Научная статья
УДК 519.862.6
EDN IQSEDQ
DOI 10.17150/2713-1734.2023.5(4).457-475

М.П. Базилевский
Иркутский государственный университет путей сообщения,
г. Иркутск, Российская Федерация

Математическое моделирование с помощью множественно-полносвязных линейных регрессий

Аннотация. Работа посвящена синтезу традиционных моделей множественной линейной регрессии с моделями полносвязной линейной регрессии. Замечено, что эти модели в определенном смысле дополняют друг друга — недостатки множественной регрессии компенсирует полносвязная, а недостатки полносвязной — множественная. Оценки множественной регрессии при частичной мультиколлинеарности факторов неустойчивы, а при полной — вовсе не существуют. Для применения полносвязной регрессии в таких условиях нет никаких преград, а ее оценки, наоборот, не существуют при полном отсутствии линейной зависимости между факторами. Оцененная полносвязная регрессия представляет собой уравнение прямой в пространстве, в отличие от множественной, представляющей собой гиперплоскость в пространстве, поэтому найти оценки полносвязной регрессии можно имея в распоряжении всего два наблюдения. Рассмотрен алгоритм оценивания полносвязных регрессий методом максимального правдоподобия. По выборке объема 21 построена множественно-полносвязная линейная регрессия пассажирских железнодорожных перевозок в Иркутской области, содержащая 23 входных переменных. В процессе построения удалось справиться и с мультиколлинеарностью, и обеспечить в модели протекцию всех входных переменных. Построенная множественно-полносвязная регрессия адекватна и полностью удовлетворяет содержательному смыслу решаемой задачи, поэтому дана ее интерпретация. По результатам проведенного исследования можно сделать вывод, что связка множественных и полносвязных регрессий может быть весьма эффективна при решении задач анализа данных.

Ключевые слова. Множественная регрессия, полносвязная регрессия, метод наименьших квадратов, метод максимального правдоподобия, мультиколлинеарность, корреляция, интерпретация, пассажирские перевозки.

Информация о статье. Дата поступления: 23 октября 2023 г.; дата принятия к публикации: 20 ноября 2023 г.; дата онлайн-размещения: 12 декабря 2023 г.

Original article

M.P. Bazilevskiy
Irkutsk State Transport University,
Irkutsk, Russian Federation

Mathematical Modeling Using Multiple and Fully-Connected Linear Regressions

Abstract. This article is devoted to the synthesis of traditional multiple linear regression models with fully connected linear regression models. It has been noted that these models, in a certain sense, complement each other — the disadvantages of multiple regression are compensated by the fully connected one, and the disadvantages of the fully connected one are compensated by the multiple one. Multiple regression estimates with partial multicollinearity of factors are unstable, and with complete multicollinearity they do not exist at all. There are no obstacles to the use of fully connected regression in such conditions, and its estimates, on the contrary, do not exist in the complete absence of linear dependence between factors. The estimated fully connected regression is an equation of a line in space, as opposed to a multiple regression, which is a hyperplane in space, so you can find estimates of a fully connected regression with only two observations at your disposal. An algorithm for estimating fully connected regressions using the maximum likelihood method is considered. Based on a sample size of 21, a multiple and fully connected linear regression of passenger rail traffic in the Irkutsk region was constructed, containing 23 input variables. During the construction process, it was possible to cope with multicollinearity and ensure protection of all input variables in the model. The constructed multiple and fully connected regression is adequate and fully satisfies the substantive meaning of the problem being solved, therefore its interpretation is given. Based on the results of the study, we can conclude that a combination of multiple and fully connected regressions can be very effective in solving data analysis problems.

Keywords. Multiple regression, fully connected regression, ordinary least squares method, maximum likelihood method, multicollinearity, correlation, interpretation, passenger transportation.

Article info. Received 23 October, 2023; Accepted 20 November, 2023; Available online 12 December, 2023.

Введение

В регрессионном анализе [1; 2] самой простой моделью принято считать модель множественной линейной регрессии:

$$y_i = \alpha_0 + \alpha_1 x_{i1} + \alpha_2 x_{i2} + \ldots + \alpha_m x_{im} + \varepsilon_i, \ i = \overline{1, n}, \qquad (1)$$

где m — число объясняющих переменных; n — объем выборки; y_i — i-е значение объясняемой переменной y; x_{ij} — i-е значение j-й объясняющей переменной; $\alpha_0, \alpha_1, \ldots, \alpha_m$ — неизвестные параметры; ε_i — i-я ошибка аппроксимации.

Самым эффективным методом оценивания неизвестных параметров модели (1) является метод наименьших квадратов (МНК), суть которого состоит в минимизации суммы квадратов ошибок $\sum_{i=1}^{n} \varepsilon_i^2$. Связка линейной регрессии (1) и МНК многие годы успешно применяется при решении различных прикладных задач анализа данных [3–5]. Тем не менее, в следующих двух ситуациях МНК-оценки модели (1) вовсе получить невозможно.

1. При совершенной мультиколлинеарности [6; 7], т.е. при наличии линейной функциональной зависимости между объясняющими переменными. На практике же чаще приходится сталкиваться с частичной мультиколлинеарностью, которая означает наличие сильной корреляции между регрессорами. В последнем случае оценки регрессии с помощью МНК вычислить все же мож-

но, но они получаются искаженными. Поэтому оцененную при наличии частичной мультиколлинеарности линейную регрессию (1) можно применять для прогнозирования, но никак не для интерпретации, поскольку такая модель представляет собой некую смесь более простых моделей. Возникает проблема, связанная с разделением такой смеси. Эта проблема сопряжена еще и с тем, что даже если, например, установлена сильная корреляция между двумя объясняющими переменными x_1 и x_2, то не до конца понятно, какую зависимость между ними нужно построить — x_1 от x_2 или x_2 от x_1, поскольку применение в обоих случаях МНК даст разные уравнения. Также не ясно, как связывать эту зависимость с переменной y. Быть может поэтому при обнаружении мультиколлинеарности рекомендуют попросту исключить из модели сильно коррелирующие переменные. По нашему мнению, такой подход не оправдан и не решает проблему, особенно в ситуации, когда все объясняющие переменные сильно коррелируют с y. Исключение приводит к потере информативности и возможности изучения совместного влияния всех переменных на выходной показатель. Существуют, конечно, и другие известные методы борьбы с мультиколлинеарностью, не требующие исключения переменных, например, метод главных компонент [7; 8] и гребневая (ридж) регрессия [7; 9]. Но в первом случае возникает проблема с интерпретацией главных компонент, а во втором случае — с выбором параметра регуляризации.

2. При условии, когда число неизвестных параметров $(m + 1)$ больше числа наблюдений n. Рекомендуется [10], чтобы при оценивании модели (1) число наблюдений было как минимум в 4 раза больше числа неизвестных параметров.

