

Министерство науки и высшего образования Российской Федерации

Федеральное государственное бюджетное образовательное учреждение
высшего образования

**«ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
СИСТЕМ УПРАВЛЕНИЯ И РАДИОЭЛЕКТРОНИКИ» (ТУСУР)**

Кафедра автоматизации обработки информации (АОИ)

ПРИКЛАДНАЯ СТАТИСТИКА

Методические указания к лабораторным работам,
практическим занятиям и организации самостоятельной работы
для студентов направления
«Бизнес-информатика» (уровень бакалавриата)

Лепихина Зинаида Павловна

Прикладная статистика: Методические указания к лабораторным работам, практическим занятиям и организации самостоятельной работы для студентов направления «Бизнес-информатика» (уровень бакалавриата) / З.П.Лепихина. – Томск, 2018. – 78 с..

© Томский государственный университет систем управления и радиоэлектроники, 2018

© Лепихина З.П., 2018

Оглавление

1	Введение.....	5
2	Методические указания к лабораторным работам.....	6
	2.1 Лабораторная работа «Расчет и анализ статистических показателей».....	6
	2.2 Лабораторная работа «Анализ и прогнозирование временных рядов».....	11
	2.3 Лабораторная работа «Построение и анализ типологии объектов».....	16
	2.4 Лабораторная работа «Анализ взаимосвязи нечисловых признаков».....	24
3	Методические указания к проведению практических занятий.....	31
	3.1 Практическое занятие «Прикладная статистика как наука. Категории статистики. Статистическое исследование».....	31
	3.2 Практическое занятие «Статистические величины и методы их исчисления».....	33
	3.3 Практическое занятие «Статистические группировки».....	39
	3.4 Практическое занятие «Анализ ранговых корреляций».....	45
4	Методические указания для организации самостоятельной работы.....	51
	4.1 Общие положения.....	51
	4.2 Проработка лекционного материала.....	51
	4.3 Самостоятельное изучение тем теоретической части курса.....	52
	4.3.1 Тема: Основы выборочного метода.....	52
	4.3.2 Тема: Представление статистических данных.....	54
	4.3.3 Тема: Факторный анализ как метод снижения размерности.....	56
	4.4 Домашнее задание.....	58
	4.4.1 Тема: Организация государственной статистики в Российской Федерации.....	58
	4.4.2 Тема: Выборочный метод: расчет объема и ошибок выборки.....	60
	4.6 Подготовка к контрольным работам.....	62
	4.7 Подготовка к практическим занятиям.....	63
	4.8 Подготовка к лабораторным работам.....	63

	4.9 Индивидуальное задание «Инструментальные средства статистического анализа данных.....	64
5	Рекомендуемые источники.....	66
	Приложение 1.....	67
	Приложение 2.....	69
	Приложение 3.....	70
	Приложение 4.....	71
	Приложение 5.....	73
	Приложение 6.....	75
	Приложение 7.....	76

1 Введение

Цель изучения дисциплины «Прикладная статистика» — дать студентам представление о содержании статистики как научной дисциплины, познакомить с ее основными понятиями, методами получения статистической информации о социально-экономических явлениях и ее обобщения, научить использовать современные методы и технологии обработки данных для решения задач анализа данных и выявления тенденций в развитии социально-экономических процессов.

Задачи изучения дисциплины — сформировать у студентов знания, умения и навыки в использовании методов получения статистической информации; развить навыки и способности студентов к применению современных теоретических и эмпирических моделей для решения конкретных задач анализа данных; сформировать умение правильно выбирать математический аппарат и инструментальные средства для обработки результатов исследования в соответствии с его целями, задачами, гипотезами и имеющимися данными; развить необходимые навыки работы с компьютером как средством управления информацией, решения конкретных задач, возникающих при исследовании различных объектов.

В данных Методических указаниях содержится:

- краткое изложение теоретического материала по теме, варианты заданий и порядок выполнения лабораторных работ;
- примеры решения типовых задач для подготовки к практическим занятиям и задачи, решение которых необходимо для закрепления изученного материала;
- рекомендации по организации самостоятельной работы.

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.

При самостоятельной работе и подготовке к лабораторным и практическим занятиям студенту следует повторить теоретический материал по конспекту лекций и источникам, приведенным в разделе «Рекомендуемая литература», а также пользоваться информацией, представленной в статистических сборниках, в научной литературе и Интернете

2 Методические указания к проведению лабораторных работ

2.1 Лабораторная работа «Расчет и анализ статистических показателей»

Цель работы

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

Форма проведения

Выполнение индивидуального задания.

Форма отчетности

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 (M_e) — 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:

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

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

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

Если данные сгруппированы

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

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

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

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

Варианты заданий

Исследован возраст посетителей клуба. Исходные данные приведены в Приложении 1. В соответствии с номером варианта выбрать исходные данные.

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

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

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

Контрольные вопросы и задания

- 1) Дайте определение средней величины.
- 2) Перечислите виды степенных средних величин.
- 3) По какой формуле вычисляется средняя величина в случае «сырых» данных?
- 4) Как отражаются на графике среднее, мода, медиана?
- 5) По какой формуле вычисляется мода в интервальном вариационном ряду?
- 6) Как определяется мода в дискретном вариационном ряду?
- 7) Сравните значения средних величин в трех случаях.
- 8) Объясните порядок вычисления медианы в интервальном вариационном ряду.
- 9) Что такое дециль?
- 10) Как определяется третий квартиль?

2.2 Лабораторная работа «Анализ и прогнозирование временных рядов»

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 (динамической) 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 :

$$\Delta y_{\text{б}i} = y_i - y_1.$$

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

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

Темп роста базисный (в процентах) определяется по формуле

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

Темп роста цепной (в процентах) определяется по формуле

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

Темп прироста базисный (в процентах) определяется по формуле

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

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

$$\bar{\text{Тр}} = \sqrt[n-1]{y_n : y_1} \times 100\%$$

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

$$\bar{\text{Тп}} = \bar{\text{Тр}} - 100.$$

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

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

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:

$$\Delta y_i = const .$$

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

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

где a_0 и a_1 — параметры уравнения;

t — обозначение времени

При вычисления параметров функции на основе требований метода наименьших квадратов получаются 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 2. 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 8.

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.

- 1) Вычислить
 - базисные и цепные абсолютные приросты,
 - базисные и цепные темпы роста,
 - базисные и цепные темпы прироста мигрантов

- 2) Вычислить средние показатели
 - среднегодовую численность мигрантов
 - среднегодовой абсолютный прирост,
 - среднегодовой темп роста,
 - среднегодовой темп прироста

Задание 2

- 1) Вычислить прогнозное значение на следующий 2017 год, используя средний абсолютный прирост

Задание 3

- 1) Провести выравнивание временного ряда методами
 - укрупнения интервалов (перейти к трехгодовым периодам);
 - трехзвенной скользящей средней.
- 2) Построить графики исходных и выровненных значений

Задание 4

- 1) построить линейную модель для выбранных данных, вычисляя коэффициенты a_0 и a_1
- 2) провести оценку модели по критерию Фишера,
- 3) рассчитать коэффициент детерминации.
- 4) построить график динамики исходных и выровненных значений.
- 5) вычислить прогнозное значение на следующий год.

Контрольные вопросы и задания.

- 1) Дайте определения показателей динамики.
- 2) Объясните выбор формул для расчета показателей динамики.
- 3) Укажите на графике абсолютные цепные и базисные приросты
- 4) Объясните результаты построения линейных моделей.
- 5) Какая из полученных средствами MS Excel моделей точнее?
- 6) Запишите формулу пятизвенной скользящей средней
- 7) Каким показателем оценивается точность модели?

2.3 Лабораторная работа «Построение и анализ типологии объектов»

Цель работы

Закрепление теоретического материала и получение практических навыков исчисления статистических показателей. Построение и анализ типологии объектов. Исследование динамики макроэкономических показателей.

