

Федеральное государственное бюджетное образовательное учреждение
высшего образования "Красноярский государственный медицинский
университет имени профессора В.Ф. Войно-Ясенецкого"
Министерства здравоохранения Российской Федерации

Кафедра медицинской кибернетики и информатики

Медицинская статистика

**Сборник методических указаний для обучающихся к практическим занятиям по направлению
подготовки 34.03.01 Сестринское дело (очная форма обучения)**

Красноярск

2022

Медицинская статистика : сборник методических указаний для обучающихся к практическим занятиям по направлению подготовки 34.03.01 Сестринское дело (очная форма обучения) / сост. М.С. Апанович, Е.Н. Галушина. - Красноярск : тип. КрасГМУ, 2022.

Составители:

к.ф.-м.н. М.С. Апанович

к.ф.-м.н. Е.Н. Галушина

Сборник методических указаний к практическим занятиям предназначен для аудиторной работы обучающихся. Составлен в соответствии с ФГОС ВО 2017 по направлению подготовки 34.03.01 Сестринское дело (очная форма обучения), рабочей программой дисциплины (2022 г.) и СТО СМК 8.3.12-21. Выпуск 5.

Рекомендован к изданию по решению ЦКМС (Протокол № 10 от 26 мая 2022 г.)

© ФГБОУ ВО КрасГМУ
им.проф. В.Ф. Войно-Ясенецкого
Минздрава России, 2022

1. Тема № 1. Общие понятия о статистических методах исследования в медицине. Программное обеспечение для статистического анализа

2. Разновидность занятия: комбинированное

3. Методы обучения: объяснительно-иллюстративный, метод проблемного изложения

4. Значение темы (актуальность изучаемой проблемы): Подавляющее большинство программного обеспечения, используемого психологами в исследовательских проектах, относятся к категории закрытых коммерческих продуктов. Текстовые документы верстаются в редакторе Word от компании Microsoft, статистический анализ производится в SPSS от компании IBM. Хотя наряду с ними существуют свободное бесплатное программное обеспечение, не уступающее им по функциональности.

5. Цели обучения

- **обучающийся должен знать** принципы работы современных информационных технологий, современные методики сбора и обработки информации, необходимой для проведения научного исследования, **уметь** понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности, анализировать полученные в результате вычислений статистические величины (абсолютные, средние, относительные показатели, показатели динамического ряда) на основании сравнения в динамике и с имеющимися нормативами, **владеть** современными информационными технологиями для решения профессиональных задач, навыками анализа полученных в результате вычислений статистических величин

6. Место проведения и оснащение занятия:

- **место проведения занятия:** компьютерный класс №1 (З-03)

- **оснащение занятия:** видеопроектор, доска магнитно-маркерная, комплект учебной мебели на посадочные места, локальный сетевой сервер, персональные компьютеры, экран

7. Аннотация (краткое содержание темы)

Понятие "статистика" можно определить различными способами:

- **СТАТИСТИКА** — наука, изучающая количественные стороны массовых явлений в конкретных условиях места и времени и разрабатывающая теоретические положения и методы, используемые статистической практикой.
- **СТАТИСТИКА** — данные о различных явлениях общественной жизни.
- **СТАТИСТИКА** — отрасль практической деятельности по сбору, обработке и анализу статистических данных.

Статистическое исследование – это анализ и обработка, как правило, большого количества элементов, иногда доходящее до десятков и сотен миллионов. Например, чтобы установить среднее количество детей в российской семье, строго говоря, необходимо обследовать все семьи страны. Такая работа потребует больших временных и материальных затрат и будет неэффективной. Поэтому чаще всего обработке подвергают не все элементы, а только небольшую их часть.

Измерение — операция для определения отношения одного объекта к другому. Измерение реализуется за счет приписывания объектам значений так, чтобы отношения между значениями отражали отношения между объектами. К примеру, мы измеряем рост двух людей (объект измерения — рост). Получив значения 170 и 185 см. мы можем точно сказать, что один человек выше другого. Данный вывод был получен благодаря измерению роста. Таким образом, отношение между объектами было передано с помощью чисел.

Всё, что мы измеряем, называется переменными.

Переменная — свойство, которое может менять своё значение. Рост является свойством всех людей, но у каждого он разный, а значит, является переменной. Пол так же является переменной, но может принимать всего 2 значения.

Шкала измерения — ограничение типа отношений между значениями переменных, накладываемое на результаты измерений. Чаще всего, шкала измерения зависит от инструмента измерения.

К примеру, если переменной является цвет глаз, то мы не можем сказать, что один человек больше или меньше другого по этой переменной, мы так же не можем найти среднее арифметическое цвета. Если переменной является порядок (именно порядок) рождения детей в семье, то мы можем сказать, что первый ребенок однозначно старше второго, но не можем сказать, на сколько он старше (отношения «больше/меньше»). Имея результаты теста интеллекта, мы можем однозначно сказать, на сколько, один человек интеллектуальнее другого.

С. Стивенс рассматривал четыре шкалы измерения:

- **Номинативная шкала** (номинальная, шкала наименований) – состоит в присваивании какому-либо свойству или признаку определенного обозначения или символа.

При измерении в этой шкале осуществляется классификация или распределение на непересекающиеся классы. С величинами, измеряемыми в шкале наименований, можно выполнять только одну операцию — проверку их совпадения или несовпадения. При построении шкалы наименований главными являются качественные различия, а количественные не принимаются во внимание. Поэтому числа, используемые в качестве обозначений классов эквивалентности в этих шкалах, не отражают количественных различий выраженности изучаемого признака.

Примеры: пол, варианты ответов испытуемых, привитые/не привитые и т.д.

- **Порядковая шкала (или ранговая)** классифицирует совокупность измеренных признаков по принципу «больше-меньше», «выше-ниже», «сильнее-слабее».

Примеры: оценка состояния больного (неудовлетворительно, удовлетворительно, хорошо, отлично), и т.д.

- **Интервальная шкала** — это шкала, классифицирующая по принципу «больше на определенное количество единиц — меньше на определенное количество единиц». Каждое из возможных значений признака отстоит от другого на равном расстоянии.

Примеры: тестовые шкалы.

- **Шкала отношений** обладает всеми свойствами интервальной шкалы и имеет твердо фиксированный нуль, который означает полное отсутствие свойства.

Примеры: рост; вес; число реакций; показатель силы; выносливости.

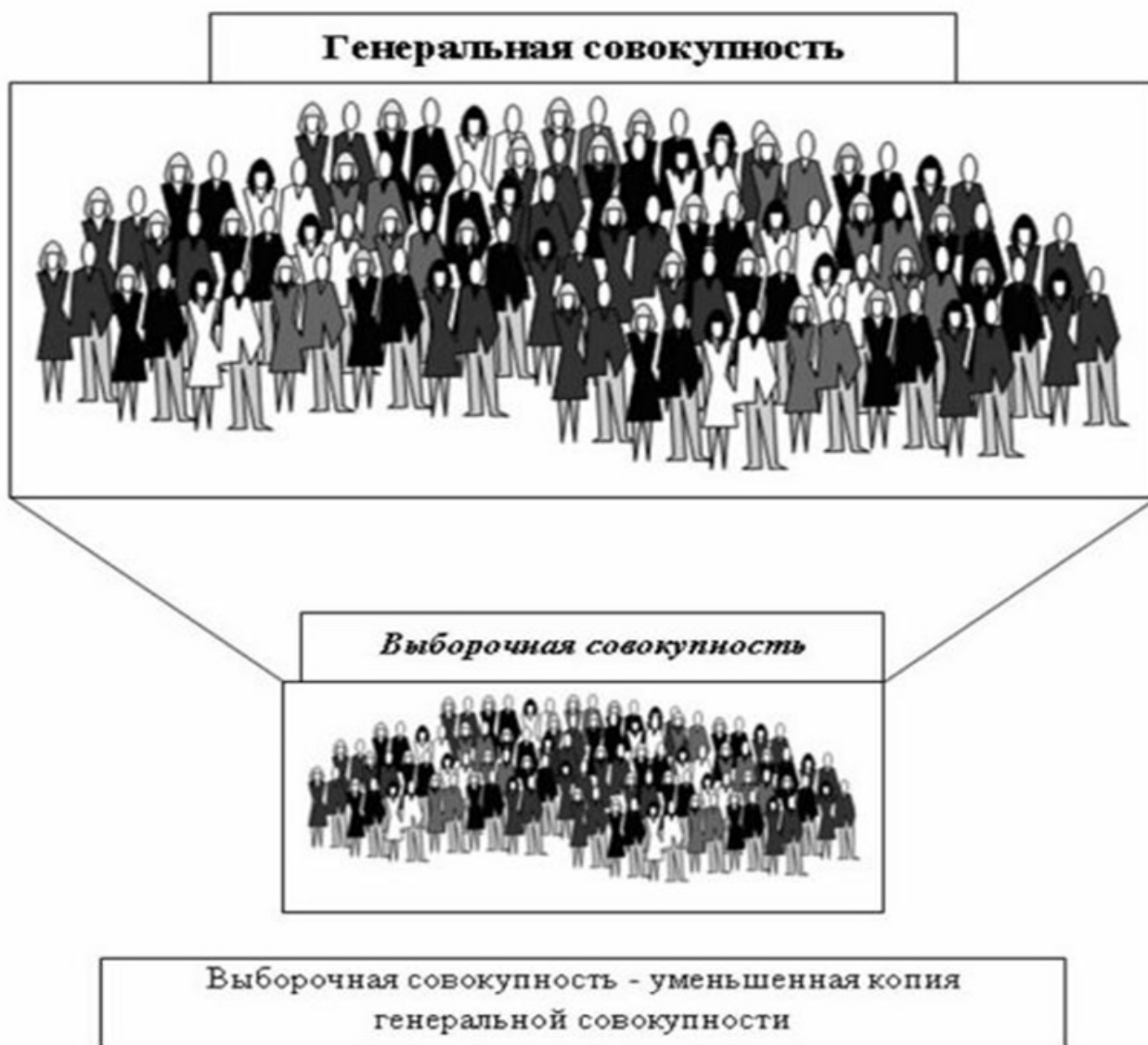
Генеральная совокупность — полная совокупность элементов, подвергающихся статистическому исследованию.

Выборка — часть генеральной совокупности элементов, которая охватывается экспериментом (наблюдением, опросом).



Для того, чтобы сформированная выборочная совокупность правильно отражала свойства генеральной совокупности, она должна удовлетворять условиям:

- объем выборки должен быть достаточно велик;
- выборка должна быть репрезентативной, то есть представительной. Любая выборка может быть репрезентативной лишь в каких-то определенных, но не всех отношениях.



Любая переменная имеет свой минимум и максимум, между которыми располагаются ее значения. Расстояние от минимума до максимума называется областью значений переменной, однако мы можем разбить это расстояние на равные отрезки и получить множество ограниченных интервалов значений. Распределение вероятностей позволяет нам судить о том, какова вероятность появления переменной в каждом из интервалов её значения.

Свободное программное обеспечение для статистического анализа

Статистический анализ данных требует большого количества сложных вычислений, поэтому выполнение статистического анализа невозможно без вычислительной техники и специализированного программного обеспечения.

Рассмотрим свободное программное обеспечение для статистического анализа данных. Напомним, что свободное программное обеспечение - это программное обеспечение, в отношении которого права пользователя («свободы») на неограниченную установку, запуск, а также свободное использование, изучение, распространение и изменение (совершенствование) защищены юридически авторскими правами при помощи свободных лицензий либо на это программное обеспечение нет исключительных прав.

PSPP - это свободное программное обеспечение для статистического анализа данных. PSPP позволяет вычислять описательные статистики, производить проверку статистических гипотез, строить уравнения линейной и логистической регрессии, производить факторный и кластерный анализ, использовать непараметрические статистические критерии и т.д.

PSPP имеются как графический пользовательский интерфейс, так и традиционный интерфейс командной строки.

PSPP позволяет обмениваться данными с другим программным обеспечением: Gnumeric, Libre Office, Open Office.

Отчёты могут быть представлены в форматах файлов: txt, Post Script, pdf, Open Document, html. Позволяет импортировать данные из электронных таблиц, текстовых файлов и баз данных. Поддерживает возможность открывать, анализировать и редактировать несколько наборов данных одновременно.

PSPP доступен в различных операционных системах.

Установка PSPP в операционной системе Windows:

1. Зайдите с помощью браузера на сайт <http://www.gnu.org/software/pspp/get.html>
2. В разделе Binaries нажмите на ссылку **Another windows version** для MS Windows
3. В таблице Downloads нажмите на ссылку с самым новым инсталлятором (в таблице он указан первой строкой), соответствующий разрядности Вашей операционной системе (32-битной или 64-битной)
4. Загрузка должна начаться автоматически.

Запустите загруженный файл. Мастер установки поможет вам установить программу.

Ввод данных в PSPP

После запуска программы откроется начальное окно программы.

Данные вводятся в таблицу по столбцам. В шапке таблицы указываются названия переменных. В каждой строке значения переменной для одного наблюдения (пациента, измерения и т.д.).

Для того чтобы добавился столбец для новой переменной нужно ввести первое значение. По умолчанию переменным дается название Var_. Это название можно изменить на вкладке Variable View.

- Name - имя переменной;
- Type - тип переменной (numeric - числовой, comma - числовой с запятой в качестве разделителя целой и дробной части, dot - числовой с точкой в качестве разделителя целой и дробной части, scientific notation - экспоненциальный, date - дата, dollar - доллар, custom currency - произвольная валюта, string - строковый);
- Width - ширина поля;
- Decimals - количество знаков после запятой.

Другие поля не являются обязательными для заполнения.

R — язык программирования для статистической обработки данных и работы с графикой, а также свободная программная среда вычислений с открытым исходным кодом в рамках проекта GNU. Язык создавался как аналогичный языку S, разработанному в Bell Labs и является его альтернативной реализацией, хотя между языками есть существенные отличия, но в большинстве своём код на языке S работает в среде R.

Изначально R был разработан сотрудниками статистического факультета Оклендского университета Россом Айхэкой (англ. Ross Ihaka) и Робертом Джентлменом (англ. Robert Gentleman) (первая буква их имён — R), на момент 2011 года язык и среда поддерживаются и развиваются организацией R Foundation.

R широко используется как статистическое программное обеспечение для анализа данных и фактически стал стандартом для статистических программ.

R доступен под лицензией GNU GPL. Распространяется в виде исходных кодов, а также откомпилированных приложений под ряд операционных систем: FreeBSD, Solaris и другие дистрибутивы Unix и Linux, Microsoft Windows, Mac OS X.

В R используется интерфейс командной строки, хотя доступны и несколько графических интерфейсов пользователя, например пакет R Commander, RKWard, RStudio, Weka, Rapid Miner, KNIME, а также средства интеграции в офисные пакеты.

В 2010 году R вошёл в список победителей конкурса журнала InfoWorld в номинации на лучшее открытое программное обеспечение для разработки приложений.

Установка R в операционной системе Windows

1. Зайдите с помощью браузера на сайт <http://cran.gis-lab.info/>
2. Нажмите на ссылку **Download R for Windows**
3. Выберите в Subdirectories ссылку base:
4. Нажмите ссылку Download 3.0.2 for Windows
5. Загрузка исполняемого файла начнется автоматически.

6. Запустите скаченный файл, следуйте советам мастера установки.

Импорт данных в R

Импортируемый файл имеет следующий формат:

1 строка: названия переменных, разделенных запятой. В последующих строках значения переменных для каждого наблюдения (пациента, измерения и т.д.), также разделенных запятой.

Импорт данных осуществляется с помощью функции

```
read.csv ("название файла")
```

При этом данные файла будут выведены на рабочее пространство.

Данные файла можно присвоить какой-нибудь переменной:

```
data = read.csv ("название файла")
```

В этом случае данные на рабочее пространство не будут выведены.

Если теперь написать командой имя переменной, в которую мы записали данные из файла, то они будут выведены на экран.

8. Вопросы по теме занятия

1. Что такое статистика?
2. Что такое генеральная совокупность?
3. Что такое выборка?
4. Как нужно производить выборку из генеральной совокупности?
5. Какие шкалы используются для измерения признаков?
6. Перечислите свободное программное обеспечение для статистического анализа
 - 1) R, PSPP, Gnuplot;

9. Тестовые задания по теме с эталонами ответов

1. ШКАЛА ИЗМЕРЕНИЯ — ЭТО:
 - 1) ограничение типа отношений между значениями переменных, накладываемое на результаты измерений;
 - 2) совокупность измеренных признаков;
 - 3) полная совокупность элементов;
 - 4) часть совокупности элементов, которая охватывается экспериментом;
 - 5) функция, которая однозначно определяет вероятность;
2. СТАТИСТИЧЕСКОЕ ИССЛЕДОВАНИЕ - ЭТО:
 - 1) анализ и обработка, как правило, большого количества элементов, иногда достигающее до десятков и сотен миллионов;
 - 2) способ научного познания сущности психических явлений и их закономерностей;
 - 3) система логических и последовательных методологических, методических и организационно-технических процедур в социологии для получения научных знаний о социальных явлениях, а также их процедур, процессов;
 - 4) форма бизнес-исследования и направление прикладной социологии, которое фокусируется на понимании поведения, желаний и предпочтений потребителей, конкурентов и рынков в диктуемой рынком экономике;
 - 5) проверка одного человека или группы людей, которые, хотя, вероятно, являются здоровыми, но принадлежат к группе риска развития какой-либо определенной болезни;
3. С УВЕЛИЧЕНИЕМ ОБЪЕМА ВЫБОРКИ ОШИБКА РЕПРЕЗЕНТАТИВНОСТИ:
 - 1) уменьшается;
 - 2) остается постоянной;
 - 3) зависит от выборки;
 - 4) сначала уменьшается, потом увеличивается;
 - 5) увеличивается;
4. ШКАЛЫ ИЗМЕРЕНИЯ, КОТОРЫЕ РАССМАТРИВАЛ С. СТИВЕНС:
 - 1) номинативная, порядковая, интервальная, шкала отношений;
 - 2) номинативная, порядковая, интегральная;
 - 3) балльная, порядковая, интервальная, шкала отношений;
 - 4) номинативная, порядковая, интервальная, шкала связей;
 - 5) численная, порядковая, интервальная, шкала отношений;
5. РОСТ ПРИЗЫВНИКОВ (167, 180, 190 .CM) ОТНОСЯТСЯ К СЛЕДУЮЩЕМУ ТИПУ ДАННЫХ:
 - 1) количественному;
 - 2) качественному;

- 3) дискретному;
 - 4) логическому;
 - 5) вариантному;
6. ЯЗЫК ПРОГРАММИРОВАНИЯ ДЛЯ СТАТИСТИЧЕСКОЙ ОБРАБОТКИ ДАННЫХ И РАБОТЫ С ГРАФИКОЙ, А ТАКЖЕ СВОБОДНАЯ ПРОГРАММНАЯ СРЕДА ВЫЧИСЛЕНИЙ С ОТКРЫТЫМ ИСХОДНЫМ КОДОМ В РАМКАХ ПРОЕКТА GNU, РАЗРАБОТАННЫЙ РОССОМ АЙХЭКОЙ И РОБЕРТОМ ДЖЕНТЛМЕНОМ:
- 1) R;
 - 2) GNUPLLOT;
 - 3) PSPP;
 - 4) S;
 - 5) SPSS;

10. Ситуационные задачи по теме с эталонами ответов

1. В качестве одной из переменной исследования вы рассматриваете Национальность (1-русский, 2-украинец, 3-казах и т.д.).

Вопрос 1: К какому типу данных относится эта переменная?;

Вопрос 2: В чем особенность этой шкалы?;

- 1) Категориальные (качественные) данные;
 - 2) Номинативная шкала (номинальная, шкала наименований) – состоит в присваивании какому-либо свойству или признаку определенного обозначения или символа. При измерении в этой шкале осуществляется классификация или распределение на непересекающиеся классы. С величинами, измеряемыми в шкале наименований, можно выполнять только одну операцию — проверку их совпадения или несовпадения. При построении шкалы наименований главными являются качественные различия, а количественные не принимаются во внимание. Поэтому числа, используемые в качестве обозначений классов эквивалентности в этих шкалах, не отражают количественных различий выраженности изучаемого признака.;
2. В качестве одной из переменной исследования вы рассматриваете место, занятое на соревновании (1,2, 3)

Вопрос 1: К какому типу данных относится эта переменная?;

Вопрос 2: В чем особенность этой шкалы?;

- 1) Ранговые данные;
 - 2) Порядковая шкала (или ранговая) классифицирует совокупность измеренных признаков по принципу «больше-меньше», «выше-ниже», «сильнее-слабее»;
3. PSPP - это свободное программное обеспечение для статистического анализа данных. PSPP позволяет вычислять описательные статистики, производить проверку статистических гипотез, строить уравнения линейной и логистической регрессии, производить факторный и кластерный анализ, использовать непараметрические статистические критерии и т.д.

