**МИНИСТРЕСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ**

**федеральное государственное бюджетное образовательное учреждение высшего профессионального образования**

**«МОСКОВСКИЙ ПЕДАГОГИЧЕСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»**

**(МПГУ)**

 **ИНСТИТУТ ФИЗИЧЕСКОЙ КУЛЬТУРЫ, СПОРТА И ЗДОРОВЬЯ**

РАБОЧАЯ ПРОГРАММА

**КОМПЬЮТЕРНАЯ ОБРАБОТКА ДАННЫХ ЭКСПЕРИМЕНТАЛЬНОГО ИССЛЕДОВАНИЯ**

Код и наименование направления подготовки:

49.03.02 Физическая культура для лиц с отклонениями в состоянии здоровья (Адаптивная физическая культура)

Квалификация (степень) выпускника: БАКАЛАВР

Форма обучения: ОЧНАЯ

Москва 2016 **Введение**

Цели и задачи дисциплины «Компьютерная обработка данных экспериментального исследования»:

Дисциплина демонстрирует возможности использования математических методов в экспериментальных исследованиях. Основной целью является практическое освоение математических методов обработки экспериментальных данных, а также получение навыков использования компьютерных пакетов для статистической обработки данных.

Требования к уровню освоения программы:

В соответствии с требованиями ФГОС 3+ к уровню подготовки бакалавра и реализацией перечисленных выше целей и задач предполагается формирование следующих основных знаний и умений:

* формирование навыков использования аппарата математической статистики для обработки экспериментальных данных;
* изучение теоретических основ проверки статистических гипотез;
* практическое освоение основных математических методов обработки экспериментальных данных;
* знакомство с компьютерными пакетами статистической обработки данных.

В результате изучения дисциплины «Компьютерная обработка данных экспериментального исследования» у студентов происходит формирование **профессиональной компетенции ПК-26**, в соответствии с ФГОС 3+, а именно:

«**Способность** проводить обработку результатов исследований с использованием методов математической статистики, информационных технологий, формулировать и представлять обобщения и выводы».

Взаимосвязь с другими дисциплинами ООП:

Освоение курса «Компьютерная обработка данных экспериментального исследования» предполагает хорошую подготовку по ряду дисциплин математического, естественнонаучного и общепрофессионального циклов, а именно:

1. общие математические и естественнонаучные дисциплины - математика и информатика;
2. общепрофессиональные дисциплины - основы психодиагностики, педагогики и психологии.

**1 Объем дисциплины, виды учебной работы и формы контроля**

Таблица 1.1 – Дневная форма обучения

|  |  |  |
| --- | --- | --- |
| Вид учебной работы | Всего часов | Часов по семестрам |
| 5 семестр |
| Аудиторные занятия:- лекции;- практические занятия;- семинары;- лабораторные работы. | 36 | 34 |
| 18 | 18 |
| 18 | 18 |
|  |  |
|  |  |
| Самостоятельная работа:- курсовой проект;- расчетно-графическая работа;- реферат;- контрольная работа;- освоение и работа с ППС- прочее | 72 | 72 |
|  |  |
|  |  |
|  |  |
|  |  |
| 72 | 72 |
|   |   |
| Всего | 108 | 108 |
| Вид итогового контроля | зачет | зачет |

**2 Содержание дисциплины**

Таблица 2.1 – Содержание теоретических занятий

| Тема теоретических занятий | Трудоемкость в часах |
| --- | --- |
| Очная форма | Заочная форма |
| ауд. | СРС | ауд. | СРС |
| 1. Основные понятия математической статистики. Генеральная совокупность и выборка из генеральной совокупности. Понятие выборочного метода.
2. Вариационный ряд, статистический ряд, гистограмма, полигон частот.
3. Числовые выборочные характеристики: среднее, дисперсия, мода, медиана, коэффициенты ассиметрии и эксцесса.
4. Корреляционный анализ: основные понятия. Различные коэффициенты корреляции. Выбор коэффициента корреляции в зависимости от измерительных шкал
5. Теоретические основы проверки статистических гипотез. Нулевая и альтернативная гипотезы. Критерии для проверки гипотез. Понятие статистики критерия, уровня значимости, критической области.
6. Проверка значимости коэффициентов корреляции.
7. Непараметрические критерии: критерий знаков, критерий Вилкоксона, критерий Манна-Уитни.
8. Однофакторный дисперсионный анализ
9. Многомерный статистический анализ
 | 1224342 | 333334512 |  |  |
| **Всего** | 18 | 36 |  |  |

Таблица 2.2 - Содержание практических занятий

| Тема практических занятий | Трудоемкость в часах |
| --- | --- |
| Очная форма | Заочная форма |
| ауд. | СРС | ауд. | СРС |
| 1. Способы представления выборки: вариационный ряд, статистический ряд, группированный статистический ряд, гистограмма
2. Числовые выборочные характеристики
3. Корреляционный анализ. Вычисление коэффициентов корреляции Пирсона, Спирмена, точечно-бисериальной корреляции, корреляции знаков, ассоциации. Проверка их значимости
4. Непараметрические критерии. Критерии сдвига: критерий знаков, критерий Вилкоксона для двух зависимых выборок, критерий Фридмана для нескольких зависимых выборок
5. Критерий Манна-Уитни для двух независимых выборок. Критерий Крускала-Уоллиса для нескольких независимых выборок
6. Многофункциональный критерий Фишера
 | 333333 | 668565 |  |  |
| **Всего** | 18 | 36 | 0 | 0 |
| **Итого:** | 36 | 72 |  |  |

 **3 Учебно-методическое обеспечение**

***3.1 Список рекомендуемой литературы***

*3.1.1 Основная литература*

1 Е. Сидоренко "Методы математической обработки в психологии",

С-Петербург, "Речь", 2001.

1. В. Боровиков. Statistica. Искусство анализа данных на компьютере: для профессионалов. СПб.: Питер,- 2003.
2. А. Гмурман “Руководство по решению задач теории вероятностей и

математической статистики”, Москва, “Высшая школа”, 1999.

3 М. Грабарь, К. Краснянская, “Применение математической статистики

в педагогических исследованиях”, Москва¸ “Педагогика”, 1977.

1. Гласс Д., Стенли Д. Статистические методы в педагогике и психологии. М.: Прогресс, 1976. – 495 с.

*3.1.2 Дополнительная литература*

1. А.Д. Наследов, «Математические методы психологического исследования». СПб. :Речь. -2004.
2. В. П**.** Боровиков.**,** Г. И. Ивченко Учебник по математической статистике с упражнениями в системе Statistica.**.** М.: StatSoft, 2003.

3 М. Грабарь, К. Краснянская, “Применение математической статистики

в педагогических исследованиях”, Москва¸ “Педагогика”, 1977.

4 Ю. Тюрин, А. Макаров, “Анализ данных на компьютере”, Москва,

“Финансы и статистика”, 1995.

* 1. ***Список методических рекомендаций и указаний***
		1. *Список методической литературы к лабораторным работам*

1 Е. Сидоренко "Методы математической обработки в психологии",

С-Петербург, "Речь", 2001.

1. <http://www.exponenta.ru/soft/Statist/Statist.asp>
	1. ***Рекомендации по использованию информационных технологий***

Программное обеспечение для поддержки курса «**«**Компьютерная обработка данных экспериментального исследования**»**» включает в себя следующие программно-методические комплексы:

1. Приложение Excel
2. Приложение Statistica
3. Приложение Studia

 Программное обеспечение работает под управлением операционной системы ОС Windows.

* 1. ***Педагогические контрольные (испытательные) материалы***

*3.4.1 Методические материалы для тестирования знаний студентов*

1 Варианты контрольных работ

*Задания контрольных работ*

Задание 1

Группа испытуемых (студентов физфака СпбГУ) проходила подготовку перед началом работы на тренажере. Испытуемые должны были решать задачи по выбору оптимального типа взлетно-посадочной полосы для заданного типа самолета. Следующая таблица показывает кол-во допущенных ошибок в тренировочной сессии, а также уровни вербального и невербального интеллекта, измеренные по шкале Векслера.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| № | **Испытуемый** | Кол-во ошибок | Показатель верб. интеллекта | Показатель неверб. интеллекта |
|  1 | И.К. | 29 | 131 | 106 |
|  2 | Р.О. | 54 | 132 | 90 |
|  3 | И.С. | 13 | 121 | 95 |
|  4 | Л.П. | 8 | 127 | 116 |
|  5 | Н.О. | 14 | 136 | 127 |
|  6 | Е.Р. | 26 | 124 | 107 |
|  7 | Г.Е. | 9 | 134 | 104 |
|  8 | О.Р. | 20 | 136 | 102 |
|  9 | Н.Ш. | 2 | 132 | 111 |
|  10 | Г.К. | 17 | 136 | 99 |

а) Вычислить коэффициенты Пирсона и Спирмена между кол-вом ошибок и уровнем вербального интеллекта и проверить их значимость при уровне значимости α=0.05.

б) Вычислить коэффициенты Пирсона и Спирмена между кол-вом ошибок и уровнем невербального интеллекта и проверить их значимость при уровне значимости α=0.05.

Задание 2

1. Перед началом игры двум группам детей были показаны два мультфильма: группе А мультипликационный боевик и группе В сказка. После этого группа исследователей наблюдала поведение детей и оценивала их уровень арессивности (количество агрессивных поступков). Результаты были представлены в порядковой шкале:

|  |  |
| --- | --- |
|  А | 26 22 19 21 14 18 29 17 11 34 |
|  В | 16 10 8 13 19 11 13 9 21 |

При уровне значимости α=0.05 с помощью критерия Манна-Уитни проверить гипотезу о влиянии просмотренных мультфильмов на уровень агрессивности детей.

Задание 3

Для следующей выборки определить выборочные :

 а) среднее; б) дисперсию; в) моду; г) медиану;

 д) получить вариационный ряд; е) получить статистический ряд.

 7 3 3 6 4 5 1 2 1 3 4 5 1 2 1 3

Задание 4

Для заданной выборки выполнить:

 а) группировку с числом интервалов =7;

 б) найти выборочные среднее и дисперсию;

 г) построить гистограмму.

