

Министерство науки и высшего образования Российской Федерации

Томский государственный университет
систем управления и радиоэлектроники

З. П. Лепихина

ПРИКЛАДНОЙ СТАТИСТИЧЕСКИЙ АНАЛИЗ

Методические указания к лабораторным работам
и организации самостоятельной работы
для студентов направления «Государственное и муниципальное управление»
(уровень магистратуры)

Томск
2022

УДК 311
ББК 22.172
Л48

Рецензент:

Сидоров А.А., заведующий кафедрой автоматизации обработки информации
Томского государственного университета
систем управления и радиоэлектроники, канд. экон. наук, доцент

Лепихина Зинаида Павловна

Л48 Прикладной статистический анализ: Методические указания к лабораторным работам и организации самостоятельной работы для студентов направления «Государственное и муниципальное управление» (уровень магистратуры) / З. П. Лепихина. – Томск : Томск. гос. ун-т систем упр. и радиоэлектроники, 2022. – 45с.

Методические указания к лабораторным работам и организации самостоятельной работы по дисциплине «Прикладной статистический анализ» содержат необходимые разъяснения по форме организации и содержанию лабораторных работ и самостоятельной деятельности студентов. Выполнение данных указаний будет способствовать успешному освоению дисциплины, в том числе выработке навыков по получению статистической информации, ее обобщения и анализа с использованием методического инструментария, математического аппарата и прикладных цифровых решений для информационно-аналитической поддержки принятия управленческих решений.

Для студентов высших учебных заведений, обучающихся по направлению «Государственное и муниципальное управление» (уровень магистратуры).

Одобрено на заседании кафедры АОИ, протокол № 1 от 20.01.2022

УДК 311
ББК 22.172

© Лепихина З. П., 2022
© Томск. гос. ун-т систем упр.
и радиоэлектроники, 2022

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ.....	4
1 МЕТОДИЧЕСКИЕ УКАЗАНИЯ К ЛАБОРАТОРНЫМ РАБОТАМ.....	5
1.1 Лабораторная работа «Сбор и представление статистических данных».....	5
1.2 Лабораторная работа «Первичный анализ социально-экономической информации».....	6
1.3 Лабораторная работа «Построение и исследование типологии социально-экономических объектов методом статистической группировки».....	10
1.4 Лабораторная работа «Построение типологии социально-экономических объектов методами кластер-анализа».....	12
1.5 Лабораторная работа «Методы исследования взаимосвязей нечисловых данных».....	15
1.6 Лабораторная работа «Исследования взаимосвязей числовых данных методами аналитической группировки и корреляционно-регрессионного анализа».....	20
1.7 Лабораторная работа «Анализ и прогнозирование динамики развития социально-экономических процессов».....	24
2 МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОРГАНИЗАЦИИ САМОСТОЯТЕЛЬНОЙ РАБОТЫ.....	28
2.1 Общие положения.....	28
2.2 Проработка лекционного материала.....	28
2.3 Самостоятельное изучение тем теоретической части курса.....	29
2.3.1 Тема: Выборочный метод в статистике.....	29
2.3.2 Тема: Представление статистических данных.....	30
2.3.3 Тема: Изучение периодической составляющей временного ряда.....	31
2.3.4 Тема: Факторный анализ.....	32
2.3.4 Тема: Инструментальные средства статистического анализа данных.....	33
2.4 Подготовка к контрольным работам.....	34
2.5 Подготовка к лабораторным работам.....	35
2.6 Подготовка к промежуточной аттестации.....	35
Список рекомендуемых источников.....	36
Приложение 1.....	37
Приложение 2.....	38
Приложение 3.....	40
Приложение 4.....	41
Приложение 5.....	43
Приложение 6.....	44
Приложение 7.....	45

ВВЕДЕНИЕ

Цель дисциплины «Прикладной статистический анализ» — формирование математической культуры магистрантов, знаний и навыков применения статистического методического инструментария и прикладных цифровых решений в диагностике, анализе социально-экономических проблем для принятия эффективных управленческих решений.

Задачи изучения дисциплины:

- сформировать у студентов знания, умения и навыки получения статистической информации;
- выработать умение критически оценивать информацию, осуществлять её верификацию и структуризацию, в том числе с использованием цифровых технологий;
- изучить методологию и методики исчисления важнейших статистических аналитических показателей социально-экономических процессов;
- развить способность правильно выбирать статистические методы обработки и анализа результатов исследования в соответствии с его целями, задачами, гипотезами и имеющимися данными для достижения профессиональных целей;
- выработать необходимые навыки использования методов и моделей прикладного статистического анализа и прикладных цифровых решений для достижения профессиональных целей.
- развить понимание принципов цифровых технологий, необходимые навыки работы с компьютером как средством управления информацией, решения конкретных задач, возникающих при принятии управленческих решений.

Прикладной статистический анализ разрабатывает и систематизирует понятия, приемы, математико-статистические методы и модели, предназначенные для обработки социально-экономической информации с целью её удобного представления и интерпретации, получения научных и практических выводов.

В данных Методических указаниях содержится:

- краткое изложение теоретического материала по теме, варианты заданий и порядок выполнения лабораторных работ. Лабораторные работы выполняются с использованием табличного процессора LibreOffice Calc (OpenOffice Calc, MS Excel). Форма контроля выполнения лабораторной работы: демонстрация преподавателю расчетов и результатов анализа, собеседование, ответы на вопросы, выполнение дополнительных заданий;
- рекомендации по организации самостоятельной работы. При самостоятельной работе и подготовке к лабораторным занятиям студенту следует повторить теоретический материал по конспекту лекций и источникам, приведенным в разделе «Список рекомендуемых источников», а также пользоваться информацией, представленной в статистических сборниках, в научной литературе и сети Интернет.

1 МЕТОДИЧЕСКИЕ УКАЗАНИЯ К ПРОВЕДЕНИЮ ЛАБОРАТОРНЫХ РАБОТ

1.1 Лабораторная работа «Сбор и представление статистических данных»

Цель работы

Получение практических навыков самостоятельного поиска официальной статистической информации, формирования массива данных социально-экономических показателей в на компьютере.

Форма проведения

Выполнение индивидуального задания.

Форма отчетности

Защита работы, устный опрос, демонстрация расчетов, выполнение дополнительных заданий.

Теоретические основы

Статистическая информация может быть первичной и вторичной. Первичная информация, как правило, получается в ходе собственного наблюдения или эксперимента. Источником вторичной информации, используемой при принятии управленческих решений, является Росстат. На занятии студент должен познакомиться на официальном портале Росстата с целями, задачами и направлениями деятельности организации, структурой портала, найти данные по заданному преподавателем кругу вопросов

Одним из источников официальной информации, отражающей явления и процессы, произошедшие в экономической и социальной жизни РФ, являются ежегодные сборники «Регионы России. Социально-экономические показатели», подготовленные Федеральной службой государственной статистики. В них публикуются статистические данные о социально-экономическом положении субъектов Российской Федерации, широко представлена информация о развитии экономики регионов, ее отраслей и секторов за определенный период времени.

При подготовке сборников использованы данные, получаемые органами государственной статистики от предприятий, организаций, населения в ходе проведения статистических наблюдений, переписей, выборочных обследований, данные министерств и ведомств Российской Федерации, а также информация, получаемая от организаций, которые проводят обследования, опросы по сбору сведений экономического и социального характера.

С 2014 года статистический сборник «Регионы России. Социально-экономические показатели» издается только в электронном виде и размещается на официальном Интернет-портале Росстата.

Варианты заданий

В соответствии с номером варианта выбрать исходные данные.

Вариант	Показатель
1.	Численность населения на 1 января 2022 г., тыс. человек
2.	Продукция сельского хозяйства - всего, млрд руб.
3.	Валовой региональный продукт в 2020г. ,млн руб.
4.	Среднемесячная номинальная начисленная заработная плата работников организаций, руб.
5.	Инвестиции в основной капитал, млрд руб.
6.	Среднедушевые денежные доходы (в месяц), руб
7.	Ввод в действие жилых домов, тыс. м ² общей площади жилых помещений
8.	Оборот розничной торговли, млрд руб

Порядок выполнения работы

Часть 1.

– Ознакомиться с деятельностью Федеральной службы государственной статистики, его структурой управления, организацией системы сбора информации в отраслевом и территориальном разрезе, перечнем публикуемых периодических журналов, сборников и документов, организацией Интернет-портала.

– Ознакомиться с Содержанием сборника «Регионы России. Социально-экономические показатели. 2022».

Часть 2.

– Из таблицы 1.1 ОСНОВНЫЕ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИЕ ПОКАЗАТЕЛИ в 2021 г сборника «Регионы России. Социально-экономические показатели. 2022» сформировать на компьютере в табличном процессоре рабочую таблицу из двух столбцов (субъект и показатель) и 14 строк (14 субъектов по собственному усмотрению), содержащую значение показателя (по вариантам).

- Провести сортировку субъектов по значению показателя
- Вычислить среднее значение показателя по совокупности
- Построить диаграммы распределения признака.
- Построить вариационный ряд с тремя равными интервалами.
 - а) вычислить центры интервала.
 - б) вычислить средние значения признака в интервале
 - в) построить диаграммы распределения признака.
- Сделать выводы о «сильных» и «слабых» регионах по значению исследуемого признака.

Контрольные вопросы и задания

- 1) Какая основная функция Росстата?
- 2) В чьем ведении в настоящее время находится Федеральная служба государственной статистики?
- 3) Когда и на основании чего Государственный комитет Российской Федерации по статистике преобразован в Федеральную службу государственной статистики?
- 4) Укажите полное название Томскстата и адрес Web-сайта
- 5) Найдите на сайте Росстата и укажите данные о размере среднемесячной номинальной начисленной заработной платы работников организаций по Российской Федерации за ноябрь 2022 года.
- 6) Сколько тематических разделов имеет сборник «Регионы России. Социально-экономические показатели. 2022».
- 7) Укажите название раздела и номер таблицы, содержащей информацию о численности государственных гражданских (муниципальных) служащих государственных органов и органов местного самоуправления

1.2 Лабораторная работа «Первичный анализ социально-экономической информации»

Цель работы

Закрепление теоретического материала и получение практических навыков исчисления обобщающих статистических показателей. Первичный анализ данных.

Форма проведения

Выполнение индивидуального задания.

Форма отчетности

Защита работы, устный опрос, демонстрация расчетов, выполнение дополнительных заданий.

Теоретические основы

Средняя величина — обобщающий показатель, характеризующий типический уровень признака в расчете на единицу совокупности в конкретных условиях места и времени.

Средняя величина обобщает количественную вариацию признака, то есть в средних величинах погашаются индивидуальные различия единиц совокупности, обусловленные случайными обстоятельствами. Средняя величина всегда именованная, имеет ту же размерность, что и признак у отдельных единиц совокупности.

Средняя величина должна вычисляться по однородной совокупности. Если совокупность неоднородна, то *общие средние*, рассчитанные по всей совокупности, должны подкрепляться *групповыми средними*, дающими характеристику размера явления, складывающуюся в конкретных условиях данной группы. Средние величины делятся на две большие категории:

степенные средние (к ним относятся средняя арифметическая, средняя гармоническая, средняя геометрическая, средняя квадратическая);

структурные средние (мода и медиана).

Средней арифметической величиной называется такое среднее значение признака, при вычислении которого общий объем признака в совокупности сохраняется неизменным. При ее вычислении общий объем признака мысленно распределяется поровну между всеми единицами совокупности.

Если данные представлены в виде списка, то среднее значение вычисляется по формуле простой средней

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n}$$

Если данные сгруппированы и представлены в виде дискретного или интервального вариационного ряда, то средняя величина должна рассчитываться по формуле взвешенной средней

$$\bar{X} = \frac{\sum x_i f_i}{\sum f_i}$$

где x_j — j -я градация признака или середина j -го интервала;

f_j — частота j -й градации или j -го интервала.

Структурные средние — *медиана* и *мода* — характеризуют величину значения признака, занимающего определенное значение в ранжированном вариационном ряду.

Мода (M_o) — наиболее часто встречающееся значение признака в совокупности. В дискретном ряду мода определяется без вычислений как значение признака с наибольшей частотой.

В интервальном ряду модальным является интервал с наибольшей частотой. Внутри него определяется *точечная мода*

$$M_o = x_{M_o} + l_{M_o} \cdot \frac{f_{M_o} - f_{M_o-1}}{(f_{M_o} - f_{M_o-1}) + (f_{M_o} - f_{M_o+1})},$$

где X_{M_o} — нижняя граница модального интервала;

f_{M_o} — частота в модальном интервале;

f_{M_o-1} — частота в предыдущем интервале;

f_{M_o+1} — частота в следующем за модальным интервале;

l_{M_o} — величина модального интервала.

