

Титульный лист

Национальный исследовательский университет «Высшая школа экономики»

Московский институт электроники и математики им. А.Н. Тихонова

Образовательная программа: Инфокоммуникационные технологии и системы связи

Отчет по Модульному заданию №1

«Пространственные данные»

по майнору «Прикладной статистический анализ»

Студент: Шембель Даниил Альбертович

Группа: БИТ213

МОСКВА 2022

Оглавление

Вступление	3
Используемый показатель для анализа	3
Описание показателя, выбранного для исследования	3
Анализ вариационных рядов	5
Основные числовые характеристики одномерных количественных данных	8
График	13
Проверка тенденции временного ряда	13
Вывод	14
Источники:	14
Приложение 1:	15
Приложение 2:	21

Вступление

Большой объем перевозок- индикатор уровня развития экономики. Автомобильные грузоперевозки- это одно из самых динамично развивающихся направлений транспорта. Грузооборот является отражением экономической активности государства.

Используемый показатель для анализа

В качестве исследуемых данных первой части были взяты показатели грузооборота автомобильного транспорта на коммерческой основе по Субъектам РФ за 2010 год. (тыс. тонно-км).

Описание показателя, выбранного для исследования

Грузооборот — экономический показатель работы транспорта (показатель объёма перевозок грузов), равный произведению массы перевозимого за определённое время груза на расстояние перевозки.

Грузооборот зависит от дальности перевозки и массы груза, он измеряется в тонно-километрах.

Графический анализ

Для графического анализа было построено 4 диаграммы (2 для неранжированных и 2 для ранжированных).







Построение вариационных рядов

Для анализа исходных данных потребовалось построение вариационного интервального ряда. Для выполнения этого я воспользовался Microsoft Excel. Для начала я рассчитал размах исходного ряда, количество элементов в нем. По этим данным я выполнил следующие расчеты:

Анализ вариационных рядов

Воспользовавшись формулой [Стерджеса](#) для определения количества интервалов:

$n_i = 1 + \log_2 N$, где n_i -количество интервалов, а N - объем выборки, получаем: $n_i=7$, после этого рассчитаем шаг:

$h = \frac{Max-min}{n_i}$, где h -шаг, Max - максимальное значение в исходном вариационном ряду, min -минимальное значение в исходном вариационном ряду, а n_i - число интервалов, полученное ранее. Получим: $h=198\ 101$, количество интервалов $n=7$.

По Стердженсу:									Относителная частота		Относительная частота		середина*част
				н	в	Частота	середина	кумулята	частота	кумулята	частота	кумулята	
мин	177			1	-98 874	198 278		47	49702,15	47	0,594937	0,594937	2336001,05
макс	1 447 065			2	198 278	396 379		12	297 328	59	0,151899	0,746835	3567940,8
R	1 446 889			3	396 379	594 480		6	495 429	65	0,075949	0,822785	2972576,4
n	79			4	594 480	792 581		8	693 530	73	0,101266	0,924051	5548243,2
h	198101			5	792 581	990 682		2	891 631	75	0,025316	0,949367	1783262,8
ср. знач	263 653			6	990 682	1 188 783		2	1 089 732	77	0,025316	0,974684	2179464,8

По методу квадратного корня:

$h = R/\sqrt{N}$, где R-размах вариации, N- объем выборки. Получим:

$h=162787,675$, количество интервалов $n=8$.

По формуле [СКОТТА](#) получим:

$h = 3,5 * s * N^{\frac{-1}{3}}$, где h-шаг, s- среднее отклонение, N-объем выборки. Получим:

$h=267\ 707$, количество интервалов $n=5$.

				н	в	Частота	середина	кумулята	частота	кумулята	частота	кумулята	
мин	177			1	-133676,6	134030,4		0	176,9	0			0
макс	1 447 065			2	134030,4	401737,4		38	267883,9	38			38
R	1 446 889			3	401737,4	669444,4		0	535590,9	38			38
n	79			4	669444,4	937151,4		21	803297,9	59			59
h	267707			5	937151,4	1204858,4		0	1071004,9	59			59

По формуле [Фридмана Диакониса](#) получим:

$h = 2 * (IQ) * N^{\frac{-1}{3}}$, где h- шаг, IQ- разница между нижним и верхним квартилем, N- количество измерений в исходном ряду. Получим: $h=97\ 660$, количество интервалов $n=15$.

