

Федеральное государственное бюджетное научное учреждение
«Федеральный исследовательский центр фундаментальной и трансляционной
медицины»

УТВЕРЖДАЮ

Директор ФИЦ ФТМ

_____ М.И. Воевода

02 июня 2023 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
«ВИДЫ ИССЛЕДОВАНИЙ И АДЕКВАТНЫЕ ИМ МЕТОДЫ СТАТИСТИЧЕСКОГО
АНАЛИЗА В МЕДИЦИНЕ»

Специальность:

31.08.07 Патологическая анатомия

Уровень высшего образования:

подготовка кадров высшей квалификации

Новосибирск

2023 г.

Рабочая программа дисциплины «Виды исследования и адекватные им методы статистического анализа в медицине» разработана в соответствии с Федеральным государственным образовательным стандартом высшего образования по специальности 31.08.07 Патологическая анатомия (уровень подготовки кадров высшей квалификации), утверждённым приказом Министерства науки и высшего образования Российской Федерации от 02.02.2022 № 110.

Разработчики программы:

Митрофанов И.М., доктор медицинских наук, ведущий научный сотрудник лаборатории патогенеза соматических заболеваний НИИЭКМ ФИЦ ФТМ.

Рабочая программа дисциплины рассмотрена и одобрена на заседании Ученого совета ФИЦ ФТМ (протокол № 6 от 01 июня 2023 г.)

1. Цель и задачи изучения дисциплины, требования к результатам освоения дисциплины

Цель изучения дисциплины:

Цели изучения дисциплины - формирование знаний теоретических основ медицинской статистики и использование различных статистических методов в научно-исследовательской и практической деятельности обучающихся.

Задачи дисциплины:

- формирование системы знаний по статистической обработке данных в медицине, позволяющих обучающемуся успешно работать в избранной сфере деятельности;
- использование математических методов для обработки информации и анализа медико-биологических данных и экспериментального материала при проведении клинических исследований;
- ознакомление с методами систематизации экспериментального материала и принципами доказательной медицины при интерпретации научных фактов.

Место дисциплины в структуре образовательной программы

Дисциплина «Виды исследования и адекватные им методы статистического анализа в медицине» относится к факультативным дисциплинам части образовательной программы, формируемой участниками образовательных отношений.

Требования к результатам освоения дисциплины (модуля):

Формирование универсальных, общепрофессиональных и профессиональных компетенций у обучающихся в рамках изучения дисциплины (модуля) предполагает овладение системой теоретических знаний по выбранной специальности и формирование соответствующих умений и (или) владений.

Соответствующих умений и (или) владений.		
Код и наименование компетенции, индикатора достижения компетенции	Планируемые результаты обучения по дисциплине	
УК-1. Способен критически и системно анализировать, определять возможности и способы применения достижения в области медицины и фармации в профессиональном контексте		
УК-1.1. Анализирует достижения в области медицины и фармации в профессиональном контексте	Знать:	законы формальной логики, методы критического анализа и синтеза
	Уметь:	анализировать любую ситуацию и находить верные решения
	Владеть:	логическими методами и приемами при анализе информации, полученной в ходе комплексного клинико-морфологического исследования
УК-1.2. Определяет возможности и способы применения достижения в области медицины и фармации в	Знать:	методы и способы оценки возможности и вариантов применения современных достижений в области медицины и фармации.
	Уметь:	определять возможности и способы применения достижений в области медицины

профессиональном контексте		и фармации в профессиональном контексте; сформулировать проблему, выделить ключевые цели и задачи по ее решению
	Владеть:	методами и приёмами системного анализа достижений в области медицины и фармации для их применения в профессиональном контексте
УК-2. Способен разрабатывать, реализовывать проект и управлять им		
УК-2.1. Участвует в разработке и управлении проектом	Знать:	- статистические методы сбора, обработки, анализа и прогнозирования данных; - методы и принципы постановки задач (конкретная, измеримая, достижимая, значимая, ограниченная во времени) и способы их реализации; - основные подходы к организации проектной деятельности.
	Уметь:	- анализировать данные из множественных источников и оценивать качество и достоверность полученной информации по явным и неявным признакам; - применять программное обеспечение (текстовые, графические, табличные и аналитические приложения, приложения для визуального представления данных) для работы с информацией; - собирать, анализировать, систематизировать сведения и данные, документировать требования к проектам; - вести деловые переговоры с целью согласования взаимных интересов участников проекта; - разрабатывать алгоритмы, модели, схемы проекта; - принимать решения при разработке и реализации проекта; - выполнять проектные работы; - оценивать результаты реализации проектной деятельности
	Владеть:	- навыком сбора и анализа исходных данных, необходимых для оценки реализуемости проекта; - навыком определения соответствия целей и задач проекта; - методами и принципами организации проектной деятельности.
УК-2.2. Выполняет задачи в зоне своей ответственности в соответствии с запланированными результатами и точками	Знать:	- процессы и методы управления проектами; - методы определения сроков реализации проектов.
	Уметь:	- осуществлять планирование проекта; - управлять процессом реализации проекта.

контроля, при необходимости корректирует способы решения задач	Владеть:	- навыками планирования проекта, организации исполнения проекта, контроля этапов реализации проекта, оценки результатов проекта.
ПК-3. Способен осуществлять научно-исследовательскую и педагогическую деятельность на основе научных знаний		
ПК-3.1. Планирует научно-исследовательскую деятельность	Знать:	- основные приемы и принципы планирования и протоколирования научных исследований; - этапы проведения научно-исследовательской работы.
	Уметь:	- планировать, организовать самостоятельный исследовательский процесс;
	Владеть:	- навыком проектирования научно-исследовательской работы; - навыком определения объекта и метода исследования.
ПК-3.2. Осуществляет научно-исследовательскую деятельность	Знать:	- методы статистической обработки данных исследования; - понятия абсолютных и относительных показателей; - методы визуализации данных; - виды диаграмм; - технологию публичного выступления; - методологию публичной дискуссии.
	Уметь:	- выбирать метод для статистической обработки; - анализировать и сравнивать полученные данные; - выбирать необходимый вид диаграмм и осуществлять построение графиков; - подготовить доклад в форме презентации с использованием мультимедийной техники; - публично представить полученные результаты; - дискутировать с коллегами по теме научно-исследовательской работы.
	Владеть:	- навыком выбора метода статистической обработки; - навыком анализа и сравнения полученных данных; - приемами и способами визуализации полученных данных; - навыком подготовки презентации по полученным данным с использованием мультимедийной техники; - навыком публичного представления полученных данных; - навыком проведения дискуссии по теме научно-исследовательской работы.

2. Объем дисциплины (модуля) по видам учебной работы

Виды учебной работы	Всего часов	Объем по годам обучения	
		1	2
Контактная работа обучающегося с преподавателем по видам учебных занятий (КР):	24		24
Лекция (Л)	2		2
Семинарское/практическое занятие (СПЗ)	22		22
Самостоятельная работа обучающегося, в том числе подготовка к промежуточной аттестации (СР)	12		12
Вид промежуточной аттестации: Зачет (З), Зачет с оценкой (ЗО), Экзамен (Э)	зачет		
Общий объем дисциплины в часах	36		
Общий объем дисциплины в зачетных единицах	2	2	

3. Содержание дисциплины

Тема 1. Сравнительное исследование (исследование случай-контроль) - анализ различий между группами.

1.1. Сравнительное исследование в психологии и клинической медицине. Примеры. Особенности независимых и зависимых переменных в сравнительном исследовании.

1.2. Структура данных для однофакторного дисперсионного анализа несвязанных выборок. Исходные допущения. Формулировка нулевой гипотезы. Процедура проверки нулевой гипотезы с использованием распределения. F-критерий и t-критерий Стьюдента. Вычисление вероятности ошибки первого рода с использованием таблицы критических значений критерия. Особенности однофакторного дисперсионного анализа с неравными группами. Проверка однородности дисперсий. Критерий Ливена. Процедура проверки нулевой гипотезы. Последствия нарушений исходных допущений.

1.3. Непараметрические аналоги дисперсионного анализа для несвязанных выборок: H-критерий Крускала-Уоллиса, T-критерий Уайта, T-критерий Манна-Уитни, U-критерий Манна-Уитни.

Тема 2. Сравнительное исследование с количеством групп более двух - методы множественного сравнения.

2.1. Значение множественный сравнений в статистическом анализе данных. Поправка Бонферрони. Общий принцип неравенства Бонферрони. Процедура вычисления. Критерий Ньюмена-Кейлса. Общая характеристика метода. Процедура вычисления. Критерий Тьюки. Общая характеристика метода. Процедура вычисления. Отличия получаемых результатов по сравнению с критерием Ньюмена-Кейлса.

2.2. Особенности применения методов множественного сравнения с контролем. Критерий Даннета. Процедура вычисления.

2.3. Особенности применения критериев Ньюмена-Кейлса и Тьюки при непараметрическом анализе. Критерий Данна для непараметрического сравнения выборок разного объема.

Тема 3. Анализ качественных признаков.

3.1. Роль качественных признаков при проведении клинического исследования. Доли (эмпирические частоты), как статистическая оценка качественных признаков. Сравнение долей. Точность оценки долей. Z-критерий. Поправка Йейтса на непрерывность.

3.2. Соответствие эмпирических частот ожидаемым частотам. Таблицы сопряженности, вычисление ожидаемых частот. Критерий хи-квадрат. Процедура вычисления.

3.3. Множественные сравнения качественных признаков. Преобразование таблиц сопряженности. Использование поправки Бонферрони.

