Лекция 9. Множественная линейная регрессия

Буре В.М., Грауэр Л.В.

ШАД

Санкт-Петербург, 2013

Содержание

Спецификация модели

Рассмотрим следующая модель наблюдений, связывающую значения некоторого наблюдаемого показателя y и объясняющих переменных $x = (x_1, \dots, x_m)^T$:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik} + \varepsilon_i, \quad i = 1, \dots, n, \tag{1}$$

где $\beta^T = (\beta_0, \dots, \beta_k)$ - неизвестные параметры, x_{ij} - значения объясняющих факторов, ε_i — ненаблюдаемая случайная компонента, j — номер переменной, i — номер наблюдения.

Будем предполагать, что имеется n наблюдений показателя y_i, точно известны значения объясняющих переменных $x_i = (x_{i1}, \dots, x_{ik})$ в каждом из наблюдений, причем в модель наблюдений (1) входит значение ненаблюдаемой случайной компоненты ε_i. Ненаблюдаемая компонента может представлять собой ошибку измерений или заменять совокупное действие других переменных, значения которых не учитываются в модели наблюдений (1).

Сформулируем основные предположения регрессионного анализа, которые относятся к случайным компонентам ε_i, $i = 1, \ldots, n$.

Первая группа

- Случайные величины ε_1, $i = 1, \ldots, n$ образуют так называемый слабый белый шум, т. е. последовательность центрированных ($E\varepsilon_i = 0$, $i = 1, \ldots, n$) и некоррелированных ($E(\varepsilon_l \varepsilon_u) = 0$ при $l \neq u$) случайных величин с одинаковыми дисперсиями σ^2 ($E\varepsilon_i^2 = \sigma_i^2$, $i = 1, \ldots n$).

Вторая группа

- Совместное распределение случайных величин ε_i, $i = 1, \ldots, n$ является нормальным распределением с нулевым вектором математических ожиданий и ковариационной матрицей $\sigma^2 E_n$, т. е. случайный вектор $\varepsilon^T = (\varepsilon_1, \ldots, \varepsilon_n) \sim N(0, \sigma^2 E_n)$, где E_n — единичная матрица порядка $n \times n$.

Модель наблюдений (1) можно записать в матричном виде:

$$Y = X\beta + \varepsilon, \tag{2}$$

где $Y = (y_1, \dots, y_n)^T$, $\beta = (\beta_0, \beta_1, \dots, \beta_k)^T$, $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^T$,

$$X = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \dots & \dots & \dots & \dots & \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{pmatrix}$$

— матрица порядка $n \times (k+1)$.

Метод наименьших квадратов

В модели наблюдений (1), (2) неизвестными являются параметры

$$\sigma, \beta_0, \beta_1, \ldots, \beta_k.$$

Рассмотрим в качестве процедуры оценивания неизвестных параметров метод наименьших квадратов.

Введем обозначение: $X_r = (x_{1r}, x_{2r}, \ldots, x_{nr})^T$, $r = 1, \ldots, k$ — столбец матрицы X, тогда

$$X = (X_0, X_1, \ldots, X_K),$$

где $X_0 = (1, 1, \ldots, 1)^T$.

В качестве минимизируемого критерия рассмотрим

$$(Y - X\beta)^T (Y - X\beta) = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - \ldots - \beta_k x_{ik})^2. \qquad (3)$$

Оценки, получаемые из условия минимума (3), называют оценками метода наименьших квадратов.

Для минимизации выражения (3) по неизвестным параметрам β воспользуемся геометрическими представлениями. Рассмотрим в n-мерном пространстве $\mathbb{R}^n$ линейное подпространство $L(X_0, X_1, \ldots, X_k)$, в котором вектора $X_0, X_1, \ldots, X_k$ являются базисом. При этом, конечно, в дальнейшем будем предполагать, что они линейно независимы. Пусть $\hat{Y}$ представляют собой ортогональную проекцию вектора Y на подпространство $L(X_0, X_1, \ldots, X_k)$. Очевидно, что для любого другого вектора $\tilde{Y} \in L(X_0, \ldots, X_k)$ будет выполняться

$$(Y - \hat{Y})^T (Y - \hat{Y}) \leqslant (Y - \tilde{Y})^T (Y - \tilde{Y}),$$

неравенство следует из условия ортогональности вектора $Y - \hat{Y}$ подпространству $L(X_0, \ldots, X_k)$.