Противоположной по смыслу по отношению ко множественной регрессии (1) является предложенная автором модель полносвязной линейной регрессии [11–13]:

$$x_{ij} = x_{ij}^* + \varepsilon_i^{(x_j)}, \; i = \overline{1, n}, \, j = \overline{1, m}, \tag{2}$$

$$x_j^* = a_j + b_j x_m^*, \; j = \overline{1, m-1}, \tag{3}$$

где x^*_{ij} — неизвестное i-е истинное значение j-й взаимосвязанной переменной; a_j, b_j, $j = \overline{1, m-1}$ — неизвестные параметры; $\varepsilon_i^{(x_j)}$ — i-я ошибка j-й переменной. Термин «взаимосвязанные переменные» введен потому, что все они с одной стороны входные (объясняемые), а с другой — выходные (объясняющие). Уравнения (2) означают, что все переменные, в отличие от множественной регрессии (1), содержат некие ошибки, причиной возникновения которых могут быть неточности, допущенные при их измерении. Но существуют так называемые истинные значения переменных x^*_{ij}, которые неизвестны, но связаны между собой линейными функциональными зависимостями (3).

Полносвязная регрессия (2), (3) применяется при как можно более сильной корреляции взаимосвязанных переменных и может быть оценена при наличии всего хотя бы двух наблюдений. Даже если среди переменных $x_1, x_2, ..., x_m$ будут те, которые связаны между собой линейными функциональными зависимостями, то никаких преград для применения полносвязной регрессии нет.

Если ошибки $\varepsilon_i^{(x_j)}$, $i = \overline{1, n}$, $j = \overline{1, m}$, — случайные величины, распределенные по нормальному закону с нулевыми математическими ожиданиями и постоянными дисперсиями, т.е. $\varepsilon_i^{(x_j)} \sim N\left(0, \sigma^2_{\varepsilon^{(x_j)}}\right)$, $i = \overline{1, n}$, $j = \overline{1, m}$, то для оценки полносвязной регрессии методом максимального правдоподобия нужно решить следующую оптимизационную задачу [14]:

$$\lambda_1 \sum_{i=1}^{n} \left(\varepsilon_i^{x_1}\right)^2 + \lambda_2 \sum_{i=1}^{n} \left(\varepsilon_i^{x_2}\right)^2 + ... + \lambda_{m-1} \sum_{i=1}^{n} \left(\varepsilon_i^{x_{m-1}}\right)^2 +,$$

$$+ \sum_{i=1}^{n} \left(\varepsilon_i^{x_m}\right)^2 \to \min, \tag{4}$$

где коэффициенты $\lambda_j > 0$, $j = \overline{1, m-1}$, есть отношения дисперсий ошибок переменных $\lambda_j = \sigma^2_{\varepsilon^{(x_m)}} \big/ \sigma^2_{\varepsilon^{(x_j)}}$.

При наличии только двух переменных x_1 и x_2 полносвязная регрессия (2), (3) вырождается в регрессию Деминга [15; 16], которая находит широкое применение в клинической химии [17; 18]. В работе [19] регрессия Деминга применена в геохимии.

Таким образом, множественная (1) и полносвязная (2), (3) регрессии в определенном смысле дополняют друг друга — недостатки одной из них компенсируются достоинствами другой. Поэтому научный интерес вызывает исследование возможности комбинирования множественных и полносвязных регрессий при решении конкретных прикладных задач. Целью данной работы является построение на основе синтеза множественных и полносвязных регрессий модели пассажирских железнодорожных перевозок в Иркутской области.

1. Полносвязная линейная регрессия

Переменная x_m в правой части равенств (3) называется связующей. В [13] доказана теорема, согласно которой выбор связующей переменной в полносвязной регрессии (2), (3) не влияет на решение оптимизационной задачи (4).

Если в задаче (4) коэффициенты λ_j известны, то оценки полносвязной регрессии находятся по следующему алгоритму [11–13].

1. Численно решается нелинейная система

$$b_p\left(D_{x_m}+\sum_{j=1}^{m-1}\lambda_j^2 b_j^2 D_{x_j}+2\sum_{j_1=1}^{m-2}\sum_{j_2=j_1+1}^{m-1}\lambda_{j_1}\lambda_{j_2}b_{j_1}b_{j_2}K_{x_{j_1}x_{j_2}}+2\sum_{j=1}^{m-1}\lambda_j b_j K_{x_j x_m}\right)=$$

$$=\left(1+\sum_{j=1}^{m-1}\lambda_j b_j^2\right)\left(\sum_{j=1}^{m-1}\lambda_j b_j K_{x_j x_p}+K_{x_m x_p}\right),\ p=\overline{1,\ m-1}, \qquad (5)$$

где символом D обозначены дисперсии, а K — ковариации переменных. Численный метод решения системы (5) подробно описан в [12]. В результате решения будут найдены оценки $\tilde{b}_1, \tilde{b}_2, ..., \tilde{b}_{m-1}$ параметров $b_1, b_2, ..., b_{m-1}$.

2. По формулам $\tilde{a}_j=\overline{x_j}-\tilde{b}_j\cdot\overline{x_m},\ j=\overline{1,\ m-1}$, определяются оценки параметров $a_1, a_2, ..., a_{m-1}$.

3. Вычисляются оценки истинных значений переменной x_m по формулам

$$\tilde{x}_{im}^*=A_0+A_1x_{i1}+A_2x_{i2}+...+A_mx_{im},\ i=\overline{1,\ n}, \qquad (6)$$

где $A_0=-\left(1+\sum_{j=1}^{m-1}\lambda_j\tilde{b}_j^2\right)^{-1}\sum_{j=1}^{m-1}\lambda_j\tilde{a}_j\tilde{b}_j,\qquad A_m=\left(1+\sum_{j=1}^{m-1}\lambda_j\tilde{b}_j^2\right)^{-1},$

$A_j=\left(1+\sum_{j=1}^{m-1}\lambda_j\tilde{b}_j^2\right)^{-1}\lambda_j\tilde{b}_j,\ j=\overline{1,\ m-1}.$

Тогда оцененная полносвязная регрессия представляет собой множество взаимосвязей между всеми возможными парами переменных:

$$\tilde{x}_j^*=\tilde{a}_j+\tilde{b}_j\tilde{x}_m^*,\ j=\overline{1,\ m-1}, \qquad (7)$$

где значения переменной $\tilde{x}_m^*$ находятся по формуле (6).

