А.Г. Белов

ВЕРОЯТНОСТНО-СТАТИСТИЧЕСКИЙ ПОДХОД К МЕТОДУ НАИМЕНЬШИХ КВАДРАТОВ

Широкое и часто формальное применение метода наименьших квадратов (МНК) при оценивании параметров моделей процессов в различных областях естествознания заставляет более подробно исследовать основные положения возникновения и использования данного метода. В работе, на примере линейной множественной регрессионной модели, описывается вероятностно-статистический подход в оценке параметров модели с помощью МНК. Он позволил получить ряд общих вероятностных и статистических формул для оценок параметров модели вне зависимости от допустимого совместного распределения исследуемых величин, глубже оценить риски в оценке параметров, связанные с неточностями задания модели, указать место и роль знания теоретического и эмпирического распределения ошибок измерения.

1. Постановка задачи

На практике часто с помощью экспериментального наблюдения исследуется зависимость

$$y = g(\mathbf{x}) \tag{1}$$

неслучайных количественных показателей некоторого объекта, переменной "отклика" $y \in R^1$ от факторной переменной $\mathbf{x} = (x_1, \ldots, x_m)^T \in R^m$. При этом разным вариантам проведения эксперимента (активный или пассивный) соответствует своя природа искажения наблюдаемых показателей и, соотвественно, методы обработки экспериментальных данных. Для пассивного эксперимента характерно неуправляемое изменение значений факторов от опыта к опыту, в то время как при активном эксперименте значения факторов для каждого опыта выбираются исследователем. Если значение какого-либо фактора не изменяется от опыта к опыту, то его значение (без уменьшения общности) можно считать равным единице. Примером такого фактора может являться наличие систематической ошибки при проведении эксперимента.

В любом случае для установления зависимости (1) проводят эксперимент с большим числом n опытов и статистические данные состоят из множества наблюдавшихся значений факторов и соответствующих значений "откликов" и имеют вид $(x_{i1}, \ldots, x_{im},\ y_i)$, $i = 1, \ldots, n$.

Однако из-за присутствия случайности при проведении опыта каждому вектору значений $\mathbf{x}_i$ факторов соответствует в общем случае множество значений y_{ij} зависимой переменной. Все это позволяет интерпретировать экспериментально исследуемую зависимость (1) как стохастическую, а наблюдаемые величины $\mathbf{x}$, $y(\mathbf{x})$ как случайные величины (с.в.) $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_m)^T, \eta|\boldsymbol{\xi}$, с функциями плотности распределения $f_{\boldsymbol{\xi}}(\mathbf{x})$, $f_{\eta|\boldsymbol{\xi}}(y|\mathbf{x})$, соответственно. В дальнейшем ограничимся рассмотрением лишь абсолютно непрерывных либо дискретных распределений и используем общее обозначение $f_{\boldsymbol{\xi}\eta}(\mathbf{x}, y)$ для совместной плотности — в первом случае и совместной вероятности $P\{\boldsymbol{\xi} = \mathbf{x},\ \eta = y\}$ — во втором, называя в любом случае функцию $f_{\boldsymbol{\xi}\eta}(\mathbf{x}, y)$ плотностью, определяемой в абсолютно непрерывном случае как $f_{\boldsymbol{\xi}\eta}(\mathbf{x}, y) = f_{\eta|\boldsymbol{\xi}}(y|\mathbf{x}) f_{\boldsymbol{\xi}}(\mathbf{x})$.

При таких условиях одним из способов определения исходной зависимости (1) является использование следующих фактов: во-первых, искомое истинное значение y (1) при заданном $\mathbf{x}$ находится в "центре" распределения значений условной с.в. $\eta|\boldsymbol{\xi}$ и может быть приближено, например, ее условным математическим ожиданием (м.о.) $E(\eta|\boldsymbol{\xi} = \mathbf{x})$, называемым регрессией η на $\boldsymbol{\xi}$ и являющейся функцией от $\mathbf{x}$; во-вторых, поиск или приближение неизвестной истинной функциональной зависимости $g(\mathbf{x})$ (1) производится внутри параметризованного семейства функций $h(\mathbf{x}; \beta)$, где $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)^T \in R^k$.

Таким образом, задача определения зависимости (1) по экспериментальым данным заменяется задачей определения по значениям регрессии $E(\eta|\boldsymbol{\xi})$ значений параметров в системе уравнении

$$E(\eta|\boldsymbol{\xi} = \mathbf{x}_i) = h(\boldsymbol{\xi} = \mathbf{x}_i; \boldsymbol{\beta}),\ i = 1, \ldots, n. \qquad (2)$$

Однако в решении (2) имеются две трудности: во-первых, для вычисления $E(\eta|\boldsymbol{\xi})$ надо абсолютно точно знать закон распределения $f_{\eta|\boldsymbol{\xi}}(y|\mathbf{x})$ или для ее оценки иметь достаточно большой объем экспериментальных измерений для каждого значения факторной переменной и, во-вторых, явное вычисление $E(\eta|\boldsymbol{\xi})$ может оказаться трудной аналитической задачей. Чтобы этого избежать может быть предложен другой подход, основанный на оптимальном свойстве м.о., реализуемом в МНК.

2. МНК

Как известно [3, с.505], регрессия $E(\eta|\boldsymbol{\xi})$ является наилучшей, в среднем квадратичном, из всех возможных функций приближением к наблюдаемой величине, если измерять расстояние вдоль оси "откликов", то есть справедливо равенство

$$E(\eta - E(\eta|\boldsymbol{\xi}))^2 = \min_{g(\boldsymbol{\xi})} E(\eta - g(\boldsymbol{\xi}))^2.$$

Поскольку вместо перебора всевозможных функций $g(\boldsymbol{\xi})$ на практике ограничиваются функциями некоторого параметризованного семейства $h(\boldsymbol{\xi};\boldsymbol{\beta})$, то вместо $E(\eta - g(\boldsymbol{\xi}))^2$ рассматривают выражение $E(\eta - h(\boldsymbol{\xi};\boldsymbol{\beta}))^2$, которое (в случае распределения, при котором регрессия $E(\eta|\boldsymbol{\xi})$ существует) может быть записано в виде [1, с.302]

$$E(\eta - h(\boldsymbol{\xi};\boldsymbol{\beta}))^2 = E(\eta - E(\eta|\boldsymbol{\xi}))^2 + 2E\left[(\eta - E(\eta|\boldsymbol{\xi}))(E(\eta|\boldsymbol{\xi}) - h(\boldsymbol{\xi};\boldsymbol{\beta}))\right] + \\ + E(E(\eta|\boldsymbol{\xi}) - h(\boldsymbol{\xi};\boldsymbol{\beta}))^2.$$

