ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«БАШКИРСКИЙ ГОСУДАРСТВЕННЫЙ МЕДИЦИНСКИЙ УНИВЕРСИТЕТ»
МИНИСТЕРСТВА ЗДРАВООХРАНЕНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
(ФГБОУ ВО БГМУ МИНЗДРАВА РОССИИ)

Кафедра медицинской физики и информатики

МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ
по самостоятельной внеаудиторной работе

Дисциплина: медико-биологическая статистика и информационные технологии в здравоохранении

Направление подготовки 34.04.01 Управление сестринской деятельностью

Квалификация Магистр

Магистр

Курс 2 Семестр 4

Уфа

Рецензенты:

Заведующий кафедрой управления сестринской деятельностью ФГАОУ ВО «Российский университет дружбы народов Имени Патриса Лумумбы», д.м.н., профессор, И.В. Радыш.

Работодатель:

Президент Региональной общественной организации «Профессиональной ассоциации специалистов с высшим сестринским, средним медицинским и фармацевтическим образованием Республики Башкортостан» Э.Ю. Ахметшина.

Автор: доцент Войтик В.В.

Утверждена на заседании № 8 кафедры медицинской физики и информатики, от «16» апреля 2024 г.

Темы:

1. Основы понятия медико-биологической статистики
2. Статистическая проверка гипотез
3. Корреляционный и регрессивный анализ зависимостей между случайными величинами
4. Анализ временных рядов.
5. Дисперсионный анализ
6. Статистические методы обработки результатов экспериментальных измерений
7. Применение пакета Statistica для анализа зависимостей.

Тема № 1: Основы понятия медико-биологической статистики (14ч)

Цель изучения темы: познакомить обучающихся с процессом обработки результатов измерений. Дать представление о генеральной и выборочной совокупности. Отработать практические навыки по определению доверительных пределов и исчислению ошибок выборки.

Задачи: Создание методов сбора и группировки обрабатываемого статистического материала, полученного в результате наблюдений за случайными процессами. Разработка методов анализа полученных статистических данных. Получение выводов по данным наблюдений.

Обучающийся должен знать:
- задачи математической статистики, алгоритм обработки статистических данных;
- формы представления статистических данных;
- понятия: случайная величина, генеральная совокупность, выборка, объем и размах выборки, статистический ряд, полигон, гистограмма
-виды погрешностей измерений
-нормальный закон распределения
-доверительная вероятность и доверительный интервал;

должен уметь:
- представлять статистические данные в форме вариационного и статистического рядов;
- графически представлять результаты обработки статистических данных;
- строить гистограммы и полигоны частот
- давать ответы на тестовые задания,

должен сформировать компетенции: УК-1, ОПК-4.

Задания для самостоятельной контактной/внеконтактной работы обучающихся по указанной теме:

1. Ознакомиться с теоретическим материалом по теме.
2. Ответить на вопросы.

1. Каковы основные задачи математической статистики?
2. Что называется генеральной и выборочной совокупностями для исследуемой случайной величины?
3. В чем сущность выборочного метода?
4. Как получают повторную и бесповторную выборки?
5. Какая выборка называется репрезентативной, однородной?
6. В чем заключается первичная обработка статистического материала?
7. Что такое частота появления варианты в выборке?
8. Как получают относительную частоту варианты в выборке?
9. Как получают вариационный ряд распределения?
10. Что такое группированный статистический ряд?
11. Как построить по данной выборке дискретный и интервальный сгруппированные статистические ряды?
12. Что такое полигон частот?
13. Как построить многоугольник распределения относительных частот?
14. Как построить гистограмму распределения плотностей относительных частот?
15. Дайте определение моды и медианы выборки.

3. Проверить свои знания с использованием тестового контроля.

1. Дайте определение следующих терминов: a) измерение, единство измерений; b) физическая величина, единица измерения физической величины; c) погрешность, абсолютная погрешность, относительная погрешность, приведённая погрешность.
2. Виды измерения: прямые, косвенные, совместные и совокупные. Дайте определения.
3. Поясните на примере, как определяется абсолютная, относительная и приведённая погрешность средства измерения.
4. Что называется «нормирующим значением»? Поясните на примере, как находится нормирующее значение, в случае если шкала средства измерения содержит нулевую отметку, не содержит нулевую отметку.
5. Чем отличаются понятия: «погрешность средства измерения» и «погрешность результата»?
6. Что такое инструментальные и методические погрешности?
7. Что такое основная и дополнительная погрешности? Чем нормальные условия отличаются от рабочих?
8. Что такое систематические и случайные погрешности?
9. Что такое промахи? Перечислите и опишите критерии для исключения промахов.

Формы контроля освоения заданий по самостоятельной контактной/внеконтактной работы обучающихся по данной теме: тестовые задания и контрольные вопросы.

Рекомендуемая литература:
См. в приложении.

Тема№2: Статистическая проверка гипотез (14ч)

Цель изучения темы: получить навыки выявления погрешностей в результатах наблюдений, статистической обработки результатов наблюдений отдельных групп, определения средневзвешенных статистических характеристик групп неравноточных наблюдений; представления результатов измерений; оценки формы и вида законов экспериментальных распределений физических величин; записи результатов измерений.
Задачи: Создание методов сбора и группировки обрабатываемого статистического материала, полученного в результате наблюдений за случайными процессами. Разработка методов анализа полученных статистических данных. Получение выводов по данным наблюдений.

Обучающийся должен знать:
- задачи математической статистики, алгоритм обработки статистических данных;
- формы представления статистических данных;
- понятия: случайная величина, генеральная совокупность, выборка, объем и размах выборки, статистический ряд, полигон, гистограмма
-виды погрешностей измерений
-нормальный закон распределения
-доверительная вероятность и доверительный интервал;
должен уметь:
- представлять статистические данные в форме вариационного и статистического рядов;
- графически представлять результаты обработки статистических данных;
- строить гистограммы и полигоны частот
- давать ответы на тестовые задания,
должен сформировать компетенции: УК-1, ОПК-4.

Задания для самостоятельной контактной/внеконтактной работы обучающихся по указанной теме:

1. Ознакомиться с теоретическим материалом по теме.
2. Ответить на вопросы.

1. Какие измерения называют прямыми и какие – косвенными?
2. Что представляют собой абсолютная и относительная погрешности?
3. Какая погрешность показывает точность измерений?
4. Чем обусловлено появление систематических погрешностей?
5. Какие погрешности называют случайными? Укажите их источники.
6. Что такое промах? Как можно его обнаружить?
7. Для чего измерение проводят несколько раз?
8. Что означает класс точности прибора?
9. Как определяют систематическую погрешность в прямых измерениях?
10. Почему стараются вести измерения в правой части шкалы прибора?
11. Сколько значащих цифр указывают в погрешности и в результате?
12. Запишите результат измерений в виде доверительного интервала.
13. Для чего в заголовок графы таблицы выносят общий множитель и единицу величины?
14. Какой интервал называют доверительным?
15. Что такое доверительная вероятность измерений?
16. По какой формуле рассчитывают среднее квадратическое отклонение случайной величины?
17. Чему равна величина систематической составляющей СКО?
18. В каких случаях рост числа измерений не приводит к увеличению точности? Чем обусловлена погрешность в этих случаях?
19. Для чего используют коэффициент Стьюдента? Чем определяется его значение?
20. Как рассчитывают доверительный интервал при прямых измерениях?
21. Каким образом находят относительную погрешность результата косвенных измерений?
22. По какой формуле вычисляют ширину доверительного интервала искомой величины в косвенных измерениях? оля.
23. Чем определяется доверительная вероятность для такого интервала?
24. Почему при вычислении погрешности в косвенных измерениях можно отбросить те из погрешностей прямых измерений δ_{x_i}, которые не превышают 1/3 (или даже половину) от максимальной из них?

1. Дайте определение следующих терминов: a) измерение, единство измерений; b) физическая величина, единица измерения физической величины; c) погрешность, абсолютная погрешность, относительная погрешность, приведённая погрешность.
2. Виды измерения: прямые, косвенные, совместные и совокупные. Дайте определения.
3. Поясните на примере, как определяется абсолютная, относительная и приведённая погрешность средства измерения.
4. Что называется «нормирующим значением»? Поясните на примере, как находится нормирующее значение, в случае если шкала средства измерения содержит нулевую отметку, не содержит нулевую отметку.
5. Чем отличаются понятия: «погрешность средства измерения» и «погрешность результата»?
6. Что такое инструментальные и методические погрешности?
7. Что такое основная и дополнительная погрешности? Чем нормальные условия отличаются от рабочих?
8. Что такое систематические и случайные погрешности?
9. Что такое промахи? Перечислите и опишите критерии для исключения промахов.

1. Точечная оценка параметра распределения признака, вычисленная по выборке, характеризуется: а) одним числом; б) средним значением признака; в) точкой на прямой; г) результатами выборки.
2. Оценкой параметра называется: а) приближенное случайное значение параметра генеральной совокупности, которое определяется по всем данным генеральной совокупности; б) приближенное случайное значение параметра генеральной совокупности, которое определяется по данным выборки; в) приближенное неслучайное значение параметра генеральной совокупности, которое определяется по данным выборки.
3. Оценка называется несмещенной, если: а) она сходится по вероятности при $n\to\infty$ к истинному значению параметра; б) она обладает по сравнению с другими наименьшей дисперсией; в) ее математическое ожидание равно истинному значению параметра.
4. Оценка называется состоятельной, если: а) она обладает по сравнению с другими наименьшей дисперсией; б) ее математическое ожидание равно истинному значению параметра; в) она сходится по вероятности при $n\to\infty$ к истинному значению параметра.
5. Оценка называется эффективной, если: а) она обладает по сравнению с другими оценками наименьшей дисперсией; б) ее математическое ожидание равно истинному значению параметра; в) она сходится по вероятности при $n\to\infty$ к истинному значению параметра.
6. Отметьте неправильный ответ. Качество точечной оценки параметра распределения признака характеризуется: а) несмещенностью; б) эффективностью; в) состоятельностью; г) случайностью.
7. Точечная оценка – это: а) оценка параметра генеральной совокупности интервалом, в который этот параметр с заданной вероятностью попадет; б) оценка параметра генеральной совокупности параметром, рассчитанным на основе выборки; в) расчет вероятности попадания точки в заданный интервал; г) расчет вероятности некоторого события.
8. Интервальная оценка – это: а) оценка параметра генеральной совокупности параметром, рассчитанным на основе выборки; б) нахождение интервала, в который попадает наудачу брошенная точка; в) оценка интервала вероятностей, с которыми может происходить некоторое событие; г) оценка параметра генеральной совокупности интервалом, в который этот параметр с заданной вероятностью попадет.
9. Точечная оценка, математическое ожидание которой равно оцениваемому параметру при любом объеме выборки, называется: а) смещенной; б) несмещенной; в) состоятельной; г) эффективной; д) несостоятельной. 10. Точечная оценка, которая имеет

наименьшую дисперсию среди всех возможных несмещенных оценок того же параметра, называется: а) эффективной; б) неэффективной; в) состоятельной; г) несостоятельной; д) центральной.
11. Точечная оценка, математическое ожидание которой не равно оцениваемому параметру при любом объеме выборки, называется: а) смещенной; б) несмещенной; в) состоятельной; г) эффективной; д) несостоятельной.
12. Математическое ожидание оценки параметра равно: а) параметру; б) выборочному среднему значению; в) выборочной дисперсии; г) нулю.
13. Несмещенная и состоятельная оценка генеральной дисперсии: а) выборочная дисперсия; б) размах признака; в) исправленная выборочная дисперсия; г) приближенное значение дисперсии.
14. При увеличении объема выборки точность оценки: а) уменьшается; б) увеличивается; в) не изменяется; г) может уменьшаться, а может и увеличиваться.
15. При увеличении надежности оценки ее точность: а) уменьшается; б) увеличивается; в) не изменяется; г) может уменьшаться, а может и увеличиваться.
16. В результате измерений некоторой физической величины одним прибором (без систематических ошибок) получены следующие результаты (в мм): 11, 13, 15. Тогда оценка дисперсии измерений равна:
а) 4; б) 13; в) 8; г) 3.
17. Несмещенная оценка математического ожидания признака:

а) $\bar{x} = x_1 + x_2 + \ldots + x_n$; б) $\bar{x}_в = \frac{1}{n}\sum_{i=1}^{n} x_i$;

в) $\bar{x}_в = \frac{1}{n-1}\sum_{i=1}^{n} x_i$; г) $\bar{x}_в = \frac{n}{n-1}\sum_{i=1}^{n} x_i$.

