

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ
«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»

Московский институт электроники и математики им. А.Н. Тихонова
Шембель Даниил Альбертович, группа БИТ213

Модульная работа №2

по дисциплине «Первичная обработка и
представление статистических данных»

Тема: «Двумерные данные»

Дата сдачи отчета: 7.12.22

Москва 2022 г.

Содержание

Оглавление

Содержание.....	2
Тема исследования.....	3
Актуальность темы исследования	3
Исследуемые показатели.....	3
Корреляционный анализ.....	3
Регрессионный анализ	3
Линейная модель	4
Степенная модель.....	4
Гиперболическая модель	5
Сравнение моделей	5
Заключение	6
Приложения.....	6
Таблица исходных данных	6
Рис. 1 Корреляционное облако	8
Рис. 3 Степенная линия тренда MS Excel	9
Рис. 5 Полученная гиперболическая линия тренда.....	10
Литература	10
Интернет Источники.....	10

Тема исследования

В качестве темы было выбрано нахождение зависимости между отношением заработной платы (отношение заработной платы в регионе к среднероссийской) в субъектах и миграционным приростом Субъекта в 2020 году.

Актуальность темы исследования

В экономически развитых странах внутренняя миграция населения оказывает существенное влияние на формирование и функционирование внутренних рынков труда и является основной силой перераспределения населения между регионами в ответ на экономические изменения. Поэтому в развитых странах внутренняя миграция населения способствует более полному использованию национальных и региональных трудовых ресурсов, сглаживает региональные различия в уровне развития, способствует устойчивому социально-экономическому развитию. Исследования, проведенные на основе российских данных, показывают, что региональные различия в доходах и уровне безработицы со временем не только не уменьшаются, но даже увеличиваются. Современная Россия характеризуется очень неравномерным территориальным размещением производительных сил. Наличие региональной замкнутости региональных рынков труда приводят к устойчивым карманам длительно безработных, хотя многие регионы уже сталкиваются с нехваткой рабочей силы в определенных отраслях и профессиях. Об остроте проблемы увеличения внутри территориальной миграции населения в России можно судить, в том числе, по большим межрегиональным колебаниям заработной платы между регионами.

Исследуемые показатели

Отношение заработной платы по субъектам- отношение средней заработной платы субъекта к среднероссийской заработной плате. (Безразмерная величина)

Миграционный прирост населения в субъекте. (Кол-во человек)

Корреляционный анализ

По корреляционному облаку ([Рис. 1 Приложения](#)) можно сказать, что наблюдается слабая прямая корреляция. При проведении расчётов со значимостью $\alpha=0,05$, найденный коэффициент R-квадрат это подтверждает (Лист Регрессионный анализ MS Excel, R-квадрат=0,06). Рассчитаем, коэффициент корреляции на основе исходных данных (Лист линейная регрессия, коэффициент корреляции= 0,25), по полученному числу можно судить о том, что слабая взаимосвязь всё-таки имеет место быть.

Регрессионный анализ

Линейная модель

Построим линейную модель регрессии по исходным данным. Для этого рассчитаем $x \cdot y$ и x^2 , после чего рассчитаем коэффициент S_x :

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

, а затем коэффициент b_1 :

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

и коэффициент b_0 :

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

Запишем получившееся уравнение:

$$y = -2418,18 + 5083,811x$$

Сравним полученную модель с линией тренда линейной модели в MS Excel (Рис. 2 Приложения). Как видно уравнения полностью совпадают. Рассчитаем МНК для данной модели, чтобы впоследствии сравнить с МНК нелинейных моделей, чтобы понять какая из них наиболее точная. Получим МНК=1000593366.