 20,3 15,4 17,2 19,2 23,3 18,1 21,0 15,3 16,8 13,2 20,4 16,5 19,7 20,5

 14,3 20,1 16,8 14,7 20,8 19,5 15,3 19,3 17,8 16,2 15,7 22,8 21,9 12,5

 10,1 21,1 18,3 14,7 14,5 18,1 18,4 13,9 19,1 18,5 20,2 23,8 16,7 20,4

**ПРИЛОЖЕНИЕ А**

**Карта учебно-методического обеспечения**

**Дисциплины**  **«**Компьютерная обработка данных экспериментального исследования**»**

**Для направления (специальности) 49.03.02 Физическая культура для лиц с отклонениями в состоянии здоровья (адаптивная физическая культура)**

**Форма обучения** очная

**Всего часов** 108 (3ЗЕ) из них: лекций 18 часов, практических занятий 18 часов, СРС 72 часа

**Вид индивидуальной работы**  консультации

**Семестр** 5

**Таблица 1** Обеспечение дисциплины учебными изданиями

| Библиографическое описание издания (автор, наименование, вид, место и год издания, кол. стр.) | Вид занятия, в котором используется | Число часов, обеспечиваемое изданием | Кол. экз. в библ. (на каф.) | Примечание |
| --- | --- | --- | --- | --- |
| 1 Е. Сидоренко "Методы математической обработки в психологии",С-Петербург, "Речь", 2001. | ЛекцииПрактические занятия | 1818 | 30 |  |

**Таблица 2** Обеспечение дисциплины учебно-методическими изданиями

| Библиографическое описание издания (автор, наименование, вид, место и год издания, кол. стр.) | Вид занятия, в котором используется | Число часов, обеспечиваемое изданием | Кол. экз. в библ. (на каф.) | Примечание |
| --- | --- | --- | --- | --- |
| 1.Рабочая программа по дисциплине «Математические методы в психологии» для студентов специальности 020400 психология2.Распечатки заданий к лабораторным и практическим занятиям1. <http://www.exponenta.ru/soft/Statist/Statist.asp>
 | ЛекцииПрактические занятияЛекцииПрактические занятияЛекцииПрактические занятия | 181818181818 | 110 | Есть наКафедре |

Учебно-методическое обеспечение дисциплины 100%

Действительно для 20\_\_\_/20\_\_\_\_ учебного года

**Вопросы к зачету:**

1. Измерительные шкалы.
2. Понятие выборки. Представление выборок.
3. Графическое представление выборок.
4. Меры центральной тенденции
5. Меры изменчивости
6. Меры связи. Коэффициент корреляции Пирсона.
7. Меры связи. Коэффициент ранговой корреляции Спирмена.
8. Меры связи. Коэффициент ассоциации.
9. Общая схема применения и ограничения корреляционно-регрессионного анализа.
10. Основные понятия теории статистического вывода (подход Неймана-Пирсона к проверке статистических гипотез, понятие критерия, критической области, ошибок 1-го и 2-го рода, априорной модели ситуации)
11. Основные понятия теории статистического вывода (схема проверки статистических гипотез в психологии).
12. Виды критериев.
13. Параметрические критерии. Примеры.
14. Непараметрические критерии. Примеры.
15. Многофункциональные критерии. Пример.

**Задания для самостоятельной работы студентов**

***Для очной формы обучения***

**Теоретическая часть (36часов):**

1. Теоретические основы проверки статистических гипотез. Нулевая и альтернативная гипотезы. Критерии для проверки гипотез. Понятие статистики критерия, уровня значимости, критической области. Проверка значимости коэффициентов корреляции.
2. Непараметрические критерии: критерий знаков, критерий Вилкоксона, критерий Манна-Уитни.
3. Однофакторный дисперсионный анализ
4. Многомерный статистический анализ

***Литература:***

1. А.Д. Наследов, «Математические методы психологического исследования». СПб. :Речь. -2004.

**Практическая часть (36 часов):**

*Выполняется с использованием пакета Statistica и Excel*

**Задачи для самостоятельного решения:**

* 1. Определите верхнюю и нижнюю квартиль, выборочную асимметрию и эксцесс для данных измерений роста групп студентов: 164, 160, 157, 166, 162, 160, 161, 159, 160, 163, 170, 171.
	2. Найти наиболее популярный туристический маршрут из четырех реализуемых фирмой, если за неделю последовательно были реализованы следующие маршруты: 1, 3, 3, 2, 1, 1, 4, 4, 2, 4, 1, 3, 2, 4, 1, 4, 4, 3, 1, 2, 3, 4, 1, 1, 3.
	3. В рабочей зоне производились замеры концентрации вредного вещества. Получен ряд значений (в мг./м3): 12, 16, 15, 14, 10, 20, 16, 14, 18, 14, 15, 17, 23, 16. Необходимо определить основные выборочные характеристики.
	4. Проверяется гипотеза о том, что изучение курса математики способствует формированию у учеников приемов логического мышления. Для этого был проведен следующий эксперимент. Ученикам 7 класса давали по 5 задач, использующих одни и те же приемы логического мышления в конце сентября и в конце мая следующего года. Работа предлагалась одним и тем же учащимся. Получились следующие результаты:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Число решеныхзадач в сентябре |  1 |  1 |  1 |  2  |  0 |  0 |  0 |
| Число решеных задач в мае |  1 |  0 |  0 |  2 |  1 |  0 |  1 |

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  0 |  0 |  1 |  0 |  3 |  2 |  1 |  0 |  0 |  1 |  0 |  1 |  0 |
|  0 |  0 |  1  |  1 |  3 |  2 |  0 |  1 |  1 |  1 |  1 |  1 |  1 |

Проверить, отвергается или принимается указанная гипотеза при уровне значимости α=0.05.

***Литература:***

1. В. Боровиков. Statistica. Искусство анализа данных на компьютере: для профессионалов. СПб.: Питер,- 2003.
2. <http://www.exponenta.ru/soft/Statist/Statist.asp>

**Методические рекомендации по преподаванию предмета**

**«Компьютерная обработка данных экспериментального исследования»**

Курс «Компьютерная обработка данных экспериментального исследования» входит в блок дисциплин,направления (специальности) 49.03.02 Физическая культура для лиц с отклонениями в состоянии здоровья (адаптивная физическая культура).

*Цели обучения*

Изучение курса **«**Компьютерная обработка данных экспериментального исследования**»** способствует формированию у студентов общей исследовательской культуры, что позволяет правильно описывать интересующие явления, точнее мыслить и точнее выражаться, обобщать результаты наблюдений и исследований, представлять эти результаты в удобном для понимания виде, делать более точные выводы, предсказывать результаты и находить причины, нередко скрытые от наблюдателя. Кроме того, изучение курса способствует лучшему пониманию профессиональной литературы, где часто используются понятия теории вероятностей, статистики, умению разработать свои варианты тестов, проводить статистическую обработку тестовых результатов, создавать тестовый профиль испытуемых, находить меру связи и влияния между различными переменными характеристиками

*Задачи обучения*

* повышение математической культуры студентов и приобретение знаний, умений и навыков статистической обработки данных исследований.
* развитие способности системно мыслить, умение перерабатывать большие объемы информации и вычленять главное;
* получение навыков работы с компьютерными пакетами статистической обработки данных.

Построение курсов математических дисциплин должно отвечать требованиям конкретного направления и не быть стандартным не только по составу математических дисциплин, но и по конкретному содержанию каждой дисциплины. Преподаватели курса должны иметь ясное представление о направлении, на котором обучаются студенты и тому, какие именно математические знания и навыки могут понадобиться будущим бакалаврам в их практической деятельности.

 Целью любого педагогического эксперимента является эмпирическое подтверждение или опровержение гипотезы исследования и/или справедливости теоретических результатов, то есть обоснование того, что предлагаемое педагогическое воздействие (например, новые содержание, формы, методы, средства обучения и т.д.) более эффективно (или, возможно, наоборот – менее эффективно). Для этого, как минимум, необходимо показать, что, будучи примененным к тому же объекту (например – к группе учащихся), оно дает другие результаты, чем применение традиционных педагогических воздействий.
Для этого выделяется экспериментальная группа, которая сравнивается с контрольной группой. Различие эффектов педагогических воздействий будет обосновано, если две эти группы, первоначально совпадающие по своим характеристикам, различаются после реализации педагогических воздействий. Следовательно, требуется провести два сравнения и показать, что при первом сравнении (до начала педагогического эксперимента) характеристики экспериментальной и контрольной группы совпадают, а при втором (после окончания эксперимента) – различаются.

 Так как объектом педагогического эксперимента, как правило, являются люди (учащиеся, учителя, сотрудники и руководители органов управления образованием и т.д.), а каждый человек индивидуален, то говорить о совпадении или различии характеристик экспериментальной и контрольной групп можно лишь в чисто формальном, статистическом смысле. Для того, чтобы выяснить, являются ли совпадения или различия случайными, используются статистические методы, которые позволяют на основании данных, полученных в результате эксперимента, принять обоснованное решение о совпадениях или различиях.

Развитие информационных технологий и вычислительной техники привело к появлению стандартных программ, которые позволяют студентам гуманитарных направлений решать сложные задачи статистической обработки результатов практических исследований. Освоение технологии работы с такими программами является основной задачей курса.

В результате изучения дисциплины «Компьютерная обработка данных экспериментального исследования» у студентов происходит формирование **профессиональной компетенции ПК-26**, в соответствии с ФГОС 3+, а именно:

«**Способность** проводить обработку результатов исследований с использованием методов математической статистики, информационных технологий, формулировать и представлять обобщения и выводы».

***Примерное содержание курса:***

Виды измерительных шкал. Основные понятия математической статистики. Генеральная совокупность и выборка из генеральной совокупности. Понятие выборочного метода. Вариационный ряд, статистический ряд, гистограмма, полигон частот. Числовые выборочные характеристики: среднее, дисперсия, мода, медиана, коэффициенты ассиметрии и эксцесса. Корреляционный анализ: основные понятия. Различные коэффициенты корреляции. Выбор коэффициента корреляции в зависимости от измерительных шкал. Теоретические основы проверки статистических гипотез. Нулевая и альтернативная гипотезы. Критерии для проверки гипотез. Понятие статистики критерия, уровня значимости, критической области. Проверка значимости коэффициентов корреляции. Непараметрические критерии: критерий знаков, критерий Вилкоксона, критерий Манна-Уитни, критерий Фридмана. Однофакторный дисперсионный анализ. Многофункциональный критерий Фишера. Многомерные статистические методы.

***Практическая составляющая предмета.***

Практическая составляющая включает в себя решение задач статистической обработки данных как вручную, так и на компьютере с помощью статистических пакетов.

***Формы организации обучения:***

1. **Бескомпьютерное обучение**. Лекции и практические занятия проводятся без деления на подгруппы.
2. **Организация компьютерной поддержки курса**. Проводятся лекции, практические занятия в группе и лабораторные занятия в компьютерном классе с делением на подгруппы.

Занятия проводятся по 2 часа в неделю – 1 час практических занятий в компьютерном классе и 1 час лекций.

