

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ**
Федеральное государственное автономное образовательное учреждение
высшего образования
«Пермский государственный национальный исследовательский университет»

Кафедра математических и естественнонаучных дисциплин

МЕТОДЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ
Методические рекомендации
для самостоятельной работы по изучению дисциплины для студентов
специальности 44.02.03 Педагогика дополнительного образования

Пермь 2022

ОБЩИЕ ПОЛОЖЕНИЯ О САМОСТОЯТЕЛЬНОЙ РАБОТЕ СТУДЕНТОВ

Самостоятельная работа по дисциплине «Методы математической статистики» – это педагогически управляемый процесс самостоятельной деятельности студентов, обеспечивающий реализацию целей и задач по овладению необходимым объемом знаний, умений и навыков, опыта творческой работы и развитию профессиональных интеллектуально-волевых, нравственных качеств будущего специалиста.

Выделяют два вида самостоятельной работы:

- аудиторная, выполняется на занятиях под руководством преподавателя и по его заданию;
- внеаудиторная, выполняется студентом по заданию преподавателя, но без его непосредственного участия.
- Основные виды аудиторной самостоятельной работы студентов при изучении дисциплины «Методы математической статистики»:
 - Выполнение практических работ;
 - выполнение письменных заданий, тестирование;
 - выступление с сообщением по новому материалу;
 - конспектирование, работа с литературными источниками;
 - работа в сети Интернет;
 - выполнение самостоятельных работ и др.

Основные виды внеаудиторной самостоятельной работы студентов при изучении дисциплины «Методы математической статистики»:

- работа с учебником;
- конспектирование отдельного вопроса пройденной темы;
- работа со справочной литературой;
- подготовка рефератов и презентаций по темам;
- изготовление наглядных пособий и моделей;
- работа в сети Интернет и др.

Самостоятельная работа студентов проводится с целью:

- систематизации и закрепления полученных знаний и практических умений и навыков студентов;
- углубления и расширения теоретических и практических знаний;
- формирования умений использовать специальную, справочную литературу, Интернет;
- развития познавательных способностей и активности студентов, творческой инициативы, самостоятельности, ответственности и организованности;
- формирования самостоятельности мышления, способностей к саморазвитию, самосовершенствованию и самореализации;
- развития исследовательских знаний.
- Лимит времени для проведения самостоятельной работы студентов аудиторно отводится преподавателем непосредственно на занятии, для каждого вида работы определенный.

Основной формой контроля самостоятельной работы студента являются практические работы, защита презентаций и рефератов на занятиях.

Самостоятельные работы являются важным средством проверки уровня знаний, умений и навыков.

Внеаудиторная самостоятельная работа обучающихся – планируемая учебная, учебно-исследовательская, проектная работа, выполняемая за рамками расписания учебных занятий по заданию и при методическом руководстве преподавателя, но без его непосредственного участия и является обязательной для каждого студента.

Целью самостоятельной работы студентов является:

- обеспечение профессиональной подготовки выпускника в соответствии с ФГОС СПО;
- формирование и развитие общих компетенций, определённых в ФГОС СПО;
- формирование и развитие профессиональных компетенций, соответствующих основным видам профессиональной деятельности.

Задачами, реализуемыми в ходе проведения внеаудиторной самостоятельной работы обучающихся, в образовательной среде являются:

- систематизация, закрепление, углубление и расширение полученных теоретических знаний и практических умений студентов;
- развитие познавательных способностей и активности студентов: творческой инициативы, самостоятельности, ответственности и организованности;
- формирование самостоятельности мышления: способности к саморазвитию, самосовершенствованию и самореализации;
- овладение практическими навыками применения информационно-коммуникационных технологий в профессиональной деятельности;
- развитие исследовательских умений.

Контроль результатов самостоятельной работы обучающихся может осуществляться в пределах времени, отведенного на обязательные учебные занятия и самостоятельную работу по дисциплине «Методы математической статистики» и может проходить в письменной, устной или смешанной форме с предоставлением изделия или продукта творческой деятельности.

Критериями оценки результатов внеаудиторной самостоятельной работы обучающегося являются:

- уровень освоения учебного материала;
- умение использовать теоретические знания и умения при выполнении практических задач;
- уровень сформированности общих и профессиональных компетенций.

Методические рекомендации по составлению конспекта

Внимательно прочитайте текст. Уточните в справочной литературе непонятные слова. При записи не забудьте вынести справочные данные на поля конспекта.

Выделите главное, составьте план.

Кратко сформулируйте основные положения текста, отметьте аргументацию автора.

Законспектируйте материал, четко следуя пунктам плана. При конспектировании старайтесь выразить мысль своими словами. Записи следует вести четко, ясно.

Грамотно записывайте цитаты. Цитируя, учитывайте лаконичность, значимость мысли.

Методические рекомендации по решению задач

В процессе изучения математики наряду с некоторыми теоретическими сведениями студенты овладевают и закрепляют способы решения задач. Обычно с такими способами знакомит сам преподаватель, показывая решение задач по темам. Наиболее эффективным при этом является такой подход, при котором преподаватель раскрывает перед студентами технологию решения задачи, показывает, чем мотивировано применение некоторого метода решения, чем обусловлен выбор того или иного пути.

Работа над задачей тоже может быть полностью самостоятельной работой студентов. Она преследует несколько целей:

- продолжить формирование умений самостоятельно изучать текст, который в данном случае представляет собой задачу;
- обучить рассуждениям;
- обучить оформлению решения задач. К тому же студенты будут знать, что у них имеется образец рассуждений и оформления задачи, к которому они могут обратиться при решении другой задачи или при проверке правильности своего решения.

