

Национальный исследовательский университет «Высшая школа экономики»
Факультет коммуникаций, медиа и дизайна
Образовательная программа: Прикладная математика и информатика

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

«Статистические методы анализа двумерных данных на примере анализа
взаимосвязи между численностью населения и численностью работников
государственных органов по регионам РФ»

Работу выполнили
студентка 2 курса
Ульянова Александра Игоревна
Преподаватель:
Грачева Светлана Сергеевна

Москва, 2022 г.

Содержание

Введение.....	3
Корреляционный анализ.....	4
Регрессионный анализ.....	8
Заключение.....	14
Список литературы.....	15
Приложения.....	16

Введение

В данной работе я решила исследовать взаимосвязь между следующими двумя дискретными показателями:

- 1) Независимый показатель: оценка численности постоянного населения по субъектам Российской Федерации на 1 января 2020 года, чел
- 2) Результативный показатель: численность сотрудников государственных органов и органов местного самоуправления по субъектам Российской Федерации на 2020 год, чел.

Первый из показателей является одним из ключевых показателей для каждого субъекта РФ. Именно он одним из первых приходит на ум, когда речь заходит о регионах нашей страны, о сравнении их между собой, о интерпретации их экономических успехов. Второй показатель несколько похож на первый, однако характеризует численность более узкой группы лиц.

Гипотеза состоит в том, что между этими двумя признаками имеется сильная взаимосвязь. Число государственных и муниципальных служащих, скорее всего, пропорционально численности населения для большинства регионов. Вероятно, также присутствуют аномальные субъекты РФ, где отношения выше обозначенных показателей сильно отличается от среднего.

Актуальность темы исследования вызвана тем, что в настоящее время многие граждане РФ недовольны быстротой и качеством услуг, оказываемых государственными органами. Изучение численности сотрудников государственных органов и органов местного самоуправления может помочь в выявлении проблем и различных трендов, а в дальнейшем возможно в оптимизации их работы.

Данные были взяты с сайта Федеральной службы государственной статистики. Выбор именно этого ресурса был связан с тем, что он предоставляет официальные и самые новые данные, на которые опираются органы власти, научное сообщество и СМИ. Также Росстат собирает информацию, которой зачастую нет в других государственных базах данных или в административных источниках. Исходный массив данных находится в Приложении 1.

Целью работы являются: изучение исходных данных, выявление наличия или отсутствия взаимосвязи между исследуемыми признаками, изучение и подробное описание этой взаимосвязи, характеристика численности сотрудников государственных органов исходя из полученных результатов

Задачи работы:

- Проведение корреляционного анализа
- Проведение регрессионного анализа
- Подведение итогов

Корреляционный анализ

Корреляционный анализ — метод обработки статистических данных, с помощью которого измеряется теснота связи между двумя или более переменными. В данном случае мной была исследована взаимосвязь между двумя признаками, описанными во введении.

Построение корреляционного облака

Первым этапом стало построение корреляционного облака, или как его еще называют, диаграммы рассеяния. Она представляет из себя точечную диаграмму, где на оси X указаны значения независимого признака, а на оси Y — соответствующие значения результативного признака. Данная диаграмма позволяет визуальное представить изучаемый массив данных и сделать некоторые предварительные выводы о наличии или отсутствии взаимосвязи между изучаемыми признаками. Полученное корреляционное облако представлено на Рис.1

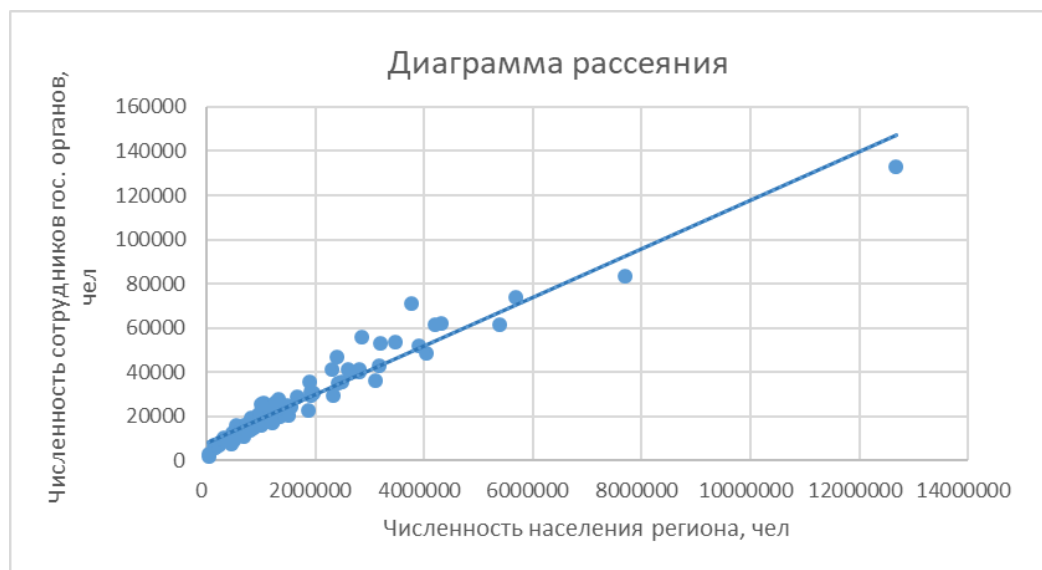


Рис.1 Диаграмма рассеяния

Для большей наглядности на данной диаграмме отмечена линия тренда. Линия тренда представляет собой геометрическое отображение средних значений анализируемых показателей, полученное в данном случае с помощью линейной функции.