3.4. Точный критерий Фишера. Процедура вычисления.

Тема 4. Когортное (лонгитюдное) исследование - анализ повторных наблюдений.

4.1. Особенности последовательного (лонгитюдного) исследования. Преимущества по сравнению со сравнительным.

4.2. Понятие о дисперсионном анализе связанных выборок. Парный t-критерий Стьюдента. Процедура вычисления. Дисперсионный анализ для связанных выборок. Процедура вычисления.

4.3. Особенности множественного сравнения для повторных наблюдений.

4.4. Непараметрические аналоги дисперсионного анализа для связанных выборок. Критерий Уилкоксона для двух наблюдений. Критерий Фридмана. Множественные сравнения повторных наблюдений.

Тема 5. Корреляционное исследование - меры связи, регрессионный анализ, корреляционный анализ. Различные меры корреляционной связи.

5.1. Корреляционное исследование - особенности, причины широкого использования в медицине и психологии. Задача предсказания на основе имеющихся статистических данных. Зависимая и независимая переменные (предикторы). Уравнение регрессии. Расчетные процедуры определения коэффициентов уравнения регрессии.

5.2. Условия оптимальности простой линейной модели регрессии. Метод наименьших квадратов. Вычисление дисперсии ошибки оценки для линейной регрессии. Сравнение двух линий регрессии. Определение значимости различий. Методы проверки линейности. Стратификация независимой переменной. Нелинейное оценивание связей между переменными.

5.3. Коэффициент корреляции Пирсона, его вычисление, свойства, область изменения. Диаграмма рассеивания. Интерпретация коэффициента корреляции Пирсона. Корреляционное отношение. Коэффициент детерминации r^2 . Примеры неадекватной интерпретации корреляции. Связь коэффициента корреляции с регрессией и стандартным отклонением.

5.4. Зависимость мер связи переменных от уровня измерения. Меры связи для номинального уровня переменных (коэффициент Фи Пирсона, коэффициенты сопряженности Пирсона, Чупрова), процедура вычисления, особенности применения. Ранговый коэффициент корреляции Спирмена. Общая характеристика, процедуры вычисления.

Тема 6. Исследование рисков в медицинских исследованиях.

6.1. Факторы риска (ФР) - определение понятия. Особенности факторов риска: продолжительный латентный период, высокая распространенность ФР, множественность причин и эффектов ФР. Применение показателей риска. Диагностика риска, ее роль в профилактике заболеваний.

6.2. Исследования риска в когортных исследованиях. Сравнение рисков. Добавочный риск. Относительный риск (relative risk) и отношение рисков. (risk ratio). Интерпретация

оценки индивидуального риска. Популяционный риск в эпидемиологических исследованиях.

6.3. Исследования риска в исследованиях случай-контроль. Отношение шансов (odds ratio) для оценки рисков в клинических исследованиях. Общая характеристика метода. Особенности подготовки данных. Процедура вычисления.

Тема 7. Классификационные и прогностические методы многомерной статистики.

7.1. Факторные планы в психологии и медицине. Двух- и многофакторный дисперсионный анализ. Формулировка задачи. Требования к подготовке данных.

7.2. Дискриминантный анализ. Формулировка задачи. Требования к подготовке данных.

7.3. Отношения шансов и логистическая регрессия. Формулировка задачи. Требования к подготовке данных. Процедура анализа.

Тема 8. Методы многомерного анализа структуры взаимосвязи переменных.

8.1. Кластерный анализ данных. Формулировка задачи. Требования к подготовке данных. Восходящий и нисходящий алгоритм кластеризации. Особенности кластерного корреляционного анализа.

8.2. Факторный анализ. Виды факторного анализа. Анализ главных компонент. Формулировка задачи. Требования к подготовке данных.

Тема 9 Окончательный дизайн научного исследования: интерпретация и представление результатов анализа в научных публикациях.

9.1. Экспериментальный отчет. Название темы. Аннотация. Раздел «методика» в экспериментальном отчете: описание испытуемых, экспериментального материала, оборудования, процедуры эксперимента. Стандарты обработки данных. Принципы и процедуры выбора статистического метода, какой статистический критерий выбрать? Нормативы представления результатов анализа данных научно-клинических исследований.

9.2. Виды представления результатов анализа: текстовый, табличный, графический, наглядно-образный (схематический) и аналитический. Преимущества и недостатки отдельных видов представления результатов.

9.3. Графическое представление результатов, его цели. Критерии выбора формы графического представления. Ошибки в построении рисунков и графиков. Требования к построению диаграмм. Литературные источники и приложения. Значение, цель и требования к экспериментальному отчету.

4. Учебно-тематический план дисциплины

Номер раздела, темы	Наименование разделов, тем	Количество часов					Форма контроля
		Всего	КР	Л	СПЗ	СР	
Тема 1	Сравнительное исследование (исследование случай-контроль) - анализ различий между группами	4	3		3	1	Перечень контрольных вопросов
Тема 2	Сравнительное исследование с количеством групп более двух - методы множественного сравнения	4	3		3	1	Перечень контрольных вопросов

Тема 3	Анализ качественных признаков	4	3		3	1	Перечень контрольных вопросов
Тема 4	Когортное (лонгитюдное) исследование - анализ повторных наблюдений	4	3	1	2	1	Перечень контрольных вопросов
Тема 5	Корреляционное исследование - меры связи, регрессионный анализ, корреляционный анализ. Различные меры корреляционной связи	4	3	1	2	1	Перечень контрольных вопросов
Тема 6	Исследование рисков в медицинских исследованиях	4	2		2	2	Перечень контрольных вопросов
Тема 7	Классификационные и прогностические методы многомерной статистики	4	2		2	2	Перечень контрольных вопросов
Тема 8	Методы многомерного анализа структуры взаимосвязи переменных	4	3		3	1	Перечень контрольных вопросов
Тема 9	Окончательный дизайн научного исследования: интерпретация и представление результатов анализа в научных публикациях	4	2		2	2	Перечень контрольных вопросов
Общий объем		36	24	2	22	12	

5. Учебно-методическое обеспечение самостоятельной работы обучающихся

Цель самостоятельной работы обучающихся заключается в глубоком усвоении учебного материала и в развитии навыков самообразования. Самостоятельная работа включает: работу с текстами, основной и дополнительной литературой, учебно- методическими пособиями, нормативными материалами, в том числе материалами Интернета, а также проработку конспектов лекций, написание рефератов, участие в работе семинаров, научных конференциях.

Контроль самостоятельной работы осуществляется на семинарских (практических занятиях) занятиях.

6. Оценочные средства для проведения текущего контроля успеваемости и промежуточной аттестации обучающихся

Примерные оценочные средства, включая оценочные задания для проведения текущего контроля успеваемости и промежуточной аттестации обучающихся по дисциплине представлены в Приложении 1 Оценочные средства по дисциплине.

7. Учебно-методическое и информационное обеспечение дисциплины

7.1. Основная литература

1. Боровиков В. П. Statistica стат. анализ и обраб. данных в среде Windows : [справ. и учеб. пособие] / Боровиков В. П., Боровиков И. П.. -М. ФИЛИНЪ 1997.- 592, VIII с.
2. Гланц С. Медико-биологическая статистика. М., 1999. – 459 с.
3. Дружинин В. Н. Экспериментальная психология: учеб. для высш. учеб. заведений по направлению и специальностям психологии. -2-е изд., доп. . - СПб. Питер 2007. - 318 с.
3. Реброва О. Статистический анализ медицинских данных. Применение пакета программ STATISTICA / О. Реброва. – М., 2002.

7.2. Дополнительная литература

1. Глинский В. В. Статистический анализ учеб. пособие / Глинский В. В., Ионин В. Г. - М. ФИЛИНЪ 1998. -264 с.
2. Немов Р. Психология Кн. 3 Психодиагностика : введение в науч. психологическое исслед. с элементами мат. статистики учеб. для высш. пед. учеб. заведений : в 3-х кн. / Р. С. Немов. - М. ВЛАДОС 2005.- 631 с.
3. Основы научных исследований учеб. для ВУЗов / В.И. Крутов, И.М. Грушко, В.В. Попова и др. – М., Высш.шк., 1989. – 400 с.
3. Тюрин Ю.Н. Статистический анализ данных на компьютере [учеб. пособие] / Тюрин Ю. Н. ; под ред. В. Э. Фигурнова. - М. ИНФРА-М 1998. - 528 с.

7.3. Перечень ресурсов информационно-телекоммуникационной сети «Интернет»

1. Официальный сайт ФИЦ ФТМ: адрес ресурса – <https://frcftm.ru/>, на котором содержатся сведения об образовательной организации и ее подразделениях, локальные нормативные акты, сведения о реализуемых образовательных программах, их учебно-методическом и материально-техническом обеспечении, а также справочная, оперативная и иная информация.
- 2.Springer Journals [Электронный ресурс]: база данных – Режим доступа: <https://link.springer.com/> – Доступ открыт со всех компьютеров сети Центра.
- 3.Федеральная электронная медицинская библиотека [Электронный ресурс]: электронно-библиотечная система (ЭБС) / ЦНМБ Первого Московского государственного медицинского Центра им. И.М.Сеченова. – Режим доступа: <http://feml.scsml.rssi.ru/feml> – Свободный доступ.
- 4.eLIBRARY.RU [Электронный ресурс]: электронная библиотека – URL: <http://www.elibrary.ru/>. – Доступ к подписке журналов открыт со всех компьютеров сети Центра.
- 5.Российская государственная библиотека [Электронный ресурс]: сайт. – Режим доступа: <http://www.rsl.ru> – Свободный доступ.
- 6.Consilium Medicum [Электронный ресурс]: сайт. – Режим доступа: <http://www.consilium-medicum.com/> – Свободный доступ.
- 7.PubMed: US National Library of Medicine National Institutes of Health [Электронный ресурс]: сайт. – URL: <http://www.ncbi.nlm.nih.gov/pubmed>.
- 8.ScienceDirect. Ресурсы открытого доступа [Электронный ресурс]: сайт. – Режим доступа: <http://www.sciencedirect.com/science/jrnallbooks/open-access>.
- 9.КиберЛенинка: научная электронная библиотека [Электронный ресурс]: сайт. – Режим доступа: <http://cyberleninka.ru/> – Свободный доступ.

