Оглавление

Предисловие

Данное учебное пособие предназначается для студентов и аспирантов математических специальностей (математика, прикладная математика). Поэтому в нем, в основном, дано описание математических моделей, которые традиционно используются в курсах под названием "Эконометрика". В силу этого основное внимание уделяется аккуратному описанию самих моделей и изучению их свойств. Большинство результатов приведено с полными доказательствами. Исключение составляет глава, посвященная моделям с гетероскедастичностью и автокорреляцией в ошибках, т.к. необходимые здесь методы значительно сложнее тех, что излагаются в пособии. Эта особенность выделяет данное пособие из большого числа учебников и учебных пособий, посвященных той же теме и написанных для разных категорий студентов и научных работников и исходящих из разных методологических и методических позиций. Именно это и позволяет рекомендовать его именно студентам-математикам.

Как отмечалось выше, пособие посвящено описанию математических моделей и их свойств. Рассматриваемые по ходу изложения примеры носят чисто иллюстративный характер. Автор, который много лет работает со студентами различных университетов, отлично понимает что обучение практическому применению описанных моделей есть отдельная, очень непростая и деликатная задача. Она требует рассмотрения большого числа хорошо подобранных реальных примеров, которые иллюстрируют многочисленные особенности работы с реальными данными, тщательного анализа имеющихся данных, проверки выполнения ограничений, накладываемых на модели, обучения работе с существующими пакетами прикладных программ по статистике и многого другого. В последние годы появилось много публикаций с критикой методов обучения студентов применению статистических методов к решению прикладных задач, особенно в экономике, социологии, психологии и других "нетехнических"науках. Особенно ярко это выражено в книге В.Н. Тутубалина ([8]). Со многими их возражениями автор данного пособия абсолютно согласен.Но, тем не менее, сами математические модели не виноваты в их неправильном применении! Более того, эта критика подталкивает к более тщательному изучению их свойств и условий применимости хотя бы теми студентами, которые в будущем должны будут создавать новые модели подобного типа.

В заключение напомним, что данное пособие носит вводный характер и в нем дано описание только некоторых базовых моделей. Автор предполагает написать вторую часть пособия, посвященную и другим более продвинутым моделям и методам, что позволит избежать упрека в том, что мы обучаем студентов только "убогой"эконометрике.

При написании данного пособия автор использовал материалы, собранные на семинарах по предодаванию начального и продвинутых курсов эконометрики, которые были организованы при Российской Экономической Школе. Организаторам и участникам этих семинаров автор многим обязан и выражает свою искреннюю благодарность.

Глава 1

Экономические и эконометрические модели

Принципиальная идея при изучении экономики состоит в том, что экономические переменные связаны между собой и оказывают влияние друг на друга.

Примеры. 1) Величина спроса есть функция цены.

2) Затраты на изготовление некоторого продукта есть функция от объема производства.

3) Потребительские расходы зависят от личных доходов.

В реальных задачах связано большое число переменных. Например, спрос на товар есть функция от цены, величины доходов и цен на конкурирующие и дополняющие товары. Экономическая теория состоит не из отдельных соотношений. Они группируются вместе, образуя так называемую экономическую модель изучаемого явления. *Модель* – это формализация идей и знаний, относящихся к определенному явлению.

Пример. Традиционная модель спроса и предложения должна объяснить соотношения между ценой и объемом выпуска, характерные для некоторого определенного рынка. Она содержит три уравнения:

-уравнение спроса,
-уравнение предложения,
-уравнение реакции рынка.

В эти уравнения входят две основные переменные–цена и объем продукции, а также другие переменные–потребительский доход, цены на иные товары и т.п. Т.о. в наши уравнения кроме *основных* переменных, которые *объясняются* в рамках модели, входят и некоторые "внешние"переменные. В этом смысле модель является "неполной".

В отличие от физических наук в социальных и экономических науках сильно ограничена возможность экспериментирования. В силу этого трудно отделить и изучить влияние отдельных факторов.

Все экономические модели имеют некоторые общие черты:

1) предполагается, что поведение экономических переменных определяется с помощью *совместных* и *одновременных* операций с некоторым числом экономических соотношений,

2) модель есть упрощение сложной действительности, но улавливает главные характеристики изучаемого объекта,

3) предполагается, что с помощью модели удается *предсказать* будущее поведение системы и *влиять* на нее.

После того, как мы отобрали некоторое количество переменных, которые будут использоваться в нашей модели, и, используя экономическую теорию, отобрали соотношения, которые образуют нашу модель, необходимо уточнить форму связи этих переменных в каждом соотношении. Этот этап называется *спецификацией* модели. Связи могут быть линейными, логарифмическими, полиномиальными и т.п. Обычно экономическая теория указывает только некоторые общие свойства этих соотношений, не конкретизируя их форму.

Чтобы уточнить структуру *эконометрической модели*, рассмотрим более подробно следующий пример.

Простейшая макромодель

Из экономической теории в качестве основы взяты следующие положения:

1) потребление есть возрастающая функция от имеющегося в наличие дохода, но растет медленнее, чем сам доход,

2) объем инвестиций есть возрастающая функция национального дохода и убывающая функция характеристик государственного регулирования (например, нормы процента),

3) национальный доход есть сумма потребительских, инвестиционных и государственных закупок товаров и услуг.

Эти положения выделяют некоторое число экономических переменных, но не определяют *конкретной формы* их связи. Кроме того, влияние одних переменных на другие может запаздывать во времени. Обычно выбирают наиболее простую форму связи, которая реализует сформулированные экономические положения. В нашем случае это приводит, например, к следующей системе соотношений.

$$C_t = \alpha_0 + \alpha_1(Y_t - T_t) \ , \tag{1.1}$$

$$\begin{aligned} I_t &= \beta_1 Y_{t-1} + \beta_2 R_t \ , & (1.2) \\ Y_t &= C_t + I_t + G_t \ . & (1.3) \end{aligned}$$

где

$$0 < \alpha_1 < 1,\ \beta_1 > 0,\ \beta_2 < 0 \ .$$

В этих соотношениях:
C–потребление,
I–инвестиции,
Y–национальный доход,
G–государственные закупки товаров и услуг,
T–подоходный налог,
R–инструмент государственного регулирования.

Заметим, что здесь присутствуют разные типы переменных:
C_t, I_t, Y_t–*текущие эндогенные* переменные,
T_t, R_t, G_t–*текущие экзогенные* переменные,
Y_{t-1}–*лаговая эндогенная* переменная.

Эндогенные переменные объясняются в рамках модели, экзогенные являются "внешними"и играт роль ограничений (часто управляемых). Экзогенные и лаговые эндогенные переменные называются *предопределенными*.

Уравнения (1)-(3) задают так называемую *структурную* форму модели.

Выделим текущие эндогенные переменные:

$$\begin{aligned} C_t &= \frac{\alpha_0}{1-\alpha_1} + \frac{\alpha_1\beta_1}{1-\alpha_1} Y_{t-1} + \frac{\alpha_1\beta_2}{1-\alpha_1} R_t + \frac{\alpha_1}{1-\alpha_1}(G_t - T_t), & (1.4) \\ I_t &= \beta_1 Y_{t-1} + \beta_2 R_t \ , & (1.5) \\ Y_t &= \frac{\alpha_0}{1-\alpha_1} + \frac{\beta_1}{1-\alpha_1} Y_{t-1} + \frac{\beta_2}{1-\alpha_1} R_t \end{aligned}$$

$$+\frac{1}{1-\alpha_1}G_t - \frac{\alpha_1}{1-\alpha_1}T_t\ . \tag{1.6}$$

Это *приведенная* форма модели. Она позволяет в явном виде увидеть зависимость текущих эндогенных переменных от предопределенных переменных.

Теоретические модели обычно не годятся для непосредственного использования в эконометрических исследованиях. Это связано с тем, что мы отобрали только небольшое число переменных, которые, на наш взгляд, являются основными. Но и потребление и инвестиции зависят еще от большого числа других переменных. Сколько бы факторов мы не учли, всегда останется что-нибудь еще, что влияет на наши переменные, но не учтено в наших уравнениях. Разумное решение состоит в том, чтобы признать следующий факт: нам частично незвестны факторы, определяющие потребление и инвестиции. Мы можем только *оценить* их вероятное влияние, считая их случайными. Поэтому, строго говоря, уравнения (1) – (2) следует переписать в следующем виде:

$$\begin{aligned} C_t &= \alpha_0 + \alpha_1(Y_t - T_t) + \varepsilon_{1t}\ , && (1.7)\\ I_t &= \beta_1 Y_{t-1} + \beta_2 R_t + \varepsilon_{2t}\ , && (1.8)\\ Y_t &= C_t + I_t + G_t\ , && (1.9) \end{aligned}$$

где ε_1 и ε_2 есть результаты воздействия неучтенных факторов, которые мы интерпретируем как *случайную ошибку*. На самом деле механизм образования ошибки (невязки) может быть и не стохастическим и мы не можем применять методы теории вероятностей. Но в эконометрике принимается гипотеза о том, что, *ошибки являются случайными величинами* с некоторым распределением вероятностей.

Уравнения (7) – (9) задают так называемую *эконометрическую модель*. Параметрами этой модели являются неизвестные параметры α_0, α_1, β_1, β_2 и распределение ошибок ε_1 и ε_1 (или их параметры, если распределения принадлежат некоторым параметрическим семействам).

В рамках описанной выше эконометрической модели необходимо решить несколько задач. Перечислим некоторые из них.

1) Адекватна ли предложенная форма модели имеющимся экономическим данным?

2) Оценить параметры модели по данным.

3) Сформулировать правила прогноза будущих значений для некоторых переменных.

Глава 2

Понятие о корреляционном и регрессионном анализе

2.1 Основные задачи корреляционного и регрессионного анализа

Как мы отмечали выше, основная задача эконометрики с точки зрения математики это построение математической модели, в которой отражена взаимосвязь интересующих нас экономических величин. Эта связь редко выражается в чисто функциональном виде. Чаще всего она носит стохастический характер, имея форму определенной тенденции. Это обусловлено тем, что мы принимаем во внимание только часть факторов, влияющих на изучаемую величину. Например, известно, что себестоимость продукции зависит от объема производства. Но на нее влияют также потери от брака, ассортимент продукции, технология производства, используемое сырье, структура цен и многое другое . Влияние всех неучтенных факторов мы собираем вместе и записываем с помощью некоторой стохастической возмущающей переменной. В эконометрике *предполагается*, что это случайная величина, обладающая некоторым вероятностным распределением. Тогда мы приходим к задаче оценки влияния одних величин на другие и поиску функциональной связи между ними в условиях присутствия возмущающего воздействия.

Корреляция в широком смысле означает связь между изучаемыми процессами и явлениями. При этом мы имеем две задачи:

1) выявление наличия и определение силы связи,
2) выбор вида и формы связи.

В математической статистике методы решения первой задачи образуют раздел, называемый *корреляционным анализом.* При этом используются те или иные *коэффициенты корреляции.* Определение формы связи осуществляется методами *регрессионного анализа.*

Необходимо отметить, что методами статистики, используя экпериментальные данные, мы не подтверждаем *причинную* связь тех или иных величин. Мы только определяем, что между ними существует определенная связь. Но она может быть обусловлена и влиянием каких-то общих третьих величин. Вопрос о причинной связи должен изучаться методами конкретных наук, относящихся к данной области исследования: физики, химии, биологии, экономики и т.д. Но наличие связи между некоторыми величинами может подтолкнуть нас на поиск истинной причины взаимосвязи величин и подсказать правильное направление такого поиска. В качестве курьеза приведем два примера сильной корреляционной связи, за которой явно не стоит никакой причинной связи. В Германии была отмечена тесная положительная корреляция между количеством импортируемых апельсинов и числом смертных случаев от онкологических заболеваний за 50 лет. Другой пример– корреляция между числом аистов, свивших гнезда в южных районах Швеции, и рождаемостью в эти же годы в Швеции.

После того, как обнаружено наличие достаточно сильной связи между величинами, мы приходим к задаче определения вида этой связи. В классическом регрессионном анализе рассматривается связь между одной *зависимой* переменной Y, называемой *откликом* системы, и одной или несколькими *независимыми* переменными $X_1, \ldots, X_p$, называемыми *факторами*. Эта связь представляется в виде

$$Y = f(X_1, \ldots, X_p) + Z \ ,$$

где Z описывает влияние неучтенных факторов и ошибок измерений. Мы хотим найти такую функцию f, которая наилучшим образом учитывает влияние факторов $X_1, \ldots, X_p$. Из курса теории вероятностей известно, что в случае, когда качество аппроксимации измеряется с помощью среднеквадратического расстояния, наилучшей функцией является *функция регрессии*, т.е. условное математическое ожидание с.в. Y при условии, что $X_1 = x_1, \ldots, X_p = x_p$:

$$y = g(x_1, \ldots, x_p) := M(Y|X_1 = x_1, \ldots, X_p = x_p) \ .$$

Основными задачами регрессионного анализа являются следующие:

1) выбор класса моделей,
2) получение оценок параметров,
3) проверка гипотез о параметрах,
4) проверка адекватности модели,
5) проверка выполнения основных предположений.

2.2 Коэффициенты корреляции

Как было отмечено выше, при исследовании зависимости между случайными величинами вначале необходимо выяснить сам факт наличия такой связи, а затем измерить насколько она сильна. Обычно это делается с помощью тех или иных коэффициентов корреляции. В этом разделе мы рассматриваем различные типы коэффициентов корреляции и описываем их свойства.

2.2.1 Простой коэффициент корреляции

Этот коэффициент применяется для измерения величины линейной связи между двумя случайными величинами.

Определение 1 . *(Простым) коэффициентом корреляции с.в.* ξ_1 *и* ξ_2 *называется число*

$$\rho = \rho(\xi_1, \xi_2) := \frac{cov(\xi_1, \xi_2)}{\sqrt{D(\xi - 1, \xi_2)}} \ .$$

В курсе теории вероятностей были доказаны следующие свойства коэффициента корреляции.

Теорема 1 .

1) $|\rho(\xi_1, \xi_2)| \leq 1$.

2) Если ξ_1 *и* ξ_2 *некоррелированы, то* $\rho(\xi_1, \xi_2) = 0$.

3) Пусть $\eta_1 = a_1\xi_1 + b_1$, $\eta_2 = a_2\xi_2 + b_2$, $a_1 \cdot a_2 \neq 0$. *Тогда*

$$\rho(\eta_1, \eta_2) = \rho(\xi_1, \xi_2) \cdot sign(a_1 \cdot a_2) \ ,$$

т. е. коэффициент корреляции не меняется при линейных преобразованиях (изменении шкалы измерений).

4) Если $\rho(\xi_1, \xi_2) = +1$, *то* $\xi_2 = a \cdot \xi_1 + b$ *где* $a > 0$, $b \in R^1$.

Если $\rho(\xi_1, \xi_2) = -1$, *то* $\xi_2 = a \cdot \xi_1 + b$ *где* $a < 0$, $b \in R^1$.

В случае $\rho > 0$ говорят, что ξ_1 и ξ_2 *положительно коррелированы.* Если $\rho < 0$, то говорят, что ξ_1 и ξ_2 *отрицательно коррелированы.*

Для прояснения смысла коэффициента корреляции рассмотрим следующую задачу. Пусть мы имеем две случайные величины Y и X и хотим объяснить поведение Y с помощью X. Для этого найдем наилучшее линейное приближение

$$\hat{Y} = \alpha + \beta \cdot X \tag{2.1}$$

для случайной величины Y с помощью случайной величины X (см. Гильбертово пространство случайных величин). Ошибку аппроксимации обозначим

$$e = Y - \hat{Y} \ . \tag{2.2}$$

По лемме о перпендикуляре для поиска коэффициентов α и β можно использовать соотношения:

$$\begin{array}{rclcl} (e, 1) & := & M(e \cdot 1) & = & M[(Y - \alpha - \beta X) \cdot 1] \\ & & & = & M(Y) - \alpha - \beta M(X) = 0 \ , \\ (e, X) & := & M(e \cdot X) & = & M[(Y - \alpha - \beta X) \cdot X] \\ & & & = & M(YX) - \alpha M(X) - \beta M(X^2) = 0 \ . \end{array}$$

Решая эти уравнения, получаем

$$\begin{gathered} \beta = \frac{cov(Y, X)}{D(X)} = \rho \cdot \frac{\sigma_Y}{\sigma_X} \\ \alpha = M(Y) - \rho \cdot \frac{\sigma_Y}{\sigma_X} \cdot M(X) \ . \end{gathered} \tag{2.3}$$

Таким образом, наилучшее приближение $\hat{Y}$ имеет вид

$$\hat{Y} = M(Y) + \rho \cdot \frac{\sigma_Y}{\sigma_X}(X - M(X)) \ . \tag{2.4}$$

Если перейти к центрированным величинам $Y_c = Y -$

$M(Y)$ и $X_c = X - M(X)$, то последнее уравнение принимает вид

$$\hat{Y}_c = \rho \cdot \frac{\sigma_Y}{\sigma_X} \cdot X_c \ . \tag{2.5}$$

Заметим, что

$$\sigma_Y^2 = D(Y) = D(Y_c) = \|Y_c\|^2 \ , \ \sigma_X^2 = D(X) = D(X_c) = \|X_c\|^2 \ .$$

Тогда из соотношения (5) следует

$$D(\hat{Y}_c) = \rho^2 \cdot D(Y) \ . \tag{2.6}$$

В силу некоррелированности $\hat{Y}_c$ и e

$$D(Y_c) = D(\hat{Y}_c) + D(e) \ .$$

Тогда

$$D(e) = (1 - \rho^2) \cdot D(Y) \ . \tag{2.7}$$

Таким образом, величина ρ^2 показывает какую часть дисперсии Y можно объяснить линейным влиянием с. в. X. Аналогично, величина $(1 - \rho^2)$ показывает какую часть дисперсии (степени изменчивости) с.в. Y нельзя объяснить (линейным) влиянием с.в. X и необходимо привлекать другие факторы (переменные).

Прежде, чем искать аппроксимацию Y через X в форме (1), нужно убедиться, что X влияет на Y, т. е. $\rho \neq 0$. Для аккуратной проверки этого предположения необходимо дополнительное ограничение, что случайный вектор (Y, X) имеет двумерное нормальное распределение. Мы хотим проверить гипотезу

$$H_0: \ \rho = 0$$

против альтернативы

$$H_1: \ \rho \neq 0 \ .$$

Пусть имеем повторную выборку $(X_1, Y_1), \ldots, (X_N, Y_N)$ из двумерного нормального распределения. Для проверки гипотезы H_0 о коэффициенте корреляции ρ естественно использовать *выборочный коэффициент корреляции*

$$R := \frac{S_{YX}}{S_X \cdot S_Y} \ , \tag{2.8}$$

где

$$S_{YX} = \frac{1}{N} \sum_{j=1}^{N} (Y_j - \bar{Y})(X_j - \bar{X})$$

– выборочная ковариация,

$$S_Y^2 = \frac{1}{N} \sum_{j=1}^{N} (Y_j - \bar{Y})^2$$

– выборочная дисперсия для Y,

$$S_X^2 = \frac{1}{N} \sum_{j=1}^{N} (X_j - \bar{X})^2$$

– выборочная дисперсия для X. Удобнее перейти к величине

$$t_{N-1} = \frac{R}{\sqrt{1 - R^2}} \cdot \sqrt{N - 1} \ . \tag{2.9}$$

Отметим, что величина

$$t_{N-1}^2 = \frac{R^2}{1 - R^2} \cdot (N - 1)$$

измеряет соотношение объясненной и необъясненной частей дисперсии с. в. Y.

Теорема 2 . *Если гипотеза H_0 верна, то t_{N-1} имеет распределение Стьюдента с $N-1$ степенью свободы.*

Доказательство. Выше было показано, что имеет место представление

$$Y_j = \alpha + \beta \cdot X_j + e_j \ , \ j = 1, \ldots, N \ ,$$

причем

$$\beta = \rho \cdot \frac{\sigma_Y}{\sigma_X} \ ,$$

а случайные величины e_j независимы между собой и от X_j и имеют нормальное распределение со средним ноль и одинаковой дисперсией σ^2. Отсюда мы видим, что наша задача эквивалентна задаче проверки гипотезы о равенстве нулю коэффициента β. В следующем параграфе будет показано, что при справедливости этой гипотезы условное распределение с.в. t_{N-1} при фиксированных значениях величин X_j, $j = 1, \ldots, N$, есть распределение Стьюдента с $N-1$ степенью свободы. Но тогда и безусловное распределение будет таким же.

Далее, как обычно, для заданного α находим критическую константу $t_{N-1}(\alpha)$:

$$P(|t_{N-1}| > t_{N-1}(\alpha)) = \alpha \ .$$

Если реально полученное $|t_{N-1}|$ будет больше $t_{N-1}(\alpha)$, то гипотеза H_0 неверна и величины Y и X следует признать зависимыми. В противном случае говорят, что эксперимент не показал значимого проявления корреляции (зависимости) Y и X.

Замечание. В ППП по статистике проверка гипотез проводится несколько иным (хотя и эквивалентным) образом. Для реально полученного значения t^*_{N-1} статистики t_{N-1} находят вероятность

$$P(|t_{N-1}| > |t^*_{N-1}|) = p \ ,$$

так называемый *достигнутый уровень значимости* (p-value), и сравнивают ее с заданным уровнем значимости α (например, 0.05).

Если мы не можем гарантировать нормального распределения случайного вектора (Y, X), то можно воспользоваться тем, что при больших N с. в. t_{N-1} имеет асимптотически нормальное распределение, если X и Y независимы.

Если мы выявили, что корреляция есть, то хотелось бы выяснить величину коэффициента корреляции, например, построить для ρ доверительный интервал. Для решения этой задачи Р. Фишер предложил рассмотреть с.в.

$$Z = \frac{1}{2} \ln \frac{1+R}{1-R} \ .$$

Было доказано, что даже при небольших N эта величина имеет приближенно нормальное распределение со средним

$$\frac{1}{2} \ln \frac{1+\rho}{1-\rho} + \frac{\rho}{2(N-1)}$$

и дисперсией

$$\frac{1}{N-3} \ .$$

Это позволяет строить доверительные интервалы и проверять гипотезы для ρ.

Всюду далее мы будем использовать парные коэффициенты корреляции между различными переменными из набора $Y, X_1, \ldots, X_m$. Обозначим для единообразия $X_0 = Y$ и введем матрицу $CORR = (\rho_{ij})$, состоящую из парных коэффициентов корреляции. Ясно, что $\rho_{ii} = 1$ для $i = \overline{0, m}$.

2.2.2 Множественный коэффициент корреляции

Этот коэффициент применяется для измерения величины линейной связи одной зависимой переменной Y от нескольких объясняющих переменных $X_1, \ldots, X_m$.

Пусть мы имеем случайный вектор $(Y, X_1, \ldots, X_m)$ и , как и ранее, ищем наилучшее линейное приближение

$$\hat{Y} = \alpha + \beta_1 \cdot X_1 + \ldots + \beta_m \cdot X_m \tag{2.10}$$

с. в. Y с помощью с. в. $X_1, \ldots, X_m$.

Определение 2 . *Множественным коэффициентом корреляции с. в. Y с набором с. в. $X_1, \ldots, X_m$ называется число*

$$\rho_{Y.X_1,\ldots,X_m} := \rho(Y, \hat{Y}) \ , \tag{2.11}$$

где $\hat{Y}$ есть наилучшее линейное приближение Y с помощью с. в. $X_1, \ldots, X_m$.

Обозначим

$$e = Y - \hat{Y}$$

ошибку аппроксимации. В силу леммы о перпендикуляре $\hat{Y}$ и e некоррелированы. Тогда

$$D(Y) = D(\hat{Y}) + D(e) \ .$$

Далее

$$cov(Y, \hat{Y}) = D(\hat{Y})$$

и

$$\rho(Y, \hat{Y}) = \frac{D(\hat{Y})}{\sqrt{D(Y)D(\hat{Y})}} \ ,$$

или

$$\rho^2(Y, \hat{Y}) = \frac{D(\hat{Y})}{D(Y)} \ . \tag{2.12}$$

Таким образом, квадрат множественного коэффициента корреляции показывает какую часть дисперсии (степени изменчивости) с. в. Y можно объяснить совокупным линейным влиянием с. в. $X_1, \ldots, X_m$. Аналогично $1 - \rho^2(Y, \hat{Y})$

показывает какую часть дисперсии Y нельзя объяснить влиянием (изменением) с. в. $X_1, \ldots, X_m$.

Рассмотрим некоторые свойства множественного коэффициента корреляции.

Теорема 3 .

1) $0 \leq |\rho_{Y.X_1,\ldots,X_m}| \leq 1$.

2) $|\rho_{Y.X_1,\ldots,X_m}|$ *равен максимальному по модулю простому коэффициенту корреляции между* Y *и* $c_0 + c_1X_1 + \ldots + c_mX_m = \tilde{Y}$ *и достигается на с.в.* $\hat{Y}$, *которая является наилучшим линейным приближением.*

3) Если $\rho_{Y.X_1,\ldots,X_m} = 0$, *то* $\hat{Y} = M(Y)$.

Если $\rho_{Y.X_1,\ldots,X_m} = 1$, *то* $Y = \hat{Y}$.