			н	в	Частота	середина	кумулята
мин	177	1	-48653,1	49006,9	0	176,9	0
макс	1 447 065	2	49006,9	146666,9	26	97836,9	26
R	1 446 889	3	146666,9	244326,9	0	195496,9	26
n	79	4	244326,9	341986,9	16	293156,9	42
h	97660	5	341986,9	439646,9	0	390816,9	42
		6	439646,9	537306,9	14	488476,9	56
		7	537306,9	634966,9	0	586136,9	56

8	634966,9	732626,9	1	683796,9	57
9	732626,9	830286,9	0	781456,9	57
10	830286,9	927946,9	4	879116,9	61
11	927946,9	1025606,9	0	976776,9	61
12	1025606,9	1123266,9	2	1074437	63
13	1123266,9	1220926,9	0	1172097	63
14	1220926,9	1318586,9	3	1269757	66
15	1318586,9	1416246,9	0	1367417	66

Далее ведется работа с вариационно-интервальным рядом по Стерджесу.

Основные числовые характеристики одномерных количественных данных

Средние структурные величины:

Средние структурные величины	Исходный ряд	Интервальный ряд
ср. арифметическое	263 653	266118,19
медиана	137 242	2525611
мода	#Н/Д	157809,3

Для расчета медианы, моды и среднего арифметического в исходном ряду были использованные встроенные Excel функции, а для расчета тех же величин для вариационного интервального ряда были использованы следующие формулы:

$$M_o = x_o + h_o * \frac{f_{M_o} - f_{M_o-1}}{(f_{M_o} - f_{M_o-1}) + (f_{M_o} - f_{M_o+1})}, \text{ где } M_o - \text{мода, } x_o - \text{начало}$$

модального интервала, f_{M_o-1} - частота интервала, перед модальным, f_{M_o+1} - частота интервала, после модального, h_o - шаг.

$$M_e = x_e + h_e * \frac{\frac{\sum f_i}{2} - S_{Me-1}}{f_{Me}}$$

, где M_e - медиана, x_0 -начало медианного интервала, f_{me-1} -частота интервала, перед медианным, f_{me+1} -частота интервала, после медианного, h_e - шаг.

Медиана \neq мода \neq среднее арифметическое, следовательно, распределение не является нормальным. Мода < медиана < среднее арифметическая, следовательно, наблюдается левосторонняя скошенность.

Показатели вариации:

Показатель вариации	Для исходного ряда	Для интервального ряда
Дисперсия	8,87154E+11	266118,194
Среднеквадратическое отклонение	875867,189	327 704
Размах вариации	1 446 889	1 446 889
Квартили		
1 квартиль	37286,65	-526029
3 квартиль	379338,3	-371262
IQR	342051,6	154766,4
Коэффициент вариации	3	1
Квартильный показатель вариации Гальтона	171025,8	77383,2

Коэффициент вариации большой ($>0,2$), следовательно, выборка сильно вариабельна. $IQR \neq$ медиане, что еще раз подтверждает отсутствие нормального распределения.

Показатели формы распределения:

Показатель формы распределения	Для исходного ряда	Для интервального ряда
Коэф. Асимметрии	0,19	-1,239150047
Коэф. эксцесса	0,219620854	75,14910425

Коэффициент эксцесса положителен и >10 , следовательно, распределение остроконечно. При этом коэффициент асимметрии около 0, что показывает, что значения расположены достаточно симметрично относительно математического ожидания.

Аналитические показатели изменения уровней ряда динамики

Для этой части работы были взяты данные об индексе изменения фондовооруженности сельского хозяйства в период с 2008 по 2021 год. Данные представлены в приложении 2, и являются временными.

Средняя хронологическая:

средняя хронологическая

104,6

Определение абсолютных приростов:

формулы для расчетов:

$$\Delta y_t = y_t - y_{t-1}, T_t = \frac{y_t}{y_{t-1}} * 100\%$$

$$y_t^b = y_t - y_b, T_t^b = \frac{y_t}{y_b} * 100\%$$

t	Индекс	Абсолютный прирост	
		Цепной	Базисный
2008	104,0	-	-
2009	103,8	-0,2	-0,2
2010	101,6	-2,2	-2,4
2011	101,8	0,2	-2,2
2012	103,6	1,8	-0,4
2013	103,6	0,0	-0,4
2014	103,3	-0,3	-0,7
2015	103,7	0,4	-0,3
2016	104,0	0,3	0,0
2017	108,9	4,9	4,9
2018	107,0	-1,9	3,0

2019	108,5	1,5	4,5
2020	104,3	-4,2	0,3
2021	106,6	2,3	2,6

Темп роста:

		Темп роста	
t	Индекс	Цепной	Базисный
2008	104,0	-	-
2009	103,8	99,8	99,8
2010	101,6	97,9	97,7
2011	101,8	100,2	97,9
2012	103,6	101,8	99,6
2013	103,6	100,0	99,6
2014	103,3	99,7	99,3
2015	103,7	100,4	99,7
2016	104,0	100,3	100,0
2017	108,9	104,7	104,7
2018	107,0	98,3	102,9
2019	108,5	101,4	104,3
2020	104,3	96,1	100,3
2021	106,6	102,2	102,5

Темп прироста:

		Темп прироста	
t	Индекс	Цепной	Базисный
2008	104,0	-	-
2009	103,8	-0,2	-0,2
2010	101,6	-2,1	-2,3

График



Проверка тенденции временного ряда

Таблица без 3-ёх последних годов:

2008	104,0	-	-
2009	103,8	-0,2	-0,2
2010	101,6	-2,2	-2,4
2011	101,8	0,2	-2,2
2012	103,6	1,8	-0,4
2013	103,6	0,0	-0,4
2014	103,3	-0,3	-0,7
2015	103,7	0,4	-0,3
2016	104,0	0,3	0,0
2017	108,9	4,9	4,9
2018	107,0	-1,9	3,0

Сделаем прогноз:

0,3	107,3	107,6	107,9	По абсолютному приросту
1,002589	107,277	107,5547	107,8331	По среднему темпу роста

Удалённые значения:

2019	108,5
2020	104,3
2021	106,6

Прогноз не совпал с действительностью.

Вывод

В итоге был сделан анализ данных вариационного ряда, также был построен интервальный ряд разными способами (Скотт, Фридман-Диаконис, квадратный корень). Дополнительными структурными и вариационными величинами удалось подтвердить скошенность видимую на графическом представлении данных.

Во второй части работы был проведен анализ исходного временного ряда, в том числе абсолютный прирост, темп роста, темп прироста, средняя хронологическая. Также была попытка восстановления нескольких лет по анализу предыдущих, но она не увенчалась успехом, что говорит о трудности предсказания значений данного временного ряда, но несмотря на это была также сделана попытка прогнозирования на следующие три года.

Источники:

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.

Анализ данных: учебник для вузов / В. С. Мхитарян [и др.]; под редакцией В. С. Мхитаряна. — Москва: Издательство Юрайт, 2020. — ISBN 978-5-534-00616-2.

<https://rosstat.gov.ru/> - Федеральная служба государственной статистики.

Приложение 1:

	2010
Белгородская область	683 767
Брянская область	550 139
Владимирская область	137 242
Воронежская область	163 042
Ивановская область	91 424
Калужская область	43 713
Костромская область	34 432
Курская область	135 976
Липецкая область	64 297
Московская область	2 012 647
Орловская область	451 650
Рязанская область	74 180

Смоленская область	740 063
Тамбовская область	36 743
Тверская область	356 124
Тульская область	59 813
Ярославская область	207 598
г. Москва	5 878 614
Республика Карелия	196 003
Республика Коми	672 527
Архангельская область	201 144
<p>в том числе:</p> <p>Ненецкий автономный округ</p>	177
<p>Архангельская область (без Ненецкого автономного округа)</p>	200 967

Вологодская область	128 004
Калининградская область	681 483
Ленинградская область	724 778
Мурманская область	26 180
Новгородская область	246 716
Псковская область	99 189
г. Санкт-Петербург	4 181 479
Республика Адыгея	87 521
Республика Калмыкия	10 568
Краснодарский край	1 447 065
Астраханская область	43 362
Волгоградская область	234 335
Ростовская область	794 998
Республика Дагестан	5 044

Республика Ингушетия	950
Кабардино- Балкарская Республика	3 659
Карачаево- Черкесская Республика	1 955
Республика Северная Осетия-Алания	5 551
Чеченская Республика	6 527
Ставропольский край	122 780
Республика Башкортостан	678 989
Республика Марий Эл	37 830
Республика Мордовия	205 886
Республика Татарстан	1 141 113

Удмуртская Республика	179 330
Чувашская Республика	97 787
Пермский край	413 087
Кировская область	179 852
Нижегородская область	354 151
Оренбургская область	133 570
Пензенская область	47 331
Самарская область	1 316 957
Саратовская область	717 827
Ульяновская область	544 016
Курганская область	59 622
Свердловская область	634 233
Тюменская область	2 364 824

<p>В том числе:</p> <p>Ханты-Мансийский автономный округ - Югра</p>	1 110 743
<p>Ямало-Ненецкий автономный округ</p>	799 711
<p>Тюменская область (без Ханты-Мансийского и Ямало-ненецкого автономных округов)</p>	454 370
<p>Челябинская область</p>	206 578
<p>Республика Алтай</p>	10 052
<p>Республика Тыва</p>	2 329
<p>Республика Хакасия</p>	40 929

Алтайский край	66 100
Красноярский край	213 926
Иркутская область	222 110
Кемеровская область	402 553
Новосибирская область	243 501
Омская область	95 661
Томская область	139 928
Республика Бурятия	19 280
Республика Саха (Якутия)	42 462
Забайкальский край	138 840
Камчатский край	529
Приморский край	101 772

Приложение 2:

Индекс изменения фондовооруженности														
	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021 ¹⁾
сельское хозяйство	104.0	103.8	101.6	101.8	103.6	103.6	103.3	103.7	104.0	108.9	107.0	108.5	104.3	106.6