18. Оценкой генеральной средней признака является: а) выборочное среднее значение; б) среднее значение признака; в) наибольшее значение признака; г) математическое ожидание.
19. Какие из точечных оценок являются смещенными оценками: а) выборочное среднее; б) исправленная выборочная дисперсия; в) выборочная дисперсия; г) асимметрия; д) исправленное среднее квадратичное отклонение;
20. Среднее значение выборки является: а) несмещенной оценкой математического ожидания; б) смещенной оценкой математического ожидания; в) смещенной оценкой дисперсии; г) несмещенной оценкой дисперсии.
21. Выборочная дисперсия, определяемая по формуле

$$D_в = \frac{\sum_{1}^{n}(x_i - \bar{x}_в)^2}{n}$$

является: а) несмещенной оценкой дисперсии генеральной совокупности; б) смещенной оценкой дисперсии генеральной совокупности; в) либо смещенной, либо несмещенной оценкой (в зависимости от условий проведения опыта) дисперсии генеральной совокупности.
22. Чтобы оценка дисперсии генеральной совокупности была несмещенной, необходимо выборочную дисперсию:

а) умножить на $\frac{n}{n-1}$; б) умножить на $\frac{n-1}{n}$;

в) разделить на $n-1$.

23. Несмещенная оценка дисперсии признака:

а) $S^2 = \frac{n}{n-1}\sum_{i=1}^{n}(x_i - \bar{x}_в)^2$; б) $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x}_в)^2$;

в) $S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_в)^2$; г) $S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_в)$.

24. Проведено четыре измерения (без систематических ошибок) некоторой случайной величины (в мм): 5, 6, 9, 12. Оценка математического ожидания равна: а) 8,25; б) 8,5 ; в) 7; г) 8.
25. Символ γ в формуле доверительного интервала означает: а) оценка параметра; б) доверительный интервал; в) объем выборки; г) доверительная вероятность.
26. Длина доверительного интервала уменьшается с увеличением: а) выборочных значений; б) объема выборки; в) доверительной вероятности; г) выборочного среднего.
27. Длина доверительного интервала с увеличением объема выборки: а) уменьшается; б) увеличивается; в) не изменяется; г) колеблется.
28. Длина доверительного интервала с увеличением доверительной вероятности: а) изменяется; б) уменьшается; в) увеличивается; г) постоянна.
29. Точечная оценка математического ожидания нормального распределения равна 12. Тогда его интервальная оценка может иметь вид:
а) (11,4; 12); б) (12; 12,6); в) (11,4; 12,6); г) (11,4; 11,5).
30. Дана интервальная оценка (8,4; 9,2) математического ожидания нормально распределенного количественного признака. Найти точечную оценку математического ожидания.
а) 8,8; б) 8,6; в) 9,0; г) 8,75.
31. Доверительный интервал для математического ожидания при известной дисперсии σ^2 нормально распределенной генеральной совокупности будет:

а) $\bar{x}_в - x_\gamma \dfrac{\sqrt{n}}{\sigma} < a < \bar{x} + x_\gamma \dfrac{\sqrt{n}}{\sigma}$, где $\Phi(x_\gamma) = \dfrac{\gamma}{2}$;

б) $\bar{x}_в - x_\gamma \dfrac{\sigma}{\sqrt{n}} < a < \bar{x}_в + x_\gamma \dfrac{\sigma}{\sqrt{n}}$, где $\Phi(x_\gamma) = \dfrac{\gamma}{2}$;

в) $\bar{x}_в - x_\gamma \dfrac{\sigma}{\sqrt{n-1}} < a < \bar{x}_в + x_\gamma \dfrac{\sigma}{\sqrt{n-1}}$, где $\Phi(x_\gamma) = \dfrac{\gamma}{2}$.

32. Доверительный интервал для математического ожидания при неизвестной генеральной дисперсии имеет вид:

а) $\bar{x}_в - t(\alpha, n-1)\dfrac{S}{\sqrt{n}} < a < \bar{x}_в + t(\alpha, n-1)\dfrac{S}{\sqrt{n}}$;

б) $S\gamma_1 < \sigma < S\gamma_2$;

в) $\bar{x}_в - t(\alpha, n)\dfrac{S}{\sqrt{n}} < a < \bar{x}_в + t(\alpha, n)\dfrac{S}{\sqrt{n}}$;

г) $\bar{x}_в - t(\alpha, n-1)\dfrac{\sigma}{\sqrt{n}} < a < \bar{x}_в + t(\alpha, n-1)\dfrac{\sigma}{\sqrt{n}}$.

33. Доверительный интервал $(V_в - \delta, V_в + \delta)$ для параметра V определяется:

а) по заданному значению δ и значению $V_в$, которое находится из соотношения $P(|V_в - V| < \delta) = \gamma$;

б) по определенному из выборки $V_в$ и значению δ, которое находится из соотношения $P(|V_в - V| < \delta) = \gamma$;

в) по заданной доверительной вероятности γ и по ее выборочным данным δ и $V_в$.

34. Дана интервальная оценка (5,1; 6,7) математического ожидания нормально распределенного количественного признака. Найти точность этой оценки. а) 0,8; б) 1,6; в) 5,9; г) 0,85.

Формы контроля освоения заданий по самостоятельной контактной/внеконтактной работы обучающихся по данной теме: тестовые задания и контрольные вопросы.
Контрольные вопросы по теме

1. Дайте определение точечной статистической оценки.

2. Какая оценка параметра распределения называется точечной?
3. Какими свойствами обладает выборочное среднее x?
4. Какими свойствами обладает выборочная дисперсия S^2?
5. Какая числовая характеристика выборки является несмещенной для математического ожидания?
6. Какая числовая характеристика выборки является несмещенной для дисперсии?
7. Что понимается под термином «интервальная оценка параметра распределения»?
8. Дайте определение доверительного интервала.
9. Что такое точность оценки и надежность оценки?
10. Что называется доверительной вероятностью? Какие значения она принимает?
11. Как изменится длина доверительного интервала, если увеличить: 1) объем выборки, 2) доверительную вероятность? Ответ обоснуйте.
12. Запишите формулу для нахождения доверительного интервала математического ожидания нормально распределенной случайной величины, если генеральная дисперсия: 1) известна; 2) неизвестна.

Рекомендуемая литература:
См. в приложении.

Тема№3: Корреляционный и регрессивный анализ зависимостей между случайными величинами (14 ч)

Цель изучения темы:

Основная цель данной темы сформировать у обучающихся знания и навыки о методах математической, статистической обработки результатов исследований и их анализа. Формирование понятий статистической гипотезы, статистического критерия и навыков проверки гипотезы о нормальном распределении генеральной совокупности.

Задачи: сформировать у обучающихся знания и навыки о методах математической, статистической обработки результатов исследований и их анализа. Формирование понятий статистической гипотезы, статистического критерия и навыков проверки гипотезы о нормальном распределении генеральной совокупности.

Обучающийся должен знать:
методы проверки статистических гипотез о параметрах распределений и согласии с теоретическим распределением;
1. Определение статистической гипотезы;
2. Примеры статистических гипотез.
3. Понятия нулевой и конкурирующей гипотез;
4. Ошибки первого и второго рода. Примеры.
5. Определение статистического критерия, критической области и области принятия решений.
6. Критерий согласия Пирсона.
должен уметь:
- проверять гипотезы о значениях и равенстве параметров различных распределений,
- проверять гипотезы о согласии эмпирического распределения с теоретическим;
- проверять влияние изучаемых факторов любой природы на исследуемую переменную
должен владеть: Навыками формулировки и проверки статистических гипотез, соответствующих данным изучаемой задачи;

должен сформировать компетенции: УК-1, ОПК-4.

Задания для самостоятельной контактной/внеконтактной работы обучающихся по указанной теме:
1.Ознакомиться с теоретическим материалом по теме.
2.Ответить на вопросы.

1. *Что называется статистической гипотезой?*
2. *Сформулируйте задачу статистической проверки гипотезы.*
3. *Приведите примеры задач на проверку гипотез и их математические формулировки.*
4. *Назовите шаги логической схемы проверки статистической гипотезы.*
5. *Поясните смысл понятий "ошибка первого рода", "ошибка второго рода", "мощность критерия".*
6. *Что такое уровень значимости?*
7. *Что такое критерий согласия? Каков смысл статистики критерия?*
8. *Зачем нужно знать закон распределения статистики критерия?*
9. *В чем отличие одностороннего и двухстороннего критериев, простой и сложной гипотез?*
10. *Как зависит ширина области принятия гипотезы от уровня значимости?*
11. *Как определяются критические границы для одностороннего и двухстороннего критерия при заданной величине уровня значимости?*

12. *Приведите примеры практических задач, когда необходимо проверить гипотезу о равенстве математических ожиданий или дисперсий.*
13. *Что общего в методике построения доверительных интервалов и проверки статистических гипотез?*
14. *Приведите примеры критериев согласия и области их применения.*
15. *Какие критерии согласия применимы для непрерывных распределений?*

3. Решить следующие задачи

Пример 1. Установлено, что в прошлом году покупатель за одно посещение магазина в среднем осуществлял покупки на сумму 956 руб. В этом году на основе случайной выборки 76 посетителей было найдено, что средняя цена покупки при одном посещении составила 1021 руб. Предполагается, что темп роста трат покупателя есть случайная величина, распределенная по нормальному закону с σ=427 руб. Можно ли на основе этой информации сделать вывод о том, что за год среднее количество денег, которые тратит за одно посещение магазина покупатель, фактически не изменилось? Принять уровень значимости 5 %.

Решение:

Дисперсия генеральной совокупности известна, используем z-распределение.

$H_0 : a_0 = 956; \quad H_1 : a_0 \neq 956.$

Статистическим критерием для проверки этой гипотезы является нормированная нормально распределенная случайная величина

Наблюдаемое значение критерия $z_{\text{набл}} = \dfrac{(\bar{x} - a_0)\sqrt{n}}{\sigma}$.

Условием задачи заданы следующие экспериментальные значения:

$$\bar{x} = 1021; \ \sigma = 427; \ n = 76.$$

При справедливости нулевой гипотезы поведение этого критерия можно приближенно описать стандартным нормальным законом распределения. Вычислим наблюдаемое значение критерия на основе экспериментальных значений:

$$z_{\text{набл}} = \frac{(\bar{x} - a_0)\sqrt{n}}{S} = \frac{(1021 - 956)\sqrt{76}}{427} = 1{,}33.$$

В соответствии с приведенной выше таблицей критериев критических областей найдем границы двусторонней критической области $\Phi(z_{\text{кр}}) = \dfrac{1-\alpha}{2} = 0{,}475;$ и в соответствии с таблицей прил. 3 $z_{\text{кр}} = 1{,}96.$

$$z_{\text{набл}} < z_{\text{кр}}.$$

Поскольку наблюдаемое значение критерия попало в область естественных значений критерия, то в результате решения задачи следует сделать такой вывод: принимаем с уровнем доверия в 95 % утверждение о том, что средняя сумма денег, которую оставляет в гипермаркете покупатель за одно посещение, значимо за год не изменилась.