Однако второе слагаемое в этом выражении равно нулю

$$\begin{array}{l} E\left[(\eta - E(\eta|\boldsymbol{\xi}))(E(\eta|\boldsymbol{\xi}) - h(\boldsymbol{\xi};\boldsymbol{\beta}))\right] = \\ E_{\xi}\left\{E_{\eta}\left[(\eta - E(\eta|\boldsymbol{\xi}))(E(\eta|\boldsymbol{\xi}) - h(\boldsymbol{\xi};\boldsymbol{\beta}))|\boldsymbol{\xi} = \mathbf{x}\right]\right\} = \\ E_{\xi}\left\{(E(\eta|\boldsymbol{\xi}) - h(\boldsymbol{\xi};\boldsymbol{\beta}))E_{\eta}\left[(\eta - E(\eta|\boldsymbol{\xi}))|\boldsymbol{\xi} = \mathbf{x}\right]\right\} = \\ E_{\xi}\left\{(E(\eta|\boldsymbol{\xi}) - h(\boldsymbol{\xi};\boldsymbol{\beta}))(E(\eta|\boldsymbol{\xi}) - E(\eta|\boldsymbol{\xi}))\right\} = 0. \end{array}$$

Тогда для любых $\boldsymbol{\beta} = (\beta_1, ..., \beta_k)^T$ получаем

$$E(\eta - h(\boldsymbol{\xi};\boldsymbol{\beta}))^2 = E(\eta - E(\eta|\boldsymbol{\xi}))^2 + E(E(\eta|\boldsymbol{\xi}) - h(\boldsymbol{\xi};\boldsymbol{\beta}))^2.$$

В последнем равенстве первое слагаемое не зависит от вектора параметров, так что второе слагаемое достигает своего минимума при тех же значениях $\widehat{\boldsymbol{\beta}} = (\widehat{\beta}_1, \ldots, \widehat{\beta}_k)^T$, что и левая часть. Отсюда непосредственно вытекает, что в случае, когда кривая регрессии $E(\eta|\boldsymbol{\xi} = \mathbf{x})$ есть кривая семейства $h(\mathbf{x};\boldsymbol{\beta})$, то она совпадает с кривой функции $h(\mathbf{x};\widehat{\boldsymbol{\beta}})$ Таким образом, вычисление параметров можно производлить как на основе $E(\eta|\boldsymbol{\xi})$, так и с.в. η, решая соответствующие задачи минимизации

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} E(E(\eta|\boldsymbol{\xi}) - h(\boldsymbol{\xi};\boldsymbol{\beta}))^2,$$

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} E(\eta - h(\boldsymbol{\xi};\boldsymbol{\beta}))^2. \tag{3}$$

При этом функция $h(\mathbf{x};\widehat{\boldsymbol{\beta}})$ даст наилучшее приближение к кривой регрессии $E(\eta|\boldsymbol{\xi})$ в среднем квадратичном в семействе функций $h(\boldsymbol{\xi};\boldsymbol{\beta})$, если измерять расстояние вдоль оси "отликов" и поэтому часто называется средней квадратичной регрессией, а метод решения задач минимизации (3) называют МНК. Часто задачи (3) приходиться решать численно, но для семейства линейных функций их можно решить явно.

3. Оценка параметров

В случае линейной средней квадратичной регрессии посредством k факторов необходимо найти наилучшую линейную аппроксимацию $h(\mathbf{x};\boldsymbol{\beta}) = \sum\limits_{j=1}^{k}\beta_j x_j = \boldsymbol{\beta}^T\mathbf{x}$ с.в. η (или регрессии $E(\eta|\boldsymbol{\xi})$) $\mathbf{x} = (x_1,\ldots,x_k)^T$, для которой м.о. величины $(\eta - h(\boldsymbol{\xi};\boldsymbol{\beta}))^2$ (или $(E(\eta|\boldsymbol{\xi}) - h(\boldsymbol{\xi};\boldsymbol{\beta}))^2$) принимает наименьшее возможное значение. Для минимизируемой функции (3) справедливы равенства

$$
\begin{aligned}
&M(\boldsymbol{\beta}) = E(\eta - h(\boldsymbol{\xi};\boldsymbol{\beta}))^2 = E\left(\eta - \boldsymbol{\beta}^T\boldsymbol{\xi} - E\eta + E\eta - \boldsymbol{\beta}^T E\boldsymbol{\xi} + \boldsymbol{\beta}^T E\boldsymbol{\xi}\right)^2 = \\
&= E\left((\eta - E\eta) - \boldsymbol{\beta}^T(\boldsymbol{\xi} - E\boldsymbol{\xi}) + (E\eta - \boldsymbol{\beta}^T E\boldsymbol{\xi})\right)^2 = \\
&= \boldsymbol{\beta}^T D\boldsymbol{\xi}\boldsymbol{\beta} - 2\boldsymbol{\beta}^T cov(\boldsymbol{\xi},\eta) + D\eta + (E\eta - \boldsymbol{\beta}^T E\boldsymbol{\xi})^2 = \\
&= \boldsymbol{\beta}^T D\boldsymbol{\xi}\boldsymbol{\beta} + \boldsymbol{\beta}^T(E\boldsymbol{\xi})(E\boldsymbol{\xi})^T\boldsymbol{\beta} - 2\boldsymbol{\beta}^T cov(\boldsymbol{\xi},\eta) - 2\boldsymbol{\beta}^T E\boldsymbol{\xi}E\eta + D\eta + (E\eta)^2 = \\
&= \boldsymbol{\beta}^T E(\boldsymbol{\xi}\boldsymbol{\xi}^T)\boldsymbol{\beta} - 2\boldsymbol{\beta}^T E(\boldsymbol{\xi}\eta) + E(\eta^2),
\end{aligned}
$$

где $E(\boldsymbol{\xi}\boldsymbol{\xi}^T) = \|E(\xi_i\xi_j)\|_1^k$, $E\boldsymbol{\xi} = (E\xi_1,\ldots,E\xi_k)^T$, $D\boldsymbol{\xi} = E(\boldsymbol{\xi}\boldsymbol{\xi}^T) - (E\boldsymbol{\xi})(E\boldsymbol{\xi})^T$, $D\eta = E(\eta^2) - (E\eta)^2$, $cov(\boldsymbol{\xi},\eta) = (cov(\xi_1,\eta),\ldots,cov(\xi_k,\eta))^T = E(\boldsymbol{\xi}\eta) - E\boldsymbol{\xi}E\eta$. Необходимое условие экстремума $\dfrac{\delta}{\delta\boldsymbol{\beta}}M(\boldsymbol{\beta}) = 0$ приводит к равенству, называемому нормальной системой уравнений, вида

$$E(\boldsymbol{\xi}\boldsymbol{\xi}^T)\boldsymbol{\beta} = E(\boldsymbol{\xi}\eta).$$