Как видно из получившейся диаграммы, все точки располагаются недалеко от линии тренда, а значит предварительно можно сказать, что связь между изучаемыми признаками есть. Так же можно сказать, что имеет место положительная линейная зависимость, о чем нам говорит положительный наклон линии тренда. Угол наклона близок к 45° , а значит связь между признаками сильная. В дальнейшем наши предположения можно будет проверить, подсчитав коэффициент корреляции.

Также по данной диаграмме можно увидеть, что большинство регионов располагаются в нижнем левом углу диаграммы, которому соответствуют значения X от 0 до 6 млн. и значения Y от 0 до 80 тыс. Наблюдаются также несколько регионов, значения показателей в которых сильно выделяются. Это город Москва, значения признаков для которого равны 12678079 и 133151 соответственно и Московская область, со значениями признаков 7690863 и 83549. Объяснить большое количество населения в этих регионах можно тем, что в них расположена Московская агломерация, куда приезжает множество людей в поисках работы и более комфортных условий для жизни. Большое же число государственных служащих объясняется расположением в Москве органов федеральной власти РФ.

Расчет линейного коэффициента корреляции

Следующими этапом стал подсчет линейного коэффициента корреляции. Коэффициент линейной корреляции известен как коэффициент корреляции Пирсона и отражает направление и силу линейной зависимости между двумя переменными x и y. Он возвращает значение от -1 до +1. При этом -1 указывает на сильную отрицательную корреляцию, на +1 – на сильную положительную корреляцию.

Расчет был произведен двумя способами: с помощью функции КОРРЕЛ в MS Excel и по формуле $r = \frac{\bar{x}\bar{y} - \bar{x}*\bar{y}}{S_x*S_y}$, где $S_x = \sqrt{\overline{x^2} - \bar{x}^2}$

Для подсчета по формуле, обозначенной выше, пришлось посчитать некоторые дополнительные показатели: ΣX – сумма значений независимого признака по всем регионам, ΣY – сумма значений результативного признака по всем регионам, ΣXY – сумма произведений признаков по всем регионам, ΣX^2 – сумма квадратов значений независимого признака и ΣY^2 – сумма квадратов значений результативного признака. Полученные результаты представлены в таблице 1.

Таблица 1. Промежуточные вычисления

ΣX	ΣY	ΣXY	ΣX^2	ΣY^2
149011821	2322714	7.15088E+12	5.41513E+14	99540853222

С помощью этих значений коэффициент корреляции был вычислен как

$$r = \frac{n*\Sigma XY - \Sigma x*\Sigma y}{\sqrt{n*\Sigma x^2 - (\Sigma x)^2}*\sqrt{n*\Sigma y^2 - (\Sigma y)^2}},$$

где n – размер массива исходных данных, т.е. в нашем случае 85

Результат оказался одинаковым при обоих способах расчета, что говорит об их правильности. Коэффициент корреляции получился равным 0.97. Он принадлежит промежутку от 0.9 до 1, а это означает, что между двумя изучаемыми признаками имеется очень сильная положительная связь.

С чем может быть связана такая большая зависимость между численностью населения и численностью сотрудников государственных органов? Во-первых, чем больше людей проживает в регионе, тем больше государственных служащих необходимо для оказания различных услуг: оформления документов, юридической поддержки и так далее. Во-вторых, некоторые органы власти расположены не во всех регионах, а только в крупных, например, только в Москве, или только в административных центрах федеральных округов. Это увеличивает число их сотрудников относительно других регионов.

Так же интерес представляют регионы, расположенные на графике далеко от линии тренда. К таким относятся, например, Тюменская область (значения $x = 3756536$, $y = 70961$), Красноярский край ($x = 2866255$, $y = 55966$) и г. Москва ($x = 12678079$, $y = 133151$). Такое расположение характеризует их как регионы, где наблюдается некоторый дисбаланс численности населения и численности сотрудников государственных учреждений. Так как точки, отвечающие этим регионам, расположены выше линии тренда, можно сделать вывод об наличии там очень большого числа гос. служащих.



Рис.2 Диаграмма рассеяния с выделяющимися регионами

Объяснить, почему именно эти регионы выделяются среди остальных, не представляет труда. Тюменская область – крупный нефтедобывающий регион, где расположено множество государственных компаний из этой отрасли. Следовательно, там необходимо много государственных служащих для организации и контроля их работы. Красноярский же край – регион, где ведется добыча многих других полезных ископаемых. Большое число работников государственных органов тут обусловлено тем же. В столице России обратная ситуация – число сотрудников государственных органов меньше, чем могло бы быть исходя из численности населения. Связано это может быть с тем, что численность населения Москвы очень быстро и постоянно растет, в то время как у государственных органов потребность в новых кадрах растет не так быстро.

Проверка значимости коэффициента корреляции

Проверим значимость коэффициента корреляции с помощью t-критерия Стьюдента. Для этого посчитаем значение коэффициента t по следующей формуле:

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

Тут r – коэффициент корреляции, n – размер массива исходных данных, k – количество факторных признаков, включенных в модель. Было получено значение

$$t_{\text{расч.}} = 0.97 * \sqrt{\frac{85 - 1 - 1}{1 - 0.97^2}} \approx 36.35$$

Воспользуемся таблицей значений t-критерия Стьюдента для разных уровней значимости (Приложение 2), чтобы проверить значимость коэффициента корреляции. Посмотрим на значение в таблице на строке 85, так как 85 – размер массива наших исходных данных, и во втором столбце, так как нам достаточно уровня значимости 0.05. Найденное значение - 1,9882679. Так как посчитанное значение оказалось больше табличного, то коэффициент корреляции признается значимым.