Для реализации этой формы обучения необходимо наличие MS Office, Excel, а также пакета статистической обработки данных (например, Statistica, Studia).

При организации компьютерной поддержки курса необходимо соблюдать требования к оборудованию кабинета вычислительной техникой, санитарные нормы организации труда за компьютером (см. СанПиН 2.2.2.542-96).

Рабочую программу по курсу «Компьютерная обработка данных экспериментального исследования» можно найти на кафедре.

Содержание курса реализовано в объеме 108 часов (3 зачетные единицы ЗЕ). При этом основной акцент должен быть сделан на практическом применении методов статистической обработки данных на конкретных задачах и использовании пакета Excel, Statistica.

Рекомендуемая литература:

1 Е. Сидоренко "Методы математической обработки в психологии",

С-Петербург, "Речь", 2001.

1. В. Боровиков. Statistica. Искусство анализа данных на компьютере: для профессионалов. СПб.: Питер,- 2003.

**Задания для практических занятий**

**1. Способы представления выборок**

**Выборкой** *х1, ..., хn* объема *n* из генеральной совокупности называется *n* значений признака X.

**Вариационным рядом** *х(1)≤ х(2)≤ ... ≤ х(n)* называется выборка, записанная в порядке возрастания ее элементов.

**Статистический ряд.**

Если выбрать из выборки несовпадающие элементы и для каждого указать частоту встречаемости, получаем статистический ряд, который записывают в виде таблицы:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| x | x | x | x | … | x |
| n | n | n | n | … | n |

Сумма частот должна быть равна объему выборки: .

**Группированный статистический ряд.**

При большом объеме выборки производят группировку выборки. Для этого разбивают диапазон значений выборки на равные интервалы и подсчитывают для каждого интервала **частоту**- количество наблюдений, попавших в него. Группированный статистический ряд записывают в виде таблицы:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| x | [a,a] | (a,a] | (a,a] | … | (a,a] |
| n | n | n | n | … | n |

Сумма частот должна быть равна объему выборки: .

a - минимальное значение выборки

a - максимальное значение выборки

Длина интервала вычисляется по формуле: L = , где k-количество интервалов. Длину интервала можно округлять.

**Гистограмма.**

Графическое представление распределения частот по интервалам называют гистограммой. При построении гистограммы на оси ОХ откладывают интервалы разбиения, на каждом интервале строят прямоугольник высотой, равной  , где n - объем выборки,

n-частота для i-того интервала, l - длина интервала. Сумма площадей прямоугольников должна быть равна 1.

Пример гистограммы:

 

 a a a a a

**Задачи:**

1.   Построить выборочные функции распределения (относительные и накопленные частоты) для роста в см. 20 студентов: 181, 169, 178, 178, 171, 179, 172, 181, 179, 168, 174, 167, 169, 171, 179, 181, 181, 183, 172, 176.

2.   Найдите распределение по абсолютным частотам для следующих результатов тестирования в баллах: 79, 85, 78, 85, 83, 81, 95, 88, 97, 85 (используйте границы интервалов 70, 80, 90).

**2. Числовые выборочные характеристики**

**Выборочное среднее:**

* =*

**Выборочная** **медиана -** центральныйчлен вариационного ряда.

Если объем выборки n - нечетно, выборочная медиана равна среднему арифметическому двух центральных элементов вариационного ряда. Обозначение: Me**

**Выборочная мода** - это элемент выборки, имеющий наибольшую частоту встречаемости. Обозначение : Mo**

**Выборочная дисперсия (**мера разброса от среднего)**:**

 *s2=*

Другая формула для вычислений:

*s2==*

**Выборочное средне-квадратическое отклонение** ơ *=*

**Выборочный** **коэффициент ассиметрии**:

*A = ;*

**Выборочный** **коэффициент эксцесса**:

*E = -*3

**Задачи:**

1.       Наблюдение посещаемости четырех внеклассных мероприятий в экспериментальном (20 человек) и контрольном (30 человек) классах дали значения (соответственно): 18, 20, 20, 18 и 15, 23, 10, 28. Требуется найти среднее значение, стандартное отклонение, медиану и квартили этих данных.

2.       Найти среднее значение, медиану, стандартное отклонение и квартили результатов бега на дистанцию 100 м у группы студентов (с): 12,8; 13,2; 13,0; 12,9; 13,5; 13,1.

 **3. Корреляционный анализ. Вычисление коэффициентов корреляции Пирсона, Спирмена, точечно-бисериальной корреляции, корреляции знаков, ассоциации. Проверка их значимости**

**Корреляционная зависимость** двух признаков - это **зависимость в среднем.**

Основным инструментом выяснения корреляционной зависимости является коэффициент корреляции.

Значение коэффициента корреляции лежит между -1 и +1. Если наблюдается тенденция возрастания одной величины при росте другой, то говорят о положительной коррелированности величин, если наблюдается тенденция увеличения одной величины при уменьшении другой, то говорят об отрицательной коррелированности величин. Если значение коэффициента корреляции равно 0, то корреляционная зависимость отсутствует.

Оценку коэффициента корреляции дает **выборочный коэффициент корреляции**. По его значению можно проверить, равен истинный коэффициент корреляции 0, или нет (проверка значимости коэффициента корреляции).

Выбор коэффициента корреляции определяется типом измерительных шкал для обоих признаков:

|  |  |  |
| --- | --- | --- |
| Измерительная шкала для1-го признака | Измерительная шкала для2-го признака | Коэффициент корреляции |
| Количественная (интервальная или отношений) | Количественная (интервальная или отношений) | Пирсона |
| Количественная или порядковая | Количественная или порядковая | Спирмена |
| Качественная с двумя значениями | Количественная или порядковая | Точечно- бисериальный |
| Качественная с двумя значениями | Качественная с двумя значениями | Ассоциации (четырехпольный)  |

Замечание: коэффициент корреляции Пирсона исследует только линейную корреляционную зависимость и используется только в случае, когда оба признака имеют нормальное распределение.

**Формулы для вычисления выборочных коэффициентов корреляции:**

1. *Коэффициент корреляции Пирсона:*

 **

где *и -* средние выборочные для X и Y.

 * и -* средне-квадратические отклонения для X и Y.

 *N -* объем выборки.

1. *Коэффициент корреляции Спирмена:*

 **

где - квадрат разности рангов значений признаков X и Y для i-того испытуемого;  *N -* объем выборки.

**Получение коэффициента корреляции Пирсона в Microsoft Excel**

* Ввести данные в два столбца.
* Нажать на кнопку fX (мастер функций). Выбрать класс функций Статистические. Выбрать функцию КОРРЕЛ или ПИРСОН.
* Выделить мышкой сначала первый столбец (Массив 1), потом второй столбец (Массив 2). Нажать ОК. В ячейке, на которую указывает курсор, появится значение выборочного коэффициента корреляции Пирсона. Значимость коэффициента корреляции проверяется с помощью таблицы критических значений коэффициентов корреляции (см. ниже).

**Проверка значимости коэффициентов корреляции.**

В таблице приведены критические значения коэффициентов корреляции для уровня значимости α=0.05. Объем выборки N= десятки\*10+ единицы. Если абсолютная величина выборочного коэффициента корреляции > критического значения, то коэффициент корреляции значим (есть корреляционная зависимость).

|  |  |
| --- | --- |
|  | единицы |
| десятки |  | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 0 |  |  |  |  |  | 0.94 | 0.85 | 0.78 | 0.72 | 0.68 |
| 1 | 0.64 | 0.61 | 0.58 | 0.56 | 0.54 | 0.52 | 0.50 | 0.48 | 0.47 | 0.46 |
| 2 | 0.45 | 0.44 | 0.43 | 0.42 | 0.41 | 0.40 | 0.39 | 0.38 | 0.38 | 0.37 |
| 3 | 0.36 | 0.36 | 0.36 | 0.34 | 0.34 | 0.33 | 0.33 | 0.33 | 0.32 | 0.32 |
| 4 | 0.31 |  |  |  |  |  |  |  |  |  |

**Задачи**

1.  10 менеджеров оценивались по методике экспертных оценок психологических характеристик личности руководителя (см. Психологические тесты. Т.2. Под ред. А.А. Карелина. - М., ВЛАДОС, 1999, стр. 99). 15 экспертов производили оценку каждой психологической характеристики по пятибалльной системе (см. табл. 3). Психолога интересует вопрос, в какой взаимосвязи находятся эти характеристики руководителя между собой.

|  |  |  |  |
| --- | --- | --- | --- |
| Испытуемые п/п  | тактичность  | требовательность  | критичность  |
| 1  | 70  | 18  | 36  |
| 2  | 60  | 17  | 29  |
| 3  | 70  | 22  | 40  |
| 4  | 46  | 10  | 12  |
| 5  | 58  | 16  | 31  |
| 6  | 69  | 18  | 32  |
| 7  | 32  | 9  | 13  |
| 8  | 62  | 18  | 35  |
| 9  | 46  | 15  | 30  |
| 10  | 62  | 22  | 36  |

2.   Определите, влияет ли фактор образования на уровень зарплаты сотрудников фирмы на основании следующих данных.

|  |  |
| --- | --- |
| Образование  | Зарплата сотрудников  |
| Высшее  | 3200 3000 2600 2000 1900 1900  |
| Среднее спец.  | 2600   2000 2000 1900 1800 1700  |
| среднее  | 2000 2000 1900 1800 1700 1700  |

3.   Определите, влияет ли фактор образования на уровень зарплаты сотрудников фирмы на основании следующих данных

|  |  |
| --- | --- |
| Образование  | Зарплата сотрудников  |
| Высшее  | 3200 3000 2600 2000 1900 1900  |
| Среднее спец.  | 2600   2000 2000 1900 1800 1700  |
| среднее  | 2000 2000 1900 1800 1700 1700  |

**4. Непараметрические критерии. Критерии сдвига: критерий знаков, критерий Вилкоксона для двух зависимых выборок, критерий Манна-Уитни для двух независимых выборок.**

1. Перед началом игры двум группам детей были показаны два мультфильма: группе А мультипликационный боевик и группе В сказка. После этого группа исследователей наблюдала поведение детей и оценивала их уровень агрессивности (количество агрессивных поступков). Результаты были представлены в порядковой шкале:

|  |  |
| --- | --- |
|  А | 26 22 19 21 14 18 29 17 11 34 |
|  В | 16 10 8 13 19 11 13 9 21 |

При уровне значимости α=0.05 с помощью критерия Манна-Уитни проверить гипотезу о влиянии просмотренных мультфильмов на уровень агрессивности детей.

