

ОСНОВЫ РЕГРЕССИОННОГО АНАЛИЗА

После обнаружения *стохастических связей* между изучаемыми переменными величинами (см. 6) исследователь приступает к математическому описанию интересующих его зависимостей. Для достижения этой цели ему необходимо решить следующие задачи:

- 1) подобрать класс функций, в котором целесообразно искать наилучшую (в определенном смысле) аппроксимацию искомой зависимости;
- 2) найти *оценки* для неизвестных значений параметров, входящих в уравнение искомой зависимости;
- 3) установить адекватность полученного уравнения искомой зависимости;
- 4) выявить наиболее информативные *входные переменные (факторы)*.

Совокупность перечисленных задач и составляет предмет исследований *регрессионного анализа*.

1. Исходные предположения

Во многих прикладных задачах требуется построить *математическую модель*, связывающую *входные переменные (факторы)* X_1, \dots, X_p и *выходное переменное (отклик)* Y . В дальнейших рассуждениях будем предполагать, что Y является случайной величиной при каждом фиксированном наборе $\vec{x} = (x_1, \dots, x_p)$ значений переменных $\vec{X} = (X_1, \dots, X_p)$. В этом случае искомая *математическая модель* может быть представлена в следующем виде:

$$Y = f(\vec{x}) + \varepsilon(\vec{x}), \quad (7.1)$$

где $f(\vec{x})$ — скалярная функция, $\varepsilon(\vec{x})$ — случайная ошибка, т.е. случайная составляющая, порожденная либо действием случайных факторов, не включенных в набор X_1, \dots, X_p , либо случайными ошибками измерений величины $f(\vec{x})$, либо и тем и другим одновременно.

Будем считать, что для каждого \vec{x} математическое ожидание $\varepsilon(\vec{x})$ равно нулю, т.е. отсутствует *систематическая погрешность* модели. Следовательно для условного математического ожидания $\bar{y}(\vec{x}) = M(Y | \vec{X} = \vec{x})$ выходного переменного Y при условии, что вектор входных переменных \vec{X} принял значение \vec{x} , согласно (7.1), имеем $\bar{y}(\vec{x}) = f(\vec{x})$.

Функцию $f(\vec{x})$, описывающую зависимость условного среднего значения $\bar{y}(\vec{x})$ выходного переменного Y от заданных фиксированных значений входных переменных X_1, \dots, X_p , называют *функцией регрессии* (или *регрессией*).

Функция регрессии полностью определена, если известен условный закон распределения выходного переменного Y при условии, что $\vec{X} = \vec{x}$. Поскольку в реальных ситуациях никогда не располагают такой информацией, то обычно ограничиваются поиском подходящей аппроксимации $f_a(\vec{x})$ для $f(\vec{x})$, основываясь на *статистических данных* вида (\vec{x}^i, y_i) , $i = \overline{1, n}$, где $\vec{x}^i = (x_1^i, \dots, x_p^i)$. Эти данные есть результат n независимых наблюдений y_1, \dots, y_n случайной величины Y при значениях входных переменных $\vec{x}^1 = (x_1^1, \dots, x_p^1)$, $\vec{x}^2 = (x_1^2, \dots, x_p^2)$, ..., $\vec{x}^n = (x_1^n, \dots, x_p^n)$, т.е. результат специально организованного эксперимента.

Говоря о подходящей аппроксимации функции $f(\vec{x})$ — *модели регрессии*, нужно, во-первых, задать класс *допустимых моделей регрессии* \mathcal{F} , т.е. класс функций, среди которых будем искать наилучшую аппроксимирующую функцию $f_a(\vec{x})$, и, во-вторых, выбрать *критерий*, по которому будем получать наилучшую аппроксимирующую функцию $f_a(\vec{x})$ из заданного класса \mathcal{F} .

Чтобы задать критерий, используют функцию $\rho(\varepsilon_f(\vec{X}))$, где $\varepsilon_f(\vec{X}) = f(\vec{X}) - f_a(\vec{X})$ — случайная величина, а $\rho(u)$ — некоторая неотрицательная функция аргумента u , как правило, *неубывающая* и *выпуклая*, например $\rho = u^2$ или $\rho = |u|$.

Функцию $f_a(x)$ считают наилучшей аппроксимирующей функцией из заданного класса \mathcal{F} , если она обеспечивает минимальное значение функционала

$$\Delta(f_a) = M\rho(\varepsilon_f(\vec{X})) \quad \text{или} \quad \Delta_n(f_a) = \frac{1}{n} \sum_{i=1}^n \rho(\varepsilon_f(\vec{x}^i)),$$

где усреднение проводится по всем возможным значениям случайного вектора \vec{X} в первом равенстве и по всем имеющимся наблюдениям — во втором.

В случае функции $\rho(u) = u^2$ получаемую *регрессию* называют *средней квадратичной*, а метод, реализующий минимизацию функционала $\Delta_n(f_a)$, принято называть *методом наименьших квадратов (МНК)*. Далее будем рассматривать только этот тип регрессии. Поэтому, говоря о регрессии, будем опускать слова „средняя квадратичная“.

В дальнейшем будем предполагать, что класс \mathcal{F} допустимых моделей регрессии можно задать некоторым параметрическим семейством функций, т.е. представить в виде $\mathcal{F}_\beta = \{f_a(\vec{x}; \vec{\beta})\}$, $\vec{\beta} \in \mathbb{R}^m$. Тогда задача отыскания наилучшей аппроксимации для $f(\vec{x})$ сводится к определению таких значений параметров $\vec{\beta}$, при которых $\Delta_n(f_a)$ достигает минимума.

