Национальный исследовательский университет «Высшая школа экономики»

Факультет компьютерных наук

Образовательная программа: Программная инженерия

Отчет по Домашнему заданию №1

«Предварительный, корреляционный и кластерный анализы»

по майнору «Классификация статистических данных»

«Статистическое исследование зависимости здоровья и социально-экономического благополучия детей в возрасте от 3 до 15 лет в субъектах Российской Федерации от систематических занятий физической культурой и спортом за 2023 год»

Работу выполнил:

студент 2 курса

Артемьев Александр Вячеславович

Преподаватель:

Арефьева Валерия Александровна

Содержание

Оглавление

[1. Введение 4](#_Toc163287909)

[1.1 Актуальность темы исследования 4](#_Toc163287910)

[1.2 Цели исследования 4](#_Toc163287911)

[1.3 Задачи исследования 4](#_Toc163287912)

[1.4 Гипотезы 4](#_Toc163287913)

[2. Используемые показатели для анализа 5](#_Toc163287914)

[2.1 Выбор данных 5](#_Toc163287915)

[2.2 Список используемых показателей для анализа 5](#_Toc163287916)

[2.3 Описание показателей, выбранных для исследования 5](#_Toc163287917)

[3. Предварительный анализ данных 6](#_Toc163287918)

[3.1 Графическое представление исходных данных 6](#_Toc163287919)

[**3.1.1** **Точечное распределение** 6](#_Toc163287920)

[**3.1.2** **Построение листовых диаграмм** 9](#_Toc163287921)

[**3.1.3** **Box plot** 12](#_Toc163287922)

[3.2 Характеристики СВ 14](#_Toc163287923)

[3.3 Z – преобразование 15](#_Toc163287924)

[3.4 Расчет IQR 18](#_Toc163287925)

[3.5 Заключение 18](#_Toc163287926)

[4. Корреляционный анализ 19](#_Toc163287927)

[4.1 Построение полей корреляции 19](#_Toc163287928)

[4.2 Матрица парных коэффициентов 23](#_Toc163287929)

[4.3 Построение полей корреляции без выбросов 25](#_Toc163287930)

[4.4 Матрица парных коэффициентов без выбросов 27](#_Toc163287931)

[4.5 Выводы о исследуемых связях 28](#_Toc163287932)

[5. Кластерный анализ 29](#_Toc163287933)

[5.1 Метод ближнего соседа 29](#_Toc163287934)

[5.2 Метод дальнего соседа 33](#_Toc163287935)

[5.3 Метод межгрупповой связи 37](#_Toc163287936)

[5.4 Метод медианной связи 41](#_Toc163287937)

[5.5 Метод центр тяжести 45](#_Toc163287938)

[5.6 Использование метода к-средних для классификации объектов 49](#_Toc163287939)

[6. Список литературы 52](#_Toc163287940)

[7. Используемые информационные источники 52](#_Toc163287941)

[8. Приложения 52](#_Toc163287942)

# **Введение**

## **Актуальность темы исследования**

В наше время наблюдается широкое обсуждение и активная пропаганда занятий спортом среди детей со стороны родителей, друзей, рекламных кампаний и государственных программ, поскольку современные дети все чаще придерживаются сидячего образа жизни и проявляют недостаточную физическую активность. Стремительный темп технологического развития и широкое распространение гаджетов приводят к тому, что многие дети проводят большую часть времени перед экранами устройств, уменьшая время на активные игры и занятия на свежем воздухе.

При этом существует множество утверждений о том, что регулярные занятия физической культурой и спортом способствуют улучшению здоровья детей и формированию их активного образа жизни. Однако научные данные о влиянии спорта на здоровье детей требуют дополнительного исследования и подтверждения. Поэтому проведение исследования, направленного на проверку фактического влияния занятий спортом на здоровье детей, является крайне актуальным и имеет важное значение для разработки обоснованных рекомендаций и стратегий по поддержке здоровья и физической активности детей.

## **Цели исследования**

* Определить, на какой тип здоровья детей в возрасте от 3 до 15 лет в большей степени оказывает влияние занятие спортом, а именно, определить, влияет ли систематическая физическая активность на физическое здоровье, психическое благополучие, общее состояние здоровья.
* Исследовать влияние систематических занятий физической культурой и спортом на уровень заболеваемости детей злокачественными новообразованиями.
* Проанализировать взаимосвязь между уровнем бедности и численностью детей, занимающихся физической культурой и спортом

## **Задачи исследования**

* ***Предварительный анализ данных***
* ***Корреляционный анализ***
* ***Кластерный анализ***

## **Гипотезы**

* Уровень физической активности у детей прямо коррелирует с улучшением общего состояния здоровья, что проявляется в увеличении количества детей, отнесенных к I и II группам здоровья.
* Дети, занимающиеся физической активностью, имеют меньший уровень психических расстройств и расстройств поведения по сравнению с теми, кто ведет сидячий образ жизни, то есть обратная зависимость
* Систематические занятия физической культурой и спортом у детей в возрасте от 3 до 15 лет снижают риск заболеваний злокачественными новообразованиями, обратная зависимость
* Увеличение численности детей, систематически занимающихся физической культурой и спортом, приводит к снижению уровня бедности в регионе.

# Используемые показатели для анализа

## Выбор данных

Дата сет был составлен, используя данные с официального сайта Росстата за 2023 год. Проводить анализ с наблюдениями за 2024 год не имеет смысла, так как он только начался на момент написания данной работы. Был выбран в качестве источника сайт Росстата [1], поскольку такие данные имеют официальный статус и являются официальной статистикой страны, а также из-за широкого обхвата и разнообразия данных. Исходные данные представлены в Приложении 1.

## Список используемых показателей для анализа

В данном исследовании используются 5 переменных:

1. Численность детей в возрасте 3-15 лет, систематически занимающихся физической культурой и спортом, человек (независимая переменная)
2. Заболеваемость детей злокачественными новообразованиями, человек (результативный)
3. Заболеваемость детей психическими расстройствами и расстройствами поведения, человек (результативный)
4. Доля детей с I и II группой здоровья, % (результативный)
5. Уровень бедности, % (результативный)

## Описание показателей, выбранных для исследования

1. Выбранный показатель в качестве независимой переменной – *Численность детей в возрасте 3-15 лет, систематически занимающихся физической культурой и спортом*.

* Показатель характеризует количество детей, занимающихся спортом, указанного возраста, в конкретном субъекте.
* Единица измерения – количество человек.
* Количественные дискретные данные

1. Выбранный показатель в качестве результативного признака – *Заболеваемость детей злокачественными новообразованиями*.

* Показатель характеризует количество документально зарегистрированных случаев у детей злокачественного образования в данном субъекте РФ.
* Единица измерения - количество человек.
* Количественные дискретные данные.

1. Выбранный показатель в качестве результативного признака – *Заболеваемость детей психическими расстройствами и расстройствами поведения*.

* Показатель характеризует количество документально зарегистрированных случаев психических расстройств у детей в данном субъекте РФ (состоит под наблюдением на конец отчетного года).
* Единица измерения - количество человек.
* Количественные дискретные данные.

1. Выбранный показатель в качестве результативного признака – *Доля детей с I и II группой здоровья*.

* Показатель характеризует процент детей в регионе попадающих в I группу здоровья (здоровые дети с нормальным физическим и психическим развитием) и II группу здоровья(входят дети, у которых нет хронических заболеваний, но есть некоторые функциональные нарушения)
* Единица измерения – процент населения детей.
* Количественные непрерывные данные.

1. Выбранный показатель в качестве результативного признака – *Уровень бедности*.

* Показатель характеризует уровень материального благополучия населения в конкретном регионе или субъекте.
* Единица измерения – процентное отношение количества людей, живущих за чертой бедности, к общему численности населения в данном регионе.
* Количественные непрерывные данные

# **Предварительный анализ данных**

Для начала стоит отметить, что найденная выборка полная, то есть нет пустых/некорректных значений данных

## **Графическое представление исходных данных**

### **Точечное распределение**

Сначала построим точечные диаграммы, по каждой из 5 переменных, построение будем осуществлять с помощью Excel.

Рисунок 1. Точечная диаграмма 1ого признака

В целом можем заметить, что больше всего наблюдений признака находится до 400 000, а также можем сразу отметить 2 явных выброса для Москвы и Московской области, но пока не понятно являются ли выбросами точки в диапазоне 600 000 – 800 000.

Структур очень похожа на диаграмму для 1ого признака, видны явные выбросы в виде Москвы и Московской области.

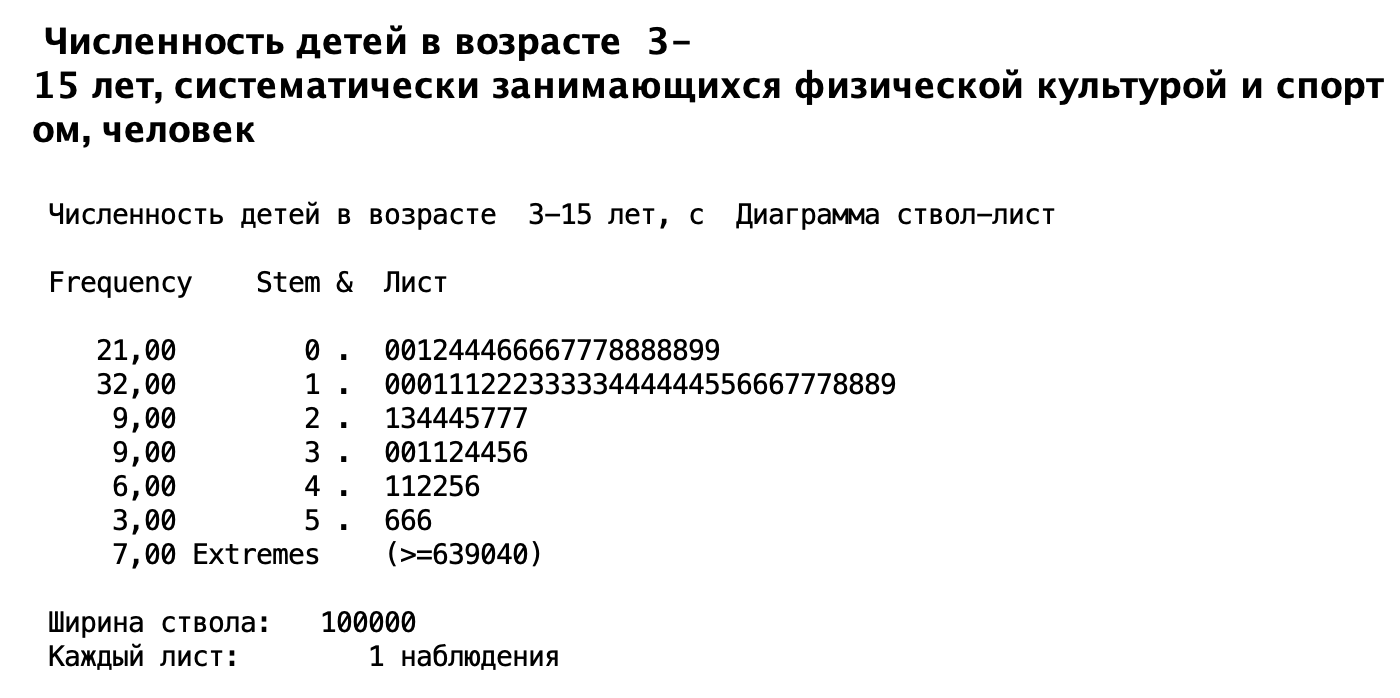
Для 3ого признака, точечная диаграмма кардинально изменилась по сравнению с предыдущими, данные более равномерно распределены внутри выборки, и предположительно как минимум 4 аномальных значения > 8000. Из точечной диаграммы в республике Дагестан больше всего встречается детей с психическими расстройствами 11 200, а в Калужской области и Ненецком автономном округе меньше всего 79 и 61 соответственно.

Не считая выбросов, значения выборки распределены равномерно и изменяются от 60 до 90. Причем для процентного показател, очевидно, данные не могут быть больше 100 %, данные 2 выброса Тюменская и Архангельская области которые считают свой показатель как сумма процентных показателей субъектов внутри них, поэтому их необходимо просто исключить, либо находить долю детей для каждой области. Также стоит отметить явный минимум – Ненецкий автономный округ 51,4 %.

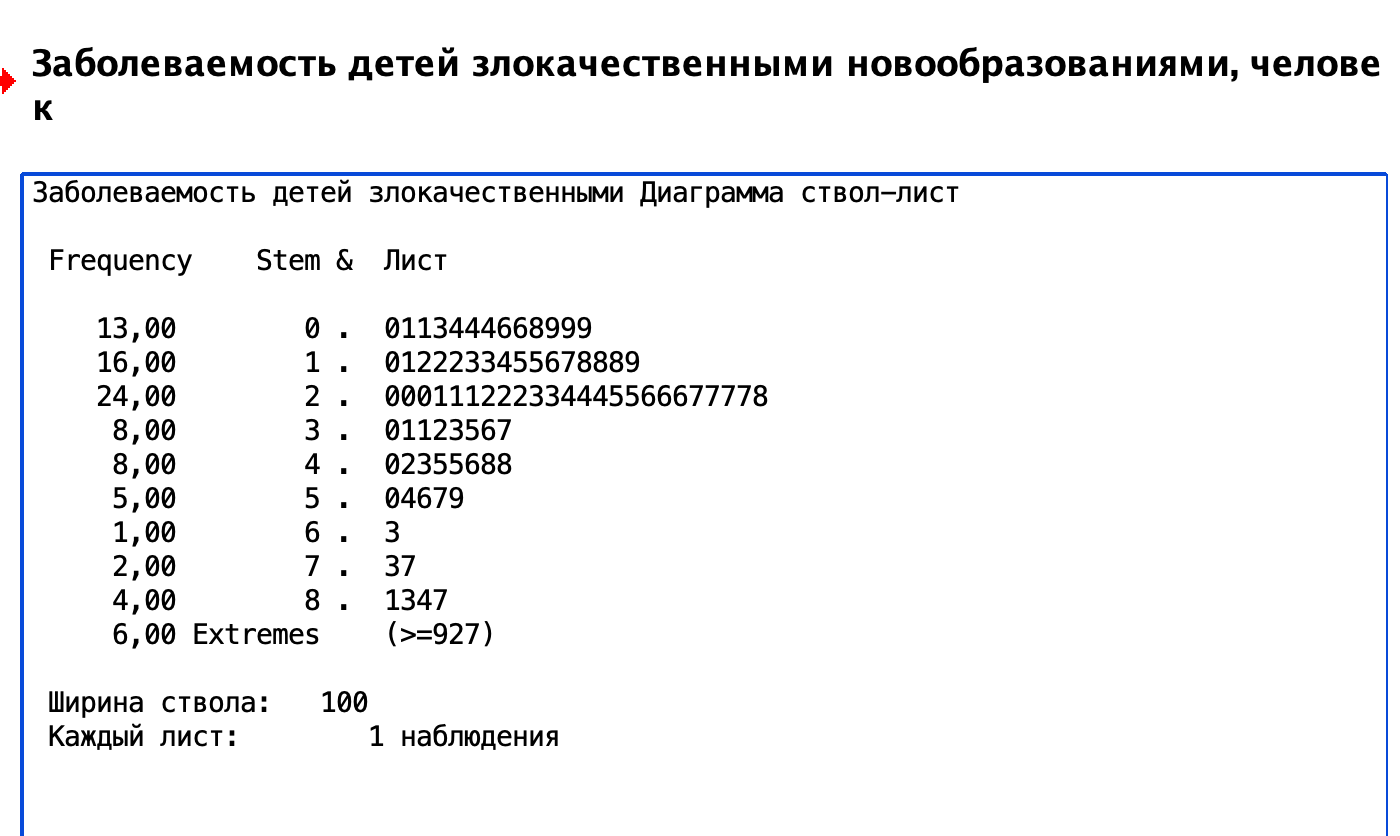
Данные располагаются в диапазоне от 5 до 20 %, предположительно остальные значения могут быть аномальными. Причем к центру то есть к 10-15 % значение признака встречается чаще всего, ассоциативно похоже на нормальное распределение.