Проверим гипотезу о значимости уравнения регрессии, для этого рассчитаем F-критерий Фишера:

$F = \frac{R^2}{1-R^2} * (n - k - 1) = 2,1132$, что меньше табличного значения при значимости $\gamma=0,05$, следовательно, модель значима, теперь проверим коэффициенты:

$$t = \frac{b_1}{S_{b_1}} \text{ и } t = \frac{b_0}{S_{b_0}}, \text{ где } S_{b_1} \text{ и } S_{b_0}:$$

$S_{b_1} = \frac{S_y}{\sqrt{\sum(x_i - \bar{x})^2}}$ и $S_{b_0} = S_y \sqrt{\frac{\sum x}{n * \sum(x_i - \bar{x})^2}}$, где S_y -стандартная ошибка оценки:

$$S_y = \sqrt{\frac{\sum(y_i - \hat{y})^2}{n-k-1}}, \text{ получим:}$$

$t_{b_1} = 2,361561392$, $t_{b_0} = 0,763218126$, что меньше табличного значения при значимости $\gamma=0,05$, следовательно, практически невероятно, что найденные значения параметров обусловлены только случайными совпадениями.

Степенная модель

Построим степенную модель, для этого рассчитаем следующие коэффициенты: $\lg x$, $\lg y$, $\lg x * \lg y$, $\lg^2 x$, в свою очередь с их помощью

рассчитаем:

$$S_{\lg x}^2 = \overline{lg^2 x} - (\overline{lg x})^2$$
$$b_1 = \frac{\overline{lg x * lg y} - \overline{lg x} * \overline{lg y}}{S_{\lg x}^2}$$
$$lg b_0 = \overline{lg y} - b_1 * \overline{lg x}$$

После чего можем записать уравнение модели в виде:

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

$\hat{y} = 1495,874 * x^{1,458008}$, сравним полученную модель с версией MS Excel.
([Рис. 3 Приложения](#))

Уравнения не совпадают. Рассчитаем МНК модели:
МНК=1034044074.

Гиперболическая модель

Построим гиперболическую модель, для этого рассчитаем коэффициенты, с учётом того, что произведена замена $z = \frac{1}{x}$

$$S_z^2 = \overline{z^2} - \bar{z}^2$$
$$b_1 = \frac{\overline{zy} - \bar{z} * \bar{y}}{S_z^2}$$
$$b_0 = \bar{y} - b_1 \bar{z}$$

В конечном итоге получим уравнение вида:

$$\hat{y} = b_0 + b_1 * \frac{1}{x}$$
$$\hat{y} = 8577,881 - 5185,04 * \frac{1}{x}$$

Сравним полученную модель с полиномиальной линией тренда MS Excel, так как MS Excel не предлагает нам гиперболическую линию тренда. ([Рис. 4 и Рис. 5 Приложения](#))

Несмотря на разные типы линий, они достаточно близки, что может служить подтверждением правильности гиперболической модели. Рассчитаем МНК модели:

МНК=988313406,2

Сравнение моделей

Сравним модели по МНК:

МНК_{лин}= 1 000 593 366

МНК_{степ}=1 034 044 074

МНК_{гип}=988 313 406,2

Наименьшая МНК-оценка у гиперболической модели, следовательно, её можно признать наиболее оптимальной, второй по малости оказалась МНК-оценка линейной модели, но при этом стоит выделить, что у всех трёх моделей очень близкая МНК.

Заключение

В результате исследования была выявлена слабая корреляция между отношением средней зарплаты в Субъекте РФ к среднероссийской заработной плате и миграционным приростом данного Субъекта, тем не менее была исключена вероятность воздействия случайных величин, следовательно, можно сделать вывод, что относительная зарплата в Субъекте РФ может служить одним из факторов его миграционного прироста, скорее всего подобными факторами могут служить географическое положение Субъекта и климат, но это лишь гипотеза. Также в ходе работы были построены 3 регрессионные модели 1 линейная и 2 нелинейных (степенная и гиперболическая), по МНК-оценке гиперболическая модель оказалась наиболее оптимальной из всех 3-ёх.