Медиана (Me) — значение признака, которое делит *упорядоченную* последовательность его значений на две равные по численности части. В итоге у одной половины единиц

совокупности значение признака не превышает медианного уровня, а у другой — не меньше его.

Положение медианы определяется ее номером $N_{Me} = (n + 1) / 2$, где n — число единиц в совокупности.

В интервальном вариационном ряду для нахождения медианы применяется формула

$$Me = x_{Me} + l_{Me} \cdot \frac{\sum f_j - f'_{Me-1}}{f_{Me}}$$

где Me — медиана;

x_{Me} — нижняя граница интервала, в котором находится медиана;

l_{Me} — величина медианного интервала;

k — число интервалов;

f'_{Me-1} — накопленная частота в интервале, предшествующем медианному;

f_{Me} — частота в медианном интервале.

Аналогично медиане вычисляются значения признака, делящие совокупность на четыре равные части — *квартили* Q . Второй квартиль совпадает с медианой, а первый и третий рассчитываются по формулам:

$$Q_1 = X_{Q_1} + h_{Q_1} \cdot \frac{\sum f_j - f'_{Q_1-1}}{f_{Q_1}};$$

$$Q_3 = X_{Q_3} + h_{Q_3} \cdot \frac{3 \sum f_j - f'_{Q_3-1}}{4 f_{Q_3}}.$$

Значения признака, делящие ряд на пять равных частей называются квинтилями, на десять частей — децилями, на сто частей — перцентилями.

Для измерения вариации признака применяются различные абсолютные и относительные показатели.

Размах или *амплитуда вариации* — абсолютная разность между максимальным и минимальным значениями признака изучаемой совокупности

$$R = X_{\max} - X_{\min}.$$

Точнее характеризуют вариацию показатели, учитывающие колеблемость всех значений признака.

Среднее линейное отклонение $d = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{n} \quad i = 1, 2, \dots, n.$

Если данные сгруппированы, то используется формула

$$d_{\text{гр}} = \frac{\sum_{j=1}^k |x_j - \bar{x}| f_j}{\sum_{j=1}^k f_j} \quad j = 1, 2, \dots, k,$$

где k — число групп;

f_j — число единиц совокупности j -й группы.

Дисперсия вычисляется по формуле

$$D_x = \sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

Если данные сгруппированы, то дисперсия рассчитывается по формуле

$$D_{x_{\text{гр}}} = \sigma^2 = \frac{\sum_{j=1}^k (x_j - \bar{x})^2 f_j}{\sum f_j}.$$

Стандартное отклонение (среднее квадратическое отклонение) равно квадратному корню из дисперсии, т.е. $\sigma = \sqrt{D_x}$.

Для оценки интенсивности вариации и для сравнения ее в различных совокупностях и для разных признаков *необходимы относительные показатели вариации*. Они вычисляются как отношения абсолютных показателей силы вариации к средней арифметической величине признака:

- относительный размах вариации $\rho: \rho = R : \bar{x}$;
- относительное линейное отклонение: $m = d : \bar{x}$;
- коэффициент вариации $K_{\text{var}} = \sigma : \bar{x}$.

Варианты заданий

В соответствии с номером варианта выбрать исходные данные. Исходные данные, характеризующие размер заработной платы в тыс.руб., приведены в Приложении 1.

Порядок выполнения работы

- 1). На основе *исходных* данных определить:
 - а) среднее значение показателя, моду и медиану
 - б) размах вариации, среднее линейное отклонение, дисперсию, стандартное отклонение, коэффициент вариации
- 2). На основе исходных данных построить *дискретный* вариационный ряд и определить:
 - а) среднее значение показателя, моду и медиану
 - б) размах вариации, среднее линейное отклонение, дисперсию, стандартное отклонение, коэффициент вариации
 - в) первый и третий квартили
- 3). На основе исходных данных построить *интервальный* вариационный ряд с равными интервалами. Число интервалов задано в каждом варианте:

Вариант 1 Число групп $k=3$	Вариант 2 Число групп $k=4$	Вариант 3 Число групп $k=3$	Вариант 4 Число групп $k=4$	Вариант 5 Число групп $k=3$
Вариант 6 Число групп $k=4$	Вариант 7 Число групп $k=3$	Вариант 8 Число групп $k=4$	Вариант 9 Число групп $k=3$	Вариант 10 Число групп $k=4$

- 4) Определить в построенном интервальном вариационном ряду:
 - а) среднее значение показателя, моду и медиану
 - б) размах вариации, среднее линейное отклонение, дисперсию, стандартное отклонение, коэффициент вариации
 - в) первый и третий квартили
 - г) постройте график распределения посетителей по возрасту.
- 5). Дать содержательную интерпретацию полученным результатам.

Контрольные вопросы и задания

- 1) Дайте определение средней величины.
- 2) Как отражаются на графике среднее, мода, медиана?
- 3) По какой формуле вычисляется мода в интервальном вариационном ряду?
- 4) Как определяется мода в дискретном вариационном ряду?
- 5) Сравните значения средних величин в трех случаях.
- 6) Объясните порядок вычисления медианы в интервальном вариационном ряду.
- 7) Как определяется третий квартиль?

1.3 Лабораторная работа «Построение и исследование типологии социально-экономических объектов методом статистической группировки»

Цель работы

Закрепление теоретического материала и получение практических навыков исчисления статистических показателей. Построение и анализ типологии объектов методом типологической статистической группировки.

Форма проведения

Выполнение индивидуального задания.

Форма отчетности

Защита работы, устный опрос, демонстрация расчетов, выполнение дополнительных заданий.

Теоретические основы

Группировка – это распределение единиц по группам в соответствии со следующим принципом: различия между единицами, отнесенными к одной группе, должны быть меньше, чем между единицами, отнесенными к разным группам.

Группировка проводится с целью установления статистических связей и закономерностей, построения описания объекта, выявления структуры изучаемой совокупности.

Типологическая группировка служит для выделения социально-экономических типов. Последовательность ее построения следующая:

- 1) называются те типы явлений, которые могут быть выделены;
- 2) выбирается группировочный признак, формирующий описание типов;
- 3) устанавливаются границы интервалов группировочного признака;
- 4) группировка оформляется в таблицу, определяется численность каждой группы, рассчитываются сводные показатели по группам (групповые средние, показатели вариации).

Оценка качества группировки делается на основе вычисления коэффициента детерминации R^2 , характеризующего долю межгрупповой дисперсии в полной.

Коэффициент детерминации R^2 определяется на основе *правила сложения дисперсий*. Если совокупность разбита на группы, то полная (общая) дисперсия признака σ^2 может быть определена как сумма межгрупповой дисперсии $\sigma_{м.гр.}^2$ и средней из групповых дисперсий $\overline{\sigma_j^2}$:

$$\sigma^2 = \overline{\sigma_j^2} + \sigma_{м.гр.}^2.$$

Полная дисперсия признака вычисляется по обычной формуле:

$$\sigma^2 = \frac{\sum_{l=1}^n (x_l - \bar{x})^2}{n},$$

где n – число объектов в совокупности

x_l – значение признака у l -го объекта, $l=1, 2, \dots, n$

Обозначим $\overline{x_j}$ — среднее значение признака в группе j ;

f_j — число наблюдений в группе j .

Межгрупповая дисперсия вычисляется по формуле

$$\sigma_{м.гр.}^2 = \frac{\sum (\overline{x_j} - \bar{x})^2 f_j}{\sum f_j},$$

где \bar{x} — среднее значение признака в совокупности.

Коэффициент детерминации R^2 вычисляется по формуле

$$R^2 = \frac{\sigma_{м.сп.}^2}{\sigma^2},$$

Коэффициент детерминации изменяется от 0 до 1. Если значение R^2 близко к 1, то группировка построена «правильно».

Порядок выполнения работы и варианты заданий

В соответствии с номером варианта определить номер показателя субъектов федерации Сибирского федерального округа, представленные в таблице «Основные социально-экономические показатели в 2016 г.» Приложения 2

Вариант	Номер показателя	Вариант	Номер показателя
	1	6.	6
	2	7.	7
	3	8.	8
	4	9.	9
	5	10.	10

2) Сформировать рабочую таблицу, содержащую название региона и указанный в варианте показатели социально-экономического развития

Название региона	Показателя-
1.	
...	...
12.	

Провести сортировку регионов по значению **показателя**

Провести по всей совокупности *регионов* расчет

- среднего значения (функция СРЗНАЧ)
- дисперсии (функция ДИСПР)
- стандартного отклонения (функции СТАНДОТКЛОНП, КОРЕНЬ)
- коэффициента вариации

Провести группировку регионов по значению *показателя* выделив 3 группы: «Малые», «Средние», «Крупные». Границы группировочного показателя задать самостоятельно. Построить таблицу (табл.1) вида

Таблица 1. Статистические характеристики группировки

Группа	Интервалы признака– фактора x	Частота группы f_j	Показатели		
			Среднее	Дисперсия	Коэффициент вариации
Малые					
Средние					
Крупные					

Рассчитать для каждой группы определить и занести в табл.1:

- частоту группы,
- групповые средние значения показателя
- групповые дисперсии показателя,
- групповые коэффициенты вариации показателей;

Вычислить межгрупповую дисперсию.

Вычислить коэффициент детерминации и сделать вывод о качестве построенной группировки.

Контрольные вопросы и задания.

- 1) Сформулируйте основной принцип группировки.
- 2) Какой показатель оценивает однородность совокупности объектов?

- 3) Объясните выбор значений границ группировочного признака-фактора.
- 4) Запишите формулу вычисления межгрупповой дисперсии.
- 5) Можно ли признать построенную группировку «правильной»?

1.4 Лабораторная работа «Построение типологии социально-экономических объектов методами кластер-анализа»

Цель работы

Закрепление теоретического материала и получение практических навыков исчисления статистических показателей. Построение и анализ типологии объектов методами многомерной группировки.

Форма проведения

Выполнение индивидуального задания.

Форма отчетности

Защита работы, устный опрос, демонстрация расчетов, выполнение дополнительных заданий.

Теоретические основы

Методы многомерной классификации (группировки) позволяют проводить разбиение совокупности на основе множества признаков. В общей постановке задача классификации объектов заключается в том, чтобы некоторую совокупность n объектов, статистически представленную в виде матрицы X , разбить на сравнительно небольшое число k (заранее известно или нет) однородных в определенном смысле групп (классов, типов, кластеров, таксонов).

Одним из методов многомерной классификации является кластер-анализ (англ. The cluster – группа, пучок, куст, т.е. объединение каких-то однородных объектов, явлений).

Каждый объект является точкой в признаковом пространстве, которое представляет собой область варьирования всех признаков совокупности изучаемых явлений. Расстояния между точками определяет «схожесть» объектов: чем ближе точки, тем более похожи (однородны) объекты по своим характеристикам. В качестве расстояния будем рассматривать евклидово расстояние:

$$d_E(x_i, x_j) = \sqrt{\sum_{k=1}^p (x_{ki} - x_{kj})^2} ;$$

Принцип работы *иерархических агломеративных процедур* состоит в последовательном объединении групп элементов сначала самых близких, а затем все более отдаленных друг от друга.

В агломеративно – иерархических алгоритмах процесс объединения объектов в группы совершается последовательно за $n - 1$ шагов (если объединяются все n объектов).

На первом шаге в матрице расстояний (различий) D находится *минимальный элемент* d_{ij} и объекты i и j объединяются в один кластер $i + j$, состоящий из двух единиц – объектов. После этого матрица различий изменяется. Из нее выбрасываются две строки и два столбца, содержащие расстояния от i и j до остальных объектов, но добавляется одна строка и один столбец с расстоянием от кластера $i + j$ до остальных объектов. При пересчете матрицы расстояние $d_{i+j,k}$ между объединенным кластером ($i + j$) и любым из остальных кластеров k вычисляются по определенному правилу, которое определяет *алгоритм*. Наиболее часто применяются алгоритмы минимальной (одной) связи или «ближайшего соседа»: «дальнего соседа», среднего связывания, Варда, и другие иерархические алгоритмы, реализованные в статистических пакетах.

Далее, на каждом шаге процедура повторяется, т.е. находится минимальный элемент в матрице расстояний и соответствующие кластеры объединяются в один. Итогом работы алгоритма является *иерархическое дерево (дендрограмма)*, отражающая последовательность создания вариантов кластеризации на $n, (n-1), \dots, 2, 1$ групп.

Последовательные процедуры кластер-анализа рассмотрим на примере метода k -средних. В отличие от иерархических алгоритмов в последовательных процедурах на каждом шаге обрабатываются одно наблюдение.

Пусть наблюдения X_1, X_2, \dots, X_n надо разбить на k ($k \ll n$) классов, однородных в смысле некоторой метрики.

