

Национальный исследовательский университет «Высшая школа экономики»
Факультет компьютерных наук
Образовательная программа: Программная инженерия

Отчет по Домашнему заданию №2
«Двумерные данные»
по майнору «Прикладной статистический анализ»

«Статистические методы анализа двумерных данных на
примере исследования зависимости количества
зарегистрированных преступлений в субъектах Российской
Федерации от средних расходов на алкогольную продукцию,
табачные изделия и наркотики за 2022 год»

Работу выполнил:
студент 2 курса
Артемьев Александр Вячеславович
Преподаватель:
Грачёва Светлана Сергеевна

Москва, 2023г.

Содержание

Оглавление

1. Введение.....	3
1.1 Актуальность темы исследования	3
1.2 Обоснование выбора темы	3
1.3 Цели исследования	3
1.4 Задачи исследования.....	3
1.5 Гипотезы.....	4
2. Используемые показатели для анализа.....	5
2.1 Список используемых показателей для анализа	5
2.2 Описание показателей, выбранных для исследования	5
3. Выбор актуальных данных.....	5
4. Исследование выбранных данных на выбросы	6
5. Корреляционный анализ.....	7
6. Регрессионный анализ	10
6.1 Линейная модель	10
6.2 Гиперболическая модель	12
6.3 Степенная модель.....	14
6.4 Сравнение построенных моделей.....	15
7. Общие выводы по работе.....	16
8. Список литературы	16
9. Используемые информационные источники	16
10. Приложения.....	17

1. Введение

1.1 Актуальность темы исследования

С течением времени изменяется уровень преступности и появляются новые виды правонарушений, что оказывает огромное влияние на различные сферы общества:

- Заболеваемость и смертность населения
- Экономические кризисы и спад производства
- Материальный и моральный вред, причиненный обществу
- Падение качества жизни населения

Исследование зависимостей между социально-экономическими показателями и уровнем преступности в регионах Российской Федерации является актуальным и необходимым, ведь нахождение факторов, объясняющих целую структуру правонарушений, поможет разработать эффективные меры по снижению количества преступлений. Что напрямую окажет положительное влияние на все общество в целом. Исследование проводится на данных за 2022 год, что добавляет актуальности работе.

1.2 Обоснование выбора темы

Такое исследование может быть полезным для разработки государственной политики в области контроля за распространением алкоголя, табака и наркотиков, а также для оптимизации деятельности правоохранительных органов в различных регионах России. Статистический анализ данных позволит выявить возможные связи между рассматриваемыми показателями. Данная тема имеет практическую значимость и может быть полезной для различных заинтересованных сторон, включая правоохранительные органы, государственные органы и общественные организации.

1.3 Цели исследования

- Выявление возможной связи между уровнем потребления алкоголя, табака и наркотиков и количеством зарегистрированных преступлений в субъектах Российской Федерации.

1.4 Задачи исследования

- Выбрать актуальные данные по уровню преступности в субъектах РФ
- Выбрать актуальные данные по расходам на алкоголь, табачные изделия и наркотики в субъектах РФ
- Найти аномальные значения в выбранных данных
- Провести корреляционный анализ, для установления характера связи признаков
- Построить двумерные регрессионные статистические модели и выявить оптимальную, найти МНК-оценки для построенных моделей.
- Сделать вывод по проведенному исследованию, обобщить результаты, подтвердить или опровергнуть гипотезы, сформулированные в начале работы.

1.5 Гипотезы

- Существует сильная положительная связь между признаками
- В наиболее развитых субъектах, исследуемая связь будет меньше,
- Увеличение потребления алкоголя, табака и наркотиков может привести к увеличению преступности.
- Линейная модель не будет оптимальной из рассматриваемых в работе

2. Используемые показатели для анализа

2.1 Список используемых показателей для анализа

В данном исследовании используются два признака:

1. Средние расходы на алкогольную продукцию, табачные изделия и наркотики, за 2022 год, % (независимая переменная)
2. Зарегистрировано преступлений на 10000 чел., за 2022 год, (результативный)

2.2 Описание показателей, выбранных для исследования

Выбранный показатель в качестве независимой переменной – средние расходы на алкоголь, табак и наркотики.

- Показатель характеризует какой процент расходов уходит на алкоголь, табачные изделия и наркотики в среднем по населению субъекта.
- Единица измерения – средний процент расходов на алкоголь, табак и наркотики в субъекте РФ.
- Количественные дискретные данные

Выбранный показатель в качестве результативного признака – уровень преступности.

- Показатель характеризует количество документально зарегистрированных преступлений на 10000 человек населения в данном субъекте РФ.
- Единица измерения - количество преступлений на 10000 человек населения рассматриваемого субъекта.
- Количественные дискретные данные.

Было принято решение использовать именно этот относительный, а не абсолютный, показатель для исследования уровня преступности. Для абсолютного показателя – количество преступлений, проводимое исследование нецелесообразно, так как для двух субъектов с существенным отличием в населении, существенно бы отличался уровень преступности, а зависимый признак мог быть одинаковым, так как он относительный и измеряется в процентах.

3. Выбор актуальных данных

К сожалению, не удалось найти открытых источников с актуальными данными по уровню преступности на 10000 человек, поэтому было принято решение взять уже рассматриваемые ранее данные с сайта государственной прокуратуры Российской Федерации[1] по количеству преступлений и данные с сайта Федеральной службы государственной статистики по численности населения[3] и объединить их в таблицу 1(Приложение 1) , затем заполнить столбец таблицы искомого признака. Для этого достаточно для каждого субъекта подставить значения в формулу:

$$X = \frac{\text{Количество преступлений}}{\text{Население субъекта}} * 10000$$

После чего мы формируем таблицу 2 со значениями нашего признака – уровня преступности и со значениями результирующего признака ([Приложение 2](#)). Данные результирующего признака взяты с сайта Федеральной службы государственной статистики[2] по доходам, расходам и потреблению домашних хозяйств в 2022 году .

4. Исследование выбранных данных на выбросы

Первым этапом исследования на выбросы было построение корреляционного облака (рис.1) и его анализ.