### **Построение листовых диаграмм**

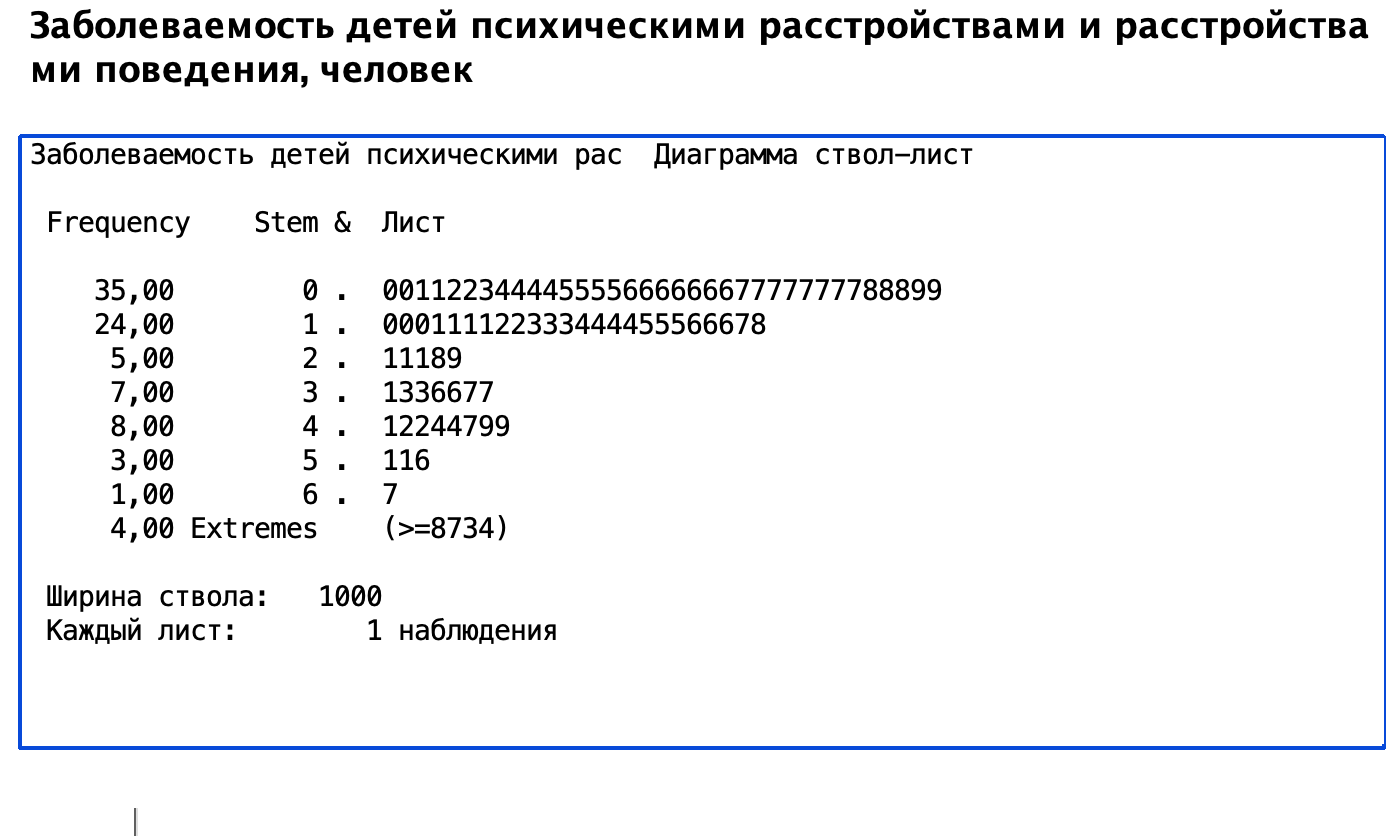
Stemplot построим с помощью spss и построенные диаграммы будут храниться, в файле Stem.spv. Будем строить диаграммы в таком же порядке, как расположены наши признаки:



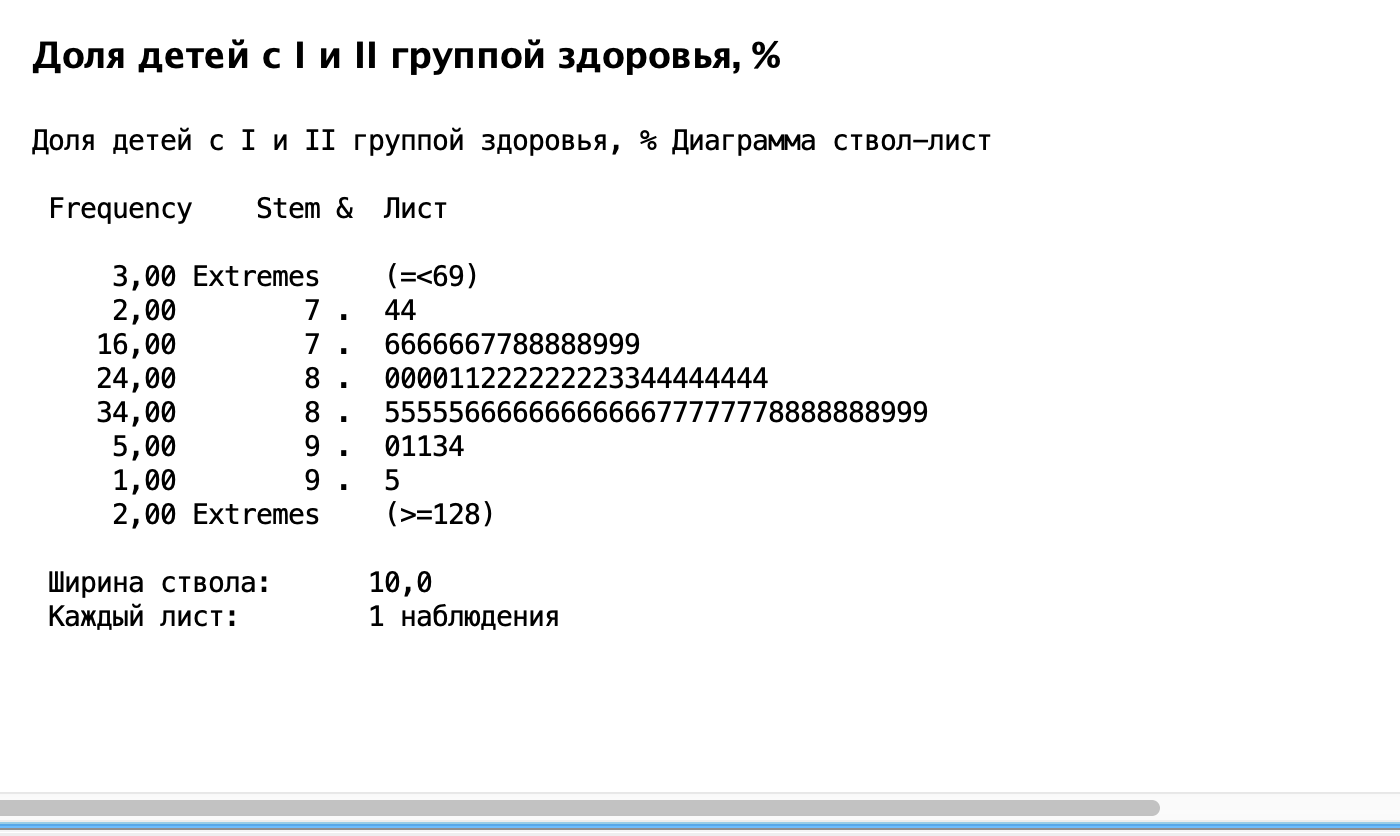
Можем отметить 2 наблюдения, больше всего частота у первых двух интервалов, что также видно из соответствующей точечной диаграммы и встречается 7 выбросов, что уже следует не так тривиально из точечной диаграммы



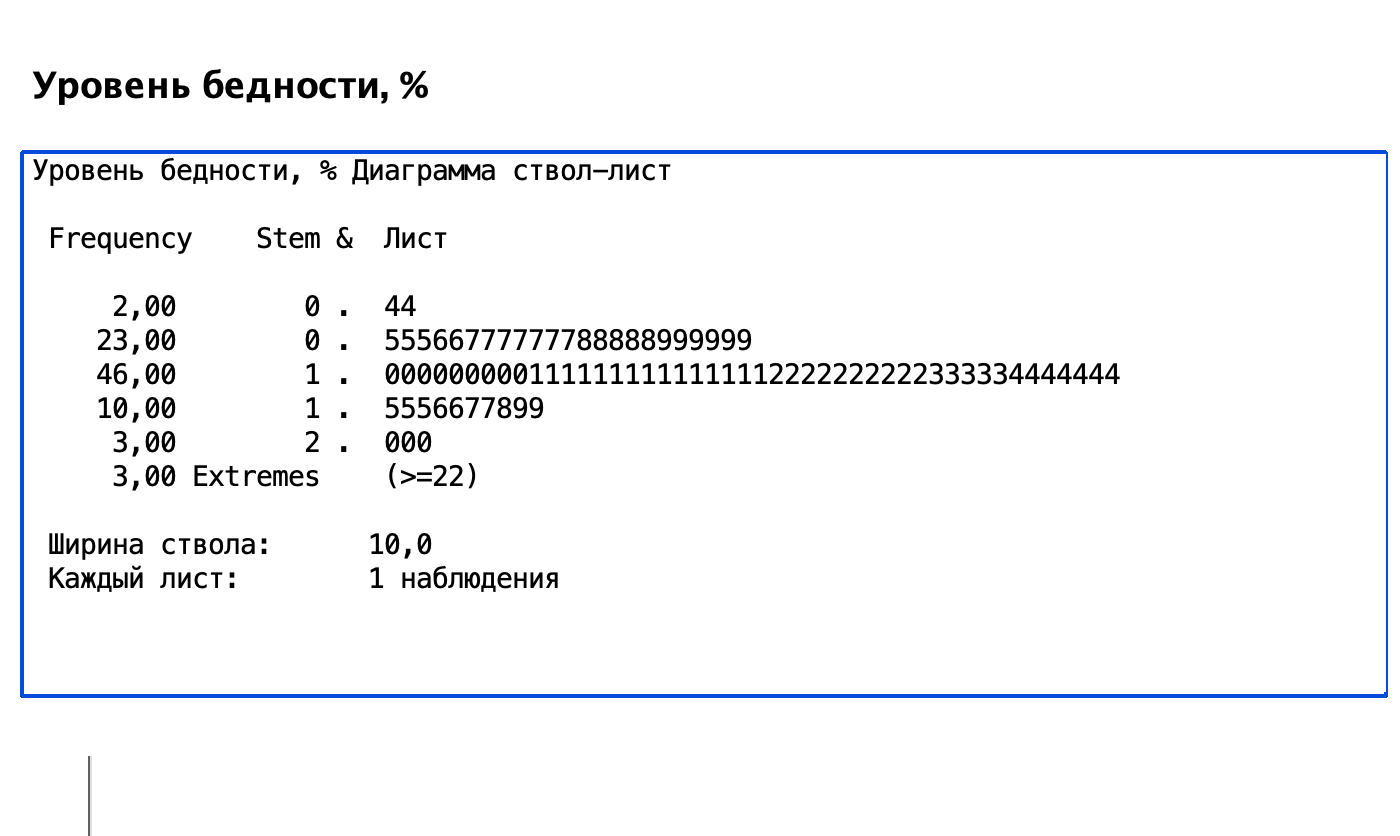
Больше блоков по сравнению с предыдущим признаком, то есть различных stem из за чего частота распределена равномернее, хотя существует явный интервал с модой, куда попадают 24 наблюдения



Явный перевес для не больших значений количества заболеваний – 59, что почти ¾ от размера выборки. 4 аномальных значения, которые были определены при построении точечной диаграммы действительно являются выбросами.

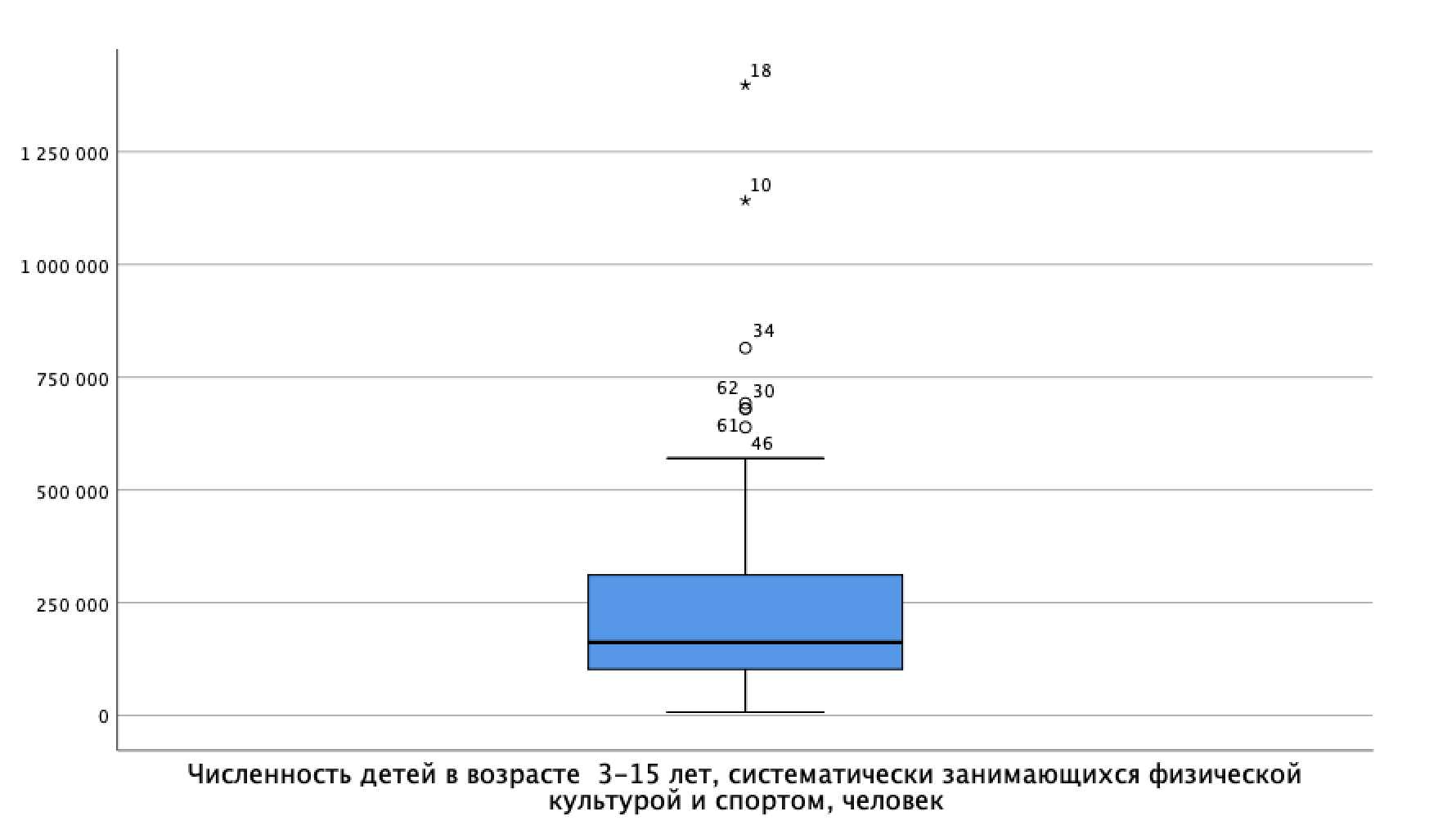


Первый признак, который имеет выбросы и меньше и больше интерквартильного размаха. Распределение СВ, судя по листовой диаграмме, близко к логнормальному.

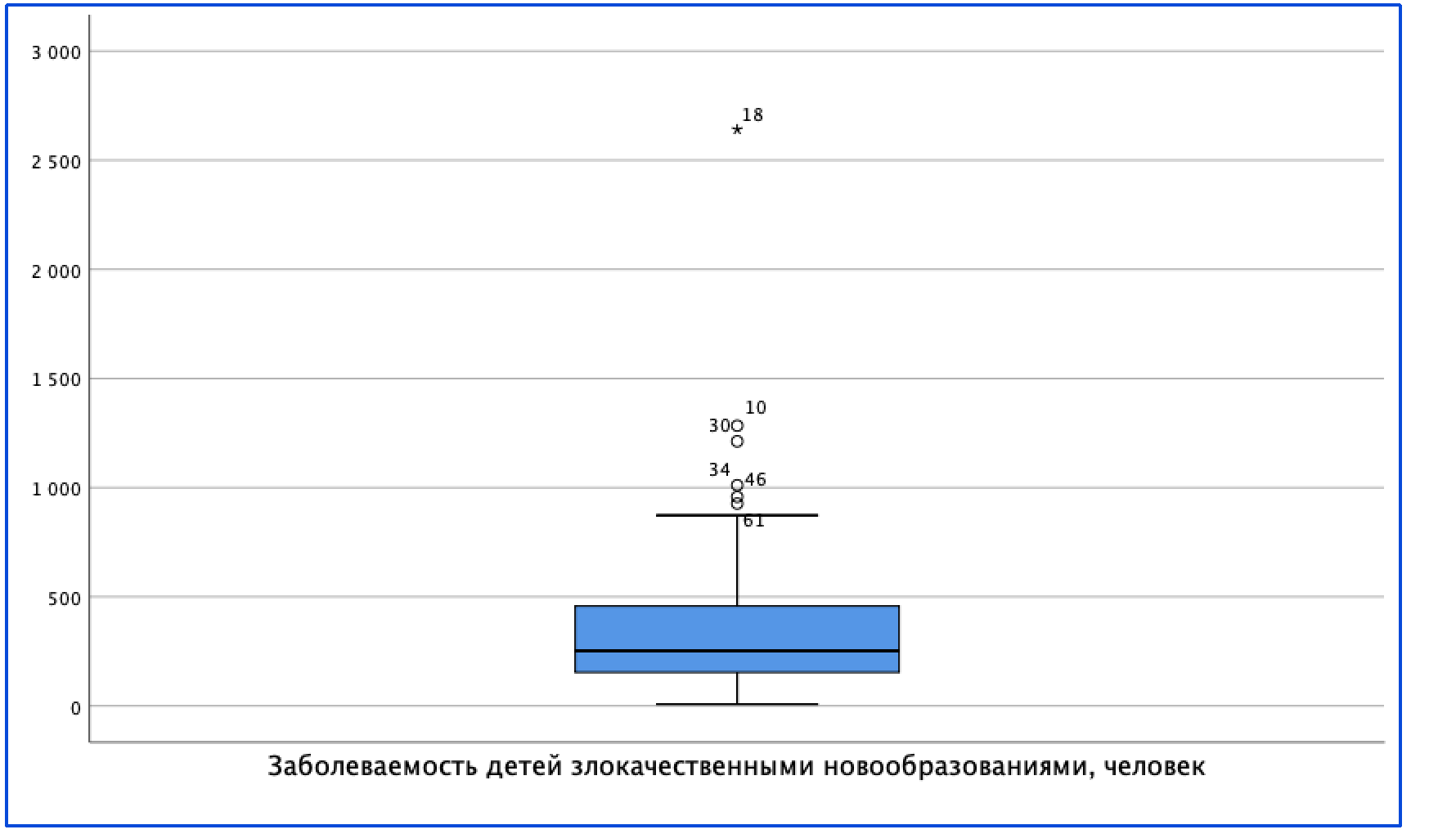


Исследуя построенную диаграмму, распределение действительно близко к нормальному, и все значения сконцентрированы в близи середины размаха вариации. Также по точечной диаграмме казалось, что обе точки 20,9 и 30,5 будут аномальными, но в итоге 20,9 попадает в интерквартильный размах.