Форма проведения

Выполнение индивидуального задания.

Форма отчетности

Устный опрос, демонстрация расчетов, выполнение дополнительных заданий.

Теоретические основы

Группировка – это распределение единиц по группам в соответствии со следующим принципом: различия между единицами, отнесенны-

ми 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

$$\sigma_{\text{м.сп.}}^2 = \frac{\sum (\bar{x}_j - \bar{x})^2 f_j}{\sum f_j},$$

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

Коэффициент детерминации R^2 вычисляется по формуле

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

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

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 расстояния $d_{i+j,k}$ 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:

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

На первом шаге «извлекается» точка X_{k+1} и выясняется, к какому из эталонов она ближе, то есть рассчитываются расстояния от точки X_{k+1} до каждого эталона. Этот ближайший эталон заменяется новым эталоном – центром тяжести старого и присоединенной точки – с увеличением ве-

са, а остальные эталоны не изменяются. Затем появляется следующая точка, и опять выясняется, к какому из эталонов она ближе и т.д.

На ν -ом шаге извлекается $X_{k+\nu}$ и алгоритм пересчета эталонных точек следующий

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

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

$$i = \overline{1, k}.$$

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

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

Задание 1.

1) В соответствии с номером варианта определить номер показателя субъектов федерации Сибирского федерального округа, представленные в таблице «Основные социально-экономические показатели в 2016 г.»
Приложения 4

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

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

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

- 2) Провести сортировку регионов по значению **показателя**
- 3) Провести по всей совокупности *регионов* расчет
 - среднего значения (функция СРЗНАЧ)
 - дисперсии (функция ДИСПР)
 - стандартного отклонения (функции СТАНДОТКЛОНП, КО-РЕНЬ)
 - коэффициента вариации
- 4) 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 при- знака– фактора x	Частота группы f_j	Показатели		
			Среднее	Дисперсия	Коэффициент вариации
Малые					
Средние					
Крупные					

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

- частоту группы,
 - групповые средние значения показателя
 - групповые дисперсии показателя,
 - групповые коэффициенты вариации показателей;
- 6) Вычислить межгрупповую дисперсию.
- 7) Вычислить коэффициент детерминации и сделать вывод о качестве построенной группировки.

Задание 2.

1) В соответствии с номером варианта определить номера таблицы исходных данных, номера **двух показателей** субъектов федерации СФО, представленные в таблицах Приложения 5, и исследуемый иерархический алгоритм.

Вариант	Номер таблицы в Приложении 9	x-признак-фактор (номер показателя)	У-признак-результат-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»
11.	1	7	1	«дальнего соседа»
12.	2	7	1	«медианной связи»

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

Название региона	Показатель (фактор) X	ВРП (результат) У
1.		
...
5.		

3) Построить точечную диаграмму в пространстве 2-х признаков.

- 4) Провести кластерный анализ, используя указанный в варианте *иерархический алгоритм*. Определить вариант разбиения **на 3 класса** и результаты разбиения записать в табл.2. Построить дендрограмму.
- 5) 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

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

- 6) Сравнить результаты, полученные двумя алгоритмами кластер-анализа.
- 7) *По результатам метода К-средних:*
 - а. внести в рабочую таблицу дополнительный столбец с номером кластера и провести сортировку по номеру кластера;
 - б. построить точечную диаграмму в пространстве двух признаков с изображением принадлежности регионов кластерам (выделить разным цветом);
- 8) 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 значений границ группировочного признака-фактора.
- 4) Запишите формулу вычисления межгрупповой дисперсии.
- 5) Можно ли признать построенную группировку «правильной»?
- 6) Объясните выбор эталонных точек в алгоритме К-средних.
- 7) Охарактеризуйте группы, полученные в результате применения алгоритма К-средних.
- 8) Сравните результаты типологической группировки и кластер-анализа.

2.4 Лабораторная работа «Анализ взаимосвязи нечисловых признаков»

Цель работы

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

Форма проведения

Выполнение индивидуального задания.

Форма отчетности

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:

a	b	$a+b$
c	d	$c+d$
$a+c$	$b+d$	n

Суммы $a+b$, $a+c$, $c+d$, $b+d$ называются маргинальными суммами. Если между A и B не существует никакой связи, т.е. если обладание признаком A не связано с обладанием признаком B , то доля индивидов с признаком A среди индивидов, обладающих признаком B , должна быть равна доле индивидов с признаком A среди индивидов, не обладающих признаком B .

Таким образом, по определению признаки **независимы** в данной совокупности из n наблюдений, если

$$\frac{a}{a+c} = \frac{b}{b+d} = \frac{a+b}{n}. \quad (2.2.1)$$

Соотношение (2.2.1) можно переписать в виде:

$$a = \frac{(a+b)(a+c)}{n}$$

Это теоретическая частота в предположении независимости.

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

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

Меры связи в таблицах 2x2

Коэффициент ассоциации

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

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;

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

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

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

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

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

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

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 . Например, при уровне $\alpha = 0,05$ для $df = 1$ $\chi^2 = 3.84$.

Шаг 4. Сравниваем фактическое и табличное значения χ^2 . Если $\chi_{\text{ф}}^2 \geq \chi_{df, \alpha}^2$, 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 $\alpha = 0,05$ ($\alpha = 5\%$). Таблица значений χ^2 при $\alpha = 0,05$ для различных степеней свободы df приведена в Приложении 6.

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, однако, при полной связи между переменными только в том случае, если $r = c$. Во всех других случаях, даже при полной связи $T^2 < 1$.

Варианты заданий и порядок выполнения

Исходные данные приведены в Приложении 7.

Задание 1. В Табл.1 Приложения 7 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 7 для анализа 20 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

2) Построить четырехклеточную таблицу вида

ПОЛ	НАПИТОК	
	1-pepsi	2-cola
Мужской (1)	<i>a</i>	<i>b</i>
Женский (2)	<i>c</i>	<i>d</i>

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

4) 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.

5) Провести анализ таблицы 2x2:

- установить **наличие** связи,
- рассчитать коэффициент ассоциации
- рассчитать коэффициент контингенции

6) Объяснить полученные результаты

Задание 2. В Табл.2 Приложения 7 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 СВЕРСНИКАМИ- ЮНОШАМИ (1-легко, 2-спокойно 3-трудно)

4 КАК СКЛАДЫВАЮТСЯ ВЗАИМООТНОШЕНИЯ СО СВЕРСНИКАМИ- 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

- 2) Провести сортировку данных по первой переменной (вопрос разреза), а затем по второй (вопрос анализа).
- 3) Построить таблицу сопряженности $r \times c$.
- 4) Провести одномерный анализ данных: рассчитать
 - a. частоты по вопросу разреза;
 - b. частоты по вопросу анализа;
 - c. относительные частоты по каждому признаку;
 - d. построить диаграммы распределения респондентов по значениям признаков (гистограммы, секторные; линейные и др.)
- 5) Провести анализ таблицы $r \times c$:
 - a. рассчитать значение Хи-квадрат;
 - b. установить наличие связи;
 - c. рассчитать коэффициент Фи;
 - d. рассчитать коэффициент Крамера;
 - e. рассчитать коэффициент Чупрова.
- 6) Дать содержательную интерпретацию полученным результатам.

Контрольные вопросы и задания

- 1) Что такое «дихотомические» признаки? Приведите пример.

- 2) Какие показатели характеризуют силу связи в четырех-клеточных таблицах?
- 3) Если связь между признаками отсутствует, какое значение принимает коэффициент ассоциации?
- 4) Как определить теоретическое значение Хи-квадрат критерия?
- 5) В каком случае значения коэффициентов Крамера и Чупрова совпадают?
- 6) Как определить значение числа степеней свободы таблицы сопряженности?
- 7) Объясните, что означает «уровень значимости» и по какому принципу выбирается его значение.