Доказательство. Свойство 1) следует непосредственно из определения, т.к. таким свойством обладает простой коэффициент корреляции.

Пусть мы имеем произвольное линейное приближение $\tilde{Y} = c_0+c_1X_1+\ldots+c_mX_m$ с.в. Y с помощью с.в. $X_1, \ldots, X_m$. Если перейти к центрированным величинам (что не меняет коэффициента корреляции), то простой коэффициент корреляции между с.в. $\tilde{Y}$ и Y равен косинусу угла между этими элементами в гильбертовом пространстве L_2 случайных величин с конечными вторыми моментами. Из линейной алгебры известно, что этот косинус принимает максимальное по абсолютной величине значение на проекции элемента Y на соответствующее подпространство, т.е. на наилучшем линейном приближении.

Если множественный коэффициент корреляции равен нулю, то (после перехода к центрированным величинам), в силу свойства 2) мы получаем, что $Y - M(Y)$ ортогонален ко всем с.в. $X_1, \ldots, X_m$. Отсюда следует, что $\hat{Y} = M(Y)$.

Если множественный коэффициент корреляции равен 1, то, в силу свойства 4) теоремы 1, мы имеем $a \cdot \hat{Y} + b = Y$.

Так как $\hat{Y}$ есть проекция Y, то $a = 1$, $b = 0$.

Для множественного коэффициента корреляции можно строить доверительные интервалы и проверять гипотезы. Но соответствующие процедуры выглядят гораздо сложнее, чем для простого коэффициента корреляции. Мы рассмотрим некоторые из них позднее.

2.2.3 Частный коэффициент корреляции

Часто необходимо знать силу "чистого"влияния на зависимую переменную Y одного из факторов X_k, когда устранено влияние всех остальных факторов. Для этого используется частный коэффициент корреляции, который измеряет силу линейной связи между двумя переменными, когда устранено влияние других.

Пусть мы имеем случайный вектор $(Y, X_1, \ldots, X_m)$. Выделим некоторый фактор X_k и обозначим через C совокупность всех остальных факторов. Обозначим через Y^c и X_k^c наилучшие линейные оценки с.в. Y и X_k с помощью с.в. из набора C. Введем с.в.

$$Z_Y = Y - Y^c \ , \ Z_{X_k} = X_k - X_k^c \ .$$

Определение 3 . *Частным коэффициентом корреляции с.в. Y и X_k называется парный коэффициент корреляции с.в. Z_Y и Z_{X_k}, т.е.*

$$\rho_{YX_k \cdot C} := \rho(Z_Y, Z_{X_k}) \ .$$

Рассмотрим некоторые свойства этого коэффициента.

Теорема 4 . *1)* $|\rho_{YX_k \cdot C}| \leq 1$.

2) $1 - \rho^2_{Y \cdot X_1, \ldots, X_k} = (1 - \rho^2_{Y \cdot X_1, \ldots, X_{k-1}})(1 - \rho^2_{YX_k \cdot X_1, \ldots, X_{k-1}})$.

3) Пусть

$$\hat{Y} = \alpha + \beta_1 X_1 + \ldots + \beta_m X_m$$

есть наилучшее линейное приближение Y с помощью с.в. $X_1, \ldots, X_k$. Тогда

$$\beta_k = \rho_{YX_k \cdot C} \cdot \frac{\sigma_{Y\cdot C}}{\sigma_{X_k \cdot C}} \ ,$$

где $\sigma^2_{X_k \cdot C}$ и $\sigma^2_{Y\cdot C}$ – условные (остаточные) дисперсии с.в. X_k и Y, когда устранено влияние с.в. из набора C.

4) Если X_k и X_l некоррелированы с набором C, то $\rho_{kl\cdot C} = \rho_{kl}$.

Доказательство. Свойство 1) является очевидным.

Для доказательства свойства 2) рассмотрим линейные пространства L_{k-1} и L_k, порожденные набором с.в. $C_{k-1} = \{X_0 = 1, X_1, \ldots, X_{k-1}\}$ и $C_k = \{X_0, X_1, \ldots, X_k\}$ соответственно, а также ортогональное дополнение $L_{k-1,k}$ пространства L_{k-1} в L_k. Пусть $\hat{Y}_k$ и $\hat{Y}_{k-1}$ есть проекции элемента Y на пространства L_k и L_{k-1} соответственно, $e_k = Y - \hat{Y}_k$, $e_{k-1} = Y - \hat{Y}_{k-1}$. Тогда

$$Y = \hat{Y}_{k-1} + e_{k-1} = \hat{Y}_k + e_k = \hat{Y}_k + (e_k - e_{k-1}) + e_{k-1} \ ,$$

где все слагаемые в соответствующих суммах являются ортогональными.

Без ограничения общности все величины можно считать центрированными. Как показано выше,

$$\|e_k\|^2 = (1 - \rho^2_{Y\cdot C_k})\|Y\|^2 \ , \ \|e_{k-1}\|^2 = (1 - \rho^2_{Y\cdot C_{k-1}})\|Y\|^2 \ . \tag{2.13}$$

Предположим, что $X_k \notin L_{k-1}$, и представим его в виде

$$X_k = \hat{X}_k + X_k^{\perp} \ ,$$

где $\hat{X}_k \in L_{k-1}$, $X_k^{\perp} \in L_{k-1,k}$. Тогда по определению частного коэффициента корреляции

$$\rho_{YX_k\cdot C_{k-1}} = \rho(Y - \hat{Y}_{k-1}, X_k - \hat{X}_k) = \rho(e_{k-1}, X_k^{\perp}) \ . \tag{2.14}$$

Пространство $L_{k-1,k}$ является одномерным и может быть порождено как элементом $X_k^{\perp}$, так и элементом $e_{k-1} - e_k$. Тогда

$$\rho(e_{k-1}, X_k^{\perp}) = \rho(e_{k-1}, e_{k-1} - e_k) =$$

$$= \frac{Cov(e_{k-1}, e_{k-1} - e_k)}{\sqrt{D(e_{k-1}) \cdot D(e_{k-1} - e_k)}} = \sqrt{\frac{D(e_{k-1} - e_k)}{D(e_{k-1})}} \ .$$

В силу ортогональности

$$D(e_{k-1}) = D(e_{k-1} - e_k) + D(e_k) \ .$$

Из последних двух соотношений получаем

$$\begin{aligned} \|e_k\|^2 = D(e_k) &= (1 - \rho^2(e_{k-1}, X_k^{\perp})) \cdot D(e_{k-1}) \\ &= (1 - \rho^2(e_{k-1}, X_k^{\perp})) \cdot \|e_{k-1}\|^2 \ . \end{aligned} \quad (2.15)$$

Из соотношений (13)-(15) получаем свойство 2).

Таким образом, $\rho_{YX_k \cdot X_1,\ldots,X_{k-1}}$ показывает долю остаточной дисперсии Y, "объясненную"добавлением переменной X_k к набору $\{X_1, \ldots, X_{k-1}\}$.

Свойство 3) достаточно доказать только для последнего коэффициента β_m, так как перестановкой факторов задачу всегда можно свести к этому случаю. Далее, без ограничения общности можно считать, что все величины центрированы. Пусть L и L_C есть линейные пространства, порожденные случайными величинами $X_1, \ldots, X_m$ и $X_1, \ldots, X_{m-1}$ соответственно, $L_C^{\perp}$ – ортогональное дополнение L_C в L. Обозначим через $\hat{Y}$ и $\hat{Y}_C$ проекции элемента Y на линейные пространства L и L_C, а через $X_{m,C}$ проекцию элемента X_m на пространство L_C. Ясно, что элемент $X_m^{\perp} = X_m - X_{m,C}$ лежит в одномерном пространстве $L_C^{\perp}$, которое ортогонально L_C. Имеют место следующие ортогональные разложения:

$$Y = (Y - Y_C) + Y_C \ , \ X_m = (X_m - X_{m,C}) + X_{m,C} \ .$$

В силу сказанного выше

$$\begin{aligned}\hat{Y} &= \beta_1 X_1 + \ldots + \beta_{m-1} X_{m-1} + \beta_m X_m \\ &= \tilde{\beta}_1 X_1 + \ldots + \tilde{\beta}_{m-1} X_{m-1} + \beta_m X_m^{\perp} ,\end{aligned}$$

где коэффициент β_m один и тот же в обоих представлениях. В силу отмеченной ортогональности этот коэффициент можно найти проектируя вектор $Y - Y_C$ на вектор $X_m^{\perp} = X_m - X_{m,C}$, т.е. решая для них задачу о простой регрессии. Решение этой задачи было найдено выше при рассмотрении простого коэффициента корреляции. А именно, было показано, что

$$\beta_m = \rho(Y - Y_C, X_m - X_{m,C}) \frac{\sigma_{Y-Y_C}}{\sigma_{X_m - X_{m,C}}} = \rho_{YX_m \cdot C} \cdot \frac{\sigma_{Y \cdot C}}{\sigma_{X_m \cdot C}} .$$

Из свойства 3) следует, что

$$\beta_k = 0 \;\Leftrightarrow\; \rho_{YX_k \cdot C} = 0 .$$

Заметим, что условия $\rho_{YX_k} = 0$ недостаточно, т.к. линейное влияние X_k на Y может быть опосредованным, через некоторые третьи величины.

Свойство 4) почти очевидно. Если X_k и X_l некоррелированы с набором C, то (для случая центрированных величин) их проекции (наилучшие линейные приближения!) на линейное пространство, порожденное величинами из набора C, равны нулю. А тогда определения частного коэффициента корреляции и простого коэффициента корреляции совпадают.

В заключение кратко обсудим вопрос о более эффективном вычислении множественного и частных коэффициентов корреляции. В гильбертовом пространстве L_2 все определяется только моментами первого и второго порядка. Это означает, что частные коэффициенты корреляции можно вычислить через парные.

Назовем парные коэффициенты корреляции ρ_{kl} коэффициентами корреляции нулевого порядка. Выделим номера k, l, p и обозначим через C набор из остальных с.в. Можно показать, что

$$\rho_{kl\cdot p} = \frac{\rho_{kl} - \rho_{kp}\rho_{lp}}{\sqrt{(1-\rho_{kp}^2)(1-\rho_{lp}^2)}}$$

– коэффициенты корреляции *первого порядка*, и

$$\rho_{kl\cdot p,C} = \frac{\rho_{kl\cdot C} - \rho_{kp\cdot C}\rho_{lp\cdot C}}{\sqrt{(1-\rho_{kp\cdot C}^2)(1-\rho_{lp\cdot C}^2)}} \ .$$

Существует еще один метод вычисления множественного и частного коэффициентов корреляции через парные, используя определители матриц. Обозначим через $CORR_{ij}$ алгебраическое дополнение элемента ρ_{ij} в матрице $CORR$. Тогда

$$\rho_{Y\cdot X_1,\ldots,X_m} = 1 - \frac{CORR}{CORR_{00}} \ ,$$

и

$$\rho_{ij\cdot C} = -\frac{CORR_{ij}}{\sqrt{CORR_{ii}CORR_{jj}}} \ .$$

В реальных задачах истинные коэффициенты корреляции неизвестны и заменяются на их выборочные аналоги, которые вычисляются по набору экспериментальных данных.

2.3 Пример

В этом разделе мы рассмотрим характерный пример использования коэффициентов корреляции при анализе взаимосвязи различных показателей экономической деятельности.

Пример. По итогам года на 37 однородных предприятиях легкой промышленности регистрировались следующие показатели их работы:

$Y = X_0$ – среднемесячная характеристика качества ткани (в баллах),

X_1 – среднемесячное число профилактических наладок автоматической линии,

X_2 – среднемесячное число обрывов нити.

Были вычислены следующие выборочные парные коэффициенты корреляции:

$$R_{01} = 0.105,\ R_{02} = -0.024,\ R_{12} = -0.996.$$

Эти данные показывают слабую связь между качеством ткани и числом профилактических наладок и числом обрывов нити, что противоречит опыту работы в данной области. Но если вычислить частные коэффициенты корреляции

$$R_{01.2} = 0.907,\ R_{02.1} = -0.906,$$

то мы видим сильную корреляционную связь. Полученные выше результаты обусловлены тем, что факторы X_1 и X_2 сильно отрицательно коррелированы и их взаимное влияние маскирует вклад каждого из факторов. Общий вклад факторов является значительным, что показывает величина множественного коэффициента корреляции

$$R^2_{Y.X_1X_2} = 0.823.$$

Глава 3

Простая линейная регрессия

В этой главе подробно анализируются все основные задачи регрессионного анализа в простейшей ситуации, когда мы изучаем зависимость переменной Y от одной независимой переменной (фактора) X.

3.1 Подгонка кривой

Рассмотрим вначале одну вспомогательную задачу, решение которой будет постоянно использоваться в дальнейшем. Пусть мы имеем N совместных измерений величин X и Y: $(X_1, Y_1), \ldots, (X_N, Y_N)$. Мы хотим найти некоторую функцию $y = f(x)$ из заданного (как правило параметрического) класса $\mathcal{F}$, которая наилучшим образом аппроксимирует имеющийся набор данных. Если данных достаточно много, а число параметров невелико, то, скорее всего, ни одна из функций f из класса $\mathcal{F}$ не пройдет через все экспериментальные точки. Образуются некоторые ошибки (невязки) $e_j = Y_j - f(X_j)$.

Найти наилучшую аппроксимирующую кривую в такой ситуации означает, что необходимо в том или ином смысле *минимизировать* невязки $\{e_j\}$. Но их много — целый набор $e = (e_1, \ldots, e_N)$. Нужен какой-то интегральный критерий Q, учитывающий всю совокупность остатков e. Ниже мы рассмотрим несколько вариантов таких критериев.

1) *Сумма квадратов остатков.*

Пусть

$$Q_1(e) = \sum_{j=1}^{N} e_j^2 = \sum_{j=1}^{N} [Y_j - f(X_j)]^2. \tag{3.1}$$

Если $y = f(x, \theta_1, \ldots, \theta_m) = f(x, \theta)$, то для минимизации Q необходимо вычислить первые производные (если они существуют) и приравнять их к нулю. Это приводит нас к некоторой системе уравнений. Достоинство этого метода состоит в том, что он приводит к несложным вычислительным процедурам при оценке параметров. Как мы увидим позднее, оценки параметров обладают хорошими статистическими свойствами и удается построить статистическую теорию для проверки разнообразных гипотез. Главным недостатком метода является то, что оценки параметров очень чувствительны к "выбросам" – одна грубая ошибка, например, неверная запись в наблюдении, сильно меняет оценки параметров.

2) *Сумма модулей остатков.*

Пусть

$$Q_2(e) = \sum_{j=1}^{N} |e_j|. \tag{3.2}$$

При таком выборе критерия оценки параметров будут меньше реагировать на выбросы. Но сама процедура оценки параметров будет гораздо сложнее с аналитической точки зрения, в силу недифференцируемости Q. Решение, как правило, находится численными методами. Кроме того, решение может быть неединственным.

3) *Функция Хьюбера.*

Пытаясь объединить достоинства двух предыдущих методов, Хьюбер (Huber) предложил следующий критерий:

$$Q_3(e) = \sum_{j=1}^{N} g(e_j), \tag{3.3}$$

где

$$g(x) = \begin{cases} x^2, & \text{если } |x| \le c, \\ 2c \cdot |x| - c^2, & \text{если } |x| \ge c. \end{cases}$$

Таким образом, при небольших значениях e_j мы берём e_j^2, а при больших — $|e_j|$. Функция g подобрана так, что она является дифференцируемой.

Наиболее популярным является первый метод. Соответствующий метод оценки параметров называется *методом наименьших квадратов (МНК)*. Рассмотрим его более подробно для случая, когда класс $\mathcal{F}$ состоит из линейных функций

$$f(x) = \alpha + \beta \cdot x. \tag{3.4}$$

В этом случае выражение (1) принимает вид:

$$Q(e) = \sum_{j=1}^{N} [Y_j - (\alpha + \beta \cdot X_j)]^2. \tag{3.5}$$

Нам необходимо найти наименьшее значение функции $Q(e) = Q(e, \alpha, \beta)$ по параметрам α и β. Запишем необходимое условие экстремума (так называемое *условие первого порядка*):

$$\begin{aligned} \tfrac{\partial Q}{\partial \alpha} &= -2 \sum_{j=1}^{N} [Y_j - (\alpha + \beta \cdot X_j)] = 0, \\ \tfrac{\partial Q}{\partial \beta} &= -2 \sum_{j=1}^{N} [Y_j - (\alpha + \beta \cdot X_j)] \cdot X_j = 0. \end{aligned} \tag{3.6}$$

После несложных преобразований получаем

$$
\begin{aligned}
&N \cdot \alpha + \left(\sum_j X_j\right) \cdot \beta = \sum_j Y_j, \\
&\left(\sum_j X_j\right) \cdot \alpha + \left(\sum_j X_j^2\right) \cdot \beta = \sum_j X_j Y_j.
\end{aligned} \tag{3.7}
$$

Уравнения (7) называются *системой нормальных уравнений.*

Решения системы (7) имеют вид

$$
\hat{\beta} = \frac{N \cdot \sum_j X_j Y_j - \left(\sum_j X_j\right) \cdot \left(\sum_j Y_j\right)}{N \cdot \sum_j X_j^2 - \left(\sum_j X_j\right)^2} = \frac{S_{XY}}{S_X^2}, \tag{3.8}
$$

$$
\hat{\alpha} = \bar{Y} - \hat{\beta} \cdot \bar{X}, \tag{3.9}
$$

где

$$
\begin{aligned}
&\bar{X} = \tfrac{1}{N} \sum_j X_j, \quad \bar{Y} = \tfrac{1}{N} \sum_j Y_j, \\
&S_{XY} = \tfrac{1}{N} \sum_j (X_j - \bar{X})(Y_j - \bar{Y}), \\
&S_X^2 = \tfrac{1}{N} \sum_j (X_j - \bar{X})^2, \quad S_Y^2 = \tfrac{1}{N} \sum_j (Y_j - \bar{Y})^2.
\end{aligned} \tag{3.10}
$$

Полезно привести геометрическую интерпретацию полученных результатов. Обозначим

$$
\vec{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_N \end{pmatrix}, \quad Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_N \end{pmatrix}, \quad \mathbf{1} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}, \quad X = \begin{pmatrix} 1 & X_1 \\ \vdots & \vdots \\ 1 & X_N \end{pmatrix}.
$$

Мы ищем наилучшую линейную оценку $\hat{Y} = \hat{\alpha} \cdot \mathbf{1} + \hat{\beta} \cdot \vec{X}$ для вектора Y в линейном пространстве $L = \{c_0 \cdot \mathbf{1} + c_1 \cdot \vec{X}\}$.

По лемме о перпендикуляре нужно найти проекцию Y на L. Обозначим

$$\hat{\theta} = \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix}, \quad \hat{Y} = X \cdot \hat{\theta}, \quad e = Y - \hat{Y}.$$

Систему нормальных уравнений (7) можно записать в виде

$$\mathbf{1}^T \cdot e = 0, \quad \vec{X}^T \cdot e = 0$$

или

$$X^T \cdot e = \mathbf{0}. \tag{3.11}$$

Фактически это формулировка леммы о перпендикуляре. Отсюда получаем

$$X^T \cdot (Y - X \cdot \theta) = X^T \cdot Y - (X^T X) \cdot \theta = \mathbf{0}.$$

Если матрица $X^T X$ невырождена, то

$$\hat{\theta} = (X^T X)^{-1} \cdot X^T Y. \tag{3.12}$$

3.2 Классическая линейная модель

3.2.1 Описание модели

Основная модель, с которой мы имеем дело в эконометрике в простейшем случае имеет вид

$$Y_j = f(X_j) + \varepsilon_j, \quad j = \overline{1, N}.$$

Всюду далее в этой главе мы будем предполагать, что выполнены следующие

Основные предположения о модели

1) Модель линейна по параметрам, т. е.

$$Y_j = \alpha + \beta \cdot X_j + \varepsilon_j,$$

2) X_j измерены без ошибок, т. е. это неслучайные величины,
3) $M(\varepsilon_j) = 0$,
4) $D(\varepsilon_j) = \sigma^2$ для всех j,
5) $\mathrm{cov}(\varepsilon_j, \varepsilon_k) = 0$, $j \neq k$,
6) ε_j имеют нормальное распределение.

Отсюда, в частности следует, что случайные величины Y_j независимы и имеют нормальное распределение со средним $\alpha + \beta \cdot X_j$ и дисперсией σ^2.

В нашей модели есть три параметра α, β и σ^2, которые необходимо оценить.

3.2.2 Метод наименьших квадратов

Для оценки параметров α и β мы применим МНК, описанный в предыдущем параграфе. Таким образом, нам необходимо минимизировать выражение

$$Q = \sum_{j=1}^{N} [Y_j - (\alpha + \beta \cdot X_j)]^2.$$

Напомним, что решение этой экстремальной задачи имеет вид

$$\hat{\beta} = \frac{S_{XY}}{S_X^2}, \tag{3.13}$$

$$\hat{\alpha} = \bar{Y} - \hat{\beta} \cdot \bar{X}, \tag{3.14}$$

где $\bar{X}$, $\bar{Y}$, S_{XY}, S_X^2 вычисляются по формулам (10).

Преобразуем (13) и (14) так, чтобы получить выражения более удобные для дальнейшего исследования. Обозначим

$$w_j = \frac{X_j - \bar{X}}{\sum\limits_k (X_k - \bar{X})^2}. \tag{3.15}$$

Нетрудно проверить, что

$$\sum_j w_j = 0, \quad \sum_j w_j^2 = \frac{1}{\sum\limits_k (X_k - \bar{X})^2}. \tag{3.16}$$

Производя несложные преобразования, получаем

$$\begin{aligned}\hat{\beta} = \frac{\sum\limits_j (X_j - \bar{X})(Y_j - \bar{Y})}{\sum\limits_k (X_k - \bar{X})^2} &= \sum_j w_j \cdot (Y_j - \bar{Y}) = \sum_j w_j \cdot Y_j = \\ &= \sum_j w_j(\alpha + \beta \cdot X_j + \varepsilon_j) = \beta \sum_j w_j \cdot X_j + \sum_j w_j \cdot \varepsilon_j = \\ &= \beta \sum_j w_j (X_j - \bar{X}) + \sum_j w_j \cdot \varepsilon_j = \beta + \sum_j w_j \cdot \varepsilon_j,\end{aligned}$$

то есть

$$\hat{\beta} = \sum_j w_j \cdot Y_j = \beta + \sum_j w_j \cdot \varepsilon_j. \tag{3.17}$$

Аналогичными преобразованиями получаем (задача!)

$$\hat{\alpha} = \sum_j \left[\frac{1}{N} - w_j \cdot \bar{X}\right] \cdot Y_j = \alpha + \sum_j \left[\frac{1}{N} - w_j \cdot \bar{X}\right] \cdot \varepsilon_j. \tag{3.18}$$

В качестве оценки дисперсии σ^2 предлагается следующее выражение

$$\hat{\sigma^2} = S^2 = \frac{1}{N-2} \sum_{j=1}^{N} e_j^2 = \frac{1}{N-2} \sum_{j=1}^{N} [Y_j - \hat{Y}_j]^2. \tag{3.19}$$

3.3 Свойства оценок

В данном разделе будет показано, что полученные выше оценки обладают целым рядом хороших свойств.

Определение 1 *Оценка $\hat{\gamma}$ некоторого параметра γ, построенная по наблюдениям $\{Y_j\}$ называется линейной, если она имеет вид*

$$\hat{\gamma} = \sum_j c_j \cdot Y_j, \tag{3.20}$$

где $\{c_j\}$ — некоторые константы.

Предложение 1 *Оценки МНК $\hat{\alpha}$ и $\hat{\beta}$ параметров α и β являются линейными и несмещёнными.*

Доказательство. Оценки $\hat{\alpha}$ и $\hat{\beta}$ определяются по формулам (17) и (18). Из их записи видно, что $\hat{\alpha}$ и $\hat{\beta}$ являются линейными. Далее

$$M(\hat{\beta}) = \beta + \sum_j w_j \cdot M(\varepsilon_j) = \beta,$$

т. е. это несмещённая оценка. Доказательство для $\hat{\alpha}$ аналогично (задача!).

