Министерство образования и науки Российской Федерации

Федеральное государственное бюджетное образовательное учреждение
высшего образования
**«ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
СИСТЕМ УПРАВЛЕНИЯ И РАДИОЭЛЕКТРОНИКИ» (ТУСУР)**

Кафедра автоматизации обработки информации (АОИ)

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Методические указания к лабораторным работам
и организации самостоятельной работы
для студентов направления «Программная инженерия»
(уровень бакалавриата)

Томск 2018

Смыслова Зинаида Александровна
Теория вероятностей и математическая статистика: Методические указания к лабораторным работам и организации самостоятельной работы для студентов направления «Программная инженерия» (уровень бакалавриата) / З.А.Смыслова. – Томск, 2018. – 54 с.

Оглавление

1 Введение

Программой изучения дисциплины предусмотрено проведение лабораторных работ, самостоятельная работа по изучению теоретической части и подготовке к лабораторным работам, самостоятельная подготовка к экзамену.

Лабораторные работы предназначены

- для отработки навыков применения изученных моделей и методов при решении практических (ситуационных) задач;
- для закрепления умения пользоваться расчетными формулами, теоремами, таблицами при решении статистических задач;
- для формирования способности и готовности применять статистические методы при обработке результатов измерений;
- для получения опыта использования компьютерной техники при представлении, анализе и интерпретации статистических данных.

Задание студенту (команде студентов) формулируется в терминах некоторой предметной области. Первый этап работы состоит в формализации задачи, выборе метода решения, установлении последовательности шагов решения. Второй этап лабораторной работы состоит в выборе программного инструментария для решения задачи и реализации задачи на выбранном или рекомендованном преподавателем программном средстве. Третий этап работы заключается в анализе полученных результатов, оформлении отчета и защите лабораторной работы.

Отчет по лабораторной работе должен содержать

- титульный лист;
- вариант задания;
- расчетные формулы и результаты расчетов;
- графический и справочный материал;
- выводы по работе.

При оценке лабораторной работы учитывается содержание отчета, правильность применения расчетных формул, а также качество выполнения и оформления, своевременность сдачи и умение студента обосновывать и защищать сделанные выводы.

Целями самостоятельной работы является систематизация, закрепление и расширение теоретических знаний, приобретение навыков практической и исследовательской деятельности. Реализация этих целей достигается в процессе теоретической подготовки, подготовки к лабораторным и контрольным работам, подготовки к экзамену.

2 Методические указания к лабораторным работам

2.1 Лабораторная работа «Описательная статистика»

Цель работы

Знакомство с методами описательной статистики, получение навыков первоначальной обработки данных средствами табличного процессора *LibreOffice*.

Форма проведения

Выполнение индивидуального задания. Варианты заданий выдаются преподавателем каждому студенту в виде файла для табличного процессора.

Форма отчетности

Защита отчета.

Задание на лабораторную работу

Для выборочных данных своего варианта выполнить следующую обработку, пояснив полученные результаты:

1) по выборке найти среднее арифметическое, выборочную дисперсию, выборочное среднее квадратичное отклонение, проиллюстрировать выполнение правила «3σ»;

2) по вариационному ряду найти моду, медиану, размах выборки, оценить среднее квадратичное отклонение с помощью размаха;

3) найти верхнюю и нижнюю выборочные квартили, пояснить их смысл; оценить симметричность распределения с помощью первого коэффициента Пирсона (если необходимо, использовать значение моды, найденное по сгруппированному ряду);

4) составить сгруппированный статистический ряд, оценить математическое ожидание и дисперсию по сгруппированному ряду, сравнить эти значения с найденными по выборке;

5) построить гистограмму; найти модальный интервал и сравнить его середину с найденным по вариационному ряду значением моды;

6) составить сгруппированный статистический ряд с накопленными частотами, построить эмпирическую функцию распределения; найти медианный интервал, сравнить середину интервала с найденным по вариационному ряду значением медианы.

Теоретические основы

Перед выполнением лабораторной работы повторите теоретический материал о выборочном методе математической статистики ([2], глава 1; [3], глава 12).

Замечания к пунктам 1 – 6 задания на лабораторную работу.

К пункту 1. Среднее арифметическое обозначается $\bar{x}$ и вычисляется по формуле $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$; выборочная дисперсия $\hat{D} = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$. Корень квадратный из выборочной дисперсии называется выборочным средним квадратическим отклонением $\hat{\sigma}$. Правило «трех сигма» выполняется для большинства унимодальных законов распределения, т.е. для выборок из таких генеральных совокупностей: более 99 % выборочных значений лежат в интервале $\left(\bar{x} - 3\hat{\sigma}; \bar{x} + 3\hat{\sigma}\right)$. Появление выборочных значений за пределами интервала радиуса 3σ может свидетельствовать об аномальности этих значений ([2], глава 3, п. 3.4), либо о том, что генеральная совокупность не является нормальной.

К пунктам 2, 3. Оценка моды $\hat{M}$ - варианта с наибольшей частотой. Медиана делит выборку на две части: половина вариант меньше медианы, половина — больше. Можно найти три числа q_1, q_2, q_3, которые аналогичным образом делят выборку на четыре равные части. Эти числа называются квартилями. Число q_2 совпадает с медианой, q_1 называется нижней, а q_3 — верхней квартилью. Размах $R = x_{(n)} - x_{(1)}$ равен разности максимальной и минимальной вариант. Этой характеристикой пользуются при работе с малыми выборками. Первый коэффициент асимметрии Пирсона: $A_1 = \frac{\bar{x} - \hat{M}}{\hat{\sigma}}$.

К пункту 4. Сведения о синтаксисе функции ЧАСТОТА. Имя и параметры: ЧАСТОТА(*массив_данных*; *массив_карманов*).

Массив_данных - это выборочные данные, для которых рассчитываются частоты попадания в «карманы». Если массив_данных не содержит значений, то функция ЧАСТОТА возвращает нули.

Массив_карманов – это массив правых концов интервалов, в которых *группируются значения массива данных.*

Функция ЧАСТОТА возвращает распределение частот в виде вертикального массива, причем количество элементов в возвращаемом массиве на единицу больше числа элементов в массиве карманов. Дополнительный элемент в возвращаемом массиве содержит количество значений, больших, чем правая граница последнего интервала. Для вызова функции сначала выделяют область, куда попадут результаты вычисления, задают значения аргументов, затем выходят нажатием сочетания клавиш *Ctrl+Shift+Enter*.

К пункту 5. Для сгруппированной выборки находят модальный интервал, т.е. интервал с наибольшей частотой. В качестве моды можно взять середину модального интервала.

К пункту 6. Для нахождения медианы по сгруппированному ряду накопленных частот выделяют промежуток, на котором накопленная частота становится больше или равной 0,5 (а левее этого промежутка – меньше 0,5). За медиану принимают середину найденного интервала.

Порядок выполнения работы

Прочитать задание на лабораторную работу, вспомнить понятия математической статистики из предыдущего семестра. Записать основные расчетные формулы; составить предварительный отчет.

Для пунктов задания 1), 2), 3) найти выборочные характеристики, используя встроенные статистические функции: СРЗНАЧ; МАКС; МИН; МЕДИАНА; КВАРТИЛЬ и пр.

Для пункта задания 5) использовать встроенную функцию ЧАСТОТА; построить гистограмму (в меню: Мастер диаграмм).

Сравнить результаты оценок числовых характеристик по вариационному ряду и по сгруппированной выборке, сделать выводы; закончить оформление отчета. Защитить работу.

Варианты заданий

Варианты заданий формируются преподавателем для каждого студента индивидуально на основе данных федеральной службы государственной статистики (www.gks.ru/) и выдаются студентам в виде файла или ссылки на соответствующую страницу сайта.

Контрольные вопросы

1) Что означают термины «генеральная совокупность» и «выборка»?

2) В чем суть выборочного метода?
3) Какая оценка называется несмещенной
4) Какая оценка называется состоятельной?.
5) Приведите пример несмещенной и состоятельной оценки.
6) Приведите пример смещенной и состоятельной оценки.
7) Поясните правило правило «3σ».
8) Что такое аномальные наблюдения?
9) Что такое «вариационный ряд»?
10) Какие оценки удобно находить по вариационному ряду?
11) Какие параметры указываются при обращении к функции КВАРТИЛЬ?
12) Как строится сгруппированный статистический ряд?
13) Поясните, как использовать встроенную функцию ЧАСТОТА.
14) Как называется оценка плотности распределения, построенная по сгруппированному статистическому ряду?
15) Как называется оценка функции распределения, построенная по сгруппированному статистическому ряду?
16) Поясните, как найти медиану по сгруппированному статистическому ряду.

2.2 Лабораторная работа «Проверка статистических гипотез»

Цель работы

Изучение процедуры проверки параметрических гипотез, проверка гипотез об однородности данных.

Форма проведения

Работа в парах. Решение ситуационных задач.

Форма отчетности

Защита отчета. Отчет оформляется в виде файла *MathCad*.

Теоретические основы

При подготовке к лабораторной работе повторите тему «Проверка статистических гипотез» по конспекту лекций или по литературе ([2], глава 3, п. 3.4; [3], глава 17, параграфы 1, 2; [4], глава 19, параграфы 1 – 6). Справочный материал для проверки параметрических гипотез приведен в приложении А1, более подробно эта информация изложена в [4], глава 19, пп. 8 – 13, 18, 19.

В качестве дополнительного задания студентам предлагается задача на проверку гипотезы об однородности двух выборок, которую рекомендуется решать, используя критерий Вилкоксона ([4], глава 19, параграф 27; [3], глава 17, параграф 6).

Порядок выполнения работы

1) Прочитайте условия задач, выясните, какую гипотезу требуется проверить.
2) Сформулируйте основную и альтернативную гипотезы, задайте уровень значимости.
3) С помощью приложения А1 подберите статистику и нарисуйте примерный вид плотности распределения.
4) Найдите границы критической области, используя встроенные функции пакета *MathCad.*
5) Рассчитайте наблюдаемое значение статистики и примите решение.
6) Оформите отчет и защитите его перед преподавателем.

Варианты заданий

Примерные варианты заданий (задача 1) приводятся в приложении А2. Для задачи 2 (дополнительной) генерируются две выборки заданного объема из нормальной генеральной совокупности с заданными параметрами.

Контрольные вопросы

1) Как обозначаются основная и альтернативная гипотезы?
2) Дайте определение ошибки первого рода.
3) Дайте определение ошибки второго рода.
4) Какую роль при проверке параметрических гипотез играет уровень значимости?
5) Какую роль при проверке параметрических гипотез играет альтернативная гипотеза?
6) Какую задачу решают критерии согласия? Назовите известные Вам критерии согласия.
7) Какую задачу решают критерии однородности? Назовите известные Вам критерии однородности.
8) Опишите критерий согласия Пирсона (постановка задачи, условия применения, принятие решения)
9) Опишите критерий знаков (постановка задачи, условия применения, принятие решения)

10) Опишите критерий Вилкоксона (постановка задачи, условия применения, принятие решения)

2.3 Лабораторная работа «Метод Монте-Карло»

Цель работы

Изучение метода Монте-Карло (ММК) на примере вычисления определенного интеграла.