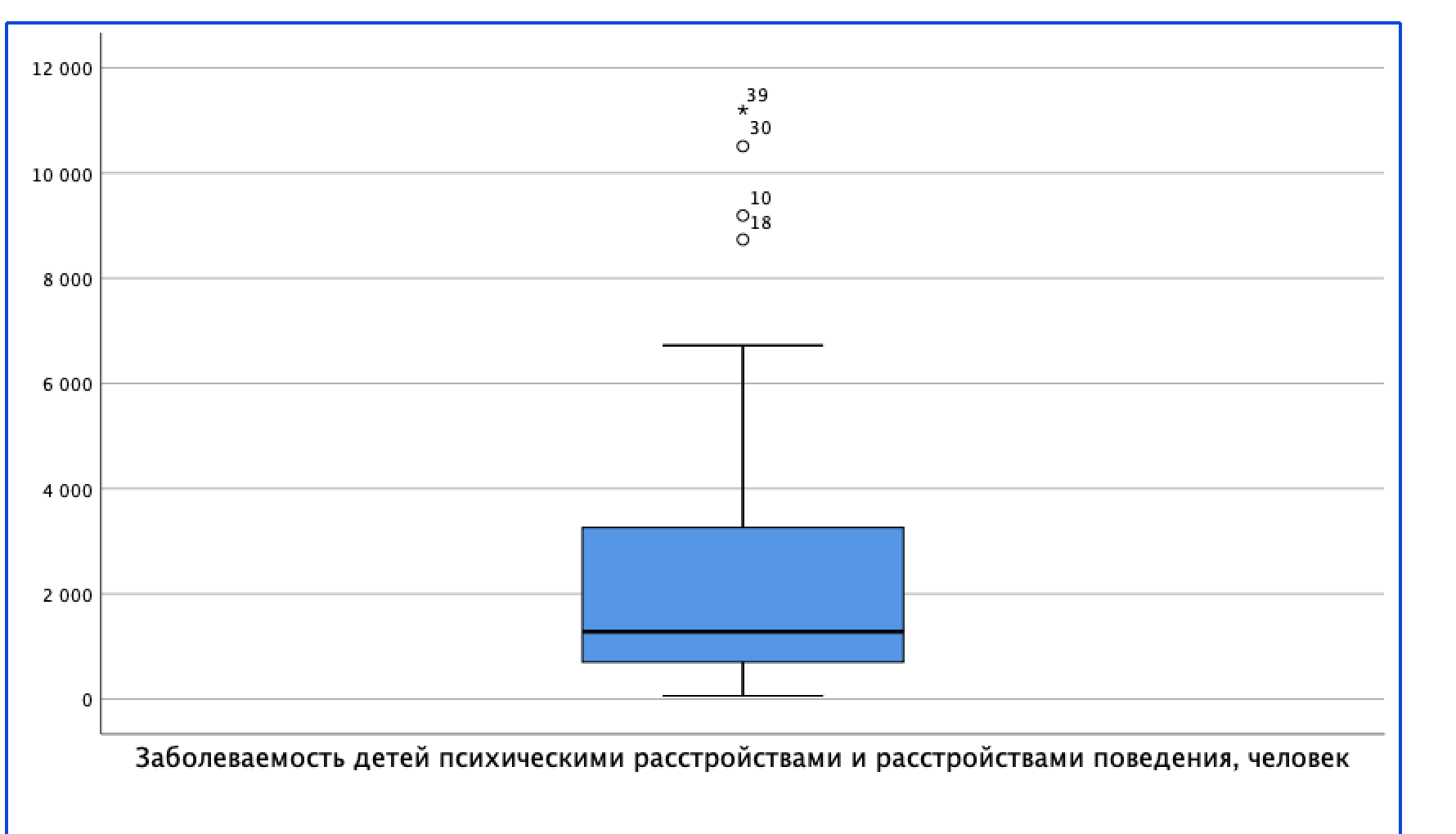
### **Box plot**

Теперь уже проанализируем только выбросы, исследуя ящички с усами. 

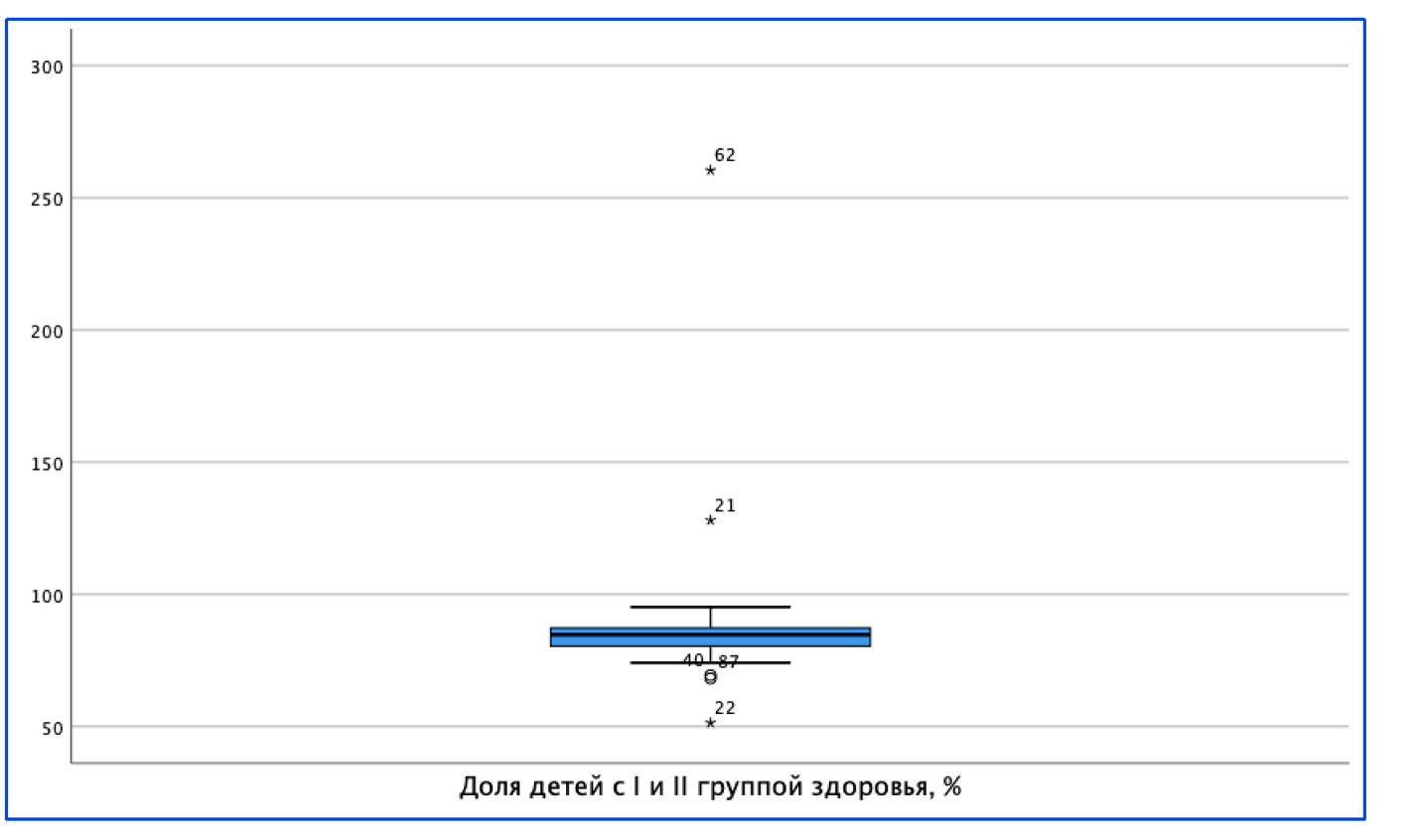
Для 1ого признака мы получаем 7 аномальных значений, 2 из которых предсказуемы – Москва и Московская область и еще 5 тоже крупные субъекты РФ, поскольку показатель показывает количество абсолютное, а не относительное, то для крупнейших городов этот показатель может быть аномально большим.



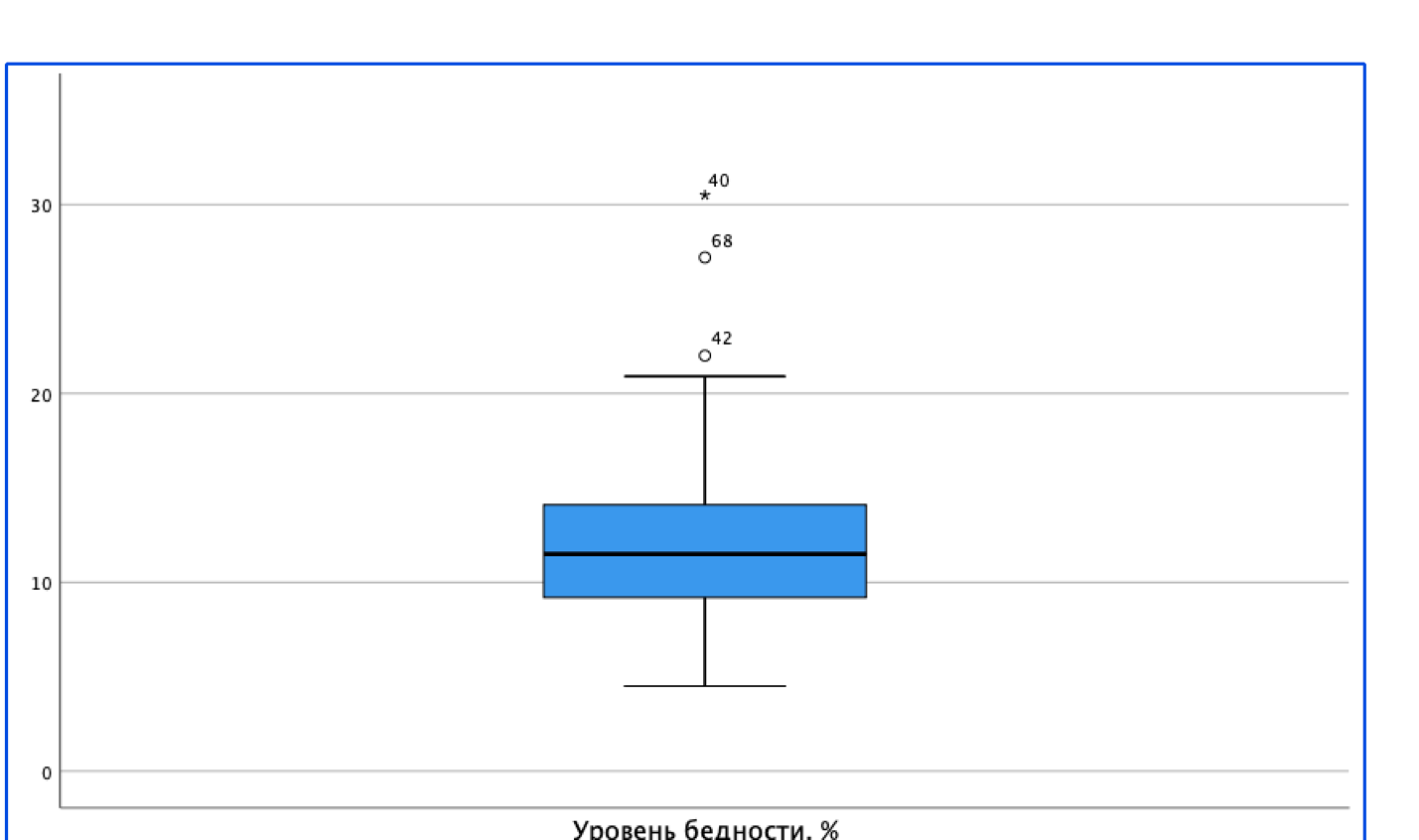
Для данного признака, получаем 6 аномальных значений причем большинство из них такие же – крупнейшие субъекты РФ.



Для данного показателя, республика Дагестан занимает первое место, что удивительно, учитывая соотношение население Москвы, Санкт- Петербурга и Дагестана.



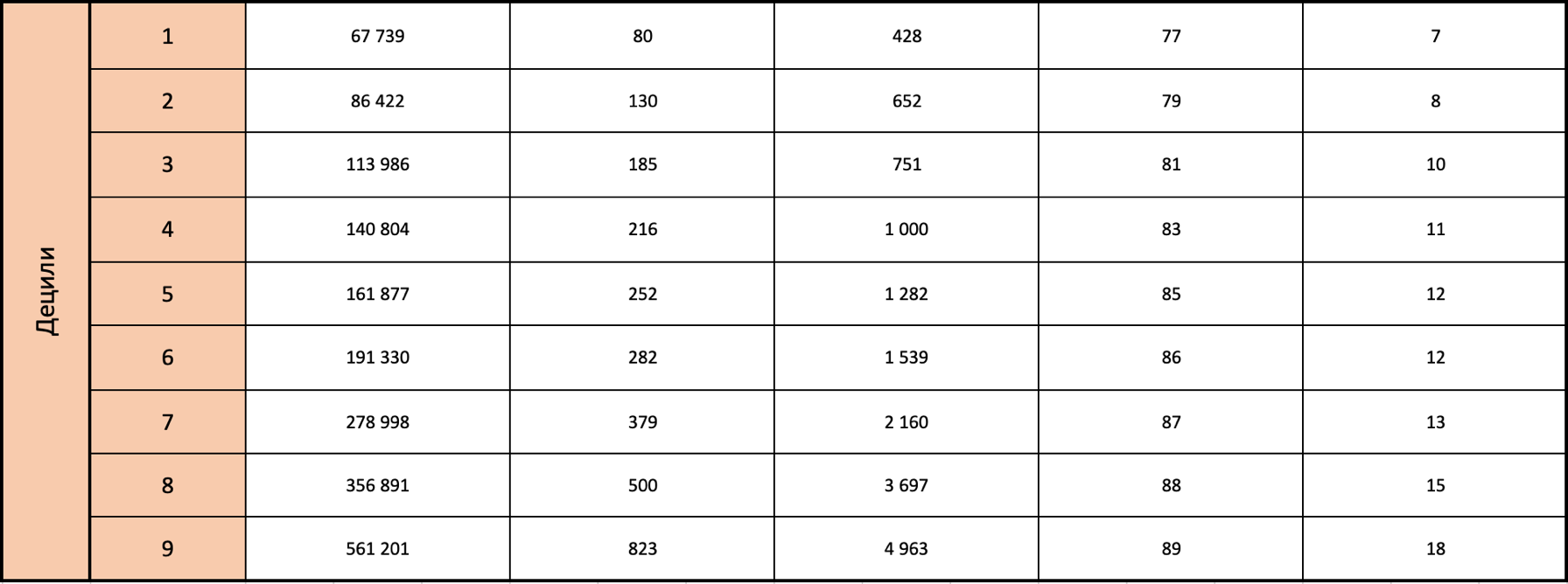
5 выбросов, 2 из которых – некорректные значения в таблице, которые необходимо будет удалить при следующих исследованиях, 2 автономных округа и Ингушетия, вероятно малочисленное население и слабое развитие влияет на значение данного признака, причем все остальные города имеют значение наблюдений довольно схожие из одного интервала от 74 до 95. Также возможно суровые климатические условия влияют на низкую долю здоровых детей.



3 выброса, которые обуславливаются в первую очередь слабой инфраструктурой, высоким уровнем преступности и развитием данных регионов.

## Характеристики СВ

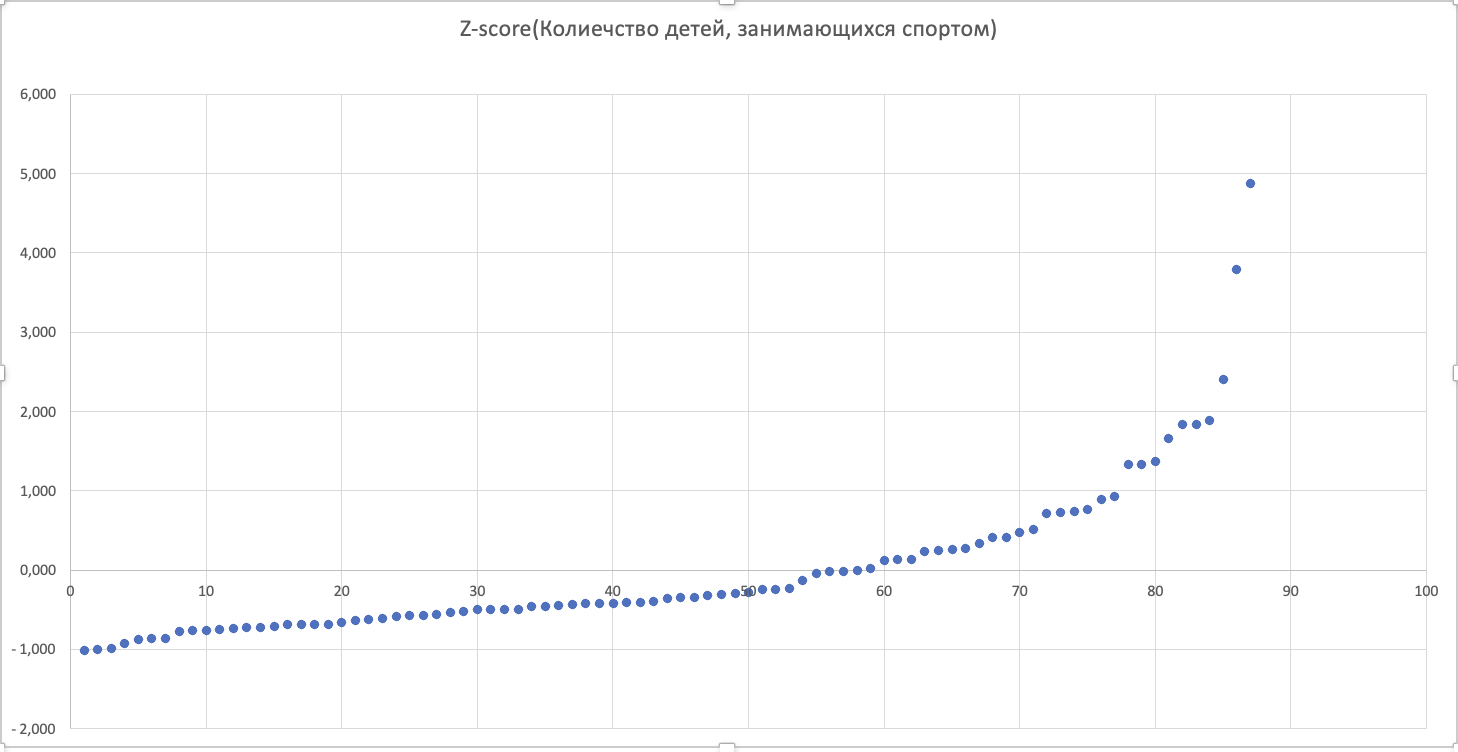
Данные характеристики рассчитаны с помощью формул в Excel, все формулы доступны в файле, так же многие характеристики посчитаны и в файле SPSS. Проанализируем полученные результаты: коэффициент вариации для 1-3 признаков очень большой: он в 3 раза больше допустимого значения для нормальных данных( 33 %). Поэтому можно сделать вывод о неоднородности информации и исключении самых больших значений. Это коррелирует с тем что для этих признаков мы нашли больше выбросов чем для 4ого и 5ого. Анализируя полученные значения моды, можно сделать вывод о уникальности реализаций 1 признака, и что у остальных признаков есть повторяющиеся значения.



