinear Models*, *Fixed Nonlinear Regression* набираем разнообразные преобразования: *X***2, *SQRT*(*X*), *e****X*, *OK*. *Descriptive …*, *Review*, *Correlations*. Видим $\hat{\rho}\left(\sqrt{y}, x^2\right) = 0.826 > \hat{\rho}(y, x) = 0.802$, поэтому оцениваем регрессию по модели $\sqrt{Y} = \beta_0 + \beta_1 x^2 + E$ или $\tilde{Y} = \beta_0 + \beta_1 \tilde{x} + E$. *Quick*, *Var-s…*, получим $\widehat{\beta_0} = 0.645$, $\widehat{\beta_1} = 0.0109$, т.е.

$$\hat{\tilde{y}} = 0.645 + 0.0109\tilde{x}. \qquad (25)$$

Коэффициент детерминации надо считать в линеаризованной модели:

$$R^2 = \frac{\sum_n \left(\hat{\tilde{y}}^n - \tilde{\tilde{y}}\right)^2}{\sum_n \left(\tilde{y}^n - \tilde{\tilde{y}}\right)^2} = (0.826)^2 = 0.682,\ df = 1,\ 3\ F = 6.44,\ sl = 0.085.$$

Эти показатели лучше, чем при регрессии Y на x. Из (25) получаем $\hat{y} = (0.645 + 0.0109x^2)^2$. График этой функции:



Коэффициенты эластичности в нелинейных моделях

Пусть $\breve{y} = f\left(\vec{x}\right)$.

$d\breve{y}_{x_k} \cdot \frac{100\%}{\breve{y}} = \frac{\partial f\left(\vec{x}\right)}{\partial x_k} \cdot \frac{x_k}{\breve{y}} \{dx_k \cdot \frac{100\%}{x_k}\}$. Выражение в {} показывает проценты изменения x_k . Если это выбрать = 1, то видно, что величина

$$Э_{yx_k} = \frac{\partial f\left(\vec{x}\right)}{\partial x_k} \cdot \frac{x_k}{\breve{y}}$$

– коэффициент эластичности y по x_k показывает, на сколько (примерно) процентов изменится результирующая величина y при увеличении x_k на 1 %.

Если заменить $\vec{\beta} \to \hat{\vec{\beta}}$, $\overset{\vee}{y} \to \hat{y}$, то получится формула оценённого коэффициента эластичности $\hat{\Theta}_{yx_k}$, если дополнительно $\vec{x}$ заменим на $\overline{\vec{x}}$, $\hat{y}$ на $\overline{y}$, получим:

$$\overline{\Theta}_{yx_k} = \frac{\partial \hat{f}\left(\overline{\vec{x}}\right)}{\partial x_k} \cdot \frac{\overline{x}_k}{\overline{y}}$$ – средний коэффициент эластичности.

5.1. Эффект масштаба и кривая обучения (опыта)

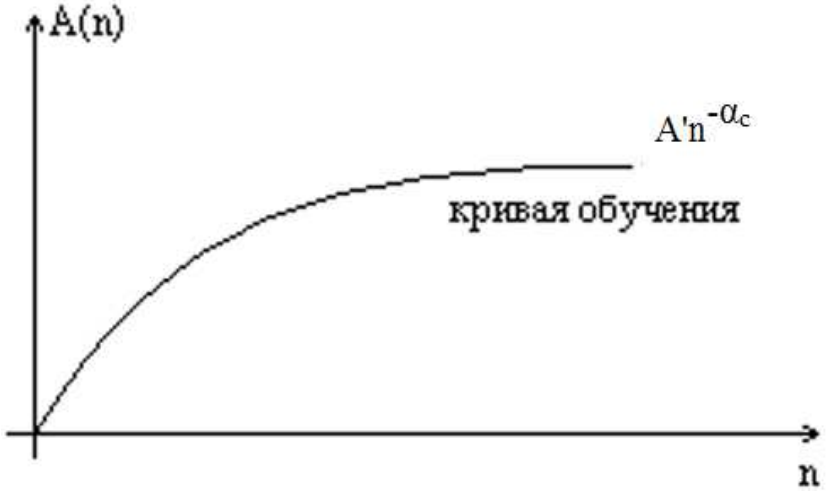
Пусть значение всех факторов производства на предприятии выросло в μ раз, а состояние технических знаний не изменилось. Если объем производства при этом вырастет в $\nu = \mu^r$, то r называется отдачей от масштаба, $r - 1$ – эффект от масштаба.

Считаем применимой производственную функцию Кобба–Дугласа

$$y = A x_1^{\alpha_1} x_2^{\alpha_2} \ldots x_K^{\alpha_K},$$

x_k – факторы производства, A – характеризует состояние технических знаний. Тогда $\nu = \mu^{\alpha_1+\alpha_2+\ldots+\alpha_K}$, т.е. $r = \alpha_1 + \alpha_2 + \ldots + \alpha_K$.

Состояние технических знаний зависит от n – накопленного выпуска продукции (опыт, обучение). $A = A(n) = A' n^{-\alpha_c}$ – кривая обучения. $\alpha_c < 0$, $-\alpha_c$ – эластичность выпуска по накопленному выпуску (опыту).



Если предприятие работает в режиме минимизации суммарных издержек

$$C = \sum_{k=1}^{K} p_k x_k ,$$

p_k – цены на факторы производства при данном объёме выпуска y, то можно показать, что оптимальные дефлятированные удельные издержки c удовлетворяют уравнению:

$$c = k(n)^{\frac{\alpha_c}{r}} (y)^{\frac{1}{r}-1} .$$

Поэтому, с целью оценки входящих сюда параметров изучается модель вида

$$c^t = k(n^t)^{\frac{\alpha_c}{r}} (y^t)^{\frac{1}{r}-1} e^{e^t}, \quad t = 0,...,T-1, \tag{26}$$

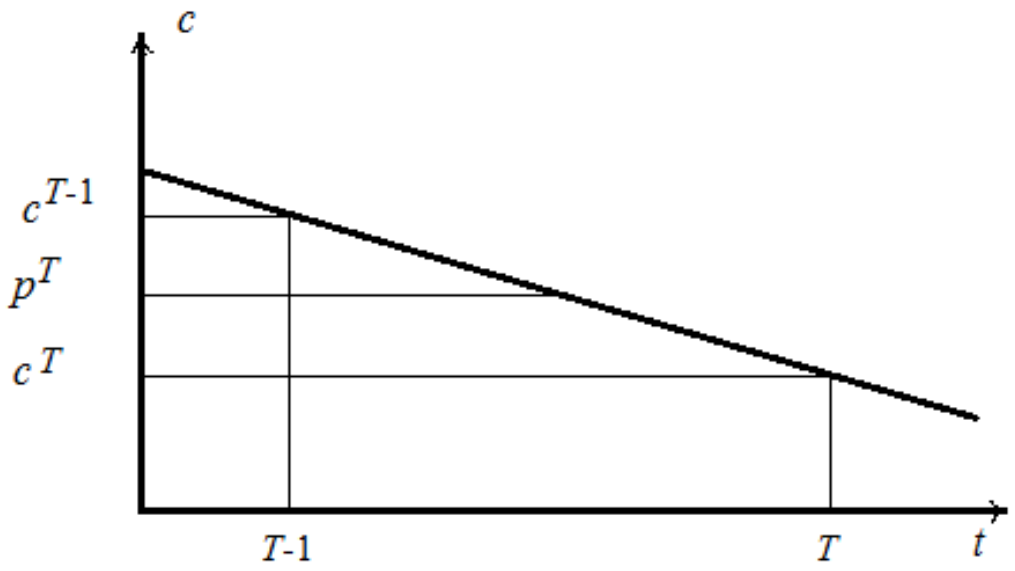
где c^t – дефлятированные удельные издержки в периоде времени t;

y^t – выпуск в периоде времени t;

n^t – накопленный выпуск до периода времени t;

e^t – случайная ошибка.

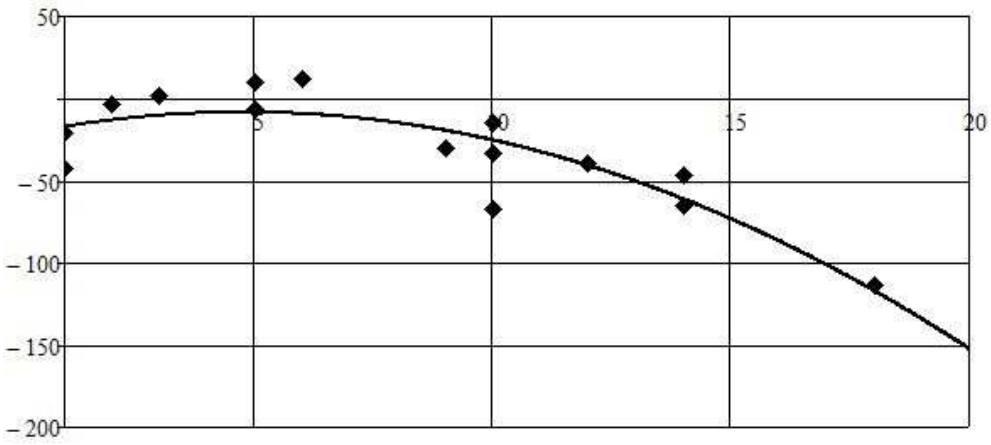
Цель – прогнозируя снижение в наступающем периоде c^T (спрогнозировав предварительно n^T и y^T), установить уже сейчас цену unit price for your production p^T ниже издержек сейчас c^{T-1} (но выше, чем в конце периода, чтобы в целом за период была прибыль). Это поможет завоевать рынок, увеличить объём выпуска и пройти скорейшее дальнейшее обучение.



Для линеаризации достаточно взять логарифм от обеих частей (26).

5.2. Параболическая регрессия

Если диаграмма рассеивания имеет вид, подобный приведённому на рисунке, то функцию регрессии следует считать полиномом (параболической функцией) некоторой степени K (на рисунке $K = 2$) от переменной x.