Предложение 2 *Случайный вектор $(\hat{\alpha}, \hat{\beta})$ имеет следующую матрицу ковариаций:*

$$\frac{\sigma^2}{\sum\limits_j (X_j - \bar{X})^2} \begin{pmatrix} X^2 & -\bar{X} \\ -\bar{X} & 1 \end{pmatrix}. \tag{3.21}$$

Доказательство. Используя формулу (17), получаем

$$D(\hat{\beta}) = \sum_j w_j^2 \cdot \sigma^2 = \frac{\sigma^2}{\sum\limits_j (X_j - \bar{X})^2}.$$

Аналогично, в силу (18)

$$D(\hat{\alpha}) = \sum_j \left(\frac{1}{N} - w_j \cdot \bar{X}\right)^2 \cdot \sigma^2 =$$

$$= \sigma^2 \cdot \left(\frac{1}{N} - 2 \cdot \frac{1}{N} \cdot \bar{X} \cdot \sum_j w_j + \bar{X}^2 \cdot \sum_j w_j^2\right) =$$

$$= \sigma^2 \cdot \frac{\sum\limits_j (X_j - \bar{X})^2 + N \cdot \bar{X}^2}{N \cdot \sum\limits_j (X_j - \bar{X})^2} = \sigma^2 \cdot \frac{\frac{1}{N}\sum\limits_j X_j^2}{\sum\limits_j (X_j - \bar{X})^2} =$$

$$= \sigma^2 \cdot \frac{\overline{X^2}}{\sum\limits_j (X_j - \bar{X})^2}.$$

Далее,

$$\operatorname{cov}(\hat{\alpha}, \hat{\beta}) = \sum_j w_j \left[\tfrac{1}{N} - w_j \cdot \bar{X}\right] \cdot \sigma^2 = \sigma^2 \cdot \bar{X} \cdot \sum_j w_j^2 =$$
$$= -\sigma^2 \cdot \frac{\bar{X}}{\sum_j (X_j - \bar{X})^2}.$$

Задача. Доказать, что матрица ковариаций для вектора $(\hat{\alpha}, \hat{\beta})$ вычисляется по следующей формуле:

$$\sigma^2 \cdot (X^T X)^{-1}. \tag{3.22}$$

Предложение 3 *Предположим, что выполнены условия*

$$\sum_j (X_j - \bar{X})^2 \to \infty, \tag{3.23}$$

$$\sum_j X_j^2 / [N \cdot \sum_j (X_j - \bar{X})^2] \to 0. \tag{3.24}$$

Тогда оценки $\hat{\alpha}$, $\hat{\beta}$ параметров α, β являются состоятельными.

Доказательство. Выше было показано, что $M(\hat{\alpha}) = \alpha$, $M(\hat{\beta}) = \beta$. Из условий (23), (24) и предложения 2 следует, что $D(\hat{\alpha}) \to 0$, $D(\hat{\beta}) \to 0$ при $N \to \infty$. Далее, применяя неравенство Чебышева, получаем сходимость по вероятности, т. е. состоятельность оценок $\hat{\alpha}$, $\hat{\beta}$.

Определение 2 *Оценка $\hat{\gamma}$ параметра γ называется оптимальной в среднем квадратическом в некотором классе оценок Γ, если для любой другой оценки $\tilde{\gamma} \in \Gamma$ имеет место неравенство*

$$M(\hat{\gamma} - \gamma)^2 \leq M(\tilde{\gamma} - \gamma)^2. \tag{3.25}$$

Теорема 1 *(Гаусса-Маркова) Пусть выполнены условия 1–5 основных ограничений. Тогда оценки МНК $\hat{\alpha}$, $\hat{\beta}$ являются оптимальными в среднем квадратическом в классе всех линейных несмещённых оценок параметров α, β.*

Доказательство. Линейность и несмещённость $\hat{\alpha}$ и $\hat{\beta}$ доказаны при доказательстве предложения 1. Докажем их оптимальность. Мы приведём доказательство только для $\hat{\beta}$. Для $\hat{\alpha}$ доказательство проводится аналогично (задача!).

Пусть

$$\tilde{\beta} = \sum_j c_j \cdot Y_j$$

есть некоторая другая линейная несмещённая оценка для β. В силу несмещённости

$$M(\tilde{\beta}) = \sum_j c_j \cdot M(Y_j) = \alpha \cdot \sum_j c_j + \beta \cdot \sum_j c_j \cdot X_j = \beta$$

для любых значений α и β. Отсюда следует

$$\sum_j c_j = 0, \quad \sum_j c_j \cdot X_j = 1. \tag{3.26}$$

В частности, условия (26) выполнены для весов $\{w_j\}$ оценки $\hat{\beta}$. Далее,

$$\hat{\beta} - \beta = \sum_j w_j \cdot \varepsilon_j,$$
$$\tilde{\beta} - \hat{\beta} = \sum_j (c_j - w_j) \cdot \varepsilon_j.$$

Отсюда легко следует, что

$$M[(\tilde{\beta} - \hat{\beta})(\hat{\beta} - \beta)] = M\left[\left(\sum_j (c_j - w_j)\varepsilon_j\right)\left(\sum_k w_k \cdot \varepsilon_k\right)\right] \stackrel{5)}{=}$$
$$= \sum_j (c_j - w_j) \cdot w_j \cdot \sigma^2 = \frac{\sigma^2}{\sum\limits_k (X_k - \bar{X})^2} \cdot \sum_j (c_j - w_j)(X_j - \bar{X}) \stackrel{(26)}{=} 0.$$

Используя последний результат, мы получаем

$$M[(\tilde{\beta} - \beta)^2] = M\Big[\big((\tilde{\beta} - \hat{\beta}) + (\hat{\beta} - \beta)\big)^2\Big] =$$
$$= M[(\tilde{\beta} - \hat{\beta})^2] + M[(\hat{\beta} - \beta)^2] \geq M[(\hat{\beta} - \beta)^2],$$

т.е. $\hat{\beta}$ является оптимальной в среднем квадратическом.

В заключение приведём следующий результат для оценки $\widehat{\sigma^2}$ дисперсии σ^2, доказательство которого будет получено позднее в более общей ситуации множественной линейной регерссии.

Предложение 4 *Пусть выполнены условия 1)–6) основных предположений. Тогда оценка $\widehat{\sigma^2}$ для дисперсии σ^2 ошибок ε_j является несмещённой и состоятельной.*

3.4 Доверительные интервалы для параметров

Для построения доверительных интервалов и проверки гипотез о параметрах недостаточно знать только первые и вторые моменты для оценок. Для вычисления вероятностей (доверительных уровней и уровней значимости) необходимо знать их распределение. Поэтому всюду далее мы предполагаем, что выполнены не только свойства 1)–5) основных предположений, но и свойство 6).

Предложение 5 *Вектор оценок* $(\hat{\alpha}, \hat{\beta})$ *параметров* (α, β) *имеет двумерное нормальное распределение со средним* (α, β) *и матрицей ковариаций*

$$\frac{\sigma^2}{\sum\limits_j (X_j - \bar{X})^2} \cdot \begin{pmatrix} \overline{X^2} & -\bar{X} \\ -\bar{X} & 1 \end{pmatrix}.$$

Доказательство. Вектор наблюдаемых значений $Y = (Y_1, \ldots, Y_N)^T$ имеет многомерное нормальное распределение. Вектор $(\hat{\alpha}, \hat{\beta})$ получен из Y линейным преобразованием и, следовательно, также имеет нормальное распределение. Утверждение о векторе средних и матрице ковариаций доказано в предложениях 1 и 2.

Предложение 6 *Векторы оценок* $(\hat{\alpha}, \hat{\beta})$ *и остатков* e *некоррелированы (независимы). В частности,* $(\hat{\alpha}, \hat{\beta})$ *и* S^2 *независимы.*

Доказательство. Так как векторы $(\hat{\alpha}, \hat{\beta})$ и e имеют нормальное распределение, то для доказательства их независимости достаточно показать, что их компоненты некоррелированы. Покажем это на примере $\hat{\beta}$ и e_k. Выше было показано, что $\hat{\beta} = \beta + \sum\limits_j w_j \cdot \varepsilon_j$. Далее,

$$\begin{aligned} e_k = Y_k - \hat{Y}_k = \alpha + \beta \cdot X_k + \varepsilon_k - \hat{\alpha} - \hat{\beta} \cdot X_k = \\ = \alpha + \beta \cdot X_k + \varepsilon_k - \bar{Y} + \hat{\beta} \cdot \bar{X} - \hat{\beta} \cdot X_k = \\ = \alpha + \beta \cdot X_k + \varepsilon_k - \alpha - \beta \cdot \bar{X} - \bar{\varepsilon} + \hat{\beta} \cdot \bar{X} - \hat{\beta} \cdot X_k = \\ = (\beta - \hat{\beta})(X_k - \bar{X}) + (\varepsilon_k - \bar{\varepsilon}) = \\ = -\left(\sum_i w_i \varepsilon_i\right)(X_k - \bar{X}) + (\varepsilon_k - \bar{\varepsilon}). \end{aligned}$$

Отсюда следует, что $M(e_k) = 0$ и

$$\begin{aligned} \mathrm{cov}(\hat{\beta}, e_k) = -\sum_i \sum_j w_i w_j \cdot \mathrm{cov}(\varepsilon_i, \varepsilon_j) \cdot (X_k - \bar{X}) \\ + \sum_j w_j \cdot \mathrm{cov}(\varepsilon_j, \varepsilon_k - \bar{\varepsilon}) = \\ = -\sum_j w_j^2 \cdot \sigma^2 \cdot (X_k - \bar{X}) - \tfrac{1}{N} \sum_j w_j \cdot \sigma^2 = \\ = -w_k \cdot \sigma^2 - 0 + w_k \cdot \sigma^2 = 0. \end{aligned}$$

Следующий результат доказывается несколько сложнее. Так как его доказательство в двумерном случае никак не упрощается, то мы рассмотрим его позднее для случая общей многомерной регрессии.

Предложение 7 *Случайная величина* $(N-2)S^2/\sigma^2 = \sum\limits_j e_j^2/\sigma^2$ *имеет* χ^2*-распределение с* $(N-2)$ *степенями свободы.*

Теперь мы имеем все необходимое для построения доверительных интервалов для параметров α и β. Оценка $\hat{\beta}$ имеет нормальное распределение со средним β и дисперсией $\sigma^2 \left(\sum\limits_j (X_j - \bar{X})^2\right)^{-1}$. В сочетании с предложениями 5 и 6 это даёт нам то, что случайная величина

$$T_\beta = \frac{\hat{\beta} - \beta}{\sigma} \cdot \sqrt{\sum_j (X_j - \bar{X})^2} / (S/\sigma) = \frac{\hat{\beta} - \beta}{S_\beta}$$

имеет распределение Стьюдента с $(N-2)$ степенями свободы. Здесь

$$S_\beta^2 = \frac{S^2}{\sum\limits_j (X_j - \bar{X})^2} = \frac{1}{N} \cdot \frac{S^2}{S_X^2}. \tag{3.27}$$

Аналогично можно показать, что случайная величина

$$\frac{\hat{\alpha} - \alpha}{S_\alpha},$$

где

$$S_\alpha^2 = \frac{1}{N} \cdot \frac{S^2}{S_X^2} \cdot \overline{X^2}, \tag{3.28}$$

имеет распределение Стьюдента с $(N-2)$ степенями свободы.

Величины S_β и S_α называются *стандартными ошибками (SE)* оценки параметров β и α соответственно.

Действуя далее стандартным образом, мы получаем следующий результат.

Теорема 2 *Доверительные интервалы уровня γ для параметров α и β имеют вид*

$$\hat{\alpha} - t_{N-2}(\gamma) \cdot S_\alpha < \alpha < \hat{\alpha} + t_{N-2}(\gamma) \cdot S_\alpha, \tag{3.29}$$

$$\hat{\beta} - t_{N-2}(\gamma) \cdot S_\beta < \beta < \hat{\beta} + t_{N-2}(\gamma) \cdot S_\beta, \tag{3.30}$$

где константы $t_{N-2}(\gamma)$ находятся из таблиц распределения Стьюдента с $(N-2)$ степенями свободы.

3.5 Проверка гипотез о параметрах

Полученные выше результаты позволяют нам построить критерии для проверки гипотез о параметрах α и β. Типичная гипотеза, с которой мы встречаемся, имеет вид

$$H_0: \quad \beta = \beta_0$$

против альтернативы

$$H_1: \quad \beta \neq \beta_0.$$

Критерий для проверки такой гипотезы строится на основе статистики

$$T_\beta = \frac{\hat{\beta} - \beta_0}{S_\beta},$$

которая при верной гипотезе H_0 имеет распределение Стьюдента с $(N-2)$ степенями свободы. Критическая зона соответствующего критерия имеет вид:

$$|T_\beta| > t_{N-2}(\alpha), \tag{3.31}$$

где константа $t_{N-2}(\alpha)$ находится из таблиц распределения Стьюдента по заданному уровню значимости α.

Критерий для проверки гипотез о параметре α строится аналогично.

Наиболее часто мы встречаемся с гипотезой $H_0: \quad \beta = 0$. Это означает, что мы проверяем гипотезу о том, влияет ли фактор X на изучаемую величину Y. Поэтому в этом случае говорят, что мы проверяем *значимость влияния* фактора X на величину Y. Если мы не подтвердили значимого влияния фактора X, то его можно исключить из модели.

3.6 Проверка адекватности модели

После того, как мы построили нашу линейную модель, необходимо убедиться в том, насколько хорошо она описывает поведение изучаемой нами величины Y. В частности, насколько хорошо мы объясняем изменение Y с помощью линейного влияния X.

Для этого рассмотрим следующее разложение:

$$\sum_j (Y_j - \bar{Y})^2 = \sum_j (\hat{Y}_j - \bar{Y})^2 + \sum_j (Y_j - \hat{Y}_j)^2. \qquad (3.32)$$

Задача. Используя факт о том, что вектор $\hat{Y}$ ортогонален к пространству L порожденному векторами **1** и $\vec{X}$, доказать соотношение (32).

Введём следующие обозначения:

$$TSS = \sum_j (Y_j - \bar{Y})^2$$

— *полная* сумма квадратов (total sum of squares),

$$RSS = \sum_j (\hat{Y}_j - \bar{Y})^2$$

— *объяснённая* сумма квадратов (regression sum of squares),

$$ESS = \sum_j (Y_j - \hat{Y}_j)^2 = \sum_j e_j^2$$

— *остаточная* сумма квадратов (error sum of squares).

Интуитивно понятно, что модель следует признать хорошей, если величина

$$R^2 = \frac{RSS}{TSS} \qquad (3.33)$$

близка к 1.

Определение 3 *Величина R^2 называется* **коэффициентом детерминации**.

Она показывает долю *объяснённой* дисперсии величины Y за счёт линейного влияния фактора X.

Задача. Показать, что R совпадает с выборочным коэффициентом корреляции, т. е.

$$R = \frac{S_{XY}}{S_X \cdot S_Y}.$$

Чем ближе R^2 к 1, тем лучше модель. Но мы уже знаем из курса математической статистики, что такие понятия как близко или далеко, много или мало имеют относительный смысл, только по сравнению с чем-то. Величина ESS оценивает уровень случайных ошибок. Модель будет хороша, если то, что она объясняет, достаточно велико на фоне случайных ошибок.

Рассмотрим вновь гипотезу

$$H_0 : \quad \beta = 0$$

против альтернативы

$$H_1 : \quad \beta \neq 0.$$

Теперь эта гипотеза означает, что модель не объясняет поведения Y на фоне влияния других неучтённых факторов.

Предложение 8 *Если верна гипотеза H_0 (и конечно основные предположения 1)–6)), то*

1. *случайные величины RSS и ESS независимы;*
2. *случайные величины RSS и ESS имеют χ^2-распределения с одной и $(N-2)$ степенями свободы соответственно;*
3. *случайная величина*

$$F = \frac{RSS/(2-1)}{ESS/(N-2)} = (N-2) \cdot \frac{R^2}{1-R^2} \tag{3.34}$$

имеет распределение Фишера с $(1, N-2)$ степенями свободы.

В следующей главе этот результат будет доказан в более общей ситуации множественной линейной регрессии.

Далее по таблицам распределения Фишера для заданного α находим константу $F(\alpha)$ такую, что

$$P_0(F > F(\alpha)) = \alpha.$$

Если реально полученное значение $F_{\text{Н}}$ статистики F будет больше $F(\alpha)$, то гипотезу H_0 следует отвергнуть и признать линейное влияние X на Y значимым. В противном случае говорят, что не выявлено значимого влияния X на Y.

Замечание. 1. Можно показать, что в рассматриваемом случае (один фактор X) имеет место соотношение $F = T_\beta^2$, т. е. рассмотренный ранее критерий о значимости влияния отдельного фактора (t-критерий Стьюдента) и проверка адекватности модели в целом эквивалентны.

2. Формула (34) показывает также, что приведённый выше критерий эквивалентен критерию проверки гипотезы $\rho(X, Y) = 0$.

3.7 Итоговый пример

В данном разделе мы на некотором модельном примере показываем весь цикл статистического исследования в рамках одномерной модели регрессии.

По данным годовых отчётов десяти машиностроительных предприятий провести регрессионный анализ зависимости производительности труда y (тыс. руб. на чел.) от объёма производства x (млн. руб.).

Y_j	2.1	2.8	3.2	4.5	4.8	4.9	5.5	6.5	12.1	15.1
X_j	3	4	5	5	5	5	6	7	15	20

Предполагаем, что имеет место линейная модель (на основе визуального исследования)

$$Y_j = \alpha + \beta \cdot X_j + \varepsilon_j.$$

По методу наименьших квадратов оценки $\hat{\theta} = (\hat{\alpha}, \hat{\beta})^T$ получаются по формуле

$$\hat{\theta} = (X^T X)^{-1} X^T Y.$$

В нашем случае

$$X^T X = \begin{pmatrix} 10 & 75 \\ 75 & 835 \end{pmatrix}.$$

Обратная матрица равна

$$(X^T X)^{-1} = \begin{pmatrix} 0.306422 & -0.0275229 \\ -0.0275229 & 0.0036697 \end{pmatrix}.$$

Далее

$$X^T Y = \begin{pmatrix} 61.5 \\ 666.5 \end{pmatrix}.$$

Окончательно получаем

$$\hat{\theta} = \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} = \begin{pmatrix} 0.500917 \\ 0.753211 \end{pmatrix}.$$

Таким образом линия регрессии имеет вид

$$\hat{y} = 0.50092 + 0.75321 \cdot x.$$

В нашей задаче есть ещё один параметр — дисперсия σ^2 ошибок измерений ε_j. Её оценка находится по формуле

$$S^2 = \frac{1}{N-2} \sum_{j=1}^{N} [Y_j - \hat{Y}_j]^2 = 0.49809.$$

Отсюда оценка среднеквадратического отклонения равна

$$S = 0.70576.$$

Прежде, чем двигаться дальше, оценим значимость линейного влияния фактора X на исследуемый показатель Y.

$$\begin{aligned} TSS &= \sum_j (Y_j - \bar{Y})^2 = RSS + ESS \\ &= \sum_j (\hat{Y}_j - \bar{Y})^2 + \sum_j (Y_j - \hat{Y}_j)^2, \\ 158.485 &= 154.597 + 3.888. \end{aligned}$$

Наблюдаемое значение F_H вычисляется по формуле

$$F_H = \frac{1 \cdot RSS}{\frac{1}{N-2} ESS} = \frac{154.957}{0.486} = 318.06.$$

При гипотезе $H_0 : \quad \beta = 0$ вероятность получить такое уклонение равна 0.00000... (см. таблицы). Таким образом мы приходим к выводу, что линейное влияние фактора X следует признать значимым. Коэффициент детерминации

$$R^2 = 0.9755,$$

т. е. более чем на 97% поведение Y можно объяснить линейным влиянием X.

Оценим значимость каждого из коэффициентов регрессии. Стандартные ошибки определения коэффициентов α и β равны соответственно 0.3859 и 0.0422. Тогда величина статистик Стьюдента и вероятности полученных уклонений равны 1.29797 и 17.8343; 0.23047 и 0.00000.... Таким образом, коэффициент β значимо отличается от нуля, напротив, коэффициент α может равняться и 0. Доверительные интервалы уровня 0.95 для α и β имеют вид

$$\begin{aligned} &-0.389 < \alpha < 1.391, \\ &0.656 < \beta < 0.851. \end{aligned}$$

Обычно результаты оценки линии регрессии записывают в виде:

$$\begin{array}{ccccc} y & = & 0.50092 & + & 0.75321 & \cdot x \\ SE & & 0.3859 & & 0.0422 & \\ t & & 1.298 & & 17.83 & \\ \alpha & & 0.23 & & 0.000 & \end{array}$$

Глава 4

Множественная линейная регрессия

В этой главе рассматривается ситуация, когда есть одна количественная характеристика Y, поведение которой мы изучаем, и несколько факторов $X_1, \ldots, X_m$, с помощью которых мы пытаемся объяснить поведение Y.

4.1 Классическая линейная модель.

Мы начинаем с описания классичесской линейной модели и далее даем решение основных задач, в рамках этой модели. Изложение во многом повторяет то, что мы уже знаем из предыдущей главы, где рассматривалась та же задача, но был только один объясняющий фактор X.

Пусть мы имеем одну зависимую (объясняемую в рамках нашей модели) переменную Y и m независимых объясняющих переменных $X_1, \ldots, X_m$, которые мы будем называть *факторами* или *предикторами*. Предположим, что мы имеем N одновременных измерений всех этих величин и они связаны соотношением

$$Y_j = g(X_{j1}, \ldots, X_{jm}) + \varepsilon_j \ , j = 1, \ldots, N \ , \qquad (4.1)$$

где ε_j описывает влияние всех остальных неучтенных факторов, интерпретируется как возмущающий член или ошибка измерений и является *случайной величиной*.

Всюду далее при анализе модели мы предполагаем, что выполнены следующие *основные предположения*:

1) модель линейна *по параметрам*, т.е.

$$Y_j = \alpha + \beta_1 X_{j1} + \ldots + \beta_m X_{jm} + \varepsilon_j \ , \qquad (4.2)$$

2) факторы измерены без ошибок, т.е. $\{X_{jk}\}$ – неслучайные величины,

3) $M(\varepsilon_j) = 0$ для всех j, т.е. ошибки не содержат систематической составляющей,

4) $D(\varepsilon_j) = \sigma_j^2 = \sigma^2$ для всех j (*условие гомоскедастичности*),

5) ошибки ε_j и ε_k независимы (достаточно некоррелированности) для $j \neq k$,

6) ошибки имеют нормальное распределение.

Формулы, которые мы будем выписывать, довольно громоздки. Поэтому всюду далее мы используем следующие матричные обозначения:

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \ldots \\ Y_N \end{pmatrix}, \ \vec{X}_k = \begin{pmatrix} X_{1k} \\ X_{2k} \\ \ldots \\ X_{Nk} \end{pmatrix}, \ \theta = \begin{pmatrix} \alpha \\ \beta_1 \\ \ldots \\ \beta_m \end{pmatrix} = \begin{pmatrix} \theta_0 \\ \theta_1 \\ \ldots \\ \theta_m \end{pmatrix},$$

$$\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \ldots \\ \varepsilon_N \end{pmatrix}, \vec{X}_0 = \begin{pmatrix} 1 \\ 1 \\ \ldots \\ 1 \end{pmatrix},$$

$$X = (X_{jk}) = \begin{pmatrix} X_{10} & X_{11} & \dots & X_{1k} & \dots & X_{1m} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ X_{j0} & X_{j1} & \dots & X_{jk} & \dots & X_{jm} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ X_{N0} & X_{N1} & \dots & X_{Nk} & \dots & X_{Nm} \end{pmatrix} .$$

В этих обозначениях соотношение (2) можно переписать в более компактном виде:

$$Y = X \cdot \theta + \varepsilon \ , \tag{4.3}$$

или

$$Y = \theta_0 \cdot \vec{X}_0 + \ldots + \theta_m \cdot \vec{X}_m \ .$$

Используя основные предположения о нашей модели, нетрудно вычислить математическое ожидание и матрицу ковариаций для вектора наблюдений Y:

$$M(Y) = X \cdot \theta \ , \ \Sigma_Y = \Sigma_\varepsilon = \sigma^2 \cdot E \ . \tag{4.4}$$

Перечислим *основные задачи*, которые необходимо решить, при исследовании описанной выше модели:

1) оценка параметров модели,
2) проверка гипотез о параметрах,
3) отбор значимых факторов,
4) проверка адекватности модели,
5) проверка выполнения основных предположений.

4.2 Оценка параметров. Метод наименьших квадратов

Нам необходимо так подобрать параметры в уравнении (2), чтобы построенная модель в определенном смысле наилучшим образом аппроксимировала экспериментальные данные. Мы уже обсуждали эту проблему при анализе простой линейной регрессии и выяснили, что *метод наименьших квадратов* (МНК) является довольно простой в аналитическом смысле процедурой, он дает оценки параметров с хорошими статистическими свойствами, для него существует развитая теория проверки разнообразных гипотез. В силу этого мы будем использовать его и в случае множественной линейной регрессии.

Согласно МНК для оценки параметров необходимо решить следующую экстремальную задачу:

$$\begin{aligned} Q(\theta) &= \sum_{j=1}^{N} [Y_j - (\theta_0 + \theta_1 X_{j1} + \ldots + \theta_m X_{jm})]^2 \\ &= (Y - X\theta)^T \cdot (Y - X\theta) = \|Y - X\theta\|^2 \to \min_\theta \end{aligned} \tag{4.5}$$

$Q(\theta)$ есть неотрицательная квадратическая форма по θ. Поэтому решение задачи существует и, в невырожденном случае, является единственным. Чтобы его найти, продифференцируем $Q(\theta)$ по параметрам и приравняем эти производные к нулю. Это приводит нас к следующей линейной системе уравнений:

$$X^T X\theta = X^T Y \tag{4.6}$$

которая называется *системой нормальных уравнений.*
Пусть матрица $X^T X$ *невырождена* (эквивалентно матрица X имеет ранг $m+1$). Это означает, что векторы $\vec{X}_0, \vec{X}_1, \ldots \vec{X}_m$ линейно независимы. Тогда решение системы (6) существует и единственно и его можно записать в виде:

$$\hat{\theta} = (X^T X)^{-1} X^T Y = \theta + (X^T X)^{-1} \varepsilon \ . \tag{4.7}$$

Говорят, что $\hat{\theta}$ являются *оценками по МНК* для параметров θ уравнения регрессии. Используя эти оценки, мы вычисляем *предсказанные значения* $\hat{Y} = X \cdot \hat{\theta}$ и *остатки* $e = Y - \hat{Y}$.