3 Методические указания к проведению практических занятий

3.1 Практическое занятие «Прикладная статистика как наука. Категории статистики. Статистическое исследование»

Цель занятия

Цель проведения практического занятия – изучение и закрепление теоретического материала; приобретение умения составить план статистического исследования состояния экономической, социальной, политической среды и сформировать систему характеризующих её показателей; получение навыков нахождения и содержательной интерпретации статистической информации.

Рекомендации по подготовке к занятию

При подготовке к занятию студенту следует повторить теоретический материал по конспекту лекций и следующим источникам:

Годин, А.М. Статистика [Электронный ресурс] : учебник / А.М. Годин. — Электрон. дан. — Москва : Дашков и К, 2017. — с.9-54. — Режим доступа: <https://e.lanbook.com/book/93468>.

Лепихина З.П. Статистика: Учебное пособие/ З. П. Лепихина; Федеральное агентство по образованию, Томский государственный университет систем управления и радиоэлектроники. - Томск: ТУСУР, 2005. – с.5-44.

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

В процессе изучения темы важно уяснить, что в настоящее время термин «статистика» употребляется в трех значениях: специальная научная дисциплина, отрасль практической деятельности, статистические данные. Следует обратить особое внимание на понятие «вариация признака», которая выражает индивидуальные особенности единицы совокупности и обусловлена различным сочетанием влияющих факторов. Именно наличие вариации предопределяет необходимость статистики. Статистика как наука исследует не отдельные факты, а массовые социально-экономические явления и процессы, выступающие как множество отдельных факторов, обладающих как индивидуальными, так и общими признаками. Здесь необходимо подчеркнуть связь статистики с законом больших чисел.

Совокупность приемов, способов и методов изучения количественных сторон массовых общественных явлений образует статистическую методологию, которая используется в статистических исследованиях.

При изучении *основных категорий* статистики необходимо подобрать несколько конкретных примеров для иллюстрации понятий, например, рост студента Иванова – это признак, а средний рост мужчин – это статистический показатель.

Особое внимание следует обратить на программно-методологические вопросы организации статистического исследования.

Порядок проведения занятия

Занятие проводится в форме семинара. Студентам формулируются вопросы и задания, проводится устный опрос. Выступления студентов обсуждаются и оцениваются преподавателем

Варианты заданий

Вопросы для обсуждения

- a) Предмет, метод и задачи прикладной статистики.
- b) Дайте определения основным категориям статистики. Приведите примеры статистической совокупности, количественных и качественных признаков, статистических показателей.
- c) Что такое «статистическая закономерность»?
- d) Какие задачи и функции выполняет Росстат?
- e) Перечислите стадии статистического исследования. Раскройте их содержание.
- f) Охарактеризуйте виды, формы, способы статистического наблюдения.
- g) Какие ошибки присущи выборочному исследованию?
- h) Что такое «репрезентативность» выборки?
- i) Дайте характеристику способам формирования выборки.

Задание.

Требуется рассмотреть указанное в варианте явление (процесс) с точки зрения статистического наблюдения и определить его форму, вид и способ.

Пример выполнения задания.

Предприятие предоставляет в налоговые органы баланс предприятия за 2017 год.

Ответ:

форма статистического наблюдения – общегосударственная отчетность, вид (по охвату явления) – сплошное наблюдение, вид (по частоте) – периодическое наблюдение, способ – документальный.

Варианты для самостоятельного выполнения.

- a) Всероссийская перепись населения 2010 г.
- b) Предприятие предоставляет в органы государственной статистики «Отчет промышленного предприятия о выполнении плана по труду за IV квартал 2017 года».
- c) Проводится инвентаризация имущества на предприятии.
- d) Проводится проверка качества партии швейных изделий: в случайном порядке отбирается десятая часть этой партии и каждое изделие тщательно осматривается.
- e) При мониторинге цен на сахар каждый понедельник регистратор фиксирует цены у трех крупнейших продавцов, поставляющих сахар в 90% магазинов города.
- f) Для изучения мнения студентов кафедра разместила на сайте анкету с просьбой ответить на вопросы.

3.2 Практическое занятие «Статистические величины и методы их исчисления»

Цель занятия

Цель проведения практического занятия – изучение и закрепление теоретического материала; приобретение умения выбрать адекватный метод исчисления статистических показателей, характеризующих состояния экономической, социальной, политической среды; получение навыков обработки данных, представления и содержательной интерпретации полученных результатов расчетов.

Рекомендации по подготовке к занятию

При подготовке к занятию студенту следует повторить теоретический материал по конспекту лекций и следующим источникам:

Годин, А.М. Статистика [Электронный ресурс] : учебник / А.М. Годин. — Электрон. дан. — Москва : Дашков и К, 2017. — с.72-118. — Режим доступа: <https://e.lanbook.com/book/93468>.

Лепихина З.П. Статистика: Учебное пособие/ З. П. Лепихина; Федеральное агентство по образованию, Томский государственный университет систем управления и радиоэлектроники. - Томск: ТУСУР, 2005. – с.44-81.

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

При изучении абсолютных величин следует обратить внимание, что они отражают физические размеры явлений или объем (численность) совокупности и всегда являются именованными числами.

Для аналитических целей в статистической практике широко применяются относительные величины, представляющие результат деления одного показателя на другой и выражающие соотношения между количественными характеристиками. При расчете относительных величин показатель, с которым проводится сравнение, является делителем и называется основанием или базой сравнения. Относительный показатель выражается в коэффициентах (разах), процентах, промилле и т.д. Необходимо уяснить назначение каждого вида относительных величин и метод их расчета.

Для оценки достигнутого уровня изучаемого показателя, при расчете нормативов, при анализе и планировании применяются средние величины. Средняя величина — обобщающий показатель, характеризующий типический уровень признака в расчете на единицу совокупности в конкретных условиях места и времени. В средних величинах проявляется статистическая закономерность. Средняя величина должна вычисляться по однородной совокупности. В зависимости от представления исходных данных бывают простые и взвешенные степенные средние величины, которые рассчитываются по соответствующим формулам. При выборе способа расчета (формулы) средней взвешенной величины необходимо прежде всего определить, какой признак является осредняемым, а какая величина выступает в качестве веса.

Порядок проведения занятия

Занятие проводится в традиционной форме, направлено на получение и закрепление навыков вычисления статистических величин. Студентам формулируются вопросы и задания, проводится устный опрос или решение студентами задачи. Ответы студентов и правильность решения задачи обсуждаются и оцениваются преподавателем

Варианты заданий

Вопросы для обсуждения

- а) Приведите примеры абсолютных статистических величин.
- б) Как связаны относительные величины динамика, планового задания и выполнения плана?

- с) Дайте определение вариационного ряда. Приведите пример ранжированного вариационного ряда.
- д) Дайте определение открытого и закрытого интервала. Каким способом можно закрыть открытый интервал?
- е) Что такое средняя величина? Где применяются средние величины?
- ф) Перечислите виды средних величин.
- г) Изменится ли средняя арифметическая величина, если все ее веса уменьшить в 100 раз?
- h) Каковы особенности определения моды и медианы в дискретном и интервальном вариационных рядах?
- i) Можно ли считать совокупность однородной, если коэффициент вариации равен 0,07?
- j) Как изменится значение среднего квадратического отклонения, если все значения признака увеличить в 2 раза?

Задание.

Требуется определить метод расчета статистической величины, выбрать формулу для расчета и провести вычисления.

Пример 1 выполнения задания

В таблице приведены данные социологического исследования - распределение посетителей театра по возрасту (графы 1 и 2).