8. Материально-техническое обеспечение дисциплины

8.1. Перечень помещений и оборудования для проведения учебных занятий по дисциплине

№	Помещение (площадь)	Перечень оборудования	Адрес, наименование организации
2	Конференц-зал, ком. 611	Ноутбук Asus F3Ka TL60 (1 шт.); проектор Epson EB-93 (1 шт.).	Институт молекулярной патологии и патоморфологии ФИЦ ФТМ
3	Помещение для самостоятельной работы обучающихся № 409	Технические средства обучения: Компьютер Asus – 1 шт. Компьютер Interlink – 4 шт. Специализированная мебель: Стол компьютерный – 10 шт. Кресло компьютерное- 10 шт. Стол – 1 шт. Стул – 1 шт. Перечень лицензионного программного обеспечения: Linux Mint, Libre Office Помещение обеспечено доступом к сети «Интернет» и электронной информационно-образовательной среде ФИЦ ФТМ.	Учебно-методический центр ФИЦ ФТМ

9. Методические указания для обучающихся по изучению дисциплины

Преподавание дисциплины осуществляется в соответствии с Федеральным государственным образовательным стандартом высшего образования.

Основными формами получения и закрепления знаний по данной дисциплине являются занятия лекционного и практического типа, самостоятельная работа обучающегося, в том числе под руководством преподавателя, прохождение контроля.

Учебный материал по дисциплине разделен на 9 тем.

Изучение дисциплины, согласно учебному плану, предполагает самостоятельную работу обучающихся. Самостоятельная работа включает в себя изучение учебной, учебно-методической и специальной литературы, её конспектирование, подготовку к семинарам (практическим занятиям), текущему контролю успеваемости и промежуточной аттестации (зачету).

Текущий контроль успеваемости по дисциплине (модулю) и промежуточная аттестация осуществляются в соответствии с Положением о проведении текущего контроля успеваемости и организации и проведения промежуточной аттестации обучающихся, устанавливающим формы проведения промежуточной аттестации, ее периодичность и систему оценок.

10. Методические рекомендации преподавателю по организации учебного процесса по дисциплине

Преподавание дисциплины осуществляется в соответствии с Федеральными государственными образовательными стандартами высшего образования.

При изучении дисциплины (модуля) рекомендуется использовать следующий набор средств и способов обучения:

- рекомендуемую основную и дополнительную литературу;
- задания для подготовки к семинарам (практическим занятиям) – вопросы для обсуждения и др.;
- задания для текущего контроля успеваемости (задания для самостоятельной работы обучающихся);
- вопросы и задания для подготовки к промежуточной аттестации по итогам изучения дисциплины, позволяющие оценить знания, умения и уровень приобретенных компетенций.

При проведении занятий лекционного и семинарского типа, в том числе в форме вебинаров и on-line курсов необходимо строго придерживаться учебно-тематического плана дисциплины (модуля), приведенного в разделе 4 данного документа. Необходимо уделить внимание рассмотрению вопросов и заданий, включенных в оценочные задания, при необходимости, решить аналогичные задачи с объяснением алгоритма решения.

Следует обратить внимание обучающихся на то, что для успешной подготовки к текущему контролю успеваемости и промежуточной аттестации нужно изучить материалы основной и дополнительной литературы, список которых приведен в разделе 7 данной рабочей программы дисциплины (модуля) и иные источники, рекомендованные в подразделах «Перечень ресурсов информационно-телекоммуникационной сети «Интернет» и «Перечень профессиональных баз данных и информационных справочных систем», необходимых для изучения дисциплины (модуля).

Текущий контроль успеваемости и промежуточная аттестация осуществляются в соответствии с Положением о проведении текущего контроля успеваемости и организации и проведения промежуточной аттестации обучающихся, устанавливающим формы проведения промежуточной аттестации, ее периодичность и систему оценок, с которыми необходимо ознакомить обучающихся на первом занятии.

Приложение 1

**ОЦЕНОЧНЫЕ СРЕДСТВА ПО ДИСЦИПЛИНЕ
«ВИДЫ ИССЛЕДОВАНИЙ И АДЕКВАТНЫЕ ИМ МЕТОДЫ СТАТИСТИЧЕСКОГО
АНАЛИЗА В МЕДИЦИНЕ»**

Специальность:
31.08.07 Патологическая анатомия

Уровень высшего образования:
подготовка кадров высшей квалификации

Новосибирск

2023 г.

1.Перечень компетенций, формируемых в процессе изучения дисциплины

Код и наименование компетенции, индикатора достижения компетенции	Планируемые результаты обучения по дисциплине	
УК-1. Способен критически и системно анализировать, определять возможности и способы применения достижения в области медицины и фармации в профессиональном контексте		
УК-1.1. Анализирует достижения в области медицины и фармации в профессиональном контексте	Знать:	законы формальной логики, методы критического анализа и синтеза
	Уметь:	анализировать любую ситуацию и находить верные решения
	Владеть:	логическими методами и приемами при анализе информации, полученной в ходе комплексного клинико-морфологического исследования
УК-1.2. Определяет возможности и способы применения достижения в области медицины и фармации в профессиональном контексте	Знать:	методы и способы оценки возможности и вариантов применения современных достижений в области медицины и фармации.
	Уметь:	определять возможности и способы применения достижений в области медицины и фармации в профессиональном контексте; сформулировать проблему, выделить ключевые цели и задачи по ее решению
	Владеть:	методами и приёмами системного анализа достижений в области медицины и фармации для их применения в профессиональном контексте
УК-2. Способен разрабатывать, реализовывать проект и управлять им		
УК-2.1. Участвует в разработке и управлении проектом	Знать:	- статистические методы сбора, обработки, анализа и прогнозирования данных; - методы и принципы постановки задач (конкретная, измеримая, достижимая, значимая, ограниченная во времени) и способы их реализации; - основные подходы к организации проектной деятельности.
	Уметь:	- анализировать данные из множественных источников и оценивать качество и достоверность полученной информации по явным и неявным признакам; - применять программное обеспечение (текстовые, графические, табличные и аналитические приложения, приложения для визуального представления данных) для работы с информацией;

		- собирать, анализировать, систематизировать сведения и данные, документировать требования к проектам; - вести деловые переговоры с целью согласования взаимных интересов участников проекта; - разрабатывать алгоритмы, модели, схемы проекта; - принимать решения при разработке и реализации проекта; - выполнять проектные работы; - оценивать результаты реализации проектной деятельности
	Владеть:	- навыком сбора и анализа исходных данных, необходимых для оценки реализуемости проекта; - навыком определения соответствия целей и задач проекта; - методами и принципами организации проектной деятельности.
УК-2.2. Выполняет задачи в зоне своей ответственности в соответствии с запланированными результатами и точками контроля, при необходимости корректирует способы решения задач	Знать:	- процессы и методы управления проектами; - методы определения сроков реализации проектов.
	Уметь:	- осуществлять планирование проекта; - управлять процессом реализации проекта.
	Владеть:	- навыками планирования проекта, организации исполнения проекта, контроля этапов реализации проекта, оценки результатов проекта.
ПК-3. Способен осуществлять научно-исследовательскую и педагогическую деятельность на основе научных знаний		
ПК-3.1. Планирует научно-исследовательскую деятельность	Знать:	- основные приемы и принципы планирования и протоколирования научных исследований; - этапы проведения научно-исследовательской работы.
	Уметь:	- планировать, организовать самостоятельный исследовательский процесс;
	Владеть:	- навыком проектирования научно-исследовательской работы; - навыком определения объекта и метода исследования.
ПК-3.2. Осуществляет научно-исследовательскую деятельность	Знать:	- методы статистической обработки данных исследования; - понятия абсолютных и относительных показателей; - методы визуализации данных; - виды диаграмм; - технологию публичного выступления; - методологию публичной дискуссии.

	Уметь:	- выбирать метод для статистической обработки; - анализировать и сравнивать полученные данные; - выбирать необходимый вид диаграмм и осуществлять построение графиков; - подготовить доклад в форме презентации с использованием мультимедийной техники; - публично представить полученные результаты; - дискутировать с коллегами по теме научно-исследовательской работы.
	Владеть:	- навыком выбора метода статистической обработки; - навыком анализа и сравнения полученных данных; - приемами и способами визуализации полученных данных; - навыком подготовки презентации по полученным данным с использованием мультимедийной техники; - навыком публичного представления полученных данных; - навыком проведения дискуссии по теме научно-исследовательской работы.

ОЦЕНОЧНЫЕ СРЕДСТВА ПО ТЕМАМ ДИСЦИПЛИНЫ

Вид оценочного средства: контрольные вопросы

1. Сравнительное исследование в психологии и клинической медицине.