Приложения

Таблица исходных данных

	x	y
Белгородская область	0,729238	802
Брянская область	0,622195	-285
Владимирская область	0,686351	-3024
Воронежская область	0,707327	348
Ивановская область	0,566434	-868
Калужская область	0,856848	1599
Костромская область	0,627532	1382
Курская область	0,697355	1052
Липецкая область	0,716539	-778
Московская область	1,130921	14365
Орловская область	0,620559	-1880
Рязанская область	0,710093	337
Смоленская область	0,645431	-1237
Тамбовская область	0,604998	395

Тверская область	0,702653	-177
Тульская область	0,796373	-982
Ярославская область	0,7366	-1961
г. Москва	1,949011	9076
Республика Карелия	0,905675	626
Республика Коми	1,113197	-3614
Архангельская область	1,08856	0
Вологодская область	0,833106	-1579
Калининградская область	0,713754	7708
Ленинградская область	0,940441	21166
Мурманская область	1,346506	-3619
Новгородская область	0,665492	685
Псковская область	0,613431	725
г.Санкт-Петербург	1,337391	1026
Республика Адыгея	0,626383	-99
Республика Калмыкия	0,6235	160
Республика Крым	0,665725	1429
Краснодарский край	0,749825	17553
Астраханская область	0,757343	-2427

