Изображение выглядит как текст, Шрифт, логотип, символ

Автоматически созданное описание**Федеральное государственное автономное образовательное учреждение высшего образования**

**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»**

Факультет:

Высшая школа бизнеса

Образовательная программа: Бизнес информатика

Название дисциплины:

«Прикладной статистический анализ»

Отчет о выполнении модульного домашнего задания №1

«Статистические методы анализа одномерных количественных данных на примере денежного дохода в Астраханской области»

Выполнила: Студентка группы ББИ226

Зайдес Анна Александровна

Москва, 2023г.

**Содержание**

1. [Введение 3](#_Toc149670472)
2. [Описание показателя 5](#_Toc149670473)
3. [Построение и анализ вариационного ряда 8](#_Toc149670474)
   1. [Построение интервального вариационного ряда для непрерывных количественных данных по формуле Стерджеса. 8](#_Toc149670475)
4. [Основные числовые характеристики одномерных количественных данных 11](#_Toc149670476)
5. [Аналитические показатели изменения уровней ряда динамики 12](#_Toc149670477)
   1. [Описание данных 12](#_Toc149670478)
   2. [Определение средних и построение прогноза 13](#_Toc149670479)
6. [Выводы 15](#_Toc149670480)
7. [Литература 16](#_Toc149670481)
8. [Используемые информационные источники 16](#_Toc149670482)
9. [Приложение 1 17](#_Toc149670483)

# Введение

**Цель**: провести анализ пространственных данных и временного ряда.

**Задачи:**

Описать показатель, выбранный для исследования с выводами:

* экономическая интерпретация (что характеризует, единицы измерения, основные закономерности);
* характеристика по шкалам;
* графический анализ (нарисовать график и проанализировать изменения, указать на наличие/отсутствие аномальных наблюдений);

Анализ вариационных рядов с выводами:

* построение интервального вариационного ряда для непрерывных количественных
* данных по формуле Стерджеса и определение наиболее оптимальной ширины
* интервала (отсутствие интервалов с нулевыми частотами/интервалов с
* минимальными частотами)
* сделать вывод об оптимальной ширине интервала для характеристики распределения признака
* графическое представление исходных данных

Основные числовые характеристики одномерных количественных данных (для дискретного и интервального ряда) с выводами, еще раз подтверждающими

* графическое распределение признака
* средние/структурные величины:
* средняя;
* медиана;
* мода;
* показатели вариации:
* дисперсия;
* среднее квадратическое отклонение;
* размах вариации;
* квартили;
* интерквартильный размах;
* коэффициент вариации;
* квартильный показатель вариации Ф. Гальтона
* показатели формы распределения:
* коэффициент асимметрии;
* коэффициент эксцесса.

Аналитические показатели изменения уровней ряда динамики с выводами:

* Указание используемого показателя для анализа (с единицами измерения);
* Описание показателя, выбранного для исследования с выводами:
* экономическая интерпретация (что характеризует, единицы измерения; основные закономерности);
* характеристика по шкалам;
* графический анализ (нарисовать график тренда и проанализировать изменения, указать на наличие/отсутствие аномальных наблюдений, т.е. отклонений от тренда);
* рассчитать среднюю хронологическую;
* определение средних абсолютных приростов:
* цепных;
* базисных;
* определение средних темпов роста:
* цепных;
* базисных;
* определение средних темпов прироста:
* цепных;
* базисных;
* построение прогноза на 3 уровня вперед на основе:
* среднего абсолютного прироста;
* среднего темпа роста;
* построение графика исходных данных и добавить на график полученные прогнозные значения;
* проверка тенденции временного ряда:
* удалить из ряда данных показатели последних 3 лет и на основе среднего абсолютного прироста/среднего темпа роста восстановить удаленные показатели;
* сделать вывод о соответствии реальных и восстановленных показателей.

**Актуальность:** в качестве пространственных данных были выбраны данные выборочного обследования домохозяйств Астраханской области во 2-м квартале 2021 года, а именно денежный доход в тыс. руб. Эти данные очень важны для анализа, ведь их показатели влияют на решения многих государственных и социальных служб. Например, процентная ставка на кредиты, стоимость тех или иных услуг и так далее.

В качестве временных данных были выбраны средние номинальные заработные платы, чтобы работа осуществлялась примерно в одной теме.

# Описание показателя

Данные показывают на доход населения в Астраханской области в тыс. руб. Используется шкала отношений, так как значения непрерывны, можно измерить различие между ними и отношение, а также есть начала отсчета, то есть 0.