Ответ:

Сравнительное исследование в психологии и клинической медицине — это методологический подход, направленный на сравнение различных аспектов, методов, теорий или результатов в этих двух областях. Несмотря на различия в предметных областях, оба направления используют сравнительные исследования для выявления особенностей, эффективности и взаимосвязей.

Основные особенности сравнительных исследований в психологии и клинической медицине:

1. Цели и задачи:

- В психологии: сравнение теорий, методов терапии, психологических характеристик различных групп или популяций.
- В клинической медицине: оценка эффективности лекарственных препаратов, методов лечения, диагностических процедур или профилактических программ.

2. Методы:

- В психологии: экспериментальные дизайны, кросс-секционные и лонгитюдные исследования, качественные методы.
- В клинической медицине: рандомизированные контролируемые испытания (РКИ), когортные исследования, мета-анализы.

3. Объекты сравнения:

- В психологии: различные психологические теории, методы терапии (например, когнитивно-поведенческая терапия vs. гуманистическая терапия), психологические показатели.
- В клинической медицине: препараты, хирургические вмешательства, диагностические тесты.

4. Значение:

- В психологии: помогает определить наиболее эффективные подходы к терапии и понять психологические механизмы.
- В клинической медицине: способствует выбору оптимальных методов лечения и улучшению качества медицинской помощи.

Общие черты:

- Оба направления используют статистические методы для анализа данных.
- Оба требуют строгой методологической базы для обеспечения надежности и валидности результатов.
- Результаты сравнительных исследований способствуют развитию практики и теории в обеих областях.

Отличия:

- Область предмета исследования (психологические процессы vs. физиологические или медицинские вмешательства).
- Типы данных (психологические тесты и опросники vs. медицинские показатели и лабораторные анализы).
- Цели применения (понимание поведения и мышления vs. лечение заболеваний).

В целом, сравнительные исследования являются важным инструментом для развития обеих дисциплин, позволяя выявлять наиболее эффективные подходы и расширять научное понимание человеческого поведения и здоровья.

2. Непараметрические аналоги дисперсионного анализа для несвязанных выборок: Н-критерий Крускала-Уоллиса, Т-критерий Уайта, Т-критерий Манна-Уитни, U-критерий Манна-Уитни.

Ответ:

Непараметрические аналоги дисперсионного анализа для несвязанных выборок включают следующие критерии:

1. Критерий Крускала-Уоллиса (Н-критерий) — используется для сравнения трех или более независимых групп по их медианам, аналог дисперсионного анализа (ANOVA). Он основан на ранжировании данных и не требует нормальности распределения.
2. Критерий Уайта (Т-критерий Уайта) — применяется для сравнения двух независимых выборок, особенно когда данные не соответствуют предположениям параметрических тестов. Он основан на ранжировании и является непараметрическим аналогом t-критерия для независимых выборок.
3. Критерий Манна-Уитни (U-критерий) — также предназначен для сравнения двух независимых выборок по их медианам. Он основан на ранжировании данных и является альтернативой t-тесту при ненормальных распределениях.
4. U-критерий Манна-Уитни — то же самое, что и выше; иногда его называют просто U-критерием Манна-Уитни.

Эти критерии позволяют делать выводы о различиях между группами без предположений о нормальности распределений и гомоскедастичности, что делает их ценными инструментами в непараметрической статистике при анализе несвязанных выборок.

3. Значение множественных сравнений в статистическом анализе данных. Поправка Бонферрони. Общий принцип, неравенство Бонферрони. Процедура вычисления. Критерий Ньюмена-Кейлса. Общая характеристика метода. Процедура вычисления. Критерий Тьюки. Общая характеристика метода. Процедура вычисления. Отличия получаемых результатов по сравнению с критерием Ньюмена-Кейлса.

Ответ:

Множественные сравнения в статистическом анализе данных — это набор методов, предназначенных для проведения одновременной проверки нескольких гипотез. Их использование важно для контроля уровня ошибочного отклонения нулевой гипотезы при множественных тестах, поскольку при большом числе сравнений вероятность получения ложноположительных результатов возрастает.

Поправка Бонферрони

- **Общий принцип:**

Поправка Бонферрони — это один из самых простых и широко используемых методов коррекции уровня значимости при множественных сравнениях. Она основана на контроле семейного уровня ошибки (family-wise error rate, FWER).

Идея заключается в том, чтобы разделить общий уровень значимости (например, $\alpha = 0,05$) на число проведенных сравнений m . Тогда для каждого отдельного теста устанавливается более строгий уровень значимости:

$$\alpha' = m\alpha$$

- **Неравенство Бонферрони:**

Если все индивидуальные гипотезы проверяются при уровне значимости α' , то вероятность хотя бы одного ложноположительного результата не превышает α . Формально:

$$P(\text{хотя бы одна ошибка}) \leq m \times \alpha' = \alpha$$

- **Процедура вычисления:**

1. Определить исходный уровень значимости α .
2. Разделить его на число сравнений m , получив скорректированный уровень $\alpha' = m\alpha$.
3. Выполнить каждое сравнение с этим уровнем.
4. Отклонять нулевую гипотезу только если p -значение меньше α' .

Критерий Ньюмена-Кейлса

- **Общая характеристика метода:**

Критерий Ньюмена-Кейлса предназначен для проведения множественных сравнений после анализа дисперсии (ANOVA). Он позволяет определить, какие именно группы отличаются друг от друга по средним значениям. Метод основан на парных сравнениях и корректирует уровень значимости для учета множественных тестов.

- **Процедура вычисления:**

- После проведения ANOVA и обнаружения значимых различий между группами, выполняются парные сравнения между всеми группами.
- Для каждого сравнения рассчитывается критическая разница (например, с помощью критерия Ньюмена-Кейлса):

$$|X_i - X_j| > q_{\alpha, k, N} \times \text{spooled} \times \sqrt{n_i + n_j}$$

-
-

где:

- $q_{\alpha, k, N}$ — квантиль распределения studentized range (распределения Ньюмена-Кейлса) при уровне α , числе групп k , и общем числе наблюдений N ;
- spooled — объединенная стандартная ошибка;
- n_i, n_j — размеры выборок групп.
- Критерий Ньюмена-Кейлса обеспечивает контроль семейной ошибки при множественных парных сравнениях.

Критерий Тьюки

- Общая характеристика метода:
Критерий Тьюки также предназначен для множественных сравнений после ANOVA. Он использует корректировку уровня значимости на основе распределения studentized range и обеспечивает контроль FWER. Метод более мощный по сравнению с поправкой Бонферрони при большом числе групп.
- Процедура вычисления:
 - Аналогично критерию Ньюмена-Кейлса, после подтверждения наличия различий между группами выполняются парные сравнения.
 - Для каждого сравнения проверяется гипотеза о равенстве средних с помощью критерия Тьюки: разница между средними должна превышать критическую величину из распределения studentized range:

$$|X_i - X_j| > q_{\alpha, k, N \times \text{error}}$$

где:

- $\text{error} = \text{spooled}/2/n$
- (при равных размерах выборок);
- остальные параметры аналогичны.

Отличия результатов по сравнению с критерием Ньюмена-Кейлса

- Мощность и строгость:
Критерий Тьюки обычно более мощный (чувствительный) к обнаружению различий по сравнению с критерием Ньюмена-Кейлса при одинаковых условиях, особенно при большом числе групп. Это связано с тем, что он использует более точное распределение studentized range для определения критической величины.
- Область применения:
Оба метода применимы к множественным парным сравнениям после ANOVA. Однако критерий Тьюки лучше подходит для ситуаций с большим числом групп и когда важна высокая мощность теста.
- Контроль ошибок:
Оба метода обеспечивают контроль семейной ошибки (FWER), но критерий Тьюки делает это более эффективно за счет использования распределения studentized range вместо простого деления уровня значимости.

Итоговая характеристика

Метод	Основное назначение	Контроль ошибок	Мощность	Особенности
Поправка Бонферрони	Коррекция уровня значимости при множественных тестах	Контроль FWER через деление α	низкая при большом числе тестов	Простая в реализации; консервативная
Критерий Ньюмена-Кейлса	Множественные парные сравнения после ANOVA	Контроль FWER через распределение studentized range	выше чем у Бонферрони	Использует таблицы квантилей; чувствителен к числу групп
Критерий Тьюки	Множественные парные сравнения после ANOVA	Контроль FWER через распределение studentized range	наиболее мощный среди перечисленных	Более точное определение критической разницы

4. Особенности применения методов множественного сравнения с контролем. Критерий Даннета. Процедура вычисления.

Ответ:

Методы множественного сравнения с контролем предназначены для проведения статистических тестов при множественных гипотезах, чтобы снизить риск ошибочного отклонения хотя бы одной истинной нулевой гипотезы (ошибка типа I). Одним из таких методов является критерий Даннета.

Особенности применения методов множественного сравнения с контролем:

- Контроль уровня ошибки: Основная задача — снизить вероятность ошибочного обнаружения значимых различий при множественных тестах, например, контролировать семейный уровень ошибки (family-wise error rate, FWER).
- Использование корректировок: Обычно применяются корректировки р-значений или уровней значимости для каждого теста, чтобы общий риск оставался в пределах заданного уровня (например, $\alpha = 0.05$).
- Применение к группам: Эти методы особенно полезны при сравнении нескольких групп или условий, чтобы избежать ложных положительных результатов.

Критерий Даннета (Dunnett's test):

Это статистический метод для множественных сравнений, когда одна контрольная группа сравнивается с несколькими экспериментальными группами. Он позволяет определить, есть ли статистически значимые различия между каждой экспериментальной группой и контрольной при сохранении контроля уровня ошибки.

