

Налоги и налогообложение

*Правильная ссылка на статью:*

Апалькова Т.Г., Левченко К.Г. Статистический анализ дифференциации регионов РФ по налоговым доходам консолидированных бюджетов средствами языка R // Налоги и налогообложение. 2024. № 3. DOI:

10.7256/2454-065X.2024.3.70590 EDN: RUQJWX URL: [https://nbpublish.com/library\\_read\\_article.php?id=70590](https://nbpublish.com/library_read_article.php?id=70590)

## Статистический анализ дифференциации регионов РФ по налоговым доходам консолидированных бюджетов средствами языка R

**Апалькова Тамара Геннадьевна**

ORCID: 0000-0001-8094-1588

кандидат экономических наук

доцент, кафедра математики, Финансовый университет при Правительстве Российской Федерации

125993, Россия, г. Москва, пр-т Ленинградский, 49

✉ [apalkova.t.g@yandex.ru](mailto:apalkova.t.g@yandex.ru)



**Левченко Кирилл Геннадиевич**

ORCID: 0009-0008-7380-3388

кандидат физико-математических наук

доцент, кафедра математики, Финансовый университет при Правительстве Российской Федерации

125167, Россия, Москва, г. Москва, ул. Ленинградский Проспект, 49

✉ [kglevchenko@fa.ru](mailto:kglevchenko@fa.ru)



[Статья из рубрики "Прогнозирование и планирование в налогообложении"](#)

### DOI:

10.7256/2454-065X.2024.3.70590

### EDN:

RUQJWX

### Дата направления статьи в редакцию:

27-04-2024

### Дата публикации:

08-05-2024

**Аннотация:** Предметная область настоящей статьи – применение методов описательной

статистики и многомерной классификации для описания региональных особенностей формирования налоговых составляющих доходов консолидированных бюджетов РФ. В работе преследуется цель продемонстрировать простоту и эффективность применения математико-статистических методов и функционала языка с открытым кодом R для решения задач структурного анализа налоговых поступлений, выявления региональной специфики, сравнительного анализа регионов с точки зрения налоговых доходов. Аппарат математической статистики, реализованный в языке R, в частности, открывает широкие возможности для классификации субъектов налогообложения, в том числе многомерной, существенно облегчая процедуры анализа, ранжирования и планирования. Описанный в статье функционал может быть использован в процессе формирования и корректировки налоговой политики на разных уровнях. Возможности аппарата математической статистики в сочетании с инструментальными методами языка R раскрываются на примере классификационного анализа регионов РФ. При этом в качестве классификационных признаков выбраны абсолютные и относительные значения налоговых доходов в доходах региональных бюджетов. Рассматривается классификация по принадлежности к федеральному округу и исследование "естественного" расслоения методом кластерного анализа. Аппарат математической статистики и, особенно, инструментарий языка R применяются в исследованиях подобного рода неоправданно редко, несмотря на простоту использования и отсутствия необходимости в специальной подготовке, эти обстоятельства определяют актуальность настоящей статьи. Агрегирование по федеральным округам позволило выделить: Уральский федеральный округ как лидирующий по среднерегиональной доле налоговых доходов в доходной части бюджета и Северо-Кавказский федеральный округ, характеризующийся наименьшими среднерегиональными вкладами налоговых платежей в региональные бюджеты. Анализ естественного расслоения регионов РФ по их относительным налоговым вкладам в консолидированные бюджеты дал возможность выделить группы: наиболее типичных регионов, дотационных регионов, регионов-доноров и регионов, в которых сосредоточены наиболее дорогие активы предприятий Российской Федерации

### **Ключевые слова:**

язык R, математическая статистика, региональные налоговые доходы, многомерная классификация, кластерный анализ, описательная статистика, налоговые доходы бюджета, статистика сбора налогов, дотационные регионы, методы анализа налогообложения

### **Введение**

Изучение публикаций, исследующих особенности, структуру и дифференциацию налоговых платежей и сборов [\[1-5\]](#) показал, что математико-статистический анализ в этих работах встречается скорее редко и в большинстве случаев сводится к анализу динамики [\[2, 3\]](#). Однако в сочетании с современными техническими средствами методы математической статистики представляют собой мощный инструмент, позволяющий осветить процессы сбора, начисления, уплаты налогов с разных сторон и выявить таким образом наиболее проблемные и, напротив, позитивные аспекты этих процессов. Настоящая статья преследует цель продемонстрировать возможности, эффективность и простоту средств визуализации, описательной статистики, агрегирования и кластеризации языка программирования с открытым кодом R в налоговом анализе.