Форма проведения

Выполнение индивидуального задания средствами пакета *MathCad*.

Форма отчетности

Защита отчета. Приложение к отчету - файл *MathCad*..

Теоретические основы

При подготовке к лабораторной работе следует повторить сведения о равномерном и нормальном законах распределения случайной величины ([1], глава 2, параграф 30), а также формулировку закона больших чисел (ЗБЧ) и центральной предельной теоремы (ЦПТ) ([1], глава 4, параграфы 41, 42). С общей идеей метода Монте-Карло и алгоритмами моделирования случайных величин можно познакомиться по конспекту лекций или по литературе ([3], глава 2, параграфы 5, 6; [4], глава 21, параграфы 1-3).

Порядок выполнения работы

Лабораторная работа состоит из двух частей:

Часть 1 – Генерация случайных чисел указанным методом и исследование качества полученной последовательности (2 часа).

Часть 2 – Вычисление интеграла простейшим методом с заданной точностью и заданной доверительной вероятностью (2 часа).

Порядок выполнения *Части 1*:

1) Выбрав произвольно начальные значения, сгенерировать n=100 случайных чисел (нечетные варианты – методом Неймана, четные – методом Лемера).

2) Изобразить полученные числа в виде точечного графика на плоскости и оценить визуально качество последовательности.

3) Оценить качество последовательности с помощью критерия Пирсона, используя статистику

$$K = \sum_{i=1}^{k} \frac{\left(n_i^* - np_i\right)^2}{np_i} \quad \approx \quad \chi^2_{(k-r-1)}$$

4) Выбрать другие начальные значения и повторить п.1)-3) два-три раза.

5) Результаты работы оформить в виде таблицы:

Таблица 2.1 - Качество генерации

№ опыта	Начальные значения	Визуальная. оценка	Кнабл	Ктабл	Вывод
1 2 3					

Порядок выполнения *Части 2*:

1) Привести интеграл к виду, требуемому для вычисления с помощью ММК.

$$I = \int_a^b h(x) \cdot f(x)\, dx\,,$$

где $f(x)$ - плотность распределения с.в. $X \sim R(a,b)$.

2) Оценить количество опытов, необходимое для вычисления интеграла с заданной точностью и заданной доверительной вероятностью

3) Используя стандартный датчик (*rnd*), сгенерировать *n* случайных чисел $\gamma_i \in (0;1)$ и преобразовать их в $x_i = a + (b-a) \cdot \gamma_i$

4) Вычислить оценку $\hat{I}$ для интеграла I :

$$\hat{I} = \frac{1}{n} \sum_{i=1}^{n} h(x_i)$$

5) Вычислить оценку дисперсии

$$\hat{D} = \frac{1}{n} \sum_{i=1}^{n} \left(h^2(x_i)\right) - \hat{I}^2$$

6) Используя оценку дисперсии и зная количество опытов, применить определение доверительного интервала и центральную предельную теорему:

$$P\{|\hat{I} - I| < \varepsilon\} = 2\Phi\left(\frac{\varepsilon\sqrt{\mathrm{n}}}{\sigma}\right) = \beta,$$

затем найти точность расчета ε из уравнения

$$2\Phi\left(\frac{\varepsilon\sqrt{\mathrm{n}}}{\sqrt{\hat{D}}}\right) = \beta$$

и сравнить с заданной точностью ε_0.

7) Провести вычисления несколько раз, добиваясь требуемой точности или минимизируя значение *n*.

8) Оформить отчет, оформив результаты вычислений в таблицу.

Таблица 2.2 – Вычисление интеграла

№ опыта	Оценка интеграла	Оценка дисперсии	Оценка точности	Вывод
1 2 3				

Ответ запмсать в виде доверительного интервала.

Варианты заданий

Приводятся варианты заданий (Приложение А3).

Контрольные вопросы

1) Какие задачи решаются методом Монте-Карло?
2) Перечислите способы получения случайных чисел.
3) Укажите достоинства и недостатки генерации случайных чисел с помощью таблиц.
4) Укажите достоинства и недостатки генерации случайных чисел с помощью физического датчика.
5) В чем заключается метод середин квадратов?
6) В чем заключается метод вычетов?
7) На каких теоремах основан метод Монте-Карло?
8) Расшифруйте сокращение ЗБЧ и сформулируйте соответствующую теорему.
9) Расшифруйте сокращение ЦПТ и сформулируйте соответствующую теорему.
10) Что такое вероятная ошибка метода Монте-Карло?

11) Как точность вычислений зависит от числовых характеристик моделируемой СВ?

12) За счет чего можно повысить точность вычислений?

13) Сравните по точности простейший и геометрический методы Монте-Карло.

14) Что такое трудоемкость метода Монте-Карло?

15) Какую случайную величину будем генерировать для вычисления интеграла I = $\int_0^{\pi/2} \cos x \, dx$ простейшим методом Монте-Карло? Приведите этот интеграл к виду, необходимому для применения метода Монте-Карло.

16) Оцените количество опытов, необходимых для вычисления интеграла I с точностью ε=0,04 и доверительной вероятностью γ=0,95.

17) Изложите идею простейшего метода для этого интеграла. Поясните на рисунке.

18) Изложите идею геометрического метода для этого интеграла. Поясните на рисунке.

19) Вычисление интеграла $\int_0^{\pi/2} \cos x \, dx$ производится методом Монте-Карло на основании 1000 испытаний. Какую максимальную погрешность вычислений можно гарантировать с надежностью 97,22%?.

2.4 Лабораторная работа «Дисперсионный анализ»

Цель работы

Изучение методов однофакторного дисперсионного анализа.

Форма проведения

Выполнение индивидуального задания средствами табличного процессора.

Форма отчетности

Защита отчета.

Теоретические основы

Постановка задачи однофакторного дисперсионного анализа как частного случая статистической гипотезы приведена в [3], глава 17, параграф 8. Более подробное изложение можно найти в [4], глава 20.

В **классической** схеме однофакторного дисперсионного анализа на некоторый количественный признак *X* (случайную величину *X*) действует фактор F, имеющий r уровней. Дисперсионный анализ позволяет ответить на вопрос: влияет ли фактор F на измеряемый признак *X?* На каждом уровне проводится ряд наблюдений исследуемого признака, которые рассматриваются как независимые выборочные значения из генеральных совокупностей $X_1, X_2, ..., X_r$, распределенных по нормальному закону с одинаковыми, хотя и неизвестными дисперсиями, и математическими ожиданиями $m_1, m_2, ..., m_r$. При этом предполагается, что ошибки наблюдений распределены нормально с нулевым математическим ожиданием и одинаковыми дисперсиями. Задача дисперсионного анализа формулируется как задача о равенстве всех математических ожиданий

$$\begin{cases} H_o: \ m_1 = m_2 = .. m_r \\ H_1: \ m_i \neq m_j \text{для некоторых} i, j \end{cases}$$

Основная идея дисперсионного анализа состоит в переходе от задачи сравнения средних на всех уровнях к эквивалентной ей задаче сравнения дисперсий: «факторной» дисперсии, оценивающей разброс, вносимый в результате воздействия фактора F, и «остаточной» дисперсии, оценивающей разброс, возникший в результате случайных причин. Для этого сумма квадратов отклонений наблюдаемых значений от их общего среднего по всей таблице $\bar{x}$ раскладывается на две части (см. [4]) в виде основного дисперсионного тождества:

$$\sum_{i=1}^{r} \sum_{j=1}^{k_i} \left(x_{ij} - \bar{x}\right)^2 = \sum_{i=1}^{r} k_i \left(\bar{x}_i - \bar{x}\right)^2 + \sum_{i=1}^{r} \sum_{j=1}^{k_i} \left(x_{ij} - \bar{x}_i\right)^2$$

В этом тождестве левая часть обозначается $Q^2_{общ}$ и служит для оценки общего разброса в наблюдаемых данных, первое слагаемое в правой части Q^2_F - для оценки «факторного» разброса, второе слагаемое правой части Q^2_ε - для оценки «случайного» разброса в данных.

С помощью этих сумм квадратов находятся оценки факторной D_F и остаточной дисперсий D_ε, для чего соответствующая сумма квадратов делится на число степеней свободы: $\hat{D}_F = \dfrac{Q^2_F}{r-1}$, $\hat{D}_\varepsilon = \dfrac{Q^2_\varepsilon}{n-r}$, где n – общее количество наблюдений. Задача о равенстве средних на всех уровнях эквивалентна задаче о равенстве дисперсий:

$$\begin{cases} H_O: \ D_F = D_\varepsilon \\ H_1: D_F > D_\varepsilon \end{cases}.$$

Здесь основная гипотеза утверждает, что факторная дисперсия отличается от остаточной незначимо, т.е. разброс, вносимый фактором, практически не отличается от разброса, обусловленного случайными причинами, следовательно, и средние значения на разных уровнях отличаются незначимо; поэтому в исходной постановке справедлива гипотеза H_O. Если же верной является альтернативная гипотеза, то влияние фактора значимо отличается от случайного, и в исходной постановке справедлива гипотеза H_1.

Для сравнения дисперсий рассматривается статистика $K = \dfrac{\hat{D}_F}{\hat{D}_\varepsilon}$, имеющая распределение Фишера $F_{(r-1,n-r)}$. Критическая область является правосторонней, поэтому если наблюдаемое значение статистики *Кнабл* превышает или равно табличному *Ктабл*, то основная гипотеза отвергается в пользу альтернативной, т.е. влияние фактора на результирующий признак значимо.

Классическая схема однофакторного дисперсионного анализа требует нормальности распределения исследуемого признака и ошибок наблюдений. Если эти условия не выполняются, то для решения задачи о равенстве средних более предпочтительными являются непараметрические методы, которые позволяют проверить гипотезу о равенстве средних с минимальными требованиями к выборочным данным: предполагается, что ошибки наблюдений независимы и имеют непрерывное распределение.