Вопрос 1: Какой пользовательский интерфейс имеет программа PSPP?;

Вопрос 2: В формате каких файлов могут быть представлены отчеты PSPP?;

- 1) PSPP имеются как графический пользовательский интерфейс, так и традиционный интерфейс командной строки.;
- 2) Отчеты могут быть представлены в форматах файлов: txt, Post Script, pdf, Open Document, html.;

11. Примерная тематика НИРС по теме

1. Обзор свободного программного обеспечения для статистического анализа данных
2. Обзор платного программного обеспечения для статистического анализа данных
3. Способы измерения различных данных в медицинском исследовании

12. Рекомендованная литература по теме занятия

- **обязательная:**

Медик, В. А. [Математическая статистика в медицине](#) : учебное пособие для вузов : в 2 т. / В. А. Медик, М. С. Токмачев. - 2-е изд., перераб. и доп. - Москва : Юрайт, 2023. - Т. 1. - 471 с. - Текст : электронный.

Медик, В. А. [Математическая статистика в медицине](#) : учебное пособие для вузов : в 2 т. / В. А. Медик, М. С. Токмачев. - 2-е изд. - Москва : Юрайт, 2023. - Т. 2. - 347 с. - Текст : электронный.

- **дополнительная:**

Наркевич, А. Н. [Статистические методы исследования в медицине и биологии](#) : учеб. пособие / А. Н. Наркевич, К. А. Виноградов, К. В. Шадрин ; Красноярский медицинский университет. - Красноярск : КрасГМУ, 2018. - 109 с. - Текст : электронный.

Малугин, В. А. [Математическая статистика](#) : учебное пособие для вузов / В. А. Малугин. - Москва : Юрайт, 2023. - 218 с. - Текст : электронный.

[Основы статистического анализа в медицине](#) : учебное пособие / ред. А. В. Решетников. - Москва : Медицинское информационное агентство, 2020. - 176 с. - Текст : электронный.

- **электронные ресурсы:**

База знаний WolframAlpha (<https://www.wolframalpha.com/examples/Statistics.html>)

Дистанционные курсы по статистике, анализу данных (<https://welcome.stepik.org/ru>)

Онлайн-калькулятор для статистики (http://math.semestr.ru/group/group_manual.php)

1. Тема № 2. Описательные статистики (в интерактивной форме)

2. Разновидность занятия: комбинированное

3. Методы обучения: объяснительно-иллюстративный, частично-поисковый (эвристический)

4. Значение темы (актуальность изучаемой проблемы): Описательная статистика занимает ключевое место в статистическом анализе, поскольку она занимается обработкой эмпирических данных, их систематизацией, наглядным представлением в форме графиков и таблиц, а также их количественным описанием посредством основных статистических показателей.

5. Цели обучения

- **обучающийся должен знать** методы анализа медико-статистической информации и интерпретации результатов состояния здоровья пациента (населения), **уметь** анализировать полученные в результате вычислений статистические величины (абсолютные, средние, относительные показатели, показатели динамического ряда) на основании сравнения в динамике и с имеющимися нормативами, представлять данные с помощью программного обеспечения, **владеть** навыками анализа полученных в результате вычислений статистических величин, навыками представления и интерпретирования данных с помощью программного обеспечения

6. Место проведения и оснащение занятия:

- **место проведения занятия:** компьютерный класс №1 (З-03)

- **оснащение занятия:** видеопроектор, доска магнитно-маркерная, комплект учебной мебели на посадочные места, локальный сетевой сервер, персональные компьютеры, экран

7. Аннотация (краткое содержание темы)

Описательная статистика - статистика, которая занимается обработкой эмпирических (первичных) данных, их систематизацией, наглядным представлением в форме графиков и таблиц, а также их количественным описанием посредством основных статистических показателей.

Описательная статистика - первый шаг анализа данных, она не предоставляет сведения и закономерности, присущие всей генеральной совокупности, она просто предоставляет факты и расчеты, основанные на собранных (полученных в ходе наблюдения/эксперимента/т.д.) данных.

Описательная статистика предполагает:

- Графическое представление данных.
- Табличное представление данных.
- Вычисление статистических показателей.

Статистические показатели разделяют на две группы: меры центральной тенденции и меры рассеяния.

Мера центральной тенденции — число, служащее для описания множества значений числом (одним). К таким показателям относят:

- Арифметическое среднее - сумма всех наблюдаемых значений, делённая на их количество.
- Взвешенная средняя арифметическая - среднее значение, учитывающее весовые коэффициенты для каждого значения.
- Мода - наиболее часто встречающаяся величина признака в данной совокупности.
- **Медиана**, или срединное значение наблюдаемой совокупности, - это величина варьирующего признака, которая находится в середине ряда, расположенного в порядке возрастания или убывания.
- Квантиль в математической статистике — такое число, что заданная случайная величина не превышает его лишь с фиксированной вероятностью.
- Минимум - число, которое не больше всех остальных признаков совокупности.
- Максимум - число, которое не меньше всех остальных признаков совокупности.

Меры рассеяния

- Размах вариации — разность между наибольшим и наименьшим значениями варьирующего признака
- **Дисперсия случайной величины** — мера разброса данной случайной величины, т.е. ее отклонения от среднего
- **Среднее квадратическое отклонение** вычисляется как корень из дисперсии
- **Квартильный размах** — это интервал, в котором вокруг медианы сосредоточилось 50% респондентов. Он равен разности значений 75-й квантили и 25-й квантили

8. Вопросы по теме занятия

1. Что такое описательная статистика?
2. Перечислите статистические показатели, относящиеся к мерам центральной тенденции.
3. Перечислите статистические показатели, относящиеся к мерам рассеяния.
4. Что предполагает описательная статистика?

5. Что такое мера центральной тенденции?

9. Тестовые задания по теме с эталонами ответов

1. СЕРЕДИННАЯ, ЦЕНТРАЛЬНАЯ ВАРИАНТА, ДЕЛЯЩАЯ ВАРИАЦИОННЫЙ РЯД ПОПОЛАМ, НА ДВЕ РАВНЫЕ ЧАСТИ:

- 1) мода;
- 2) высота;
- 3) медиана;
- 4) апофема;
- 5) биссектриса;

2. КВАРТИЛЬ ДЕЛИТ РАСПРЕДЕЛЕНИЕ НА:

- 1) 6 частей;
- 2) 5 частей;
- 3) 4 части;
- 4) 3 части;
- 5) 2 части;

3. КАК ПРАВИЛО, АНАЛИЗ ДАННЫХ НАЧИНАЕТСЯ С:

- 1) вычисления описательных статистик;
- 2) вычисления выборки;
- 3) выбора порядковой шкалы;
- 4) выбора интервальной шкалы;
- 5) выбора генеральной совокупности;

4. КВАРТИЛЬНЫЙ РАЗМАХ - ЭТО:

- 1) интервал, в котором вокруг медианы сосредоточилось 50% респондентов;
- 2) интервал, в котором вокруг медианы сосредоточилось 25% респондентов;
- 3) интервал, в котором вокруг медианы сосредоточилось 75% респондентов;
- 4) интервал, в котором вокруг медианы сосредоточилось 100% респондентов;
- 5) интервал, в котором вокруг моды сосредоточилось 50% респондентов;

5. МЕДИАНА - ЭТО:

- 1) срединное значение ранжированного вариационного ряда;
- 2) разность между наибольшим и наименьшим значениями варьирующего признака;
- 3) мера разброса данной случайной величины, т.е. ее отклонения от среднего;
- 4) наиболее часто встречающаяся величина признака в данной совокупности;
- 5) объем совокупности;

10. Ситуационные задачи по теме с эталонами ответов

1. При определении степени выраженности некоторого медицинского показателя в опытной группе были получены следующие результаты. Опытная группа — 18, 15, 16, 11, 14, 15, 16, 16, 16, 22, 17, 12, 11, 12, 18, 19, 20

Вопрос 1: Чему равна дисперсия выборки?;

Вопрос 2: Чему равно среднее значение выборки?;

- 1) 9,82;
- 2) 15,76;

2. Дан ряд {3, 4, 5, 4, 4, 6, 2, 6, 7, 8, 4, 7}

Вопрос 1: Вычислите для него моду;

Вопрос 2: Вычислите медиану;

Вопрос 3: Вычислите 3 квартиль;

1) Упорядочим данный массив по возрастанию: {2, 3, 4, 4, 4, 4, 5, 6, 6, 7, 7, 8}. Модой (наиболее часто встречающееся значение в ряду данных) будет являться значение 4: $M_o = 4$;

2) Медианой (значение признака, делящее выборку на две равные части) будет является среднее значение между 4 и 4: $M_e = 4$;

3) Для определения квартили делим совокупность на 4 равных части. 3 квартиль находится между значениями 6 и 7. Ее значение равно полусумме — 6,5);

11. Примерная тематика НИРС по теме

1. Способы измерения различных данных в медицинском исследовании
2. Описательная статистика и автоматические функции электронных таблиц
3. Сводные статистические характеристики

12. Рекомендованная литература по теме занятия

- **обязательная:**

Медик, В. А. [Математическая статистика в медицине](#) : учебное пособие для вузов : в 2 т. / В. А. Медик, М. С. Токмачев. - 2-е изд., перераб. и доп. - Москва : Юрайт, 2023. - Т. 1. - 471 с. - Текст : электронный.

Медик, В. А. [Математическая статистика в медицине](#) : учебное пособие для вузов : в 2 т. / В. А. Медик, М. С. Токмачев. - 2-е изд. - Москва : Юрайт, 2023. - Т. 2. - 347 с. - Текст : электронный.

- **дополнительная:**

Наркевич, А. Н. [Статистические методы исследования в медицине и биологии](#) : учеб. пособие / А. Н. Наркевич, К. А.

Виноградов, К. В. Шадрин ; Красноярский медицинский университет. - Красноярск : КрасГМУ, 2018. - 109 с. - Текст : электронный.

Малугин, В. А. [Математическая статистика](#) : учебное пособие для вузов / В. А. Малугин. - Москва : Юрайт, 2023. - 218 с. - Текст : электронный.

[Основы статистического анализа в медицине](#) : учебное пособие / ред. А. В. Решетников. - Москва : Медицинское информационное агентство, 2020. - 176 с. - Текст : электронный.

- электронные ресурсы:

База знаний WolframAlpha (<https://www.wolframalpha.com/examples/Statistics.html>)

Дистанционные курсы по статистике, анализу данных (<https://welcome.stepik.org/ru>)

Онлайн-калькулятор для статистики (http://math.semestr.ru/group/group_manual.php)

1. Тема № 3. Организация статистического исследования

2. **Разновидность занятия:** комбинированное

3. **Методы обучения:** объяснительно-иллюстративный, частично-поисковый (эвристический)

4. **Значение темы** (актуальность изучаемой проблемы): В настоящее время во многих областях жизни резко возросла потребность в компьютерной статистической обработке различных данных. Кроме традиционного использования в диссертационных работах, статистика стала интенсивно применяться в студенческих курсовых и дипломных работах, маркетинговых исследованиях в области здравоохранения. Для статобработки используются разнообразные компьютерные программы. Одной из самых распространенных программ является Statistica for Windows, часто используется MicroSoft Excel

5. Цели обучения

- **обучающийся должен знать** способы получения информации из различных информационных, библиографических ресурсов, современные методики сбора и обработки информации, необходимой для проведения научного исследования, теоретические основы и технологии организации научно-исследовательской и проектной деятельности, в том числе особенности проведения конкурсов российскими и международными научными фондами; требования к оформлению проектных и исследовательских работ, конкурсной документации, **уметь** осуществлять поиск профессиональной, достоверной, качественной информации в сети интернет с использованием специализированных научных и медицинских библиотек и баз данных, организовывать проведение научных конференций, выставок, конкурсов профессионального мастерства, иных конкурсов и мероприятий в условиях цифровой экономики; организовывать научно-исследовательские, проектные работы, консультировать участников работы на всех этапах ее проведения; использовать отечественный и зарубежный опыт и результаты собственных научных исследований в процессе руководства научно-исследовательской, проектной работой, применять современные методики сбора и обработки информации, необходимой для проведения научного исследования, **владеть** навыками поиска информации, в том числе и медицинской, в сети интернет, научно-методическими основами организации научно-исследовательской, проектной деятельности; навыками оценивания качества выполнения и оформления проектных, научно-исследовательских работ с применением цифровых технологий, современными методиками сбора и обработки информации, необходимой для проведения научного исследования

6. Место проведения и оснащение занятия:

- **место проведения занятия:** компьютерный класс №1 (З-03)

- **оснащение занятия:** видеопроектор, доска магнитно-маркерная, комплект учебной мебели на посадочные места, локальный сетевой сервер, персональные компьютеры, экран

7. Аннотация (краткое содержание темы)

Исследование («следование изнутри»):

- в предельно широком смысле — поиск новых знаний или систематическое расследование с целью установления фактов;
- в более узком смысле *исследование* — научный метод(процесс) изучения чего-либо;
- результат такого действия (*исследования*), научный труд, документ с описанием изученного объекта или чего то.

Этапы научного исследования

Любое конкретное исследование может быть представлено в виде ряда этапов.

- Выбор темы исследования.
- Определение объекта и предмета исследования.
- Определение цели и задач.
- Формулировка названия работы.
- Разработка гипотезы.
- Составление плана исследования.
- Работа с литературой.
- Выбор методов исследования.
- Организация условий проведения исследования.
- Проведение исследования (сбор материала).
- Обработка результатов исследования.
- Формулирование выводов.
- Оформление работы.

Выбор темы исследования

Научное исследование всегда предполагает решение какой-либо научной проблемы. Недостаточность знаний, фактов, противоречивость научных представлений создают основания для проведения научного исследования. Постановка научной проблемы предполагает:

- обнаружение существования такого дефицита;
- осознание потребности в устранении дефицита;
- формулирование проблемы.

Объект и предмет исследования

Объект исследования – это то, что будет взято для изучения и исследования. Объектом исследования может быть процесс или явление действительности. Обычно название объекта исследования содержится в ответе на вопрос: **что рассматривается?**

Предмет— это особая проблема, отдельные стороны объекта, его свойства и особенности, которые, не выходя за рамки исследуемого объекта, будут исследованы в работе. Предмет исследования более конкретен и включает только те связи и отношения, которые подлежат непосредственному изучению в данной работе. Обычно предмет исследования содержится в ответе на вопрос: **что изучается?**

Из сказанного следует, что объектом выступает то, что исследуется, а предметом — то, что в этом объекте получает научное объяснение. Именно предмет исследования определяет тему исследования.

Цели и задачи

Исходя из объекта и предмета можно приступить к определению цели и задач исследования. **Цель формулируется кратко и предельно точно**, в смысловом отношении выражая, то основное, что намеревается сделать исследователь, к какому конечному результату он стремится. Цель конкретизируется иррационально в задачах исследования.

Задач ставится несколько, и каждая из них четкой формулировкой раскрывает ту сторону темы, которая подвергается изучению. Определяя задачи, необходимо учитывать их взаимную связь. Иногда невозможно решить одну задачу, не решив предварительно другую. Каждая поставленная задача должна иметь решение, отраженное в одном или нескольких выводах.

- **Первая задача**, как правило, связана с выявлением, уточнением, углублением, методологическим обоснованием сущности, структуры изучаемого объекта.
- **Вторая** связана с анализом реального состояния предмета исследования.
- **Третья задача** связана с преобразованиями предмета исследования, т. е. выявлением путей и средств повышения эффективности совершенствования исследуемого явления или процесса.
- **Четвертая** — с опытно-экспериментальной проверкой эффективности предлагаемых преобразований.

Задачи следует формулировать четко и лаконично. Как правило, каждая задача формулируется в виде поручения: «Изучить...», «Разработать...», «Выявить...», «Установить...», «Обосновать...», «Определить...», «Проверить...», «Доказать...» и т.п.

Формулировка названия работы

Определив тему и конкретные задачи, уточнив объект и предмет исследования, можно дать первый вариант формулировки названия работы. Название работы рекомендуется формулировать по возможности кратко, точно в соответствии с ее содержанием. Необходимо помнить, что в названии должен быть отражен предмет исследования.

Разработка гипотезы

Гипотеза — познавательная конструкция, включающая как предположение, так и способы его фальсификации; обобщающая, прогнозирующая получение определенных новых фактов и систематизирующая их.

Любая гипотеза рассматривается как первоначальная канва и отправная точка для исследований, которая может подтвердиться или не подтвердиться.

План исследования

План исследования представляет собой намеченную программу действий, которая включает все этапы работы с

определением календарных сроков их выполнения. План необходим для того, чтобы правильно организовать работу и придать ей более целеустремленный характер. Кроме того, он дисциплинирует, заставляет работать в определенном ритме.

В процессе работы первоначальный план можно детализировать, пополнять и изменять.

Работа с литературой

Место данного этапа работы определено условно, поскольку реально работа с литературой начинается в процессе выбора темы и продолжается до конца исследования. Эффективность работы с литературными источниками зависит от знания определенных правил их поиска, соответствующей методики изучения и конспектирования. Под «литературным источником» понимается документ, содержащий какую-либо информацию (монография, статья, тезисы, книга, диссертации и т. п.).

Подбор исследуемых

Любое исследование в конечном счете является сравнительным. Сравнить можно результаты экспериментальной группы с результатами контрольной группы.

Можно сравнивать и результаты «сегодняшних» исследований с результатами, которые были получены на тех же людях, но раньше. Правда, в этом случае потребуются фактический материал результативности предшествующей деятельности, который позволил бы провести статистические расчеты на достоверность различий.

Наконец, сравнивать можно результаты, полученные на данной группе людей, с теми стандартами, которые существуют в науке (например, сравнивать уровень физического развития 10-летних детей «своей» школы со стандартами, характеризующими физическое развитие детей этого возраста в определенных географических и климатических зонах страны).

Экспериментатору нужно решить **две задачи: кого выбрать** в качестве исследуемых и **сколько их надо выбрать**.

Выбор методов исследования

Метод исследования – это способ получения сбора, обработки или анализа данных. В исследованиях широко применяются различные методы научного познания из других областей науки и техники. С одной стороны, это явление можно считать положительным, так как оно дает возможность изучить исследуемые вопросы комплексно, рассмотреть многообразие связей и отношений, с – другой это разнообразие затрудняет выбор методов, соответствующих конкретному исследованию.

Основным ориентиром для выбора методов исследования могут служить его задачи. Именно задачи, поставленные перед работой, определяют способы их разрешения, а стало быть, и выбор соответствующих методов исследования. При этом важно подбирать такие методы, которые были бы адекватны своеобразию изучаемых явлений.

В практике проведения исследований, направленных на решение задач того или иного процесса или явления, наибольшее распространение получили следующие методы:

- анализ научно-методической литературы, документальных и архивных материалов;
- опрос (беседа, интервью и анкетирование);
- контрольные испытания (тестирование);
- хронометрирование;
- экспертное оценивание;
- педагогическое наблюдение;
- педагогический эксперимент;
- методы математической обработки.

Перечисленные группы методов тесно связаны между собой. Они не могут применяться изолированно. Например, для проведения наблюдения или эксперимента необходимо предварительно получить информацию о том, что уже есть в практике и теории изучаемого процесса или явления, т. е. воспользоваться методами анализа научно-методической литературы или опроса. Полученный в процессе исследования фактический материал не будет достоверен без методов математической обработки.

Организация условий проведения исследования

Организация педагогического эксперимента связана с планированием его проведения, которое определяет последовательность всех этапов работы, а также с подготовкой всех условий, обеспечивающих полноценное исследование. Сюда входят подготовка соответствующей обстановки, приборов, средств, инструктаж помощников, планирование наблюдения, выбор экспериментальных и контрольных групп, оценка всех особенностей экспериментальной базы и т.д.

Проведение исследования

На этом этапе работы с помощью выбранных методов исследования собирают необходимые эмпирические данные для проверки выдвинутой гипотезы.

Обработка результатов исследования

Первичная обработка данных. Результаты каждого исследования важно обрабатывать по возможности тотчас же по его окончании, пока память экспериментатора может подсказать те детали, – которые почему-либо не зафиксированы, но представляют интерес для понимания существа дела. При обработке собранных данных может оказаться, что их или недостаточно, или они противоречивы и поэтому не дают оснований для окончательных выводов. В таком случае исследование необходимо продолжить, внося в него требуемые дополнения.

В большинстве случаев обработку целесообразно начать с составления таблиц (сводных таблиц) полученных данных.

И для ручной, и для компьютерной обработки в исходную сводную таблицу чаще всего заносят начальные данные. В последнее время преимущественной формой математико-статистической обработки стала компьютерная, поэтому в таблицу целесообразно внести все интересующие вас признаки в форме десятичного числа, т.е. предварительно пересчитать минуты в десятичные доли часа, секунды – в десятичные доли минуты, количество месяцев – в десятичную долю года и т. д. Это необходимо, поскольку формат данных для большинства используемых компьютерных программ накладывает свои ограничения.