График 1. Точечный график исходных неранжированных данных

График 2. Столбиковая диаграмма исходных неранжированных данных

**Вывод:** уже можем заметить некоторые выбросы, например как сильно выделяется точки в окрестностях 600 и 500 тысяч. Также по такому графику можно примерно назвать моду – промежуток от 100 до 200 тысяч, но трудно назвать медиану и среднее значение.

График 3. Точечный график исходных ранжированных данных

График 4. Столбиковая диаграмма исходных ранжированных данных

**Вывод:** по ранжированным данным можноуже примерно понять медиану – число близкое к 200 тысячам. И также на графиках ранжированных данных хорошо видны выбросы. Поэтому можно сделать вывод, что большие доходы имеют единицы, а большинство людей имеют средние доходы (100-200 тыс. руб.).

# Построение и анализ вариационного ряда

# Построение интервального вариационного ряда для непрерывных количественных данных по формуле Стерджеса.

После расчета ширина интервала по Стерджесу мы можем построить сам вариационный ряд, в данном случае из 10 интервалов (Приложение 1). В данном вариационном ряде отсутствуют интервалы с нулевыми частотами, что уже хорошо, но есть интервалы, где частота минимальна – 1. Для наглядности можно обратиться к построенным графикам:

График 5. Полигон интервального ряда

На полигоне мы можем увидеть резкий спад и острую вершину, что говорит о не очень оптимальной ширине ряда. Для большей наглядности также прикладываю гистограмму

График 6. Гистограмма ИВР

График 7. Полигон накопленных относительных частот ИВР

Данный полигон представляет собой слишком пологую линию под конец графика, что также говорит о неоптимальной ширине. Для больше наглядности также прикладываю Кумуляту.

График 8. Кумулята накопленных относительных частот ИВР

График 9. Огива ИВР

**Вывод:** при ручном подборе оптимальной ширины интервала было найдено значение 148, при котором медиана и мода почти равны. Было замечено, что при уменьшении ширины ряда появляются нулевые частоты из-за того, что присутствуют явные выбросы, поэтому пришлось повышать ширину, чтобы избавиться от нулей. Но тогда график выходит еще более резкий (График 9). Зато лучше видно какой доход у подавляющего большинства и что больше 400 тысяч рублей зарабатывают совсем единицы. Поэтому, я считаю, что вычисленная ширина по Стерджесу наиболее оптимальна для данного ряда, так как из-за выбросов не сделать лучше.

График 10. Гистограмма ИВР при подобранной ширине

# Основные числовые характеристики одномерных количественных данных

В таблице 1 и 2 представлены все числовые характеристики для исходного ряда и ИВР соответственно. Все расчеты можно посмотреть в Приложении 1.

Таблица 1. Основные числовые характеристики исходного ряда

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ***А. Моментные показатели*** | | ***В. Порядковые показатели*** | | ***С. Показатели вариации*** | |
| **Центральный момент 2ого порядка** | 7266,85 | **1й квартиль** | 125,90 | **средняя арифметическая** | 187,60 |
| **Центральный момент 3ого порядка** | 826528,05 | **3й квартиль** | 236,35 | **медиана** | 174,44 |
| **Центральный момент 4ого порядка** | 340940786,18 | **IQR** | 110,45 | **мода** | 153,53 |
| **Коэффициент ассиметрии Ac** | 1,340967643 | **d1** | 93,46 | **дисперсия** | 7266,85 |
| **Коэффициент эксцесса Ek** | 3,534917923 | **d9** | 297,22 |  |  |
|  |  | **Децильный коэффицент** | 3,18 |  |  |

Таблица 2. Основные числовые характеристики вариационного ряда

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ***А. Моментные показатели*** | | ***В. Порядковые показатели*** | | ***С. Показатели вариации*** | |
| **Центральный момент 2ого порядка** | 7584,01 | **1й квартиль** | 121,89 | **средняя арифметическая** | 187,60 |
| **Центральный момент 3ого порядка** | 787963,29 | **3й квартиль** | 225,51 | **медиана** | 174,44 |
| **Центральный момент 4ого порядка** | 57242709591,38 | **IQR** | 103,62 | **мода** | 153,53 |
| **Коэффициент ассиметрии Ac** | 1,193047883 | **d1** | 87,52 |  |  |
| **Коэффициент эксцесса Ek** | 992,2282703 | **d9** | 297,28 |  |  |
|  |  | **Децильный коэффицент** | 3,40 |  |  |

**Вывод:** на основе данных из таблиц мы можем сделать вывод, что у ряда правосторонняя скошенность (так как Ас > 0 в обеих таблицах) – что мы и можем увидеть на графике 5 и графике 3, если посмотреть на него сбоку. По коэффициенту эксцесса можно сказать, что ряд островершинный (так как Ек > 0) – что также видно на данных графиках. По показателю IQR мы можем видеть, что туда попадает больше минимальных значений, что подтверждает правостороннюю скошенность. Среднее арифметическое больше медианы и моды, что подтверждает еще раз скошенность вправо.