Мы оценили параметры уравнения регрессии. Нам нужно оценить еще один параметр модели, а именно, дисперсию σ^2 ошибок измерений. Если бы мы знали значения ε_j, $j = 1, \ldots, N$, ошибок измерений, то величина

$$\frac{1}{N}\sum_{j=1}^{N}\varepsilon_j^2$$

была бы несмещенной и, в силу закона больших чисел, состоятельной оценкой для σ^2. К сожалению, ошибки измерений являются ненаблюдаемыми величинами и, поэтому, не могут быть использованы для построения оценок. Но, как мы неоднократно убедимся в дальнейшем, остатки e_j, $j = 1, \ldots, N$, во многом наследуют свойства ошибок измерений. Поэтому, в качестве оценки дисперсии σ^2 ошибок измерений предлагается взять величину

$$\hat{\sigma^2} = S^2 = \frac{1}{N-(m+1)}\sum_{j=1}^{N} e_j^2\ . \tag{4.8}$$

Изменение нормировки произведено для того, чтобы получить несмещенную оценку (сравни с оценкой дисперсии с помощью исправленной выборочной дисперсии, построенной по повторной выборке!).

Заметим, что фактически мы решаем задачу о наилучшей линейной оценке $\hat{Y}$ вектора Y в линейном пространстве L, порожденном векторами $\vec{X}_0$, $\vec{X}_1, \ldots, \vec{X}_m$. По лемме о перпендикуляре $e^T \cdot \vec{X}_k = 0$, $k = 0, 1, \ldots, m$. В частности, при $k = 0$ получаем $\sum e_j \cdot 1 = 0$. Кроме того, $e^T \cdot \hat{Y} = (e, \hat{Y}) = \sum_{j=1}^{N} e_j \cdot \hat{Y}_j = 0$.

4.3 Свойства оценок параметров

В этом разделе исследуются те свойства оценок параметров модели, которые не зависят от предположения о виде распределения ошибок измерений. Начнем с некоторого вспомогательного результата, доказательство которого предлагается в виде задачи.

Лемма 1 . *Пусть Y есть некоторый случайный вектор с конечными моментами второго порядка, а A – неслучайная матрица. Образуем новый случайный вектор $Z = A \cdot Y$. Тогда*

$$M(Z) = A \cdot M(Y) \ , \ \Sigma_Z = A \cdot \Sigma_Y \cdot A^T \ .$$

Пусть мы имеем некоторый вектор Y измерений и на его основе строим оценку $\hat{\gamma}$ вектора параметров γ.

Определение 4 . *Оценка $\hat{\gamma}$ параметра γ называется линейной (по Y), если она имеет вид $\hat{\gamma} = A \cdot Y$, где A – некоторая неслучайная матрица.*

Лемма 2 . *Пусть выполнены условия 1)-5) основных предположений. Линейная оценка $\tilde{\theta} = A \cdot Y$ вектора параметров θ является несмещенной тогда и только тогда, когда выполнено условие*

$$A \cdot X = E \ . \tag{4.9}$$

Доказательство. В силу леммы 1, определения несмещенности и, используя основные предположения о модели, имеем

$$M(\tilde{\theta}) = AM(Y) = AX\theta + AM(\varepsilon) = AX\theta = \theta$$

для всех θ. Отсюда следует соотношение (9).

Предложение 1 . *Пусть выполнены условия 1)-5) основных предположений, тогда оценка по МНК $\hat{\theta}$ для вектора параметров θ является линейной и несмещенной.*

Доказательство. Из соотношения (7) мы видим, что оценка по МНК является линейной с матрицей $A = (X^T X)^{-1} X^T$. Проверим условие несмещенности (9):

$$AX = (X^T X)^{-1} X^T X = E \ .$$

Предложение 2 . *Пусть выполнены условия 1)-5) основных предположений. Тогда матрица ковариаций оценки по МНК $\hat{\theta}$ для параметров θ имеет вид*

$$\Sigma_{\hat{\theta}} = \sigma^2 \cdot (X^T X)^{-1} \ . \tag{4.10}$$

Доказательство. В силу леммы 1 и используя основные предположения о модели, получаем

$$\Sigma_{\hat{\theta}} = (X^T X)^{-1} X^T \Sigma_Y X (X^T X)^{-1} =$$
$$(X^T X)^{-1} X^T \cdot \sigma^2 E \cdot X (X^T X)^{-1} = \sigma^2 (X^T X)^{-1} \ .$$

Предложение 3 . *Пусть выполнены условия 1)-5) основных предположений о модели и, кроме того, диагональные элементы матрицы $(X^T X)^{-1}$ стремятся к нулю при $N \to \infty$. Тогда оценка по МНК $\hat{\theta}$ для параметров θ уравнения модели является состоятельной.*

Доказательство легко следует из несмещенности оценки, предложения 2 и неравенства Чебышева.

Основным результатом этого раздела является следующая

Теорема 1 *(Гаусса-Маркова). Пусть выполнены условия 1)-5) основных предположений о модели. Тогда оценка по МНК $\hat{\theta}$ является оптимальной в среднем квадратическом оценкой в классе всех линейных несмещенных оценок параметров θ. (Best Linear Unbiased Estimator = BLUE).*

Доказательство. Выше было доказано, что $\hat{\theta} = AY$ является несмещенной оценкой. Пусть $\tilde{\theta} = BY = (A + C)Y$ есть некоторая другая несмещенная оценка. Из условия несмещенности (9) легко следует, что $CX = 0$. Далее, используя лемму 1, получаем

$$\Sigma_{\tilde{\theta}} = (A + C)\sigma^2 E(A + C)^T = \sigma^2(A + C)(A^T + C^T) =$$

$$\sigma^2(AA^T + CC^T + CA^T + AC^T) \ .$$

В силу предложения 2 $\sigma^2 AA^T = \sigma^2(X^T X)^{-1} = \Sigma_{\hat{\theta}}$ и

$$CA^T = (AC^T)^T = CX(X^T X)^{-1} = 0 \ .$$

Отсюда следует, что

$$\Sigma_{\tilde{\theta}} - \Sigma_{\hat{\theta}} = \sigma^2 CC^T \ .$$

Матрица CC^T является неотрицательно определенной. Из линейной алгебры известно, что в этом случае диагональные элементы матрицы $\Sigma_{\tilde{\theta}}$ не меньше соответствующих диагональных элементов матрицы $\Sigma_{\hat{\theta}}$.

В заключение исследуем свойства оценки дисперсии ошибок измерений.

Предложение 4 *. Пусть выполнены условия 1)-6) основных предположений. Тогда оценка по МНК*

$$S^2 = \frac{1}{N - (m + 1)} \sum_{j=1}^{N} e_j^2$$

является несмещенной и состоятельной оценкой для дисперсии σ^2 ошибок измерений.

Доказательство. Как отмечалось выше, мы имеем ортогональное разложение вектора Y в пространстве R^N:

$$Y = \hat{Y} + e \ ,$$

где $\hat{Y} \in L$, $e \perp L$. Подпространство L имеет размерность $m+1$, если матрица X имеет ранг $m+1$, а ортогональное дополнение к нему имеет размерность $N-(m+1)$. Так как случайный вектор Y имеет невырожденное распределение в R^N, то его проекции на любые ненулевые подпространства невырождены. Далее, случайный вектор Y имеет многомерное нормальное распределение со средним $X \cdot \theta \in L$ и матрицей ковариаций $\Sigma_Y = \Sigma_\varepsilon = \sigma^2 \cdot I_N$. Тогда его ортогональная проекция e на подпространство размерности $N-(m+1)$ имеет многомерное нормальное распределение с нулевым средним и матрицей ковариаций $\Sigma_e = \sigma^2 \cdot I_{N-(m+1)}$ в этом подпространстве. Далее

$$\sum_{j=1}^{N} e_j^2 = \sum_{k=1}^{N-(m+1)} \eta_k^2 \ ,$$

где $\eta = (\eta_1, \ldots, \eta_{N-(m+1)})^T$ есть представление случайного вектора e через координаты относительно некоторого ортонормального базиса в $L^\perp$. Тогда

$$E(S^2) = \frac{1}{N-(m+1)} E\left(\sum_{j=1}^{N} e_j^2\right)$$

$$= \frac{1}{N-(m+1)} E\left(\sum_{j=1}^{N-(m+1)} \eta_k^2\right) =$$

$$= \frac{(N-(m+1)\cdot\sigma^2}{N-(m+1)} = \sigma^2 \ ,$$

т.е. оценка S^2 для величины σ^2 является несмещенной.

Далее, так как случайные величины $\eta_1, \ldots, \eta_{N-(m+1)}$ независимы, одинаково распределены и $E(\eta_k^2) = \sigma^2 < \infty$ для всех k, то в силу закона больших чисел

$$S^2 = \frac{1}{N-(m+1)} \sum_{j=1}^{N-(m+1)} \eta_k^2 \xrightarrow{P} \sigma^2 \ , \ N \to \infty \ .$$

4.4 Построение доверительных интервалов

Выше мы исследовали такие свойства оценок параметров, которые опирались только на свойства 1)-5) основных ограничений и не требовали информации о распределении ошибок измерений ε. Дальнейший анализ свойств построенной модели требует предположения 6), что ошибки измерений имеют нормальное распределение. Всюду далее в этом параграфе мы предполагаем, что выполнены условия 1)-6) основных предположений без специального упоминания об этом.

Основой для всех дальнейших результатов является следующая

Теорема 2 . *Пусть $\hat{\theta}$ есть оценка по МНК для вектора параметров уравнения модели. Тогда имеют место следующие свойства:*

1) $\hat{\theta}$ имеет многомерное нормальное распределение со средним θ и матрицей ковариаций $\sigma^2 \cdot (X^T X)^{-1}$,

2) случайная величина $(\hat{\theta} - \theta)^T X^T X(\hat{\theta} - \theta)/\sigma^2$ имеет χ^2-распределение с $(m + 1)$ степенью свободы,

3) случайная величина $\sum\limits_j (Y_j - \hat{Y}_j)^2/\sigma^2 = \sum\limits_j e_j^2/\sigma^2$ имеет χ^2-распределение с $(N - m - 1)$ степенями свободы,

4) случайные величины S^2 и $\hat{\theta}$ (или $\hat{Y}$) независимы.

Доказательство. Случайный вектор Y имеет многомерное нормальное распределение со средним $X\theta$ и матрицей ковариаций $\Sigma_Y = \sigma^2 I_N$. Оценка $\hat{\theta}$ параметра θ по МНК имеет вид

$$\hat{\theta} = (X^T X)^{-1} X^T Y = AY \ ,$$

т.е. является линейным преобразованием вектора Y. Тогда в силу теоремы B1 случайный вектор $\hat{\theta}$ имеет многомерное нормальное распределение со средним

$$E(\hat{\theta}) = A \cdot E(Y) = (X^T X)^{-1} X^T (X\theta) = \theta$$

и матрицей ковариаций

$$\sigma_{\hat{\theta}} = A\Sigma_Y A^T = (X^T X)^{-1} X^T \cdot \sigma^2 I_N \cdot X (X^T X)^{-1} = \sigma^2 (X^T X)^{-1} .$$

Первое утверждение доказано.

В силу теоремы B4 и первого утверждения теоремы случайная величина

$$(\hat{\theta} - \theta)^T \Sigma_{\hat{\theta}}^{-1} (\hat{\theta} - \theta) = (\hat{\theta} - \theta)^T (X^T X)(\hat{\theta} - \theta)/\sigma^2$$

имеет χ^2-распределение с $m+1$ степенями свободы. Второе утверждение доказано.

Как было показано выше (смотри доказательство предложения 4), случайный вектор e есть ортогональная проекция вектора Y на подпространство размерности $N - (m+1)$ и, следовательно, имеет в нем невырожденное многомерное нормальное распределение с нулевым средним и матрицей ковариаций $\sigma^2 I_{N-(m+1)}$. В координатах относительно ортонормированного базиса этого подпространства его можно записать в виде $\eta = (\eta_1, \ldots, \eta_{N-(m+1)})^T$, где все η_k независимы, имеют нормальное распределение и $E(\eta_k) = 0$, $D(\eta_k) = \sigma^2$. Далее,

$$S^2/\sigma^2 = \sum_{j=1}^{N} e_j^2/\sigma^2 = \sum_{k=1}^{N-(m+1)} \eta_k^2/\sigma^2 \ ,$$

где все случайные величины η_k/σ независимы и имеют стандартное нормальное распределение. Третье утверждение доказано.

Случайный вектор Y имеет ортогональное разложение $Y = \hat{Y} + e$. Отсюда следует, что случайные векторы $\hat{Y}$ и e независимы. Так как случайный вектор $\hat{\theta}$ представляет собой координаты вектора $\hat{Y}$ в некотором базисе соответствующего пространства, т.е. является функцией от $\hat{Y}$, то случайные векторы $\hat{\theta}$ и e также независимы. Отсюда следует, что случайный вектор $\hat{\theta}$ и случайная величина S^2 (как функция от e!) независимы. Четвертое утверждение и вся теорема в целом доказаны.

Обозначим $B = (b_{ik}) = (X^T X)^{-1}$. Тогда оценка $\hat{\theta}_k$ параметра θ_k имеет нормальное распределение со средним θ_k и дисперсией $\sigma^2 \cdot b_{kk}$. В силу сформулированной выше теоремы случайная величина

$$T_k = \frac{\hat{\theta}_k - \theta_k}{\sigma\sqrt{b_{kk}}} : \frac{S}{\sigma} = \frac{\hat{\theta}_k - \theta_k}{S\sqrt{b_{kk}}} == \frac{\hat{\theta}_k - \theta_k}{S_k}$$

имеет распределение Стьдента с $N-m-1$ степенями свободы. Величина S_k называется *стандартной ошибкой* оценки параметра θ_k. Далее для заданного *уровня доверия* γ по таблицам распределения Стьюдента мы находим такую константу $t(\gamma)$, для которой

$$P(|T_k| < t(\gamma)) = \gamma \ .$$

Тогда доверительный интервал для параметра γ строиться по правилу:

$$\hat{\theta} - S_k \cdot t(\gamma) < \theta_k < \hat{\theta} + S_k \cdot t(\gamma) \ .$$

4.5 Проверка значимости влияния отдельного фактора

Полученные в предыдущем параграфе результаты позволяют решить еще одну важную с практической точки зрения задачу, а именно, проверить гипотезу о том, оказывает ли некоторый фактор X_k, включенный в нашу модель, значимое влияние на поведение изучаемой величины Y. Формально отсутствие такого влияния означает, что этот фактор на самом деле отсутствует, т.е. $\theta_k = 0$, и его можно исключить из модели.

Рассмотрим вначале задачу критерии проверки гипотезы

$$H_0 : \ \theta_k = \theta_{k0}$$

против альтернативы

$$H_1 : \ \theta_k \neq \theta_{k0} \ .$$

Как отмечалось выше, при верной гипотезе H_0 статистика

$$T_k = \frac{\hat{\theta}_k - \theta_{k0}}{S_k}$$

имеет распределение Стьюдента с $N - m - 1$ степенями свободы. Для заданного уровня значимости α по таблицам находим критическую константу $t(\alpha)$, для которой

$$P(|T_k| > t(\alpha)) = \alpha \ .$$

Тогда правило проверки сформулированной выше гипотезы имеет вид:

1) если реально полученное значение статистики T_k будет больше критической константы $t(\alpha)$, то мы отвергаем гипотезу H_0 и принимаем H_1,

2) в противном случае мы говорим, что гипотеза H_0 не противоречит экспериментальным данным.

Важным частным случаем является гипотеза

$$H_0 : \ \theta_k = 0$$

против альтернативы

$$H_1 : \ \theta_k \neq 0 \ .$$

В этом случае мы проверяем *значимость влияния фактора* X_k на исследуемую величину Y. Иногда говорят, что мы проверяем гипотезу о *значимости коэффициента* θ_k.

Если какой-либо коэффициент (фактор) признан незначимым, то его удаляют из модели и заново оценивают все коэффициенты. Это связано с тем, что при удалении или добавлении одного или нескольких факторов оценки коэффициентов при других факторах, в общем случае, меняются.

Построенный нами критерий известен в статистической литературе под названием *T-критерия Стьюдента.*

4.6 Проверка линейных гипотез

Построенный выше T-критерий Стьюдента позволяет установить незначимость влияния отдельно взятого фактора. Но если с его помощью мы выяснили незначимость двух и более факторов, то это вовсе не означает, что все они могут быть удалены из модели. Их совместное влияние может оказаться и значимым. Так бывает, например, когда в модель включены сильно коррелирующие факторы. В этом разделе мы рассмотрим общую процедуру, которая позволит проверить незначимость не только одного, но и группы факторов, а также решить и более сложные задачи.

4.6.1 Общая линейная гипотеза

Пусть, как и ранее, мы имеем модель измерений

$$Y = X \cdot \theta + \varepsilon \tag{4.11}$$

и выполнены условия 1)-6) основных ограничений. Мы хотим проверить гипотезу:

$$H_0: \ A\theta = a \ , \tag{4.12}$$

где A есть неслучайная $p \times (m+1)$-матрица, а a – неслучайный p-мерный вектор-столбец. Например, при проверке значимости отдельного фактора мы проверяли гипотезу $\theta_k = 0$. Соотношение (12) показывает, что на параметры уравнения регрессии накладывается p линейных соотношений. Далее мы будем предполагать, что $p \leq m+1$ и $rank(R) = p$, т.е. все соотношения линейно независимы.

Гипотеза (12) означает, что среди $m+1$ исходных параметров можно выделить некоторые p параметров, которые можно линейно выразить через остальные $(m+1)-p$. Если это сделать, то в линейном пространстве L_{UR}, порожденном векторами $\vec{X}_0, \ldots, \vec{X}_m$, будет выделено некоторое подпространство L_R. Заметим, что набор используемых факторов при этом изменится.

Сначала оценим модель (11) без учета ограничений (12). При этом исходный вектор наблюдений Y может быть представлен в виде

$$Y = \hat{Y}_{UR} + e_{UR} \ ,$$

где $\hat{Y}_{UR}$ есть ортогональная проекция Y на подпространство L_{UR}, e_{UR} – вектор остатков, который ортогонален подпространству L_{UR}.

Далее оценим модель (11) с учетом ограничений (12). В этом случае вектор наблюдений Y имеет представление

$$Y = \hat{Y}_R + e_R \ ,$$

где $\hat{Y}_R$ есть ортогональная проекция Y на подпространство L_R, e_R – соответствующий вектор остатков.

Так как $L_R \subset L_{UR}$, то вектор $e_R - e_{UR}$ является ортогональным к вектору e_{UR} (школьная теорема о трех перпендикулярах!). Окончательно получаем следующее представление:

$$Y = \hat{Y}_R + (e_R - e_{UR}) + e_{UR} \ , \qquad (4.13)$$

где все три слагаемых взаимно ортогональны.

Случайный вектор Y имеет многомерное нормальное распределение в R^N со средним $X\theta$ и матрицей ковариаций $\sigma^2 I_N$. Если верна гипотеза H_0, то $X\theta \in L_R$. В этом случае:

1) случайный вектор $\hat{Y}_R$ имеет невырожденное многомерное нормальное распределение в пространстве размерности $(m+1)-p$ со средним $X\theta$ и матрицей ковариаций $\sigma^2 I_{(m+1)-p}$,

2) случайный вектор $(e_R - e_{UR})$ имеет невырожденное многомерное нормальное распределение в пространстве размерности p с нулевым средним и матрицей ковариаций $\sigma^2 I_p$,

3) случайный вектор e_{UR} имеет невырожденное многомерное нормальное распределение в пространстве размерности $m+1$ с нулевым средним и матрицей ковариаций $\sigma^2 I_{N-(m+1)}$.

В силу ортогональности случайных векторов в представлении (13) они являются некоррелированными и, следовательно, независимыми. Обозначим через $ESS_{UR} = |e_{UR}|^2$ и $ESS_R = |e_R|^2$ остаточные суммы квадратов в задачах оценки модели (11) без учета и с учетом ограничений (12) соответственно. Используя ортогональность, получаем $ESS_R - ESS_{UR} = |e_R - e_{UR}|^2$. Из свойств 2) и 3) и независимости $e_R - e_{UR}$ и e_{UR} следует, что случайные величины ESS_{UR}/σ^2 $(ESS_R - ESS_{UR})/\sigma^2$ независимы и имеют χ^2-распределения с $N-(m+1)$ и p степенями свободы соответственно. Тогда случайная величина

$$F = \frac{(ESS_R - ESS_{UR})/p}{ESS_{UR}/(N-(m+1))}$$

при верной гипотезе H_0 имеет распределение Снедекора-Фишера с $(p, N-(m+1))$ степенями свободы.

Далее, для заданного уровня значимости α ($= 0.1$, 0.05, 0.01) находим по таблицам критическую константу $F_\alpha > 0$:

$$P_0(F > F_\alpha) = \alpha \ .$$

Если реально наблюдаемое значение статистики F_{obs} статистики F окажется больше F_α, то гипотеза H_0 отвергается. В противном случае говорят, что она не противоречит экспериментальным данным.

Описанный выше критерий основан на МНК и работает только при выполнении основных ограничений классической линейной модели.

Далее мы рассматриваем несколько важных частных случаев общей линейной гипотезы

4.6.2 Проверка значимости влияния отдельного фактора

Выше мы уже рассматривали эту задачу в параграфе 4.5. Для ее решения применялся T-критерий Стьюдента. Как отмечалось ранее, формально задача сводится к проверке гипотезы

$$H_0 : \ \theta_k = 0 \ ,$$

против альтернативы

$$H_1 : \ \theta_k \neq 0 \ .$$

Это частный случай общей линейной гипотезы вида (12) и к нему можно применить описанный выше F-критерий. В случае простой линейной регрессии T-критерий и F-критерий эквивалентны. Но для множественной регрессии это, вообще говоря, два разных критерия, т.к. F-критерий учитывает чистое влияние фактора X_k.

4.6.3 Проверка значимости совместного влияния группы факторов

Как легко понять из заголовка данного раздела, мы рассматриваем задачу проверку наличия или отсутствия значимого влияния группы факторов $X_{i_1}, \ldots, X_{i_p}$, $1 \leq i_1 < \ldots < i_p \leq m$. Формально это приводит нас к задаче проверки гипотезы

$$H_0: \ \theta_{i_1} = \ldots = \theta_{i_p} = 0$$

против альтернативы H_1, что хотя бы один из этих параметров отличен от нуля. Фактически мы имеем p линейных ограничений на параметры и вновь можно применить описанный выше F-критерий.

Если гипотеза H_0 не противоречит экспериментальным данным, то мы можем исключить из нашей модели *все* факторы $X_{i_1}, \ldots, X_{i_p}$, так как они не оказывают существенного влияния на поведение изучаемой величины Y.

4.6.4 Проверка адекватности модели

Понятие адекватности модели зависит от того, что именно мы от нее хотим. В данном разделе мы проверяем то, насколько хорошо можно объяснить поведение (изменение) величины Y с помощью совокупного линейного влияния факторов $X_1, X_2, \ldots, X_m$.

В качестве нулевой гипотезы выбирается предположение

$$H_0: \ \theta_1 = \ldots = \theta_m = 0 \ ,$$

т.е., выбранные факторы не оказывают влияния на переменную Y. Тогда альтернатива

$$H_1: \ \theta_k \neq 0 \quad k$$

означает, что совокупное влияние факторов является существенным (значимым).

Отметим, что в этом случае мы рассматриваем только те факторы, которые действительно меняются, и оставляем в стороне фактор X_0. В этом случае удобнее перейти к центрированным наблюдениям. Спроектируем вектор исходных наблюдений Y на вектор $\vec{X}_0$. В результате получаем $\bar{Y} \cdot \vec{X}_0$. Рассмотрим вектор $Y^{(0)} = Y - \bar{Y} \cdot \vec{X}_0$ центрированных наблюдений, компоненты которого равны $Y_j^{(0)} = Y_j - \bar{Y}$, $j = 1, \ldots, N$.

Далее, используя МНК, оценим, как обычно, линейную модель (11). Пусть $\hat{Y}$ есть вектор предсказанных значений. По построению векторы $\hat{Y} - \bar{Y} \cdot \vec{X}_0$ и $Y - \hat{Y}$ ортогональны, т.е. мы имеем следующее ортогональное разложение:

$$Y^{(0)} = Y - \hat{Y} + \hat{Y} - \bar{Y} \cdot \vec{X}_0 \ .$$

Отсюда мы получаем соотношение (теорема Пифагора!):

$$TSS = \sum_{j=1}^{N} (Y_j - \bar{Y})^2 = \sum_{j=1}^{N} (Y_j - \hat{Y}_j)^2 + \sum_{j=1}^{N} (\hat{Y}_j - \bar{Y})^2 = ESS + RSS \ ,$$

где

TSS – *полная сумма квадратов*,

ESS – *остаточная сумма квадратов*,

RSS – *объясненная (регрессионная) сумма квадратов*.

