ПРИМЕНЕНИЕ РОБАСТНЫХ МЕТОДОВ ФАКТОРНОГО АНАЛИЗА ДЛЯ ВЫЯВЛЕНИЯ ИННОВАЦИОННОГО ПОТЕНЦИАЛА СУБЪЕКТОВ РФ

Горяинова Е.Р., Мамонтова Д.А.
Национальный исследовательский университет «Высшая школа экономики», Москва, Россия
el-goryainova@mail.ru, damamontova@edu.hse.ru

Аннотация. Рассмотрена задача построения рейтинга инновационного развития субъектов РФ. Построение проведено с помощью методов факторного анализа, основанного на робастных оценках корреляционных матриц наблюдаемых социально-экономических показателей, распределение которых не является гауссовским. Выделены три общих фактора, по оцененным факторным баллам определены лидеры.

Ключевые слова: факторный анализ, метод минимальных остатков, робастные оценки корреляционной матрицы, MCD-оценка, оценка Олива-Хокинса, факторные баллы.

Введение

В современных экономических и финансовых исследованиях используются большие объемы данных, где каждый объект описывается множеством тесно взаимосвязанных показателей. Это создает необходимость описания объектов меньшим числом характеристик при максимальном сохранении важной информации. Такие новые латентные (не наблюдаемые напрямую) характеристики принято называть общими факторами.

Введение общих факторов имеет, в основном, следующие цели: построение наиболее информативных интегративных показателей, изучение структуры взаимосвязей исходных переменных, выявление объектов с наиболее выраженными общими факторами и сокращение объема данных для хранения.

Модели факторного анализа представлены в работах ([1–5]). Отправной точкой факторного анализа, позволяющего строить обобщенные показатели, является оценка корреляционной матрицы исходных данных. Однако, как показало компьютерное моделирование (например, в [6, 7]), использование классической оценки корреляции Пирсона при сжатии данных в условиях наличия аномальных наблюдений или при распределениях с «тяжёлыми хвостами» приводит к искажению реальной картины. При этом было продемонстрировано, что применение робастных оценок корреляционных матриц для зашумленных данныхповышает устойчивость и точность метода главных компонент.

В данной работе анализируются 12 экономических показателей 85 субъектов РФ за 2023 год, распределение которых оказалось отличным от гауссовского. Применяя методы факторного анализа, основанного на робастных оценках корреляционных матриц, выявлены три общих фактора, интерпретированных как инновационный потенциал регионов РФ. Этим факторам даны содержательные названия «Научное развитие», «Инновационное развитие бизнеса» и «Уровень цифровизации». В соответствии с методом Бартлетта для каждого из 85 субъектов были вычислены факторные баллы по всем выявленным общим факторам. На основе полученных факторных значений определены регионы-лидеры инновационного развития по каждому из трёх указанных факторов.

1. Задача факторного анализа

Факторный анализ, позволяющий строить некоррелированные латентные факторы, которые без существенной потери информации могут описывать наблюдаемые коррелированные многомерные показатели, включает несколько этапов. К ним относятся: определение числа общих факторов; оценивание корреляционной матрицы исходных показателей; оценивание матрицы нагрузок исходных показателей на общие факторы; процедуру вращения; интерпретацию построенных общих факторов и оценивание факторных баллов для каждого исследуемого объекта.

Рассмотрим основную модель факторного анализа. Пусть $\mathbf{X} = (X_1, \dots, X_r)$ – r-мерный вектор наблюдаемых показателей у каждого из n наблюдаемых объектов. Требуется найти такие некоррелированные показатели $f_1, \dots, f_k$, $k < r$, которые объясняют максимально возможную долю изменчивости наблюдаемых показателей $X_1, \dots, X_r$.