Непременным условием усвоения новых теоретических сведений и овладения новыми приемами решения задач является выполнение студентами тренировочных упражнений, в ходе которого приобретенные знания становятся полным достоянием студентов. Как известно, существуют две формы организации такой тренировочной работы – фронтальная работа и самостоятельная работа. Фронтальная работа на занятиях по методам математической статистики – это традиционная, давно сложившаяся форма. Схематически ее можно описать так: один из студентов выполняет задание на доске, остальные выполняют это же задание в тетрадях. Самостоятельная работа студентов на уроке состоит в выполнении без помощи преподавателя и товарищей задания.

Большие возможности для подготовки студентов к творческому труду и самостоятельному пополнению знаний имеет самостоятельное выполнение

заданий. В этом случае студент без какой-либо помощи должен наметить пути решения, правильно выполнить все построения, преобразования, вычисления и т. п. В таком случае мысль студента работает наиболее интенсивно. Он приобретает практический навык работы в ситуации, с которой ему неоднократно придется сталкиваться в последующей трудовой деятельности. Вместе с тем самостоятельная работа студентов на занятиях по дисциплине «Методы математической статистики» имеет и свои недостатки. Усилия студента могут оказаться напрасными и не привести к результату, если он недостаточно подготовлен к решению поставленной задачи. Студент не слышит комментариев к решению, а рассуждения, которые он проводит мысленно, могут быть не всегда правильными и достаточно полными, причем возможности обнаружить это студент не имеет. Вообще при самостоятельном выполнении заданий мыслительные процессы не могут быть проконтролированы преподавателем. Поэтому даже верный ответ может оказаться случайным. Исправление ошибок, допущенных при самостоятельной работе, происходит в ходе ее проверки по окончании всей работы. Поэтому, выполняя упражнение самостоятельно, студент, не усвоивший материал, может повторять одну и ту же ошибку от примера к примеру и невольно закрепить неправильный алгоритм.

Методические рекомендации по выполнению практических занятий

Для того чтобы практические занятия приносили максимальную пользу, необходимо помнить, что упражнение и решение ситуативных задач проводятся по вычитанному на лекциях материалу и связаны, как правило, с детальным разбором отдельных вопросов лекционного курса. Следует подчеркнуть, что только после усвоения лекционного материала с определенной точки зрения (а именно с той, с которой он излагается на лекциях) он будет закрепляться на практических занятиях как в результате обсуждения и анализа лекционного материала, так и с помощью решения ситуативных задач. При этих условиях студент не только хорошо усвоит материал, но и научится применять его на практике, а также получит дополнительный стимул (и это очень важно) для активной проработки лекции.

При самостоятельном решении поставленных задач нужно обосновывать каждый этап действий, исходя из теоретических положений курса. Если обучающийся видит несколько путей решения проблемы (задачи), то нужно сравнить их и выбрать самый рациональный. Полезно до начала решения поставленных задач составить краткий план решения проблемы (задачи). Решение проблемных задач или примеров следует излагать подробно, нужно сопровождать комментариями, схемами, чертежами и рисунками, инструкциями по выполнению.

Следует помнить, что решение каждой учебной задачи должно доводиться до окончательного логического ответа, которого требует условие, и по возможности с выводом. Полученный результат следует проверить способами, вытекающими из существа данной задачи.

Методические рекомендации по написанию контрольной работы

Контрольная работа – промежуточный метод проверки знаний обучающегося с целью определения конечного результата в обучении по данной теме или разделу. Она призвана систематизировать знания, позволяет повторить и закрепить материал. При выполнении студенты ограничены во времени, могут использовать любые учебные пособия, консультации преподавателя.

Методические рекомендации по написанию реферата

Реферат необходимо сдать в печатном виде на листе формата А4, выполненном шрифтом Times New Roman 14 пунктов.

Требования, предъявляемые к реферату:

Реферат (доклад) должен быть оформлен в MS Word, шрифт текста Times New Roman, 14 пт., интервал 1.

1. Титульный лист
2. Содержание
3. Введение
4. Основная часть реферата
5. Заключение
6. Список используемой литературы

Если возникнут затруднения в процессе работы, обратитесь к преподавателю.

Методические рекомендации по подготовке сообщения

Сообщение – это сокращенная запись информации, в которой должны быть отражены основные положения текста, сопровождающиеся аргументами, 1–2 самыми яркими и в то же время краткими примерами.

Сообщение составляется по нескольким источникам, связанным между собой одной темой. Вначале изучается тот источник, в котором данная тема изложена наиболее полно и на современном уровне научных и практических достижений. Записанное сообщение дополняется материалом других источников.

Этапы подготовки сообщения:

1. Прочитайте текст.
2. Составьте его развернутый план.
3. Подумайте, какие части можно сократить так, чтобы содержание было понято правильно и, главное, не исчезло.
4. Объедините близкие по смыслу части.
5. В каждой части выделите главное и второстепенное, которое может быть сокращено при конспектировании.
6. При записи старайтесь сложные предложения заменить простыми.

Тематическое и смысловое единство сообщения выражается в том, что все его компоненты связаны с темой первоисточника.

Сообщение должно содержать информацию на 3-5 мин. и сопровождаться презентацией, схемами, рисунками, таблицами и т.д.

Методические рекомендации по составлению презентаций

Требования к презентации

На первом слайде размещается:

название презентации;

автор: ФИО, группа, название учебного учреждения (соавторы указываются в алфавитном порядке);

год.

На втором слайде указывается содержание работы, которое лучше оформить в виде гиперссылок (для интерактивности презентации).