Величина TSS оценивает степень общей изменчивости показателя Y как за счет тех факторов, что включены в модель, так и за счет неучтенных факторов. Величина RSS оценивает степень изменчивости показателя Y, вызванной влиянием (изменением) факторов, включенных в модель. Величина ESS оценивает степень изменчивости Y, вызванной влиянием неучтенных факторов.

Аналогично тому, как это было сделано при рассмотрении общей линейной гипотезы, может быть доказана следующая

Теорема 3 . *Пусть верна гипотеза H_0. Тогда*

1) случайные величины ESS и RSS независимы,

2) случайные величины ESS/σ^2 и RSS/σ^2 имеют χ^2-распределения с $N-m-1$ и m степенями свободы,

3) случайная величина

$$F = \frac{RSS/m}{ESS/(N-(m+1))}$$

имеет распределение Снедекора-Фишера с m и $N-(m+1)$ степенями свободы.

Далее проверка гипотезы проводится по стандартной схеме: если реально полученное значение статистики F будет больше критической константы, найденной из таблиц, то совокупное линейное влияние факторов признается значимым; в противном случае влияние факторов признается незначимым. В последнем случае необходимо выбрать новый набор факторов и построить новую модель.

Важной интегральной оценкой качества подгонки модели к имеющимся экспериментальным данным является следующая характеристика.

Определение 5 . *Величина*

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

называется коэффициентом детерминации.

Коэффициент детерминации показывает качество подгонки регрессионной модели к реально наблюдаемым значениям величины Y. Нетрудно показать, что

$$F = \frac{N-m-1}{m} \cdot \frac{R^2}{1-R^2} \ ,$$

т.е. коэффициент детерминации напрямую связан со статистикой F-критерия для проверки адекватности модели.

4.6.5 Оценка вклада факторов

Обычно при построении модели мы включаем в нее достаточно большое число факторов, часть из которых, возможно, и не оказывает влияния на изучаемый величину Y, либо дублирует другие факторы в силу сильной корреляции с ними. В подобных случаях коэффициент θ_k перед соответствующим фактором X_k должен равняться нулю. Но так как мы имеем только оценку $\hat{\beta}_k$ этого коэффициента, то скорее всего мы получим нечто близкое к нулю, но все-таки отличное от него. Таким образом мы сталкиваемся с проблемой проверки предположения о том, что некоторый фактор не оказывает влияния или, эквивалентно, что соответствующий коэффициент равен нулю.

Выше мы уже рассматривали эту задачу. Здесь она обсуждается с другой точки зрения, что не меняет, конечно, ее математической сути.

Чтобы упростить изложение, мы рассмотрим случай, когда мы имеем только два фактора. И так, мы рассматриваем модель измерений

$$Y_j = \alpha + \beta_1 X_{j1} + \beta_2 X_{j2} + \varepsilon_j \ . \qquad (4.14)$$

Нам необходимо проверить гипотезу

$$H_0 : \ \beta_2 = 0 \ ,$$

против альтернативы

$$H_1 : \ \beta_2 \neq 0 \ .$$

Вначале мы оцениваем уравнение регрессии в полной (длинной) модели без учета проверяемой гипотезы. При этом мы получим регрессию $\hat{Y}_{UR}$ и соответствующую остаточную сумму квадратов ESS_{UR}. Затем мы оцениваем уравнение регрессии в короткой модели, в которой отсутствует фактор X_2. Это дает нам новую регрессию $\hat{Y}_R$ и соответствующую ей остаточную сумму квадратов ESS_R. Ясно, что $ESS_{UR} < ESS_{UR}$. Но при верной гипотезе H_0 эти величины должны быть близки. Более того, уменьшение остаточной суммы квадратов за счет учета фактора X_2 должно быть мало по сравнению с общей ошибкой измерений в нашей модели, которая измеряется с помощью величины ESS_{UR}.

Эти интуитивные рассуждения можно аккуратно обосновать с помощью следующей теоремы.

Теорема 4 . *Пусть верна гипотеза* H_0. *Тогда:*

1) случайные величины $ESS_R - ESS_{UR}$ *и* ESS_{UR} *независимы,*

2) случайная величина

$$F = (N-3)\frac{ESS_R - ESS_{UR}}{ESS_{UR}}$$

имеет распределение Фишера с $(1, N-3)$ *степенями свободы.*

Фактически эта теорема уже была доказана выше.

Теперь мы можем описать критерий для проверки гипотезы H_0. По таблицам распределения Фишера для заданного уровня значимости α находим критическую константу $F(\alpha)$, для которой при верной гипотезе H_0 имеет место соотношение

$$P(F > F(\alpha)) = \alpha \ .$$

Если реально наблюдаемое значение величины F окажется больше $F(\alpha)$, то мы отвергаем гипотезу H_0 и принимаем гипотезу H_1 на заданном уровне значимости. В противном случае мы говорим, что H_0 не противоречит экспериментальным данным.

Если какой-то фактор оказался незначимым, то его следует удалить и пересчитать модель заново, так как оценки для параметров в новом модели будут, вообще говоря, другими. Эту тему мы обсудим позднее подробнее.

Объясним полученный результат несколько подробнее. Можно показать, что имеет место следующее тождество:

$$TSS = RSS_R + (ESS_R - ESS_{UR}) + ESS_{UR} \ .$$

Первое слагаемое описывает вклад первого фактора X_1 (в присутствии X_2). Второе слагаемое описывает чистый вклад второго фактора X_2, когда устранено влияние X_1. Третье слагаемое описывает влияние других (неучтенных) факторов. Если второй фактор не влияет, то его чистый вклад должен быть незначимым на фоне ошибок измерений. Заметим, что это разложение зависит от порядка учета факторов.

4.7 Пример

В этом разделе мы рассмотриваем некоторый модельный пример, который иллюстрирует применение описанных выше процедур. Пусть изучается индекс импорта товаров и услуг Y в зависимости от индекса валового внутреннего продукта X_1 и отношения индексов цен на импортные товары и валового внутреннего продукта. Были получены следующие статистические данные за 9 лет.

Y	X_1	X_2
100	100	100
106	104	99
107	106	110
120	111	126
110	111	113
116	115	103
123	120	102
133	124	103
137	126	98

Мы рассматриваем следующую модель

$$Y_j = \alpha + \beta_1 X_{j1} + \beta_2 X_{j2} + \varepsilon_j \ .$$

Применяя МНК, получаем следующие результаты:

$$\begin{array}{ccccccc} y & = & \text{-49.34} & + & 1.364\ x_1 & + & 0.114\ x_2 \\ SE & & 24.06 & & 0.143 & & 0.143 \\ t & & \text{-2.0507} & & 9.5299 & & 0.7943 \\ \alpha & & 0.0862 & & 0.0001 & & 0.4573 \end{array}$$

Кроме того, $\hat{\sigma^2} = S^2 = 12.9236$, $\hat{\sigma} = 3.5949$. Наблюдаемое значение статистики $F = 45.7825$, что соответствует достигнутому уровню значимости $\alpha = 0.0002$. Это указывает на сильное совместное линейной влияние факторов. $R^2 = 0.938502$, то есть выбранные факторы объясняют почти 94% изменения величины Y.

Анализируя остатки, получаем

	SS	DF	Mean Sq	F-ratio	P-value
RSS	1183.35	2	591.67	45.78	0.0002
ESS	77.54	6	12.92		
TSS	1260.89	8			
X_1	1175.19	1	1175.19	90.93	0.0001
X_2	8.16	1	8.16	0.63	0.4654

Отсюда мы видим, что первый фактор вносит значимый вклад, второй – незначимый.

О значимости вклада факторов можно судить и по величине частного коэффициента корреляции. В нашем случае $\rho_{YX_1.X_2} = 0.96852$, $\rho_{YX_2.X_1} = 0.30846$, то есть, чистое влияние первого фактора велико, а второго – нет.

Все указывает на то, что второй фактор незначимый и разумно устранить его из модели. Оценивая новую модель, где второй фактор устранен, получаем:

$$\begin{array}{ccccc} y & = & \text{-35.05} & + & 1.345\ x_1 \\ SE & & 15.55 & & 0.137 \\ t & & \text{-2.25} & & 9.800 \\ \alpha & & 0.059 & & 0.000 \end{array}$$

Причем, $R^2 = 0.92$, $F = 96$, $p = 0.000$.

4.8 Спецификация модели

Всюду ранее мы предполагали, что мы имеем дело с *правильной* моделью, которая имеет вид

$$Y = X \cdot \theta + \varepsilon\ ,$$

и набор объясняющих факторов $X_1, \ldots, X_m$ (это и есть спецификация модели!) выбран правильно. Кроме того, предполагается, что выполнены основные ограничения 1-6 классической линейной модели. Про такую правильную или истинную модель мы будем говорить, что она *порождает наблюдаемые данные*. Но при построении модели в реальной задаче мы не знаем истинного набора объясняющих факторов. Поэтому часть *существенных* переменных может быть пропущена, а часть *несущественных*, наоборот, включена в модель. В этом разделе мы посмотрим как это влияет на оценки параметров. Для простоты рассуждений и обозначений мы будем рассматривать случай, когда на Y могут влиять только два фактора. Но полученные результаты справедливы и в общем случае.

4.8.1 Исключена существенная переменная

Предположим, что наблюдаемые данные порождаются процессом

$$Y_j = \alpha + \beta_1 \cdot X_{j1} + \beta_2 \cdot X_{j2} + \varepsilon_j \ , \tag{4.15}$$

а мы оцениваем (рассматриваем) модель

$$Y_j = \alpha + \beta_1 \cdot X_{j1} + \varepsilon_j \ . \tag{4.16}$$

Модель (15) будем называть *длинной* (L), а модель (16) – *короткой* (S).

Обозначим $X_{j0} \equiv 1$, $X_S = (\vec{X}_0, \vec{X}_1)$, $X_L = (\vec{X}_0, \vec{X}_1, \vec{X}_2)$ – матрицы планов эксперимента в короткой и длинной моделях, $\theta_S = (\alpha, \beta_1)^T$ – вектор параметров в оцениваемой (короткой) модели. Тогда, применяя МНК, мы получаем оценку

$$\hat{\theta}_S = (X_S^T X_S)^{-1} X_S^T Y = Z_S \cdot Y \ . \tag{4.17}$$

В частности

$$\hat{\beta}_{1,S} = \frac{S_{YX_1}}{S_{X_1}^2} \ .$$

В силу (15) и (17)

$$Y = X_S \cdot \theta_S + \vec{X}_2 \cdot \beta_2 + \varepsilon \ ,$$

$$\hat{\theta}_S = \theta_S + (X_S^T X_S)^{-1} X_S^T \cdot \vec{X}_2 \cdot \beta_2 + (X_S^T X_S)^{-1} X_S^T \cdot \varepsilon \ ,$$

и

$$M(\hat{\theta}_S) = \theta_S + (X_S^T X_S)^{-1} X_S^T \cdot \vec{X}_2 \cdot \beta_2 \ .$$

Если $\beta_2 \neq 0$, то, вообще говоря, $M(\hat{\theta}_S) \neq \theta$, т. е. оценки получились *смещенными*. Исключением является случай, когда $X_S^T \cdot \vec{X}_2 = 0$, т. е. $(\vec{X}_0, \vec{X}_2) = (\vec{X}_0, \vec{X}_2) = 0$ – фактор X_2 ортогонален факторам X_0 и X_1. В этом случае оценки по МНК в короткой и длинной моделях совпадают.

Обозначим

$$\begin{aligned} \hat{Y}_S &= X_S \cdot \hat{\theta}_S = X_S \cdot (X_S^T X_S)^{-1} \cdot X_S^T \cdot Y = P_S \cdot Y \ , \\ e_S &= Y - \hat{Y} = (I_N - P_S) \cdot Y = P_S^{\perp} \cdot Y = P_S^{\perp} \cdot \vec{X}_2 \cdot \beta_2 + P_S^{\perp} \cdot \varepsilon \ , \end{aligned}$$

где P_S и $P_S^{\perp}$ есть проекторы в пространстве R^N на подпространство L_S, порожденное векторами $\vec{X}_0, \vec{X}_1$, и ортогональное к нему подпространство $L_S^{\perp}$ соответственно. Случайный вектор e_S имеет многомерное нормальное распределение в $L_S^{\perp}$ со средним $P_S^{\perp} \cdot \vec{X}_2 \cdot \beta_2$ и матрицей ковариаций $\sigma^2 \cdot I_{N-2}$. Возьмем, как обычно, в качестве оценки дисперсии ошибок σ^2 величину

$$\hat{\sigma}_S^2 = \frac{|e_S|^2}{N-2} \ .$$

В силу сказанного (смотри также доказательство предложения 4)

$$M(\hat{\sigma^2}_S) = \sigma^2 + \frac{1}{N-2} \cdot |P_S^{\perp} \cdot \vec{X}_2 \cdot \beta_2|^2 = \sigma^2 + \beta_2^2 \cdot \frac{1}{N-2} \sum_{j=1}^{N} (P_S^{\perp} \cdot \vec{X}_2)_j^2 \ .$$

Если $\beta_2 \neq 0$, то $\hat{\sigma^2}_S$ является смещенной оценкой для σ^2 и, в среднем, она завышена. Это приводит к тому, что в некоторых статистических процедурах мы можем приходить к ложным выводам. Например, значимые факторы признаются незначимыми, т.е. занижается эффект их влияния.

Сравним дисперсии оценок для β_1 в истинной модели (15) и оцениваемой модели (16). Оценка $\hat{\theta}_S = Z_S \cdot Y$ для параметра θ_S в модели (16) имеет ковариационную матрицу $\sigma^2 \cdot Z_S \cdot Z_S^T = \sigma^2 \cdot (X_S^T \cdot X_S)^{-1}$. Так как, по существу, мы имеем дело с простой линейной регрессией, то (смотри предложение 2 из главы 3)

$$D(\hat{\beta}_1) = \frac{\sigma^2}{\sum\limits_j (X_{j1} - \bar{X}_1)^2} = \frac{\sigma^2}{N \cdot S^2_{X_1}} \ . \tag{4.18}$$

В длинной модели оценка $\hat{\theta}_L$ вектора параметров имеет

$$\hat{\theta}_L = (X_L^T X_L)^{-1} \cdot X_L^T \cdot Y \ ,$$

причем она имеет среднее $\theta_L = (\alpha, \beta_1, \beta_2)^T$ и матрицу ковариаций $\sigma^2 \cdot (X_L^T \cdot X_L)^{-1}$. Прямым вычислением, используя определители, отсюда можно получить

$$D(\hat{\beta}_1) = \frac{\sigma^2}{\sum\limits_j (X_{j1} - \bar{X}_1)^2} \cdot \frac{1}{(1 - r^2_{X_1,X_2})} \ , \tag{4.19}$$

где r_{X_1,X_2} – выборочный коэффициент корреляции для факторов X_1 и X_2. Сравнение (18) и (19) показывает, что дисперсия оценки для параметра β_1 в короткой модели меньше, чем в длинной. Но это не должно вводить в заблуждение, т.к. оценка $\hat{\beta}_{1,S}$ является смещенной и ее точность следует оценивать с помощью величины

$$M[(\hat{\beta}_{1,S} - \beta_1)^2] = D(\hat{\beta}_{1,S}) + [(Z_S \cdot \vec{X}_2)_2 \cdot \beta_2]^2 \ . \tag{4.20}$$

Сравнивать величины (19) и (20) непросто и, кроме того, вторая величина зависит от β_2.

Таким образом, в короткой модели оценка β_1 смещена, но обладает меньшей дисперсией. Оценка для дисперсии σ^2 всегда смещена в сторону завышения.

4.8.2 Включена несущественная переменная

Рассмотрим теперь противоположный случай. Истинная *модель, порождающая данные*, есть

$$Y = \alpha + \beta_1 \cdot X_{j1} + \varepsilon_j \ , \tag{4.21}$$

а мы *оцениваем модель*

$$Y = \alpha + \beta_1 \cdot X_{j1} + \beta_2 \cdot X_{j2} + \varepsilon_j \ . \qquad (4.22)$$

На самом деле, модель (22) также является истинной, т.к. $\beta_2 = 0$. Но для ее оценки мы используем другие процедуры. Если построить оценки $\hat{\beta}_{1,L}$, $\hat{\beta}_{2,L}$ и $\hat{\sigma}^2_L$ для модели (22), то $M(\hat{\beta}_{1,L}) = \beta_1$, $M(\hat{\sigma}^2_L) = \sigma^2$, т. е. оценки несмещенные. Но, как показано выше, дисперсия оценки $\hat{\beta}_{1,L}$ больше дисперсии оценки $\hat{\beta}_{1,S}$ за счет включения несущественной переменной.

4.9 Выбор модели

Если у нас есть несколько моделей, например, (15) и (16), то мы должны выбрать одну из них. Обычно это делают на основе коэффициента детерминации R^2. Но R^2 возрастает с увеличием числа факторов независимо от того сколько их в истинной модели. Отсюда получается, что чем больше факторов, тем лучше. R^2 является оценкой для множественного коэффициента корреляции, причем эта оценка смещенная. Несмещенной оценкой будет

$$R^2_{adj} = 1 - (1 - R^2)\frac{N - 1}{N - m - 1}$$

– *исправленный коэффициент детерминации.* Этот коэффициент сначала растет с ростом m, а затем убывает. Поэтому существует некоторое оптимальное значение. Но и этот показатель не является универсальным. Кроме того, для одного и того же числа факторов m разные наборы факторов дают разные результаты.

Поэтому для выбора наилучшей модели рекомендуется перебрать все возможные модели и выбрать ту, для которой будет приемлемой сумма квадратов остатков.

4.10 Сравнение двух регрессий

Разработанная выше техника проверки линейных гипотез для параметров линейной регрессии позволяет решить еще одну полезную задачу. Предположим, что одну и ту же регрессионную модель мы оцениваем для двух различных наборов данных, т.е. они получены из разных генеральных совокупностей. Например, мы можем изучать зависимость величины заработной платы в зависимости от различных факторов (возраста, стажа, уровня образования и т.п.) для мужчин и женщин. В этом случае естественным образом возникает задача сравнения этих двух регрессий, т.е. мы хотим проверить гипотезу о том, что мы оцениваем одно и то же уравнение. Формально это выглядит следующим образом. Пусть мы имеем два независимых набора данных для одной и той же задачи, по которым мы строим регрессионые модели с одними и теми же факторами. Для первого набора данных мы рассматриваем модель измерений

$$Y_j = \beta_1' X_{j1} + \beta_2' X_{j2} + \ldots + \beta_m' X_{jm} + \varepsilon_j' \ , \ j = 1, \ldots, N_1 \ , \quad (4.23)$$

для второго набора данных мы имеем, аналогично, следующую модель измерений:

$$Y_j = \beta_1'' X_{j1} + \beta_2'' X_{j2} + \ldots + \beta_m'' X_{jm} + \varepsilon_j'' \ , \ j = N_1 + 1, \ldots, N_1 + N_2 \ . \quad (4.24)$$

Тогда задача о равенстве этих двух регрессий сводится к проверке гипотезы:

$$\mathbf{H_0} : \beta_k' = \beta_k'' \ , \ k = 1, \ldots, m, \quad \sigma' = \sigma'' \ , \quad (4.25)$$

против альтернативы, что хотя бы один коэффициент в этих двух регрессиях разный.

Рассмотрим вспомогательную задачу, в которой мы имеем $N_1 + N_2$ независимых измерений следующего вида

$$Y_j = \beta'_1 \tilde{X}_{j,1} + \ldots + \beta'_m \tilde{X}_{j,m} + \beta''_1 \tilde{X}_{j,m+1} + \beta''_m \tilde{X}_{j,2m} + \varepsilon'_j \ , \quad (4.26)$$

где измерения $\tilde{X}_{j,k}$ новых факторов определены по правилу

$$\tilde{X}_{j,k} = \begin{cases} X_{j,k} & , \quad 1 \le j \le N_1, \\ 0 & , \quad N_1 + 1 \le j \le N_1 + N_2 \end{cases}$$

для $1 \le k \le m$, и

$$\tilde{X}_{j,k} = \begin{cases} 0 & , \quad 1 \le j \le N_1, \\ X_{j,k-m} & , \quad N_1 + 1 \le j \le N_1 + N_2 \end{cases}$$

для $m + 1 \le k \le 2m$. Ошибки $\varepsilon_j, \ j = 1, \ldots, N_1 + N_2$, есть независимые случайные величины, которые имеют нормальное распределение с нулевым средним и дисперсией σ^2. Тогда гипотеза (25) есть обычная линейная гипотеза для модели (26), которая задает m линейных ограничений.

В силу специфики измерений факторов в модели (26) оценки параметров в ней по МНК без учета ограничений сводятся к независимой оценке параметров в моделях (23) и (24). Причем остаточная сумма квадратов ESS_{UR} в модели (26) равна сумме $ESS_1 + ESS_2$ остаточных сумм квадратов в моделях (23) и (24).

Если мы учитываем ограничения, то, фактически, оценивается модель вида (23) (или 24), но по всей совокупности $N_1 + N_2 = N$ измерений. При этом мы получаем некоторую остаточную сумму квадратов ESS_R.

Из общей теории проверки линейных гипотез мы знаем, что при верной гипотезе H_0 случайные величины ESS_{UR}/σ^2 $= (ESS_1 + ESS_2)/\sigma^2$ и $(ESS_R - ESS_{UR})/\sigma^2$ независимы и имеют χ^2-распределения с $N - 2m$ и m степенями свободы соответственно, а случайная величина

$$F = \frac{(ESS_R - ESS_{UR})/m}{ESS_{UR}/(N-2m)}$$

имеет распределение Снедекора-Фишера с $(m, N-2m)$ степенями свободы. Далее, используя статистику F, критерий для проверки H_0 строится по стандартной схеме. Полученная процедура известна как *тест Чоу* (Chow). Заметим, что эта задача может быть обобщена на случай нескольких регрессий.

Пример. Для иллюстрации изучаемой темы рассмотрим следующий пример, который предлагался на семинаре "Преподавание начального курса эконометрики"(Arthur van Soest, Tilburg University).

Изучается разница в оплате мужчин и женщин в Нидерландах. На основании опроса 75 мужчин и 75 женщин, работавших на полную ставку, была получена информация о их зарплате (w), а также значения двух факторов: возраст (age) и уровень образования (edu).

Оценим регрессию величины зарплаты на факторы возраста и уровня образования для мужчин. Получаем уравнение регрессии

$$w^{(1)} = -3.37 + 0.479 \cdot age + 3.943 \cdot edu \ ,$$

причем оценка свободного члена оказалась незначимой, а оценки факторов значимыми. Остаточная сумма квадратов равна $ESS_1 = 5672.328$.

Аналогичное уравнение регрессии для женщин оказалось следующим:

$$w^{(2)} = -0.20 + 0.414 \cdot age + 2.305 \cdot edu \ ,$$

остаточная сумма квадратов равна $ESS_2 = 1788.344$.

Отсюда получаем, что остаточная сумма квадратов в модели без ограничений равна $ESS_{UR} = 7460.672$.

Уравнение регрессии, оцененное по всей выборке имеет вид

$$w = -3.06 + 0.478 \cdot age + 3.254 \cdot edu \ ,$$

остаточная сумма квадратов в модели с учетом ограничений равна $ESS_R = 8080.113$.

Наблюдаемое значение статистики F равно

$$\frac{(ESS_R - ESS_{UR})/m}{ESS_{UR}/(N-2m)} = 3.342 \ .$$

Такое значение статистики F при числе степеней свободы (3, 144) является значимым для уровня значимости теста, равного 0.05.

4.11 Ортогональная матрица плана эксперимента

Как мы отмечали выше, при удалении и добавлении факторов оценки параметров при тех факторах, которые присутствуют в обеих моделях, будут, вообще говоря, различными. Но в одном специальном случае коэффициенты меняться не будут. Назовем *матрицей плана эксперимента* матрицу X. Эта матрица называется *ортогональной*, если ортогональны столбцы этой матрицы. В этом случае матрица $X^T X$ будет диагональной и нормальная система уравнений при оценке параметров распадается на отдельные уравнения для каждого фактора. Поэтому добавление или удаление такого ортогонального фактора не меняет уравнения для вычисления оценки коэффициента при другом факторе.

Ортогональная матрица плана эксперимента оказывается особенно полезной в случае, когда мы не уверены в правильной спецификации модели. В этом случае как при потере существенной переменной, так и при включении несущественной переменной все полученные оценки для параметров уравнения регрессии и дисперсии ошибок измерений будут несмещенными и не будет происходит изменения точности построенных оценок.