$\breve{y} = \beta_0 + \beta_1 x + \beta_2 (x)^2 + \ldots + \beta_K (x)^K$. Заменяя $(x)^k = \tilde{x}_k$, получаем КЛММР:

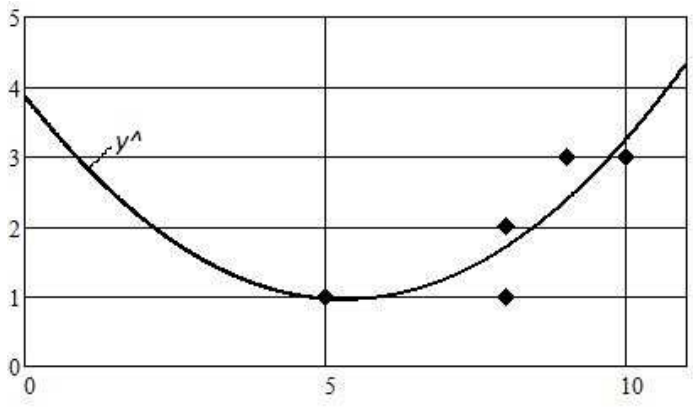
$\breve{y} = \beta_0 + \beta_1 \tilde{x}_1 + \beta_2 \tilde{x}_2 + \ldots + \beta_K \tilde{x}_K$. Оценки коэффициентов $\hat{\tilde{\beta}}_{МНК} = \left(\tilde{z}^T \tilde{z}\right)^{-1} \tilde{z}^T \vec{y}$, где

$$\tilde{z} = \begin{pmatrix} 1 & \tilde{x}_1^0 & \tilde{x}_2^0 & \cdots & \tilde{x}_K^0 \\ 1 & \tilde{x}_1^1 & \tilde{x}_2^1 & \cdots & \tilde{x}_K^1 \\ - & - & - & - & - \\ 1 & \tilde{x}_1^{N-1} & \tilde{x}_2^{N-1} & \cdots & \tilde{x}_K^{N-1} \end{pmatrix} = \begin{pmatrix} 1 & x^0 & (x^0)^2 & \cdots & (x^0)^K \\ - & - & - & - & - \\ 1 & x^{N-1} & (x^{N-1})^2 & \cdots & (x^{N-1})^K \end{pmatrix}.$$

Пример 5.4 (продолжение примера 2.1)

Оценим квадратичную функцию регрессии.

$$\tilde{z} = \begin{pmatrix} 1 & 8 & 8^2 \\ 1 & 10 & 10^2 \\ 1 & 5 & 5^2 \\ 1 & 8 & 8^2 \\ 1 & 9 & 9^2 \end{pmatrix} \qquad \hat{\tilde{\beta}} = \begin{pmatrix} 3.88 \\ -1.1 \\ 0.104 \end{pmatrix} \qquad \hat{y} = 3.88 - 1.1x + 0.104x^2.$$



Нелинеаризуемые модели

Если не удаётся получить линейную по параметрам $\vec{\beta}$ модель, то работают с исходной нелинейной зависимостью $Y = f(\vec{x}, \vec{\beta}) + E$, где f – известная функция, методом (нелинейного) МНК:

$$\hat{\vec{\beta}}_{МНК} = \arg\min_{\hat{\vec{\beta}}} \sum_{n=0}^{N-1} \left(y^n - f(\vec{x}^n, \hat{\vec{\beta}}) \right)^2, \qquad (27)$$

но теперь целевая функция не квадратична по $\vec{\beta}$ и минимизация усложняется. Чаще всего в правую часть (27) подставляют все данные и численно минимизируют по $\hat{\vec{\beta}}$.

5.3. Упражнения

Упражнение 5.1

Обсудить смысл ошибок и константы β_0 в примере 5.1.

Упражнение 5.2

В условиях примера 5.3 найти оценённый коэффициент эластичности объёма продаж по затратам на рекламу при x = 5 и 10 тыс. руб. и средний коэффициент эластичности.

Упражнение 5.3

По 20 регионам страны изучается зависимость уровня безработицы y (%) от индекса роста потребительских цен x (в % к предыдущему году). Информация о логарифмах исходных показателей представлена в таблице.

Показатель	*ln x*	*ln y*
Выборочное среднее	0.6	1.0
s	0.4	0.2

Известна также оценка коэффициента корреляции между логарифмами исходных показателей: 0.8.

1. Оценить функцию регрессии y на x в степенной форме.
2. Найти оценённый коэффициент эластичности при x = 20 %. Если индекс роста цен увеличится на 1 %, каков ожидается уровень безработицы?
3. Найти коэффициент детерминации.
4. Проверить гипотезу о равенстве всех коэффициентов (кроме константы) нулю.

Упражнение 5.4

Повторить в *Statistica* в двух процедурах: *Fixed Nonlinear Regression* и *Multiple Regression* результаты примера 5.4. Найти коэффициент детерминации и уровень значимости данных против гипотезы о равенстве всех коэффициентов регрессии (кроме константы) нулю.

6. Категоризованные зависимые переменные

Иногда зависимая переменная по своему смыслу может принимать лишь дискретный набор значений или категорий. Биномиальный случай – когда две категории – и общее введение см., например, [13, 11, 3].

6.1. Неупорядоченная мультиномиальная логит-модель

Аналогично биномиальной логит-модели,

$$P\{Y^n = j\} = \frac{e^{\vec{x}^{nT}\vec{\beta}^j}}{1 + \sum_{k=1}^{J} e^{\vec{x}^{nT}\vec{\beta}^k}} \qquad (28)$$

– вероятности попасть результирующей величине в *n*-м наблюдении в ту или иную категорию *j*. Всего имеется *J* + 1 категория: 1-я, … *J*-я, *J* + 1-я (референсная). Категории нумеруются, например, в порядке появления их в данных (сверху вниз). Вероятности зависят от факторов $\vec{x}^n$ через коэффициенты $\vec{\beta}^j$. В каждой категории свои коэффициенты, чтобы разные наборы факторов давали максимум вероятности «своей», соответствующей категории. Для последней $\vec{\beta}^{J+1} = \vec{0}$. Для неотрицательности – экспонента, сумма должна быть равна 1. Отсюда имеем вид (28).

Пример 6.1

Исследуется возможность определения типа корпуса телефона по его цене. Типы: классический, «раскладушка», слайдер. Данные имеют вид:

1	3720	classic
2	3900	classic
3	3900	classic
4	3960	rasklad
5	3960	rasklad
6	4200	rasklad
7	4260	classic
8	4260	rasklad
9	4350	classic
10	4350	slider
11	4410	classic
12	4410	rasklad
13	4500	classic
14	4560	classic
15	4650	classic

16	4800	classic
17	4890	classic
18	4950	slider
19	4950	classic
20	4950	rasklad
21	4950	rasklad
22	5040	rasklad
23	5100	classic
24	5190	slider
25	5340	classic
26	5340	classic
27	5340	slider
28	5400	classic
29	5400	rasklad
30	5550	rasklad

31	5640	classic
32	5790	classic
33	5940	classic
34	6090	rasklad
35	6090	rasklad
36	6480	rasklad
37	6780	rasklad
38	6930	rasklad
39	6930	rasklad
40	7080	rasklad
41	7080	rasklad
42	7170	rasklad
43	7230	rasklad
44	7320	rasklad
45	7380	rasklad

46	7680	slider
47	8010	rasklad
48	8160	classic
49	8370	rasklad
50	8520	rasklad
51	9750	slider
52	9840	slider
53	9990	rasklad
54	10680	rasklad
55	10740	rasklad
56	10980	rasklad
57	11040	slider
58	11130	rasklad
59	11670	rasklad

Statistics => *Analyzing Data* => *Generalized Linear/Nonlinear Models* => *Quick* => *Multinomial logit model* => *OK* => *Variables* => *Dep* (указать переменную с категориями), *Cont. predictors* (указать переменную(ые) с количественными факторами) => *OK* => *OK*. *Summary – Estimates* (оценки векторов бета).

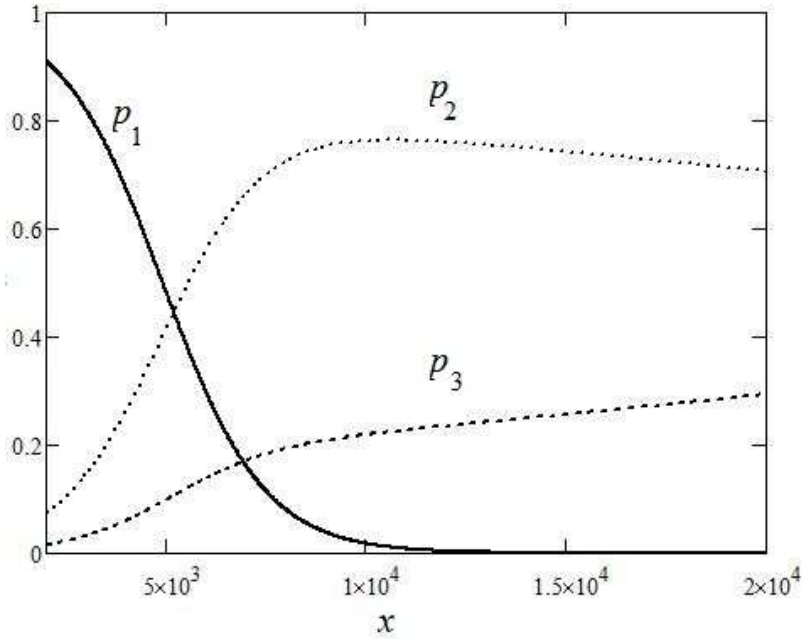
Категории: 1 – *classic*, 2 – *rasklad*, 3 – *slider*.

$\hat{\beta}_0^1 = 5.697$, $\hat{\beta}_1^1 = -0.000822$, $\hat{\beta}_0^2 = 1.617$, $\hat{\beta}_1^2 = -0.000037$, $\hat{\beta}_0^3 = 0$, $\hat{\beta}_1^3 = 0$.

Зная это, можно оценить вероятности того, что первый телефон (с ценой 3720) должен иметь классический тип корпуса: $\frac{e^{5.697-0.000822\times 3720}}{1+18.35}=0.721,$ аналогично с типом «раскладушка» = 0.227 и типом слайдер = 0.052. Это можно увидеть: *Resid*.1 – *Predicted values* (вероятности объекту принадлежать каждой категории).

Таким образом, анализ позволяет отвечать на вопросы: каковы вероятности объекту с заданными факторами принадлежать каждой категории? В каких областях значений факторов вероятнее оказаться в заданной категории?

В этом примере вероятней принадлежать типу «раскладушка» для цен от 5197 до 43702 (см. рисунок).