Ложная корреляция

Стоит заметить, что коэффициент корреляции устанавливает лишь статистические взаимосвязи между исследуемыми признаками. Некорректно говорить о наличии какой-либо причинно-следственной связи между численностью населения и численностью сотрудников государственных органов, основываясь на посчитанных мною данных. Не исключен вариант, что корреляция изучаемых признаков вызвана третьим фактором, который не очевиден во время исследования, или, как его называют, смешивающим фактором. Для выяснения причинно-следственной связи необходимо дальнейшее изучение других близких по тематике признаков, что не является целью данной работы.

Выводы

В ходе корреляционного анализа мы построили диаграмму рассеяния, по которой сделали предположение о сильной прямой статистической связи между численностью населения региона и численностью сотрудников государственных органов. В дальнейшем был подсчитан коэффициент корреляции, который оказался положительным и по модулю близким к единице, чем подтвердил сделанные ранее предположения.

Регрессионный анализ

Регрессионный анализ представляет собой набор статистических процессов для оценки взаимосвязей между зависимой переменной и одной или несколькими независимыми переменными. Зависимость изучаемых переменных пытаются приблизить функцией, например, линейной или степенной, после чего оценивают точность приближения одним из множества способов.

Наиболее распространенной формой регрессионного анализа является линейная регрессия, при которой находят линию (или более сложную линейную комбинацию), которая наиболее точно соответствует данным в соответствии с определенным математическим критерием. Например, метод обычных наименьших квадратов вычисляет единственную прямую, которая минимизирует сумму квадратов различий между истинными данными и этой линией.

Построение линейной модели

Построим линейную регрессию для интересующих нас признаков. Будем приближать зависимость функцией вида $y = a + bx$.

Для расчетов коэффициентов a и b используем следующие формулы:

$$a = \frac{\sum y_i \sum x_i^2 - \sum x_i \sum x_i y_i}{n \sum x_i^2 - \sum x_i \sum x_i}$$
$$b = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - \sum x_i \sum x_i}$$

Воспользовавшись расчетами, полученными в Таблице 1, а также значением $n = 85$, получим

$$a = 8068.04,$$

$$b = 0.01$$

Значит искомая функция имеет вид $y = 8068.04 + 0.01x$

Для проверки правильности расчетов я также использовала модуль “Анализ данных” в MS Excel. Его функционал позволяет в том числе вычислить нужные нам коэффициенты. Полученные данным способом значения совпали со значениями, посчитанными вручную.

При приближении такой линейной функцией становится минимально возможной сумма наименьших квадратов, получаемая по формуле:

$$S = \sum_{i=1}^n (y_i - \hat{y}_x)^2$$

Здесь y_i – реальное значение результативного признака, а \hat{y}_x – расчетное значение результативного признака.

Значение S получилось равным 2772631166.08. Теперь мы можем использовать это значение для оценки качества построенной регрессионной модели.

Интерпретация линейной модели

В линейной модели регрессии коэффициенты a и b имеют следующий смысл: a – свободный член уравнения регрессии, т.е. это значение y при $x = 0$, b же показывает, на какую величину в абсолютном выражении изменяется « y » при изменении « x ».

В нашем случае коэффициент b , равный 0.01, говорит о том, что при изменении численности населения региона на 100 человек приводит в среднем к увеличению численности сотрудников государственных органов на 1 человека.

Оценка качества регрессионной модели

Для того, чтобы понять, насколько точно на самом деле построенная нами модель приближает исходные данные, посчитаем два следующих показателя: среднее отклонение и среднюю ошибку аппроксимации.

1) Среднее отклонение

В теории вероятностей и статистике среднеквадратическое отклонение — наиболее распространённый показатель рассеивания значений случайной величины относительно её математического ожидания. Вычисляется по

формуле
$$S = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_x)^2}{n}}$$

2) Средняя ошибка аппроксимации

Обозначает среднее отклонение расчетных значений от фактических и вычисляется по формуле $\bar{A} = \frac{1}{n} * \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} * 100$. Считается, что модель хорошо приближает исходные данные, если значение данного показателя не превышает 12 - 15%

Полученные в ходе вычислений значения показателей представлены на Рис. 2

Среднее отклонение	5711.32123
Средняя ошибка аппроксимации	20.893736

Рис. 3 Значения показателей для оценки качества модели

Видим, что среднее отклонение достаточно небольшое (с учетом того, что наши исходные признаки измеряются очень большими числами). Средняя ошибка аппроксимации немного превышает 15%, что говорит о неточности модели. Видно, что линейная функция приближает исходные данные достаточно хорошо, но все же не так хорошо, как нам бы хотелось.

Интерпретация коэффициента детерминации

Теперь посчитаем коэффициент детерминации. Он представляет собой долю вариации зависимой переменной, которая предсказуема из независимой переменной. Для линейной зависимости коэффициент детерминации равен квадрату коэффициента корреляции, то есть в нашем случае $R^2 = 0.97^2 \sim 0.94$

Это означает, что в 94% случаев изменения x приводят к изменению y . Другими словами, точность подбора уравнения регрессии – высокая.