Группы посетителей по возрасту, лет x_i	Число посетителей f_i	Середина интервала, лет x'_i	$x'_i f_i$	Накопленные частоты f'_i
1	2	3	4	5
10–15	6	12,5	75,0	6
15–20	9	17,5	157,5	15
20–25	20	22,5	450,0	35
25–30	41	27,5	1127,5	76
30–35	26	32,5	845,0	102
35–40	21	37,5	787,5	123
40–45	14	42,5	595,0	137
45–50	5	47,5	237,5	142
50–55	1	52,5	52,5	143
ИТОГО	143		4327,5	15

Задание:

- а) Вычислить средние величины.
- б) Определить, однородна ли совокупность по возрасту

Решение

а) Вычислим среднее значение признака «возраст» по формуле средней арифметической взвешенной. Для удобства вычислений дополним исходную таблицу столбцами 3-5, в которых будем помещать промежуточные расчеты.

Так как исходный ряд – интервальный, то при вычислении среднего значения используется серединное значение интервала, равное полусумме нижней и верхней границ интервала:

$$(10 + 15)/2 = 12,5; (15 + 20)/2 = 17,5 \text{ и т.д.}$$

Вычисленные значения середины интервала записываем в графу 3. В графе 4 запишем значения произведения $x'_i f_i$. Тогда итог графы 4 равен числителю в формуле среднего. Среднее значение возраста равно

$$\bar{X} = \frac{\sum x_i f_i}{\sum f_i} = \frac{4327,5}{143} = 30,3 \text{ лет.}$$

Медианным является среднее из 143 значений, т.е. семьдесят второе от начала ряда значение урожайности, оно находится в четвертом интервале.

$$Me = 25 + 5 \cdot \frac{72 - 35}{41} = 29,5 \text{ года;}$$

Если упорядоченная последовательность значений признака делится на 4 равных по численности части, то вычисляются квартили. Номер первого квартиля

$N_{Q1} = (143+1)/4=36$. По накопленным частотам определяем, что первый квартиль находится в четвертом интервале:

$$Q_1 = 25 + 5 \cdot \frac{35,75 - 35}{41} = 25,09 \text{ года;}$$

Номер третьего квартиля $N_{Q3} = 3 \cdot (143+1)/4=108$. Он находится в шестом интервале:

$$Q_3 = 35 + 5 \cdot \frac{107,25 - 102}{21} = 36,25 \text{ года}$$

Модальным является четвертый интервал, так как он имеет максимальную частоту. Точечную моду определяем по формуле.

$$M_0 = 25 + \frac{41 - 20}{(41 - 20) + (41 - 26)} = 27,9 \text{ года}$$

б) Показателем однородности совокупности может служить коэффициент вариации v : $v = \sigma : \bar{x}$.

Дисперсия вычисляется по формуле

$$D = \frac{\sum_{j=1}^k (x_j - \bar{x})^2 f_j}{\sum f_j} = \frac{10183,9}{143} = 71,2.$$

Среднее квадратическое отклонение равно квадратному корню из дисперсии, т.е.

$$\sigma = \sqrt{D} = 8,44 \text{ лет.}$$

Находим коэффициент вариации

$$v = \frac{\sigma}{\bar{x}} = \frac{8,44}{30,3} = 0,279 \text{ или } v = 27,9\%.$$

Так как коэффициент вариации менее 30%, то можно говорить об однородности опрошенной совокупности посетителей театра

Пример 2 выполнения задания.

Из 44 студентов специальности ГМУ сдали сессию на «отлично» 12 человек, а из 142 студентов специальности АСОИУ — 32 человека. Вычислить все возможные относительные величины.

Решение:

Относительная величина структуры:

$OBS_{\text{Стр}} = 12/44 = 0,273$ или 27,3%. Таким образом, доля отличников составляет 27,3%.

Относительная величина сравнения

$OBS_{\text{Ср}} = 142/44 = 3,23$, то есть численность студентов специальности АСОИУ в 3,23 раза больше численности студентов специальности ГМУ.

Варианты для самостоятельного выполнения.

а) По данным, представленным в таблице из примера 1, постройте график (столбиковую диаграмму) распределения посетителей по возрасту.

б) Относительная величина динамики показателя объем производства равна 1,03. Что это означает?

в) «В студенческой группе девушек было в 3 раза больше, чем парней». Какая относительная величина используется в этом предложении?

г) Зная значения относительных величин планового задания и выполнения плана, какую относительную величину можно определить?

д) В 2010 году магазин продал 5 тонн сахара по цене 20 рублей за кг. В 2011 году планировалось получить сумму на 20 тыс. рублей больше.

Для этого цена на сахар в 2011 году была увеличена на 10%. Фактически было продано 4,5 тонн сахара. Охарактеризуйте выполнение плана.

ф) На основании следующих данных по двум сельскохозяйственным кооперативам необходимо определить, в каком из них и насколько выше средняя урожайность зерновых культур.

Культура	СХК «Заря»		СХК «Луговое»	
	Валовый сбор, ц	Урожайность, ц/га	Посевная площадь, га	Урожайность, ц/га
Пшеница	32500	25	1540	20
Рожь	1620	18	120	19
Ячмень	13640	22	460	18
Просо	1650	15	80	13
ИТОГО	49410	—	2200	—

г) В таблице приведены данные о числе посещений кинотеатра работниками фирмы за год

Число посещений	2	3	4	6	7
Число работников	4	6	9	4	2

Определите среднее число посещений?

h) Расстояние между городами А и Б автобус проехал со скоростью 60 км/час. а при возвращении из Б в А автобус ехал со скоростью 70 км/час. Какова средняя скорость автобуса за весь рейс? Какой вид средней величины применяется?

i) Имеются данные о размерах заработной платы на предприятиях области.

Размер заработной платы, тыс.руб	Число предприятий
до 10	20
10-20	40
20-40	30
40 и выше	10

Определите среднее значение признака «заработная плата», моду и медиану.

3.3 Практическое занятие «Статистические группировки»

Цель занятия

Цель проведения практического занятия – изучение и закрепление теоретического материала; приобретение умения проводить группировку в соответствии с характеристикой состояния экономической, социальной, политической среды; получение навыков оценки построенной группировки и содержательной интерпретации полученной типологии объектов и связи описывающих объекты признаков.

Рекомендации по подготовке к занятию

При подготовке к занятию студенту следует повторить теоретический материал по конспекту лекций и следующим источникам:

Годин, А.М. Статистика [Электронный ресурс] : учебник / А.М. Годин. — Электрон. дан. — Москва : Дашков и К, 2017. — с.55-72, 143-174. — Режим доступа: <https://e.lanbook.com/book/93468>.

Лепихина З.П. Статистика: Учебное пособие/ З. П. Лепихина; Федеральное агентство по образованию, Томский государственный университет систем управления и радиоэлектроники. - Томск: ТУСУР, 2005. – с.109-124.

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

При изучении темы необходимо вспомнить стадии статистического исследования, а также виды вариационных рядов и способы их построения.

Вторым этапом является сводка и группировка данных, полученных на этапе статистического наблюдения.

Сводка – это комплекс последовательных операций по обобщению конкретных единичных фактов, образующих совокупность, для выявления типичных закономерностей, присущих изучаемому явлению.

Группировкой называется разбиение единиц изучаемой совокупности на однородные группы по определенным существенным признакам. При изучении темы важно уяснить, что группировки не являются самоцелью, а позволяют выявить социально-экономические типы явления, изучить структуру явлений и изменение структуры, выявить связи и зависимости между признаками. Различие задач определяет виды группировок: типологические, структурные, аналитические. Оценка качества группировок проводится на основе правила сложения дисперсий. Необходимо уметь объяснить связь между величиной коэффициента детерминации и силой связи факторного и результирующего признаков.

Следует также уяснить понятие статистической связи и проработать основы методов изучения статистической связи признаков.