Другой важный случай, когда бывает полезной ортогональная матрица плана, это ситуация мультиколлинеарности. В этом случае матрица $(X^T X)$ имеет определеитель, близкий к нулю, т.е. она практически вырождена. В такой ситуации обычная процедура оценки параметров становится сильно неустойчивой: небольшое изменение начальных данных приводит к сильному изменению величины полученных оценок. Как правило при этом наблюдается незначимость нескольких факторов. Исправить ситуацию можно устранив один или несколько факторов. Особенно легко это можно делать в случае ортогональной матрицы плана, т.к. факторы можно удалять не меняя оценок коэффициентов при других факторах. Более того, по величине значимости факторов, можно отобрать те, которые нужно удалить в первую очередь.

Глава 5

Фиктивные переменные

До сих пор мы неявно предполагали, что наши факторы $X_1, \ldots, X_m$ меняются непрерывно. Но часто возникают ситуации, когда нужно учесть некоторые *качественные* факторы, которые меняются дискретно, например, разделить студентов по их уровню знаний на экзаменах, учесть влияние времени года на продажу товара, выяснить влияние наличия или отсутствия некоторого признака. Простейший подход к таким задачам состоит в том, чтобы построить несколько уравнений, соответствующих разным значениям качественного признака, а затем сравнить их. Но более эффективным является введение так называемых *фиктивных (dummy)* переменных.

Пример 1. Изучается зависимость зарплаты Y от некоторого набора факторов $X_1, \ldots, X_m$. Обычно модель имеет вид

$$Y_j = X_{j1}\beta_1 + \ldots + X_{jm}\beta_m + \varepsilon_j \ .$$

Предположим, что мы решили выяснить влияет ли наличие высшего образования на величину зарплаты. Для этого введем еще один фактор d, который равен 1, если высшее образование есть, и равен 0 в противном случае.

Новая модель имеет вид

$$Y_j = \alpha d_j + X_{j1}\beta_1 + \ldots + X_{jm}\beta_m + \varepsilon_j \ .$$

Для проверки значимости влияния наличия высшего образования на величину зарплаты нужно проверить гипотезу $H_0 : \alpha = 0$. Сама величина α показывает величину прибавки при переходе из одной категории в другую.

Если интересующая нас качественная переменная принимает более двух значений, то можно ввести дополнительную переменную с несколькими дискретными значениями, например, 0, 1, 2, 3. Но, так как перед такой переменной будет стоять некоторый единый коэффициент, то мы неявно предполагаем, что переходы из категории 0 в категорию 1 и из категории 1 в категорию 2 приводят к одинаковым изменениям изучаемой переменной Y. Часто по смыслу задачи видно, что это не так. В таких случаях удобнее ввести несколько бинарных переменных для кодировки различных уровней качественного фактора. Продемонстрируем это на следующем примере.

Пример 2. Изучается объем продаж мороженого в зависимости от некоторых факторов $X_1, \ldots, X_m$, например, цены на мороженое, цен на сопутствующие товары и т.д. Ясно, что объемы продаж мороженого зимой и летом будут разными. Поэтому хотелось бы изучить зависимость объемов продаж в зависимости от времени года. Для этого введем три (почему не четыре?!) бинарных переменных d_1, d_2, d_3, которые равны 1 соответственно зимой, весной и летом и равны нулю в противном случае. Тогда модель измерений будет иметь вид

$$Y_j = d_{j1}\alpha_1 + d_{j2}\alpha_2 + d_{j3}\alpha_3 + X_{j1}\beta_1 + \ldots + X_{jm}\beta_m + \varepsilon_j \ .$$

Рассмотрим еще несколько примеров.

Пример 3. В 1974 году наблюдалось резкое снижение расходов на пребретение автомобилей в США. Это было связано с арабо-израильской войной 1973 года, которая привела к увеличению цен на бензин. Затем расходы вновь начали расти. Мы предполагаем, что функция спроса имела сдвиг в 1974 году. Чтобы учесть этот эффект, мы вводим вспомогательную переменную d, которая равна 0 до 1974 года и равна 1 после 1974 года. Если Y есть расходы на автомобили, а X есть личный располагаемый доход, то можно рассмотреть следующую модель:

$$Y_t = \alpha + \beta \cdot X_t + \delta \cdot d_t + \varepsilon_t \ .$$

Здесь t обозначает год, т.е. мы имеем модель временного ряда. По имеющим данным о расходах на автомобили в США с 1963 по 1982 годы была проведена оценка этой модели и получены следующие результаты.

$$y = 0.57 + 0.035 \cdot x - 4.40 \cdot d \ .$$

Мы видим, что $\alpha = 0.57$ до 1974 года и $\alpha + \delta = -3.83$ после 1974 года. Гипотеза $H_0 : \delta = 0$ отвергается на 5% уровне.

В заключение рассмотрим еще раз пример о зависимости величины зарплаты в Голландии от пола работника.

Пример 4. Изучается зависимость величины зарплаты в Голландии от возраста (стажа работы) и пола работника. Для решения этой задачи используются данные, собранные в Голландии в 1997 году. Для исследования были отобраны 75 мужчин и 75 женщин. Введем фиктивную переменную S, равную 1 для женщин и 0 для мужчин. Оценивалась линейная регрессия вида

$$W_j = \alpha + \beta_1 \cdot age_j + \beta_2 \cdot S_j + \varepsilon_j \ .$$

Результаты оценки модели собраны в таблице

coef	estimates	SE	t	p
α	6.11	2.39	2.56	0.011
β_1	0.53	0.06	8.58	0.000
β_2	- 3.73	1.35	- 2.76	0.0065

Таким образом, мы получили следующее уравнение регресии:

$$\hat{W} = 6.11 + 0.53 \cdot age - 3.73 \cdot s \ .$$

Коэфициент детерминации равен $R^2 = 0.39$. Из таблицы мы видим, что коэффициенты β_1 и β_2 являются значимыми на уровне 0.01. Модель объясняет поведение переменной W на 39 процентов. При прочих равных условиях различие в зарплате у мужчин и женщин равно 3.73.

Глава 6

Обобщения классической линейной модели

6.1 Обзор результатов для классической линейной модели

Прежде, чем переходить к обобщениям классической линейной модели, напомним основные результаты, полученные в рамках этой модели.

Пусть мы имеем одну зависимую переменную и m факторов (объясняющих переменных). Проводится N измерений, причем предполагается, что имеет место следующая модель

$$Y_j = g(X_{j1}, \ldots, X_{jm}) + \varepsilon_j \ .$$

Всюду далее в этом параграфе предполагается, что выполнены следующие *основные ограничения*:

1) модель линейна по параметрам

$$Y_j = \alpha + \beta_1 X_{j1} + \ldots + \beta_m X_{jm}) + \varepsilon_j \tag{6.1}$$

или, в матричной форме,

$$Y = X\theta + \varepsilon \ ,$$

где $X_{j0} \equiv 1$, $\theta = (\alpha, \beta_1, \dots, \beta_m)^T$;

2) X – неслучайная матрица;

3) $\{\varepsilon_j\}$ – независимы (некоррелированы);

4) $M(\varepsilon_j) = 0$;

5) $D(\varepsilon_j) = \sigma^2$ для всех j (гомоскедастичность);

Из (3)-(5) следует, что

$$\Sigma_\varepsilon = M(\varepsilon \cdot \varepsilon^T) = \sigma^2 \cdot E \ , \tag{6.2}$$

6) Ошибки $\{\varepsilon_j\}$ имеют нормальное распределение.

Основной задачей является оценка параметров θ и дисперсии ошибок σ^2. Если матрица плана эксперимента $X^T X$ невырождена (эквивалентно X имеет ранг $m+1$), то существуют оценки по МНК и они имеют вид

$$\hat{\theta} = (X^T X)^{-1} X^T Y = \theta + X^T X^{-1} X^T \varepsilon \ . \tag{6.3}$$

Оценка дисперсии σ^2 находится по формуле

$$\hat{\sigma}^2 = \frac{1}{N-(m+1)} \sum_{j=1}^{N} (Y_j - \hat{Y}_j)^2 \ . \tag{6.4}$$

Полученные оценки обладают следующими свойствами:

1) оценка $\hat{\theta}$ – линейна;

2) $\hat{\theta}$ – несмещенная оценка для θ;

3) $\hat{\theta}$ – оптимальная в среднем квадратическом оценка в классе всех линейных несмещенных оценок;

4) если $\frac{1}{N} X^T X \to \Sigma_X$ при $N \to \infty$, где Σ_X – невырожденная матрица размера $m+1$, то $\hat{\theta} = \hat{\theta}_N$ – состоятельная оценка, т. е. $\hat{\theta}_N \to \beta$, $N \to \infty$;

5) $\hat{\sigma}^2$ – несмещенная и состоятельная оценка для σ^2.

Предположение о нормальности используется при построении доверительных интервалов и проверке гипотез.

6.2 Модель со стохастическими факторами

В реальных ситуациях мы не можем *управлять* проведением экперимента и факторы $X_1, \ldots, X_m$ сами являются случайными величинами. Ниже будет показано, что при выполнении некоторых ограничений это не нарушает основных результатов, полученных для классической линейной модели.

6.2.1 Описание модели

Пусть, как и ранее, проведено N измерений и выполнены следующие ограничения:

1) модель линейна по параметрам

$$Y = X\theta + \varepsilon \ ,$$

2) X –случайная матрица, но $X_{j0} \equiv 1$, причем (без ограничения общности) $M(X_{jk}) = 0$;
3) X и ϵ – независимы (некоррелированы);
4) $M(\varepsilon) = M(\varepsilon|X) = 0$;
5) $\Sigma_\varepsilon = M(\varepsilon\varepsilon^T) = M(\varepsilon\varepsilon^T|X) = \sigma^2 \cdot E$;
6) $(Y, X_1, \ldots, X_m)$ имеют невырожденное нормальное распределение.

В силу закона больших чисел получаем

$$\frac{1}{N}\varepsilon^T\varepsilon = \frac{1}{N}\sum_{j=1}^{N}\varepsilon_j^2 \xrightarrow{P} \sigma^2 \qquad (6.5)$$

$$\frac{1}{N}X^T X \xrightarrow{P} \Sigma_{XX} \ , \qquad (6.6)$$

где Σ_{XX} – невырожденная квадратная матрица размера $m+1$,

$$\frac{1}{N} X^T \varepsilon \xrightarrow{P} 0 \ . \tag{6.7}$$

6.2.2 Оценки параметров и их свойства

Применяя обычный МНК, мы получаем оценки для θ и σ^2:

$$\hat{\theta} = (X^T X)^{-1} X^T Y = \theta + X^T X^{-1} X^T \varepsilon \ , \tag{6.8}$$

$$\hat{\sigma}^2 = \frac{1}{N-(m+1)} \sum_{j=1}^{N} (Y_j - \hat{Y}_j)^2 \ . \tag{6.9}$$

Покажем, что эти оценки обладают всеми свойствами, которые были доказаны в рамках классической линейной модели.

Пусть выполнены условия 1)-6), сформулированные выше. Тогда справедливы следующие свойства.

1) $\hat{\theta}$ – линейна по параметрам

Это свойство есть прямое следствие вида оценки $\hat{\theta}$ (см. (8)).

2) $\hat{\theta}$ – несмещенная оценка параметра θ.

Для доказательства нужно применить формулу полного математического ожидания к выражению (8) и использовать свойство 4).

3) $\hat{\theta} = \hat{\theta}_N$ – состоятельная оценка.

Для доказательства используем явный вид (8) для оценки параметра θ, а также соотношение (7):

$$\hat{\theta} = (X^T X)^{-1} X^T Y = \theta + X^T X^{-1} X^T \varepsilon =$$

$$= \theta + \left(\frac{1}{N} X^T X^{-1}\right) \cdot \left(\frac{1}{N} X^T \varepsilon\right)$$

$$\xrightarrow{P} \theta + (\Sigma_{XX})^{-1} \cdot 0 = \theta \ .$$

Важным моментом последнего доказательства является то, что X и ε являются некоррелированными. Если это не так, то мы получаем

$$\frac{1}{N} X^T \varepsilon \xrightarrow{P} \Sigma_{X\varepsilon} \ ,$$

где $\Sigma_{X\varepsilon}$ есть вектор ковариаций факторов $X_1, \ldots, X_m$ и ошибки ε. Если какая-то из компонент этого вектора не равна нулю, т. е. соответствующий фактор коррелирует с ошибкой измерения, то для такой же компоненты вектора оценок параметров не выполняется свойство состоятельности.

4) $\hat{\theta}$ оптимальна в среднем квадратическом в классе всех линейных несмещенных оценок.

Мы уже доказали это свойство для случая, когда матрица X неслучайна. Для завершения доказательства нужно применить формулу полного математического ожидания (Задача!).

5) $\hat{\sigma}^2$ – несмещенная и состоятельная оценка для σ^2.

6.2.3 Метод инструментальных переменных

Выше, при доказательстве состоятельности оценок, было отмечено, что это свойство нарушается, если факторы коррелируют с ошибкой измерений. В подобной ситуации применяют иак называемый *метод инструментальных переменных*. Суть этого метода состоит в том, что мы заменяем "плохие"(коррелирующие с ошибкой) факторы на некоторые новые переменные, которые тесно связаны (сильно коррелируют) с теми переменными, которые мы удаляем, но сами не коррелируют с ошибкой.

Пусть мы имеем матрицу Z размера $N\times(m+1)$, $Z_{j0}\equiv 1$, измерений некоторых новых измерений, причем выполнены следующие условия:

$$\frac{1}{N}Z^T\varepsilon \xrightarrow{P} 0 \ , \tag{6.10}$$

т. е. новые переменные не коррелируют с ошибками,

$$\frac{1}{N}X^TX \xrightarrow{P} \Sigma_{ZX} \ , \tag{6.11}$$

где Σ_{ZX} – невырожденная квадратная матрица размера $m+1$, т. е. новые и старые переменные коррелируют и могут быть выражены друг через друга.

В качестве новой оценки для параметра θ предлагается взять

$$\tilde{\theta} = (Z^TX)^{-1}Z^TY = \theta + (Z^TX)^{-1}Z^T\varepsilon \tag{6.12}$$

Покажем, что эта оценка является состоятельной. В силу свойств (10) и (11) получаем

$$\tilde{\theta} = \theta + \left(\frac{1}{N}Z^TX\right)^{-1}\left(\frac{1}{N}Z^T\varepsilon\right) \xrightarrow{P} \theta + \Sigma_{ZX}^{-1}\cdot 0 = \theta \ .$$

Реально матрица Z строиться из матрицы X заменой некоторых переменных, которые коррелируют с ошибками, на экзогенные переменные, которые всегда *предполагаются* некоррелированными с ошибками.

6.3 Обобщенный метод наименьших квадратов

В классической линейной модели предполагалось, что

$$Y = X\theta + \varepsilon$$

и ошибки измерений имеют следующую матрицу ковариаций

$$\Sigma_\varepsilon = M(\varepsilon\varepsilon^T) = \sigma^2 \cdot E \ ,$$

что эквивалентно тому, что ошибки в разных измерениях некоррелированы и имеют одинаковые дисперсии. Во многих реальных экономических задачах это не так. Интуитивно понятно, что, как правило, большим значениям X_{jk} соответствует и больший разброс значений Y_j. При анализе временных рядов естественно предположить, что Y_j и Y_{j-1} (эквивалентно ε_j и ε_{j-1}) коррелированы. Для описания подобных ситуаций рассмотрим более сложную модель. Предположим, что

$$\Sigma_\varepsilon = \sigma^2 \cdot \Omega \ , \tag{6.13}$$

где Ω – *известная* симметричная положительно определенная $N \times N$-матрица, σ^2 – *неизвестный* параметр.

Если для оценки параметров θ применить обычный МНК, то получим оценку

$$\hat{\theta} = (X^T X)^{-1} X^T Y = \theta + X^T X^{-1} X^T \varepsilon \ .$$

Если выполнены основные ограничения классической линейной модели с заменой свойства (2) на (13), то оценка θ будет несмещенной и состоятельной, но, как мы увидим ниже, она не будет эффективной (оптимальной в среднем квадратическом). Для получения оптимальной оценки сведем задачу к классическому случаю. Так как Ω есть известная симметричная положительно определенная матрица, то можно найти невырожденную матрицу P размера $N \times N$ такую, что

$$\Omega = P \cdot P^T \ . \tag{6.14}$$

Отметим, что матрица P определяется не единственным образом. Положим

$$Y_* = P^{-1} Y \ , \ X_* = P^{-1} X \ , \ \varepsilon_* = P^{-1} \varepsilon \ . \tag{6.15}$$

Для новых переменных мы имеем модель

$$Y_* = X_* \theta + \varepsilon_* \tag{6.16}$$

и

$$\Sigma_{\varepsilon_*} = M(\varepsilon_* \varepsilon_*^T) = M(P^{-1} \varepsilon \varepsilon^T (P^T)^{-1}) =$$
$$= P^{-1} \sigma^2 \Omega (P^T)^{-1} = \sigma^2 P^{-1} P P^T (P^T)^{-1} = \sigma^2 \cdot E \ .$$

Таким образом, мы вновь имеем классическую линейную модель. Применяя к ней обычный МНК, т. е. минимизируя по θ квадратичную форму

$$(Y_* - X_* \theta)^T (Y_* - X_* \theta) = (Y - X\theta)^T \cdot \Omega^{-1} (Y - X\theta) \ , \tag{6.17}$$

в результате мы получаем оценку

$$\begin{aligned} \theta^* &= (X_*^T X_*)^{-1} X_*^T Y_* \\ &= (X^T \Omega^{-1} X)^{-1} X^T \Omega^{-1} Y \\ &= \theta + (X^T \Omega^{-1} X)^{-1} X^T \Omega^{-1} \varepsilon \ . \end{aligned} \tag{6.18}$$

Такой метод получения оценок парметров θ называется *обобщенным МНК*. Оценка θ^* называется *оценкой обобщенного МНК* или *оценкой Эйткена*. Она является линейной, несмещенной, состоятельной, оптимальной в среднем квадратическом в классе всех линейных несмещенных оценок, имеет многомерное нормальное распределение со средним θ и матрицей ковариаций $\sigma^2 \cdot (X^T \Omega^{-1} X)^{-1}$.

Для оценки σ^2, как и раньше, используем остатки $e_* = Y_* - \hat{Y}_* = Y_* - X_* \theta^*$. Тогда по формуле (4) получаем

$$\begin{aligned} \sigma_*^2 &= \frac{1}{N - (m+1)} \sum_{j=1}^{N} (Y_{*j} - \hat{Y}_{*j})^2 \\ &= \frac{1}{N - (m+1)} (Y_* - \hat{Y}_*)^T (Y_* - \hat{Y}_*) \ . \end{aligned} \tag{6.19}$$

Это несмещенная и состоятельная оценка для σ^2. Применение обычного МНК приведет к смещенным оценкам для дисперсии, что может исказить результаты анализа модели. Как и ранее можно строить доверительные интервалы и проверять гипотезы о параметрах, но с соответствующими изменениями (нужно использовать Y_*, X_* и θ^* !).

В частном случае, когда Ω диагональная матрица с элементами $w_1, \ldots, w_N$ квадратичная форма (17) переходит в следующую

$$\sum_{j=1}^{N}[Y_j - (\alpha + \beta_1 X_{j1} + \ldots + \beta_m X_{jm})]^2 \cdot \frac{1}{w_j} \; .$$

Поэтому в данной ситуации говорят о *взвешенном МНК*.

Если матрица Ω неизвестна, то пытаются ее оценить. В общем случае Ω содержит $N(N+1)/2$ неизвестных параметров. Оценить их, имея только N измерений, невозможно. Для получения разумных результатов необходимы те или иные предположения о структуре Ω. Некоторые примеры будут рассмотрены ниже.

Глава 7

Гетероскедастичность и автокорреляция ошибок

Все основные результаты, полученные выше, были выведены для линейной модели

$$Y = X\theta + \varepsilon \ , \tag{7.1}$$

в предположении, что вектор ошибок ε имеет матрицу ковариаций

$$\Sigma_\varepsilon = \sigma^2 \cdot E \ . \tag{7.2}$$

В предыдущей главе мы несколько обобщили нашу модель, предполагая, что

$$\Sigma_\varepsilon = \sigma^2 \cdot \Omega \ , \tag{7.3}$$

где Ω есть известная симметричная положительно определенная матрица. Но в реальных задачах эта матрица Ω, как правило, неизвестна. Оценить ее по N измерениям невозможно, т. к. она содержит $N(N+1)/2$ неизвестных параметров. Чтобы задача стала решаемой, необходимо сделать те или иные предположения о матрице Ω. В этой главе мы рассмотрим два подхода к решению этой задачи, имеющие важные практические приложения.

7.1 Гетероскедастичность

7.1.1 Описание модели

В классической линейной модели ошибки измерений ε_j предполагаются независимыми и имеющими одинаковые дисперсии $\sigma^2 = M(\varepsilon_j^2)$ (*гомоскедастичность*). Пусть мы вновь имеем линейную модель (1), а относительно ошибок ε_j предположим, что они независимы, имеют нормальное распределение, $M(\varepsilon_j) = 0$, но дисперсии $\sigma_j^2 = M(\varepsilon_j^2)$ будут, вообще говоря, разными. В этом случае матрица ковариаций Σ_ε будет диагональной, но ее элементы различны. Именно этой случай и называется *гетероскедастичностью* модели.

Если $\sigma_j^2 = \sigma^2 \cdot \omega_j$, где $\sigma > 0$ есть неизвестный параметр, а ω_j – известные веса, то можно применить взвешенный МНК. Полученные таким образом оценки будут несмещенными, состоятельными и оптимальными в среднем квадратическом. Оценка для σ^2, построенная по остаткам e_j, будет несмещенной и состоятельной. Но в реальных задачах веса ω_j неизвестны.

Если мы применим в нашей ситуации обычный МНК, то оценки $\hat{\theta}$ для параметров θ останутся несмещенными и состоятельными, но уже не будут оптимальными. Еще хуже обстоит дело с оценкой дисперсии. Применяя обычную оценку для σ^2, мы оценим нечто среднее, т. к. все дисперсии различны и непонятно, что именно мы оцениваем. Обычные оценки для дисперсий случайного вектора $\hat{\theta}$ будут смещенными и несостоятельными. Поэтому основанные на них процедуры построения доверительных интервалов и проверки гипотез больше неприменимы и дают неверные результаты.

7.1.2 Модели гетероскедастичности

Ниже мы рассматриваем две модели гетероскедастичности, наиболее часто используемые на практике.

В первой модели мы предполагаем, что *ошибки пропорциональны одной из независимых переменных* X_l, т. е.

$$\sigma_j^2 = \sigma^2 X_{jl}^2 \tag{7.4}$$

для всех j. Фактически, мы находимся в ситуации, когда применим взвешенный МНК, т. к. веса $\omega_j = |X_{jl}|^2$ нам известны. Нужно перейти к новым переменным

$$Y_j^* = Y_j/|X_{jl}| \ , \ X_{jk}^*/|X_{jl}| \ , \ k = 0, 1, \ldots, m \ ,$$

и применить обычный МНК.

Существует простой *визуальный* метод, позволяющий *догадаться*, что мы находимся в описанной выше ситуации. Все наблюдения упорядочиваются в порядке возрастания (убывания) переменной $|X_l|$ (обычно $|X_l| = X_l$). Затем применяют обычный МНК. Находим остатки e_j и визуально исследуем их график. Если величина остатков увеличивается (уменьшается), то это подтверждает нашу гипотезу. Далее мы производим описанное выше преобразование и повторяем все заново, но с новыми переменными. Если вновь полученные остатки ведут себя нерегулярно, то это новое подтверждение нашей гипотезы о причинах гетероскедастичности модели.

Но столь простая ситуация встречается нечасто. Обычно требуется более сложная модель. Предположим, что

$$\sigma_j^2 = \gamma_0 + \gamma_1 Z_{j1} + \ldots + \gamma_p Z_{jp} \ , \tag{7.5}$$

где $Z_1, \ldots, Z_p$ есть некоторые экзогенные переменные (например, какие-то из наших объясняющих факторов). В этом случае оценка модели производится следующим образом.

1) Оцениваем исходную модель (1) по обычному МНК и находим остатки e_j.

2) Используя остатки e_j, оцениваем регрессию

$$e_j^2 = \gamma_0 + \gamma_1 Z_{j1} + \ldots + \gamma_p Z_{jp} + u_j \ , \qquad (7.6)$$

т. е. берем e_j^2 как аппроксимацию для σ_j^2. Далее, в качестве оценок для σ_j^2 берем предсказанные значения для e_j^2 в этой модели:

$$\hat{\sigma_j^2} = \hat{\gamma}_0 + \hat{\gamma}_1 Z_{j1} + \ldots + \hat{\gamma}_p Z_{jp} \ . \qquad (7.7)$$

3) Вновь оцениваем исходную модель (1), используя взвешенный МНК с весами $\omega_j = \hat{\sigma_j^2}$.

4) По новым остаткам оцениваем дисперсии оценок $\hat{\theta}$ параметром модели.

Такой метод оценивания называется *двухшаговым* МНК. Можно показать, что полученным таким образом оценки для дисперсий оценок параметров $\hat{\theta}$ будут состоятельными.

Отметим, что частными случаями этого метода являются описанный выше случай, когда ошибки пропорциональны одному из факторов, и *модель Глейзера*, где

$$\sigma_j^2 = \gamma_0 + \gamma_1 X_{jl}^{\alpha} \ . \qquad (7.8)$$

Здесь X_l – один из факторов, $0 < \alpha \leq 1$.