Проведём нормализацию координат вектора $\mathbf{X}$. По имеющейся n-мерной выборке построим выборочное среднее $\bar{X}_{i\cdot} = \frac{1}{n}\sum_{j=1}^{n} X_{ij}$ и несмещённую выборочную дисперсию $s_i^2 = \frac{1}{n-1}\sum_{j=1}^{n} (X_{ij} -$

$\bar{X}_{i\cdot})^2$ компоненты $i = 1, \dots, r$ вектора **X**. Центрированно-нормированнуюкомпоненту i-ю компоненту вектора **X**обозначим как $x_i = \frac{X_i - \bar{X}_{i\cdot}}{s_i}$, и определим вектор $\vec{x} = (x_1, \dots, x_r)$.

Представим теперь вектор $\vec{x}$ в виде следующей линейной комбинации

$$\begin{cases} x_1 = l_{11} f_1 + \cdots + l_{1k} f_k + \varepsilon_1 \\ \vdots \\ x_r = l_{r1} f_1 + \cdots + l_{rk} f_k + \varepsilon_r \end{cases}, \tag{1}$$

где $f_1, \dots, f_k$ – центрированно-нормированные некоррелированные общие факторы; $\varepsilon_1, \dots, \varepsilon_r$ – центрированные частные (или специфические) факторы. При этом предполагается, что частные факторы некоррелированы между собой, некоррелированы с общими факторами $f_1, \dots, f_k$, и имеют некоторые неизвестные конечные дисперсии $\mathbb{D}\varepsilon_j = \nu_j < \infty, j = 1, \dots, r$.

Учитывая условие некоррелированности общих факторов и свойство линейности коэффициента корреляции, нетрудно видеть, что коэффициент корреляции между m-м показателем x_m и общим фактором f_q есть $\rho(x_m, f_{\mathrm{q}}) = l_{mq}, m = 1, \dots, r, q = 1, \dots, k$. Значения l_{mq}, $m = 1, \dots, r, q = 1, \dots, k$ принято называть *нагрузками* m-го показателя на q-й фактор.

Представим соотношение (1) в компактной матричной форме. Введём детерминированную матрицу нагрузок **L** размера $r \times k$ с элементами $l_{mq}, m = 1, \dots, r, q = 1, \dots, k$; случайный вектор $\vec{f} = (f_1, \dots, f_k)$ общих факторов и случайный вектор частных факторов $\vec{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_r)$. В этих обозначениях система (1) примет следующий вид:

$$\vec{x} = \mathbf{L}\vec{f} + \vec{\varepsilon}. \tag{2}$$

Для получения представления (2) необходимо найти матрицу нагрузок **L** и дисперсии частных факторов $\varepsilon_1, \dots, \varepsilon_r$.

Следует отметить, что представление вектора признаков $\vec{x}$ в форме (2) не является единственным. Рассмотрим, например, произвольную ортогональную матрицу **B** размера $k \times k$. Учитывая свойство ортогональности матрицы **B**, соотношение (2) можно теперь записать следующим образом:

$$\vec{x} = (\mathbf{L}\mathbf{B}^T)(\mathbf{B}\vec{f}) + \vec{\varepsilon} = \mathbf{L}_1 \vec{f}^{(1)} + \vec{\varepsilon}.$$

В последнем уравнении вектор $\vec{f}^{(1)} = \mathbf{B}\vec{f}$ будет представлять собой «новый» вектор общих факторов, а матрица $\mathbf{L}_1 = \mathbf{L}\mathbf{B}^T$ – «новую» матрицу нагрузок. Известно, что умножение на ортогональную матрицу осуществляет поворот осей системы координат на некоторый угол. Поэтому вектор общих факторов $\vec{f}$ и матрица нагрузок **LB** модели (2) определяются лишь с точностью до ортогонального преобразования (поворота). По этой причине в факторном анализе возникает ещё одна задача – вращение факторов. Суть этой задачи состоит в том, чтобы подобрать такое преобразование, которое позволило бы дать наилучшую содержательную интерпретацию выявленных общих факторов. Основными методами процедуры вращения являются метод Varimax, который минимизирует количество признаков, связанных с каждым из общих факторов, и метод Quartimax, минимизирующий количество общих факторов, необходимых для объяснения признака. Данные методы реализованы в статистических программных продуктах.