Действительно, для любого вектора $\tilde{Y} \in L(X_0, \ldots, X_k)$ выполняется $(\tilde{Y} - \hat{Y}) \perp (Y - \hat{Y})$, так как $(\tilde{Y} - \hat{Y}) \in L(X_0, \ldots, X_k)$ и $(Y - \hat{Y}) \perp L(X_0, \ldots, X_k)$. Но тогда

$$(Y - \tilde{Y})^T (Y - \tilde{Y}) = (Y - \hat{Y} - (\tilde{Y} - \hat{Y}))^T (Y - \hat{Y} - (\tilde{Y} - \hat{Y})) =$$
$$= (Y - \hat{Y})^T (Y - \hat{Y}) + (\tilde{Y} - \hat{Y})^T (\tilde{Y} - \hat{Y})$$

Откуда следует

$$(Y - \tilde{Y})^T (Y - \tilde{Y}) \geqslant (Y - \hat{Y})^T (Y - \hat{Y}),$$

причем, равенство возможно только, если $\tilde{Y} = \hat{Y}$.
Для того, чтобы вектор $Y - \hat{Y}$ был ортогонален подпространству $L(X_0, X_1, \ldots, X_k)$ необходимо и достаточно, чтобы

$$(Y - \hat{Y}) \perp X_i, \quad i = 0, 1, \ldots, k$$

или, чтобы

$$(Y - \hat{Y})^T X = 0 \qquad (4)$$

Так как вектор $\hat{Y} \in L(X_0, X_1, \ldots, X_k)$, то существует набор коэффициентов, обозначим их через $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$, что

$$\hat{Y} = \sum_{r=0}^{k} \hat{\beta}_r X_r = X\hat{\beta},$$

где $\hat{\beta} = (\hat{\beta}_0, \ldots, \hat{\beta}_k)^T$.

Тогда равенство (4) равносильно равенству

$$X^T Y - X^T X \hat{\beta} = 0$$

или равенству

$$(X^T X)\hat{\beta} = X^T Y. \tag{5}$$

Так как, по нашему предположению, столбцы матрицы X линейно независимы, то отсюда следует, что $n \geqslant k$ и определитель $|X^T X| \neq 0$. Следовательно, оценки метода наименьших квадратов $\hat{\beta}$ неизвестных параметров β имеют вид:

$$\hat{\beta} = (X^T X)^{-1} X^T Y. \tag{6}$$

Для удобства дальнейших преобразований введем ряд новых матриц и изучим их свойства. Из формулы (6) следует, что ортогональная проекция произвольного вектора Y на $L(X_0, \ldots, X_K)$ имеет вид $\hat{Y} = X\hat{\beta} = X(X^TX)^{-1}X^TY$, но тогда матрица

$$P = X(X^TX)^{-1}X^T$$

представляет собой матрицу ортогонального проектирования на пространство $L(X_0, \ldots, X_k)$. Рассмотрим подпространство $L^\perp$ ортогональное подпространству $L(X_0, \ldots, X_k)$. Любой вектор Y из $\mathbb{R}^n$ однозначно представим в виде:

$$Y = \hat{Y} + (Y - \hat{Y}),$$

где $\hat{Y} \in L(X_0, \ldots, X_k)$, $Y - \hat{Y} \in L^\perp$. Из предыдущего ясно, что

$$Y - \hat{Y} = Y - X(X^TX)^{-1}X^TY = (E_n - X(X^TX)^{-1}X^T)Y.$$

Следовательно, матрица

$$P^\perp = E_n - P$$

является матрицей ортогонального проектирования на подпространство $L^\perp$.

Действительно, для любого вектора $Y \in \mathbb{R}^n$ получаем

$$P^{\perp}Y = Y - PY = Y - \hat{Y},$$

но вектор $Y - \hat{Y} \perp L(X_0, \ldots, X_k)$, следовательно, вектор $Y - \hat{Y} \in L^{\perp}$.
Нетрудно проверить справедливость следующих свойств:

1. $P^T = P$, $(P^{\perp})^T = P^{\perp}$.
2. $P^2 = P$, $(P^{\perp})^2 = P^{\perp}$.
3. $PP^{\perp} = P^{\perp}P = 0$.
4. $P + P^{\perp} = E_n$.
5. $PX_j = X_j$, $j = 0, 1, \ldots k$, $P^{\perp}X_j = 0$, $j = 0, 1, \ldots, k$.

Введем вектор

$$\hat{\varepsilon} = Y - \hat{Y} = P^{\perp} Y, \tag{7}$$

который будем называть **вектором остатков**.
Следует иметь в виду, что после нахождения коэффициентов $\hat{\beta}$ можно рассмотреть следующую функцию, которую собственно и называют линейной регрессией:

$$\hat{y}(x) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \ldots + \hat{\beta}_k x_k. \tag{8}$$

Построенная линейная регрессия представляет собой эмпирическую зависимость между изучаемым показателем y и объясняющими переменными $x = (x_1, \ldots, x_k)^T$, поэтому естественно называть остатками разности

$$y_i - \hat{y}(x_i) = \hat{\varepsilon}_i, \quad i = 1, \ldots, n, \tag{9}$$

которые не объясняются построенной эмпирической моделью (8).
Вектор, составленный из разностей $\hat{\varepsilon}_i$, и представляет собой вектор $\hat{\varepsilon} = (\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_n)$.