При проверке гипотезы H_O с помощью **критерия Краскела-Уоллеса** выборочные значения x_{ij} заменяются их рангами x'_{ij}. Напомним, что ранг – это число, соответствующее порядковому номеру наблюдаемого значения в данной выборке, если наблюдаемые значения расположить по возрастанию. Для каждого уровня вычисляется средний ранг $R_i = \frac{1}{k_i} \sum_{j=1}^{k_i} x'_{ij}, \quad i=\overline{1,r},$ и сравнивается с общим средним рангом, который, в предположении справедливости гипотезы H_O, равен $R = \frac{n+1}{2}$. Сравнение проводится по статистике Краскела-Уоллеса $K = \frac{12}{n(n+1)} \sum_{i=1}^{r} k_i \left(R_i - R\right)^2$, критические точки которой зависят от количества наблюдений *n*, количества уровней *r* и количества наблюдений k_i на каждом уровне. Критическая область в данной задаче является правосторонней, поэтому если $K_{набл} \geq K_{табл}$, то гипотеза H_O отвергается. При больших значениях *n* распределение статистики Краскела-Уоллеса приближается к распределению χ^2_{r-1}.

Порядок выполнения работы

1) Получить у преподавателя таблицу наблюдений (строки – уровни фактора).
2) Изучить теоретическую часть [4, 3]. Ответить на вопросы.
3) Решить задачу по классической схеме.
4) Решить эту же задачу методом Краскела-Уоллеса.
5) Проверить результат с помощью встроенного модуля «Статистика» табличного процессора. Проанализировать результаты.
6) Написать отчет и защитить его перед преподавателем. Для отчета придумать содержательную постановку задачи дисперсионного анализа, указав, что является измеряемой величиной, а что – фактором, и какие значения принимают уровни фактора.

Варианты заданий

Приводятся варианты заданий (Приложение А4).

Контрольные вопросы

1) Какой вид имеет таблица наблюдений дисперсионного анализа (ДА)?
2) Какие требования предъявляются к экспериментальным данным в классической схеме ДА?
3) Сформулируйте основную и альтернативную гипотезы ДА (о средних).
4) Сформулируйте гипотезу, эквивалентную основной гипотезе ДА (о дисперсиях).
5) В чем основная идея ДА?
6) Как оценивается межгрупповой разброс данных?
7) Как оценивается внутригрупповой разброс данных?
8) Запишите основное дисперсионное тождество.
9) Какая статистика используется при проверке гипотезы о дисперсиях?
10) В каких случаях применяется непараметрический анализ (какие требования предъявляются к экспериментальным данным)?
11) В чем состоит метод Краскела-Уоллиса?
12) Каким распределением аппроксимируется статистика Краскела-Уоллиса при большом объеме выборки?

2.5 Лабораторная работа «Корреляционный анализ»

Цель работы

Знакомство с числовыми коэффициентами, предназначенными для выявления связи между двумя переменными; оценка силы корреляционной связи.

Форма проведения

Выполнение индивидуального задания средствами *MathCad*.

Форма отчетности

Защита отчета.

Теоретические основы

Перед выполнением лабораторной работы рекомендуется повторить определение и свойства коэффициента корреляции Пирсона ([1], глава 3, параграф 36; [4], глава 14, параграф 17). Полезно сравнить свойства коэффициента корреляции Пирсона и рангового коэффициента корреляции Спирмена ([3], глава 17, параграф 10; [4], глава 19, параграф 25). Ознакомьтесь с формой представления сгруппированных двумерных наблюдений (корреляционные таблицы), с выборочным корреляционным отношением и его свойствами ([4], глава 18, параграфы 5, 11 – 13).

Расчетные формулы для несгруппированных данных

Коэффициент корреляции Спирмена:

$$\hat{r}_s = 1 - \frac{6\sum d_i^2}{n^3 - n}$$

где d_i – разности рангов двух рядов наблюдений.

Статистика для проверки значимости коэффициента корреляции Спирмена:

$$\frac{\hat{r}_s\sqrt{n-2}}{\sqrt{1-\hat{\ }_s^{2}}} \sim T_{n-2}$$

Расчетные формулы для сгруппированных данных

Коэффициент корреляции Пирсона:

$$\hat{r}_{xy} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{k} n_{ij}\, x_i y_j - n\cdot\bar{x}\cdot\bar{y}}{\sqrt{\left(\sum_{i=1}^{m}\left(n(X = x_i)\cdot x_i^2\right) - n\cdot\bar{x}^2\right)\left(\sum_{j=1}^{k}\left(n(Y = y_j)\cdot y_j^2\right) - n\cdot\bar{y}^2\right)}}$$

Статистика для проверки значимости коэффициента корреляции Пирсона:

$$\frac{\hat{r}_{xy}\sqrt{n-2}}{\sqrt{1-\hat{r}_{xy}^2}} \approx T_{n-2}$$

Расчет корреляционного отношения:

$$\hat{\eta}^2 = \frac{\sum_{i=1}^{m} n(X = x_i) \cdot \left(\bar{y}(X = x_i) - \bar{y}\right)^2}{\sum_{j=1}^{k} n(Y = y_j) \cdot \left(y_j - \bar{y}\right)^2}$$

Статистика для проверки гипотезы о линейности связи:

$$\frac{\left(\hat{\eta}^2 - r_{xy}^2\right)(n-m)}{\left(1-\hat{\eta}^2\right)(m-2)} \approx F_{(m-2,\, n-m)}$$

Порядок выполнения работы

Лабораторная работа состоит из двух частей:

Часть 1 – Измерение корреляции несгруппированных данных.

Часть 2 – Исследование корреляции сгруппированных данных.

Порядок выполнения *Части 1*:

1) Получить набор несгруппированных данных (фактор – отклик).
2) Рассчитать коэффициент корреляции Спирмена и проверить его значимость.
3) Измерить связь между переменными с помощью коэффициента корреляции Пирсона (встроенная функция corr)и сравнить значения коэффициентов.

Порядок выполнения *Части 2*:

4) Получить корреляционную таблицу.
5) Найти оценку коэффициента корреляции Пирсона.
6) Найти выборочное корреляционное отношение.
7) Проверить гипотезу о линейности корреляционной связи.
8) Проанализировать результаты.
9) Написать отчет и защитить его перед преподавателем.

Варианты заданий

Приводятся варианты заданий для несгруппированных данных (Приложение А5) и для сгруппированных данных (Приложение А6).

Контрольные вопросы

1) Какие значения может принимать коэффициент корреляции Спирмена?

2) При каких условиях применяется коэффициент корреляции Спирмена?
3) Перечислите свойства коэффициента корреляции Спирмена.
4) При каких условиях применяется коэффициент корреляции Пирсона?
5) Как зависит сила линейной связи между переменными от величины коэффициента корреляции?
6) Перечислите свойства коэффициента корреляции Пирсона.
7) Как проверить значимость коэффициента корреляции?
8) Что такое корреляционное отношение?
9) Какими свойствами оно обладает?
10) Как проверяется линейность связи с помощью сравнения коэффициента корреляции и корреляционного отношения?

2.6 Лабораторная работа «Регрессионный анализ»

Цель работы

Получение навыка построения и оценки качества регрессионных моделей.

Форма проведения

Выполнение индивидуального задания.

Форма отчетности

Защита отчета.

Теоретические основы

Теоретический материал приведен в [3], глава 18, параграфы 1 - 4; [2], глава 4, параграф 25). При подготовке к лабораторной работе необходимо уяснить разницу между функциональной и стохастической зависимостью ([4], глава 18, параграф 1) и познакомиться с основным методом построения регрессионных моделей – методом наименьших квадратов.

Предположим, что нам необходимо описать в виде некоторой функции взаимосвязь двух переменных X и Y (X — фактор, независимая переменная: Y — отклик, зависимая переменная): $Y = f(X)$. По результатам наблюдений мы можем оценить эту зависимость приближенно (в силу воздействия неучтенных факторов, случайных причин, ошибок измерения): $y = f(x) + \varepsilon$, где ε — случайная переменная, называемая возмущением. Предполагается, что среднее значение возмущения равно нулю: $M(\varepsilon) = 0$. При этом для каждого значения $X = x$ мы имеем случайную переменную Y со средним значением (математическим ожиданием) $f(x)$. Функция $f(x)$ называется ***функцией регрессии*** случайной переменной Y на X, а график этой функции — ***линией регрессии***. Уравнение регрессии позволяет определить, каким в среднем будет значение отклика Y при том или ином значении фактора X.

Согласно методу наименьших квадратов сумма квадратов отклонений экспериментальных значений отклика $y_i, i = 1, 2, \ldots, n$ от модельных $f(x_i; b_0, b_1, \ldots, b_k)$ должна быть минимальной:

$$Q(b_0, \ldots, b_k) = \sum_{i=1}^{n} \left(y_i - f(x_i; b_0, \ldots, b_k) \right)^2 \to \min.$$

Этот критерий позволяет получить оценки коэффициентов уравнения парной (простой) регрессии.

В случае, когда одна переменная (отклик) является функцией нескольких переменных (факторов) $y = \varphi\left(x_1, x_2, \ldots x_k\right)$ применяется многомерный регрессионный анализ. Как и в случае парной регрессии, основным методом определения коэффициентов модели является метод наименьших квадратов. Рассмотрим матричную форму записи этого метода в случае линейной формы связи отклика и факторов.

Пусть уравнение связи имеет вид $y = b_0 + b_1 x_1 + \ldots + b_k x_k$, и проведено n наблюдений отклика y и факторов $x_1, x_2, \ldots x_k$. Результаты наблюдений представим в виде матриц

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \ldots \\ y_n \end{bmatrix} \quad \text{и} \quad X = \begin{bmatrix} x_{10} \ x_{11} \ldots x_{1k} \\ x_{20} \ x_{21} \ldots x_{2k} \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ x_{n0} \ x_{n1} \ldots x_{nk} \end{bmatrix}.$$

Первый столбец в матрице X занимает фиктивная переменная $x_o \equiv 1$, она введена для удобства записи уравнения регрессии. Столбец коэффициентов регрессии обозначим $b = \begin{bmatrix} b_0 \\ b_1 \\ \ldots \\ b_k \end{bmatrix}$. Тогда в матричной форме уравнение примет вид

$Y = Xb + \varepsilon$, где $\varepsilon^T = \left(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n\right)^T$ - вектор невязок, характеризующий отклонение экспериментальных данных Y от модельных $y = Xb$.

Метод наименьших квадратов требует минимизировать отклонение экспериментальных данных от модели, т.е.

$$Q(b) = \varepsilon^T \cdot \varepsilon = (y - Y)^T (y - Y) \to \min$$

$$Q(b) = (Y - Xb)^T (Y - Xb) = Y^T Y - b^T X^T Y - Y^T Xb + b^T X^T Xb$$

Пользуясь правилами векторно-матричного дифференцирования:

1) $\nabla_b \left(b^T A\right) = A$,

2) $\nabla_b (Ab) = A^T$,

3) $\nabla_b \left(b^T Ab\right) = Ab + A^T b$,

найдем производную функции *Q(b)* и приравняем ее нулю:

$$\frac{\partial Q}{\partial b} = -X^T Y - X^T Y + X^T Xb + X^T Xb = 0 \quad => \quad X^T Xb = X^T Y \quad =>$$

$$\tilde{b} = \left(X^T X\right)^{-1} X^T Y$$

Эти формулы позволяют найти оценки коэффициентов множественной линейной регрессии в матричной форме. Для построения доверительных интервалов и проверки адекватности модели нам потребуется остаточная дисперсия:

$$s_\varepsilon^2 = \frac{Q(\tilde{b})}{n - (k+1)}$$

Обозначим $C = \left(X^T X\right)^{-1}$, тогда доверительные интервалы для коэффициентов можно найти по формулам

$$b_i = \tilde{b}_i \pm t_{\left(\frac{\alpha}{2}; n - (k+1)\right)} s_\varepsilon \sqrt{c_{ii}} \quad i = 0, 1, \ldots, k$$

Для проверки адекватности модели используется статистика $K = \dfrac{S_y^2}{S_\varepsilon^2}$, имеющая распределение Фишера $F(n-1, n-k-1)$.