Математическая обработка данных. Для определения способов математико-статистической обработки, прежде всего, необходимо оценить характер распределения по всем используемым параметрам. Для параметров, имеющих нормальное распределение или близкое к нормальному, можно использовать методы параметрической статистики, которые во многих случаях являются более мощными, чем методы непараметрической статистики. Достоинством последних является то, что они позволяют проверять статистические гипотезы независимо от формы распределения.

Важнейшими статистическими характеристиками являются:

- а) средняя арифметическая
- б) среднее квадратическое отклонение
- в) коэффициент вариации

Ориентируясь на эти характеристики нормального распределения, можно оценить степень близости к нему рассматриваемого распределения.

Одной из наиболее часто встречающихся задач при обработке данных является оценка достоверности различий между двумя или более рядами значений. В математической статистике существует ряд способов для ее решения. Компьютерный вариант обработки данных стал в настоящее время наиболее распространенным. Во многих прикладных статистических программах есть процедуры оценки различий между параметрами одной выборки или разных выборок. При полностью компьютеризованной обработке материала нетрудно в нужный момент использовать соответствующую процедуру и оценить интересующие различия.

Формулирование выводов

Выводы – это утверждения, выражающие в краткой форме содержательные итоги исследования, они в тезисной форме отражают то новое, что получено самим автором. Частой ошибкой является то, что автор включает в выводы общепринятые в науке положения – уже не нуждающиеся в доказательствах.

Решение каждой из перечисленных во введении задач должно быть определенным образом отражено в выводах.

Оформление работы

Основанная задача данного этапа работы представить полученные результаты в общедоступной и понятной форме, позволяющей сравнивать их с результатами других исследователей и использовать в практической деятельности. Поэтому оформление работы должно соответствовать требованиям, предъявляемым к работам, направляемым в печать.

(Волохова Е. С. Основные этапы научного исследования // Молодой ученый. — 2016. — №6. — С. 755-757. — URL <https://moluch.ru/archive/110/26991/>).

8. Вопросы по теме занятия

1. Что такое научное исследование?
2. Перечислите основные этапы научного исследования
 - 1) Любое конкретное исследование может быть представлено в виде ряда этапов. Выбор темы исследования. Определение объекта и предмета исследования. Определение цели и задач. Формулировка названия работы. Разработка гипотезы. Составление плана исследования. Работа с литературой. Выбор методов исследования. Организация условий проведения исследования. Проведение исследования (сбор материала). Обработка результатов исследования. Формулирование выводов. Оформление работы.;
3. Что включает в себя этап "Работа с литературой"?
4. Что включает в себя этап "формулирование выводов"?
5. Что включает в себя этап "Оформление работы"?

9. Тестовые задания по теме с эталонами ответов

1. ШКАЛА ИЗМЕРЕНИЯ — ЭТО:
 - 1) ограничение типа отношений между значениями переменных, накладываемое на результаты измерений;
 - 2) совокупность измеренных признаков;
 - 3) полная совокупность элементов;
 - 4) часть совокупности элементов, которая охватывается экспериментом;
 - 5) функция, которая однозначно определяет вероятность;
2. СТАТИСТИЧЕСКОЕ ИССЛЕДОВАНИЕ - ЭТО:
 - 1) анализ и обработка, как правило, большого количества элементов, иногда доходящее до десятков и сотен миллионов;
 - 2) способ научного познания сущности психических явлений и их закономерностей;
 - 3) система логических и последовательных методологических, методических и организационно-технических процедур в социологии для получения научных знаний о социальных явлениях, а также их процедур, процессов;
 - 4) форма бизнес-исследования и направление прикладной социологии, которое фокусируется на понимании поведения, желаний и предпочтений потребителей, конкурентов и рынков в диктуемой рынком экономике;
 - 5) проверка одного человека или группы людей, которые, хотя, вероятно, являются здоровыми, но принадлежат к группе риска развития какой-либо определенной болезни;
3. С УВЕЛИЧЕНИЕМ ОБЪЕМА ВЫБОРКИ ОШИБКА РЕПРЕЗЕНТАТИВНОСТИ:
 - 1) уменьшается;
 - 2) остается постоянной;
 - 3) зависит от выборки;
 - 4) сначала уменьшается, потом увеличивается;
 - 5) увеличивается;
4. РОССИЙСКАЯ НАУЧНАЯ ЭЛЕКТРОННАЯ БИБЛИОТЕКА, ИНТЕГРИРОВАННАЯ С РОССИЙСКИМ ИНДЕКСОМ НАУЧНОГО ЦИТИРОВАНИЯ:
 - 1) Krasgmu;
 - 2) Elibrary;
 - 3) Google Scholar;
 - 4) PubMed;
 - 5) Medscape;
5. ОБЪЕКТ ИССЛЕДОВАНИЯ - ЭТО:
 - 1) то, что будет взято для изучения и исследования;
 - 2) особая проблема, отдельные стороны объекта, его свойства и особенности, которые, не выходя за рамки исследуемого объекта, будут исследованы в работе;
 - 3) познавательная конструкция, включающая как предположение, так и способы его фальсификации; обобщающая, прогнозирующая получение определенных новых фактов и систематизирующая их;
 - 4) способ получения сбора, обработки или анализа данных. В исследованиях широко применяются различные методы научного познания из других областей науки и техники;
 - 5) утверждения, выражающие в краткой форме содержательные итоги исследования, они в тезисной форме отражают то новое, что получено самим автором;

10. Ситуационные задачи по теме с эталонами ответов

1. Интернет на сегодняшний день представляет собой огромное скопление разнообразной информации, значительная часть которой является не достоверной

Вопрос 1: Расскажите, какие сайты в Интернете содержат достоверную медицинскую информацию?;

Вопрос 2: Почему этим сайтам можно доверять?;

Вопрос 3: Найдите в Интернете не менее 6-ти профессиональных медицинских новостей за последний месяц;

- 1) Среди сайтов, предоставляющих доступ к базам данных, безусловным лидером является всем известный PubMed MEDLINE (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=PubMed>). На сайте Medscape (<http://www.medscape.com/>) представлено большое количество журналов, доступ к которым предоставляется бесплатно после заполнения регистрационной формы. Портал Medbioworld (<http://www.medbioworld.com/>) является одной из наиболее полных коллекций ссылок на источники профессиональной медицинской информации в сети Интернет. Данный ресурс содержит более 25000 ссылок на медицинские журналы, профессиональные медицинские ассоциации, медицинские словари, нозологические базы данных, клинические испытания, руководства;
- 2) Им можно доверять, так как они являются специализированными порталами, предоставляющими профессиональную медицинскую информацию, которым доверяет весь мир;
- 3) Для поиска профессиональных медицинских новостей можно воспользоваться порталом Medbioworld (<http://www.medbioworld.com/>). На главной странице представлена ссылка на профессиональные медицинские новости. . Зайти по ссылке и ознакомиться с последними новостями текущего месяца. News. Professional Medical News. Professional Medical News. Health eLineHealth eLine. 2011-01-07. • Moxifloxacin beats clindamycin for some odontogenic infections. • Newer antipsychotics overused, U.S. study suggests. • First-in-class antibiotic has high risk of adverse events and death;

2. Темой вашего исследования является "Профилактика простудных заболеваний среди детей дошкольного возраста закаливанием"

Вопрос 1: Определите объект исследования;

Вопрос 2: Определите предмет исследования;

- 1) простудные заболевания среди детей дошкольного возраста;
- 2) закаливание как средство профилактики простудных заболеваний;

11. Примерная тематика НИРС по теме

1. Непараметрические критерии согласия
2. Критерий омега-квадрат (фон Мизеса)
3. Методика проверки статистических гипотез

12. Рекомендованная литература по теме занятия

- обязательная:

Медик, В. А. [Математическая статистика в медицине](#) : учебное пособие для вузов : в 2 т. / В. А. Медик, М. С. Токмачев. - 2-е изд., перераб. и доп. - Москва : Юрайт, 2023. - Т. 1. - 471 с. - Текст : электронный.

Медик, В. А. [Математическая статистика в медицине](#) : учебное пособие для вузов : в 2 т. / В. А. Медик, М. С. Токмачев. - 2-е изд. - Москва : Юрайт, 2023. - Т. 2. - 347 с. - Текст : электронный.

- дополнительная:

Наркевич, А. Н. [Статистические методы исследования в медицине и биологии](#) : учеб. пособие / А. Н. Наркевич, К. А. Виноградов, К. В. Шадрин ; Красноярский медицинский университет. - Красноярск : КрасГМУ, 2018. - 109 с. - Текст : электронный.

Малугин, В. А. [Математическая статистика](#) : учебное пособие для вузов / В. А. Малугин. - Москва : Юрайт, 2023. - 218 с. - Текст : электронный.

[Основы статистического анализа в медицине](#) : учебное пособие / ред. А. В. Решетников. - Москва : Медицинское информационное агентство, 2020. - 176 с. - Текст : электронный.

- электронные ресурсы:

База знаний WolframAlpha (<https://www.wolframalpha.com/examples/Statistics.html>)

Дистанционные курсы по статистике, анализу данных (<https://welcome.stepik.org/ru>)

Онлайн-калькулятор для статистики (http://math.semestr.ru/group/group_manual.php)

1. Тема № 4. Визуализация статистических данных (в интерактивной форме)

2. Разновидность занятия: комбинированное

3. Методы обучения: объяснительно-иллюстративный, исследовательский

4. Значение темы (актуальность изучаемой проблемы): Для многих наборов статистических данных существует необходимость в представлении информации в максимально наглядной форме. К таким средствам визуализации относятся R, MS Excel.

5. Цели обучения

- **обучающийся должен знать** способы представления статистических данных, **уметь** представлять данные с помощью программного обеспечения, **владеть** навыками представления и интерпретирования данных с помощью программного обеспечения

6. Место проведения и оснащение занятия:

- **место проведения занятия:** компьютерный класс №1 (3-03)

- **оснащение занятия:** видеопроектор, доска магнитно-маркерная, комплект учебной мебели на посадочные места, локальный сетевой сервер, персональные компьютеры, экран

7. Аннотация (краткое содержание темы)

ВИЗУАЛИЗАЦИЯ ДАННЫХ

Визуализация данных - это отображение больших массивов числовой и семантической информации в виде графических объектов.

Истоки представления данных в виде таблиц, диаграмм и карт прослеживаются с древнейших времён. Ощутимая потребность в качественном представлении информации стала возникать в эпоху Возрождения, с появлением больших количеств данных и визуальной информации из географии, астрономии, геометрии, статистики и других наук.

В первой половине XIX века наблюдался значительный рост работ, в которых использовалось графическое отображение данных. К середине века были изобретены все основные типы представления данных: столбчатые и круговые диаграммы, гистограммы, линейные графики, графики временных рядов, контурные диаграммы и т. д.

Тенденция роста пошла на спад в начале XX века, уступив место точной математике. Тем не менее, именно в этот период стали появляться учебники и курсы по графическим методам представления данных, а сами графики стали использоваться не только для представления результатов, но и для исследования информации и выдвижения гипотез в астрономии, физике, биологии и других науках.

Новый виток визуализация получила в третьей четверти XX века. Этому способствовали три события:

- Появление работы Джона Тьюки, посвящённой разведочному анализу данных.
- Появление книги Жака Бертена (*Jacques Bertin*) «Графическая семиология» (фр. *Sémiologie graphique*)
- Возможность визуализации данных с помощью вычислительных машин: появление эффективных средств вывода (перьевых графопостроителей, графических терминалов), а также эргономичных средств ввода данных в компьютер (кодирующего планшета, мыши).

КЛАССИФИКАЦИЯ

По цели представления данных визуализация делится на презентационную (англ. «presentation», «explanation») и исследовательскую (англ. «exploration»). Презентационная визуализация предназначена для представления данных некоторой аудитории (например, в рамках научной работы, доклада или аналитического обзора в новостях). Исследовательская визуализация предназначена для анализа и обработки набора данных, например, с целью обнаружения закономерностей в них.

Существуют также гибридные презентационно-исследовательские формы визуализации данных. В этом случае целью является всё та же презентация заложенной информации, однако человеку предоставляется возможность подробно изучать показываемый набор данных посредством интерактивных элементов, например, накладывая какие-либо ограничения на данные.

Визуализация как этап анализа данных

Подсистема визуализации данных является важной составной частью качественных систем интеллектуального анализа данных, особенно ориентированных на обработку больших объёмов информации. В системах бизнес-аналитики визуализация может использоваться на всех этапах процесса обработки данных:

- Визуализация исходных данных. Этот этап полезен для оценки степени соответствия ожиданиям и пригодности данных к анализу, выдвижения гипотез о закономерностях и необходимых процедурах

первичной обработки.

- Визуализация выборки, загруженной в систему обработки.
- Визуализация результатов первичной обработки.
- Визуализация промежуточных результатов.
- Визуализация окончательных результатов.

В отличие от обычного графического интерфейса, эти средства обеспечивают:

- краткость (англ.*concision*) — способность одновременного отображения большого числа разнотипных данных;
- относительность (англ.*relativity*) и близость (англ. *proximity*) — способность демонстрировать в результатах запроса кластеры, относительные размеры групп, схожесть и различие групп, выпадающие значения (англ. *outliers*);
- концентрацию и контекст (англ.*focus with context*) — взаимодействие в некотором выбранном объектом с возможностью просмотра его положения и связей с контекстом;
- масштабируемость (англ.*zoomability*) — способность легко и быстро перемещаться между микро- и макропредставлением;
- ориентацию на «правое полушарие» — предоставление пользователю не только заранее установленных методов работы с данными (обеспечивающими его намеренные и спланированные подходы к поиску нужной информации), но и поддержка его интуитивных, импровизационных когнитивных процессов идентификации закономерностей.

ПОЛИГОН РАСПРЕДЕЛЕНИЯ

Полигон распределения (дословно - многоугольник распределения) - ломаная, которая строится в прямоугольной системе координат.



Величина признака откладывается на оси абсцисс, соответствующие частоты (абсолютные или относительные) - по оси ординат. Полученные точки соединяют отрезками прямых и получают полигон распределения.

ГИСТОГРАММА ЧАСТОТ

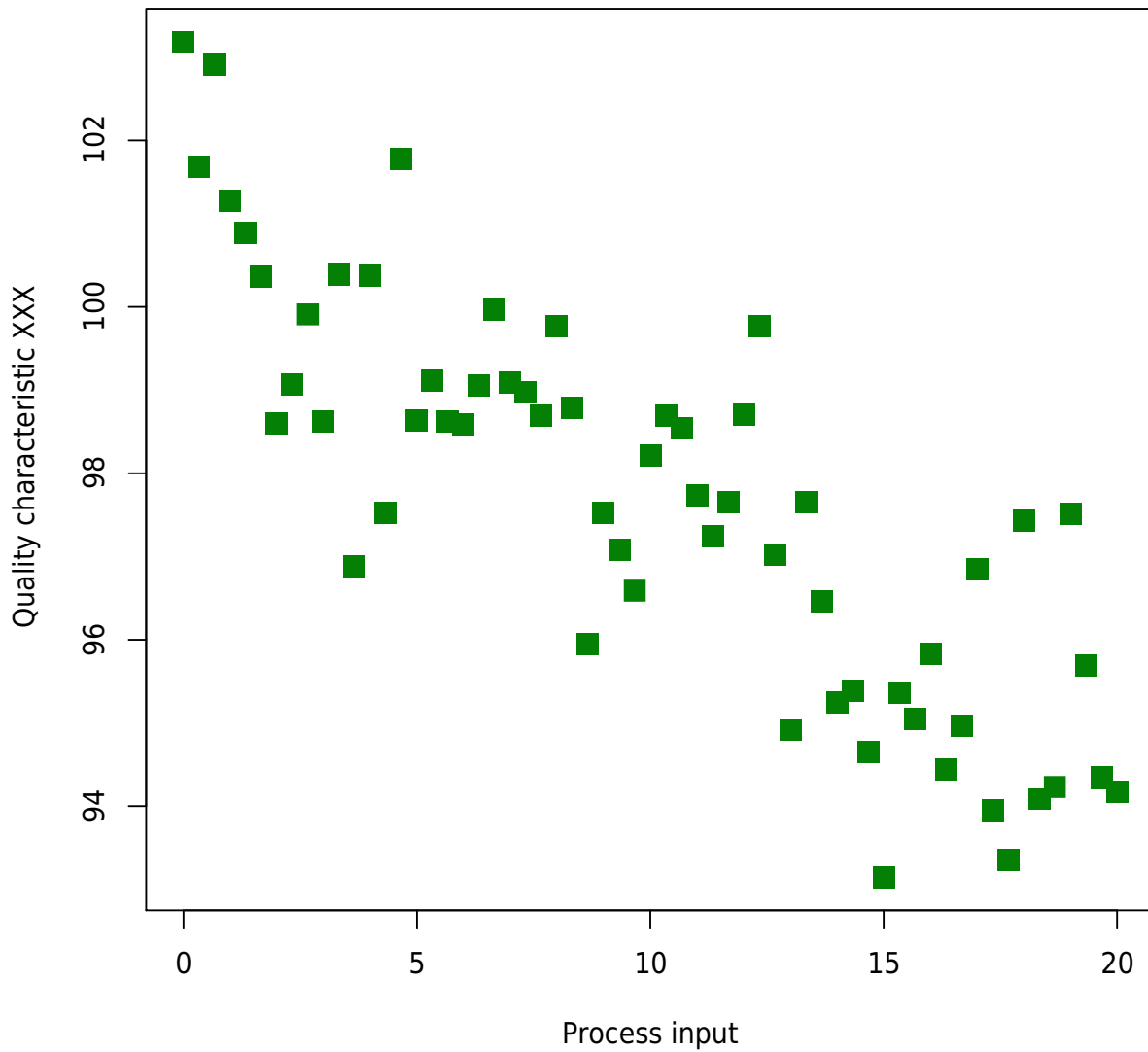
Гистограммой частот (частостей) называют ступенчатую фигуру, состоящую из прямоугольников, основания которых расположены на оси Ox и длины их равны длинам частичных интервалов (h), а высоты равны количеству элементов выборки, попавших в этот интервал.

Для построения гистограммы по горизонтальной оси откладываются границы интервалов группирования данных, а по вертикальной оси частоты попадания наблюдений в интервалах.

ДИАГРАММА РАССЕЯНИЯ

Диаграмма рассеяния — математическая диаграмма, изображающая значения двух переменных в виде точек на декартовой плоскости.

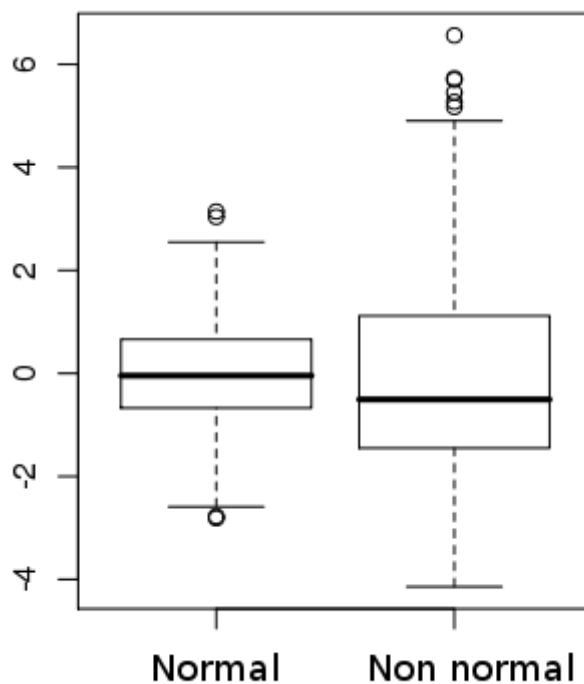
Scatterplot for quality characteristic XXX



На диаграмме рассеяния каждому наблюдению (или элементарной единице набора данных) соответствует точка, координаты которой (в декартовой системе координат) равны значениям двух каких-то параметров этого наблюдения.

ДИАГРАММА РАЗМАХА

Диаграмма размаха (англ. box-and-whiskers diagram or plot, box plot) — график, использующийся в описательной статистике, компактно изображающий одномерное распределение вероятностей.



Чтобы нарисовать ящик для одной группы про исходные данные необходимо знать всего три характеристики: Первый квартиль. Медиану. Третий квартиль. Иногда к “обязательному” набору добавляют следующие дополнительные: Минимум. Максимум. Среднее значение.

8. Вопросы по теме занятия

1. Что такое гистограмм частот?
2. Что такое Boxplot (ящичковая диаграмма, диаграмма размаха)?
 - 1) Ящичковая диаграмма — график, использующийся в описательной статистике, компактно изображающий одномерное распределение вероятностей.;
3. Что такое визуализация данных?
4. Что такое полигон частот?
5. Что такое диаграмма рассеяния?