Нетрудно заметить, что

$$\hat{\varepsilon} = P^{\perp} Y = P^{\perp}(X\beta + \varepsilon) = P^{\perp}\varepsilon. \tag{10}$$

Величины $\hat{\varepsilon}_i$, $i = 1, \ldots, n$, можно, следовательно, рассматривать, как оценки ненаблюдаемых величин ε_i, $i = 1, \ldots, n$.

Свойство оценок метода наименьших квадратов

Как нетрудно заметить, оценки метода наименьших квадратов обладают свойством несмещенности:

$$E\hat{\beta} = (X^T X)^{-1} X^T EY = (X^T X)^{-1} X^T X\beta = \beta.$$

Вычислим ковариационную матрицу оценок метода наименьших квадратов, считая, что выполнена первая группа предположений регрессионного анализа:

$$D\hat{\beta} = E\{(\hat{\beta} - \beta)(\hat{\beta} - \beta)^T\} = E\{(X^T X)^{-1} X^T \varepsilon\varepsilon^T X (X^T X)^{-1}\} =$$
$$= (X^T X)^{-1} X^T \sigma^2 E_n X (X^T X)^{-1} = \sigma^2 (X^T X)^{-1}.$$

Оценки метода наименьших квадратов при условии выполнения предположений из первой группы являются наилучшими линейными несмещенными оценками. Линейность понимается в том смысле, что оценки имеют вид $\hat{\beta} = AY$, где $A = (X^T X)^{-1} X^T$ — матрица коэффициентов, т.е. речь идет о линейности по наблюдениям.

Следуя общему определению эффективности оценок, заметим, что оценки $\hat{\beta}$ являются наилучшими линейными несмещенными оценками в том смысле, что для любых других оценок $\tilde{\beta}$ вида $\tilde{\beta} = CY$, обладающих свойством несмещенности, оказывается, что разность дисперсионных матриц оценок $D\tilde{\beta} - D\hat{\beta}$ представляет собой неотрицательно определенную матрицу, т.е. для любого вектора $\gamma \in \mathbb{R}^{k+1}$ выполняется

$$\gamma^T(D\tilde{\beta} - D\hat{\beta})\gamma \geqslant 0.$$

Отсюда, в частности, сразу следует, что дисперсии оценок $\tilde{\beta}_i$ не меньше, чем дисперсии оценок $\hat{\beta}_i$, т.е.

$$D\tilde{\beta}_i \geqslant D\hat{\beta}_i, \quad i = 0, 1, 2, \ldots, k.$$

Теорема 1 (Теорема Гаусса-Маркова)

Пусть выполнены все условия из первой группы предположений регрессионного анализа, тогда оценки метода наименьших квадратов $\hat{\beta}$ являются наилучшими линейными несмещенными оценками.

Доказательство

Пусть $\tilde{\beta} = CY$ — произвольные линейные несмещенные оценки вектора β. Из условия несмещенности следует

$$E\tilde{\beta} = CX\beta = \beta.$$

Так как это равенство должно выполняться для любого вектора β, то условие несмещенности равносильно матричному равенству $CX = E_{k+1}$.

Введем в рассмотрение матрицу D, определяемую равенством

$$D = C - (X^T X)^{-1} X^T.$$

Как легко видеть, $DX = 0$. Очевидно, что

$$D\tilde{\beta} = E\{(\tilde{\beta} - \beta)(\tilde{\beta} - \beta)^T\} = E\{C\varepsilon\varepsilon^T C^T\} = \sigma^2 CC^T.$$

Заметим, что

$$CC^T = (D + (X^T X)^{-1} X^T)(D + (X^T X)^{-1} X^T)^T = DD^T + (X^T X)^{-1}.$$

Но тогда для произвольного вектора $\gamma \in \mathbb{R}^{k+1}$:

$$\gamma^T (D\tilde{\beta} - D\beta)\gamma = \sigma^2 \gamma^T DD^T \gamma = \sigma^2 (\gamma^T D)(\gamma^T D)^T \geqslant 0.$$

Оценим дисперсию одиночного наблюдения σ^2. Для этого вначале найдем ковариационную матрицу $D\hat{\varepsilon}$. Как следует из формулы (6)

$$D\hat{\varepsilon} = E\{P^{\perp}\varepsilon\varepsilon^T P^{\perp}\} = \sigma^2 P^{\perp}.$$