Перейдём к нахождению матрицы нагрузок. Для определения матрицы нагрузок также известно несколько методов. Среди них: метод максимального правдоподобия [8], предполагающий гауссовость наблюдений; метод главных факторов [5] и метод минимальных остатков [3]. Метод максимального правдоподобия некорректно применять в нашей прикладной задаче, поскольку предварительный анализ показывает, что гипотеза о гауссовости наблюдений отвергается. Поэтому мы использовали метод минимальных остатков (MinRes), предложенный Харманом в [3]. Кратко опишем его суть.

Пусть число общих факторов k известно (способы выбора k обсуждаются ниже). Поскольку факторы $f_1, \dots, f_k$ некоррелированы, корреляционная матрица **R** вектора **X** согласно в соответствии с представлением (2) выражается следующим образом:

$$\mathbf{R} = \mathbf{L}\mathbf{L}^T + \mathbf{D}, \tag{3}$$

где **D** – диагональная матрица размера $r \times r$ с диагональными элементами $\nu_{\mathrm{q}} = \mathbb{D}\varepsilon_q$, $q = 1, \dots, r$.

Идея метода минимальных остатков состоит в том, чтобы выбрать такие факторные нагрузки, при которых сумма квадратов расхождений между наблюдаемыми корреляциями исходных показателей и предсказанными значениями этих корреляций согласно модели (3), будет минимальной. Таким образом, в рамках метода MinRes нахождение $k(k-1)/2$ внедиагональных элементов матрицы $\mathbf{L}$ сводится к решению следующей задачи минимизации

$$f(\mathbf{L}) = \sum_{i=1}^{r-1} \sum_{j=i+1}^{r} \left(\rho_{ij} - \sum_{p=1}^{k} l_{ip} l_{jp} \right)^2 \to \min_{\mathbf{L}}.$$

Алгоритм нахождения аргминимума функции $f(\mathbf{L})$ с учётом ограничения

$$\sum_{j=1}^{k} l_{ij}^2 \leq 1, \quad i = 1, \ldots, r$$

подробно описан в работе [9]. После нахождения факторных нагрузок остаточные дисперсии $\mathbb{D}\varepsilon_q$, $q = 1, \ldots, r$ определяются в соответствии с формулой (3).

Необходимым условием для применения метода минимальных остатков является знание количества общих факторов k. Для определения числа k можно использовать разные подходы: 1) метод, основанный на выборе числа факторов, которые описывают достаточно большую долю дисперсии исходных признаков; 2) метод «каменистой осыпи» или 3) метод Кайзера. В нашем исследовании был применён метод Кайзера, суть которого заключается в том, чтобы выбирать количество общих факторов равным числу собственных значений корреляционной матрицы $\mathbf{R}$, превышающих единицу. Поскольку сумма собственных значений матрицы $\mathbf{R}$ равна количеству показателей r, то фактор с собственным значением меньше единицы вносит вклад в общую изменчивость меньший, чем один исходный показатель. Поэтому такой фактор исключается из рассмотрения в качестве значимой обобщённой характеристики. Отметим также, что в некоторых задачах исследователю бывает априори известно количество общих факторов. В таких ситуациях метод минимальных остатков, требующий в качестве входных данных число общих факторов и корреляционную матрицу наблюдаемых показателей, представляется удобным выбором.

Для применения любого метода определения числа общих факторов и любого метода факторного анализа требуется знать корреляционную матрицу наблюдаемых показателей. На практике эта матрица неизвестна, и её требуется оценить по имеющейся выборке. Рассмотрим несколько методов оценивания.