Графическое представление

Представим наши расчеты графически с помощью уже известной нам диаграммы рассеяния. Изобразим на одном графике как исходные данные, так и данные, полученные благодаря линейной модели регрессии. Результат представлен на Рис.3

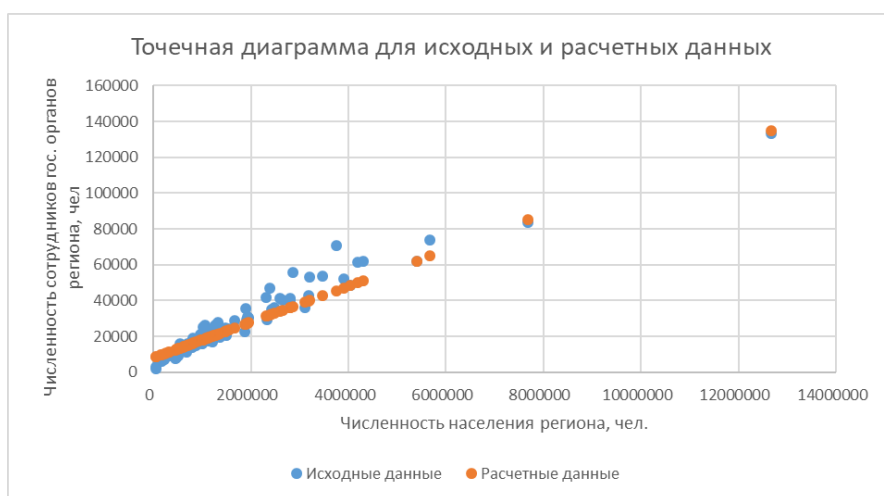


Рис.4 Графическое представление исходных данных, и данных, рассчитанных по линейной модели

Видно, что все точки, отвечающие за расчетные данные, располагаются на одной прямой. Это логично, ведь мы получали данные на основе линейной функции, графиком которой является прямая. Большинство оранжевых точек располагаются там же, где располагалось и большинство синих – в левой нижней части графика. Это говорит о правильности наших построений. Но также встречаются точки, у которых одинаковая координата x , но различные координаты y . Это вызвано погрешностью, которую дает приближение линейной функцией.

Проверка гипотезы о значимости уравнения регрессии

Целью расчетов является ответ на вопрос, являются ли полученные коэффициенты регрессии a и b статистически значимыми? Другими словами, отличны ли они от нуля и существенно ли влияют на результативный признак?

Для проверки значимости уравнения регрессии воспользуемся F-критерием Фишера. Рассчитаем его значение с использованием посчитанного нами ранее коэффициента детерминации по следующей формуле:

$$F = \frac{r^2}{1 - r^2} * (n - 2)$$

Получим значение $F = \frac{0.94^2}{1 - 0.94^2} * (85 - 2) \approx 630.1$

Число степеней свободы числителя равно количеству объясняющих переменных (т.е. для линейной модели вида $y = ax + b$ число степеней свободы равно 1). Число степеней свободы знаменателя равно $n-k-1$, где n – количество экспериментальных точек, k – количество объясняющих переменных, т.е. $85-1-1 = 83$. Воспользуемся таблицей значений F-критерия Фишера при уровне значимости $\alpha = 0,05$ (Приложение 3). Найденное табличное значение – 3.98. Наше значение сильно превышает табличное, а значит уравнение регрессии можно считать значимым.

Проверим также значимость коэффициента уравнения регрессии $b = 0.01$. Заметим, что он не равен нулю, а это говорит о значимости этого коэффициента.

Построение нелинейных моделей регрессии

Попытаемся подобрать уравнение регрессии, которое будет приближать исходные данные лучше, чем полученное нами ранее линейное уравнение.

Начнем с построения гиперболической модели. Уравнение в этой модели будет иметь следующий вид:

$$\hat{y} = b_0 + b_1 * \frac{1}{x}$$

Воспользуемся методом линеаризации нелинейных моделей путем замены переменных. Для этого введем переменную $z = \frac{1}{x}$. Относительно этой переменной наше уравнение уже будет иметь линейный вид, так что мы сможем посчитать значения b_0, b_1 аналогично тому, как мы считали значения a и b при построении линейного уравнения. Получим значения $b_0 = 31023.5$, $b_1 = -2288873156$, $R^2 = 0.13$

Для проверки полученных результатов также воспользуемся пакетом “Анализ данных” в MS Excel. Более подробно с процессом построения модели можно ознакомиться на листе “Гиперболическая модель” расчетного файла, прилагаемого к работе.

Значит наше гиперболическое уравнение имеет вид:

$$\hat{y} = 31023.5 - 2288873156 * \frac{1}{x}$$

Посмотрим, как выглядит наша модель, если отобразить ее на графике.