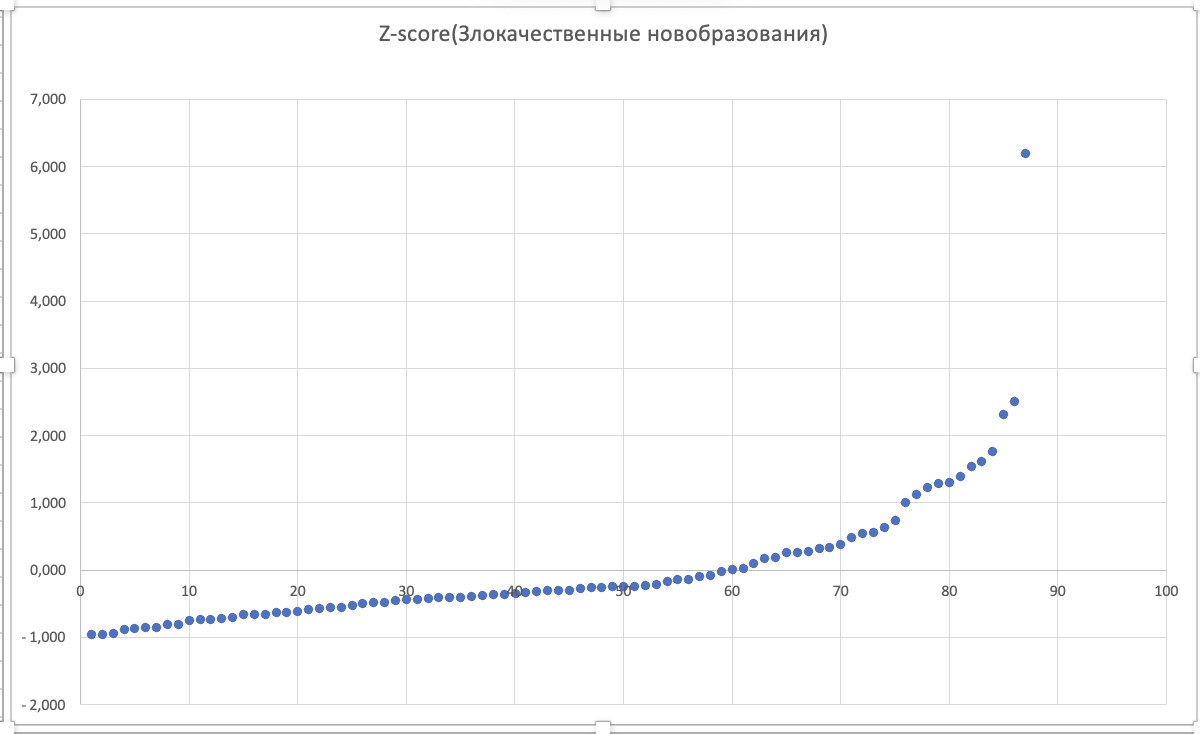
Исследуя квартили и децили и зная размах вариации, можно вывести заключения о распределение данных. Так, например для 2ого признака, 75% наблюдений находятся <458 а размах вариации 2633, доказывая наличие выбросов. Аналогичные заключения для 1ого и 3ого признака. Для доля детей с 1ой и 2ой групп здоровья, анализируя, что соседние децили почти не отличаются друг от друга, заключаем близость к равномерному распределению, что также присутствует у уровня бедности, но более большие переходы между децилями.

## Z – преобразование

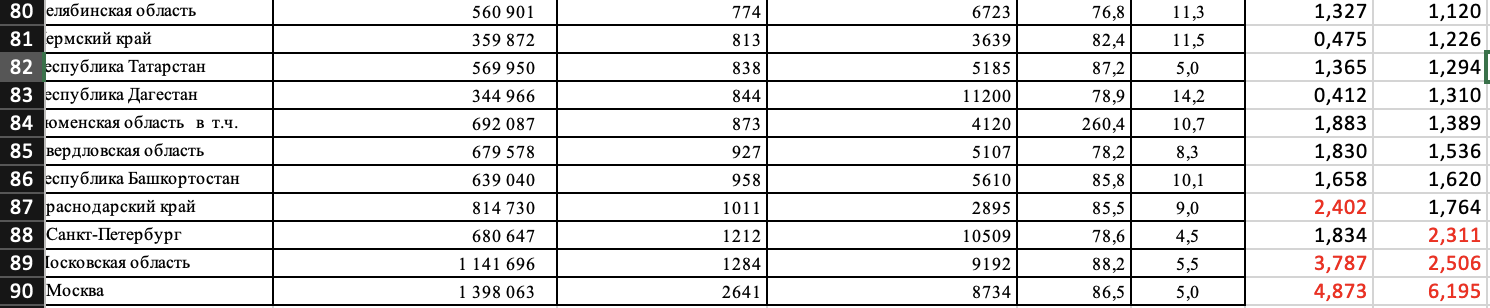
Для нахождения выбросов с помощью z-score, рассчитаем z-score для каждого наблюдения и построим график отсортированных z-score( Приложение 5,Расчеты в excel). Затем визуально по графику определяем резкий скачок/рост, откуда и начнут идти выбросы, либо по отсортированным данным это будет резкое изменение z – score.

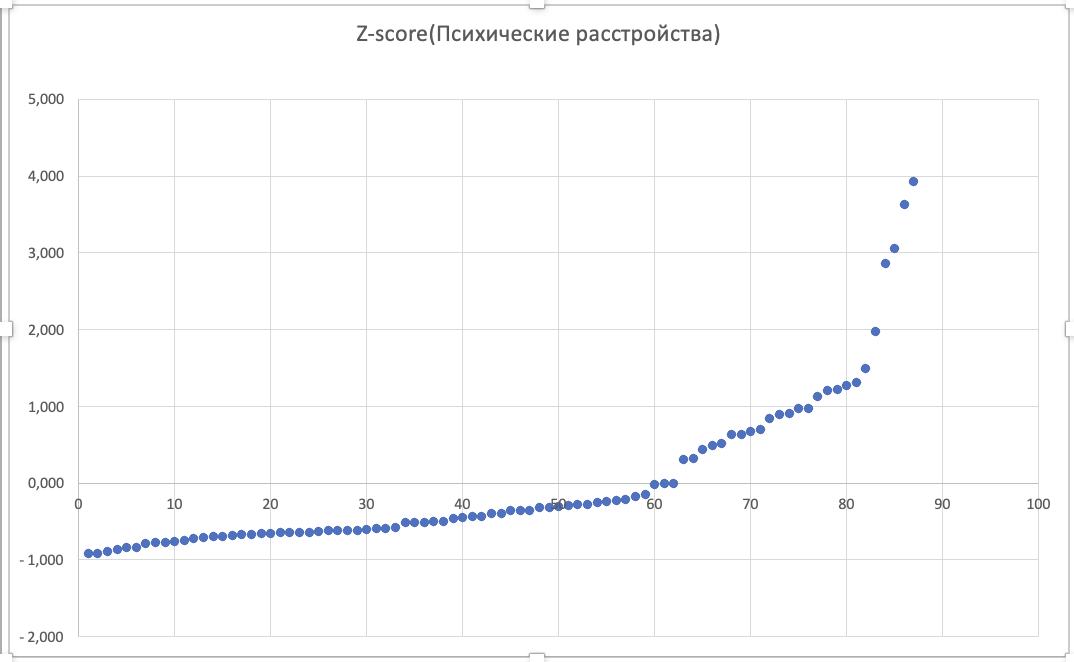


Существенный скачок после 84, то есть 3 аномальных значения.



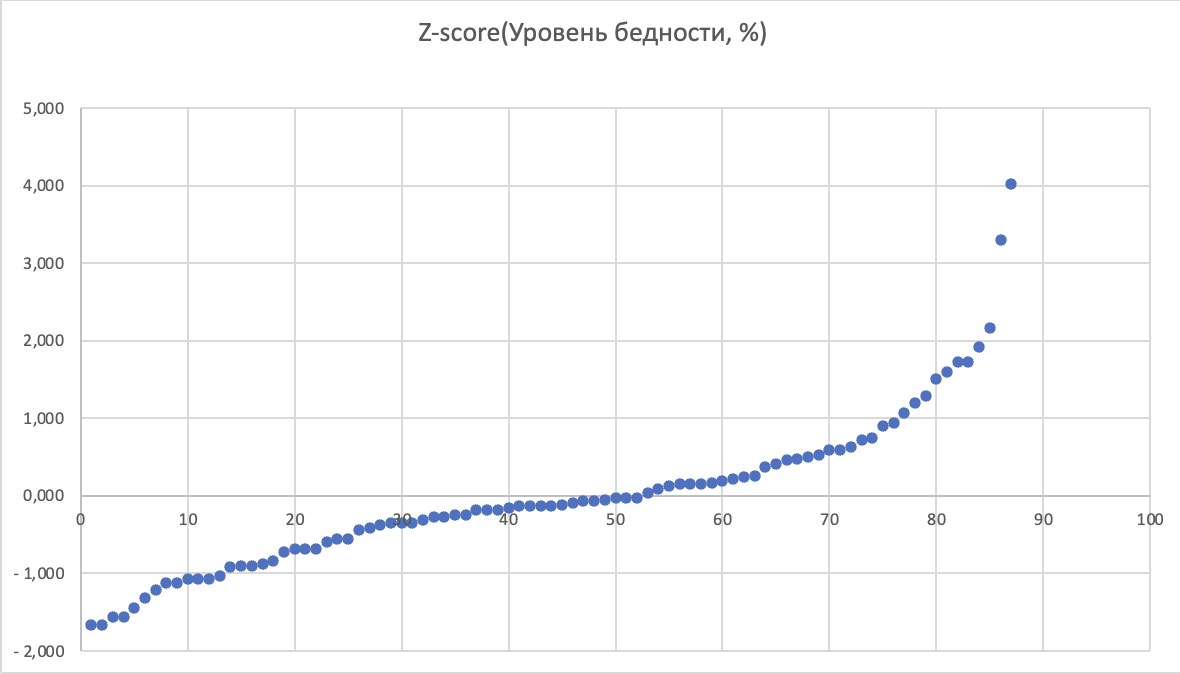
Похожая ситуация на первый случай, аналогично отнесем 85-87 к аномальным значениям где начинаются существенные изменения как на графике так и на данных.





Для данного признака, 1ый скачок виден уже после 60ого элемента, но если смотреть в целом на кривую она довольно плавно растет, существенно изгибаясь после 81 элемента



Для признак доли детей в 1ой и 2ой группе, у нас выделяется только 2 числа значение которых некорректны, то есть они,очевидно, аномальны. Соответственно выбросы будут 86 и 87 элементы.

Ярко выражены 2 выброса 86 и 87, как для прошлого признака, и на графике и на данных; все выбросы помечены красным цветом в файле excel . Результаты анализа z-score в Приложении 7

## Расчет IQR

Зная квартили найдем IQR для каждого из 5 признаков(Приложение 6).

После этого определим жесткие и мягкие выбросы, жесткие(Приложение 2):

Мягкие(Приложение 3):

Для этого построим вспомогательную таблицу в excel и определим аномальные значения.

Также найдем выбросы по правилу 3 сигм(Приложение 4).

## Заключение

Придём к логическому завершению проделанной работы 1 блока. Цель предварительного анализа данных была в том, чтобы нормализовать и стандартизовать данные. Давайте пройдемся по каждой выборке и сделаем вывод:

1. *Численность детей в возрасте 3-15 лет, систематически занимающихся физической культурой и спортом*.

При первичном анализе имело 7 выбросов – крупные регионы РФ с большим населением, в таких субъектах больше детей, то есть и больше детей, занимающихся спортом. По лист диаграмме и рассчитанным децилям можно сказать, что большинство значений не велико (до 200 000). Москва и Московская область являются жесткими выбросами, и также это выбросы по z-score и 3 сигма, то есть их необходимо удалить в любом случае. Краснодарский край, было принято решение тоже удалить, хоть это и мягкий выброс, но его значение 814 730 слишком велико по сравнению с другими городами

1. *Заболеваемость детей злокачественными новообразованиями*

При первичном анализе было 6 выбросов, также крупные регионы РФ. Коэффициент вариации 102%, что говорит о неоднородности данных, также как и для 1ого признака. По лист диаграмме, можно сделать вывод что данные более равномерно распределены внутри размаха вариации, по сравнению с 1ым признаком. Для дальнейших исследований необходимо удалить Москву, так как она встречается в каждом алгоритме нахождения выбросов. Но также удали 2 мягких выброса Московскую область и Санкт Петербург, так как их значения признака >1000, а в среднем значения наблюдений 100 – 200.

1. *Заболеваемость детей психическими расстройствами и расстройствами поведения*

59 из 87 элементов в первых двух stems, первоначально определили 4 выброса. Первые 3 признака в общем можно охарактеризовать правосторонней асимметрией и большой частью данных в малых значениях. Все 4 выброса оказались мягкие, но Московская область, Санкт- Петербург, Дагестан оказались выбросами по 3 сигма правилу, поэтому было принято решение оставить из 4 выбросов только Москву

1. *Доля детей с I и II группой здоровья*.

Распределение близко к нормальному, в выборке встретились 2 некорректных значения процентов, которые не имеет смысл сравнивать с другими субъектами , поэтому было принято решение их убрать, кроме того почти все данные лежат в интервале 70-95 %. Из данных мы уберем Тюменскую область, Ненецкий автономный округ, Архангельскую область как жесткие выбросы и имевшие аномальные значения.

1. *Уровень бедности*.

Распределение близко к логнормальному, изначально было 2 выброса, при дальнейшем исследовании мы поняли что они не экстремальные, но они не проходят ни по z-score ни по 3 сигма, поэтому пришли к выводу убрать Тыву и Ингушетию

# Корреляционный анализ

В этом блоке мы изучим и следуем связи между переменными. Поскольку мы придерживаемся целей, поставленных в начале проекта и нам необходимо проверить достоверность гипотез, то мы оставим те же 5 переменных, которые исследовались на выбросы. То есть актуальность и ценность исследования связи именно этих переменных уже была объяснена в начале работы.

## Построение полей корреляции

Выдвинуто 4 гипотезы о зависимости каждой переменной от количества детей, занимающиеся физической культурой, то есть построим 4 корреляционных поля в Excel.

Сначала мы разберем связь *Численность детей в возрасте 3-15 лет, систематически занимающихся физической культурой и спортом* и *Заболеваемость детей злокачественными новообразованиями.* Каждая точка на графике — это субъект РФ. Здесь мы еще не избавились от выбросов. Отлично видно прямую положительную очень сильную связь двух переменных, точки расположены вдоль линии тренда, и заметны отдалённые выбросы, от которых мы не избавились. Можно предположить, что коэффициент корреляции у такой связи 0,9 – 1. Вывод о положительном знаке связи, можно сделать из положительности коэффициента линии тренда.

Следующая зависимость численности детей с психологическими расстройствами от количества детей спортсменов. Можно заметить схожие черты с предыдущим графиком, только этот в свою очередь более размыт по плоскости, но точки безусловно располагаются вдоль линии тренда причем образуя широкую полосу. Полоса шире чем в предыдущем примере, что говорит о менее сильном коэффициенте корреляции(0,65 – 0,75). Также можно понять что связь положительная, то есть с увеличением числа спортсменов увеличивается число психологически больных. Также отчетливо видны выбросы, которые располагаются дальше от линии тренда, чем основной массив данных.

На следующем графике большинство точек располагаются в одном квадрате, от 0 до 200 000 детей – спортсменов и от 50 до 100 процент здоровых детей. Здесь связь(она положительная) прослеживается намного слабее, чем в предыдущих случаях, но если анализировать линию тренда на участке от 0 до 600 000, то достаточно большой массив данных располагается вдоль нее. Если избавиться от выбросов, то связь будет более явная. Предположительно связь будет слабая, поскольку большинство точек располагается в одном месте, указанном выше, а вдоль линии тренда располагается небольшое количество субъектов.

И с логической точки зрения и анализируя данный график. Наш первый признак слабее всего влияет на уровень бедности. Данные хаотично разбросаны, но присутствует слабая отрицательная связь. Предположение слабой связи, подкреплено тем, что если убрать пару точек на интервале 600 000 – 800 000 то останется просто куча не связанных точек и несколько выбросов в дали, куда идет сейчас линия тренда. Также часть выбросов находится выше 20% уровня бедности.

## Матрица парных коэффициентов

Построение матрицы парных коэффициентов выполним в excel где в клетке i j будет стоять коэфф. корреляции (=КОРРЕЛ) зависимости i переменной от j. Иными словами число от -1 до 1, и чем ближе оно по модулю к 1, тем сильнее связь.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Численность детей в возрасте 3-15 лет, систематически занимающихся физической культурой и спортом, человек** | **Заболеваемость детей злокачественными новообразованиями, человек** | **Заболеваемость детей психическими расстройствами и расстройствами поведения, человек** | **Доля детей с I и II группой здоровья, %** | **Уровень бедности, %** |
| **Численность детей в возрасте 3-15 лет, систематически занимающихся физической культурой и спортом, человек** | 1 | 0,952049589 | 0,793579282 | 0,20168897 | -0,351945714 |
| **Заболеваемость детей злокачественными новообразованиями, человек** | 0,952049589 | 1 | 0,821169247 | 0,141856907 | -0,336824449 |
| **Заболеваемость детей психическими расстройствами и расстройствами поведения, человек** | 0,793579282 | 0,821169247 | 1 | 0,053722761 | -0,259993421 |
| **Доля детей с I и II группой здоровья, %** | 0,20168897 | 0,141856907 | 0,053722761 | 1 | -0,014728691 |
| **Уровень бедности, %** | -0,351945714 | -0,336824449 | -0,259993421 | -0,014728691 | 1 |

Теперь сравним наши предположения о связи между переменными и *Численность детей в возрасте 3-15 лет, систематически занимающихся физической культурой и спортом, человек* с матрицей парных коэффициентов.

1. **Заболеваемость детей злокачественными новообразованиями, человек – 0,952 – сильная положительная связь**
2. **Заболеваемость детей психическими расстройствами и расстройствами поведения, человек – 0,8 сильная положительная связь**
3. **Доля детей с I и II группой здоровья, % - 0,2 – слабая положительная связь**
4. **Уровень бедности, % - (-0,35) – умеренная отрицательная связь**

Предположение по 1,2 и 3 ой связи оказались верными. 4ая связь оказалась не слабой, а средней.

## Построение полей корреляции без выбросов

Для данного графика визуально почти ничего не поменялось, более отчетливо стало видно, что точки лежат вдоль линии тренда, но изначально была сильная положительная связь, и сейчас она тоже сильна положительная, поэтому в таком случае удаление выбросов не влияет на корреляцию, видимо коэффициент корреляции либо не изменится либо изменится на небольшую константу, которая сохранит направление и силу связи этих двух переменных

Для рассматриваемых переменных после удаления выбросов, оставшиеся точки сконцентрировались вдоль линии тренда и связь стала более сильная. То есть она осталась положительная, но увеличился коэффициент корреляции, поскольку точки в дали от линии тренда теперь удалены.

На полученном графике независимо от количество детей – спортсменов, доля детей с 1ой и 2ой группой здоровья располагаются хаотично, следовательно связь будет также слабой, и знак связи поменялся с положительного на отрицательный

Аналогично предыдущим рассуждениям, теперь связь ослабела, наблюдения располагаются хаотично линия тренда почти параллельная оси детей-спортсменов. Множество точек располагаются вдали от линии тренда и знак поменялся на минус.