Процедура вычисления критерия Даннета:

1. Определение исходных данных:
 - Средние значения по группам: $\bar{X}_0$ — контрольная группа; $\bar{X}_i$ — остальные группы.
 - Общая дисперсия или стандартная ошибка средних.
 - Количество групп: k (одна контрольная + $k-1$ экспериментальных).
2. Расчет стандартной ошибки (SE):

$$SE = \sqrt{n \cdot 2 \cdot MSE}$$

где:

- MSE — средняя квадратичная ошибка из дисперсионного анализа,
 - n — число наблюдений в каждой группе (предполагается равное).
2. Определение критической границы:
 - Используется таблица критических значений для распределения Стюдента-Стюдента или распределения Тьюнка с учетом числа групп и степени свободы.
 - В случае Даннета используют таблицы или функцию для определения критического значения $q_{k,df}$, где:
 - k — число сравниваемых групп,
 - df — степени свободы по ошибке.

3. Расчет границы значимых различий:

$$|\bar{X}_i - \bar{X}_0| > q_{k,df} \times SE$$

4. Интерпретация:

- Если разница между средними превышает рассчитанную границу, то различие считается статистически значимым при контроле уровня ошибки.

Кратко:

Шаг Действие

Формула / описание

1 Определить средние и дисперсии

Средние по группам и MSE из ANOVA

2 Вычислить стандартную ошибку

$SE = \sqrt{n \cdot 2 \cdot MSE}$

Шаг	Действие	Формула / описание
3	Определить критическое значение	Использовать таблицы или функцию для q_k, df
4	Рассчитать границу различия	$q_k, df \times SE$
5	Сделать выводы о значимости различий	Проверить каждую разницу между контрольной и экспериментальной группой

Итог: Критерий Даннета позволяет проводить множественные сравнения с контролем уровня ошибки за счет использования специального критерия на основе распределения Тьюнка и учитывает число сравниваемых групп. Это обеспечивает более строгий контроль ошибок по сравнению с простыми построчными тестами.

5. Особенности применения критериев Ньюмена-Кейлса и Тьюки при непараметрическом анализе.

Критерий Данна для непараметрического сравнения выборок разного объема.

Ответ: Критерии Ньюмена-Кейлса и Тьюки являются важными инструментами непараметрического анализа, используемыми для сравнения нескольких групп без предположения о форме распределения данных. Ниже представлены особенности их применения и особенности критерия Данна при сравнении выборок разного объема.

Критерий Ньюмена-Кейлса (Kruskal-Wallis)

- Назначение: Используется для сравнения трех или более независимых групп по ранжированным данным.
- Особенности применения:
 - Не требует предположений о нормальности распределения.
 - Работает с ранжированными значениями, что делает его устойчивым к выбросам.
 - Особенно полезен при наличии гетерогенности дисперсий.
- Процедура:
 - Объединяют все наблюдения из групп, присваивают им ранги.
 - Вычисляют сумму рангов по каждой группе.
 - Рассчитывают статистику H , которая примерно следует распределению χ^2 с $k-1$ степенями свободы (где k — число групп).
- Особенности:
 - Позволяет определить, есть ли статистически значимые различия между группами, но не указывает, какие именно.

Критерий Тьюки

- Назначение: Предназначен для сравнения нескольких групп по ранжированным данным.
- Особенности применения:
 - Используется в случаях, когда необходимо провести множественные сравнения после обнаружения значимых различий (например, при использовании критерия Ньюмена-Кейлса).
 - Позволяет определить, между какими именно группами есть статистически значимые различия.
- Процедура:
 - После проведения критерия Ньюмена-Кейлса или другого общего теста проводят попарные сравнения с поправкой на множественные тесты (например, поправка Бонферрони).
- Особенности:
 - Обеспечивает контроль уровня ошибки при множественных сравнениях.

Критерий Данна для непараметрического сравнения выборок разного объема

- Назначение: для сравнения двух или более независимых выборок с разным объемом по их медианам или другим центральным тенденциям.
- Особенности применения:
 - Не требует равных размеров выборок.
 - Основан на ранжировании всех данных и сравнении их медианных позиций.
 - В случае двух выборок с разными объемами используется модифицированный критерий Данна, учитывающий размеры выборок.
- Процедура:
 - Объединяют все наблюдения из обеих выборок и присваивают им ранги.
 - Рассчитывают медианы или средние ранги для каждой выборки.
 - Статистика критерия определяется как разница между медианными рангами или средними рангами двух групп, скорректированная на размеры выборок.
- Особенности:
 - Позволяет делать выводы о различиях в центральных тенденциях двух выборок независимо от их размеров и формы распределений.

Итог:

Критерии Ньюмена-Кейлса и Тьюки широко применяются в непараметрическом анализе для сравнения нескольких групп без предположений о нормальности. Они позволяют выявлять наличие различий между группами на основе ранговых данных. Критерий Данна расширяет возможности для сравнения двух или более выборок разного объема, учитывая их размеры и распределения. В совокупности эти методы обеспечивают гибкий и надежный инструментарий для анализа данных в ситуациях, когда параметры распределений неизвестны или нарушены предположения классических методов параметрического анализа.

6. Роль качественных признаков при проведении психологического и клинического исследования. Доли (эмпирические частоты), как статистическая оценка качественных признаков. Сравнение долей. Точность оценки долей. Z-критерий. Поправка Йейтса на непрерывность.

Ответ:

Роль качественных признаков при проведении психологического и клинического исследования

Качественные признаки (категориальные переменные) играют важную роль в психологических и клинических исследованиях, поскольку позволяют классифицировать объекты или субъекты по различным признакам, таким как пол, диагноз, тип терапии, уровень образования и др. Их использование помогает выявлять связи между категориями, определять распространенность определенных характеристик в выборке и делать выводы о популяции.

Доли (эмпирические частоты) как статистическая оценка качественных признаков

Эмпирическая доля (частота) — это отношение числа наблюдений с определенной категорией к общему числу наблюдений. Она служит оценкой вероятности принадлежности объекта к данной категории и используется для описания распределения качественного признака в выборке.

Например, если из 100 пациентов 30 имеют диагноз А, то эмпирическая доля для диагноза А равна 0,3 или 30%.

Сравнение долей

Для анализа различий между группами по качественным признакам используют сравнение долей. Например, сравнивают доли пациентов с определенным диагнозом в двух группах лечения. Статистические методы позволяют определить, есть ли статистически значимые различия между этими долями.

Точность оценки долей

Точность оценки зависит от размера выборки и вариации данных. Обычно для оценки

доверительного интервала для доли используют формулу:
Доверительный интервал $= p \pm Z_{\alpha/2} \sqrt{p(1-p)}$

где:

- p — эмпирическая доля,
- n — объем выборки,
- $Z_{\alpha/2}$ — критическое значение стандартного нормального распределения для выбранного уровня доверия (например, 1,96 для 95%).

Чем больше объем выборки, тем выше точность оценки.

Z-критерий (критерий 3-test)

3-критерий используется для проверки гипотез о равенстве двух пропорций (долей) в больших выборках. Например, чтобы проверить гипотезу о равенстве долей двух групп по какому-либо качественному признаку.

Формула для 3-критерия при сравнении двух долей:

$$Z = \frac{p_1 - p_2}{\sqrt{p(1-p)(n_1 + n_2)}}$$

где:

- p_1, p_2 — эмпирические доли в группах,
- $p = \frac{n_1 x_1 + n_2 x_2}{n_1 + n_2}$ — объединенная доля,
- n_1, n_2 — размеры групп.

Если полученное значение Z превышает критическое значение из таблицы (например, 1,96 при уровне значимости 0,05), то гипотеза о равенстве долей отвергается.

Поправка Йейтса на непрерывность

При использовании хи-квадрат теста для сравнения двух пропорций в небольших выборках рекомендуется применять поправку Йейтса на непрерывность. Она уменьшает вероятность ложноположительных результатов при малых объемах данных.

Корректировка заключается в вычислении статистики хи-квадрат по формуле:

$$\chi^2 Y^2 = \frac{(a+b)(c+d)(a+c)(b+d) - (ad-bc)^2}{2n}$$

где a, b, c, d — частоты в таблице сопряженности 2×2 .

Эта поправка делает тест более консервативным и уменьшает риск ошибок первого рода при малых размерах выборки.

7. Соответствие эмпирических частот ожидаемым частотам. Таблицы сопряженности, вычисление ожидаемых частот. Критерий хи-квадрат. Процедура вычисления.

Ответ:

1. Таблицы сопряженности и эмпирические частоты

Таблица сопряженности — таблица, в которой представлены наблюдаемые (эмпирические) частоты для сочетаний двух или более признаков.

Например, для двух признаков (А и В) с категориями:

	B1	B2	... Bk	ИТОГО по А
A1	n_{11}	n_{12}	... n_{1k}	$n_{1\cdot}$
A2	n_{21}	n_{22}	... n_{2k}	$n_{2\cdot}$
...	...	...	...	...
Am	n_{m1}	n_{m2}	... n_{mk}	$n_{m\cdot}$
ИТОГИ	$n_{\cdot 1}$	$n_{\cdot 2}$	... $n_{\cdot k}$	N

где:

- n_{ij} — эмпирическая частота в ячейке (i, j) ,
- $n_{i\cdot}$ — сумма по строке i ,
- $n_{\cdot j}$ — сумма по столбцу j ,
- N — общее число наблюдений.