Выражения (7) можно представить в виде:

$$\frac{\tilde{x}_1^*-\tilde{a}_1}{\tilde{b}_1}=\frac{\tilde{x}_2^*-\tilde{a}_2}{\tilde{b}_2}=...=\frac{\tilde{x}_{m-1}^*-\tilde{a}_{m-1}}{\tilde{b}_{m-1}}=\tilde{x}_m^*. \qquad (8)$$

По выражению (8) можно сделать вывод, что при оценивании полносвязной регрессии строится уравнение прямой в пространстве, в отличие от множественной регрессии, при оценивании которой строится гиперплоскость в пространстве. Этим же объясняется и то, что для построения полносвязной регрессии достаточно всего двух разных точек m-мерного пространства.

В [12] для измерения суммарного аппроксимационного качества полносвязной регрессии введен аддитивный коэффициент детерминации

$$R^2_{\text{add}} = \sum_{j=1}^{m} R^2_{x_j}, \tag{9}$$

где R^2_{xj} — коэффициент детерминации парной линейной регрессии $\tilde{x}_j^*$ от x_j.

Если в задаче (4) коэффициенты λ_j неизвестны, то их можно назначить как отношения дисперсий переменных:

$$\lambda_1 = D_{x_m} / D_{x_1},\ \lambda_2 = D_{x_m} / D_{x_2},\ \ldots,\ \lambda_{m-1} = D_{x_m} / D_{x_{m-1}}. \tag{10}$$

В той же работе установлено, что значение критерия (9) при коэффициентах λ_j, назначенных по формулам (10), будет максимальным.

Вопрос — велико ли смещение оценок полносвязной регрессии при назначении коэффициентов λ_j по формулам (10) — исследован в [20]. Пусть матрица дисперсий-ковариаций переменных имеет вид:

$$V = \begin{pmatrix} D_{x_1} & K_{x_1x_2} & \ldots & K_{x_1x_m} \\ K_{x_1x_2} & D_{x_2} & \ldots & K_{x_2x_m} \\ \ldots & \ldots & \ldots & \ldots \\ K_{x_1x_m} & K_{x_2x_m} & \ldots & D_{x_m} \end{pmatrix}.$$

Введем матрицу M, полученную из матрицы V поэлементным делением в ней первых $(m-1)$ строк на последнюю строку, а также m точек $P_1, P_2, \ldots, P_m$, координатами которых являются элементы столбцов матрицы M. В [20] доказана следующая теорема.

Теорема. Если в матрице M элементы каждой строки одного знака, то для любых $\lambda_j > 0$, $j = \overline{1, m-1}$, оценки параметров b_j, $j = \overline{1, m-1}$, полносвязной регрессии (2), (3) для метода (4) всегда лежат внутри открытого выпуклого m-угольника с вершинами в точках P_j, $j = \overline{1, m}$, расположенного в том ортанте $(m-1)$-мерного пространства, в котором $K_{x_jx_m} \cdot b_j > 0$, $j = \overline{1, m-1}$.

Из теоремы следует, во-первых, поскольку при сильной корреляции всех переменных площадь выпуклого m-угольника стремится к нулю, то в такой ситуации назначение коэффициентов λ_j по формулам (10) сводит к минимуму разницу между вычисленными и несмещенными оценками полносвязной регрессии. Во-вторых, если в матрице M элементы строки одного знака, то для любых $\lambda_j > 0$, $j = \overline{1, m-1}$, оценки параметров b_j, $j = \overline{1, m-1}$, полносвязной регрессии (2), (3) для метода (4) всегда согласуются со знаками соответствующих коэффициентов корреляции $r_{x_jx_m}$, $j = \overline{1, m-1}$. Таким образом, оценки полносвязной регрессии в любом случае корректно интерпретируются. Механизм такой интерпретации описан в [14].

2. Моделирование

Моделирование пассажирских железнодорожных перевозок является актуальной научной задачей, поэтому ее решению посвящено множество научных работ [21; 22]. В большинстве из них результатом моделирования является оцененная с помощью МНК модель множественной линейной регрессии (1). Иногда исследователи и вовсе ограничиваются только лишь проведением корреляционного анализа. При этом при оценивании множественной регрессии для устранения мультиколлинеарности часто применяется банальное исключение переменных, которое делает полученную модель весьма ограниченной. Покажем, как с помощью комбинирования множественных и полносвязных регрессий можно справляться с мультиколлинеарностью, обеспечивая при этом протекцию всех переменных.

Для построения модели в качестве выходной переменной y было выбрано отправление пассажиров железнодорожным транспортом общего пользования (тыс. человек) в Иркутской области. С использованием базы данных Федеральной службы государственной статистики (https://rosstat.gov.ru/) были собраны статистические данные за период с 2000 по 2020 гг. для переменной y и шестидесяти двух переменных x_1, x_2, ..., x_{62}, предположительно влияющих на нее. Затем из этого списка переменных были отобраны те, которые довольно тесно коррелируют с переменной y и у которых знак коэффициента корреляции не противоречит содержательному смыслу решаемой задачи. Таких переменных оказалось 23. Их описание и величины их коэффициентов корреляции с y представлены в табл. 1.

Таблица 1

Список переменных

Группа	Переменная	Описание переменной	Корреляция
Численность и состав населения	x_2	население в трудоспособном возрасте (в процентах от общей численности населения)	0.964
	x_3	численность рабочей силы (тыс. чел.)	0.819
	x_5	численность пенсионеров (тыс. чел.)	–0.887
Уровень жизни	x_6	среднемесячная номинальная начисленная заработная плата работников организаций (р.)	–0.888
	x_7	средний размер назначенных пенсий (р.)	–0.927
	x_8	число собственных легковых автомобилей на 1 000 чел. населения (ш.)	–0.872
Образование	x_{11}	численность студентов, обучающихся по программам подготовки квалифицированных рабочих, служащих (тыс. чел.)	0.879
	x_{13}	численность студентов, обучающихся по программам бакалавриата, специалитета, магистратуры (тыс. чел.)	0.893

Окончание табл. 1

Группа	Переменная	Описание переменной	Корреляция
Экономика и финансы	x_{14}	валовой региональный продукт (млн р.)	–0.890
	x_{15}	инвестиции в основной капитал (млн р.)	–0.852
	x_{16}	доходы консолидированного бюджета (млн р.)	–0.854
	x_{17}	расходы консолидированного бюджета (млн р.)	–0.853
Предприятия и организации	x_{19}	сальдированный финансовый результат (прибыль минус убыток) деятельности организаций (млн р.)	–0.887
Торговля	x_{31}	оборот розничной торговли (млн р.)	–0.857
	x_{32}	оборот оптовой торговли (млн р.)	–0.899
Транспорт	x_{37}	плотность автомобильных дорог общего пользования с твердым покрытием (км путей на 1 000 км^2 территории)	–0.931
	x_{38}	число автобусов общего пользования на 100 000 чел. населения	–0.790
	x_{39}	среднегодовая номинальная начисленная заработная плата работников транспорта (р.)	–0.891
Информационные и коммуникационные технологии	x_{41}	организации, имеющие Веб-сайт (в процентах от общего числа обследованных организаций)	–0.841
Наука и инновации	x_{43}	затраты на технологические инновации (млн р.)	–0.831
Цены и тарифы	x_{44}	стоимость фиксированного набора потребительских товаров и услуг (р.)	–0.890
	x_{53}	средние потребительские цены на проезд в купейном вагоне скорого нефирменного поезда дальнего следования (в расчете на 100 км пути) за декабрь (р.)	–0.800
	x_{54}	средние потребительские цены на проезд в купейном вагоне скорого фирменного поезда дальнего следования (в расчете на 100 км пути) за декабрь (р.)	–0.822