Обозначая матрицу $\mathbf{V} = E(\boldsymbol{\xi}\boldsymbol{\xi}^T)$ и предполагая, что ее определитель $|\mathbf{V}| \neq 0$, после умножения последнего равенства слева на обратную $\mathbf{V}^{-1}$ получим

$$\widehat{\boldsymbol{\beta}} = \mathbf{V}^{-1}E(\boldsymbol{\xi}\eta). \tag{4}$$

Поскольку матрица вторых производных равна $\dfrac{\delta^2}{\delta\boldsymbol{\beta}^2}M(\boldsymbol{\beta}) = 2\mathbf{V}$, а невырожденная симметричная матрица $\mathbf{V}$ является положительно определенной, то, учитывая (4), получим минимальное значение $M(\boldsymbol{\beta})$

$$M(\widehat{\boldsymbol{\beta}}) = E(\eta^2) - E(\boldsymbol{\xi}\eta)^T\mathbf{V}^{-1}E(\boldsymbol{\xi}\eta). \tag{5}$$

Для различных линейных моделей получим соответствующие представления формул (4), (5).

4. Модели

4.1. Модель 1

Рассмотрим однофакторную линейную регрессионную модель, $k=1$, когда на "отклик" влияет один фактор, значение которого меняется от опыта к опыту, а именно

$$E(\eta|\xi)=\beta\xi.$$

Тогда, очевидно, (4), (5) принимают соответственно вид

$$\widehat{\beta}=\frac{E(\xi\eta)}{E(\xi^2)}=\frac{cov(\xi,\eta)+E\xi E\eta}{D\xi+(E\xi)^2}=\frac{\rho\sigma_\xi\sigma_\eta+E\xi E\eta}{D\xi+(E\xi)^2},$$
$$M(\widehat{\beta})=E(\eta^2)-\frac{(E(\xi\eta))^2}{E(\xi^2)},$$

где $\rho=\dfrac{cov(\xi,\eta)}{\sigma_\eta\sigma_\xi}$ — коэффициент корреляции, $\sigma_\xi=\sqrt{D\xi}$ и $\sigma_\eta=\sqrt{D\eta}$ — средние квадратические отклонения. Отметим, что если усреднить левую и правую части модели по с.в. ξ, то получим $E_\xi(E_\eta(\eta|\xi))=\beta E\xi$, а учитывая $E_\xi(E_\eta(\eta|\xi))=E\eta$ получим другую оценку параметра, не являющуюся в общем случае наилучшей в среднем квадратичном, а именно

$$\widetilde{\beta}=\frac{E\eta}{E\xi},$$
$$M(\widetilde{\beta})=\left(\frac{E\eta}{E\xi}\right)^2E(\xi^2)-2\frac{E\eta}{E\xi}E(\xi\eta)+E(\eta^2).$$

Для доказательства $M(\widehat{\beta})<M(\widetilde{\beta})$ достаточно сравнить величины

$$-\frac{(E(\xi\eta))^2}{E(\xi^2)}+2\frac{E\eta}{E\xi}E(\xi\eta),\ \left(\frac{E\eta}{E\xi}\right)^2E(\xi^2).$$

Но первое выражение представимо в виде

$$\left(\frac{E\eta}{E\xi}\right)^2E(\xi^2)-\left(\frac{E(\xi\eta)}{\sqrt{E(\xi^2)}}-\frac{E\eta}{E\xi}\sqrt{E(\xi^2)}\right)^2,$$

что и доказывает требуемое. В случае постоянства ξ оценки совпадают $\widehat{\beta}=\widetilde{\beta}=\dfrac{E\eta}{\xi}$.

4.2. Модель 2

В случае однофакторной модели с неизменяемым значением фактора, $\xi = 1$, имеющей вид

$$E(\eta|\xi) = \beta,$$

из (4), (5) имеем

$$\begin{aligned}
&\widehat{\beta} = E(\eta), \\
&M(\widehat{\beta}) = E(\eta^2) - (E\eta)^2 = D\eta.
\end{aligned}$$

4.3. Модель 3

Для двухфакторной модели, в которой значение одного из факторов не изменяется, называемой часто простой или "парной" линейной регрессией,

$$E(\eta|\boldsymbol{\xi}) = \beta_1 + \beta_2\xi,$$

имеем

$$\boldsymbol{\xi} = \begin{pmatrix} 1 \\ \xi \end{pmatrix}, \ \mathbf{V} = \begin{pmatrix} 1 & E\xi \\ E\xi & E(\xi^2) \end{pmatrix}, \ \mathbf{V}^{-1} = \frac{1}{D\xi}\begin{pmatrix} E(\xi^2) & -E\xi \\ -E\xi & 1 \end{pmatrix},$$

$$E(\boldsymbol{\xi}\eta) = \begin{pmatrix} E(\eta) \\ E(\xi\eta) \end{pmatrix}.$$

Подставляя в (4), (5) получим широко известные равенства

$$\begin{aligned}
&\widehat{\beta}_1 = \frac{E\eta E(\xi^2) - E\xi E(\xi\eta)}{D\xi} = E\eta - \widehat{\beta}_2 E\xi, \\
&\widehat{\beta}_2 = \frac{E(\xi\eta) - E\xi E\eta}{D\xi} = \frac{cov(\xi,\eta)}{D\xi} = \frac{\rho\sigma_\eta}{\sigma_\xi}, \\
&M(\widehat{\boldsymbol{\beta}}) = \sigma_\eta^2(1 - \rho^2).
\end{aligned}$$

4.4. Модель 4

Для двухфакторной модели, $k = 2$, в которой на "отклик" влияют два изменяемых фактора

$$E(\eta|\boldsymbol{\xi}) = \beta_1\xi_1 + \beta_2\xi_2,$$

имеем

$$\mathbf{V} = \begin{pmatrix} E(\xi_1^2) & E(\xi_1\xi_2) \\ E(\xi_1\xi_2) & E(\xi_2^2) \end{pmatrix}, \ \mathbf{V}^{-1} = \frac{1}{|\mathbf{V}|}\begin{pmatrix} E(\xi_2^2) & -E(\xi_1\xi_2) \\ -E(\xi_1\xi_2) & E(\xi_1^2) \end{pmatrix},$$

$$E(\boldsymbol{\xi}\eta) = \begin{pmatrix} E(\xi_1\eta) \\ E(\xi_2\eta) \end{pmatrix},$$