Порядок выполнения работы

1) Получить у преподавателя набор экспериментальных данных.
2) Изучить теоретическую часть [1 - 3]. Ответить на вопросы.
3) Найти коэффициенты регрессионной модели матричным способом, оценить их точность.
4) Проверить адекватность модели
5) Проанализировать результаты
6) Написать отчет и защитить его перед преподавателем.

Варианты заданий

Приводятся варианты заданий для построения множественной линейной регрессии (Приложение А7).

Контрольные вопросы

1) Какие задачи решает регрессионный анализ (РА)?
2) Как называются независимые переменные в РА? Зависимые?
3) Что описывает вектор невязок?
4) Как в матричной форме записать уравнение модели?
5) Для чего вводится фиктивная переменная x_0?
6) Изложите идею метода наименьших квадратов.
7) Запишите систему нормальных уравнений.
8) Сформулируйте правила векторного дифференцирования.
9) Какие требования предъявляются к экспериментальным данным?
10) Как проверяется адекватность модели?

2.7 Лабораторная работа «Временные ряды»

Цель работы

Знакомство с методами исследования модели временного ряда.

Форма проведения

Выполнение индивидуального задания.

Форма отчетности

Защита отчета.

Теоретические основы

Для подготовки к лабораторной работе рекомендуется [6]. Построение регрессионной модели проводится с помощью модуля «Статистика» табличного процессора. Полученные остатки исследуются с помощью трех методов.

Критерий **Гольдфельда-Квандта** позволяет ответить на вопрос о наличии гетероскедастичности остатков. Гетероскедастичность означает, что остатки модели регрессии не являются постоянными, вследствие чего доверительные интервалы будут ненадежными. Проверка осуществляется в предположении о нормальном распределении остатков регрессии и пропорциональности их дисперсий значениям фактора. Последовательность остатков упорядочивается по возрастанию фактора и выделяются две группы остатков: третья часть в начале последовательности и третья часть – в конце. По этим группам рассчитываются суммы квадратов остатков: $S_1 = \sum_{i=1}^{k} \varepsilon_i^2$ и $S_3 = \sum_{i=n-k+1}^{n} \varepsilon_i^2$, где n – общее количество наблюдений, $k \approx n/3$.

Основная гипотеза $H_0 : S_1 = S_3$ означает отсутствие, альтернатива $H_1 : S_3 > S_1$ - наличие гетероскедастичности (рост дисперсий остатков с ростом значений фактора). Для проверки гипотезы рассматривается статистика, имеющая распределение Фишера $F_{(k\text{-}1,\, k\text{-}1)}$. Если $F_{факт} = \dfrac{S_3}{S_1} > F_{табл}$ (где $F_{табл} = F_{\alpha,k-1,k-1}$, α – выбранный уровень значимости), то гипотеза H_0 об отсутствии гетероскедастичности отклоняется.

Критерий **серий** применяется для исследования вопроса о случайности остатков временного ряда. В ранговом критерии “восходящих” и “нисходящих” серий формируется последовательность знаков "+" или "-", на основе которой и принимается решение.

Пусть для временного ряда x_t $\left(t=\overline{0,n}\right)$ построена модель и найдены остатки ε_t $\left(t=\overline{0,n}\right)$.Опишем последовательность шагов критерия.

Шаг 1. Формулируем основную и альтернативную гипотезы:

$$\begin{cases} H_0 : остатки\ \varepsilon_t\ случайны \\ H_1 : остатки\ \varepsilon_t\ не\ являются\ случайными \end{cases}$$

Шаг 2. Задаем уровень значимости α.

Шаг 3. Формируем последовательность знаков.

Для этого рассматриваем разности $\varepsilon_t - \varepsilon_{t-1}$ $\left(t=\overline{1,n}\right)$ и ставим "+", если разность положительна, ставим "-", если она отрицательна. Если при некотором *t* оказалось, что $\varepsilon_t = \varepsilon_{t-1}$, то не ставим никакой знак.

Шаг 4. В качестве статистики рассматривается пара $\left(S, K_{\max}\right)$, где S – количество серий в последовательности знаков (серия – набор идущих подряд одинаковых знаков), а $K_{\max}$ – количество знаков в самой длинной серии. Наличие длинных серий в последовательности знаков – довод против гипотезы H_0.

Шаг 5. Принятие решения.

Если одновременно выполняются условия $\begin{cases} S > S_{гр}(n) \\ K_{\max} < K_{гр}(n) \end{cases}$, то гипотеза H_0 может быть принята с вероятностью ошибки первого рода *α*. Если хотя бы одно из неравенств нарушено, гипотеза H_0 отвергается, т.е. остатки ε_t $\left(t=\overline{0,n}\right)$ нельзя считать статистически независимыми.

Границы критической области:

$$S_{гр}(n) = \frac{2n-1}{3} - 1{,}96 \cdot \sqrt{\frac{16n-29}{90}} \qquad \text{при } \alpha=0{,}05.$$

$$K_{гр}(n) = \begin{cases} 5, при \;\; n \le 26, \\ 6, при \;\; 26 < n \le 153, \\ 7, при \;\; 153 < n \le 1170. \end{cases}$$

Критерий **Дарбина-Уотсона** отвечает на вопрос о наличии автокорреляции в остатках. Наблюдаемое значение статистики Дарбина-Уотсона рассчитывается по формуле $DW = \dfrac{\sum(\varepsilon_i - \varepsilon_{i-1})^2}{\sum \varepsilon_i^2}$ и сравнивается с табличными границами d_L и d_U для заданного числа наблюдений n. Если DW ∈ $[0; \;\; d_L)$, то автокорреляция положительна; если DW ∈$(4-d_L; \;\; 4]$, то автокорреляция отрицательна, DW ∈ $(d_U; \; 4-d_U)$ - автокорреляция отсутствует. В остальных случаях вопрос об автокорреляции остатков остается открытым.

Порядок выполнения работы

В соответствии с номером варианта скопируйте в табличный процессор данные временного ряда и выполните следующие задания.

1) Обратитесь к модулю «Статистика» табличного процессора, подберите с его помощью наилучшую модель, прокомментируйте лист отчета, выведите остатки временного ряда.
2) Проанализируйте остатки временного ряда графически, исследуйте их на наличие гетероскедастичности с помощью теста Гольдфельда-Квандта.
3) Проанализируйте остатки временного ряда по критерию серий
4) Проанализируйте остатки временного ряда с помощью статистики Дарбина-Уотсона. Сделайте вывод.
5) Напишите отчет.

Варианты заданий

Приводятся варианты заданий для лабораторной работы (Приложение А8).

Контрольные вопросы

1) Как называется основной метод построения регрессионной модели? Опишите его суть.
2) Какие требования предъявляются к экспериментальным данным в методе наименьших квадратов?
3) Каков смысл требования $M(X^T\varepsilon)=0$?
4) Каков смысл требования $D(\varepsilon_i)=\sigma_\varepsilon^2$?
5) Что означает условие $M(\varepsilon\varepsilon^T)=0$?
6) В чем суть гетероскедастичности?
7) Каковы последствия гетероскедастичности остатков?
8) Какие методы применяются для выявления гетероскедастичности?
9) Опишите схему теста Голдфелда-Кванта.
10) Для чего применяется взвешенный МНК и в чем его суть?
11) Каковы последствия автокорреляции?
12) Опишите алгоритм выявления автокорреляции остатков с использованием критерия Дарбина-Уотсона.
13) Для чего применяется критерий серий. Опишите его суть.

2.8 Лабораторная работа «Цепи Маркова»

Цель работы

Закрепление теоретических знаний и получение навыка исследования с помощью марковских цепей.

Форма проведения

Выполнение индивидуального задания. Решение ситуационных задач

Форма отчетности

Защита отчета.

Теоретические основы

Теоретический материал приведен в [7]. В электронном курсе опубликованы теоретические сведения о цепях Маркова, разобраны примеры решения задач, перечислены вопросы для подготовки, приведены варианты заданий для лабораторной работы.

Порядок выполнения работы

Задача 1. Опишите ситуацию первой задачи с помощью цепи Маркова (объясните, почему можно использовать такую модель). Нарисуйте орграф, охарактеризуйте состояния цепи Маркова (ЦМ). Последовательно возводя матрицу перехода в 5, 10, 15.. степень, исследуйте поведение цепи при неограниченном увеличении времени наблюдения. Если предельные вероятности существуют, найдите их.

Задача 2. По данной матрице перехода поглощающей цепи Маркова нарисуйте орграф, занумеруйте состояния, запишите матрицу перехода в каноническом виде. Найдите фундаментальную матрицу, укажите, каким состояниям соответствуют ее строки и столбцы. Сделайте выводы о среднем времени нахождения ЦМ в каждом из непоглощающих состояний. Матричным способом найдите финальные вероятности поглощения в каждом из поглощающих состояний.

Варианты заданий

Приводятся варианты заданий для лабораторной работы (Приложение А9).

Контрольные вопросы

1) Что такое цепь Маркова (ЦМ)?

2) В чем состоит свойство стохастичности матрицы перехода ЦМ?
3) Что такое поглощающее состояние ЦМ?
4) Как найти распределение вероятностей через m шагов от начала наблюдения?
5) Сформулируйте теорему о вероятности перехода в эргодическое состояние.
6) Что такое поглощающая ЦМ?
7) Как получить канонический вид матрицы перехода?
8) Что показывают элементы подматриц Q и R
9) Сформулируйте теорему о свойствах подматрицы Q.
10) Сформулируйте теорему об элементах фундаментальной матрицы.
11) Сформулируйте теорему о вероятности поглощения в данном поглощающем состоянии (что показывают элементы матрицы B)?
12) Что такое эргодическая ЦМ?
13) Как найти предельные вероятности?

3 Методические указания к самостоятельной работе

3.1 Теоретическая подготовка

Самостоятельная работа над теоретическим материалом направлена на освоение основных понятий и методов теории вероятностей и математической статистики. Теоретическая подготовка включает в себя не только проработку лекционного материала, но и самостоятельное изучение тем (вопросов) теоретической части дисциплины.