Порядок проведения занятия

Занятие проводится в традиционной форме, направлено на получение и закрепление навыков построения и анализа статистических группировок. Студентам формулируются вопросы и задания, проводится устный опрос или решение студентами задачи. Ответы студентов и правильность решения задачи обсуждаются и оцениваются преподавателем

Варианты заданий

Вопросы для обсуждения

- а) Укажите различия между простой и сложной группировкой статистических данных.
- б) Сформулируйте принцип построения группировок.
- в) Приведите примеры простой и сложной группировок.
- д) Какой вид статистических таблиц можно применить для группировки по двум признакам?
- е) Сформулируйте назначение типологической группировки?
- ф) Какой вывод можно сделать, если квадратичный показатель структурных сдвигов равен нулю?
- г) Перечислите задачи аналитической группировки.
- h) Объясните, почему коэффициент детерминации принимает близкие к единице значения, если наблюдается сильная связь факторного и результативного признаков.
- и) Какие методы измерения связи вам известны?

Задание.

Требуется построить группировку, провести ее оценку и объяснить результаты вычислений.

Пример выполнения задания

С целью установления зависимости между урожайностью и сортом винограда в одном из хозяйств на основе выборки определили урожай на 10 кустах винограда.

Наименование сорта винограда	Число проверенных кустов	Урожай винограда с каждого куста, кг				
		куст №1	куст №2	куст №3	куст №4	куст №5
Сорт «А»	3	6	5	7	–	–
Сорт «Б»	5	7	6	8	5	9
Сорт «В»	2	9	7	–	–	–

Исчислите общую, межгрупповую и среднюю из групповых (частных) дисперсий. Определите связь между сортом и его урожайностью.

Решение. Если совокупность разбита на группы, то дисперсия признака σ^2 может быть определена как сумма межгрупповой дисперсии $\delta_{м.гр.}^2$ и средней из групповых дисперсий $\overline{\sigma_i^2}$:

$$\sigma^2 = \overline{\sigma_i^2} + \delta_{м.гр.}^2$$

$$\overline{\sigma_i^2} = \frac{\sum(\sigma_i^2 f_i)}{\sum f_i},$$

где $\sigma_i^2 = \frac{\sum(x - \bar{x}_i)^2}{f_i}$ — дисперсия признака в группе i (групповая дисперсия);

x — индивидуальное значение признака;

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

f_i — число наблюдений в группе i .

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

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

Отношение межгрупповой дисперсии к общей дает возможность измерить вариацию результативного признака за счет факторного, то есть признака, положенного в основание группировки, и тем самым судить о связи между изучаемыми признаками:

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

где R^2 — коэффициент детерминации.

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

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

Эмпирическое корреляционное отношение варьирует от 0 до 1.

При $\eta = 0$ связи нет, при $\eta = 1$ — связь полная.

Применим правило сложения дисперсии к исходным данным.

1. Групповые средние, т.е. средняя урожайность по каждому сорту винограда равны:

$$\bar{x}_A = \frac{\sum x}{n} = \frac{6+5+7}{3} = 6 \text{ кг};$$

$$\bar{x}_B = \frac{7+6+8+5+9}{5} = 7 \text{ кг};$$

$$\bar{x}_C = \frac{9+7}{2} = 8 \text{ кг}.$$

2. Определим среднюю урожайность винограда по хозяйству:

$$\bar{x} = \frac{\sum(x_i \cdot f_i)}{\sum f_i} = \frac{6 \cdot 3 + 7 \cdot 5 + 8 \cdot 2}{10} = 6,9 \text{ кг}.$$

3. Определим групповую (частную) дисперсию урожайности для каждого сорта отдельно:

$$\sigma_A^2 = \frac{(6-6)^2 + (5-6)^2 + (7-6)^2}{3} = 0,667;$$

$$\sigma_B^2 = \frac{(7-7)^2 + (6-7)^2 + (8-7)^2 + (5-7)^2 + (9-7)^2}{5} = 2,000;$$

$$\sigma_C^2 = \frac{(8-9)^2 + (8-7)^2}{2} = 1,000.$$

4. Средняя из частных дисперсий:

$$\overline{\sigma_i^2} = \frac{0,667 \cdot 3 + 2,000 \cdot 5 + 1,000 \cdot 2}{3+5+2} = \frac{14,001}{10} = 1,4.$$

5. Межгрупповая дисперсия:

$$\delta_{м.гр.}^2 = \frac{(6-6,9)^2 \cdot 3 + (7-6,9)^2 \cdot 5 + (8-6,9)^2 \cdot 2}{3+5+2} = \frac{4,9}{10} = 0,49.$$

6. Определяем общую дисперсию урожайности по всей совокупности, используя правило сложения дисперсий:

$$\sigma^2 = 1,40 + 0,49 = 1,89.$$

Проверим этот вывод путем расчета общей дисперсии обычным способом:

$$\sigma^2 = \frac{\sum (x - \bar{x})^2}{n} = \frac{(6 - 6,9)^2 + (5 - 6,9)^2 + (7 - 6,9)^2 + (7 - 6,9)^2 + (6 - 6,9)^2 + (8 - 6,9)^2 + (5 - 6,9)^2 + (9 - 6,9)^2 + (9 - 6,9)^2 + (7 - 6,9)^2}{10} = 1,89.$$

7. Определим коэффициент детерминации — R^2 :

$$R^2 = \frac{0,49}{1,89} = 0,26 \text{ или } 26\%.$$

Таким образом, только на 26% вариация урожайности обусловлена различиями между сортами, а на 74% — другими факторами (характером почвы, поливом и т.п.).

8. Определяем эмпирическое корреляционное отношение:

$$\eta = \sqrt{\frac{0,49}{1,89}} = \sqrt{0,26} \approx 0,5.$$

Следовательно, можно утверждать, что связь умеренная.

Варианты для самостоятельного выполнения

а) Имеются следующие данные о размере рекламного модуля в газете и его цене

Размер, кв.см	56	75	64	81	100	125	105	98	120	150	176	160
Цена, тыс.руб	1,7	2,3	1,9	2,4	3,0	3,6	3,1	2,9	3,4	4,0	5,0	4,2

Исследуйте связь между ценой и размером рекламного модуля с помощью аналитической группировки.

б) Заработная плата членов бригады А составляет (тыс.руб.): 5, 2, 6, 7, а в бригаде Б: 8, 4, 3. Для установления различия в уровнях заработной платы работников бригад определите коэффициент детерминации.

в) Определите значение коэффициента детерминации, если полная дисперсия признака равна 2,4, а средняя из групповых дисперсий равна 0,6. Дайте характеристику полученному результату.

d) Имеются данные об образовании лиц, стоящих на учете в службе занятости, в тыс.чел

Образование	Число безработных, тыс.чел	
	2016	2017
Высшее	10	20
Среднее	15	30

Вычислите линейный (абсолютный) показатель структурных сдвигов.

e) Имеются следующие данные о стаже работы рабочих.

Рабочий, № п\п	Стаж, число лет
1	1
2	8
3	5
4	12
5	6
6	6
7	9
8	5
9	4
10	1
11	2
12	2
13	4
14	8
15	15
16	16

Исследуйте структуру совокупности рабочих по стажу при помощи типологической группировки, выделив группы работников со стажем «до трех лет», «от 3 до 10 лет», «свыше 10 лет». Постройте диаграммы, характеризующие структуру.

f) Из 100 работников фирмы 20 мужчин и 10 женщин имеют стаж до 5 лет, стаж работы от 5 до 10 лет имеют 30 мужчин и 20 женщин, остальные 10 мужчин и 10 женщин работают в фирме 10 и более лет. Есть ли связь между полом и стажем? Обоснуйте выбор метода установления взаимосвязи и охарактеризуйте результаты вычислений.

3.4 Практическое занятие «Анализ ранговых корреляций»

Цель занятия

Закрепление теоретического материала и получение практических навыков расчета и анализа коэффициентов ранговой корреляции на примере оценок согласованности мнений экспертов.

Рекомендации по подготовке к занятию

При подготовке к занятию студенту следует повторить теоретический материал по конспекту лекций и следующим источникам:

Рекомендуемые источники

Прогнозирование и планирование в условиях рынка: учеб. пособие / Т.Н.Бабич, И.А.Козьева, Ю.В.Вертакова, Э.Н.Кузьбожев.— М.: ИНФРА-М, 2018. — с.203-212. — Режим доступа: <http://znanium.com/catalog/product/944382>.