где $|\mathbf{V}| = E\xi_1^2 E\xi_2^2 - (E(\xi_1\xi_2))^2$. Подставляя в (4), (5) получим

$$\widehat{\beta}_1 = \frac{E(\xi_1\eta)E(\xi_2^2) - E(\xi_2\eta)E(\xi_1\xi_2)}{E(\xi_1^2)E(\xi_2^2) - (E(\xi_1\xi_2))^2}, \ \widehat{\beta}_2 = \frac{E(\xi_2\eta)E(\xi_1^2) - E(\xi_1\eta)E(\xi_1\xi_2)}{E(\xi_1^2)E(\xi_2^2) - (E(\xi_1\xi_2))^2},$$
$$M(\widehat{\boldsymbol{\beta}}) = E(\eta^2) - (E(\xi_1\eta))^2 E(\xi_2^2) + 2E(\xi_1\eta)E(\xi_2\eta)E(\xi_1\xi_2) - (E(\xi_2\eta))^2 E(\xi_1^2)$$

4.5. Модель 5

Наконец, рассмотрим $k-$ факторную модель, в которой один из факторов не изменяется, а оставшиеся $(k-1)$ фактор изменяются от опыта к опыту, воздействуя на "отклик"

$$E(\eta|(1,\boldsymbol{\xi})) = \beta_1 + \boldsymbol{\beta}^T\boldsymbol{\xi},$$

где $\boldsymbol{\xi} = (\xi_2, \ldots, \xi_k)^T$, $\boldsymbol{\beta} = (\beta_2, \ldots, \beta_k)^T$. В этом случае имеем

$$\mathbf{V} = \begin{pmatrix} 1 & E\boldsymbol{\xi}^T \\ E\boldsymbol{\xi} & E(\boldsymbol{\xi}\boldsymbol{\xi}^T) \end{pmatrix}, \ E(\boldsymbol{\xi}\eta) = \begin{pmatrix} E\eta \\ E(\boldsymbol{\xi}\eta) \end{pmatrix}.$$

Для обращения $\mathbf{V}$ используем один из вариантов широко известных формул Фробениуса блочного обращения матриц [2, с.60]

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}\mathbf{H}^{-1}\mathbf{C}\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B}\mathbf{H}^{-1} \\ -\mathbf{H}^{-1}\mathbf{C}\mathbf{A}^{-1} & \mathbf{H}^{-1} \end{pmatrix},$$

где $\mathbf{H} = \mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}$. Тогда обратная $\mathbf{V}$ матрица имеет вид

$$\mathbf{V}^{-1} = \begin{pmatrix} 1 + E\boldsymbol{\xi}^T[D\boldsymbol{\xi}]^{-1}E\boldsymbol{\xi} & -E\boldsymbol{\xi}^T[D\boldsymbol{\xi}]^{-1} \\ -[D\boldsymbol{\xi}]^{-1}E\boldsymbol{\xi} & [D\boldsymbol{\xi}]^{-1} \end{pmatrix}.$$

Подставляя в (4), (5) получим соотношения [3, с.509]

$$\widehat{\beta}_1 = E\eta - \widehat{\boldsymbol{\beta}}^T E\boldsymbol{\xi}, \ \widehat{\boldsymbol{\beta}} = [D\boldsymbol{\xi}]^{-1} cov(\boldsymbol{\xi}, \eta),$$
$$M(\widehat{\beta}_1, \widehat{\boldsymbol{\beta}}) = D\eta(1 - \rho^2_{\eta\boldsymbol{\xi}}),$$

где $\rho^2_{\eta\boldsymbol{\xi}} = \dfrac{cov(\boldsymbol{\xi},\eta)^T[D\boldsymbol{\xi}]^{-1}cov(\boldsymbol{\xi},\eta)}{D\eta}$ — множественный коэффициент корреляции.

Таким образом, полученная формула (4), как и приведенные в моделях ее известные аналоги, применима для оценки параметров средней квадратической линейной множественной регрессии при известном допустимом совместном распределении с.в. $\boldsymbol{\xi}$ и η. Однако свойства получаемых оценок зависит от многих факторов, описанных в специальной литературе по статистическому анализу. Приведем примеры для некоторых известных теоретических и эмпирических распределений.

5. Примеры

5.1. Пример 1

Пусть случайный вектор (ξ,η) имеет двумерное невырожденное (собственное) нормальное распределение

$$\mathcal{N}_2\left(\boldsymbol{\mu} = (\mu_\xi,\mu_\eta)^T,\ \ \boldsymbol{\Sigma} = \begin{pmatrix} \sigma^2_\xi & \rho\sigma_\xi\sigma_\eta \\ \rho\sigma_\xi\sigma_\eta & \sigma^2_\eta \end{pmatrix}\right),$$

где $|\rho| < 1$. Тогда, как известно [3, с.87], функция регрессии η на ξ линейная

$$E(\eta|\xi = x) = \left(\mu_\eta - \mu_\xi\frac{\sigma_\eta}{\sigma_\xi}\rho\right) + \frac{\sigma_\eta}{\sigma_\xi}\rho x = \mu_\eta + \frac{\sigma_\eta}{\sigma_\xi}\rho(x-\mu_\xi).$$

Этот же результат получается с учетом формул оценок параметров для модели 3 регрессии η на ξ.

Примером указанного двумерного нормального распределения случайного вектора (ξ,η) в экспериментальной практике может являться случайный эксперимент, в котором на исход $i-$ого опыта y_i, описываемого с.в. η, влияют случайные неконтролируемый фактор x_i (с.в. ξ) и случайная независимая от x_i аддитивная ошибка измерения e_i (или неконтролируемый случайный фактор, с.в. ε, независимая от ξ). При этом предполагается, что наблюдаемые величины связаны равенством

$$y_i(x_i) = \beta_1 + \beta_2 x_i + e_i,$$

а соответствующие левой и правой частям с.в. эквивалентны

$$\eta|\xi \sim \beta_1 + \beta_2\xi + \varepsilon,$$

где с.в. $\varepsilon \sim \mathcal{N}_1(0, \sigma_\eta^2(1-\rho^2))$, с.в. $\xi \sim \mathcal{N}_1(\mu_\xi, \sigma_\xi^2)$, а $\beta_1 = \left(\mu_\eta - \mu_\xi \frac{\sigma_\eta}{\sigma_\xi}\rho\right)$ и $\beta_2 = \frac{\sigma_\eta}{\sigma_\xi}\rho$. Нетрудно убедиться, что функция плотности вероятностей случайного вектора (ξ, η) соответствует двумерному нормальному распределению $\mathcal{N}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

$$f_{\xi\eta}(x,y) = f_{\eta|\xi}(y|x) f_\xi(x) = \frac{1}{2\pi\sigma_\xi\sigma_\eta\sqrt{1-\rho^2}} \times$$
$$\times exp\left\{ -\frac{1}{2(1-\rho^2)} \left[\frac{(x-\mu_\xi)^2}{\sigma_\xi^2} - \frac{2\rho(x-\mu_\xi)(y-\mu_\eta)}{\sigma_\xi\sigma_\eta} + \frac{(y-\mu_\eta)^2}{\sigma_\eta^2} \right] \right\}.$$