Следует отметить, что проблема выбора параметрического семейства функций \mathcal{F}_β , являясь ключевой в *регрессионном анализе*, не имеет, к сожалению, формализованных процедур для своего решения. Иногда выбор определяют на основе *экспериментальных данных* (\vec{x}^i, y_i) , $i = \overline{1, n}$ (см. пример 7.1), чаще — из теоретических соображений.

Например, если известно, что скорость протекания химической реакции между некоторыми компонентами пропорцио-

нальна объему исходного вещества, то объем вещества $V(t)$ в момент t изменяется по экспоненциальному закону

$$V(t) = \theta_0 e^{-\theta_1(t-t_0)}, \quad t > t_0,$$

где θ_0, θ_1 — неизвестные параметры модели, которые нужно оценить наилучшим образом по результатам наблюдений, а t_0 — начальный момент времени.

К сожалению, такие случаи редки. Более реальной является ситуация, когда о механизме явления ничего не известно и можно лишь предполагать, что искомая функция $f(\vec{x})$ является достаточно гладкой. Тогда аппроксимирующая ее функция $f_a(\vec{x})$ может быть представлена в виде линейной комбинации некоторого набора линейно независимых **базисных функций** $\{\psi_k(\vec{x})\}$, $k = \overline{0, m-1}$, т.е. в виде

$$f_a(\vec{x}; \vec{\beta}) = \vec{\beta}^T \vec{\psi}(\vec{x}) = \sum_{k=0}^{m-1} \beta^k \psi_k(\vec{x}), \quad (7.2)$$

где $\vec{\beta} = (\beta_0 \ \beta_1 \ \dots \ \beta_{m-1})^T$ — вектор неизвестных параметров; $\vec{\psi} = (\psi_0 \ \psi_1 \ \dots \ \psi_{m-1})^T$ — вектор базисных функций (известных заранее); m — число неизвестных параметров β_k , в общем случае неизвестная величина, уточняемая в ходе построения модели.

Следует заметить, что, согласно (7.2), функция $f_a(\vec{x}) = f_a(\vec{x}; \vec{\beta})$ является **линейной** по параметрам, представленным вектором $\vec{\beta}$. Поэтому в рассматриваемом случае говорят о **модели, линейной по параметрам**.

Другими словами, исходный класс функций \mathcal{F} , содержащий истинную функцию регрессии $f(\vec{x})$, заменяют некоторым классом $\mathcal{F}_\beta = \{f_a(\vec{x}; \vec{\beta})\}$, $\vec{\beta} \in \mathbb{R}^m$, более простых по структуре функций, представимых в виде (7.2), и задача сводится к наилучшей оценке вектора неизвестных параметров $\vec{\beta} = (\beta_0, \beta_1, \dots, \beta_{m-1})^T$.

При такой постановке задачи общая погрешность

$$\Delta = \sqrt{\sum_{i=1}^n (y_i - f_a(\bar{x}^i))^2}$$

от аппроксимации результатов наблюдений

$$y_i = f(\bar{x}^i) + \varepsilon_i, \quad i = \overline{1, n},$$

полученных в эксперименте, значениями функции $f_a(\bar{x}) \in \mathcal{F}_\beta$ обусловлена рассеянием отклика Y относительно истинной регрессии $f(\bar{x})$, т.е. величиной

$$\Delta_\varepsilon = \sqrt{\sum_{i=1}^n (y_i - f(\bar{x}^i))^2}$$

и систематической погрешностью аппроксимации, связанной с заменой исходного класса функций \mathcal{F} более узким $\mathcal{F}_\beta \in \mathcal{F}$:

$$\Delta_a = \sqrt{\sum_{i=1}^n (f(\bar{x}^i) - f_a(\bar{x}^i))^2}.$$

Следовательно, приближение $f(\bar{x}) \approx f_a(\bar{x}; \vec{\beta})$ (см. 7.2) нужно понимать в том смысле, что систематической погрешностью Δ_a при замене класса \mathcal{F} на \mathcal{F}_β можно пренебречь по сравнению со случайной погрешностью Δ_ε . Именно на сопоставлении этих двух типов погрешностей и основаны правила проверки адекватности модели $f_a(\bar{x}; \hat{\vec{\beta}}) = \hat{f}_a(\bar{x})$, где вектор параметров $\vec{\beta}$ заменен значением вектора оценок $\hat{\vec{\beta}}$.

Одним из наиболее распространенных аппроксимирующих классов функций \mathcal{F}_β является класс полиномов, в котором в качестве базисных функций выбраны степени переменных x_1, \dots, x_p .

Простейшей полиномиальной моделью является модель 1-го порядка, линейная по всем переменным:

$$f_a(x) = \sum_{k=1}^{m-1} \beta_k x_k, \quad m \leq p+1,$$

Метод наименьших квадратов

Матрицы F и Y в *линейной регрессионной модели* (7.6) содержат всю информацию, получаемую в результате эксперимента. По этим данным нам нужно оценить вектор неизвестных параметров $\beta = (\beta_0, \beta_1, \dots, \beta_{m-1})$. Для получения оценок, как отмечалось выше, будем использовать *метод наименьших квадратов*. Предварительно сформулируем предположения, лежащие в его основе.

1. $M\varepsilon_i = 0$, $i = 1, n$, т.е. систематическая погрешность модели отсутствует.

2. $M(\varepsilon_i \varepsilon_j) = 0$, $i \neq j$, т.е. случайные ошибки некоррелированы (это ограничение можно снять, если матрица ковариаций $D(\vec{\varepsilon})$ вектор-столбца ошибок известна*).

3. $D\varepsilon_i = M\varepsilon_i^2 = \sigma^2$, $i = 1, n$, т.е. в любых точках факторного пространства X^p случайные ошибки имеют одинаковую дисперсию.

4. Значения x_i переменных X_i , $i = 1, p$, в процессе эксперимента измеряются без ошибок.