### Данные для моделирования.

Исходными данными послужили сведения об исполнении консолидированных бюджетов субъектов Российской Федерации за 2020 год (Минфин России. Данные об исполнении консолидированных бюджетов субъектов Российской Федерации), представленные источником в формате файла .xlsx. Импортировать данные такого формата в среду R можно при помощи функции `readxl::read_excel()`. Здесь и далее по тексту запись перед двойным двоеточием обозначает название пакета, к которому принадлежит функция, указанная после двойного двоеточия. В упомянутом файле содержатся сведения о доходах и расходах консолидированного бюджета, доходы дифференцируются на налоговые и неналоговые, отдельно выделены три наиболее «весомых» для каждого региона налога: на прибыль, на имущество и на доходы физических лиц. На рисунке 1 демонстрируется фрагмент исходного файла с небольшим изменением: после столбца «Регион» добавлен столбец «Округ», строки с промежуточным и итоговым суммированием (по округам и стране в целом) удалены. Таким образом сформированные данные носят название «прямоугольных»: каждая строка содержит сведения об одном объекте, каждый столбец отвечает за конкретный признак. Помимо полученных извне в набор данных были добавлены несколько синтетических признаков: относительный вклад налоговых поступлений (общих и по каждому налогу) в доходную часть бюджетов.

Регион	Округ	Итого доходов	Налоговые и неналоговые доходы
Белгородская область	ЦФО	133 462 852	97 940 945
Брянская область	ЦФО	85 935 007	39 449 257
Владимирская область	ЦФО	90 131 841	62 827 803
Воронежская область	ЦФО	168 139 487	118 661 445
Ивановская область	ЦФО	66 837 973	30 818 609
Калужская область	ЦФО	93 722 432	67 731 121
Костромская область	ЦФО	46 991 542	27 239 855
Курская область	ЦФО	83 635 069	58 046 204
Липецкая область	ЦФО	85 868 944	62 054 647
Московская область	ЦФО	805 479 821	683 311 560
Орловская область	ЦФО	48 823 652	27 252 150

Рисунок 1 – Набор исходных данных (фрагмент)

**Агрегирование данных.** В случае, если для исследователя представляет интерес различие налоговых поступлений в консолидированный бюджет в разрезе федеральных округов, можно определить средние значения по каждому округу. Для этой цели применимы базовые функции: `aggregate()` – в случае, когда нужно посчитать средние сразу по нескольким признакам, или `tapply()` <sup>[6]</sup> – если средние вычисляются по одному признаку. Рассмотрим пример применения функции `aggregate()` <sup>[7]</sup> для вычисления средних по федеральным округам значений суммарных налоговых доходов (таб. 1), а также – средних долей налоговых доходов (суммарных и по видам) в доходной части бюджета (таб. 2).

В общем виде синтаксис функции `aggregate()` выглядит следующим образом:

`aggregate(x = <усеченный, или полный набор данных>,`

`by=list(<один, или несколько группировочных признаков>),`

FUN = mean)

Таблица 1- Среднерегиональные значения доходов консолидированных бюджетов по видам доходов и федеральным округам, млн рублей

Округ	Итого доходов	Налоговые и неналоговые доходы	Налоговые доходы	Налог на прибыль организаций	Налог на имущество организаций	Налог на доходы физических лиц
ДВФО	119511257	73507483	67380456	20812532	7739126	26092894
ПФО	149479227	99578519	93573818	22474054	7927675	37393626
СЗФО	149216638	116133709	109882505	31548774	10284799	48888251
СКФО	92213275	29249889	27493643	3830534	2475400	12663845
СФО	150846949	101698591	95735467	29090442	7086183	36871352
УФО	232046876	192116831	182897714	62714623	30552349	65740254
ЦФО	283274516	234648498	215238958	66308098	14350978	99215141
ЮФО	150454181	89772487	84085745	18571173	8863769	33242862