Проработка конспекта лекций играет важную роль при выстраивании структуры дисциплины и выделения важных аспектов изучаемого материала. Конспект оказывает помощь студенту при подготовке к лабораторным работам и лекционным занятиям; в нем должны быть выделены основные положения, определения и формулы. Контроль изучения теоретического материала проводится на каждой лекции в виде теста (теоретического опроса).

Примерные темы опросов

1) Вероятностные пространства.
2) Алгебра событий.
3) Свойства вероятностей.
4) Дискретная случайная величина.
5) Непрерывная случайная величина.
6) Важные законы распределения случайных величин.
7) Числовые характеристики и их свойства.
8) Системы случайных величин.
9) Независимость и некоррелированность случайных величин.
10) Свойства коэффициента корреляции.
11) Функция регрессии и ее свойства.
12) Предельные теоремы теории вероятностей.
13) Оценки параметров и их свойства.
14) Статистические гипотезы.
15) Генерация псевдослучайных чисел.
16) Вычисление интегралов методом Монте-Карло.
17) Однофакторный дисперсионный анализ.
18) Корреляционный анализ.
19) Регрессионный анализ.
20) Временные ряды.
21) Эргодические цепи Маркова.

22) Поглощающие цепи Маркова.
23) Случайные процессы.
24) Стационарность и эргодичность.

Некоторые темы изучаемой дисциплины рассматриваются на лекциях в обзорном порядке, затем выносятся для более детальной проработки на самостоятельное изучение (самоподготовку).

Примерные темы для самоподготовки

1) Случайные события: независимость событий в совокупности.
2) Случайные величины: распределение Пуассона как предельное для биномиального.
3) Способы повышения точности метода Монте-Карло.
4) Сравнение критериев согласия по мощности.
5) Корреляционный анализ: сравнение числовых характеристик.
6) Случайные процессы: гармонический анализ.

Изучая темы, вынесенные для самостоятельной проработки, необходимо работать с карандашом и бумагой, составляя план прочитанного материала и решая задачи и упражнения, предложенные в учебнике. При работе с разными источниками важно привести в систему обозначения и термины, которые могут отличаться в разных учебниках. Если при изучении материала проводится сравнительный анализ условий применения или свойств характеристик, результаты такого сравнения желательно представит в компактной форме (в виде таблицы или диаграммы). Контроль самостоятельного изучения материала осуществляется в форме проверки конспекта, собеседования с преподавателем или доклада на занятии.

Подготовка доклада начинается с выбора темы и обсуждения содержания доклада с преподавателем. Затем в течение недели студент осуществляет самостоятельный поиск в интернете или работает с литературой, рекомендованной преподавателем. Через неделю студент обсуждает с преподавателем план доклада и продолжает подготовку текста доклада и презентации. Оценка и рецензирование доклада проводится студентами и преподавателем совместно в устной или письменной форме.

Примерные темы докладов

1) Парадоксы теории вероятностей.
2) Применение экспоненциального распределения.
3) Двумерное нормальное распределение.
4) Графическое представление информации в статистике.

5) Исследование свойств оценок распределения.
6) Ранговые методы статистики.

Теоретическая подготовка включает в себя подготовку к коллоквиуму, который проводится в конце года обучения и рассматривает вопросы, вызывающие затруднения у студентов. Тема коллоквиума «Регрессионный анализ для временных рядов и случайные процессы общего вида». Теоретический материал можно повторить по [4, 5, 6]. Вопросы для подготовки к коллоквиуму делятся на две части: по теме «Временные ряды» (ВР) и по теме «Случайные процессы» (СП). В зависимости от имеющегося времени коллоквиум может быть проведен по двум темам одновременно или разбит на две части.

Вопросы по теме «Временные ряды»

1) Какие компоненты рассматриваются при построении модели ВР?
2) Что такое тренд ВР?
3) В чем разница между циклической и сезонной компонентами ВР?
4) Какими особенностями должны обладать критерии, применяющиеся при проверке гипотез о компонентах модели ВР?
5) Какую гипотезу проверяют с помощью критерия Фостера-Стюарта? Опишите суть критерия.
6) Какую гипотезу проверяют с помощью критерия поворотных точек? Опишите суть критерия.
7) Какую задачу решают с помощью критерия серий? Опишите суть критерия.
8) Какую задачу решают с помощью статистики Дарбина-Уотсона? В чем недостаток этого метода?
9) Какие методы используются при построении модели ВР?
10) Поясните термин «гетероскедастичность». Какие методы используются для выявления гетероскедастичности?
11) Для чего используется метод инструментальных переменных?
12) Для чего используется взвешенный метод наименьших квадратов?
13) Для чего используется авторегрессионное преобразование?
14) Какие методы применяются в задаче прогнозирования ВР?

Вопросы по теме «Случайные процессы»

1) Как описать случайный процесс общего вида?
2) Что такое сечение СП? Реализация СП? Какие законы распределения необходимо задать для описания СП?
3) Какие основные характеристики рассматриваются для СП общего вида?

4) Какими свойствами обладают математическое ожидание и корреляционная функция случайного процесса общего вида?

5) Что такое стационарные (в узком и в широком смысле) СП?

6) Каковы свойства корреляционной функции стационарного случайного процесса (ССП)?

7) Что такое нормированная корреляционная функция ССП и какими свойствами она обладает?

8) Что такое интервал корреляции ССП? Как его найти?

9) Что такое эргодические СП? Сформулируйте достаточные условия эргодичности.

10) Спектральное представление ССП. Что показывает спектральная плотность?

11) Взаимосвязь спектральной плотности и корреляционной функции (уравнения Винера-Хинчина).

12) Комплексная форма записи спектрального разложения, уравнения Винера-Хинчина в комплексной форме.

13) Что такое эффективная ширина спектра ССП?

14) Для чего используется частота Найквиста? Сформулируйте теорему Котельникова.

3.2 Подготовка к лабораторным работам

Лабораторные работы позволяют получить навыки

- представления и обработки статистических данных;
- освоения алгоритма проверки статистических гипотез;
- анализа зависимостей в группах статистических данных;
- построения и анализа вероятностных и статистических моделей.

Для подготовки к лабораторной работе необходимо изучить теоретический материал по теме работы, проработать основные понятия, ответить на контрольные вопросы, составить предварительный отчет, выписав туда основные формулы для расчетов по теме лабораторной работы.

По результатам выполнения лабораторной работы оформляется отчет, который защищается перед преподавателем, группой студентов или студентом-экспертом. Отчет о лабораторной должен содержать:

- титульный лист, оформленный по стандарту ОС ТУСУР;
- номер варианта;
- условия всех задач;
- решение каждой задачи с необходимыми пояснениями и формулами.

При защите отчета студент должен знать используемые термины, уметь формулировать определения и теоремы, давать пояснения к решению.

3.3 Подготовка к контрольным работам

Контрольные работы по дисциплине «Теория вероятностей и математическая статистика» проводятся в течение всего учебного года и имеют сквозную нумерацию. Контрольные работы 1 – 3 проводятся в третьем семестре и описаны в [8]. Для промежуточного контроля умений и навыков студентов в четвертом семестре проводится четыре контрольных работы.

Контрольная работа 4 «Основы математической статистики»

Контрольная работа 4 проводится после выполнения двух лабораторных работ по темам «Основы математической статистики» и «Проверка статистических гипотез». После выполнения этих работ студент должен знать основные понятия математической статистики, уметь пользоваться оценками параметров, знать их свойства, уметь формулировать статистические гипотезы и выбирать подходящие статистики для проверки гипотез.

Вариант контрольной работы содержит три задачи. Задача 1 проверяет умение студента строить критическую область и пользоваться соответствующими таблицами распределения. Задача 2 проверяет умение студента сформулировать математическую постановку данной ситуации и применять алгоритм проверки параметрической гипотезы. При решении задач 1 и 2 рекомендуется использовать справочный материал (приложение А). Задача 3 требует применения критерия согласия (Пирсона) или критерия однородности (Вилкоксона). Пример варианта контрольной работы приведен ниже.

ВАРИАНТ 1

Задача 1. Нарисовать критическую область для проверки гипотезы $H_0 : m_1 = m_2$ при альтернативе $H_1 : m_1 \neq m_2$. Найти ее границы для уровня значимости $\alpha = 0{,}02$ и числе степеней свободы $\nu = 5$. Предполагается нормальное распределение генеральной совокупности.

Задача 2. Точность наладки станка-автомата, металлические стержни, характеризуется дисперсией длины стержня. Если эта величина будет больше *400 мкм*^{2}, станок останавливается для наладки. Выборочная дисперсия длины *15* случайно отобранных стержнейй из продукции станка оказалась равной $S^2 = 680$ *мкм*^{2}. Нужно ли производить наладку станка, если уровень значимости $\alpha = 0{,}01$?

Задача 3. Метод получения случайных чисел был применен 250 раз, при этом получены следующие результаты (i – цифра, n_i – частота ее появления)

i	0	1	2	3	4	5	6	7	8	9
n_i	27	18	23	31	21	23	28	25	22	32

Можно ли считать, что примененный метод действительно дает случайные числа?

Контрольная работа 5 «Метод Монте-Карло и дисперсионный анализ»

Варианты заданий контрольной работы формируются на основе лабораторных работ по соответствующим темам (приложения А3, А4), выполнение которых и является подготовкой к контрольной работе. Вариант контрольной работы содержит три задачи.

Первая задача является частью лабораторной работы «Метод Монте-Карло». При решении задачи требуется оценить количество опытов для вычисления интеграла, сформулировать теоремы, на которых основан метод и пояснить оценивание разброса с помощью характеристики «размах».

Во второй задаче необходимо применить указанный метод генерации псевдослучайных чисел и найти период полученной числовой последовательности. Начальные значения подобраны так, что период будет коротким (5 – 6 различных значений), поэтому если последовательность получается более длинной, следует вернуться и проверить расчеты с самого начала.

Третья задача описывает ситуацию, которую необходимо сформулировать на языке дисперсионного анализа. Поскольку в задаче приведен небольшой набор опытов и не делается предположения о нормальном распределении данных, следует применить ранговый дисперсионный анализ. Критические точки вычислять с помощью таблицы распределения «хи-квадрат».

ВАРИАНТ 1

Задача 1. Представить интеграл в виде, удобном для применения метода Монте-Карло. Указать распределение случайной величины, которую необходимо генерировать для вычисления. Оценить количество опытов при доверительной вероятности β=0,94 и требуемой точности ε=0,03.

$$I = \int_{-\frac{\pi}{6}}^{\frac{\pi}{6}} |\sin 3x| dx$$

Задача 2. Определить период и записать последовательность различных значений псевдослучайных чисел, полученных методом вычетов с начальными значениями m_0=1, M=11, K=14

Задача 3. На четырех малых предприятиях по одной технологии производятся комплектующие детали для основного производства. В таблице приведены данные о производительности труда (в условных единицах). Зависит ли производительность труда от номера предприятия?