1. Для 10 человек была предложена специальная диета. После двухнедельного питания по этой диете масса их тел изменилась следующим образом:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Масса до диеты | 68 | 80 | 92 | 81 | 70 | 79 | 78 | 66 | 57 | 76 |
| Масса послеДиеты | 60 | 84 | 87 | 79 | 74 | 71 | 72 | 67 | 57 | 70 |

Проверить гипотезу о действии диеты на массу тела по критерию знаков и по критерию Вилкоксона при уровне значимости α=0.05.

**5. Многофункциональный критерий Фишера.**

***Сравнить доли эффектов с помощью многофункционального критерия Фишера.***

1. Проверяется гипотеза о том, что изучение курса математики способствует формированию у учеников приемов логического мышления. Для этого был проведен следующий эксперимент. Ученикам 7 класса давали по 5 задач, использующих одни и те же приемы логического мышления в конце сентября и в конце мая следующего года. Работа предлагалась одним и тем же учащимся. Получились следующие результаты:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| НаписавшиеЗадачи в сентябре |  1 |  1 |  1 |  0  |  0 |  0 |  0 |
| Написавшие Задачи в мае |  1 |  0 |  0 |  1 |  1 |  0 |  1 |

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  0 |  0 |  1 |  0 |  1 |  1 |  1 |  0 |  0 |  1 |  0 |  1 |  0 |
|  0 |  0 |  1  |  1 |  1 |  1 |  0 |  1 |  1 |  1 |  1 |  1 |  1 |

Проверить, отвергается или принимается указанная гипотеза при уровне значимости α=0.05.

**6. Однофакторный дисперсионный анализ**

*Гипотезы:*

H0: различия *между* группами являются не более выраженными, чем случайные различия *внутри* каждой группы

H1: различия *между* группами являются более выраженными, чем случайные различия *внутри* каждой группы.

Последовательность операций в однофакторном дисперсионном анализе для несвязанных выборок:

1. подсчитаем **SSфакт** - вариативность признака, обусловленную действи­ем исследуемого фактора. Часто встречающееся обозначе­ние SS - сокращение от "суммы квадратов" (sum of squares). Это со­кращение чаще всего используется в переводных источниках (см., на­пример: Гласс Дж., Стенли Дж., 1976).

, (1)

где Тс – сумма индивидуальных значений по каждому из условий. Для нашего примера 43, 37, 24 (см. табл. 1);

с – количество условий (градаций) фактора (=3);

n – количество испытуемых в каждой группе (=6);

N – общее количество индивидуальных значений (=18);

- квадрат общей суммы индивидуальных значений (=1042=10816)

Отметим разницу между , в которой все индивидуальные значения сначала возводятся в квадрат, а потом суммируются, и , где индивидуальные значения сначала суммируются для получения об­щей суммы, а потом уже эта сумма возводится в квадрат.

По формуле (1) рассчитав фактическую вариативность признака, получаем:



2. подсчитаем **SSобщ** – общую вариативность признака:

 (2)

3. подсчитаем случайную (остаточную) величину **SSсл**, обусловленную неучтенными факторами:

 (3)

4.      **число степеней свободы** равно:

=3-1=2 (4)

 

5.      **«средний квадрат»** или **математическое ожидание суммы квадратов,** усредненная величина соответствующих сумм квадратов SS равна:

**** (5)

****

6.      значение **статистики критерия** **Fэмп** рассчитаем по формуле:

 (6)

Для нашего примера имеем*: Fэмп=15,72/2,11=7,45*

7.      определим **Fкрит** по статистическим таблицам

8. если **Fэмп<** Fкрит, то нулевая гипотеза принимается, в противном случае принимается альтернативная гипотеза.

***Задание:***

1.   Исследователь сравнивает эффективность четырех разных методик обучения производственным навыкам. Для этой цели из всех выпускников ПТУ выбраны четыре группы учащихся, обучавшиеся, соответственно четырьмя разными методами. Эффективность методик оценивалась по сумме обработанных учащимися деталей в течение одного дня (см. табл. 5). Проверить гипотезу об отсутствии влияния регулируемого фактора (методик обучения) на продуктивность деятельности ученика

Таблица 5.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| № учащихся  | 1 группа  | 2 группа  | 3 группа  | 4 группа  |
| 1  | 60  | 75  | 60  | 95  |
| 2  | 80  | 66  | 80  | 85  |
| 3  | 75  | 85  | 65  | 100  |
| 4  | 80  | 80  | 60  | 80  |
| 5  | 85  | 70  | 86  |    |
| 6  | 70  | 80  | 75  |    |
| 7  |    | 90  |    |    |

**Лабораторная работа № 1.**

**Способы представления выборок**

***Получение гистограммы и таблицы частот в пакете Statistica.***

* Statistics → Basic Statistics/Tables (Основные статистики ) → Frecuensy Tables (таблица частот).
* В появившемся окне щелкнуть по кнопке Variables. Выбрать щелчком мыши столбец(переменную) для анализа нажать кнопку ОК.
* С помощью вкладки Advanced задать количество интервалов (No. of exact intervals).
* Для получения таблицы частот нажать кнопку Summary: Frecuensy Tables. Для получения гистограммы нажать кнопку Histograms.

Для заданных выборок выполнить:

 а) группировку с числом интервалов =7;

 б) построить гистограмму.

 **1.** 203 154 172 192 233 181 210 153 168 132 204 165 197 205

 143 201 168 147 208 195 153 193 178 162 157 22,8 21,9 12,5

 101 21,1 18,3 14,7 14,5 18,1 18,4 13,9 19,1 18,5 20,2 23,8 16,7 20,4

 19,5 17,2 19,6 17,8 21,3 17,5 19,4 17,8 13,5 17,8 11,8 18,6 19,1

**2.** 38 60 41 51 33 42 45 21 53 60 68 52 47 46 49 49 14 57 54 59 77 47 28 48 58

 32 42 58 61 30 61 35 47 72 41 45 44 55 30 40 67 65 39 48 43 60 54 42 59 50

**3.** 13,4 14,7 15,2 15,1 13,0 8,8 14,0 17,9 15,1 16,5 16,6 14,2 16,3 14,6

 11,7 16,4 15,1 17,6 14,1 18,8 11,6 13,9 18,0 12,4 17,2 14,5 16,3 13,7

 15,5 16,2 8,4 14,7 15,4 11,3 10,7 16,9 15,8 16,1 12,3 14,0 15,7 21,9

 14,4 17,7 15,4 10,9 18,2 17,3 15,2 16,7 17,3 12,1 19,2

**4**. 1,9 3,1 1,3 0,7 3,2 1,1 2,9 2,7 2,7 4,0 1,7 3,2 0,9 0,8 3,1 1,2 2,6 1,9

 2,3 3,2 4,1 1,3 2,4 4,5 2,5 0,9 1,4 1,6 2,2 3,1 1,5 1,1 2,3 4,3 2,1 0,7

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

 4,5 2,1 3,1 5,1 1,1 1,9 0,9 3,1 0,9 3,1 3,3 2,8 2,5 4,0 4,3 1,1 2,1 3,8

 4,6 3,8 2,3 3,9 2,4 4,1 0,9 4,2

 **5.** 17,2 19,2 23,3 18,1 21,0 15,3 16,8 13,2 20,4 16,5 19,7 20,5 20,3 15,4

 14,3 20,1 16,8 14,7 20,8 19,5 15,3 19,3 17,8 16,2 15,7 22,8 21,9 12,5

 10,1 21,1 18,3 14,7 14,5 18,1 18,4 13,9 19,1 18,5 20,2 23,8 16,7 20,4

 19,5 17,2 19,6 17,8 21,3 17,5 19,4 17,8 13,5 17,8 19,1 11,8 18,6

**6.** 11,7 16,4 15,1 17,6 14,1 18,8 11,6 13,9 18,0 12,4 17,2 14,5 16,3 13,7

 15,5 16,2 8,4 14,7 15,4 11,3 10,7 16,9 15,8 16,1 12,3 14,0 15,7 21,9

 13,4 14,7 15,2 15,1 13,0 8,8 14,0 17,9 15,1 16,5 16,6 14,2 16,3 14,6

 14,4 17,7 15,4 10,9 18,2 17,3 15,2 16,7 17,3 12,1 19,2

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

 4,5 2,1 3,1 5,1 1,1 1,9 0,9 3,1 0,9 3,1 3,3 2,8 2,5 4,0 4,3 1,1 2,1 3,8

 1,9 3,1 1,3 0,7 3,2 1,1 2,9 2,7 2,7 4,0 1,7 3,2 0,9 0,8 3,1 1,2 2,6 1,9

 2,3 3,2 4,1 1,3 2,4 4,5 2,5 0,9 1,4 1,6 2,2 3,1 1,5 1,1 2,3 4,3 2,1 0,7

 4,6 3,8 2,3 3,9 2,4 4,1 4,2

**8.** . 10,1 21,1 18,3 14,7 14,5 18,1 18,4 13,9 19,1 18,5 20,2 23,8 16,7 20,4

 20,3 15,4 17,2 19,2 23,3 18,1 21,0 15,3 16,8 13,2 20,4 16,5 19,7 20,5

 14,3 20,1 16,8 14,7 20,8 19,5 15,3 19,3 17,8 16,2 15,7 22,8 21,9 12,5

 19,5 17,2 19,6 17,8 21,3 17,5 19,4 17,8 13,5 17,8 11,8 18,6 19,1

**9.** 13,4 14,7 15,2 15,1 13,0 8,8 14,0 17,9 15,1 16,5 16,6 14,2 16,3 14,6

 11,7 16,4 15,1 17,6 14,1 18,8 11,6 13,9 18,0 12,4 17,2 14,5 16,3 13,7

 15,5 16,2 8,4 14,7 15,4 11,3 10,7 16,9 15,8 16,1 12,3 14,0 15,7 21,9

 14,4 17,7 15,4 10,9 18,2 17,3 15,2 16,7 17,3 12,1 19,2 13,4 14,7

**10.** 1,9 3,1 1,3 0,7 3,2 1,1 2,9 2,7 2,7 4,0 1,7 3,2 0,9 0,8 3,1 1,2 2,6 1,9

 2,3 3,2 4,1 1,3 2,4 4,5 2,5 0,9 1,4 1,6 2,2 3,1 1,5 1,1 2,3 4,3 2,1 0,7

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

 4,5 2,1 3,1 5,1 1,1 1,9 0,9 3,1 0,9 3,1 3,3 2,8 2,5 4,0 4,3 1,1 2,1 3,8

 4,6 3,8 2,3 3,9 2,4 4,1 4,2 0,9

**Лабораторная работа № 2.**

**Числовые выборочные характеристики**

***Получение числовых выборочных характеристик в пакете Statistica.***

* Statistics → Basic Statistics/Tables (Основные статистики ) → Destriptive Statistics (Описательные статистики). Дважды щелкнуть по нему либо выделить и нажать кнопку OK.
* В появившемся окне щелкнуть по кнопке Variables. Выбрать щелчком мыши столбец(переменную) для анализа нажать кнопку ОК.
* С помощью вкладки Advanced выбрать числовые характеристики, которые нужно получить Mean (среднее выборочное), Median(медиана), Mode (мода), Variance (дисперсия), Standart Deviation(средне-квадратическое отклонение), Skewness (коэф. ассиметрии), Kurtosis (коэф. эксцесса).