Описанный выше экспериментальный пример может быть использован для моделирования двумерного нормального распределения. Такой метод моделирования совпадает с наиболее распространенным алгоритмом [6, с.37] моделирования многомерного нормального распределения, основанного на, полученном при разложении Холесского ковариационной матрицы $\boldsymbol{\Sigma}$, линейном преобразовании вектора стандартных нормально распределенных с.в.

Действительно, пусть задан случайный вектор $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_k)^T$ со случайными нормально распределенными компонентами $\nu_i \sim \mathcal{N}_1(0,1)$. Поскольку ковариационная матрица $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^T > 0$, то как известно существует и единственна нижняя треугольная матрица $\mathbf{L}$ такая, что $\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}^T$ (разложение Холесского [5]). Тогда случайный вектор $(\xi, \eta)^T \sim \mathbf{L}\boldsymbol{\nu} + \boldsymbol{\mu}$ будет иметь двумерное нормальное распределение $\mathcal{N}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. В нашем $k = 2$ случае

$$\mathbf{L} = \begin{pmatrix} \sigma_\xi & 0 \\ \rho\sigma_\eta & \sigma_\eta\sqrt{1-\rho^2} \end{pmatrix}, \ \xi \sim l_{11}\nu_1 + \mu_\xi, \ \eta \sim l_{21}\nu_1 + l_{22}\nu_2 + \mu_\eta \ .$$

Тогда справедливы соотношения

$$\eta|\xi \sim \left(\mu_\eta - \mu_\xi \frac{l_{21}}{l_{11}}\right) + \frac{l_{21}}{l_{11}}\xi + l_{22}\nu_2 =$$
$$= \left(\mu_\eta - \mu_\xi \frac{\sigma_\eta}{\sigma_\xi}\rho\right) + \frac{\sigma_\eta}{\sigma_\xi}\rho\xi + \sigma_\eta\sqrt{1-\rho^2}\nu_2 = \beta_1 + \beta_2\xi + \varepsilon.$$

5.2. Пример 2

Пусть случайный вектор $(\boldsymbol{\xi}, \eta) \in R^k$ имеет $k-$мерное собственное нормальное распределение с вектором средних $\boldsymbol{\mu} = (\boldsymbol{\mu}_{\boldsymbol{\xi}}^T, \mu_\eta)^T$ и ковариационной матрицей в блочном представлении

$$\mathbf{\Sigma} = \|\sigma_{ij}\|_1^k = \begin{pmatrix} D\boldsymbol{\xi} & cov(\boldsymbol{\xi},\eta) \\ cov(\eta,\boldsymbol{\xi}^T) & D\eta \end{pmatrix},$$

где $\boldsymbol{\mu}_{\boldsymbol{\xi}}=(\mu_{\xi_1},\ldots,\mu_{\xi_{k-1}})^T,\ \boldsymbol{\xi}=(\xi_1,\ldots,\xi_{k-1})^T,\ D\boldsymbol{\xi}=\|cov(\xi_i,\xi_j)\|\in R^{(k-1)\times(k-1)}$, $cov(\boldsymbol{\xi},\eta)\in R^{(k-1)\times 1},\ cov(\eta,\boldsymbol{\xi}^T)\in R^{1\times(k-1)}$. Тогда функция регрессии η на $\boldsymbol{\xi}$ линейна [3, с.88]

$$E(\eta|\boldsymbol{\xi}=\mathbf{x}) = \mu_\eta - \sum_{i=1}^{k-1}(x_i-\mu_{\xi_i})\frac{\sigma^{ik}}{\sigma^{kk}},\ \mathbf{x}^T=(x_1,\ldots,x_{k-1}),$$

и может быть описана моделью 5 с коэффициентами

$$\beta_1 = \mu_\eta - \widehat{\boldsymbol{\beta}}^T\boldsymbol{\mu}_{\boldsymbol{\xi}},\ \boldsymbol{\beta} = -\frac{1}{\sigma^{kk}}\boldsymbol{\sigma},$$

где $\mathbf{\Sigma}^{-1}=\|\sigma^{ij}\|_1^k,\ \boldsymbol{\sigma}=(\sigma^{1k},\ldots,\sigma^{(k-1)k})^T$. Однако формулы оценок параметров для модели 5 дают возможность другого представления для коэффициентов регрессии $E(\eta|\boldsymbol{\xi})$ $k-$мерного нормального распределения

$$\beta_1 = \mu_\eta - \widehat{\boldsymbol{\beta}}^T\boldsymbol{\mu}_{\boldsymbol{\xi}},\ \boldsymbol{\beta} = [D\boldsymbol{\xi}]^{-1}cov(\boldsymbol{\xi},\eta).$$

Примером указанного k - мерного нормального распределения случайного вектора $(\boldsymbol{\xi},\eta)$ в экспериментальной практике может являться случайный эксперимент, в котором на исход $i-$ого опыта y_i, описываемого с.в. η, влияют случайные неконтролируемые $(k-1)$ факторы $\mathbf{x}_i^T=(x_{i1},\ldots,x_{i(k-1)})$ (случайный вектор $\boldsymbol{\xi}$) и случайная независимая от $\mathbf{x}_i$ аддитивная ошибка измерения e_i (или неконтролируемый случайный фактор, с.в. ε, независимая от $\boldsymbol{\xi}$). При этом предполагается, что наблюдаемые величины связаны равенством

$$y_i(\mathbf{x}_i) = \beta_1 + \boldsymbol{\beta}^T\mathbf{x}_i + e_i,$$

а соответствующие левой и правой частям с.в. эквивалентны

$$\eta|\boldsymbol{\xi} \sim \beta_1 + \boldsymbol{\beta}^T\boldsymbol{\xi} + \varepsilon,$$