Таблица 2- Среднерегиональные значения относительных вкладов налоговых доходов в консолидированные бюджеты по видам доходов и федеральным округам

Округ	Налог на прибыль организаций	Налог на имущество организаций	Налог на доходы физических лиц	Налоговые доходы
ДВФО	0,15	0,06	0,21	0,53
ПФО	0,13	0,05	0,24	0,59
СЗФО	0,16	0,09	0,25	0,63
СКФО	0,03	0,03	0,13	0,27
СФО	0,14	0,04	0,22	0,54
УФО	0,26	0,12	0,27	0,75
ЦФО	0,16	0,05	0,26	0,64
ЮФО	0,10	0,05	0,19	0,48

Более наглядным сравнение средних по округам значений делают графики, в данном случае наиболее уместны столбиковые диаграммы (рис. 1 и 2). На рисунках 1 и 2 построены графики для средних по каждому округу величин налоговых доходов, использованы агрегированные данные таблиц 1 и 2 соответственно. В качестве инструмента визуализации может быть использована функция `ggplot2::ggplot()` в сочетании одной из функций быстрого доступа: `geom_bar()`, или `geom_col()` [8-10]. При этом, в качестве аргументов функций быстрого доступа следует указать переменные отвечающие за координаты *x* и *y* – федеральный округ и средние по округу налоговые доходы бюджета (для рис. 1), или средние по округу доли налоговых доходов в бюджете (для рис. 2) соответственно. Кроме того, функция `geom_bar()` в данной ситуации требует дополнительно указания значения аргумента `stat='identity'`, что позволит отражать по оси ординат не частоты, а значения координаты *y*. Функция `geom_col()` этого дополнительного аргумента не требует.

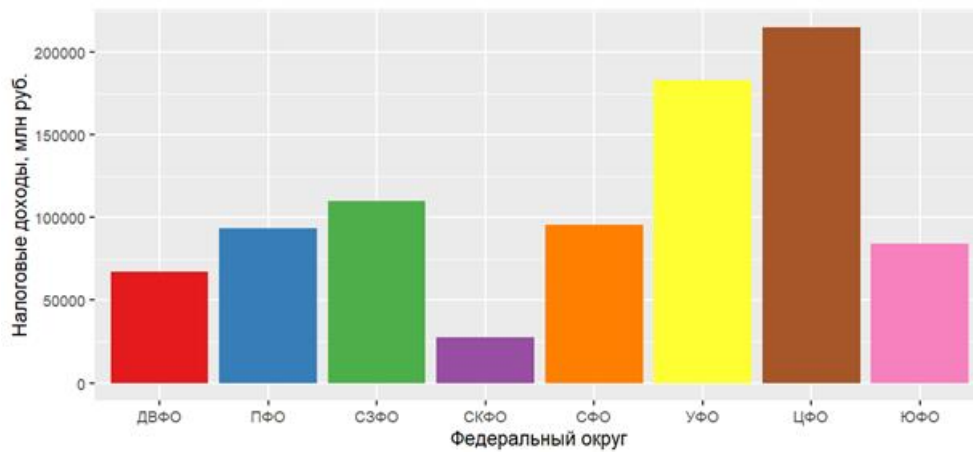


Рисунок 1 – Среднерегиональные значения налоговых доходов по федеральным округам, млн рублей