Смысл алгоритма состоит в последовательном уточнении эталонных точек – центров классов $E^{(v)} = \{e_1^{(v)}, e_2^{(v)}, \dots, e_k^{(v)}\}$, v - номер итерации. При этом эталонным точкам приписываются «веса» $\Omega^{(v)} = \{\omega_1^{(v)}, \omega_2^{(v)}, \dots, \omega_k^{(v)}\}$, которые пересчитываются на каждом шаге.

Реализация алгоритма происходит в два этапа. *На первом этапе* пересчитываются эталонные точки, *на втором этапе* производится разбиение объектов на k классов по числу эталонных точек.

Этап 1.

В качестве *нулевого* приближения примем первые k точек (объектов) исходной совокупности:

$$e_i^{(0)} = X_i, \quad \omega_i^{(0)} = 1, \quad i = \overline{1, k}.$$

На первом шаге «извлекается» точка X_{k+1} и выясняется, к какому из эталонов она ближе, то есть рассчитываются расстояние от точки X_{k+1} до каждого эталона. Этот ближайший эталон заменяется новым эталоном – центром тяжести старого и присоединенной точки – с увеличением веса, а остальные эталоны не изменяются. Затем появляется следующая точка, и опять выясняется, к какому из эталонов она ближе и т.д.

На v -ом шаге извлекается X_{k+v} и алгоритм пересчета эталонных точек следующий

$$e_i^{(v)} = \begin{cases} \frac{\omega_i^{(v-1)} e_i^{(v-1)} + X_{k+v}}{\omega^{(v-1)} + 1}, & \text{если } d(X_{k+v}, e_i^{(v-1)}) = \min_{1 \leq j \leq k} d(X_{k+v}, e_j^{(v-1)}), \\ e_i^{(v-1)}, & \text{иначе.} \end{cases}$$

$$\omega_i^{(v)} = \begin{cases} \omega_i^{(v-1)} + 1, & \text{если } d(X_{k+v}, e_i^{(v-1)}) = \min_{1 \leq j \leq k} d(X_{k+v}, e_j^{(v-1)}), \\ \omega_i^{(v-1)}, & \text{иначе.} \end{cases} \quad i = \overline{1, k}.$$

Максимальное число итераций - $n - k$. Пересчет эталонов заканчивается, если задано число итераций, либо когда эталоны перестают «колебаться», то есть $\max d(e_k^v, e_k^{v-1}) \leq \varepsilon$.

Этап 2.

Процесс разбиения исходной совокупности объектов на классы следующий:

извлекается точка X_i ($i = \overline{1, n}$) и вычисляются расстояния от нее до всех e_j ($j = \overline{1, k}$) эталонов. Если $d(X_i, e_s) = \min_{1 \leq j \leq k} d(X_i, e_j)$, то точка X_i включается в класс, образованный эталоном e_s .

В результате последовательного просмотра все точки все точки будут разбиты на заданное число классов. Результатом является алфавит классификации, то есть списки объектов, входящих в каждый кластер.

Порядок выполнения работы и варианты заданий

В соответствии с номером варианта определить номера таблицы исходных данных, номера двух показателей субъектов федерации СФО, представленные в таблицах Приложения 3, и исследуемый иерархический алгоритм.

Вариант	Номер таблицы в Приложении 3	х-признак- (номер показателя)	У-признак ВРП	Иерархический алгоритм
1	1	2	1	«ближайшего соседа»
2	2	2	1	«дальнего соседа»
3	1	3	1	«медианной связи»
4	2	3	1	«ближайшего соседа»
5	1	4	1	«дальнего соседа»
6	2	4	1	«медианной связи»
7	1	5	1	«ближайшего соседа»
8	2	5	1	«дальнего соседа»
9	1	6	1	«медианной связи»
10	2	6	1	«ближайшего соседа»

Сформировать рабочую таблицу, содержащую названия регионов и указанные в варианте показатели социально-экономического развития регионов СФО, следующего вида

Название региона	Показатель 1 <i>X</i>	Показатель 2 <i>У</i>
1.		
...
5.		

Построить точечную диаграмму в пространстве 2-х признаков.

Провести кластерный анализ, используя указанный в варианте *иерархический алгоритм*. Определить вариант разбиения на 3 класса и результаты разбиения записать в табл.2. Построить дендрограмму.

Провести кластерный анализ на 3 класса, используя метод К-средних. Эталонные точка задать самостоятельно. Сделать максимальное число итераций. Результаты разбиения записать в табл.2.

Таблица 2. Варианты разбиения регионов на классы

Метод	Названия регионов, входящих в классы		
	Класс № 1	Класс №2	Класс №3
Иерархический метод			
к-средних			

Сравнить результаты, полученные двумя алгоритмами кластер-анализа.

По результатам метода К-средних:

- внести в рабочую таблицу дополнительный столбец с номером кластера и провести сортировку по номеру кластера;
- построить точечную диаграмму в пространстве двух признаков с изображением принадлежности регионов кластерам (выделить разным цветом);

Дать содержательную интерпретацию результатов кластер-анализа (оценить: уровень развития регионов по классификационным признакам, наличие «естественного расслоения», различие средних значений, однородность групп, взаимосвязь признаков)

Контрольные вопросы и задания.

- 1) Сформулируйте основной принцип группировки.
- 2) Какой показатель оценивает однородность совокупности объектов?
- 3) Объясните выбор эталонных точек в алгоритме К-средних.
- 4) Охарактеризуйте группы, полученные в результате применения алгоритма К-средних.
- 5) Сравните результаты кластеризации иерархического метода и к-средних.

1.5 Лабораторная работа «Методы исследования взаимосвязей нечисловых данных»

Цель работы

Закрепление теоретического материала и получение практических навыков расчета и анализа показателей взаимосвязи нечисловых данных. Обработка данных социологического исследования.

Форма проведения

Выполнение индивидуального задания.

Форма отчетности

Защита работы, устный опрос, демонстрация расчетов, выполнение дополнительных заданий.

Теоретические основы

Анализ четырехклеточных таблиц.

Рассмотрим генеральную совокупность из n объектов (индивидов) в которой классификация произведена на основании наличия или отсутствия двух дихотомических (бинарных) признаков A и B . Тогда количества попаданий в 4 возможные подгруппы могут быть представлены таблицей 2×2 (четырёхклеточная) и ее записывают в виде:

Признак А	Признак В		Сумма
	есть В	нет В	
есть А	a	b	$a + b$
нет А	c	d	$c + d$
Сумма	$a + c$	$b + d$	n

В случае отсутствия связи между признаками частота в клетке a принимает значение $a = \frac{(a+b)(a+c)}{n}$ - теоретическая частота в предположении независимости.

Если фактическая частота в клетке a равна теоретической, то *связи между признаками нет*.

Если для какой-либо таблицы выполнено неравенство $a > \frac{(a+b)(a+c)}{n}$, означающее, что доля признака A среди B больше, чем среди не B , то признаки A и B называют *положительно связанными* или просто связанными.

Если имеем противоположное неравенство, то есть $a < \frac{(a+b)(a+c)}{n}$, то будем говорить, что A и B *отрицательно связаны*.

Меры связи в таблицах 2x2

1. Коэффициент ассоциации

$$Ka = \frac{ad - bc}{ad + bc}.$$

2. Коэффициент контингенции:

$$Kk = \frac{ad - bc}{\sqrt{(a+b)(a+c)(b+d)(c+d)}}.$$

Коэффициенты изменяются в диапазоне

$$-1 \leq Ka, Kk \leq +1$$

Если признаки независимы, то $Ka, Kk=0$;

$Ka, Kk = +1$ в случае полной положительной связи,
 $Ka, Kk = -1$, если полная отрицательная связь между признаками.

Анализ таблиц $r \times c$

Рассмотрим совокупность из n объектов, каждый из которых описан двумя признаками, измеренными в шкале наименований. Первый признак имеет r градаций (располагается по строкам таблицы сопряженности), второй признак имеет c градаций (располагается в столбцах таблицы сопряженности). Задача анализа взаимосвязи заключается в установлении при заданном значении *уровня значимости* α наличия или отсутствия статистической связи (проверка гипотезы H_0) и измерении силы связи.

Классическим тестом, используемым в тех случаях, когда данные расклассифицированы в таблице с двумя входами, является χ^2 - тест К.Пирсона:

$$\chi^2 = \sum \sum \frac{(n_{ij} - \frac{n_{i.} \cdot n_{.j}}{n})^2}{\frac{n_{i.} \cdot n_{.j}}{n}},$$

где n_{ij} - наблюдаемая частота (число объектов) в ячейке (i, j) таблицы;

$e_{ij} = \frac{n_{i.} \cdot n_{.j}}{n}$ - теоретически ожидаемая частота (по H_0 - предположение о статистической независимости рассматриваемых переменных) в этой ячейке;

$i = 1, 2, \dots, r$ - r - число строк;

$j = 1, 2, \dots, c$ - c - число столбцов;

$n_{i.} = \sum_{j=1}^c n_{ij}$ - сумма по строке;

$n_{.j} = \sum_{i=1}^r n_{ij}$ - сумма по столбцу;

$n = \sum_{i=1}^r n_{i.} = \sum_{j=1}^c n_{.j} = \sum_{i=1}^r \sum_{j=1}^c n_{ij}$ - общее число объектов или объем выборки.

Число степеней свободы для таблицы сопряженности $r \times c$ равно $df = (r-1)(c-1)$.

Заметим, что для таблицы 2×2 $df = 1$.

Установление статистической связи между признаками проводится по правилам проверки гипотез.

Шаг 1. Выдвигается гипотеза H_0 : статистическая связь между признаками отсутствует, то есть обладание первым признаком никак не связано с обладанием вторым признаком.

Шаг 2. По исходной таблице сопряженности вычисляем по формуле (1.2) фактическое значение χ^2 .

Шаг 3. Вычисляем число степеней свободы df . Задаем уровень значимости α . В таблице распределения статистики $\chi_{df, \alpha}^2$ отыскиваем значения этой величины при заданном уровне значимости α и при вычисленном числе степеней свободы df (таблица)

Таблица - Теоретические значения χ^2 при $\alpha = 0,05$

Число степеней свободы df	1	2	3	4	5	6	7	8	9	10
Значение χ^2	3,84	5,99	7,81	9,49	11,07	12,59	14,07	15,51	16,92	18,31

Шаг 4. Сравниваем фактическое и табличное значения χ^2 . Если $\chi^2 \geq \chi_{df,\alpha}^2$, то гипотеза H_0 на данном уровне значимости α может быть отвергнута. То есть можно утверждать, что с вероятностью $1-\alpha$ статистическая связь между признаками существует.

Вероятность того, что, отвергая H_0 , мы совершаем ошибку (ошибка первого рода) численно равна уровню значимости α , задаваемому при проверке гипотезы.

В социально-экономических исследованиях часто принимают уровень значимости $\alpha = 0,05$ ($\alpha = 5\%$).

Меры связи для таблиц $r \times c$

1. Средняя квадратичная сопряженность (коэффициент Фи)

$$\Phi = \sqrt{\frac{\chi^2}{n}}, \quad 0 \leq \Phi^2 \leq \min(r-1, c-1).$$

2. Коэффициент Крамера $C = \sqrt{C^2}$, где величина

$$C^2 = \frac{\chi^2}{n \min(r-1, c-1)} = \frac{\Phi^2}{\min(r-1, c-1)},$$

верхним пределом которой является 1, не зависимо от того, равны ли r и c .

3. Коэффициент Чупрова $T = \sqrt{T^2}$, где

$$T^2 = \frac{\chi^2}{n\sqrt{(r-1)(c-1)}} = \frac{\Phi^2}{\sqrt{(r-1)(c-1)}}$$

Верхним пределом изменения коэффициента Чупрова является 1, которая достигается при полной связи между переменными только в том случае, если $r = c$. Во всех других случаях, даже при полной связи $T^2 < 1$.

Взаимосвязь порядковых переменных

Под *ранговой корреляцией* понимается статистическая связь между порядковыми переменными. Эта связь анализируется на основе исходных статистических данных, представленных упорядочениями (ранжировками) n рассматриваемых объектов по разным свойствам.

Методы анализа ранговых корреляций часто используются в экспертных обследованиях для оценки согласованности мнений экспертов и построения интегральной (совокупной) оценки признака.

Пусть даны две ранжировки

$$x^{(1)} = (x_1^{(1)}, x_2^{(1)}, \dots, x_n^{(1)})$$

$$x^{(2)} = (x_1^{(2)}, x_2^{(2)}, \dots, x_n^{(2)}).$$

Ранговый коэффициент корреляции Спирмена вычисляется по формуле:

$$\rho = 1 - \frac{6}{n^3 - n} \sum_{i=1}^n (x_i^{(1)} - x_i^{(2)})^2,$$

Ранговый коэффициент корреляции Кендалла:

$$\tau = 1 - \frac{4\nu(x^{(1)}, x^{(2)})}{n(n-1)},$$

где $\nu(x^{(2)}, x^{(1)})$ - минимальное число обменов соседних элементов последовательности $x^{(2)}$, необходимое для приведения ее к упорядочению $x^{(1)}$.