## Матрица парных коэффициентов без выбросов

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Численность детей в возрасте 3-15 лет, систематически занимающихся физической культурой и спортом, человек** | **Заболеваемость детей злокачественными новообразованиями, человек** | **Заболеваемость детей психическими расстройствами и расстройствами поведения, человек** | **Доля детей с I и II группой здоровья, %** | **Уровень бедности, %** |
| **Численность детей в возрасте 3-15 лет, систематически занимающихся физической культурой и спортом, человек** | 1 | 0,95138226 | 0,8569112 | -0,0806419 | -0,196019 |
| **Заболеваемость детей злокачественными новообразованиями, человек** | 0,95138226 | 1 | 0,8751045 | -0,1082807 | -0,2159259 |
| **Заболеваемость детей психическими расстройствами и расстройствами поведения, человек** | 0,8569112 | 0,8751045 | 1 | -0,1446263 | -0,1491107 |
| **Доля детей с I и II группой здоровья, %** | -0,0806419 | -0,1082807 | -0,1446263 | 1 | 0,15906863 |
| **Уровень бедности, %** | -0,196019 | -0,2159259 | -0,1491107 | 0,15906863 | 1 |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

1. **Заболеваемость детей злокачественными новообразованиями, человек – 0,951 – сильная положительная связь**
2. **Заболеваемость детей психическими расстройствами и расстройствами поведения, человек – 0,856 сильная положительная связь**
3. **Доля детей с I и II группой здоровья, % - (-0,08) – слабая отрицательная связь**
4. **Уровень бедности, % - (-0,196) – умеренная отрицательная связь**

После построения матрицы парных коэффициентов без выбросов можно прийти к следующему выводу: сильная связь, которая была изначально она осталась, причем в случае заболеваний детей с психическими расстройствами она даже стала сильнее. То есть очистка от выбросов помогла усилить связь между признаками, так как не нужные наблюдение были удалены. А для слабой связи и умеренной удаление выбросов только ослабила эти связи, отсюда следует что первоначальны эти связи основывались на том, что аномальные значения хорошо коррелирует на стыке рассматриваемых признаков, а после их удаления уже «чистый» дата сет проявил настающую связь, которая оказалась намного слабее, так для 4 признака – детей с 1ой и 2ой группой здоровья связь вообще почти стала = 0.

## Выводы о исследуемых связях

1,2) Наблюдается сильная прямая связь между количеством детей, которые занимаются спортом и заболеваемостью детей злокачественными образованиями, что опровергает поставленную гипотезу. Вероятно, большое количество занятий спортом может коррелировать с заболеваемостью злокачественными образованиями. Больший объем физической активности может быть связан с увеличенным контактом с окружающей средой, включая воздействие вредных веществ, загрязнение воздуха и контакт с токсичными веществами. Увеличенная физическая активность может повысить риск травм и травматических повреждений. Интенсивные тренировки могут привести к повышенному стрессу на организм, что может оказать негативное воздействие на иммунную систему и увеличить риск развития рака. Люди, которые активно занимаются спортом, могут иметь специфические диетические привычки и образ жизни, которые могут повлиять на риск развития рака. Например, некоторые спортсмены могут прибегать к употреблению большего количества пищевых добавок или употреблять больше белков, что может повысить риск развития рака. Поэтому действительно в теории большое количество спортсменов может напрямую коррелировать с количество больных с новообразованиями и больных с психологическими расстройствами. ( Это не говорит о том что спорт – это вред, видимо конкретно на эту область заболеваний он не влияет положительно)

3)Мы пришли к тому, что доля здоровых детей не зависит от количества спортсменов, это можно аргументировать тем, что здоровье зависит от множества факторов, не только от занятий спортом и заниматься спортом и иметь хорошее здоровье - это два различных аспекта, которые могут быть независимыми друг от друга. Некоторые дети могут иметь здоровые образы жизни без занятий спортом, например, за счет правильного питания или активного образа жизни в целом.

4) Мы получили что количество детей, занимающихся спортом обратно пропорционально слабо влияет на бедность, это можно объяснить, что семьи с низким уровнем дохода могут сталкиваться с финансовыми ограничениями, которые могут затруднять участие их детей в спортивных клубах, тренировках и других мероприятиях. Это может привести к снижению количества занимающихся спортом детей в более бедных регионах.

# Кластерный анализ

В данном этапе мы исследуем методы кластеризации данных, а именно исследуем различные подходы к разделению данных на группы для нахождения наиболее подходящего способа группировки. Кластерный анализ представляет собой набор методик, направленных на классификацию элементов или событий в группы, обладающие высокой степенью внутреннего сходства, известные как кластеры. Основными видами кластерного анализа являются иерархический, предпочтительный для анализа малых объемов данных, и метод k-средних, используемый для обработки больших данных.

Начнем с рассмотрения иерархического анализа. В его рамках, каждый объект изначально рассматривается как отдельный кластер, и затем, на первом этапе, наиболее близкие кластеры объединяются. Этот процесс объединения продолжается до достижения определенного исследователем количества кластеров.

В качестве ключевых показателей для нашего анализа будут использованы следующие данные о субъектах РФ:

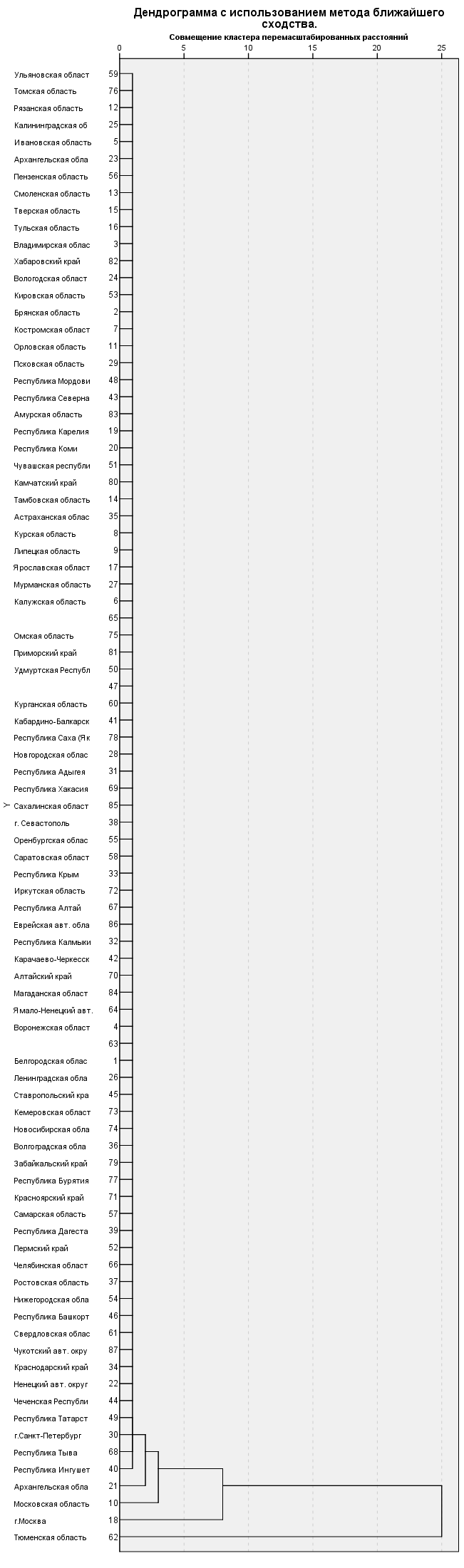
* Численность детей в возрасте 3-15 лет, которые регулярно занимаются физической культурой и спортом;
* Заболеваемость детей злокачественными новообразованиями;
* Заболеваемость детей психическими расстройствами и расстройствами поведения, находящихся под наблюдением на конец отчетного года;
* Доля детей с I и II группой здоровья;
* Уровень бедности.

Эти показатели позволят провести комплексный анализ состояния здоровья и социального благополучия детского населения различных регионов, выявить взаимосвязи и группировать регионы по схожим характеристикам.

## Метод ближнего соседа

Метод ближайшего соседа в кластеризации используется для вычисления расстояния между кластерами на основе наименьшего расстояния между парой объектов, принадлежащих различным кластерам, с применением квадрата евклидова расстояния. При использовании программы SPSS с выбранным "Методом ближайшего соседа", полученные дендрограмма и матрица близости иллюстрируют, как происходит объединение объектов. Объекты с минимальным расстоянием объединяются в первую очередь. Для определения оптимального числа кластеров анализируется таблица, ищется момент резкого увеличения коэффициента расстояния. В нашем случае, применив вычитание 86-78=8, мы определяем, что оптимальное количество кластеров составляет восемь. Этот вывод подтверждается визуальным осмотром дендрограммы и порядок агломерации, указывая на то, что данные в определенном диапазоне формируют кластеры по одному и тому же принципу.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Порядок агломерации (кластеров)** | | | | | | |
| Этап | Объединенный кластер | | Коэффициенты | Этап первого появления кластера | | Следующий этап |
| Кластер 1 | Кластер 2 | Кластер 1 | Кластер 2 |
| 1 | 59 | 76 | ,011 | 0 | 0 | 13 |
| 2 | 12 | 25 | ,015 | 0 | 0 | 7 |
| 3 | 43 | 83 | ,016 | 0 | 0 | 6 |
| 4 | 45 | 73 | ,019 | 0 | 0 | 33 |
| 5 | 3 | 82 | ,020 | 0 | 0 | 11 |
| 6 | 19 | 43 | ,023 | 0 | 3 | 27 |
| 7 | 5 | 12 | ,024 | 0 | 2 | 13 |
| 8 | 15 | 16 | ,026 | 0 | 0 | 20 |
| 9 | 23 | 56 | ,027 | 0 | 0 | 16 |
| 10 | 67 | 86 | ,031 | 0 | 0 | 15 |
| 11 | 3 | 24 | ,035 | 5 | 0 | 20 |
| 12 | 20 | 51 | ,046 | 0 | 0 | 27 |
| 13 | 5 | 59 | ,047 | 7 | 1 | 16 |
| 14 | 7 | 11 | ,048 | 0 | 0 | 34 |
| 15 | 32 | 67 | ,050 | 0 | 10 | 53 |
| 16 | 5 | 23 | ,050 | 13 | 9 | 17 |
| 17 | 5 | 13 | ,052 | 16 | 0 | 26 |
| 18 | 65 | 75 | ,052 | 0 | 0 | 21 |
| 19 | 29 | 48 | ,055 | 0 | 0 | 29 |
| 20 | 3 | 15 | ,056 | 11 | 8 | 26 |
| 21 | 65 | 81 | ,061 | 18 | 0 | 30 |
| 22 | 28 | 31 | ,070 | 0 | 0 | 47 |
| 23 | 55 | 58 | ,075 | 0 | 0 | 42 |
| 24 | 8 | 9 | ,081 | 0 | 0 | 28 |
| 25 | 47 | 60 | ,081 | 0 | 0 | 45 |
| 26 | 3 | 5 | ,085 | 20 | 17 | 31 |
| 27 | 19 | 20 | ,088 | 6 | 12 | 29 |
| 28 | 8 | 17 | ,089 | 24 | 0 | 35 |
| 29 | 19 | 29 | ,092 | 27 | 19 | 34 |
| 30 | 50 | 65 | ,093 | 0 | 21 | 44 |
| 31 | 3 | 53 | ,097 | 26 | 0 | 32 |
| 32 | 2 | 3 | ,100 | 0 | 31 | 38 |
| 33 | 45 | 74 | ,102 | 4 | 0 | 60 |
| 34 | 7 | 19 | ,103 | 14 | 29 | 37 |
| 35 | 8 | 27 | ,106 | 28 | 0 | 36 |
| 36 | 6 | 8 | ,107 | 0 | 35 | 43 |
| 37 | 7 | 80 | ,108 | 34 | 0 | 38 |
| 38 | 2 | 7 | ,115 | 32 | 37 | 39 |
| 39 | 2 | 14 | ,121 | 38 | 0 | 41 |
| 40 | 41 | 78 | ,131 | 0 | 0 | 45 |
| 41 | 2 | 35 | ,139 | 39 | 0 | 43 |
| 42 | 33 | 55 | ,146 | 0 | 23 | 54 |
| 43 | 2 | 6 | ,148 | 41 | 36 | 44 |
| 44 | 2 | 50 | ,152 | 43 | 30 | 46 |
| 45 | 41 | 47 | ,154 | 40 | 25 | 46 |
| 46 | 2 | 41 | ,158 | 44 | 45 | 47 |
| 47 | 2 | 28 | ,162 | 46 | 22 | 49 |
| 48 | 4 | 63 | ,179 | 0 | 0 | 50 |
| 49 | 2 | 69 | ,190 | 47 | 0 | 51 |
| 50 | 1 | 4 | ,201 | 0 | 48 | 57 |
| 51 | 2 | 85 | ,203 | 49 | 0 | 52 |
| 52 | 2 | 38 | ,205 | 51 | 0 | 55 |
| 53 | 32 | 42 | ,213 | 15 | 0 | 56 |
| 54 | 33 | 72 | ,282 | 42 | 0 | 55 |
| 55 | 2 | 33 | ,285 | 52 | 54 | 56 |
| 56 | 2 | 32 | ,298 | 55 | 53 | 58 |
| 57 | 1 | 26 | ,303 | 50 | 0 | 62 |
| 58 | 2 | 70 | ,320 | 56 | 0 | 59 |
| 59 | 2 | 84 | ,328 | 58 | 0 | 61 |
| 60 | 36 | 45 | ,333 | 0 | 33 | 62 |
| 61 | 2 | 64 | ,335 | 59 | 0 | 64 |
| 62 | 1 | 36 | ,345 | 57 | 60 | 64 |
| 63 | 46 | 61 | ,357 | 0 | 0 | 74 |
| 64 | 1 | 2 | ,363 | 62 | 61 | 65 |
| 65 | 1 | 79 | ,374 | 64 | 0 | 66 |
| 66 | 1 | 77 | ,385 | 65 | 0 | 68 |
| 67 | 39 | 52 | ,390 | 0 | 0 | 70 |
| 68 | 1 | 71 | ,425 | 66 | 0 | 69 |
| 69 | 1 | 57 | ,462 | 68 | 0 | 70 |
| 70 | 1 | 39 | ,475 | 69 | 67 | 71 |
| 71 | 1 | 66 | ,492 | 70 | 0 | 72 |
| 72 | 1 | 37 | ,530 | 71 | 0 | 73 |
| 73 | 1 | 54 | ,571 | 72 | 0 | 74 |
| 74 | 1 | 46 | ,726 | 73 | 63 | 75 |
| 75 | 1 | 87 | ,800 | 74 | 0 | 76 |
| 76 | 1 | 34 | ,849 | 75 | 0 | 77 |
| 77 | 1 | 22 | ,948 | 76 | 0 | 78 |
| 78 | 1 | 44 | ,970 | 77 | 0 | 79 |
| 79 | 1 | 49 | 1,193 | 78 | 0 | 80 |
| 80 | 1 | 30 | 1,278 | 79 | 0 | 81 |
| 81 | 1 | 68 | 1,291 | 80 | 0 | 82 |
| 82 | 1 | 40 | 1,827 | 81 | 0 | 83 |
| 83 | 1 | 21 | 3,257 | 82 | 0 | 84 |
| 84 | 1 | 10 | 4,932 | 83 | 0 | 85 |
| 85 | 1 | 18 | 15,811 | 84 | 0 | 86 |
| 86 | 1 | 62 | 54,608 | 85 | 0 | 0 |

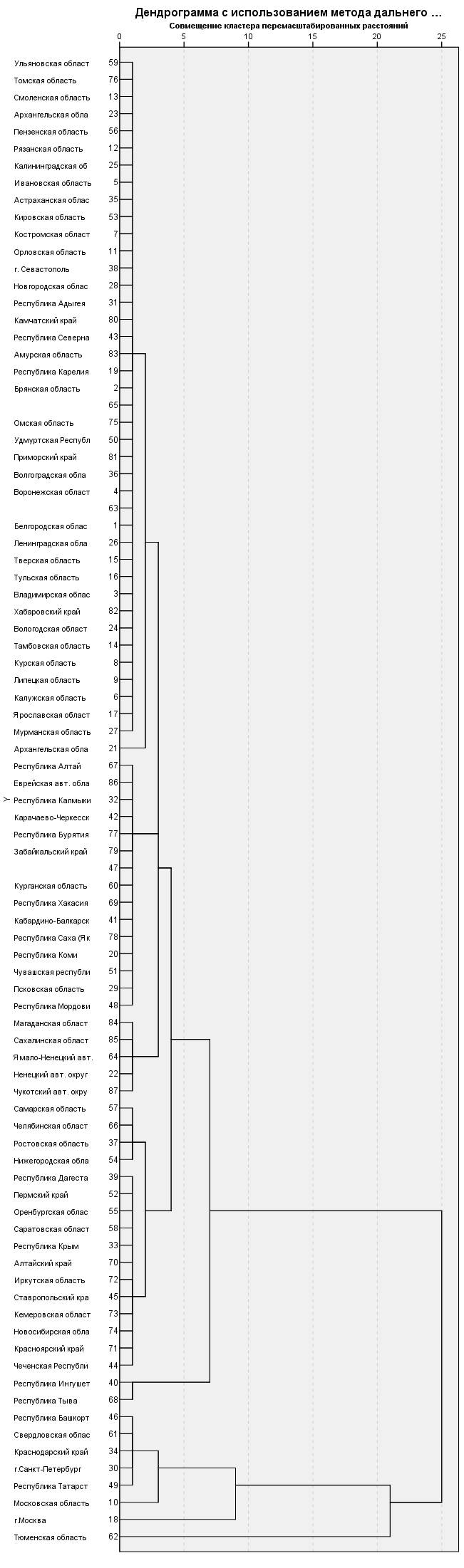


## Метод дальнего соседа

Применяя квадрат евклидова расстояния для измерения интервалов между объектами, метод дальнего соседа в анализе кластеризации идентифицирует расстояние между двумя кластерами, основываясь на максимальном промежутке между их наиболее удалёнными элементами, с каждый элемент представляющий собой отдельный кластер. В ходе работы с программой SPSS и выбором опции "Метод дальнего соседа", мы получаем дендрограмму и матрицу близости, которые иллюстрируют поэтапное увеличение расстояний между объектами. Из этих данных становится возможным определение оптимального количества кластеров, для чего требуется обнаружить момент значительного увеличения расстояния между кластерами, а затем выполнить вычитание номера этой стадии агломерации из общего числа объектов. В данном случае, принимая во внимание корректировку на величину (86-78), выводится, что идеальное число кластеров составляет восемь, 78-86=8. Этот вывод находит подтверждение в визуальном анализе дендрограммы, которая также указывает на оптимальное количество кластеров равное восьми.

При детальном рассмотрении дендрограммы, где кластеры обозначены как субъекты Российской Федерации, выделяются только три кластера, разграниченных между собой заметными промежутками. Это демонстрирует устойчивую взаимосвязь между данными внутри каждого кластера. Заключительное объединение кластеров происходит между группами данных, между которыми наблюдается резкий скачок расстояний, и теми, расстояния между которыми до этого были более или менее одинаковыми, что подчёркивает логику выбора в методе дальнего соседа, основанного на идентификации максимального расстояния между элементами кластера.