7.1.3 Коррекция на гетероскедастичность

Как мы отмечали выше, если к модели (1) в условиях гетероскедастичности применить обычный МНК, то оценка $\hat{\theta}$ для вектора параметров θ будет несмещенной и состоятельной. Но обычная оценка $S^2(X^TX)^{-1}$ для матрицы ковариаций с. в. $\hat{\theta}$ будет смещенной и, главное, несостоятельной. Чтобы исправить этот дефект, используют поправки на гетероскедастичность. Наиболее часто используют *стандартные ошибки в форме Уайта* (White, 1980). Уайт предложил следующий вариант оценки для матрицы ковариаций с. в. $\hat{\theta}$:

$$\hat{\Omega}_{\hat{\theta}} = N(X^T X)^{-1} \cdot \left(\frac{1}{N} \sum_{j=1}^{N} e_j^2 X_j X_j^T \right) (X^T X)^{-1} , \qquad (7.9)$$

где $X_j = (X_{j0}, X_{j1}, \ldots, X_{jm})$, $e_j = Y_j - \hat{Y}_j$. Уайт показал, что это состоятельная оценка для матрицы ковариаций оценки $\hat{\theta}$. Доказательство этого факта несколько сложнее и выходит за рамки данного пособия. Любознательные студенты могут обратиться к оригинальной работе Уайта.

Если ошибки ε_j коррелированы, но не более чем на L шагов, т. е. $Cov(\varepsilon_j, \varepsilon_k) = 0$, если $|j - k| > L$, то существует другой вариант коррекции на гетероскедастичность: *стандартные ошибки в форме Невье-Веста* (Newey-West, 1987). Мы не будем рассматривать их здесь. Отметим только, что вновь получается состоятельная оценка для матрицы ковариаций.

7.1.4 Тесты на гетероскедастичность

Если есть подозрение о том, что дисперсии ошибок $\sigma_j^2 = M(\varepsilon_j^2)$ будут различными в разных измерениях, то хотелось бы проверить это предположение с помощью некоторой формальной процедуры. При этом мы проверяем нулевую гипотезу

$$H_0 : \ \sigma_1^2 = \sigma_1^2 = \ldots = \sigma_N^2 \qquad (7.10)$$

против альтернативы, что дисперсии не равны. Каждый конкретный тест настроен на определенный тип альтернативы.

Тест Голдфельда-Куандта

Это один из наиболее ранних и популярных тестов (Goldfeld, Quandt, 1956). Он применяется тогда, когда есть подозрение, что ошибки измерений пропорциональны одной из независимых переменных X_l (см. выше). В этом случае альтернатива имеет вид:

$$H_1: \ \sigma_j^2 = \sigma^2 X_{jl}^2 \ . \qquad (7.11)$$

Проверка гипотезы H_0 против альтернативы H_1 вида (12) проводится следующим образом.

1) Все переменные упорядочиваются по возрастанию переменной $|X_l|$.

2) Из всех данных исключают средние d измерений (обычно $d \sim N/4$).

3) Строят две независимых регрессии по первым $(N - d)/2$ наблюдениям и последним $(N - d)/2$ наблюдениям и находим векторы остатков e_1 и e_2.

4) По остаткам e_1 и e_2 вычисляют суммы квадратов остатков $Q_1 = e_1^T e_1$ и $Q_1 = e_1^T e_1$ и статистику $F = Q_2/Q_1$.

5) Если верна гипотеза H_0, то статистика F имеет распределение Фишера с $(N - d)/2, (N - d)/2$ степенями свободы.

6) Для заданного α находим константу $F_\alpha > 0$:

$$P_0(F > F_\alpha) = \alpha \ .$$

7) Если реально полученное значение $F_{\text{Н}}$ будет больше F_α, то принимаем гипотезу H_1. В противном случае H_0 не противоречит экспериментальным данным.

Если будет принято решение о наличии гетероскедастичности (т. е. H_1), то необходимо произвести преобразование (5) и заново оценить параметры уравнения регрессии, но с помощью новых переменных. Доказательство сформулированных выше утверждений проводится аналогично тому, как это делалось выше при проверке линейных гипотез и, поэтому, предлагается читателю в качестве задачи.

Тест Бреуша-Пагана

Этот тест применяется в ситуации, когда для дисперсий ошибок используется модель

$$H_1 : \ \sigma_j^2 = \gamma_0 + \gamma_1 Z_{j1} + \ldots + \gamma_p Z_{jp} \ , \qquad (7.12)$$

где $Z_1, \ldots, Z_p$ – некоторые экзогенные переменные. Процедура проверки выглядит следующим образом.

1) Оцениваем исходную модель (1) и находим вектор остатков e.

2) Строим оценку для σ^2 по формуле

$$\hat{\sigma}^2 = \frac{1}{N} \sum_j e_j^2 \ .$$

3) Оцениваем регрессионное уравнение

$$\frac{e_j^2}{\hat{\sigma}^2} = \gamma_0 + \gamma_1 Z_{j1} + \ldots + \gamma_p Z_{jp} + u_j \ ,$$

и находим *для этой модели* сумму квадратов остатков ESS.

4) Если верна гипотеза H_0, то статистика $ESS/2$ имеет χ^2-распределение с $(p+1)$ степенью свободы.

Далее процедура проверки проводится стандартным образом.

Если принято решение о наличии гетероскедастичности, то для оценки модели (1) применяют двухшаговый МНК (см. выше).

Тест Уайта

Этот тест имеет более сложную структуру. Поэтому мы рассматриваем его только для случая $p = 2$, т.е. когда используется два фактора. Таким образом мы рассматриваем модель измерений

$$Y_j = \theta_0 + \theta_1 X_{j1} + \theta_2 X_{j2} + \varepsilon_j \ .$$

В качестве альтернативы рассматривается следующая модель для дисперсий

$$H_1 : \ \gamma_0 + \gamma_1 X_{j1} + \gamma_2 X_{j2} + \gamma_{11} X_{j1}^2 + \gamma_{22} X_{j2}^2 + \gamma_{12} X_{j1} X_{j2} \ . \quad (7.13)$$

Процедура проверки проводится следующим образом.

1) Применяем к основной модели обычный МНК и находим вектор остатков e.

2) Оцениваем регрессию

$$e_j^2 = \gamma_0 + \gamma_1 X_{j1} + \gamma_2 X_{j2} + \gamma_{11} X_{j1}^2 + \gamma_{22} X_{j2}^2 + \gamma_{12} X_{j1} X_{j2} + u_j \ .$$

3) Для последней регрессии вычисляем выборочный коэффициент детерминации R^2.

4) При верной гипотезе H_0 величина $N \cdot R^2$ имеет χ^2-распределение с 5 степенями свободы.

Далее процедура проверки проводится стандартным образом.

Если принято решение о наличии гетероскедастичности, то в данном случае, в отличие от предыдущих двух тестов, неизвестно преобразование переменных, после применения которого обычный МНК дает хорошие оценки. Обычно в этом случае проводят коррекцию ошибок в форме Уайта.

7.2 Автокорреляция ошибок

7.2.1 Постановка задачи

Во многих экономических задачах, особенно при анализе временных рядов, ошибки измерений оказываются коррелированными. Для этого есть несколько причин:

1) влияние на ошибки факторов, не включенных в модель (неправильная спецификация модели),

2) реальная корреляция ошибок.

В общем случае задача оценивания уравнения регрессии в присутствии корреляции ошибок является достаточно сложной. Поэтому мы рассматриваем следующий частный случай. Пусть, как и ранее, мы рассматриваем линейную модель

$$Y = X\theta + \varepsilon \ ,$$

но теперь ошибки измерений удовлетворяют соотношению

$$\varepsilon_j = \rho \cdot \varepsilon_{j-1} + \delta_j \ , \ |\rho| < 1 \ , \tag{7.14}$$

где с. в. δ_j являются независимыми и одинаково распределенными, $M(\delta_j) = 0$, $D(\delta_j) = \sigma^2$. Модель такого типа называется *авторегрессией первого порядка*. Предположим, что с.в. ε_0 независима от последовательности $\{\delta_j\}$, $M(\varepsilon_0) = 0$ и $D(\varepsilon_0) = \sigma^2/(1-\rho^2)$. Используя соотношение (14), нетрудно показать,что

$$M(\varepsilon_j) = 0 \ , \ D(\varepsilon_j) = \sigma_\varepsilon^2 = \frac{\sigma^2}{1-\rho^2} \ , \ Cov(\varepsilon_j, \varepsilon_{j+m}) = \rho^m \cdot \sigma_\varepsilon^2 \ . \tag{7.15}$$

Последовательность случайных величин $\{\varepsilon_j\}$, удовлетворяющая соотношения (15), называется *стационарной в широком смысле*. Таким образом, для ошибок измерений мы получам следующую матрицу ковариаций:

$$\Sigma_\varepsilon = M(\varepsilon\varepsilon^T) = \sigma_\varepsilon^2 \cdot \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{N-1} \\ \rho & 1 & \rho & \dots & \rho^{N-2} \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \rho^{N-1} & \rho^{N-2} & \rho^{N-3} & \dots & 1 \end{pmatrix} = \sigma_\varepsilon^2 \cdot \Omega(\rho)\ . \qquad (7.16)$$

Если в этой ситуации применить обычный МНК, то оценки для параметров θ будут несмещенными и состоятельными, но уже не будут оптимальными. Но самое главная оценка для дисперсии σ_ε^2 окажется смещенной вниз и, как следствие, несостоятельной. Это приведет к ошибкам при построении доверительных интервалов и проверке гипотез о параметрах.

Если величина параметра ρ известна, то можно применить обобщенный МНК или обычный МНК к новой модели

$$Y_j^* = Y_j - \rho Y_{j-1} = \theta_0(1-\rho) + \theta_1(X_{j1} - \rho X_{j-1,1}) + \dots + \theta_m(X_{jm} - \rho X_{j-1,m}) + \delta_j\ , \qquad (7.17)$$

где $j = 2, \dots, N$. В качестве начального значения берут

$$Y_1^* = \sqrt{1-\rho^2}Y_1 = \sqrt{1-\rho^2}\theta_0 + \sqrt{1-\rho^2}\theta_1 X_{j1} + \dots + \sqrt{1-\rho^2}\theta_m X_{jm} + \sqrt{1-\rho^2}\varepsilon_1 \qquad (7.18)$$

Для этой модели выполнены все основные предположения классической линейной модели и применим обычный МНК. Преобразование (17) называется *авторегрессионным преобразованием* исходной модели. Уравнение (18) гарантирует стационарность ошибок.

Но в реальных задачах, как правило, величина ρ неизвестна. Тогда для оценки параметров нужно применять более сложные методы.

7.2.2 Оценивание в модели с авторегрессией

Существует несколько методов оценивания исходной модели (1) в присутствии автокорреляции при незвестном ρ. Мы приведем краткое описание трех из них, которые наиболее часто используются в эконометрической практике.

Процедура Кохрейна-Оркатта (Cochrane-Orcutt)

1) Применяем обычный МНК к исходной модели (1) и находим вектор остатков e.

2) Рассматриваем модель регрессии

$$e_j = \rho e_{j-1} + u_j \ , \ j \geq 2 \ ,$$

и находим оценку r параметра ρ по обычному МНК.

3) Подставляем эту оценку в авторегрессионное преобразование (17-18) и оцениваем остальные параметры в модели (17-18) по обычному МНК.

4) Находим новый вектор остатков $e = Y - X\hat{\theta}$, где $\hat{\theta}$ – оценка параметров θ, полученная на предыдущем шаге.

5) Далее повторяем шаги 2)-4) пока не будет достигнута нужная точность при оценке ρ. Можно также фиксировать число итераций.

Процедура Хилдрета-Лу (Hildreth-Lu)

1) Отрезок $(-1, 1)$ разбивается на мелкую сетку возможных значений ρ.

2) Для каждого значения ρ производится авторегрессионное преобразование (17-18).

3) Проводится оценка параметров пребразованной системы (при фиксированном ρ).

4) Выбирается то значение ρ, при котором остаточная сумма квадратов будет минимальной.

5) В окрестности найденного ρ берется более мелькая сетка и процедура повторяется.

Процедура заканчивается, когда будет достигнута нужная точность или проведено заданное число итераций.

Процедура Дарбина (Durbin)

1) Пребразованное уравнение (17) записывается в виде

$$\begin{aligned} Y_j &= \theta_0(1-\rho) + \rho Y_{j-1} + \theta_1 X_{j1} - \rho\theta_1 X_{j-1,1}) + \ldots \\ &\quad +\theta_m X_{jm} - \rho\theta_m X_{j-1,m}) + \delta_j \;, \end{aligned} \tag{7.19}$$

т. е. Y_{j-1} включают в число объясняющих факторов.

2) Последнее уравнение оценивают по обычному МНК и находят оценки r и $\hat{\gamma}$ для параметров ρ и $\gamma = \rho\theta_k$. В качестве оценки для параметра θ_k берут $\hat{\theta_k} = \hat{\gamma}/r$.

Можно улучшить оценки $\hat{\theta_k}$, если подставить найденное значение r в уравнения (17-18) и вновь оценить их по обычному МНК.

Замечание. В принципе мы могли бы применить обычный МНК к модели (17-18), рассматривая в качестве искомых параметры $\rho, \theta_0, \ldots, \theta_m$. Но параметры входят в модель нелинейно и мы имеем более сложную задачу нелинейной программирования. Существуют различные методы решения задач такого типа. Описанные выше процедуры есть три примера таких методов.

Т. к. полученные оценки являются, по-существу, нелинейными, то исследование их свойств более сложная задача и здесь обсуждаться не будет.

7.2.3 Критерий Дарбина-Уотсона

Хотелось бы иметь критерий, с помощью которого можно проверить наличие или отсутствие автокорреляции в ошибках измерений. Все критерии такого типа основаны на предположении, что если корреляции присутствует в ошибках, то она должна проявляться и в остатках, полученным после оценки модели.

Рассмотрим, как и ранее, следующую модель:

$$Y = X\theta + \varepsilon \ ,$$

причем

$$\varepsilon_j = \rho\varepsilon_{j-1} + \delta_j \ ,$$

где δ_j – н. о. р. с. в., $M(\delta_j) = 0$, $D(\delta_j) = \sigma^2$, ε_0 не зависит от $\{\delta_j\}$, $M(\varepsilon_0) = 0$, $D(\varepsilon_0) = \sigma^2/(1-\rho^2)$.

Мы хотим проверить гипотезу

$$H_0: \ \rho = 0$$

(– нет корреляции) против альтернативы

$$H_1: \ \rho > 0$$

(– есть положительная корреляция). Дарбин и Уотсон (Durbin, Watson 1951) предложили использовать следующую статистику для проверки этих гипотез:

$$DW = \frac{\sum\limits_{j=2}^{N}(e_j - e_{j-1})^2}{\sum\limits_{j=1}^{N} e_j^2} \ , \tag{7.20}$$

где $e_j = Y_j - \hat{Y}_j$, $\hat{Y} = X\hat{\theta}$. Посмотрим, как ведет себя эта статистика при больших N:

$$\begin{aligned} DW &= \frac{\sum\limits_{j=2}^{N} e_j^2 + \sum\limits_{j=2}^{N} e_{j-1}^2 - 2\sum\limits_{j=2}^{N} e_j e_{j-1}}{\sum\limits_{j=1}^{N} e_j^2} = \frac{2\sum\limits_{j=1}^{N} e_j^2 - e_1^2 - e_N^2 - 2\sum\limits_{j=2}^{N} e_j e_{j-1}}{\sum\limits_{j=1}^{N} e_j^2} \\ &= 2\left[1 - \frac{\sum\limits_{j=2}^{N} e_j e_{j-1}}{\sum\limits_{j=1}^{N} e_j^2}\right] - \frac{e_1 + e_N}{\sum\limits_{j=1}^{N} e_j^2} \approx 2(1-r) \ . \end{aligned}$$

Т. о., если верна H_0, то $DW \approx 2$, а при гипотезе H_1 DW должна быть меньше 2.

Для проверки H_0 против альтернативы H_1 необходимо найти критическую константу $d^* = d^*(\alpha)$:

$$P(DW < d^*|H_0) = \alpha \ .$$

Но, к сожалению, распределение статистики DW, а значит и величина d^*, зависит не только от N и α, но и значений факторов X_k, которые в каждой конкретной задаче свои. Дарбин и Уотсон показали, что существуют две границы d_l и d_u такие, что

1) если $DW < d_l$, то $DW < d^*$ и можно принять H_1,

2) если $DW > d_u$, то $DW > d^*$ и можно принять $H_)$.

Но есть зона неопределенности(d_l, d_u), где мы не знаем какое принять решение.

Эти границы зависят только от N, m и α и могут быть затабулированы. Например, для $N = 100$, $m = 1$ и $\alpha = 0.05$ мы имеем $d_l = 1.65$, $d_u = 1.69$.

Глава 8

Задача прогноза в линейных системах

8.1 Постановка задачи

Одной из целей построения эконометрической модели является использование ее, в дальнейшем, для оценки изучаемого экономического показателя Y при некотором сочетании факторов $X_1, \ldots, X_m$, т.е. решение *задачи прогноза.*

Уточним стоящую перед нами задачу. Пусть мы имеем N измерений

$$Y_j = \theta_0 + \theta_1 X_{j1} + \ldots + \theta_m X_{jm} + \varepsilon_j \ , \ j = \overline{1, N}, \qquad (8.1)$$

интересующего нас показателя Y при определенных сочетаниях факторов, причем выполнены основные ограничения 1-6 классической линейной модели. Мы прогнозируем, что в некоторой новой ситуации ("в будущем") факторы примут значения $X_{N+1,1}, \ldots, X_{N+1,m}$. Необходимо построить оценку для нового значения Y_{N+1}. Существуют две разновидности этой задачи:

1) оценить среднее значение $M(Y_{N+1})$,

2) оценить саму величину Y_{N+1}.
Кроме того, нужно не только построить оценку, т.е. сделать прогноз, но и оценить точность этого прогноза, т.е. построить доверительный интервал.

8.2 Оценка среднего значения

В этом случае оценке подлежит числовой параметр

$$a_{N+1} = M(Y_{N+1}) = \theta_0 + \theta_1 X_{N+1,1} + \ldots + \theta_m X_{N+1,m} \ . \quad (8.2)$$

Решение почти очевидно. Используя обычный метод МНК, находим оценку $\hat{\theta}$ вектора параметров θ. Обозначим через X_j вектор строку $(1, X_{j1}, \ldots, X_{jm})$. В качестве оценки величины a_{N+1} предлагается взять

$$\hat{a}_{N+1} = X_{N+1} \cdot \hat{\theta} \ . \quad (8.3)$$

Отметим, что так как $\hat{\theta}$ есть линейная по Y оценка, то $\hat{a}_{N+1}$ также является линейной и несмещенной оценкой для a_{N+1}. Аналогично тому, как это было сделано ранее при оценке параметра θ, можно доказать следующий результат.

Теорема 1 . *В классе всех линейных по Y и несмещенных оценок оценка $\hat{a}_{N+1}$, построенная по формуле (3), является оптимальной в среднем квадратическом, т.е. обладает наименьшей дисперсией.*

Оценим точность построенного прогноза. Ранее было доказано, что случайный вектор $\hat{\theta}$ имеет многомерное нормальное распределение со средним θ и матрицей ковариаций $\sigma^2 \cdot (X^T X)^{-1}$. Оценка $\hat{a}_{N+1} = X_{N+1} \cdot \hat{\theta}$ есть линейное преобразование случайного вектора $\hat{\theta}$. Поэтому, эта случайная величина имеет нормальное распределение со средним $X_{N+1}\theta$ и дисперсией $\sigma^2 \cdot X_{N+1} \cdot (X^T X)^{-1} X_{N+1}^T$. В качестве оценки дисперсии σ^2 вновь используем величину

$$S^2 = \frac{1}{N-(m+1)} \sum_{j=1}^{N} [Y_j - \hat{Y}_j]^2 \ , \tag{8.4}$$

где $\hat{Y}_j = X_j \cdot \hat{\theta}$. Случайные величины $\hat{a}$ и S^2 независимы и $(N-(m+1))S^2/\sigma^2$ имеет χ^2-распределение с $(N-(m+1))$ степенями свободы. Отсюда получаем, что случайная величина

$$T = \frac{\hat{a}_{N+1} - a_{N+1}}{\sigma\sqrt{X_{N+1}(X^T X)^{-1} X_{N+1}^T}} : \frac{S}{\sigma} \tag{8.5}$$

имеет распределение Стьюдента с $(N-(m+1))$ степенями свободы. По заданному уровню доверия γ по таблицам распределения Стьюдента находим такую константу $t(\gamma) > 0$, чтобы

$$P(|T| < t(\gamma)) = \gamma \ . \tag{8.6}$$

Окончательно для a_{N+1} получаем следующий доверительный интервал:

$$\hat{a}_{N+1} - t(\gamma) \cdot S \cdot \sqrt{X_{N+1}(X^T X)^{-1} X_{N+1}^T} < a_{N+1} < \\ \hat{a}_{N+1} + t(\gamma) \cdot S \cdot \sqrt{X_{N+1}(X^T X)^{-1} X_{N+1}^T} \ . \tag{8.7}$$

8.3 Оценка значения прогнозируемой величины

В этом случае нам нужно оценить не $a_{N+1} = M(Y_{N+1})$, а само будущее значение $Y_{N+1} = a_{N+1} + \varepsilon_{N+1}$. В качестве оценки предлагается вновь взять $\hat{Y}_{N+1} = \hat{a}_{N+1} = X_{N+1} \cdot \hat{\theta}$. Но теперь мы оцениваем не a_{N+1}, а Y_{N+1}. Поэтому нас интересует разность

$$\hat{Y}_{N+1} - Y_{N+1} = x_{N+1} \cdot \hat{\theta} - a_{N+1} - \varepsilon_{N+1} \ . \qquad (8.8)$$

Отметим, что $\hat{Y}_{N+1} - Y_{N+1}$ имеет среднее, равное нулю, (аналог несмещенности) и $\hat{Y}_{N+1}$ есть линейная по Y оценка.

Теорема 2 . *В классе всех линейных и несмещенных оценок $\tilde{Y}$ для величины Y_{N+1} оценка $\hat{Y}_{N+1}$, построенная по формуле (3), является оптимальной, т.е.*

$$M(\hat{Y}_{N+1} - Y_{N+1})^2 \leq M(\tilde{Y}_{N+1} - Y_{N+1})^2 \ . \qquad (8.9)$$

Для построения доверительного интервала нужно дословно повторить все рассуждения, приведенные в пункте 2, но для разности $\hat{Y}_{N+1} - Y_{N+1}$ вместо $\hat{a}_{N+1} - a_{N+1}$. Едиственное различие состоит в том, что $D(\hat{Y}_{N+1} - Y_{N+1}) = \sigma^2 \cdot (X_{N+1} \cdot (X^T X)^{-1} \cdot X_{N+1}^T + 1)$ за счет слагаемого $-\varepsilon_{N+1}$. Таким образом, доверительный интервал для Y_{N+1} имеет вид:

$$\hat{Y}_{N+1} - t(\gamma) \cdot S \cdot \sqrt{X_{N+1}(X^T X)^{-1} X_{N+1}^T + 1} < Y_{N+1} <$$
$$\hat{Y}_{N+1} + t(\gamma) \cdot S \cdot \sqrt{X_{N+1}(X^T X)^{-1} X_{N+1}^T + 1} \ (8.10)$$

Мы рассмотрели только простейшую ситуацию классической линейной модели. В реальных задачах нужно учитывать то, что факторы могут быть стохастическими, а также присутствие гетероскедастичности и автокорреляции ошибок.

Дополнение A. Гильбертово пространство случайных величин

В этом разделе мы выделим некоторое подмножество случайных величин и покажем как в нем можно ввести геометрическую структуру. Далее мы решим важную с точки зрения приложений задачу о наилучшей линейной оценке.

Задача о наилучшем линейном приближении в конечномерном пространстве

Для лучшего понимания задач, рассматриваемых в этом разделе, мы начнем с задачи о наилучшем линейном приближении в конечномерном пространстве.

Пусть мы имеем конечномерное линейное пространство R^N. Мы предполагаем, что в этом пространстве выделен некоторый базис и любой элемент $x \in R^N$ можно записать в виде $x = (x_1, \ldots, x_N)^T$. Далее, в этом пространстве задается обычным образом *скалярное произведение*, т.е. любым двум элементам $x, y \in R^N$ приписывается число (x, y) по правилу:

$$(x, y) := \sum_{j=1}^{N} x_j \cdot y_j \ .$$

Это позволяет нам определить *длину* элемента $x \in R^N$:

$$\|x\| := \sqrt{(x, x)} = \sqrt{\sum_{j=1}^{N} x_j^2} \ ,$$

и расстояние между элементами $x, y \in R^N$:

$$\rho(x, y) := \|x - y\| \ .$$

Таким образом, R^N становится *метрическим пространством*, в котором можно определить понятие сходимости, открытые и замкнутые множества и т.п. Кроме того, определяется важное для нас понятие *ортогональности*. А именно, два элемента $x, y \in R^N$ будем называть *ортогональными*, если $(x, y) = 0$.