Для совпадающих ранжировок $\rho, \tau = 1$, для противоположных $\rho, \tau = -1$. Таким образом, $-1 \leq \rho, \tau \leq 1$.

Если необходимо измерить связь между *несколькими* (двумя и более) переменными (экспертами), то вычисляется коэффициент конкордации (согласованности) по формуле

$$W(m) = \frac{12}{m^2(n^3 - n)} \sum_{i=1}^n \left(\sum_{j=1}^m x_i^{(j)} - \frac{m(n+1)}{2} \right)^2,$$

где m - число порядковых переменных (число ранжировок или число экспертов);

n - число объектов или длина ранжировки (объем выборки);

$j=1, 2, \dots, m$ - номера экспертов.

Коэффициент конкордации обладает свойствами:

а) $0 \leq W \leq 1$;

б) $W = 1$ при совпадении всех m ранжировок.

Варианты заданий

Исходные данные приведены в Приложении 4.

Задание 1. В табл.1 Приложения 4 приведены данные 50 респондентов о предпочитаемых напитках. При этом данные закодированы следующим образом: первая переменная ПОЛ (1- мужской, 2-женский), НАПИТОК (1-pepsi, 2-cola).

Порядок выполнения работы

1) В соответствии с вариантом отобрать из исходных данных Табл.1 Приложения 4 для анализа 30 респондентов (строк).

№ варианта	1	2	3	4	5
Строки	2-31	5-34	10-39	15-44	20-49
№ варианта	6	7	8	9	10
Строки	3-32	6-35	11-40	16-45	21-50

Построить четырехклеточную таблицу вида

ПОЛ	НАПИТОК	
	1-pepsi	2-cola
Мужской (1)	a	b
Женский (2)	c	d

Для удобства подсчета частот провести сортировку данных по первой переменной, а затем по второй.

Провести одномерный анализ данных: рассчитать

– частоты (число мужчин, число женщин, число любителей пепси, число любителей колы),

– относительные частоты по каждому признаку,

– построить диаграммы различных видов (гистограммы, секторные; линейные и др.

Провести анализ таблицы 2x2:

– установить наличие связи,

– рассчитать коэффициент ассоциации

– рассчитать коэффициент контингенции

Объяснить полученные результаты

Задание 2. В табл.2 Приложения 4 приведены данные социологического опроса студентов. Приведены ответы 30 респондентов на 11 вопросов.

Текст вопросов и Кодировка ответов:

1. С каким настроением вы смотрите в будущее (1-оптимистично, 2-спокойно 3-пессимистично)

2. Нужно ли молодым людям знать историю своих предков? (1-обязательно, 2-необязательно)

- 3 Как складываются взаимоотношения со сверстниками- юношами (1-легко, 2-спокойно 3-трудно)
- 4 Как складываются взаимоотношения со сверстниками- девушками (1-легко, 2-спокойно 3-трудно)
5. Хотели бы вы иметь личное огнестрельное оружие?(1-да, 0-нет)
- 6 Насколько вы согласны, что жизненный успех – это высокое служебное положение, карьера (1-нет, 2-скорее нет 3-скорее да 4 - да)
- 7 Насколько вы согласны, что жизненный успех – это богатство, финансовое благосостояние (1-нет, 2-скорее нет 3-скорее да 4 - да)
- 8 Насколько вы согласны, что жизненный успех – это удачный брак (1-скорее «нет», 2-скорее «да»)
9. Пол (1-м, 2-ж)
10. Возраст
11. ВУЗ

Порядок выполнения работы

1)В соответствии с вариантом отобрать для анализа 2 вопроса (столбца).

№ варианта	1	2	3	4	5	6	7	8	9	10
Вопрос разреза	9	10	11	9	10	11	9	10	11	9
Вопрос анализа	1	2	6	4	5	7	8	3	2	8

Провести сортировку данных по первой переменной (вопрос разреза), а затем по второй (вопрос анализа).

Построить таблицу сопряженности $r \times c$.

Провести одномерный анализ данных: рассчитать

- частоты по вопросу разреза;
- частоты по вопросу анализа;
- относительные частоты по каждому признаку;
- построить диаграммы распределения респондентов по значениям признаков (гистограммы, секторные; линейные и др.)

Провести анализ таблицы $r \times c$:

- рассчитать значение Хи-квадрат;
- установить наличие связи,
- рассчитать коэффициент Фи;
- рассчитать коэффициент Крамера;
- рассчитать коэффициент Чупрова.

Дать содержательную интерпретацию полученным результатам.

Задание 3

На основе таблицы сопряженности, полученной в п.3 Задания 2, сформировать ранжировки, принимая в качестве экспертов значения (градации) признака-разреза. Перейти в таблице сопряженностей к рангам, приписывая ранг 1 максимальному значению в строке.

Провести расчет возможных ранговых коэффициентов корреляции.

Сделать выводы относительно статистической связи и согласованности мнений экспертов.

Контрольные вопросы и задания

- 1) Что такое «дихотомические» признаки? Приведите пример.
- 2) Какие показатели характеризуют силу связи в четырехклеточных таблицах?

- 3) Если связь между признаками отсутствует, какое значение принимает коэффициент ассоциации?
- 4) Как определить теоретическое значение Хи-квадрат критерия?
- 5) Объясните понятие «уровень значимости».
- 6) В каком случае значения коэффициентов Крамера и Чупрова совпадают?
- 7) Как определить значение числа степеней свободы таблицы сопряженности?
- 8) Есть ли согласованность между построенными ранжировками
- 9) Поясните различие понятий «статистическая связь» и «согласованность».

1.6 Лабораторная работа «Исследования взаимосвязей числовых данных методами аналитической группировки и корреляционно-регрессионного анализа»

Цель работы

Закрепление теоретического материала и получение практических навыков исследования зависимости числовых признаков на компьютере.

Форма проведения

Выполнение индивидуального задания.

Форма отчетности

Защита работы, устный опрос, демонстрация расчетов, выполнение дополнительных заданий.

Теоретические основы

Группировка – это распределение единиц по группам в соответствии со следующим принципом: различия между единицами, отнесенными к одной группе, должны быть меньше, чем между единицами, отнесенными к разным группам.

Оценка качества группировки делается на основе вычисления коэффициента детерминации R^2 , характеризующего долю межгрупповой дисперсии в полной:

$$R^2 = \frac{\delta_{м.гр.}^2}{\sigma^2}$$

Полная дисперсия признака x вычисляется по обычной формуле:

$$\sigma^2 = \frac{\sum_{l=1}^n (x_l - \bar{x})^2}{n}$$

где n – число объектов в совокупности; x_l – значение признака у l -го объекта, $l=1, 2, \dots, n$

Межгрупповая дисперсия вычисляется по формуле

$$\delta_{м.гр.}^2 = \frac{\sum (\bar{x}_i - \bar{x})^2 f_i}{\sum f_i},$$

где \bar{x} — среднее значение признака в совокупности.

\bar{x}_i — среднее значение признака в группе i ;

f_i — число наблюдений в группе i .

Типологическая группировка служит для выделения социально-экономических типов

Часть 1. Аналитическая группировка характеризует взаимосвязь между двумя признаками, один из которых рассматривается как результат (ЗАВИСИМАЯ ПЕРЕМЕННАЯ), другой – как фактор (НЕЗАВИСИМАЯ ПЕРЕМЕННАЯ).

Группировка строится по признаку - фактору, например, как интервальный вариационный ряд или как типологическая группировка. Затем для каждой группы вычисляются соответствующие статистические показатели (средние, дисперсии, коэффициенты вариации)

как для факторного, так и результирующего признака. Если «качественная» группировка по фактору приводит к «качественной» группировке по результирующему признаку, то можно предположить наличие связи между признаками.

Задача аналитической группировки состоит в том, чтобы увидеть, есть ли связь между признаками или нет; прямая связь или обратная; линейная или нелинейная, сильная или слабая.

Для ответа на вопрос о виде связи вычисляются коэффициенты

$$b_{xy} = \frac{\bar{y}_2 - \bar{y}_1}{\bar{x}_2 - \bar{x}_1}; \quad b_{xy} = \frac{\bar{y}_3 - \bar{y}_2}{\bar{x}_3 - \bar{x}_2} \text{ и т.д.}$$

Полученные значения показывают величину изменения результата на единицу изменения фактора. Величина b_{xy} равна тангенсу угла наклона отрезка прямой к оси x . Если $b_{xy} = const$, то y линейно зависит от x , если $b_{xy} \neq const$, то связь нелинейная. Величины b_{xy} - показатели силы связи, характеризуют прирост результата на единицу изменения фактора.

Для оценки силы связи проводится расчет коэффициента детерминации по признаку-результату R^2 и эмпирического корреляционного отношения r .

$$R_y^2 = \frac{D_{м.р.-y}}{D_y},$$

где R_y^2 — коэффициент детерминации, вычисленный по признаку-результату y .

Коэффициент детерминации изменяется от 0 до 1. Если значение R^2 близко к 1, то связь между результативным и факторным признаком существенна.

Для характеристики тесноты связи берется также показатель r — эмпирическое корреляционное отношение, рассчитываемое как

$$r = \sqrt{R^2} = \sqrt{\frac{D_{м.р.-y}}{D_y}}.$$

Эмпирическое корреляционное соотношение варьирует от -1 до 1.

При $r = 0$ связи нет, при $r = 1$ — связь прямая полная, $r = -1$ — связь обратная полная

Для характеристики силы связи используют шкалу Чеддока

Показатели тесноты связи	0,1-0,3	0,3-0,5	0,5-0,7	0,7-0,9	0,9-0,99
Характеристика силы связи	слабая	умеренная	заметная	высокая	весьма высокая

Часть 2 Корреляционный анализ

Для исследования взаимосвязи двух количественных переменных можно применять коэффициент корреляции Пирсона (r -Пирсона).

Коэффициент корреляции r -Пирсона характеризует существование **линейной связи** между двумя величинами. Если связь криволинейная, то он не будет работать. Формула расчета коэффициента корреляции Пирсона следующая:

$$r_{xy} = \frac{\sum (x_i - \bar{x}) \times (y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \times \sum (y_i - \bar{y})^2}}$$

Коэффициент корреляции изменяется в пределах от -1 до +1, если нет связи то $r_{xy}=0$

Если $r_{xy}=0$, то связь отсутствует; если $r_{xy} = 1$, то связь — функциональная;

если $r_{xy} < 0$, то связь — обратная; $r_{xy} > 0$, то связь — прямая.

Как и в случае корреляционного отношения для интерпретации показателей тесноты связи применяется шкала Чеддока.

Часть 3 Регрессионный анализ.

Уравнение простой (парной) регрессии является статистической моделью связи двух количественных переменных и определяет функциональную зависимость результативного признака от факторного:

Рассмотрим случай **парной линейной** регрессии. Если результативный признак с увеличением факторного признака *равномерно* возрастает или убывает, то такая зависимость является линейной и выражается уравнением прямой:

$$\hat{y} = a_0 + a_1x$$

где a_0 и a_1 — параметры уравнения;

На практике построение линейной регрессии сводится к оценке параметров уравнения, то есть по исходным данным необходимо определить численные значения a_0 и a_1

В уравнении прямой параметр a_0 экономического смысла не имеет, параметр a_1 является *коэффициентом регрессии* и показывает изменение результативного признака при изменении факторного на единицу.

В лабораторной работе для построения уравнения регрессии (трендовой модели) будем использовать специальный аппарат табличного процессора для аппроксимации экспериментальных данных с последующим подбором аппроксимирующей функции (линии тренда) $y=f(t)$ <Добавить линию тренда> .

Варианты заданий

Исходные данные представлены в таблице Приложения 1. В соответствии с номером варианта скопировать в Excel **3 столбца**: названия региона, столбец с номером признака-фактора, столбец с номером 1-результат.

Задача состоит в исследовании зависимости валового внутреннего продукта от заданного в варианте фактора.

Определить признак-фактор и признак-результат.

Вариант	x-признак-фактор (номер показателя)	y-признак-результат-ВРП
1.	2	1
2.	3	1
3.	4	1
4.	5	1
5.	6	1
6.	7	1
7.	4	1
8.	5	1

Порядок выполнения работы

Сформировать исходную таблицу, содержащую названия регионов и указанные в варианте показатели социально-экономического развития регионов СФО (Приложение 5).