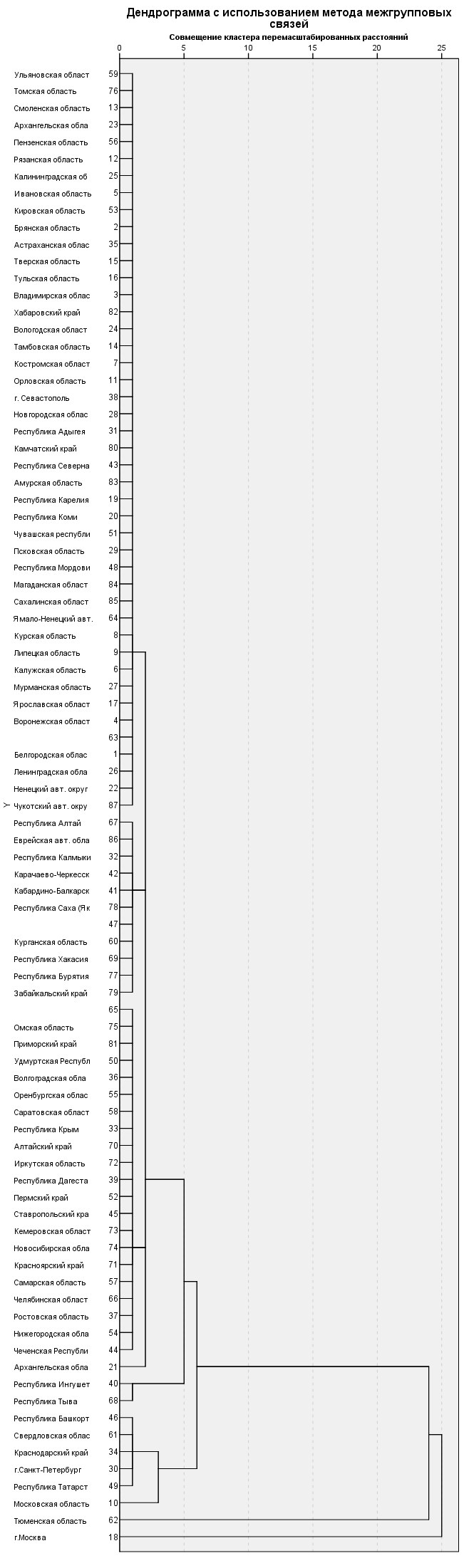
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Порядок агломерации (кластеров)** | | | | | | |
| Этап | Объединенный кластер | | Коэффициенты | Этап первого появления кластера | | Следующий этап |
| Кластер 1 | Кластер 2 | Кластер 1 | Кластер 2 |
| 1 | 59 | 76 | ,011 | 0 | 0 | 18 |
| 2 | 12 | 25 | ,015 | 0 | 0 | 9 |
| 3 | 43 | 83 | ,016 | 0 | 0 | 15 |
| 4 | 45 | 73 | ,019 | 0 | 0 | 27 |
| 5 | 3 | 82 | ,020 | 0 | 0 | 12 |
| 6 | 15 | 16 | ,026 | 0 | 0 | 26 |
| 7 | 23 | 56 | ,027 | 0 | 0 | 23 |
| 8 | 67 | 86 | ,031 | 0 | 0 | 19 |
| 9 | 5 | 12 | ,037 | 0 | 2 | 23 |
| 10 | 20 | 51 | ,046 | 0 | 0 | 33 |
| 11 | 7 | 11 | ,048 | 0 | 0 | 42 |
| 12 | 3 | 24 | ,051 | 5 | 0 | 26 |
| 13 | 65 | 75 | ,052 | 0 | 0 | 37 |
| 14 | 29 | 48 | ,055 | 0 | 0 | 33 |
| 15 | 19 | 43 | ,059 | 0 | 3 | 34 |
| 16 | 28 | 31 | ,070 | 0 | 0 | 30 |
| 17 | 55 | 58 | ,075 | 0 | 0 | 28 |
| 18 | 13 | 59 | ,077 | 0 | 1 | 39 |
| 19 | 32 | 67 | ,080 | 0 | 8 | 43 |
| 20 | 8 | 9 | ,081 | 0 | 0 | 24 |
| 21 | 47 | 60 | ,081 | 0 | 0 | 36 |
| 22 | 50 | 81 | ,093 | 0 | 0 | 37 |
| 23 | 5 | 23 | ,126 | 9 | 7 | 39 |
| 24 | 6 | 8 | ,130 | 0 | 20 | 46 |
| 25 | 41 | 78 | ,131 | 0 | 0 | 52 |
| 26 | 3 | 15 | ,132 | 12 | 6 | 32 |
| 27 | 45 | 74 | ,141 | 4 | 0 | 56 |
| 28 | 33 | 55 | ,171 | 0 | 17 | 55 |
| 29 | 4 | 63 | ,179 | 0 | 0 | 48 |
| 30 | 28 | 80 | ,188 | 16 | 0 | 54 |
| 31 | 17 | 27 | ,204 | 0 | 0 | 46 |
| 32 | 3 | 14 | ,208 | 26 | 0 | 57 |
| 33 | 20 | 29 | ,210 | 10 | 14 | 52 |
| 34 | 2 | 19 | ,228 | 0 | 15 | 54 |
| 35 | 35 | 53 | ,230 | 0 | 0 | 49 |
| 36 | 47 | 69 | ,298 | 21 | 0 | 58 |
| 37 | 50 | 65 | ,300 | 22 | 13 | 50 |
| 38 | 1 | 26 | ,303 | 0 | 0 | 48 |
| 39 | 5 | 13 | ,304 | 23 | 18 | 49 |
| 40 | 70 | 72 | ,320 | 0 | 0 | 55 |
| 41 | 84 | 85 | ,328 | 0 | 0 | 59 |
| 42 | 7 | 38 | ,340 | 11 | 0 | 61 |
| 43 | 32 | 42 | ,353 | 19 | 0 | 73 |
| 44 | 46 | 61 | ,357 | 0 | 0 | 62 |
| 45 | 39 | 52 | ,390 | 0 | 0 | 67 |
| 46 | 6 | 17 | ,405 | 24 | 31 | 57 |
| 47 | 77 | 79 | ,422 | 0 | 0 | 66 |
| 48 | 1 | 4 | ,456 | 38 | 29 | 69 |
| 49 | 5 | 35 | ,464 | 39 | 35 | 70 |
| 50 | 36 | 50 | ,465 | 0 | 37 | 71 |
| 51 | 57 | 66 | ,492 | 0 | 0 | 53 |
| 52 | 20 | 41 | ,531 | 33 | 25 | 58 |
| 53 | 37 | 57 | ,559 | 0 | 51 | 64 |
| 54 | 2 | 28 | ,737 | 34 | 30 | 61 |
| 55 | 33 | 70 | ,741 | 28 | 40 | 63 |
| 56 | 45 | 71 | ,757 | 27 | 0 | 63 |
| 57 | 3 | 6 | ,807 | 32 | 46 | 69 |
| 58 | 20 | 47 | ,880 | 52 | 36 | 66 |
| 59 | 64 | 84 | ,891 | 0 | 41 | 76 |
| 60 | 22 | 87 | ,948 | 0 | 0 | 76 |
| 61 | 2 | 7 | ,977 | 54 | 42 | 70 |
| 62 | 34 | 46 | 1,172 | 0 | 44 | 72 |
| 63 | 33 | 45 | 1,360 | 55 | 56 | 67 |
| 64 | 37 | 54 | 1,412 | 53 | 0 | 77 |
| 65 | 30 | 49 | 1,646 | 0 | 0 | 72 |
| 66 | 20 | 77 | 1,653 | 58 | 47 | 73 |
| 67 | 33 | 39 | 1,793 | 63 | 45 | 74 |
| 68 | 40 | 68 | 1,827 | 0 | 0 | 83 |
| 69 | 1 | 3 | 1,926 | 48 | 57 | 71 |
| 70 | 2 | 5 | 1,953 | 61 | 49 | 75 |
| 71 | 1 | 36 | 2,464 | 69 | 50 | 75 |
| 72 | 30 | 34 | 3,107 | 65 | 62 | 79 |
| 73 | 20 | 32 | 3,236 | 66 | 43 | 80 |
| 74 | 33 | 44 | 3,918 | 67 | 0 | 77 |
| 75 | 1 | 2 | 4,660 | 71 | 70 | 78 |
| 76 | 22 | 64 | 4,706 | 60 | 59 | 81 |
| 77 | 33 | 37 | 6,402 | 74 | 64 | 82 |
| 78 | 1 | 21 | 7,132 | 75 | 0 | 80 |
| 79 | 10 | 30 | 13,054 | 0 | 72 | 84 |
| 80 | 1 | 20 | 13,361 | 78 | 73 | 81 |
| 81 | 1 | 22 | 15,702 | 80 | 76 | 82 |
| 82 | 1 | 33 | 17,900 | 81 | 77 | 83 |
| 83 | 1 | 40 | 32,960 | 82 | 68 | 86 |
| 84 | 10 | 18 | 48,091 | 79 | 0 | 85 |
| 85 | 10 | 62 | 114,323 | 84 | 0 | 86 |
| 86 | 1 | 10 | 136,939 | 83 | 85 | 0 |



## Метод межгрупповой связи

Применение метода межгрупповой связи в SPSS для анализа кластеризации показало, что идеальное число кластеров равняется 8. Это определение основано на анализе скачка роста коэффициента расстояния между кластерами, отмеченного на 86-м шаге агломерации. Таким образом, из общего количества исследуемых объектов (78), вычитая номер шага агломерации, на котором наблюдается скачок (86), получаем оптимальное количество кластеров - 8.

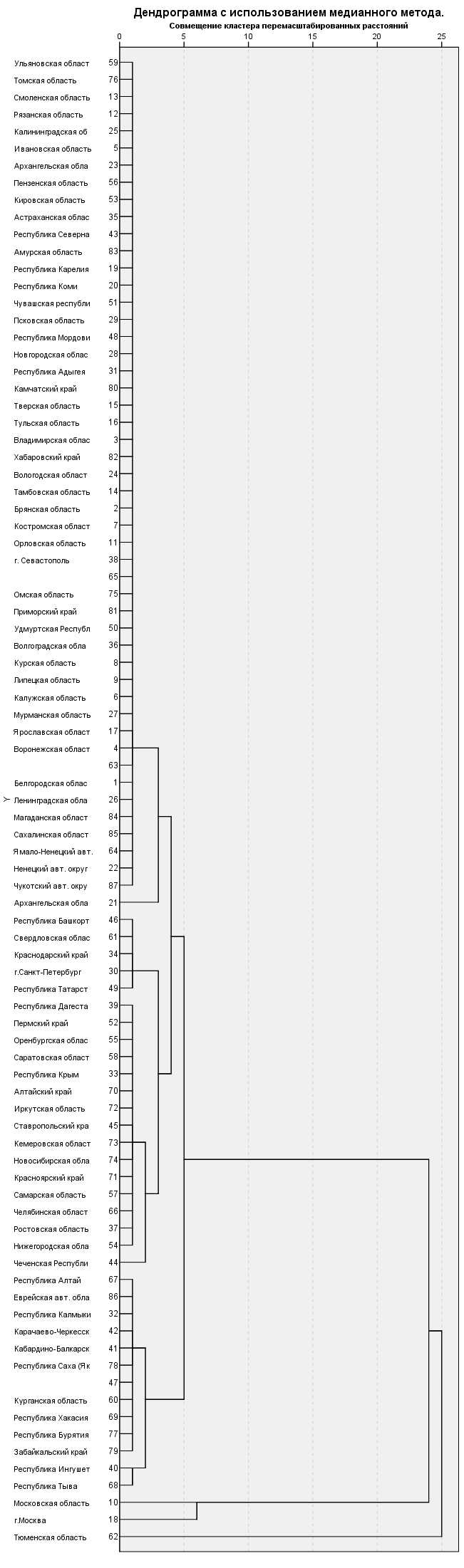
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Порядок агломерации (кластеров)** | | | | | | |
| Этап | Объединенный кластер | | Коэффициенты | Этап первого появления кластера | | Следующий этап |
| Кластер 1 | Кластер 2 | Кластер 1 | Кластер 2 |
| 1 | 59 | 76 | ,011 | 0 | 0 | 16 |
| 2 | 12 | 25 | ,015 | 0 | 0 | 8 |
| 3 | 43 | 83 | ,016 | 0 | 0 | 10 |
| 4 | 45 | 73 | ,019 | 0 | 0 | 26 |
| 5 | 3 | 82 | ,020 | 0 | 0 | 11 |
| 6 | 15 | 16 | ,026 | 0 | 0 | 24 |
| 7 | 23 | 56 | ,027 | 0 | 0 | 23 |
| 8 | 5 | 12 | ,031 | 0 | 2 | 23 |
| 9 | 67 | 86 | ,031 | 0 | 0 | 17 |
| 10 | 19 | 43 | ,041 | 0 | 3 | 36 |
| 11 | 3 | 24 | ,043 | 5 | 0 | 24 |
| 12 | 20 | 51 | ,046 | 0 | 0 | 28 |
| 13 | 7 | 11 | ,048 | 0 | 0 | 41 |
| 14 | 65 | 75 | ,052 | 0 | 0 | 20 |
| 15 | 29 | 48 | ,055 | 0 | 0 | 28 |
| 16 | 13 | 59 | ,064 | 0 | 1 | 31 |
| 17 | 32 | 67 | ,065 | 0 | 9 | 42 |
| 18 | 28 | 31 | ,070 | 0 | 0 | 33 |
| 19 | 55 | 58 | ,075 | 0 | 0 | 30 |
| 20 | 65 | 81 | ,079 | 14 | 0 | 37 |
| 21 | 8 | 9 | ,081 | 0 | 0 | 25 |
| 22 | 47 | 60 | ,081 | 0 | 0 | 40 |
| 23 | 5 | 23 | ,085 | 8 | 7 | 29 |
| 24 | 3 | 15 | ,093 | 11 | 6 | 32 |
| 25 | 6 | 8 | ,119 | 0 | 21 | 35 |
| 26 | 45 | 74 | ,122 | 4 | 0 | 58 |
| 27 | 41 | 78 | ,131 | 0 | 0 | 51 |
| 28 | 20 | 29 | ,132 | 12 | 15 | 36 |
| 29 | 5 | 53 | ,150 | 23 | 0 | 31 |
| 30 | 33 | 55 | ,159 | 0 | 19 | 56 |
| 31 | 5 | 13 | ,169 | 29 | 16 | 38 |
| 32 | 3 | 14 | ,172 | 24 | 0 | 53 |
| 33 | 28 | 80 | ,175 | 18 | 0 | 54 |
| 34 | 4 | 63 | ,179 | 0 | 0 | 47 |
| 35 | 6 | 27 | ,179 | 25 | 0 | 39 |
| 36 | 19 | 20 | ,186 | 10 | 28 | 54 |
| 37 | 50 | 65 | ,209 | 0 | 20 | 50 |
| 38 | 2 | 5 | ,215 | 0 | 31 | 45 |
| 39 | 6 | 17 | ,240 | 35 | 0 | 63 |
| 40 | 47 | 69 | ,244 | 22 | 0 | 51 |
| 41 | 7 | 38 | ,272 | 13 | 0 | 59 |
| 42 | 32 | 42 | ,273 | 17 | 0 | 70 |
| 43 | 1 | 26 | ,303 | 0 | 0 | 47 |
| 44 | 70 | 72 | ,320 | 0 | 0 | 56 |
| 45 | 2 | 35 | ,321 | 38 | 0 | 53 |
| 46 | 84 | 85 | ,328 | 0 | 0 | 60 |
| 47 | 1 | 4 | ,343 | 43 | 34 | 63 |
| 48 | 46 | 61 | ,357 | 0 | 0 | 66 |
| 49 | 39 | 52 | ,390 | 0 | 0 | 69 |
| 50 | 36 | 50 | ,414 | 0 | 37 | 64 |
| 51 | 41 | 47 | ,419 | 27 | 40 | 61 |
| 52 | 77 | 79 | ,422 | 0 | 0 | 61 |
| 53 | 2 | 3 | ,444 | 45 | 32 | 59 |
| 54 | 19 | 28 | ,464 | 36 | 33 | 62 |
| 55 | 57 | 66 | ,492 | 0 | 0 | 57 |
| 56 | 33 | 70 | ,499 | 30 | 44 | 64 |
| 57 | 37 | 57 | ,545 | 0 | 55 | 67 |
| 58 | 45 | 71 | ,587 | 26 | 0 | 68 |
| 59 | 2 | 7 | ,608 | 53 | 41 | 62 |
| 60 | 64 | 84 | ,613 | 0 | 46 | 71 |
| 61 | 41 | 77 | ,692 | 51 | 52 | 70 |
| 62 | 2 | 19 | ,739 | 59 | 54 | 74 |
| 63 | 1 | 6 | ,889 | 47 | 39 | 71 |
| 64 | 33 | 36 | ,899 | 56 | 50 | 76 |
| 65 | 22 | 87 | ,948 | 0 | 0 | 77 |
| 66 | 34 | 46 | 1,011 | 0 | 48 | 75 |
| 67 | 37 | 54 | 1,026 | 57 | 0 | 68 |
| 68 | 37 | 45 | 1,041 | 67 | 58 | 69 |
| 69 | 37 | 39 | 1,401 | 68 | 49 | 76 |
| 70 | 32 | 41 | 1,443 | 42 | 61 | 79 |
| 71 | 1 | 64 | 1,493 | 63 | 60 | 74 |
| 72 | 30 | 49 | 1,646 | 0 | 0 | 75 |
| 73 | 40 | 68 | 1,827 | 0 | 0 | 83 |
| 74 | 1 | 2 | 1,847 | 71 | 62 | 77 |
| 75 | 30 | 34 | 1,873 | 72 | 66 | 82 |
| 76 | 33 | 37 | 1,960 | 64 | 69 | 78 |
| 77 | 1 | 22 | 3,366 | 74 | 65 | 79 |
| 78 | 33 | 44 | 3,388 | 76 | 0 | 80 |
| 79 | 1 | 32 | 4,087 | 77 | 70 | 80 |
| 80 | 1 | 33 | 4,804 | 79 | 78 | 81 |
| 81 | 1 | 21 | 6,381 | 80 | 0 | 83 |
| 82 | 10 | 30 | 9,136 | 0 | 75 | 84 |
| 83 | 1 | 40 | 15,915 | 81 | 73 | 84 |
| 84 | 1 | 10 | 20,672 | 83 | 82 | 85 |
| 85 | 1 | 62 | 87,357 | 84 | 0 | 86 |
| 86 | 1 | 18 | 93,323 | 85 | 0 | 0 |



## Метод медианной связи

Используя метод медианной связи для кластеризации, где расстояние между объектами определяется как квадрат евклидова расстояния, и применяя программное обеспечение SPSS с выбором указанного метода, были получены результаты, отраженные в дендрограмме и матрице сходства. Анализируя эти результаты, особое внимание уделяется увеличению расстояния между объектами на каждом этапе слияния. Чтобы определить оптимальное количество кластеров, необходимо обратить внимание на внезапный рост коэффициента расстояния. В данном случае, значительный скачок расстояния произошел на 86-м шаге, что позволяет заключить, что идеальное количество кластеров составляет 86-78 = 8, что, учитывая контекст, означает наличие 8 кластеров. Такой вывод также подтверждается визуальным изучением дендрограммы. В результате анализа и сравнения пяти различных методов кластеризации можно утверждать, что оптимальное число кластеров равно 8.