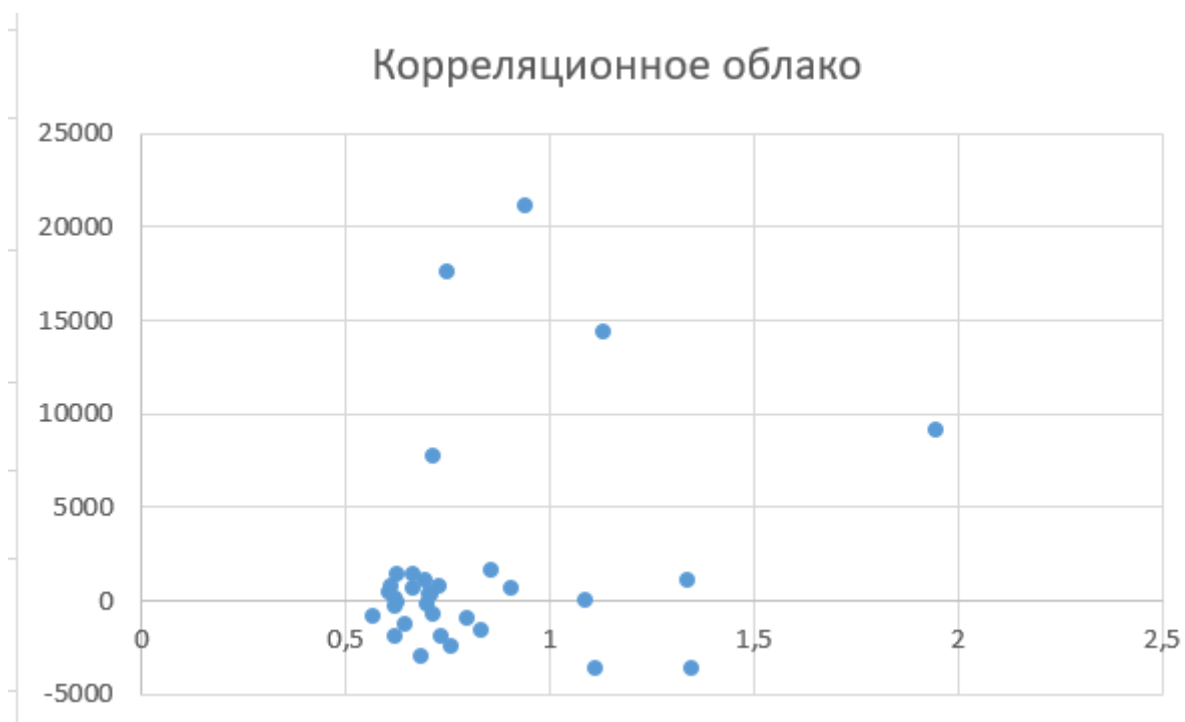


Рис. 1 Корреляционное облако

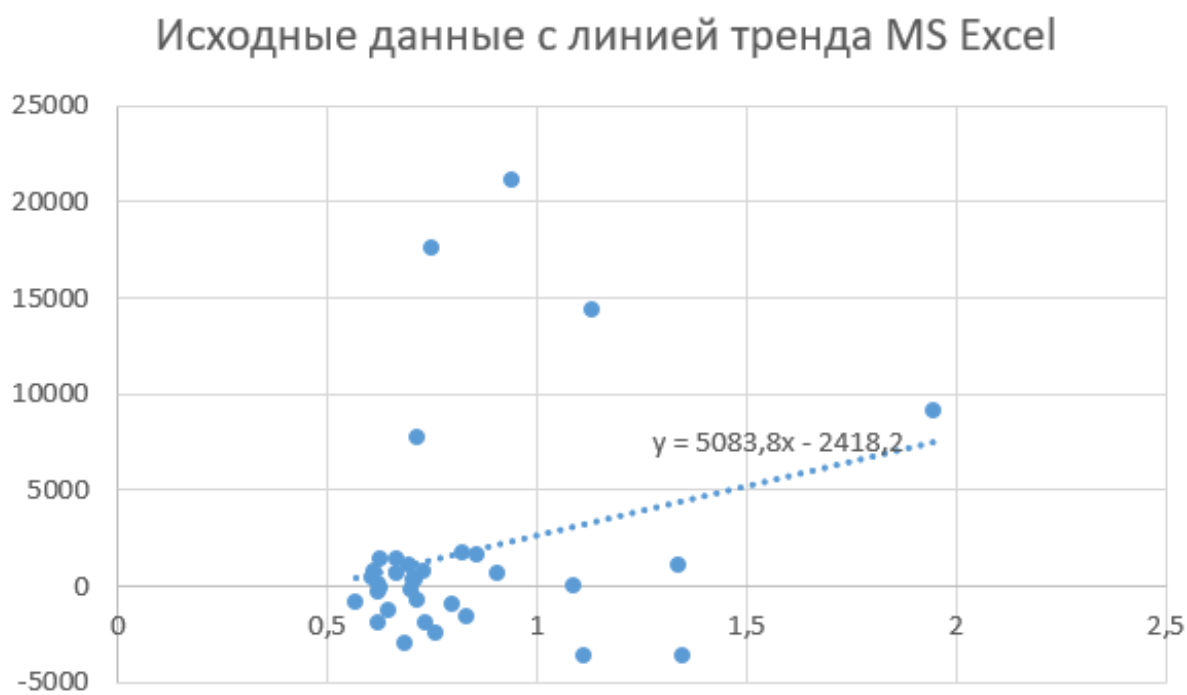


Рис. 2 Линейная линия тренда MS Excel

Степенная линия тренда MS Excel

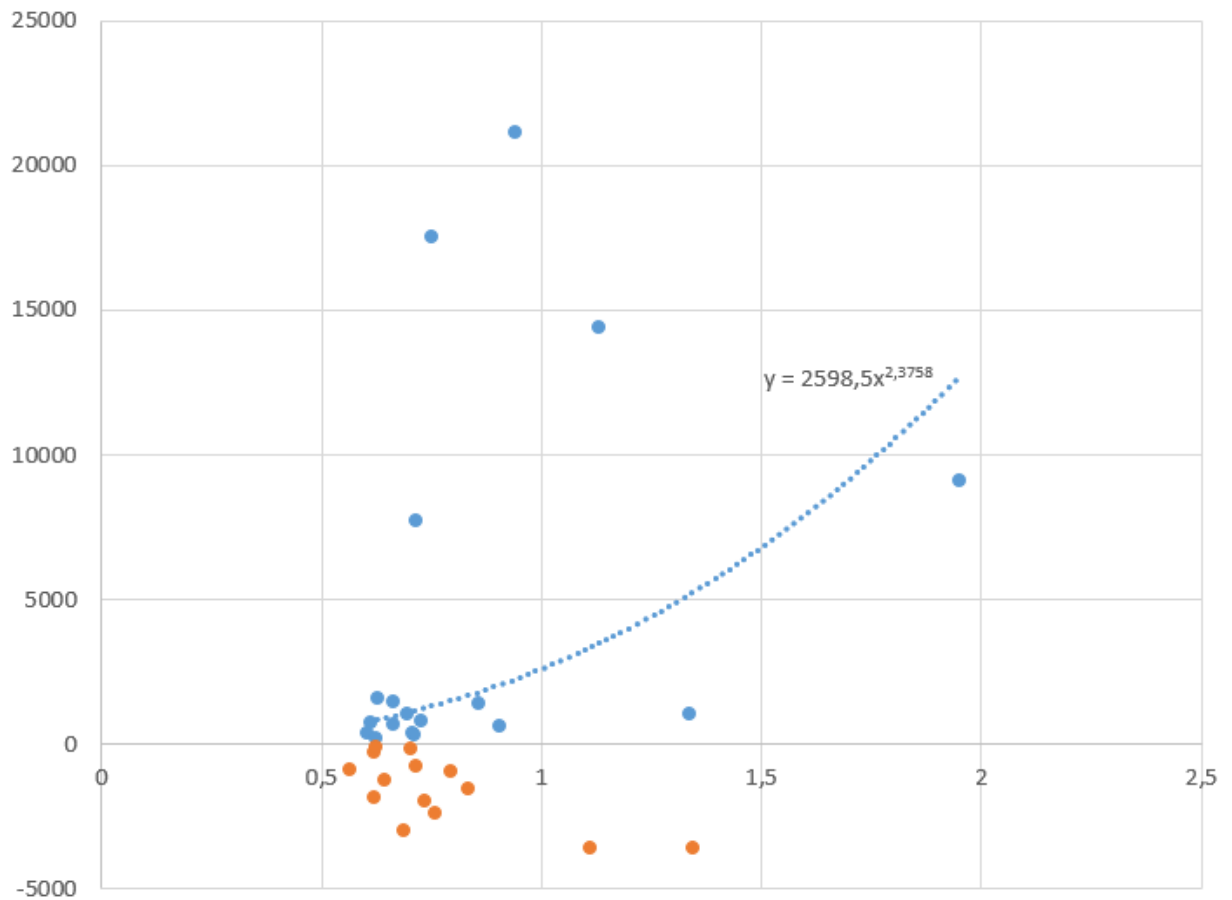


Рис. 3 Степенная линия тренда MS Excel

Исходные данные с полиномиальной линией тренда MS Excel