1.1. Оценивание корреляционных матриц

Многочисленные исследования (например, [10, 11]) показывают, что реальные данные часто содержат аномальные наблюдения (выбросы). Это обуславливает использование робастных методов, устойчивых к различного рода загрязнениям данных. Одной из количественных мер робастности оценок является *пороговая точка (breakdown point)* ε^*, формальное определение которой дано в [12] (с. 309). Пороговая точка ε^* определяет наименьшую долю выбросов в выборке, при которой оценка может принимать сколь угодно большие (неограниченные) значения. Оценки, с пороговой точкой близкой к нулю, не являются устойчивыми к засорениям, а оценки с пороговой точкой близкой к 0.5, обладают максимальной устойчивостью.

Оценка Пирсона. В традиционном факторном анализе в качестве оценки корреляционной матрицы $\mathbf{R}$ выбирают матрицу $\widehat{\mathbf{R}}$, элементами которой являются выборочные коэффициенты корреляции Пирсона. Эта оценка является эффективной в случае, когда выборка имеет гауссовское распределение. Однако, как показано в [13] (с. 271), пороговая точка такой оценки составляет $1/n$, где n – объём выборки, и стремится к нулю при $n \to \infty$.

*Оценка*MCD *(Minimum Covariance Determinant)*. Данная оценка была предложена Руссо и Лероем в [13]. Она определяется как выборочная корреляционная матрица Пирсона, построенная на тех h наблюдениях, для которых детерминант выборочной ковариационной матрицы является наименьшим среди всех возможных подвыборок объёма h. Если в качестве h выбрать $h = \left\lfloor \frac{n+r+1}{2} \right\rfloor$, то пороговая точка такой оценки примет максимальное значение $\varepsilon^* = \frac{n+1-h}{n}$. В [14] предложен быстрый алгоритм Fast-MCD для вычисления MCD-оценок. В настоящее время этот алгоритм реализован во многих современных статистических пакетах. Отметим, что при $h = n$ оценка MCD совпадает с оценкой Пирсона.

Оценка Олива-Хокинса. Построение оценки Олива-Хокинса (OH-оценка) [15, 16] является двухэтапной процедурой. Сначала строят выборочную оценку по половине наблюдений с наименьшим расстоянием Махаланобиса от центра и выборочную оценку по половине наблюдений, ближайших (в смысле евклидова расстояния) к покоординатной медиане. Затем по определённому правилу , описанному в [15], выбирают лучшую из двух этих оценок. В [16] показано, что такая оценка является робастной в смысле близости её пороговой точки к 0.5.

1.2. Вычисление факторных баллов

После оценивания факторных нагрузок можно оценить значения общих факторов $f_1, \ldots, f_k$ для каждого наблюдения. Метод оценивания был предложен Бартлеттом. Для каждого объекта $i = 1, \ldots, n$ рассматривается модель линейной регрессии:

$$X_j^{(i)} = f_1^{(i)} l_{j1} + \cdots + f_k^{(i)} l_{jk} + \varepsilon_j, \quad j = 1, \ldots, r, \tag{4}$$

где l_{jm} – элементы построенной матрицы нагрузок **L**, $X_j^{(i)}$ – наблюдаемое значение j-й компоненты i-го объекта, ε_j – ненаблюдаемые погрешности с нулевым средним, а $f_1^{(i)}, \ldots, f_k^{(i)}$ – неизвестные параметры.

Применяя метод наименьших квадратов для оценивания параметров $f_1^{(i)}, \ldots, f_k^{(i)}$, получаем оценки общих факторов для каждого объекта. Назовём такие оценки *факторными баллами* объекта.Таким образом, можно считать, что наблюдение, имеющее наивысший факторный балл f_j, является тем объектом, у которого латентный общий фактор f_j выражен наиболее сильно по сравнению с другими наблюдаемыми объектами.

Отметим также, что для оценивания параметров $f_1^{(i)}, \ldots, f_k^{(i)}$ в модели (4) можно было использовать и иные методы оценивания параметров линейной регрессионной модели.