Предприятие	Производительность труда					
1	50	53	58	62	60	57
2	54	46	50	64	59	63
3	52	48	51	70	62	61
4	60	55	56	58	54	51

Контрольная работа 6 «Корреляционный и регрессионный анализ»

Варианты заданий контрольной работы формируются на основе соответствующих лабораторных работ, выполнение которых и является подготовкой к контрольной работе. Вариант контрольной работы содержит три задачи.

Первая задача проверяет умение студентов измерять корреляцию между двумя рядами данных. Здесь количество опытов мало и предположения о нормальности распределения не делается, поэтому следует рассчитать коэффициент корреляции Спирмена и проверить его значимость. Если уровень значимости в задании не указан, во всех задачах контрольной работы следует выполнять проверку при $\alpha = 0{,}05$.

Вторая задача требует знания свойств корреляционного отношения. В задаче требуется сформулировать математическую постановку для проверки гипотезы о линейности связи, подобрать статистику с известным законом распределения, построить критическую область и принять статистическое решение.

В третьей задаче проверяется знание основного метода построения регрессионных моделей и умение применить матричную форму записи для решения задачи регрессионного анализа. При дефиците времени пункты г) и д) можно считать дополнительными заданиями

ВАРИАНТ 1

Задача 1. При приеме на работу семи кандидатам на вакантные должности было предложено два теста. Результаты тестирования (в баллах) приведены в таблице. Оценить корреляцию между тестами на уровне значимости 0,05.

Тест	Кандидат						
	1	2	3	4	5	6	7
1	31	82	25	26	53	30	29
2	21	55	8	27	32	42	26

Задача 2. По 56 выборочным данным (в корреляционной таблице 8 значений фактора и 6 значений отклика) была получена оценка коэффициента корреляции $r_{yx} = -0{,}64$ и оценка корреляционного отношения $\eta_{yx}^2 = 0{,}54$. Можно ли описать зависимость y от x линейной моделью?

Задача 3. Предполагается, что зависимость y от x может быть описана линейной функцией. Задание: а) постройте диаграмму рассеяния по выборочным данным и проиллюстрируйте метод наименьших квадратов графически; б) запишите уравнение регрессии для данной таблицы наблюдений в матричной форме; в) найдите оценки коэффициентов матричным способом; г) проверьте адекватность модели; д) постройте 95% доверительный интервал для линии регрессии.

x	-3	-2	-1	1	2
y	-5	-3	-1,5	1	3

Контрольная работа 7 «Временные ряды и цепи Маркова»

Варианты заданий контрольной работы формируются на основе лабораторных работ (приложения А8, А9) и самостоятельного решения задач по двум указанным темам. Вариант контрольной работы содержит две задачи.

Первая посвящена теме «Временные ряды» и может содержать задание на проверку гипотезы о существовании тренда, о наличии периодической составляющей в модели ВР, гипотезы о гетероскедастичности остатков, гипотезы о случайности остатков или гипотезы о наличии автокорреляции в остатках временного ряда. Студенту необходимо повторить алгоритм проверки перечисленных гипотез.

Вторая задача выбирается случайным образом из заданий на лабораторную работу по теме «Цепи Маркова» [7] и может содержать как задачу на исследование эргодической цепи Маркова, так и задачу на поглощающую цепь Маркова.

ВАРИАНТ 1

Задача 1. В таблице представлена динамика выпуска продукции Королевства кривых зеркал (у.е.).

Проверить гипотезы о структуре модели временного ряда
а) по критерию Фостера-Стюарта,
б) по критерию поворотных точек.

Год	Выпуск	Год	Выпуск	Год	Выпуск
2001	11172	2006	13471	2011	23298
2002	14150	2007	13617	2012	26570
2003	14004	2008	16356	2013	23080
2004	13088	2009	20037	2014	23981
2005	12518	2010	21748	2015	23446

Задача 1. Частица на прямой может иметь координаты x=1, 2, 3, 4. Каждую секунду частица может совершать единичные скачки влево или вправо с вероятностями соответственно 0,2 и 0,8. Из положения x=1 частица с вероятностью 0,7 переходит в точку x=2 и с вероятностью 0,3 остается на месте; а из положения x=4 она с вероятностью 0,6 остается на месте, а с вероятностью 0,4 переходит в положение x=3. Задание: а) нарисовать орграф блужданий частицы; б) составить матрицу перехода цепи Маркова; в) найти предельные вероятности.

3.4 Подготовка к экзамену

Экзамен подводит итог двухсеместровому курсу теории вероятностей и математической статистики. Для получения оценки «удовлетворительно» студенту достаточно ответить на вопросы тестового характера (темы вопросов приведены выше в п.3.1).

Примеры вопросов

Вопрос 1. Подброшены две монеты. Для события "выпал хотя бы один герб" противоположным является событие

А) выпала ровно одна цифра;
Б) выпала хотя бы одна цифра;
В) выпало две цифры;
Г) выпало два герба.
Д) выпало две цифры или выпало два герба

Вопрос 2. Если вероятность события А не меняется при наступлении события В, то эти события

А) невозможные;
Б) несовместные;
В) независимые;
Г) неполные.

Вопрос 3. В таблице распределения случайной величины X пропущено значение ***p***, равное

X	2	3	4
P	0,1	***p***	0,3

А) 0 ;   Б) 0,2 ;   В) 0,4;   Г) 0,6 ; Д) 0,8 ;   Е) 1.

Вопрос 4. Функция распределения $F(x)$ случайной величины X НЕ обладает свойством

A) стремится к 1 при $x \to \infty$;
B) неотрицательна;
C) не убывает;
D) площадь под кривой равна 1.

Вопрос 5. Случайная величина X – количество подбрасываний игральной кости до выпадения первой «шестерки» - распределена по закону

А) $Bin(6,\frac{1}{6})$; Б) $Bin(6,\frac{1}{2})$; В) $G(6)$; Г) $G\left(1/6\right)$; Д) $N(6,\ 1/6)$

Вопрос 6. Дана матрица распределения системы дискретных СВ (X,Y): $\left[p_{jk}\right]\left(j=\overline{1,n}; k=\overline{1,m}\right)$. Сумма элементов i-ого столбца равна

А) единице;

Б) вероятности $P(X=x_i)$;

В) вероятности $P(Y=y_i)$;

Г) условной вероятности $P(X\,|\,Y=y_i)$;

Д) условной вероятности $P(Y\,|\,X=x_i)$.

Вопрос 7. Если случайные величины некоррелированы, то их ковариация

А) равна +1;
Б) равна -1;
В) равна   0;
Г) не существует/

Вопрос 8. Закон больших чисел утверждает, что

А) при больших значениях случайной величины ее математическое ожидание постоянно;
Б) при большом значении среднего арифметического дисперсия случайной величины мала;

В) при большом количестве опытов значение среднего арифметического случайной величины примерно равно математическому ожиданию;

Г) при большом количестве опытов среднее арифметическое случайной величины можно описать нормальным законом распределения.

Вопрос 9. Гистограмма – это

А) выборка, упорядоченная по возрастанию;

Б) точечная оценка параметра генеральной совокупности;

В) интервальная оценка параметра генеральной совокупности;

Г) оценка функции распределения генеральной совокупности;

Д) оценка плотности распределения генеральной совокупности.

Вопрос 10. Оценка параметра называется несмещенной, если

А) ее математическое ожидание равно нулю;

Б) ее дисперсия равна нулю;

В) ее математическое ожидание равно значению параметра;

Г) ее дисперсия равна дисперсии параметра.

Вопрос 11. Однофакторный дисперсионный анализ отвечает на вопрос:

А) оказывает ли влияние признак X на фактор F?

Б) оказывает ли влияние фактор F на признак X?

В) какой уровень фактора F оказывает влияние на признак X?

Г) на какой уровень фактора F оказывает влияние признак X?

Вопрос 12. В методе Монте-Карло базовая случайная величина – это случайная величина, распределенная

А) по стандартному нормальному закону;

Б) по закону Стьюдента;

В) по равномерному закону;

Г) по закону больших чисел.

Вопрос 13. Для оценки качества регрессионной модели НЕ применяется

А) коэффициент детерминации;

Б) интервал корреляции;

В) корреляционное отношение;

Г) остаточная дисперсия.

Вопрос 14. Если коэффициент корреляции Пирсона равен (-1), то корреляционное отношение

А) равно +1;

Б) равно -1;

В) меньше 1;

Г) больше 1;

Д) равно 0.

Вопрос 15. Если при корреляционное отношение равно единице, то связь

А) условна;

Б) незначима;

В) линейна;

Г) нелинейна;

Д) функциональна.

Вопрос 16. Корреляционная функция СТАЦИОНАРНОГО случайного процесса - это

А) случайная функция времени;

Б) неслучайная функция времени;

В) неслучайная функция, зависящая от промежутка между сечениями;

Г) неслучайная периодическая функция;

Д) постоянная величина.

Вопрос 17. Состояние цепи Маркова, вероятность выхода из которого равна нулю, называется

А) поглощающим;

Б) транзитным;

В) стационарным;

Г) эргодическим.

Вопрос 18. Спектральная плотность и корреляционная функция стационарного случайного процесса связаны уравнениями

А) Колмогорова;

Б) Винера-Хинчина;

В) Маркова;

Г) Фостера-Стюарта;

Д) Гольдфельда-Квандта.

Вопрос 19. Для устранения гетероскедастичности используется

А) взвешенный метод наименьших квадратов;

Б) авторегрессионное преобразование;

В) метод инструментальных переменных;

Г) рекуррентный метод наименьших квадратов.

Вопрос 20. Для проверки существования периодической составляющей применяют критерий

А) Дарбина-Уотсона;

Б) Фостера-Стюарта;

В) точек роста;

Г) поворотных точек.

Для получения оценки «хорошо» студенту необходимо продемонстрировать умения и навыки, приобретенные при решении задач вероятностного и статистического характера. В билете три задачи, аналогичные задачам индивидуального задания 6 ([8]), заданиям лабораторных работ, задачам контрольных работ 4 - 7. Тематика задач приведена ниже.

Задача А

1) Применение неравенства Чебышева и центральной предельной теоремы для суммы и среднего арифметического независимых СВ.
2) Применение неравенства Чебышева и центральной предельной теоремы для относительной частоты события.
3) Применение теорем Муавра-Лапласа для расчета вероятностей биномиального закона распределения.
4) Построение интервальной оценки. Влияние метода отбора на точность оценки.
5) Определение количества опытов, необходимых для получения оценки с заданной точностью и заданной доверительной вероятностью
6) Проверка параметрических гипотез (о числовых характеристиках НГС, о сравнении параметров НГС, о генеральной доле).
7) Проверка гипотезы о виде распределения по критерию Пирсона.
8) Проверка гипотезы об однородности данных (критерий Вилкоксона).