 Нажать кнопку Summary: Destriptive Statistics.

Задание 1.

Для следующих выборок определить выборочные :

 а) среднее;

 б) дисперсию;

 в) моду;

 г) медиану;

 д) коэффициенты ассиметрии и эксцесса;

 е) получить вариационный ряд;

 ж) получить статистический ряд.

1. 7 3 3 6 4 5 1 2 1 3 4 5 1 2 1 3
2. 5 6 8 2 3 1 1 4 1 2 3 1 1 4 1 2
3. 1 2 3 4 3 3 2 3 4 5 1 2 3 4 3 3
4. 7 2 5 7 7 6 4 6 5 3 5 7 7 6 4 6
5. 1 2 3 4 3 4 5 4 2 1 4 3 4 5 4 2
6. 3 6 7 2 3 1 1 4 1 2 2 3 1 1 4 1
7. 2 4 3 3 2 3 4 5 1 2 2 3 4 5 1 2
8. 7 2 5 7 7 6 4 6 5 3 4 5 4 2 1 5
9. 2 3 4 3 4 5 4 2 1 5 6 5 3 4 4 2
10. 6 1 5 4 6 5 3 4 4 2 1 5 4 6 5 3
11. 7 3 3 6 4 5 1 2 1 3 8 2 3 1 1 4
12. 5 6 8 2 3 1 1 4 1 2 5 6 8 2 3 1
13. 1 2 3 4 3 3 2 3 4 5 3 3 2 3 4 5
14. 7 2 5 7 7 6 4 6 5 3 7 2 5 7 7 6
15. 1 2 3 4 3 4 5 4 2 1 3 4 3 4 5 4
16. 3 6 7 2 3 1 1 4 1 2 3 6 7 2 3 1
17. 2 4 3 3 2 3 4 5 1 2 2 3 4 5 1 2
18. 7 2 5 7 7 6 4 6 5 3 7 7 6 4 6 5
19. 2 3 4 3 4 5 4 2 1 5 4 3 4 5 4 2
20. 6 1 5 4 6 5 3 4 4 2 6 5 3 4 4 2

**Лабораторная работа 3.**

**Корреляционный анализ. Вычисление коэффициентов корреляции Пирсона, Спирмена, точечно-бисериальной корреляции, корреляции знаков, ассоциации. Проверка их значимости**

***Получение коэффициентов корреляции в пакете STATISTICA.***

1. *Выборочный коэффициент корреляции Пирсона.*

* Statistics → Basic Statistics/Tables (Основные статистики ) → Correlation matrices (Корреляционные матрицы). Дважды щелкнуть по нему либо выделить и нажать кнопку OK. На экране появится окно Pearson Product-Moment Correlation (Корреляция Пирсона)
* Нажмите на кнопку Two lists (Два списка). После чего откроется окно выбора переменных. Щелкните по кнопке OK для подтверждения вашего выбора и возврата к окну Pearson Product-Moment Correlation
* Нажмите на кнопку Summary Correlation matrix.

Появляется матрица попарных корреляций в которой красным цветом выделяются значимые коэффициенты корреляции при уровне значимости α=0ю05.

 2. *Выборочный коэффициент корреляции Спирмена.*

* Statistics → Nonparametrics (Непараметрические ) → Correlation (Spearman …)
* С помощью кнопки Variables выберите столбцы для анализа .
* Нажмите на кнопку Spearman rank R или Spearman R.

Появляется список попарных корреляций, в котором красным цветом выделяются значимые коэффициенты корреляции при уровне значимости α=0.05.

Коэффициент корреляции Спирмена : Spiarman R.

Уровень значимости : p-level. Если p-level< α , где α -уровень значимости, то коэффициент корреляции значим.

**Использование пакета *STATISTICA*для вычисления коэффициента корреляции *Пирсона***

**Шаг 1.** Запустите систему **STATISTICA**. Переключитесь в модуль **Основные статистики**. Нажмите кнопку **Open data (Открыть файл данных)** и откройте файл **lab2Pirson.sta**.



**Рис.1**

**Шаг 2.** В стартовой панели модуля **Basic Statistic (Основные статистики)** выберите пункт **Correlation matrices (Корреляционные матрицы)**.



**Рис.2**

 Дважды щёлкните по ней, либо высветите и нажмите кнопку **Ok**. На экране появится окно **Product-Moment Correlation (Корреляция Пирсона)**.



**Рис.3 Окно Product-Moment Correlation (Корреляция Пирсона).**

**Шаг 3.** Щёлкните по кнопке **Two lists (Два списка)**. После чего откроется окно выбора переменных. Выберите переменные как показано на рисунке ниже.



**Рис.4 Окно выбора переменных для построения корреляционной матрицы.**

Таким образом, мы определили два списка переменных **X** **- First variables list (Первый список переменных)** и **Y** - **Second variables list (Второй список переменных)**. Мы хотим подсчитать корреляции между переменной Y и переменной X. Щёлкните на кнопке **Ok** для подтверждения вашего выбора и возврата к окну **Product-Moment Correlation**.

**Шаг 4.** В окне **Product-Moment Correlation** нажмите кнопку **Summary Correlation matrix**. Система произведёт вычисления и спустя секунду вы увидите корреляционную матрицу на экране.



**Рис.5 Корреляционная матрица для переменной X и переменной Y.**

В нашей корреляционной матрице красным цветом автоматически выделены коэффициенты значимые на уровне p<0,05. Именно на эти коэффициенты следует обратить наибольшее внимание. Грубо говоря, зависимость между переменными с выделенными красным цветом коэффициентами корреляции наиболее значимая. В нашем случае переменная Y значимо зависима от переменной X. Коэффициент корреляции между этими переменными равен 0,94. Так как 0,94>0, то мы можем считать, что при возрастании переменной X переменная Y так же возрастает.

Полезно просмотреть зависимость между переменными Y и X графически.

**Шаг 5.** В окне **Product-Moment Correlation** нажмите кнопку **2D scatterplot (2D диаграмма рассеяния).** После этого появится окно диаграммы рассеяния для выбранных переменных.



**Рис.6 График распределения переменной Y от переменной X**

Из графика отчётливо видно, что зависимость является почти линейной - прямая хорошо "ложится" на данные. На графике дана лучшая прямая. Как бы мы не меняли коэффициент наклона, подгонка будет только хуже.

**Использование пакета *STATISTICA*для вычисления коэффициента ранговой корреляции *Спирмена***

**Шаг 1.** Запустите систему **STATISTICA**. Переключитесь в модуль **Основные статистики**. Нажмите кнопку **Open data (Открыть файл данных)** и откройте файл **lab2Spirmen1.sta**.



**Рис.1**

**Шаг 2**. Откройте модуль Statistic → Nonparametrics (Непараметрические) → Correlation (Spearman…) (Корреляция Спирмена) и нажмите клавишу **OK**

|  |  |
| --- | --- |
|  |  |

**Рис. 2**

**Шаг 3.** Кнопкой Variables (переменные) выберите столбцы для анализа и нажмите **ОК**

|  |  |
| --- | --- |
|  |  |

**Рис. 3**

**Шаг 4.** Нажмите кнопку Spearman rank R или Spearman R.



**Рис. 4**

В появившемся окне найдите коэффициент корреляции Спирмена (Spearman R), если он выделен красным цветом, то полученное значение является значимым для уровня значимости 0,05. В нашем случае коэффициент ранговой корреляции Спирмена равен 0,916084 и выделен красным цветом.

**Задание 1.**

Группа испытуемых (студентов физфака СпбГУ) проходила подготовку перед началом работы на тренажере. Испытуемые должны были решать задачи по выбору оптимального типа взлетно-посадочной полосы для заданного типа самолета. Следующая таблица показывает кол-во допущенных ошибок в тренировочной сессии, а также уровни вербального и невербального интеллекта, измеренные по шкале Векслера.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| № | Испытуемый | Кол-во ошибок | Показатель верб. интеллекта | Показатель неверб. интеллекта |
|  1 | И.К. | 29 | 131 | 106 |
|  2 | Р.О. | 54 | 132 | 90 |
|  3 | И.С. | 13 | 121 | 95 |
|  4 | Л.П. | 8 | 127 | 116 |
|  5 | Н.О. | 14 | 136 | 127 |
|  6 | Е.Р. | 26 | 124 | 107 |
|  7 | Г.Е. | 9 | 134 | 104 |
|  8 | О.Р. | 20 | 136 | 102 |
|  9 | Н.Ш. | 2 | 132 | 111 |
|  10 | Г.К. | 17 | 136 | 99 |

а) Вычислить коэффициенты Пирсона и Спирмена между кол-вом ошибок и уровнем вербального интеллекта и проверить их значимость при уровне значимости α=0.05.

б) Вычислить коэффициенты Пирсона и Спирмена между кол-вом ошибок и уровнем невербального интеллекта и проверить их значимость при уровне значимости α=0.05.

в) Вычислить коэффициенты Пирсона и Спирмена между уровнем вербального и невербального интеллекта и проверить их значимость при уровне значимости α=0.05.

г) Получить уравнение прямой регрессии кол-ва ошибок на уровень невербального интеллекта. Построить график прямой регрессии. Определить допустимый уровень невербального интеллекта, при котором количество ошибок не превысит 20.

д) Получить уравнение прямой регрессии кол-ва ошибок на уровень вербального интеллекта. Построить график прямой регрессии. Определить допустимый уровень вербального интеллекта, при котором количество ошибок не превысит 20.

