

Министерство науки и высшего образования РФ
ФГБОУ ВО «Ульяновский государственный университет»
Передовая инженерная школа «ФармИнжиниринг»

Саенко В.В.

МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ
СТУДЕНТОВ ПО ДИСЦИПЛИНЕ «БИОМЕДИЦИНСКАЯ СТАТИСТИКА»

для магистрантов Передовой инженерной школы «ФармИнжиниринг»
всех форм обучения

Ульяновск, 2024

Методические указания для самостоятельной работы студентов по дисциплине «Биомедицинская статистика» / составитель: В. В. Саенко. - Ульяновск: УлГУ, 2024.

Настоящие методические указания предназначены для магистрантов Передовой инженерной школы «ФармИнжиниринг» всех форм обучения, изучающих дисциплину «Биомедицинская статистика». В методических указаниях приведена литература по дисциплине, основные темы курса и вопросы в рамках каждой темы, рекомендации по изучению теоретического материала, контрольные вопросы для самоконтроля. Магистрантам они будут полезны при подготовке к практическим занятиям и к зачету по данной дисциплине.

Рекомендовано к введению в образовательный процесс координационным советом Передовой инженерной школы «ФармИнжиниринг» УлГУ (протокол № 2 от «05» июня 2024 г.).

Оглавление

Литература для изучения дисциплины	4
Тема 1. Введение в дисциплину «Биомедицинская статистика»	5
Тема 2. Основы доказательной медицины	5
Тема 3. Случайные величины. Основные виды распределений	5
Тема 4. Системы случайных величин	6
Тема 5. Выборочный метод.....	6
Тема 6. Оценка параметров распределения	6
Тема 7. Основы проверки статистических гипотез	7
Тема 8. Основы дисперсионного анализа	7
Тема 9. Анализ зависимостей	7
Тема 10. Непараметрические критерии	8
Приложение 1 Задачи для самоконтроля.....	8

Литература для изучения дисциплины

Основная литература

- [1] Клинецвич, С. И. Биомедицинская статистика : учебно-методическое пособие / С. И. Клинецвич, А. В. Копыцкий, А. К. Пашко. — Гродно : ГрГМУ, 2023. — 180 с. — ISBN 978 985 595 843 8. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/419039>
- [2] Медик, В. А. Математическая статистика в медицине в 2 т. Том 1 : учебное пособие для вузов / В. А. Медик, М. С. Токмачев. — 2-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2024. — 471 с. — (Высшее образование). — ISBN 978-5-534-07583-0. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/537501>
- [3] Медик, В. А. Математическая статистика в медицине в 2 т. Том 2 : учебное пособие для вузов / В. А. Медик, М. С. Токмачев. — 2-е изд. — Москва : Издательство Юрайт, 2024. — 347 с. — (Высшее образование). — ISBN 978-5-534-11958-9. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/540631>

Дополнительная литература

- [4] Биостатистика : учебное пособие / Д. Н. Бегун, Е. Л. Борщук, Т. В. Бегун [и др.]. — Оренбург : ОрГМУ, 2020. — 117 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/176339>
- [5] Омельченко, В. П. Информатика, медицинская информатика, статистика : учебник / В. П. Омельченко, А. А. Демидова. - Москва : ГЭОТАР-Медиа, 2021. - 608 с. - ISBN 978-5-9704-5921-8. - Текст : электронный // ЭБС "Консультант студента" : [сайт]. - URL : <https://www.studentlibrary.ru/book/ISBN9785970459218.html>
- [6] Илясов, Л. В. Биомедицинская аналитическая техника : учебное пособие для вузов / Л. В. Илясов. — 2-е изд., испр. и доп. — Москва : Издательство Юрайт, 2024. — 332 с. — (Высшее образование). — ISBN 978-5-534-13163-5. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/543230>
- [7] Илясов, Л. В. Биомедицинская измерительная техника : учебное пособие для вузов / Л. В. Илясов. — 2-е изд., испр. и доп. — Москва : Издательство Юрайт, 2024. — 329 с. — (Высшее образование). — ISBN 978-5-534-13079-9. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/543231>
- [8] Леонов, С. А. Статистические методы анализа в здравоохранении. Краткий курс лекций / Леонов С. А. , Вайсман Д. Ш. , Моравская С. В, Мирсков Ю. А. - Москва : Менеджер здравоохранения, 2011. - 172 с. - ISBN 978-5-903834-11-2. - Текст : электронный // ЭБС "Консультант студента" : [сайт]. - URL : <https://www.studentlibrary.ru/book/ISBN9785903834112.html>
- [9] Гмурман Владимир Ефимович. Теория вероятностей и математическая статистика : учеб. пособие для вузов / Гмурман Владимир Ефимович. - 10-е изд., стер. - Москва : Высшая школа, 2004. - 479 с. : ил. - Предм. указ. в конце кн. - ISBN 5-06-004214-6 (в пер.).
- [10] Кремер Наум Шевелевич. Теория вероятностей и математическая статистика : учебник для экон. спец. вузов / Кремер Наум Шевелевич. - 2-е изд., перераб. и доп. - Москва : ЮНИТИ-ДАНА, 2006. - 573 с. - ISBN 5-238-00573-3 (в пер.)