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. – с.69-75.

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

В порядковой шкале результатом измерения является приписывание каждому объекту некоторой *условной числовой метки*, обозначающей место этого объекта в ряду из всех n анализируемых объектов, упорядоченном по возрастанию (убыванию) степени проявления в них k -го изучаемого свойства.

Под *ранговой корреляцией* понимается статистическая связь между порядковыми переменными.

Методы анализа ранговых корреляций часто используются в экспертных обследованиях для оценки согласованности мнений экспертов и построения интегральной (совокупной) оценки признака. Если наблюдается согласие мнений, то можно определять общие (суммарные или средние) оценки положения или уровня развития объектов.

При подготовке к практическому занятию студенту следует проработать теоретический материал по лекциям, просмотреть рекомендуемую литературу, прорешать задачи. Студент должен уметь обосновать выбор расчетных формул, провести анализ результатов расчетов.

Порядок проведения занятия

Занятие проводится в традиционной форме, направлено на получение и закрепление навыков выбора и техники расчетов показателей. Студентам формулируются вопросы и задания, проводится устный опрос или решение студентами задачи. Ответы студентов и правильность решения задачи обсуждаются и оцениваются преподавателем

Варианты заданий

Вопросы для обсуждения

- Какие измерительные шкалы вы знаете?
- В чем специфика порядковой шкалы?
- Какой одномерный анализ возможен в порядковой шкале?
- Приведите примеры порядковых переменных.
- Дайте определение ранга.
- Запишите формулы ранговых коэффициентов корреляции Спирмена и Кендалла?
- Какие значения может принимать коэффициент конкордации?

Задание.

Требуется определить согласованность мнений экспертов..

Пример выполнения задания

а) Два эксперта проранжировали 10 предприятий с точки зрения эффективности.

Эксперт 1	1	2	3	4	5	6	7	8	9	10
Эксперт 2	2	3	1	4	6	5	9	7	8	10

Коэффициент Спирмена вычисляется по формуле:

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

Подставляем числовые значения

$$\rho_{kj} = 1 - \frac{6}{1000 - 10} (1^2 + 1^2 + 2^2 + 0 + 1^2 + 1^2 + 2^2 + 1^2 + 1^2 + 0) = 0,915$$

Ранговый коэффициент корреляции Кендалла.

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

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

Для вычисления коэффициента Кендалла подсчитаем число обменов соседних элементов второй ранжировки, необходимое для ее приведения к упорядочению по возрастанию, так как первая ранжировка — возрастающая:

Запишем вторую ранжировку в исходном виде

2 3 1 4 6 5 9 7 8 10

Переставляем соседние элементы:

1) 2 1 3 4 6 5 9 7 8 10 (переставили 1 и 3)

2) 1 2 3 4 6 5 9 7 8 10 (переставили 1 и 2)

3) 1 2 3 4 5 6 9 7 8 10 (переставили 5 и 6)

4) 1 2 3 4 5 6 7 9 8 10 (переставили 7 и 9)

5) 1 2 3 4 5 6 7 8 9 10 (переставили 9 и 8)

Таким образом, за пять шагов упорядочили вторую ранжировку, то есть $v(x^{(2)}, x^{(1)}) = 5$.

Подставляем числовые значения в формулу коэффициента Кендалла:

$$\tau = 1 - \frac{4 \cdot 5}{10 \cdot 9} = 0,778.$$

Очевидно, для совпадающих ранжировок оба коэффициента $\rho, \tau = 1$, для противоположных $\rho, \tau = -1$.

В данном случае оба коэффициента указывают на согласованность мнений экспертов.

б) По трем порядковым переменным $x^{(1)}, x^{(2)}, x^{(3)}$ произведено упорядочение 10 объектов.

	$x^{(1)}$	$x^{(2)}$	$x^{(3)}$	Σ
1	1	2,5	2	5,5
2	4,5	1	1	6,5
3	2	2,5	4,5	9
4	4,5	4,5	4,5	13,5
5	3	4,5	4,5	12
6	7,5	8	4,5	20
7	6	9	8	23
8	9	6,5	8	23,5
9	7,5	10	8	25,5

10	10	6,5	10	26,5
----	----	-----	----	------

Так как число экспертов более двух и имеются объединенные ранги, то следует воспользоваться модифицированной формулой коэффициента конкордации:

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

где $T^{(j)}$ вычисляется по формуле

$$T^{(l)} = \frac{1}{12} \sum_{t=1}^{m^{(l)}} [(n_t^l)^3 - n_t^l]$$

где $m^{(l)}$ - число групп неразличимых рангов, $n_t^{(l)}$ - число элементов (рангов), входящих в t -ю группу неразличимых рангов.

Вычислим числитель:

$$S = \sum_{i=1}^{10} \left(\sum_{j=1}^3 x_i^{(j)} - \frac{3 \cdot 11}{2} \right)^2 = (-11)^2 + (-10)^2 + (-7,5)^2 + (-3)^2 + (-4,5)^2 + (3,5)^2 + (6,5)^2 + 7^2 + 9^2 + 10^2 = 591.$$

$$T^{(1)} = \frac{1}{12} (2^2 - 2) \cdot 2 = 1, \text{ т.к. } m^{(1)} = 2, n_1^{(1)} = n_2^{(1)} = 2;$$

$$T^{(2)} = \frac{1}{12} (2^3 - 2) \cdot 3 = 1,5, \text{ т.к. } m^{(2)} = 3, n_1^{(2)} = 2, n_2^{(2)} = n_3^{(2)} = 2;$$

$$T^{(3)} = \frac{1}{12} ((4^3 - 4) + (3^3 - 3)) = 7, \text{ т.к. } m^{(3)} = 2, n_1^{(3)} = 4, n_2^{(3)} = 3.$$

$$W(3) = \frac{591}{\frac{1}{12} 3^2 (10^2 - 10) - 3(1 + 1,5 + 7)} = 0,828.$$

Вычисляем статистику:

$$\chi_{\phi}^2 = m(n-1) \cdot W(3) = 3 \cdot 9 \cdot 0,828 = 22,36.$$

Оценим значимость полученного коэффициента. Показано, что при $n > 7$ величина $m(n-1) \cdot W(m)$ при отсутствии связи распределена приближенно по χ^2 распределению с $(n-1)$ степенями свободы. Если ока-

жется, что $m(n-1) \cdot W(m) > \chi_{\alpha(n-1)}^2$, то гипотеза об отсутствии связи должна быть отвергнута с уровнем α .

Теоретическое значение при $\alpha = 0,01$ и числе степеней свободы

$df = 9 \chi_m^2 = 21,67$. Так как $\chi_{\phi}^2 > \chi_m^2$, то гипотеза об отсутствии связи отвергается, т.е. согласованность признаков значима.