Отметим, что предположения 2 и 3 можно объединить и представить в следующем виде:

$$D\vec{\varepsilon} = \sigma^2 I_n,$$

где I_n — единичная матрица порядка n .

Четвертое предположение означает, что, согласно соотношениям (7.3), верны равенства

$$MY_i = \sum_{k=0}^{m-1} \beta_k \psi_k(\bar{x}^i), \quad DY_i = D\varepsilon_i = \sigma^2, \quad i = \overline{1, n},$$

которые в матричной записи имеют вид

$$MY = F\vec{\beta}, \quad DY = \sigma^2 I_n.$$

Подчеркнем, что никаких предположений о законе распределения случайных величин Y_i , $i = 1, n$, мы пока не делаем.

Теорема 7.1. Пусть $M = F^T F$ — невырожденная матрица. Несмещенной эффективной оценкой в классе всех линейных

оценок для параметра $\vec{\beta} = (\beta_0 \ \beta_1 \ \dots \ \beta_{m-1})^T$ в линейной регрессионной модели (7.6) является **оценка метода наименьших квадратов (МНК-оценка)**, определяемая матричным равенством

$$\widehat{\beta}(\vec{Y}_n) = (F^T F)^{-1} F^T Y. \quad \# \quad (7.7)$$

Поясним идею метода наименьших квадратов и происхождение формулы (7.7). Докажем несмещенность и эффективность оценки $\widehat{\beta}(\vec{Y}_n)$ в классе линейных оценок.

Пусть отклик Y зависит лишь от одного фактора X ($p = 1$), а искомая функция регрессии $M(Y|x) = f(x)$ имеет график, изображенный пунктирной линией на рис. 7.4. Функция $f(x)$ нам не известна, известны лишь значения отклика y_1, \dots, y_n , полученные в эксперименте при значениях факторов x_1, \dots, x_n (на рис. 7.4 точки (x_i, y_i) , $i = \overline{1, n}$, отмечены „крестиками“).

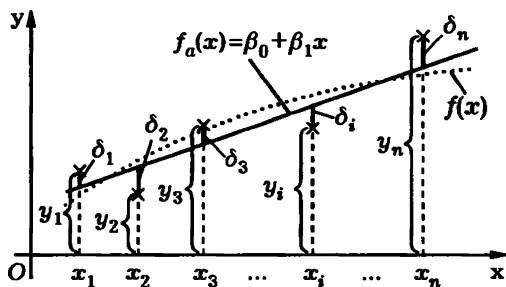


Рис. 7.4

Неизвестную функцию $f(x)$ на основании характера расположения экспериментальных точек (они визуально расположены вдоль прямой) естественно аппроксимировать линейной функцией $f_a(x; \vec{\beta}) = \beta_0 + \beta_1 x$. Отклонения $\delta_i = y_i - (\beta_0 + \beta_1 x_i)$, $i = \overline{1, n}$, ординат экспериментальных точек (x_i, y_i) от любой прямой $f_a(x; \vec{\beta}) = \beta_0 + \beta_1 x$ называют **невязками**.

В общем случае для линейной регрессионной модели (7.6)

$$\delta_i = y_i - \sum_{k=0}^{m-1} \beta_k \psi_k(x_i), \quad i = \overline{1, n},$$

и невязку δ_i можно рассматривать как реализацию случайной ошибки $\varepsilon_i = \varepsilon(x_i)$, $i = \overline{1, n}$.

Согласно методу наименьших квадратов, оценку $\widehat{\vec{\beta}}(\vec{Y}_n) = (\widehat{\beta}_0(\vec{Y}_n) \dots \widehat{\beta}_{m-1}(\vec{Y}_n))^T$ вектора параметров $\vec{\beta} = (\beta_0 \dots \beta_{m-1})^T$ выбирают так, чтобы сумма квадратов невязок δ_i была минимальной, т.е.

$$\Delta(\vec{\beta}) = \sum_{i=1}^n \delta_i^2 = \sum_{i=1}^n \left(y_i - \sum_{k=0}^{m-1} \beta_k \psi_k(x^i) \right)^2 \rightarrow \min,$$

или, что то же самое,

$$\Delta(\vec{\beta}) = \delta^T \delta = (Y - F\vec{\beta})^T (Y - F\vec{\beta}) = \Delta(\vec{\beta}) \rightarrow \min,$$

где $\delta = (\delta_1, \dots, \delta_n)$ — вектор невязок.

Необходимым условием экстремума функции $\Delta(\vec{\beta})$ переменных $\beta_0, \beta_1, \dots, \beta_{m-1}$ (а, следовательно, и условием существования МНК-оценки параметра $\vec{\beta}$), как известно, является равенство нулю ее частных производных [V], т.е.

$$\frac{\partial \Delta}{\partial \beta_\nu} = -2 \sum_{i=1}^n \psi_\nu(\vec{x}^i) \left(y_i - \sum_{k=0}^{m-1} \psi_k(\vec{x}^i) \beta_k \right) = 0, \quad \nu = \overline{0, m-1}.$$

Эту систему можно представить, используя матричную запись $-2F^T(Y - F\vec{\beta}) = 0$, или

$$F^T F \vec{\beta} = F^T Y. \quad (7.8)$$

Из геометрических соображений очевидно, что решением системы (7.8) является точка минимума функции $\Delta(\beta)$, в чем можно убедиться непосредственно, воспользовавшись достаточным условием экстремума [V]. Систему линейных алгебраических уравнений (7.8) называют **системой нормальных уравнений** Гаусса. Она всегда имеет решение (хотя не всегда единственное).