9. Тестовые задания по теме с эталонами ответов

1. ГРАНИЦАМИ ЯЩИКА В ДИАГРАММЕ РАЗМАХА СЛУЖАТ:
 - 1) первый и третий квартили;
 - 2) первый и второй квартили;
 - 3) второй и третий квартили;
 - 4) первый и четвертый квартили;
 - 5) второй и четвертый квартили;
2. ОТОБРАЖЕНИЕ БОЛЬШИХ МАССИВОВ ЧИСЛОВОЙ И СЕМАНТИЧЕСКОЙ ИНФОРМАЦИИ В ВИДЕ ГРАФИЧЕСКИХ ОБЪЕКТОВ:
 - 1) визуализация данных;
 - 2) формулирование данных;
 - 3) вывод данных;
 - 4) формализация данных;
 - 5) семантизация данных;
3. ПО ЦЕЛИ ПРЕДСТАВЛЕНИЯ ДАННЫХ ВИЗУАЛИЗАЦИЯ ДЕЛИТСЯ НА:
 - 1) представительную и изобретательную;
 - 2) презентационную и изобретательную;
 - 3) презентационную и исследовательскую;
 - 4) представительную и исследовательскую;
 - 5) презентационную и естественную;
4. ГРАФИК ЯЩИК С УСАМИ (ДИАГРАММА РАЗМАХА, BOXPLOT, ЯЩИЧКОВАЯ ДИАГРАММА) БЫЛ РАЗРАБОТАН:
 - 1) Джоном Тьюки;
 - 2) Герман Конринг;
 - 3) Готфрид Ахенваль;
 - 4) Карл Герман;
 - 5) Уильям Петти;

5. МАТЕМАТИЧЕСКАЯ ДИАГРАММА, ИЗОБРАЖАЮЩАЯ ЗНАЧЕНИЯ ДВУХ ПЕРЕМЕННЫХ В ВИДЕ ТОЧЕК НА ДЕКАРТОВОЙ ПЛОСКОСТИ:

- 1) диаграмма рассеяния;
- 2) гистограмма;
- 3) полигон;
- 4) диаграмма размаха;
- 5) boxplot;

10. Ситуационные задачи по теме с эталонами ответов

1. Вам необходимо изобразить на диаграмме медиану, нижний и верхний квартили, минимальное и максимальное значение выборки и выбросы.

Вопрос 1: Какой тип диаграммы Вам нужно использовать?;

Вопрос 2: С помощью какой компьютерной программы можно построить такой тип диаграммы?;

1) Boxplot (ящичковая диаграмма, диаграмма размаха);

2) Можно воспользоваться программой R. Для этого необходимо воспользоваться командой: `attach(data)`, где `data` – переменная, в которой хранятся данные. Далее пишется команда: `boxplot(data)`;

2. Исследуется взаимосвязь между медицинскими показателями А и В. А: 132 122 121 118 118 115 112 111 110 104 102 100 99 98, В: 100 111 115 125 175 105 120 110 147 180 155 153 200 210

Вопрос 1: Какой тип диаграммы следует применить для визуализации имеющихся данных?;

Вопрос 2: Какую команду для построения данной диаграммы нужно ввести в R, если эти данных хранятся в переменной `data`, в столбцах А и В, соответственно;

1) Диаграмма рассеяния;

2) `plot(data$A~data$B)`;

11. Примерная тематика НИРС по теме

1. Инфографика
2. Генеративный дизайн
3. Даталогия

12. Рекомендованная литература по теме занятия

- обязательная:

Медик, В. А. [Математическая статистика в медицине](#) : учебное пособие для вузов : в 2 т. / В. А. Медик, М. С. Токмачев. - 2-е изд., перераб. и доп. - Москва : Юрайт, 2023. - Т. 1. - 471 с. - Текст : электронный.

Медик, В. А. [Математическая статистика в медицине](#) : учебное пособие для вузов : в 2 т. / В. А. Медик, М. С. Токмачев. - 2-е изд. - Москва : Юрайт, 2023. - Т. 2. - 347 с. - Текст : электронный.

- дополнительная:

Наркевич, А. Н. [Статистические методы исследования в медицине и биологии](#) : учеб. пособие / А. Н. Наркевич, К. А. Виноградов, К. В. Шадрин ; Красноярский медицинский университет. - Красноярск : КрасГМУ, 2018. - 109 с. - Текст : электронный.

Малугин, В. А. [Математическая статистика](#) : учебное пособие для вузов / В. А. Малугин. - Москва : Юрайт, 2023. - 218 с. - Текст : электронный.

[Основы статистического анализа в медицине](#) : учебное пособие / ред. А. В. Решетников. - Москва : Медицинское информационное агентство, 2020. - 176 с. - Текст : электронный.

- электронные ресурсы:

База знаний WolframAlpha (<https://www.wolframalpha.com/examples/Statistics.html>)

Дистанционные курсы по статистике, анализу данных (<https://welcome.stepik.org/ru>)

Онлайн-калькулятор для статистики (http://math.semestr.ru/group/group_manual.php)

1. **Тема № 5.** Проверка статистических гипотез (в интерактивной форме)

2. **Разновидность занятия:** комбинированное

3. **Методы обучения:** объяснительно-иллюстративный, частично-поисковый (эвристический)

4. **Значение темы** (актуальность изучаемой проблемы): Статистическая гипотеза представляет собой некоторое предположение о законе распределения случайной величины или о параметрах этого закона, формулируемое на основе выборке. Важно уметь не только подтверждать гипотезы, но и формулировать их.

5. **Цели обучения**

- **обучающийся должен знать** методы анализа медико-статистической информации и интерпретации результатов состояния здоровья пациента (населения), **уметь** формулировать проблемы и гипотезы, **владеть** навыками проверки гипотез о распределении по различным критериям

6. **Место проведения и оснащение занятия:**

- **место проведения занятия:** компьютерный класс №1 (З-03)

- **оснащение занятия:** видеопроектор, доска магнитно-маркерная, комплект учебной мебели на посадочные места, локальный сетевой сервер, персональные компьютеры, экран

7. **Аннотация** (краткое содержание темы)

СТАТИСТИЧЕСКИЕ ГИПОТЕЗЫ

Любое исследование направлено на определение некоторой характеристики или выявление связи между признаками. Полученные в экспериментах выборочные данные всегда ограничены и носят случайный характер. В исследованиях невозможно учесть все возможные влияния на элементы выборки, поэтому оценка параметров генеральной совокупности, сделанная на основании выборочных данных, всегда будет сопровождаться погрешностью, и поэтому подобного рода оценки должны рассматриваться как предположительные, а не как окончательные утверждения.

Существует два вида гипотез:

- научные;
- статистические.

Научная гипотеза — это предполагаемое решение некоторой проблемы. Она обычно формулируется в виде теоремы.

Статистическая гипотеза — некоторое утверждение относительно неизвестного параметра или какой-либо характеристики генеральной совокупности. Например, среднее значение генеральной совокупности равно 125 или коэффициент корреляции равен 0.

Для проверки статистических гипотез используются статистические критерии, которые представляют собой некоторое правило, по которому мы делаем вывод о правильности или неправильности рассматриваемой статистической гипотезы.

Проверяемую статистическую гипотезу принято называть **основной (или нулевой) гипотезой** (обозначается H_0), а противоречащую ей гипотезу — **альтернативной (или конкурирующей)** гипотезой (обозначается H_1). Нулевая гипотеза обычно заключается в предположении о том, что изучаемое вмешательство не оказывает значимого воздействия на генеральные параметры распределения. Изначально предполагается, что вмешательство не влияет. Любые различия между изучаемыми группами объясняются случайностью, и становится задача опровергнуть это предположение.

ОШИБКИ

Ошибку, заключающуюся в том, что нулевая гипотеза отвергается, тогда как она в действительности верна, называют **ошибкой первого рода**: истинная нулевая гипотеза отклонена.

Ошибку, состоящую в том, что нулевая гипотеза не отвергается, тогда как она в действительности не верна, называют **ошибкой второго рода**: ложная нулевая гипотеза не может быть отклонена.

Вероятность совершить ошибку второго рода обозначается β .

Величина $1-\beta$ называется **мощностью критерия**; она равна вероятности отвергнуть неверную гипотезу.

Чаще всего множество возможных значений критерия принадлежит некоторому интервалу. Интервалом является и критическая область. Граничные точки критической области называются критическими точками. Критические точки выбираются таким образом, чтобы при выбранном уровне значимости, а мощность критерия ($1 - \beta$) была наибольшей.

СТАТИСТИЧЕСКИЕ КРИТЕРИИ

Проверка гипотез осуществляется с помощью критериев статистической оценки различий. "Статистический критерий - это решающее правило, обеспечивающее надежное поведение, то есть принятие истинной и отклонение ложной гипотезы с высокой вероятностью" (Суходольский Г.В.)

Если «нулевая» гипотеза формулируется о равенстве ($X_1 = X_2$), то для проверки используется двусторонний критерий.

Если же «нулевая» гипотеза формулируется о неравенстве, то возможны три варианта:

- если $X_1 \neq X_2$, то используется двусторонний критерий;
- если $X_1 > X_2$ или $X_1 < X_2$, то односторонний критерий.

Мощность критерия - это его способность выявлять различия или отклонять гипотезу, если она неверна.

ПАРАМЕТРИЧЕСКИЕ КРИТЕРИИ

Параметрические критерии — это некоторые функции от параметров совокупности, они служат для проверки гипотез об этих параметрах или для их оценивания.

Параметрические критерии включают в формулу расчета параметры распределения, т.е. средние и дисперсии.

НЕПАРАМЕТРИЧЕСКИЕ КРИТЕРИИ

Непараметрические критерии — это некоторые функции от функций распределения или непосредственно от вариационного ряда наблюдавшихся значений изучаемого случайного явления.

Непараметрические критерии не включают в формулу расчета параметров распределения и основаны на оперировании частотами или рангами.

СТАТИСТИЧЕСКАЯ ЗНАЧИМОСТЬ

Уровень статистической значимости результата исследования (p-уровень) - это количественно выраженная вероятность, свидетельствующая о том, что результаты достоверны.

Надежность (достоверность) непосредственно связана с репрезентативностью выборки, т.е. с тем, насколько уверенно данные, полученные по выборке, позволяют судить о соответствующих параметрах генеральной совокупности.

Уровнем значимости называется вероятность ошибочного отклонения нулевой гипотезы.

В таблицах критических значений обычно приводятся значения критериев, соответствующих уровням значимости $p \leq 0,05$ и $p \leq 0,01$, иногда для $p \leq 0,001$.

Используя методы математической статистики, исследователь получает так называемую эмпирическую статистику (Кэмп. - коэффициент эмп.). Кэмп - это условное название.

Полученную эмпирическую статистику необходимо сравнить с двумя критическими величинами, которые соответствуют рассмотренным выше уровням статистической значимости. Критические величины для используемых коэффициентов находятся в специальных таблицах.

ОСЬ ЗНАЧИМОСТИ

Ось значимости представляет собой прямую, на левом конце которой располагается 0 и слева направо идет увеличение числового ряда. Левая зона называется зоной незначимости, правая - зона значимости, а промежуточная зоной неопределенности. Границами всех трех зон являются для $p=0,05$ и $p=0,01$.



Для принятия статистического решения нужно на оси значимости отметить эмпирическое значение коэффициента. Если Кэмп. попало в зону незначимости, то принимается H_0 . Если Кэмп. попало в зону неопределенности то принимается H_1 на 5% уровне статистической значимости. Если Кэмп. попало в зону значимости, то принимается H_1 на 1% уровне статистической значимости.

ЭТАПЫ ПРИНЯТИЯ СТАТИСТИЧЕСКОГО РЕШЕНИЯ

- Формулировка нулевой и альтернативной гипотез
- Определение объема выборки N
- Выбор соответствующего уровня значимости или вероятности отклонения нулевой гипотезы. Это может быть величина меньшая или равная 0,05 (5% уровень значимости). В зависимости от важности исследования можно выбрать уровень значимости в 0,1% или даже в 0,001%.
- Выбор статистического метода, который зависит от типа решаемой задачи.
- Вычисление соответствующего эмпирического значения по экспериментальным данным, согласно выбранному статистическому методу.
- Нахождение по таблице для выбранного статистического метода критических значений, соответствующих уровню значимости для $P=0,05$ и для $P=0,01$.
- Построение оси значимости и нанесение на нее табличных критических значений и эмпирического значения. Для этого целесообразно пользоваться каждый раз приведенными выше рисунками.
- Формулировка принятия решения.

РЕКОМЕНДАЦИИ К ВЫБОРУ КРИТЕРИЯ РАЗЛИЧИЙ

- Определить, является ли выборка связанной (зависимой) или несвязанной (независимой)
- Определить однородность - неоднородность выборки.
- Оценить объем выборки и, зная ограничения каждого критерия по объему, выбрать соответствующий критерий
- Целесообразнее всего начинать работу с выбора наименее трудоемкого критерия
- Если используемый критерий не выявил различия - следует применить более мощный, но одновременно и более трудоемкий критерий
- Если в распоряжении исследователя имеется несколько критериев, то следует выбирать те из них, которые наиболее полно используют информацию, содержащуюся в экспериментальных данных.
- При малом объеме выборки следует увеличивать величину уровня значимости (не менее 1%), так как небольшая выборка и низкий уровень значимости приводят к увеличению вероятности принятия ошибочных решений.

Распределением признака называется закономерность встречаемости разных его значений (Плохинский Н.А.)

- Любая переменная (признак) имеет свой минимум и максимум, между которыми располагаются ее значения.
- Расстояние от минимума до максимума называется областью значений переменной, однако мы можем разбить это расстояние на равные отрезки и получить множество ограниченных интервалов значений.
- Распределение вероятностей позволяет нам судить о том, какова вероятность появления переменной в каждом из интервалов её значения.

Классификация распределений

- Равномерное распределение - когда все значения встречаются одинаково (или почти одинаково) часто.
 - Симметричное распределение - когда одинаково часто встречаются крайние значения.
 - Нормальное распределение - симметричное распределение, у которого крайние значения встречаются редко и частота постепенно повышается от крайних к серединным значениям признака.
 - Асимметричные распределения - левосторонние (с преобладанием частот малых значений), правосторонние (с преобладанием частот больших значений). К понятию формы распределения мы еще не раз вернемся, прежде всего - в связи с использованием в психологии нормального распределения как особого эталона - стандарта.
- и др.

Нормальное распределение

- Наиболее распространенным распределением является нормальное распределение. Оно часто используется для приближенного описания многих случайных явлений, в которых на интересующий нас признак оказывает воздействие большое количество независимых случайных факторов, среди которых нет резко выделяющихся.
- Нормальное распределение однозначно распределяется, если мы указываем значения двух его параметров: 1) среднее значение; 2) дисперсии.
- График нормального распределения называется кривой Гаусса и является симметричным относительно среднего значения.

Способы проверки на «нормальность» распределения

1. Сравнение характеристик набора данных со свойствами нормального распределения

- Все характеристики среднего значения нормального распределения — математическое ожидание, мода и медиана — совпадают друг с другом.
- Кривая распределения частот и значений совершенно симметрична по отношению к средней, т. е. слева и справа от нее лежит 50% вариантов; в интервале от $M - \sigma$ до $M + \sigma$ находится 68,26% всех вариантов; в интервале от $M - 2\sigma$ до $M + 2\sigma$ лежит 95,44% вариантов.
- Межквартильный размах нормального распределения равен 1.33σ

2. Построение графика нормального распределения

- В графическом способе результаты испытаний сортируют в вариационный ряд. Для каждого результата $x_{\{i\}}$ находят накопленную частоту: $W = i/(n+1)$, i - номер результата в ряду, n - объём испытаний.
- По рассчитанным значениям накопленных частот находят квантили нормального распределения (Q) с таким же средним значением и стандартным отклонением, как и у выборки.
- Далее по имеющимся данным наносят точки на плоскость в координатах $x-Q$. Если значения $x_{\{i\}}$ являются квантилями распределения того же вида, что и $Q_{\{i\}}$, между ними должна быть зависимость, близкая к линейной. Тогда, если точки на графике аппроксимируются прямой линией лишь с небольшими отклонениями от неё, то результаты испытаний достаточно хорошо описываются выбранным теоретическим распределением.
- При существенных отклонениях хотя бы некоторых точек от прямой, распределение результатов испытаний нельзя считать соответствующим выбранному теоретическому.

3. **Критерий Шапиро-Уилка** используется для проверки гипотезы о нормальном распределении. Этот критерий надёжен при $8 \leq n \leq 50$ (существует модифицированный критерий Шапиро-Уилка, применимый при n до 2000), и является более мощным, чем другие, критерии, т.е. даёт наименьшую вероятность принять нулевую гипотезу, когда на самом деле верна альтернативная, в данном случае - принять гипотезу о нормальном распределении, когда распределение не соответствует нормальному.

8. Вопросы по теме занятия

1. Что такое статистическая гипотеза?
2. Какие бывают ошибки при проверке статистических гипотез?
3. Что такое уровень статистической значимости?
4. Какие критерии называют параметрическими?

5. Какие критерии называют непараметрическими?
6. Что такое распределение случайной величины?
7. Какие виды распределения случайной величины бывают?
8. Что такое нормальное распределение?

9. Тестовые задания по теме с эталонами ответов

1. КРИТЕРИЙ, ИСПОЛЬЗУЕМЫЙ ДЛЯ ПРОВЕРКИ РАСПРЕДЕЛЕНИЯ НА НОРМАЛЬНОСТЬ:

- 1) критерий Манна-Уитни;
- 2) критерий Шапиро-Уилка;
- 3) критерий знаков;
- 4) критерий Стьюдента;
- 5) критерий Вилкоксона;

2. ВИД РАСПРЕДЕЛЕНИЯ, ПРИ КОТОРОМ ВСЕ ЗНАЧЕНИЯ ВСТРЕЧАЮТСЯ ОДИНАКОВО (ИЛИ ПОЧТИ ОДИНАКОВО) ЧАСТО:

- 1) равномерное;
- 2) левостороннее;
- 3) правостороннее;
- 4) нормальное;
- 5) симметричное;

3. СТАТИСТИЧЕСКАЯ ГИПОТЕЗА - ЭТО:

- 1) некоторое утверждение относительно неизвестного параметра или какой-либо характеристики;
- 2) предполагаемое решение некоторой проблемы;
- 3) алгоритм решения некоторой проблемы;
- 4) некоторое утверждение относительно неизвестных данных;
- 5) некоторое утверждение относительно известных данных;

4. УРОВЕНЬ СТАТИСТИЧЕСКОЙ ЗНАЧИМОСТИ РЕЗУЛЬТАТА ИССЛЕДОВАНИЯ - ЭТО:

- 1) количественно выраженная вероятность, свидетельствующая о том, что результаты достоверны;
- 2) количественно выраженная вероятность, свидетельствующая о том, что результаты недостоверны;
- 3) связан с репрезентативностью выборки;
- 4) вероятность ошибочного отклонения нулевой гипотезы;
- 5) вероятность ошибочного отклонения альтернативной гипотезы;

5. ГРАНИЧНЫЕ ТОЧКИ КРИТИЧЕСКОЙ ОБЛАСТИ НАЗЫВАЮТСЯ:

- 1) критическими точками;
- 2) граничными точками;
- 3) начальными точками;
- 4) конечными точками;
- 5) экстремумами критической области;

10. Ситуационные задачи по теме с эталонами ответов

1. Проводится исследование о действии некоторого препарата группа А на людей разных возрастных групп. Обследуемые разбиты на две группы: обследуемые в возрасте до 60 лет и старше 60 лет.

Вопрос 1: Что такое статистическая гипотеза?;

Вопрос 2: Сформулируйте нулевую гипотезу для этого исследования?;

- 1) Статистическая гипотеза — некоторое утверждение относительно неизвестного параметра или какой-либо характеристики генеральной совокупности.;
- 2) Нулевая гипотеза обычно заключается в предположении о том, что изучаемое вмешательство не оказывает значимого воздействия на генеральные параметры распределения. В данном случае, что препарат А оказывает одинаковое влияние на обе возрастные группы.;

2. При исследовании взаимосвязи роста заболеваемости и температуры окружающей среды вы получили некоторый набор данных. Для статистической обработки вам нужно определить распределение полученных данных.

Вопрос 1: Что такое распределение?;

Вопрос 2: Какими способами можно проверить, что распределение нормальное?;

- 1) Распределением признака называется закономерность встречаемости разных его значений;
- 2) Сравнение характеристик набора данных со свойствами нормального распределения; построение графика нормального распределения; критерий Шапиро-Уилка. Вычисление частот, описательной статистики;

11. Примерная тематика НИРС по теме

1. Логнормальное распределение
2. Аддитивный белый гауссовский шум
3. Статистическая мощность

12. Рекомендованная литература по теме занятия

- обязательная:

Медик, В. А. [Математическая статистика в медицине](#) : учебное пособие для вузов : в 2 т. / В. А. Медик, М. С. Токмачев. - 2-е изд., перераб. и доп. - Москва : Юрайт, 2023. - Т. 1. - 471 с. - Текст : электронный.