Варианты для самостоятельного выполнения

а) Три эксперта оценили инвестиционную привлекательность 5 регионов следующим образом:

	<i>Регион 1</i>	<i>Регион 2</i>	<i>Регион 3</i>	<i>Регион 4</i>	<i>Регион 5</i>
<i>Иванов</i>	1	2	5	4	3
<i>Перов</i>	1	3	4	5	2
<i>Сонин</i>	1	2	4	5	3

б) Два эксперта проранжировали 7 проектов программ развития туризма

	<i>Проект 1</i>	<i>Проект 2</i>	<i>Проект 3</i>	<i>Проект 4</i>	<i>Проект 5</i>	<i>Проект 6</i>	<i>Проект 7</i>
<i>Эксперт-1</i>	5	4	1	3	2	6	7
<i>Эксперт-2</i>	5	4	1	3	2	6	7

с) Два эксперта проранжировали 6 проектов программ развития туризма

<i>Эксперт 1</i>	<i>Эксперт 2</i>
2	1,5
1	1,5
3	3
4	4
6	5
5	6

д) Четыре эксперта проранжировали 3 бизнес-проекта

	<i>Эксперт 1</i>	<i>Эксперт 2</i>	<i>Эксперт 3</i>	<i>Эксперт 4</i>
<i>Проект 1</i>	1	1	3	3
<i>Проект 2</i>	2	2	2	2
<i>Проект 3</i>	3	3	1	1

е) Три эксперта проранжировали 5 проектов прогноза по достоверности

	<i>Эксперт 1</i>	<i>Эксперт 2</i>	<i>Эксперт 3</i>
<i>Проект 1</i>	1	1	2
<i>Проект 2</i>	2	2	2
<i>Проект 3</i>	3	3	2
<i>Проект 4</i>	4,5	5	5
<i>Проект 5</i>	4,5	4	4

f) Два эксперта проранжировали 4 бизнес-проекта

	<i>Проект 1</i>	<i>Проект 2</i>	<i>Проект 3</i>	<i>Проект 4</i>
<i>Эксперт 1</i>	1	2	3	4
<i>Эксперт 2</i>	4	3	2	1

4 Методические указания для организации самостоятельной работы

4.1 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 «Статистика» включает следующие виды его активности:

1. проработка лекционного материала;
2. изучение тем теоретической части дисциплины, вынесенных для самостоятельной проработки;
3. выполнение домашнего задания;
4. подготовка к контрольным работам;
5. подготовка к практическим занятиям;
6. подготовка к лабораторным работам;
7. выполнение индивидуального задания;
8. подготовка к экзамену.

4.2 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 рекомендованной литературы следует начинать с основных рекомендованных преподавателем глав и разделов учебников и учебных пособий, а затем переходить к нормативно-правовым актам, 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 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.

4.3 Самостоятельное изучение тем теоретической части курса

4.3.1 Тема: Основы выборочного метода

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 всю 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

Годин, А.М. Статистика [Электронный ресурс] : учебник / А.М. Годин. — Электрон. дан. — Москва : Дашков и К, 2017. — с.125-142. — Режим доступа: <https://e.lanbook.com/book/93468>.

Лепихина З.П. Статистика: Учебное пособие/ З. П. Лепихина; Федеральное агентство по образованию, Томский государственный универси-

тет систем управления и радиоэлектроники. - Томск: ТУСУР, 2005. – с.87-108.

Лацкевич, Н.В. Статистика [Электронный ресурс] : учебное пособие / Н.В. Лацкевич, С.А. Дещеня, Т.Н. Бессонова. — Электрон. дан. — Минск : "Вышэйшая школа", 2015. — с.75-87. — Режим доступа: <https://e.lanbook.com/book/75126>.

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

4.3.2 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 до-

кументации, но общие принципы построения таблиц и графиков сохраняются.

Рекомендуемые источники

Годин, А.М. Статистика [Электронный ресурс] : учебник / А.М. Годин. — Электрон. дан. — Москва : Дашков и К, 2017. — с. 72-77. — Режим доступа: <https://e.lanbook.com/book/93468>.

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 «Рекомендуемые источники».

4.3.3 Тема: Факторный анализ как метод снижения размерности

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:

- 1) отыскание скрытых, но объективно существующих закономерностей, которые определяются воздействием внутренних и внешних причин на изучаемый процесс;

2) сжатие информации путем описания процесса при помощи общих факторов, число которых значительно меньше количества первоначально взятых признаков;

3) выявление и изучение статистической связи признаков с факторами;

4) 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 ото-*

бражением, а ее коэффициенты - факторными нагрузками. $P = (p_{ij})$ - 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

Методы и средства комплексного анализа данных [Электронный ресурс] / Кулаичев А.П., 4-е изд., перераб. и доп. - М.: НИЦ ИНФРА-М, 2016. - с.315-349 с. — Режим доступа: <http://znanium.com/catalog/product/548836>

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

4.4 Домашнее задание

4.4.1 Тема: Организация государственной статистики в Российской Федерации

Цель домашнего задания

Изучение системы государственной статистики в РФ, формирование навыков самостоятельного поиска официальной статистической информации.

Порядок выполнения и содержание работ

При рассмотрении статистики как отрасли практической деятельности (статистический учет) по сбору, обработке, анализу и публикации массовых цифровых данных о самых различных явлениях и процессах

общественной жизни следует отметить, что эту деятельность на профессиональном уровне осуществляет государственная статистика.

Федеральная служба государственной статистики (Росстат) является федеральным органом исполнительной власти, осуществляющим функции по формированию официальной статистической информации о социальном, экономическом, демографическом и экологическом положении страны, а также функции по контролю и надзору в области государственной статистической деятельности на территории Российской Федерации.

При выполнении домашнего задания студенту необходимо ознакомиться с деятельностью Росстата, его структурой управления, организацией системы сбора информации в отраслевом и территориальном разрезах, перечнем публикуемых периодических журналов, сборников и документов, организацией Интернет-портала.

В результате изучения темы студент должен написать отчет объемом 4-5 страниц и уметь ответить на вопросы преподавателя и выполнить задание по поиску заданной преподавателем информации на сайте Росстата:

1. Когда была организована Федеральная служба государственной статистики?
2. В чьем ведении она находится в настоящее время?
3. Перечислите основные задачи.
4. Какие функции выполняет?
5. Как организована служба в региональном разрезе?
6. Приведите пример публикации (название сборника или документа) статистической информации.

Рекомендуется также ознакомиться с историей развития статистики в России и организацией статистики в зарубежных странах.

Основным информационным ресурсом при выполнении домашнего задания является официальный сайт Росстата. Студенту важно отобрать из большого объема материала необходимую информацию для развернутых и обоснованных ответов на вопросы.

Рекомендуемые источники

Федеральная служба государственной статистики: Официальный сайт [Электронный ресурс] — Режим доступа: <https://gks.ru>

Годин, А.М. Статистика [Электронный ресурс] : учебник / А.М. Годин. — Электрон. дан. — Москва : Дашков и К, 2017. — с.11-33. — Режим доступа: <https://e.lanbook.com/book/93468>. Лепихина З.П. Статистика: Учебное пособие/ З. П. Лепихина; Федеральное агентство по образова-

нию, Томский государственный университет систем управления и радиоэлектроники. - Томск: ТУСУР, 2005. – с.13-20.

4.4.2 Тема: Выборочный метод: расчет объема и ошибок выборки

Цель домашнего задания

Закрепление теоретического материала и получение практических навыков выбора метода и расчет ошибок и объема выборочной совокупности для выборок различных типов.

Порядок выполнения и содержание работ

При изучении темы «Основы выборочного метода» важными практическими вопросами являются определение способа отбора единиц совокупности, вычисление ошибок выборки и построение доверительных интервалов выборочных характеристик, расчет необходимого объема выборки.

Для выполнения домашнего задания студенту необходимо проработать теоретический материал (см. п.4.3.1 настоящих Методических указаний) и самостоятельно дома решить и представить преподавателю решение трех задач. Студент должен указать вид выборки, уметь обосновать выбор расчетных формул, провести анализ результатов расчетов. Правильность решения задач и защита результатов решения обсуждается и оценивается преподавателем.

Решение типовой задачи и примеры задач

Задача. Для изучения скорости автомобилей при подъеме на гору методом случайной повторной выборки была зафиксирована скорость двухсот автомобилей. В результате установлена средняя скорость 30 км/час при среднем квадратическом отклонении 4 км/час. С вероятностью 0,997 определите пределы, в которых находится средняя скорость движения в генеральной совокупности.

Решение:

По условию задачи известны:

$n = 200$ автомобилей – объем выборочной совокупности;

$S = 4$ км/час – среднее квадратическое отклонение

$t = 3$ – коэффициент кратности средней ошибки выборки (при $p = 0,997$);

$\bar{x} = 30$ км/час – средняя скорость.