Пример 2. Установлено, что в прошлом году покупатель за одно посещение гипермаркета в среднем осуществлял покупки на сумму 956 руб. В этом году на основе случайной выборки 76 посетителей было найдено, что средняя цена покупки при одном посещении магазина составила 1021 руб. со стандартным отклонение для одного покупателя 427 руб. Можно ли на основе этой информации сделать вывод о том, что

за год среднее количество денег, которые тратит за одно посещение магазина покупатель, фактически не изменилось? Принять уровень значимости 5 %. (Полагаем, что сумма покупок меняется по нормальному закону распределения).

Решение:

Дисперсия генеральной совокупности неизвестна, используем t-распределение.

$H_0 : a_0 = 956; \quad H_1 : a_0 \neq 956.$

Статистическим критерием для проверки этой гипотезы является нормированная нормально распределенная случайная величина

$$t = \frac{(\bar{x} - a)\sqrt{n}}{S}.$$

Величина t имеет распределение Стьюдента с $k = n - 1$ степенями свободы.

Наблюдаемое значение критерия $t_{\text{набл}} = \dfrac{(\bar{x} - a_0)\sqrt{n}}{S}$.

Условием задачи заданы следующие экспериментальные значения:

$$\bar{x} = 1021; \; S = 427; \; n = 76.$$

$$t_{\text{набл}} = \frac{(\bar{x} - a)\sqrt{n}}{S} = \frac{(1021 - 956)\sqrt{76}}{427} = 1{,}33.$$

Далее по таблице критических точек распределения Стьюдента, по заданному уровню значимости α, помещенному в верхней строке таблицы, и числу степеней свободы $k = n - 1$ находим критическую точку $t_{\text{кр}}, (\alpha; k) = 1{,}99$.

$$t_{\text{набл}} < t_{\text{кр}}.$$

Поскольку по-прежнему наблюдаемое значение критерия попало в область принятия нулевой гипотезы, то ранее полученный вывод не изменился, т. е. мы принимаем нулевую гипотезу об отсутствии значимых изменений в той сумме денег, которую в среднем тратит покупатель за одно посещение гипермаркета.

Пример 3. По утверждению руководства фирмы, средний размер дебиторского счета равен 187,5 тыс. руб. Ревизор составляет случайную выборку из 10 счетов и обнаруживает, что средняя арифметическая выборки равна 175 тыс. руб. при среднем квадратическом отклонении 35 тыс. руб. Выяснить, может ли оказаться в действительности правильным объявленный размер дебиторского счета. Принять уровень значимости равным 0,05.

Решение:

Имеем: $\bar{x} = 175; \; a_1 = 187{,}5; \; n = 10; \; \sigma = 35; \; \alpha = 0{,}05.$

$H_0 : a_1 = a_2; \quad H_1 : a_1 \neq a_2.$

Так как дисперсия неизвестна, то для проверки гипотезы H_0 воспользуемся распределением Стьюдента.

Тогда наблюдаемое значение критерия

$$t_{\text{набл}} = \frac{(\bar{x} - a_1)\sqrt{n}}{\sigma} = \frac{(175 - 187{,}5)\sqrt{10}}{35} = -1{,}129.$$

По таблице распределения Стьюдента (прил. 4) найдем $t_{\text{двус.кр}}(0{,}05;9) = 2{,}26.$

Так как $|t_{\text{набл}}| < t_{\text{двус.кр}}$, то гипотезу H_0 о среднем размере дебиторского счета можно принять на уровне доверия 0,95.

Гипотеза о равенстве среднего значения числу

Пример 1. Утверждается, что шарики для подшипников, изготовленные автоматическим станком, имеют средний диаметр 10 мм. Используя односторонний критерий с α=0,05, проверить эту гипотезу, если в выборке из n шариков средний диаметр оказался равным 10,3 мм, а дисперсия известна и равна 1 мм.

ЗАДАНИЕ. Утверждается, что шарики для подшипников, изготовленные автоматическим станком, имеют средний диаметр $d_0 = 10$ мм. Используя односторонний критерий с $\alpha = 0{,}05$, проверить эту гипотезу, если в выборке из $n = 16$ шариков средний диаметр оказался равным 10,3 мм, а дисперсия известна и равна $\sigma^2 = 1$ мм^2.

РЕШЕНИЕ.
Нулевая гипотеза: $H_0 : \mu = 10$. Альтернативная гипотеза (односторонняя) $H_1 : \mu > 10$
Вычисляем наблюдаемое значение критерия

$$U_{\text{набл}} = \frac{(\bar{x} - \mu)}{\sigma}\sqrt{n} = \frac{10{,}3 - 10}{1}\sqrt{16} = 1{,}2 \ .$$

По таблице функции Лапласа найдем критическую точку для односторонней критической области (при гипотезе $H_1 : \mu > 10$) по уровню значимости $\alpha = 0{,}05$:

$\Phi(u_{кр}) = \dfrac{1 - 2\alpha}{2} = 0{,}45$, откуда $u_{кр} = 1{,}645$.

Так как $U_{\text{набл}} = 1{,}2 < 1{,}645 = u_{кр}$, то нулевую гипотезу можно принять, можно считать что средний диаметр действительно $d_0 = 10$ мм.

Пример 2. Продавец утверждает, что средний вес пачки чая составляет 100 г. Из партии извлечена выборка и взвешена. Вес каждой пачки - см. таблицу вариантов. Не противоречит ли это утверждению продавца? Доверительная вероятность 99%. Вес пачек чая распределен нормально.

	Выборка
5	98, 104, 97, 97, 101, 100, 99, 101, 99, 98

Решение.

Вычислим показатели выборки.

$$\bar{x} = \frac{1}{n}\sum x_i n_i = \frac{1}{10}994 = 99,4.$$

Исправленная выборочная дисперсия

$$S^2 = \frac{1}{n-1}\sum\left(x_i - \bar{x}\right)^2 n_i = \frac{1}{9}42,4 \approx 4,71.$$

Выборочное исправленное среднеквадратичное отклонение:

$S \approx 2,171$.

Расчеты в таблице:

x_i	n_i	$x_i n_i$	$(x_i - \bar{x})^2 n_i$
97	2	194	11,52
98	2	196	3,92
99	2	198	0,32
100	1	100	0,36
101	2	202	5,12
104	1	104	21,16
Сумма	**10**	**994**	**42,4**

Введем нулевую гипотезу $H_0 : a = 100$ при конкурирующей гипотезе $H_1 : a \neq 100$.

Вычисляем наблюдаемое значение критерия

$$T_{набл} = \frac{(\bar{x} - a)}{S}\sqrt{n} = \frac{(99,4 - 100)}{2,171}\sqrt{10} \approx -0,87.$$

По таблице критических точек распределения Стьюдента найдем критическую точку по уровню значимости $\alpha = 0,01$ и числу степеней свободы $k = 9$, откуда $t_{кр} = 3,25$.

Так как $\left|T_{набл}\right| = 0,87 < 3,25 = t_{кр}$, то нулевую гипотезу о равенстве среднего веса 100 г можно принять.

Ответ: Гипотеза принимается.

Формы контроля освоения заданий по самостоятельной контактной/внеконтактной работы обучающихся по данной теме: тестовые задания и контрольные вопросы.

Рекомендуемая литература:

См. в приложении.

Тема №4: Анализ временных рядов (14 ч)

Цель изучения темы:

Основная цель данной темы сформировать у обучающихся знания и навыки о методах математической, статистической обработки результатов исследований и их анализа. Формирование понятий статистической гипотезы, статистического критерия и навыков проверки гипотезы о нормальном распределении генеральной совокупности.

Задачи: сформировать у обучающихся знания и навыки о методах математической, статистической обработки результатов исследований и их анализа. Формирование понятий статистической гипотезы, статистического критерия и навыков проверки гипотезы о нормальном распределении генеральной совокупности.

Обучающийся должен знать:

методы проверки статистических гипотез о параметрах распределений и согласии с теоретическим распределением;

1. Определение статистической гипотезы;
2. Примеры статистических гипотез.
3. Понятия нулевой и конкурирующей гипотез;

4. Ошибки первого и второго рода. Примеры.
5. Определение статистического критерия, критической области и области принятия решений.
6. Критерий согласия Пирсона.

должен уметь:
- проверять гипотезы о значениях и равенстве параметров различных распределений,
- проверять гипотезы о согласии эмпирического распределения с теоретическим;
- проверять влияние изучаемых факторов любой природы на исследуемую переменную

должен владеть: Навыками формулировки и проверки статистических гипотез, соответствующих данным изучаемой задачи;

должен сформировать компетенции: УК-1, ОПК-4.

Задания для самостоятельной контактной/внеконтактной работы обучающихся по указанной теме:

1. Ответить на контрольные вопросы:

2. *Что называется статистической гипотезой?*
3. *Сформулируйте задачу статистической проверки гипотезы.*
4. *Приведите примеры задач на проверку гипотез и их математические формулировки.*
5. *Назовите шаги логической схемы проверки статистической гипотезы.*
6. *Поясните смысл понятий "ошибка первого рода", "ошибка второго рода", "мощность критерия".*
7. *Что такое уровень значимости?*
8. *Что такое критерий согласия? Каков смысл статистики критерия?*
9. *Зачем нужно знать закон распределения статистики критерия?*
10. *В чем отличие одностороннего и двухстороннего критериев, простой и сложной гипотез?*
11. *Как зависит ширина области принятия гипотезы от уровня значимости?*
12. *Как определяются критические границы для одностороннего и двухстороннего критерия при заданной величине уровня значимости?*
13. *Приведите примеры практических задач, когда необходимо проверить гипотезу о равенстве математических ожиданий или дисперсий.*
14. *Что общего в методике построения доверительных интервалов и проверки статистических гипотез?*
15. *Приведите примеры критериев согласия и области их применения.*
16. *Какие критерии согласия применимы для непрерывных распределений?*

2. Решить задачи.

Пример 4. На двух токарных станках обрабатывается втулки. Отобраны две пробы: $n_1 = 11$ штук – из втулок, обработанных на первом станке, и $n_2 = 14$ штук – из втулок, обработанных на втором станке. По данным этих выборок найдены исправленные выборочные дисперсии $S_x^2 = 0{,}76$ и $S_y^2 = 0{,}38$. Проверить при уровне значимости $\alpha = 0{,}05$ гипотезу H_0 об одинаковой точности этих двух станков при альтернативной гипотезе H_1 о том, что второй станок точнее, чем первый.

Решение:

Найдем отношение большей исправленной дисперсии к меньшей:

$$F_{\text{набл}} = \frac{S_Б^2}{S_М^2} = \frac{0{,}76}{0{,}38} = 2\,.$$

По таблице (прил. 5), по уровню значимости $\alpha = 0{,}05$ и числам степеней свободы $k_1 = n_1 - 1 = 11 - 1 = 10$, $k_2 = n_2 - 1 = 14 - 1 = 13$ находим критическую точку

$$F_{кр},(0{,}05;10;13) = 2{,}67.$$

Так как $F_{набл} < F_{кр}$ – нет оснований отвергнуть гипотезу о равенстве генеральных дисперсий. Другими словами, выборочные исправленные дисперсии различаются незначимо.

Пример 5. На кафедре математики утверждают, что в прошлом году более половины студентов второго курса сдали экзамен на пятерки и четверки. Усомнившись, несколько студентов решили провести исследование, в ходе которого из 30 опрошенных студентов лишь 12 сдали экзамен по математике на пятерки и четверки. Есть ли основания думать, что кафедра математики лукавит? Проверить гипотезу на уровне значимости 1 %.

Решение:

$H_0: p \geq 0{,}5$; $H_1: p < 0{,}5$.

Задан уровень значимости $\alpha = 0{,}01$.