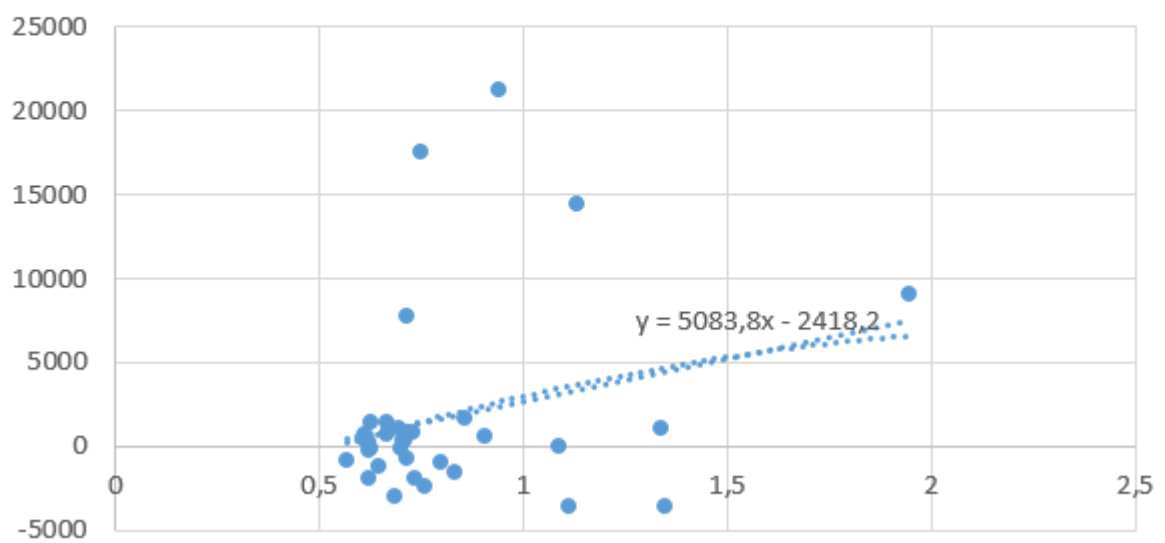


Рис. 4 Полиномиальная линия тренда MS Excel

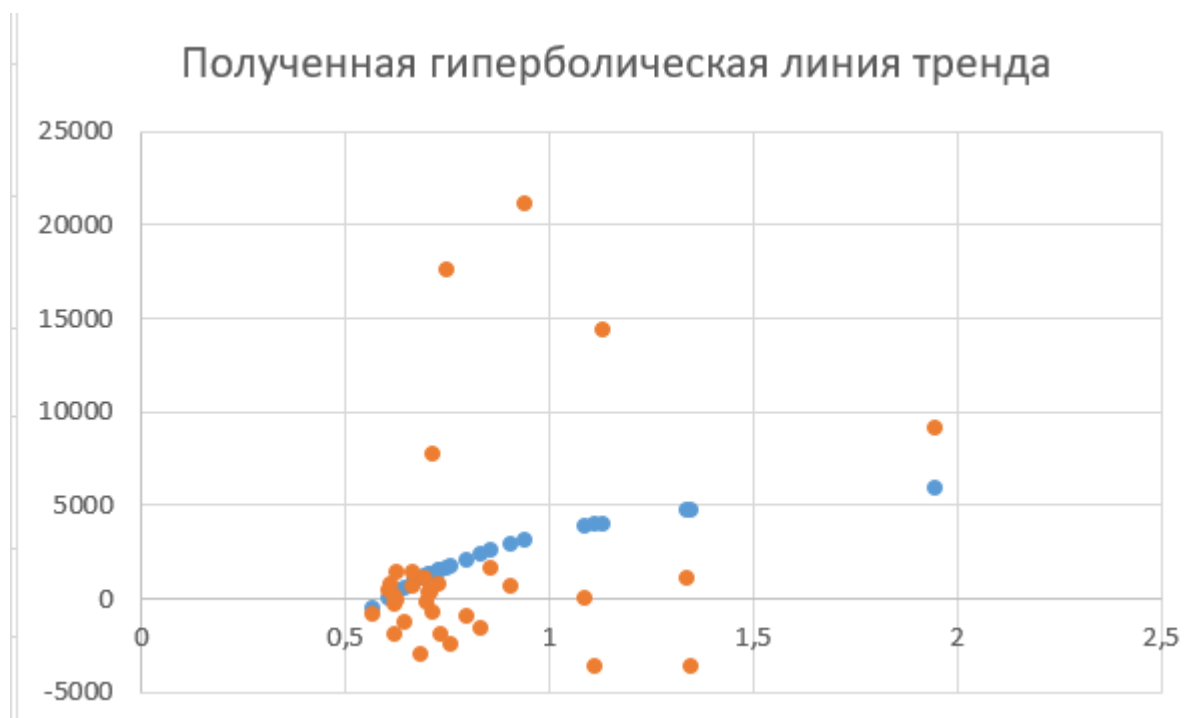


Рис. 5 Полученная гиперболическая линия тренда

Литература

Анализ данных: учебник для академического бакалавриата / В. С. Мхитарян [и др.] ; под ред. В. С. Мхитаряна. — М. : Издательство Юрайт, 2016. — 490 с. — (Серия : Бакалавр. Академический курс). — ISBN 978-5-9916-5591-0. — Режим доступа : www.biblio-online.ru/book/AF1D197F-1759-422E-9593-8B43E2D1093B

Интернет Источники

<https://www.rbc.ru>

<https://stats.hh.ru/>

<https://rosstat.gov.ru/folder/10705>