Рис.5 Графическое представление исходных данных, и данных, рассчитанных по гиперболической модели

На графике видно, что расчетные данные сильно отличаются от исходных. Более того, некоторые полученные нами значения результативного признака оказываются отрицательными, что не соответствует смыслу нашего признака. По данной точечной диаграмме можно сказать, что гиперболическая модель очень плохо приближает наши исходные данные. Это же подтверждается при анализе посчитанного раньше значения коэффициента детерминации $R^2 = 0.13$. Значение коэффициента близко к 0

Следующим построим степенную модель регрессии. Ее уравнение имеет вид

$$\hat{y} = b_0 x^{b_1} \Leftrightarrow \log \hat{y} = \log b_0 + b_1 * \log x$$

Второе из представлений этого уравнения является линейным относительно $\log x$, а значит предварительно подсчитав значения $\log x$, $\log y$ мы сможем воспользоваться пакетом “Анализ данных” в MS Excel и построить модель линейной регрессии. Получим следующие результаты:

$$b_0 = 0.98, b_1 = 0.71, R^2 = 0.95$$

$$\text{Значит наше степенное уравнение имеет вид } \hat{y} = 0.98 * x^{0.71}$$

Полученные параметры нашей модели имеют следующую экономическую интерпретацию: увеличение X на 1% приводит к увеличению Y в среднем на 0.71%.

Посмотрим, как построенная нами модель выглядит на графике.

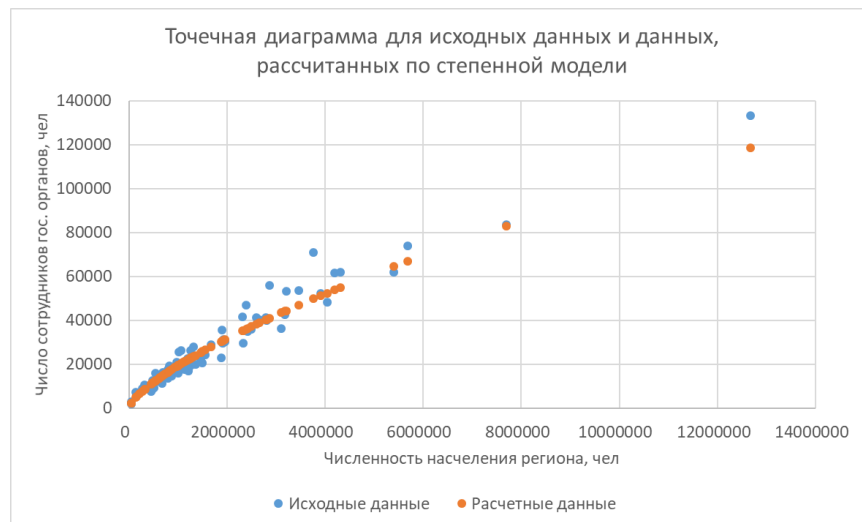


Рис. 6 Графическое представление исходных данных, и данных, рассчитанных по степенной модели

Посмотрев на диаграмму, можно увидеть, что расчетные данные достаточно точно отображают исходные. Это же подтверждает коэффициент детерминации $R^2 = 0.95$. Заметим, что этот коэффициент ближе к единице, чем коэффициент, полученный при построении линейной модели, а значит степенная функция приближает исходные данные немного лучше, чем линейная.

МНК-оценка моделей

Оценим наши модели с точки зрения метода наименьших квадратов. Более подробно про этот метод было сказано в начале раздела “Регрессионный анализ”. Воспользуемся формулой ниже и посчитаем значение S для каждой из построенных моделей. Полученные результаты представлены в Таблице 2.

$$S = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_x)^2}{n}}$$

	Линейная модель	Гиперболическая модель	Степенная модель
S	5711.32	19186.25	4750

Таблица 2. МНК-оценка моделей

Из таблицы выше видно, что наименьшее значение S получено в степенной модели. Это означает, что она точнее всего приближает наши данные. Это согласуется с выводами, сделанными нами ранее при сравнении коэффициента детерминации.

Заключение

В данной работе мной была исследована взаимосвязь между оценкой численности постоянного населения по субъектам Российской Федерации на 1 января 2020 года и численностью сотрудников государственных органов и органов местного самоуправления по субъектам Российской Федерации на 2020 год. Был проведен корреляционный анализ с построением корреляционного облака и подсчетом коэффициента корреляции. Был также произведен регрессионный анализ: были построены регрессионные модели нескольких видов, была проведена оценка качества каждой из них, модели были сравнены между собой.

Выводы по каждому отдельному шагу анализа указаны после вычисления каждого показателя в ходе работы. Если говорить о выводах в целом, из полученных результатов можно судить о том, что два исследуемых показателя имеют сильную положительную статистическую взаимосвязь. По мере увеличения численности населения в регионе численность сотрудников государственных органов тоже растет. Подтвердилось предположение, которое было сделано мной во введении.

Несколько выделяются значения исследуемых признаков в следующих регионах: г. Москва, Красноярский край, Тюменская область. Связано это либо со слишком быстрым темпом роста населения (в случае столицы), либо с сильным развитием промышленности и нефтедобывающей отрасли.

Лучше всего исходные данные приближаются степенной моделью регрессии. Немногим хуже их можно также приблизить с помощью линейной функции, а вот гиперболическая функция абсолютно не отражает исходные данные.