где с.в. $\varepsilon\sim\mathcal{N}_1(0,\sigma_\eta^2(1-\rho_{\eta\boldsymbol{\xi}}^2))$, с.в. $\boldsymbol{\xi}\sim\mathcal{N}_{k-1}(\boldsymbol{\mu}_{\boldsymbol{\xi}},D\boldsymbol{\xi})$. Нетрудно убедиться, что функция плотности вероятностей случайного вектора $(\boldsymbol{\xi},\eta)$ соответствует k-мерному нормальному распределению $\mathcal{N}_k(\boldsymbol{\mu},\mathbf{\Sigma})$

$$f_{\boldsymbol{\xi}\eta}(\mathbf{x},y) = f_{\eta|\boldsymbol{\xi}}(y|\mathbf{x})f_{\boldsymbol{\xi}}(\mathbf{x}) = (2\pi)^{-k/2}|\mathbf{\Sigma}|^{-1/2}exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}.$$

Описанный выше экспериментальный пример также может быть использован для моделирования k-мерного нормального распределения.

5.3. Пример 3

В экспериментальной практике статистическим аналогом теоретического совместного распределения случайного вектора $(\boldsymbol{\xi}, \eta)$ часто является распределение дискретного случайного вектора, принимающего n независимых равноточных значений $(x_{i1}, \ldots, x_{ik},\ y_i),\ i = 1, \ldots, n$, с вероятностями, равными $\dfrac{1}{n}$.

Примером такой ситуации может являться случайный эксперимент, в котором на исход $i-$ого опыта y_i, описываемого с.в. η_i, влияют неслучайные контролируемые факторы $x_{i1}, \ldots, x_{ik}$ и случайная аддитивная ошибка измерения e_i (или ряд неконтролируемых факторов), описываемых с.в. $\varepsilon_i,\ i = 1, \ldots, n$, некоррелированными, имеющими $E\varepsilon_i = 0$ и одинаковые дисперсии $D\varepsilon_i = \sigma^2 > 0$. В силу равноточности величины $e_1, \ldots, e_n$ можно интерпретировать как значения одной и той же с.в. ε, описываемой некоторым вероятностным распределением с функцией плотности $f_\varepsilon(z)$, и которой эквивалентны с.в. $\varepsilon_i,\ i = 1, \ldots, n$. Аналогично, величины $y_1, \ldots, y_n$ можно считать значениями одной с.в. η. В матричных обозначениях величину $\mathbf{y} = (y_1, \ldots, y_n)^T \in R^n$, описывающую результат эксперимента можно представить в виде

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}, \tag{6}$$

где $\mathbf{e} = (e_1, \ldots, e_n)$, матрица $\mathbf{X} = \|\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(k)}\| \in R^{n \times k}$, назывыаемая матрицей плана, составлена из векторов-столбцов $\mathbf{x}^{(j)} = (x_{1j}, \ldots, x_{nj})^T,\ j = 1, \ldots, k$, наблюдаемых значений с.в. $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_k)$. Тогда в силу равенства значений следует эквивалентность с.в. $\eta_i|\mathbf{x}_i \sim \boldsymbol{\beta}^T\mathbf{x}_i + \varepsilon_i$ или $\eta|\boldsymbol{\xi} \sim \boldsymbol{\beta}^T\boldsymbol{\xi} + \varepsilon$. Отсюда следует взаимосвязь плотностей $f_{\eta|\boldsymbol{\xi}}(z) = f_\varepsilon(z - \boldsymbol{\beta}^T\boldsymbol{\xi})$.

В этом случае, если $\varphi = E\varphi(\boldsymbol{\xi}; \eta)$ есть некоторая теоретическая характеристика наблюдаемых с.в., то ее статистический аналог вычисляется по формуле

$$\widehat{\varphi}_n = \frac{1}{n}\sum_{i=1}^{n} \varphi(\mathbf{x}_i; y_i),$$

где суммирование происходит по всем выборочным значениям. Тогда заменой теоретических характеристик на эмпирические, например, $E(\xi_i\xi_j) \sim \frac{1}{n}\sum_{l=1}^{n} x_{li}x_{lj}$, $E(\xi_i\eta) \sim \frac{1}{n}\sum_{l=1}^{n} x_{li}y_l$, легко могут быть получены расчетные формулы оценок параметров для любой линейной модели. Минимизируемый функционал (3) и формула (4) принимут вид [3, с.460]

$$
\begin{aligned}
&\widehat{\boldsymbol{\beta}}_n = \arg\min_{\boldsymbol{\beta}} \frac{1}{n}\sum_{i=1}^{n}(y_i - \boldsymbol{\beta}^T\mathbf{x}_i)^2 = \arg\min_{\boldsymbol{\beta}}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^T(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}),\\
&\widehat{\boldsymbol{\beta}}_n = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y},
\end{aligned}
\qquad (7)
$$

где $\mathbf{x}_i = (x_{i1}, \ldots, x_{ik})^T,\ i = 1, \ldots, n$. Для приведенных в качестве примеров моделей также могут быть получены широко известные формулы оценок параметров:
для модели 1

$$
\widehat{\beta}_n = \frac{\sum\limits_{i=1}^{n} x_i y_i}{\sum\limits_{i=1}^{n} x_i^2},
$$

или в виде (7) с матрицей плана $\mathbf{X} = \|\mathbf{x}^{(1)}\| \in R^{n\times 1}$, где $\mathbf{x}^{(\mathbf{1})} = (x_1, \ldots, x_n)^T$;
для модели 2

$$
\widehat{\beta}_n = \frac{1}{n}\sum_{i=1}^{n} y_i,
$$

или в виде (7) с матрицей плана $\mathbf{X} = \|\mathbf{1}\| \in R^{n\times 1}$, где $\mathbf{1} = (1, \ldots, 1)^T \in R^n$;
для модели 3 [4, с.383]

$$
\widehat{\beta}_{1n} = \overline{y} - \widehat{\beta}_{2n}\overline{x},\ \widehat{\beta}_{2n} = \frac{n\sum\limits_{i=1}^{n} x_i y_i - \sum\limits_{i=1}^{n} x_i \sum\limits_{i=1}^{n} y_i}{n\sum\limits_{i=1}^{n} x_i^2 - \left(\sum\limits_{i=1}^{n} x_i\right)^2},
$$