Пусть матрица $F^T F$ имеет обратную матрицу $(F^T F)^{-1}$ (для этого необходимо и достаточно, чтобы $\text{rang } F$ был равен числу столбцов матрицы F). Тогда, умножая обе части равенства (7.8) слева на матрицу $(F^T F)^{-1}$, приходим к формуле (7.7), которая дает единственное решение системы (7.8).

Если матрица $F^T F$ не имеет обратной (случай, когда rang матрицы F меньше числа m ее столбцов), то МНК-оценка параметра $\vec{\beta}$ существует, но не является единственной*.

Несмещенность оценки $\vec{\beta}(\vec{Y}_n)$, заданной равенством (7.7), и эффективность в классе всех линейных несмещенных оценок непосредственно следуют из исходных предположений для метода наименьших квадратов. Действительно,

$$\begin{aligned} M\widehat{\vec{\beta}}(\vec{Y}_n) &= M((F^T F)^{-1} F^T Y) = \\ &= (F^T F)^{-1} F^T M Y = (F^T F)^{-1} F F^T \vec{\beta} = \vec{\beta}, \end{aligned}$$

т.е. $\widehat{\vec{\beta}}(\vec{Y}_n)$ — несмещенная оценка для β . Докажем ее эффективность.

Пусть LY — произвольная линейная несмещенная оценка для $\vec{\beta}$. Тогда из равенства

$$M(LY) = LMY = LF\vec{\beta} = \vec{\beta}$$

получаем $LF = I_n$ и

$$\begin{aligned} D(LY) &= M(LY - LMY)^2 = ML(Y - MY)^2 L^T = \\ &= LDY L^T = L\sigma^2 I_n L^T = \sigma^2 LL^T. \end{aligned}$$

Наша задача минимизировать диагональные элементы матрицы LL^T , которые с точностью до множителя σ^2 являются дисперсиями оценок параметров β_k , $k = \overline{0, m-1}$. Для этого рассмотрим равенство

$$LL^T = (M^{-1}F^T)(M^{-1}F^T)^T + (L - M^{-1}F^T)(L - M^{-1}F^T)^T,$$

в справедливости которого можно убедиться непосредственно, перемножив матрицы в правой его части с учетом равенств $LF = I_n$ и $M = F^T F$. Поскольку диагональные элементы матрицы вида AA^T являются неотрицательными, то можно утверждать, что диагональные элементы матрицы LL^T будут минимальными, если $L = M^{-1}F^T$, т.е. оценка $\widehat{\vec{\beta}}(\vec{Y}_n)$ является эффективной в классе всех линейных оценок.

Итак, по теореме 7.1 МНК-оценки являются наилучшими в указанном выше смысле в классе линейных оценок. Тем самым равенство (7.5) определяет наилучшую *модель регрессии* для выбранных базисных функций и значений $\hat{\beta}_k$, $k = \overline{0, m-1}$, найденных по методу наименьших квадратов, которую будем записывать (обозначив $\hat{f}_a(\vec{x}) = \hat{y}(\vec{x})$) в виде

$$\hat{y}(\vec{x}) = \sum_{k=0}^{m-1} \hat{\beta}_k \psi_k(\vec{x}). \quad (7.9)$$

Случайную величину

$$\hat{Y}(\vec{x}) = \sum_{k=0}^{m-1} \hat{\beta}_k(\vec{Y}_n) \psi_k(\vec{x})$$

будем называть *оценкой среднего значения отклика* Y .

Согласно (7.9), можно определить оценки $\hat{Y}_i = \hat{Y}(\vec{x}^i)$ среднего значения отклика (условного математического ожидания отклика) в каждой точке \vec{x}^i факторного пространства:

$$\hat{Y}_i = \sum_{k=0}^{m-1} \hat{\beta}_k(\vec{Y}_n) \psi_k(\vec{x}^i), \quad i = \overline{1, n}.$$

При этом, если ввести *матрицу* $\hat{Y} = (\hat{Y}_1, \dots, \hat{Y}_n)^T$ *оценок среднего значения отклика*, то

$$\hat{Y} = F \hat{\beta}(\vec{Y}_n). \quad (7.10)$$

Замечание 7.2. В ряде случаев интерес представляют не сами параметры $\beta_0, \beta_1, \dots, \beta_{m-1}$ в линейной регрессионной модели (7.6), а их некоторые линейные комбинации, т.е. новый вектор параметров $\vec{\alpha} = (\alpha_0, \alpha_1, \dots, \alpha_{q-1})$, $q \leq m$, связанный с вектором $\vec{\beta} = (\beta_0, \dots, \beta_{m-1})^T$ соотношением $\vec{\alpha} = A \vec{\beta}$, где A — некоторая матрица типа $q \times m$.

Для вектора $\vec{\alpha}$ МНК-оценка $\widehat{\vec{\alpha}}(\vec{Y}_n)$ определяется равенством

$$\widehat{\vec{\alpha}}(\vec{Y}_n) = A\widehat{\vec{\beta}}(\vec{Y}_n), \quad (7.11)$$

где $\widehat{\vec{\beta}}(\vec{Y}_n)$ — МНК-оценка для $\vec{\beta}$. #

Укажем теперь правило определения ковариационной матрицы $\Sigma(\widehat{\vec{\beta}})$ МНК-оценки $\widehat{\vec{\beta}}(\vec{Y}_n)$ вектора параметров $\vec{\beta}$. Это правило будет вытекать как частный случай из следующей теоремы.

Теорема 7.2. Пусть $\vec{\beta}$ — m -мерный вектор-столбец линейной регрессионной модели (7.6), A — произвольная матрица типа $q \times m$, где $1 \leq q \leq m$, а матрица $F^T F$ является обратимой (т.е. $\det(F^T F) \neq 0$). Тогда для вектора $\vec{\alpha} = A\vec{\beta}$ МНК-оценка $\widehat{\vec{\alpha}}(\vec{Y}_n)$, определяемая равенством (7.11), является несмещенной оценкой с матрицей ковариаций

$$\Sigma(\widehat{\vec{\alpha}}) = \sigma^2 A(F^T F)^{-1} A^T, \quad (7.12)$$

где σ^2 — дисперсия отклика.