6.2. Упорядоченная мультиномиальная логит-модель

Здесь категории упорядочены (по смыслу): 1-я, … J-я, J + 1-я. Например: лейтенант, майор, полковник.

$$P\{Y^n = 1\} = P\{Y^* \le 1\} = \Lambda(\beta_0^1 + \beta_1 x^n)$$

$$P\{Y^n = 2\} = P\{1 < Y^* \le 2\} = \Lambda(\beta_0^2 + \beta_1 x^n) - \Lambda(\beta_0^1 + \beta_1 x^n)$$

……..

$$P\{Y^n = J+1\} = 1 - P\{Y^n = J\} - \ldots - P\{Y^n = 1\} = 1 - \Lambda(\beta_0^J + \beta_1 x^n)$$

Y^* – ненаблюдаемая «полезность» для продвижения по категориям, а $\Lambda(\ldots)$ – логистическая функция.

Продолжение примера 6.1

Если считать эти категории упорядоченными. *Statistics* => *Analyzing Data* => *Generalized Linear/Nonlinear Models* => *Quick* => *Ordinal logit model* => *OK* => *Variables* => *Dep* (указать переменную с категориями), *Cont. predictors* (указать переменную(ые) с количественными факторами) => *OK* => *OK*. *Summary* => *Estimates* (оценки β_0^j и вектора коэффициентов при факторах).

$$\hat{\beta}_0^1 = 1.886, , \hat{\beta}_0^2 = 4.876, \hat{\beta}_1 = -0.000414.$$

Зная это, можно оценить вероятности того, что первый телефон (с ценой 3720) должен иметь классический корпус: $\frac{1}{1+e^{-(1.886-0.000414\times 3720)}} = 0.586$, аналогично, с типом «раскладушка» = $\frac{1}{1+e^{-3.336}} - 0.586 = 0.379$. Это можно увидеть: *Resid*.1 – *Predicted values* (вероятности объекту принадлежать каждой категории).

Ясно, что обычная (биномиальная) логит-модель есть частный случай упорядоченной мультиномиальной логит-модели.

6.3. Модель регрессии Пуассона

Применяется для описания результирующей величины, которая есть дискретная случайная величина с распределением Пуассона, то есть имеющая смысл числа успехов, каждый из которых маловероятен, при большом числе независимых испытаний. Примеры таких случайных величин: 1) число клиентов, разместивших депозит за день в банке, 2) число заболевших (на тысячу человек) в регионе. Считаем, что параметр распределения (а он совпадает с математическим ожиданием самой величины) подвержен влиянию нескольких

(K) факторов (объясняющих переменных). Т.е. в каждом n-м наблюдении был свой параметр a^n, определяемый $\vec{x}^n$. Модель:

$$Y^n = \exp(\vec{x}^{nT}\vec{\beta}) + E^n, \qquad P\{Y^n = y^n\} = \frac{(a^n)^{y^n}}{(y^n)!}e^{-a^n}$$

экспонента использована для получения неотрицательных значений, член ошибки можно выбрать и мультипликативно, т.е. в показателе (если нет нулевых y^n). Очевидно,

$$M(Y^n \mid \vec{x}^n) = \exp(\vec{x}^{nT}\vec{\beta}) = a^n. \tag{29}$$

Функция правдоподобия $L = \prod_{n=1}^{N} \frac{(a^n)^{y^n}}{(y^n)!}e^{-a^n}$. Беря логарифм и подставляя a^n, получим оптимизационную задачу:

$$\hat{\vec{\beta}}_{ML} = \arg\max \sum_{n=1}^{N} (y^n(\vec{x}^{nT}\hat{\vec{\beta}}) - \exp(\vec{x}^{nT}\hat{\vec{\beta}}) - \ln(y^n!)).$$

Так как каждый фактор по (29) входит мультипликативно, легко судить во сколько раз изменится отклик за счёт каждого из них.

О проверке значимости полученных оценок см.[13, стр. 13 – 14].

В пакете Statistica оценивание методом максимального правдоподобия реализовано: Statistics → Advanced Linear/Nonlinear Models → Generalized Linear/Nonlinear Models → Poisson log model. Значимость отдельных коэффициентов даёт Type 1 LR test.

При отсутствии 0-х результирующих значений можно логарифмированием линеаризовать модель и применить обычный МНК.

7. Системы линейных одновременных уравнений

До сих пор мы анализировали одно уравнение, отражающее интересующие нас связи. Более сложные задачи приводят к необходимости рассмотрения систем взаимосвязанных уравнений.

Пример 7.1

Равновесная модель спроса-предложения. Пусть $\widetilde{y}_1^{\,t}$ – объем предложения некоторого товара в момент *t*, он же и объём спроса на него (равновесие), $\widetilde{y}_2^{\,t}$ – цена товара, $\tilde{x}^t$ – среднедушевой доход в тот же момент. $(t = 0,1,\ldots,T-1)$.

Перейдём к центрованным переменным: $x^t = \tilde{x}^t - \bar{\tilde{x}}$, $\bar{\tilde{x}} = \frac{1}{T}\sum_{t=0}^{T-1} \tilde{x}^t$. y_1, y_2 – аналогично. Можно предположить, исходя из общеэкономических соображений (структуры задачи):

$$\begin{cases} \textit{Уравнение предложения}: y_1^t = b_1 y_2^t + e_1^t,\ b_1 > 0 \\ \textit{Уравнение спроса}: y_1^t = b_2 y_2^t + c x^t + e_2^t,\ b_2 < 0, c > 0 \end{cases}. \qquad (30)$$

В модели (30) значение цены y_2 и спроса-предложения y_1 формируются «внутри» модели, тогда как среднедушевой доход *x* задаётся «извне» и в итоге определяет y_1, y_2. Поэтому y_1, y_2 называют *эндогенными* переменными, а *x* – *экзогенной* переменной.

Если попытаться оценить коэффициент b_1 *структурной формы* (30), работая только с 1-м уравнением, считая y_2 объясняющей переменной, то эта оценка МНК будет смещённой и несостоятельной. Это есть следствие того, что ошибка e_1 зависит от объясняющей переменной y_2, что видно из второго уравнения (30), где y_2 зависит от y_1, а значит, и от e_1 [13]. Поэтому применим другую процедуру.

7.1. Косвенный метод наименьших квадратов

Предопределённые (объясняющие) переменные – это все экзогенные переменные + эндогенные лаговые переменные (т.е. эндогенные переменные, измеренные в прошлые, по отношению к текущему, объясняемому моменту времени, т.е. уже известные, заданные).

Пример 7.2

$$\begin{cases} I^t = a + bY^{t-1} + cR^t + E_1^t \\ Y^t = d + eC^t + E_2^t \end{cases},$$

Y^t – совокупный выпуск в году t, C^t – совокупное потребление в году t,

I^t – инвестиции в году t, R^t – банковская ставка в году t.

Модель предназначена для объяснения величин I^t, Y^t – эндогенные, C^t, R^t – экзогенные (задаются извне произвольно, управляемо, планируемо), предопределённые: C^t, R^t, Y^{t-1}.

Если все эндогенные переменные явно линейно выражены через предопределённые переменные и случайные ошибки

$$\vec{y}^t = \Pi\vec{x}^t + \vec{e}^{*t},$$

то это называется *приведённой формой* СОУ (системы одновременных уравнений), где П – матрица коэффициентов.

В *косвенном МНК*:

- Составляют приведённую форму модели и оценивают параметры её каждого уравнения обычным МНК. Так как каждое уравнение ПСОУ удовлетворяет требованиям КЛММР, оценки эти будут несмещёнными и состоятельными.
- Путём алгебраических преобразований выражают назад коэффициенты структурной формы через коэффициенты приведённой формы, получая тем самым оценки коэффициентов структурной формы.

Продолжение примера 7.1

Вычтем из первого уравнения (30) второе:

$0 = y_2^t\left(b_1 - b_2\right) - cx^t + \left(e_1^t - e_2^t\right)$, отсюда

$$\begin{cases} y_2^t = \dfrac{c}{b_1 - b_2}x^t + \dfrac{e_2^t - e_1^t}{b_1 - b_2} \\ y_1^t = \dfrac{b_1 c}{b_1 - b_2}x^t + \dfrac{b_1 e_2^t - b_2 e_1^t}{b_1 - b_2} \end{cases}, \tag{31}$$

что является приведённой формой СОУ. Оцениваем коэффициенты: $\widehat{\pi}_1, \widehat{\pi}_2$ по МНК. Разделив π_1 на π_2, найдём $\widehat{b_1} = \dfrac{\widehat{\pi}_1}{\widehat{\pi}_2}$. Но уравнений (31) недостаточно, чтобы выразить b_2 и c через π_1 и π_2.

Иногда таким образом удаётся найти все коэффициенты структурной формы, а иногда не все. Если все коэффициенты структурного уравнения однозначно восстанавливаются по коэффициентам приведённой формы, то уравнение называется *точно идентифицируемым*. Если коэффициенты восстанавливаются неоднозначно (могут быть разными при одной приведённой форме), то уравнение называется *сверхидентифицируемым*. Если коэффициенты нельзя восстановить, то уравнение *неидентифицируемо*.

7.2. Двухшаговый метод наименьших квадратов

Этот метод используется для идентификации сверхидентифицируемого уравнения. Назовём эндогенную переменную в левой части структурного уравнения зависимой, все переменные в правой части – объясняющие. Они делятся на предопределённые и эндогенные. Назовём инструментальными переменными переменные, которые будем использовать для «замены» эндогенных в правой части. «Замена» означает, что мы оценим их регрессии на инструментальные (1-й шаг), а затем при оценке регрессии зависимой переменной на объясняющие, используем их значения, найденные по оценённым регрессиям (2-й шаг). В инструментальные переменные лучше

включать все предопределённые переменные из всех уравнений. Это будет соответствовать оцениванию по приведённым уравнениям в косвенном МНК.

Условие размерности для идентификации

Ясно, что для успешного объяснения («замены») эндогенных переменных в уравнении число прочих экзогенных должно быть >= их числа. Пусть есть всего m эндогенных переменных. Если число эндогенных переменных в правой части уравнения максимально, то оно = $m - 1$, тогда прочих экзогенных должно быть >= $m - 1$. Если в уравнение не входит j эндогенных переменных от их максимального набора, то число не включённых в уравнение экзогенных должно быть >= $m - 1 - j$, то есть общее число не включённых всяких переменных опять должно быть >= $j + (m - 1 - j) = m - 1$. Таким образом в идентифицируемом уравнении должно отсутствовать $m - 1$ или более переменных. Если больше, то оно будет сверхидентифицируемым, если равно, то, скорее всего, точно идентифицируемым. Но это необходимые условия.