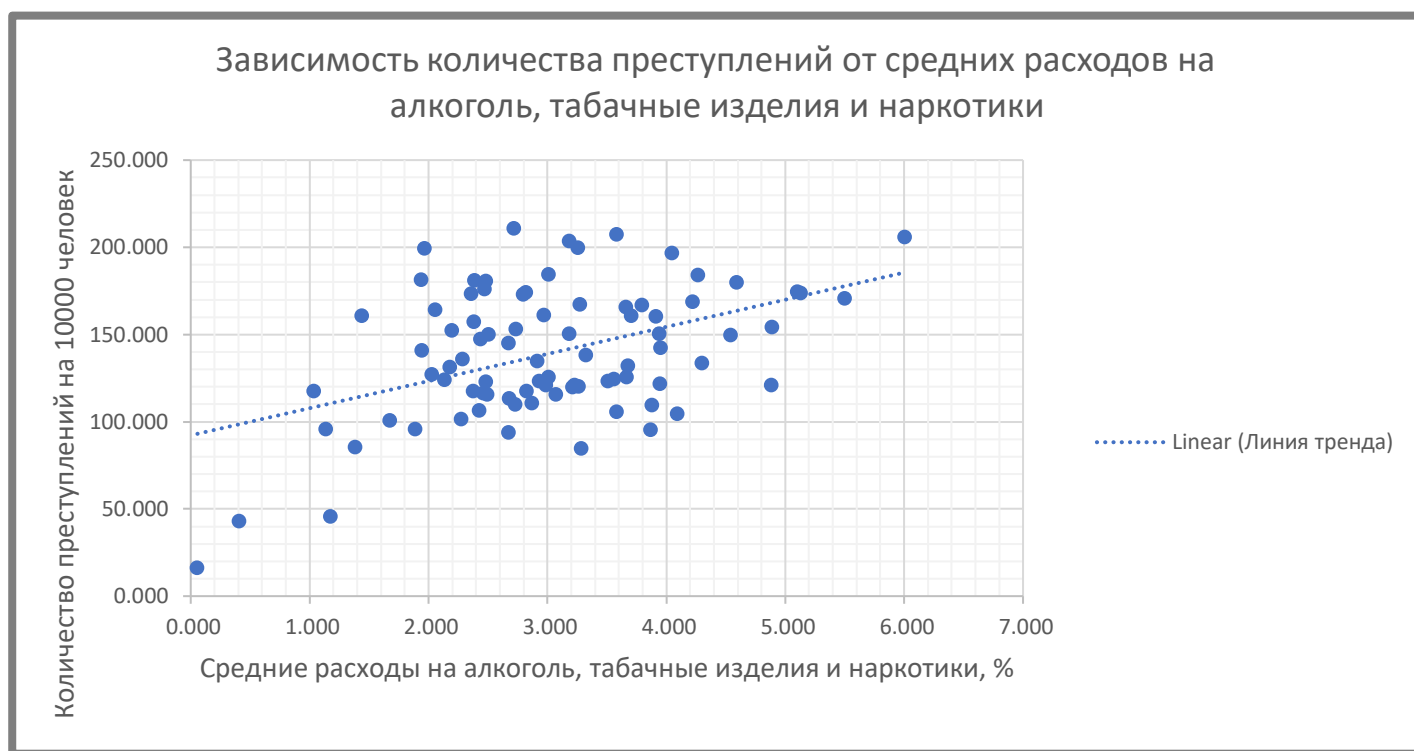


Рисунок 1. Корреляционное облако

Визуализируя исходные данные в виде точечной диаграммы, становится легче их анализировать. По первоначальному анализу рис. 1 без коэффициента корреляции можно сделать следующий вывод: связь между коэффициентами определенно есть, но эта слабая связь. Примерно 60 % точек сконцентрировано вблизи линии тренда, и значительная часть наблюдений хаотично располагается на графике, множество точек беспорядочно распределены в прямоугольнике с расходами от 2 до 4 % и количеством преступлений от 100 до 200 на 10000 чел.

Было принято решение избавиться от выбросов, рассматривая субъекты по федеральным округам. Для этого были сгруппированы субъекты РФ по

округам и составлены таблицы 3 – 10 ([Приложение 3](#)) . В каждой таблице представлена информация о субъектах в рассматриваемом федеральном округе его значении уровня преступности в 1 столбце и значение расходов во 2 столбце. Значение не из таблицы, выделенное цветом – коэффициент корреляции.

Далее я исключил из дальнейшего рассмотрения округа с очень слабой корреляцией (коэффициент корреляции <0.1). Отсутствие связи с изучаемым признаком может возникнуть по нескольким причинам.

1) На преступность влияет очень много факторов: высокий уровень безработицы, расслоение населения по имущественному признаку, сохранения значительного числа негативных фоновых явлений, в частности, проституции, бродяжничества, игромании, алкоголизма; интенсивная миграция и другие. Поэтому в крупнейших федеральных округах: Центральном и Северо-Западном влияние рассматриваемого коэффициента может не быть так сильно заметно, о чем говорит коэффициент корреляции близкий к 0, поэтому нет смысла рассматривать эти округа в дальнейшем.

2) Также округу может быть свойственна преступность, не зависящая от рассматриваемого фактора, так, например, в Сибирском округе имеет большой удельный вес коррупционная преступность и преступность экономической направленности, налоговые преступления, преступления в финансовой кредитной сфере.

Сибирский, Центральный, Северо-Западный федеральные округа в дальнейшем не будут рассматриваться. Итоговые данные сформированы в таблице 11([Приложение 4](#)).

5. Корреляционный анализ

Для проведения корреляционного анализа, сначала построим корреляционное облако и исследуем его (рис. 2)

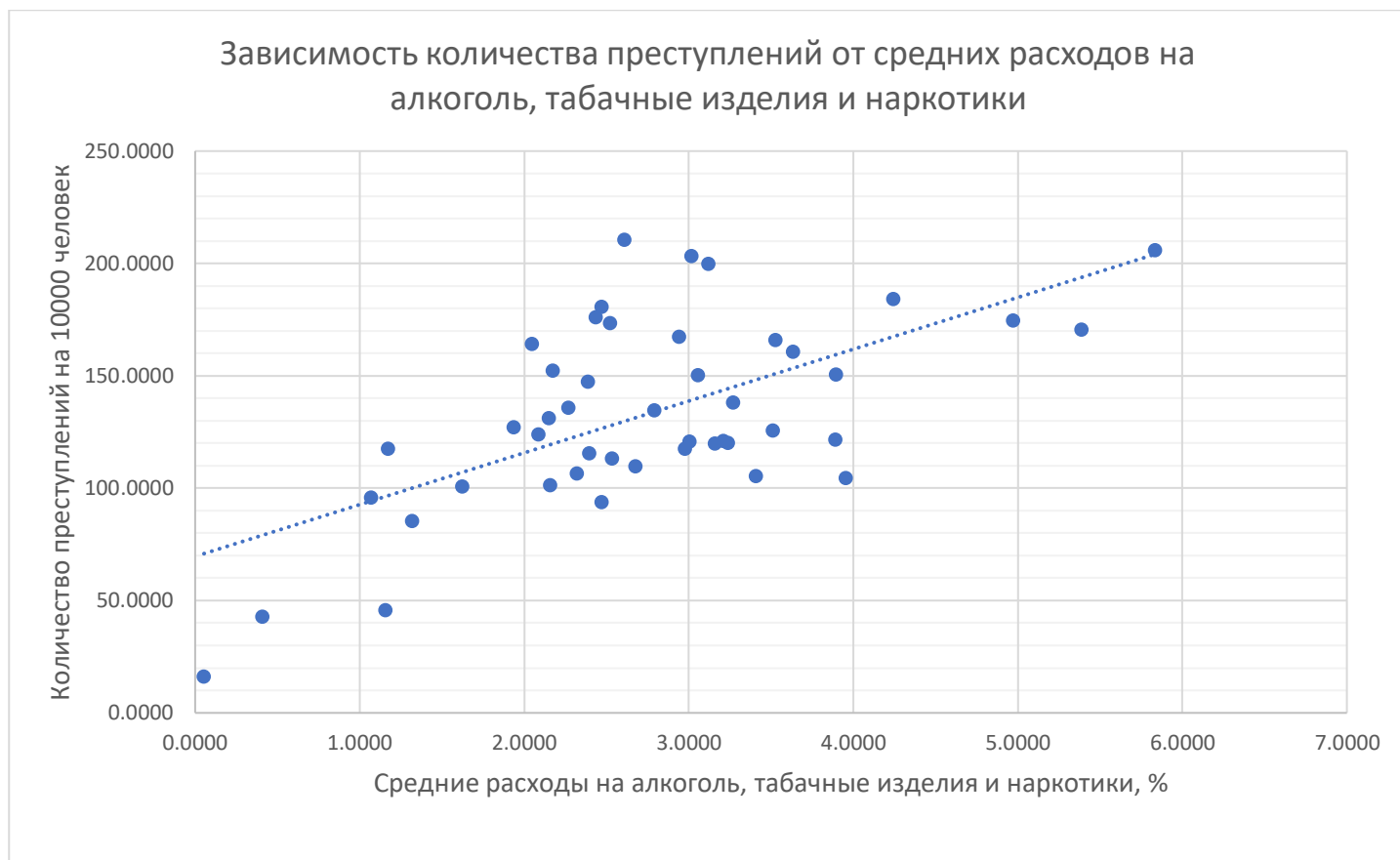


Рисунок 2. Корреляционное облако, для новых данных

Здесь намного сильнее заметна связь между показателями, наблюдения сконцентрированы вдоль линии тренда, понятно, что есть точки, которые выбиваются из тренда, что может нам говорить не о сильной, скорее о средней связи, и корреляционный коэффициент должен быть около 0,65 - 0,75, причем положительный (следует из направления линии тренда). После избавления от выбросов мы пришли от хаотично разбросанного набора точек, к структурированному массиву наблюдений со средней связью. Следующим шагом найдем коэффициент корреляции. Для этого построим таблицу со вспомогательными данными ([Приложение 5](#)), затем воспользуемся формулой

Формула 2. Среднеквадратическое отклонение

$$S = \sqrt{D} = \sqrt{\overline{x^2} - \bar{x}^2}$$

и найдем среднеквадратическое отклонение по каждой переменной X (средние расходы) и Y (уровень преступности).

Получим значения соответственно $S_x = 1,133$ и $S_y = 41,439$. Осталось найти сам коэффициент корреляции по формуле 3. $r = 0,63$.