◀ Согласно (7.7) и (7.6), имеем

$$\begin{aligned} \widehat{\vec{\beta}}(\vec{Y}_n) &= (F^T F)^{-1} F^T (F\vec{\beta} + \vec{\varepsilon}) = \\ &= (F^T F)^{-1} (F^T F)\vec{\beta} + (F^T F)^{-1} F^T \vec{\varepsilon} = \vec{\beta} + (F^T F)^{-1} F^T \vec{\varepsilon}. \end{aligned}$$

Отсюда для оценки $\widehat{\vec{\alpha}}(\vec{Y}_n) = A\widehat{\vec{\beta}}(\vec{Y}_n)$ параметра $\vec{\alpha} = A\vec{\beta}$ имеем представление

$$\begin{aligned} \widehat{\vec{\alpha}}(\vec{Y}_n) &= A\widehat{\vec{\beta}}(\vec{Y}_n) = A\vec{\beta} + A(F^T F)^{-1} F^T \vec{\varepsilon} = \\ &= \vec{\alpha} + A(F^T F)^{-1} F^T \vec{\varepsilon}, \quad (7.13) \end{aligned}$$

из которого вытекает несмещенность оценки $\widehat{\vec{\alpha}}(\vec{Y}_n)$, так как, согласно исходным предположениям метода наименьших квад-

ратов, $M\vec{\varepsilon} = \vec{0}$ и, следовательно,

$$M\widehat{\vec{\alpha}}(\vec{Y}_n) = \vec{\alpha} + A(F^T F)^{-1} F^T M\vec{\varepsilon} = \vec{\alpha}.$$

Далее, матрица ковариаций вектора МНК-оценок $\widehat{\vec{\alpha}}(\vec{Y}_n)$ в силу несмещенности оценки $\vec{\alpha}$ имеет вид

$$\Sigma(\widehat{\vec{\alpha}}) = M\left(\left(\widehat{\vec{\alpha}}(\vec{Y}_n) - \vec{\alpha}\right)\left(\widehat{\vec{\alpha}}(\vec{Y}_n) - \vec{\alpha}\right)^T\right).$$

Используя представление (7.13), преобразуем выражение для $\Sigma(\widehat{\vec{\alpha}})$ следующим образом:

$$\begin{aligned} \Sigma(\widehat{\vec{\alpha}}) &= M\left(\left(A(F^T F)^{-1} F^T \vec{\varepsilon}\right)\left(A(F^T F)^{-1} F^T \vec{\varepsilon}\right)^T\right) = \\ &= M\left(A(F^T F)^{-1} F^T \vec{\varepsilon} \vec{\varepsilon}^T F(F^T F)^{-1} A^T\right) = \\ &= A(F^T F)^{-1} F^T M(\vec{\varepsilon} \vec{\varepsilon}^T) F(F^T F)^{-1} A^T. \end{aligned}$$

(При переходе к правой части мы воспользовались правилом транспонирования произведения матриц $(AB)^T = B^T A^T$ [III] и симметричностью матрицы $(F^T F)^{-1}$.)

Если теперь учесть, что, согласно исходным предположениям метода наименьших квадратов, $M(\vec{\varepsilon} \vec{\varepsilon}^T) = I_n \sigma^2$, то

$$\begin{aligned} \Sigma(\widehat{\vec{\alpha}}) &= A(F^T F)^{-1} F^T I_n \sigma^2 F(F^T F)^{-1} A^T = \\ &= \sigma^2 A(F^T F)^{-1} (F^T F) (F^T F)^{-1} A^T = \sigma^2 A(F^T F)^{-1} A^T, \end{aligned}$$

что и доказывает представление (7.12). ►

Следствие 7.1. Если $A = I_m$, т.е. $\vec{\alpha} = A\vec{\beta} = \vec{\beta}$, то

$$\Sigma(\widehat{\vec{\beta}}) = \sigma^2 C, \tag{7.14}$$

где $C = (F^T F)^{-1}$ — **дисперсионная матрица Фишера**.

Следствие 7.2. Дисперсия оценки $\widehat{Y}(\vec{x})$ среднего значения отклика в произвольной точке \vec{x} факторного пространства \mathcal{X}^p

определяется по формуле

$$\mathbf{D}\hat{Y}(\vec{x}) = \sigma^2 \psi^T(\vec{x}) C \psi(\vec{x}), \quad (7.15)$$

где $\psi^T(\vec{x}) = (\psi_0(\vec{x}), \dots, \psi_{m-1}(\vec{x}))$.

◀ Действительно, согласно (7.9) и (7.14), имеем

$$\begin{aligned} \mathbf{D}\hat{Y}(\vec{x}) &= \mathbf{D}(\psi^T(\vec{x})\hat{\beta}(\vec{Y}_n)) = \\ &= \mathbf{M}(\psi^T(\vec{x})(\hat{\beta}(\vec{Y}_n) - \vec{\beta})(\hat{\beta}(\vec{Y}_n) - \vec{\beta})^T \psi(\vec{x})) = \\ &= \psi^T(\vec{x}) \mathbf{M}((\hat{\beta}(\vec{Y}_n) - \vec{\beta})(\hat{\beta}(\vec{Y}_n) - \vec{\beta})^T) \psi(\vec{x}) = \\ &= \psi^T(\vec{x}) \Sigma(\hat{\beta}) \psi(\vec{x}) = \sigma^2 \psi^T(\vec{x}) C \psi(\vec{x}). \quad \blacktriangleright \end{aligned}$$