Рассмотрим сумму квадратов остатков:

$$\sum_{i=1}^{n} \hat{\varepsilon}_i^2 = \hat{\varepsilon}^T\hat{\varepsilon} = tr\{\hat{\varepsilon}\hat{\varepsilon}^T\}.$$

Вычислим математическое ожидание:

$$E\hat{\varepsilon}^T\hat{\varepsilon} = Etr\{\hat{\varepsilon}\hat{\varepsilon^T}\} = trE\{\hat{\varepsilon}\hat{\varepsilon^T}\} = tr\{E(P^{\perp}\varepsilon\varepsilon^T P^{\perp})\} = \sigma^2 trP^{\perp}.$$

Заметим, что

$$trP^{\perp} = tr\{E_n - X(X^TX)^{-1}X^T\} = n - tr\{X(X^TX)^{-1}X^T\} =$$
$$= n - tr\{(X^TX)(X^TX)^{-1}\} = n - k - 1.$$

При выводе формулы мы воспользовались свойством $tr(AB) = tr(BA)$, если оба произведения имеют смысл.

В результате проведенных выкладок получаем, что статистика

$$S^2 = \frac{\sum\limits_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-k-1} = \frac{1}{n-k-1}\hat{\varepsilon}^T\hat{\varepsilon}$$

является несмещенной оценкой дисперсии σ^2.

Рассмотрим матрицу $E\{(\hat{\beta}-\beta)\hat{\varepsilon}^T\}$, элементы которой представляют собой ковариации случайных величин $\hat{\beta}_j$ и $\hat{\varepsilon}_i$, $j = 0, 1, \ldots, k$,
$i = 1, 2, \ldots, n$.
Из свойств матрицы $P^{\perp}$ следует

$$E\{(\hat{\beta}-\beta)\hat{\varepsilon}^T\} = (X^TX)^{-1}X^TE\{\varepsilon\varepsilon^T\}P^{\perp} = \sigma^2(X^TX)^{-1}X^TP^{\perp} = 0.$$

Найденная несмещенная оценка S^2 одиночной дисперсии σ^2 позволяет построить несмещенные оценки всех ковариаций вектора оценок $\hat{\beta}$.

Действительно, как было получено ранее, $D\hat{\beta} = \sigma^2(X^TX)^{-1}$, но теперь, заменяя σ^2 на S^2, получаем несмещенные оценки всех ковариаций и дисперсий вектора $\hat{\beta}$, другими словами, элементы матрицы $\hat{\sigma}^2(X^TX)^{-1}$ являются несмещенными оценками дисперсий и ковариаций оценок $\hat{\beta}_j$, $j = 0, \ldots, k$.

Лемма 1

Пусть выполнены все условия из первой и второй групп предположений регрессионного анализа, тогда статистика $(n-k-1)S^2/\sigma^2$ *подчиняется распределению* χ^2 *с* $n-k-1$ *степенями свободы.*

Доказательство

Ранее было установлено, что $\hat{\varepsilon} = P^{\perp}\varepsilon$, где $P^{\perp}$ — матрица ортогонального проектирования на подпространстве $L^{\perp}$. Очевидно, что размерность подпространства $L(X_0, \ldots, X_k)$ равна $k+1$. Но тогда размерность подпространства $L^{\perp}$ равна $n-k-1$. Выберем в подпространстве $L^{\perp}$ ортонормированный базис $l_1, \ldots, l_{n-k-1}$. Пусть $l = (l_1, \ldots, l_{n-k-1})$. Тогда вектор $\hat{\varepsilon}$ единственным образом представим в базисе l:

$$\hat{\varepsilon} = \sum_{r=1}^{n-k-1} \eta_r l_r = l\eta,$$

где $\eta^T = (\eta_1, \ldots, \eta_{n-k-1})$.

Очевидно, что

$$\eta = l^T\hat{\varepsilon}.$$

В силу второй группы предположений вектор $\hat{\varepsilon}$ подчиняется многомерному нормальному распределению, но тогда η также подчиняется многомерному нормальному распределению в виду линейности преобразования.

Нетрудно увидеть, что $E\eta = I^T E\widehat{\varepsilon} = 0$,

$$D\eta = I^T D\hat{\varepsilon} I = \sigma^2 I^T P^{\perp} I = \sigma^2 I^T I = \sigma^2 E_{n-k-1},$$

так как $P^{\perp}$ матрица ортогонального проектирования на $L^{\perp}$, но все $l_i \in L^{\perp}$, $i = 1, \ldots, n-k-1$. Кроме того,

$$\hat{\varepsilon}^T \hat{\varepsilon} = \eta^T I^T I \eta = \eta^T \eta.$$

Следовательно, статистики $\hat{\varepsilon}^T \hat{\varepsilon}/\sigma^2$ и $\eta^T \eta/\sigma^2$ распределены одинаково, но из определения распределения χ^2 получаем, что статистика $\eta^T \eta/\sigma^2$ подчиняется распределению χ^2 с $n-k-1$ степенями свободы.