Формула 3. Коэффициент корреляции

$$r = \frac{\overline{xy} - \bar{x} * \bar{y}}{S_x * S_y}$$

Теперь проверим значимость (для уровня значимости 0.05) найденного коэффициента. Сформулируем две гипотезы.

H0 – связь между признаками X и Y отсутствует

H1 – связь между признаками X и Y присутствует

Для проверки гипотезы найдем $t_{\text{набл.}}$. По формуле 4:

Формула 4. Расчетное значение t-критерия Стьюдента

$$t_{\text{набл.}} = r * \sqrt{\frac{n - k - 1}{1 - r^2}}$$

Где $n = 46$ – количество наблюдений, $k = 1$ - количество факторных признаков, число степеней свободы $= n - k - 1 = n - 2$. Получаем расчетное значение t – критерия $= 5,38$. Табличное значение t -критерия при уровне значимости $= 0.05$ и числе степеней свободы $= 44$ можно найти в файле с математико-статистическими таблицами [4]. Стоит заметить, что при увеличении числа степеней свободы табличное значение критерия уменьшается, то есть $t(40, 0.05) > t(44, 0.05) > t(60, 0.05)$. Значит табличное значение зажато между числами: $2.021 > t(44, 0.05) > 2.000$. Отсюда следует что расчетное значение t – критерия оказывается больше табличного и коэффициент корреляции признается значим.

Выводы:

Построив диаграмму рассеивания, мы сделали предположение о наличии средней связи между двумя величинами. Коэффициент оказался чуть меньше, чем в предположении, но раз $0.5 < r < 0.7$ наблюдается положительная средняя связь между рассматриваемыми показателями. После удостоверились в том, что найденный коэффициент значим с вероятностью $0.95(1 - 0.05)$. Если понизить уровень значимости до 0.001, то наш коэффициент $= 5.38 > 3.551$, то есть с вероятностью 0.999 коэффициент корреляции будет значим, на практике видимо почти всегда.

Результаты о наличии корреляционной связи между признаками приводят к необходимости более подробно изучить эту связь и построить модель аппроксимации.

Уровень преступности действительно зависит от уровня расходов, выделяемых на алкоголь, табачные изделия и наркотики причем довольно существенно раз связь средняя. Но видимо есть множество других факторов, которые влияют на уровень преступности, раз связь не сильная.

6. Регрессионный анализ

6.1 Линейная модель

Для построения линейной регрессионной модели, используя те же обозначения X для независимой переменной и Y для результативного признака.

1. Найдем средние значения для X и Y : $X_{\text{ср.}} = 2.75$ и $Y_{\text{ср.}} = 133.004$.
2. После чего вычислим вспомогательные данные столбцы таблицы 13 ([Приложение 6](#)) $X*Y$, $X-X_{\text{ср.}}$, $(X-X_{\text{ср.}})^2$.
3. Найдем $(XY)_{\text{ср.}} = 385.319$ и $X_{\text{ср.}} * Y_{\text{ср.}} = 365.715$.
4. Последним шагом нашли дисперсию как среднее значение столбца $(X-X_{\text{ср.}})^2 = 1.284$.

Следующий этап – расчет коэффициентов регрессии по формулам 5 и 6:

Формула 6. Старший коэффициент уравнения регрессии

$$b_1 = \frac{\overline{xy} - \bar{x}\bar{y}}{S_x^2}$$

Формула 5. Свободный член уравнения линейной модели

$$b_0 = \bar{y} - b_1 \bar{x}$$

Получили $b_1 = 23.049$ и $b_0 = 69.628$. Уравнение модели имеет следующий вид:

$$\hat{y} = b_0 + b_1 x = 69.628 + 23.049x$$

Уравнение 1. Линейная регрессионная модель

Коэффициент детерминации - доля дисперсии зависимой переменной, объясняемая рассматриваемой моделью. Для модели линейной регрессии с одним признаком x коэффициент детерминации равен квадрату обычного коэффициента корреляции между y и x . $R^2 = r^2 = 0.63 * 0.63 = 0.397$. Коэффициент детерминации является показателем устойчивости модели, чем он выше, тем лучше модель.

Свободный член $b_0 = 69.628$ показывает средние расходы на алкоголь, табачные изделия и наркотики при уровне преступности = 0. Значения коэффициента b_1 показывает на сколько единиц в среднем изменится уровень преступности (Y) при изменении средних расходов (X) на 1%.

Для построения линейной модели рассчитываем столбец $Y_{\text{регр}}$. В таблице со вспомогательными данными ([Приложение 6](#)). Построенная модель представлена на рисунке 3.

Модель линейной регрессии для аппроксимации зависимости уровня преступности от расходов на алк., таб. изд., нарк.

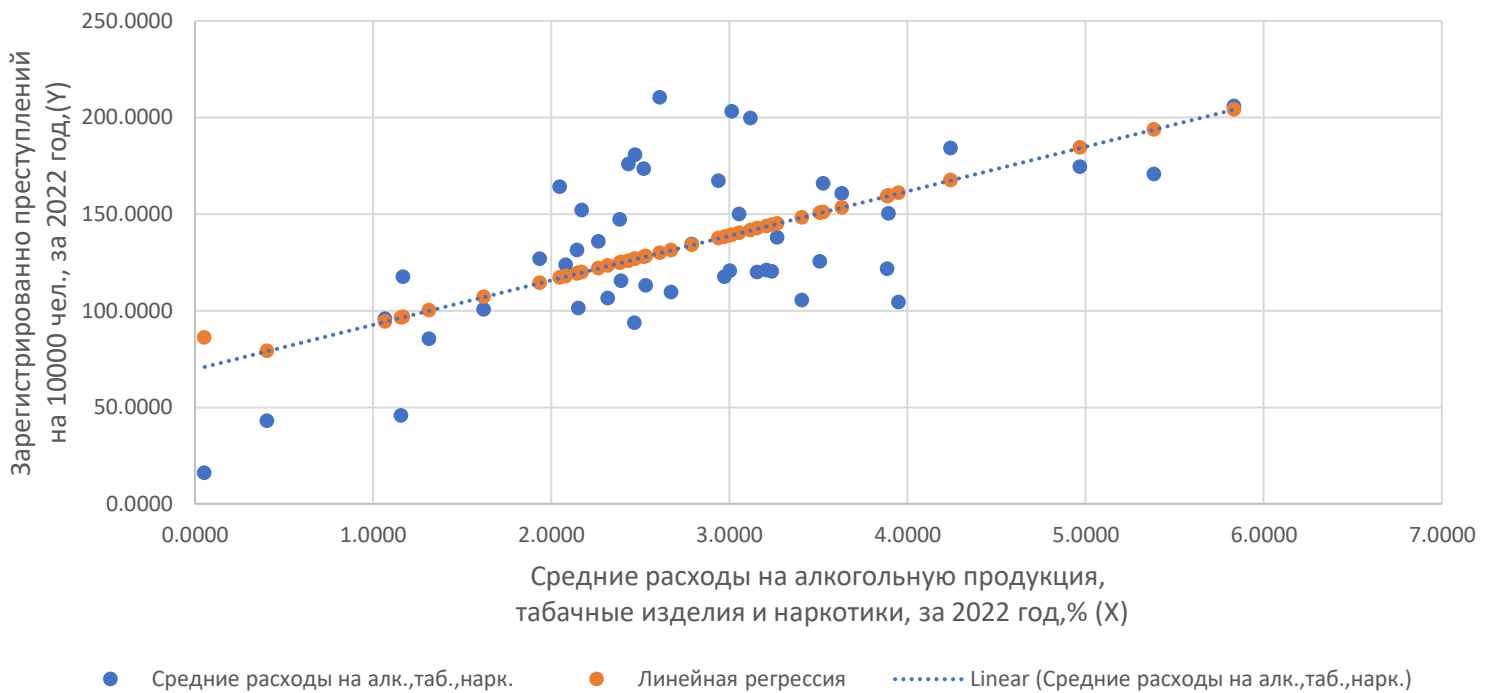


Рисунок 3. Линейная регрессионная модель

Следующим шагом рассчитываем МНК-оценку по формуле 7:

$$S = \sqrt{\frac{\sum_{i=1}^N (y_i - y_{\text{перп.}})^2}{N}}$$

Формула 7. МНК-оценка

$S = 32.805$. На данном этапе трудно сделать вывод по найденной МНК - оценке, так как она напрямую зависит от значений y , то есть если бы мы все y умножили на константу, то S тоже умножилась на эту константу, таким образом даже если бы линия достаточно хорошо аппроксимировала данные, но данные имели большое значение мы получили такое же значение коэффициента. Следующий этап – проверка значимости уравнения регрессии. Формулируем гипотезы:

H_0 – уравнение регрессии не значимо

H_1 – уравнение регрессии значимо

Определим расчетный F-критерий Фишера по следующей формуле:

Формула 8. F-критерий Фишера

$$F = \frac{r_{xy}^2}{1 - r_{xy}^2} * (n - 2) = \frac{0.397}{1 - 0.397} * 44 = 29.013$$

Теперь находим t критическое в таблице распределения Фишера-Снедекора, для 46 (число степеней свободы = 44) элементов выборки оно

будет зажато между: $t(30) > t(44) > t(60)$, то есть $4.17 > t(44) > 4$. Значит для F – критерия расчетного = 29.013 > 4.17, уравнение регрессии значимо с вероятностью ошибки 0.05 – уровень значимости.