Пусть теперь в R^N выделено некоторое замкнутое подмножество L и некоторый элемент $y \in R^N$. Будем говорить, что элемент $\hat{y} \in R^N$ является *наилучшим приближением элемента y во множестве* L, если:

1) $\hat{y} \in L$,

2) $\|y - \hat{y}\| \le \|y - x\|$ для любого $x \in L$.

В силу замкнутости L такое наилучшее приближение существует.

Наиболее интересным для нас является случай, когда L есть линейное подпространство в R^N. В этой ситуации справедлива следующая важная

Лемма о перпендикуляре. *Элемент $\hat{y}$ является наилучшим приближением элемента y в линейном пространстве L тогда и только тогда, когда*

1) $\hat{y} \in L$,

2) $(y - \hat{y}, x) = 0$ *для любого* $x \in L$.

Отсюда следует, что *ошибка приближения* $e = y - \hat{y}$ есть вектор, ортогональный ко всем элементам пространства L, а само наилучшее приближение $\hat{y}$ есть проекция вектора y на подпространство L. Эта геометрическая интерпретация поможет нам далее более наглядно представлять себе полученные результаты.

Гильбертово пространство случайных величин

Расмотрим множество L_2, состоящее из случайных величин ξ таких, что $M(|\xi|^2) < \infty$. Используя неравенство Коши-Буняковского, нетрудно показать, что L_2 это линейное пространство (задача!). Для каждой пары $\xi, \eta \in L_2$ определим число

$$(\xi, \eta) := M(\xi \cdot \eta) \ .$$

Будем говорить, что случайные величины ξ и η совпадают **почти наверное** (п.н.), если $P(\xi \neq \eta) = 0$. В дальнейшем мы не будем различать такие случайные величины и объединим их в один класс. Таким образом, мы рассматриваем L_2 как совокупность классов эквивалентности.

Задача. Введенный выше функционал обладает следующими свойствами:

a) $(\xi, \xi) \geq 0, \ (\xi, \xi) = 0 \Leftrightarrow \xi = 0$ п.н.,

b) $\forall \xi, \eta \in L_2, \ (\xi, \eta) = (\eta, \xi)$,

c) $\forall \xi_1, \xi_2, \eta \in L_2, \ c_1, \ c_2 \in R^1 \ (c_1\xi_1 + c_2\xi_2, \eta) = c_1(\xi_1, \eta) + c_2(\xi_2, \eta)$.

Функционал, который приписывает каждой паре элементов из линейного пространства некоторое число так, что выполнены свойства a)-c), называется **скалярным произведением**. Используя скалярное произведение, можно определить **норму** произвольного элемента $\xi \in L_2$:

$$\|\xi\| := \sqrt{(\xi, \xi)} = (M(|\xi|^2))^{1/2} .$$

Норма элемента ξ определяет его длину. Будем говорить, что последовательность с.в. $\{\xi_n\}$ **сходится в среднем квадратическом** (с.к.) к с.в. ξ, если

$$\|\xi_n - \xi\|^2 = M(|\xi_n - \xi|^2) \to 0,\ n \to \infty .$$

В этом случае мы будем писать $\xi_n \stackrel{..}{\to} \xi$. Можно показать, что пространство L_2 полно относительно такой сходимости, т.е. всякая фундаментальная последовательность в L_2 имеет предел. Линейное пространство, в котором определено скалярное произведение и которое полно относительно порожденной этим скалярным произведением сходимости, называется **гильбертовым пространством**. Таким образом, мы показали, что L_2 – гильбертово пространство.

Выше мы отмечали, что норма вектора определяет его длину. Но в геометрии нам необходимо измерять не только длины, но и углы. Используя аналогию с понятием скалярного произведения в планеметрии, определим косинус угла α между элементами $\xi, \eta \in L_2$ по правилу

$$\cos\alpha = \frac{M(\xi\eta)}{\sqrt{M|\xi|^2 \cdot M|\eta|^2}} = \frac{(\xi, \eta)}{\|\xi\| \cdot \|\eta\|} .$$

Назовем ξ и η **ортогональными**, если $(\xi, \eta) = M(\xi\eta) = 0$.

Замечание. Пусть $M\xi = M\eta = 0$. Тогда имеет место следующее:

1) $M(\xi\eta) = cov(\xi, \eta) = 0$, т.е. понятие ортогональности совпадает с понятием некоррелированности;

2) $\cos\alpha = \rho(\xi, \eta)$, т.е. косинус угла между случайными величинами совпадает с коэффициентом корреляции.

Задача о наилучшей линейной оценке

Далее, используя введенные выше понятия, мы решим очень важную с практической точки зрения задачу. Пусть $\mathcal{L} \subset L_2$ – некоторое линейное подпространство, которое замкнуто относительно сходимости в среднем квадратическом. Пусть далее $\eta \in L_2$ – произвольный элемент.

Определение 1 . *Случайная величина $\hat{\eta} \in L_2$ называется* **наилучшим приближением с.в. η в пространстве $\mathcal{L}$**, *если*

1) $\hat{\eta} \in \mathcal{L}$,

2) $\|\eta - \hat{\eta}\|^2 = M|\eta - \hat{\eta}|^2 \leq M|\eta - \xi|^2 = \|\eta - \xi\|^2$, $\forall \xi \in \mathcal{L}$.

Лемма о перпендикуляре. *С.в. $\hat{\eta}$ является наилучшим приближением с.в. η в линейном пространстве $\mathcal{L}$ тогда и только тогда, когда*

1) $\hat{\eta} \in \mathcal{L}$,

2) $(\eta - \hat{\eta}, \xi) = M[(\eta - \hat{\eta})\xi] = 0, \ \forall \xi \in \mathcal{L}$.

Доказательство. Пусть выполнено свойство 2 и $\xi \in \mathcal{L}$. Тогда

$$\begin{aligned}\|\eta - \xi\|^2 = M|\eta - \xi|^2 = M|(\eta - \hat{\eta}) + (\hat{\eta} - \xi)|^2 = \\ = M|\eta - \hat{\eta}|^2 + M|\hat{\eta} - \xi|^2 + M[(\eta - \hat{\eta})(\hat{\eta} - \xi)] = \\ = M|\eta - \hat{\eta}|^2 + M|\hat{\eta} - \xi|^2 + 0 \geq M|\eta - \hat{\eta}|^2 \ .\end{aligned}$$

Здесь мы использовали тот факт, что $\hat{\eta} - \xi \in \mathcal{L}$.

Обратно, пусть мы знаем, что $\hat{\eta}$ наилучшее приближение для η в $\mathcal{L}$. Возьмем произвольные $\xi \in \mathcal{L}$ и $\lambda \in R^1$. В силу оптимальности $\hat{\eta}$ получаем

$$M|\eta - \hat{\eta}|^2 \leq M|\eta - \hat{\eta} + \lambda\xi|^2 =$$

$$M|\eta-\hat{\eta}|^2+2M[(\eta-\hat{\eta})\xi]\cdot\lambda+M|\xi|^2\cdot\lambda^2 \ .$$

Последнее выражение есть полином второго порядка от λ, у которого коэффициент при λ^2 положительный и который принимает минимальное значение при $\lambda = 0$. Из курса математического анализа мы знаем, что в этом случае коэффициент при λ, т.е. $M[(\eta-\hat{\eta})\xi]$ равен нулю. А это и есть условие 2.

Дадим геометрическую интерпретацию полученному результату. Свойство 2 означает, что $\eta-\hat{\eta}$ является **перпендикуляром** к подпространству $\mathcal{L}$ (эквивалентно: $\hat{\eta}$ – **проекция** η на $\mathcal{L}$).

Иногда с.в. $\hat{\eta}$ называют **условным математическим ожиданием в широком смысле** с.в. η относительно подпространства $\mathcal{L}$.

Рассмотрим несколько более общую ситуацию. Пусть $\mathcal{L} = \xi_0 + \mathcal{L}_0$, где $\xi_0 \in L_2$ – фиксированный элемент, а $\mathcal{L}_0$ – замкнутое линейное подпространство в L_2. Если $\xi_0 \neq 0$, то $\mathcal{L}$ не является линейным пространством и называется **гиперплоскостью**. Пусть далее $\eta \in L_2$ – некоторая с.в. Определение наилучшего приближения $\hat{\eta}$ для с.в. η в гиперплоскости $\mathcal{L}$ дословно повторяет определение 1. Нетрудно доказать следующий результат.

Задача. $\hat{\eta}$ является наилучшим приближением для η в гиперплоскости $\mathcal{L}$, если

1) $\hat{\eta} \in \mathcal{L}$,

2) $(\eta-\hat{\eta}, \xi-\hat{\eta}) = M[(\eta-\hat{\eta}, \xi-\hat{\eta})] = 0$, $\forall \xi \in \mathcal{L}$.

Применим эти результаты к следующей задаче. Пусть $\xi, \eta \in L_2$–две случайные величины. Необходимо найти наилучшее линейное приближение с.в. η с помощью случайной величины ξ. В этом случае $\mathcal{L} = \{c_1 + c_2\xi \ , \ c_1, c_2 \in R^1\}$. Это двумерное линейное подпространство в L_2, базисом которого являются с.в. $\xi_1 = 1$ п.н. и $\xi_2 = \xi$. Наилучшее приближение должно иметь вид $\hat{\eta} = \hat{c}_1 + \hat{c}_2\xi$, т.е. мы должны найти $\hat{c}_1$ и $\hat{c}_2$. По лемме о перпендикуляре $(\eta - \hat{\eta}, \theta) = 0,\ \forall\theta \in \mathcal{L}$. В частности,

$$(\eta - \hat{\eta}, \xi_1) = M[(\eta - \hat{\eta}) \cdot 1] = M\eta - \hat{c}_1 - \hat{c}_2 M\xi = 0 \ ,$$

$$(\eta - \hat{\eta}, \xi_2) = M[(\eta - \hat{\eta}) \cdot \xi] = M(\eta\xi) - \hat{c}_1 M\xi - \hat{c}_2 M(\xi^2) = 0 \ .$$

Решая эти уравнения относительно $\hat{c}_1$ и $\hat{c}_2$, получим:

$$\hat{c}_2 = \frac{cov(\xi, \eta)}{D(\xi)} = \rho(\xi, \eta) \cdot \frac{\sigma_\eta}{\sigma_\xi} \ ,$$

$$\hat{c}_1 = M\eta - \rho(\xi, \eta) \cdot \frac{\sigma_\eta}{\sigma_\xi} \cdot M\xi \ ,$$

где $\sigma_\eta^2 = D(\eta)$, $\sigma_\eta^2 = D(\eta)$. Таким образом, окончательно получаем

$$\hat{\eta} = M\eta + \rho(\xi, \eta) \cdot \frac{\sigma_\eta}{\sigma_\xi} \cdot (\xi - M\xi) \ .$$

Дополнение B. Многомерное нормальное распределение

Нормальное распределение играет важную роль в теории вероятностей и математической статистике. В этом приложении мы дадим определение и перечислим основные свойства многомерного нормального распределения.

Определение 1 . *Случайный вектор $\xi = (\xi_1, \ldots, \xi_n)^T$ имеет* **многомерное нормальное распределение**, *если он обладает плотностью распределения следующего вида:*

$$\rho_\xi(x_1, \ldots, x_n) = (2\pi)^{-\frac{n}{2}} det(A) \exp\{-\frac{1}{2}(A(x-m), x-m)\} =$$

$$(2\pi)^{-\frac{n}{2}} det(A) \exp\{-\frac{1}{2}\sum_{i,j=1}^{n} a_{ij}(x_i - m_i)(x_j - m_j)\} \ ,$$

где $m = (m_1, \ldots, m_n)^T \in R^n$, $A = (a_{ij})$ – симметричная положительно определенная $n \times n$-матрица.

В дальнейшем для случайного вектора ξ, который имеет многомерное нормальное распределение с параметрами (m, A), мы будем использовать обозначение $\xi \sim N(m, \Sigma)$, где $\Sigma = A^{-1}$.

Многие свойства многомерного нормального распределения легче получить, используя его характеристическую функцию. Напомним, что *характеристической функцией* случайного вектора $\xi = (\xi_1, \ldots, \xi_n)^T$ называется комплекснозначная функция $\varphi_\xi(t)$, заданная в R^n по правилу

$$\varphi_\xi(t) = M\left[e^{i\cdot(\xi,t)}\right] ,$$

$t = (t_1, \ldots, t_n)^T \in R^n$. Можно показать, что для случайного вектора $\xi \sim N(m, \Sigma)$ его характеристическая функция имеет вид

$$\begin{aligned}\varphi_\xi(t) &= \exp\{i\cdot(m,t) - \frac{1}{2}(\Sigma t, t)\} \\ &= \exp\{i\cdot\sum_{j=1}^{n} m_j \cdot t_j - \frac{1}{2}\sum_{p,q=1}^{n} \sigma_{pq}\cdot t_p \cdot t_q\} .\end{aligned}$$

Выясним смысл параметров многомерного нормального распределения. Для этого вычислим математические ожидания и ковариации компонент вектора ξ. Хорошо известно, что начальные моменты случайного вектора можно вычислить, используя характеристическую функцию. В частности,

$$\frac{\partial}{\partial t_j}\varphi_\xi(t)\,|_{t=0} = i\cdot M(\xi_j) ,$$

$$\frac{\partial^2}{\partial t_p \partial t_q}\varphi_\xi(t)\,|_{t=0} = -M(\xi_p\cdot\xi_q) .$$

Отсюда для многомерного нормального распределения с параметрами (m, Σ) прямым вычислением легко получаем

$$M(\xi_j) = m_j\ ,\ Cov(\xi_p, \xi_q) = \sigma_{pq} .$$

Таким образом, m есть *вектор средних*, а Σ – *матрица ковариаций.*

Будем называть многомерное нормальное распределение в R^n *стандартным*, если оно имеет нулевые средние и единичную матрицу ковариаций $\Sigma = I_n$, т.е. компоненты вектора ξ некоррелированы (на самом деле, независимы) и имеют дисперсии равные единице.

Многомерное распределение в R^n будем называть *невырожденным*, если его матрица ковариаций Σ является невырожденной. В противном случае распределение называется *вырожденным*. Данное выше определение многомерного нормального распределения относилось именно к невырожденному случаю. Если распределение вырождено, то существуют такие линейное подпространство L в R^n и неслучайный вектор $a \in R^n$, что распределение вектора ξ сосредоточено на множестве $L + a$, т.е. случайный вектор ξ с вероятностью 1 принадлежит этому множеству. Размерность пространства L совпадает с рангом матрицы Σ. В этом случае определение 1 справедливо только в координатах соответствующего подпространтсва. А вот формула (2) для характеристической функции справедлива и в вырожденном случае.

Рассмотрим несколько основных свойств многомерного нормального распределения, которые систематически будут использоваться в нашем курсе.

Теорема 1 (B1). *Если $C : R^n \to R^k$–линейное отображение R^n на R^k, $\xi \sim N(m, \Sigma)$, то $C\xi \sim N(Cm, C\Sigma C^T)$.*

Доказательство. Вычислим характеристическую функцию случайного вектора $C\xi$.

$$\varphi_{C\xi}(t) = M(\exp\{i(C\xi, t)\}) = M(\exp\{i(\xi, C^T t)\}) =$$
$$= \exp\{i(m, C^T t) - \frac{1}{2}(\Sigma C^T t, C^T t)\} =$$
$$= \exp\{i(Cm, t) - \frac{1}{2}((C\Sigma C^T)t, t)\} \ .$$

Последнее выражение есть характеристическая функция многомерного нормального распределения со средним Cm и ковариационной матрицей $C\Sigma C^T$. Теорема 1 доказана.

Следствие 1 . *Пусть случайный вектор ξ имеет стандартное нормальное распределение в R^n, а C есть ортогональная проекция R^n на R^k, $k < n$. Тогда случайный вектор $\eta = C\xi$ имеет стандартное нормальное распределение в R^k.*

Доказательство. В силу теоремы 1 случайный вектор η имеет многомерное нормальное распределение в R^k со средним $C \cdot 0 = 0$ и матрицей ковариаций $\Sigma_\eta = C \cdot I_n \cdot C^T = C \cdot C^T = I_k$, т.е. имеет стандартное нормальное распределение.

Разобьем случайный вектор ξ на два подвектора размерности n_1 и n_2: $n_1 + n_2 = n$, т.е. $\xi = (\xi', \xi'')$. Это порождает разложения вектора средних $m = (m', m'')$ и матрицы ковариаций

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} .$$

Теорема 2 *(***(B2)** . *Если $\xi = (\xi_1, \ldots, \xi_n)^T \sim N(m, \Sigma)$, то*

1) $\xi' \sim N(m', \Sigma_{11})$,

2) ξ' и ξ''–независимы тогда и только тогда, когда

$$\Sigma_{12} = \Sigma_{21}^T = 0.$$

Доказательство. Пусть $t = (t_1, \ldots, t_{n_1}, 0, \ldots, 0)^T = ((t')^T, 0, \ldots, 0)^T$. Тогда

$$\varphi_{\xi'}(t') = \varphi_\xi(t) = \exp\{i(m, t) - \frac{1}{2}(\Sigma t, t)\} =$$

$$= \exp\{i(m', t') - \frac{1}{2}(\Sigma_{11} t', t')\} .$$

Последнее выражение есть характеристическая функция многомерного нормального распределения со средним m' и матрицей ковариаций Σ_{11}.

Пусть теперь $t = (t_1, \ldots, t_{n_1}, t_{n_1+1}, \ldots, t_n)^T = ((t')^T, (t'')^T)^T$. Векторы ξ' и ξ'' независимы тогда и только тогда, когда для всех $t \in R^n$

$$\varphi_\xi(t) = \varphi_{\xi'}(t') \cdot \varphi_{\xi''}(t'') \ .$$

Используя первое утверждение теоремы последнее соотношение можно записать в виде:

$$\exp\{i(m,t) - \frac{1}{2}(\Sigma t, t)\} =$$

$$\exp\{i(m',t') - \frac{1}{2}(\Sigma_{11}t', t')\} \cdot \exp\{i(m'',t'') - \frac{1}{2}(\Sigma_{22}t'', t'')\} \ .$$

А это равенство имеет место для всех $t \in R^n$ тогда и только тогда, когда $\Sigma_{12} = \Sigma_{21}^T = 0$.

Теорема 3 (B3). *Если $\xi = (\xi_1, \ldots, \xi_n)^T \sim N(m, \Sigma)$, то условное распределение ξ'' при условии $\xi' = x'$ является многомерным нормальным распределением с вектором средних*

$$m'' + \Sigma_{21}\Sigma_{11}^{-1}(x' - m')$$

и матрицей ковариаций

$$\Sigma_{22.1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \ .$$

Доказательство. Будем рассматривать ξ' и ξ'' как векторы в R^n, дополнив недостающие координаты нулями. Положим $\eta = ((\eta')^T, (\eta'')^T)^T$, где

$$\begin{aligned} \eta' &= \xi' \ , \\ \eta'' &= \xi'' - \Sigma_{21} \cdot \Sigma_{11}^{-1} \cdot \xi' \ . \end{aligned}$$

Случайный вектор η получен из случайного вектора ξ невырожденным линейным преобразованием с матрицей

$$C = \begin{pmatrix} I_{n_1} & 0 \\ -\Sigma_{21} \cdot \Sigma_{11}^{-1} & I_{n_2} \end{pmatrix} .$$

В силу теоремы 1 он имеет многомерное нормальное распределение со средним $((m')^T, (m'' - \Sigma_{21} \cdot \Sigma_{11}^{-1} \cdot m')^T)^T$ и матрицей ковариаций

$$\begin{aligned} \Sigma_\eta &= C \cdot \Sigma_\xi \cdot C^T \\ &= \begin{pmatrix} I_{n_1} & 0 \\ -\Sigma_{21} \cdot \Sigma_{11}^{-1} & I_{n_2} \end{pmatrix} \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \begin{pmatrix} I_{n_1} & -\Sigma_{21} \cdot \Sigma_{11}^{-1} \\ 0 & I_{n_2} \end{pmatrix} \\ &= \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} - \Sigma_{21} \cdot \Sigma_{11}^{-1} \cdot \Sigma_{12} \end{pmatrix} . \end{aligned}$$

В силу теорем 1 и 2 случайные векторы η' и η'' независимы и имеют матрицы ковариаций Σ_{11} и $\Sigma_{22} - \Sigma_{21} \cdot \Sigma_{11}^{-1} \cdot \Sigma_{12}$, соответственно. Далее, в силу независимости условное распределение η'' при условии $\eta' = \xi' = x'$ совпадает с безусловным и является многомерным нормальным со средним $m'' - \Sigma_{21} \cdot \Sigma_{11}^{-1} \cdot m'$ и матрицей ковариаций $\Sigma_{22} - \Sigma_{21} \cdot \Sigma_{11}^{-1} \cdot \Sigma_{12}$. Тогда условное распределение $\xi'' = \eta'' + \Sigma_{21} \cdot \Sigma_{11}^{-1} \cdot \xi'$ при условии $\xi' = \eta' = x'$ совпадает с условным распределением $\eta'' + \Sigma_{21} \cdot \Sigma_{11}^{-1} \cdot x'$ при том же условии и является многомерным нормальным распределением в R^{n_2} со средним $m'' + \Sigma_{21} \cdot \Sigma_{11}^{-1} \cdot (x' - m')$ и матрицей ковариаций $\Sigma_{22} - \Sigma_{21} \cdot \Sigma_{11}^{-1} \cdot \Sigma_{12}$.

Теорема 3 доказана.

Во многих задачах теории вероятностей и математической статистики часто используется следующий результат.

Теорема 4 (В4). *Если случайный вектор ξ имеет невырожденное нормальное распределение в R^n с вектором средних m и матрицей ковариаций Σ, то случайная величина*

$(\xi-m)^T\Sigma^{-1}(\xi-m)$ имеет χ^2-распределение с n степенями свободы.

Доказательство. Из линейной алгебры известно, что если Σ есть симметричная невырожденная положительно определенная матрица, то существует такая невырожденная квадратная матрица P, что $\Sigma = P \cdot P^T$. Определим случайный вектор $\eta = P^{-1}(\xi - m)$. В силу теоремы 1 он имеет многомерное нормальное распределение со средним 0 и матрицей ковариаций $P^{-1}\Sigma(P^{-1})^T = P^{-1}PP^T(P^T)^{-1} = I_n$, т.е. стандартное нормальное распределение. Тогда случайная величина

$$\begin{aligned}
\eta^T \cdot \eta &= (P^{-1}(\xi - m))^T(P^{-1}(\xi - m)) \\
&= (\xi - m)^T((P^{-1})^T P^{-1})(\xi - m) \\
&= (\xi - m)^T\Sigma^{-1}(\xi - m)
\end{aligned}$$

имеет по определению χ^2-распределение с n степенями свободы.

Литература

[1] Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики. Учебник для вузов. – М.: ЮНИТИ, 1998. – 1022 с.

[2] Айвазян С.А., Мхитарян В.С. Практикум по прикладной статистике и эконометрике (учебное пособие). – М.: МЭСИ, 1998. – 158 с.

[3] Доугерти К. Введение в эконометрику. – М.: ИНФРА-М, 1997. – 402 с.

[4] Замков О.О. Эконометрические методы в макроэкономическом анализе: Курс лекций. – М.: ГУ ВШЭ, 2001. – 122 с.

[5] Кремер Н.Ш., Путко Б.А. Эконометрика: Учебник для вузов. – М.: ЮНИТИ-ДАНА, 2002. – 311 с.

[6] Магнус Я.Р., Катышев П.К., Пересецкий А.А. Эконометрика. Начальный курс. – М.: Дело, 2000. – 400 с.

[7] Орлов А.И. Эконометрика. Учебник. – М.: Экзамен, 2002. – 576 с.

[8] Тутубалин В.Н. Эконометрика: образование которое нам не нужно. – М.: Фазис, 2004. – 168 с.

[9] Эконометрика: Учебник/ Под ред. И.И. Елисеевой. – М.: Финансы и статистика, 2001. – 344 с.

[10] Green W.H., Econometric Analysis, second edition. – Macmillan Publishing Company, 1993.

[11] Durbin J., Watson G.S., Testing for Serial Correlation in Least-Squares Regression. – Biometrica, 1951, v. 38, pp. 159-177.

[12] Newey W., West K., A Simple Positive Semi-Definite, Heteroscedasticity and Autocorrelation Consistent Covariance Matrix. – Econometrica, 1987, v. 55, pp. 703-708.

[13] Pindyck R.S., Rubinfeld D.L., Econometric Models and Economic Forecasts, third edition. – McGraw-Hill, 1991.

[14] White H., A Heteroscedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity. – Econometrica, 1980, v. 48, no. 4, pp. 817-838.

[15] http://www.nsu.ru/ef/tsy/ecmr/index.htp

[16] Econometric Links Econometric Journal (Ресурс в Интернет)

[17] Econometric Sources (U. of Illinois) (Ресурс в Интернет)

[18] Econometric Resourceson the Net (Kane) (Ресурс в Интернет)

[19] http://www.statistics.com/

[20] Ресурсы по статистике и эконометрике (Ресурс в Интернет)