или в виде (7) с матрицей плана $\mathbf{X} = \|\mathbf{1}, \mathbf{x}^{(2)}\| \in R^{n\times 2}$, где $\overline{y} = \frac{1}{n}\sum\limits_{i=1}^{n} y_i$, $\overline{x} = \frac{1}{n}\sum\limits_{i=1}^{n} x_i$;
для модели 4 [3, с.520]

$$
\widehat{\beta}_{1n} = \frac{\sum\limits_{i=1}^{n} x_{i2}^2 \sum\limits_{i=1}^{n} x_{i1} y_i - \sum\limits_{i=1}^{n} x_{i1}x_{i2} \sum\limits_{i=1}^{n} x_{i2} y_i}{\sum\limits_{i=1}^{n} x_{i1}^2 \sum\limits_{i=1}^{n} x_{i2}^2 - \left(\sum\limits_{i=1}^{n} x_{i1}x_{i2}\right)^2},
$$

$$
\widehat{\beta}_{2n} = \frac{\sum\limits_{i=1}^{n} x_{i1}^2 \sum\limits_{i=1}^{n} x_{i2} y_i - \sum\limits_{i=1}^{n} x_{i1}x_{i2} \sum\limits_{i=1}^{n} x_{i1} y_i}{\sum\limits_{i=1}^{n} x_{i1}^2 \sum\limits_{i=1}^{n} x_{i2}^2 - \left(\sum\limits_{i=1}^{n} x_{i1}x_{i2}\right)^2},
$$

или в виде (7) с матрицей плана $\mathbf{X} = \|\mathbf{x}^{(1)}, \mathbf{x}^{(2)}\| \in R^{n\times 2}$;

для модели 5

$$\widehat{\beta}_{1n} = \overline{y} - \widehat{\boldsymbol{\beta}}_n^T \overline{\mathbf{x}},\ \widehat{\boldsymbol{\beta}}_n = \left(n\mathbf{X}^T\mathbf{X} - \overline{\mathbf{x}}\ \overline{\mathbf{x}}^T\right)^{-1} (n\mathbf{X}^T\mathbf{y} - \overline{\mathbf{x}}\ \overline{y}),$$

где матрица $\mathbf{X} = \|\mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(k)}\| \in R^{n\times(k-1)}$, $\overline{\mathbf{x}} = (\overline{x}_1, \ldots, \overline{x}_k)^T$ — вектор-столбец средних значений $\overline{x}_j = \dfrac{1}{n}\sum\limits_{i=1}^{n} x_{ij}$, $j = 2, \ldots, k$, или в виде (7) с матрицей плана [5, с.50] $\mathbf{X} = \|\mathbf{1}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(k)}\| \in R^{n\times k}$.

5.4. Пример 4

Рассмотрим теперь случай, аналогичный примеру 3, с тем отличием, что измерения проводяться с различной от опыта к опыту точностью. На практике это имеет место тогда, когда точность разных приборов для измерения величины отклика различная, при этом среднее значение отклика для заданного фактора не изменяется. Тогда в уравнении неравноточных наблюдений (6) величины $e_1, \ldots, e_n$ нельзя интерпретировать как значения одной и той же с.в. ε, а лишь как описываемых некоррелированными центрированными соответствующими с.в. $\varepsilon_1, \ldots, \varepsilon_n$ с $E\varepsilon_i = 0$ и разными дисперсиями $D\varepsilon_i = \sigma_i^2 > 0, i = 1, \ldots, n$. В связи с этим наблюдаемые значения $y_1, \ldots, y_n$ являются значениями соответствующих с.в. $\eta_1|\boldsymbol{\xi}, \ldots, \eta_n|\boldsymbol{\xi}$ с диагональной ковариационной матрицей $\boldsymbol{\Sigma} = diag(\sigma_1^2, \ldots, \sigma_n^2)$ и может быть представлена в виде $\boldsymbol{\Sigma} = \mathbf{L}^2$, где $\mathbf{L} = diag(\sigma_1, \ldots, \sigma_n)$. В этом случае статистическим аналогом теоретического совместного распределения случайного вектора $(\boldsymbol{\xi}, \eta)$ не может являться распределение значений $(x_{i1}, \ldots, x_{ik},\ y_i)$, $i = 1, \ldots, n$ с вероятностями $\dfrac{1}{n}$. Однако простыми преобразованиями пример 4 можно свести к примеру 3.

Действительно, из равенства (6) наблюдаемых значений следует эквивалентность с.в. $\eta_i|\boldsymbol{\xi} \sim \boldsymbol{\beta}^T\boldsymbol{\xi} + \varepsilon_i$. Преобразуем с.в. поделив их на σ_i. Тогда для вновь образованных с.в. $\acute{\boldsymbol{\eta}}|\acute{\boldsymbol{\xi}} = \mathbf{L}^{-1}\boldsymbol{\eta}|\boldsymbol{\xi}$, $\acute{\boldsymbol{\xi}} = \mathbf{L}^{-1}\boldsymbol{\xi}$, $\acute{\boldsymbol{\varepsilon}} = \mathbf{L}^{-1}\boldsymbol{\varepsilon}$ справедливо $\acute{\boldsymbol{\eta}}|\acute{\boldsymbol{\xi}} \sim \boldsymbol{\beta}^T\acute{\boldsymbol{\xi}} + \acute{\boldsymbol{\varepsilon}}$, где $E\acute{\boldsymbol{\varepsilon}} = \mathbf{0}$ и $D\acute{\boldsymbol{\varepsilon}} = \mathbf{I}_n$, $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n)^T$, $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^T$, $\mathbf{0}$ — нулевой вектор, а $\mathbf{I}_n$ — единичная матрица. Результаты обновленных равноточных наблюдений $(\acute{x}_{i1}, \ldots, \acute{x}_{ik},\ \acute{y}_i)$, $i = 1, \ldots, n$, можно представить как

$$\acute{\mathbf{y}} = \acute{\mathbf{X}}\boldsymbol{\beta} + \acute{\boldsymbol{e}},$$

где $\acute{\mathbf{y}} = \mathbf{L}^{-1}\mathbf{y}$, $\acute{\mathbf{X}} = \mathbf{L}^{-1}\mathbf{X}$, $\acute{\mathbf{e}} = \mathbf{L}^{-1}\mathbf{e}$, $\mathbf{y} = (y_1, \ldots, y_n)^T$, $\mathbf{e} = (e_1, \ldots, e_n)^T$.