Проверка значимости коэффициента уравнения регрессии. Формулируем гипотезы:

H0 – коэффициент регрессии не значим

H1 – коэффициент регрессии значим

Мы можем найти $t_{\text{набл.}}$ по следующей формуле:

Формула 9. t-критерий расчетный

$$t_{\text{набл.}} = \frac{|P|}{S_p}$$

Где P значение параметра, а S_p – стандартное отклонение параметра. Но на самом деле мы получим такое же значение, что при расчете $t_{\text{набл.}}$ для проверки значимости коэффициента корреляции = 5.386. $t_{\text{крит.}}$ будет также находится по таблице Стьюдента и будет зажато между 2.021 и 2, где каждое значение меньше 5.386. Приходим к выводу что коэффициент регрессии значим.

Вывод по линейной модели:

Модель линейной регрессии слабо приближает реальные данные. Это можно понять как из построенной точечной диаграммы, где множество наблюдений находятся далеко от линии тренда, так и с помощью коэффициента детерминации = $0.397 < 0.5$. Модель считается достаточно хорошей при коэффициенте > 0.8 . У нас же коэффициент почти в два раза меньше допустимой нижней границы. Если бы линейная модель достаточно точно аппроксимировала наши данные, это означало бы, что между X и Y линейная зависимость и коэффициент корреляции был бы достаточно высокий 0.85 - 1, наш же коэффициент 0.63. Как само уравнение регрессии так и его коэффициент оказались значимы.

6.2 Гиперболическая модель

Для построения гиперболической модели, понадобилось рассчитать определенные вспомогательные данные ([Приложение 7](#)) в следующем порядке:

1. Сначала выполнили замену $z = 1/x$, то есть задача свелась к нахождению линейной регрессии для Z и Y.
2. Затем нашли средние значения Z и Y, которые равны соответственно 0.843 и 133.004
3. Построили столбец ZY и соответственно нашли среднее значений этого столбца = 58.509.
4. Зная $Z_{\text{ср.}}$, построили столбцы $(Z - Z_{\text{ср.}})$ и $(Z - Z_{\text{ср.}})^2$.
5. Последним шагом нашли дисперсию = 7.159 и произведение $Z_{\text{ср.}} * Y_{\text{ср.}} = 112.079$

Следующий этап - расчет коэффициентов регрессии, используя формулы 10 и 11:

Формула 10. Старший коэффициент при уравнении гиперболической модели

$$b_0 = \bar{y} - b_1 \bar{z}$$

Формула 11. Свободный член в уравнение гиперболической модели

$$b_1 = \frac{\overline{zy} - \bar{z}\bar{y}}{S_z^2}$$

Получили $b_1 = -7.483$ и $b_0 = 139.309$. Коэффициент детерминации также равен квадрату коэффициенту корреляции $= 0,48315^2 = 0,233$. (коэффициент корреляции рассчитывается для столбцов Z и Y) После чего выполним обратную замену $X = 1/Z$ и получим уравнение гиперболической модели Уравнение модели имеет следующий вид:

Уравнение 2. Уравнение гиперболической модели

$$\hat{y} = b_0 + \frac{b_1}{x} = 139,309 - \frac{7,483}{x}$$

На рисунке 4 отображена гиперболическая регрессионная модель:

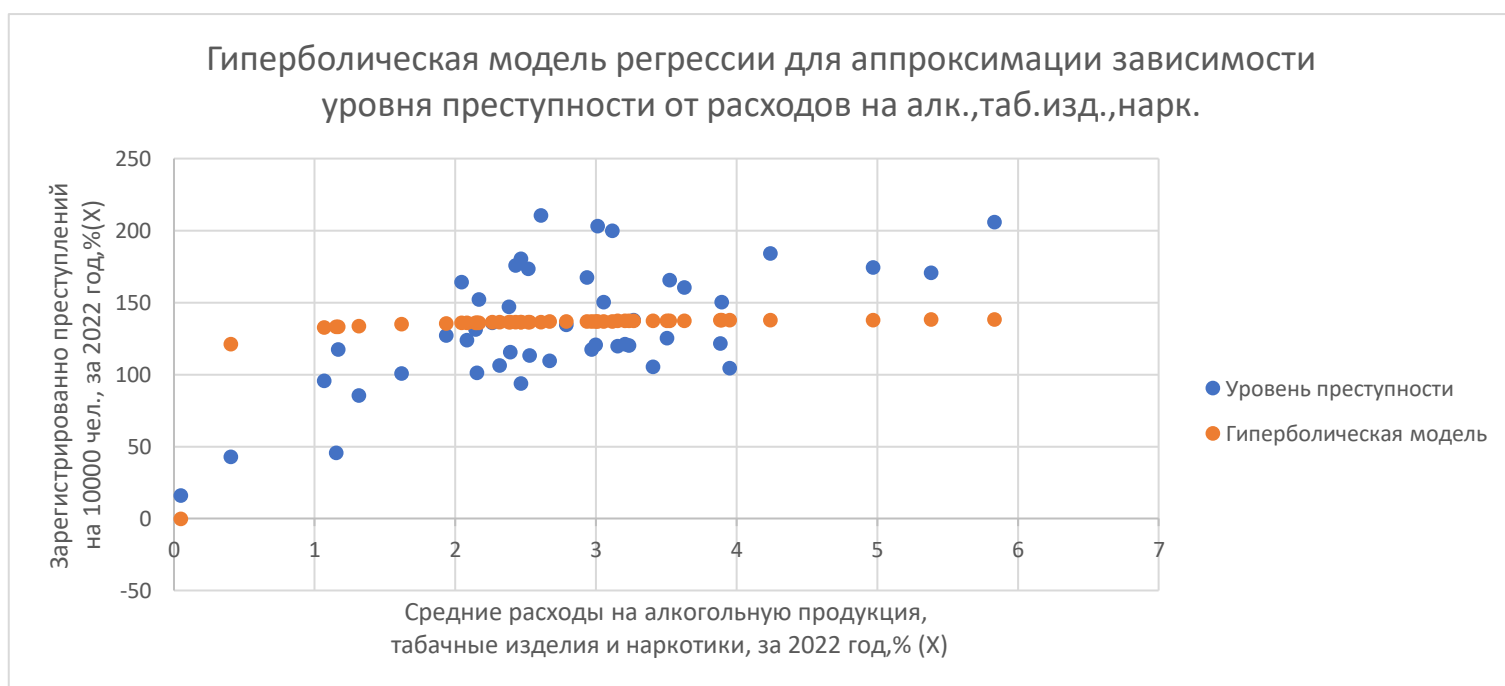


Рисунок 4. Гиперболическая регрессионная модель

Далее вычислим МНК-оценку по формуле 12, $S = 36,281$

Формула 12. МНК-оценка

$$S = \sqrt{\frac{\sum_{i=1}^N (y_i - y_{\text{перп.}})^2}{N}}$$

6.3 Степенная модель

Для построения степенной модели, понадобилось рассчитать определенные вспомогательные данные ([Приложение 8](#)) в следующем порядке:

1. Сначала выполнили замену $x_1 = \log x$ и $y_1 = \log y$, то есть задача свелась к нахождению линейной регрессии для X_1 и Y_1 .
2. Затем нашли средние значения X_1 и Y_1 , после чего построили столбец $X_1 Y_1$ и соответственно нашли среднее значений этого столбца.
3. Зная $X_1 \text{ср.}$ Построили столбцы $(X_1 - X_1 \text{ср.})$ и $(X_1 - X_1 \text{ср.})^2$.
4. Последним шагом нашли дисперсию и произведение $X_1 \text{ср.} * Y_1 \text{ср.}$.