**Задание 2.**

. Вычислить точечно-бисериальный коэф. корреляции между уровнем агрессивности и полом подростков (1-мальчики, 2-девочки) и проверить его значимость при уровне значимости α=0.05.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| агресс | 86 |  69 |  72 |  65 |  113 |  65 |  118 |  45 |  141 |  104 |  41 |  50 |
| пол |  1 |  1 |  1 |  1 |  1 |  1 |  1 |  1 |  1 |  1 |  1 |  1 |

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 36 |  55 |  40 |  22 |  58 |  16 |  7 |  9 |  16 |  26 |  15 |  20 |  15 |
|  2 |  2 |  2 |  2 |  2 |  2 |  2 |  2 |  2 |  2 |  2 |  2 |  2 |

 **Задание 3.**

Вычислить коэф. корреляции Спирмена между уровнем вербального и невербального интеллекта (шкала Векслера) и проверить его значимость при уровне значимости α=0.05.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Верб. |  131 |  132 |  121 |  127 |  136 |  124 |  134 |  136 |  132 |  136 |
| Неверб. |  106 |  90 |  95 |  116 |  127 |  107 |  104 |  102 |  111 |  99 |

 **Задание 4.**

 Вычислить коэф. корреляции знаков между строгостью родителей и агрессивностью детей и проверить его значимость при уровне значимости α=0.05.

(+ - есть, - -нет):

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Агресс |  **+** |  **-** |  **-** |  **+** |  **+** |  **-** |  **-** |  **-** |  **+** |  **-** |  **+** |  **+** |  **+** |  **-** |  **-** |
| Строг |  **+** |  **-** |  **-** |  **-** |  **+** |  **-** |  **-** |  **-** |  **+** |  **-** |  **+** |  **-** |  **+** |  **+** |  **-** |

 **Задание 5.**

 Вычислить коэф. ассоциации между дальтонизмом и полом (выборка 1000 человек) и проверить его значимость при уровне значимости α=0.05.

:

|  |  |  |
| --- | --- | --- |
|  |  Мужчины |  Женщины  |
| Дальтоники |  38 |  6 |
| Недальтоники |  442 |  514 |

**Задание 6.**

В следующей таблице представлены ранги ценностей, полученные при обследовании пары мать-дочь.

|  |  |  |
| --- | --- | --- |
| Ценности | Ранг ценностейв иерархии матери  | Ранг ценностейв иерархии дочери |
| 1. Активная деятельная жизнь | 15 | 15 |
| 2. Жизненная мудрость | 1 | 3 |
| 3. Здоровье | 7 | 14 |
| 4. Интересная работа | 8 | 12 |
| 5. Красота природы и искусство | 16 | 17 |
| 6. Любовь | 11 | 10 |
| 7. Материально-обеспеченная жизнь | 12 | 13 |
| 8. Наличие хороших друзей | 9 | 11 |
| 9. Общественное признание | 17 | 5 |
| 10. Познание | 5 | 1 |
| 11. Продуктивная жизнь | 2 | 2 |
| 12. Развитие | 6 | 8 |
| 13. Развлечения | 8 | 18 |
| 14. Свобода | 4 | 6 |
| 15. Счастливая семейная жизнь | 14 | 4 |
| 16. Счастье других | 13 | 16 |
| 17. Творчество | 10 | 9 |
| 18. Уверенность в себе | 3 | 7 |

Вычислить коэффициент Спирмена между ценностной ориентацией матери и дочери и проверить его значимость при уровне значимости α=0.05.

**Лабораторная работа 4.**

**Непараметрические критерии. Критерии сдвига: критерий знаков, критерий Вилкоксона для двух зависимых выборок, критерий Манна-Уитни для двух независимых выборок.**

**Непараметрические методы сравнения выборок.**

* Statistics → Nonparametrics (Непараметрические )
* Выбрать метод:
* Для случая двух зависимых выборок: Comparing two dependent samples (variables)
* Для случая двух независимых выборок со значениями в одном столбце (второй столбец содержит номер выборки - группирующая переменная)::

 Comparing two independent samples (groups)

* Для случая нескольких зависимых выборок: Comparing multiple dep. samples (variables)
* Для случая нескольких независимых выборок со значениями в одном столбце (второй столбец содержит номер выборки - группирующая переменная)::

 Comparing multiple ind. samples (groups)

 Нажать кнопку ОК.

* В появившемся окне щелкнуть по кнопке Variables. Выбрать щелчком мыши столбцы (переменные) для анализа , в случае наличия группирующей переменной слева выбрать столбец значений признака, справа - группирующую переменную. В случае одного списка переменных первая выделяется щелчком мыши остальные -щелчком мыши при нажатой клавише CTRL Нажать кнопку ОК.
* Щелкнуть по кнопке Summary …

 Появляются выборочные значения статистики соответствующего критерия и значение p- уровня значимости. Если p<α , где α - заданный уровень значимости, то различие в распределении выборок значимо..

1. Перед началом игры двум группам детей были показаны два мультфильма: группе А мультипликационный боевик и группе В сказка. После этого группа исследователей наблюдала поведение детей и оценивала их уровень агрессивности (количество агрессивных поступков). Результаты были представлены в порядковой шкале:

|  |  |
| --- | --- |
|  А | 26 22 19 21 14 18 29 17 11 34 |
|  В | 16 10 8 13 19 11 13 9 21 |

При уровне значимости α=0.05 с помощью критерия Манна-Уитни проверить гипотезу о влиянии просмотренных мультфильмов на уровень агрессивности детей.

1. Для 10 человек была предложена специальная диета. После двухнедельного питания по этой диете масса их тел изменилась следующим образом:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Масса до диеты | 68 | 80 | 92 | 81 | 70 | 79 | 78 | 66 | 57 | 76 |
| Масса послеДиеты | 60 | 84 | 87 | 79 | 74 | 71 | 72 | 67 | 57 | 70 |

Проверить гипотезу о действии диеты на массу тела по критерию знаков и по критерию Вилкоксона при уровне значимости α=0.05.

1. Проверяется гипотеза о том, что изучение курса математики способствует формированию у учеников приемов логического мышления. Для этого был проведен следующий эксперимент. Ученикам 7 класса давали по 5 задач, использующих одни и те же приемы логического мышления в конце сентября и в конце мая следующего года. Работа предлагалась одним и тем же учащимся. Получились следующие результаты:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Число решенныхЗадач в сентябре |  1 |  1 |  1 |  2  |  0 |  0 |  0 |
| Число решенных Задач в мае |  1 |  0 |  0 |  2 |  1 |  0 |  1 |

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  0 |  0 |  1 |  0 |  3 |  2 |  1 |  0 |  0 |  1 |  0 |  1 |  0 |
|  0 |  0 |  1  |  1 |  3 |  2 |  0 |  1 |  1 |  1 |  1 |  1 |  1 |

Проверить, отвергается или принимается указанная гипотеза при уровне значимости α=0.05.

1. Участникам психологического эксперимента, моделирующего деятельность воздушного диспетчера, был измерен уровень вербального и невербального интеллекта с помощью методики Д. Векслера. Было обследовано 26 юношей в возрасте от 18 до 24 лет (средний возраст 20,5 лет). 14 из них были студентами физического факультета, а 12 – студентами психологического факультета Ленинградского университета (СпбГУ). Показатели вербального интеллекта представлены в таблице №1, а показатели невербального интеллекта представлены в таблице №2.

Используя критерий Манна-Уитни, определите:

1.Можно ли утверждать, что одна из групп превосходит другую по уровню вербального интеллекта?

2.Можно ли утверждать, что одна из групп превосходит другую по уровню невербального интеллекта? Сформулируйте для каждого случая H0 и H1 .

**Таблица№1**

|  |  |
| --- | --- |
| Студенты-физики | Студенты-психологи |
| №п/п | Код имени испытуемого | Показатель вербального интеллекта | №п/п | Код имени испытуемого | Показатель вербального интеллекта |
| 1 | И.А. | 132 | 1 | Н.Т. | 126 |
| 2 | К.А. | 134 | 2 | О.В. | 127 |
| 3 | К.Е. | 124 | 3 | Е.В. | 132 |
| 4 | П.А. | 132 | 4 | Ф.О. | 120 |
| 5 | С.А. | 135 | 5 | И.Н. | 119 |
| 6 | Ст.А. | 132 | 6 | И.Ч. | 126 |
| 7 | Т.А. | 131 | 7 | И.В. | 120 |
| 8 | Ф.А. | 132 | 8 | К.О. | 123 |
| 9 | Ч.И. | 121 | 9 | Р.Р. | 120 |
| 10 | Ц.А. | 127 | 10 | Р.И. | 116 |
| 11 | См.А. | 136 | 11 | О.К. | 123 |
| 12 | К.Ан. | 129 | 12 | Н.К. | 115 |
| 13 | Б.Л. | 136 |  |  |  |
| 14 | Ф.В. | 136 |  |  |  |

Таблица№2

|  |  |
| --- | --- |
| Студенты-физики | Студенты-психологи |
| №п/п | Код имени испытуемого | Показатель невербального интеллекта | №п/п | Код имени испытуемого | Показатель невербального интеллекта |
| 1 | И.А. | 111 | 1 | Н.Т. | 113 |
| 2 | К.А. | 104 | 2 | О.В. | 107 |
| 3 | К.Е. | 107 | 3 | Е.В. | 123 |
| 4 | П.А. | 90 | 4 | Ф.О. | 122 |
| 5 | С.А. | 115 | 5 | И.Н. | 117 |
| 6 | Ст.А. | 107 | 6 | И.Ч. | 112 |
| 7 | Т.А. | 106 | 7 | И.В. | 105 |
| 8 | Ф.А. | 107 | 8 | К.О. | 108 |
| 9 | Ч.И. | 95 | 9 | Р.Р. | 111 |
| 10 | Ц.А. | 116 | 10 | Р.И. | 114 |
| 11 | См.А. | 127 | 11 | О.К. | 102 |
| 12 | К.Ан. | 115 | 12 | Н.К. | 104 |
| 13 | Б.Л. | 102 |  |  |  |
| 14 | Ф.В. | 99 |  |  |  |

1. В исследовании изучалась проблема психологических барьеров при обращении в службу знакомств у мужчин и женщин. В эксперименте участвовали 17 мужчин и 23 женщины в возрасте от 17 до 45 лет (средний возраст 32,5 года). Испытуемые должны были отметить на отрезке точку, соответствующую интенсивности внутреннего сопротивления, которое им пришлось преодолеть, чтобы обратиться в службу знакомств. Длина отрезка, отражающая максимально возможное сопротивление, составляла 100 мм. В таблице №1 приведены показатели интенсивности сопротивления, выраженные в миллиметрах. Можно ли утверждать, что мужчинам приходится преодолевать субъективно более мощное сопротивление?

|  |  |
| --- | --- |
| Группа 1 – мужчины (n1=17) | Группа 2 – женщины (n2=23) |
| 1 | 81 | 1 | 70 |
| 2 | 80 | 2 | 66 |
| 3 | 73 | 3 | 66 |
| 4 | 72 | 4 | 63 |
| 5 | 72 | 5 | 63 |
| 6 | 69 | 6 | 61 |
| 7 | 69 | 7 | 60 |
| 8 | 65 | 8 | 54 |
| 9 | 65 | 9 | 47 |
| 10 | 62 | 10 | 43 |
| 11 | 60 | 11 | 41 |
| 12 | 54 | 12 | 40 |
| 13 | 54 | 13 | 39 |
| 14 | 43 | 14 | 38 |
| 15 | 30 | 15 | 38 |
| 16 | 26 | 16 | 35 |
| 17 | 26 | 17 | 30 |
|  |  | 18 | 27 |
|  |  | 19 | 25 |
|  |  | 20 | 23 |
|  |  | 21 | 17 |
|  |  | 22 | 10 |
|  |  | 23 | 9 |

1. В эксперименте по исследованию интеллектуальной настойчивости 22 испытуемым предъявлялись сначала разрешимые четырехбуквенные, пятибуквенные и шестибуквенные анаграммы, а затем неразрешимые анаграммы, время работы над которыми не ограничивалось. Эксперимент проводился индивидуально с каждым испытуемым. Использовалось четыре комплекта анаграмм. Показатели длительности попыток (в секундах) в решении неразрешимых анаграмм представлены в таблице №1. Все испытуемые были юношами-студентами технического вуза в возрасте от 20 до 22 лет.