Название региона	Показатель (фактор)	ВРП (результат)
	x	y
...

Провести *сортировку регионов* по значению **фактора** – **x** (с расширением диапазона)

Провести по всей совокупности для *каждого признака* расчет

- среднего значения (функция СРЗНАЧ)
- дисперсии (функция ДИСПР)

- стандартного отклонения (функции СТАНДОТКЛОНП или КОРЕНЬ из дисперсии)
- коэффициента вариации (равен отношению стандартного отклонения к среднему значению)

Построить Столбиковую диаграмму распределения регионов по признаку

Аналитическая группировка

- 1) Используя столбиковую диаграмму, провести типологическую группировку регионов по значению **фактора**, выделив 3 группы: «Малые», «Средние», «Крупные».
- 2) Для каждой группы определить и занести в табл. 1:
 - частоту группы,
 - групповые средние значения показателей x и y
 - групповые дисперсии показателей x и y
 - групповые коэффициенты вариации показателей x и y .

Таблица 1. Статистические характеристики группировки

Группа	Интервалы признака–фактора	Частота группы f_j	Признак – фактор x			Признак – результат y		
			Среднее	Дисперсия	Коэффициент вариации	Среднее	Дисперсия	Коэффициент вариации
Малые								
Средние								
Крупные								

- 1) Вычислить по факторному признаку межгрупповую дисперсию
- 2) Вычислить по факторному признаку коэффициент детерминации и сделать вывод о качестве построенной по x группировки
- 3) Рассчитать межгрупповую дисперсию по признаку-**результату**
- 4) Рассчитать коэффициент детерминации R^2 по признаку-**результату** и сделать вывод о качестве группировки по y и силе связи между признаками
- 5) Вычислить значение эмпирического корреляционного отношения по признаку – результату и сделать выводы
- 6) Построить точечную диаграмму зависимости результата от фактора.

Корреляционный анализ

- 1) Определить коэффициент корреляции r_{xy} , (функции **КОРРЕЛ** или **ПИРСОН**)
- 2) Сделать выводы по результатам корреляционного анализа (сила связи по Чеддоку)

Регрессионный анализ

- 1) Построить точечную диаграмму зависимости результата от фактора.
- 2) На точечной диаграмме зависимости $y=f(x)$ добавить линию *линейного* тренда $\hat{y} = f(x)$, поместить на графике уравнение и коэффициент R^2
- 3) Сделать общие выводы по результатам исследования зависимости ВРП от фактора.

Контрольные вопросы и задания.

- 1) Дайте определения статистической группировки.
- 2) Каковы задачи аналитической группировки?
- 3) Как вычисляется коэффициент детерминации?.
- 4) Какие значения может принимать коэффициент корреляции?
- 5) Каким показателем оценивается качество группировки?
- 6) Сформулируйте задачи корреляционного и регрессионного анализов.
- 7) Объясните результаты построения регрессионной модели.
- 8) Существует ли зависимость ВРП от исследуемого фактора??

1.7 Лабораторная работа «Анализ и прогнозирование динамики развития социально-экономических процессов»

Цель работы

Закрепление теоретического материала и получение практических навыков вычисления показателей динамики. Построение моделей временного ряда.

Форма проведения

Выполнение индивидуального задания.

Форма отчетности

Устный опрос, демонстрация расчетов, выполнение дополнительных заданий.

Теоретические основы

Временной (динамический) ряд - ряд расположенных в хронологической последовательности значений статистических показателей. Каждый временной ряд включает два элемента: момент или период времени и конкретное значение показателя (уровень ряда). Уровни ряда обычно обозначают латинской буквой y , а моменты или периоды времени, к которым они относятся, - буквой t .

Пусть n — число уровней ряда и нумерация ряда начинается с 1 (единицы).

Базисный абсолютный прирост $\Delta y_{\text{б}}$ исчисляется как разность между сравниваемым уровнем y_i и уровнем, принятым за постоянную базу сравнения y_1 :

$$\Delta y_{\text{б}} = y_i - y_1.$$

Цепной абсолютный прирост $\Delta y_{\text{ц}}$ — разность между сравниваемым уровнем y_i и уровнем, который ему предшествует y_{i-1} :

$$\Delta y_{\text{ц}} = y_i - y_{i-1}$$

Темп роста базисный (в процентах) определяется по формуле

$$\text{Тр}_{\text{б}} = (y_i : y_1) \cdot 100$$

Темп роста цепной (в процентах) определяется по формуле

$$\text{Тр}_{\text{ц}} = (y_i : y_{i-1}) \cdot 100.$$

Темп прироста базисный (в процентах) определяется по формуле

$$\text{ТПр}_{\text{б}} = \text{Тр}_{\text{б}} - 100$$

Средний уровень ряда (\bar{y}) динамики характеризует типическую величину абсолютных уровней. Метод расчета среднего уровня ряда динамики зависит от вида временного ряда.

Для *интервального* временного ряда абсолютных показателей с равными периодами времени средний уровень ряда \bar{y} рассчитывается по формуле простой арифметической:

$$\bar{y} = \frac{\sum y_i}{n} = \frac{y_1 + y_2 + \dots + y_n}{n}.$$

где n — число уровней ряда.

В *моментном* ряду динамики с равностоящими датами времени средний уровень определяется по формуле средней хронологической

$$\bar{y} = \frac{\frac{1}{2} y_1 + y_2 + \dots + \frac{1}{2} y_n}{n - 1}.$$

Показатель *среднего абсолютного прироста* можно определить по формуле

$$\Delta \bar{y} = \frac{y_n - y_1}{n - 1}.$$

Средний темп роста можно определить по абсолютным уровням ряда динамики по формуле

$$\bar{T}_p = \sqrt[n]{y_n : y_1} \times 100\%$$

Для получения *средних темпов прироста* $\bar{T}_п$ в процентах используется зависимость:

$$\bar{T}_п = \bar{T}_p - 100.$$

Прогнозирование на следующий $(n+1)$ период времени осуществляется по формуле:

$$\hat{y}_{n+1} = y_n + \bar{\Delta}y,$$

При этом предполагается, что период упреждения составляет 1-2 временных интервала, в течение которых тенденция развития сохраняется.

Динамический ряд теоретически может быть представлен в виде совокупности *трех составляющих*:

- 1) *тренд* — основная тенденция развития динамического ряда (тенденция к росту или к снижению);
- 2) циклические (периодические) колебания, в том числе сезонные;
- 3) случайные колебания.

На практике для непосредственного выявления и изучения тренда в рядах динамики используются три основных метода:

- метод укрупнения интервалов;
- метод скользящей средней;
- метод аналитического выравнивания.

Метод укрупнения интервалов заключается в том, что исходные уровни ряда заменяются средними значениями, вычисленными на более длинных временных интервалах. Например, переходим от помесечных данных к поквартальным или от годовых данных к пятилетним и т.д.

В методе *трехзвенной скользящей средней* сглаженные уровни ряда вычисляются последовательно по формуле

$$\bar{y}_i = \frac{y_{i-1} + y_i + y_{i+1}}{3};$$

При *аналитическом выравнивании* ряда динамики фактический уровень изучаемого показателя оценивается как функция времени (трендовая модель, уравнение регрессии)

$$y = f(t) + \varepsilon,$$

где $f(t) = \hat{y}$ — уровень, определяемый тенденцией развития;

ε — случайное или циклическое отклонение от тенденции.

Подбор адекватной функции осуществляется методом наименьших квадратов — минимальностью отклонений суммы квадратов между теоретическими \hat{y}_i и эмпирическими y_i уровнями:

$$\sum (\hat{y}_i - y_i)^2 = \min.$$

В простейшем случае динамический ряд характеризуется *равномерным развитием*. Для этого типа динамики характерны постоянные цепные абсолютные приросты:

$$\Delta y_i = \text{const}.$$

Основная тенденция развития в рядах динамики со стабильными приростами отображается уравнением линейной функции:

$$\hat{y}_i = a_0 + a_1 t,$$

где a_0 и a_1 — параметры уравнения;

t — обозначение времени

Для вычисления параметров функции на основе требований метода наименьших квадратов получают следующие формулы расчета коэффициентов

$$a_0 = \frac{\sum y \sum t^2 - \sum ty \sum t}{n \sum t^2 - \sum t \sum t};$$

$$a_1 = \frac{n \sum ty - \sum t \sum y}{n \sum t^2 - \sum t \sum t}.$$

Построив уравнение регрессии, проводят оценку его надежности. Это делается посредством критерия Фишера (F). Фактический уровень ($F_{\text{факт}}$) сравнивается с теоретическим (табличным) значением:

$$F_{\text{факт}} = \frac{\sigma_f^2 (n - k)}{\sigma_{\text{ост}}^2 (k - 1)},$$

где k — число параметров функции, описывающей тенденцию;
 n — число уровней ряда.

σ_f^2 - факторная дисперсия, которая вычисляется по формуле

$$\sigma_f^2 = \frac{\sum (\hat{y}_i - \bar{y})^2}{n}$$

Остаточная дисперсия $\sigma_{\text{ост}}^2$ определяется по формуле

$$\sigma_{\text{ост}}^2 = \frac{\sum (y_i - \hat{y}_i)^2}{n}.$$

По правилу сложения дисперсий общая дисперсия

$$\sigma_y^2 = \frac{\sum (y_i - \bar{y})^2}{n} = \sigma_f^2 + \sigma_{\text{ост}}^2.$$

$F_{\text{факт}}$ сравнивается с $F_{\text{теор}}$ при $\nu_1 = (k - 1)$, $\nu_2 = (n - k)$ степенях свободы и уровне значимости α (обычно $\alpha = 0,05$). Таблица теоретических значений приведена в Приложении 6. Если $F_{\text{факт}} > F_{\text{теор}}$, то уравнение регрессии значимо, т.е. основная модель адекватна фактической временной тенденции.

Для оценки точности модели вычисляют коэффициент детерминации:

$$R^2 = \frac{\sigma_f^2}{\sigma_y^2}, \quad 0 \leq R^2 \leq 1.$$

Если значение коэффициента детерминации R^2 близко к 1, то модель близка к реальному процессу.

Подставляя в полученное уравнение модели значение времени t_k ($k \notin i = 1, 2, \dots, n$), получаем точечный прогноз (оценку прогнозного) значения

$$\hat{y}_k = a_0 + a_1 \cdot t_k.$$

Порядок выполнения работы и варианты заданий

Исходные данные о прибывших в Россию из стран дальнего зарубежья представлены в таблице Приложения 7.

В соответствии с номером варианта определить из таблицы исходных данных страну

№ варианта	Страна	№ варианта	Страна
Вариант 1	Австралия	Вариант 6	США
Вариант 2	Австрия	Вариант 7	Турция
Вариант 3	Вьетнам	Вариант 8	Финляндия
Вариант 4	Германия	Вариант 9	Франция
Вариант 5	Греция	Вариант 10	Израиль

Порядок выполнения работы

Задание 1.

Вычислить

- базисные и цепные абсолютные приросты,
- базисные и цепные темпы роста,
- базисные и цепные темпы прироста показателя

Вычислить средние показатели

- среднегодовую численность мигрантов
- среднегодовой абсолютный прирост,
- среднегодовой темп роста,
- среднегодовой темп прироста

Задание 2

Вычислить прогнозное значение на следующий 2017 год, используя средний абсолютный прирост.

Задание 3

Провести выравнивание временного ряда методами

- укрупнения интервалов (перейти к трехгодичным периодам);
- трехзвенной скользящей средней.

Построить графики исходных и выровненных значений

Задание 4

Построить линейную модель исходных данных, вычисляя коэффициенты a_0 и a_1 .

Провести оценку модели по критерию Фишера,

Рассчитать коэффициент детерминации.

Построить график динамики исходных и выровненных значений.

Вычислить прогнозное значение на следующий год.

Контрольные вопросы и задания.

- 1) Объясните выбор формул для расчета показателей динамики.
- 2) Укажите на графике абсолютные цепные и базисные приросты.
- 3) Запишите формулу пятизвенной скользящей средней.
- 4) Каким показателем оценивается точность модели?
- 5) Сколько параметров имеет линейная модель временного ряда?
- 6) Объясните результаты построения линейных моделей.
- 7) Какая из полученных моделей точнее?
- 8) Совпадают ли прогнозные значения показателя, рассчитанные разными методами?

2 МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОРГАНИЗАЦИИ САМОСТОЯТЕЛЬНОЙ РАБОТЫ

2.1 Общие положения

Цель самостоятельной работы по дисциплине – закрепление и углубление теоретических знаний; формирование умения работать с научной и технической литературой и осуществлять самостоятельный поиск информации; развитие научно-исследовательских и творческих способностей; приобретение навыков расчётно-аналитической работы.