Как видно по табл. 1, только малая часть переменных оказывает прямое влияние на x_2 — это население в трудоспособном возрасте x_2 и численность рабочей силы x_3, которые формируют потребность в перемещении к месту работы и обратно [21], а также численности студентов x_{11} и x_{13}, которые формируют аналогичную потребность в перемещении к месту учебы и обратно. Остальные переменные оказывают на y обратное влияние. Сюда в первую очередь относятся группы — уровень жизни населения (переменные x_6, x_7, x_8), экономика и финансы (переменные x_{14}, x_{15}, x_{16}, x_{17}), торговля (переменные x_{31}, x_{32}), предприятия и организации (переменная x_{19}) и стоимость фиксированного набора потребительских товаров и услуг x_{44}. Точно такие же тенденции обнаружены в работе [22]. В той же работе отмечено, что снижение спроса на поездки железнодорожным транспортом при улучшении материального достатка у населения объясняется тем, что увеличивается конкурентное влияние в сегменте местного сообщения и развивается рынок личного автотранспорта. Увеличение численности пенсионеров x_5 снижает численность рабочей силы, что приводит к снижению спроса на железнодорожные перевозки. Развитие автотранспорта (переменные x_{37}, x_{38}, x_{39}) способствует увеличению автомобильных перевозок и снижению перевозок железнодорожным транспортом. Развитие информационных и коммуникационных технологий (переменная x_{41}) приводит к появлению удобных и доступных в режиме реального времени приложений, с помощью которых можно быстро организовать заказ автотранспорта практически в любой точке области, что снижает потребность в железнодорожном транспорте. Затраты на технологические инновации x_{43}, по-видимому, больше способствуют развитию автомобильного транспорта, нежели железнодорожного. Увеличение цен на проезд в вагонах x_{53} и x_{54} естественным образом снижает спрос на пассажирские железнодорожные перевозки.

Наличие такого большого числа факторов, негативно влияющих на пассажирские железнодорожные перевозки в Иркутской области, очень четко согласуется с тем фактом, что за период с 2009 г. по 2019 г. их объемы снизились примерно в 2 раза — с 22 427 тыс. чел. до 11 907 тыс. чел..

В результате была сформирована выборка данных объема $n = 21$, содержащая одну объясняемую переменную y и 23 объясняющих. В этом случае оценить с помощью МНК модель множественной линейной регрессии (1) невозможно, поскольку число наблюдений меньше числа неизвестных параметров. Поэтому было принято решение с помощью корреляционной матрицы сформировать кластер высоко коррелирующих переменных (КВКП), который был бы использован для построения полносвязной регрессии.

Для формирования КВКП решалась следующая задача булевого линейного программирования:

$$\sum_{j=1}^{23} \delta_j \to \max, \tag{11}$$

$$\left| r_{x_i x_j} \right| \geq r\left(\delta_i + \delta_j - 1\right), \ (i, j) \in \left\{ (s_1, s_2) \,|\, \left| r_{x_{s_1} x_{s_2}} \right| \leq r \right\}, \tag{12}$$

$$\delta_j \in \{0, 1\}, \ j = \overline{1, 23}, \ j = \overline{1,23}, \tag{13}$$

где δ_j — бинарная переменная, равная 1, если j-я переменная входит в КВКП, и 0, если не входит; $r_{x_i x_j}$ — значение коэффициента корреляции между переменными x_i и x_j; r — наименьшее значение коэффициента корреляции между переменными, входящими в КВКП.

ISSN 2713-1734

Число r было выбрано равным 0,75. В результате решения задачи (11)–(13) был сформирован КВКП, включающий 21 переменную: x_2, x_3, x_5, x_6, x_7, x_8, x_{11}, x_{14}, x_{15}, x_{16}, x_{17}, x_{19}, x_{31}, x_{32}, x_{37}, x_{39}, x_{41}, x_{43}, x_{44}, x_{53}, x_{54}.

После чего с использованием переменных, входящих в КВКП, оценивалась модель полносвязной линейной регрессии (2), (3) с отношениями дисперсий ошибок (10). В качестве связующей была выбрана переменная x_6. Для оценивания был использован специально разработанный на языке программирования hansl эконометрического пакета Gretl скрипт «full connect 2», реализующий численный метод, подробно описанный в [12]. Точность сходимости была выбрана равной 0.0001. В результате работы скрипта указанная точность была достигнута за 250 итераций. Оцененная модель полносвязной регрессии имеет следующий вид:

$$\frac{\tilde{x}_2^* - 64.22}{-0.000192} = \frac{\tilde{x}_3^* - 1330.519}{-0.00307} = \frac{\tilde{x}_5^* - 658.992}{0.00236} = \frac{\tilde{x}_7^* + 813.399}{0.345} =$$
$$= \frac{\tilde{x}_8^* - 134.755}{0.00318} = \frac{\tilde{x}_{11}^* - 35.194}{-0.00056} = \frac{\tilde{x}_{14}^* + 50921.5}{32.55} =$$
$$= \frac{\tilde{x}_{15}^* + 23099.88}{7.902} = \frac{\tilde{x}_{16}^* - 2695.628}{4.466} = \frac{\tilde{x}_{17}^* - 1096.412}{4.617} =$$
$$= \frac{\tilde{x}_{19}^* + 48321.48}{7.653} = \frac{\tilde{x}_{31}^* - 36117.56}{7.633} = \frac{\tilde{x}_{32}^* + 58371.42}{18.809} =$$
$$= \frac{\tilde{x}_{37}^* - 12.137}{0.000453} = \frac{\tilde{x}_{39}^* - 1566.423}{1.161} = \frac{\tilde{x}_{41}^* - 10.505}{0.000855} =$$
$$= \frac{\tilde{x}_{43}^* + 5840.38}{0.792} = \frac{\tilde{x}_{44}^* - 1854.28}{0.297} = \frac{\tilde{x}_{53}^* - 44.048}{0.00503} =$$
$$= \frac{\tilde{x}_{54}^* - 67.563}{0.00655} = \tilde{x}_6^*, \qquad (14)$$

где

$$\tilde{x}_6^* = 17215.34 - 222.594x_2 - 13.456x_3 + 21.104x_5 + 0.146x_7 +$$
$$+ 13.697x_8 - 85.643x_{11} + 0.00154x_{14} + 0.00624x_{15} + 0.011x_{16} +$$
$$+ 0.0106x_{17} + 0.00606x_{19} + 0.00657x_{31} + 0.00267x_{32} + 99.104x_{37} +$$
$$+ 0.0528x_{43} + 0.170x_{44} + 9.402x_{53} + 7.16x_{54} + 0.0507x_6. \qquad (15)$$

Поскольку в матрице M, составленной для выбранных переменных, элементы каждой строки оказались одного знака, то, как видно, оценки в знаменателях дробей равенств (14) и коэффициенты при переменных в уравнении (15), согласно приведенной выше теореме, согласуются со знаками соответствующих коэффициентов корреляции переменных с x_6.