Медик, В. А. [Математическая статистика в медицине](#) : учебное пособие для вузов : в 2 т. / В. А. Медик, М. С.

Токмачев. - 2-е изд. - Москва : Юрайт, 2023. - Т. 2. - 347 с. - Текст : электронный.

- дополнительная:

Наркевич, А. Н. [Статистические методы исследования в медицине и биологии](#) : учеб. пособие / А. Н. Наркевич, К. А. Виноградов, К. В. Шадрин ; Красноярский медицинский университет. - Красноярск : КрасГМУ, 2018. - 109 с. - Текст : электронный.

Малугин, В. А. [Математическая статистика](#) : учебное пособие для вузов / В. А. Малугин. - Москва : Юрайт, 2023. - 218 с. - Текст : электронный.

[Основы статистического анализа в медицине](#) : учебное пособие / ред. А. В. Решетников. - Москва : Медицинское информационное агентство, 2020. - 176 с. - Текст : электронный.

- электронные ресурсы:

База знаний WolframAlpha (<https://www.wolframalpha.com/examples/Statistics.html>)

Дистанционные курсы по статистике, анализу данных (<https://welcome.stepik.org/ru>)

Онлайн-калькулятор для статистики (http://math.semestr.ru/group/group_manual.php)

1. **Тема № 6.** Корреляционный анализ (в интерактивной форме)

2. **Разновидность занятия:** упражнение

3. **Методы обучения:** объяснительно-иллюстративный, частично-поисковый (эвристический)

4. **Значение темы** (актуальность изучаемой проблемы): Корреляционный анализ – один из базовых методов математической статистики, представляющий вычислительную сложность для исследователя. Современное программное обеспечение решает эту проблему.

5. **Цели обучения**

- **обучающийся должен знать** методы анализа медико-статистической информации и интерпретации результатов состояния здоровья пациента (населения), **уметь** выбирать способ вычисления коэффициента корреляции, определять характер и силу корреляционной связи, ошибку и достоверность коэффициента корреляции и формулировать выводы, **владеть** навыками определения корреляционной взаимосвязи между признаками

6. **Место проведения и оснащение занятия:**

- **место проведения занятия:** компьютерный класс №1 (3-03)

- **оснащение занятия:** видеопроектор, доска магнитно-маркерная, комплект учебной мебели на посадочные места, локальный сетевой сервер, персональные компьютеры, экран

7. **Аннотация** (краткое содержание темы)

Корреляционный анализ – метод, позволяющий обнаружить связь между несколькими случайными величинами.

Термин корреляция ввёл французский палеонтолог Жорж Кювье в XVIII веке. Он разработал «закон корреляции» частей и органов живых существ, с помощью которого можно восстановить облик ископаемого животного, имея в распоряжении лишь часть его останков. В статистике слово «корреляция» первым стал использовать английский биолог и статистик Фрэнсис Гальтон в конце XIX века.

Взаимосвязи на языке математики обычно описываются при помощи функций, которые графически изображаются в виде линий.

Если изменение одной переменной на одну единицу всегда приводит к изменению другой переменной на одну и ту же величину, функция является **линейной** (график ее представляет прямую линию); любая другая связь – **нелинейная**.

Если увеличение одной переменной связано с увеличением другой, то связь – **положительная** (прямая); если увеличение одной переменной связано с уменьшением другой, то связь – **отрицательная** (обратная).

В качестве числовой характеристики вероятностной связи используются коэффициенты корреляции.

Статистика говорит о корреляции между двумя переменными и указывает силу связи при помощи некоторого критерия взаимосвязи, который получил название **коэффициента корреляции**. Этот коэффициент, всегда обозначаемый латинской буквой r , может принимать значения между -1 и +1. Если абсолютное значение коэффициента корреляции находится ближе к 1, то это означает наличие сильной связи, а если ближе к 0, то слабой.

Метод вычисления коэффициента корреляции

Типы шкал		Мера связи
Переменная X	Переменная Y	
Интервальная (шкала отношений)	Интервальная (шкала отношений)	Коэффициент Пирсона
Ранговая, интервальная (или отношений)	Ранговая, интервальная (или отношений)	Коэффициент Спирмена
Ранговая	Ранговая	Коэффициент Кендалла
Дихотомическая	Дихотомическая	Коэффициент ϕ , четырёхполевая корреляция
Дихотомическая	Ранговая	Рангово-бисериальный коэффициент

r -Пирсона (Pearson r) применяется для изучения взаимосвязи двух метрических переменных, измеренных на одной и той же выборке. При этом характер взаимосвязи должен быть линейным и монотонным (смотрим диаграмму рассеяния) и характеристики переменных распределены нормально.

СТАТИСТИЧЕСКИЕ ГИПОТЕЗЫ

$H_0: r_{XY} = 0$. Коэффициент корреляции генеральной совокупности равен 0.

$H_1: r_{XY} \neq 0$. Коэффициент корреляции генеральной совокупности не равен 0.

Рассчитывается t -значение со степенями свободы $df=N-2$, на основе которого рассчитывает p -значимость.

- Если $|T_{эмп}| > t_{кр}$, то нулевую гипотезу отвергают, и выборочный коэффициент корреляции значимо отличается от нуля, а X и Y коррелированы, т.е. связаны линейной зависимостью.
- Если $|T_{эмп}| \leq t_{кр}$, то нет оснований отвергать нулевую гипотезу и говорят, что выборочный коэффициент корреляции незначим, а X и Y некоррелированы, т.е. не связаны линейной зависимостью.

8. Вопросы по теме занятия

1. Что такое корреляция?
2. Кто ввёл термин «корреляция»?
3. Что понимается под положительной и отрицательной корреляцией?
4. Каковы градации силы коэффициента корреляции?
5. Перечислите способы нахождения коэффициентов корреляции.

9. Тестовые задания по теме с эталонами ответов

1. ФРАНЦУЗСКИЙ ПАЛЕОНТОЛОГ, КОТОРЫЙ ВПЕРВЫЕ ВВЁЛ ТЕРМИН «КОРРЕЛЯЦИЯ»:

- 1) Фрэнсис Гальтон;
- 2) Жорж Кювье;
- 3) Тома Корнель;
- 4) Жан Расин;
- 5) Жан де Лонуа;

2. ПРОМЕЖУТОК, В КОТОРОМ ИЗМЕНЯЕТСЯ КОЭФФИЦИЕНТ КОРРЕЛЯЦИИ:

- 1) [-2;0];
- 2) [-1;0];
- 3) [-1;1];
- 4) [0;1];
- 5) [0;2];

3. ЗНАЧЕНИЕ КОЭФФИЦИЕНТА КОРРЕЛЯЦИИ, КОТОРОЕ ПОКАЗЫВАЕТ, ЧТО ИССЛЕДУЕМЫЕ ПРИЗНАКИ НЕ ВЗАИМОСВЯЗАНЫ:

- 1) -1;
- 2) -0,5;
- 3) 0;
- 4) 0,5;
- 5) 1;

4. СТАТИСТИЧЕСКАЯ ВЗАИМОСВЯЗЬ ДВУХ ИЛИ НЕСКОЛЬКИХ СЛУЧАЙНЫХ ВЕЛИЧИН (ЛИБО ВЕЛИЧИН, КОТОРЫЕ МОЖНО С НЕКОТОРОЙ ДОПУСТИМОЙ СТЕПЕНЬЮ ТОЧНОСТЬЮ СЧИТАТЬ ТАКОВЫМИ):

- 1) корреляция;
- 2) дисперсия;
- 3) ковариация;
- 4) регрессия;
- 5) стандартное отклонение;

5. ДЛЯ ВЫЯВЛЕНИЯ ВЗАИМОСВЯЗИ МЕЖДУ КОЛИЧЕСТВЕННЫМИ ИЛИ КАЧЕСТВЕННЫМИ ПОКАЗАТЕЛЯМИ, ЕСЛИ ИХ МОЖНО РАНЖИРОВАТЬ, ПРИМЕНЯЕТСЯ:

- 1) коэффициент ранговой корреляции Кендалла;
- 2) линейный коэффициент корреляции;
- 3) коэффициент корреляции Спирмена;
- 4) коэффициент корреляции Фишера;
- 5) коэффициент корреляции Кювье;

10. Ситуационные задачи по теме с эталонами ответов

1. Вы исследовали взаимосвязь между двумя медицинскими показателями, выраженных в интервальной шкале. При этом вы получили коэффициент корреляции Пирсона -0,42; а коэффициент корреляции Кендалла -0,38

Вопрос 1: Как можно проинтерпретировать полученные значения?;

Вопрос 2: Какой из этих коэффициентов лучше подходит для данной ситуации?;

- 1) Результаты довольно хорошо согласуются друг с другом, и их значения свидетельствуют о наличии слабой отрицательной взаимосвязи;
- 2) Коэффициент корреляции Пирсона;

2. Необходимо определить взаимосвязь двух медицинских показателей X : 24 27 26 21 20 31 26 22 20 18 30 29 24 26 и Y : 100 115 117 119 134 94 105 103 111 124 122 109 110 86, выбранных из генеральной совокупности с нормальным распределением данных.

Вопрос 1: Каким коэффициентом корреляции стоит воспользоваться для исследования взаимосвязи данных?;

Вопрос 2: Вычислите этот коэффициент;

Вопрос 3: Сделайте вывод о корреляции данных в зависимости от полученного значения коэффициента корреляции;

- 1) Коэффициентом корреляции Пирсона, так как данные представлены в шкале отношений и отображены из

генеральной совокупности с нормальным распределением.;

2) -0.4211;

3) В соответствии с таблицей значений величин коэффициента корреляции делаем вывод о том, что $r = -0.421$ это слабая по силе отрицательная корреляция;

11. Примерная тематика НИРС по теме

1. Содержательная интерпретация корреляции
2. Количественная интерпретация корреляции
3. Автокорреляционная функция

12. Рекомендованная литература по теме занятия

- обязательная:

Медик, В. А. [Математическая статистика в медицине](#) : учебное пособие для вузов : в 2 т. / В. А. Медик, М. С. Токмачев. - 2-е изд., перераб. и доп. - Москва : Юрайт, 2023. - Т. 1. - 471 с. - Текст : электронный.

Медик, В. А. [Математическая статистика в медицине](#) : учебное пособие для вузов : в 2 т. / В. А. Медик, М. С. Токмачев. - 2-е изд. - Москва : Юрайт, 2023. - Т. 2. - 347 с. - Текст : электронный.

- дополнительная:

Наркевич, А. Н. [Статистические методы исследования в медицине и биологии](#) : учеб. пособие / А. Н. Наркевич, К. А. Виноградов, К. В. Шадрин ; Красноярский медицинский университет. - Красноярск : КрасГМУ, 2018. - 109 с. - Текст : электронный.

Малугин, В. А. [Математическая статистика](#) : учебное пособие для вузов / В. А. Малугин. - Москва : Юрайт, 2023. - 218 с. - Текст : электронный.

[Основы статистического анализа в медицине](#) : учебное пособие / ред. А. В. Решетников. - Москва : Медицинское информационное агентство, 2020. - 176 с. - Текст : электронный.

- электронные ресурсы:

База знаний WolframAlpha (<https://www.wolframalpha.com/examples/Statistics.html>)

Дистанционные курсы по статистике, анализу данных (<https://welcome.stepik.org/ru>)

Онлайн-калькулятор для статистики (http://math.semestr.ru/group/group_manual.php)

1. Тема № 7. Регрессионный анализ (в интерактивной форме)

2. Разновидность занятия: комбинированное

3. Методы обучения: объяснительно-иллюстративный, частично-поисковый (эвристический)

4. Значение темы (актуальность изучаемой проблемы): Регрессионный анализ - один из базовых методов математической статистики, позволяющий изучать взаимосвязь между переменными. Для успешного применения этого метода необходимо понимать принцип работы этого метода.

5. Цели обучения

- **обучающийся должен знать** методы анализа медико-статистической информации и интерпретации результатов состояния здоровья пациента (населения), **уметь** составлять уравнение регрессии (линейной, нелинейной, парной, множественной) и интерпретировать коэффициенты уравнения регрессии, **владеть** навыками регрессионного анализа

6. Место проведения и оснащение занятия:

- **место проведения занятия:** компьютерный класс №1 (3-03)

- **оснащение занятия:** видеопроектор, доска магнитно-маркерная, комплект учебной мебели на посадочные места, локальный сетевой сервер, персональные компьютеры, экран

7. Аннотация (краткое содержание темы)

Регрессионный анализ предназначен для изучения взаимосвязи одной переменной (зависимой, результирующей) и нескольких других переменных (независимых, исходных).

Эта взаимосвязь математически выражается в уравнении. Поэтому провести регрессионный анализ данных - это, по сути, означает найти уравнение зависимости одной переменной от других.

Исходные данные для регрессионного представляют собой таблицу (матрицу) размерностью $N \times P$ следующего вида:

№	X_1	X_2	...	X_p
1	X_{11}	X_{12}	...	X_{1p}
2	X_{21}	X_{22}	...	X_{2p}
...
N	X_{N1}	X_{N2}	...	X_{Np}

Строки этой таблицы соответствуют объектам (испытуемым), а столбцы - переменным. Все переменные при этом должны быть измерены в количественной шкале. Одна из переменных определяется исследователем как зависимая, а остальные (или часть их) - как независимые переменные. Допускается, что для некоторых объектов значения зависимой переменной неизвестны, и их определение (оценка) может составлять важный результат анализа.

Регрессионный анализ может применяться как для решения прикладных задач, так и в исследовательских целях. Обычно регрессионный анализ применяется для изучения возможности предсказания некоторого результата (обучения, деятельности) по ряду предварительно измеренных характеристик.

В зависимости от количества независимых переменных различают:

- Парную регрессию, когда независимая переменная одна
- Множественную регрессию, когда независимых переменных несколько

В зависимости от вида уравнения регрессии

- Линейную регрессию, когда зависимость переменных выражается прямой
- Нелинейную регрессию - все остальные зависимости

Парная линейная регрессия - это регрессионная модель зависимости одной переменной от другой с линейной функцией зависимости.

Общий вид уравнения парной линейной регрессии имеет вид:

$$y = ax + b + \epsilon,$$

где x - независимая переменная, y - зависимая, a, b - коэффициенты регрессии, ϵ - отклонение предсказаний модели от эмпирических значений.

Зависимая переменная (y) — это переменная, описывающая процесс, который мы пытаемся предсказать или понять.

Независимые переменные (x) — это переменные, используемые для моделирования или прогнозирования значений

зависимых переменных. В уравнении регрессии они располагаются справа от знака равенства и часто называются объяснительными переменными. Зависимая переменная - это функция независимых переменных.

Коэффициенты регрессии (a, b) — это коэффициенты, которые рассчитываются в результате выполнения регрессионного анализа. Вычисляются величины для каждой независимой переменной, которые представляют силу и тип взаимосвязи независимой переменной по отношению к зависимой.

Невязки. Существует необъяснимое количество зависимых величин, представленных в уравнении регрессии как случайные ошибки ε .

Графическое выражение регрессионного анализа называют линией регрессии. Линия регрессии выражает наилучшее предсказание зависимой переменной Y по независимым переменным X.

$$a = r \frac{\sigma_Y}{\sigma_X}, b = Y_{cp} - aX_{cp}$$

где r - коэффициент корреляции X и Y, σ_X - среднее квадратическое отклонение X, σ_Y - среднее квадратическое отклонение Y, X_{cp} - среднее значение X, Y_{cp} - среднее значение.

Условия применения регрессионного анализа

- Сравнимые переменные должны быть измерены в шкале интервалов или отношений
- Предполагается, что переменные имеют нормальный закон распределения
- Число варьирующих признаков в сравниваемых переменных должно быть одинаковым

Анализ качества эмпирического уравнения регрессии включает в себя:

- проверку статистической значимости коэффициентов уравнения
- проверку общего качества уравнения регрессии
- проверку выполнимости предварительных условий, на основании которых построено уравнения (например, проверка выполнимости предпосылок метода наименьших квадратов)

Для коэффициентов регрессионного уравнения проверка их уровня значимости осуществляется по t-критерию Стьюдента и по критерию F Фишера.

Рассмотрим оценку достоверности показателей регрессии для линейного уравнения

$$y = ax + b + \varepsilon$$

Для этого типа уравнений оценивают по t-критерию Стьюдента только величину коэффициента b с использованием вычисления величины Tф по следующим формулам:

$$T_f = b / S_{b_{yx}}$$

где

$$S_{b_{yx}} = \sqrt{\frac{(1 - r^2) \sum (x_i - x_{cp})^2}{(n - 2) \sum (y_i - y_{cp})^2}}$$

Величина Tф позволяет оценить уровень значимости коэффициента b уравнения регрессии Y по X.

Для оценки качества подбора линейной функции рассчитывается квадрат линейного коэффициента корреляции r^2 , который называется коэффициентом детерминации. Коэффициент детерминации характеризует долю дисперсии результативного признака y, объясняемую регрессией, в общей дисперсии результативного признака.

Проверить значимость уравнения регрессии - значит установить, соответствует ли математическая модель, выражающая зависимость между переменными, экспериментальным данным и достаточно ли включенных в уравнение объясняющих переменных (одной или нескольких) для описания зависимой переменной.

Множественный регрессионный анализ (МРА) предназначен для изучения взаимосвязи одной переменной (зависимой, результирующей) и нескольких других переменных (независимых, исходных).

Множественная регрессия во многом аналогична простой (парной) регрессии.

В общем случае уравнение регрессии представляет собой сложный полином, описывающий зависимость сразу между несколькими переменными. Такое уравнение множественной регрессии имеет вид:

$$y = a_0 + a_1x_1 + a_2x_2 + \dots + a_px_p + \varepsilon,$$

где x_1, \dots, x_p — интересующие независимые переменные, а y — зависимая переменная, ε — отклонение предсказаний модели от эмпирических значений.

Главное требование к исходным данным - отсутствие линейных взаимосвязей между переменными, когда одна переменная является линейной производной другой переменной. Таким образом, нельзя пользоваться суммой переменных или их средним арифметическим наряду с самими переменными! Соответственно, недопустимы переменные, коэффициент корреляции которых с любой другой переменной равен 1. Следует избегать включения в анализ переменных, корреляция между которыми близка к 1, так как сильно коррелирующая переменная не несет для анализа новой информации, добавляя излишний «шум».

Для проведения множественной линейной регрессии необходимо, чтобы:

- Сравнимые переменные должны быть измерены в шкале интервалов или отношений
- Предполагается, что переменные имеют нормальный закон распределения
- Число варьирующих признаков в сравниваемых переменных должно быть одинаковым

Линейная регрессия в R: lm()

Аргумент	Описание
formula	Пишется в виде Зависимая переменная ~ независимые переменные Вес ~ Рост
data	Название переменной, где хранятся данные для регрессионного анализа

Линейная регрессия в PSPP: Анализ - Регрессия - Линейная

Нелинейная регрессия — это вид регрессионного анализа, в котором экспериментальные данные моделируются функцией, являющейся нелинейной комбинацией параметров модели и зависящей от одной и более независимых переменных.

Нелинейность регрессии может быть обусловлена двумя причинами :

- нелинейность по объясняющей переменной;
- нелинейность по коэффициентам регрессии.

Примером такой нелинейности может служить уравнение регрессии вида (гиперболическая регрессия):

$$y(x) = a_0 + a_1\sqrt{x}$$

В этом случае, вводя новую переменную $Z = X^{1/2}$, приходим к линейной регрессии $y = b_0 + b_1Z$, коэффициенты b_0, b_1 , которой вычисляются на основе метода наименьших квадратов. Вычислив коэффициенты, возвращаемся к исходному нелинейному уравнению регрессии.

К такому классу нелинейных регрессий относятся уравнения, в которых зависимая переменная нелинейным образом зависит от коэффициентов регрессии.

Примеры таких нелинейных регрессионных моделей могут служить функции

- степенная;
- показательная;
- экспоненциальная.

В общем случае, чтобы привести нелинейные зависимости к линейной, используют методы линеаризации:

1. Замена переменных.
2. Логарифмирование обеих частей уравнения.
3. Комбинированный.

$y=f(x)$	Преобразование	Метод линеаризации
$y = b x^a$	$Y = \ln(y); X = \ln(x)$	Логарифмирование

$y = b e^{ax}$	$Y = \ln(y); X = x$	Комбинированный
$y = 1/(ax+b)$	$Y = 1/y; X = x$	Замена переменных
$y = x/(ax+b)$	$Y = x/y; X = x$	Замена переменных
$y = a \ln(x)+b$	$Y = y; X = \ln(x)$	Комбинированный
$y = a + bx + cx^2$	$x_1 = x; x_2 = x^2$	Замена переменных
$y = a + bx + cx^2 + dx^3$	$x_1 = x; x_2 = x^2; x_3 = x^3$	Замена переменных
$y = a + b/x$	$x_1 = 1/x$	Замена переменных
$y = a + \sqrt{x}b$	$x_1 = \sqrt{x}$	Замена переменных

Нелинейный регрессионный анализ в MS Excel

Команда «Добавить линию тренда» используется для выделения тренда (медленных изменений) при анализе временных рядов. Однако эту команду можно использовать и для построения уравнения регрессии, рассматривая в качестве времени t независимую переменную x . Эта команда позволяет построить следующие регрессии: линейную; полиномиальную; логарифмическую; степенную; экспоненциальную.