Продолжение примера 7.1

$m - 1 = 2 - 1 = 1$, а во второе уравнение в (30) не входит 0 переменных, значит, оно будет неидентифицируемым.

Пример 7.3

По данным, приведённым в таблице, оценить параметры структурной модели вида:

$$\begin{cases} Y_1 = b_{12}Y_2 + c_{11}X_1 + c_{12}X_2 + E_1 \\ Y_2 = b_{21}Y_1 + c_{22}X_2 + c_{23}X_3 + E_2 \\ Y_3 = b_{31}Y_1 + c_{33}X_3 + E_3 \end{cases}.$$

Период времени	Темпы прироста, %					Безработных x_1, %	$\hat{y}_1$
	Зарплаты y_1	Цен y_2	Дохода y_3	Цен на импорт x_2	Населения x_3		
1	2	6	10	2	1	1	2.22
2	3	7	12	3	2	2	4.07
3	4	8	11	1	5	3	4.53
4	5	5	15	4	3	2	4.36
5	6	4	14	2	3	3	5.08
6	7	9	16	2	4	4	6.51
7	8	10	18	3	4	5	8.4

В первом уравнении отсутствуют Y_3, X_3, т.е. две переменные, $m - 1 = 3 - 1 = 2$, поэтому уравнение идентифицируемо. Аналогично идентифицируемо и второе. В третьем отсутствуют три переменные, поэтому оно сверхидентифицируемо.

Во втором уравнении Y_1 коррелирует с E_2 (т.к. по первому Y_1 зависит от Y_2, а значит, по второму от E_2). Это обычная ситуация с СОУ, поэтому применим двухшаговый метод наименьших квадратов. Сначала оцениваем регрессию Y_1 на X_1, X_2, X_3. Получаем $\hat{Y}_1 = 1.51X_1 + 0.395X_2 - 0.0795X_3$. Находим $\hat{y}_1^n$ (см. таблицу). Оцениваем регрессию Y_3 на $\hat{Y}_1$ и X_3, аналогично остальное:

$$\begin{cases} \hat{Y}_3 = 2.42\hat{Y}_1 + 0.235X_3 \\ \hat{Y}_2 = 0.721\hat{Y}_1 + 0.413X_2 + 0.684X_3 \\ \hat{Y}_1 = -0.127\hat{Y}_2 + 1.65X_1 + 0.484X_2 \end{cases}. \qquad (32)$$

Шляпки в правых частях можно убрать.

В пакете *SPSS* есть процедура 2*SLS*. Она сразу позволяет найти, например второе уравнение в (32), если указать: зависимая Y_2, объясняющая Y_1, X_2, X_3, инструментальные X_1, X_2, X_3. Ясно, что если список инструментальных равен списку объясняющих, то имеем обычный МНК (нет 1-й стадии, т.к. нет эндогенных переменных). Ясно, что если список объясняющих есть подмножество списка инструментальных, то лишние инструментальные игнорируются.

Если применить простой МНК для оценки второго уравнения, то получим $\hat{Y}_2 = 0.117Y_1 + 0.824X_2 + 1.335X_3$ – неверные, смещённые оценки.

Упражнение 7.1

Модель денежного рынка:

$$R^t = a_1 + b_{11}M^t + b_{12}Y^t + e_1{}^t$$

$$Y^t = a_2 + b_{21}R^t + b_{22}I^t + e_2{}^t ,$$

где R – процентная ставка, Y – ВВП, M – денежная масса, I – внутренние инвестиции.

Записать приведённую форму модели. Выразить коэффициенты структурной формы через коэффициенты приведённой формы.

8. Временные ряды

Временной ряд – совокупность измерений некоторой величины (*Y*), производимых по мере возрастания времени. Обычно измерения производятся через равные промежутки времени и нумеруются

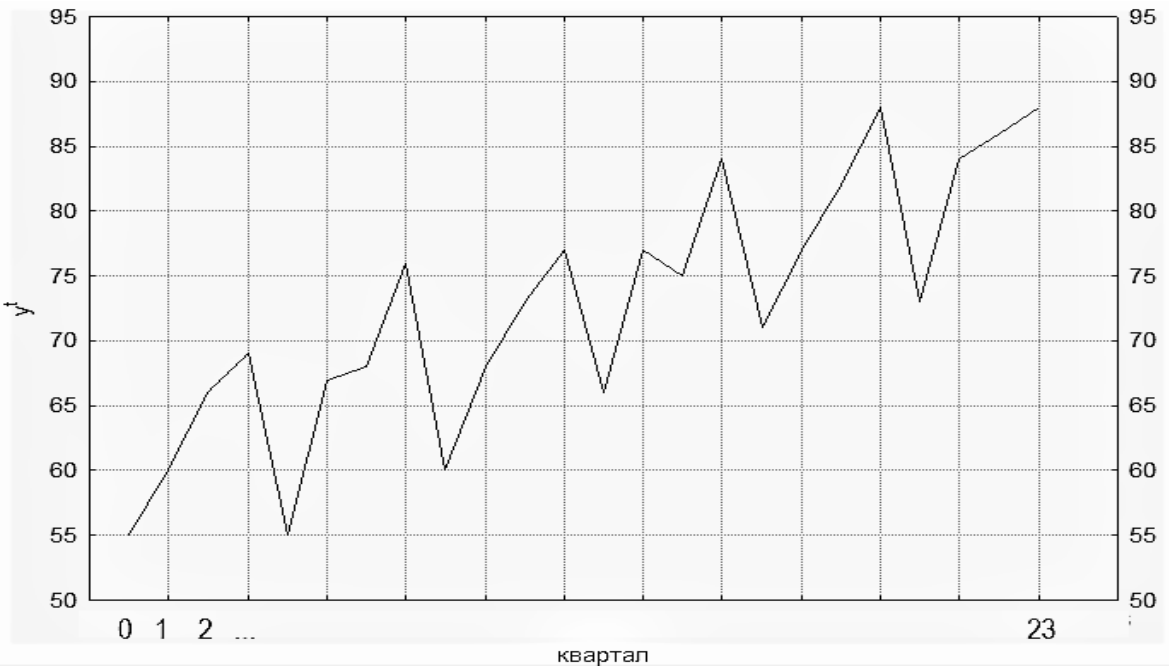
$$y^0, y^1, ..., y^t, ..., y^{T-1} \ . \qquad (1.1)$$

Пример 8.1

Пусть есть ежеквартальные данные об объёме экспорта из РФ в сотнях млрд. долларов за 6 лет (24 квартала) y^t.

t	y^t	y^t_{ma}	dif^t	$\hat{s}^t$ (*сезонные факторы*)	y^t_{adj}
(весна) 0	55	–	–	– 8.42	63.42
(лето) 1	60	–	–	0.18	59.82
(зима) 2	66	62.5	3.5	1.67	64.33
(осень) 3	69	62.5	6.5	6.58	…
(в) 4	55	64.25	– 9.25	– 8.42	…
(л) 5	67	64.75	2.25	0.18	…
(з) 6	68	66.5	1.5	1.67	…
(о) 7	76	67.75	8.25	…	…
8	…	…	-8	…	…
…	…	…	…	…	…
20	73	81.75	…	…	…
21	84	82.75	…	…	…
22	86	82.75	…	…	…
23	88	–	–	…	…

На графике объём экспорта из РФ выглядит так:



Как спрогнозировать дальнейшие значения (в будущем) в 24-м, 25-м и др. кварталах?

Пример важности прогноза: если бы знать потребность в электроэнергии через 10 лет, то сейчас можно было бы начать строительство нужных электростанций (т.к. оно длится примерно столько).

Видно, что в поведении экспорта есть линейный рост, но есть и сезонные колебания. Поэтому примем модель:

$$Y^t = tc^t + s^t + E^t,$$

где *tc* – тренд-цикл;

s – сезонная компонента;

E – ошибка или иррегулярный член.

Нужно выделить каждое слагаемое в этой сумме, оценить их, тогда можно будет делать прогнозы. Прежде всего устраним (выделим) сезонную компоненту. Для этого применим движущееся среднее (*moving average*) с окном, равным периоду сезонной компоненты *m*.

$$y^t_{ma} = \frac{1}{m}\left(y^{t-\frac{m}{2}} + \ldots + y^{t-1} + y^t + y^{t+1} + \ldots + y^{t+\frac{m}{2}-1} \right) \qquad \text{при } m \text{ чётном.}$$

$$y_{ma}^2 = \frac{1}{4}\left(y^0 + y^1 + y^2 + y^3\right) = \frac{1}{4}(55 + 60 + 66 + 69) = 62.5,$$

$y_{ma}^3 = \frac{1}{4}\left(y^1 + y^2 + y^3 + y^4\right)$, … и т.д. y_{ma}^t – сглаженный ряд без сезонных колебаний (может быть нелинейным в глобальном поведении, что было бы потеряно, если сразу к исходному ряду подгонять линейный тренд!), поэтому ряд разностей $dif^t = y^t - y_{ma}^t$ представляет собой влияние сезонных добавок. Их нужно усреднить по имеющимся соответствующим одноимённым сезонам:

$$\tilde{s}^4 = \tfrac{1}{5}(dif^4 + dif^8 + dif^{12} + dif^{16} + dif^{20}) =$$
$$\tfrac{1}{5}(-9.25 - 8.0 - \ldots - 8.75) = -7.8 = \tilde{s}^0 .$$

Аналогично $\tilde{s}^1, \tilde{s}^2, \tilde{s}^3$. Но принято считать, что $\sum_{t=0}^{m-1} s^t = 0$, поэтому $\hat{s}^t = \tilde{s}^t - \frac{1}{m}\sum_{t=0}^{m-1} \tilde{s}^t$. Получим значения – 8.42, 0.18, 1.67, 6.58, которые периодически повторяем. Когда сезонная компонента оценена, устраним её:

$$y_{adj}^t = y^t - \hat{s}^t .$$

Adjusted – исправленный ряд, в нём остались тренд-цикл и ошибка.

Выделим из y_{adj}^t тренд-цикл в форме линейного тренда, т.е. считаем $y_{adj}^t = a + bt + e^t$. МНК даёт $\hat{a} = 59.75, \hat{b} = 1.036$. Остатки $\hat{e}^t = y_{adj}^t - \hat{a} - \hat{b}t$. Оценка дисперсии ошибок $\hat{\sigma}^2 = \frac{1}{T-2}\sum_{t=0}^{T-1}(\hat{e}^t)^2$, $\hat{\sigma}^2 = 2.88$.