Коэффициенты детерминации линейных регрессий полученных латентных переменных от наблюдаемых факторов составляют $R^2_{x_2} = 0.840$, ${}^2_3 = 0.811$, $R^2_{x_5} = 0.977$, $R^2_{x_7} = 0.991$, $R^2_{x_8} = 0.856$, $R^2_{x_{11}} = 0.942$, ${}^2_{14} = 0.981$, $R^2_{x_{15}} = 0.968$, $R^2_{x_{16}} = 0.964$, $R^2_{x_{17}} = 0.959$, $R^2_{x_{19}} = 0.911$, $R^2_{x_{31}} = 0.984$, $R^2_{x_{32}} = 0.985$, $R^2_{x_{37}} = 0.882$, $R^2_{x_{39}} = 0.993$, $R^2_{x_{41}} = 0.931$, $R^2_{x_{43}} = 0.821$, $R^2_{x_{44}} = 0.991$, $R^2_{x_{53}} = 0.928$, $R^2_{x_{54}} = 0.920$, $R^2_{x_6} = 0.995$. Как видно, все эти коэффициенты превышают значение 0,8, что говорит о высокой степени схожести полученных латентных переменных с наблюдаемыми.

В табл. 2 приведены точечные оценки параметров b_j, а также нижние и верхние границы их возможных значений для любых $\lambda_j > 0$. По табл. 2 можно сделать вывод, что точка, имеющая координаты найденных оценок параметров b_j полносвязной регрессии, расположена примерно в середине выпуклого многоугольника с вершинами P_j, т.е. нет существенной разницы между полученными и несмещенными оценками.

Таблица 2

Список переменных

Параметр	Точечная оценка	Нижняя граница	Верхняя граница
b_2	–0.000192	–0.00023	–0.00018
b_3	–0.00307	–0.00376	–0.00280
b_5	0.00236	0.00220	0.00246
b_7	0.345	0.33540	0.35671
b_8	0.00318	0.00283	0.00379
b_{11}	–0.00056	–0.00060	–0.00050
b_{14}	32.55	31.43520	33.63993
b_{15}	7.902	7.66866	8.12194
b_{16}	4.466	4.31676	4.61273
b_{17}	4.617	4.50237	4.85398
b_{19}	7.653	7.32231	8.41651
b_{31}	7.633	7.37538	7.87127
b_{32}	18.809	18.48376	19.15541
b_{37}	0.000453	0.00043	0.00052
b_{39}	1.161	1.13717	1.18375
b_{41}	0.000855	0.00078	0.00093
b_{43}	0.792	0.72111	0.95087

Окончание табл. 2

Параметр	Точечная оценка	Нижняя граница	Верхняя граница
b_{44}	0.297	0.28756	0.30204
b_{53}	0.00503	0.00455	0.00544
b_{54}	0.00655	0.00589	0.00718

Затем строилась модель множественной линейной регрессии зависимости y от $\tilde{x}_6^*$, x_{13} и x_{38}. Отметим, что для этого выполнены все требуемые условия — число наблюдений $n = 21$ превышает число неизвестных параметров $(m + 1) = 4$ более чем в 5 раз, а наиболее тесная корреляция (с коэффициентом –0.76) наблюдается между переменными $\tilde{x}_6^*$ и x_{13}, которая позволяет судить о незначительном эффекте мультиколлинеарности. Результаты оценивания в пакете Gretl с помощью МНК множественной регрессии представлены на рис. 1.

```
Модель 1: МНК, использованы наблюдения 2000-2020 (T = 21)
Зависимая переменная: y2

             Коэффициент    Ст. ошибка   t-статистика  P-значение
  --------------------------------------------------------------
  const      15199,6        3148,70          4,827      0,0002    ***
  x6            -0,164910      0,0413693    -3,986      0,0010    ***
  x13          118,276        22,4703        5,264      6,34e-05  ***
  x38          -45,0281       13,0074       -3,462      0,0030    ***

Среднее зав. перемен    19484,43   Ст. откл. зав. перемен   6419,474
Сумма кв. остатков      41478094   Ст. ошибка модели        1562,014
R-квадрат               0,949674   Испр. R-квадрат          0,940793
F(3, 17)                106,9332   Р-значение (F)           3,10e-11
Лог. правдоподобие     -182,0073   Крит. Акаике             372,0146
Крит. Шварца            376,1927   Крит. Хеннана-Куинна     372,9214
Параметр rho            0,219724   Стат. Дарбина-Вотсона    1,555638
```

Рис. 1. Результаты оценивания множественной регрессии

Как видно по рис. 1, для уровня значимости $\alpha = 0.01$ все коэффициенты значимы. Коэффициент детерминации регрессии $R = 0.9497$, что подтверждает ее высокую адекватность. Дополнительно были найдены коэффициенты вздутия дисперсии $VIF_{\tilde{x}_6^*} = 3.045$, $VIF_{x_{13}} = 2.448$, $VIF_{x_{38}} = 1.952$, как оказалось, не превышающие значение 10, что говорит об отсутствии в модели мультиколлинеарности. Уравнение полученной множественной регрессии имеет вид:

$$\tilde{y} = 15199.6 - 0.165\tilde{x}_6^* + 118.276x_{13} - 45.028x_{38}. \qquad (16)$$

Структурная спецификация полученной модели (14)–(16) изображена на рис. 2. Как видно, в результате ее оценивания сформировалось множество из двадцати одной латентной переменной, все пары которых связаны между собой линейными функциональными зависимостями. Все эти 210 взаимозависимостей можно получить с помощью равенств (14). Каждая латентная переменная представляет собой линейную комбинацию из двадцати одной исходной переменной. Все такие комбинации можно получить обычной подстановкой уравнения (15) в зависимости (14). При этом знаки коэффициентов в линейных комбинациях будут совпадать со знаками соответствующих коэффициентов корреляции. На зависимую переменную влияет переменная x_{13}, x_{38} и любая из латентных переменных. В нашем случае и в модели, и на рис. 2 связующей выбрана $\tilde{x}_6^*$.