Следующий этап - расчет коэффициентов регрессии, используя формулы 13 и 14:

Формула 13. Старший коэффициент в уравнении регрессионной модели

$$b_1 = \frac{\overline{x_1 y_1} - \overline{x_1} * \overline{y_1}}{S_{x_1}^2}$$

Формула 14. Свободный член в уравнении регрессии

$$b_0 = 10^{\overline{y_1} - b_1 \overline{x_1}}$$

Получили $b_1 = 0.522$ и $b_0 = 78.873$. Коэффициент детерминации также равен квадрату коэффициенту корреляции $= 0,8529^2 = 0,727$. (коэффициент корреляции рассчитывается для столбцов X_1 и Y_1) После чего выполним обратную замену $X = 10^{(X_1)}$ и $Y = 10^{(Y_1)}$ и получим уравнение степенной модели, рисунок 6:

Степенная модель регрессии для аппроксимации зависимости уровня преступности от расходов на алк., таб. изд., нарк.

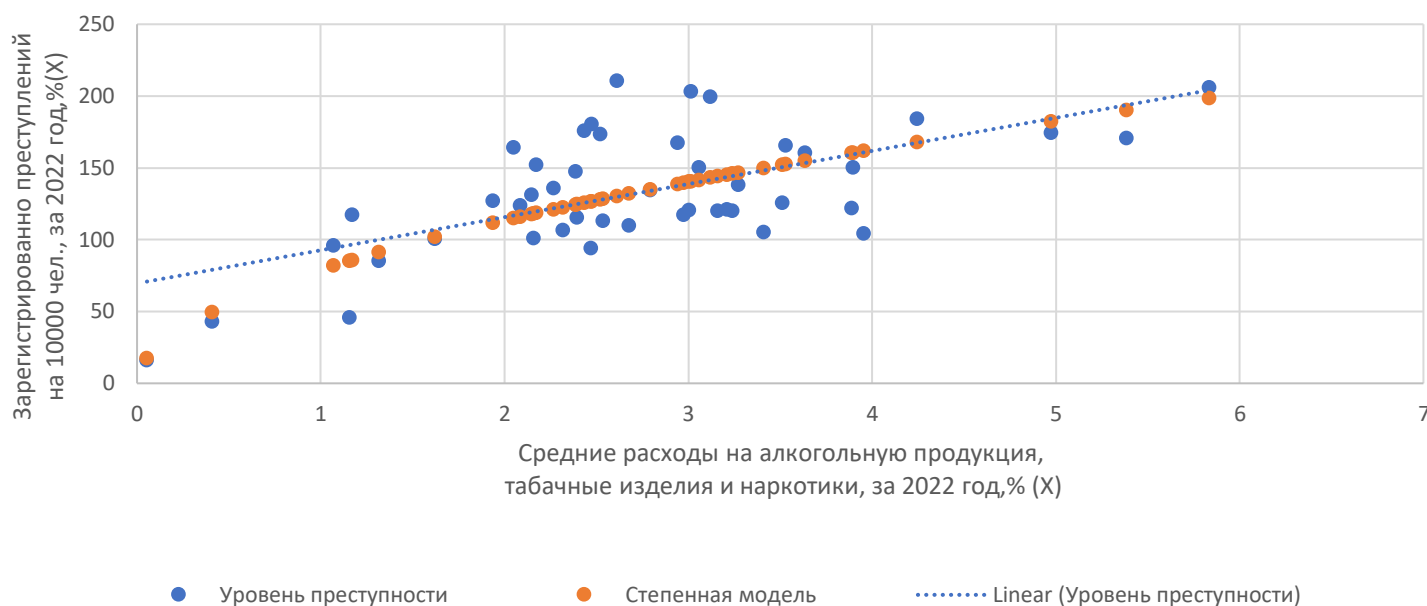


Рисунок 5. Степенная регрессионная модель

Далее вычислим МНК-оценку по формуле 15, $S = 30,696$

$$S = \sqrt{\frac{\sum_{i=1}^N (y_i - y_{\text{перп.}})^2}{N}}$$

6.4 Сравнение построенных моделей

В следующей таблице приведены данные для сравнения построенных моделей аппроксимации:

	Линейная модель	Степенная модель	Гиперболическая модель
МНК-Оценка	32,805	30,696	36,281
R2	0,397	0,727	0,233

Рисунок 6. Сравнение МНК-оценки и коэффициента детерминации

Для построения нелинейных регрессионных моделей (в частности, степенной и гиперболической) вычисления можно свести к построению линейной модели с помощью замены переменных. После построения моделей методом линеаризации я вычислил МНК - оценку для каждой модели. Сравнив полученные результаты, можно сделать выводы: Самая оптимальная и устойчивая модель оказалась Степенная так как она имеет наименьшее значение в МНК оценке и наибольший показатель R2 соответственно. Анализируя графики построенных аппроксимаций, можно заметить, что каждая модель приближается почти к линии на интервале от 2 до 4 %

расходов на алкоголь, табачные изделия и наркотики. Поэтому существенное отличие построенных моделей заключается в точности приближений значений вне этого интервала, и как можно заметить из графика у степенной функции это точность больше других. Гипербола с точки зрения оптимальности и устойчивости занимает последнее место среди 3 претендентов (наибольший МНК и наименьший R²), вне интервала (2,4) крайние значения находятся суммарно дальше, чем у линейной и степенной. Проблема линейной модели, заключалась в том, что сложно приблизить все значения выборки в одну линию, с этой точки зрения степенная модель оптимальнее. Коэффициент детерминации у степенной примерно в 1,75 раз больше, чем у линейной модели, тем самым проделанная работа привела к нахождению более подходящей модели аппроксимации.

7. Общие выводы по работе

Я выбрал актуальные данные по населению субъектов РФ и количеству преступлений по субъектам РФ, после чего рассчитал количество преступлений по субъектам РФ на 10000 населения. После чего исследовал зависимость уровня преступности от средних расходов на алкоголь, табачные изделия и наркотики. Первым этапом произошел анализ выборки на выбросы, после чего были удалены Сибирский, Северо-Западный и Центральный федеральные округа, на этом этапе подтверждается гипотеза о слабой связи для наиболее развитых городов. Убрав выбросы, провел корреляционный анализ, где была установлена средняя положительная связь, чем подтверждается гипотеза об увеличении преступности при увеличении расходов на алкоголь, табак и наркотики и опровергается гипотеза о сильной связи между признаками. Можно сделать вывод о существовании других факторов, влияющих на уровень преступности. Коэффициент корреляции оказался значимым. Последний этап был построение регрессионных моделей, сравнивая МНК-оценки и коэффициент детерминации приходим к выводу, что лучшая из рассматриваемых моделей является степенная, а худшая гиперболическая. Подтвердилась гипотеза о линейной модели, действительно при лучшей модели аппроксимации – линейной, не будет других факторов, от которых зависит преступность, что, очевидно, неверно.