Лемма 2

Пусть выполнены предположения первой и второй групп, тогда справедливы утверждения:

1. *Вектор оценок $\hat{\beta}$ подчиняется многомерному нормальному распределению, $\hat{\beta} \sim N(\beta, \sigma^2 (X^T X)^{-1})$.*
2. *Статистика $(n-k-1)S^2/\sigma^2$ подчиняется распределению χ^2 с $(n-k-1)$ степенями свободы и взаимно независима с вектором оценок $\hat{\beta}$.*

Доказательство

Первое утверждение сразу следует из линейности явного выражения $\hat{\beta}$ и второй группы предположений:

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

линейное преобразование не меняет тип нормального распределения.

Вектор математических ожиданий и дисперсионная матрица были вычислены ранее. Второе утверждение следует из леммы 1 и того факта, что $E\{(\hat{\beta}-\beta)\hat{\varepsilon}^T\}=0$.
Действительно, как нетрудно заметить

$$\begin{pmatrix} \hat{\beta} \\ \hat{\varepsilon} \end{pmatrix} = \begin{pmatrix} (X^TX)^{-1}X^T \\ P^{\perp} \end{pmatrix} Y.$$

Следовательно, вектор $(\hat{\beta}^T, \hat{\varepsilon}^T)$ подчиняется многомерному нормальному распределению, но тогда вектора $\hat{\beta}$ и $\hat{\varepsilon}$ взаимно независимы.

Построение доверительных интервалов и проверка статистических гипотез

При выполнении всех условий из первой и второй групп предположений регрессионного анализа справедлива лемма 2 предыдущего параграфа, из которой сразу следует, что статистика

$$\frac{\widehat{\beta}_j - \beta_j}{s\sqrt{[(X^TX)^{-1}]_{(j+1)(j+1)}}} \sim T_{n-k-1}, \;\; j = 0, \ldots, k, \tag{11}$$

где $[(X^TX)^{-1}]_{(j+1)(j+1)}$ — элемент стоящий на главной диагонали в строке $j+1$ и столбце $j+1$ матрицы $(X^TX)^{-1}$, распределение T_{n-k-1} — распределение Стьюдента с $n-k-1$ степенями свободы.
Доказательство следует из определения распределения Стьюдента и леммы 2.
Из формулы (11) следует формула для доверительного интервала с уровнем доверия $1-\alpha$ для любого параметра β_j, $j = 0, 1, \ldots, k$.

Доверительный интервал имеет вид:

$$P\left\{\hat{\beta}_j - t_{1-\frac{\alpha}{2},n-k-1} S\sqrt{[(X^TX)^{-1}]_{(j+1)(j+1)}} < \beta_j < \right.$$
$$\left. < \hat{\beta}_j + t_{1-\frac{\alpha}{2},n-k-1} S\sqrt{[(X^TX)^{-1}]_{(j+1)(j+1)}}\right\} = 1-\alpha,$$

где $t_{1-\frac{\alpha}{2},n-k-1}$ — квантиль уровня $1-\frac{\alpha}{2}$ распределения Стьюдента T_{n-k-1}.

Важное значение имеет проверка гипотез статистической значимости найденных оценок $\widehat{\beta}_j$, $j = 0, 1, \ldots, n$.
Если доверительный интервал для параметра β_j с уровнем доверия $1-\alpha$ содержит нуль, т. е. если концы доверительного интервала имеют разный знак, то нельзя отклонить гипотезу H_0: $\beta_j = 0$, следовательно, в этом случае найденная оценка $\hat{\beta}_j$ не может быть признана статистически значимой на уровне значимости α.
Отсюда следует правило проверки статистической значимости оценки $\hat{\beta}_j$ или, что тоже самое, правило проверки гипотезы H_0: $\beta_j = 0$.

Выберем уровень значимости α и вычислим статистику

$$t_{\beta_j} = \frac{\hat{\beta}_j}{s\sqrt{[(X^T X)^{-1}]_{(j+1)(j+1)}}}.$$

Если $|t_{\beta_j}| > t_{1-\frac{\alpha}{2},n-k-1}$, то гипотеза H_0 отклоняется, и оценка $\hat{\beta}_j$ признается статистически значимой на уровне значимости α.
Если $|t_{\beta_j}| \leqslant t_{1-\frac{\alpha}{2},n-k-1}$, то гипотеза H_0 не отклоняется, и оценка $\hat{\beta}_j$ признается статистически незначимой на уровне значимости α.