2. Вычисление ожидаемых частот

Если предполагается, что признаки независимы (гипотеза нулевого равенства), то ожидаемые частоты для каждой ячейки рассчитываются по формуле:

$$E_{ij} = N(n_{i\cdot}) \times (n_{\cdot j})$$

где:

- E_{ij} — ожидаемая частота в ячейке (i, j) ,
- $n_{i\cdot}$ — сумма по строке i ,
- $n_{\cdot j}$ — сумма по столбцу j ,
- N — общее число наблюдений.

3. Критерий хи-квадрат

Используется для проверки соответствия эмпирических данных теоретической модели (например, независимости признаков).

Статистика хи-квадрат:

$$\chi^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(n_{ij} - E_{ij})^2}{E_{ij}}$$

где сумма берется по всем ячейкам таблицы.

Критерий нулевой гипотезы:

"Эмпирические частоты соответствуют ожидаемым при условии независимости признаков."

4. Процедура вычисления критерия хи-квадрат

1. Построение таблицы сопряженности: заполните таблицу наблюдаемыми частотами n_{ij} .
2. Вычисление сумм по строкам и столбцам: найдите $n_{i\cdot}$, $n_{\cdot j}$, и общее число наблюдений N .
3. Расчет ожидаемых частот: для каждой ячейки используйте формулу:

$$E_{ij} = N n_{i\cdot} \times n_{\cdot j}$$

4. Вычисление статистики хи-квадрат:

$$\chi^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(n_{ij} - E_{ij})^2}{E_{ij}}$$

5. Определение степени свободы:

$$df = (I - 1)(J - 1)$$

6. Сравнение полученного значения с критическим значением из таблицы хи-квадрат при выбранном уровне значимости (α).

Итог

Если рассчитанное значение χ^2 превышает критическое, гипотеза о независимости отвергается; если нет — гипотеза принимается.

8. Множественные сравнения качественных признаков. Преобразование таблиц сопряженности. Использование поправки Бонферрони. Точный критерий Фишера. Процедура вычисления.

Ответ:

Множественные сравнения качественных признаков — важная задача в анализе таблиц сопряженности, особенно когда необходимо проверить гипотезы о равенстве пропорций или ассоциациях между несколькими группами.

Ниже приведено описание основных методов и процедур:

1. Таблицы сопряженности и их преобразование

Таблица сопряженности — это таблица, показывающая распределение двух или более качественных переменных. Например, для двух признаков с двумя уровнями (бинарные признаки) таблица может выглядеть так:

	Признак 2: Да Признак 2: Нет Всего	
Признак 1: Да a	b	a + b
Признак 1: Нет c	d	c + d

	Признак 2: Да		Признак 2: Нет	Всего
Всего	a + c	b + d	N	

Преобразование таблиц включает:

- Объединение строк или столбцов для анализа более крупных групп.
- Матричное преобразование для получения новых переменных.
- Кросс-табуляцию для сравнения нескольких признаков.

2. Множественные сравнения и поправка Бонферрони

При множественных тестах риск ошибочного отклонения нулевой гипотезы (ошибка типа I) увеличивается. Для контроля уровня значимости используют поправку Бонферрони:

$$\alpha_{\text{корр}} = m\alpha$$

где:

- α — желаемый уровень значимости (например, 0.05),
- m — число проведённых независимых тестов.

Пример: если проводится 10 тестов при $\alpha=0.05$, то для каждого теста порог значимости станет:

$$\alpha_{\text{корр}} = 10 \cdot 0.05 = 0.005$$

Это уменьшает вероятность ложных положительных результатов.

3. Точный критерий Фишера (Fisher's Exact Test)

Используется для анализа таблиц сопряженности размером 2x2, особенно при малых выборках.

Процедура вычисления:

Для таблицы:

	Да		Нет	Всего
Группа 1	a	b	n_1	
Группа 2	c	d	n_2	
Всего	m_1	m_2	N	

где:

- a, b, c, d — частоты,
- $n_1 = a + b$,
- $n_2 = c + d$,
- $m_1 = a + c$,
- $m_2 = b + d$,
- $N = a + b + c + d$.

Формула вероятности получения такой таблицы при условии фиксированных сумм:

$$p = \frac{(n_1! n_2! m_1! m_2!)}{N!} \frac{N!}{(a! b! c! d!)} = \frac{(n_1! n_2! m_1! m_2!)}{N!} \frac{N!}{(a! b! c! d!)}$$

где (kn) — биномиальный коэффициент.

Для определения значимости обычно используют двусторонний или односторонний p-value, суммируя вероятности всех таблиц с равными или более экстремальными значениями по гипотезе.

4. Процедура вычисления множественных сравнений с использованием критерия Фишера и поправки Бонферрони

Шаги:

1. Построение таблиц сопряженности: для каждой пары признаков или групп.
2. Вычисление p-value: с помощью точного критерия Фишера.
3. Корректировка p-value: применяя поправку Бонферрони:

$$p_{\text{корр}} = p \times m$$

где m — число проведённых тестов.

4. Интерпретация результатов: если скорректированный p-value < выбранного уровня (α), гипотеза отвергается.

Итог: Использование преобразований таблиц сопряженности позволяет структурировать данные для анализа; точный критерий Фишера обеспечивает корректную оценку при малых выборках; а поправка Бонферрони помогает контролировать ошибку первого рода при множественных сравнениях.

9. Особенности последовательного (лонгитюдного) исследования. Преимущества по сравнению со сравнительным.

Ответ:

Особенности последовательного (лонгитюдного) исследования заключаются в следующем:

1. Длительность и многократность измерений: Исследование проводится на одних и тех же участниках или группах в течение определенного периода времени с несколькими точками измерения. Это позволяет отслеживать динамику изменений.
2. Изучение развития и причинно-следственных связей: Позволяет выявлять закономерности развития, причинно-следственные связи и временные последовательности событий или изменений.
3. Контроль за изменениями во времени: обеспечивает возможность наблюдения за изменениями внутри субъектов, а не только между группами.

Преимущества по сравнению с сравнительным (кросс-секционным) исследованием:

- Обнаружение причинных связей: Лонгитюдные исследования лучше подходят для установления причинно-следственных связей, поскольку позволяют проследить, что происходит раньше — изменение переменной или результат этого изменения.
- Изучение динамики развития: Позволяют понять, как меняются показатели со временем у одних и тех же участников, что невозможно при кросс-секционных подходах.
- Уменьшение влияния индивидуальных различий: Так как измерения проводятся на одних и тех же участниках, уменьшается влияние межгрупповых различий, характерных для сравнительных исследований.
- Обеспечение более точных данных о развитии процессов: Позволяют выявлять закономерности и тенденции, которые могут быть скрыты при разовом исследовании.

В целом, лонгитюдные исследования ценны для изучения процессов развития, динамики изменений и причинно-следственных связей во времени.

10. Понятие о дисперсионном анализе связанных выборок. Парный t-критерий Стьюдента. Процедура вычисления. Дисперсионный анализ для связанных выборок. Процедура вычисления.

Ответ:

Дисперсионный анализ связанных выборок и парный t-критерий Стьюдента — это статистические методы, используемые для сравнения средних значений в зависимых (связанных) группах данных.

1. Понятие о дисперсионном анализе связанных выборок

Дисперсионный анализ (ANOVA) предназначен для проверки гипотезы о равенстве средних значений в нескольких группах. В случае связанных выборок (или зависимых выборок), наблюдения в одной группе связаны с наблюдениями в другой, например, измерения до и после воздействия на одних и тех же объектов.

При анализе связанных выборок используют повторные измерения или парные данные. Основная идея — учитывать внутреннюю связанность данных, чтобы повысить чувствительность теста.

2. Парный t-критерий Стьюдента

Это статистический критерий для проверки гипотезы о равенстве средних двух связанных выборок.

- Гипотезы:

- $H_0: \mu_1 = \mu_2$ (средние значения равны)
- $H_1: \mu_1 \neq \mu_2$
- Расчет:
 1. Для каждого объекта вычисляется разность между двумя измерениями: $d_i = x_{i1} - x_{i2}$.
 2. Вычисляется среднее разностей: $\bar{d} = \frac{1}{n} \sum_{i=1}^n d_i$.
 3. Определяется стандартное отклонение разностей: $s_d = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (d_i - \bar{d})^2}$.
- .
- Вычисляется t-статистика: $t = \bar{d} / (s_d / \sqrt{n})$
- Критерий принятия решения: сравниваем полученное значение t с критическим значением из таблицы распределения Стьюдента при заданном уровне значимости α и степенях свободы $df = n - 1$.

3. Процедура вычисления парного t-критерия

Шаги:

1. Собрать пары данных (x_{i1}, x_{i2}) для каждого объекта.
2. Вычислить разности d_i .
3. Найти среднее разностей $\bar{d}$.
4. Определить стандартное отклонение разностей s_d .
5. Рассчитать t по формуле.
6. Определить критическое значение t из таблицы при выбранном уровне значимости и $df = n - 1$.
7. Сделать вывод о гипотезе:
 - Если $|t| > t_{\text{критическое}}$, отвергаем H_0 — есть статистически значимая разница.
 - Если $|t| \leq t_{\text{критическое}}$, нет оснований отвергнуть H_0 .

4. Дисперсионный анализ для связанных выборок

В случае более чем двух связанных групп используют повторные меры дисперсионного анализа (ANOVA для связанных выборок). Он позволяет проверить гипотезу о равенстве средних во всех группах одновременно.

- Основная идея: разделить вариацию на между-объектную и внутри-объектную.
- Внутри-объектная вариация учитывает различия внутри одного объекта по разным условиям.