На последнем слайде указывается список используемой литературы в соответствии с требованиями, интернет-ресурсы указываются в последнюю очередь.

Оформление слайдов	
Стиль	<ul style="list-style-type: none">– необходимо соблюдать единый стиль оформления;– нужно избегать стилей, которые будут отвлекать от самой презентации;– вспомогательная информация (управляющие кнопки) не должны преобладать над основной информацией (текст, рисунки)
Фон	<ul style="list-style-type: none">– для фона выбираются более холодные тона (синий или зеленый)
Использование цвета	<ul style="list-style-type: none">– на одном слайде рекомендуется использовать не более трех цветов: один для фона, один для заголовков, один для текста;– для фона и текста используются контрастные цвета;– особое внимание следует обратить на цвет гиперссылок (до и после использования)
Анимационные эффекты	<ul style="list-style-type: none">– нужно использовать возможности компьютерной анимации для представления информации на слайде;– не стоит злоупотреблять различными анимационными эффектами; анимационные эффекты не должны отвлекать внимание от содержания информации на слайде
Представление информации	
Содержание информации	<ul style="list-style-type: none">– следует использовать короткие слова и предложения;– времена глаголов должно быть везде одинаковым;– следует использовать минимум предлогов, наречий, прилагательных;

	– заголовки должны привлекать внимание аудитории
Расположение информации на странице	– предпочтительно горизонтальное расположение информации; – наиболее важная информация должна располагаться в центре экрана; – если на слайде располагается картинка, надпись должна располагаться под ней
Шрифты	– для заголовков не менее 24; – для остальной информации не менее 18; – шрифты без засечек легче читать с большого расстояния; – нельзя смешивать разные типы шрифтов в одной презентации; – для выделения информации следует использовать жирный шрифт, курсив или подчеркивание того же типа; – нельзя злоупотреблять прописными буквами (они читаются хуже, чем строчные).
Способы выделения информации	Следует использовать: – рамки, границы, заливку – разные цвета шрифтов, штриховку, стрелки – рисунки, диаграммы, схемы для иллюстрации наиболее важных фактов
Объем информации	– не стоит заполнять один слайд слишком большим объемом информации: люди могут одновременно запомнить не более трех фактов, выводов, определений. – наибольшая эффективность достигается тогда, когда ключевые пункты отражаются по одному на каждом отдельном слайде.
Виды слайдов	Для обеспечения разнообразия следует использовать разные виды слайдов: с текстом, с таблицами, с диаграммами.

Критерии оценки презентации

Критерии оценки	Содержание оценки
1. Содержательный критерий	правильный выбор темы, знание предмета и свободное владение текстом, грамотное использование научной терминологии, импровизация, речевой этикет
2. Логический критерий	стройное логико-композиционное построение речи, доказательность, аргументированность
3. Речевой критерий	использование языковых (метафоры, фразеологизмы, пословицы, поговорки и т.д.) и неязыковых (поза, манеры и пр.) средств выразительности; фонетическая организация речи, правильность ударения, четкая дикция, логические ударения и пр.
4. Психологический критерий	взаимодействие с аудиторией (прямая и обратная связь), знание и учет законов восприятия речи,

	использование различных приемов привлечения и активизации внимания
5. Критерий соблюдения дизайн-эргономических требований к компьютерной презентации	соблюдены требования к первому и последним слайдам, прослеживается обоснованная последовательность слайдов и информации на слайдах, необходимое и достаточное количество фото- и видеоматериалов, учет особенностей восприятия графической (иллюстративной) информации, корректное сочетание фона и графики, дизайн презентации не противоречит ее содержанию, грамотное соотнесение устного выступления и компьютерного сопровождения, общее впечатление от мультимедийной презентации

Методические рекомендации по использованию ФГИС «Моя школа»

1. Для входа на страницу ФГИС «Моя Школа» необходимо перейти по ссылке: <https://myschool.edu.ru/>.

2. Для функциональной роли «Педагогический работник» на главной странице доступны все разделы и функциональные элементы Системы.

Вариационные ряды и их характеристики. Основы математической теории выборочного метода.

Самостоятельная работа

Решение заданий на определение характеристик вариационных рядов

Цель: *знать характеристики вариационных рядов.*

Методические рекомендации

Мода

Мода (Mo) – наиболее часто встречающееся значение признака. В предыдущем примере (ранжированный ряд уровня личностной тревожности) мы имеем две моды: $Mo_1 = 36$ и $Mo_2 = 45$ (эти значения переменной встречаются трижды, в то время как все остальные – по 1 или 2 раза). В зависимости от того, сколько значений признака удовлетворяют определению моды, различают *мономодальные* (имеющие одну моду), *бимодальные* (имеющие две моды) и *полимодальные* распределения (имеют более чем две моды), а также *распределения, не имеющие моды* (все значения признака встречаются примерно с одинаковой частотой). В бимодальном и полимодальном распределениях, в свою очередь, можно определить наибольшую и наименьшую моды.

В тех случаях, когда анализируются таблицы сгруппированных частот исследуемого признака, как правило, определяется *модальный класс*, т. е. тот

класс распределения, в который попадает наибольшее количество частот (значений признака). Так, для иллюстрации зачерненный столбец на рис. 3.1, а соответствует модальному классу.

Мода не является достаточно строгой мерой центральной тенденции, поскольку она не учитывает характера распределения переменных, а значит может использоваться лишь в предварительных выводах и прогнозах. Кроме того, необходимо использовать моду только для больших объемов выборок, поскольку для малых она недостаточно информативна.