Формулы (7.14) и (7.15) содержат неизвестный параметр σ^2 — дисперсию отклика Y . Поэтому требуется правило определения оценки параметра σ^2 . Такое правило устанавливает следующая теорема. Однако прежде чем формулировать теорему, отметим, что случайную величину

$$\hat{\varepsilon}^T \hat{\varepsilon} = (Y - F\hat{\beta}(\vec{Y}_n))^T (Y - F\hat{\beta}(\vec{Y}_n)), \quad (7.16)$$

где $\hat{\varepsilon} = Y - F\hat{\beta}(\vec{Y}_n)$ — случайный вектор, а $\hat{\beta}(\vec{Y}_n)$ — МНК-оценка вектора параметров $\vec{\beta}$ линейной регрессионной модели (7.6), называют *остаточной суммой квадратов*.

Теорема 7.3. Если выполнены исходные предположения метода наименьших квадратов и ранг матрицы базисных функций F типа $n \times m$ равен m , то несмещенная оценка для *остаточной дисперсии* σ^2 определяется по формуле

$$S^2(\vec{Y}_n) = \frac{1}{n-m} (Y - F\hat{\beta}(\vec{Y}_n))^T (Y - F\hat{\beta}(\vec{Y}_n)), \quad (7.17)$$

где $\hat{\beta}(\vec{Y}_n)$ — МНК-оценка вектора параметров $\vec{\beta}$ линейной регрессионной модели (7.6).

◀ Из равенства (7.6) и предположений о случайной составляющей модели следует:

$$\begin{aligned} \mathbf{M}((Y - F\vec{\beta})^T(Y - F\vec{\beta})) &= \mathbf{M}(\vec{\varepsilon}^T \vec{\varepsilon}) = \\ &= \mathbf{M}\left(\sum_{i=1}^n \varepsilon_i^2\right) = \sum_{i=1}^n \mathbf{D}\varepsilon_i = n\sigma^2. \end{aligned} \quad (7.18)$$

Рассмотрим равенства

$$\begin{aligned} (Y - F\vec{\beta})^T(Y - F\vec{\beta}) &= \\ &= (Y - F\hat{\beta}(\vec{Y}_n) + F(\hat{\beta}(\vec{Y}_n) - \vec{\beta}))^T (Y - F\hat{\beta}(\vec{Y}_n) + F(\hat{\beta}(\vec{Y}_n) - \vec{\beta})) = \\ &= (Y - F\hat{\beta}(\vec{Y}_n))^T(Y - F\hat{\beta}(\vec{Y}_n)) + (F(\hat{\beta}(\vec{Y}_n) - \vec{\beta}))^T F(\hat{\beta}(\vec{Y}_n) - \vec{\beta}) + \\ &+ (Y - F\hat{\beta}(\vec{Y}_n))^T F(\hat{\beta}(\vec{Y}_n) - \vec{\beta}) + (F(\hat{\beta}(\vec{Y}_n) - \vec{\beta}))^T (Y - F\hat{\beta}(\vec{Y}_n)). \end{aligned}$$

Поскольку $\hat{\beta}(\vec{Y}_n)$ — МНК-оценка вектора параметров $\vec{\beta}$, то, согласно (7.7), $\hat{\beta}(\vec{Y}_n) = (F^T F)^{-1} F^T Y$, и, как следствие, имеем

$$\begin{aligned} (F(\hat{\beta}(\vec{Y}_n) - \vec{\beta}))^T (Y - F\hat{\beta}(\vec{Y}_n)) &= \\ &= (\hat{\beta}(\vec{Y}_n) - \vec{\beta})^T F^T (Y - F(F^T F)^{-1} F^T Y) = \\ &= (\hat{\beta}(\vec{Y}_n) - \vec{\beta})^T (F^T Y - F^T F(F^T F)^{-1} F^T Y) = 0 \end{aligned}$$

и, кроме того,

$$(Y - F\hat{\beta}(\vec{Y}_n))^T F(\vec{\beta} - \hat{\beta}(\vec{Y}_n)) = ((F(\hat{\beta}(\vec{Y}_n) - \vec{\beta}))^T (Y - F\hat{\beta}(\vec{Y}_n)))^T = 0.$$

Таким образом,

$$\begin{aligned} (Y - F\vec{\beta})^T(Y - F\vec{\beta}) - (F(\hat{\beta}(\vec{Y}_n) - \vec{\beta}))^T F(\hat{\beta}(\vec{Y}_n) - \vec{\beta}) &= \\ &= (Y - F\hat{\beta}(\vec{Y}_n))^T (Y - F\hat{\beta}(\vec{Y}_n)). \end{aligned} \quad (7.19)$$

Воспользовавшись свойствами следа матриц и следствием 7.1, получим

$$\begin{aligned}
 \mathbf{M}\left((F\widehat{\beta}(\vec{Y}_n) - \vec{\beta})^T F(\widehat{\beta}(\vec{Y}_n) - \vec{\beta})\right) &= \\
 &= \mathbf{M}\left((\widehat{\beta}(\vec{Y}_n) - \vec{\beta})^T (F^T F)(\widehat{\beta}(\vec{Y}_n) - \vec{\beta})\right) = \\
 &= \mathbf{M}\left(\text{tr}\left((F^T F)(\widehat{\beta}(\vec{Y}_n) - \vec{\beta})(\widehat{\beta}(\vec{Y}_n) - \vec{\beta})^T\right)\right) = \\
 &= \text{tr}\left((F^T F)\mathbf{M}\left((\widehat{\beta}(\vec{Y}_n) - \vec{\beta})(\widehat{\beta}(\vec{Y}_n) - \vec{\beta})^T\right)\right) = \\
 &= \text{tr}\left((F^T F)\Sigma(\widehat{\beta}(\vec{Y}_n))\right) = \text{tr}\left((F^T F)(F^T F)^{-1}\sigma^2\right) = \\
 &= \sigma^2 \text{tr} I_m = m\sigma^2.
 \end{aligned}$$