- В выбранном листе MS Excel ввести по столбцам исходные данные $\{x_i, y_i\}, i=1, \dots, n$
- По этим данным построить график в декартовой системе координат
- Установить курсор на построенном графике, сделать щелчок правой кнопкой и в появившемся контекстном меню выполнить команду «Добавить линию тренда»
- В появившемся диалоговом окне активизировать закладку «Тип» и выбрать нужное уравнение регрессии
- «Включить» необходимые для нас опции:
 - «Показать уравнение на диаграмме» - на диаграмме будет показано выбранное уравнение регрессии с вычисленными коэффициентами.
 - «Поместить на диаграмму величину достоверности аппроксимации (R^2)» - на диаграмме будет показана значение индекса детерминации R_{xy}^2 , которое можно использовать для проверки значимости построенной регрессии с помощью F - теста. Если по построенному уравнению регрессии необходимо выполнить прогноз, то нужно указать число периодов прогноза.
- После задания всех перечисленных опций щелкнуть на кнопке «Заккрыть» и на диаграмме появиться формула построенного уравнения регрессии и значение индекса детерминации R_{xy}^2

8. Вопросы по теме занятия

1. Что такое регрессия?

1) Регрессия - величина, выражающая зависимость среднего значения случайной величины Y от значений случайной величины X ;

2. Что такое регрессивный анализ?

3. Как называют независимые переменные?

4. Как выглядит уравнение регрессии?

5. Для решения каких задач применяется множественный регрессионный анализ?

6. Что выступает в качестве оценки силы связи между зависимой переменной и совокупностью объясняющих переменных?

9. Тестовые задания по теме с эталонами ответов

1. ВЕЛИЧИНА, ВЫРАЖАЮЩАЯ ЗАВИСИМОСТЬ СРЕДНЕГО ЗНАЧЕНИЯ СЛУЧАЙНО ВЕЛИЧИНЫ Y ОТ ЗНАЧЕНИЙ СЛУЧАЙНОЙ ВЕЛИЧИНЫ X :

- 1) корреляция;
- 2) ковариация;
- 3) дисперсия;
- 4) регрессия;
- 5) констатация;

2. ТЕРМИН, КОТОРЫЙ ИСПОЛЬЗУЮТ ДЛЯ ОБОЗНАЧЕНИЯ НЕЗАВИСИМЫХ ПЕРЕМЕННЫХ В РЕГРЕССИОННОМ АНАЛИЗЕ:

- 1) критерии;
- 2) константы;
- 3) коэффициенты;
- 4) факторы;
- 5) предикторы;

3. КРИТЕРИЙ, ПРИМЕНЯЮЩИЙСЯ ИСКЛЮЧИТЕЛЬНО ДЛЯ ВЫБОРА ИЗ НЕСКОЛЬКИХ СТАТИСТИЧЕСКИХ МОДЕЛЕЙ ПРИ МНОЖЕСТВЕННОЙ РЕГРЕССИОННОМ АНАЛИЗЕ:

- 1) критерий согласия;
- 2) информационный критерий;
- 3) критерий Манна-Уитни;

- 4) критерий Колмогорова-Смирнова;
- 5) критерий значимости;
4. МЕТОД, С ПОМОЩЬЮ КОТОРОГО МОЖНО ПОЛУЧИТЬ ПАРАМЕТРЫ ДЛЯ ПАРНОЙ ЛИНЕЙНОЙ РЕГРЕССИИ:
 - 1) метод Монте-Карло;
 - 2) метод наименьших квадратов;
 - 3) метод проб и ошибок;
 - 4) метод математического моделирования;
 - 5) методы дифференциального и интегрального исчисления;
5. РАСПРЕДЕЛЕНИЕ СТАТИСТИКИ, КОТОРУЮ ИСПОЛЬЗУЮТ ДЛЯ ПРОВЕРКИ ЗНАЧИМОСТИ КОЭФФИЦИЕНТА ДЕТЕРМИНАЦИИ:
 - 1) распределение Стьюдента;
 - 2) нормальное распределение;
 - 3) распределение Гаусса;
 - 4) распределение Фишера;
 - 5) распределение Пуассона;

10. Ситуационные задачи по теме с эталонами ответов

1. Вам необходимо исследовать влияние одной или нескольких независимых переменных X_1, X_2, \dots, X_n на зависимую переменную Y .

Вопрос 1: Какой статистический метод Вы будете использовать?;

Вопрос 2: Какие величины необходимо вычислить в ходе применения данного метода;

- 1) Регрессионный анализ;
 - 2) Выборочное среднее, ковариационная матрица;
2. В ходе исследования эффективности деятельности психолога-практика и психолога-исследователя, которую можно прогнозировать на основе разных характеристик: A — готовность к контактам, B — общая интеллектуальность, N — ненасыщаемость контактами с другими людьми, N — умение поддерживать контакт, были получены следующие уравнения зависимостей: для психолога-практика $\text{Эфф.} = 0.72 \cdot A + 0.28 \cdot B + 0.2 \cdot N + 0,28N$; для психолога-исследователя $\text{Эфф.} = 0.32 \cdot A + 0.68 \cdot B + 0.43 \cdot N$

Вопрос 1: Как называются эти уравнения зависимостей в статистическом анализе?;

Вопрос 2: Какой вывод можно сделать, исходя из уравнений зависимостей, о значимости для психологов интенсивности общения?;

- 1) Уравнения множественной регрессии;
- 2) Для психолога-исследователя не характерно наличие интенсивного общения, в то время как для психолога-практика интенсивное общение оказывается самым значимым качеством;

11. Примерная тематика НИРС по теме

1. Информационный критерий Акаике (AIC)
2. Метод максимального правдоподобия
3. Метод инструментальных переменных
4. Метод моментов

12. Рекомендованная литература по теме занятия

- обязательная:

Медик, В. А. [Математическая статистика в медицине](#) : учебное пособие для вузов : в 2 т. / В. А. Медик, М. С. Токмачев. - 2-е изд., перераб. и доп. - Москва : Юрайт, 2023. - Т. 1. - 471 с. - Текст : электронный.

Медик, В. А. [Математическая статистика в медицине](#) : учебное пособие для вузов : в 2 т. / В. А. Медик, М. С. Токмачев. - 2-е изд. - Москва : Юрайт, 2023. - Т. 2. - 347 с. - Текст : электронный.

- дополнительная:

Наркевич, А. Н. [Статистические методы исследования в медицине и биологии](#) : учеб. пособие / А. Н. Наркевич, К. А. Виноградов, К. В. Шадрин ; Красноярский медицинский университет. - Красноярск : КрасГМУ, 2018. - 109 с. - Текст : электронный.

Малугин, В. А. [Математическая статистика](#) : учебное пособие для вузов / В. А. Малугин. - Москва : Юрайт, 2023. - 218 с. - Текст : электронный.

[Основы статистического анализа в медицине](#) : учебное пособие / ред. А. В. Решетников. - Москва : Медицинское информационное агентство, 2020. - 176 с. - Текст : электронный.

- электронные ресурсы:

База знаний WolframAlpha (<https://www.wolframalpha.com/examples/Statistics.html>)

Дистанционные курсы по статистике, анализу данных (<https://welcome.stepik.org/ru>)

Онлайн-калькулятор для статистики (http://math.semestr.ru/group/group_manual.php)

1. **Тема № 8.** Основные медико-статистические показатели и исследование их в динамике (в интерактивной форме)
2. **Разновидность занятия:** комбинированное
3. **Методы обучения:** объяснительно-иллюстративный, частично-поисковый (эвристический)
4. **Значение темы** (актуальность изучаемой проблемы): Расчет медико-статистических показателей имеет огромное значение для анализа социальных, медицинских, клинических явлений, а также прогнозирования и планирования различных лечебно-профилактических и организационных мероприятий.
5. **Цели обучения**
 - **обучающийся должен знать** методы анализа медико-статистической информации и интерпретации результатов состояния здоровья пациента (населения), **уметь** понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности, анализировать полученные в результате вычислений статистические величины (абсолютные, средние, относительные показатели, показатели динамического ряда) на основании сравнения в динамике и с имеющимися нормативами, **владеть** современными информационными технологиями для решения профессиональных задач, навыками анализа полученных в результате вычислений статистических величин
6. **Место проведения и оснащение занятия:**
 - **место проведения занятия:** компьютерный класс №1 (З-03)
 - **оснащение занятия:** видеопроектор, доска магнитно-маркерная, комплект учебной мебели на посадочные места, локальный сетевой сервер, персональные компьютеры, экран
7. **Аннотация** (краткое содержание темы)

Динамический ряд — ряд, представляющий собой значения однородных величин, характеризующих изменения явления во времени (определенной хронологической последовательности).

Каждый динамический ряд содержит две составляющие:

1. Показатели периодов времени (годы, кварталы, месяцы, дни или даты);
2. Показатели, характеризующие исследуемый объект за временные периоды или на соответствующие даты, которые называют уровнями ряда.

Динамические ряды могут быть представлены только однородными величинами: абсолютными, относительными или средними величинами.

Различают:

- **Динамический интервальный ряд** содержит значения показателей за определенные периоды времени. В интервальном ряду уровни можно суммировать, получая объем явления за более длительный период, или так называемые накопленные итоги.
- **Динамический моментный ряд** отражает значения показателей на определенный момент времени (дату времени). В моментных рядах исследователя может интересовать только разность явлений, отражающая изменение уровня ряда между определенными датами, поскольку сумма уровней здесь не имеет реального содержания. Накопленные итоги здесь не рассчитываются.

Показатели динамического ряда

Для анализа динамических рядов вычисляют показатели: абсолютный прирост, темп роста, темп прироста, значение 1 % прироста.

Абсолютный прирост - равен разности между каждым последующим (изучаемым) и предыдущим уровнями.

Темп роста - представляет собой процентное отношение каждого последующего (изучаемого) уровня к предыдущему.

Темп прироста - это процентное соотношение последующего (изучаемого) абсолютного прироста к предыдущему уровню и показатель, на сколько процентов изменилось явление.

Сглаживание динамического ряда

Динамический ряд не всегда состоит из уровней, последовательно изменяющихся в сторону снижения или увеличения, что затрудняет возможность проследить закономерность, делает ряд не в достаточной степени наглядным. В таких случаях производят сглаживание динамического ряда.

Способы сглаживания динамического ряда:

- Укрупнение интервалов производят путем суммирования данных за определенный период времени.

- Вычисление групповой средней путем вычисления средней величины за определенный промежуток времени.
- Вычисление скользящей средней путем вычисления среднего уровня данного времени и соседних с ним.
- По способу наименьших квадратов по специальной формуле.

Измерение сезонных колебаний

Часто обнаруживается, что заболеваемость (смертность) от отдельных болезней проявляет значительную «сезонность», т. е. в одни месяцы года возрастает и снижается в другие. Например, простудные заболевания - в холодные месяцы, а кишечные заболевания - в жаркие месяцы. Для определения сезонности следует пользоваться отношением среднедневного числа заболеваний (смертности) в каждом месяце среднедневному годовому числу заболеваний. Это отношение для удобства выражается в процентах.

Статистический показатель — количественная характеристика некоторой статистической совокупности, численное выражение внутренней сущности изучаемого явления.

Выделяют следующие группы статистических показателей:

- абсолютные,
- относительные,
- средние.

Абсолютные показатели отражают объем или размер изучаемого события или явления, процесса, выраженного в соответствующих единицах измерения в конкретных условиях места и времени.

Относительные показатели - это результат сравнения (отношения, деления) двух абсолютных величин.

Различают:

- экстенсивные показатели;
- интенсивные показатели;
- показатели соотношения;
- показатели наглядности;
- показатели координации;
- показатели правдоподобия.

Экстенсивные показатели отражают внутреннюю структуру явления, разбиение его на составные части, удельный вес каждой части в целом. Например, к экстенсивным показателям относятся структура заболеваемости, инвалидности, смертности, коечного фонда, врачебных специальностей и др. Рассчитывается показатель по следующей формуле и выражается в процентах (%):

Экстенсивный показатель = Часть явления × 100% / Целое явление

Интенсивные показатели характеризуют уровень, распространенность какого-либо явления в среде, непосредственно связанного с ней. Эти показатели рассчитываются, как правило, для анализа здоровья населения, где в качестве среды берется численность населения, а в качестве явления — число рождений, заболеваний, смертей и др. К ним относятся показатели заболеваемости, рождаемости, смертности населения и др. Рассчитываются по следующей формуле и выражаются в промилле (‰), децимилле (^), сантимилле (^о):

Интенсивный показатель = Целое явление × 1000 / Среда.

Различия между интенсивным и экстенсивным показателями

Экстенсивный показатель	Интенсивный показатель
Не сравнивают, или сравнивают с большой осторожностью, зная сущность сравниваемых явлений	Сравнивают между собой
Надо иметь целое явление (общее число заболеваний) без среды и часть его (число заболеваний определенной нозологии)	Необходимо иметь среду (численность населения) и явление, произошедшее в данной среде (число больных с определенным заболеванием)
С изменением целого явления его часть может не изменяться	С изменением среды изменяется явление
Явление не связано со средой	Явление связано со средой
Отвечает на вопрос: «Какая часть?»	Отвечает на вопрос: «Как часто?»

Показатель соотношения. Характеризует соотношение между двумя не связанными между собой совокупностями (обеспеченность населения койками, врачами, дошкольными учреждениями, соотношение родов и абортотв,

соотношение врачей и медицинских сестер и др.).

Коэффициенты правдоподобия - это числовые соотношения одноименных показателей структуры, рассчитанные на двух разных совокупностях.

Показатели координации характеризуют соотношение частей целого между собой.

Показатель наглядности характеризует отношение каждой из сравниваемых величин к исходному уровню, который принимается за условную величину (обычно за 100).

Средние величины - обобщающая характеристика признака в статистической совокупности. К ним относятся мода (M_o), медиана (M_e), средняя арифметическая (M).

МЕДИКО-ДЕМОГРАФИЧЕСКИЕ ПОКАЗАТЕЛИ

Демография (от греч. *demos* - народ и *grapho* - писать, изображать) - наука о населении в его общественном развитии.

Население (народонаселение) - совокупность людей, объединенных общностью проживания в пределах страны, части ее территории, группы стран, всего мира.

Медицинская демография - изучает взаимосвязь воспроизводства населения с медико-социальными факторами и разрабатывает на этой основе меры, направленные на обеспечение наиболее благоприятного развития демографических процессов и улучшения здоровья населения.

Воспроизводство населения - изменение численности, состава и размещения населения, обусловленное рождаемостью, смертностью, браками и разводами.

Рождаемость - число родившихся живыми, приходящихся на 1000 населения.

Общий показатель рождаемости = число родившихся живыми \times 1000 / численность населения

Смертность - число умерших, приходящихся на 1000 населения.

Общий показатель смертности = число умерших \times 1000 / численность населения.

Показатель естественного прироста - число естественного прироста, приходящегося на 1000 населения или разница между показателем рождаемости и смертности.

Показатель естественного прироста = Показатель рождаемости - показатель смертности

Средняя продолжительность предстоящей жизни - это число лет, которое в среднем предстоит прожить данному поколению, родившемуся в данном году при условии, что на протяжении их жизни сохраняются повозрастные показатели смертности данного года.

Плодовитость (фертильность) - число детей, родившихся живыми, приходящееся на 1000 женщин в возрасте 15 - 49 лет (детородный возраст).

Плодовитость = число родившихся живыми \times 1000 / численность женщин в возрасте 15 - 49 лет

Повозрастная плодовитость - это число детей, родившихся живыми у женщин фертильного возраста, относящееся к численности женщин данного возраста.

Повозрастная плодовитость = число родившихся живыми \times 1000 / повозрастная численность женщин в возрасте 15 - 49 лет

Брачная плодовитость - число рождений, приходящееся на 1000 замужних женщин в возрасте 15-49 лет.

Показатель брачной плодовитости = число родившихся живыми у женщин, состоящих в браке \times 1000 / численность женщин в возрасте 15 - 49 лет, состоящих в браке

Повозрастная брачная плодовитость - число детей, родившихся живыми у женщин данного возраста, к числу женщин данного возраста, состоящих в браке

Показатель повозрастной брачной плодовитости = число родившихся живыми у женщин данного возраста, состоящих в браке \times 1000 / численность женщин в возрасте 15 - 49 лет, состоящих в браке

Младенческая смертность - число умерших в возрасте до 1 года, приходящееся на 1000 родившихся (или 1/3 в предыдущем году + 2/3 в данном году).

Годовой показатель младенческой смертности = $\frac{\text{число умерших детей до 1 года в данном календарном году} \times 1000}{\text{число родившихся в данном году}}$

или

Годовой показатель младенческой смертности = $\frac{\text{число умерших детей до 1 года в данном календарном году} \times 1000}{(\text{число родившихся за } 2/3 \text{ данного года} + \text{число родившихся за } 1/3 \text{ предыдущего года})}$

Мертворождаемость - это число гибели плода от 22 недель беременности до рождения, приходящееся на 1000 живорожденных

Мертворождаемость = $\frac{\text{число родившихся мертвыми}}{(\text{число родившихся мертвыми} + \text{число родившихся живыми})}$

Показатели смертности по полу = $\frac{\text{число умерших мужчин/женщин} \times 1000}{\text{численность мужского/женского населения}}$

и др.

Показатель заболеваемости - это уровень и структура заболеваний среди населения

ОСНОВНЫЕ ПОКАЗАТЕЛИ ЗАБОЛЕВАЕМОСТИ

Первичная заболеваемость = $\frac{\text{Число впервые выявленных заболеваний за год} \times 1000}{\text{численность населения}}$

Болезненность = $\frac{\text{Число всех имеющихся заболеваний у населения за год} \times 1000}{\text{численность населения}}$

Патологическая пораженность = $\frac{\text{Число заболеваний, выявленных на медосмотрах} \times 100}{\text{численность населения, прошедших медосмотр}}$

Специальные показатели заболеваний по возрастным группам = $\frac{\text{Число заболеваний у лиц определенного возраста} \times 1000}{\text{численность населения того же возраста}}$

Структура заболеваемости = $\frac{\text{Число заболеваний определенной болезни} \times 100}{\text{Число всех имеющихся заболеваний у населения за год}}$

ПОКАЗАТЕЛИ ИНВАЛИДИЗАЦИИ НАСЕЛЕНИЯ

Инвалидность - это постоянная (или длительная) потеря или значительное ограничение трудоспособности

Различают несколько групп инвалидности:

- I группа - лица с полной потерей трудоспособности
- II группа - лица со значительной утратой трудоспособности
- III группа - лица с ограничением трудовых функций.

Причины инвалидности: общее заболевание, профессиональное заболевание, трудовое увечье, инвалидность с детства, инвалидность до начала трудовой деятельности, ранение и заболевание на военной службе.