Самостоятельная работа студента по дисциплине «Прикладной статистический анализ» включает следующие виды его активности:

- проработка лекционного материала;
- изучение тем теоретической части дисциплины, вынесенных для самостоятельной проработки;
- подготовка к контрольным работам;
- подготовка к лабораторным работам;
- выполнение индивидуального задания;
- подготовка к экзамену.

2.2 Проработка лекционного материала

При проработке лекционного материала по каждой теме студент должен внимательно ознакомиться с конспектом лекций, а затем для углубленного изучения материала следует обратиться к литературным источникам (учебникам, учебным пособиям, монографиям, статьям, статистическим сборникам), а также материалам, размещенным в сети Интернет. Для закрепления материала темы необходимо ответить на предлагаемые в пособиях вопросы и прорешать задачи по теме.

При изучении каждой темы целесообразно:

- 1) ознакомиться с методическим обеспечением изучаемой дисциплины, включающей тематический план и программу курса;
- 2) руководствоваться рекомендованной нормативной базой и учебной литературой, которая имеется в фондах библиотеки;
- 3) использовать возможности сайта библиотеки университета и другие информационные ресурсы Интернета;
- 4) прочитать соответствующую теме главу учебника;
- 5) доработать конспект лекции.

При изучении учебного материала темы студенту необходимо, прежде всего, разобраться в основанных понятиях и терминах данной темы. Для этого рекомендуется использовать различные источники информации, в том числе учебные пособия, монографии, периодические издания, законодательные и нормативные документы, статистические материалы, информацию государственных органов власти и управления, органов местного самоуправления, переводные издания, а также труды зарубежных авторов.

Изучение рекомендованных источников следует начинать с основных рекомендованных преподавателем глав и разделов учебников и учебных пособий, а затем переходить к научным монографиям и материалам периодических изданий. При этом полезно делать выписки и конспекты наиболее интересных материалов, что способствует более глубокому осмыслению материала и лучшему его запоминанию. Такая практика учит отделять в тексте главное от второстепенного, а также позволяет проводить систематизацию и сравнительный анализ изучаемой информации, что важно в условиях большого количества разнообразных по качеству и содержанию сведений.

Проработка пройденного лекционного материала является наиболее важным видом самостоятельной работы. Чем глубже и полнее проработан материал, тем легче при выполнении других видов самостоятельной работы. Систематическая, регулярная работа над пройденным лекционным материалом, начиная с первого занятия, является необходимым условием для понимания материалов последующих лекций и усвоения материалов практических и лабораторных занятий.

2.3 Самостоятельное изучение тем теоретической части курса

2.3.1 Тема: Выборочный метод в статистике

Перечень вопросов, подлежащих изучению

- 1) Сущность выборочного метода.
- 2) Генеральная и выборочная совокупности.
- 3) Ошибки выборочного наблюдения.
- 4) Виды отбора единиц в выборочную совокупность.
- 5) Определение необходимой численности выборки.

Методические рекомендации по изучению

Тема «Основы выборочного метода» особенно сильно взаимосвязана такими темами дисциплины как статистическое наблюдение, статистические величины и показатели вариации. Она также связана с курсами математики и теории вероятностей (закон больших чисел, теоремы Чебышева, Бернулли и др.). Важными практическими вопросами являются определение способа отбора единиц совокупности, вычисление ошибок выборки и построение доверительных интервалов выборочных характеристик, расчет необходимого объема выборки.

Студенту следует уяснить понятия генеральной и выборочной совокупностей, репрезентативности. Совокупность единиц, из которых производится отбор, принято называть генеральной совокупностью. Совокупность отобранных единиц из генеральной совокупности называется выборочной совокупностью.

При любом способе отбора должен соблюдаться принцип: каждой единице генеральной совокупности обеспечивается одинаковая вероятность (возможность) быть выбранной. Единица отбора совпадает с единицей наблюдения. Этим обеспечивается возможность распространения выводов, сделанных на основе выборочного наблюдения, на все генеральную совокупность.

Различают среднюю и предельную ошибки выборки. Расчет ошибок позволяет решить одну из главных проблем организации выборочного наблюдения – оценить представительность (репрезентативность) выборочной совокупности. Величина средней ошибки выборки рассчитывается дифференцировано в зависимости от способа отбора (повторный или бесповторный) и процедуры (вида) выборки по соответствующим формулам. Правильный выбор метода отбора единиц в выборочную совокупность (простая случайная выборка, механическая выборка, стратифицированная выборка, гнездовая выборка, комбинированная выборка, многоступенчатая выборка, квотная и др.) имеет большое значение при практическом применении выборочного метода.

При изучении темы следует большое внимание уделить решению типовых задач, представленных в рекомендованной литературе.

Рекомендуемые источники

Дудин, М. Н. Статистика [Электронный ресурс]: учебник и практикум для вузов / М. Н. Дудин, Н. В. Лясников, М. Л. Лезина. — Москва : Издательство Юрайт, 2021. — 374 с. — (Высшее образование). — ISBN 978-5-9916-8908-3. — Текст : электронный // ЭБС Юрайт [сайт]. [Электронный ресурс]: — Режим доступа: <https://urait.ru/bcode/470169>.

Лепихина З.П. Статистика: Учебное пособие/ З. П. Лепихина; Федеральное агентство по образованию, Томский государственный университет систем управления и радиоэлектроники. - Томск: ТУСУР, 2005. – с.87-108.

Миркин, Б. Г. Введение в анализ данных : учебник и практикум / Б. Г. Миркин. — Москва: Издательство Юрайт, 2022. — 174 с. — (Высшее образование). — ISBN 978-5-9916-5009-0. — Текст : электронный // Образовательная платформа Юрайт [сайт] [Электронный ресурс]: — Режим доступа: <https://urait.ru/bcode/469306>.

При необходимости рекомендуется ознакомиться с другими источниками, приведенными в разделе «Список рекомендуемых источников».

2.3.2 Тема: Представление статистических данных

Перечень вопросов, подлежащих изучению

- 1) Вариационные ряды.
- 2) Статистические таблицы.
- 3) Статистические графики.

Методические рекомендации по изучению

Изучение темы следует начать с повторения понятий статистического признака, показателя, вариации. Так как статистика изучает массовые стохастические явления, то вопрос правильного представления результатов наблюдения и результатов обработки данных имеет большое значение.

Вариационный ряд – совокупность значений варьирующего признака и соответствующих им численностей единиц совокупности. Следует различать дискретные и интервальные вариационные ряды. Необходимо обратить внимание на способы определения числа интервалов, способы «закрытия» открытых интервалов. При изучении вариационных рядов надо обратить внимание на понятия «интервал» и «группа»: построение интервального вариационного ряда – это группировка единиц совокупности.

Результаты статистического исследования представляются в виде статистических таблиц и графиков.

Статистическая таблица — система строк и столбцов, в которых в определенной последовательности и связи излагается статистическая информация о социально-экономических явлениях.

Следует различать подлежащее и сказуемое статистической таблицы.

В подлежащем указывается характеризуемый объект — либо единица совокупности, либо группы единиц, либо совокупность в целом. В сказуемом дается характеристика подлежащего, обычно в количественной форме в виде системы показателей.

По характеру подлежащего статистические таблицы подразделяются на простые, групповые и комбинационные.

По характеру сказуемого статистические таблицы делятся на таблицы с простой разработкой сказуемого и таблицы со сложной разработкой сказуемого.

Студент должен при оформлении таблиц соблюдать следующие *общие* правила.

Обязателен заголовок таблицы, в котором указывается, к какой категории и к какому времени относится таблица. В таблице не должно быть лишних линий. Может быть горизонтальная черта, отделяющая итоговую строку. Вертикальные линии могут быть, а могут отсутствовать. Заголовки граф содержат названия показателей без сокращения слов и единиц измерения. Общие единицы измерения могут быть вынесены в заголовок таблицы. Итоговая строка завершает таблицу и располагается внизу таблицы. Иногда итоговая строка бывает первой, в этом случае второй строкой идет строка «в том числе» или «из них». Цифровые сведения записываются в пределах каждой графы с одной и той же степенью точности.

Статистические графики представляют собой условные изображения числовых величин и их соотношений посредством линий, геометрических фигур, рисунков или географических карт-схем.

Графики обязательно сопровождаются заголовками, в которых указывается, какой показатель изображен, в каких единицах измерения, по какой территории и за какое время он

определен. На графике должен быть указан масштаб — мера перевода числовой величины в графическую.

По способу построения статистические графики делятся на диаграммы (линейные, объемные, плоскостные, радиальные, точечные, фигурные), картограммы и картодиаграммы.

Среди плоскостных диаграмм часто используются столбиковые диаграммы, на которых величина столбика соответствует значению показателя. Линейные графики обычно используются для представления динамики показателя.

Для иллюстрации структуры совокупности используется секторная диаграмма. Вся совокупность принимается за 100 процентов, ей соответствует вся площадь круга, а площади секторов соответствуют частям совокупности.

Студент должен иметь в виду, что конкретная организация (например, ТУСУР) может разработать внутренние стандарты оформления документации, но общие принципы построения таблиц и графиков сохраняются.

Рекомендуемые источники

Дудин, М. Н. Статистика [Электронный ресурс]: учебник и практикум для вузов / М. Н. Дудин, Н. В. Лясников, М. Л. Лезина. — Москва : Издательство Юрайт, 2021. — 374 с. — (Высшее образование). — ISBN 978-5-9916-8908-3. — Текст : электронный // ЭБС Юрайт [сайт]. [Электронный ресурс]: — Режим доступа: <https://urait.ru/bcode/470169>.

Лепихина З.П. Статистика: Учебное пособие/ З. П. Лепихина; Федеральное агентство по образованию, Томский государственный университет систем управления и радиоэлектроники. - Томск: ТУСУР, 2005. – с.44-59.

Лацкевич, Н.В. Статистика [Электронный ресурс] : учебное пособие / Н.В. Лацкевич, С.А. Дещеня, Т.Н. Бессонова. — Электрон. дан. — Минск : "Вышэйшая школа", 2015. — с.29-31, 48-49. — Режим доступа: <https://e.lanbook.com/book/75126>.

При необходимости рекомендуется ознакомиться с другими источниками, приведенными в разделе «Список рекомендуемых источников».

2.3.3 Тема: Изучение периодической составляющей временного ряда

Перечень вопросов, подлежащих изучению

- 1) Составляющие временного ряда.
- 2) Сущность периодических (циклических, сезонных) колебаний.
- 3) Индекс сезонных колебаний.
- 4) Метод гармонического анализа.

Методические рекомендации по изучению

Динамический ряд теоретически может быть представлен в виде совокупности *трех составляющих*:

1) *тренд* — основная тенденция развития динамического ряда (тенденция к росту или к снижению);

2) *циклические (периодические) колебания*, в том числе сезонные;

3) случайные колебания.

Если в анализируемой временной последовательности наблюдаются устойчивые отклонения от тенденции (как в большую, так и в меньшую сторону), то можно предположить наличие в ряду динамики некоторых (одного или нескольких) *колебательных процессов*, то есть *циклические (периодические) колебания*

При изучении темы следует изучить методы выявления и оценки периодических колебаний временного ряда. Наличие периодической составляющей оценивается с помощью индексов сезонности, а моделирование осуществляется с помощью гармонического анализа.

Индексы сезонности показывают, во сколько раз фактический уровень ряда в момент или интервал времени t больше среднего уровня либо уровня, вычисляемого по уравнению функции тенденции $f(t)$. Следует обратить внимание, что способы определения индексов сезонности зависят от наличия или отсутствия основной тенденции.

Для анализа и моделирования внутригодовой динамики социально-экономических явлений (или явлений с 12-ти интервальным циклом) могут применяться *гармоники ряда Фурье*. В зависимости от числа гармоник аналитическое выражение изменений уровней ряда динамики выражается разными формулами.

Если применяется одна первая гармоника ряда Фурье, то уравнение модели (1) принимает вид:

$$\hat{y}_t = a_0 + a_1 \cos t + b_1 \sin t$$

Параметры модели определяются по методу наименьших квадратов. Оценка адекватности модели проводится по критерию Фишера. Для оценки точности модели вычисляют коэффициент детерминации: Студенту следует провести аналогию с построением линейной модели тренда.

Рекомендуемые источники

Дудин, М. Н. Статистика [Электронный ресурс]: учебник и практикум для вузов / М. Н. Дудин, Н. В. Лясников, М. Л. Лезина. — Москва : Издательство Юрайт, 2021. — 374 с. — (Высшее образование). — ISBN 978-5-9916-8908-3. — Текст : электронный // ЭБС Юрайт [сайт]. [Электронный ресурс]: — Режим доступа: <https://urait.ru/bcode/470169>.