Тема 1. Введение в дисциплину «Биомедицинская статистика»

Основные вопросы темы:

1. Задачи, решаемые с помощью биомедицинской статистики. Особенности организации медико-биологических исследований.
2. Типы данных. Качественные, порядковые и количественные показатели. Шкалы измерения данных.
3. Формирование выборок, рандомизация. Вероятностный подход. Не вероятностный подход.
4. Определение ошибок репрезентативности для долей, определение необходимых объёмов выборок. Понятие о цензурированных (неполных) данных.
5. Этапы статистического исследования. Виды сбора данных.

Рекомендации по изучению темы:

Материал темы 1 изложен в Главе 1 учебника [1]

Тема 2. Основы доказательной медицины.

Основные вопросы темы:

1. Базовые понятия в областях диагностики, лечения и литературного анализа.
2. Фазы клинических испытаний. Понятие конечной точки в клиническом испытании.
3. Таблицы сопряжённости 2x2 в описательных и экспериментальных исследованиях.

Рекомендации по изучению темы:

Материал темы 2 изложен в Главе 2 учебника [1]

Тема 3. Случайные величины. Основные виды распределений

Основные вопросы темы:

1. Функция распределения и плотность вероятности.
2. Дискретные случайные величины и их числовые характеристики. Биномиальное распределение. Распределение Пуассона.
3. Дискретные распределения. Вырожденное распределение. Дискретное равномерное распределение. Отрицательное биномиальное распределение. Гипергеометрическое распределение. Распределение Пуассона.
4. Непрерывные случайные величины. Равномерное распределение. Нормальное распределение.
5. Числовые характеристики распределений.
6. Показательное распределение. Двойное показательное распределение. Распределение Шарлье. Гамма-распределение. Распределение гиперболического косинуса. Распределение гиперболического синуса. Логнормальное распределение.

Рекомендации по изучению темы:

Материал темы 3 изложен в Главе 2 учебника [2]

Тема 4. Системы случайных величин

Основные вопросы темы:

1. Функция распределения и плотность вероятности системы случайных величин.
2. Зависимые и независимые случайные величины.
3. Моменты системы случайных величин. Ковариация.
4. Свойства математического ожидания. Свойства дисперсии. Свойства корреляционных моментов.
5. Независимость и некоррелированность случайных величин. Коэффициент корреляции.
6. Система двух случайных величин.
7. Функции случайных величин: линейная функция одной случайной величины; линейная функция двух случайных величин; произведение двух случайных величин; частное от деления двух случайных величин.
8. Распределение Пирсона. Распределение Стьюдента. Распределение Фишера.
9. Условное распределение. Регрессия. Среднеквадратическая регрессия.

Рекомендации по изучению темы:

Материал темы 4 изложен в Главе 3 учебника [2]

Тема 5. Выборочный метод.

Основные вопросы темы:

1. Основные понятия выборочного метода.
2. Выборочное распределение и его характеристики.
3. Преобразование выборок.
4. Графический метод представления статистических данных.
5. Методика выравнивания статистических рядов.
6. Критерии согласия.
7. Приближенная проверка гипотезы и нормальном распределении.