# Аналитические показатели изменения уровней ряда динамики

# Описание данных

В качестве временных данных были взяты показатели с 1993 по 2022 год средней номинальной заработной платы (Приложение 1). Был взят такой ряд, чтобы далеко не отходить от темы. Единицы измерения – рубли. Экономическая интерпретация та же, что и была описана выше. Так как данные почти такие же, то и шкала отличаться не будет, то есть в данном случае шкала отношений.

График 11. График тренда

**Вывод:** по данному графику видно, что средняя зарплата с каждым годом растет. Сразу можно увидеть, что самый большой прирост был в 2022, а наименьшие изменения наблюдались в 90ых, 2008 – 2010 и в 2015, скорее всего такие показатели были вызваны государственным кризисом

# Определение средних и построение прогноза

Первым делом была посчитана средняя хронологическая - 20141,36 рублей. (Все расчеты можно посмотреть в Приложении 1). После подсчетов абсолютных приростов, темпов роста и темпов прироста (цепных и базисных) была составлена таблица средних значений:

Таблица 3. Средние значения

|  |  |  |  |
| --- | --- | --- | --- |
| **Средние показатели** | **Абсолютного прироста** | **Темпа роста** | **Темпа прироста** |
| **Цепной** | 2251,01 | 127,37% | 27,37% |
| **Базисный** | 9737,96 | 72,90% | -27,10% |

Теперь имея эти данные, можно сделать прогноз. Все показатели можно посмотреть в Приложении 1. Ниже представлен график с прогнозом на основе двух показателей – среднего абсолютного прироста и среднего.

График 12. График прогноза показателей

**Вывод**: по моему мнению прогноз составленный на основе темпа роста больше похож на правду, так как средний абсолютный прирост не очень учитывает тенденцию последних лет.

Давайте проверим наши прогнозы для 2020 – 2022 годов.

|  |  |  |  |
| --- | --- | --- | --- |
| **По прогнозу** | **на основе абс. Прироста** | **на основе темпа роста** | **настоящие данные** |
| 2020 | 49719,01 | 60457,98 | 51344,00 |
| 2021 | 51970,02 | 77002,76 | 56545,00 |
| 2022 | 54221,03 | 98075,15 | 65338,00 |

Можем увидеть, что прогноз на основе абсолютного прироста слишком дает слишком меленькие значения, а на основе темпа роста дает слишком большие значения, это все потому, что дельты обоих показателей очень разные, поэтому ни на один из данных прогнозов нельзя полагаться. Но можно заметить, что среднее арифметическое двух прогнозов дает прогноз, похожий на правду на следующий год. Если провести прогноз по такому алгоритму, то на 2023 год можно ожидать среднюю заработную плату в размере около 76 тыс. рублей.

# Выводы

В данной работе были освоены навыки анализа интервальных и временных рядов. Я научилась составлять прогнозы, строить графики и анализировать их, а также по числам определять, как будет выглядеть диаграмма. Что касается экономических выводов: если считать, что в среднем в домохозяйстве доход приносят 2 человека, то средние доходы в Астраханской области больше двух средних зарплат за прошедший год, это может говорить о нескольких вещах:

1. Было взято слишком мало домохозяйств для вычисления точных значений
2. Доход домохозяйств составляет не только заработная плата (например это могут быть деньги со сдачи квартиры или другие законные/незаконные формы дохода)
3. В Астраханской области живут более обеспеченные люди, чем в целом по стране

Также в выводе хочется сказать, что очень жаль, что большие доходы имеют лишь единицы, но так много средних и маленьких показателей, но можно лишь надеется на то, что прогнозы окажутся правдой и средняя заработная плата будет расти!

# Литература

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.

# Используемые информационные источники

<https://finance.rambler.ru/money/45669288-chto-budet-s-dohodami-grazhdan-i-pochemu-oni-vazhny-dlya-rosta-ekonomiki/>

<http://sophist.hse.ru/hse/nindex.shtml>

# Приложение 1

[Приложение.xlsx](file:///D:\Desktop\мои%20файлы\учеба%20ВШЭ\майнор\Модульная%20работа_приложение.xlsx)