8. Список литературы

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>

9. Используемые информационные источники

1. Сайт Государственной прокуратуры Российской Федерации : <https://epp.genproc.gov.ru/web/gprf/activity/crimestat>
2. Федеральная служба государственной статистики: Доходы, расходы и потребление домашних хозяйств в 2022 году https://rosstat.gov.ru/bgd/regl/b22_102/Main.htm

3. Федеральная служба государственной статистики: Численность населения Российской Федерации по муниципальным образованиям
<https://rosstat.gov.ru/compendium/document/13282>

4. Математико-статистические таблицы
<https://docs.google.com/document/d/17-MRyhmovzDaGOz6qavIwCCN6vNzcMvD/edit>

10. Приложения

Приложение 1

Таблица 1. Нахождения количества преступлений на 10000 человек субъекта РФ

Субъект РФ	Зарегистрировано преступлений ,шт.	Численность населения преступлений ,шт	Зарегистрировано преступлений на 10000 чел., за 2022 год
Амурская область	16253	772525	210,388
Республика Алтай	4586	221559	206,988
Еврейская автономная область	3164	153831	205,680
Республика Бурятия	19955	982629	203,078
Забайкальский край	20816	1043467	199,489
Республика Тыва	6619	332609	199,002
Республика Карелия	11850	603067	196,496
Республика Коми	14797	803477	184,162
Челябинская область	62844	3418606	183,829
Новгородская область	10615	586129	181,103
Кемеровская область	47071	2604272	180,745
Приморский край	33604	1863011	180,375
Республика Хакасия	9493	528338	179,677
Сахалинская область	8511	484177	175,783
Камчатский край	5450	312704	174,286
Томская область	18578	1068304	173,902
Тверская область	21352	1230190	173,567
Магаданская область	2386	137767	173,191
Алтайский край	39177	2268179	172,724
Чукотский автономный округ	853	50040	170,464
Новосибирская область	46814	2780292	168,378
Курганская область	13456	805510	167,049
Мурманская область	12075	724452	166,678
Удмуртская Республика	24583	1484460	165,602
Хабаровский край	21312	1298978	164,067
Иркутская область	37867	2357134	160,648
Смоленская область	14611	909856	160,586
Пермский край	41027	2556852	160,459

Красноярский край	45561	2849169	159,910
Архангельская область	16786	1069782	156,910
Ненецкий автономный округ	686	44540	154,019
Вологодская область	17411	1139499	152,795
Тюменская область	23591	1552148	151,989
Самарская область	47060	3131720	150,269
Кировская область	18519	1234780	149,978
Ярославская область	18368	1227383	149,652
Ленинградская область	28569	1911586	149,452
Волгоградская область	36031	2449781	147,078
Калужская область	14656	1012844	144,701
Псковская область	8714	613356	142,071
Тамбовская область	13799	980984	140,665
Ростовская область	57246	4153763	137,817
Краснодарский край	77125	5687378	135,607
Республика Татарстан	52189	3886395	134,286
Воронежская область	30460	2287678	133,148
Костромская область	8188	620776	131,899
Республика Саха (Якутия)	13000	992115	131,033
Республика Башкортостан	50751	4001678	126,824
Нижегородская область	39404	3144254	125,321
Липецкая область	13940	1113680	125,171
Курская область	13448	1083584	124,107
Саратовская область	29200	2360959	123,679
Калининградская область	12645	1027678	123,044
Владимирская область	16254	1323659	122,796
Омская область	23074	1879548	122,764
Свердловская область	51822	4264340	121,524
Ставропольский край	33569	2780204	120,743
Ивановская область	11791	976918	120,696
Ямало-Ненецкий автономный округ	6657	552117	120,572
Севастополь	6269	522057	120,083
Ханты-Мансийский автономный округ	20376	1702240	119,701
Астраханская область	13062	1114322	117,219
Республика Северная Осетия-Алания	8065	688124	117,203
Санкт Петербург	62971	5377503	117,101
Брянская область	13549	1168771	115,925
Орловская область	8243	714094	115,433
Республика Марий Эл	7741	671455	115,287
Оренбургская область	21742	1924578	112,970

Тульская область	15826	1432570	110,473
Республика Крым	20779	1896393	109,571
Москва	138180	12635466	109,359
Чувашская Республика	12730	1198429	106,222
Республика Мордовия	8107	770673	105,194
Пензенская область	13274	1274062	104,186
Ульяновская область	12159	1203969	100,991
Республика Калмыкия	2693	267756	100,577
Карачаево-Черкесская Республика	4436	464219	95,558
Московская область	74054	7768878	95,321
Белгородская область	14588	1531917	95,227
Республика Адыгея	4388	468340	93,693
Кабардино-Балкарская Республика	7416	870487	85,194
Рязанская область	9154	1085152	84,357
Республика Дагестан	14357	3153857	45,522
Республика Ингушетия	2233	524058	42,610
Чеченская Республика	2416	1516387	15,933

Приложение 2

Таблица 2. Исходные данные по субъектам РФ: X - количество зарегистрированных преступлений, Y - средние расходы на алкоголь, табак, наркотики.

Субъект РФ	Средние расходы на алкогольную продукцию, табачные изделия и наркотики, за 2022 год, % (X)	Зарегистрировано преступлений на 10000 чел., за 2022 год (Y)
Чеченская Республика	0,056	15,933
Республика Ингушетия	0,411	42,610
Республика Дагестан	1,176	45,522
Рязанская область	3,289	84,357
Кабардино-Балкарская Республика	1,386	85,194
Республика Адыгея	2,677	93,693
Белгородская область	3,869	95,227
Московская область	1,890	95,321
Карачаево-Черкесская Республика	1,138	95,558
Республика Калмыкия	1,679	100,577
Ульяновская область	2,280	100,991
Пензенская область	4,094	104,186
Республика Мордовия	3,586	105,194

Чувашская Республика	2,427	106,222
Москва	3,880	109,359
Республика Крым	2,730	109,571
Тульская область	2,870	110,473
Оренбургская область	2,683	112,970
Республика Марий Эл	2,496	115,287
Орловская область	3,076	115,433
Брянская область	2,466	115,925
Санкт Петербург	2,379	117,101
Республика Северная Осетия-Алания	1,038	117,203
Астраханская область	2,828	117,219
Ханты-Мансийский автономный округ	3,216	119,701
Севастополь	3,265	120,083
Ямало-Ненецкий автономный округ	2,991	120,572
Ивановская область	4,885	120,696
Ставропольский край	3,231	120,743
Свердловская область	3,949	121,524
Омская область	2,487	122,764
Владимирская область	2,935	122,796
Калининградская область	3,510	123,044
Саратовская область	2,139	123,679
Курская область	3,565	124,107
Липецкая область	3,012	125,171
Нижегородская область	3,669	125,321
Республика Башкортостан	2,029	126,824
Республика Саха (Якутия)	2,185	131,033
Костромская область	3,677	131,899
Воронежская область	4,303	133,148
Республика Татарстан	2,919	134,286
Краснодарский край	2,290	135,607
Ростовская область	3,329	137,817
Тамбовская область	1,946	140,665
Псковская область	3,955	142,071
Калужская область	2,674	144,701
Волгоградская область	2,440	147,078
Ленинградская область	4,544	149,452
Ярославская область	2,506	149,652
Кировская область	3,188	149,978
Самарская область	3,941	150,269

Тюменская область	2,202	151,989
Вологодская область	2,740	152,795
Ненецкий автономный округ	4,892	154,019
Архангельская область	2,385	156,910
Красноярский край	3,917	159,910
Пермский край	3,705	160,459
Смоленская область	1,443	160,586
Иркутская область	2,976	160,648
Хабаровский край	2,058	164,067
Удмуртская Республика	3,664	165,602
Мурманская область	3,795	166,678
Курганская область	3,275	167,049
Новосибирская область	4,222	168,378
Чукотский автономный округ	5,505	170,464
Алтайский край	2,798	172,724
Магаданская область	2,363	173,191
Тверская область	5,134	173,567
Томская область	2,821	173,902
Камчатский край	5,102	174,286
Сахалинская область	2,473	175,783
Республика Хакасия	4,595	179,677
Приморский край	2,485	180,375
Кемеровская область	2,392	180,745
Новгородская область	1,941	181,103
Челябинская область	4,267	183,829
Республика Коми	3,014	184,162
Республика Карелия	4,048	196,496
Республика Тыва	1,970	199,002
Забайкальский край	3,257	199,489
Республика Бурятия	3,185	203,078
Еврейская автономная область	6,010	205,680
Республика Алтай	3,583	206,988
Амурская область	2,720	210,388

Приложение 3

Таблица 3. Северо-Западный федеральный округ

Северо-Западный федеральный округ		
Санкт-Петербург	117,101	2,4
Республика Коми	184,162	2,9
Республика Карелия	196,496	4,0

Псковская область	142,071	3,8
Новгородская область	181,103	1,9
Ненецкий автономный округ	154,019	4,8
Мурманская область	166,678	3,8
Ленинградская область	149,452	4,4
Калининградская область	123,044	3,5
Вологодская область	152,795	2,7
Архангельская область	156,910	2,4