Таким образом, построенная модель (14) – (16) включает в себя как полносвязную, так и множественную регрессию. Поэтому назовем ее моделью множественно-полносвязной линейной регрессии.

Если подставить (15) в (16), то получим уравнение, содержащее все 23 исходных объясняющих переменных, причем, в этом уравнении знаки всех коэффициентов при факторах будут согласованы со знаками соответствующих коэффициентов корреляции из табл. 1.

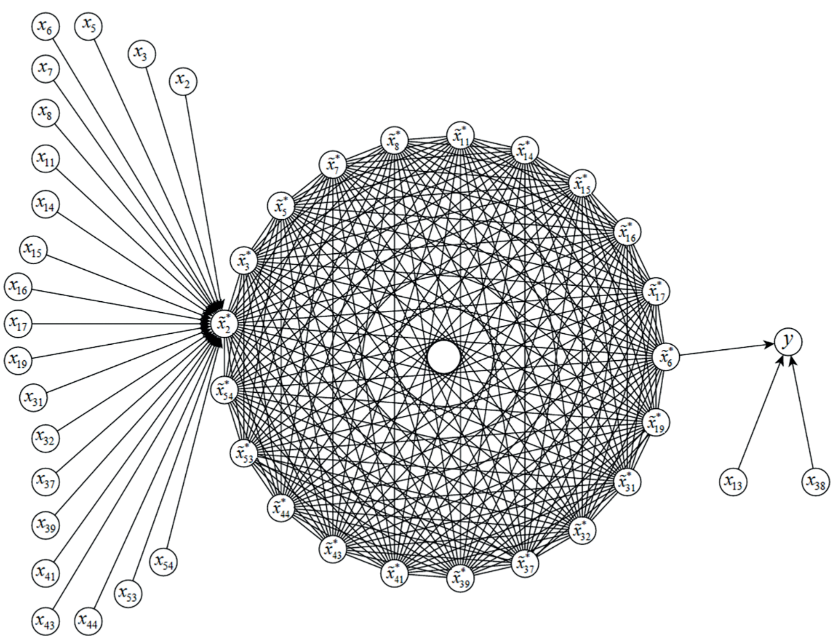

Рис. 2. Структурная спецификация модели

С учетом теоремы, доказанной в [14], дадим интерпретацию построенной множественно-полносвязной модели (14)–(16).

1. При увеличении численности студентов x_{13} на 10 тыс. чел. (при неизменных значениях остальных переменных) отправление

пассажиров железнодорожным транспортом общего пользования y увеличивается в среднем на 1 183 тыс. чел.

2. При увеличении автобусов общего пользования на 100 000 чел. населения x_{38} на 10 ед. (при неизменных значениях остальных переменных) отправление пассажиров железнодорожным транспортом общего пользования y уменьшается в среднем на 450 тыс. чел.

3. Увеличение среднемесячной номинальной начисленной зарплаты работников организаций x_6 на 1 000 р. означает, что параллельно в регионе произошли следующие изменения: население в трудоспособном возрасте x_2 уменьшилось в среднем на 0.192 %; численность рабочей силы x_3 уменьшилась в среднем на 3.07 тыс. чел.; численность пенсионеров x_5 увеличилась в среднем на 2.36 тыс. чел.; средний размер назначенных пенсий x_7 увеличился в среднем на 345 р.; число собственных легковых автомобилей на 1 000 чел. населения x_8 увеличилось в среднем на 3.18 ед.; численность студентов, обучающихся по программам подготовки квалифицированных рабочих, x_{11} уменьшилась в среднем на 0.56 тыс. чел.; валовой региональный продукт x_{14} увеличился в среднем на 32 550 млн р.; инвестиции в основной капитал x_{15} увеличились в среднем на 7 902 млн р.; доходы консолидированного бюджета x_{16} увеличились в среднем на 4 466 млн р.; расходы консолидированного бюджета x_{17} увеличились в среднем на 4 617 млн р.; сальдированный финансовый результат деятельности организаций x_{19} увеличился в среднем на 7 653 млн р.; оборот розничной торговли x_{31} увеличился в среднем на 7 633 млн р.; оборот оптовой торговли x_{32} увеличился в среднем на 18 809 млн р.; плотность автомобильных дорог общего пользования с твердым покрытием x_{37} увеличилась в среднем на 0.453 км; среднегодовая номинальная начисленная заработная плата работников транспорта x_{39} увеличилась в среднем на 1 161 р.; число организаций, имеющих Веб-сайт, x_{41} увеличилось в среднем на 0.855 %; затраты на технологические инновации x_{43} увеличились в среднем на 792 млн р.; стоимость фиксированного набора потребительских товаров и услуг x_{44} увеличилась в среднем на 297 р.; средние потребительские цены на проезд в купейном вагоне скорого нефирменного поезда дальнего следования x_{53} увеличились в среднем на 5.03 р.; средние потребительские цены на проезд в купейном вагоне скорого фирменного поезда дальнего следования x_{54} увеличились в среднем на 6.55 р. Указанные изменения переменных (при неизменных значениях переменных x_{13} и x_{38}) приведут к уменьшению отправления пассажиров железнодорожным транспортом общего пользования y в среднем на 165 тыс. чел.

Построенная множественно-полносвязная модель (14)–(16) может быть использована и для прогнозирования. При этом значения переменных x_{13} и x_{38} можно задавать любыми, поскольку считается, что они не связаны между собой. А значения оставшихся переменных, входящих в полносвязную часть, из-за того, что все они взаимосвязаны между собой, нужно задавать следующим образом.

1. Задать прогнозное значение любой переменной, входящей в полносвязную часть.

2. Для заданного значения с помощью формул (14) найти прогнозные значения остальных переменных из полносвязной части.

В противном случае, если задать значения входящих в полносвязную часть переменных произвольно, нарушив их пропорциональность, полученные прогнозы могут оказаться неточными.

Заключение

В настоящем исследовании рассмотрены вопросы синтеза традиционных моделей множественной линейной регрессии с моделями полносвязной линейной регрессии. Построена множественно-полносвязная линейная регрессионная модель пассажирских железнодорожных перевозок в Иркутской области.

Резюмируя, хотелось бы обратить внимание на следующие моменты.

1. Моделирование проводилось по выборке, содержащей 23 объясняющих переменных и 21 наблюдение. Оценить по этой выборке традиционную модель множественной линейной регрессии невозможно. И даже если бы наблюдений было в 2 раза больше, то из-за эффекта мультиколлинеарности, скорее всего, были бы искажены знаки оценок регрессии. В любом случае для построения модели пришлось бы прибегать к исключению переменных, т.е. терять возможность изучать их совместное влияние на выходной показатель. При оценивании же множественно-полносвязной регрессии пассажирских перевозок (14)–(16) удалось обеспечить в ней протекцию абсолютно всех переменных. При этом практически полностью удалось справиться с негативным эффектом мультиколлинеарности, о чем свидетельствует тот факт, что абсолютно в любом выражении, полученном с помощью уравнений (14)–(16), знаки коэффициентов при объясняющих переменных согласуются со знаками соответствующих коэффициентов корреляции. Интерпретация построенной модели гораздо богаче, чем была бы при использовании множественной регрессии.