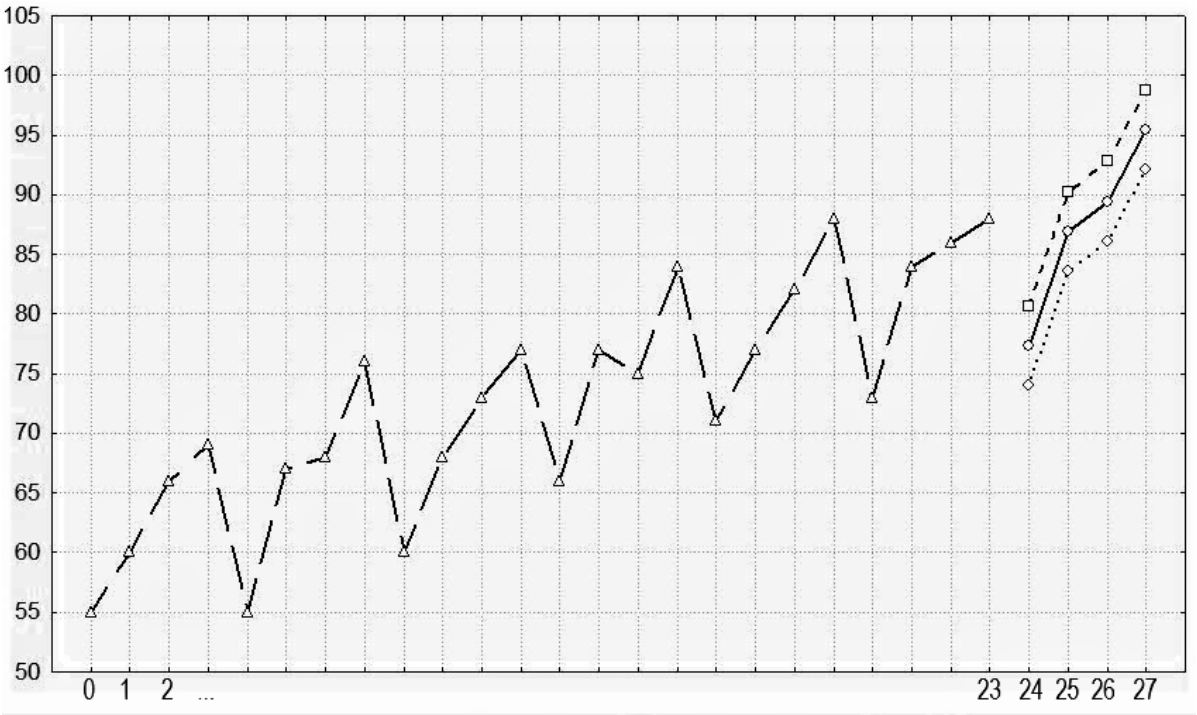
Прогноз в момент $t > T$ (T – обычно велико):

$$y_*^t = \hat{a} + \hat{b}t + \hat{s}^t \pm 2\hat{\sigma} \qquad (33)$$

– 95 %-ый интервал для прогноза. Например,

$$y_*^{24} = 59.75 + 1.036 \cdot 25 - 8.42 \pm 2 \cdot 1.7 = 77.23 \pm 3.4 .$$

Изобразим на графике исходный ряд и прогноз с доверительным интервалом.



Если y^t_{adj} окажется не очень линейно, то к нему можно подогнать полиномиальный тренд.

8.1. Фурье-анализ временных рядов

Имеющиеся T наблюдений временного ряда (1.1) (отстоящие через одинаковую единицу времени) можно представить суммой периодических составляющих:

$$y^t = \sum_{k=0}^{T-1} c^k e^{i\frac{2\pi kt}{T}}, \tag{34}$$

где

$$c^k = \frac{1}{T}\sum_{t=0}^{T-1} y^t e^{-i\frac{2\pi kt}{T}}. \tag{35}$$

(35), (34) – прямое и обратное дискретное преобразование Фурье.

Действительно, до множим (35) на $e^{i\frac{2\pi kt'}{T}}$ и просуммируем по k. П.ч. примет вид $\sum_{t=0}^{T-1} y^t \frac{1}{T}\sum_{k=0}^{T-1} e^{-i\frac{2\pi k(t-t')}{T}}$. Рассмотрим

$$\frac{1}{T}\sum_{k=0}^{T-1} e^{-i\frac{2\pi k(t-t')}{T}} = \left\{ \begin{array}{c} \dfrac{1+1+\ldots+1}{T} = 1, t = t' \\ \left| \sum_{k=0}^{T-1} q^k = \dfrac{1-q^T}{1-q} \right| = \dfrac{1-e^{-i2\pi(t-t')}}{1-e^{-i\frac{2\pi(t-t')}{T}}} = \dfrac{0}{\neq 0} = 0, 0 < |t-t'| < T \end{array} \right\} =$$

$$= \delta_{t,t'}, t, t' \in [0, T-1].$$

Т.о. $\sum_{k=0}^{T-1} c^k e^{i\frac{2\pi k t'}{T}} = y^{t'}$ или (34).

Видно, что в (34) $y^{t+T} = y^t$, значит (34) даёт бесконечное продолжение исходных наблюдений с периодом *T*. Можно получить вещественные формулы, эквивалентные (34, 35).

$$y^t = c^0 + \sum_{k=1}^{[\frac{T-1}{2}]} (c_c^k \cos\frac{2\pi k t}{T} + c_s^k \sin\frac{2\pi k t}{T}) + \begin{cases} c^{T/2}(-1)^t, \textit{если } T - \textit{четное} \\ 0, \textit{ если } T - \textit{нечетное} \end{cases}, \qquad (36)$$

$$c^0 = \frac{1}{T}\sum_{t=0}^{T-1} y^t, \qquad (37)$$

$$c_c^k = \frac{2}{T}\sum_{t=0}^{T-1} y^t \cos\frac{2\pi k t}{T}, \quad c_s^k = \frac{2}{T}\sum_{t=0}^{T-1} y^t \sin\frac{2\pi k t}{T}, \qquad (38)$$

$$c^{T/2} = \frac{1}{T}\sum_{t=0}^{T-1} y^t (-1)^t. \qquad (39)$$

Знак […] в (36) означает целую часть числа.

Зависимость квадрата модуля коэффициентов (35) разложения временного ряда от *k* называется *периодограммой*:

$$U(k) = \frac{1}{T}\left| \sum_{t=0}^{T-1} y^t e^{-i\frac{2\pi k t}{T}} \right|^2. \qquad (40)$$

Каков смысл чисел *k*? По (34) наблюдённая зависимость от времени (*t*) разложена на сумму гармоник (синусов и косинусов) с частотами *k* / *T* (число периодов в единицу времени). В силу периодичности (35) с периодом *T*, *k* в (34) можно брать от – (*T* / 2) + 1 до *T* / 2 (пусть *T* чётно, как принято считать при анализе периодограммы). Таким образом, в разложение входят частоты от 0 до 0.5, или входят периоды от бесконечности до 2 единиц времени (число разных –

T / 2). Наблюдая периодограмму, можно выявить скрытые периодичности во временном ряде (с частотами, где находятся её пики). Частота колебания – это число периодов в единицу времени (наблюдения отстоят на одну единицу времени). Например, частота 0.25 соответствует значению 4 периода (число единиц времени, требующих на полный цикл).

После этого можно оценить регрессионную модель y^t, например, на $x_1{}^t = cos(2\pi f_1 t), x_2{}^t = sin(2\pi f_1 t), \; x_3{}^t = cos(2\pi f_2 t), x_4{}^t = sin(2\pi f_2 t)$. Здесь f_1 и f_2 частоты, соответствующие важнейшим максимумам на периодограмме. Они необязательно должны совпадать с гармоническими. Это называется *модель циклических компонент*.

Гармоническая регрессия

Если мы уверены, что наш временной ряд обладает (зашумлённой) периодичностью с периодом m – период сезонности, где T кратно m, то соответствующее точное представление (36), где T заменено на m, берём за истинную функцию регрессии и оцениваем её коэффициенты по МНК. Это называется *гармоническая регрессия*. Результат будет:

$$\hat{y}^t = \hat{c}^0 + \sum_{k=1}^{[\frac{m-1}{2}]} (\hat{c}_c^k \cos\frac{2\pi kt}{m} + \hat{c}_s^k \sin\frac{2\pi kt}{m}) + \begin{cases} \hat{c}^{m/2}(-1)^t, если \;\; m - четное \\ 0, \; если \;\; m - нечетное \end{cases}, \quad (41)$$

где

$$\hat{c}^0 = \frac{1}{T}\sum_{t=0}^{T-1} y^t \,, \quad (42)$$

$$\hat{c}_c^k = \frac{2}{T}\sum_{t=0}^{T-1} y^t \cos\frac{2\pi kt}{m}, \quad \hat{c}_s^k = \frac{2}{T}\sum_{t=0}^{T-1} y^t \sin\frac{2\pi kt}{m}, \quad (43)$$

$$\hat{c}^{m/2} = \frac{1}{T}\sum_{t=0}^{T-1} y^t (-1)^t \,. \quad (44)$$

Продолжение примера 8.1

Применить гармоническую регрессию для оценки сезонной компоненты по ряду dif^t.

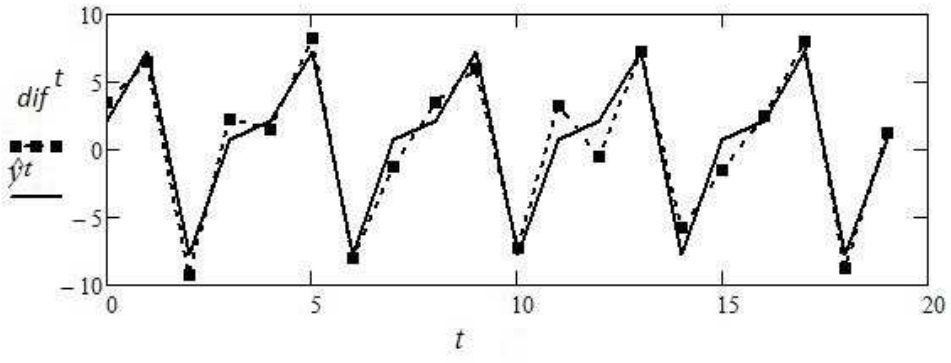
Из имеющихся отсчётов dif^t можно собрать 5 полных периодов сезонности, начинающихся с «зима».