0,001

Таблица 4. Сибирский федеральный округ

Сибирский федеральный округ

Томская область	173,902	2,8
Республика Хакасия	179,677	4,3
Республика Тыва	199,002	1,9
Республика Алтай	206,988	3,5
Омская область	122,764	2,4
Новосибирская область	168,378	4,0
Красноярский край	159,910	3,8
Кемеровская область	180,745	2,3
Иркутская область	160,648	2,9
Алтайский край	172,724	2,7

0,053

Таблица 5. Дальневосточный федеральный округ

Дальневосточный федеральный округ

Чукотский автономный округ	170,464	5,4
Хабаровский край	164,067	2,0
Сахалинская область	175,783	2,4
Республика Саха (Якутия)	131,033	2,1
Республика Бурятия	203,078	3,0
Приморский край	180,375	2,5
Магаданская область	173,191	2,5
Камчатский край	174,286	5,0
Забайкальский край	199,489	3,1
Еврейская автономная область	205,680	5,8

Амурская область	210,388	2,6
------------------	---------	-----

0,280

Таблица 6. Южный федеральный округ

Южный федеральный округ

Севастополь	120,083	3,2
Ростовская область	137,817	3,3
Республика Крым	109,571	2,7
Республика Калмыкия	100,577	1,6
Республика Адыгея	93,693	2,5
Краснодарский край	135,607	2,3
Волгоградская область	147,078	2,4
Астраханская область	117,219	3,0

0,297

Таблица 7. Северо-Кавказский федеральный округ

Северо-Кавказский федеральный округ

Чеченская Республика	15,933	0,1
Ставропольский край	120,743	3,2
Республика Северная Осетия-Алания	117,203	1,2
Республика Ингушетия	42,610	0,4
Республика Дагестан	45,522	1,2
Карачаево- Черкесская Республика	95,558	1,1
Кабардино- Балкарская Республика	85,194	1,3

0,757

Таблица 8. Центральный федеральный округ

Центральный федеральный округ		
Ярославская область	149,652	2,5
Тульская область	110,473	2,8
Тверская область	173,567	4,9
Тамбовская область	140,665	1,9
Смоленская область	160,586	1,4
Рязанская область	84,357	3,2
Орловская область	115,433	2,9
Московская область	95,321	1,8
Москва	109,359	3,9
Липецкая область	125,171	2,7
Курская область	124,107	3,4
Костромская область	131,899	3,6
Калужская область	144,701	2,6
Ивановская область	120,696	4,8
Воронежская область	133,148	4,0
Владимирская область	122,796	2,8
Брянская область	115,925	2,4
Белгородская область	95,227	3,7

0,019

Таблица 9. Уральский федеральный округ

Уральский федеральный округ		
Ямало-Ненецкий автономный округ	120,572	3,0
Челябинская область	183,829	4,2
Ханты-Мансийский автономный округ	119,701	3,2
Тюменская область	151,989	2,2

Свердловская область	121,524	3,9
Курганская область	167,049	2,9

0,171

Таблица 10. Приволжский федеральный округ

Приволжский федеральный округ

Чувашская Республика	106,222	2,3
Ульяновская область	100,991	2,2
Удмуртская Республика	165,602	3,5
Саратовская область	123,679	2,1
Самарская область	150,269	3,9
Республика Татарстан	134,286	2,8
Республика Мордовия	105,194	3,4
Республика Марий Эл	115,287	2,4
Республика Башкортостан	126,824	1,9
Пермский край	160,459	3,6
Пензенская область	104,186	4,0
Оренбургская область	112,970	2,5
Нижегородская область	125,321	3,5
Кировская область	149,978	3,1

0,417

Приложение 4

Таблица 11. Новая выборка, без выбросов

Субъект РФ	Средние расходы на алкогольную продукцию, табачные изделия и наркотики, за 2022 год, % (X)	Зарегистрировано преступлений на 10000 чел., за 2022 год, (Y)
Чукотский автономный округ	5,3858	170,4636
Хабаровский край	2,0488	164,0674
Сахалинская область	2,4343	175,7828
Республика Саха (Якутия)	2,1485	131,0332
Республика Бурятия	3,0158	203,0777
Приморский край	2,4720	180,3747

Магаданская область	2,5211	173,1910
Камчатский край	4,9727	174,2862
Забайкальский край	3,1204	199,4888
Еврейская автономная область	5,8357	205,6803
Амурская область	2,6112	210,3880
Севастополь	3,2403	120,0827
Ростовская область	3,2720	137,8172
Республика Крым	2,6756	109,5712
Республика Калмыкия	1,6220	100,5766
Республика Адыгея	2,4693	93,6926
Краснодарский край	2,2679	135,6073
Волгоградская область	2,3867	147,0785
Астраханская область	2,9756	117,2193
Чеченская Республика	0,0536	15,9326
Ставропольский край	3,2121	120,7429
Республика Северная Осетия-Алания	1,1717	117,2027
Республика Ингушетия	0,4071	42,6098
Республика Дагестан	1,1576	45,5220
Карачаево-Черкесская Республика	1,0705	95,5583
Кабардино-Балкарская Республика	1,3174	85,1937
Чувашская Республика	2,3187	106,2224
Ульяновская область	2,1563	100,9910
Удмуртская Республика	3,5290	165,6023
Саратовская область	2,0853	123,6786
Самарская область	3,8965	150,2689
Республика Татарстан	2,7912	134,2864
Республика Мордовия	3,4103	105,1938
Республика Марий Эл	2,3940	115,2870
Республика Башкортостан	1,9379	126,8243
Пермский край	3,6336	160,4590
Пензенская область	3,9535	104,1865
Оренбургская область	2,5336	112,9702
Нижегородская область	3,5115	125,3207
Кировская область	3,0571	149,9781
Ямало-Ненецкий автономный округ	3,0041	120,5723
Челябинская область	4,2437	183,8293
Ханты-Мансийский автономный округ	3,1585	119,7011

Тюменская область	2,1730	151,9894
Свердловская область	3,8900	121,5241
Курганская область	2,9404	167,0494

Приложение 5

Таблица 12. Вспомогательные данные для нахождения коэффициента корреляции

XY	X^2	Y^2
918,076	29,006	29057,849
336,137	4,197	26918,126
427,914	5,926	30899,601
281,525	4,616	17169,699
612,449	9,095	41240,537
445,890	6,111	32535,023
436,630	6,356	29995,111
866,669	24,727	30375,689
622,487	9,737	39795,789
1200,285	34,055	42304,369
549,358	6,818	44263,116
389,102	10,499	14419,848
450,935	10,706	18993,582
293,167	7,159	12005,841
163,139	2,631	10115,661
231,353	6,097	8778,306
307,546	5,143	18389,341
351,039	5,697	21632,071
348,798	8,854	13740,355
0,855	0,003	253,848
387,834	10,317	14578,857
137,322	1,373	13736,476
17,347	0,166	1815,594
52,697	1,340	2072,256
102,297	1,146	9131,398
112,235	1,736	7257,965
246,298	5,376	11283,197
217,768	4,650	10199,176
584,412	12,454	27424,124
257,907	4,348	15296,385
585,519	15,183	22580,731

374,821	7,791	18032,838
358,748	11,630	11065,730
275,999	5,731	13291,081
245,776	3,756	16084,402
583,047	13,203	25747,101
411,900	15,630	10854,817
286,225	6,419	12762,271
440,062	12,331	15705,269
458,505	9,346	22493,441
362,212	9,025	14537,672
780,123	18,009	33793,217
378,082	9,976	14328,353
330,267	4,722	23100,769
472,729	15,132	14768,102
491,199	8,646	27905,518

Приложение 6

Таблица 13. Вспомогательные данные для построения линейной регрессионной модели

$X*Y$	$X-X_{cp.}$	$(X - X_{cp.})^2$	$Y_{перг.}$
0,855	-2,696	7,269	86,079
17,347	-2,343	5,488	79,011
52,697	-1,592	2,535	96,309
112,235	-1,432	2,051	99,992
231,353	-0,280	0,079	126,542
102,297	-1,679	2,820	94,302
163,139	-1,128	1,272	107,014
217,768	-0,593	0,352	119,328
411,900	1,204	1,449	160,751
358,748	0,661	0,437	148,232
246,298	-0,431	0,186	123,071
293,167	-0,074	0,005	131,297
286,225	-0,216	0,047	128,025
275,999	-0,356	0,126	124,807
137,322	-1,578	2,490	96,633