Медиана

Медиана (Md) – значение, которое делит упорядоченное множество данных (ранжированный ряд) пополам так, что одна половина значений оказывается больше, а другая – меньше медианы. Медиана – среднее значение ранжированного ряда. Если число значений нечетное, то медиана соответствует среднему члену ряда, если четное, то медиана есть среднее между двумя центральными значениями (в предыдущем примере $Md = 41,5$).

Медиана соответствует 50-му перцентилю, 5-му децилю или 2-му квартилю в группе данных, т. е. $Md = P_{50} = D_5 = Q_2$.

Мода и медиана не учитывают разброса данных, и переменные, лежащие в стороне от центра, не влияют на их величину.

Среднее арифметическое значение

Среднее арифметическое значение, или просто *среднее* (\bar{x}), равно сумме переменных, деленной на их число.

Для несгруппированных переменных среднее арифметическое вычисляется по формуле:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} \sum x_i.$$

Для сгруппированных переменных можно воспользоваться другой формулой – среднее будет соответствовать сумме произведений средних значений каждого класса и частоты встречаемости значения признака в данном классе:

$$\bar{x} = \frac{\bar{x}_1 f_1 + \bar{x}_2 f_2 + \dots + \bar{x}_N f_N}{n} = \frac{1}{n} \sum \bar{x}_i f_i.$$

Среднее арифметическое может использоваться и для тех признаков, для которых не найден способ количественного измерения (шкала порядка). Для этого в качестве x_i используются ранговые числа, а среднее принято называть *непараметрическим средним*.

Взвешенное среднее арифметическое используется в тех случаях, когда разные составляющие имеют разный «удельный вес» в формировании общей совокупности:

$$\bar{x}_w = \frac{x_1 p_1 + x_2 p_2 + \dots + x_N p_N}{n} = \frac{\sum x_i p_i}{n},$$

или: $\bar{x}_w = \frac{x_1 p_1 + x_2 p_2 + \dots + x_N p_N}{p_1 + p_2 + \dots + p_N}$, где n – объем выборки, N – число классов.

Среднее геометрическое значение

Среднее геометрическое значение используется для вычисления центральной тенденции при прогрессивно возрастающих квантилях (когда распределение значений переменной имеет выраженную положительную (правостороннюю) асимметрию).

Формула среднего геометрического:

$$x_g = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n} = (x_1 \cdot x_2 \cdot \dots \cdot x_n)^{\frac{1}{n}}.$$

Среднее отклонение

Среднее отклонение (MD) – параметрическая мера изменчивости, предложенная в свое время Г. Т. Фехнером. Среднее отклонение равно сумме отклонений от среднего значения (или, другими словами, сумме расстояний между x_i и \bar{x}), взятых по модулю:

$$MD = \frac{|x_1 - \bar{x}| + |x_2 - \bar{x}| + \dots + |x_n - \bar{x}|}{n} = \frac{\sum |x_i - \bar{x}|}{n}.$$

Дисперсия

Дисперсия представляет собой сумму квадратов отклонений от среднего (сумму квадратов расстояний между x_i и \bar{x}):

$$\sigma^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n-1} = \frac{\sum (x_i - \bar{x})^2}{n-1}.$$

Среднеквадратичное (стандартное) отклонение

Стандартное отклонение (σ) соответствует квадратному корню из дисперсии. Наряду с дисперсией является одной из наиболее часто используемых мер вариабельности признака.

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{\sum x_i^2 - (\sum x_i)^2 / n}{n-1}}.$$

(5.9)

Коэффициент вариации

Коэффициент вариации (V) есть отношение стандартного отклонения к среднему арифметическому значению, выраженное в процентах:

$$V = \frac{\sigma}{\bar{x}} \cdot 100\%$$

Используя методические рекомендации, выполните задания:

1. Вычислить:

- моду;
- медиану;
- среднее арифметическое;
- среднее геометрическое;
- среднее отклонение;
- дисперсию;
- среднеквадратичное (стандартное) отклонение;

- коэффициент вариации.

Вариант 1.

Для поступления в ВУЗ необходимо сдать 3 экзамена. Каждый экзамен оценивается по 10-ти бальной системе. После всех экзаменов абитуриенты данной группы, состоящей из 30 человек, имели следующие баллы:

17, 15, 13, 30, 28, 17, 19, 25, 26, 23, 21, 29, 26, 20, 18, 19, 23, 24, 29, 25, 27, 30, 12, 21, 16, 18, 28, 16, 16, 30.

Вариант 2.

Проведено измерение роста 30 учеников данного класса, в ходе которого были получены следующие данные (в см.):

151, 153, 144, 149, 155, 150, 163, 160, 154, 143, 147, 152, 141, 143, 151, 153, 156, 162, 148, 156, 152, 143, 149, 147, 157, 160, 159, 162, 154, 143.

Вариант 3.

Проведено измерение массы 30 учеников данного класса, в ходе которого были получены следующие данные:

51, 45, 39, 50, 48, 38, 60, 63, 52, 46, 41, 47, 46, 39, 50, 54, 44, 47, 51, 53, 55, 47, 45, 42, 40, 49, 43, 44, 53, 58.

Вариант 4.

В выпускных классах средней школы подсчитали средний балл успеваемости каждого из 30 учеников. В результате получили следующие данные:

3,6; 3,4; 3,9; 3,6; 3,8; 3,7; 4,0; 4,2; 4,9; 4,6; 3,3; 4,7; 3,9; 4,3; 3,8; 5,0; 3,2; 4,5; 4,8; 4,5; 4,9; 3,8; 3,1; 3,7; 4,1; 3,6; 4,4; 4,5; 4,6; 4,1.