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.
2. Основы теории статистики : [учеб. пособие] / В. В. Полякова, Н. В. Шаброва ; М-во образования и науки Рос. Федерации, Урал. федер. ун-т. — 2-е изд., испр. и доп. — Екатеринбург : Изд-во Урал. ун-та, 2015. — 148 с. ISBN 978-5-7996-1520-8

Приложения

Приложение 1

Численность населения на 1 января 2020 года и число сотрудников
государственных органов на 2020 год по субъектам РФ

<i>Регионы</i>	<i>Численность населения на 1 января 2020 года, чел (X)</i>	<i>Число сотрудников государственных органов на 2020 год, чел (Y)</i>
Белгородская область	1549151	24204
Брянская область	1192491	22597
Владимирская область	1358416	19788
Воронежская область	2324205	29430
Ивановская область	997135	15939
Калужская область	1002575	18482
Костромская область	633385	14011
Курская область	1104008	18825
Липецкая область	1139371	17644
Московская область	7690863	83549
Орловская область	733498	15006
Рязанская область	1108847	18107
Смоленская область	934889	18504
Тамбовская область	1006748	18095
Тверская область	1260379	21951
Тульская область	1466127	21656
Ярославская область	1253389	20954
г.Москва	12678079	133151
Республика Карелия	614064	13117
Республика Коми	820473	19262
Архангельская область	1136535	21198
Ненецкий автономный округ	44111	1914
Вологодская область	1160445	21566
Калининградская область	1012512	25438
Ленинградская область	1875872	22703
Мурманская область	741404	14866
Новгородская область	596508	12174
Псковская область	626115	14315
г.Санкт-Петербург	5398064	61781
Республика Адыгея	463088	8320
Республика Калмыкия	271135	8878
Республика Крым	1912622	29514
Краснодарский край	5675462	74013
Астраханская область	1005782	17765
Волгоградская область	2491036	35851

Ростовская область	4197821	61437
г.Севастополь	449138	7586
Республика Дагестан	3110858	36082
Республика Ингушетия	507061	9254
Кабардино-Балкарская Республика	868350	14690
Карачаево-Черкесская Республика	465528	9829
Республика Северная Осетия- Алания	696837	16057
Чеченская Республика	1478726	24842
Ставропольский край	2803573	39975
Республика Башкортостан	4038151	48351
Республика Марий Эл	679417	11080
Республика Мордовия	790197	13650
Республика Татарстан	3902888	52168
Удмуртская Республика	1500955	20666
Чувашская Республика	1217818	16897
Пермский край	2599260	41341
Кировская область	1262402	26295
Нижегородская область	3202946	53257
Оренбургская область	1956835	30315
Пензенская область	1305563	20015
Самарская область	3179532	42677
Саратовская область	2421895	35027
Ульяновская область	1229824	18400
Курганская область	827166	17772
Свердловская область	4310681	61793
Тюменская область	3756536	70961
Ханты-Мансийский автономный округ - Югра	1674676	28810
Ямало-Ненецкий автономный округ	544444	15986
Челябинская область	3466369	53512
Республика Алтай	220181	6894
Республика Тыва	327383	9033
Республика Хакасия	534262	12288
Алтайский край	2317153	41537
Красноярский край	2866255	55966
Иркутская область	2391193	46959
Кемеровская область	2657854	40039
Новосибирская область	2798170	41094
Омская область	1926665	31181

Томская область	1079271	19924
Республика Бурятия	985937	17068
Республика Саха (Якутия)	971996	20863
Забайкальский край	1059700	26324
Камчатский край	313016	10520
Приморский край	1895868	35603
Хабаровский край	1315643	28025
Амурская область	790044	17435
Магаданская область	140149	7160
Сахалинская область	488257	12648
Еврейская автономная область	158305	5947
Чукотский автономный округ	50288	2913