Вычислим значение выборочной доли (или найдем точечную оценку вероятности биномиального закона распределения, т. е. вероятности того, что случайно выбранный студент получит хорошую или отличную оценку):

$$w = \frac{k}{n} = \frac{12}{30} = 0{,}4.$$

Найдем наблюдаемое значение критерия:

$$z_{набл} = \frac{0{,}4 - 0{,}5}{\sqrt{\dfrac{0{,}5(1-0{,}5)}{30}}} = -1{,}095.$$

В соответствии с приведенной выше таблицей критериев критических областей найдем границы левосторонней критической области $\Phi(z_{кр}) = \dfrac{1-2\alpha}{2} = 0{,}49$; и в соответствии с таблицей прил. 3 $z_{кр} = 2{,}33$.

Полученное значение статистики не попало в критическую область. Мы принимаем основную гипотезу. У нас нет оснований сомневаться в истинности утверждения кафедры математики о результатах экзамена.

Пример 6. Компания год назад провела исследование и выяснила, что 5 % покупателей заинтересованы в выпуске нового продукта. Спустя год после начала выпуска, компания провела новое исследование, в ходе которого из 6000 опрошенных 335 положительно отнеслись к выпуску нового продукта. На 2 % уровне значимости определить, возрос ли интерес покупателей к новому продукту?

Решение:

$H_0: p \geq 0{,}05$; $H_1: p < 0{,}05$.

Задан уровень значимости $\alpha = 0{,}02$.

Вычислим значение выборочной доли (или найдем точечную оценку вероятности биномиального закона распределения, т. е. вероятности того, что случайно выбранный студент получит хорошую или отличную оценку):

$$w = \frac{k}{n} = \frac{335}{6000} = 0{,}056.$$

Найдем наблюдаемое значение критерия:

$$z_{\text{набл}} = \frac{0{,}056 - 0{,}05}{\sqrt{\frac{0{,}05(1 - 0{,}05)}{6000}}} = 2{,}14.$$

В соответствии с приведенной выше таблицей критериев критических областей найдем границы левосторонней критической области $\Phi(z_{\text{кр}}) = \frac{1 - 2\alpha}{2} = 0{,}48$; и в соответствии с таблицей прил. 3 $z_{\text{кр}} = 2{,}05$.

$$t_{\text{набл}} > t_{\text{кр}}.$$

Полученное значение статистики попало в критическую область. Мы отклоняем основную гипотезу. Таким образом, в результате исследования выявлено, что интерес покупателей к новой марке возрос.

Формы контроля освоения заданий по самостоятельной контактной/внеконтактной работы обучающихся по данной теме: тестовые задания и контрольные вопросы.

Рекомендуемая литература:

См. в приложении.

Тема № 5: Дисперсионный анализ (14 ч)

Цель изучения темы:
Основная цель данной темы сформировать у обучающихся знания и навыки о методах математической, статистической обработки результатов исследований и их анализа. Формирование понятий статистической гипотезы, статистического критерия и навыков проверки гипотезы о нормальном распределении генеральной совокупности.

Задачи: сформировать у обучающихся знания и навыки о методах математической, статистической обработки результатов исследований и их анализа. Формирование понятий статистической гипотезы, статистического критерия и навыков проверки гипотезы о нормальном распределении генеральной совокупности.
Обучающийся должен знать:
методы проверки статистических гипотез о параметрах распределений и согласии с теоретическим распределением;
1. Определение статистической гипотезы;
2. Примеры статистических гипотез.
3. Понятия нулевой и конкурирующей гипотез;
4. Ошибки первого и второго рода. Примеры.
5. Определение статистического критерия, критической области и области принятия решений.
6. Критерий согласия Пирсона.

должен уметь:
- проверять гипотезы о значениях и равенстве параметров различных распределений,
- проверять гипотезы о согласии эмпирического распределения с теоретическим;
- проверять влияние изучаемых факторов любой природы на исследуемую переменную

должен владеть: Навыками формулировки и проверки статистических гипотез, соответствующих данным изучаемой задачи;
должен сформировать компетенции: УК-1, ОПК-4.
Задания для самостоятельной контактной/внеконтактной работы обучающихся по указанной теме:
1.Ознакомиться с теоретическим материалом по теме.
2.Ответить на вопросы.

Контрольные вопросы по теме «Проверка статистических гипотез»
1. Какая гипотеза называется статистической? Приведите пример.
2. Какая статистическая гипотеза называется нулевой? Альтернативной? Приведите примеры.
3. Что такое критерий значимости?
4. Что такое уровень значимости? Как он связан с доверительной вероятностью?
5. Что такое критическая область критерия?
6. Поясните смысл ошибок первого и второго рода, возникающих при проверке гипотез.
7. Какие критерии называются односторонними и двусторонними?
8. Приведите пример Н0 и Н1 гипотез.
9. Какие выводы делает исследователь, если гипотеза Н0 отклоняется?
10. Какие выводы делает исследователь, если гипотеза Н0 принимается?
11. Как связаны вид альтернативной гипотезы и тип критической области?
12. Какой области (допустимых значений или критической) принадлежит Кэмп, если делается вывод, что выборочные данные не противоречат данной гипотезе Н0 о генеральной совокупности?
13. Какой области (допустимых значений или критической) принадлежит Кэмп, если делается вывод, что выборочные данные не согласуются с выдвинутой гипотезой?
14. Какие критерии называются параметрическими?
15. Дайте постановку задачи, для решения которой применяется критерий Стьюдента.

16. При каких условиях применяется критерий Стьюдента?
17. Какое условие необходимо проверить до начала применения критерия Стьюдента при малых выборках?
18. Опишите последовательность действий применения критерия Стьюдента для независимых выборок.
19. Дайте описание нулевой гипотезы в задаче о сравнении средних значений признака в двух независимых выборках

Тестовые задания по теме
«Проверка статистических гипотез»

1. Нулевая гипотеза – это: а) выдвинутая гипотеза; б) гипотеза, противоречащая выдвинутой;
в) гипотеза о равенстве нулю генерального среднего; г) гипотеза, которая никогда не выполняется.
2. Конкурирующая гипотеза – это: а) гипотеза, противоречащая выдвинутой; б) гипотеза, которая никогда не выполняется; в) гипотеза, совпадающая с выдвинутой; г) гипотеза о равенстве нулю генерального среднего.
3. Если основная гипотеза имеет вид а=20: то конкурирующей может быть гипотеза:
а) $H_1 : a = 4$; б) $H_1 : a \geq 2$;
в) $H_1 : a > 20$; г) $H_1 : a \leq 20$.
4. Случайную величину, которая служит для проверки нулевой гипотезы, называют: а) среднеквадратичным отклонением; б) дисперсией; в) статистическим критерием; г) наблюдаемым значением критерия.
5. Значение статистического критерия, вычисленное по данным выборки, называют: а) генеральным значением критерия; б) выборочным значением критерия; в) реальным значением критерия; г) наблюдаемым значением критерия.
6. Область значений статистического критерия, когда нулевая гипотеза отвергается, называется: а) критической областью; б) полупрямой; в) интервалом; г) областью принятия гипотезы.
7. Ошибкой второго рода называют ошибку, состоящую в том, что: а) будут приняты и нулевая, и конкурирующая гипотезы; б) будет принята неправильная гипотеза; в) будет отвергнута правильная гипотеза; г) будут отвергнуты и нулевая, и конкурирующая гипотезы.
8. Ошибкой первого рода называют ошибку, состоящую в том, что: а) будет отвергнута правильная гипотеза; б) будут приняты и нулевая, и конкурирующая гипотезы; в) будут отвергнуты и нулевая, и конкурирующая гипотезы; г) будет принята неправильная гипотеза.
9. Вероятность совершить ошибку первого рода называется: а) надежностью; б) среднеквадратичным отклонением; в) доверительным интервалом; г) уровнем значимости.
10. Эмпирическое значение критерия принадлежит области допустимых значений. На уровне значимости α правильное статистическое решение: 1) выборочные данные не противоречат основной гипотезе; 2) выборочные данные противоречат основной гипотезе; 3) основная гипотеза принимается; 4) основная гипотеза отклоняется.
Ответы:
а) 1 и 2; б) 1 и 3; в) 2 и 4; г) 2 и 3.
11. Эмпирическое значение критерия принадлежит критической области. На уровне значимости α принимается статистическое решение: 1) выборочные данные не согласуются с основной гипотезой; 2) выборочные данные противоречат основной гипотезе; 3) основная гипотеза принимается; 4) основная гипотеза отклоняется в пользу конкурирующей гипотезы.
Ответы:

а) 1 и 2; б) 1 и 3; в) 2 и 4; г) 2 и 3.

12. Вероятность статистического решения отклонить верную гипотезу называют: а) уровень значимости; б) уровень доверия; в) мощность критерия; г) ошибка второго рода.

13. Вероятность статистического решения принять верную альтернативную гипотезу: а) уровень значимости; б) уровень доверия; в) мощность критерия; г) ошибка второго рода.

14. Критическая область отклонения нулевой гипотезы в пользу альтернативной гипотезы о генеральных средних $H_1 : a_x > a_y$:

а) двусторонняя; б) правосторонняя;
в) левосторонняя ; г) не содержит точку $(a_x; a_y)$.

15. Область допустимых значений критерия – это

а) область принятия гипотезы H_0;
б) область принятия гипотезы H_1;
в) область отвержения гипотезы H_0;
г) область отвержения гипотезы H_1.

16. Критерий Стьюдента применяется для статистической оценки различия: а) генеральных средних значений признака; б) выборочных средних значений признака; в) генеральных дисперсий признака; г) выборочных дисперсий признака.

17. При проверке гипотезы о теоретическом законе распределения наблюдаемое значение критерия сравнивают с критической точкой распределения: а) Стьюдента; б) Фишера; в) Пирсона; г) Гаусса; д) нормального.

18. При проверке гипотезы о равенстве математических ожиданий при известных дисперсиях наблюдаемое значение критерия сравнивают с критической точкой распределения: а) Стьюдента; б) Фишера; в) Пирсона; г) Гаусса; д) нормального.

19. Что не надо делать при проверке статистической гипотезы о равенстве математических ожиданий: а) определить основную гипотезу; б) найти медиану; в) задать уровень значимости или доверительной вероятности; г) найти выборочное среднее, объем выборки, выборочное среднее квадратичное отклонение; д) вычислить наблюдаемое значение критерия.

20. Отметьте не менее двух правильных ответов. Пусть для гипотезы о средних $H_0 : a_x = a_y$ альтернативная гипотеза H_1 имеет вид $a_x > a_y$. По критерию Стьюдента значение $t_{\text{эмп}} = 1{,}2$, а для уровня значимости 0,05 значение $t_{\text{крит}} = 2{,}015$. Принимается следующее статистическое решение:

1) с ошибкой 0,05 нет оснований для отклонения гипотезы Н0 о незначимости различий между генеральными средними; 2) с ошибкой 0,05 можно считать различие между генеральными средними статистически незначимым (недостоверным), что объясняется случайными причинами; 3) с достоверностью 0,95 принимается гипотеза Н1 о том, что генеральное среднее выборки X больше генеральной средней выборки Y; 4) средние двух совокупностей значимо различаются с достоверностью 0,95. Различие средних значений не может быть объяснено случайными причинами.

Ответы:

а) 1 и 2; б) 1, 2, 3; в) 1, 2, 4; г) 2 и 3.

21. Отметьте не менее двух правильных ответов. Для гипотезы «о средних» $H_0 : a_x = a_y$ альтернативная гипотеза H_1 имеет вид $a_x > a_y$. По критерию Стьюдента $t_{\text{эмп}} = 2{,}2$, а для уровня значимости 0,05 значение $t_{\text{крит}} = 1{,}697$. Статистическое решение …

1) с достоверностью 0,95 принимается гипотеза H_1 о том, что «генеральное среднее выборки X больше генеральной средней выборки Y;

2) с достоверностью 0,95 генеральные средние совокупностей значимо различаются; имеющееся различие средних значений не может быть объяснено случайными причинами;

3) с ошибкой 0,05 нет оснований для отклонения гипотезы H_0 о незначимости различий между генеральными средними;

4) с ошибкой 0,05 можно говорить о незначимом различии между генеральными средними, что объясняется случайностью выборок.