Вариант 5.

У 29 юношей-студентов физического факультета ВУЗа был измерен уровень интеллектуального развития. В результате были получены следующие данные:

132, 134, 124, 132, 135, 132, 131, 132, 121, 127, 136, 129, 136, 136, 127, 133, 126, 131, 138, 120, 128, 133, 134, 129, 137, 132, 135, 124, 129

Вариант 6.

Проведено измерение роста 30 студентов данной группы, в ходе которого были получены следующие данные (в см.):

161, 166, 183, 181, 190, 155, 170, 169, 173, 158, 192, 183, 176, 159, 161, 167, 163, 172, 176, 178, 159, 160, 163, 178, 175, 182, 161, 164, 171, 180

Самостоятельная работа

Решение заданий по теме «Проверка статистических гипотез»

Цель: *уметь осуществлять проверку статистических гипотез.*

Методические рекомендации

Непараметрический критерий Q Розенбаума

Q-критерий Розенбаума основан на сравнении «наложенных» друг на друга ранжированных рядов значений двух независимых переменных. При этом не анализируется характер распределения признака внутри каждого ряда – в

16	7	7	7	7	6	6										
17	7	7	7	7	7	7	7									
18	7	7	7	7	7	7	7	7								
19	7	7	7	7	7	7	7	7	7							
20	7	7	7	7	7	7	7	7	7	7						
21	8	7	7	7	7	7	7	7	7	7	7					
22	8	7	7	7	7	7	7	7	7	7	7	7				
23	8	8	7	7	7	7	7	7	7	7	7	7	7			
24	8	8	8	8	8	8	8	8	8	7	7	7	7	7		
25	8	8	8	8	8	8	8	8	8	8	7	7	7	7	7	
26	8	8	8	8	8	8	8	8	8	8	7	7	7	7	7	7
Уровень значимости 0,99																
<i>n</i>	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
11	9															
12	9	9														
13	9	9	9													
14	9	9	9	9												
15	9	9	9	9	9											
16	9	9	9	9	9	9										
17	10	9	9	9	9	9	9									
18	10	10	9	9	9	9	9	9								
19	10	10	10	9	9	9	9	9	9							
20	10	10	10	10	9	9	9	9	9	9						
21	11	10	10	10	9	9	9	9	9	9	9					
22	11	11	10	10	10	9	9	9	9	9	9	9				
23	11	11	10	10	10	10	9	9	9	9	9	9	9			
24	12	11	11	10	10	10	10	9	9	9	9	9	9	9		
25	12	11	11	10	10	10	10	10	9	9	9	9	9	9	9	
26	12	12	11	11	10	10	10	10	10	9	9	9	9	9	9	9

U-критерий Манна – Уитни

В отличие от критерия Розенбаума, U-критерий Манна – Уитни основан на определении зоны перекрытия между двумя ранжированными рядами, т. е. чем меньше зона перекрытия, тем достовернее различия между выборками. Для этого используется специальная процедура преобразования интервальных шкал в ранговые.

Рассмотрим алгоритм вычислений по U-критерию на примере предыдущей задачи.

Для более экономичной работы рекомендуется построение рабочей таблицы следующего вида (табл. 7.2).

Таблица 7.2

<i>x, y</i>	R_{xy}	R_{xy}^*	R_x	R_y
1	2	3	4	5
26	1	1		1
28	2	2,5	2,5	
28	3	2,5		2,5
30	4	4	4	

32	5	5,5		5,5
32	6	5,5		5,5
33	7	7		7
34	8	9	9	
34	9	9	9	
34	10	9		9
35	11	11,5	11,5	
35	12	11,5		11,5
36	13	13	13	
37	14	14	14	
38	15	15		15
39	16	16,5	16,5	
39	17	16,5		16,5
40	18	18,5	18,5	
40	19	18,5		18,5
41	20	20,5	20,5	
41	21	20,5		20,5
42	22	23	23	
42	23	23	23	
42	24	23		23
43	25	25,5	25,5	
43	26	25,5		25,5
44	27	27,5	27,5	
44	28	27,5		27,5
45	29	29	29	
46	30	30	30	
		Σ	276,5	188,5

Рекомендуется следующий порядок заполнения таблицы и соответствующих вычислений:

1. Из двух независимых выборок строим единый ранжированный ряд. В данном случае значения для обеих выборок идут «вперемешку», столбец 1 (x, y). В целях упрощения дальнейшей работы (в том числе и в компьютерном варианте) следует значения для разных выборок отмечать разным шрифтом (или разным цветом) с учетом того, что в дальнейшем мы будем их разносить по разным столбцам.
2. Преобразуем интервальную шкалу значений в порядковую (для этого переобозначаем все значения ранговыми числами от 1 до 30, столбец 2 (R_{xy})).
3. Вводим поправки на связанные ранги (одинаковые значения переменной обозначаются одним и тем же рангом при условии, что сумма рангов не изменяется, столбец 3 (R_{xy}^*). На этом этапе рекомендуется подсчитать суммы рангов во 2-м и 3-м столбце (если все поправки введены верно, то эти суммы должны быть равны).
4. Разносим ранговые числа в соответствии с их принадлежностью к той или иной выборке (столбцы 4 и 5 (R_x и R_y)).
5. Проводим вычисления по формуле:

$$U = n_x n_y + \frac{n_x(n_x + 1)}{2} - T_x, \quad (7.1)$$

где T_x – наибольшая из ранговых сумм; n_x и n_y , соответственно, объемы выборок. В данном случае следует иметь в виду, что если $T_x < T_y$, то обозначения x и y следует сменить на обратные.