2. По результатам проведенного исследования можно сделать вывод, что связка множественных и полносвязных регрессий может быть весьма эффективна при решении любых задач анализа данных. Недостатки множественной регрессии компенсирует полносвязная, а недостатки полносвязной — множественная. Если все объясняющие переменные коррелируют слабо, то необходимо оценивать множественную регрессию, если сильно — полносвязную, а если смешанным образом, то необходимо комбинировать

множественную регрессию с полносвязными. В любом случае при наличии достаточного числа наблюдений всегда можно значительно ослабить мультиколлинеарность, не жертвуя объясняющими переменными.

Список использованной литературы

1. Montgomery D.C. Introduction to linear Regression Analysis / D.C. Montgomery, E.A. Peck, G.G. Vining. — New York : John Wiley, 2001. — 680 p.

2. Gunst R.F. Regression Analysis and its Application: a Data-Oriented Approach / R.F. Gunst, R.L. Mason. — New York : M. Dekker, 1980. — 402 p.

3. The influence of the COVID-19 Pandemic on Stock Market Returns in Indonesia Stock Exchange / A. Herwany, E. Febrian, M. Anwar, A. Gunardi // The Journal of Asian Finance, Economics and Business. — 2021. — Vol. 8, No. 3. — P. 39–47.

4. Niftiyev I. Dutch Disease Effects in the Azerbaijan Economy: Results of Multivariate Linear Ordinary Least Squares (OLS) Estimations / I. Niftiyev // HSE Economic Journal. — 2021. — Vol. 25, No. 2. — P. 309–346.

5. Aloisio A. Dynamic Identification of a Masonry Façade From Seismic Response Data Based on an Elementary Ordinary Least Squares Approach / A. Aloisio, R. Alaggio, M. Fragiacomo // Engineering Structures. — 2019. — Vol. 197. — P. 109415.

6. Kim J.H. Multicollinearity and Misleading Statistical Results / J.H. Kim // Korean journal of anesthesiology. — 2019. — Vol. 72, No. 6. — P. 558–569.

7. Shrestha N. Detecting multicollinearity in regression analysis / N. Shrestha // American Journal of Applied Mathematics and Statistics. — 2020. — Vol. 8, No. 2. — P. 39–42.

8. Gwelo A.S. Principal Components to Overcome Multicollinearity Problem / A.S. Gwelo // Oradea Journal of Business and Economics. — 2019. — Vol. 4, No. 1. — P. 79–91.

9. Schreiber-Gregory D.N. Ridge Regression and Multicollinearity: An in-Depth Review / D.N. Schreiber-Gregory // Model Assisted Statistics and Applications. — 2018. — Vol. 13, No. 4. — P. 359–365.

10. Носков С.И. Технология моделирования объектов с нестабильным функционированием и неопределенностью в данных / С.И. Носков. — Иркутск : РИЦ ГП «Облинформпечать», 1996. — 320 с.

11. Базилевский М.П. Многофакторные модели полносвязной линейной регрессии без ограничений на соотношения дисперсий ошибок переменных / М.П. Базилевский // Информатика и её применения. — 2020. — Т. 14. — № 2. — С. 92–97.

12. Базилевский М.П. Метод выпрямления искаженных из-за мультиколлинеарности коэффициентов в регрессионных моделях / М.П. Базилевский // Информатика и её применения. — 2021. — Т. 15, № 2. — С. 60–65.

13. Базилевский М.П. Исследование поведения относительных вкладов переменных в общую детерминацию в оцененном на основе метода выпрямления искаженных коэффициентов регрессионном уравнении / М.П. Базилевский. — DOI 10.55648/1998-6920-2022-16-1-89-96. — EDN CDWXVL // Вестник СибГУТИ. — 2022. — № 1 (57). — С. 89–96.

14. Bazilevskiy M.P. Interpretation of Parameter Estimates for Fully connected Linear Regression Models / M.P. Bazilevskiy // International Journal of Open Information Technologies. — 2023. — Vol. 11, No. 10. — P. 21–25.

15. Deming W.E. Statistical Adjustment of Data / W.E. Deming. — New York : Wiley, 1948. — 269 p.

16. Тимофеев В.С. Идентификация зависимостей признаков стохастической природы на основе регрессии Деминга / В.С. Тимофеев, В.Ю. Щеколдин, А.Ю. Тимофеева // Информатика и её применения. — 2013. — Т. 7, № 2. — С. 60–68.

17. Sample Multiplexing: Increased Throughput for Quantification of Total Testosterone in Serum by Liquid Chromatography-Tandem Mass Spectrometry / J.D. Colletti, M.M. Redor-Goldman, A.E. Pomperada [et al.] // Clinical Chemistry. — 2020. — Vol. 66, No. 9. — P. 1181–1189.

18. Commutability of External Quality Assessment Materials for point-of-care Glucose Testing Using the Clinical and Laboratory Standards Institute and International Federation of Clinical Chemistry approaches / Y. Wang, M. Plebani, L. Sciacovelli [et al.] // Journal of clinical laboratory analysis. — 2020. — Vol. 34, No. 8. — P. e23327.

19. Смирнов М.Б. Зависимости между основными структурно-групповыми параметрами состава нефтей Западной Сибири по данным ЯМР / М.Б. Смирнов, Н.А. Ванюкова. — DOI 10.7868/S002824211405009. — EDN SMMNZJ // Нефтехимия. — 2014. — Т. 54. — № 5. — С. 355–365.

20. Базилевский М.П. Идентификация областей возможных оценок параметров моделей полносвязной линейной регрессии / М.П. Базилевский. — DOI 10.17759/mda.2023130304. — EDN JZJRBI // Моделирование и анализ данных. — 2023. — Т. 13, № 3. — С. 52–65.

21. Криклевская Л.Ю. Моделирование взаимовлияния развития транспорта и социально-экономического развития Забайкальского края / Л.Ю. Криклевская. — EDN YAUMVN // Проблемы социально-экономического развития Сибири. — 2018. — № 2. — С. 43–55.

22. Муктепавел С.В. Анализ макроэкономических факторов, влияющих на объем пассажирских перевозок в местном сообщении / С.В. Муктепавел. — EDN VMFSZX // Вестник Научно-исследовательского института железнодорожного транспорта. — 2016. — Т. 75, № 1. — С. 53–59.

References

1. Montgomery D.C., Peck E.A., Vining G.G. *Introduction to linear Regression Analysis*. New York, John Wiley, 2001. 680 p.