5. Процедура вычисления дисперсионного анализа для связанных выборок

Шаги:

1. Собрать данные по всем группам и объектам.
2. Вычислить общие средние по всем данным.
3. Для каждого объекта вычислить его среднее значение по всем условиям.
4. Рассчитать сумму квадратов между объектами (SS_{between}), внутри объектов (SS_{within}), а также общую сумму квадратов (SS_{total}).
5. Определить степени свободы:
 - $df_{\text{between}} = k - 1$, где k — число условий
 - $df_{\text{within}} = n(k - 1)$, где n — число объектов
6. Вычислить средние квадраты:
 - $MS_{\text{between}} = SS_{\text{between}} / df_{\text{between}}$
 - $MS_{\text{within}} = SS_{\text{within}} / df_{\text{within}}$
7. Рассчитать F-статистику:

$$F = \frac{MS_{\text{within}}}{MS_{\text{between}}}$$

8. Сравнить полученное значение F с критическим из таблицы F-распределения при заданных уровнях α и соответствующих степенях свободы.

Если F превышает критическое значение, делается вывод о наличии статистически значимых различий между группами.

ИТОГ:

- Парный t-критерий применяется для сравнения двух зависимых выборок; он основан на анализе разностей парных наблюдений.
- Дисперсионный анализ для связанных выборок расширяет этот подход на более чем две группы, позволяя одновременно проверять гипотезу о равенстве средних во всех группах с учетом внутренней связанности данных.

11. Корреляционное исследование - особенности, причины широкого использования в медицине и психологии. Задача предсказания на основе имеющихся статистических данных. Зависимая и независимая переменные (предикторы). Уравнение регрессии. Расчетные процедуры определения коэффициентов уравнения регрессии.

Ответ:

Корреляционное исследование — это статистический метод, направленный на выявление и измерение степени и направления связи между двумя или более переменными. В медицине и психологии оно широко используется для определения взаимосвязей между различными факторами, например, между уровнем стресса и состоянием здоровья или между образом жизни и психологическим благополучием.

Особенности корреляционного исследования:

- Позволяет установить наличие связи между переменными.
- Не предполагает причинно-следственной связи.
- Используется для предварительной оценки взаимосвязей перед проведением более сложных исследований.
- Может включать расчет коэффициента корреляции (например, коэффициента Пирсона).

Причины широкого использования в медицине и психологии:

- Возможность анализа больших объемов данных.
- Помогает выявить потенциальные факторы риска или защитные факторы.
- Простота проведения и интерпретации.
- Используется для формирования гипотез для дальнейших исследований.

Задача предсказания на основе статистических данных: На основе известных значений одних переменных (предикторов) можно предсказать значения других переменных (зависимых). Например, по уровню физической активности предсказать уровень артериального давления.

Зависимая и независимая переменные (предикторы):

- Зависимая переменная — это та, которую мы пытаемся предсказать или объяснить (например, уровень тревожности).
- Независимая переменная — это фактор или предиктор, который влияет на зависимую переменную (например, стрессовые ситуации).

Уравнение регрессии: Линейное уравнение регрессии имеет вид: $Y = a + bX + \varepsilon$ где:

- Y — зависимая переменная,
- X — независимая переменная,
- a — свободный член (константа),
- b — коэффициент регрессии (наклон),
- ε — случайная ошибка.

Расчетные процедуры определения коэффициентов уравнения регрессии:

1. Метод наименьших квадратов: минимизация суммы квадратов отклонений наблюдаемых значений Y_i от значений, предсказанных моделью $a + bX_i$.
2. Формулы для коэффициентов:
 - Коэффициент наклона: $b = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}$
 - Свободный член: $a = \bar{Y} - b\bar{X}$ где $\bar{X}$ и $\bar{Y}$ — средние значения по выборке.

Эти процедуры позволяют определить линию наилучшего соответствия данным и использовать ее для предсказаний новых значений зависимой переменной при известных значениях независимой.

12. Условия оптимальности простой линейной модели регрессии. Метод наименьших квадратов. Вычисление дисперсии ошибки оценки для линейной регрессии. Сравнение двух линий регрессии. Определение значимости различий.

Ответ:

1. Условия оптимальности простой линейной модели регрессии

Модель: $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, i = 1, \dots, n$

Условия оптимальности (метод наименьших квадратов):

- Минимизация суммы квадратичных ошибок: $S(\beta_0, \beta_1) = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2$
- Экстремум функции S:
 - Производные по параметрам равны нулю:

$$\frac{\partial S}{\partial \beta_0} = 0, \frac{\partial S}{\partial \beta_1} = 0$$

- Решение системы нормальных уравнений:

$$n\beta^0 + \beta^1 \sum x_i = \sum y_i$$

$$\beta^0 \sum x_i + \beta^1 \sum x_i^2 = \sum x_i y_i$$

2. Метод наименьших квадратов

- Цель: найти оценки параметров (β^0, β^1) , минимизирующие сумму квадратичных ошибок.
- Решение:
 - Оценки:

$$\beta^1 = \frac{n \sum x_i^2 - (\sum x_i)^2}{n \sum x_i y_i - (\sum x_i)(\sum y_i)}, \beta^0 = \bar{y} - \beta^1 \bar{x}$$

- где $\bar{x} = \frac{1}{n} \sum x_i, \bar{y} = \frac{1}{n} \sum y_i$.

3. Вычисление дисперсии ошибки оценки для линейной регрессии

- Оценка дисперсии ошибок:

$$(n-2) \frac{1}{(n-2)} \sum_{i=1}^n (y_i - \hat{y}_i)^2 = n - 2 \text{RSS}$$

где $\text{RSS} = \sum_{i=1}^n (y_i - (\beta^0 + \beta^1 x_i))^2$.

- Дисперсия оценки наклона:

$$\text{Var}(\beta^1) = \frac{s^2}{S_{xx}}, S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2$$

- Дисперсия оценки свободного члена:

$$\text{Var}(\beta^0) = \frac{s^2}{n} (1 + \frac{\bar{x}^2}{S_{xx}})$$

4. Сравнение двух линий регрессии и определение значимости различий

Задача: проверить, существенно ли отличаются две модели или два коэффициента.

Методы:

- Тесты гипотез для коэффициентов:

Для проверки значимости коэффициента β_j :

Гипотеза:

$$H_0: \beta_j = 0; H_a: \beta_j \neq 0.$$

Статистика:

$$t_j = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)}, SE(\hat{\beta}_j) = \sqrt{\text{Var}(\hat{\beta}_j)}$$

- .

Сравниваем с критическим значением из t-распределения с $n-2$ степенями свободы.

- Сравнение двух линий регрессии (например, по разным группам):

Используют F-тест для сравнения моделей или разницы между наклонами:

- Проверка равенства наклонов:

Гипотеза:

$$H_0: \beta_1^{(1)} = \beta_1^{(2)}; H_a: \beta_1^{(1)} \neq \beta_1^{(2)}.$$

Статистика F:

$$F = \frac{\text{RSS}_{\text{unrestricted}} / (n - m)}{(\text{RSS}_{\text{restricted}} - \text{RSS}_{\text{unrestricted}}) / k},$$

где k = число ограничений, m = число параметров в полной модели.

Итог:

Для проведения анализа необходимо:

- Проверить условия линейной регрессии (независимость ошибок, гомоскедастичность, нормальность остатков).
- Оценить параметры методом МНК.
- Вычислить дисперсию ошибок и доверительные интервалы.
- Провести статистические тесты для определения значимости коэффициентов и различий между моделями.

13. Методы проверки линейности. Стратификация независимой переменной. Нелинейное оценивание связей между переменными.

Ответ:

Методы проверки линейности, стратификация независимой переменной и нелинейное оценивание связей между переменными — важные аспекты анализа данных в статистике и исследовательской практике. Ниже приведено краткое описание каждого из них:

1. Методы проверки линейности:
 - Визуальный анализ: построение графиков рассеяния (scatter plots) для оценки формы связи между переменными.
 - Остаточные диаграммы: анализ остатков модели (например, остатки против предсказанных значений) для выявления нелинейных зависимостей.
 - Статистические тесты: использование тестов на линейность, таких как тест Бреуша-Пагана или тесты на основе добавления полиномиальных или квадратичных терминов в модель.
 - Анализ моделей с полиномиальными или функциями преобразования: сравнение моделей с линейными и нелинейными компонентами по критериям качества подгонки (например, AIC, BIC).
2. Стратификация независимой переменной:
 - Метод заключается в разделении данных по уровням или диапазонам значения независимой переменной (например, по квантилям или фиксированным интервалам).
 - Цель — выявить изменение эффекта зависимой переменной в разных слоях (стратах), что помогает понять наличие или отсутствие взаимодействий и нелинейных эффектов.
 - После стратификации можно провести отдельный анализ внутри каждой страты или использовать методы стратифицированного анализа для оценки стабильности связи.
3. Нелинейное оценивание связей между переменными:
 - Использование методов, позволяющих моделировать нелинейные зависимости:
 - Полиномиальные регрессии (например, квадратичная или кубическая модель).
 - Модели с функциями ядра (например, регрессия с использованием сглаживающих сплайнов).
 - Модели машинного обучения: деревья решений, случайные леса, градиентный бустинг и нейронные сети.
 - Метод локальной регрессии (LOESS/LOWESS), который позволяет гибко моделировать нелинейные связи без предположения о конкретной форме функции.
 - Эти методы позволяют более точно моделировать сложные зависимости между переменными, если линейная модель не подходит.

14. Факторные планы в психологии и медицине. Двух- и многофакторный дисперсионный анализ. Формулировка задачи. Требования к подготовке данных.