Нетрудно аналогичным образом проверить более общую гипотезу вида H_0: $\beta_j = \beta_j^{(0)}$. Рассуждая таким же образом, получаем правило проверки гипотезы H_0 на уровне значимости α:

- Если $\frac{|\hat{\beta}_j - \beta_j^{(0)}|}{s\sqrt{[(X^T X)^{-1}]_{(j+1)(j+1)}}} > t_{1-\frac{\alpha}{2},n-k-1}$, то гипотеза H_0: $\beta_j = \beta_j^{(0)}$ отклоняется на уровне значимости α.
- Если $\frac{|\hat{\beta}_j - \beta_j^{(0)}|}{s\sqrt{[(X^T X)^{-1}]_{(j+1)(j+1)}}} \leqslant t_{1-\frac{\alpha}{2},n-k-1}$, то гипотеза H_0: $\beta_j = \beta_j^{(0)}$ принимается на уровне значимости α.

В линейном регрессионном анализе **коэффициентом детерминации** R^2 называется квадрат коэффициента корреляции между наблюдаемыми значениями показателя $Y^T = (y_1, \ldots, y_n)$ и значениями эмпирической функции $\hat{Y}^T = (\hat{y}_1, \ldots, \hat{y}_n)$.

Так как $X_0^T = (1, \ldots, 1)$ и $\hat{Y} = PY$, причем, P — матрица ортогонального проектирования на подпространство $L(X_0, \ldots, X_n)$, то $PX_0 = X_0$ или $X_0^T = X_0^T P$.
Но тогда

$$X_0^T \hat{Y} = X_0^T PY = X_0^T Y$$

или

$$\sum_{j=1}^{n} \hat{y}_i = \sum_{i=1}^{n} y_i,$$

следовательно,

$$\bar{\hat{y}} = \frac{1}{n}\sum_{i=1}^{n} \hat{y}_i = \frac{1}{n}\sum_{i=1}^{n} y_i = \bar{y}.$$

По определению

$$R^2 = \frac{\left(\sum\limits_{i=1}^{n}(y_i - \overline{y})(\hat{y}_i - \overline{y})\right)^2}{\sum\limits_{i=1}^{n}(y_i - \overline{y})^2 \sum\limits_{i=1}^{n}(\hat{y}_i - \overline{y})^2} =$$

$$= \frac{((Y - \overline{y}X_0)^T(\hat{Y} - \overline{y}X_0))^2}{(Y - \overline{y}X_0)^T(Y - \overline{y}X_0)(\hat{Y} - \overline{y}X_0)^T(\hat{Y} - \overline{y}X_0)}$$

Заметим, что

$$(Y - \overline{y}X_0)^T(\hat{Y} - \overline{y}X_0) = ((Y - \hat{Y}) + (\hat{Y} - \overline{y}X_0))^T(\hat{Y} - \overline{y}X_0) =$$
$$= (\hat{Y} - \overline{y}X_0)^T(\hat{Y} - \overline{y}X_0),$$

так как вектор $Y - \hat{Y} \in L^{\perp}$, а $\hat{Y} - \overline{y}X_0 \in L(X_0, \ldots, X_k)$ и, следовательно, вектора $Y - \hat{Y}$ и $\hat{Y} - \overline{y}X_0$ ортогональны, т.е.

$$(Y - \hat{Y})^T(\hat{Y} - \overline{y}X_0) = 0.$$

Но тогда

$$R^2 = \frac{(\hat{Y} - \bar{y}X_0)^T(\hat{Y} - \bar{y}X_0)}{(Y - \bar{y}X_0)^T(Y - \bar{y}X_0)} = 1 - \frac{\hat{\varepsilon}^T\hat{\varepsilon}}{(Y - \bar{y}X_0)^T(Y - \bar{y}X_0)}. \qquad (12)$$

Для доказательства последнего равенства заметим, что

$$\begin{aligned}(\hat{Y} - \bar{y}X_0)^T(\hat{Y} - \bar{y}X_0) &= \\ &= ((\hat{Y} - Y) + (Y - \bar{y}X_0))^T((\hat{Y} - Y) + (Y - \bar{y}X_0)) = \\ &= (-\hat{\varepsilon} + (Y - \bar{y}X_0))^T(-\hat{\varepsilon} + (Y - \bar{y}X_0)) = \\ &= -\hat{\varepsilon}^T\hat{\varepsilon} + (Y - \bar{y}X_0)^T(Y - \bar{y}X_0),\end{aligned}$$

в виду того, что

$$\hat{\varepsilon}^T(Y - \bar{y}X_0) = \hat{\varepsilon}^T Y - \bar{y}\hat{\varepsilon}^T X_0 = \hat{\varepsilon}^T(Y - \hat{Y}) + \hat{\varepsilon}^T\hat{Y} - \bar{y}\hat{\varepsilon}^T X_0 = \hat{\varepsilon}^T\widehat{\varepsilon},$$

так как $\hat{\varepsilon} \in L^{\perp}$, $\hat{Y} \in L(X_0, \ldots, X_k)$, $X_0 \in L(X_0, \ldots, X_k)$.