Источник: <https://rosstat.gov.ru>

Значение t-критерия Стьюдента при разных уровнях значимости

	0.01	0.05	0.1	0.15	0.2	0.25	0.3
1	63,6567412	12,7062047	6,3137515	4,1652998	3,0776835	2,4142136	1,9626105
2	9,9248432	4,3026527	2,9199856	2,2819306	1,8856181	1,6035675	1,3862066
3	5,8409093	3,1824463	2,3533634	1,9243197	1,6377444	1,4226253	1,2497781
4	4,6040949	2,7764451	2,1318468	1,7781922	1,5332063	1,3443976	1,1895669
5	4,0321430	2,5705818	2,0150484	1,6993626	1,4758840	1,3009490	1,1557673
6	3,7074280	2,4469119	1,9431803	1,6501732	1,4397557	1,2733493	1,1341569
7	3,4994833	2,3646243	1,8945786	1,6165917	1,4149239	1,2542787	1,1191591
8	3,3553873	2,3060041	1,8595480	1,5922214	1,3968153	1,2403183	1,1081454
9	3,2498355	2,2621572	1,8331129	1,5737358	1,3830287	1,2296592	1,0997162
10	3,1692727	2,2281389	1,8124611	1,5592359	1,3721836	1,2212554	1,0930581
11	3,1058065	2,2009852	1,7958848	1,5475598	1,3634303	1,2144602	1,0876664
12	3,0545396	2,1788128	1,7822876	1,5379565	1,3562173	1,2088525	1,0832114
13	3,0122758	2,1603687	1,7709334	1,5299196	1,3501713	1,2041462	1,0794687
14	2,9768427	2,1447867	1,7613101	1,5230951	1,3450304	1,2001403	1,0762802
15	2,9467129	2,1314495	1,7530504	1,5172280	1,3406056	1,1966893	1,0735314
16	2,9207816	2,1199053	1,7458837	1,5121302	1,3367572	1,1936854	1,0711372
17	2,8982305	2,1098156	1,7396067	1,5076598	1,3333794	1,1910471	1,0690331
18	2,8784405	2,1009220	1,7340636	1,5037077	1,3303909	1,1887115	1,0671695
19	2,8609346	2,0930241	1,7291328	1,5001888	1,3277282	1,1866293	1,0655074
20	2,8453397	2,0859634	1,7247182	1,4970355	1,3253407	1,1847614	1,0640158
21	2,8313596	2,0796138	1,7207429	1,4941938	1,3231879	1,1830764	1,0626697
22	2,8187561	2,0738731	1,7171444	1,4916196	1,3212367	1,1815487	1,0614488
23	2,8073357	2,0686576	1,7138715	1,4892769	1,3194602	1,1801572	1,0603365
24	2,7969395	2,0638986	1,7108821	1,4871358	1,3178359	1,1788845	1,0593189
25	2,7874358	2,0595386	1,7081408	1,4851713	1,3163451	1,1777160	1,0583844
26	2,7787145	2,0555294	1,7056179	1,4833625	1,3149719	1,1766394	1,0575232
27	2,7706830	2,0518305	1,7032884	1,4816916	1,3137029	1,1756443	1,0567270
28	2,7632625	2,0484071	1,7011309	1,4801434	1,3125268	1,1747218	1,0559887
29	2,7563859	2,0452296	1,6991270	1,4787048	1,3114336	1,1738642	1,0553022
30	2,7499957	2,0422725	1,6972609	1,4773647	1,3104150	1,1730649	1,0546623
31	2,7440419	2,0395134	1,6955188	1,4761131	1,3094635	1,1723181	1,0540644
32	2,7384815	2,0369333	1,6938887	1,4749418	1,3085728	1,1716189	1,0535045
33	2,7332766	2,0345153	1,6923603	1,4738431	1,3077371	1,1709628	1,0529790
34	2,7283944	2,0322445	1,6909243	1,4728105	1,3069516	1,1703459	1,0524849
35	2,7238056	2,0301079	1,6895725	1,4718382	1,3062118	1,1697649	1,0520194
36	2,7194846	2,0280940	1,6882977	1,4709212	1,3055139	1,1692167	1,0515802
37	2,7154087	2,0261925	1,6870936	1,4700547	1,3048544	1,1686986	1,0511651
38	2,7115576	2,0243942	1,6859545	1,4692348	1,3042302	1,1682082	1,0507721
39	2,7079132	2,0226909	1,6848751	1,4684578	1,3036386	1,1677433	1,0503995
40	2,7044593	2,0210754	1,6838510	1,4677204	1,3030771	1,1673020	1,0500458
41	2,7011813	2,0195410	1,6828780	1,4670197	1,3025434	1,1668826	1,0497095
42	2,6980662	2,0180817	1,6819524	1,4663529	1,3020355	1,1664834	1,0493895
43	2,6951021	2,0166922	1,6810707	1,4657177	1,3015516	1,1661030	1,0490846

44	2,6922783	2,0153676	1,6802300	1,4651119	1,3010901	1,1657402	1,0487936
45	2,6895850	2,0141034	1,6794274	1,4645335	1,3006493	1,1653936	1,0485158
46	2,6870135	2,0128956	1,6786604	1,4639807	1,3002280	1,1650624	1,0482501
47	2,6845556	2,0117405	1,6779267	1,4634518	1,2998249	1,1647454	1,0479959
48	2,6822040	2,0106348	1,6772242	1,4629453	1,2994389	1,1644418	1,0477524
49	2,6799520	2,0095752	1,6765509	1,4624598	1,2990688	1,1641507	1,0475190
50	2,6777933	2,0085591	1,6759050	1,4619940	1,2987137	1,1638714	1,0472949
51	2,6757222	2,0075838	1,6752850	1,4615468	1,2983727	1,1636032	1,0470798
52	2,6737336	2,0066468	1,6746892	1,4611170	1,2980450	1,1633454	1,0468730
53	2,6718226	2,0057460	1,6741162	1,4607037	1,2977298	1,1630975	1,0466741
54	2,6699848	2,0048793	1,6735649	1,4603059	1,2974265	1,1628588	1,0464826
55	2,6682160	2,0040448	1,6730340	1,4599228	1,2971343	1,1626289	1,0462982
56	2,6665124	2,0032407	1,6725223	1,4595535	1,2968527	1,1624073	1,0461204
57	2,6648705	2,0024655	1,6720289	1,4591974	1,2965810	1,1621936	1,0459489
58	2,6632870	2,0017175	1,6715528	1,4588538	1,2963189	1,1619873	1,0457833
59	2,6617588	2,0009954	1,6710930	1,4585219	1,2960657	1,1617881	1,0456234
60	2,6602830	2,0002978	1,6706489	1,4582013	1,2958211	1,1615955	1,0454689
61	2,6588571	1,9996236	1,6702195	1,4578913	1,2955846	1,1614094	1,0453196
62	2,6574786	1,9989715	1,6698042	1,4575914	1,2953558	1,1612293	1,0451750
63	2,6561450	1,9983405	1,6694022	1,4573011	1,2951343	1,1610550	1,0450351
64	2,6548543	1,9977297	1,6690130	1,4570201	1,2949198	1,1608861	1,0448996
65	2,6536045	1,9971379	1,6686360	1,4567478	1,2947120	1,1607226	1,0447683
66	2,6523935	1,9965644	1,6682705	1,4564838	1,2945106	1,1605640	1,0446410
67	2,6512197	1,9960084	1,6679161	1,4562278	1,2943152	1,1604102	1,0445176
68	2,6500813	1,9954689	1,6675723	1,4559795	1,2941256	1,1602609	1,0443978
69	2,6489768	1,9949454	1,6672385	1,4557384	1,2939416	1,1601161	1,0442815
70	2,6479046	1,9944371	1,6669145	1,4555042	1,2937629	1,1599754	1,0441685
71	2,6468634	1,9939434	1,6665997	1,4552768	1,2935893	1,1598387	1,0440588
72	2,6458519	1,9934636	1,6662937	1,4550557	1,2934205	1,1597058	1,0439521
73	2,6448688	1,9929971	1,6659962	1,4548408	1,2932564	1,1595766	1,0438484
74	2,6439129	1,9925435	1,6657069	1,4546317	1,2930968	1,1594509	1,0437475
75	2,6429831	1,9921022	1,6654254	1,4544282	1,2929415	1,1593286	1,0436493
76	2,6420783	1,9916726	1,6651514	1,4542302	1,2927903	1,1592095	1,0435537
77	2,6411976	1,9912544	1,6648845	1,4540374	1,2926430	1,1590936	1,0434606
78	2,6403400	1,9908471	1,6646246	1,4538495	1,2924996	1,1589806	1,0433699
79	2,6395046	1,9904502	1,6643714	1,4536665	1,2923598	1,1588705	1,0432815
80	2,6386906	1,9900634	1,6641246	1,4534881	1,2922236	1,1587632	1,0431953
81	2,6378971	1,9896863	1,6638839	1,4533141	1,2920907	1,1586586	1,0431113
82	2,6371234	1,9893186	1,6636492	1,4531444	1,2919611	1,1585565	1,0430294
83	2,6363688	1,9889598	1,6634202	1,4529788	1,2918347	1,1584569	1,0429494
84	2,6356325	1,9886097	1,6631967	1,4528173	1,2917113	1,1583597	1,0428713
85	2,6349139	1,9882679	1,6629785	1,4526595	1,2915908	1,1582648	1,0427951