6. Сравниваем полученное значение с табличным (см. Приложения, табл. IX). Вывод о достоверности различий между двумя выборками делается в случае, если $U_{\text{эсп.}} < U_{\text{кр.}}$.

В нашем примере $U = 14 \cdot 16 + \frac{16 \cdot 17}{2} - 276,5 = 83,5$. $U_{\text{эсп.}} = 83,5 > U_{\text{кр.}} = 71$.

Вывод

Различия между двумя выборками по критерию Манна – Уитни не являются статистически достоверными.

Примечания

1. Критерий Манна-Уитни не имеет практически никаких ограничений; минимальные объемы сравниваемых выборок – 2 и 5 человек (см. табл. IX Приложения).
2. Аналогично критерию Розенбаума критерий Манна-Уитни может быть использован применительно к любым выборкам независимо от характера распределения.

Критические значения критерия U Манна-Уитни для уровня значимости 0,95

n	2	3	4	5	6	7	8	9	10	11	12	13	14	15
5	0	1	2	4										
6	0	2	3	5	7									
7	0	2	4	6	8	11								
8	1	3	5	8	10	13	15							
9	1	4	6	9	12	15	18	21						
10	1	4	7	11	14	17	20	24	27					
11	1	5	8	12	16	19	23	27	31	34				
12	2	5	9	13	17	21	26	30	34	38	42			
13	2	6	10	15	19	24	28	33	37	42	47	51		
14	3	7	11	16	21	26	31	36	41	46	51	56	61	
15	3	7	12	18	23	28	33	39	44	50	55	61	66	72
16	3	8	14	19	25	30	36	42	48	54	60	65	71	77
17	3	9	15	20	26	33	39	45	51	57	64	70	77	83
18	4	9	16	22	28	35	41	48	55	61	68	75	82	88
19	4	10	17	23	30	37	44	51	58	65	72	80	87	94
20	4	11	18	25	32	39	47	54	62	69	77	84	92	100
n	16	17	18	19	20	21	22	23	24	25	26	27	28	29
4	14	15	16	17	18	19	20	21	22	23	24	25	26	27
5	19	20	22	23	25	26	28	29	31	32	33	35	36	38
6	25	26	28	30	32	34	36	37	39	41	43	45	47	48
7	30	33	35	37	39	41	44	46	48	50	53	55	57	59
8	36	39	41	44	47	49	52	55	57	60	62	65	68	70
9	42	45	48	51	54	57	60	63	66	69	72	75	79	82

10	48	51	55	58	62	65	69	72	75	79	82	86	89	93
11	54	57	61	65	69	73	77	81	85	89	93	96	100	104
12	60	64	68	72	77	81	85	90	94	98	103	107	111	116
13	65	70	75	80	84	89	94	99	103	108	113	118	122	127
14	71	77	82	87	92	97	102	107	113	118	123	128	133	139
15	77	83	88	94	100	105	111	116	122	128	133	139	144	150
16	83	89	95	101	107	113	119	125	131	137	143	150	156	162
17	89	96	102	109	115	121	128	134	141	147	154	160	167	173
18	95	102	109	116	123	130	136	143	150	157	164	171	178	185
19	101	109	116	123	130	138	145	152	160	167	174	182	189	196
20	107	115	123	130	138	146	154	161	169	177	185	193	200	208
21	113	121	130	138	146	154	162	170	179	187	195	203	212	220
22	119	128	136	145	154	162	171	180	188	197	206	214	223	232

3.

Критерий Стьюдента

В отличие от критериев Розенбаума и Манна-Уитни критерий t Стьюдента является параметрическим, т. е. основан на определении основных статистических показателей – средних значений в каждой выборке (\bar{x} и \bar{y}) и их дисперсий (σ_x^2 и σ_y^2), рассчитываемых по стандартным формулам (см. раздел 5).

Использование критерия Стьюдента предполагает соблюдение следующих условий:

1. Распределения значений для обеих выборок должны соответствовать закону нормального распределения (см. раздел 6).
2. Суммарный объем выборок должен быть не менее 30 (для $\beta_1 = 0,95$) и не менее 100 (для $\beta_2 = 0,99$).
3. Объемы двух выборок не должны существенно отличаться друг от друга (не более чем в $1,5 \div 2$ раза).

Идея критерия Стьюдента достаточно проста. Предположим, что значения переменных в каждой из выборок распределяются по нормальному закону, т. е. мы имеем дело с двумя нормальными распределениями, отличающимися друг от друга по средним значениям и дисперсии (соответственно \bar{x} и σ_x^2 , \bar{y} и σ_y^2 , см. рис. 7.1).

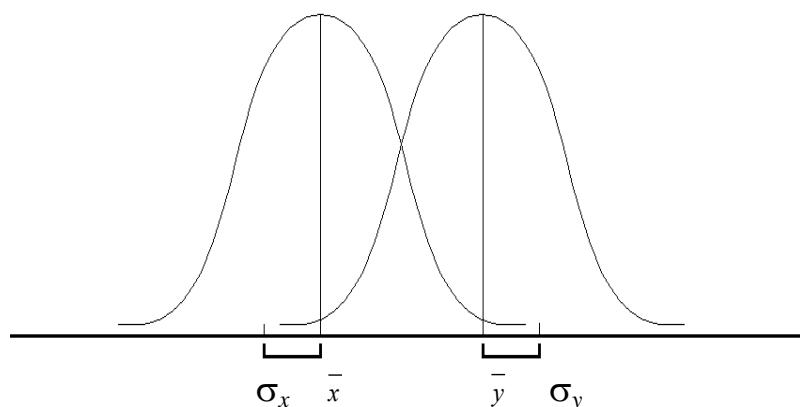


Рис. 7.1. Оценка различий между двумя независимыми выборками: \bar{x} и \bar{y} – средние значения выборок x и y ; σ_x и σ_y – стандартные отклонения

Нетрудно понять, что различия между двумя выборками будут тем больше, чем больше разность между средними значениями и чем меньше их дисперсии (или стандартные отклонения).