3.5	6.5	– 9.25	2.25	1.5	8.25	– 8	– 1.25	3.5	6
– 7.25	3.25	– 0.5	7.25	– 5.75	-1.5	2.5	8	– 8.75	1.25

Подставляя это в (42)–(44) получим

$$(\hat{c}^0\ \hat{c}_c^1\ \hat{c}_s^1\ \hat{c}^2)^T = (0.575\ \ 4.95\ \ 3.2\ \ -3.425)^T.$$

Тогда формула (41) даёт оценки периодических сезонных вкладов «зима», «осень», «весна», «лето» = 2.1, 7.2, – 7.8, 0.8. Или график на фоне «наблюдений» dif^t:



Если сравнить суммы квадратов остатков представления наблюдённой сезонной компоненты dif^t в примере 8.1 и с помощью гармонической регрессии здесь, то получим 46.9 и 40.2.

Заметим, что формулы (42)–(44) и (37)–(39) дают одинаковые значения $\hat{c}^{k'} = c^k$ при $k'/m = k/T$. Это и позволяет брать оценки коэффициентов гармонической регрессии из Фурье-анализа. В нашем случае *Statistica* в Фурье-анализе, если убрать преобразования по умолчанию, даёт результаты:

	Spectral analysis: VAR1 (Spreadsheet3) No. of cases: 20				
	Frequency	Period	Cosine Coeffs	Sine Coeffs	Periodogram
0	0,000000		1,15000	-0,00000	13,2250
1	0,050000	20,00000	0,03172	-0,03400	0,0216
2	0,100000	10,00000	0,14844	0,05398	0,2495
3	0,150000	6,66667	-0,20444	-0,24424	1,0145
4	0,200000	5,00000	0,42291	-0,30726	2,7326
5	0,250000	4,00000	4,95000	3,20000	347,4250
6	0,300000	3,33333	0,65156	0,45951	6,3568
7	0,350000	2,85714	0,78030	-1,29204	22,7822
8	0,400000	2,50000	-0,24791	0,76299	6,4361
9	0,450000	2,22222	-0,18258	0,17273	0,6317
10	0,500000	2,00000	-6,85000	-0,00000	469,2250

В *Statistica* коэффициенты (37) и (39) вычисляются в два раза больше.

Связь временных рядов со случайными процессами [13].

8.2. Упражнения

Упражнение 8.1

Повторить в *Statistica* пример 8.1.

1	55	7	68	13	66	13	66	19	82
2	60	8	76	14	77	14	77	20	88
3	66	9	60	15	75	15	75	21	73
4	69	10	68	16	84	16	84	22	84
5	55	11	73	17	71	17	71	23	86
6	67	12	77	18	77	18	77	24	88

Advanced Lin/Nonlinear Models, *Time Series/Forecasting*, *Seasonal decomposition* (*Census* 1), *Additive*, *Seasonal Lag* 4,

Advanced, *Seasonal Factors*, *Seasonally Adj. Series*, *Summary*, *Review series*, *review multiple variables*, *Plot* (все 3), *OK*. *Advanced Other transformations* & *plots*, Выбрать *Adjusted* (*season* = 4), x = *f*(x), *Trend Subtract*, *OK*. *Descriptive*, *Std.Dv*. 1.6598.

Построить график: из окна отчёта *Seasonal decomposition* скопировать период *Seasonal Factors* в новую переменную файла данных для моментов прогноза, с её помощью по (33) создать 3 переменные, пометить их и исходный ряд, правый клик *Graphs of Block Data*, *Line Plot*, *Entire Columns*.

Упражнение 8.2

Администрация банка изучает динамику депозитов физических лиц за ряд лет (млн. долл. в сопоставимых ценах). Исходные данные представлены ниже.

Время, лет	1	2	3	4	5	6	7
Депозиты	2	6	7	3	10	12	13

1. Оценить уравнение линейного тренда и дать интерпретацию его параметров.
2. Определить коэффициент детерминации.
3. Администрация банка предполагает, что среднегодовой абсолютный прирост депозитов физических лиц составляет не менее 2.5 млн. долл. Подтверждается ли это предположение данными?

Упражнение 8.3

Проверить формулу (35).

Упражнение 8.4

Из формул (34), (35) получить (36)–(39).

Упражнение 8.5

Из формулы (9) получить (41)–(44). *Указание*: рассмотреть вначале комплексные гармоники $\vec{x}^t = (e^{\frac{i2\pi 0t}{m}} e^{\frac{i2\pi 1t}{m}} ... e^{\frac{i2\pi(m-1)t}{m}})^T$, затем получить вещественное представление.

9. Задания для самостоятельной работы

Задача № 1. Парная регрессия

1. Построить диаграмму рассеяния, пронумеровать 4 первые наблюдения.

2. Найти точечные оценки параметров линейной регрессии, записать оценку функции регрессии и построить её график на диаграмме рассеяния вместе с границами 80 %-х интервалов для предсказаний.

3. Найти стандартную ошибку аппроксимации и стандартные ошибки оценок коэффициентов регрессии.

4. Найти доверительные интервалы для коэффициентов регрессии с доверительной вероятностью $\gamma = 0.8$ для чётных вариантов и 0.95 для нечётных.

5. Проверить гипотезы о равенстве отдельных коэффициентов регрессии нулю (при альтернативе не равно), т.е. рассчитать уровни значимости.

6. Найти коэффициент детерминации и на уровне значимости 0.05 проверить гипотезу о равенстве всех коэффициентов регрессии (кроме константы) нулю.

7. Найти точечное и интервальное (с надёжностью 0.9) предсказание зависимой переменной при значении объясняющей переменной, равном максимальному наблюдённому её значению, увеличенному на 10 %. Изобразить их на диаграмме рассеивания.

8. Найти средний коэффициент эластичности зависимой переменной по независимой.

9. Визуальным методом проверить гипотезу нормальной распределённость ошибок.

10. По критерию Дёрбина–Уотсона проверить гипотезу об автокоррелированности ошибок.

Указание: ручные расчёты подтвердить расчётами в *Statistica* (кроме пунктов 4 и 8). Меню *Statistics*, *Multiple Regression*. Построение графика: *Graphs => Scatterplots => Advanced => Regression bands => predicted*.

В вариантах 1–12 исследуется зависимость производительности труда y (т/ч) от уровня механизации работ x (%) по данным 14 промышленных предприятий.

Вариант 1

x	10	25	56	38	75	23	66	34	17	21	89	19	10	52
y	7	22	30	29	39	15	32	22	20	20	55	15	10	39

Вариант 2

x	65	21	21	65	44	87	22	75	25	75	22	68	32	64
y	35	13	21	23	18	26	16	30	13	32	14	22	21	26

Вариант 3

x	44	56	60	50	65	25	24	46	17	17	84	81	28	42
y	20	27	40	18	30	25	22	8	13	13	32	26	17	28

Вариант 4

x	72	58	15	62	18	28	83	63	49	49	50	65	58	10
y	33	19	5	29	19	22	44	31	15	24	22	49	32	12

Вариант 5

x	65	30	52	41	56	64	44	10	58	56	27	10	16	78
y	18	5	4	25	23	6	14	23	13	7	6	8	5	19

Вариант 6

x	24	55	60	13	30	81	68	41	82	41	38	69	69	26
y	36	24	17	7	15	49	45	19	52	40	32	35	39	16

Вариант 7

x	73	57	55	63	10	37	71	79	68	81	52	72	25	73
y	39	22	28	24	4	21	23	34	33	43	19	31	17	34

Вариант 8

x	10	71	17	64	22	13	11	61	83	44	29	62	11	21
y	26	30	11	26	6	10	9	31	44	33	26	40	10	13

Вариант 9

x	49	74	10	40	76	81	62	68	86	39	82	46	79	13
y	25	33	12	20	30	34	27	27	31	20	33	15	36	14

Вариант 10

x	63	37	73	88	33	79	18	21	34	75	75	81	31	28
y	40	24	35	53	13	47	20	15	18	41	44	48	19	15

Вариант 11

x	43	79	31	64	86	73	75	20	11	69	82	71	60	64
y	24	43	18	37	42	48	29	20	21	43	49	35	27	38

Вариант 12

x	86	76	74	79	85	85	20	47	12	46	20	27	73	62
y	29	27	26	32	32	32	11	24	4	16	6	13	25	23

В вариантах 13–30 значения независимой переменной приведены в первой строке таблицы, зависимой – во второй.

Вариант 13

8	10	15	3	9	13	2	2	13	8	19	3	16	3	16
-16	-20	-30	-7	-20	-25	-3	-2	-20	-13	-35	-3	-29	-4	-33

Вариант 14

11	12	10	13	3	3	9	1	1	18	17	4	8	12	9
-9	-14	-9	-10	-1	-1	-12	1	1	-21	-20	-4	-6	-12	-9

Вариант 15

13	2	4	18	13	9	9	10	13	12	0	2	17	14	7
15	5	8	20	15	11	12	11	14	20	1	1	20	15	7

Вариант 16

13	8	0	12	11	4	0	1	17	16	13	10	2	14	0
13	7	2	10	8	9	1	2	16	12	13	12	5	13	3

Вариант 17

7	18	7	7	14	14	4	13	7	2	4	13	13	1	0
8	19	3	11	18	17	3	14	7	3	4	18	13	2	1

Вариант 18

15	3	15	16	12	12	5	15	14	1	17	6	9	3	19
-5	0	-7	-10	-5	-5	-1	-6	-3	1	-7	-4	-1	0	-8

Вариант 19

6	12	19	3	14	15	12	13	11	10	4	9	13	12	15
12	12	20	5	14	18	17	15	12	11	7	9	14	16	17

Вариант 20

16	19	4	7	11	19	7	2	0	15	0	8	9	14	11
49	61	16	23	33	54	22	7	4	46	4	29	31	45	35

Вариант 21

7	18	0	8	4	3	10	15	4	15	11	17	10	17	4
22	53	2	24	14	14	31	37	14	40	30	47	27	47	11

Вариант 22

7	18	19	11	8	9	17	15	11	5	17	6	17	3	17
-2	-3	-3	-2	1	0	0	-1	-1	1	-2	0	-2	0	-2

Вариант 23

2	11	17	17	7	13	5	19	11	2	12	7	9	18	19
-6	-20	-30	-28	-11	-27	-11	-30	-18	-3	-23	-14	-19	-35	-32