Можно ли утверждать, что длительность попыток решения каждой из четырех неразрешимых анаграмм примерно одинакова? Сформулируйте нулевую и альтернативную гипотезу. Используя непараметрический критерий Крускала-Уоллиса, проверьте справедливость нулевой гипотезы.

Таблица№1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| №п/п | Группа 1: анаграмма ФОЛИТОН (n1=4) | Группа 2: анаграмма КАМУСТО (n2=8) | Группа 3: анаграмма СНЕРАКО (n3=6) | Группа 4: анаграмма ГРУТОСИЛ (n4=4) |
| 1 | 145 | 145 | 128 | 60 |
| 2 | 194 | 210 | 283 | 2361 |
| 3 | 731 | 236 | 469 | 2416 |
| 4 | 1200 | 385 | 482 | 3600 |
| 5 |  | 720 | 1678 |  |
| 6 |  | 848 | 2081 |  |
| 7 |  | 905 |  |  |
| 8 |  | 1080 |  |  |

**Лабораторная работа 5.**

**Многофункциональный критерий Фишера.**

***Сравнить доли эффектов с помощью многофункционального критерия Фишера.***

1. Проверяется гипотеза о том, что изучение курса математики способствует формированию у учеников приемов логического мышления. Для этого был проведен следующий эксперимент. Ученикам 7 класса давали по 5 задач, использующих одни и те же приемы логического мышления в конце сентября и в конце мая следующего года. Работа предлагалась одним и тем же учащимся. Получились следующие результаты:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| НаписавшиеЗадачи в сентябре |  1 |  1 |  1 |  0  |  0 |  0 |  0 |
| Написавшие Задачи в мае |  1 |  0 |  0 |  1 |  1 |  0 |  1 |

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  0 |  0 |  1 |  0 |  1 |  1 |  1 |  0 |  0 |  1 |  0 |  1 |  0 |
|  0 |  0 |  1  |  1 |  1 |  1 |  0 |  1 |  1 |  1 |  1 |  1 |  1 |

Проверить, отвергается или принимается указанная гипотеза при уровне значимости α=0.05.

**Лабораторная работа 6.**

**Однофакторный дисперсионный анализ**

Будем выполнять в модуле *Basic Statistics and Tables* (можно выполнять также в модуле *ANOVA/MANOVA*). Создадим таблицу с двумя столбцами *Р* и *Т* и 30 строками; в *Р* занесем данные по производительности, в *Т* - уровни *Т*: технологии *Т*0*, Т*1*, Т*2. Далее выполним:

*One - Way ANOVA (Analys Of Variances) - Analysis: Detailed Analysis Of Individual tables, Variabbles: Grouping* *variabbles* (группирующие переменные): *T, Dependent variabbles* (зависимые переменные - отклики): *P - OK - OK -* отметив *Statistics: Number of observations* (количество наблюдений), *Standart deviations* (стандартные отклонения) и *Variances* (дисперсии), получим *Summary table of means* (таблицу средних); видно, как отличаются средние в каждой из групп (при фиксированном уровне фактора *Т*) - Возвращаемся в окно *Descriptive Stats and ... Results* и выполняем *Analysis of Variance -* Наблюдаем таблицу: в столбце *SS (Sum of Squares)* *Effect* указана сумма квадратов (4), умноженная на (*k -* 1), *df =* 2 *= k -* 1 *-* число степеней свободы, *MS (Mean Square)* = 839.0 - оценка (4), *SS* = 2711 - сумма квадратов (3), умноженная на *(N - k)*, *df =* 27 *= N - k, Ms Error* = 100.4 - оценка (3), *F* = 8.35 - значение статистики (5), *p* = 0.0015 - вероятность в (7); последняя слишком мала, чтобы поверить в истинность гипотезы *Н* об отсутствии влияния фактора *Т*.

*Гипотезы:*

H0: различия *между* группами являются не более выраженными, чем случайные различия *внутри* каждой группы

H1: различия *между* группами являются более выраженными, чем случайные различия *внутри* каждой группы.

Последовательность операций в однофакторном дисперсионном анализе для несвязанных выборок:

***Задание:***

1. В эксперименте по исследованию интеллектуальной настойчивости 22 испытуемым предъявлялись сначала разрешимые четырехбуквенные, пятибуквенные и шестибуквенные анаграммы, а затем неразрешимые анаграммы, время работы над которыми не ограничивалось. Эксперимент проводился индивидуально с каждым испытуемым. Использовалось четыре комплекта анаграмм. Показатели длительности попыток (в секундах) в решении неразрешимых анаграмм представлены в таблице №1. Все испытуемые были юношами-студентами технического вуза в возрасте от 20 до 22 лет.

Можно ли утверждать, что длительность попыток решения каждой из четырех неразрешимых анаграмм примерно одинакова? Сформулируйте нулевую и альтернативную гипотезу. Проверьте справедливость нулевой гипотезы.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| №п/п | Группа 1: анаграмма ФОЛИТОН (n1=4) | Группа 2: анаграмма КАМУСТО (n2=8) | Группа 3: анаграмма СНЕРАКО (n3=6) | Группа 4: анаграмма ГРУТОСИЛ (n4=4) |
| 1 | 145 | 145 | 128 | 60 |
| 2 | 194 | 210 | 283 | 2361 |
| 3 | 731 | 236 | 469 | 2416 |
| 4 | 1200 | 385 | 482 | 3600 |
| 5 |  | 720 | 1678 |  |
| 6 |  | 848 | 2081 |  |
| 7 |  | 905 |  |  |
| 8 |  | 1080 |  |  |

**Лекции по курсу «Компьютерная обработка данных экспериментального исследования»**

**Лекция 1**

***Основные понятия математической статистики. Генеральная совокупность и выборка из генеральной совокупности. Понятие выборочного метода.***

Исходным понятием статистики является понятие **совокупность**, объединяющее обычно какое-либо множество испытуемых (учащихся) по одному или нескольким интересующим признакам. Главное требование к выделению изучаемой совокупности — это ее качественная однород­ность, например, по уровню знаний, росту, весу и другим признакам. Члены совокупности могут сравниваться меж­ду собой в отношении только того качества, которое ста­новится предметом исследования. При этом обычно абстрагируются от других неинтересующих качеств. Так, если педагога интересует успеваемость учащихся, то он не принимает во внимание, как правило, их рост, вес и другие параметры, не относящиеся непосредственно к изучаемому вопросу.

Применение большинства статистических методов основано на идее использования небольшой случайной совокупности испытуемых из общего числа тех, на которых можно было бы распространить (генерализовать) выводы, полученные в результате изучения совокупности. Эта небольшая совокупность в статистике называется **выборочной совокупностью** (или короче — **выборкой**). Главный принцип формирования выборки — это случайный отбор испытуемых из мыслимого множества учащихся, называемого **генеральной совокупностью** или **популяцией** объектов или явлений. Как по анализу элементов, содержащихся в капле крови, медики нередко судят о составе всей крови человека, так и по выборочной совокупности учащихся изучаются явления, характерные для всей генеральной совокупности.

Когда для каждого объекта в выборке измерено значение одной переменной, популяция и выборка называются **одномерными**. Если же для каждого объекта регистрируются значения двух или нескольких переменных, такие данные называются **многомерными**.

Одной из основных задач статистического анализа является получение по имеющейся выборке достоверных сведений о интересующих исследователя характеристиках генеральной совокупности. Поэтому важным требованием к выборке является ее **репрезентативность**, то есть правильная представимость в ней пропорций генеральной совокупности. Достижению репрезентативности может способствовать такая организация эксперимента, при которой элементы выборки извлекаются из генеральной совокупности случайным образом.

Обычно в статистике различают три типа значений переменных: **количественные, номинальные и ранговые**.

Значения количественных переменных являются числовыми, могут быть упорядочены и для них имеют смысл различные вычисления (например, среднее значение). На обработку количественных переменных ориентиро­вано подавляющее большинство статистических методов.

Значения номинальных переменных (например: пол, вид, цвет) являются нечисловыми, они означают принадлежность к некоторым классам и не могут быть упорядочены или непосредственно использованы в вычислениях. Для анализа номинальных переменных специально предназначены лишь избранные разделы математической статистики, например, катего­риальный анализ. Однако в ряде случаев для этой цели могут быть использованы и некоторые ранговые и количественные методы, если номинальные значения предварительно заменить на числа, обозначающие их условные коды.

Ранговые или порядковые переменные занимают промежуточное поло­жение: их значения упорядочены (состояние больного, степень предпочтения), но не могут быть с уверенностью измерены и сопо­ставлены количественно. К анализу ранговых переменных применимы так называемые ранговые методы.

**Ранг наблюдения** – это тот номер, который получит данное наблюдение в упорядоченной совокупности всех данных – после их упорядочивания по определенному  правилу (например, от большего значения к меньшим). Процедура перехода от совокупности наблюдений к последовательности их рангов называется ранжированием.

Ранговые и номинальные значения при вводе данных следует обозначать целыми числами.

**Лекция 2.**

***Вариационный ряд, статистический ряд, гистограмма, полигон частот.***

Статистическим рядом распределения называют упорядоченное распределение единиц совокупности на группы по изучаемому признаку, при этом известна численность единиц в группах или удельный вес группы в общей совокупности. В зависимости от типа от группирующего признака ряды могут быть *вариационными* (количественными) и *атрибутивными* (качественными). Таким образом, группировка данных, описанная в предыдущем параграфе, служит основой для получения статистических рядов распределения. Статистический ряд распределения является простейшей разновидностью структурной группировки по одному признаку.