Ответы:

а) 1 и 2; б) 1, 2, 3; в) 1, 2, 4; г) 2 и 3.

22. Отметьте не менее двух правильных ответов. Для гипотезы «о средних» $H_0 : a_x = a_y$ альтернативная гипотеза H_1 имеет вид $a_x > a_y$. По критерию Стьюдента $t_{\text{эмп}} = 1{,}35$, а для уровня значимости 0,05 значение $t_{\text{крит}} = 1{,}697$. Статистическое решение …

1) с достоверностью 0,95 принимается гипотеза H_1;

2) с достоверностью 0,95 имеющееся различие средних значений не может быть объяснено случайными причинами;

3) с ошибкой 0,05 нет оснований для отклонения гипотезы H_0 о незначимости различий между генеральными средними;

4) с ошибкой 0,05 можно говорить о незначимом различии между генеральными средними, что объясняется случайностью выборок.

Ответы:

а) 1 и 2; б) 1 и 3; в) 3 и 4; г) 2 и 3.

23. При проверке гипотезы о равенстве дисперсий двух нормально распределенных случайных величин наблюдаемое значение критерия сравнивают с критической точкой распределения:

а) Стьюдента; б) Фишера; в) Пирсона;
г) Гаусса; д) нормального.

24. Что не надо делать при проверке статистической гипотезы о равенстве дисперсий двух нормально распределенных случайных величин: а) определить основную гипотезу; б) применить формулу Стэрджеса; в) задать уровень значимости или доверительной вероятности; г) вычислить наблюдаемое значение критерия Фишера.

25. При проверке гипотезы о нормальном законе распределения по критерию Пирсона вероятность попадания случайной величины в i-й интервал (x_i, x_{i+1}) определяется по формуле:

а) $P_i = \Phi\left(\dfrac{x_i - \bar{x}}{\sigma_в}\right) - \Phi\left(\dfrac{x_{i+1} - \bar{x}}{\sigma_в}\right)$;

б) $P_i = \Phi\left(\dfrac{x_{i+1} - \bar{x}}{\sigma_в}\right) - \Phi\left(\dfrac{x_i - \bar{x}}{\sigma_в}\right)$;

в) $P_i = \Phi(x_{i+1}) - \Phi(x_i)$;

г) $P_i = \Phi(x_{i+1} - \bar{x}) - \Phi(x_i - \bar{x})$.

26. Что не надо делать при проверке статистической гипотезы о теоретическом законе распределения: а) определить основную гипотезу; б) найти доверительные интервалы для оценки параметров; в) задать уровень значимости или доверительной вероятности; г) найти числовые характеристики; д) вычислить наблюдаемое значение критерия.

27. Что не надо делать при проверке статистической гипотезы о теоретическом законе распределения: а) определить основную гипотезу; б) определить альтернативную гипотезу; в) задать уровень значимости или доверительной вероятности; г) найти наблюдаемое значение критерия Фишера; д) построить прямые регрессии.

3.Решить задачи.

Пример 7. Студенты двух вузов сдавали экзамены по физике. В первом вузе экзаменовалось 30 студентов, средняя оценка 52, во втором вузе – 36 студентов, средняя оценка 47 (по100-бальной системе оценок). Стандартное отклонение оценок на экзаменах, вычисленное для нескольких тысяч студентов, равно 12. Можно ли утверждать с вероятностью 95 %, что первый вуз дает подготовку по физике лучше, чем второй?

Решение:

Имеем:

$\overline{X}=52;\ \overline{Y}=47;\ n_x=30;\ n_y=36;\ \sigma_1=\sigma_2=12;\ \alpha=0{,}05.$

$H_0: a_1=a_2;\quad H_1: a_1>a_2.$

$$z_{\text{набл}}=\frac{\overline{X}-\overline{Y}}{\sqrt{\frac{\sigma_1^2}{n_x}+\frac{\sigma_2^2}{n_y}}}=\frac{52-47}{12\sqrt{\frac{1}{30}+\frac{1}{36}}}=1{,}69.$$

В соответствии с приведенной выше таблицей критериев критических областей найдем границы правосторонней критической области $\Phi(z_{\text{кр}})=\frac{1-2\alpha}{2}=0{,}45;$ и в соответствии с таблицей прил. 3 $z_{\text{кр}}=1{,}65.$

$z_{\text{набл}}>z_{\text{кр}}$, следовательно, гипотеза H_0 отвергается, т. е. подготовка студентов по физике вуза *A* лучше, чем вуза *B*.

Пример 8. Имеются независимые выборки значений нормально распределенных случайных величин

X: 2, 2, 3, 3, 4, 4, 4, 5, 5, 6

и *Y*: 3, 3, 3, 4, 4, 4, 4, 5, 5, 5, 5, 5, 7, 8, 9.

Требуется проверить для уровня значимости α = 0,1 при условии равенства генеральных дисперсий нулевую гипотезу H_0: $M(X)=M(Y)$ при конкурирующей гипотезе H_1: $M(X)\neq M(Y)$.

Решение:

Объемы выборок $m=10$, $n=15$. Вычислим выборочные средние и исправленные выборочные дисперсии:

$$\bar{x}_B=3{,}8;\ \bar{y}_B=4{,}93;\ S_x^2=1{,}73;\ S_y^2=3{,}21.$$

$$S^2=\frac{(n_x-1)S_x^2+(n_y-1)S_y^2}{n_x+n_y-2}=\frac{9\cdot1{,}73^2+14\cdot3{,}21^2}{23}=7{,}44.$$

Вычислим наблюдаемое значение критерия:

$$T_{\text{набл}}=\frac{\overline{X}-\overline{Y}}{S\sqrt{\frac{1}{n_x}+\frac{1}{n_y}}}=\frac{3{,}8-4{,}93}{2{,}73\sqrt{\frac{1}{10}+\frac{1}{15}}}=-1{,}02.$$

Критическая область – двусторонняя, $t_{\text{двуст.кр.}}(0{,}1;\,23)=1{,}71.$

Итак, $|T_{\text{набл}}|<t_{\text{двуст.кр.}}$, следовательно, нет оснований отвергнуть нулевую гипотезу – можно считать, что математические ожидания генеральных совокупностей равны.

Пример 9. В ходе проведения выборов было установлено, что в Москве из 1500 потенциальных случайно выбранных избирателей реально в выборах приняли участие 480 человек, а в Санкт-Петербурге из 1630 потенциальных избирателей на избирательные участки пришли 490 человек. На уровне значимости α = 10 % проверить гипотезу о равенстве генеральных долей избирателей в двух этих городах, реально принявших участие в выборах.

Решение:

$H_0 : W_1 = W_2$.
$H_1 : W_1 \neq W_2$.
Экспериментальные данные:
$n_1 = 1500$; $k_1 = 480$; $n_2 = 1630$; $k_2 = 490$; $\gamma = 0{,}90$; $\alpha = 0{,}10$.

Тогда $W_1 = \dfrac{480}{1500} = 0{,}32$; $W_2 = \dfrac{490}{1630} = 0{,}30$;

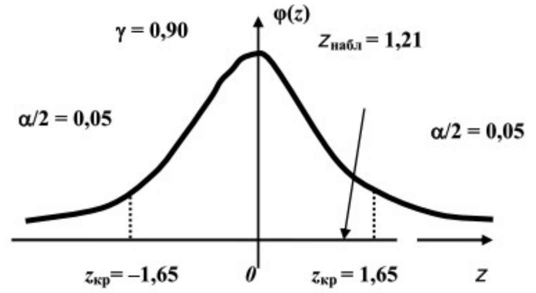
и $\hat{p} = \dfrac{m_1 + m_2}{n_1 + n_2} = \dfrac{970}{3130} = 0{,}31$.

Вычислим на основе экспериментальных данных наблюдаемое значение критерия:

$$z_{\text{набл}} = \frac{W_1 - W_2}{\sqrt{\hat{p} \cdot \left(1 - \hat{p}\right) \cdot \left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} = \frac{0{,}32 - 0{,}30}{\sqrt{0{,}31 \cdot 0{,}69 \cdot \left(\dfrac{1}{1500} + \dfrac{1}{1630}\right)}} = 1{,}21.$$

В соответствии с приведенной выше таблицей критериев критических областей найдем границы двусторонней критической области $\Phi(z_{\text{кр}}) = \dfrac{1-\alpha}{2} = 0{,}45$; и в соответствии с таблицей прил. 3 $z_{\text{кр}} = 1{,}65$.

Покажем все найденные значения на графике плотности стандартного нормального закона распределения, который описывает поведение случайной величины t при справедливости нулевой гипотезы.



Поскольку наблюдаемое значение критерия попало в область естественных для данного закона распределения значений (в данном случае стандартного нормального закона распределения), то гипотеза H0 принимается как не противоречащая экспериментальным данным с уровнем доверия 90 %, т. е. генеральные доли электората, реально принявших участие в выборах в Москве и Санкт-Петербурге, значимо не отличаются (их можно считать одинаковыми).

Пример 10. Для выборки, интервальный статистический ряд которой имеет вид

Номер интервала	Границы интервала	Эмпирические частоты
1	2–5	6
2	5–8	8
3	8–11	15
4	11–14	22
5	14–17	14
6	17–20	5

проверить при уровне значимости $\alpha = 0,05$ гипотезу:

а) о показательном; б) равномерном; в) нормальном законе распределения генеральной совокупности с помощью критерия Пирсона.

Решение:

Объем выборки $n = 70$. Будем считать вариантами середины частичных интервалов: $x_1 = 3,5$, $x_2 = 6,5,\ldots, x_6 = 18,5$.

Найдем $\bar{x}_B = 11,43$; $\sigma_B = 4,03$; $s = 4,05$.

а) Вычислим теоретические частоты в предположении о показательном распределении генеральной совокупности при $\lambda^* = \dfrac{1}{11,43} = 0,087$:

$$n_1' = 70\left(e^{-0,087\cdot 2} - e^{-0,087\cdot 5}\right) = 70\left(e^{-0,174} - e^{-0,435}\right) = 13,44;$$

аналогично $n_2' = 10,37;\ n_3' = 8,05;\ n_4' = 6,23;\ n_5' = 4,76;\ n_6' = 3,64.$

Наблюдаемое значение критерия

$$\chi^2_{\text{набл}} = \frac{(6-13,44)^2}{13,44} + \ldots + \frac{(5-3,64)^2}{3,64} = 69,02.$$

Критическая точка $\chi^2(0,05;4) = 9,5;\ \chi^2_{\text{набл}} > \chi^2_{\text{кр}}$ и гипотеза о показательном распределении отклоняется.

б) Для равномерного распределения

$$a^* = 11,43 - \sqrt{3}\cdot 4,03 = 4,45;$$
$$b^* = 11,43 + \sqrt{3}\cdot 4,03 = 18,41;$$
$$f(x) = \frac{1}{a^* - b^*} = \frac{1}{18,41 - 4,45} = 0,072.$$

Теоретические частоты:

$n_1' = 70\cdot(5 - 4,45)\cdot 0,072 = 2,77;$
$n_2' = n_3' = n_4' = n_5' = 70\cdot 3\cdot 0,072 = 15,12;$
$n_6' = 70\cdot(18,41 - 17)\cdot 0,072 = 7,1.$

Наблюдаемое значение критерия

$$\chi^2_{\text{набл}} = \frac{(6-2,77)^2}{2,77} + \ldots + \frac{(5-7,1)^2}{7,1} = 10,95.$$

Критическая точка $\chi^2(0,05;3) = 7,8;\ \chi^2_{\text{набл}} > \chi^2_{\text{кр}}$ и гипотеза о равномерном распределении отклоняется.