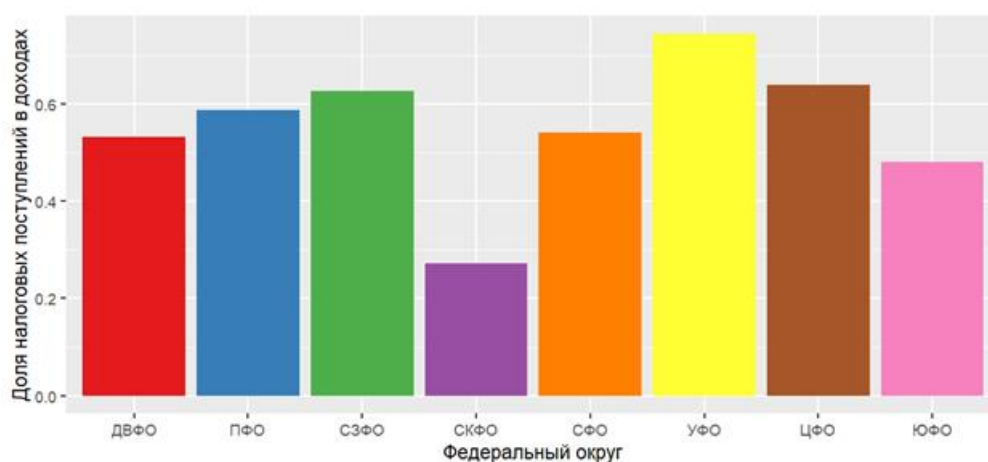


Рисунок 2 – Среднерегиональные значения относительных вкладов налоговых доходов в консолидированные бюджеты регионов по федеральным округам

При сравнении диаграмм становится очевидно, что в результате вычисления относительных показателей эффект масштаба нивелируется: для нескольких округов (Дальневосточный, Приволжский, Северо-Западный, Северный, Центральный) среднерегиональные доли налоговых поступлений в бюджеты различаются мало, Центральный округ, лидирующий по абсолютному среднему вкладу, уступает Уральскому округу по доле входу (0,64 и 0,75 от всех доходов бюджета соответственно). Регионы Северо-Кавказского Федерального округа характеризуются в среднем как наиболее низкими суммами собираемых налогов, так и минимальной долей налоговых поступлений при формировании доходной части бюджета (0,27). Последний вывод вполне закономерен, поскольку из шести дотационных регионов РФ, которые на конец 2019 года два года из трех получали дотации не менее 40% от собственных доходов, три находятся в Северо-Кавказском Федеральном округе (Приказ Министерства Финансов «Об утверждении перечня субъектов Российской Федерации в соответствии с положениями пункта 5 статьи 130 Бюджетного кодекса Российской Федерации»).

**Кластеризация.** Далее рассмотрим возможность выявления естественного расслоения регионов по совокупности признаков, характеризующих долю налоговых платежей в доходной части бюджетов. При этом учитываются платежи: по налогам на прибыль, имущество, доходы физических лиц и совокупные налоговые поступления. Наиболее подходящим методом в данном случае является кластерный анализ, обеспечивающий классификацию «без учителя» - то есть правило разбиения на кластеры априори

неизвестно [11, 12].

Процедура агломеративного кластерного анализа в языке R может быть реализована при помощи достаточно простого кода. Приведем ее поэтапно с надлежащими комментариями.

1) Вычисление расстояний между объектами осуществляется командой

`d<-dist(scale(PT))`, где PT – название набора данных, содержащего значения классификационных признаков (доли платежей по налогам в бюджете). В качестве метрики по умолчанию используется простое Евклидово расстояние:

$$d(p, q) = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}, \quad (1)$$

где  $d$  – расстояние,  $p_i$  – координата  $i$  первого объекта (значение признака  $i$ ),  $q_i$  – координата  $i$  второго объекта,  $i=1, \dots, n$ ,  $n$  – количество классификационных признаков.

2) Объединение в кластеры в рассматриваемом примере осуществляется по методу Варда, который подразумевает на каждом шаге объединение объектов, приводящее к минимальному увеличению внутригрупповой суммы квадратов. для этого выполняется команда: `hsw`

3) Далее при помощи команды `plot(hsw, cex=0.5)` строится дендрограмма и на ней выделяются кластеры, количество которых задается исследователем (в рассматриваемом примере их 4). Для этого применяется функция `rect.hclust(hsw, k = 4)`. Результат данного этапа отображается на графике (рис. 3):