Вариант 24

3	4	4	8	5	7	4	12	7	19	17	19	4	7	5
5	2	6	6	7	7	6	9	5	12	16	14	3	7	6

Вариант 25

12	5	8	4	10	0	15	19	18	4	9	8	16	8	14
-17	-3	-8	-4	-16	3	-25	-26	-27	-5	-12	-10	-21	-7	-20

Вариант 26

18	16	2	18	17	19	6	17	8	5	10	15	9	11	6
-29	-24	-4	-28	-26	-23	-9	-20	-8	-5	-12	-15	-12	-19	-6

Вариант 27

0	0	5	14	2	17	4	6	12	19	7	4	11	0	18
0	1	12	35	6	43	11	14	31	49	19	14	29	0	49

Вариант 28

3	5	4	12	11	14	14	14	2	7	17	9	13	6	15
-1	-2	-1	-8	0	-8	-7	-9	2	0	-14	-7	-12	-1	-12

Вариант 29

19	7	8	5	4	12	17	8	19	6	12	17	4	10	11
-34	-14	-11	-10	-5	-22	-33	-17	-34	-8	-22	-29	-5	-18	-23

Вариант 30

10	15	14	1	19	12	18	18	9	15	14	7	16	8	1
6	7	11	1	15	11	17	14	7	11	11	8	11	6	3

Задача № 2. Множественная регрессия

1. Провести линейный регрессионный анализ со всеми имеющимися регрессорами. При наличии сильной мультиколлинеарности, возможно, придётся уменьшить параметр *tolerance* в используемой процедуре пакета статистического анализа. Привести результаты: оценку функции регрессии, параметры оценок коэффициентов регрессии, коэффициент детерминации, уровень значимости против гипотезы о равенстве всех коэффициентов регрессии (кроме константы) нулю, стандартную ошибку аппроксимации.

2. Указать признаки отягощённости мультиколлинеарностью, привести результаты корреляционного анализа регрессоров, применить ридж-регрессию с параметром = 0.1. Привести результаты.

3. Убрать «галку» с *Ridge regression*. Применяя пошаговую регрессию вперёд, ввести в модель два регрессора, обеспечивающих наилучшее описание зависимой переменной без отягощённости мультиколлинеарностью. Привести результаты. Сравнить с п. 1 и 2. Сделать выводы.

4. На основании результатов п. 3 найти: а) средние коэффициенты эластичности зависимой переменной по независимым; б) точечное и интервальное (с надёжностью 0.9) предсказание зависимой переменной при значении важнейшей объясняющей переменной, равном максимальному наблюдённому её значению, увеличенному на 10 %, и при значении второй объясняющей переменной, равном минимальному наблюдённому её значению, уменьшенному на 15%; в) проверить гипотезу нормальной распределённости ошибок.

Указание: расчёты проводить в *Statistica* (кроме пункта 4 а)) побригадно.

В вариантах 1–12 исследуется зависимость производительности труда y (т/ч) от уровня механизации работ x_1 (%), среднего возраста работников x_2 (лет) и энерговооружённости x_3 (КВт/100 работающих) по данным 14 промышленных предприятий.

Вариант 1

x_1	32	30	36	40	41	47	56	54	60	55	61	67	69	76
x_2	33	31	41	39	46	43	34	38	42	35	39	44	40	41
x_3	300	290	350	400	400	480	500	520	590	540	600	700	700	750
y	20	24	28	30	31	33	34	37	38	40	41	43	45	48

Вариант 2

x_1	55	46	40	39	35	29	31	75	68	66	60	54	59	53
x_2	33	42	45	38	40	30	32	40	39	43	38	34	41	37
x_3	500	450	390	400	340	300	300	745	690	660	590	545	600	525
y	33	32	30	29	27	23	19	47	44	42	40	39	37	36

Вариант 3

x_1	48	57	55	61	56	62	68	70	77	42	41	37	31	33
x_2	44	35	39	43	36	40	45	41	42	47	40	42	32	34
x_3	475	560	540	620	565	625	670	700	760	420	400	375	305	325
y	34	35	38	39	41	42	44	46	49	32	31	29	25	21

Вариант 4

x_1	52	54	45	39	38	34	28	30	74	67	65	59	53	58
x_2	36	32	41	44	37	39	29	31	39	38	42	37	33	40
x_3	520	535	455	385	385	345	285	310	730	660	650	600	525	570
y	35	32	31	29	28	26	22	18	46	43	41	39	38	36

Вариант 5

x_1	43	49	58	56	62	57	63	69	71	78	34	32	38	42
x_2	48	45	36	40	44	37	41	46	42	43	35	33	43	41
x_3	425	485	580	550	610	560	620	700	700	785	350	325	385	415
y	33	35	36	39	40	42	43	45	47	50	22	26	30	32

Вариант 6

x_1	52	57	51	53	44	38	37	33	27	29	73	66	64	58
x_2	32	39	35	31	40	43	36	38	28	30	38	37	41	36
x_3	525	565	500	535	455	375	375	325	280	305	725	660	645	590
y	37	35	34	31	30	28	27	25	21	17	45	42	40	38

Вариант 7

x_1	39	43	44	50	59	57	63	58	64	70	72	79	35	33
x_2	44	42	49	46	37	41	45	38	42	47	43	44	36	34
x_3	450	425	500	465	380	400	455	390	415	480	435	440	355	340
y	31	33	34	36	37	40	41	43	44	46	48	51	23	27

Вариант 8

x_1	63	57	51	56	50	52	43	37	36	32	26	28	72	65
x_2	40	35	31	38	34	30	39	42	35	37	27	29	37	36
x_3	395	380	350	350	325	310	285	280	255	250	210	155	455	400
y	39	37	36	34	33	30	29	27	26	24	20	16	44	41

Вариант 9

x_1	64	59	65	71	73	80	36	34	40	44	45	51	60	58
x_2	46	39	43	48	44	45	37	35	45	43	50	47	38	42
x_3	500	400	500	550	500	600	350	345	420	410	480	490	500	500
y	42	44	45	47	49	52	24	28	32	34	35	37	38	41

Вариант 10

x_1	46	52	61	59	65	60	66	72	74	81	37	35	41	45
x_2	51	48	39	43	47	40	44	49	45	46	38	36	46	44
x_3	460	520	600	585	645	610	655	720	745	805	380	345	405	440
y	36	38	39	42	43	45	46	48	50	53	25	29	33	35

Вариант 11

x_1	62	30	36	50	41	47	56	54	60	55	61	67	69	66
x_2	43	51	41	39	46	43	34	38	42	25	39	44	40	41
x_3	555	400	380	460	420	450	410	450	500	400	500	550	550	500
y	5	2	2	3	3	3	4	3	3	4	4	4	4	4

Вариант 12

x_1	45	46	40	39	35	29	61	75	68	66	60	54	59	53
x_2	63	42	45	38	40	30	32	40	39	43	38	34	41	37
x_3	625	415	445	380	395	305	315	385	360	450	375	350	400	360
y	3	2	3	9	7	3	9	7	4	2	6	9	7	6

Задача № 3. Введение фиктивных переменных

1. Добавить (домыслить) правдоподобную качественную переменную к данным своего варианта задания «Парная регрессия», приписав каждому наблюдению её уровень (2 уровня в нечётных вариантах и 3 в чётных). При этом

добиться значимого влияния качественной переменной. Привести обновлённые данные в отчёте с указанием смысла всех трёх переменных.

2. Ввести в модель нужное число дихотомических фиктивных переменных, оценить параметры модели и из общей оценки записать оценки функций регрессии для каждого уровня качественной переменной отдельно.

3. Обосновать вывод о наличии влияния качественной переменной и описать его характер. Предложить объяснение причины.

4. По данным для какого-либо уровня отдельно оценить функцию регрессии. Для этого воспользоваться кнопкой *Select Cases* диалога *Multiple Linear Regression*. Сравнить результаты моделирования с таковыми п. 2. Сделать выводы.

Указание: расчёты проводить в *Statistica*.

Задача № 4. Линеаризация

1. Подбором нелинейных преобразований исходных переменных в своём варианте задания «Парная регрессия» добиться улучшения представления данных с помощью нелинейной функции регрессии.

2. Сравнить коэффициент детерминации и уровень значимости против гипотезы о равенстве всех коэффициентов регрессии (кроме константы) нулю с таковыми для линейной функции регрессии. Сделать выводы.

3. Записать оценённую нелинейную функцию регрессии и построить её график вместе с линейной функцией регрессии на диаграмме рассеяния.

4. Найти средний коэффициент эластичности зависимой переменной по независимой в полученной нелинейной модели и сравнить его с таковым в линейной.

Указание: расчёты проводить в *Statistica* (кроме п. 4). *Advanced Linear/Nonlinear Models*, *Fixed Nonlinear Regression*. Для построения графика: *Graphs*, *Scatterplots*, *Variables*, *Regression bands*, *OK*. Правый клик на график, *Graph Properties*, *Custom Function*, *Add new function*.

Задача № 4*а*. Эффект масштаба и кривая обучения (опыта)

Изучается модель вида: $c^t = k(n^t)^{\frac{\alpha_c}{r}}(y^t)^{\frac{1}{r}-1}e^{e^t}$,

где c^t – дефлятированные удельные издержки в периоде времени *t*;

y^t – выпуск (в физическом выражении) в периоде времени *t*;

n^t – накопленный выпуск (в физическом выражении) до периода времени *t*;

e^t – случайная ошибка;

r – отдача от масштаба;

– α_c – эластичность выпуска по накопленному выпуску (опыту).

Использовать данные по производству диоксида титана компанией *DuPont* и ВНП в качестве дефлятора.

Год	Выпуск	Накопл. выпуск	Издержки на тонну	ВНП
1955	112	1127	17.1	68.5
1956	124	1239	18.1	71
1957	116	1363	18.7	74
1958	101	1479	18.6	73.1
1959	125	1580	18.4	73.5
1960	134	1705	18.6	75.2
1961	149	1839	17.6	76.1
1962	168	1988	16.5	76
1963	177	2156	16.4	76.3

1. Дефлятировать удельные издержки (к ценам 1955 г.).

2. Линеаризовать модель.

3. Построить диаграммы рассеивания зависимой переменной с каждым фактором по отдельности.

4. Проверить гипотезу, что эффект масштаба равен 0.

5. Найти точечные оценки отдачи от масштаба и эластичности выпуска по накопленному выпуску и их стандартные ошибки. Изобразить кривую обучения. Интерпретировать и обсудить результаты: ситуация на предприятии. Что с техническими знаниями? Следует ли рекомендовать его экстенсивный рост или сокращение?