Равенство (12) удобно в вычислительном отношении, кроме того, из него видно, что, если построенная линейная регрессия идеально точно соответствует наблюдениям, то $Y = \hat{Y}$ и, следовательно, $\hat{\varepsilon} = 0$, но тогда $R^2 = 1$.
Наоборот, если $\hat{\beta}_1 = \ldots = \hat{\beta}_k = 0$, т.е. линейная регрессия не зависит от $x_1, \ldots, x_k$, то $R^2 = 0$. Из определения коэффициента детерминации R^2 сразу следует, что $0 \leqslant R^2 \leqslant 1$.

Наиболее важным применением коэффициента R^2 является его использование при проверке гипотезы статистической значимости линейной регрессионной модели в целом, т.е. при проверке гипотезы H_0: $\beta_1 = \ldots = \beta_k = 0$. Если указанную гипотезу H_0 нельзя отвергнуть на уровне значимости α, то это означает, что на уровне значимости α построенная линейная регрессия статистически незначима. Малые значения R^2 свидетельствуют против статистической значимости построенной линейной регрессии.

Лемма 3

Пусть выполнены обе группы основных предположений линейного регрессионного анализа, тогда в предположении справделивости гипотезы H_0: $\beta_1 = \ldots = \beta_k = 0$ *статистика*

$$F = \frac{R^2}{1 - R^2}\frac{n - k - 1}{k} \sim \mathcal{F}_{k,n-k-1}$$

подчиняется распределению Фишера со степенями свободы k *и* $n - k - 1$*.*

Доказательство

Очевидно, что

$$\frac{R^2}{1 - R^2} = \frac{(\hat{Y} - \overline{y}X_0)^T(\hat{Y} - \overline{y}X_0)}{\hat{\varepsilon}^T\hat{\varepsilon}}.$$

При доказательстве леммы 2 было показано, что вектора $\hat{\beta}$ и $\hat{\varepsilon}$ взаимно независимы, так как $\hat{Y} = X\hat{\beta}$, и $\overline{y} = \frac{1}{n}\sum_{i=1}^{n}\hat{y}_i$, то числитель и знаменатель дроби взаимно независимы.

Кроме того, уже было ранее показано, что статистика

$$\frac{(n-k-1)S^2}{\sigma^2} = \frac{1}{\sigma^2}\hat{\varepsilon}^T\hat{\varepsilon} \sim \chi^2_{n-k-1}$$

подчиняется распределению χ^2 с $n-k-1$ степенями свободы. Найдем распределение числителя в предположении справедливости гипотезы H_0: $\beta_1 = \ldots = \beta_k = 0$. Заметим, что

$$\hat{Y} = PY = X\beta + P\varepsilon = \beta_0 X_0 + P\varepsilon,$$

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{n}\hat{y}_i = \beta_0 + \frac{1}{n}X_0^T P\varepsilon = \beta_0 + \frac{1}{n}X_0^T\varepsilon$$

Тогда

$$\hat{Y} - \overline{y}X_0 = P\varepsilon - \frac{1}{n}X_0^T\varepsilon X_0 = P\varepsilon - \overline{\varepsilon}PX_0 = P(\varepsilon - \overline{\varepsilon}X_0),$$

где $\overline{\varepsilon} = \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i$, $X_0 = PX_0$.

Как видим

$$\eta = \hat{Y} - \overline{y}X_0 = P(E_n - \frac{1}{n}T_n)\varepsilon, \qquad (13)$$

где $T_n = X_0 X_0^T$ — квадратная матрица порядка $n \times n$, все элементы которой являются единицами. Нетрудно заметить, что $\eta \in L(X_0, \ldots, X_k)$, кроме того, вектор η ортогонален вектору X_0, следовательно, $\eta \in L(\tilde{X}_1, \ldots, \tilde{X}_k)$, где базисные вектора определяются следующими равенствами

$$\tilde{X}_r = X_r - \overline{x}_r X_0, \ \ \overline{x}_r = \sum_{i=1}^{n} x_{ir}/n.$$

Нетрудно убедиться, что все вектора $\tilde{X}_r$ ортогональны вектору X_0, следовательно, вектор X_0 ортогонален подпространству $L(\tilde{X}_1, \ldots, \tilde{X}_k)$. Кроме того, из (13) следует, что вектор η подчиняется многомерному нормальному распределению, $E\eta = 0$.