Таким образом, согласно (7.17)–(7.19), имеем

$$n\sigma^2 - m\sigma^2 = \mathbf{M}\left(Y - F\widehat{\beta}(\vec{Y}_n)\right)^T (Y - F\widehat{\beta}(\vec{Y}_n)),$$

откуда

$$\sigma^2 = \frac{1}{n - m} \mathbf{M}\left((Y - F\widehat{\beta}(\vec{Y}_n))^T (Y - F\widehat{\beta}(\vec{Y}_n))\right).$$

Следовательно,

$$S^2 = \frac{1}{n - m} (Y - F\widehat{\beta}(\vec{Y}_n))^T (Y - F\widehat{\beta}(\vec{Y}_n))$$

является несмещенной оценкой для σ^2 . ►

Замечание 7.3. а. Оценка S^2 остаточной дисперсии σ^2 представляет собой отношение остаточной суммы квадратов $\widehat{\varepsilon}^T \widehat{\varepsilon}$, отнесенное к числу степеней свободы $n - m$, где n — количество наблюдений $\vec{Y}_n = (Y_1, \dots, Y_n)$, представленных матрицей отклика Y , а m — число оцениваемых параметров, представленного вектором $\vec{\beta}$. Таким образом, S^2 — доля остаточной суммы квадратов линейной регрессионной модели (7.6), приходящаяся

на одну „степень свободы“. Фактически число степеней свободы равно объему случайной выборки за вычетом числа независимых линейных связей, наложенных на выборочные значения.

б. Формула (7.17) верна лишь в том случае, если есть основания считать, что выбранная линейная регрессионная модель (7.6) является верной, т.е. $MY = F\vec{\beta}$. В противном случае в остаточную сумму квадратов кроме случайных ошибок входят и систематические, а потому она может давать завышенную оценку для σ^2 .

в. Полезно обратить внимание на сходство результата (7.17) с несмещенной оценкой $S^2(\vec{Y}_n)$ дисперсии случайной величины Y по наблюдениям \vec{Y}_n , которая имеет вид

$$S^2(\vec{Y}_n) = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2.$$

Здесь также сумма квадратов отклонений Y_i от \bar{Y} делится на число степеней свободы $n - 1$, так как неизвестный параметр $MY = \mu$ заменен его оценкой \bar{Y} , т.е. на экспериментальные данные наложена одна линейная связь. #

При решении реальных задач, связанных с практическим использованием регрессионных моделей, необходимо проверять выполнение исходных предположений для метода наименьших квадратов, т.е. проводить статистический анализ регрессионной модели (см. 7.3).

Проиллюстрируем процедуру построения регрессионной модели на частных примерах, имеющих и самостоятельный интерес.

Пример 7.3. Рассмотрим случай *простой линейной регрессии*, когда отклик Y зависит от одного фактора X (т.е. $p = 1$) и в качестве приближения искомой функции регрессии выбрана функция

$$f_a(x) = \beta_0 + \beta_1 x.$$

Эту функцию получают из общей модели (7.2) при $\psi_0(x) \equiv 1$, $\psi_1(x) = x$, т.е. размерность вектора X равна $p = 1$, а число параметров $m = 2$.

Роль фактора X могут играть время (иногда часто вместо x пишут t), температура, доза лечебного препарата и т.д. Задача состоит в изучении связи между откликом Y и фактором X на основании выборки (x^i, y_i) , $i = \overline{1, n}$, полученной в результате эксперимента (вместо x^i далее будем писать x_i , так как x — скаляр).

Для конкретности будем считать, что X — это скорость движения автомобиля (в км/ч), а Y — тормозной путь (в м) по скользкой дороге до полной его остановки, и по результатам $n = 17$ замеров X и Y получены данные, представленные в табл. 7.1.

Таблица 7.1

x_i	28	29	32	35	40	44	45	51	53
y_i	0,53	0,92	1,52	2,07	2,17	3,65	3,97	5,27	5,54
x_i	58	64	65	73	75	80	83	93	
y_i	6,43	7,60	7,91	9,48	10,1	8,95	11,48	13,74	

Найдем значения оценок параметров β_0 и β_1 , дисперсии отклика и дисперсии среднего значения отклика, а также дадим прогноз для длины тормозного пути при скорости $x_0 = 120$ км/ч. В рассматриваемом примере матрицы Y и F имеют вид

$$Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad F = \begin{pmatrix} 1 & x_1 \\ \cdot & \cdot \\ 1 & x_n \end{pmatrix}.$$

Следовательно,

$$F^T F = \begin{pmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{pmatrix}, \quad F^T Y = \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i y_i \end{pmatrix}.$$

Далее случайные векторы и их реализации будем обозначать одинаково: из текста всегда ясно, о чем идет речь.