348,798	0,226	0,051	138,212
378,082	0,409	0,167	142,428
389,102	0,491	0,241	144,312
362,212	0,254	0,065	138,869
387,834	0,462	0,214	143,662
472,729	1,140	1,300	159,287
257,907	-0,664	0,441	117,691
440,062	0,762	0,580	150,563
245,776	-0,812	0,659	114,294
281,525	-0,601	0,361	119,148
374,821	0,042	0,002	133,962
307,546	-0,482	0,232	121,900
450,935	0,522	0,273	145,043
351,039	-0,363	0,132	124,639
458,505	0,307	0,095	140,091
585,519	1,147	1,315	159,437
330,267	-0,577	0,333	119,712
583,047	0,884	0,781	153,378
336,137	-0,701	0,491	116,849
584,412	0,779	0,607	150,967
491,199	0,191	0,036	137,401
918,076	2,636	6,949	193,763
436,630	-0,229	0,052	127,736
866,669	2,223	4,942	184,242
427,914	-0,315	0,099	125,736
445,890	-0,278	0,077	126,605
780,123	1,494	2,232	167,441
622,487	0,371	0,137	141,549
612,449	0,266	0,071	139,139
1200,285	3,086	9,524	204,133
549,358	-0,138	0,019	129,812

Приложение 7

Таблица 14. Вспомогательные данные для построения гиперболической регрессионной модели

$Z(= \frac{1}{X})$	$Z \cdot Y$	$Z - Z_{cp.}$	$(Z - Z_{cp.})^2$	$Y_{перп.}$	$(Y_i - Y_{перп.})^2$
18,644	297,043	17,801	316,878	-0,199	260,218

2,456	104,664	1,614	2,604	120,929	6133,915
0,864	39,324	0,021	0,000	132,845	7625,365
0,759	64,667	-0,084	0,007	133,629	2346,026
0,405	37,943	-0,438	0,192	136,279	1813,604
0,934	89,264	0,091	0,008	132,319	1351,381
0,617	62,006	-0,226	0,051	134,696	1164,143
0,464	46,835	-0,379	0,144	135,839	1214,400
0,253	26,353	-0,590	0,348	137,417	1104,250
0,293	30,845	-0,549	0,302	137,115	1018,982
0,431	45,811	-0,411	0,169	136,082	891,611
0,374	40,952	-0,469	0,220	136,513	725,847
0,395	44,588	-0,448	0,201	136,356	546,896
0,418	48,156	-0,425	0,181	136,184	436,678
0,853	100,031	0,011	0,000	132,923	247,124
0,336	39,393	-0,507	0,257	136,795	383,197
0,317	37,897	-0,526	0,277	136,940	297,192
0,309	37,059	-0,534	0,285	137,000	286,199
0,333	40,136	-0,510	0,260	136,819	263,942
0,311	37,590	-0,531	0,282	136,980	263,636
0,257	31,240	-0,586	0,343	137,386	251,594
0,480	59,310	-0,363	0,132	135,721	145,022
0,285	35,689	-0,558	0,311	137,178	140,607
0,516	65,443	-0,327	0,107	135,448	74,371
0,465	60,988	-0,377	0,142	135,827	22,977
0,358	48,111	-0,484	0,235	136,629	5,486
0,441	59,794	-0,402	0,161	136,010	0,162
0,306	42,120	-0,537	0,288	137,022	0,632
0,419	61,623	-0,424	0,180	136,174	118,901
0,327	49,058	-0,516	0,266	136,862	172,039
0,257	38,565	-0,586	0,343	137,389	165,891
0,460	69,946	-0,382	0,146	135,866	259,969
0,275	44,160	-0,567	0,322	137,250	538,655
0,488	80,081	-0,355	0,126	135,657	807,149
0,283	46,926	-0,559	0,313	137,189	807,314

0,340	56,811	-0,503	0,253	136,765	917,171
0,186	31,651	-0,657	0,432	137,920	1059,085
0,397	68,697	-0,446	0,199	136,341	1357,896
0,201	35,049	-0,642	0,412	137,805	1330,907
0,411	72,210	-0,432	0,187	136,236	1563,987
0,405	72,967	-0,438	0,192	136,282	1944,128
0,236	43,318	-0,607	0,368	137,546	2142,131
0,320	63,930	-0,522	0,273	136,911	3915,934
0,332	67,337	-0,511	0,261	136,828	4388,986
0,171	35,245	-0,671	0,451	138,027	4576,941
0,383	80,572	-0,460	0,211	136,444	5467,759

Таблица 15. Вспомогательные данные для построения степенной модели Приложение 8

$Y_1 = \log Y$	$X_1 = \log X$	$X_1 * Y_1$	$X_1 - X_{1cp.}$	$(X_1 - X_{1cp.})^2$	$Y_{\text{perp.}}$	$(Y_i - Y_{\text{perp}})^2$
1,202	-1,271	-1,528	-1,644	2,704	17,113	1,392
1,630	-0,390	-0,636	-0,764	0,584	49,326	45,111
1,658	0,064	0,105	-0,310	0,096	85,139	1569,502
1,930	0,120	0,231	-0,254	0,065	91,088	34,745
1,972	0,393	0,774	0,019	0,000	126,466	1074,088
1,980	0,030	0,059	-0,344	0,118	81,731	191,200
2,002	0,210	0,421	-0,164	0,027	101,542	0,932
2,004	0,334	0,669	-0,040	0,002	117,823	283,322
2,018	0,597	1,205	0,223	0,050	161,711	3309,057
2,022	0,533	1,077	0,159	0,025	149,698	1980,634
2,026	0,365	0,740	-0,009	0,000	122,377	260,983
2,040	0,427	0,872	0,054	0,003	131,879	497,632
2,053	0,404	0,829	0,030	0,001	128,177	231,238
2,062	0,379	0,782	0,005	0,000	124,438	83,737
2,069	0,069	0,142	-0,305	0,093	85,677	993,854
2,069	0,474	0,980	0,100	0,010	139,406	492,270
2,078	0,499	1,038	0,126	0,016	143,819	581,692
2,079	0,511	1,062	0,137	0,019	145,751	658,882
2,081	0,478	0,994	0,104	0,011	140,102	381,428
2,082	0,507	1,055	0,133	0,018	145,087	592,635
2,085	0,590	1,230	0,216	0,047	160,349	1507,386
2,092	0,319	0,668	-0,055	0,003	115,780	62,382
2,098	0,545	1,144	0,172	0,029	152,001	711,828
2,103	0,287	0,604	-0,086	0,007	111,432	236,935
2,117	0,332	0,703	-0,042	0,002	117,600	180,449
2,128	0,446	0,949	0,072	0,005	134,825	0,290
2,132	0,356	0,758	-0,018	0,000	120,970	214,254

2,139	0,515	1,101	0,141	0,020	146,494	75,292
2,168	0,378	0,819	0,004	0,000	124,240	521,590
2,176	0,485	1,056	0,112	0,012	141,389	73,773
2,177	0,591	1,286	0,217	0,047	160,489	104,442
2,182	0,337	0,735	-0,037	0,001	118,298	1135,140
2,205	0,560	1,236	0,187	0,035	154,739	32,714
2,215	0,311	0,690	-0,062	0,004	114,717	2435,508
2,219	0,548	1,215	0,174	0,030	152,396	174,396
2,223	0,468	1,041	0,095	0,009	138,544	812,587
2,232	0,731	1,632	0,357	0,128	190,050	383,625
2,239	0,402	0,899	0,028	0,001	127,845	2056,260
2,241	0,697	1,561	0,323	0,104	182,291	64,083
2,245	0,386	0,867	0,013	0,000	125,528	2525,570
2,256	0,393	0,887	0,019	0,000	126,539	2898,270
2,264	0,628	1,421	0,254	0,065	167,807	256,722
2,300	0,494	1,137	0,120	0,015	142,910	3201,193
2,308	0,479	1,106	0,106	0,011	140,388	3930,020
2,313	0,766	1,772	0,392	0,154	198,184	56,199
2,323	0,417	0,968	0,043	0,002	130,211	6428,395