Задача Б

1) Приведение интеграла к виду, удобному для применения метода Монте-Карло. Предварительная оценка количества опытов.
2) Генерация случайной величины методом вычетов; определение периода, проверка качества моделирования по критерию Пирсона.
3) Решение задачи однофакторного ДА (классическая схема).
4) Решение задачи непараметрического анализа (критерий Краскела-Уоллеса).
5) Выборочный коэффициент корреляции Пирсона. Проверка гипотезы о независимости.
6) Корреляционное отношение. Проверка гипотезы о линейности связи.
7) Коэффициент корреляции Спирмена. Проверка значимости.
8) Парная линейная регрессия. Нахождение оценок коэффициентов регрессии методом наименьших квадратов.
9) Проверка адекватности модели с помощью остаточной дисперсии.
10) Множественная линейная регрессия. Нахождение оценок параметров в матричной форме.

Задача В

1) Компоненты модели ВР. Проверка гипотезы о существовании тренда.
2) Проверка гипотезы о наличии периодической составляющей.
3) Проверка гипотезы о случайности остатков ВР.
4) Проверка гипотезы об отсутствии автокорреляции остатков ВР.
5) Выявление гетероскедастичности данных.
6) Матрица перехода и орграф однородной ЦМ. Многошаговый переход в ЦМ. Распределение вероятностей через m шагов от начала наблюдения.
7) Поглощающие ЦМ. Канонический вид матрицы перехода. Нахождение среднего времени до момента поглощения.
8) Поглощающие ЦМ. Канонический вид матрицы перехода. Нахождение вероятности поглощения в заданном поглощающем состоянии.
9) Эргодические ЦМ. Нахождение предельных вероятностей.
10) Нормированная корреляционная функция ССП, нахождение интервала корреляции.
11) Нахождение эффективной ширины спектра ССП и интервала дискретизации.

Пример практического билета на экзамен

Задача 1. Выборочное обследование распределения населения города по среднедушевому денежному доходу показало, что 45% обследованных в выборке имеют среднедушевой денежный доход не более 18 тыс. руб. В каких пределах находится доля населения, имеющая такой среднедушевой доход, во всей генеральной совокупности, если объем генеральной совокупности составляет 100000 человек, выборка не превышает 10% объема генеральной совокупности и осуществляется по методу случайного бесповторного отбора, а доверительная вероятность принимается равной 0,92?

Задача 2 Предполагается, что зависимость между переменными y и x описывается функцией $y = ax^2 + bx + c$, ошибки наблюдений независимы и имеют распределение $N(0,\sigma)$. С помощью замены переменных перейти к модели множественной линейной регрессии, записать уравнение в матричной форме и найти оценки параметров матричным способом.

x	2,5	4,5	5,0	1,5	3,5	6,0	6,5	4,0	3,5	2,0
y	0,5	1,2	1,7	0,3	0,8	2,7	3,3	1,0	0,7	0,4

Задача 3. По 34 странам оценивалась регрессия расходов на образование от валового национального продукта(ВНП). Были определены линейные регрессии вначале для 12 стран с наименьшим ВНП, а затем для 12 стран с наибольшим ВНП. Сумма квадратов отклонений в первой регрессии равна 2,68, а во второй – 3888,24. Можно ли на однопроцентном уровне значимости принять нулевую гипотезу об отсутствии гетероскедастичности? Каким методом решалась задача?

Для получения оценки «отлично» студент должен знать основные понятия теории вероятностей и математической статистики, умеет излагать их в корректной математической форме, пояснять на примерах. Ниже приводятся примеры теоретических вопросов для собеседования.

Теоретические вопросы

1) Дисперсия и ее свойства. Доказать одну из теорем.
2) Коэффициент корреляции и его свойства. Доказать одну из теорем.
3) Функция регрессии и ее свойства. Доказать основное свойство регрессии.
4) Сходимость по распределению. Интегральная теорема Муавра-Лапласа как следствие центральной предельной теоремы.
5) Сходимость по вероятности. Закон больших чисел. Доказать теорему Чебышева.
6) Сходимость по вероятности. Закон больших чисел. Доказать теорему Бернулли.
7) Метод максимального правдоподобия. Пример получения оценок и исследование свойств.
8) Геометрический метод Монте-Карло и оценка его точности.
9) Однофакторный дисперсионный анализ. Вывести основное дисперсионное тождество, обосновать переход к гипотезе о сравнении дисперсий.
10) Метод наименьших квадратов. Вывод формул для простой линейной регрессии.
11) Цепи Маркова. Доказать теорему о вероятности перехода в эргодическое множество.
12) Канонический вид матрицы перехода поглощающей цепи Маркова. Доказать теорему о свойствах элементов фундаментальной матрицы.

4 Рекомендуемые источники

1. Хрущева И.В. Теория вероятностей: Учебное пособие. - СПб.: Изд-во "Лань",2009.-304с.- ISBN 978-5-8114-0915-0 (Адрес доступа https://e.lanbook.com/reader/book/425/#1)
2. Хрущева И.В., Щербаков В.И., Леванова Д.С.Основы математической статистики и теории случайных процессов: Учебное пособие. – СПб.: Изд-во "Лань",2009.-336с. - ISBN 978-5-8114-0914-3 (Адрес доступа https://e.lanbook.com/reader/book/426/#1)
3. Буре В. М., Парилина Е. М. Теория вероятностей и математическая статистика: Учебник для вузов/- 1-е изд..- СПб.:Изд-во "Лань", 2013.-416с.- ISBN 978-5-8114-1508-3 (Адрес доступа https://e.lanbook.com/reader/book/10249/#1)
4. Гмурман В.Е. Теория вероятностей и математическая статистика.-М.: Высш.шк., 2003.- 480 с.
5. Вентцель Е.С., Овчаров Л.А Теория случайных процессов и ее инженерные приложения - М.:Академия, 2003 г.
6. Потахова, И. В. Эконометрика: Методические указания к лабораторным работам и самостоятельной работе [Электронный ресурс] / И. В. Потахова. — Томск: ТУСУР, 2018. — 60 с. — Режим доступа: https://edu.tusur.ru/publications/8138
7. Смыслова З.А. Цепи Маркова [Электронный ресурс]. - Режим доступа: https://edu.tusur.ru/lecturer/distance_courses/466
8. Смыслова З.А. Теория вероятностей и математическая статистика: Методические указания к практическим занятиям [Электронный ресурс] / З. А. Смыслова. — Томск: ТУСУР, 2018. — 68 с. — Режим доступа: https://edu.tusur.ru/publications/8316

ПРИЛОЖЕНИЕ А1

Параметрические гипотезы

Гипотеза о параметре генеральной совокупности

Гипотеза	$H_0 : a = a_0$	$H_0 : a = a_0$	$H_0 : \sigma = \sigma_0$
Предпо-ложения	НГС, σ известно	НГС, σ неизвестно	НГС *a* неизвестно
Оценки	$\hat{a} = \bar{x}$	$\hat{a} = \bar{x}$; $\hat{\sigma} = s$	$\hat{a} = \bar{x}$; $\hat{\sigma} = s$
Статис-тика	$\frac{\bar{X} - a_o}{\sigma} \cdot \sqrt{n}$	$\frac{\bar{X} - a_o}{s} \cdot \sqrt{n}$	$\frac{(n-1)s^2}{\sigma_o^2}$
Распре-деление	$N(0, 1)$	$T(n-1)$	$\chi^2_{(n-1)}$

Сравнение параметров генеральной совокупности

Гипотеза	$H_0 : a_1 = a_2$	$H_0 : a_1 = a_{20}$	$H_0 : \sigma_1 = \sigma_2$
Предпо-ложения	НГС, σ_1, σ_2 известны	НГС, $\sigma_1 = \sigma_2$, σ_1, σ_2 неизвестны	НГС *a* неизвестно
Оценки	$\hat{a}_1 = \bar{x}; \hat{a}_2 = \bar{y}$	$\hat{a}_1 = \bar{x}; \hat{a}_2 = \bar{y}$ $\hat{\sigma}_1 = s_1; \hat{\sigma}_2 = s_2$ $s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$	$\hat{a}_1 = \bar{x}; \hat{a}_2 = \bar{y}$ $\hat{\sigma}_1 = s_1; \hat{\sigma}_2 = s_2$
Статис-тика	$\frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$	$\frac{\bar{X} - \bar{Y}}{S} \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}}$	$\frac{S_1^2}{S_2^2}$
Распре-деление	$N(0, 1)$	$T(n_1 + n_2 - 2)$	$F(n_1 - 1, n_2 - 1)$

Гипотезы о генеральной доле и о сравнении генеральных долей

Гипотеза	$H_0 : p = p_0$	$H_0 : p_1 = p_2$
Предпол ожения	Схема испытаний Бернулли	Схема испытаний Бернулли
Оценки	$p^* = \frac{m}{n}$	$p_1^* = \frac{m_1}{n_1};\ p_2^* = \frac{m_2}{n_2}$ $\tilde{p} = \frac{m_1 + m_2}{n_1 + n_2}$
Статисти ка	$\frac{p^* - p_o}{\sqrt{p_o(1 - p_o)}} \cdot \sqrt{n}$	$\frac{p_1^* - p_2^*}{\sqrt{\tilde{p}(1 - \tilde{p})}} \cdot \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}}$
Распреде ление	$N(0,\ 1)$	$N(0,\ 1)$

ПРИЛОЖЕНИЕ А2

Варианты задания «Проверка статистических гипотез»

ВАРИАНТ 1

Компания, производящая средства для потери веса, утверждает, что прием таблеток в сочетании со специальной диетой позволяет сбросить в среднем 400 г веса. Случайным образом отобраны 25 человек, использующих эту терапию, и обнаружено, что в среднем еженедельная потеря в весе составила 430 г со с.к.о. 110 г. Проверьте гипотезу о том, что средняя потеря в весе составляет 400 г. Уровень значимости $\alpha = 0{,}05$.

ВАРИАНТ 2

Инженер по контролю качества проверяет среднее время горения нового вида электроламп. Для проверки случайным образом было отобрано 20 ламп, среднее время горения которых составило 1075 часов. Предположим, что среднее квадратичное отклонение времени горения для генеральной совокупности известно и составляет 100 часов. На уровне значимости $\alpha = 0{,}05$ проверьте гипотезу о том, что среднее время горения ламп более 1000 часов.

Предположим, что инженер по контролю качества не имеет информации о генеральной дисперсии и использует выборочное среднеквадратичное отклонение. Изменится ли ответ?