Первичная инвалидность = $\frac{\text{Число впервые признанных инвалидами за год} \times 10000}{\text{численность работающих}}$

Структура первичной инвалидности = $\frac{\text{Число признанных инвалидами определенной группы (болезни/причины/возраста и т.д.)} \times 100}{\text{число признанных инвалидами}}$

Общая инвалидность = $\frac{\text{число инвалидов} \times 10000}{\text{численность работающих}}$

8. Вопросы по теме занятия

1. Что такое динамический ряд?
2. Что такое статистический показатель?
3. Перечислите основные медико-демографические показатели

4. Перечислите типы динамических рядов
5. Перечислите основные показатели заболеваемости

9. Тестовые задания по теме с эталонами ответов

1. ДИНАМИЧЕСКИЙ РЯД, КОТОРЫЙ СОДЕРЖИТ ЗНАЧЕНИЯ ПОКАЗАТЕЛЕЙ ЗА ОПРЕДЕЛЕННЫЕ ПЕРИОДЫ ВРЕМЕНИ:

- 1) динамический интервальный ряд;
 - 2) динамический моментный ряд;
 - 3) динамический абсолютный ряд;
 - 4) динамический ряд отношений;
 - 5) динамический частотный ряд;
2. ОТНОШЕНИЕ ЧИСЛА ЗАБОЛЕВАНИЙ, ВЫЯВЛЕННЫХ НА МЕДОСМОТРАХ, К ОПРЕДЕЛЕННОМУ ЧИСЛУ ОСМОТРЕННЫХ - ЭТО:
- 1) заболеваемость;
 - 2) первичная заболеваемость;
 - 3) болезненность;
 - 4) патологическая пораженность;
 - 5) истинная заболеваемость;
3. ПОКАЗАТЕЛЬ ДИНАМИЧЕСКОГО РЯДА, РАВНЫЙ РАЗНОСТИ МЕЖДУ КАЖДЫМ ПОСЛЕДУЮЩИМ (ИЗУЧАЕМОМ) И ПРЕДЫДУЩИМ УРОВНЯМИ:
- 1) абсолютный прирост;
 - 2) темп роста;
 - 3) темп прироста;
 - 4) относительный прирост;
 - 5) абсолютный рост;
4. ОТНОШЕНИЕ ЧИСЛА ВСЕХ ЗАБОЛЕВАНИЙ В ДАННОМ ГОДУ К ОПРЕДЕЛЕННОЙ ЧИСЛЕННОСТИ НАСЕЛЕНИЯ - ЭТО:
- 1) заболеваемость;
 - 2) первичная заболеваемость;
 - 3) болезненность;
 - 4) патологическая пораженность;
 - 5) истинная заболеваемость;
5. ПОКАЗАТЕЛЬ НАГЛЯДНОСТИ ХАРАКТЕРИЗУЕТ:
- 1) внутреннюю структуру явления, разбиение его на составные части, удельный вес каждой части в целом;
 - 2) уровень, распространенность какого-либо явления в среде, непосредственно связанного с ней;
 - 3) соотношение между двумя не связанными между собой совокупностями;
 - 4) соотношение частей целого между собой;
 - 5) отношение каждой из сравниваемых величин к исходному уровню, который принимается за условную величину (обычно за 100);

10. Ситуационные задачи по теме с эталонами ответов

1. Численность постоянного населения г. Красноярска на 1 января 2018 года составляла 1093860 человек, на 1 января 2019 - 1096086, на 1 января 2020 - 1094548.

Вопрос 1: Вычислите абсолютный прирост численности постоянного населения;

Вопрос 2: Вычислите темп роста численности постоянного населения;

1) Абсолютный прирост в 2019 = $1096086 - 1093860 = 2226$. Абсолютный прирост в 2020 = $1094548 - 1096086 = -1538$;

2) В 2019 году темп роста = $1096086 \cdot 100 / 1093860 = 100,2\%$ В 2020 году темп роста = $1094548 \cdot 100 / 1096086 = 99,9\%$;

2. Число родившихся человек в 2019 30177, число умерших детей в возрасте до 1 года - 196.

Вопрос 1: Рассчитайте младенческую смертность;

Вопрос 2: Рассчитайте плодовитость, если число женщин в возрасте 15 - 49 лет (детородный возраст) составляет 705483;

1) Младенческая смертность = $196 \cdot 1000 / 30177 = 6,5$;

2) Плодовитость = $30177 \cdot 1000 / 705483 = 42,8$;

11. Примерная тематика НИРС по теме

1. Модели ARMA
2. Прерванные временные ряды
3. Показатели физического развития

12. Рекомендованная литература по теме занятия

- **обязательная:**

Медик, В. А. [Математическая статистика в медицине](#) : учебное пособие для вузов : в 2 т. / В. А. Медик, М. С. Токмачев. - 2-е изд., перераб. и доп. - Москва : Юрайт, 2023. - Т. 1. - 471 с. - Текст : электронный.

Медик, В. А. [Математическая статистика в медицине](#) : учебное пособие для вузов : в 2 т. / В. А. Медик, М. С.

Токмачев. - 2-е изд. - Москва : Юрайт, 2023. - Т. 2. - 347 с. - Текст : электронный.

- дополнительная:

Наркевич, А. Н. [Статистические методы исследования в медицине и биологии](#) : учеб. пособие / А. Н. Наркевич, К. А. Виноградов, К. В. Шадрин ; Красноярский медицинский университет. - Красноярск : КрасГМУ, 2018. - 109 с. - Текст : электронный.

Малугин, В. А. [Математическая статистика](#) : учебное пособие для вузов / В. А. Малугин. - Москва : Юрайт, 2023. - 218 с. - Текст : электронный.

[Основы статистического анализа в медицине](#) : учебное пособие / ред. А. В. Решетников. - Москва : Медицинское информационное агентство, 2020. - 176 с. - Текст : электронный.

- электронные ресурсы:

База знаний WolframAlpha (<https://www.wolframalpha.com/examples/Statistics.html>)

Дистанционные курсы по статистике, анализу данных (<https://welcome.stepik.org/ru>)

Онлайн-калькулятор для статистики (http://math.semestr.ru/group/group_manual.php)

1. Тема № 9. Методы сравнения нескольких групп (в интерактивной форме)

2. Разновидность занятия: комбинированное

3. Методы обучения: объяснительно-иллюстративный, частично-поисковый (эвристический)

4. Значение темы (актуальность изучаемой проблемы): Статистическая гипотеза представляет собой некоторое предположение о законе распределения случайной величины или о параметрах этого закона, формулируемое на основе выборки. Статистическая проверка гипотез проводится с помощью некоторого статистического критерия по общей логической схеме.

5. Цели обучения

- **обучающийся должен знать** методы анализа медико-статистической информации и интерпретации результатов состояния здоровья пациента (населения), **уметь** проверять гипотезы о различии групп, **владеть** навыками проверки гипотезы о параметрах совокупности данных или для их оценки параметрическими критериями

6. Место проведения и оснащение занятия:

- **место проведения занятия:** компьютерный класс №1 (3-03)

- **оснащение занятия:** видеопроектор, доска магнитно-маркерная, комплект учебной мебели на посадочные места, локальный сетевой сервер, персональные компьютеры, экран

7. Аннотация (краткое содержание темы)

Вопрос о наличии различий между выборками является одним из важнейшим вопросом, возникающем при анализе двух выборок. Обычно для этого проводят проверку статистических гипотез о принадлежности обеих выборок одной генеральной совокупности или о равенстве средних.

Если вид распределения или функция распределения выборки известны, то в этом оценка различий двух групп независимых наблюдений может решаться с использованием **параметрических критериев** статистики: либо критерия Стьюдента (t), если сравнение выборок ведется по средним значениям (X и Y), либо с использованием критерия Фишера (F), если сравнение выборок ведется по их дисперсиям. В ином случае пользуются непараметрическими критериями - критерия Манна-Уитни, Вилкоксона, критерий знаков.

Перед применением параметрических критериев нужно обязательно проверить распределение на нормальность!

ПАРАМЕТРИЧЕСКИЕ КРИТЕРИИ. Т-КРИТЕРИЙ СТЬЮДЕНТА

Данный критерий был разработан Уильямом Госсетом для оценки качества пива в компании Гиннесс. В связи с обязательствами перед компанией по неразглашению коммерческой тайны (руководство Гиннесса считало таковой использование статистического аппарата в своей работе), статья Госсета вышла в 1908 году в журнале «Биометрика» под псевдонимом «Student» (Студент).

ТРЕБОВАНИЯ К ДАННЫМ

- Исходные данные имели нормальное распределение (если выборка меньше 30).
- В случае применения двухвыборочного критерия для независимых выборок также необходимо соблюдение условия равенства дисперсий.

ОДНОВЫБОРОЧНЫЙ Т-КРИТЕРИЙ

t -критерий Стьюдента для одной выборки используется для сравнения среднего значения ($M_{ср}$) генеральной совокупности, из которой была взята выборка, с некоторой известной заранее величиной (A).

Нулевая гипотеза: $M_{ср} = A$

$$t = \frac{M - A}{\sigma \sqrt{N}}$$

где M — среднее арифметическое изучаемой выборки, A — нормативный показатель, с которым происходит сравнение, N — объем исследуемой выборки, σ^2 — дисперсия исследуемой выборки.

Формула расчет степеней свободы t-критерия Стьюдента для одной выборки: $df=N-1$.

ЧИСЛО СТЕПЕНЕЙ СВОБОДЫ

Степень свободы (degrees of freedom = df) - количество значений в итоговом вычислении статистики, способных варьироваться.

Одновыборочный t-критерий в R: t.test(Выборка, mu = A)

Одновыборочный t-критерий в PSPP: Анализ - Сравнение средних - Т-проверка по одной выборке

T-КРИТЕРИЙ ДЛЯ НЕЗАВИСИМЫХ ВЫБОРОК

H_0 : обе эти выборки происходят из нормально распределенных генеральных совокупностей с одинаковыми средними значениями:

$$H_1: M_1 \neq M_2$$

t-статистика для проверки нулевой гипотезы равна

$$t = \frac{M_1 - M_2}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}}$$

Эта статистика при справедливости нулевой гипотезы имеет распределение t(df), где

$$df = \frac{\left(\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}\right)^2}{\left(\frac{\sigma_1^2}{N_1}\right)^2 / (N_1 - 1) + \left(\frac{\sigma_2^2}{N_2}\right)^2 / (N_2 - 1)}$$

t-критерий в MS Excel: Мастер функций - Категория Статистические - функция TТЕСТ. Массив 1, Массив 2 - значения выборок

Хвосты - тип распределения (двусторонний - 2; односторонний - 1)

Тип - тип выборок (зависимые - 1, независимые - 2 или 3)

t-критерий в R: t.test(Выборка ~ группирующая переменная); t.test(Выборка 1, выборка 2)

t-критерий для независимых выборок в PSPP: Анализ - Сравнение средних - Т-проверка по одной выборке. Обязательно Назначить группы

T-КРИТЕРИЙ ДЛЯ ЗАВИСИМЫХ ВЫБОРОК

Формула для эмпирического значения критерия t-Стьюдента отражает тот факт, что единицей анализа различий является разность (сдвиг) значений признака для каждой пары наблюдений.

$$t = \frac{|M_d|}{\sigma_d / \sqrt{N}}$$

где d_i - разности пар показателей, M_d - средняя разность значений; σ_d - стандартное отклонение разностей. Степень свободы вычисляется $df = N - 1$.

t-критерий для зависимых выборок в R: t.test(Выборка ~ группирующая переменная, paired = TRUE) или t.test(Выборка 1, Выборка 2, paired = TRUE)

t-критерий для зависимых выборок в PSPP: Анализ - Сравнение средних - Т-проверка парных выборок.

Двусторонние и односторонние t-тесты в R: t.test()

Аргумент alternative:

- "two.sided" ("двухсторонний"; выбирается программой по умолчанию);
- "greater" ("больше");
- "less" ("меньше").

ПАРАМЕТРИЧЕСКИЕ КРИТЕРИИ. КРИТЕРИЙ ФИШЕРА

F-критерий Фишера является параметрическим критерием и используется для сравнения дисперсий двух вариационных рядов.

Проверяемая статистическая гипотеза H_0 : $\sigma_1^2 = \sigma_2^2$

При ее отклонении принимается альтернативная гипотеза о том, что одна дисперсия больше другой.

Формула для эмпирического значения критерия F-Фишера

$$F = \frac{\sigma_1^2}{\sigma_2^2}$$

где σ_1^2 - большая дисперсия, σ_2^2 - меньшая дисперсия.
 $df_{\text{числ}} = N_1 - 1$, $df_{\text{знам}} = N_2 - 1$.

Тест Фишера в R: var.test(Выборка ~ группирующая переменная) или var.test(Выборка 1, выборка 2)

НЕПАРАМЕТРИЧЕСКИЕ КРИТЕРИИ. КРИТЕРИЙ МАННА-УИТНИ

U-критерий Манна — Уитни (англ. Mann — Whitney U-test) — статистический критерий, используемый для оценки различий между двумя независимыми выборками по уровню какого-либо признака, измеренного количественно. Позволяет выявлять различия в значении параметра между малыми выборками

Данный метод выявления различий между выборками был предложен в 1945 году Фрэнком Уилкоксоном (F. Wilcoxon). В 1947 году он был существенно переработан и расширен Х. Б. Манном (H. B. Mann) и Д. Р. Уитни (D. R. Whitney), по именам которых сегодня обычно и называется. Этот метод определяет, достаточно ли мала зона перекрещивающихся значений между двумя рядами (ранжированным рядом значений параметра в первой выборке и таким же во второй выборке). Чем меньше значение критерия, тем вероятнее, что различия между значениями параметра в выборках достоверны.

Критерий Манна-Уитни:

- Проверяет гипотезу о равенстве средних.
- Может проверять гипотезу о сдвиге.
- Не проверяет гипотезу об однородности распределений.
- Не проверяет гипотезу о равенстве медиан.
- Является непараметрическим аналогом параметрического критерия Стьюдента. Если выборки нормальные, то лучше использовать критерий Стьюдента, так как он более мощный.

Ограничения применимости критерия Манна-Уитни:

- В каждой из выборок должно быть не менее 3 значений признака. Допускается, чтобы в одной выборке было два значения, но во второй тогда не менее пяти;
- В выборочных данных не должно быть совпадающих значений (все числа — разные) или таких совпадений должно быть очень мало.

Использование критерия Манна-Уитни

Составить единый ранжированный ряд из обеих сопоставляемых выборок, расставив их элементы по степени нарастания признака и приписав меньшему значению меньший ранг. Общее количество рангов получится равным: $N = n_1 + n_2$,

где n_1 , n_2 - количество элементов в первой и второй выборках соответственно.

- Разделить единый ранжированный ряд на два, состоящие соответственно из единиц первой и второй выборок.
- Подсчитать отдельно сумму рангов, пришедшихся на долю элементов первой выборки, и отдельно — на долю элементов второй выборки.
- Определить большую из двух ранговых сумм (T_x), соответствующую выборке с n_x элементами.
- Определить значение U-критерия Манна — Уитни по формуле:

$$U = n_1 \cdot n_2 + \frac{n_x \cdot (n_x + 1)}{2} - T_x$$

- По таблице критических значений статистической значимости определить критическое значение для данных.
- Если полученное значение U меньше табличного или равно ему, то признается наличие существенного различия между уровнем признака в рассматриваемых выборках (принимается альтернативная гипотеза).
- Если же полученное значение U больше табличного, принимается нулевая гипотеза. Достоверность различий тем выше, чем меньше значение U.
- При справедливости нулевой гипотезы критерий имеет математическое ожидание $M(U) = n_1 n_2 / 2$ и дисперсию $D(U) = n_1 n_2 (n_1 + n_2) / 12$ и при достаточно большом объеме выборочных данных ($n_1 > 19$, $n_2 > 19$) распределен

практически нормально.

НЕПАРАМЕТРИЧЕСКИЕ КРИТЕРИИ. Т-КРИТЕРИЙ ВИЛКОКСОНА

Т-Критерий Вилкоксона — непараметрический статистический тест (критерий), используемый для проверки различий между двумя выборками парных измерений. Впервые предложен Фрэнком Уилкоксоном

Критерий применим в тех случаях, когда признаки измерены, по крайней мере, в порядковой шкале. Целесообразно применять данный критерий, когда величина самих сдвигов варьирует в некотором диапазоне (10-15% от их величины). Это объясняется тем, что разброс значений сдвигов должен быть таким, чтобы появлялась возможность их ранжирования. В случае если сдвиги незначительно отличаются между собой, и принимают какие-то конечные значения, например, +1, -1 и 0, формальных препятствий к применению критерия нет, но, ввиду большого числа одинаковых рангов, ранжирование утрачивает смысл, и те же результаты проще было бы получить с помощью критерия знаков.

- Суть метода состоит в том, что мы сопоставляем абсолютные величины выраженности сдвигов в том или ином направлении.
- Для этого сначала все абсолютные величины сдвигов ранжируются, а потом суммируются ранги.
- Если сдвиги в ту или иную сторону происходят случайно, то и суммы их рангов окажутся примерно равны.
- Если же интенсивность сдвигов в одну сторону больше, то сумма рангов абсолютных значений сдвигов в противоположную сторону будет значительно ниже, чем это могло бы быть при случайных изменениях.
- Минимальное значение величины $W = n(n+1)/2$, где n - объём второй выборки. Максимальное значение величины $W = n(n+1)/2 + mn$, где n - объём второй выборки, m - объём первой выборки.
- Объём выборки — от 5 до 50 элементов. Нулевые сдвиги исключаются из рассмотрения. (Это требование можно обойти, переформулировав вид гипотезы. Например: сдвиг в сторону увеличения значений превышает сдвиг в сторону их уменьшения и тенденцию к сохранению на прежнем уровне).
- Сдвиг в более часто встречающемся направлении принято считать «типичным», и наоборот.

Алгоритм

- Составить список испытуемых в любом порядке, например, алфавитном.
- Вычислить разность между индивидуальными значениями во втором и первом замерах. Определить, что будет считаться типичным сдвигом.
- Согласно алгоритму ранжирования, проранжировать абсолютные величины разностей, начисляя меньшему значению меньший ранг, и проверить совпадение полученной суммы рангов с расчетной.
- Отметить каким-либо способом ранги, соответствующие сдвигам в нетипичном направлении. Подсчитать их сумму T .
- Определить критические значения T для данного объема выборки. Если $T_{эмп}$ меньше или равен $T_{кр}$ - сдвиг в «типичную» сторону достоверно преобладает. Фактически оцениваются знаки значений, полученных вычитанием ряда значений одного измерения из другого. Если в результате количество снизившихся значений примерно равно количеству увеличившихся, то гипотеза о нулевой медиане подтверждается.

НЕПАРАМЕТРИЧЕСКИЕ КРИТЕРИИ. G-КРИТЕРИЙ ЗНАКОВ

G-критерий Знаков предназначен для исследования определения направления сдвига в значениях исследуемого признака в двух выборках. Критерий позволяет определить изменяются ли значения переменной при переходе от одного измерения к другому в сторону улучшения или сторону ухудшения. Он не предназначен для определения интенсивности сдвигов

Ограничения критерия: количество человек в обоих замерах должно варьироваться от 5 до 300

АЛГОРИТМ

- Составить таблицу значений двух выборок.
- Парно вычесть из значений второй переменной значения первой переменной.
- Подсчитать количество нулевых сдвигов.
- Подсчитать общее количество значений (без нулевых сдвигов). Считать это число как n
- Убедиться, что количество значений в выборке варьируется от 5 до 300
- Подсчитать количество «отрицательных» и «положительных» сдвигов.
- Считать «типичными» те сдвиги, количество которых больше.
- Считать эмпирическим значением G то количество сдвигов, которых меньше.
- По таблице критические значения g -критерия Знаков определить $G_{кр}$.
- Сопоставить между собой $G_{кр}$ и $G_{эмп}$. Если $G_{эмп}$ меньше $G_{кр}$, то сдвиг и «типичную» сторону достоверен.

Непараметрический критерий сравнения двух выборок в R

Тип выборки	Тип теста	Функция
Зависимые	Критерий Вилкоксона	<code>wilcox.test(x1, x2, paired = T)</code>
Независимые	Критерий Манна-Уитни	<code>wilcox.test(x1, x2, paired = F)</code>

Проверяемая нулевая гипотеза: нет статистически значимых различий в выборках. Если в результате теста p -value меньше 0,05, то нулевая гипотеза отвергается и принимается альтернативная на уровне статистической значимости 0,05.

Дисперсионный анализ — статистический метод, направленный на поиск зависимостей в экспериментальных данных путём исследования значимости различий в средних значениях.

Суть дисперсионного анализа сводится к изучению влияния одной или нескольких независимых переменных, обычно именуемых факторами, на зависимую переменную. Зависимые переменные представлены значениями абсолютных шкал (шкала отношений). Независимые переменные являются номинативными (шкала наименований), то есть отражают групповую принадлежность, и могут иметь два или более значения (типа, градации или уровня).

Градации, соответствующие независимым выборкам объектов, называются межгрупповыми, а градации, соответствующие зависимым выборкам, — внутригрупповыми.

В зависимости от типа и количества переменных различают:

- однофакторный и многофакторный дисперсионный анализ (одна или несколько независимых переменных);
- одномерный и многомерный дисперсионный анализ (одна или несколько зависимых переменных);
- дисперсионный анализ с повторными измерениями (для зависимых выборок);
- дисперсионный анализ с постоянными факторами, случайными факторами, и смешанные модели с факторами обоих типов.

Процедура дисперсионного анализа состоит в определении соотношения систематической (межгрупповой) дисперсии к случайной (внутригрупповой) дисперсии в измеряемых данных. В качестве показателя изменчивости используется сумма квадратов отклонения значений параметра от среднего.

ОДНОФАКТОРНЫЙ ДИСПЕРСИОННЫЙ АНАЛИЗ

Принципы дисперсионного анализа (англ. analysis of variance, ANOVA) были разработаны в 1920-х гг. Рональдом Эйлером Фишером.

Одномерный однофакторный анализ может применяться для двух или нескольких независимых групп, когда все группы объединены по одному признаку. В ходе анализа проверяется нулевая гипотеза о равенстве средних. В случае анализа двух групп дисперсионный анализ тождественен двухвыборочному t -критерию Стьюдента для независимых выборок, и величина F -статистики равна квадрату соответствующей t -статистики.

Для подтверждения положения о равенстве дисперсий обычно применяется критерий Ливена (*Levene's test*). В случае отвержения гипотезы о равенстве дисперсий основной анализ неприменим. Если дисперсии равны, то для оценки соотношения межгрупповой и внутригрупповой изменчивости применяется F -критерий Фишера.