Рекомендации по изучению темы:

Материал темы 5 изложен в Главе 4 учебника [2]

Тема 6. Оценка параметров распределения

Основные вопросы темы:

1. Особенности малых выборок. Точечные оценки.
2. Точечные оценки для математического ожидания и дисперсии.
3. Распределение некоторых статистик.
4. Интервальные оценки. Доверительные интервалы.
5. Доверительные интервалы для математического ожидания.
6. Доверительные интервалы для дисперсии.
7. Доверительный интервал для разности средних.
8. Оценка вероятности по частоте.
9. Ошибка выборки. Оптимальная численность выборки.

Рекомендации по изучению темы:

Материал темы 6 изложен в Главе 5 учебника [2]

Тема 7. Основы проверки статистических гипотез

Основные вопросы темы:

1. Статистическая гипотеза.
2. Статистические критерии.
3. Сравнение дисперсии двух нормальных генеральных совокупностей.
4. Проверка гипотезы о равенстве неизвестной дисперсии конкретному значению.
5. Сравнение средних двух нормальных генеральных совокупностей при известных дисперсиях.
6. Критерий Стьюдента. Сравнение средних двух нормальных генеральных совокупностей при неизвестных одинаковых дисперсиях.
7. Сравнение выборочной средней с известной величиной. Наблюдения до и после эксперимента.
8. Мощность критерия, уровень значимости; величина различий; объем выборки.
9. Проверка гипотезы о вероятности в схеме Бернулли: одна генеральная совокупность; две генеральные совокупности.
10. Критерий хи-квадрат.
11. Критерий Кочрена сравнения дисперсий нескольких нормальных генеральных совокупностей.
12. Критерий Бартлетта сравнения дисперсий нескольких генеральных совокупностей.

Рекомендации по изучению темы:

Материал темы 7 изложен в Главе 5 учебника [2]

Тема 8. Основы дисперсионного анализа

Основные вопросы темы:

1. Основные понятия дисперсионного анализа.
2. Суммы квадратов отклонений. Общая, факторная и остаточная дисперсии.
3. Однофакторный дисперсионный анализ.
4. Однофакторный дисперсионный анализ в случае разного числа испытаний на различных уровнях.
5. Схема двухфакторного дисперсионного анализа.
6. Множественные сравнения. Критерий Стьюдента с поправкой Бонферрони.
7. Критерий Ньюмена- Кейсла.

Рекомендации по изучению темы:

Материал темы 8 изложен в Главе 7 учебника [2]

Тема 9. Анализ зависимостей

Основные вопросы темы:

1. Типы зависимостей случайных величин.
2. Выборочный коэффициент корреляции.
3. Проверка независимости признаков.
4. Проверка гипотезы о силе линейной связи двух признаков.

5. Выборочная регрессия.
6. Параметры выборочного уравнения регрессии при линейной зависимости.
7. Проверка гипотезы о параметрах уравнения регрессии.
8. Использование линейной регрессии в случае нелинейной зависимости.
9. Мера корреляционной связи. Выборочное корреляционное отношение.
10. Простейшие случаи нелинейной регрессии.

Рекомендации по изучению темы:

Материал темы 9 изложен в Главе 8 учебника [2]

Тема 10. Непараметрические критерии

1. Условия использования критериев.
2. Критерий Манна-Уитни.
3. Критерий Уилкоксона.
4. Критерий знаков.
5. Критерий Краскела-Уоллиса.
6. Критерий Фишера.
7. Гипотеза о независимости элементов выборки: критерий серий, основанный на медиане; критерий «восходящих» и «нисходящих» серий; критерий минимумов и максимумов; критерий Аббе

Рекомендации по изучению темы:

Материал темы 10 изложен в Главе 9 учебника [3]

Приложение 1 Задачи для самоконтроля

Тема 3 «Случайные величины»

1. Используя определения и свойства функции распределения, дайте обоснованный ответ на следующие вопросы:
 - a. может ли функция распределения быть отрицательной?
 - b. может ли функция распределения быть больше 1
 - c. какое ограничение имеется на вероятности p_i значений дискретной случайной величины?
 - d. какой смысл разности $F(b) - F(a)$, где $F(x)$ - функция распределения случайной величины X .
2. Случайная величина X имеет биномиальное распределение с параметрами $n = 5, p = 0.2$. Найти все возможные значения X и соответствующие им вероятности. Найти наиболее вероятное значение m_0 .
3. Случайная величина X имеет распределение Пуассона с параметром $a = 3$. Найти:
 - a) $P(X = 1)$; б) $P(X < 2)$; в) $M(X^2)$.
4. Вероятность рождения мальчика $p = 0.52$. Случайная величина X - количество девочек среди четверых детей. Найти функцию распределения $F(x)$.
5. Два студента отвечают правильно на любой вопрос Программы с вероятностью $p = 0.6$. Каждому из двух студентов задано по два вопроса. Случайная величина X