2. Применение факторного анализа к реальным данным

Данные для анализа собраны на основе ежегодной публикации Федеральной службы государственной статистики (Росстат), отражающей явления и процессы в экономической и социальной жизни Российской Федерации. Выборка содержит информацию о социально-экономических показателях 85 субъектов России (включая 3 города федерального значения: Москву, Санкт-Петербург и Севастополь) за 2023 год.

Руководствуясь публикацией ИСИЭЗ НИУ ВШЭ [17], для анализа инновационного потенциала регионов РФ были отобраны 12 показателей:
X_1: Численность персонала, занятого научными исследованиями и разработками (человек)
X_2: Внутренние затраты на научные исследования и разработки (млн рублей)
X_3: Подано патентных заявок на изобретения и модели (штук)
X_4: Выдано патентов на изобретения и модели (штук)
X_5: Численность исследователей с учеными степенями (человек)
X_6: Уровень инновационной активности организаций (%)
X_7: Удельный вес организаций, осуществлявших технологические инновации (%)
X_8: Объем инновационных товаров, работ, услуг (% от общего объема)
X_9: Процент населения, имеющего доступ в Интернет
X_{10}: Доля населения 15-74 лет, использовавшего мобильный телефон/смартфон
X_{11}: Доля населения 15-74 лет, использовавшего Интернет для заказа товаров/услуг
X_{12}: Доля населения 15-72 лет, использовавшего Интернет для получения госуслуг.

Можно заметить, что показатели X_1–X_5 характеризуют научное развитие, X_6–X_8 – инновационное развитие бизнеса, а X_9–X_{12} – уровень цифровизации экономики.

Естественно предположить, что показателями внутри каждой из трёх указанных групп окажутся сильно коррелированными. Это позволяет выдвинуть гипотезу о возможности сжатия структуры взаимосвязей до небольшого числа факторов (в данном случае – трёх факторов). Для проверки гипотезы и определения количества факторов был использован критерий Кайзера. Корреляционная матрица оценена методами Пирсона, MCD и Олива-Хокинса (OH). Во всех случаях получено по три собственных значения, превышающих единицу (см. табл. 1), что подтверждает выбор трёх общих латентных факторов.

Таблица 1. Собственные значения оцененных корреляционных матриц

Метод	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Пирсон	5,49	2,27	1,73	0,78	0,66	0,48	0,44	0,08	0,05	0,02	0,002	0,0007
MCD	4,72	2,29	1,64	0,95	0,74	0,58	0,47	0,27	0,16	0,12	0,05	0,01
OH	5,94	2,07	1,27	0,84	0,59	0,49	0,32	0,24	0,13	0,09	0,02	0,01

Следующим этапом является оценивание матрицы факторных нагрузок. Применение теста Шапиро–Уилка [18] показало, что гипотеза о нормальном распределении наблюдаемых данных отвергается на любом разумном уровне значимости. По этой причине метод максимального правдоподобия, предполагающий нормальность наблюдений, следует в данной ситуации признать некорректным. Аналогично, в качестве оценки корреляционной матрицы не подходит выборочная корреляционная матрица Пирсона, обладающая практически нулевой пороговой точкой.

В итоге нами был выбран метод минимальных остатковMinRes, использующий робастные оценки корреляционных матриц MCD и Олива-Хокинса (OH). Для улучшения интерпретируемости факторной структуры были применены ортогональные вращения Varimax и Quartimax. Поскольку эти вращения дали схожие результаты, мы приведём только матрицу факторных нагрузок после вращения Varimax.

В таблице 2 представлены факторные нагрузки показателей X_1–X_{12} на общие факторы f_1, f_2, f_3, полученные методом минимальных остатков, основанном на робастных оценках MCD и OH корреляционной матрицы наблюдений с последующим вращением факторной структуры методом Varimax.