МНОГОФАКТОРНЫЙ ДИСПЕРСИОННЫЙ АНАЛИЗ

Многофакторный анализ позволяет проверить влияние нескольких факторов на зависимую переменную. В отличие от однофакторной модели, где имеется одна межгрупповая сумма квадратов, модель многофакторного анализа включает суммы квадратов для каждого фактора в отдельности и суммы квадратов всех взаимодействий между ними.

Дисперсионный анализ в R: `aov()`

НЕПАРАМЕТРИЧЕСКИЕ КРИТЕРИИ СРАВНЕНИЯ НЕСКОЛЬКИХ ГРУПП

Критерий Краскела-Уоллиса. Данный метод - ранговый. Ранг - порядковый номер конкретного наблюдения в ряду упорядоченных по возрастанию (или убыванию - не важно) наблюдений. Проверяемая гипотеза заключается в отсутствии разницы между сравниваемыми независимыми группами (каждая группа имеет одинаковое распределение величин в совокупности).

Например, исследуется влияние нового препарата на пациентов, разделенных по возрастному принципу: дети,

взрослые, пожилые. Влияние оценивается по 10-балльной шкале. Здесь, нулевая гипотеза: препарат одинаковое воздействие вне зависимости от возрастной группы, все возникшие различия вызваны случайным фактором.

Данный критерий, как и однофакторный дисперсионный анализ, позволяет сделать заключение только о наличии или об отсутствии значимых различий. Ни один из этих методов сам по себе не позволяет сказать, где именно лежат различия.

Критерий Краскелла-Уоллиса в R: **kruskal.test()**

В случае зависимых выборок используется критерий Фридмана. В качестве нулевой гипотезы H_0 выступает следующая: "между полученными в разных условиях измерениями имеются лишь случайные различия".

Критерий Фридмана: **friedman.test(x)**

ПОПРАВКА НА МНОЖЕСТВЕННОЕ СРАВНЕНИЕ

Поправка на множественную проверку гипотез — способ устранения эффекта множественных сравнений, возникающего при необходимости построения семейства статистических выводов. Другими словами, при одновременной проверке большого числа гипотез на одном и том же наборе данных вероятность сделать неверное заключение в отношении хотя бы одной из этих гипотез значительно превышает изначально принятый уровень значимости ($p = 0.05$). Для устранения этого эффекта существует большой арсенал методов, различающихся по своей мощности и применимости в разных ситуациях.

Первым упоминанием о проблеме множественных сравнений можно найти в работе *Exposition de La Theorie Des Chances Et Des Probabilites* (1843) Антуана Августина Курно о том, что при разделении популяции на как можно большее число групп, рано или поздно найдется группа, значимо отличная от остальной совокупности. Тогда проблема была сочтена нерешаемой.

Следующая после работы Бонферрони (*Teoria statistica delle classi e calcolo delle probabilita*, 1936) волна интереса к проблеме множественного тестирования возникла в 1950-х годах в связи с работой Джона Тьюки и Генри Шеффе. Последующие работы были направлены на увеличение мощности поправок. Так, более мощная поправка Холма-Бонферрони была разработана в 1979 году. В 1995 году со статьи Бенджамини и Хохберга начались работы по FDR (доле ложных отклонений гипотез), что дало возможность тестировать большое количество гипотез.

Поправка Бонферрони

Метод поправки Бонферрони утверждает, что для уменьшения ложноположительных результатов необходимо отклонить те гипотезы, для которых p -value по критерию $p_i < \alpha/m$. Данная поправка позволяет получить групповую вероятность ошибки первого рода $\leq \alpha$.

При увеличении m в результате применения поправки Бонферрони мощность статистической процедуры резко уменьшается — шансы отклонить неверные гипотезы падают.

Метод Холма (поправка Холма — Бонферрони)

Равномерно более мощный, чем поправка Бонферрони, и решает проблему падения мощности при росте числа гипотез.

Пусть $p_1 \leq p_2 \leq \dots \leq p_m$ упорядочены, H_1, \dots, H_m - соответствующие этим p_i гипотезы. Процедура Холма определена следующим образом:

1. Если $p_1 > \alpha/(m-1+1)$, то нужно принять гипотезы H_1, \dots, H_m и остановиться. Иначе, если $p_1 < \alpha/(m-1+1)$, отвергнуть гипотезу H_1 и продолжить проверку оставшихся гипотез на уровне значимости $\alpha/(m-2+1)$
2. Если $p_2 > \alpha/(m-2+1)$, то нужно принять гипотезы H_1, \dots, H_m и остановиться. Иначе, если $p_2 < \alpha/(m-2+1)$, отвергнуть гипотезу H_2 и продолжить проверку оставшихся гипотез на уровне значимости $\alpha/(m-3+1)$
3. И т.д.

8. Вопросы по теме занятия

1. В каких случаях применяется t -критерий Стьюдента?
2. В каких случаях применяется критерий Фишера?
3. Для чего используется T -Критерий Вилкоксона?
4. Для чего используется U -критерий Манна — Уитни?
5. Для чего применяется критерий Краскела-Уоллиса?
6. В каких случаях применяется критерий Фридмана?

9. Тестовые задания по теме с эталонами ответов

1. ОГРАНИЧЕНИЕМ ДЛЯ ПРИМЕНЕНИЯ Т-КРИТЕРИЯ СТЬЮДЕНТА ЯВЛЯЕТСЯ:

- 1) неравенство средних значений;
 - 2) значительные различия среднеквадратичных отклонений;
 - 3) большие объемы выборок;
 - 4) несоответствие эмпирического распределения нормальному;
 - 5) неравенство медиан в выборках;
2. СТАТИСТИЧЕСКИЙ КРИТЕРИЙ, ИСПОЛЬЗУЕМЫЙ ДЛЯ ОЦЕНКИ РАЗЛИЧИЙ МЕЖДУ ДВУМЯ НЕЗАВИСИМЫМИ ВЫБОРКАМИ ПО УРОВНЮ КАКОГО-ЛИБО ПРИЗНАКА, ИЗМЕРЕННОГО КОЛИЧЕСТВЕННО:

- 1) критерий значимости;
 - 2) критерий Кайзера;
 - 3) критерий доли воспроизводимой дисперсии;
 - 4) критерий каменистой осыпи;
 - 5) U-критерий Манна — Уитни;
3. НЕПАРАМЕТРИЧЕСКИЙ СТАТИСТИЧЕСКИЙ ТЕСТ (КРИТЕРИЙ), ИСПОЛЬЗУЕМЫЙ ДЛЯ ПРОВЕРКИ РАЗЛИЧИЙ МЕЖДУ ДВУМЯ ВЫБОРКАМИ ПАРНЫХ ИЗМЕРЕНИЙ:

- 1) Т-Критерий Вилкоксона;
 - 2) критерий Кайзера;
 - 3) критерий доли воспроизводимой дисперсии;
 - 4) критерий каменистой осыпи;
 - 5) критерий интерпретируемости и инвариантности;
4. НЕПАРАМЕТРИЧЕСКИЙ КРИТЕРИЙ, КОТОРЫЙ ИСПОЛЬЗУЕТСЯ ДЛЯ СРАВНЕНИЯ НЕСКОЛЬКИХ (БОЛЬШЕ ДВУХ) НЕЗАВИСИМЫХ ГРУПП:

- 1) критерий Манна-Уитни;
 - 2) критерий Краскела-Уоллиса;
 - 3) критерий Фридмана;
 - 4) критерий Стьюдента для независимых выборок;
 - 5) критерий Вилкоксона;
5. НЕПАРАМЕТРИЧЕСКИЙ КРИТЕРИЙ, КОТОРЫЙ ИСПОЛЬЗУЕТСЯ ДЛЯ СРАВНЕНИЯ НЕСКОЛЬКИХ (БОЛЬШЕ ДВУХ) ЗАВИСИМЫХ ГРУПП:

- 1) критерий Манна-Уитни;
- 2) критерий Краскела-Уоллиса;
- 3) критерий Фридмана;
- 4) критерий Стьюдента для зависимых выборок;
- 5) критерий Вилкоксона;

10. Ситуационные задачи по теме с эталонами ответов

1. В ходе некоторого медицинского эксперимента исследовалось влияние препарата на показатели крови. Имеются результаты показателей группы А, где препарат не применялся, и группы В, где препарат применялся.

Вопрос 1: Какой критерий проверки различий в выборках следует применить, если известно, что обе выборки распределены не нормально?;

Вопрос 2: Сформулируйте проверяемую нулевую гипотезу;;

- 1) Так как выборки распределены не нормально и они независимые, то для выявления статистических различий в группе воспользуемся критерием Манна-Уитни;
- 2) Проверяемая нулевая гипотеза: между группами нет статистически значимых различий в исследуемом показателе (то есть препарат не оказывает влияние на показатели крови);

2. В течение пяти лет использовались три различных технологии (А, В, С) по созданию донорских органов.

Вопрос 1: Какой критерий для проверки влияния технологии на качество получаемых органов стоит выбрать, если выборки данных распределены нормально;

Вопрос 2: Какой критерий для проверки влияния технологии на качество получаемых органов стоит выбрать, если выборки данных распределены не нормально?;

- 1) Однофакторный дисперсионный анализ;;
- 2) Критерий Краскела-Уоллиса;

11. Примерная тематика НИРС по теме

1. Факторный анализ
2. Многофакторный дисперсионный анализ
3. Метод главных компонент

12. Рекомендованная литература по теме занятия

- **обязательная:**

Медик, В. А. [Математическая статистика в медицине](#) : учебное пособие для вузов : в 2 т. / В. А. Медик, М. С. Токмачев. - 2-е изд., перераб. и доп. - Москва : Юрайт, 2023. - Т. 1. - 471 с. - Текст : электронный.

Медик, В. А. [Математическая статистика в медицине](#) : учебное пособие для вузов : в 2 т. / В. А. Медик, М. С. Токмачев. - 2-е изд. - Москва : Юрайт, 2023. - Т. 2. - 347 с. - Текст : электронный.

- **дополнительная:**

Наркевич, А. Н. [Статистические методы исследования в медицине и биологии](#) : учеб. пособие / А. Н. Наркевич, К. А. Виноградов, К. В. Шадрин ; Красноярский медицинский университет. - Красноярск : КрасГМУ, 2018. - 109 с. - Текст : электронный.

Малугин, В. А. [Математическая статистика](#) : учебное пособие для вузов / В. А. Малугин. - Москва : Юрайт, 2023. - 218 с. - Текст : электронный.

[Основы статистического анализа в медицине](#) : учебное пособие / ред. А. В. Решетников. - Москва : Медицинское информационное агентство, 2020. - 176 с. - Текст : электронный.

- электронные ресурсы:

База знаний WolframAlpha (<https://www.wolframalpha.com/examples/Statistics.html>)

Дистанционные курсы по статистике, анализу данных (<https://welcome.stepik.org/ru>)

Онлайн-калькулятор для статистики (http://math.semestr.ru/group/group_manual.php)

1. Тема № 10. Систематизация изученного материала. Зачет (в интерактивной форме)

2. Разновидность занятия: комбинированное

3. Методы обучения: репродуктивный, частично-поисковый (эвристический)

4. Значение темы (актуальность изучаемой проблемы): осуществить контроль знаний, умений, проверить сформированность компетенций.

5. Цели обучения

- **обучающийся должен знать** способы получения информации из различных информационных, библиографических ресурсов, принципы работы современных информационных технологий, методы анализа медико-статистической информации и интерпретации результатов состояния здоровья пациента (населения), современные методики сбора и обработки информации, необходимой для проведения научного исследования, способы представления статистических данных, теоретические основы и технологии организации научно-исследовательской и проектной деятельности, в том числе особенности проведения конкурсов российскими и международными научными фондами; требования к оформлению проектных и исследовательских работ, конкурсной документации, **уметь** осуществлять поиск профессиональной, достоверной, качественной информации в сети интернет с использованием специализированных научных и медицинских библиотек и баз данных, понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности, анализировать полученные в результате вычислений статистические величины (абсолютные, средние, относительные показатели, показатели динамического ряда) на основании сравнения в динамике и с имеющимися нормативами, выбирать способ вычисления коэффициента корреляции, определять характер и силу корреляционной связи, ошибку и достоверность коэффициента корреляции и формулировать выводы, составлять уравнение регрессии (линейной, нелинейной, парной, множественной) и интерпретировать коэффициенты уравнения регрессии, представлять данные с помощью программного обеспечения, организовывать проведение научных конференций, выставок, конкурсов профессионального мастерства, иных конкурсов и мероприятий в условиях цифровой экономики; организовывать научно-исследовательские, проектные работы, консультировать участников работы на всех этапах ее проведения; использовать отечественный и зарубежный опыт и результаты собственных научных исследований в процессе руководства научно-исследовательской, проектной работой, проверять гипотезы о различии групп, применять современные методики сбора и обработки информации, необходимой для проведения научного исследования, формулировать проблемы и гипотезы, **владеть** навыками поиска информации, в том числе и медицинской, в сети интернет, современными информационными технологиями для решения профессиональных задач, навыками анализа полученных в результате вычислений статистических величин, навыками определения корреляционной взаимосвязи между признаками, навыками регрессионного анализа, навыками представления и интерпретирования данных с помощью программного обеспечения, научно-методическими основами организации научно-исследовательской, проектной деятельности; навыками оценивания качества выполнения и оформления проектных, научно-исследовательских работ с применением цифровых технологий, навыками проверки гипотезы о параметрах совокупности данных или для их оценки параметрическими критериями, современными методиками сбора и обработки информации, необходимой для проведения научного исследования, навыками проверки гипотез о распределении по различным критериям

6. Место проведения и оснащение занятия:

- **место проведения занятия:** компьютерный класс №1 (3-03)

- **оснащение занятия:** видеопроектор, доска магнитно-маркерная, комплект учебной мебели на посадочные места, локальный сетевой сервер, персональные компьютеры, экран

7. Аннотация (краткое содержание темы)

Преподаватель распределяет темы собеседования (вопросы к зачёту), список практических навыков, выдает тестовые задания.

8. Вопросы по теме занятия

1. Что такое генеральная совокупность?

2. Что включает в себя этап "Работа с литературой"?

3. Как выглядит уравнение регрессии?

4. Перечислите свободное программное обеспечение для статистического анализа

1) R, PSPP, Gnuplot;

5. Что такое корреляция?

6. Что такое статистика?

9. Тестовые задания по теме с эталонами ответов

1. ОБЪЕКТ ИССЛЕДОВАНИЯ - ЭТО:

1) то, что будет взято для изучения и исследования;

2) особая проблема, отдельные стороны объекта, его свойства и особенности, которые, не выходя за рамки исследуемого объекта, будут исследованы в работе;

3) познавательная конструкция, включающая как предположение, так и способы его фальсификации; обобщающая, прогнозирующая получение определенных новых фактов и систематизирующая их;

4) способ получения сбора, обработки или анализа данных. В исследованиях широко применяются различные методы научного познания из других областей науки и техники;

- 5) утверждения, выражающие в краткой форме содержательные итоги исследования, они в тезисной форме отражают то новое, что получено самим автором;
2. РОСТ ПРИЗЫВНИКОВ (167, 180, 190 .CM) ОТНОСЯТСЯ К СЛЕДУЮЩЕМУ ТИПУ ДАННЫХ:
- 1) количественному;
 - 2) качественному;
 - 3) дискретному;
 - 4) логическому;
 - 5) вариантному;
3. НЕПАРАМЕТРИЧЕСКИЙ СТАТИСТИЧЕСКИЙ ТЕСТ (КРИТЕРИЙ), ИСПОЛЬЗУЕМЫЙ ДЛЯ ПРОВЕРКИ РАЗЛИЧИЙ МЕЖДУ ДВУМЯ ВЫБОРКАМИ ПАРНЫХ ИЗМЕРЕНИЙ:
- 1) Т-Критерий Вилкоксона;
 - 2) критерий Кайзера;
 - 3) критерий доли воспроизводимой дисперсии;
 - 4) критерий каменистой осыпи;
 - 5) критерий интерпретируемости и инвариантности;
4. ФРАНЦУЗСКИЙ ПАЛЕОНТОЛОГ, КОТОРЫЙ ВПЕРВЫЕ ВВЕЛ ТЕРМИН "КОРРЕЛЯЦИЯ":
- 1) Фрэнсис Гальтон;
 - 2) Жорж Кювье;
 - 3) Тома Корнель;
 - 4) Жан Расин;
 - 5) Жан де Лонуа;
5. ШКАЛА ИЗМЕРЕНИЯ — ЭТО:
- 1) ограничение типа отношений между значениями переменных, накладываемое на результаты измерений;
 - 2) совокупность измеренных признаков;
 - 3) полная совокупность элементов;
 - 4) часть совокупности элементов, которая охватывается экспериментом;
 - 5) функция, которая однозначно определяет вероятность;

10. Ситуационные задачи по теме с эталонами ответов

1. Интернет на сегодняшний день представляет собой огромное скопление разнообразной информации, значительная часть которой является не достоверной

Вопрос 1: Расскажите, какие сайты в Интернете содержат достоверную медицинскую информацию?;

Вопрос 2: Почему этим сайтам можно доверять?;

Вопрос 3: Найдите в Интернете не менее 6-ти профессиональных медицинских новостей за последний месяц;

1) Среди сайтов, предоставляющих доступ к базам данных, безусловным лидером является всем известный PubMed MEDLINE (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=PubMed>). На сайте Medscape (<http://www.medscape.com/>) представлено большое количество журналов, доступ к которым предоставляется бесплатно после заполнения регистрационной формы. Портал Medbioworld (<http://www.medbioworld.com/>) является одной из наиболее полных коллекций ссылок на источники профессиональной медицинской информации в сети Интернет. Данный ресурс содержит более 25000 ссылок на медицинские журналы, профессиональные медицинские ассоциации, медицинские словари, нозологические базы данных, клинические испытания, руководства;

2) Им можно доверять, так как они являются специализированными порталами, предоставляющими профессиональную медицинскую информацию, которым доверяет весь мир;

3) Для поиска профессиональных медицинских новостей можно воспользоваться порталом Medbioworld (<http://www.medbioworld.com/>). На главной странице представлена ссылка на профессиональные медицинские новости. . Зайти по ссылке и ознакомиться с последними новостями текущего месяца. News. Professional Medical NewsProfessional Medical News. Health eLineHealth eLine. 2011-01-07. • Moxifloxacin beats clindamycin for some odontogenic infections. • Newer antipsychotics overused, U.S. study suggests. • First-in-class antibiotic has high risk of adverse events and death;

2. Исследуется взаимосвязь между медицинскими показателями А и В. А: 132 122 121 118 118 115 112 111 110 104 102 100 99 98, В: 100 111 115 125 175 105 120 110 147 180 155 153 200 210

Вопрос 1: Какой тип диаграммы следует применить для визуализации имеющихся данных?;

Вопрос 2: Какую команду для построения данной диаграммы нужно ввести в R, если эти данных хранятся в переменной data, в столбцах А и В, соответственно;

1) Диаграмма рассеяния;

2) `plot(data$A~data$B)`;

11. Примерная тематика НИРС по теме

1. Генетический алгоритм
2. Нейронная сеть Кохонена
3. Кластерный анализ и его роль в психологическом исследовании

12. Рекомендованная литература по теме занятия

- обязательная:

Медик, В. А. [Математическая статистика в медицине](#) : учебное пособие для вузов : в 2 т. / В. А. Медик, М. С. Токмачев. - 2-е изд., перераб. и доп. - Москва : Юрайт, 2023. - Т. 1. - 471 с. - Текст : электронный.

Медик, В. А. [Математическая статистика в медицине](#) : учебное пособие для вузов : в 2 т. / В. А. Медик, М. С. Токмачев. - 2-е изд. - Москва : Юрайт, 2023. - Т. 2. - 347 с. - Текст : электронный.

- дополнительная:

Наркевич, А. Н. [Статистические методы исследования в медицине и биологии](#) : учеб. пособие / А. Н. Наркевич, К. А. Виноградов, К. В. Шадрин ; Красноярский медицинский университет. - Красноярск : КрасГМУ, 2018. - 109 с. - Текст : электронный.

Малугин, В. А. [Математическая статистика](#) : учебное пособие для вузов / В. А. Малугин. - Москва : Юрайт, 2023. - 218 с. - Текст : электронный.

[Основы статистического анализа в медицине](#) : учебное пособие / ред. А. В. Решетников. - Москва : Медицинское информационное агентство, 2020. - 176 с. - Текст : электронный.

- электронные ресурсы:

База знаний WolframAlpha (<https://www.wolframalpha.com/examples/Statistics.html>)

Дистанционные курсы по статистике, анализу данных (<https://welcome.stepik.org/ru>)

Онлайн-калькулятор для статистики (http://math.semestr.ru/group/group_manual.php)