- число правильных ответов. Найти закон распределения $X, M(X), D(X)$. (Указание. Предварительно установите тип распределения X .)
- б. Записать первые четыре значения и соответствующие им вероятности для случайной величины X :
- имеющей геометрическое распределение с параметром $p = 0.4$;
 - имеющей отрицательное биномиальное распределение с параметрами $p = 0.2, r = 3$

Тема 4. «Системы случайных величин»

7. Случайная величина X имеет логнормальное распределение с параметрами $X \sim N(a, m)$, $a = 1, m = 2$. Найти: $M(X), D(X), As(X), P(X \in (1; 3))$.
8. Бросают две помеченные (например, различно окрашенные) игральные кости. Для данного испытания рассмотреть систему двух случайных величин, количество выпавших очков на каждой кости. Найти закон распределения этой системы.
9. Случайные величины X и Y имеют распределения: $X \sim N(1, 2), Y \sim N(3, 5)$. Найти плотность распределения суммы $Z = X + Y$ при условии, что X и Y : а) независимы; б) не коррелированы. Как связаны независимость и некоррелированность при нормальном распределении? (Указание: Воспользоваться свойством, что сумма нормально распределенных случайных величин распределена также нормально и найти параметры распределения $Z: m_2, \sigma_2$)
10. Как приближенно вычислить математическое ожидание (среднее) n независимых одинаково распределенных случайных величин X_1, X_2, \dots, X_n (указание. См. теорему Чебышева)
11. Дискретная случайная величина X имеет закон распределения:

x_i	-1	0	1	2
p_i	0.1	0.4	0.3	0.2

- Найти закон распределения случайной величины, являющейся функцией от X : а) $Y = X + 1$; б) $Z = X^2$.
12. Случайные величины X и Y подчинены нормальному закону распределения: $X \sim N(1, 3), Y \sim N(2, 4)$. Найти уравнение регрессии Y на X и X на Y в случаях: а) X и Y не коррелированы; б) X и Y коррелированы, и $cov(x, y) = -3^{0.5}$. Построить графики регрессионной зависимости.

Тема 5 «Выборочный метод»

13. Абитуриент на вступительном экзамене может получить любое количество баллов от 0 до 20. В результате наблюдения над группой из 40 абитуриентов получены следующие значения баллов: 16, 10, 12, 8, 15, 11, 17, 13, 12, 19, 16, 2, 19, 11, 14, 13, 17, 20, 19, 5, 14, 18, 9, 15, 11, 17, 19, 10, 16, 14, 17, 19, 15, 13, 20, 12, 18, 14, 16, 14. Записать эти данные в виде: а) вариационного ряда; б) статистического ряда; в) статистического распределения. Что в данном случае подразумевается под генеральной совокупностью?
14. Выборка записана в виде статистического ряда:

x_i	1	5	10
m_i	8	10	7

Найти выборочное значение \bar{x} и выборочную дисперсию \bar{s}^2

15. Имеются 22 наблюдения в различных поликлиниках области о числе лабораторных исследований на 100 посещений: 136.8, 96.1, 150.7, 185.6, 281.2, 132.8, 197.1, 93.2, 133.0, 200.2, 131.6, 123.3, 97.9, 152.9, 110.9, 104.4, 108.6, 107.4, 112.6, 118.5, 131.3, 124.3. Исходя из данной выборки найти точную оценку среднего числа исследований на 100 посещений и выборочную дисперсию.
16. Абитуриент на вступительном экзамене может получить любое количество баллов от 0 до 20. В результате наблюдения над группой из 40 абитуриентов получены следующие значения баллов: 16, 10, 12, 8, 15, 11, 17, 13, 12, 19, 16, 2, 19, 11, 14, 13, 17, 20, 19, 5, 14, 18, 9, 15, 11, 17, 19, 10, 16, 14, 17, 19, 15, 13, 20, 12, 18, 14, 16, 14. Найти моду и медиану выборки.
17. Имеются 22 наблюдения в различных поликлиниках области о числе лабораторных исследований на 100 посещений: 136.8, 96.1, 150.7, 185.6, 281.2, 132.8, 197.1, 93.2, 133.0, 200.2, 131.6, 123.3, 97.9, 152.9, 110.9, 104.4, 108.6, 107.4, 112.6, 118.5, 131.3, 124.3. Преобразовать выборку по данным задачи в выборку с четырьмя равноотстоящими вариантами. Вычислить дисперсию с поправкой Шеппарда. Объясните, чем вызвано в данной ситуации расхождение выборочной дисперсии со случаем исходных данных.
18. Из генеральной совокупности X извлечена выборка объема $n = 60$.