Нетрудно проверить, что матрица $P_1 = P(E_n - T_n/n)$ — матрица ортогонального проектирования на подпространство $L(\tilde{X}_1, \dots, \tilde{X}_k)$.
Вектор $\eta = P_1\varepsilon$, следовательно, $D\eta = \sigma^2 P_1$.
Все свойства, которые были сформулированы для матрицы P, выполняются и для матрицы P_1. Выберем в подпространстве $L(\tilde{X}_1, \dots, \tilde{X}_k)$ ортонормированный базис $d = (d_1, \dots, d_k)$, тогда

$$\eta = \sum_{r=1}^{k} \xi_r d_r = d\xi, \quad \xi = d^T \eta,$$

$$\xi = (\xi_1, \dots, \xi_k)^T \sim N(0, \sigma^2 E_k),$$

так как

$$D\xi = E\{(d^T\eta)(\eta^T d)\} = \sigma^2 d^T P_1 P_1 d = \sigma^2 d^T d = \sigma^2 E_k.$$

Кроме того,

$$\eta^T \eta = \xi^T d^T d\xi = \xi^T \xi.$$

Следовательно, распределения левой и правой частей совпадают. Но тогда из определения распределения χ^2 получаем, что статистика

$$\frac{1}{\sigma^2}(\hat{Y} - \overline{y}X_0)^T(\hat{Y} - \overline{y}X_0) \sim \chi^2_k$$

подчиняется распределению χ^2 с k степенями свободы.

Построенная линейная регрессия статистически значима на уровне α тогда и только тогда, когда гипотеза H_0: $\beta_1 = \ldots = \beta_k = 0$ отклоняется на уровне значимости α. Поэтому правило проверки статистической значимости линейной регрессии в целом сформулируем следующим образом:

- Если $F = \frac{R^2}{1-R^2}\frac{n-k-1}{k} > F_{1-\alpha;k,n-k-1}$, то гипотеза H_0: $\beta_1 = \ldots = \beta_k = 0$ отклоняется на уровне значимости α и, следовательно, построенная линейная регрессия является статистически значимой.
- Если $F \leqslant F_{1-\alpha;k,n-k-1}$, то гипотеза H_0 принимается, и, следовательно, построенная линейная регрессия является статистически незначимой, здесь $F_{1-\alpha;k,n-k-1}$ — квантиль уровня $1-\alpha$ распределения Фишера с k и $n-k-1$ степенями свободы.

Нелинейные регрессионные модели, сводяшиеся к ленийным при помощи замены переменных

Регрессионная модель от k факторов и l параметров может быть записана в виде $y = f(x_1, \ldots, x_k, \beta_1, \ldots, \beta_l)$. Среди них есть нелинейные как по факторам, так и по параметрам. Существуют нелинейные модели, которые путем преобразований могут быть приведены к линейным. Преобрзования применяются как к отклику, так и к факторам.

- *Обратное преобразование:* $Y = \beta_0 + \beta_1(1/X) + \varepsilon$.
  Замена $Z = 1/X$.
- *Логарифмическое преобразование:* $Y = \beta_0 + \beta_1 \ln X + \varepsilon$.
  Замена $Z = \ln X$.
- *Преобразование типа квадратного корня:* $Y = \beta_0 + \beta_1 \sqrt{X} + \varepsilon$.
  Замена $Z = \sqrt{X}$.

- *Мультипликативная модель:* $Y = \alpha X^{\beta}\varepsilon$.
  α, β — неизвестные параметры, ε — мультипликативная случайная ошибка, которая имеет непрерывное распределение с математическим ожиданием, равным 1, и конечной дисперсией.

$$\ln Y = \ln\alpha + \beta\ln X + \ln\varepsilon.$$

  Следует помнить, что оценки границ доверительных интравлов будут обоснованными, только если логарифм ошибок имеет нормально распределение: $\ln\varepsilon \in N(0,\sigma^2)$. Замена $Z = \sqrt{X}$.
- *Экспоненциальная модель:* $Y = \alpha e^{\beta_1 X}\varepsilon$.

$$\ln Y = \ln\alpha + \beta_1 X + \ln\varepsilon.$$

- *Обратная модель:* $Y = \dfrac{1}{\beta_0 + \beta_1 X + \varepsilon}$.

$$1/Y = \beta_0 + \beta_1 X + \varepsilon.$$

- *Обратная экспоненциальная модель:* $Y = \dfrac{1}{1 + \alpha e^{\beta_1 X + \varepsilon}}$.

$$\ln(1/Y - 1) = \ln \alpha + \beta_1 X + \varepsilon.$$

При использовании преобразований, особенно отклика, следует проверять предпосылки регрессионного анализа: независимость ошибок, нормальность их распределений.