Поскольку x_i — различные числа, то матрица $M = F^T F$ обратима, причем

$$\det M = n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 = n \sum_{i=1}^n (x_i - \bar{x})^2, \quad \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i.$$

С помощью присоединенной матрицы находим

$$M^{-1} = \frac{1}{\det M} \begin{pmatrix} \sum_{i=1}^n x_i^2 & -\sum_{i=1}^n x_i \\ -\sum_{i=1}^n x_i & n \end{pmatrix}.$$

Используя обратную матрицу $M^{-1} = (F^T F)^{-1}$, по формуле (7.7) получаем

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = M^{-1} F^T Y = \frac{1}{\det M} \begin{pmatrix} \sum_{i=1}^n x_i^2 \left(\sum_{i=1}^n y_i \right) - \sum_{i=1}^n x_i \left(\sum_{i=1}^n x_i y_i \right) \\ -\sum_{i=1}^n x_i \left(\sum_{i=1}^n y_i \right) + n \sum_{i=1}^n x_i y_i \end{pmatrix},$$

откуда

$$\hat{\beta}_0 = \frac{\sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i - \sum_{i=1}^n x_i \left(\sum_{i=1}^n x_i y_i \right)}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2}, \quad \hat{\beta}_1 = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \left(\sum_{i=1}^n y_i \right)}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2}.$$

Из последних двух равенств с помощью простых преобразований получаем

$$\begin{cases} \hat{\beta}_1 = \frac{Q_{xy}}{Q_x}, & \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}, \\ Q_x = \sum_{i=1}^n (x_i - \bar{x})^2, \\ Q_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}). \end{cases} \quad (7.20)$$

Из равенств (7.20) видно, что оценки $\hat{\beta}_0(\vec{Y}_n)$ и $\hat{\beta}_1(\vec{Y}_n)$ связаны линейной зависимостью.

Поскольку $Q_{xy}/n = \hat{K}_{xy}$ является значением оценки ковариации $Q_{xy}(\vec{X}_n, \vec{Y}_n)/n = \hat{K}_{xy}(\vec{X}_n, \vec{Y}_n)$ фактора и отклика, а $Q_x/n = \hat{\sigma}_x^2$ — значением оценки дисперсии фактора, то для значения $\hat{\beta}_1$ оценки параметра β_1 справедливо и такое представление:

$$\hat{\beta}_1 = \frac{\hat{K}_{xy}}{\hat{\sigma}_x^2} = \frac{\hat{\rho} \hat{\sigma}_x \hat{\sigma}_y}{\hat{\sigma}_x^2},$$

где $\hat{\rho}$ — значение оценки коэффициента корреляции ρ между фактором и откликом.

Таким образом, найдена модель простой регрессии

$$\hat{y}(x) = \hat{\beta}_0 + \hat{\beta}_1 x,$$

где $\hat{\beta}_0$ и $\hat{\beta}_1$ определены по формулам (7.20).

Для данных из табл. 7.1 имеем $n = 17$. Далее находим $Q_x \approx 6557$, $Q_y \approx 250,8$, $Q_{xy} \approx 1271,5$.

По формулам (7.20) вычисляем значения оценок $\hat{\beta}_0$ и $\hat{\beta}_1$:

$$\hat{\beta}_1 = 1271,5/6557 \approx 0,194, \quad \hat{\beta}_0 = 5,95 - 0,194 \cdot 55,76 \approx -4,87.$$

Следовательно, прогнозируемое значение y при $x = x_0 = 120$ равно

$$\hat{y}(120) = -4,87 + 0,194 \cdot 120 = 18,4.$$

Найдем точность оценок $\hat{\beta}_0(\vec{Y}_n)$, $\hat{\beta}_1(\vec{Y}_n)$ и $\hat{Y}(x)$. Используя формулу (7.14), получаем

$$\Sigma(\hat{\beta}) = \sigma^2 M^{-1} = \frac{\sigma^2}{\det M} \begin{pmatrix} \sum_{i=1}^n x_i^2 & -\sum_{i=1}^n x_i \\ -\sum_{i=1}^n x_i & n \end{pmatrix},$$

где

$$\det M = n \sum_{i=1}^n (x_i - \bar{x})^2.$$

Таким образом,

$$D\hat{\beta}_0(\vec{Y}_n) = \frac{\sigma^2 \sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2}, \quad D\hat{\beta}_1(\vec{Y}_n) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2},$$

$$\hat{K}(\vec{X}_n, \vec{Y}_n) = \text{cov}(\hat{\beta}_0(\vec{Y}_n), \hat{\beta}_1(\vec{Y}_n)) = -\frac{\sigma^2 \bar{x}}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

По формуле (7.15) находим $D\hat{Y}(x)$:

$$\begin{aligned} D\hat{Y}(x) &= \frac{\sigma^2}{n \sum_{i=1}^n (x_i - \bar{x})^2} (1 \ x) \begin{pmatrix} \sum_{i=1}^n x_i^2 & -\sum_{i=1}^n x_i \\ -\sum_{i=1}^n x_i & n \end{pmatrix} \begin{pmatrix} 1 \\ x \end{pmatrix} = \\ &= \sigma^2 \left(\frac{1}{n} + \frac{(x - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right). \quad (7.21) \end{aligned}$$

В точке прогноза $x = x_0 = 120$ и

$$D\hat{Y}(x_0) = \sigma^2 \left(\frac{1}{n} + \frac{(120 - 55,76)^2}{6557} \right) = 0,69\sigma^2.$$

Наконец, по формуле (7.17) находим значение S_y^2 оценки дисперсии отклика:

$$S_y^2 = \frac{1}{n-2} \sum_{i=1}^n (y_i - \hat{y}(x_i))^2 \approx \frac{1}{15} \cdot 4,2 = 0,28$$

и заменяем σ^2 на S_y^2 во всех предыдущих равенствах, где присутствует σ^2 .

Считая оценку $\hat{Y}(x)$ нормально распределенной с математическим ожиданием $M\hat{Y}(x) = f(x)$ и дисперсией $D\hat{Y}(x)$, вычисленной по формуле (7.21), можно по *правилу* „ 3σ “ указать

интервал возможных значений для Y в точке $x = x_0 = 120$, учитывая, что $\sigma \approx S_y = 0,53$:

$$(\hat{y}(x_0) - 3S_y, \hat{y}(x_0) + 3S_y) = (18,4 - 1,6; 18,4 + 1,6).$$