x_i	1	2	3
m_i	10	30	20

Найти начальные и центральные выборочные моменты до четвертого порядка включительно. Вычислить выборочные коэффициенты асимметрии и эксцесса.

Тема 6. «Оценка параметров распределения»

19. Первичный диагноз заболевания A был поставлен в клинике 150 пациентам. Из них, несмотря на проведенное в клинике лечение, в течение года после постановки диагноза умерли 27 больных. С надежностью $\gamma = 0.9$ найти интервальную оценку вероятности p – смерти в течение года после постановки первичного диагноза заболевания A больного, получившего лечение в этой клинике.
20. При тестировании лекарственного препарата на 14 испытуемых побочные эффекты выявлены в одном случае. С надежностью $\gamma = 0.95$ найти доверительный интервал для вероятности p – появления побочных эффектов у одного испытуемого. (Указание. Найдите доверительный интервал тремя способами)
21. При исследовании вопроса наследственности у больных шизофренией и эпилепсией обследовано $n_1 = 900$ больных шизофренией и $n_2 = 1100$ больных эпилепсией. Среди обследованных больных шизофренией 63 имели наследственный характер заболевания (т.е. 7%), среди обследованных больных эпилепсией – 110 (т.е. 10%). Разность, полученная по этим данным, составляет $10\% - 7\% = 3\%$. Можно ли утверждать, что наследственность более выражена при заболевании эпилепсией, чем шизофренией? Привести ответ с надежностью: а) $\gamma = 0.95$, б) $\gamma = 0.99$.
22. Из предварительного исследования известно, что среди детей дошкольного возраста с речевыми расстройствами левши составляют 15%. Найти необходимый объем выборки n чтобы предельная ошибка выборки при надежности $\gamma = 0.95$ составляла 4%.

23. При наблюдении над $n = 100$ случайным образом выбранными детьми дошкольного возраста с речевыми расстройствами установлено, что $m = 60$ из них левши. Учитывая, что процент левшей – величина случайная (для другой выборки она примет скорее всего другое значение), с надежностью 0.95 найти доверительный интервал для величины процента левшей среди детей дошкольного возраста с речевыми расстройствами.
24. При наблюдении над $n = 400$ случайным образом выбранными детьми дошкольного возраста с речевыми расстройствами установлено, что $m = 60$ из них левши. Учитывая, что процент левшей – величина случайная (для другой выборки она примет скорее всего другое значение), с надежностью 0.95 найти доверительный интервал для величины процента левшей среди детей дошкольного возраста с речевыми расстройствами.

Тема 7. «Проверка статистических гипотез»

25. Измерения одной и той же величины проведены двумя методами по несколько раз. Получены две выборки результатов:

X:	9.6	10.0	9.8	10.2	10.6
Y:	10.4	9.7	10.0	10.3	

Предполагая, что выборки независимы и извлечены из нормальных генеральных совокупностей, сравнить точность измерений при уровне значимости $\alpha = 0.1$. (Указание. $H_0: D(X) = D(Y)$, $H_1: D(X) \neq D(Y)$.)