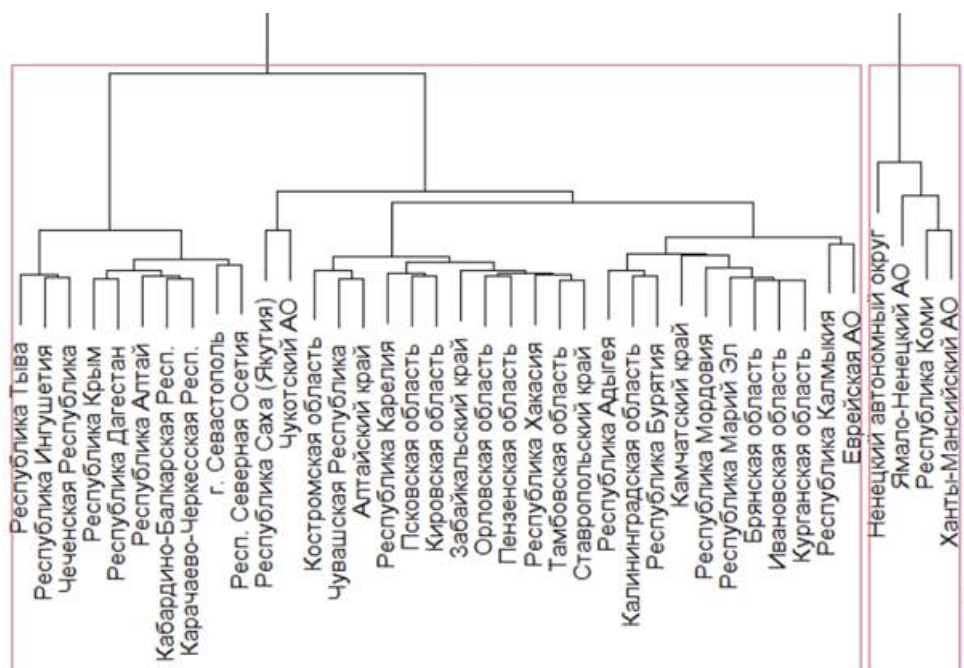


Рисунок 3 – Первый способ визуализации результатов кластерного анализа (фрагмент дендрограммы: кластеры 2 и 4)

Отметим, что окончательное количество кластеров определяется эмпирически, существуют процедуры, оценивающие качество разбиения и позволяющие сравнивать разные варианты. Рассмотрение этих процедур выходит за рамки настоящей работы и пример демонстрирует одно из возможных разбиений (4 кластера).



Рассмотренные выше функции, применяемые в процессе кластеризации, относятся к категории базовых и представляют собой набор минимально достаточных инструментов. Однако в R предусмотрены также приёмы, позволяющие облегчить восприятие результатов кластеризации. Так, следующий код:

```
library("ape") #подключение специальной библиотеки

colors = c("red", "blue", "green", "black") #определение цветов диаграммы по количеству
кластеров

groups4

plot(as.phylo(hcw), type = "fan", tip.color = colors[groups4],
label.offset = 0.5, cex = 0.7, no.margin = TRUE) #построение графика
```

является альтернативой этапу 3 и представляет дендрограмму в виде круга и выделяет объекты, относящиеся к каждому кластеру одним цветом, рис. 4.

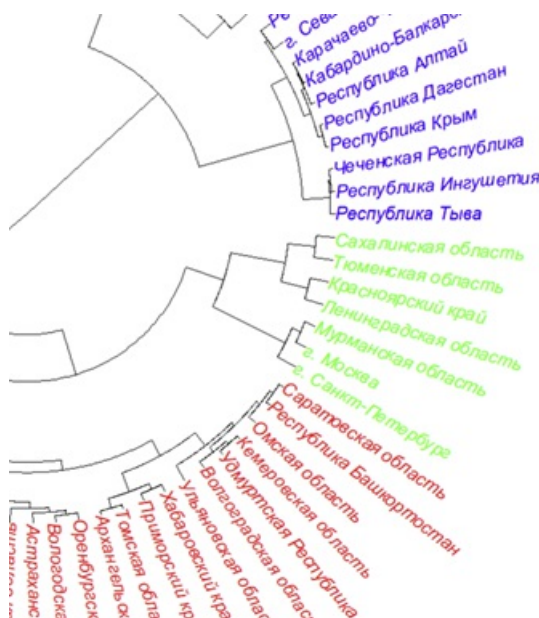


Рисунок 4 – Второй способ визуализации результатов кластерного анализа, с применением функционала библиотеки *ape* (фрагмент дендрограммы, представлены частично кластеры 1 и 2. полностью - кластер 3 )