2. Gunst R.F., Mason R.L. *Regression Analysis and its Application: a Data-Oriented Approach*. New York, M. Dekker, 1980. 402 p.

3. Herwany A., Febrian E., Anwar M., Gunardi A. The Influence of the COVID-19 Pandemic on Stock Market Returns in Indonesia Stock Exchange. *The Journal of Asian Finance, Economics and Business*, 2021, vol. 8, no. 3, pp. 39–47.

4. Niftiyev I. Dutch Disease Effects in the Azerbaijan Economy: Results of Multivariate Linear Ordinary Least Squares (OLS) Estimations. *HSE Economic Journal*, 2021, vol. 25, no. 2, pp. 309–346.

5. Aloisio A., Alaggio R., Fragiacomo M. Dynamic Identification of a Masonry Façade From Seismic Response Data Based on an Elementary Ordinary Least Squares Approach. *Engineering Structures,* 2019, vol. 197, pp. 109415.

6. Kim J.H. Multicollinearity and Misleading Statistical Results. *Korean journal of anesthesiology*, 2019, vol. 72, no. 6, pp. 558–569.

7. Shrestha N. Detecting multicollinearity in regression analysis. *American Journal of Applied Mathematics and Statistics*, 2020, vol. 8, no. 2, pp. 39–42.

8. Gwelo A.S. Principal Components to Overcome Multicollinearity Problem. *Oradea Journal of Business and Economics*, 2019, vol. 4, no. 1, pp. 79–91.

9. Schreiber-Gregory D.N. Ridge Regression and Multicollinearity: An in-Depth Review. *Model Assisted Statistics and Applications*, 2018, vol. 13, no. 4, pp. 359–365.

10. Noskov S.I. *Technology for modeling objects with unstable functioning and data uncertainty*. Irkutsk, RITs GP «Oblinformpechat'» Publ., 1996. 320 p.

11. Bazilevskii M.P. Multivariate fully connected linear regression models without restrictions on the ratios of error variances of variables. *Informatika i ee primeneniya = Computer science and its applications*, 2020, vol. 14, no. 2, pp. 92–97. (In Russian).

12. Bazilevskii M.P. A method for straightening coefficients distorted due to multicollinearity in regression models. *Informatika i ee primeneniya = Computer science and its applications*, 2021, vol. 15, no. 2, pp. 60–65. (In Russian).

13. Bazilevskii M.P. Researching the Behavior of Variables Relative Contributions to the Total Determination in Regression Equation Estimated Using the Method of Distorted Coefficients Straightening. *Vestnik SIBGUTI = Bulletin of SIBGUTI*, 2022, no. 1, pp. 89–96. (In Russian). EDN: CDWXVL. DOI: 10.55648/1998-6920-2022-16-1-89-96.

14. Bazilevskiy M.P. Interpretation of Parameter Estimates for Fully connected Linear Regression Models. *International Journal of Open Information Technologies*, 2023, vol. 11, no. 10, pp. 21–25.

15. Deming W.E. *Statistical Adjustment of Data*. New York, Wiley, 1948. 269 p.

16. Timofeev V.S, Shchekoldin V.Yu., Timofeeva A.Yu. Identification of dependencies of features of a stochastic nature based on Deming regression. *Informatika i ee primeneniya = Computer science and its applications*, 2013, vol. 7, no. 2, pp. 60–68. (In Russian).

17. Colletti J.D., Redor-Goldman M.M., Pomperada A.E. [et al.]. Sample Multiplexing: Increased Throughput for Quantification of Total Testosterone in Serum by Liquid Chromatography-Tandem Mass Spectrometry. *Clinical Chemistry*, 2020, vol. 66, no. 9. Pp. 1181–1189.

18. Wang Y., Plebani M., Sciacovelli L. [et al.]. Commutability of External Quality Assessment Materials for point-of-care Glucose Testing Using the Clinical and Laboratory Standards Institute and International Federation of Clinical Chemistry approaches. *Journal of clinical laboratory analysis*, 2020, vol. 34, no. 8, pp. e23327.

19. Smirnov M.B., Vanyukova N.A. Elations Between the Main Structural-Group Composition Parameters of Western Siberia Crude Oils According To Nmr Data. *Neftekhimiya = Petroleum Chemistry*, 2014, vol. 54, no. 5, pp. 355–365. (In Russian). EDN: SMMNZJ. DOI: 10.7868/S0028242114050098.

20. Bazilevskii M.P. Identification of Possible Estimates Areas for Parameters of Fully Connected Linear Regression Models. *Modelirovanie i analiz dannykh = Data modeling and analysis*, 2023, vol. 13, no.3, pp. 52–65. (In Russian). EDN: JZJRBI. DOI: 10.17759/mda.2023130304.

21. Kriklevskaya L.Yu. The Factor Model of Mutual Influence of Transportand Social and Economic Development of the Transbaikal Region. *Problemy sotsial'no-ekonomicheskogo razvitiya Sibiri = Problems of socio-economic development of Siberia*, 2018, no. 2, pp. 43–55. (In Russian). EDN: YAUMVN.

22. Muktepavel S.V. Analysis of Macroeconomic Factors Influencing on Local Passenger Traffic Volume. *Vestnik Nauchno-issledovatel'skogo instituta zheleznodorozhnogo transporta = Bulletin of the Research Institute of Railway Transport*, 2016, vol. 75, no. 1, pp. 53–59. (In Russian). EDN: VMFSZX.

Информация об авторе

Базилевский Михаил Павлович — кандидат технических наук, доцент, кафедра математики, Иркутский государственный университет путей сообщения, г. Иркутск, Российская Федерация, e-mail: mik2178@yandex.ru, https://orcid.org/0000-0002-3253-5697, SPIN-код: 4347-5028, AuthorID РИНЦ: 679277.

Information about the Author

Mikhail P. Bazilevskiy — PhD in Technical Sciences, Associate Professor, Department of Mathematics, Irkutsk State Transport University, Irkutsk, Russian Federation, e-mail: mik2178@yandex.ru, https://orcid.org/0000-0002-3253-5697, SPIN-Code: 4347-5028, AuthorID RSCI: 679277.

Для цитирования

Базилевский М.П. Математическое моделирование с помощью множественно-полносвязных линейных регрессий / М.П. Базилевский. — DOI 10.17150/2713-1734.2023.5(4).457-475 — EDN IQSEDQ // System Analysis & Mathematical Modeling. — 2023. — Т. 5, № 4. — С. 457–475.

For Citation

Bazilevskiy M.P. Mathematical Modeling Using Multiple and Fully-Connected Linear Regressions. *System Analysis & Mathematical Modeling*, 2023, vol. 5, no. 4, pp. 457–475. (In Russian). EDN: IQSEDQ. DOI: 10.17150/2713-1734.2023.5(4).457-475.