26. Нормальная генеральная совокупность X имеет дисперсию $D(X) = 1600$. Извлеченная из генеральной совокупности X выборка объемом $n = 64$ имеет выборочную среднюю $\bar{x} = 136.5$. При уровне значимости $\alpha = 0.01$ требуется проверить нулевую гипотезу $H_0: M(X) = 130$ при конкурирующей гипотезе $H_1: M(X) \neq 130$.
27. Установлено, что средняя масса таблетки сильнодействующего лекарства должна быть равна 0.50 мг. Выборочная проверка 121 таблетки полученной партии лекарства показала, что средняя масса таблетки этой партии $\bar{x} = 0.53$ мг. Требуется при уровне значимости $\alpha = 0.01$ проверить гипотезу $H_0: M(X) = 0.50$ при конкурирующей гипотезе $H_1: M(X) \neq 0.5$, где масса таблетки – генеральная совокупность X . В процессе многократных предварительных опытов по взвешиванию таблеток, поставляемых фармацевтической фирмой, было установлено, что масса таблеток распределена нормально со средним квадратическим отклонением $\sigma = 0.11$ мг.
28. Установлено, что средняя масса таблетки сильнодействующего лекарства должна быть равна 0.50 мг. Выборочная проверка 121 таблетки полученной партии лекарства показала, что средняя масса таблетки этой партии $\bar{x} = 0.53$ мг. Требуется при уровне значимости $\alpha = 0.01$ проверить гипотезу $H_0: M(X) = 0.50$ при конкурирующей гипотезе $H_1: M(X) \neq 0.5$, где масса таблетки – генеральная совокупность X . В процессе многократных предварительных опытов по взвешиванию таблеток, поставляемых фармацевтической фирмой, было установлено, что масса таблеток распределена нормально со средним квадратическим отклонением σ . Решить задачу при условии, что значение

- σ заранее не известно, а его оценка s найдена по имеющейся выборке: $s = 0.10$.
(Указание. Использовать критерий Стьюдента. $t_{\text{правостор.кр.}}(0.01, 120) = 2.36$)
29. Предполагается, что некоторое событие в результате испытания осуществляется с постоянной вероятностью $p_0 = 0.20$. По 100 независимым испытаниям найдена относительная частота осуществления этого события $\omega = 0.16$. Требуется при уровне значимости $\alpha = 0.05$ проверить гипотезу $H_0: p = p_0 = 0.20$ при конкурирующей гипотезе $H_1: p \neq 0.20$. (Указание. В качестве критерия проверки нулевой гипотезы используйте статистику Z вида.)
 30. В двух цехах завода работают соответственно 200 и 300 человек. Заболеваемость простудными болезнями в течение наблюдаемого периода составила 13% для первого цеха и 20% для второго. При уровне значимости $\alpha = 0.05$ проверить нулевую гипотезу о равенстве вероятностей заболеваний одного работающего в первом цеху и во втором цехах: $p_1 = p_2 = 0.13$ при конкурирующей гипотезе $p_1 \neq p_2$. (Указание. В качестве критерия проверки нулевой гипотезы используйте статистику Z вида.)
 31. В двух цехах завода работают соответственно 200 и 300 человек. Заболеваемость простудными болезнями в течение наблюдаемого периода составила 13% для первого цеха и 20% для второго. При уровне значимости $\alpha = 0.05$ проверить нулевую гипотезу о равенстве вероятностей заболеваний одного работающего в первом цеху и во втором цехах: $p_1 = p_2 = 0.13$ при конкурирующей гипотезе $p_1 \neq p_2$. Вместо гипотетической вероятности $p = 0.13$ использовать среднюю вероятность p , вычисляемую по наблюдаемым данным.
 32. В двух цехах завода работают соответственно 200 и 300 человек. Заболеваемость простудными болезнями в течение наблюдаемого периода составила 13% для первого цеха и 20% для второго. При уровне значимости $\alpha = 0.05$ проверить нулевую гипотезу о равенстве вероятностей заболеваний одного работающего в первом цеху и во втором цехах: $p_1 = p_2 = 0.13$ при конкурирующей гипотезе $p_1 \neq p_2$. Вместо гипотетической вероятности $p = 0.13$ использовать среднюю вероятность p , вычисляемую по наблюдаемым данным. Решить задачу, используя поправку Йейтса.
 33. По шести независимым выборкам одинаково объема $n = 37$, извлеченным из нормальных генеральных совокупностей, найдены исправленные выборочные дисперсии: 2.34, 2.66, 2.95, 3.65, 3.86, 4.54. Используя критерий Кочрена, требуется проверить нулевую гипотезу об однородности дисперсии. Уровень значимости $\alpha = 0.05$.
 34. По четырем независимым выборкам, объемы которых $n_1 = 17, n_2 = 20, n_3 = 15, n_4 = 16$, извлеченным из нормальных генеральных совокупностей, найдены исправленные выборочные дисперсии: 2.5, 3.6, 4.1, 5.8. Используя критерий Бартлетта, на уровне значимости 0.05 проверить гипотезу об однородности дисперсии. Оценить генеральную дисперсию.