Рассчитаем предельную ошибку *случайной повторной* выборки:

$$\Delta_{\bar{x}} = t \cdot \sqrt{\frac{S^2}{n}} = 3 \cdot \sqrt{\frac{4^2}{200}} = 0,86$$

Определим пределы генеральной средней:

$$\tilde{x} - \Delta_{\bar{x}} \leq \bar{x} \leq \tilde{x} + \Delta_{\bar{x}};$$

$$30 - 0,86 \leq \bar{x} \leq 30 + 0,86;$$

$$29,14 \leq \bar{x} \leq 30,86.$$

Следовательно, с вероятностью 0,997 можно утверждать, что средняя скорость движения автомобилей при подъеме в гору в генеральной совокупности находится в пределах от 29,14 км/час до 30,86 км/час.

Задача 1. Торговая фирма купила у производителей 480 коробок с калькуляторами, в каждой коробке упаковано по 24 калькулятора. Для проверки качества необходимо сформировать выборку. Определите объем выборки, если результат требуется гарантировать с вероятностью 0,954 и ошибкой не более 5%, а межгрупповая дисперсия равна 51.

Задача 2. По результатам контрольной проверки налоговыми службами 400 бизнес-структур, у 140 из них в налоговых декларациях не полностью указаны доходы, подлежащие налогообложению. Определите в генеральной совокупности (по всему району) долю бизнес-структур, скрывших часть доходов от уплаты налогов, с вероятностью 0,954.

Задача 3. Выборочный хронометраж работы 2% рабочих, изготавливающих одинаковые детали, показал, что по затратам времени на изготовление одной детали рабочие распределились следующим образом:

Затраты времени на изготовление 1 детали (мин)	20-24	24-28	26-32	32-36	Итого
Число изготовленных деталей	6	18	22	4	50

Определите средние затраты времени на изготовление одной детали в выборке и доверительный интервал этой средней с вероятностью 0,997 ($t=3$).

Задача 4. В 100 туристических фирмах города предполагается провести обследование среднемесячного количества реализованных путевок. По данным пробного исследования дисперсия признака равна 225.

Сколько фирм необходимо обследовать для того, чтобы с вероятностью 0,683 ошибка выборки не превышала 3 путежки?

Задача 5. Для определения среднего возраста 1200 студентов факультета необходимо провести выборочное обследование методом случайного бесповторного отбора. Предварительно установлено, что среднее квадратическое отклонение возраста студентов равно 10 годам. Сколько студентов нужно обследовать, чтобы с вероятностью 0,954 средняя ошибка выборки не превышала 3 года?

Рекомендуемые источники

Годин, А.М. Статистика [Электронный ресурс] : учебник / А.М. Годин. — Электрон. дан. — Москва : Дашков и К, 2017. — с.125-142. — Режим доступа: <https://e.lanbook.com/book/93468>.

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.

Лацкевич, Н.В. Статистика [Электронный ресурс] : учебное пособие / Н.В. Лацкевич, С.А. Дещеня, Т.Н. Бессонова. — Электрон. дан. — Минск : "Вышэйшая школа", 2015. — с.75-87. — Режим доступа: <https://e.lanbook.com/book/75126>.

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

4.5 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.

4.6 Подготовка к практическим занятиям

Практические занятия позволяют выработать навыки применения теоретических знаний для решения задач, зафиксировать в памяти студента определенные типовые действия в конкретной ситуации, запомнить формулы и алгоритмы расчетов.

Содержание практических занятий приведено в разделе 3.4 настоящих указаний. При подготовке к практическим занятиям 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.

Рекомендуется прорешать типовые задачи по теме занятия, приведенные в разделе 3.4 указаний для самостоятельного выполнения. Если возникают неясности или проблемы, то необходимо сформулировать вопросы и задать их преподавателю на консультации или в начале занятия.

4.7 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 2 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.

4.8 Индивидуальное задание «Инструментальные средства статистического анализа данных»

Цель индивидуального задания

Изучение и анализ рынка программных продуктов по статистическому анализу данных, включая методы обработки многомерных данных

Исходные данные к работе

Рынок программных продуктов по статистическому анализу данных. Вид исследования — «кабинетное». Студенту необходимо сделать обзор программных средств анализа данных на основе информации из различных источников, в том числе в глобальных компьютерных сетях.

Порядок выполнения и содержание работы

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 02-2013. Перед началом работы над рефератом следует наметить план и подобрать литературу. Прежде всего, следует пользоваться литературой, рекомендованной учебной программой, а затем расширить список источников, включая и использование специальных журналов и Интернет.

Рекомендуется следующая структура отчета:

- Титульный лист.
- Оглавление.
- Введение (дается постановка вопроса, объясняется значимость и актуальность темы, указываются цель и задачи работы, даётся характеристика используемой литературы).

- Основная часть (может состоять из разделов, которые раскрывают отдельную проблему или одну из её сторон и логически являются продолжением друг друга).
- Заключение (подводятся итоги и даются обобщённые основные выводы по теме работы, делаются рекомендации).
- Список использованных источников. В списке должно быть не менее 6-8 различных источников.

Приветствуется включение таблиц, графиков, схем, как в основном тексте, так и в качестве приложений.

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.

При оценке работы учитываются: глубина проработки материала; правильность и полнота использования источников; владение терминологией и культурой речи; оформление отчета.

Рекомендуемые источники

Компьютерные технологии анализа данных в эконометрике [Электронный ресурс]/ Д.М. Дайитбегов. - 2-е изд., испр. и доп. - М.: Вузский учебник: ИНФРА-М, 2010. — с511-530: — Режим доступа: <http://znanium.com/catalog/product/251791>

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

5 Рекомендуемые источники

- 1) Регионы России. Социально-экономические показатели. 2017: [Электронный ресурс] : Р32 Стат. сб. / Росстат. – М., 2017. – 1402 с. — Режим доступа: http://www.gks.ru/wps/wcm/connect/rosstat_main/rosstat/ru/statistics/publications/catalog/doc_1138623506156
- 2) Годин, А.М. Статистика [Электронный ресурс] : учебник / А.М. Годин. — Электрон. дан. — Москва : Дашков и К, 2017. — 412 с. — Режим доступа: <https://e.lanbook.com/book/93468>.
- 3) 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.
- 4) 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>.
- 5) Балдин, К.В. Общая теория статистики [Электронный ресурс] : учебное пособие / К.В. Балдин, А.В. Рукоусев. — Электрон. дан. — Москва : Дашков и К, 2017. — 312 с. — Режим доступа: <https://e.lanbook.com/book/93403>.
- 6) 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>
- 7) 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>
- 8) Компьютерные технологии анализа данных в эконометрике [Электронный ресурс]/ Д.М. Дайитбегов. - 2-е изд., испр. и доп. - М.: Вузовский учебник: ИНФРА-М, 2010. – с511-530: — Режим доступа: <http://znanium.com/catalog/product/251791>
- 9) 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 1

Таблица Возраст посетителей клуба

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
20	20	20	20	20
18	18	18	18	18
20	20	20	20	22
20	20	18	20	18
23	23	20	23	22
20	20	20	20	32
18	18	23	18	18
22	22	20	22	20
16	17	18	23	23
22	22	22	30	20

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
20	20	20	20	20
18	18	18	18	18
20	20	20	20	22
20	20	18	20	18
23	23	20	23	22
20	20	20	20	25
18	18	23	18	18
22	22	20	22	20
25	17	18	23	23
22	22	22	25	20

ПРИЛОЖЕНИЕ 2

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 3

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

ПРИЛОЖЕНИЕ 4

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	Среднедушевые 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 5

**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.	Среднемесячная 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.	Среднемесячная 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

ПРИЛОЖЕНИЕ 6

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

Число степеней свободы <i>df</i>	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

ПРИЛОЖЕНИЕ 7

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			

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

Продолжение таблицы 2

№ анкеты	№ вопроса										
	1	2	3.	4.	5	6.	7.	8	9	10	11
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