ВАРИАНТ 3

Компания утверждает, что новый вид зубной пасты для детей лучше предохраняет зубы от кариеса, чем зубные пасты, производимые другими фирмами. Для проверки была отобрана случайным образом группа из 400 детей, которые пользовались новым видом зубной пасты. Другая группа из 300 детей, также случайно выбранных, в это же время пользовалась другими видами зубной пасты. Было выявлено, что у 30 детей, использующих новую пасту, и 25 детей из контрольной группы появились новые признаки кариеса. Имеются ли у компании достаточные основания для утверждения о том, что новый сорт зубной пасты эффективнее предотвращает кариес, чем другие виды зубной пасты?

ВАРИАНТ 4

Компания по производству безалкогольных напитков предполагает выпустить на рынок новую модификацию популярного напитка, в котором сахар заменен сукразитом. Компания хотела бы быть уверенной в том, что не менее 70 % ее потребителей предпочтут новую модификацию напитка. Новый напиток был предложен на пробу 2000 человек, и 1422 из них сказали, что он вкуснее старого. Может ли компания отклонить предложение о том, что только 70 % всех ее потребителей предпочтут новую модификацию напитка старой? Уровень значимости 0,05.

ПРИЛОЖЕНИЕ А3

Варианты задания «Метод Монте-Карло»

Вариант	Интеграл I	β	ε
1.	$\int_{2}^{5}\sqrt{x^3+3x}\,dx$	0,92	0,02
2.	$\int_{1}^{3}\sqrt{x^3+8}\,dx$	0,94	0,02
3.	$\int_{-1}^{2}\sqrt{x^2+1}\,dx$	0,98	0,05
4.	$\int_{0,5}^{3}\sqrt{2+x^2}\,dx$	0,94	0,03
5.	$\int_{3}^{5}\sqrt{x^3+2}\,dx$	0,96	0,03
6.	$\int_{0}^{2}\sqrt{8-x^3}\,dx$	0,95	0,03
7.	$\int_{-1}^{2}\sqrt{x^4+1}\,dx$	0,96	0,05

ПРИЛОЖЕНИЕ А4

Варианты задания «Дисперсионный анализ»

ВАРИАНТ 1

25	21	17	28	29
31	35	23	22	26
26	25	24	27	23
21	20	22	30	33
24	23	21	32	20

ВАРИАНТ 2

53	52	45	43	48	54
49	55	38	47	46	49
48	50	51	54	52	46
44	47	46	49	50	42

ВАРИАНТ 3

34	41	42	35	39
40	43	44	43	47
34	37	40	38	32
40	38	39	36	37
41	42	45	48	50

ВАРИАНТ 4

53	69	67	62	61	62
51	64	56	57	66	67
53	58	52	51	53	50
59	59	61	60	60	69

ПРИЛОЖЕНИЕ А5

Варианты задания «Корреляционный анализ»

Часть 1. Несгруппированные данные (фактор - отклик)

ВАРИАНТ 1

0.944	4.467
0.97	4.05
2.273	5.685
4.107	7.325
4.516	7.864
5.978	9.382
6.248	11.61
6.281	9.913
8.548	12.84
8.946	14.79
9.315	15.06
9.4	13.9
9.54	14.8
9.6	15.2
9.72	14.56

ВАРИАНТ 2

1.353	-1.799
3.075	-5.325
4.688	-10.2
5.017	-7.426
5.189	-9.349
5.687	-12.38
6.045	-10.14
7.505	-12.3
7.941	-11.33
9.181	-13.91
9.317	-13.17
10.44	-18.34
10.5	-15.5
10.6	-16.2
10.3	-17.3

ВАРИАНТ 3

1.133	-5.82
2.211	-10.3
2.728	-14.89
4.929	-33.42
5.419	-34.05
6.662	-45.74
6.789	-46.05
7.174	-45.35
7.497	-48.97
7.537	-50.1
10.26	-69.67
10.74	-73.05
11.18	-75.34
11.3	-72.3
11.6	-70.2

ВАРИАНТ 4

1.353	1.527
3.075	0.7278
4.688	0.6984
5.017	0.06758
5.189	0.2915
5.687	0.2237
6.045	-0.147
7.505	-0.2601
7.941	-1.42
9.181	-1.865
9.317	-1.353
10.44	-2.455
10.52	-1.63
10.7	-1.82
10.82	-1.74

ВАРИАНТ 5

1.133	-5.82
2.211	-10.3
2.728	-14.89
4.929	-33.42
5.419	-34.05
6.662	-45.74
6.789	-46.05
7.174	-45.35
7.497	-48.97
7.537	-50.1
10.26	-69.67
10.74	-73.05
11.18	-75.34
11.34	-70.26
11.68	-71.22

ВАРИАНТ 6

1.133	8.652
2.211	11.5
2.728	14.14
4.929	26.59
5.419	28.41
6.662	32.78
6.789	33.93
7.174	37
7.497	35.86
7.537	35.89
10.26	50.86
10.74	49.93
11.18	52.02
11.24	46.92
11.56	48.64

ПРИЛОЖЕНИЕ А6

Варианты задания «Корреляционный анализ»

Часть 2. Сгруппированные данные (корреляционная таблица)

ВАРИАНТ 1

	Y					
X	4	9	14	19	24	29
10	2	3				
20		7	3			
50			2	50	2	
40				10	6	
50				4	7	3

ВАРИАНТ 2

	Y					
X	10	15	20	25	30	35
30	4	2				
40		6	4			
50			6	45	2	
60			2	8	6	
70				4	7	4

ВАРИАНТ 3

	Y					
X	15	20	25	30	35	40
5	2	6				
10		4	4			
15			7	35	8	
20			2	10	8	
25				5	6	3

ВАРИАНТ 4

	Y					
X	6	11	16	21	26	31
22	1	5				
32		5	3			
42			9	40	2	
52			4	11	6	
62				4	7	3

ВАРИАНТ 5

	Y					
X	10	15	20	25	30	35
6	4	2				
12		6	2			
18			5	40	5	
24			2	8	7	
30				4	7	8

ВАРИАНТ 6

	Y					
X	4	9	14	19	24	29
20				4	7	3
30				9	7	
40			4	45	5	
50		4	6			
60	1	4				

ПРИЛОЖЕНИЕ А7

Варианты задания «Регрессионный анализ»

По 12 предприятиям региона изучается зависимость выработки продукции на одного работника y (тыс. руб.) от ввода в действие новых основных фондов x_1 (% от стоимости фондов на конец года) и от удельного веса рабочих высокой квалификации в общей численности рабочих x_2 (%) (смотри таблицу своего варианта). Построить модель множественной линейной регрессии, найти оценки параметров, определить доверительные интервалы для параметров, проверить адекватность модели. Ошибки наблюдений независимы и имеют распределение $N(0,\sigma)$.

Вариант 1				**Вариант 2**			
Номер предприятия	y	x_1	x_2	Номер предприятия	y	x_1	x_2
1	6	3,6	9	1	6	3,5	10
2	6	3,6	12	2	6	3,6	12
3	6	3,9	14	3	7	3,9	15
4	7	4,1	17	4	7	4,1	17
5	7	3,9	18	5	7	4,2	18
6	7	4,5	19	6	8	4,5	19
7	8	5,3	19	7	8	5,3	19
8	8	5,3	19	8	9	5,3	20
9	9	5,6	20	9	9	5,6	20
10	10	6,8	21	10	10	6	21
11	9	6,3	21	11	10	6,3	21
12	11	6,4	22	12	11	6,4	22

ПРИЛОЖЕНИЕ А8

Варианты задания «Временные ряды»

В таблице представлены сведения о доходах Y (одинаковые для всех вариантов), расходах на промышленные товары X (по вариантам) в течение 22 месяцев

Y	Вариант №				
	1	2	3	4	5
	X	X	X	X	X
91,76	16,34	10,90	24,25	3,03	7,73
38,68	10,49	6,99	8,28	7,81	6,07
34,14	5,30	4,08	4,21	1,63	0,28
30,77	13,79	10,61	12,95	4,49	9,18
50,02	2,03	1,57	2,40	0,43	2,63
34,33	9,65	7,43	2,42	6,31	8,51
42,63	13,91	10,70	11,80	5,05	17,05
63,47	3,24	2,16	1,01	3,94	0,85
19,86	2,20	1,47	1,86	0,29	0,91
58,87	12,82	11,65	4,28	5,37	2,33
72,45	29,44	26,77	29,97	6,54	10,96
29,70	8,03	7,30	1,25	0,93	6,98
93,74	33,44	22,29	39,73	1,82	32,73
17,77	0,60	0,40	0,74	0,51	0,72
78,84	32,66	23,33	41,47	15,87	7,89
39,73	6,24	4,46	2,40	1,78	1,48
93,87	26,48	18,91	24,48	25,53	20,85
86,15	25,31	16,87	20,51	31,97	27,76
25,95	2,27	2,06	1,85	2,28	0,19
36,95	12,05	8,03	10,88	10,92	13,03
45,78	20,65	17,20	3,11	12,76	3,41
12,36	0,23	0,15	0,26	0,05	0,34

ПРИЛОЖЕНИЕ А9

Варианты задания «Цепи Маркова»

ВАРИАНТ 1

1. Частица на прямой может иметь координаты x=1,2,3,4. Каждую секунду частица может совершать единичные скачки влево или вправо с вероятностями соответственно 0,3 и 0,7. Из положения x=1 частица с вероятностью 0,7 переходит в точку x=2 и с вероятностью 0,3 остается на месте, а из положения x=4 она с вероятностью 0,7 остается на месте, а с вероятностью 0,3 перейдет в положение x=3. Составить матрицу перехода блужданий и орграф. Является ли ЦМ регулярной? Найти предельные вероятности.

2. Найти фундаментальную матрицу и матрицу B. Какую информацию о цепи Маркова содержат эти матрицы?

$$P=\begin{bmatrix} 1 & 0 & 0 & 0 \\ \frac{1}{3} & 0 & \frac{2}{3} & 0 \\ \frac{1}{5} & \frac{2}{5} & 0 & \frac{2}{5} \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

ВАРИАНТ 2

1. На окружности расположены точки $A_1, A_2, ..., A_4$ - вершины правильного четырехугольника. Частица движется из точки в точку следующим образом: из данной точки она перемещается в одну из ближайших соседних точек с вероятностью 0,5. Построить матрицу перехода и орграф данной цепи Маркова. Какова вероятность частице оказаться в этой же точке через два шага? Является ли ЦМ регулярной? Исследовать поведение системы при $t \to \infty$.

2. Найти фундаментальную матрицу и матрицу B. Какую информацию о цепи Маркова содержат эти матрицы?

$$P=\begin{bmatrix} 1 & 0 & 0 & 0 \\ \frac{1}{4} & 0 & \frac{1}{2} & \frac{1}{4} \\ \frac{2}{3} & \frac{1}{3} & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$