Тема 8. «Дисперсионный анализ»

35. На 5 самках дрозофилы поставлен опыт по изучению развития мушек из оплодотворенных яиц при разных температурах (20 °C, 25 °C, 30 °C). От каждой самки при каждой из указанных температур брали по 40 яиц. Их разделяли на две группы по 20 яиц в каждой. Таким образом, получено 30 групп по 20 яиц в каждой. Наблюдаемое значение - число неразвившихся яиц среди каждых 20. Так как

наблюдаемых групп 30, то имеется всего 30 наблюдений, каждое из которых получено от определенной самки (фактор A с 5 уровнями, по количеству самок) и при определенной температуре (фактор B с 3 уровнями, по количеству рассматриваемых температур). Все 30 наблюдений занесены в таблицу

Уровни фактор A	Уровни фактора B	20 °С	25 °С	30 °С
1	1	1; 1	0; 4	0; 1
2	11	11; 9	5; 5	10; 14
3	4	4; 3	3; 2	1; 1
4	10	10; 7	8; 6	5; 7
5	2	2; 0	2; 0	2; 4

При уровне значимости $\alpha = 0.05$ методом двухфакторного дисперсионного анализа требуется исследовать влияние факторов A (индивидуальность самки), B (температура) и AB (взаимодействие факторов индивидуальности самки и температуры) на развитие мушек дрозофилы из оплодотворенных яиц. (Указание. По данным задачи составьте таблицу, в каждой ячейке которой - сумма наблюдаемых элементов, в частности (см. данные): $T_{11} = 1 + 1 = 2$, $T_{21} = 11 + 9 = 20$ и т.д. Далее, используя формулы (7.25) - (7.34), заполните таблицу дисперсионного анализа. Сравнивая $F_{\text{набл}}$ с соответствующим $F_{\text{кр}}$ сделайте вывод о влиянии каждого из факторов и их взаимодействия на рассматриваемую случайную величину - количество неразвившихся яиц.)

36. Данные о количестве сцеженного и высосанного молока (в процентах к среднесуточному) у восьми кормящих матерей, страдающих гипогалактией, в течение 6-суточных кормлений представлены в таблице.

Час кормления	Кормящие матери							
	1-я	2-я	3-я	4-я	5-я	6-я	7-я	8-я
6	100	141	147	126	133	101	148	128
9	92	119	122	104	115	102	128	146
12	101	87	88	105	93	102	74	94
15	127	83	80	87	96	106	68	97
18	111	84	83	96	83	90	101	64
21	69	86	80	82	80	99	81	71

Необходимо проверить, действительно ли в течение суток за шесть кормлений количество сцеженного и высосанного молока у кормилиц с гипогалактией уменьшается или это различие случайно. Исследование провести при уровне значимости $\alpha = 0,01$.

Указание. Имеющиеся данные записать в таблицу представленной ниже формы по уровням фактора - часа кормления

Номер испытания i	Уровни фактора (час кормления)					
	6	9	12	15	18	21
1	100	92	101	127	11	69
2	141	119	87	83	84	86
3	147	122	88	80	83	80
4	126	104	105	87	96	82
5	133	115	93	96	83	80
6	101	102	102	106	90	99
7	148	128	74	68	101	81
8	128	146	94	97	64	71
$\bar{x}_{гр,j}$	128	116	93	93	89	81

37. Данные о количестве сцеженного и высосанного молока (в процен-тах к среднесуточному) у восьми кормящих матерей, страдающих гипогалактией, в течение 6-суточных кормлений представлены в таблице. При экспериментальном изучении было решено устранить «выделяющиеся» результаты наблюдений, используя определенную методику. Оставшиеся данные занесены в таблицу ниже.