6. Предположив прирост выпуска в 1964 году, равным среднему приросту за последние 3 года наблюдений, найти точечный и 80 %-ый интервальный прогнозы для удельных издержек в 1964 году. Какую цену за тонну можно установить на 1964 год, если рассмотреть прирост ВВП аналогично приросту выпуска?

Указание: расчёты проводить по бригадам в *Statistica* или *MathCAD*.

Задача № 5. Параболическая регрессия

1. Найти оценку функции параболической (степени 2) регрессии.

2. Построить диаграмму рассеяния, пронумеровать 4 первые наблюдения и нанести на неё график оценённой регрессии.

3. Найти коэффициент детерминации и на уровне значимости 0.05 проверить гипотезу о равенстве всех коэффициентов регрессии (кроме константы) нулю.

4. Найти точечное и интервальное (с надёжностью 0.9) предсказание зависимой переменной при значении объясняющей переменной, равной максимальному наблюдённому её значению, увеличенному на 10 %. Нанести прогнозы на диаграмму рассеяния.

5. Найти средний коэффициент эластичности зависимой переменной по независимой.

6. Визуальным методом проверить гипотезу нормальной распределённость ошибок.

7. По критерию Дёрбина–Уотсона проверить гипотезу о автокоррелированности ошибок.

Указание: расчёты проводить побригадно в *Statistica*.

В таблице для каждого варианта указаны наблюдённые значения независимой (первая строка) и зависимой переменной.

Вариант 1

14	5	15	11	10	14	14	18	15	7	1	0	15	3	13
183	-11	186	96	81	156	140	241	155	91	-46	21	216	40	160

Вариант 2

19	16	16	17	18	18	2	9	0	9	2	4	15	13	15
54	28	25	15	30	73	27	41	-17	18	40	14	36	28	50

Вариант 3

0	5	4	17	0	1	1	19	17	8	13	18	0	0	8
-3	-14	-6	56	-3	9	25	70	56	0	35	69	-1	-18	17

Вариант 4

6	6	5	3	1	19	8	3	9	4	5	16	12	0	5
10	-49	-18	-9	25	-65	-20	2	-11	10	4	-66	-30	-8	11

Вариант 5

11	2	16	1	15	5	7	9	18	17	19	15	2	15	3
-37	-25	-61	9	-53	17	25	31	-49	-70	-112	-67	-6	-33	44

Вариант 6

7	16	0	19	9	10	1	14	12	19	1	18	4	12	11
-34	38	38	-7	23	-12	20	23	2	7	-12	5	5	25	-15

Вариант 7

2	1	7	10	0	1	14	16	1	2	13	5	6	1	16
8	1	-13	-3	27	11	13	21	5	24	30	21	20	6	38

Вариант 8

16	3	4	1	7	12	15	4	2	3	16	18	11	19	15
27	-36	54	14	46	55	19	18	9	46	38	59	21	42	44

Вариант 9

19	17	4	5	3	12	14	3	18	11	17	0	12	1	9
43	82	8	14	-15	49	37	7	62	18	-1	-10	40	-10	25

Вариант 10

8	1	17	19	12	9	14	5	8	10	2	2	0	13	13
-16	-24	-61	-144	-56	-33	-61	16	-21	-28	-10	12	-22	-27	-47

Вариант 11

4	7	15	2	2	7	9	4	15	4	5	13	17	8	14
28	23	38	8	4	-3	15	-3	-30	-37	-1	10	33	27	2

Вариант 12

17	18	0	19	16	17	19	0	9	0	1	18	6	8	14
-56	-30	-2	-69	-75	-72	-62	-24	-1	-25	21	-81	-6	-25	-24

Вариант 13

3	15	13	13	8	2	2	18	9	12	12	16	7	10	9
23	136	51	77	13	28	0	114	44	68	42	126	-17	100	55

Вариант 14

9	11	4	9	8	1	4	8	17	4	4	5	15	9	5
31	81	11	51	37	4	-3	82	146	29	31	18	115	60	53

Задача № 6. Логит- и пробит-модели

1. Подобрать данные с числом факторов не менее 2 и числом наблюдений ≥ 15.

2. Оценить модель с помощью метода максимального правдоподобия, проверить значимость модели.

3. Построить 3D график.

4. Построить таблицу наблюдённых, предсказанных значений и остатков, указать Odds Ratio.

5. Найти средний маржинальный эффект какого-либо фактора.

6. Сделать прогноз для нового объекта.

Указание: (см. [13]), расчёты проводить побригадно в *Statistica*.

Задача № 7. Временные ряды (имитационное моделирование и сезонная декомпозиция)

1. Сгенерировать не менее 6 периодов сезонности временного ряда. Период сезонности m = 4 для нечётных и 7 для чётных вариантов. Формула для генерации в Variable Specifications: = a + b*V0 + S +(Rnd(1) –0.5)*c. Разумные параметры a, b, c и сезонные эффекты S подобрать самим.

2. Построить график временного ряда и показать его преподавателю.

3. Произвести аддитивную декомпозицию полученного ряда, выделив оценки сезонной компоненты и линейного тренда. Сопоставить их с S, a, b.

4. По ряду остатков оценить дисперсию ошибок, сопоставить её с параметром c, сделать точечный и интервальный прогноз (с доверительной вероятностью 0.95) на глубину в период сезонности.

5. Изобразить на графике исходный ряд и прогнозные значения с доверительными интервалами.

Указание: расчёты проводить побригадно в *Statistica*.

Задача № 8. Системы эконометрических уравнений (косвенный МНК)

По данным, представленным в таблице, построить модель вида

$$\begin{cases} y_1 = a_1 + b_{12} y_2 + c_{11} x_1 + e_1 \\ y_2 = a_2 + b_{21} y_1 + c_{22} x_2 + e_2 \end{cases}.$$

Год	Годовое потребление свинины на душу населения, фунтов, y_1	Оптовая цена за фунт, долл., y_2	Доход на душу населения, долл., x_1	Расходы по переработке мяса, % к цене, x_2
1990	60	5.1	1300	62
1991	63	4.2	1300	56
1992	65	4.0	1500	56
1993	62	5.0	1600	63
1994	66	3.8	1800	50

1. Применяя условие размерности, установить идентифицируемость каждого уравнения структурной формы.

2. Записать приведённую форму уравнений.

3. Выразить коэффициенты структурной формы через коэффициенты приведённой формы.

4. Косвенным МНК идентифицировать структурную форму модели.

5. Дать содержательное объяснение знакам коэффициентов структурной формы.

6. Найти средние коэффициенты эластичности потребления по цене и доходу.

7. Сравнить результаты косвенного МНК с простым МНК, применённым к первому уравнению структурной формы.

Указание: расчёты проводить побригадно в *Statistica*.

Задача № 9. Системы эконометрических уравнений (двухшаговый МНК)

Изучается модель вида

$$\begin{cases} y^t = a_1 + b_1(c^t + d^t) + e_1^t \\ c^t = a_2 + b_2 y^t + b_3 y^{t-1} + e_2^t \end{cases},$$

где y – валовой национальный доход;

c – личное потребление;

d – конечный спрос (помимо личного потребления).

Исходные данные представлены в таблице.

Год	d	y	c	Год	d	y	c
0	–	46.7	–	5	5.9	17.8	25.8
1	-6.8	3.1	7.4	6	44.7	37.2	8.6
2	22.4	22.8	30.4	7	23.1	35.7	30.0
3	-17.3	7.8	1.3	8	51.2	46.6	31.4
4	12	21.4	8.7	9	32.3	56.0	39.1

1. Применяя условие размерности, проверить, идентифицируемо ли каждое уравнение структурной формы.

2. Записать приведённую форму уравнений.

3. Двухшаговым МНК идентифицировать первое уравнение структурной формы модели.

Указание: расчёты проводить побригадно в *Statistica*.

Библиографический список

Основная литература

1. Айвазян С. А. Прикладная статистика и основы эконометрики / С. А. Айвазян, В. С. Мхитарян. – М. : ЮНИТИ-ДАНА, 2001.

2. Доугерти К. Введение в эконометрику / К. Доугерти. – М. : Инфра-М, 2001.

3. Тихомиров Н. П. Эконометрика / Н. П. Тихомиров, Е. Ю. Дорохина. – М. : Экзамен, 2003.

4. Магнус Я. Р. Эконометрика. Начальный курс / Я. Р. Магнус, П. К. Катышев, А. А. Пересецкий. – М. : Дело, 2005.

5. Берндт Э. Р. Практика эконометрики / Э. Р. Берндт. – М. : ЮНИТИ-ДАНА, 2005.

Дополнительная литература

6. Тюрин Ю. Н. Статистический анализ данных на компьютере / Ю. Н. Тюрин, А. А. Макаров. – М. : Инфра-М, 1998.

7. Ивченко Г. И. Математическая статистика / Г. И. Ивченко, Ю. И. Медведев. – М. : Высш. шк., 1984.

8. Пугачев В. С. Теория вероятностей и математическая статистика / В. С. Пугачев. – М. : ФИЗМАТЛИТ, 2002.

9. Руководство пользователя пакета программ *Statistica v*. 8

10. Справочник по прикладной статистике : в 2 т. / ред. Э. Ллойд, У. Ледерман. – М. : Финансы и статистика, 1990.

11. *Gatignon H. Statistical Analysis of Management Data* / *H. Gatignon*. – *Kluwer Academic Publishers*, 2003.

12. *Salvatore D. Statistics and econometrics* / *D. Salvatore, D. Reagle*. – *McGraw-Hill*, 2002.

13. Эконометрическое моделирование : учебное пособие / С. М. Бородачёв. Екатеринбург : УГТУ–УПИ, 2008.

Учебное издание

Бородачёв Сергей Михайлович

ЭКОНОМЕТРИКА

Редактор *Н. В. Рощина*

Компьютерная вёрстка *авторская*

Корректор *А. А. Пыжьянова*

Подписано в печать 08.11.2011г. Формат 60 х 84 1/16.
Бумага писчая. Плоская печать. Усл. печ. л. 4,42.
Уч.-изд. л. 3,2. Тираж 100 экз. Заказ

Редакционно-издательский отдел УрФУ
620002, Екатеринбург, ул. Мира, 19
rio@mail.ustu.ru

Ризография НИЧ УрФУ
620002, Екатеринбург, ул. Мира, 19