в) Теоретические частоты для нормального распределения:

$$n_1' = 70\cdot\left(\Phi\left(\frac{5-11,43}{4,05}\right) - \Phi\left(\frac{2-11,43}{4,05}\right)\right) = 70\cdot\left(\Phi(-1,588) - \Phi(-2,328)\right) =$$
$$= 70\cdot\left(\Phi(2,328) - \Phi(1,588)\right) = 70\cdot(0,4900 - 0,4441) = 3,2.$$

Так же вычисляются

$n_2' = 9,9;\ n_3' = 18,2;\ n_4' = 19,6;\ n_5' = 12,5;\ n_6' = 4,7.$

Наблюдаемое значение критерия

$$\chi^2_{\text{набл}} = \frac{(6-3,2)^2}{3,2} + \ldots + \frac{(5-4,7)^2}{4,7} = 3,87.$$

Критическая точка $\chi^2(0,05;\ 3) = 7,8$. Поскольку $\chi^2_{набл} < \chi^2_{кр}$, гипотеза о нормальном распределении генеральной совокупности принимается.

Формы контроля освоения заданий по самостоятельной контактной/внеконтактной работы обучающихся по данной теме: тестовые задания и контрольные вопросы.

Рекомендуемая литература:
См. в приложении.

Тема № 6: Статистические методы обработки результатов экспериментальных измерений (14 ч)

Цель изучения темы: научиться проводить статистический анализ связей между факторными и результативными признаками статистической совокупности (причинно-следственная связь) или определение зависимости параллельных изменений нескольких признаков этой совокупности от какой либо третьей величины (от общей их причины), уметь изучать особенности этой связи, определять ее размеры и направление, а также оценивать ее достоверность.

Задачи:
проводить статистический анализ связей между факторными и результативными признаками статистической совокупности (причинно-следственная связь) или определение зависимости параллельных изменений нескольких признаков этой совокупности от какой либо третьей величины (от общей их причины), уметь изучать особенности этой связи, определять ее размеры и направление, а также оценивать ее достоверность.
Для формирования профессиональных компетенций обучающийся должен **знать**:
основные понятия и методы статистической обработки результатов исследований. а) виды проявления количественных связей;
б) понятие функциональной и корреляционной зависимости;
в) характеристики коэффициента корреляции;
г) методику и порядок определения коэффициента корреляции;
Для формирования профессиональных компетенций обучающийся должен **владеть** понятийным аппаратом статистического анализа
и овладеть следующими **компетенциями:** УК-1, ОПК-4..
Для формирования профессиональных компетенций обучающийся должен **уметь**:
вычислять характеристики статистической обработки результатов исследований, решать задачи на нахождение коэффициента линейной корреляции.
1. Рассчитывать коэффициент корреляции
2. Оценивать силу, направление и достоверность полученного коэффициента корреляции и делать соответствующие выводы;
3.Давать ответы на тестовые задания,
4. Решать ситуационные статистические задачи по теме.
Задания для самостоятельной контактной/внеконтактной работы обучающихся по указанной теме:
1.Ознакомиться с теоретическим материалом по теме.
2.Ответить на вопросы.
Тесты.

1. Назовите виды зависимостей между признаками, которые могут иметь место в психологическом исследовании.
2. Какая зависимость между психологическими признаками называется статистической? Приведите пример.
3. В чем отличие терминов «корреляционная связь» и «корреляционная зависимость»?
4. Сформулируйте основные задачи корреляционного анализа.
1.Приведите пример положительной корреляционной связи между психологическими признаками. Ответ обоснуйте.
6. Приведите пример линейной положительной корреляционной связи между психологическими признаками. Ответ обоснуйте.
5.Приведите пример линейной отрицательной корреляционной связи между психологическими признаками. Ответ обоснуйте.
6.Сформулируйте свойства коэффициента корреляции Пирсона.

7.Какой вывод делает исследователь, если выборочный коэффициент корреляции Пирсона равен: 1) ; 2); 3)?

Решить задачи.

На основании 18 наблюдений установлено, что на 64% вес (X) кондитерских изделий зависит от их объема (Y). Можно ли на уровне значимости $\alpha = 0,05$ утверждать, что между X и Y существует зависимость?

РЕШЕНИЕ.

Из условия задачи имеем, что $n = 18$, $r = 0,64$.

Введем нулевую гипотезу $H_0 : r = 0$. Проверим эту гипотезу об отсутствии корреляционной зависимости (о незначимости коэффициента корреляции). Вычислим значение критерия $T_{набл} = \dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \dfrac{0,64\cdot\sqrt{16}}{\sqrt{1-0,64^2}} \approx 3,33$.

Найдем критическую точку по уровню значимости $\alpha = 0,05$ и числу степеней свободы $k = n-2 = 16$, получаем $t_{кр} = 2,12$. Так как $|T_{набл}| = 3,33 > 2,12 = t_{кр}$, следует отвергнуть нулевую гипотезу $H_0 : r = 0$, то есть корреляционная зависимость существенна, между X и Y существует зависимость.

ОТВЕТ. между X и Y существует зависимость.

Исследование 27 семей по среднедушевому доходу (***X***) и сбережениям (***Y***) дало результаты: $\overline{X}$ =82 у.е., S_x=31 у.е., $\overline{Y}$ =39 у.е., S_y=29 у.е., $\overline{XY}$ =3709 (у.е.)^2. При α=0,05 проверить наличие линейной связи между ***X*** и ***Y***. Определить размер сбережений семей, имеющих среднедушевой доход X=130у.е.

РЕШЕНИЕ.

Вычислим выборочный коэффициент корреляции:

$$r = \frac{\overline{XY} - \overline{X}\cdot\overline{Y}}{S_X\cdot S_Y} = \frac{3709 - 82\cdot 39}{31\cdot 29} \approx 0,568.$$

Проверим гипотезу о значимости коэффициента корреляции. Введем нулевую гипотезу $H_0 : r = 0$ и вычислим наблюдаемое значение критерия Стьюдента

$$t_{набл.} = \sqrt{\frac{r^2}{1-r^2}(n-2)} = \sqrt{\frac{0,568^2}{1-0,568^2}(27-2)} = 3,451.$$

Найдем критическую точку $t_{кр.}(0,05,\ 27-2) = t_{кр.}(0,05,\ 25) = 2,06$. Так как $t_{набл.} = 3,451 > 2,06 = t_{кр.}$, нулевую гипотезу следует отвергнуть, коэффициент корреляции значим и между X и Y существует линейная связь (средней силы).

Найдем уравнение линейной регрессии Y на X по формуле: $\overline{Y}_x - \overline{Y} = r\dfrac{S_y}{S_x}(X - \overline{X})$.

Получаем:

$$\overline{Y}_x - 39 = 0,568\frac{29}{31}(X - 82),$$

$$\overline{Y}_x = 0,531X - 4,571.$$

Тогда размер сбережений семей, имеющих среднедушевой доход X=130у.е. равен $\overline{Y}_x(130) = 0,531\cdot 130 - 4,571 = 64,459$ у.е.

ОТВЕТ. 64,459 у.е. Связь значима.

Пример 1. В таблице приведен ряд, устанавливающий связь между уровнем IQ и уровнем средней успеваемости студентов по математике.

X – уровень IQ	75	85	90	100	105	110	110	115	115	120	125	130	140
Y – средняя успеваемость	3,1	3,1	3,5	3,7	3,8	4,0	4,2	4,3	4,6	4,7	4,8	4,9	5,0

Существует ли взаимосвязь между уровнем IQ (признак X) и средним уровнем успеваемости по математике (признак Y)?

Решение: представим исходные данные в расчетную таблицу.

Номер п/п	x_i	y_i	x_i^2	y_i^2	$x_i y_i$
1	75	3,1	5625	9,61	232,5
2	85	3,1	7225	9,61	263,5
3	90	3,5	8100	12,25	315,0
4	100	3,7	10000	13,69	370
5	105	3,8	11025	14,44	399
6	110	4,0	12100	16,00	440
7	110	4,2	12100	17,64	462
8	115	4,3	13225	18,49	494,5
9	115	4,6	13225	21,16	529
10	120	4,7	14400	22,09	564
11	125	4,8	15625	23,04	600
12	130	4,9	16900	24,01	637
13	140	5,0	19600	25,00	700
Сумма:	**1420**	**53,7**	**159150**	**227,03**	**6006,5**

Вычислим выборочные средние:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1420}{13} = 109{,}23; \qquad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{53{,}7}{13} = 4{,}13;$$

$$\overline{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i y_i) = \frac{6006{,}5}{13} = 462{,}04.$$

Теперь вычислим значения выборочных средних квадратических отклонений:

$$S_x = \sqrt{\overline{x^2} - (\bar{x})^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i{}^2 - (\bar{x})^2} = \sqrt{\frac{1}{13}\cdot 159150 - 109{,}23^2} = 17{,}64;$$

$$S_y = \sqrt{\overline{y^2} - (\bar{y})^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} y_i{}^2 - (\bar{y})^2} = \sqrt{\frac{1}{13}\cdot 227{,}03 - 4{,}13^2} = 0{,}63.$$

Подставим в формулу:

$$r = \frac{\overline{xy} - \bar{x}\cdot\bar{y}}{S_x \cdot S_y} = \frac{462{,}04 - 109{,}23\cdot 4{,}13}{17{,}64\cdot 0{,}63} = \frac{10{,}9201}{11{,}1132} \approx 0{,}97.$$

Корреляционная связь между уровнем IQ и средним уровнем успеваемости по математике близка к линейной положительной. Чем выше уровень IQ у студентов, тем выше средний уровень успеваемости по математике, и наоборот.

Пример 2. Определить значимость выборочного коэффициента корреляции, вычисленного в примере 1.

Решение: выдвинем гипотезу H0: r=0 о том, что в генеральной совокупности отсутствует корреляция. Так как знак корреляции в результате решения примера 1 определен – корреляция положительна, то альтернативная гипотеза является односторонней вида H1 : r>0 .

Найдем эмпирическое значение t -критерия:

$$t_{\text{эмп}} = |r|\sqrt{\frac{n-2}{1-r^2}} = 0{,}97\sqrt{\frac{13-2}{1-0{,}97^2}} = 13{,}23.$$

Число степеней свободы равно $k = n - 2 = 13 - 2 = 11$, уровень значимости выберем равным $\alpha = 0{,}01$. По таблице прил. 2 находим критическое значение $t_{\text{крит}}(0{,}01;11) = 3{,}11$.

Так как $t_{\text{эмп}} > t_{\text{крит}}$, то между уровнем IQ и средним уровнем успеваемости по математике существует статистически значимая корреляция.

Пример 3. По выборке n = 100 извлеченной из двумерной нормальной генеральной совокупности (X, Y), найден выборочный коэффициент корреляции r = 0,2. По уровню значимости 0,05 проверить нулевую гипотезу о равенстве нулю генерального коэффициента корреляции при конкурирующей гипотезе H1: $r \neq 0$

Решение: найдём наблюдаемое (эмпирическое) значение критерия:

$$t_{\text{эмп}} = |r|\sqrt{\frac{n-2}{1-r^2}} = 0{,}2\sqrt{\frac{100-2}{1-0{,}2^2}} = 2{,}02.$$

Конкурирующая гипотеза имеет вид $r \neq 0$ поэтому критическая область – двусторонняя. По таблице прил. 2 при уровне значимости α=0,05 и числе степеней свободы k= n-2=98 находим критическую точку двусторонней критической области tкр(0,05; 98) =1,99. Так как tэмп>t кр – отвергаем нулевую гипотезу о равенстве нулю генерального коэффициента корреляции. Другими словами, коэффициент корреляции значимо отличается от нуля. Таким образом, X и Y коррелированны.