В случае независимых выборок коэффициент Стьюдента определяют по формуле:

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}}, \quad (7.2)$$

где n_x и n_y – соответственно численность выборок x и y .

После вычисления коэффициента Стьюдента в таблице стандартных (критических) значений t (см. Приложение, табл. X) находят величину, соответствующую числу степеней свободы $\nu = n_x + n_y - 2$, и сравнивают ее с рассчитанной по формуле. Если $t_{\text{эксп.}} \leq t_{\text{кр.}}$, то гипотезу о достоверности различий между выборками отвергают, если же $t_{\text{эксп.}} > t_{\text{кр.}}$, то ее принимают. Другими словами, выборки достоверно отличаются друг от друга, если вычисленный по формуле коэффициент Стьюдента больше табличного значения для соответствующего уровня значимости.

В рассмотренной нами ранее задаче вычисление средних значений и дисперсий дает следующие значения: $x_{\text{ср.}} = 38,5$; $\sigma_x^2 = 28,40$; $y_{\text{ср.}} = 36,2$; $\sigma_y^2 = 31,72$.

Можно видеть, что среднее значение тревожности в группе девушек выше, чем в группе юношей. Тем не менее эти различия настолько незначительны, что вряд ли они являются статистически значимыми. Разброс значений y юношей, напротив, несколько выше, чем у девушек, но различия между дисперсиями также невелики.

Подставляем значения в формулу: $t = \frac{38,5 - 36,2}{\sqrt{\frac{28,40}{16} + \frac{31,72}{14}}} = 1,14$.

Вывод

$t_{\text{эксп.}} = 1,14 < t_{\text{кр.}} = 2,05$ ($\beta_1 = 0,95$). Различия между двумя сравниваемыми выборками не являются статистически достоверными. Данный вывод вполне согласуется с таковым, полученным при использовании критериев Розенбаума и Манна-Уитни.

Другой способ определения различий между двумя выборками по критерию Стьюдента состоит в вычислении доверительного интервала стандартных отклонений. Доверительным интервалом называется среднеквадратичное (стандартное) отклонение, деленное на корень квадратный из объема выборки и умноженное на стандартное значение коэффициента Стьюдента для $n - 1$ степеней свободы (соответственно, $\frac{t_{n-1} \cdot \sigma_x}{\sqrt{n_x}}$ и $\frac{t_{n-1} \cdot \sigma_y}{\sqrt{n_y}}$).

Стандартные значения критерия Стьюдента

Число	Уровни значимости	Число	Уровни значимости
-------	-------------------	-------	-------------------

степене й свобод ы	0,95	0,99	0,999	степене й свобод ы	0,95	0,99	0,999
1	12,71	63,66	636,62	19	2,09	2,86	3,88
2	4,30	9,92	31,60	20	2,09	2,85	3,85
3	3,18	5,84	12,94	21	2,08	2,83	3,82
4	2,78	4,60	8,61	22	2,07	2,82	3,79
5	2,57	4,03	6,86	23	2,07	2,81	3,77
6	2,45	3,71	5,96	24	2,06	2,80	3,74
7	2,36	3,50	5,40	25	2,06	2,79	3,72
8	2,31	3,36	5,04	26	2,06	2,78	3,71
9	2,26	3,25	4,78	27	2,05	2,77	3,69
10	2,23	3,17	4,59	28	2,05	2,76	3,66
11	2,20	3,11	4,49	29	2,05	2,76	3,66
12	2,18	3,05	4,32	30	2,05	2,76	3,66
13	2,16	3,01	4,22	35	2,04	2,75	3,65
14	2,14	2,98	4,14	40	2,02	2,70	3,55
15	2,13	2,95	4,07	60	2,00	2,66	3,46
16	2,12	2,92	4,01	120	1,98	2,62	3,37
17	2,11	2,90	3,96	∞	1,96	2,58	3,29
18	2,10	2,88	3,92				

Критерий Фишера

Критерий Фишера (F) основан на том же принципе, что и критерий Стьюдента, т. е. предполагает вычисление средних значений и дисперсий в сравниваемых выборках. Чаще всего используется при сравнении между собой неравноценных по объему (разных по численности) выборок. Критерий Фишера является несколько более жестким, чем критерий Стьюдента, а потому более предпочтителен в тех случаях, когда возникают сомнения в достоверности различий (например, если по критерию Стьюдента различия достоверны при нулевом и недостоверны при первом уровне значимости).

Формула Фишера выглядит следующим образом:

$$F = \frac{n_x \cdot n_y}{n_x + n_y} \cdot \frac{d^2}{\sigma_z^2},$$

где $d^2 = (\bar{x} - \bar{y})^2$ и $\sigma_z^2 = \frac{\sigma_x^2(n_x - 1) + \sigma_y^2(n_y - 1)}{n_x + n_y - 2}$.

В рассматриваемой нами задаче $d^2 = 5,29$; $\sigma_z^2 = 29,94$.

Подставляем значения в формулу: $F = \frac{16 \cdot 14}{16 + 14} \cdot \frac{5,29}{29,94} = 1,32$.