4 ) на последнем этапе, когда состав кластеров идентифицирован, необходимо охарактеризовать полученные кластеры и выявить ключевые принципы полученного разбиения. В описываемом случае состав кластеров определился следующим образом :

Первый кластер: Белгородская область, Владимирская область, Воронежская область, Калужская область, Курская область, Липецкая область, Московская область, Рязанская область, Смоленская область, Тверская область, Тульская область, Ярославская область, Архангельская область, Вологодская область, Новгородская область, Краснодарский край, Астраханская область, Волгоградская область, Ростовская область, Республика Башкортостан, Республика Татарстан, Удмуртская Республика, Пермский край, Нижегородская область, Оренбургская область, Самарская область, Саратовская область, Ульяновская область, Свердловская область, Челябинская область, Иркутская область, Кемеровская область, Новосибирская область, Омская область, Томская область, Приморский край, Хабаровский край, Амурская область, Магаданская область.

Второй кластер: Брянская область, Ивановская область, Костромская область, Орловская область, Тамбовская область, Республика Карелия, Калининградская область, Псковская область, Республика Адыгея, Республика Калмыкия, Республика Крым, г. Севастополь, Республика Дагестан, Республика Ингушетия, Кабардино-Балкарская Респ., Карачаево-Черкесская Респ., Респ. Северная Осетия, Чеченская Республика, Ставропольский край, Республика Марий Эл, Республика Мордовия, Чувашская Республика, Кировская область, Пензенская область, Курганская область, Республика Алтай, Республика Тыва, Республика Хакасия, Алтайский край, Республика Бурятия, Республика Саха (Якутия), Забайкальский край, Камчатский край, Еврейская АО, Чукотский АО.

Третий кластер: г. Москва, Ленинградская область, Мурманская область, г. Санкт-Петербург, Тюменская область, Красноярский край, Сахалинская область.

Четвертый кластер: Республика Коми, Ненецкий автономный округ, Ханты-Мансийский АО, Ямало-Ненецкий АО.

Описать кластеры можно, например, вычислив средние значения признаков по каждой группе. Для этого можно воспользоваться функцией `aggregate()` со следующим набором аргументов:

`x` – фрейм, состоящий из признаков, по которым надо вычислить средние значения, в рассматриваемом примере это для налоговых поступлений в доходной части бюджета;

`by` – группирующий признак, в рассматриваемом примере это номер кластера;

`FUN` – вычисляемая характеристика, в рассматриваемом примере – среднее выборочное значение (`mean`).

Помимо описания самих кластеров, для удобства интерпретации результатов кластеризации полезно также рассчитать средние для каждого признака по всем объектам (регионам), это возможно при помощи базовой функции `colMeans()`, в качестве аргумента достаточно указать фрейм, состоящий из признаков, по которым надо вычислить средние значения.

Таким образом можно получить следующий результат:

Таблица 3 – Среднерегиональные значения относительных вкладов налоговых доходов в консолидированные бюджеты по видам доходов и кластерам

Кластер	Налоговые доходы, всего	Налог на прибыль	Налог на имущество организаций	НДФЛ
1	0,66	0,17	0,06	0,26
2	0,39	0,07	0,03	0,17
3	0,82	0,36	0,06	0,30
4	0,73	0,19	0,25	0,23
Все регионы	0,57	0,14	0,06	0,23

Результаты кластеризации можно прокомментировать следующим образом.

Характеристики кластера №1 более всего близки общим среднерегиональным.

Для кластера №2 характерны минимальные среднерегиональные доли налоговых



платежей в доходной части бюджета – это касается и суммарных налоговых доходов и доходов от каждого из трех рассматриваемых налогов в отдельности. В среднем для регионов этого кластера доля налоговых поступлений составляет 39% от доходной части бюджета, условно можно назвать эту группу «Дотационные регионы».