Ответ:

Факторные планы в психологии и медицине, а также двух- и многофакторный дисперсионный анализ — важные методы статистического анализа, позволяющие исследовать влияние нескольких факторов на изучаемую переменную.

Факторные планы в психологии и медицине

Факторные планы — это экспериментальные конструкции, в которых исследуются влияние одного или нескольких факторов (независимых переменных) на зависимую переменную. В психологии и медицине такие планы позволяют выявлять взаимодействия между факторами, оценивать их индивидуальное и совместное влияние.

- Однофакторные планы: один фактор с несколькими уровнями (например, тип терапии: медикаментозная, психологическая, комбинированная).
- Многофакторные планы: два или более факторов, каждый с несколькими уровнями (например, тип терапии и возрастная группа).

Двух- и многофакторный дисперсионный анализ

Дисперсионный анализ (ANOVA) — метод статистической проверки гипотез о равенстве средних значений. В многофакторных планах он позволяет определить:

- Влияние каждого фактора отдельно (основные эффекты).
- Взаимодействия между факторами (например, как эффект терапии зависит от возраста).

Формулировка задачи

Пример задачи: Исследовать влияние типа терапии (фактор А) и возраста пациента (фактор В) на уровень симптоматики. Требуется определить:

- Есть ли значимое различие в уровне симптоматики между разными типами терапии.
- Есть ли различия между возрастными группами.
- Есть ли взаимодействие между типом терапии и возрастом.

Требования к подготовке данных

1. Независимость наблюдений: данные по разным группам должны быть независимы.
2. Нормальность распределения: распределение зависимой переменной в каждой группе должно быть приблизительно нормальным.
3. Гомогенность дисперсий: дисперсии внутри групп должны быть примерно равны (проверяется тестами, например, Левене или Бартлетта).
4. Размер выборки: желательно иметь достаточно большие группы для повышения надежности анализа.

Перед проведением анализа рекомендуется:

- Проверить нормальность распределения с помощью тестов (Шапиро-Уилка, Колмогорова-Смирнова).
- Проверить однородность дисперсий.
- При необходимости применить преобразования данных или использовать непараметрические методы.

15. Дискриминантный анализ. Формулировка задачи. Требования к подготовке данных. Отношения шансов и логистическая регрессия. Формулировка задачи. Требования к подготовке данных.

Ответ:

Дискриминантный анализ и логистическая регрессия — это методы классификации, используемые для определения принадлежности объекта к одной из нескольких категорий на основе набора признаков.

Дискриминантный анализ

Формулировка задачи:

Задача состоит в классификации объектов по группам, основываясь на их признаках.

Например, определить, является ли пациент больным или здоровым, исходя из медицинских показателей.

Требования к подготовке данных:

- Нормальность распределения признаков: предполагается, что признаки внутри каждой группы нормально распределены.

- Гомогенность дисперсий: Дисперсии признаков в разных группах должны быть примерно равны (для классического линейного дискриминантного анализа).
- Отсутствие сильной мультиколлинеарности: Признаки не должны быть сильно коррелированы между собой.
- Недостаток пропущенных данных: необходимо заполнить или исключить пропущенные значения.
- Масштабирование признаков: в некоторых случаях рекомендуется стандартизация или нормализация признаков для повышения стабильности модели.

Отношения шансов и логистическая регрессия

Формулировка задачи:

Цель — предсказать вероятность принадлежности объекта к определенной категории (например, больной или здоровый) на основе признаков. В отличие от дискриминантного анализа, логистическая регрессия моделирует вероятность напрямую.

Отношения шансов (Odds Ratio):

Отношение шансов — это отношение вероятности события к вероятности его отсутствия:

$$\text{Odds} = \frac{p}{1-p}$$

где p — вероятность события.

Логистическая регрессия:

Модель связывает логарифм отношения шансов с линейной комбинацией признаков:

$$\log(\text{Odds}) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$$

или в виде:

$$p = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k)}}$$

Требования к подготовке данных:

- Отсутствие пропущенных значений: необходимо заполнить пропуски.
- Масштабирование признаков: хотя логистическая регрессия менее чувствительна к масштабу, стандартизация может помочь при использовании регуляризации.
- Линейность связи между признаками и логитами: желательно проверить и при необходимости преобразовать признаки (например, логарифмировать).
- Отсутствие сильной мультиколлинеарности: чтобы избежать нестабильных оценок коэффициентов.
- Баланс классов: при сильном дисбалансе может потребоваться использование методов балансировки или взвешивания.

16. Кластерный анализ данных. Формулировка задачи. Требования к подготовке данных. Восходящий и нисходящий алгоритм кластеризации. Особенности кластерного корреляционного анализа.

Ответ:

Кластерный анализ данных — это метод многомерного анализа, направленный на группировку объектов (наблюдений, элементов) в кластеры (группы) так, чтобы объекты внутри одного кластера были максимально похожи друг на друга, а объекты из разных кластеров — максимально различны.

Формулировка задачи: Дана выборка объектов с набором признаков. Требуется разбить эти объекты на группы (кластеры), чтобы внутри каждой группы объекты были более похожи друг на друга по выбранным признакам, чем на объекты из других групп. Цель — выявить скрытые структуры и закономерности в данных.

Требования к подготовке данных:

1. Очистка данных: устранение пропусков, выбросов и ошибок.
2. Нормализация или стандартизация признаков: приведение признаков к единой шкале (например, z-преобразование или масштабирование), чтобы избежать доминирования признаков с большим диапазоном.

3. Выбор признаков: исключение нерелевантных или избыточных признаков для повышения качества кластеризации.
4. Обработка категориальных признаков: преобразование в числовой формат (например, one-hot encoding).
5. Удаление мультиколлинеарных признаков: для избежания искажения расстояний.

Восходящий алгоритм кластеризации (агломеративный):

- Начинается с каждого объекта как отдельного кластера.
- На каждом шаге объединяет два наиболее похожих кластера по выбранной метрике (например, по расстоянию между центрами или по связности).
- Процесс продолжается до тех пор, пока не будет достигнуто желаемое число кластеров или пока все объекты не объединятся в один кластер.
- Визуализируется в виде дендрограммы.

Нисходящий алгоритм кластеризации (дедуктивный):

- Начинается с одного большого кластера, содержащего все объекты.
- На каждом шаге делит существующие кластеры на более мелкие подгруппы по определённым правилам.
- Продолжается до тех пор, пока не достигнуто нужное число кластеров или пока разделения не перестанут быть значимыми.
- Реже используется по сравнению с восходящими методами.

Особенности кластерного корреляционного анализа:

- Использует корреляционные показатели (например, коэффициент Пирсона) для определения сходства между объектами или переменными.
- В отличие от классической кластеризации по расстояниям, фокусируется на выявлении групп переменных или объектов с высокой корреляционной связью.
- Полезен для выявления структурных связей и взаимозависимостей в данных.
- Может применяться для кластеризации переменных (факторный анализ) или объектов с учетом их корреляционных характеристик.

17. Факторный анализ. Виды факторного анализа. Анализ главных компонент. Формулировка задачи. Требования к подготовке данных.

Ответ:

Факторный анализ — это статистический метод, используемый для выявления скрытых факторов (латентных переменных), объясняющих взаимосвязи между наблюдаемыми переменными. Он помогает упростить сложные многомерные данные, выявляя основные компоненты, которые объясняют большую часть вариации.

Виды факторного анализа:

1. Эксплораторный факторный анализ (ЭФА)
Используется для исследования структуры данных без предварительных гипотез о количестве факторов или их природе. Цель — выявить возможные факторы и понять структуру данных.
2. Конфирматорный факторный анализ (КФА)
Предполагает наличие заранее сформулированной модели факторов и проверку её соответствия данным. Используется для подтверждения гипотез о структуре факторов.
3. Общий факторный анализ
Включает моделирование общих факторов, влияющих на несколько переменных одновременно, с целью объяснить их корреляции.
4. Анализ главных компонент (Principal Component Analysis, PCA)
Хотя технически не является классическим факторным анализом, часто используется как метод снижения размерности и предварительный этап перед факторным анализом.

Анализ главных компонент (ПКА):

- Цель: уменьшить размерность данных, сохраняя максимально возможную дисперсию исходных переменных.
 - Задача: найти линейные комбинации исходных переменных (компоненты), которые объясняют максимальную часть вариации данных.
 - Формулировка задачи:
Для набора переменных $X_1, X_2, \dots, X_p$ найти векторы весов $a = (a_1, a_2, \dots, a_p)$, такие что линейная комбинация $Y = a_1 X_1 + a_2 X_2 + \dots + a_p X_p$ максимизирует дисперсию при условии нормировки $\sum a_i^2 = 1$.
-

Требования к подготовке данных:

1. Отсутствие пропущенных значений:
Необходимо либо заполнить пропуски (импутация), либо исключить такие наблюдения.
2. Нормализация и стандартизация:
Обычно переменные стандартизируют (приводят к среднему 0 и стандартному отклонению 1), чтобы обеспечить равные условия для всех переменных.
3. Проверка на адекватность данных:
 - Тест Каймана-Мейера-Олкина (КМО) — показатель пригодности данных для факторного анализа.
 - Барлеттовский тест сферы корреляций — проверка значимости корреляционной матрицы.
4. Корреляционная матрица:
Должна быть достаточно коррелированной для выявления факторов; низкая корреляция говорит о слабой структуре.
5. Выбор числа факторов/компонент:
На основе критериев (например, собственных значений > 1 , графика скритов или доли объясненной дисперсии).