Таблица 2. Матрица факторных нагрузок, оцененная методом минимальных остатков с использованием вращения Varimax

Переменная	**Фактор 1**		**Фактор 2**		**Фактор 3**	
	MCD	OH	MCD	OH	MCD	OH
X_1	0,979	0,939	0,159	0,199	-0,038	0,058
X_2	0,969	0,953	0,134	0,132	-0,033	0,079
X_3	0,972	0,97	0,37	0,34	0,1	0,113
X_4	0,958	0,962	0,126	0,113	0,114	0,135
X_5	0,98	0,949	-0,031	-0,095	-0,1	0,079
X_6	0,136	0,316	0,808	0,859	-0,116	0,015
X_7	0,232	0,372	0,712	0,8	-0,162	0,026
X_8	0,049	0,298	0,67	0,736	-0,014	0,022
X_9	0,002	-0,002	-0,056	-0,042	0,591	0,586
X_{10}	0,08	-0,036	-0,03	-0,012	0,663	0,704
X_{11}	0,009	0,177	-0,031	0,05	0,581	0,453
X_{12}	-0,163	0,056	-0,052	-0,119	0,697	0,583

Можно заметить, что показатели X_1–X_{12}чётко группируются в три общих фактора. Полученная структура представляется логичной и содержательно интерпретируемой, что позволяет каждому из факторов можно присвоить содержательное название:

Фактор 1 – «Научное развитие»;

Фактор 2 – «Инновационное развитие бизнеса»;

Фактор 3 – «Уровень цифровизации».

Действительно, первые пять показателей (X_1–X_5) имеют очень высокие нагрузки на Фактор 1, показатели X_6–X_8 демонстрируют высокие нагрузки на Фактор 2, а показатели X_9–X_{12} имеют достаточно высокие нагрузки на Фактор 3. При этом нагрузки указанных показателей на другие факторы оказались близкими к нулю. Отметим также, что несмотря на некоторые несущественные различия в оценках факторных нагрузок для MCD и OH, оба метода дают схожую качественную картину.

На заключительном этапе были восстановлены значения латентных факторов для каждого субъекта на основании модели Бартлетта (4). Высокие значения факторных баллов свидетельствуют о высокой степени развитости субъектов по соответствующему направлению, в то время как низкие или отрицательные значения указывают на её недостаточность. В таблице 3 представлены пять лидеров рейтинга по каждому из трёх указанных факторов.

Таблица 3. Топ-5 рейтинга регионов по факторам

Место	Научное развитие	Инновационное развитие бизнеса	Уровень цифровизации
1	г. Москва	Республика Татарстан	ЯНАО
2	г. Санкт-Петербург	Ростовская область	ХМАО
3	Московская область	Республика Мордовия	Республика Татарстан
4	Свердловская область	Самарская область	Республика Калмыкия
5	Новосибирская область	Нижегородская область	Чукотский АО

Важно отметить, что ИСИЭЗ НИУ ВШЭ также публикует рейтинг инновационного развития регионов РФ [17], построенный с помощью иной методологии. Однако результаты проведённого нами факторного анализа демонстрируют высокую степень согласованности с указанным экспертно-аналитическим рейтингом.

А именно, регионы, определённые как лидеры по фактору «Научное развитие», входят в топ-15 рейтинга в работе [17]. Более того, порядок их расположения (за исключением Новосибирской области) совпадает в обоих подходах.

По фактору «Инновационное развитие бизнеса» оба рейтинга выделяют Республику Татарстан как лидера. Кроме того, все субъекты, идентифицированные факторным анализом как лидирующие, входят в топ-20 рейтинга [17].

По фактору «Уровень цифровизации», включённому в индекс «Социально-экономические условия инновационной деятельности» в [17], регионы-лидеры в нашем анализе демонстрируют сопоставимые позиции в расчётах ИСИЭЗ. Более того, порядок их ранжирования полностью соответствует значениям, представленным в указанном индексе.

3. Заключение

В данной работе успешно применены методы факторного анализа с использованием робастных оценок корреляционных матриц для выявления регионов-лидеров инновационного развития среди субъектов Российской Федерации. Учитывая существенное отклонение распределения анализируемых социально-экономических показателей от гауссовского, обоснован выбор устойчивых методов оценивания (MCD и Олива-Хокинса) вместо классической корреляции Пирсона.