Кластер №3 отличается максимальными среди всех четырех кластеров вкладами в бюджет от совокупных налоговых платежей, налога на прибыль и НДФЛ. Эти показатели превышают среднерегиональные. Бюджеты регионов третьего кластера в среднем более чем на 80% наполняются за счёт налоговых платежей. Кроме того, суммарные налоговые платежи этих семи регионов составляют 37% от всех налоговых платежей регионов РФ. Условно эта группа может быть названа «Регионы-доноры».

Регионы кластера №4 отличаются максимальной долей платежей по налогу на имущество организаций в региональные бюджеты. Можно сделать вывод, что это регионы, где сосредоточены наиболее дорогие активы организаций на территории РФ.

### **Заключение**

Рассмотренный пример демонстрирует эффективность математико-статистических методов в решении задач описания особенностей регионов-налогоплательщиков. Рассмотрены два способа классификации – агрегирование по округам и выявление естественного расслоения. Выделены федеральные округа, для которых характерны минимальный и максимальный среднерегиональные относительные вклады налоговых поступлений в доходную часть бюджета. Кластерный анализ позволил выделить группу регионов-доноров и группу дотационных регионов. Следует подчеркнуть, что все необходимые расчёты – визуализация, вычисление групповых средних, кластеризация проведены при помощи небольшого набора функций языка R, простых по синтаксису, применение которых не требует глубоких знаний программирования. Преимущества использования в расчётах языка R заключаются в его высокой функциональности и финансовой доступности, выгодно отличающих его от специализированного программного обеспечения и MS Excel, а также – в синтаксисе команд, простота которого, по нашему мнению, выгодно отличает этот язык от Python.

### **Библиография**

1. Жиликов Д.И. Ретроспективный анализ налоговых доходов федерального бюджета /Жиликов Д.И., Новосельский С.О., Плахутин Ю.В., Петрушина О.В. // Экономические науки-2023-№2 (219) URL: [https://ecsn.ru/wp-content/uploads/202302\\_173.pdf](https://ecsn.ru/wp-content/uploads/202302_173.pdf) (дата обращения: 20.04.2024).
2. Васильченко А.Д. Налоговые поступления в бюджетную систему России: статистическая оценка и меры по мобилизации // Налоги и налогообложение. 2019. № 5. С.45-57. DOI: 10.7256/2454-065X.2019.5.30101 URL: [https://e-notabene.ru/ttmag/article\\_30101.html](https://e-notabene.ru/ttmag/article_30101.html)
3. Костина А.А. Статистический анализ структуры и динамики налоговых поступлений Российской Федерации. // Вестник магистратуры.-2017.-№6-1 (69)
4. Деденева Д.Б.-Анализ налоговых поступлений в бюджетную систему России. //Электронный научный журнал «Вектор экономики»-2022-№4 URL: <http://www.vectoreconomy.ru/images/publications/2022/4/taxes/Dedeneva.pdf> (дата обращения: 20.04.2024).
5. Селюков М.В.-Анализ налоговых доходов в субъектах Российской Федерации. // Сибирская финансовая школа – 2023-№1 DOI: 10.34020/1993-4386-2023-1-35-43 (дата обращения: 20.04.2024).