В табл. находим, что для уровня значимости $\beta_1 = 0,95$ и $v = n_x + n_y - 2 = 28$ критическое значение составляет 4,20.

Вывод

$F = 1,32 < F_{кр.} = 4,20$. Различия между выборками статистически недостоверны.

Стандартные значения критерия Фишера, используемые для оценки достоверности различий между двумя выборками

Степени свободы (v)	Уровень значимости			Степени свободы (v)	Уровень значимости		
	0,95	0,99	0,999		0,95	0,99	0,999
3	10,13	34,12	167,5	28	4,20	7,64	13,5
4	7,71	21,20	74,1	29	4,18	7,60	13,4
5	6,61	16,26	47,0	30	4,17	7,56	13,3
6	5,99	13,74	35,5	32	4,15	7,50	13,2
7	5,59	12,25	29,2	34	4,13	7,44	13,1
8	5,32	11,26	25,4	36	4,11	7,39	13,0
9	5,12	10,56	22,9	38	4,10	7,35	12,9
10	4,96	10,04	21,0	40	4,08	7,31	12,8
11	4,84	9,65	19,7	42	4,07	7,27	12,7
12	4,75	9,33	18,6	44	4,06	7,24	12,5
13	4,67	9,07	17,8	46	4,05	7,21	12,4
14	4,60	8,86	17,1	48	4,04	7,19	12,3
15	4,54	8,68	16,6	50	4,03	7,17	12,2
16	4,49	8,53	16,1	55	4,02	7,12	12,1
17	4,45	8,40	15,7	60	4,00	7,08	12,0
18	4,41	8,28	15,4	65	3,99	7,04	11,9
19	4,38	8,18	15,1	70	3,98	7,01	11,6
20	4,35	8,10	14,8	80	3,96	6,96	11,6
21	4,32	8,02	14,6	100	3,94	6,90	11,5
22	4,30	7,94	14,4	125	3,92	6,84	11,4
23	4,28	7,88	14,2	150	3,91	6,81	11,3
24	4,26	7,82	14,0	200	3,89	6,76	11,2
25	4,24	7,77	13,9	400	3,86	6,70	11,0
26	4,22	7,72	13,7	1000	3,85	6,66	10,9
27	4,21	7,68	13,6	∞	3,84	6,64	10,8

Используя методические рекомендации, выполните задания:

По предложенным данным задачи проверить значимы ли различия между данными несвязанными выборками: по критерию Розенбаума, Манна-Уитни, Стьюдента и Фишера.

Вариант 1.

В двух выборках приведены данные об уровне холестерина в крови для людей двух возрастных групп: 20-40 лет (X_i) и 50-70 лет (Y_i). Проверить значимы ли различия между этими группами людей по данной характеристике.

X : 153; 289; 231; 269; 244; 293; 276; 248; 240; 253; 267; 159; 249; 267; 246; 239; 235; 253; 175; 157.

Y : 201; 229; 191; 229; 210; 201; 290; 203; 389; 304; 294; 274; 261; 361; 283; 258; 243; 360; 248; 313.

Вариант 2.

В двух выборках приведены данные о кровяном давлении для людей двух возрастных групп: 20-40 лет (X) и 50-70 лет (Y). Проверить значимы ли различия между этими группами людей по данной характеристике.

X: 118; 128; 138; 136; 124; 116; 114; 104; 128; 116; 142; 153; 164; 185; 140; 110; 115; 152; 147; 149 .

Y : 138; 132; 150; 182; 148; 158; 136; 126; 160; 156; 147; 129; 158; 164; 139; 163; 157; 173; 162; 137.

Вариант 3.

В двух выборках приведены данные о массе женщин (X_I) и мужчин (Y₀ 20-40 лет. Проверить значимы ли различия в весе мужчин и женщин.

X: 58; 54; 59; 48; 61; 56; 67; 52; 58; 73; 55; 58; 63; 57; 52; 69; 54; 60; 72; 65.

Y: 78; 83; 64; 69; 78; 83; 90; 61; 71; 65; 75; 84; 79; 81; 96; 87; 91; 77; 74; 83.

Вариант 4.

В двух выборках приведены данные о весе женщин (X_I) и мужчин (Y_;) 40-60 лет. Проверить значимы ли различия в весе мужчин и женщин.

X: 60; 56; 64; 68; 53; 75; 62; 59; 67; 56; 74; 63; 72; 80; 58; 59; 54; 59; 48; 61.

Y : 75; 83; 89; 94; 87; 98; 102; 86; 94; 71; 78; 83; 64; 69; 78; 83; 90; 61; 71; 65.

Вариант 5.

В двух выборках приведены данные о весе женщин (X_I) и мужчин (Y₀ 60-80 лет. Проверить значимы ли различия в весе мужчин и женщин.

X: 62; 58; 71; 83; 64; 55; 67; 75; 61; 79; 54; 59; 48; 61; 56; 67; 52; 58; 73; 55.

Y: 73; 69; 83; 75; 94; 88; 91; 87; 76; 89; 78; 83; 64; 69; 78; 83; 90; 61; 71; 65.

История возникновения дисперсионного анализа

Цель: *Развитие интереса к предмету.*

Форма самостоятельной деятельности: подготовить реферат по предложенной теме.

Методические рекомендации

Реферат должен быть выполнен с соблюдением методических рекомендаций по написанию реферата.

Самостоятельная работа

Составление презентации на тему: «Регрессионный анализ»

Цель: *Развитие интереса к предмету, интуиции, логического мышления.*

Методические рекомендации

Презентация должна быть выполнена с соблюдением методических рекомендаций по составлению презентации.