Выявленные общие факторы – «Научное развитие», «Инновационное развитие бизнеса» и «Уровень цифровизации» – позволили комплексно охарактеризовать инновационный потенциал регионов. На основе факторных баллов определён рейтинг субъектов РФ.

Важно отметить, что несмотря на различия в методологии и наборе исходных показателей, результаты данного исследования демонстрируют высокую согласованность с независимым рейтингом ИСИЭЗ НИУ ВШЭ. Это подтверждает адекватность предложенного робастного подхода для анализа данных с аномальными наблюдениями и его эффективность для решения задач выявления скрытых структур взаимосвязей и объективного ранжирования территорий по уровню инновационного развития.

Полученные результаты создают основу для разработки адресных стратегий регионального развития и могут быть использованы при оценке эффективности государственных программ поддержки инноваций.

Литература

1. *Иберла К.* Факторный анализ. М.: Статистика, 1980. – 398 с.
2. *Лоули Д., Максвелл А.* Факторный анализ как статистический метод. М.: Мир, 1967. – 145 с.
3. *Harman H., Jones W.* Factor analysis by minimizing residuals (minres) // Psychometrika. – 1966. – Vol. 31, № 3. – P. 351–369.
4. *Харман Г.* Современный факторный анализ. М.: Статистика, 1972.
5. *Дронов С.В.* Многомерный статистический анализ. Барнаул: Алтайский государственный университет, 2003. – 213 с.
6. *Горяинов В.Б., Горяинова Е.Р.* Исследование устойчивости к аномальным наблюдениям модификаций метода главных компонент // Вестник МГТУ им. Н.Э. Баумана. Сер. Естественные науки. – 2023. – № 2. – С. 17–34.
7. *Goryainov V.B., Goryainova E.R.* Comparison of the Quality of Robust PCA versions in the Reducion of datasets with Outliers // Proceedings of the 16th International Conference Management of Large-Scale System Development (MLSD), 2023.
8. *Ивченко Г.И., Медведев Ю.И.* Введение в математическую статистику. М.: ЛКИ, 2010. – 600 с.
9. *Гольтяпин В.В.* Вычислительные аспекты метода минимальных остатков при разрешении варианта Хейвуда // Сибирский журнал индустриальной математики. – 2005. – Т. 8, № 3. – С. 55–67.

10. *Huber P.J., Ronchetti E.M.* Robust statistics. Hoboken: Wiley, 2009. – 360 p.
11. *Maronna R.A., Martin R.D., Yohai V.J. et al.* Robust statistics: Theory and methods (with R). Wiley, 2019.
12. *Maronna R., Zamar R.H.* Robust estimates of location and dispersion for high-dimensional datasets // Technometrics. – 2002. – Vol. 44, iss. 4. – P. 307–317. DOI: 10.1198/004017002188618509.
13. *Rousseeuw P.J., Leroy A.M.* Robust Regression and Outlier Detection. Chichester: Wiley, 1987. – 347 p.
14. *Rousseeuw P.J., van Driessen K.* A fast algorithm for the minimum covariance determinant estimator // Technometrics. – 1999. – Vol. 41, iss. 3. – P. 212–223. DOI: 10.2307/1270566.
15. *Olive D.J.* A resistant estimator of multivariate location and dispersion // Comput. Statist. Data Anal. – 2004. – Vol. 46, № 1. – P. 93–102.
16. *Olive D.J.* Robust multivariate analysis. Cham: Springer, 2017. – 501 p.
17. *Абашкин В.Л., Абдрахманова Г.И., Артёмов С.В. и др.* Рейтинг инновационного развития субъектов Российской Федерации. Вып. 9. М.: ИСИЭЗ ВШЭ, 2024. – 248 с.
18. *Shapiro S.S., Wilk M.B.* An analysis of variance test for normality // Biometrika. – 1965. – Vol. 52. – P. 591–611.