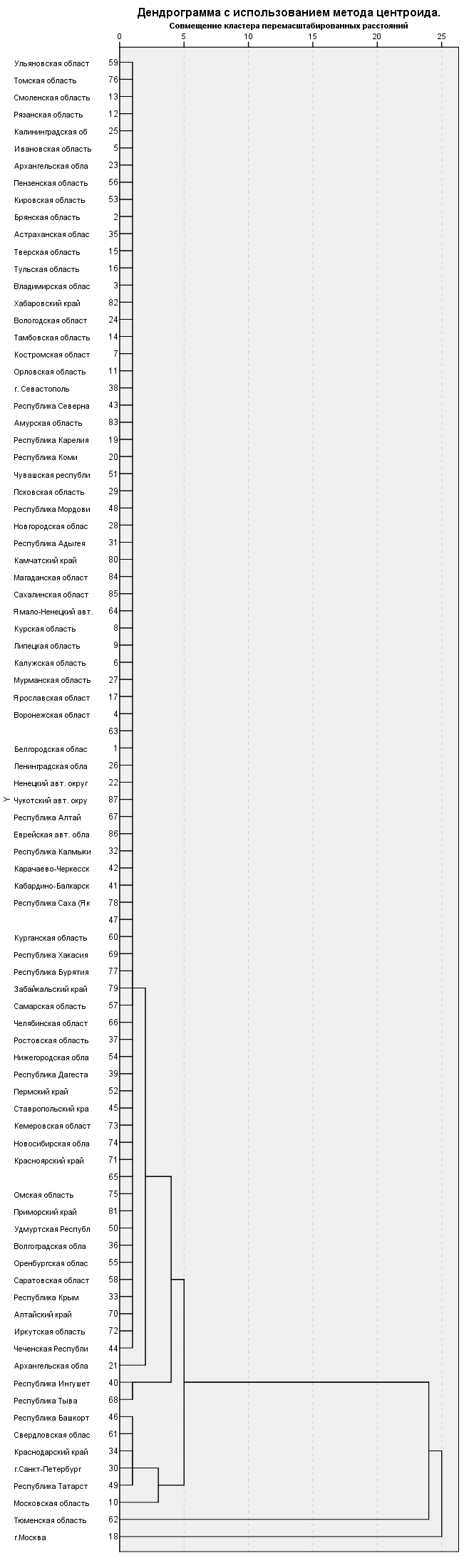
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Порядок агломерации (кластеров)** | | | | | | |
| Этап | Объединенный кластер | | Коэффициенты | Этап первого появления кластера | | Следующий этап |
| Кластер 1 | Кластер 2 | Кластер 1 | Кластер 2 |
| 1 | 59 | 76 | ,011 | 0 | 0 | 17 |
| 2 | 12 | 25 | ,015 | 0 | 0 | 7 |
| 3 | 43 | 83 | ,016 | 0 | 0 | 10 |
| 4 | 45 | 73 | ,019 | 0 | 0 | 27 |
| 5 | 3 | 82 | ,020 | 0 | 0 | 11 |
| 6 | 15 | 16 | ,026 | 0 | 0 | 24 |
| 7 | 5 | 12 | ,027 | 0 | 2 | 20 |
| 8 | 23 | 56 | ,027 | 0 | 0 | 20 |
| 9 | 67 | 86 | ,031 | 0 | 0 | 16 |
| 10 | 19 | 43 | ,037 | 0 | 3 | 31 |
| 11 | 3 | 24 | ,038 | 5 | 0 | 24 |
| 12 | 20 | 51 | ,046 | 0 | 0 | 26 |
| 13 | 7 | 11 | ,048 | 0 | 0 | 43 |
| 14 | 65 | 75 | ,052 | 0 | 0 | 18 |
| 15 | 29 | 48 | ,055 | 0 | 0 | 26 |
| 16 | 32 | 67 | ,057 | 0 | 9 | 42 |
| 17 | 13 | 59 | ,061 | 0 | 1 | 36 |
| 18 | 65 | 81 | ,066 | 14 | 0 | 33 |
| 19 | 28 | 31 | ,070 | 0 | 0 | 34 |
| 20 | 5 | 23 | ,075 | 7 | 8 | 29 |
| 21 | 55 | 58 | ,075 | 0 | 0 | 30 |
| 22 | 8 | 9 | ,081 | 0 | 0 | 25 |
| 23 | 47 | 60 | ,081 | 0 | 0 | 41 |
| 24 | 3 | 15 | ,083 | 11 | 6 | 32 |
| 25 | 6 | 8 | ,098 | 0 | 22 | 35 |
| 26 | 20 | 29 | ,106 | 12 | 15 | 31 |
| 27 | 45 | 74 | ,117 | 4 | 0 | 62 |
| 28 | 41 | 78 | ,131 | 0 | 0 | 51 |
| 29 | 5 | 53 | ,131 | 20 | 0 | 36 |
| 30 | 33 | 55 | ,140 | 0 | 21 | 50 |
| 31 | 19 | 20 | ,140 | 10 | 26 | 46 |
| 32 | 3 | 14 | ,145 | 24 | 0 | 40 |
| 33 | 50 | 65 | ,157 | 0 | 18 | 49 |
| 34 | 28 | 80 | ,158 | 19 | 0 | 46 |
| 35 | 6 | 27 | ,164 | 25 | 0 | 39 |
| 36 | 5 | 13 | ,173 | 29 | 17 | 37 |
| 37 | 5 | 35 | ,175 | 36 | 0 | 61 |
| 38 | 4 | 63 | ,179 | 0 | 0 | 44 |
| 39 | 6 | 17 | ,188 | 35 | 0 | 63 |
| 40 | 2 | 3 | ,220 | 0 | 32 | 53 |
| 41 | 47 | 69 | ,224 | 23 | 0 | 51 |
| 42 | 32 | 42 | ,240 | 16 | 0 | 72 |
| 43 | 7 | 38 | ,260 | 13 | 0 | 53 |
| 44 | 1 | 4 | ,284 | 0 | 38 | 45 |
| 45 | 1 | 26 | ,237 | 44 | 0 | 63 |
| 46 | 19 | 28 | ,291 | 31 | 34 | 54 |
| 47 | 70 | 72 | ,320 | 0 | 0 | 50 |
| 48 | 84 | 85 | ,328 | 0 | 0 | 60 |
| 49 | 36 | 50 | ,342 | 0 | 33 | 66 |
| 50 | 33 | 70 | ,347 | 30 | 47 | 64 |
| 51 | 41 | 47 | ,351 | 28 | 41 | 57 |
| 52 | 46 | 61 | ,357 | 0 | 0 | 68 |
| 53 | 2 | 7 | ,378 | 40 | 43 | 54 |
| 54 | 2 | 19 | ,374 | 53 | 46 | 61 |
| 55 | 39 | 52 | ,390 | 0 | 0 | 65 |
| 56 | 77 | 79 | ,422 | 0 | 0 | 57 |
| 57 | 41 | 77 | ,414 | 51 | 56 | 72 |
| 58 | 57 | 66 | ,492 | 0 | 0 | 59 |
| 59 | 37 | 57 | ,422 | 0 | 58 | 67 |
| 60 | 64 | 84 | ,531 | 0 | 48 | 76 |
| 61 | 2 | 5 | ,553 | 54 | 37 | 70 |
| 62 | 45 | 71 | ,597 | 27 | 0 | 64 |
| 63 | 1 | 6 | ,645 | 45 | 39 | 66 |
| 64 | 33 | 45 | ,647 | 50 | 62 | 65 |
| 65 | 33 | 39 | ,801 | 64 | 55 | 71 |
| 66 | 1 | 36 | ,851 | 63 | 49 | 70 |
| 67 | 37 | 54 | ,875 | 59 | 0 | 71 |
| 68 | 34 | 46 | ,921 | 0 | 52 | 73 |
| 69 | 22 | 87 | ,948 | 0 | 0 | 77 |
| 70 | 1 | 2 | 1,182 | 66 | 61 | 76 |
| 71 | 33 | 37 | 1,212 | 65 | 67 | 78 |
| 72 | 32 | 41 | 1,321 | 42 | 57 | 79 |
| 73 | 30 | 34 | 1,582 | 0 | 68 | 74 |
| 74 | 30 | 49 | 1,405 | 73 | 0 | 80 |
| 75 | 40 | 68 | 1,827 | 0 | 0 | 79 |
| 76 | 1 | 64 | 1,871 | 70 | 60 | 77 |
| 77 | 1 | 22 | 2,381 | 76 | 69 | 81 |
| 78 | 33 | 44 | 3,367 | 71 | 0 | 80 |
| 79 | 32 | 40 | 4,468 | 72 | 75 | 83 |
| 80 | 30 | 33 | 7,102 | 74 | 78 | 82 |
| 81 | 1 | 21 | 8,423 | 77 | 0 | 82 |
| 82 | 1 | 30 | 9,274 | 81 | 80 | 83 |
| 83 | 1 | 32 | 11,891 | 82 | 79 | 85 |
| 84 | 10 | 18 | 15,811 | 0 | 0 | 85 |
| 85 | 1 | 10 | 66,767 | 83 | 84 | 86 |
| 86 | 1 | 62 | 71,803 | 85 | 0 | 0 |



## Метод центр тяжести

Применяя метод центра тяжести в анализе кластеризации, который включает в себя расчет средних значений переменных для каждого кластера и последующее определение квадратичного евклидова расстояния между этими средними, был использован программный пакет SPSS с выбором "Метода дальнего соседа". Это привело к созданию дендрограммы и матрицы сходства, которые отображают, как расстояние между объектами возрастает на каждом этапе. Из этих данных можно вывести оптимальное количество кластеров, обратив внимание на моменты заметного увеличения коэффициента расстояния. В данном анализе, важное увеличение было зафиксировано на 86-м шаге, что предполагает, что идеальное количество кластеров составляет (86-78) = 8. Такой вывод также находит подтверждение при наблюдении за построенной дендрограммой, подтверждая, что наилучшее число кластеров равно 8.

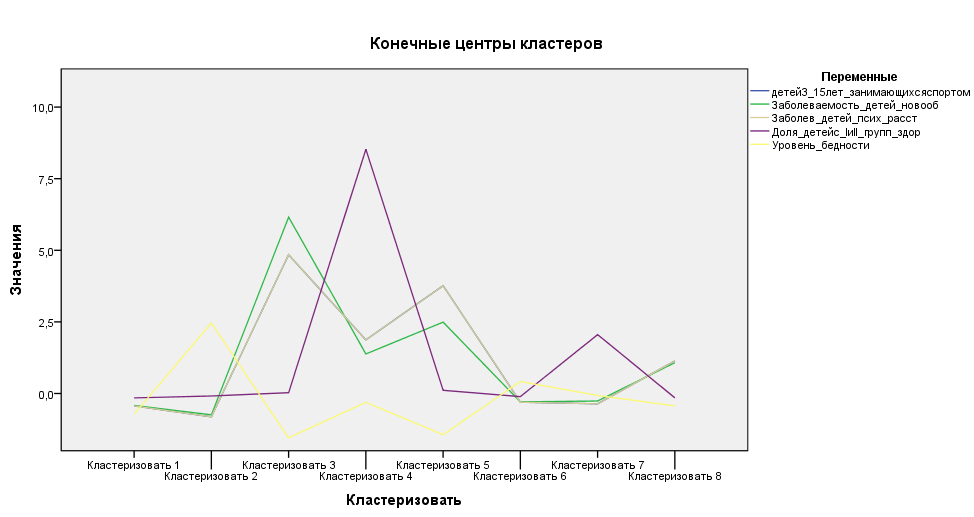
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Порядок агломерации (кластеров)** | | | | | | |
| Этап | Объединенный кластер | | Коэффициенты | Этап первого появления кластера | | Следующий этап |
| Кластер 1 | Кластер 2 | Кластер 1 | Кластер 2 |
| 1 | 59 | 76 | ,011 | 0 | 0 | 17 |
| 2 | 12 | 25 | ,015 | 0 | 0 | 7 |
| 3 | 43 | 83 | ,016 | 0 | 0 | 10 |
| 4 | 45 | 73 | ,019 | 0 | 0 | 28 |
| 5 | 3 | 82 | ,020 | 0 | 0 | 11 |
| 6 | 15 | 16 | ,026 | 0 | 0 | 21 |
| 7 | 5 | 12 | ,027 | 0 | 2 | 20 |
| 8 | 23 | 56 | ,027 | 0 | 0 | 20 |
| 9 | 67 | 86 | ,031 | 0 | 0 | 16 |
| 10 | 19 | 43 | ,037 | 0 | 3 | 30 |
| 11 | 3 | 24 | ,038 | 5 | 0 | 21 |
| 12 | 20 | 51 | ,046 | 0 | 0 | 26 |
| 13 | 7 | 11 | ,048 | 0 | 0 | 43 |
| 14 | 65 | 75 | ,052 | 0 | 0 | 18 |
| 15 | 29 | 48 | ,055 | 0 | 0 | 26 |
| 16 | 32 | 67 | ,057 | 0 | 9 | 41 |
| 17 | 13 | 59 | ,061 | 0 | 1 | 27 |
| 18 | 65 | 81 | ,066 | 14 | 0 | 38 |
| 19 | 28 | 31 | ,070 | 0 | 0 | 35 |
| 20 | 5 | 23 | ,070 | 7 | 8 | 27 |
| 21 | 3 | 15 | ,075 | 11 | 6 | 34 |
| 22 | 55 | 58 | ,075 | 0 | 0 | 32 |
| 23 | 8 | 9 | ,081 | 0 | 0 | 25 |
| 24 | 47 | 60 | ,081 | 0 | 0 | 40 |
| 25 | 6 | 8 | ,098 | 0 | 23 | 33 |
| 26 | 20 | 29 | ,106 | 12 | 15 | 30 |
| 27 | 5 | 13 | ,114 | 20 | 17 | 31 |
| 28 | 45 | 74 | ,117 | 4 | 0 | 62 |
| 29 | 41 | 78 | ,131 | 0 | 0 | 48 |
| 30 | 19 | 20 | ,136 | 10 | 26 | 51 |
| 31 | 5 | 53 | ,138 | 27 | 0 | 36 |
| 32 | 33 | 55 | ,140 | 0 | 22 | 53 |
| 33 | 6 | 27 | ,144 | 25 | 0 | 39 |
| 34 | 3 | 14 | ,144 | 21 | 0 | 47 |
| 35 | 28 | 80 | ,158 | 19 | 0 | 51 |
| 36 | 2 | 5 | ,159 | 0 | 31 | 42 |
| 37 | 4 | 63 | ,179 | 0 | 0 | 44 |
| 38 | 50 | 65 | ,185 | 0 | 18 | 52 |
| 39 | 6 | 17 | ,186 | 33 | 0 | 65 |
| 40 | 47 | 69 | ,224 | 24 | 0 | 48 |
| 41 | 32 | 42 | ,255 | 16 | 0 | 72 |
| 42 | 2 | 35 | ,256 | 36 | 0 | 47 |
| 43 | 7 | 38 | ,260 | 13 | 0 | 55 |
| 44 | 1 | 4 | ,284 | 0 | 37 | 45 |
| 45 | 1 | 26 | ,246 | 44 | 0 | 65 |
| 46 | 70 | 72 | ,320 | 0 | 0 | 53 |
| 47 | 2 | 3 | ,321 | 42 | 34 | 55 |
| 48 | 41 | 47 | ,323 | 29 | 40 | 58 |
| 49 | 84 | 85 | ,328 | 0 | 0 | 61 |
| 50 | 46 | 61 | ,357 | 0 | 0 | 67 |
| 51 | 19 | 28 | ,357 | 30 | 35 | 57 |
| 52 | 36 | 50 | ,362 | 0 | 38 | 63 |
| 53 | 33 | 70 | ,376 | 32 | 46 | 63 |
| 54 | 39 | 52 | ,390 | 0 | 0 | 70 |
| 55 | 2 | 7 | ,401 | 47 | 43 | 57 |
| 56 | 77 | 79 | ,422 | 0 | 0 | 58 |
| 57 | 2 | 19 | ,427 | 55 | 51 | 71 |
| 58 | 41 | 77 | ,458 | 48 | 56 | 72 |
| 59 | 57 | 66 | ,492 | 0 | 0 | 60 |
| 60 | 37 | 57 | ,422 | 0 | 59 | 66 |
| 61 | 64 | 84 | ,531 | 0 | 49 | 69 |
| 62 | 45 | 71 | ,557 | 28 | 0 | 64 |
| 63 | 33 | 36 | ,651 | 53 | 52 | 64 |
| 64 | 33 | 45 | ,675 | 63 | 62 | 70 |
| 65 | 1 | 6 | ,700 | 45 | 39 | 69 |
| 66 | 37 | 54 | ,850 | 60 | 0 | 75 |
| 67 | 34 | 46 | ,921 | 0 | 50 | 73 |
| 68 | 22 | 87 | ,948 | 0 | 0 | 77 |
| 69 | 1 | 64 | 1,056 | 65 | 61 | 71 |
| 70 | 33 | 39 | 1,072 | 64 | 54 | 75 |
| 71 | 1 | 2 | 1,148 | 69 | 57 | 77 |
| 72 | 32 | 41 | 1,166 | 41 | 58 | 79 |
| 73 | 30 | 34 | 1,544 | 0 | 67 | 74 |
| 74 | 30 | 49 | 1,377 | 73 | 0 | 82 |
| 75 | 33 | 37 | 1,565 | 70 | 66 | 78 |
| 76 | 40 | 68 | 1,827 | 0 | 0 | 83 |
| 77 | 1 | 22 | 2,584 | 71 | 68 | 79 |
| 78 | 33 | 44 | 2,711 | 75 | 0 | 80 |
| 79 | 1 | 32 | 3,015 | 77 | 72 | 80 |
| 80 | 1 | 33 | 2,953 | 79 | 78 | 81 |
| 81 | 1 | 21 | 4,794 | 80 | 0 | 83 |
| 82 | 10 | 30 | 8,526 | 0 | 74 | 84 |
| 83 | 1 | 40 | 13,831 | 81 | 76 | 84 |
| 84 | 1 | 10 | 17,040 | 83 | 82 | 85 |
| 85 | 1 | 62 | 84,318 | 84 | 0 | 86 |
| 86 | 1 | 18 | 89,350 | 85 | 0 | 0 |



Следовательно, после анализа данных с помощью пяти различных методов кластеризации, можно сделать вывод, что оптимальное число кластеров составляет 8

## Использование метода к-средних для классификации объектов

В этой части исследования будет проведен анализ кластеров с использованием метода k-средних, где количество кластеров определено на основе данных, полученных из иерархического кластерного анализа. Этот подход к кластеризации включает в себя инициализацию центров кластеров и последующую классификацию объектов по близости к этим центрам в рамках установленного порога. Основные преимущества метода k-средних заключаются в его простоте применения, использовании евклидова расстояния как метрики для измерения близости и в способности к наглядному представлению результатов анализа.