Час кормления	Кормящие матери								$\bar{x}_{гр,j}$
	1-я	2-я	3-я	4-я	5-я	6-я	7-я	8-я	
6	100	-	-	126	133	101	-	128	117.60
9	-	119	122	104	115	102	128	-	115.00
12	101	87	88	105	93	102	74	94	93.00
15	-	83	-	87	96	-	-	97	90.75
18	-	84	83	96	83	90	101	64	85.86
21	69	86	80	82	80	99	81	71	81.00

Применяя однофакторный дисперсионный анализ, требуется установить, имеется ли существенное различие между количеством сцеженного и высосанного молока во время шести кормлений в течение суток.

38. Имеются данные о плодовитости мышей при облучении их рентге-новскими лучами.

Группа	Число мышат от отдельных самок
--------	--------------------------------

Контрольная	10	12	11	10
Получившая дозу 100 Р	8	10	7	9
Получившая дозу 200 Р	7	9	6	4

Установить методами дисперсионного анализа, влияет ли облучение на плодовитость мышей. Если различие между группами по уровням фактора - дозы облучения существенно, то нужно провести попарное сравнение групп, используя критерий Стьюдента для множественных сравнений.

39. Данные общей заболеваемости в различных районах страны (на 1000 чел. населения) приведены в таблице.

	Сельские районы	Пригородные районы	Небольшие города	Средние города	Крупные города
Туберкулез	4.9	8.9	6.8	8.4	11.2
Прочие заболевания	21.9	20.7	21.7	20.4	20.0

Методами дисперсионного анализа требуется установить влияние фактора местности на заболеваемость населения. Проследите отдельно данные по туберкулезу. Какие напрашиваются выводы?

40. Данные о причинах смертности в разных социальных группах населения Франции за 1986 г. (на 100 тыс. занятых) приведены в таблице.

Причина смерти (фактор B)	Вид занятия (фактор A)				
	Руководители высшего звена	преподаватели	Руководители среднего звена	Сельскохозяйственные рабочие	Промышленные рабочие
Новообразование	150	140	205	290	350
Сердечно-сосудистые болезни	130	150	180	190	185
Несчастные случаи	45	30	75	175	95
Цирроз печени	15	16	33	75	95
Самоубийства	20	25	36	30	45

а) Методами однофакторного дисперсионного анализа требуется установить влияние профессионального фактора A на смертность. Если различие по группам (уровням фактора) существенное, то провести сравнения групп: а) используя критерий Стьюдента для множественных сравнений; б) используя критерий Ньюмена — Кейлса.

б) Методами однофакторного дисперсионного анализа установить влияние заболеваемости и бытового фактора (причина смерти) на смертность населения. Если различие по группам причин смертности (фактор В) существенно, то провести сравнение групп: а) используя критерий Стьюдента для множественных сравнений; б) используя критерий Ньюмена — Кейлса.

Тема 9. «Анализ зависимостей»

41. Какой смысл несет понятие «некоррелированность» случайных величин? Как связаны понятия независимости и некоррелированности? При каких условиях на случайные величины понятия некоррелированности и независимости совпадают?
42. При каких значениях коэффициента корреляции зависимость случайных величин является: а) прямой; б) обратной?
43. Знания 12 студентов по двум различным предметам оценены по сто-балльной шкале. В результате получены две выборки значений:

48	52	64	74	91	78	62	53	74	66	93	99
56	60	63	80	94	76	51	58	80	61	88	98

Найти выборочный коэффициент ранговой корреляции Спирмена между оценками.

44. Два члена жюри, оценив профессионализм 13 участников конкурса «Врач года», расставили их по местам следующим образом (проранжировали):

1-й	1	2	3	4	5	6	7	8	9	10	11	12	13
2-й	6	3	4	2	1	10	7	8	9	5	11	13	12

С помощью коэффициента ранговой корреляции Спирмена требуется оценить согласованность членов жюри. Проверить гипотезу об отсутствии согласованности: $p = 0$.

45. Имеются данные о вакцинации против гриппа и заболеваемости гриппом во время эпидемии: вакцинированных всего 500 чел., из них заболели 10; не вакцинированных 1500, из них заболели 990 чел. Требуется определить величину зависимости между проведенной вакцинацией и заболеваемостью гриппом.

	Заболели	Не заболели	Всего
Вакцинировано	10	490	500
Не вакцинировано	990	510	1500
Всего	1000	1000	2000

Используя данные таблицы, вычислить коэффициент контингенции Пирсона.