Источник: <https://studfile.net/preview/6063776/>

Таблица значений F-критерия Фишера при уровне значимости $\alpha = 0,05$

$k_1 \backslash k_2$	1	2	3	4	5	6	8	12	24	∞
1	161,45	199,50	215,72	224,57	230,17	233,97	238,89	243,91	249,04	234,52
2	18,51	19,00	19,16	19,25	19,30	19,33	19,37	19,41	19,45	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,84	8,74	8,64	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,91	5,77	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,68	4,53	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,00	3,84	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,57	3,41	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,28	3,12	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,07	2,90	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,91	2,74	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	2,95	2,79	2,61	2,40
12	4,75	3,88	3,49	3,26	3,11	3,00	2,85	2,69	2,50	2,30
13	4,67	3,80	3,41	3,18	3,02	2,92	2,77	2,60	2,42	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,53	2,35	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,64	2,48	2,29	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,59	2,42	2,24	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,55	2,38	2,19	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,34	2,15	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,48	2,31	2,11	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,28	2,08	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,42	2,25	2,05	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,40	2,23	2,03	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,38	2,20	2,00	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,36	2,18	1,98	1,73
25	4,24	3,38	2,99	2,76	2,60	2,49	2,34	2,16	1,96	1,71
26	4,22	3,37	2,98	2,74	2,59	2,47	2,32	2,15	1,95	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,30	2,13	1,93	1,67
28	4,20	3,34	2,95	2,71	2,56	2,44	2,29	2,12	1,91	1,65
29	4,18	3,33	2,93	2,70	2,54	2,43	2,28	2,10	1,90	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,09	1,89	1,62
35	4,12	3,26	2,87	2,64	2,48	2,37	2,22	2,04	1,83	1,57
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2,00	1,79	1,51
45	4,06	3,21	2,81	2,58	2,42	2,31	2,15	1,97	1,76	1,48
50	4,03	3,18	2,79	2,56	2,40	2,29	2,13	1,95	1,74	1,44
60	4,00	3,15	2,76	2,52	2,37	2,25	2,10	1,92	1,70	1,39
70	3,98	3,13	2,74	2,50	2,35	2,23	2,07	1,89	1,67	1,35
80	3,96	3,11	2,72	2,49	2,33	2,21	2,06	1,88	1,65	1,31
90	3,95	3,10	2,71	2,47	2,32	2,20	2,04	1,86	1,65	1,31
100	3,94	3,09	2,70	2,46	2,30	2,19	2,03	1,85	1,63	1,26
125	3,92	3,07	2,68	2,44	2,29	2,17	2,01	1,83	1,60	1,21
150	3,90	3,06	2,66	2,43	2,27	2,16	2,00	1,82	1,59	1,18
200	3,89	3,04	2,65	2,42	2,26	2,14	1,98	1,80	1,57	1,14
300	3,87	3,03	2,64	2,41	2,25	2,13	1,97	1,79	1,55	1,10
400	3,86	3,02	2,63	2,40	2,24	2,12	1,96	1,78	1,54	1,07
500	3,86	3,01	2,62	2,39	2,23	2,11	1,96	1,77	1,54	1,06
1000	3,85	3,00	2,61	2,38	2,22	2,10	1,95	1,76	1,53	1,03
∞	3,84	2,99	2,60	2,37	2,21	2,09	1,94	1,75	1,52	1,00