Лепихина З.П. Статистика: Учебное пособие/ З. П. Лепихина; Федеральное агентство по образованию, Томский государственный университет систем управления и радиоэлектроники. - Томск: ТУСУР, 2005. – с.44-59.

При необходимости рекомендуется ознакомиться с другими источниками, приведенными в разделе «Список рекомендуемых источников».

2.3.4 Тема: Факторный анализ

Перечень вопросов, подлежащих изучению

- 1) Сущность факторного анализа как метода снижения размерности.
- 2) Основная модель факторного анализа.
- 3) Проблемы факторного анализа и схема решения задач.
- 4) Основные методы факторного анализа.

Методические рекомендации по изучению

Факторный анализ - это метод многомерного статистического анализа, позволяющий на основе экспериментального наблюдения признаков объекта выделить группу переменных, определяющих корреляционную взаимосвязь между признаками.

При изучении темы студенту необходимо повторить математические понятия матрицы, обратной матрицы, собственные числа и векторы, а также понятия статистической связи, определения ковариации и корреляции, основные положения регрессионного анализа.

Важно понимать, что с помощью метода факторного анализа можно решить четыре основные задачи:

- отыскание скрытых, но объективно существующих закономерностей, которые определяются воздействием внутренних и внешних причин на изучаемый процесс;
- сжатие информации путем описания процесса при помощи общих факторов, число которых значительно меньше количества первоначально взятых признаков;
- выявление и изучение статистической связи признаков с факторами;
- прогнозирование хода развития процесса на основе уравнения регрессии; уравнения регрессии, построенные при помощи результатов, полученных в факторном анализе, обладают значительными преимуществами перед классическим регрессионным анализом.

Студент должен уяснить основное предположение факторного анализа, которое заключается в том, что корреляционные связи между большим числом наблюдаемых переменных определяются существованием меньшего числа гипотетических наблюдаемых переменных или факторов.

Целью метода факторного анализа является представление величины z_{ij} , то есть элемента матрицы Z , в виде линейной комбинации нескольких факторов (гипотетических величин). Таким образом, значение z_{ij} может быть выражено в виде линейной комбинации r факторов

$$z_{ij} = a_{i1} \cdot p_{1j} + a_{i2} \cdot p_{2j} + \dots + a_{ir} \cdot p_{rj},$$

где z_{ij} - стандартизированное значение i -ой переменной для j -го индивидуума, то есть

$$z_{ij} = \frac{y_{ij} - \bar{y}_i}{\sigma_i}$$

a_{il} - вычисленные факторные нагрузки (постоянные коэффициенты, которые следует определить);

$p_{1j} - p_{rj}$ - значения фактора r у j -го объекта.

Это равенство отражает основную модель факторного анализа. В матричном виде модель для всех z_{ij} запишется

$$Z = A \cdot P,$$

где Z является матрицей порядка $m \times n$ стандартизированных переменных - исходных данных. $A = (a_{il})$ является матрицей порядка $m \times r$, которую можно определить. Она называется *факторным отображением*, а ее коэффициенты - *факторными нагрузками*. $P = (p_{lj})$ - матрицей порядка $r \times n$ значений всех факторов у всех объектов.

Таким образом, матрица Z представляет собой произведение двух матриц: A и P . При этом матрица A отражает связи переменных с факторами, а P описывает отдельные объекты.

Умножив обе части на A^{-1} (обратная матрица), можно методом множественного регрессионного анализа получить значения факторов для каждого объекта

$$P = A^{-1}Z$$

Студенту следует изучить основные проблемы факторного анализа (проблему общности, проблему факторов, проблему вращения, проблему значений факторов) и идею методов их решения.

Вычислительные процедуры, отражающие содержание этих методов, реализованы в стандартных программах, которые входят в большинство пакетов статистического анализа данных.

Рекомендуемые источники

Анализ данных : учебник для вузов / В. С. Мхитарян [и др.] ; под редакцией В. С. Мхитаряна. — Москва : Издательство Юрайт, 2022. — 490 с. — (Высшее образование). — ISBN 978-5-534-00616-2. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/489100>

Миркин, Б. Г. Введение в анализ данных : учебник и практикум / Б. Г. Миркин. — Москва: Издательство Юрайт, 2022. — 174 с. — (Высшее образование). — ISBN 978-5-9916-5009-0. — Текст : электронный // Образовательная платформа Юрайт [сайт] [Электронный ресурс]: — Режим доступа: <https://urait.ru/bcode/469306>.

При необходимости рекомендуется ознакомиться с другими источниками, приведенными в разделе «Список рекомендуемых источников».

2.3.5 Тема: Инструментальные средства статистического анализа данных

Перечень вопросов, подлежащих изучению

- 1) Обзор рынка программных продуктов по статистическому анализу данных

- 2) Изучение структуры программных продуктов
- 3) Сравнение функционала программных продуктов

Методические рекомендации по изучению

Рекомендуется начать работу с изучения официальных статистических данных о состоянии данной сферы. Например, ознакомиться с разделом 19. «ИНФОРМАЦИОННЫЕ И КОММУНИКАЦИОННЫЕ ТЕХНОЛОГИИ» в статистических сборниках «Регионы России. Социально-экономические показатели» на официальном сайте Федеральной службы государственной статистики. Далее следует изучить статьи в компьютерных журналах — периодических изданиях, основной темой которых являются информационные технологии, программное и аппаратное обеспечение. Задача студента — составить общее представление о спектре имеющихся на рынке программных продуктов. По согласованию с преподавателем студент отбирает для сравнительного анализа 3–4 продукта и проводит углубленное исследование, используя специальные журналы и интернет.

Объём обзора, как правило, от 7 до 10 машинописных страниц. Студент разрабатывает и оформляет отчет в соответствии с требованиями ОС ТУСУР 02-2021. Работы студенческие по направлениям подготовки и специальностям гуманитарного профиля.

Содержание основной части отчета должно разрабатываться в направлениях:

- рассмотрение понятий «программная система», «инструментальные средства», «пакет прикладных программ», «информационная технология» и т.д., в том числе в нормативных документах;
- классификация рассматриваемых программных продуктов;
- выбор показателей для сравнения программных продуктов (функционал, стоимость и т.д.);
- сравнительный анализ рассматриваемых программных продуктов по выбранным показателям (критериям).

При изучении источников студенту следует обратить внимание на возможности использования программного продукта для решения практических задач, степень открытости кода программ, репутацию разработчика.

Рекомендуемые источники

Программные системы статистического анализа. Обнаружение закономерностей в данных с использованием системы R и языка Python : учебное пособие / В. М. Волкова, М. А. Семёнова, Е. С. Четвертакова, С. С. Вожов. — Новосибирск : НГТУ, 2017. — 74 с. — ISBN 978-5-7782-3183-2. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book>

Боровиков, В.П. Популярное введение в современный анализ данных в системе STATISTICA. Учебное пособие для вузов [Электронный ресурс] : учебное пособие / В.П. Боровиков. — Электрон. дан. — Москва : Горячая линия-Телеком, 2013. — 288 с. — Режим доступа:

При необходимости рекомендуется ознакомиться с другими источниками, приведенными в разделе «Рекомендуемые источники».

2.4 Подготовка к контрольным работам

Контрольная работа – одна из форм проверки и оценки усвоенных знаний, получения информации о характере познавательной деятельности, уровня самостоятельности и активности студентов в учебном процессе.

При подготовке к выполнению контрольной работы необходимо повторить теоретический материал по теме, основные формулы и методы решения задач на данную тему. Следует вновь просмотреть примеры и задачи, разобранные в учебниках, на лекции и практических занятиях.

Важно понять, что если студент систематически работает над пройденным материалом, начиная с первой лекции, то подготовка к контрольной работе не вызовет затруднений и много времени на нее не понадобится.

2.5 Подготовка к лабораторным работам

Лабораторные занятия являются связующим звеном теории и практики. Они позволяют углубить и закрепить теоретические знания, получаемые на лекциях, проверить теоретические положения экспериментальным путем, выработать у студентов практические умения и навыки работы с реальной статистической информацией. Одновременно они являются базой для аналитической исследовательской работы студентов.

Содержание лабораторных работ и порядок выполнения определены в разделе 1 настоящих указаний. Следует помнить, что в начале методических указаний на выполнение каждой лабораторной работы приводится краткое изложение теоретических положений, поэтому студент должен заранее самостоятельно подготовиться к лабораторной работе с использованием указанной преподавателем литературы: Подготовить ответы на контрольные вопросы, предложенные преподавателем к данной лабораторной работе.

Каждая лабораторная работа выполняется по определенной теме с указанием цели её выполнения. Как правило, в лабораторной работе студент оперирует реальными статистическими данными. Поэтому необходимо чётко знать наименование объектов и признаков, единицы измерения признаков. Студенту необходимо уяснить цель работы и при подготовке к работе, при выполнении работы и анализе результатов следовать ей.

2.6 Подготовка к промежуточной аттестации

Подготовка к промежуточной аттестации осуществляется по вопросам, приведенным в рабочей программе дисциплины.

СПИСОК РЕКОМЕНДУЕМЫХ ИСТОЧНИКОВ

- 1) Публикации Федеральной службы государственной статистики: Регионы России. Социально-экономические показатели // Стат. сб. / Росстат. – М., 2014-2022. — Режим доступа: <https://rosstat.gov.ru/folder/210>
- 2) Дудин, М. Н. Статистика [Электронный ресурс]: учебник и практикум для вузов / М. Н. Дудин, Н. В. Лясников, М. Л. Лезина. — Москва : Издательство Юрайт, 2021. — 374 с. — (Высшее образование). — ISBN 978-5-9916-8908-3. — Текст : электронный // ЭБС Юрайт [сайт]. [Электронный ресурс]: — Режим доступа: <https://urait.ru/bcode/470169>
- 3) Годин, А.М. Статистика [Электронный ресурс] : учебник / А.М. Годин. — Электрон. дан. — Москва : Дашков и К, 2017. — с.202-218. — Режим доступа: <https://e.lanbook.com/book/93468>
- 4) Программные системы статистического анализа. Обнаружение закономерностей в данных с использованием системы R и языка Python : учебное пособие / В. М. Волкова, М. А. Семёнова, Е. С. Четвертакова, С. С. Вожов. — Новосибирск : НГТУ, 2017. — 74 с. — ISBN 978-5-7782-3183-2. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book>
- 5) Боровиков, В.П. Популярное введение в современный анализ данных в системе STATISTICA. Учебное пособие для вузов [Электронный ресурс] : учебное пособие / В.П. Боровиков. — Электрон. дан. — Москва : Горячая линия-Телеком, 2013. — 288 с. — Режим доступа: <https://e.lanbook.com/book/11828>
- 6) Анализ данных : учебник для вузов / В. С. Мхитарян [и др.] ; под редакцией В. С. Мхитаряна. — Москва : Издательство Юрайт, 2022. — 490 с. — (Высшее образование). — ISBN 978-5-534-00616-2. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/489100>
- 7) Миркин, Б. Г. Введение в анализ данных : учебник и практикум / Б. Г. Миркин. — Москва: Издательство Юрайт, 2022. — 174 с. — (Высшее образование). — ISBN 978-5-9916-5009-0. — Текст : электронный // Образовательная платформа Юрайт [сайт] [Электронный ресурс]: — Режим доступа: <https://urait.ru/bcode/469306>
- 8) Лепихина З.П. Статистика: Учебное пособие/ З. П. Лепихина; Федеральное агентство по образованию, Томский государственный университет систем управления и радиоэлектроники. - Томск: ТУСУР, 2005. – 284 с.
- 9) Лацкевич, Н.В. Статистика [Электронный ресурс] : учебное пособие / Н.В. Лацкевич, С.А. Дещеня, Т.Н. Бессонова. — Электрон. дан. — Минск: "Вышэйшая школа", 2015. — 369 с. — Режим доступа: <https://e.lanbook.com/book/75126>
- 10) Балдин, К.В. Общая теория статистики [Электронный ресурс] : учебное пособие / К.В. Балдин, А.В. Рукосуев. — Электрон. дан. — Москва : Дашков и К, 2017. — 312 с. — Режим доступа: <https://e.lanbook.com/book/93403>.
- 11) Колесникова, И.В. Статистика. Практикум [Электронный ресурс] : учебное пособие / И.В. Колесникова, Г.В. Круглякова. — Электрон. дан. — Минск : "Вышэйшая школа", 2011. — 285 с. — Режим доступа: <https://e.lanbook.com/book/65326>
- 12) Лепихина, З.П. Основы социального прогнозирования: Учебное пособие/ З. П. Лепихина; Федеральное агентство по образованию, Томский государственный университет систем управления и радиоэлектроники, Кафедра автоматизации обработки информации. - Томск: ТМЦДО, 2006. – 112 с.