Исходя из данного графика конечных центров кластеров, можно сделать вывод, что восьмикластерная модель отражает разнообразие характеристик исследуемых объектов. Видно, что некоторые кластеры имеют выраженные специфические особенности, как например, кластер 4 с высоким уровнем заболеваемости детей, что может указывать на определённые риски для здоровья в этой группе. Уровень бедности достигает максимальной отметки в кластере 2 и является наименьшим в кластерах 1 и 8, что явно демонстрирует различия в экономических условиях этих групп. Это подчеркивает экономическое неравенство между кластерами и потребность в целевых социальных интервенциях. Другие кластеры, такие как 1 и 8, кажутся более однородными и с низкими средними показателями по большинству переменных. В целом, график конечных центров кластеров демонстрирует полезность метода k-средних для выявления структур в данных и для дальнейшего изучения особенностей и потребностей различных социально-экономических групп.

|  |  |  |
| --- | --- | --- |
| **Число наблюдений в каждом кластере** | | |
| Кластеризовать | 1 | 27,000 |
| 2 | 6,000 |
| 3 | 1,000 |
| 4 | 1,000 |
| 5 | 1,000 |
| 6 | 35,000 |
| 7 | 1,000 |
| 8 | 15,000 |
| Допустимо | | 87,000 |
| Пропущенные | | ,000 |

На данной таблице отчетливо видно распределение наблюдений по восьми кластерам:

Кластер 1: Включает 27,000 наблюдений, что утверждает его как второй по величине кластер в исследовании, показывая, что значительное число объектов демонстрирует общие черты, определенные выбранными параметрами кластеризации.

Кластер 2: Состоит из 6,000 наблюдений, заметно меньше по размеру, что говорит о более специализированной группе объектов, выделенных по конкретным характеристикам.

Кластеры 3, 4 и 5: Каждый с 1,000 наблюдений являются меньшими кластерами, отображая набор объектов с возможно уникальными атрибутами, отличающимися от остальных групп.

Кластер 6: Самый большой кластер с 35,000 наблюдениями, он доминирует в выборке, что свидетельствует о наиболее широком наборе общих характеристик среди объектов этой категории.

Кластер 7: Также имеет 1,000 наблюдений и, подобно другим малым кластерам, может содержать объекты с особыми свойствами.

Кластер 8: С 15,000 наблюдениями он выделяется как третий по масштабу кластер, что подчеркивает его значимость и представительность в контексте всего исследования.

Общее количество наблюдений составляет 87,000, что указывает на обширность исследованной выборки. Распределение наблюдений показывает широкий спектр вариативности.

**Результаты дисперсионного анализа**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **ANOVA** | | | | | | |
|  | Кластеризовать | | Ошибка | | F | Знач. |
| Средний квадрат | ст.св. | Средний квадрат | ст.св. |
| детей3\_15лет\_занимающихсяспортом | 10,424 | 7 | ,165 | 79 | 63,193 | ,000 |
| Заболеваемость\_детей\_новооб | 10,639 | 7 | ,146 | 79 | 72,936 | ,000 |
| Заболев\_детей\_псих\_расст | 10,425 | 7 | ,165 | 79 | 63,236 | ,000 |
| Доля\_детейс\_IиII\_групп\_здор | 11,198 | 7 | ,096 | 79 | 116,169 | ,000 |
| Уровень\_бедности | 9,154 | 7 | ,277 | 79 | 32,991 | ,000 |
| F-критерий следует использовать только для целей описания, так как кластеры выбраны так, чтобы разница между наблюдениями в разных кластерах была максимальной. Наблюдаемые уровни значимости не скорректированы для этого, и поэтому их нельзя использовать для проверки гипотезы о равенстве средних кластеров. | | | | | | |

Исходя из результатов ANOVA, можно с сказать, что все рассмотренные переменные имеют статистически значимое влияние на исследуемые кластеры. Уровень значимости для каждой из переменных значительно превышает установленный порог, что подтверждает их влияние с меньше чем вероятностью случайного возникновения таких результатов. Кроме того, фактические значения F-критерия для каждой переменной выше установленного критического уровня, что указывает на то, что межкластерная дисперсия значимо превосходит внутрикластерную. Это свидетельствует о том, что исследуемые переменные действительно образуют четко определенные группы в рамках кластеров.

**Таблица результатов теста Краскела-Уоллиса**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Статистические критерииa,b** | | | | | |
|  | детей3\_15лет\_занимающихсяспортом | Заболеваемость\_детей\_новооб | Заболев\_детей\_псих\_расст | Доля\_детейс\_IиII\_групп\_здор | Уровень\_бедности |
| Хи-квадрат | 44,050 | 41,716 | 44,050 | 8,937 | 37,833 |
| ст.св. | 6 | 6 | 6 | 6 | 6 |
| Асимптотическая значимость | ,000 | ,000 | ,000 | ,177 | ,000 |
| a. Критерий Краскела-Уоллиса | | | | | |
| b. Группирующая переменная: Номер кластера наблюдения | | | | | |

Исходя из асимптотических значений в таблице, можно сделать вывод, что решение о разделении данных на 8 кластеров является обоснованным, так как большинство переменных показывают высокую статистическую значимость. Это свидетельствует о том, что кластеризация данных эффективно разделяет исследуемые группы по ключевым измеряемым характеристикам. Следовательно, введение 8 кластеров представляется наилучшим решением для данного анализа.

# **Список литературы**

1. Статистика: учебник и практикум для академического бакалавриата / В. С. Мхитарян [и др.] ; под ред. В. С. Мхитаряна. — М. : Издательство Юрайт, 2018. — 250 с. — (Серия : Бакалавр. Академический курс). — ISBN 978-5-9916-5591-0. — Режим доступа : <https://biblio-online.ru/viewer/C6BE26A0-4B8E-47F4-B5F0-7388CCEBF6E0/statistika-v-2-ch-chast-1#page/1>
2. Елисеева, И.И., Юзбашев, М.М. (2004). "Общая теория статистики".

<https://techlibrary.ru/b/2m1m1j1s1f1f1c1a_2q.2q.,_3m1i1b1a1z1f1c_2u.2u._2w1b2a1a2g_1t1f1p1r1j2g_1s1t1a1t1j1s1t1j1l1j._2004.pdf>

# **Используемые информационные источники**

1. Сайт Федеральной службы государственной статистики:

<https://rosstat.gov.ru>

# **Приложения**

Приложение 1

Таблица 1. Исходные данные

|  |  |  |
| --- | --- | --- |
| **Субъект РФ** | **Численность детей в возрасте 3-15 лет, систематически занимающихся физической культурой и спортом, человек** | **Заболеваемость детей злокачественными новообразованиями, человек** |
| Белгородская область | 217 077 | 333 |
| Брянская область | 111 336 | 277 |
| Владимирская область | 178 841 | 249 |
| Воронежская область | 279 146 | 310 |
| Ивановская область | 149 544 | 205 |
| Калужская область | 130 597 | 194 |
| Костромская область | 84 934 | 93 |
| Курская область | 147 350 | 247 |
| Липецкая область | 120 783 | 270 |
| Московская область | 1 141 696 | 1284 |
| Орловская область | 80 765 | 119 |
| Рязанская область | 145 123 | 186 |
| Смоленская область | 124 425 | 152 |
| Тамбовская область | 130 699 | 180 |
| Тверская область | 148 699 | 266 |
| Тульская область | 172 222 | 270 |
| Ярославская область | 151 882 | 312 |
| г.Москва | 1 398 063 | 2641 |
| Республика Карелия | 86 309 | 129 |
| Республика Коми | 114 648 | 213 |
| Архангельская область | 161 877 | 265 |
| Ненецкий авт. округ | 7 275 | 13 |
| Архангельская область без авт.оруга | 154 602 | 252 |
| Вологодская область | 174 982 | 286 |
| Калининградская область | 146 987 | 220 |
| Ленинградская область | 237 480 | 200 |
| Мурманская область | 111 074 | 187 |
| Новгородская область | 77 817 | 132 |
| Псковская область | 86 876 | 122 |
| г.Санкт-Петербург | 680 647 | 1212 |
| Республика Адыгея | 68 671 | 44 |
| Республика Калмыкия | 44 265 | 37 |
| Республика Крым | 276 700 | 430 |
| Краснодарский край | 814 730 | 1011 |
| Астраханская область | 188 250 | 226 |
| Волгоградская область | 312 591 | 462 |
| Ростовская область | 561 652 | 570 |
| г. Севастополь | 64 341 | 89 |
| Республика Дагестан | 344 966 | 844 |
| Республика Ингушетия | 66 341 | 229 |
| Кабардино-Балкарская Республика | 139 743 | 204 |
| Карачаево-Черкесская Республика | 69 590 | 93 |
| Республика Северная Осетия-Алания | 101 774 | 160 |
| Чеченская Республика | 418 714 | 487 |
| Ставропольский край | 427 791 | 503 |
| Республика Башкортостан | 639 040 | 958 |
| Республика Марий Эл | 98 305 | 136 |
| Республика Мордовия | 84 854 | 170 |
| Республика Татарстан | 569 950 | 838 |
| Удмуртская Республика | 243 311 | 326 |
| Чувашская республика | 130 360 | 242 |
| Пермский край | 359 872 | 813 |
| Кировская область | 193 121 | 231 |
| Нижегородская область | 420 541 | 631 |
| Оренбургская область | 302 335 | 373 |
| Пензенская область | 166 621 | 252 |
| Самарская область | 465 663 | 731 |
| Саратовская область | 325 600 | 458 |
| Ульяновская область | 137 755 | 214 |
| Курганская область | 129 934 | 156 |
| Свердловская область | 679 578 | 927 |
| Тюменская область в т.ч. | 692 087 | 873 |
| Ханты-Мансийский авт. округ - Югра | 310 037 | 368 |
| Ямало-Ненецкий авт. округ | 103 089 | 104 |
| Тюменская область без авт.округов | 278 961 | 401 |
| Челябинская область | 560 901 | 774 |
| Республика Алтай | 40 315 | 47 |
| Республика Тыва | 77 588 | 67 |
| Республика Хакасия | 91 208 | 95 |
| Алтайский край | 343 582 | 456 |
| Красноярский край | 458 641 | 565 |
| Иркутская область | 304 593 | 597 |
| Кемеровская область - Кузбасс | 414 872 | 483 |
| Новосибирская область | 369 126 | 543 |
| Омская область | 244 271 | 427 |
| Томская область | 142 395 | 214 |
| Республика Бурятия | 164 510 | 272 |
| Республика Саха (Якутия) | 188 643 | 236 |
| Забайкальский край | 253 104 | 260 |
| Камчатский край | 44 204 | 67 |
| Приморский край | 246 617 | 354 |
| Хабаровский край | 181 984 | 300 |
| Амурская область | 108 692 | 147 |
| Магаданская область | 14 135 | 11 |
| Сахалинская область | 72 330 | 122 |
| Еврейская авт. область | 28 086 | 46 |
| Чукотский авт. округ | 9 648 | 8 |

|  |  |  |
| --- | --- | --- |
| **Заболеваемость детей психическими расстройствами и расстройствами поведения, человек** | **Доля детей с I и II группой здоровья, %** | **Уровень бедности, %** |
| 991 | 81,1 | 6,1 |
| 1628 | 84,3 | 12,0 |
| 2146 | 86,9 | 10,4 |
| 3316 | 79,9 | 7,2 |
| 554 | 82,8 | 11,5 |
| 79 | 86,8 | 7,9 |
| 577 | 80,5 | 11,0 |
| 1666 | 80,4 | 8,1 |
| 706 | 82,8 | 7,2 |
| 9192 | 88,2 | 5,5 |
| 829 | 84,1 | 11,5 |
| 694 | 79,3 | 11,8 |
| 756 | 74,1 | 12,8 |
| 648 | 88,1 | 10,5 |
| 1481 | 84,4 | 9,6 |
| 991 | 85,7 | 9,4 |
| 1677 | 76,6 | 8,8 |
| 8734 | 86,5 | 5,0 |
| 639 | 86,0 | 12,9 |
| 414 | 86,7 | 14,0 |
| 836 | 128,0 | 11,8 |
| 61 | 51,4 | 9,0 |
| 775 | 76,6 | 10,9 |
| 1444 | 83,8 | 11,0 |
| 779 | 80,7 | 12,0 |
| 1276 | 88,2 | 7,0 |
| 571 | 76,7 | 8,0 |
| 1188 | 93,3 | 11,9 |
| 748 | 89,4 | 14,3 |
| 10509 | 78,6 | 4,5 |
| 626 | 94,9 | 11,6 |
| 461 | 83,2 | 20,9 |
| 4272 | 77,9 | 14,5 |
| 2895 | 85,5 | 9,0 |
| 1014 | 74,7 | 13,8 |
| 1019 | 88,8 | 9,6 |
| 3644 | 86,5 | 10,5 |
| 406 | 80,7 | 9,0 |
| 11200 | 78,9 | 14,2 |
| 1766 | 69,4 | 30,5 |
| 1133 | 89,2 | 16,4 |
| 514 | 90,0 | 22,0 |
| 695 | 84,3 | 13,0 |
| 1362 | 87,5 | 19,4 |
| 2911 | 85,3 | 11,7 |
| 5610 | 85,8 | 10,1 |
| 752 | 79,9 | 16,2 |
| 667 | 86,9 | 15,0 |
| 5185 | 87,2 | 5,0 |
| 1543 | 84,7 | 10,2 |
| 1024 | 86,1 | 14,8 |
| 3639 | 82,4 | 11,5 |
| 1352 | 82,6 | 12,5 |
| 4423 | 81,3 | 8,0 |
| 2157 | 76,3 | 13,1 |
| 1497 | 78,0 | 11,5 |
| 3199 | 84,7 | 10,9 |
| 4785 | 77,4 | 13,2 |
| 1178 | 78,3 | 12,8 |
| 738 | 82,2 | 17,0 |
| 5107 | 78,2 | 8,3 |
| 4120 | 260,4 | 10,7 |
| 1597 | 87,0 | 7,2 |
| 679 | 87,3 | 4,5 |
| 1844 | 86,1 | 12,3 |
| 6723 | 76,8 | 11,3 |
| 371 | 87,3 | 20,0 |
| 265 | 91,0 | 27,2 |
| 1159 | 87,3 | 18,0 |
| 3710 | 88,9 | 15,4 |
| 4402 | 82,6 | 14,4 |
| 4950 | 82,3 | 14,8 |
| 4243 | 84,2 | 11,3 |
| 4982 | 85,0 | 11,3 |
| 3376 | 86,1 | 12,0 |
| 2173 | 76,3 | 12,7 |
| 1456 | 91,2 | 19,0 |
| 1282 | 89,0 | 15,5 |
| 710 | 86,6 | 17,6 |
| 437 | 88,6 | 12,8 |
| 3781 | 87,5 | 11,4 |
| 1533 | 86,7 | 10,5 |
| 1348 | 82,4 | 13,3 |
| 184 | 95,2 | 7,4 |
| 848 | 88,4 | 7,0 |
| 262 | 84,0 | 20,0 |
| 139 | 68,2 | 6,6 |

Приложение 2

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Жесткие выбросы | | | | |
|  |  |
|  | **Количество детей, занимающихся спортом** | **Злокачественные новообразования** | **Психические расстройства** | **I и II группа здоровья** | **Уровень бедности, %** |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
| Краснодарский край | 0 | 0 | 0 | 0 | 0 |  |
| Республика Дагестан | 0 | 0 | 0 | 0 | 0 |  |
| г.Санкт-Петербург | 0 | 0 | 0 | 0 | 0 |  |
| Московская область | 1 | 0 | 0 | 0 | 0 |  |
| Чукотский авт. округ | 0 | 0 | 0 | 0 | 0 |  |
| Республика Тыва | 0 | 0 | 0 | 0 | 0 |  |
| Республика Ингушетия | 0 | 0 | 0 | 0 | 0 |  |
| Архангельская область | 0 | 0 | 0 | 1 | 0 |  |
| Ненецкий авт. округ | 0 | 0 | 0 | 1 | 0 |  |
| Тюменская область в т.ч. | 0 | 0 | 0 | 1 | 0 |  |
| г.Москва | 1 | 1 | 0 | 0 | 0 |  |

Приложение 3

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Мягкие выбросы | | | | |
|  |  |
|  | **Количество детей, занимающихся спортом** | **Злокачественные новообразования** | **Психические расстройства** | **I и II группа здоровья** | **Уровень бедности, %** |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
| Краснодарский край | 1 | 0 | 0 | 0 | 0 |  |
| Республика Дагестан | 0 | 0 | 1 | 0 | 0 |  |
| г.Санкт-Петербург | 0 | 1 | 1 | 0 | 0 |  |
| Московская область | 0 | 1 | 1 | 0 | 0 |  |
| Чукотский авт. округ | 0 | 0 | 0 | 1 | 0 |  |
| Республика Тыва | 0 | 0 | 0 | 0 | 1 |  |
| Республика Ингушетия | 0 | 0 | 0 | 1 | 1 |  |
| Архангельская область | 0 | 0 | 0 | 0 | 0 |  |
| Ненецкий авт. округ | 0 | 0 | 0 | 0 | 0 |  |
| Тюменская область в т.ч. | 0 | 0 | 0 | 0 | 0 |  |
| г.Москва | 0 | 0 | 1 | 0 | 0 |  |

Приложение 4

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Правило 3 сигм | | | | |
|  |  |
|  | **Количество детей, занимающихся спортом** | **Злокачественные новообразования** | **Психические расстройства** | **I и II группа здоровья** | **Уровень бедности, %** |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 0 | 0 |  |
| Краснодарский край | 0 | 0 | 0 | 0 | 0 |  |
| Республика Дагестан | 0 | 0 | 1 | 0 | 0 |  |
| г.Санкт-Петербург | 0 | 0 | 1 | 0 | 0 |  |
| Московская область | 1 | 0 | 1 | 0 | 0 |  |
| Чукотский авт. округ | 0 | 0 | 0 | 0 | 0 |  |
| Республика Тыва | 0 | 0 | 0 | 0 | 1 |  |
| Республика Ингушетия | 0 | 0 | 0 | 0 | 1 |  |
| Архангельская область | 0 | 0 | 0 | 0 | 0 |  |
| Ненецкий авт. округ | 0 | 0 | 0 | 0 | 0 |  |
| Тюменская область в т.ч. | 0 | 0 | 0 | 1 | 0 |  |
| г.Москва | 1 | 1 | 0 | 0 | 0 |  |

Приложение 6

Приложение 7

Красным цветом выделены полученные выбросы z-score, I колонка отвечает за I признак