Контрольные вопросы:

1. Что называется функциональной зависимостью?

2. Дайте определение функции регрессии.

3. Как составляется корреляционной таблица?

4. По какой формуле вычисляется условное среднее значение случайной величины?

5. Запишите выборочное уравнение линейной регрессии.

6. Почему метод наименьших квадратов наиболее эффективен, если функция линейна относительно искомых параметров?

7. Какой величиной оценивается теснота корреляционной зависимости случайных величин?

8. Какие значения может принимать коэффициент корреляции?

9. Какая из формул выражает условное среднее значение случайной величины У

а) $\bar{y}_{x_i} = \frac{1}{m_{x_i}} \sum_{j=1}^{l} m_{ij} y_j$; б) $\bar{y}_{x_i} = \sum_{j=1}^{l} m_{ij} y_j$;

в) $\bar{y}_{x_i} = m_{x_i} \sum_{j=1}^{l} m_{ij} y_j$; г) $\bar{y}_{x_i} = m_{x_i} \sum_{j=1}^{l} m_{ij}$

10. Выберите правильную запись выборочного коэффициента линейной корреляции. У на X:

а) $r_{xy} = \frac{\overline{XY} - \bar{X}\bar{Y}}{\sigma_x^3}$; б) $r_{xy} = \frac{\overline{XY} - \bar{X}\bar{Y}}{\sigma_x \sigma_y}$;

в) $r_{yx} = \frac{\overline{XY} - \bar{X}\bar{Y}}{\sigma_x^2}$; г) $r_{xy} = \frac{\bar{X} - \bar{Y}}{\sigma_y}$.

11. Величина коэффициента корреляции двух величин, связанных линейной корреляционной зависимостью, удовлетворяет

а) $-2 \le r \le 1$; б) $-2 \le r \le 2$;

в) $-1 \le r \le \frac{1}{2}$; г) $-1 \le r \le 1$.

неравенству:

Тест.

1. Коэффициент корреляции является мерой …

1. *статистической связи между случайными величинами*
2. *вероятностной связи между случайными величинами*
3. *корреляционной связи между случайными величинами*
4. *линейной связи между случайными величинами*

2. Высокому уровню линейной связи между переменными соответствует значение коэффициента линейной корреляции ….

1. *по модулю близкое к нулю*
2. *по модулю близкое к единице*
3. *положительное больше 1*

3. Выборочный коэффициент линейной корреляции …

1. *безразмерная величина*
2. *зависит от единиц измерения признаков X и Y*
3. *имеет размерность, совпадающую с размерностью признаков X и Y*
4. *имеет размерность произведения размерностей признаков X и Y*

4. *Отметьте не менее двух правильных ответов. Задачи корреляционного анализа:*
1. установление направления корреляционной связи
2. установление формы корреляционной связи
3. измерение тесноты корреляционной связи
4. нахождение уравнения регрессии
5. *Выборочный коэффициент линейной корреляции принимает значения ...*
1. от –1 до –1
2. больше нуля
3. положительные и отрицательные
4. от нуля до единицы
6. *Отметьте не менее двух правильных ответов. Проверка значимости выборочного коэффициента корреляции основана на статистической проверке гипотезы:*
1. в генеральной совокупности отсутствует корреляция
2. отличие от нуля выборочного коэффициента корреляции объясняется только случайностью выборки
3. коэффициент корреляции значимо отличается от 0
4. отличие от нуля выборочного коэффициента корреляции не случайно
Ответы. ***1.*** *3.* ***2.*** *2.* ***3.*** *1.* ***4.*** *1, 2, 3.* ***5.*** *1.* ***6.*** *1, 3.*

Формы контроля освоения заданий по самостоятельной контактной/внеконтактной работы обучающихся по данной теме: тестовые задания и контрольные вопросы.

Рекомендуемая литература:
См. в приложении.

Тема № 7: Применение пакета Statistica для анализа зависимостей (12ч)

Цель изучения темы: сформировать у обучающихся основные понятия и процедуры использования коэффициентов корреляции при проведении исследования, знакомство с основными понятиями корреляционного анализа; анализ особенностей применения и процедуры подсчета коэффициентов корреляции; освоение пошаговой процедуры подсчета коэффициентов корреляции. Научиться определять наличие связи между признаками статистической совокупности, определять ее размеры и направление, и оценить достоверность.

Задачи: сформировать у обучающихся основные понятия и процедуры использования коэффициентов корреляции при проведении исследования, знакомство с основными понятиями корреляционного анализа; анализ особенностей применения и процедуры подсчета коэффициентов корреляции; освоение пошаговой процедуры подсчета коэффициентов корреляции. Научиться определять наличие связи между признаками статистической совокупности, определять ее размеры и направление, и оценить достоверность.

Для формирования профессиональных компетенций обучающийся должен **знать**: основные понятия и методы статистической обработки результатов исследований.
а) виды проявления количественных связей;
б) понятие функциональной и корреляционной зависимости;
в) характеристики коэффициента корреляции;
г) методику и порядок определения коэффициента корреляции;

Для формирования профессиональных компетенций обучающийся должен **владеть** понятийным аппаратом статистического анализа
и овладеть следующими **компетенциями:** УК-1, ОПК-4..

Для формирования профессиональных компетенций обучающийся должен **уметь**: вычислять характеристики статистической обработки результатов исследований,
решать задачи на нахождение коэффициента линейной корреляции.
рассчитывать коэффициент корреляции
оценивать силу, направление и достоверность полученного коэффициента корреляции и делать соответствующие выводы;
давать ответы на тестовые задания,
решать ситуационные статистические задачи по теме.

Задания для самостоятельной контактной/внеконтактной работы обучающихся по указанной теме:
1.Ознакомиться с теоретическим материалом по теме.
2.Ответить на вопросы.
1. Назовите виды зависимостей между признаками, которые могут иметь место в психологическом исследовании.
2. Какая зависимость между психологическими признаками называется статистической? Приведите пример.
3. В чем отличие терминов «корреляционная связь» и «корреляционная зависимость»?
4. Сформулируйте основные задачи корреляционного анализа.
2.Приведите пример положительной корреляционной связи между психологическими признаками. Ответ обоснуйте.
6. Приведите пример линейной положительной корреляционной связи между психологическими признаками. Ответ обоснуйте.

8.Приведите пример линейной отрицательной корреляционной связи между психологическими признаками. Ответ обоснуйте.
9.Сформулируйте свойства коэффициента корреляции Пирсона.
10.Какой вывод делает исследователь, если выборочный коэффициент корреляции Пирсона равен: 1) ; 2); 3)?
1. Коэффициент корреляции является мерой …

1) статистической связи между случайными величинами
2) вероятностной связи между случайными величинами
3) корреляционной связи между случайными величинами
4) линейной связи между случайными величинами
2. Высокому уровню линейной связи между переменными соответствует значение коэффициента линейной корреляции ….

1) по модулю близкое к нулю
2) по модулю близкое к единице
3) положительное
4) больше 1
3. Выборочный коэффициент линейной корреляции …

1) безразмерная величина
2) зависит от единиц измерения признаков X *и* Y
3) имеет размерность, совпадающую с размерностью признаков X *и* Y
4) имеет размерность произведения размерностей признаков X *и* Y
***4.** Отметьте не менее двух правильных ответов. Задачи корреляционного анализа:*

1) установление направления корреляционной связи
2) установление формы корреляционной связи
3) измерение тесноты корреляционной связи
4) нахождение уравнения регрессии
***5.** Выборочный коэффициент линейной корреляции принимает значения ...*

1) от –1 до –1
2) больше нуля
3) положительные и отрицательные
4) от нуля до единицы
***6.** Отметьте не менее двух правильных ответов. Проверка значимости выборочного коэффициента корреляции основана на статистической проверке гипотезы:*

1) в генеральной совокупности отсутствует корреляция
2) отличие от нуля выборочного коэффициента корреляции объясняется только случайностью
выборки
3) коэффициент корреляции значимо отличается от 0
4) отличие от нуля выборочного коэффициента корреляции не случайно
*Ответы. **1.** 3. **2.** 2. **3.** 1. **4.** 1, 2, 3. **5.** 1. **6.** 1, 3.*

3. Решить следующие задачи
Пример 3. По выборке n = 100 извлеченной из двумерной нормальной генеральной совокупности (X, Y), найден выборочный коэффициент корреляции r = 0,2. По уровню значимости 0,05 проверить нулевую гипотезу о равенстве нулю генерального коэффициента корреляции при конкурирующей гипотезе H1: $r \neq 0$
Решение: найдём наблюдаемое (эмпирическое) значение критерия:

$$t_{эмп}=|r|\sqrt{\frac{n-2}{1-r^2}}=0,2\sqrt{\frac{100-2}{1-0,2^2}}=2,02.$$

Конкурирующая гипотеза имеет вид $r\neq 0$ поэтому критическая область – двусторонняя. По таблице прил. 2 при уровне значимости $\alpha=0,05$ и числе степеней свободы $k= n-2=98$ находим критическую точку двусторонней критической области $t_{кр}(0,05;\ 98)=1,99$. Так как $t_{эмп}>t_{кр}$ – отвергаем нулевую гипотезу о равенстве нулю генерального коэффициента корреляции. Другими словами, коэффициент корреляции значимо отличается от нуля. Таким образом, X и Y коррелированны.

Пример 4. В таблице представлены результаты испытаний двух случайных величин X и Y (n=20).

i	x_i	y_i	i	x_i	y_i
1	4,08	2,14	11	4,31	6,29
2	6,91	3,00	12	2,34	5,52
3	7,42	1,73	13	3,82	3,11
4	3,58	4,24	14	3,98	5,70
5	5,16	3,27	15	3,24	2,60
6	5,19	2,83	16	2,88	5,13
7	4,10	4,22	17	6,19	1,44
8	5,37	4,40	18	5,86	2,20
9	5,02	2,19	19	2,67	3,58
10	6,19	3,20	20	4,36	3,90

Требуется определить выборочный коэффициент корреляции и проанализировать результаты.

Решение: вычислим необходимые для нахождения выборочного коэффициента корреляции оценки параметров распределения генеральной совокупности.