6. Математическая статистика. Практикум : учебное пособие / Т.Г. Апалькова, В.И. Глебов, С.А. Зададаев [и др.]. — Москва : ИНФРА-М, 2023. — 254 с. — (Высшее образование). — DOI 10.12737/1896790. — ISBN 978-5-16-017913-1 — Текст: электронный.-URL: <https://znanium.com/catalog/product/1896790> (дата обращения: 25.07.2023). — Режим доступа: по подписке.
7. Hadley Wickham. R for Data Science, 2nd Edition / Hadley Wickham, Mine Çetinkaya-Rundel, Garrett Grolemund// -Publisher(s): O'Reilly Media, Inc., 2023
8. Маркова, С. В., Анализ данных на языке R. : учебник и практикум / С. В. Маркова. — Москва : КноРус, 2023. — 216 с. — ISBN 978-5-406-10865-9. — URL: <https://book.ru/book/948838> (дата обращения: 17.03.2024). — Текст : электронный
9. Мастицкий С.Э. Визуализация данных с помощью ggplot2. – М.:ДМК Пресс, 2017. – 222 с.
10. Платонов В.В. Визуализация больших данных в экономических науках в условиях информационного общества // Вопросы инновационной экономики. 2020. № 4. [Электронный ресурс]. URL: Визуализация больших данных в экономических науках в условиях информационного общества-Платонов В.В. / Вопросы инновационной экономики / № 4, 2020-Первое экономическое издательство-(1economic.ru) (дата обращения: 12.12.2023)
11. Шипунов А.Б. Анализ данных с R (II). Шипунов А.Б. , Коробейников А. И., Е. М. Балдин Е. М. Электронное издание. URL: <https://inp.nsk.su/~baldin/DataAnalysis/R/R-07-datamining.pdf?ysclid=lvic8543su725025982> (дата обращения: 25.04.2024)
12. Дубров А.М. Многомерные статистические методы: учебник. Дубров А.М., Мхитарян В.С., Трошин Л.И.— М.: Финансы и статистика, 2011

## Результаты процедуры рецензирования статьи

*В связи с политикой двойного слепого рецензирования личность рецензента не раскрывается.*

*Со списком рецензентов издательства можно ознакомиться [здесь](#).*

В рецензируемой статье рассматриваются вопросы статистического анализа дифференциации регионов РФ по налоговым доходам консолидированных бюджетов средствами языка программирования R.

Методология исследования базируется на применении методов математического и статистического моделирования, методов кластерного анализа и визуализации.

Актуальность работы авторы справедливо связывают с тем, что методы математической статистики в сочетании с современными техническими средствами представляют собой мощный инструмент, позволяющий осветить процессы сбора, начисления, уплаты налогов и выявить проблемные и позитивные аспекты этих процессов.

Научная новизна работы, по мнению рецензента состоит в обосновании возможностей и эффективности применения средств визуализации, описательной статистики, агрегирования и кластеризации языка программирования с открытым кодом R в налоговом анализе. В статье выделены федеральные округа, для которых характерны минимальный и максимальный среднерегиональные относительные вклады налоговых поступлений в доходную часть бюджета, с применением методов кластерного анализа выделены группы регионов-доноров и дотационных регионов.

Структурно в статье выделены следующие разделы: Введение, Данные для моделирования, Агрегирование данных, Кластеризация, Заключение, Библиография.

В публикации приведен фрагмент набора исходных данных, отражены среднерегиональные значения доходов консолидированных бюджетов по видам доходов и федеральным округам, а также среднерегиональные значения относительных вкладов

налоговых доходов в консолидированные бюджеты по видам доходов и федеральным округам. Текст статьи сопровождается иллюстрациями, выполненными с помощью средств визуализации языка программирования R. Авторами рассмотрена возможность выявления естественного расслоения регионов по совокупности признаков, характеризующих долю налоговых платежей в доходной части бюджетов. Отражена процедура агломеративного кластерного анализа при помощи кода языка R, показаны особенности вычисления расстояний между объектами, объединения в кластеры по методу Варда, построения дендрограммы и выделения кластеров на ней, а также создания дендрограммы в виде круга. В результате исследования выделены четыре кластера, для каждого из которых получены среднерегionalные значения относительных вкладов налоговых доходов в консолидированные бюджеты по видам доходов, даны характеристики каждому кластеру. По мнению авторов статьи, преимущества использования в расчётах языка R заключаются в его высокой функциональности и финансовой доступности, выгодно отличающих его от специализированного программного обеспечения и MS Excel, а также в синтаксисе команд, простота которого, выгодно отличает этот язык от Python.

Библиографический список включает 12 источников – научные публикации отечественных и зарубежных авторов по рассматриваемой теме на русском и английском языках. В тексте публикации имеются адресные отсылки к списку литературы, подтверждающие наличие апелляции к оппонентам.

Тема статьи актуальна, материал отражает результаты проведенного авторами исследования, содержит элементы приращения научного знания, соответствует тематике журнала «Налоги и налогообложение», может вызвать интерес у читателей и рекомендуется к публикации.