Тогда минимизируемый функционал (3) и формула (4) примут вид [5, с.64]

$$\widehat{\boldsymbol{\beta}}_n = \arg\min_{\boldsymbol{\beta}} (\acute{\mathbf{y}} - \acute{\mathbf{X}}\boldsymbol{\beta})^T (\acute{\mathbf{y}} - \acute{\mathbf{X}}\boldsymbol{\beta}) = \arg\min_{\boldsymbol{\beta}} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}),$$
$$\widehat{\boldsymbol{\beta}}_n = (\acute{\mathbf{X}}^T \acute{\mathbf{X}})^{-1} \acute{\mathbf{X}}^T \acute{\mathbf{y}} = (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{y},$$

5.5. Пример 5

Аналогичный примера 4 результат получается в случае, когда ошибки коррелированы с невырожденной ковариационной матрицей ошибок $|\boldsymbol{\Sigma}| \neq 0$. В этом случае, поскольку матрица $\boldsymbol{\Sigma}$ симметричная и положительно определена, то существует невырожденная матрица $\mathbf{K} \in R^{(n \times n)}$ для которой $\boldsymbol{\Sigma} = \mathbf{K}\mathbf{K}^T$. Тогда образуя с.в. $\acute{\boldsymbol{\eta}} = \mathbf{K}^{-1}\boldsymbol{\eta}$, $\acute{\boldsymbol{\xi}} = \mathbf{K}^{-1}\boldsymbol{\xi}$ и $\acute{\boldsymbol{\varepsilon}} = \mathbf{K}^{-1}\boldsymbol{\varepsilon}$ приходим к модели с равноточными некоррелированными ошибками наблюдений в виде эквивалентных с.в. $\acute{\boldsymbol{\eta}}|\acute{\boldsymbol{\xi}} \sim \boldsymbol{\beta}^T\acute{\boldsymbol{\xi}} + \acute{\boldsymbol{\varepsilon}}$, где $E\acute{\boldsymbol{\varepsilon}} = \mathbf{0}$, $D\acute{\boldsymbol{\varepsilon}} = \mathbf{I}_n$.

В качестве матрицы $\mathbf{K}$ может быть взята матрица $\boldsymbol{\Sigma}^{-1/2}$. Также можно воспользоваться свойством любой симметрической положительно определенной матрицы $\boldsymbol{\Sigma}$ быть приведенной с помощью ортогонального преобразования, заданного ортогональной матрицей $\mathbf{H}$ (т.е. $\mathbf{H}^{-1} = \mathbf{H}^T$), к диагональному виду $\boldsymbol{\Sigma} = \mathbf{H}\boldsymbol{\Lambda}\mathbf{H}^T$, где $\boldsymbol{\Lambda} = diag(\lambda_1, \ldots, \lambda_n)$, $\lambda_i > 0$, $i = 1, \ldots, n$, — собственные (или характеристические) числа $\boldsymbol{\Sigma}$, а столбцы $\mathbf{H}$ — соответствующие числам собственные вектора, образующие ортонормированный базис. Тогда с помощью матрицы $\mathbf{K} = \mathbf{H}\boldsymbol{\Lambda}^{1/2}$ можно нормализовать любой случайный вектор $\boldsymbol{\varepsilon}$ в виде $\acute{\boldsymbol{\varepsilon}} = \mathbf{K}^{-1}(\boldsymbol{\varepsilon} - E(\boldsymbol{\varepsilon})) = \boldsymbol{\Lambda}^{-1/2}\mathbf{H}^T(\boldsymbol{\varepsilon} - E(\boldsymbol{\varepsilon}))$, где для вновь полученного вектора $\acute{\boldsymbol{\varepsilon}}$ имеем $E\acute{\boldsymbol{\varepsilon}} = \mathbf{0}$ и $D\acute{\boldsymbol{\varepsilon}} = \mathbf{I}_n$.

Отметим, что вид распределения с.в. $\acute{\boldsymbol{\varepsilon}}$, полученной из линейного преобразования с.в. $\boldsymbol{\varepsilon}$, может не совпадать с видом распределения последней. Это совпадение может иметь место среди распределений, обладающих свойством воспроизводимости по параметрам. Одним из таких уникальных распределений является нормальное. Например [3, с.51], всегда можно указать линейное преобразование, переводящее невырожденный нормально распределенный вектор в стандартный нормальный вектор. Действительно, пусть с.в. $\boldsymbol{\varepsilon} \sim \mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ и с.в. $\acute{\boldsymbol{\varepsilon}} = \mathbf{H}^T\boldsymbol{\varepsilon}$, тогда с.в. $\acute{\boldsymbol{\varepsilon}} \sim \mathcal{N}_n(\mathbf{H}^T\boldsymbol{\mu}, \boldsymbol{\Lambda})$ будет иметь некоррелированные (независимые) компоненты, а с.в.

$$\dot{\boldsymbol{\varepsilon}} = \boldsymbol{\Lambda}^{-1/2}(\acute{\boldsymbol{\varepsilon}} - \mathbf{H}^T\boldsymbol{\mu}) = \boldsymbol{\Lambda}^{-1/2}\mathbf{H}^T(\boldsymbol{\varepsilon} - \boldsymbol{\mu})$$

будет стандартным нормальным вектором $\dot{\boldsymbol{\varepsilon}} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I}_n)$. Благодаря этому качеству, в случае нормального распределения ошибок наблюдений, хорошо изучены свойства оценок параметров множественной линейной регрессии аналитический вид которых был представлен выше.

6. Заключение

Таким образом, вероятностно-статистический подход позволил исследовать основные положения возникновения и использования МНК, получить ряд общих вероятностных и статистических формул для оценок параметров линейной множественной модели вне зависимости от допустимого совместного распределения исследуемых величин при оценки параметров моделей с помощью данного метода. Хотя общий вид оценок параметров справедлив для многих распределений ошибок, однако свойства оценок уже определяются выбранной моделью и условиями эксперимента и в ряде случаев, например, нормальной множественной линейной регрессии, хорошо изучены.

Список литературы

1. *Крамер Г.* Математические методы статистики.— М.: Мир, 1976. 648 с.

2. *Гантмахер Ф. Р.* Теория матриц.— М.: Изд-во Наука. 1967. 575 с.

3. *Ивченко Г. И., Медведев Ю. И.* Введение в математическую статистику.— М.: Изд-во ЛКИ, 2010. 600 с.

4. *Кендалл М., Стьюарт А.* Статистические выводы и связи.— М.: Изд-во Наука. Главная редакция физико-математической литературы, 1973. 899 с.

5. *Себер Дж.* Линейный регрессионный анализ.— М.: Изд-во Мир, 1980. 456 с.

6. *Ермаков С. М., Михайлов Г. А.* Статистическое моделирование.— М.: Изд-во Наука. Главная редакция физико-математической литературы, 1982. 296 с.