Номер п/п	x_i	y_i	x_i^2	y_i^2	$x_i y_i$
1	4,08	2,14	16,6464	4,5796	8,7312
2	6,91	3,00	47,7481	9	20,73
3	7,42	1,73	55,0564	2,9929	12,8366
4	3,58	4,24	12,8164	17,9776	15,1792
5	5,16	3,27	26,6256	10,6929	16,8732
6	5,19	2,83	26,9361	8,0089	14,6877
7	4,10	4,22	16,81	17,8084	17,302
8	5,37	4,40	28,8369	19,36	23,628
9	5,02	2,19	25,2004	4,7961	10,9938
10	6,19	3,20	38,3161	10,24	19,808
11	4,31	6,29	18,5761	39,5641	27,1099
12	2,34	5,52	5,4756	30,4704	12,9168
13	3,82	3,11	14,5924	9,6721	11,8802
14	3,98	5,70	15,8404	32,49	22,686
15	3,24	2,60	10,4976	6,76	8,424
16	2,88	5,13	8,2944	26,3169	14,7744
17	6,19	1,44	38,3161	2,0736	8,9136
18	5,86	2,20	34,3396	4,84	12,892
19	2,67	3,58	7,1289	12,8164	9,5586
20	4,36	3,90	19,0096	15,21	17,004
Σ	92,67	70,69	467,0631	285,6699	306,9292

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i=\frac{92,67}{20}=4,6335;\quad \bar{y}=\frac{1}{n}\sum_{i=1}^{n}y_i=\frac{70,69}{20}=3,5345;$$

$$\overline{xy}=\frac{1}{n}\sum_{i=1}^{n}x_i y_i=\frac{1}{20}\cdot 306,9292=15,3465.$$

Теперь вычислим значения выборочных средних квадратических отклонений:

$$S_x=\sqrt{\overline{x^2}-(\bar{x})^2}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}{x_i}^2-(\bar{x})^2}=\sqrt{\frac{1}{20}\cdot 467,0631-4,6335^2}=1,3725;$$

$$S_y=\sqrt{\overline{y^2}-(\bar{y})^2}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}{y_i}^2-(\bar{y})^2}=\sqrt{\frac{1}{20}\cdot 285,6699-15,3465^2}=1,3381.$$

Подставим в формулу:

$$r = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{S_x \cdot S_y} = \frac{15{,}3465 - 4{,}6335 \cdot 3{,}5345}{1{,}3725 \cdot 1{,}3381} \approx -0{,}5611.$$

Корреляционная связь между случайными величинами заметная. Так как выборочный коэффициент корреляции отрицателен, то при возрастании одной случайной величины другая имеет тенденцию в среднем убывать. Найдем наблюдаемое (эмпирическое) значение критерия:

$$t_{\text{эмп}} = |r| \sqrt{\frac{n-2}{1-r^2}} = 0{,}5611 \sqrt{\frac{20-2}{1-0{,}5611^2}} = 2{,}87.$$

Конкурирующая гипотеза имеет вид $r \neq 0$ поэтому критическая область – двусторонняя. По таблице прил. 2 при уровне значимости $\alpha=0{,}05$ и числе степеней свободы $k=n-2=18$ находим критическую точку двусторонней критической области $t_{кр}(0{,}05;18)=2{,}10$. Так как $t_{эмп}>t_{кр}$ – отвергаем нулевую гипотезу о равенстве нулю генерального коэффициента корреляции. Другими словами, коэффициент корреляции значимо отличается от нуля. Таким образом, с уверенностью 95 % можем полагать, что между рассматриваемыми числовыми совокупностями существует корреляционная связь. Заметим, что наличие корреляционной связи можно утверждать даже при уровне значимости $\alpha = 0{,}02$, т. к. в этом случае $t_{кр}(0{,}02;18)=2{,}55$.

Пример 5. Определить достоверность взаимосвязи между показателями веса студента и максимального количества сгибания и разгибания рук в упоре лежа N у 10 исследуемых с помощью расчета рангового коэффициента корреляции, если данные выборок таковы:

85	75	73	77	77	81	78	90	83	80
26	20	25	22	27	28	16	15	18	24

Решение: составим вспомогательную таблицу, в которой результаты приведены в виде ранжированного ряда по весу студента.

Вес	85	75	73	77	77	81	78	90	83	80
d_x	**9**	**2**	**1**	**3,5**	**3,5**	**7**	**5**	**10**	**8**	**6**
N	26	20	25	22	27	28	16	15	18	24
d_y	**9**	**4**	**7**	**5**	**8**	**10**	**2**	**1**	**3**	**6**
$d_i = d_x - d_y$	0	–2	–6	–1,5	–4,5	–3	3	9	5	0

Заметим, т. к. четвертый и пятый испытуемый имеют равный вес, то им присваивается одинаковый ранг, равный 3,5.

$$\sum d_i^2 = 4 + 36 + 2{,}25 + 20{,}25 + 9 + 9 + 81 + 25 = 186{,}5.$$

Тогда выборочный коэффициент ранговой корреляции Спирмена

$$r_S = 1 - \frac{6 \cdot 186{,}5}{1000 - 10} = -0{,}13.$$

r

Так как $r_S = -0{,}13<0$, то между данными выборок наблюдается прямая отрицательная взаимосвязь, т. е. увеличение показателей веса вызывает снижение максимального количества сгибаний и разгибаний рук в упоре лежа в группе исследуемых. Проверим значимость найденного рангового коэффициента корреляции. Найдем критические значения коэффициента ранговой корреляции Спирмена по таблице прил. 7 для $\alpha=0{,}05$ и $n=10$:

$$(r_S)_{\text{кр}} = 0{,}64.$$

Значение выборочного коэффициента ранговой корреляции $|r_S| = 0{,}13$ меньше значения $(r_S)_{\text{крит}} = 0{,}64$. Это говорит о том, что значение $r_S = -0{,}13$ не попало в область значимости коэффициента корреляции.

Таким образом, с вероятностью 95 % можно утверждать, что выявленная зависимость недостоверна. Для уточнения результатов следовало бы повысить число исследуемых (увеличить объем выборки), а при отсутствии такой возможности высказанные оценки следует воспринимать с определенной осторожностью. Итак: 1) увеличение показателей веса вызывает снижение максимального количества сгибаний и разгибаний рук в упоре лежа в группе исследуемых; 2) с уверенностью 95 % можно говорить о том, что выявленная зависимость недостоверна.

Пример 6. 14 жителям города N были заданы два вопроса: «Считаете ли Вы, что развитие космонавтики необходимо?» и соответственно «Согласились бы Вы предоставить себя в распоряжение ученым для научных экспериментов?» (с кодировками 0 = да и 1 = нет). Результаты анкетирования представлены в таблице.

Вопрос		1		Итого
		0	1	
2	0	$a=9$	$b=5$	$a+b=14$
	1	$c=3$	$d=11$	$c+d=14$
	Итого:	$a+c=12$	$b+d=16$	$n=28$

Оценить степень зависимости желания прогресса от личного здоровья.

Решение: чтобы воспользоваться формулой для вычисления коэффициента ассоциации, переведем исходные данные в таблицу сопряженности.

$$Q=\frac{ad-bc}{ad+bc}=\frac{9\cdot 11-3\cdot 5}{9\cdot 11+3\cdot 5}=\frac{84}{114}\approx 0{,}737;$$

$$\varphi=\frac{ad-bc}{\sqrt{(b+d)(a+c)(c+d)(a+b)}}=\frac{9\cdot 11-3\cdot 5}{\sqrt{16\cdot 12\cdot 14\cdot 14}}=\frac{84}{194}\approx 0{,}433.$$

Можно утверждать, что в генеральной совокупности присутствует взаимная связь. Иными словами, существует значимая связь между желанием космических исследований и риском личного здоровья горожан.

Пример 7. 220 студентов были опрошены на предмет курения и употребления алкоголя. Результаты анкетирования представлены в таблице.

Признак		*X*		Итого
		Пью	Не пью	
Y	Курю	$a=40$	$b=60$	$a+b=100$
	Не курю	$c=80$	$d=40$	$c+d=120$
	Итого:	$a+c=120$	$b+d=100$	$n=220$

Оценить степень зависимости курения от употребления алкоголя.

Решение:

$$Q=\frac{ad-bc}{ad+bc}=\frac{40\cdot 40-60\cdot 80}{40\cdot 40+60\cdot 80}=-0{,}5;$$

$$\varphi=\frac{ad-bc}{\sqrt{(b+d)(a+c)(c+d)(a+b)}}=\frac{-3200}{12000}\approx -0{,}267.$$

Можно утверждать, что в генеральной совокупности отсутствует взаимная связь. Иными словами, не существует значимая связь между курением и употреблением алкоголя.

Формы контроля освоения заданий по самостоятельной контактной/внеконтактной работы обучающихся по данной теме: тестовые задания и контрольные вопросы.

Рекомендуемая литература:

См. в приложении.

ПРИЛОЖЕНИЕ

Перечень основной и дополнительной литературы, необходимой для освоения учебной дисциплины (модуля)

	Основная литература	
	Обмачевская, С. Н. Медицинская информатика. Курс лекций : учебное пособие для вузов / С. Н. Обмачевская. — 3-е изд., стер. — Санкт-Петербург : Лань, 2021. — 184 с. — ISBN 978-5-8114-7053-2. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/154391	Неограниченный доступ
	Зарубина, Т. В. Медицинская информатика : учебник / Зарубина Т. В. [и др.] - Москва : ГЭОТАР-Медиа, 2018. - 512 с. - ISBN 978-5-9704-4573-0. - Текст : электронный // ЭБС "Консультант студента" : [сайт]. - URL : https://www.studentlibrary.ru/book/ISBN9785970445730.html	Неограниченный доступ
	Омельченко, В. П. Информатика, медицинская информатика, статистика : учебник / В. П. Омельченко, А. А. Демидова. - Москва : ГЭОТАР-Медиа, 2021. - 608 с. - ISBN 978-5-9704-5921-8. - Текст : электронный // ЭБС "Консультант студента" : [сайт]. - URL : https://www.studentlibrary.ru/book/ISBN9785970459218.html	Неограниченный доступ
	Царик, Г. Н. Информатика и медицинская статистика / под ред. Г. Н. Царик - Москва : ГЭОТАР-Медиа, 2017. - 304 с. - ISBN 978-5-9704-4243-2. - Текст : электронный // ЭБС "Консультант студента" : [сайт]. - URL : https://www.studentlibrary.ru/book/ISBN9785970442432.html	Неограниченный доступ
	Дополнительная литература	
	Диденко Г. А. Теоретические основы медицинской информатики / Г. А. Диденко, А. А. Мукашева, О. А. Степанова. - Челябинск : ЮУГМУ, 2017. - 175 с. - Текст : электронный // ЭБС "Букап" : [сайт]. - URL : https://www.books-up.ru/ru/book/teoreticheskie-osnovy-medicinskoj-informatiki-15045004/	Неограниченный доступ
	Медицинская информатика : учебное пособие / Н. В. Маркина, Г. А. Диденко, А. А. Мукашева и др. - Челябинск : ЮУГМУ, 2017. - 145 с. - Текст : электронный // ЭБС "Букап" : [сайт]. - URL : https://www.books-up.ru/ru/book/medicinskaya-informatika-11851468/	Неограниченный доступ
	Медицинская информатика: параметрические и непараметрические методы статистики на компьютере / Н. В. Маркина, Э. И. Беленкова, Г. А. Диденко и др. - Челябинск : ТЕТА, 2022. - 138 с. - Текст : электронный // ЭБС "Букап" : [сайт]. - URL : https://www.books-up.ru/ru/book/medicinskaya-informatika-parametricheskie-i-neparametricheskie-metody-statistiki-na-kompyutere-15440733/	Неограниченный доступ
	Семенова О. Л. Медицинская информатика: в 2 ч. Часть 1 : учебное пособие / О. Л. Семенова, Н. Ю. Часовских, А. Ю.	Неограниченный доступ

	Гречишникова. - Томск : Издательство СибГМУ, 2021. - 79 с. - Текст : электронный // ЭБС "Букап" : [сайт]. - URL : https://www.books-up.ru/ru/book/medicinskaya-informatika-v-2-ch-chast-1-12564392/	
	Статистические методы в медицине и здравоохранении [Электронный ресурс] : учеб. пособие / ФГБОУ ВО «Баш. гос. мед. ун-т» МЗ РФ ; сост. Н. Х. Шарафутдинова [и др.]. - Электрон. текстовые дан. - Уфа, 2018. - Текст: электронный // БД «Электронная учебная библиотека» .- **URL:** http://library.bashgmu.ru/elibdoc/elib719.pdf	Неограниченный доступ
	Таллер В. А. Медицинская информатика / В. А. Таллер. - Витебск : ВГМУ, 2019. - 225 с. - ISBN 9789854669809. - Текст : электронный // ЭБС "Букап" : [сайт]. - URL : https://www.books-up.ru/ru/book/medicinskaya-informatika-12137206/	Неограниченный доступ
	ЭБС "Букап"	https://www.books-up.ru/
	Электронно-библиотечная система «Консультант студента» для ВПО	www.studmedlib.ru
	База данных «Электронная учебная библиотека»	http://library.bashgmu.ru

Перечень ресурсов информационно-телекоммуникационной сети «интернет», необходимых для освоения учебной дисциплины (модуля)

1. https://www.medicinform.net/ (Медицинская информационная сеть)
2. https://www.studentlibrary.ru/ (Консультант студента)