ПРИЛОЖЕНИЕ 1

Исходные данные к лабораторной работе «Первичный анализ социально-экономической информации»

Таблица Размер заработной платы, тыс. руб.

Вариант 1	Вариант 2	Вариант 3	Вариант 4	Вариант 5
23	23	18	20	30
20	20	22	18	20
25	25	32	20	20
20	20	18	20	18
20	20	20	23	18
18	18	16	20	20
20	20	20	18	18
18	18	18	22	20
20	20	24	18	20
20	20	20	22	23
23	23	23	32	20
20	20	20	18	18
18	18	18	20	22
22	22	22	23	23
18	18	16	20	30
22	22	22	30	20
32	32	23	20	18
18	18	20	20	20
20	20	24	18	20
16	17	20	18	23

Вариант 6	Вариант 7	Вариант 8	Вариант 9	Вариант 10
23	23	18	20	25
20	20	22	18	20
25	25	28	20	20
20	20	18	20	18
20	20	20	23	18
18	18	16	20	20
20	20	20	18	18
18	18	18	22	20
20	20	24	18	20
20	20	20	22	23
23	23	23	27	20
20	20	20	18	18
18	18	18	20	22
22	16	22	23	23
18	18	16	20	23
22	22	22	27	20
29	32	23	20	17
18	18	20	20	20
20	20	24	18	20
17	17	20	18	23

ПРИЛОЖЕНИЕ 2

Исходные данные к лабораторной работе «Построение и исследование типологии социально-экономических объектов методом статистической группировки».

Таблица - Основные показатели развития регионов Сибирского федерального округа

	Валовой региональный продукт в 2015 г. , млн. руб.	Численность населения на 1 января 2017 г., тыс. человек	Средне-душевые денежные доходы (в месяц), руб.	Потребительские расходы в среднем на душу населения (в месяц), руб.	Основные фонды в экономике (по полной учетной стоимости; на конец года) ²⁾ , млн. руб.
	1	2	3	4	5
Республика Алтай	41776,8	217,0	13836,9	7179,0	61628
Республика Бурятия	204156,2	984,1	15715,5	11340,0	430210
Республика Тыва	47287,3	318,6	10962,8	4944,6	47409
Республика Хакасия	171663,9	537,7	14222,8	9680,5	292915
Алтайский край	492138,9	2365,7	12499,9	9765,7	757632
Забайкальский край	248847,6	1079,0	15968,8	10572,7	650405
Красноярский край	1618166,0	2875,3	20145,5	14105,7	1815754
Иркутская область	1013542,3	2408,9	16017,2	10580,2	1975486
Кемеровская область	842618,9	2708,8	16666,0	11237,2	1406912
Новосибирская область	980850,5	2779,5	18244,1	14898,1	1229181
Омская область	617184,4	1972,7	17247,9	12663,1	725451
Томская область	473693,1	1078,9	16516,0	11199,4	863117

Продолжение Таблицы

Объем отгруженных товаров собственного производства, выполненных работ и услуг собственными силами по видам экономической деятельности, млн. руб.			Продукция сельского хозяйства - всего, млн. руб.	Инвестиции в основной капитал, млн. руб.	
добыча полезных ископаемых	обрабатывающие производства	Производство и распределение электроэнергии, газа и воды			
6	7	8	9	10	
862	1296	1654	8020	11802	Республика Алтай
12808	51115	19826	13044	41017	Республика Бурятия
3376	514	2868	4648	7033	Республика Тыва
26536	56595	21462	9371	38064	Республика Хакасия
6041	189279	31991	93784	70833	Алтайский край
40377	14365	18311	15154	51557	Забайкальский край
266636	628113	95432	68598	303885	Красноярский край
129795	299406	81275	43610	137995	Иркутская область
507993	385413	85949	38044	225131	Кемеровская область
19674	249816	61592	60425	142078	Новосибирская область
4411	529355	36013	66911	83342	Омская область
137513	100598	25617	19420	101927	Томская область

ПРИЛОЖЕНИЕ 3

Исходные данные к лабораторной работе «Построение типологии социально-экономических объектов методами кластер-анализа.

Таблица 1- Основные показатели развития регионов Сибирского федерального округа
(округленные значения)

	Валовой региональный продукт в 2014 г. , млрд. руб.	Оборот розничной торговли, млрд. руб.	Добыча полезных ископаемых млрд. руб.	Инвестиции в основной капитал, млрд. руб.	Продукция сельского хозяйства - всего, млрд. руб.	Средне-месячная номинальная начисленная заработная плата работников, тыс.руб.	обрабатывающие производства млрд. руб.
	1	2	3	4	5	6	7
1. Республика Алтай	39	14	1	12	8	15	1
2. Республика Бурятия	185	100	13	41	13	20	51
3. Республика Тыва	46	13	3	7	5	19	1
4. Республика Хакасия	160	46	26	38	9	21	56
5. Алтайский край	448	218	6	71	94	14	189

Таблица 2- Основные показатели развития регионов Сибирского федерального округа
(округленные значения)

	Валовой региональный продукт в 2014 г. , млрд. руб.	Оборот розничной торговли, млрд. руб.	Добыча полезных ископаемых млрд. руб.	Инвестиции в основной капитал, млрд. руб.	Продукция сельского хозяйства - всего, млрд. руб.	Средне-месячная номинальная начисленная заработная плата работников, тыс.руб.	обрабатывающие производства млрд. руб.
	1	2	3	4	5	6	7
1. Красноярский край	1423	3617	267	304	68	26	630
2. Иркутская область	907	225	130	138	44	23	300
3. Кемеровская область	747	287	508	225	38	20	380
4. Новосибирская область	895	368	20	142	60	20	250
5. Томская область	428	93	138	102	19	24	100

Исходные данные к лабораторной работе «Методы исследования взаимосвязей нечисловых данных».

Часть 1»

Таблица -Данные 50 респондентов о предпочитаемых напитках

№ респондента	ПОЛ	НАПИТОК	№ респондента	ПОЛ	НАПИТОК	№ респондента	ПОЛ	НАПИТОК
1	1	1	18	2	1	35	1	2
2	1	2	19	2	2	36	1	2
3	2	2	20	1	2	37	2	1
4	1	1	21	1	1	38	2	1
5	1	1	22	2	1	39	2	1
6	2	2	23	1	1	40	1	2
7	2	2	24	2	2	41	1	1
8	1	1	25	1	2	42	1	2
9	2	1	26	1	2	43	2	1
10	1	1	27	1	1	44	1	1
11	2	1	28	2	1	45	1	2
12	1	2	29	2	1	46	1	1
13	1	1	30	1	2	47	2	2
14	1	2	31	1	1	48	2	1
15	2	1	32	2	2	49	2	2
16	2	1	33	2	1	50	1	2
17	1	2	34	1	2			

Исходные данные к лабораторной работе «Методы исследования взаимосвязей нечисловых данных. Часть 2»

Таблица 2 - Ответы 30 респондентов на 11 вопросов

№ анкеты	№ вопроса										
	1	2	3.	4.	5	6.	7.	8	9	10	11
1	1	1	1	1	1	4	3	4	1	19	ТПУ
2	1	2	2	2	1	4	4	4	1	20	ТПУ
3	2	0	1	2	0	4	3	4	1	22	ТПУ
4	2	0	2	1	0	3	3	4	1	19	ТПУ
5	1	1	2	2	1	3	3	3	1	19	ТПУ
6	1	1	2	3	1	1	4	1	1	20	ТПУ
7	1	2	3	3	1	4	4	2	1	19	ТУСУР
8	1	2	1	1	0	2	3	1	1	19	ТУСУР
9	3	1	2	2	1	3	3	4	1	20	ТПУ
10	2	2	1	1	1	4	3	4	1	19	ТПУ
11	1	1	2	2	1	4	4	4	2	20	ТПУ
12	1	1	1	2	1	1	1	2	1	18	ТПУ
13	1	1	1	1	0	4	4	4	1	19	ТПУ
14	1	0	1	3	0	2	4	4	2	19	ТПУ
15	2	2	2	2	0	4	4	2	2	19	ТУСУР
16	1	2	2	1	1	2	2	2	1	19	ТУСУР
17	1	2	1	1	1	3	3	4	2	20	ТУСУР
18	3	2	2	3	1	4	3	4	2	20	ТУСУР
19	1	2	2	2	0	4	4	4	2	22	ТПУ
20	1	2	2	2	1	4	4	4	2	20	ТУСУР
21	0	2	2	2	1	1	1	1	2	20	ТУСУР
22	3	2	3	2	1	4	4	4	2	20	ТУСУР
23	3	2	1	2	0	4	4	4	2	19	ТУСУР
24	1	1	1	1	0	4	4	4	2	19	ТПУ
25	1	2	1	1	0	4	4	4	2	19	ТПУ
26	1	1	1	1	1	4	4	4	2	20	ТПУ
27	2	2	2	2	1	4	4	4	2	19	ТПУ
28	1	1	2	2	1	3	3	4	2	20	ТПУ
29	3	2	2	2	0	4	4	4	2	20	ТУСУР
30	1	1	1	2	1	4	4	4	2	22	ТУСУР

ПРИЛОЖЕНИЕ 5

Исходные данные к лабораторной работе «Исследования взаимосвязей числовых данных методами аналитической группировки и корреляционно-регрессионного анализа»

Таблица - Основные показатели развития регионов Сибирского федерального округа (2015 год)

	ВРП в 2014 г. , млрд. руб.	Численность населения на 1 января 2014г., млн. человек	Инвестиции в основной капитал, млн. руб	Средне- душевые денежные доходы (в месяц), ттыс.руб.	Средне- месячная номинальная начис-ленная заработная плата работ- ников, тыс.руб.	Продукция сельского хозяйства - всего, млрл. руб	Ввод в дей- ствие общей пло- щади жилых до- мов, тыс. м ²
	1	2	3	4	5	6	7
Республика Алтай	39	0,2	12	14	16	8	77
Республика Бурятия	185	1,0	41	16	20	13	304
Республика Тыва	47	0,3	7	11	19	5	52
Республика Хакасия	160	0,5	38	14	21	9	156
Алтайский край	448	2,4	71	12	14	94	663
Забайкальский край	228	1,1	52	16	21	15	277
Красноярский край	1423	2,9	304	20	26	69	1047
Иркутская область	907	2,4	138	16	23	44	755
Кемеровская область	747	2,7	225	17	20	38	1083
Новосибирская область	895	2,8	142	18	20	60	1505
Омская область	599	2,0	83	17	19	67	837
Томская область	428	1,1	102	17	24	19	458

Значения F-распределения Фишера

Уровень значимости $\alpha=0,05$

Число степеней свободы $\nu_1 = k-1=1$, $\nu_2 = n-k = n-2$,

где k – число параметров функции, описывающей тенденцию (для линейной функции $k=2$); n – число уровней ряда

$\nu_2 = n-2$	F_{tab}	$\nu_2 = n-2$	F_{tab}	$\nu_2 = n-2$	F_{tab}
1	162,4	11	4,84	21	4,32
2	18,51	12	4,75	22	4,30
3	10,13	13	4,67	23	4,28
4	7,71	14	4,60	24	4,26
5	6,61	15	4,55	25	4,24
6	5,99	16	4,51	26	4,22
7	5,59	17	4,45	27	4,21
8	5,32	18	4,41	28	4,19
9	5,12	19	4,38	29	4,18
10	4,96	20	4,35	30	4,17

ПРИЛОЖЕНИЕ 7

Исходные данные к лабораторной работе «Анализ и прогнозирование динамики развития социально-экономических процессов»

Таблица – Международная миграция (человек)

	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Прибыло в РФ – из стран:												
Австралия	30	28	38	31	39	49	83	78	113	71	89	82
Австрия	24	53	50	35	37	45	60	70	84	81	122	83
Вьетнам	114	157	921	714	950	921	3294	3653	3852	3854	4012	3735
Германия	3025	2900	3164	3134	2585	2621	4520	4239	4166	3743	3976	4153
Греция	200	176	260	289	240	298	614	835	995	694	557	450
Израиль	1004	1053	1094	1002	861	814	1240	1091	1132	1139	1077	900
Индия	54	72	107	66	72	110	1390	1068	1451	1850	2894	4768
Канада	99	77	118	105	98	110	192	207	226	171	189	193
Китай	432	499	1687	1177	770	1380	7063	8547	8149	10563	9043	8027
Польша	55	48	96	100	97	105	187	200	217	199	194	181
Великобритания	40	34	100	80	92	125	166	182	221	185	273	226
США	396	411	578	551	575	653	947	1122	954	1000	1084	1137
Турция	86	172	315	373	443	562	1832	2252	2755	2631	2091	1626
Финляндия	129	137	172	174	141	178	266	342	429	468	401	393
Франция	40	54	144	72	96	150	322	326	352	351	360	303