Численные или качественные значения признака называются *вариантами*, а численность каждой группы – *частотами.*

Если группирующий признак является атрибутивным, то статистический ряд называется также *атрибутивным.*

Если группирующий признак является количественным, то соответствующий статистический ряд распределения называется *вариационным.* Вариационные ряды бывают двух видов; *дискретные* и *интервальные.*

Если количественный признак меняется *пре­рывно (дискретно),* т. е. может принимать только некоторые — ча­ще целые значения (например, тарифный разряд рабочих), то число групп должно соответствовать количеству значений при­знака. Для каждого значения признака (варианты) подсчитывается чаcтота встречаемости в исходной совокупности данных. Совокупность вариант и соответствующих им частот называется *дискретным статистическим рядом*. Дискретный ряд представляется в табличной форме и графической (полигон частот, кумулята).

При построении атрибутивных рядов число групп соответст­вует числу разных значений признака, который располагаются в порядке возрастания.

Пример дискретного статистического ряда, представленного в табличной форме.

Табл. 1*.* Распределение рабочих по тарифным разрядам:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Тарифный разряд | 1 | 2 | 3 | 4 | 5 | 6 |
| Число рабочих, чел | 10 | 20 | 40 | 60 | 50 | 20 |

При *непрерывном* изменении признак принимает любые зна­чения (например, стаж работы или возраст рабочих), поэтому группы ограничиваются значениями признака в интервале «от – до». Для каждого интервала подсчитываются частоты попадания признака в этот интервал. Совокупность интервалов и соответствующих им частот называется *интервальным статистическим рядом*. Ниже описан алгоритм получения интервального статистического ряда.

Алгоритм получения интервального статистического ряда:

1. выбор группировочного признака;
2. определение размаха вариации

 R = xmax - xmin ,

 где xmax и xmin соответственно наибольшее и наименьшее значения

 признака в изучаемой совокупности;



1. определение длины интервала

Длина интервала в случае равных интервалов вычисляется по формуле

 L = 

 где k *—* количество интервалов. Обычно в случае, если признак принимает только целые значения, длину интервала округляют до ближайшего целого, большего L .

Возможен и другой способ определения длины интервала, не требующий предварительного установления числа групп. В этом случае используется формула Стерджесса для определения оптимального количества групп (интервалов):

 k =1+3,322∙ ln n

где *п –* число наблюдений. Длина интервала определяется аналогично предыдущему случаю.

1. определение границ интервалов; левой границей первого интервала является xmin , правой границей первого интервала и левой границей следующего будет xmin + L , затем к этому числу прибавляется L для получения следующей границы и т. д. до тех пор, пока очередная правая граница интервала не превысит xmax.
2. подсчет частоты попадания значений признака в каждый из интервалов. Если значение попадает на границу интервала, то его обычно приписывают правому интервалу (т.е. интервалы открыты справа и закрыты слева).

Если некоторые из групп (интервалов) оказываются пустыми, рекомендуется провести перегруппировку, например, изменить количество интервалов (при небольшом объеме совокупности).

Статистический ряд может быть представлен в виде таблицы, в одной графе которой указываются варианты или интервалы, а в другой – соответствующие им частоты. Статистические ряды, которые использовались в качестве примеров в предыдущем параграфе, были представлены в табличной форме.

Первым этапом изучения вариационного ряда является его графическое изображение.Дискретный вариационный ряд изоб­ражается в виде так называемого *полигона распределения частот*,являющегося разновидностью статистиче­ских ломаных. Для изображения интервального ряда применяют­ся *полигон распределения частот и гистограмма частот.*

Строятся графики в прямоугольной системе координат. При построении полигона частот на оси абсцисс в одинаковом мас­штабе откладываются направо в порядке возрастания значения признака (для дискретного признака) или центральные значения интервалов (для интервальных рядов); по оси ординат наносится шкала для выражения величин частот. Из точек на оси абсцисс, соответствующих величине признака, восстанавливаются пер­пендикуляры высотой, соответствующей частоте; вершины перпендикуляров соединяются отрезками прямой. Крайние точки полученной ломаной соединяются с лежащими на оси абсцисс следующими (меньшими и большими) возможными, но фактиче­ски не наблюдающимися значениями признака, частота которых, очевидно, равна 0. Замкнутая с осью абсцисс ломаная линия представляет полигон распределения частот.

Для построения *гистограммы* по оси абсцисс откладывают ве­личины интервалов, а частоты изображаются прямоугольниками, построенными на интервалах с высотой в масштабе оси ординат.

В случае неравенства интервалов гистограмма строится не по частотам или частостям, а по плотности распределения.

В ряде случаев для изображения вариационных рядов исполь­зуется *кумулятивная кривая (кумулята),*она особенно удобна для сравнения вариационных рядов. Накопленные частоты наносят­ся на чертеж в виде ординат; соединяя вершины отдельных орди­нат прямыми, получают ломаную линию, которая, начиная с ну­ля, непрерывно поднимается над осью абсцисс до тех пор пока не достигнет высоты, соответствующей общей сумме частот.

Если поменять местами оси координат в кумуляте, то получа­ем новый вид графического изображения - *огиву.*

При изучении процессов концентрации (концентрации про­изводства, концентрации капитала и др.) используется графичес­кое изображение вариационного ряда в виде кривой Лоренца. Для ее построения абсолютные показатели числа единиц в груп­пах и размер изучаемого признака выражаются в относительных показателях (в долях или процентах к итогу) и исчисляются их на­копленные значения. При построении графика на горизонталь­ной линии наносится шкала для ряда накопленных частостей, а 'на вертикальной линии — шкала для накопленных относительных величин размера-изучаемого признака. Далее наносятся точки в соответствии с накопленными значениями двух рядов. Соединив все точки прямыми линиями, получают кривую, характеризую­щую степень неравномерности распределения. Линия, соединя­ющая нижний левый угол графика с верхним правым (диагональ четырехугольника), является линией равномерного распределе­ния. Чем больше кривая отличается от диагонали, тем больше не­равномерность.

При построении графических изображений вариационного ряда большое значение имеет соотношение масштабов по оси аб­сцисс *(х)* и оси ординат (/). В этом случае следует руководство­ваться так называемым «правилом золотого сечения», в соответ­ствии с которым высота графика должна быть примерно в 1,5 ра­за меньше его основания.

**Лекция 3.**

***Числовые выборочные характеристики: среднее, дисперсия, мода, медиана, коэффициенты ассиметрии и эксцесса.***

Среднее (оценка среднего, выборочное среднее) — сумма значений переменной, деленная на n (число значений перемен­ной). Если вы имеете значения Х(1), ..., X(N), то формула для выборочного среднего имеет вид:

`х =                                                                (1)

Пример: Наблюдение посещаемости четырех внеклассных мероприятий в экспериментальном (20 учащихся) и контрольном (30) классах дали значения (соответственно): 18, 20, 20, 18 и 15, 23, 10, 28. Среднее значение посещаемости в обоих классах получается одинаковое - 19. Однако видно, что в контрольном классе этот показатель подчинен воздействию каких-то специфических факторов.

Выборочное среднее является той точкой, сумма отклонений наблюдений от которой равна 0. Формально это записывается следующим образом:

(`х  - х1) + (`х  - х2) + ... + (`х  - хn) =0

Для оценки степени разброса (отклонения) какого-то показателя от его среднего значения, наряду с максимальным и минимальным значениями, используются понятия дисперсии и стандартного отклонения.

Дисперсия выборки или выборочная дисперсия (от английского variance) – это мера изменчивости перемен­ной. Термин впервые введен  Фишером в 1918 году. Выборочная дисперсия вычисляется по формуле:

s2 =                                                                                       (2)

где `х  — выборочное среднее,

N — число наблюдений в выборке.

Дисперсия меняется от нуля до бесконечности. Крайнее значение 0 означает отсутствие изменчивости, когда значения переменной постоянны.

Стандартное отклонение, среднее квадратическое отклонение (от английского standard deviation) вычисляется как корень квадратный из дисперсии. Чем выше дисперсия или стандартное отклонение, тем сильнее разбросаны значения переменной относительно среднего.

                                                                                          (3)

Медиана разбивает выборку на две равные части. Половина значений переменной лежит ниже медианы, половина — выше. Медиана дает общее представление о том, где сосредоточены значения переменной, иными словами, где находится ее центр. В некоторых случаях, например при описании доходов населения, медиана более удобна, чем среднее.

Рассмотрим способы определения медианы при различных значениях N.  Для нахождения медианы измерения записывают в ряд по возрастанию значений. Если число измерений N нечетное, то медиана численно равна значению этого ряда, стоящему точно в середине, или на (N+1)/2 месте. Например, медиана пяти измерений: 10, 17, 21, 24, 25 – равна 21 – значению, стоящему на третьем месте (N+1)/2=(5+1)/2=3.

Если число измерений четное, то медиана численно равна среднему арифметическому значений ряда, стоящих в середине, или на N/2 и N/2+1 местах. Например, медиана восьми измерений: 5, 5, 6, 7, 8, 8, 9, 9 – равна 7,5 (7+8)/2=7,5 – среднему арифметическому значений ряда, стоящих на четвертом и пятом местах (N/2=8/2=4 и N/2+1=4+1=5).

Квартили представляют собой значения, которые делят две половины выборки (разбитые медианой) еще раз пополам (от слова кварта — четверть).

Различают верхнюю квартиль, которая больше медианы и делит пополам верх­нюю часть выборки (значения переменной больше медианы), и нижнюю квартиль, которая меньше медианы и делит пополам нижнюю часть выборки.

Нижнюю квартиль часто обозначают символом 25%, это означает, что 25% зна­чений переменной меньше нижней квартили.

Верхнюю квартиль часто обозначают символом 75%, это означает, что 75% зна­чений переменной меньше верхней квартили.

Таким образом, три точки — нижняя квартиль, медиана и верхняя квартиль - делят выборку на 4 равные части.

¼ наблюдений лежит между минимальным значением и нижней квартилью, ¼ - между нижней квартилью и медианой, ¼ - между медианой и верхней квартилью, ¼ - между верхней квартилью и максимальным значением выборки.

Мода представляет собой максимально часто встречающееся значение пере­менной (иными словами, наиболее «модное» значение переменной), например, популярная передача на телевидении, модный цвет платья или марка автомоби­ля и т. д, Сложность в том, что редкая совокупность имеет единственную моду. (Например: 2, 6, 6, 8, 9, 9, 9, 10 – мода = 9).

Если распределение имеет несколько мод, то говорят, что оно мультимодально или многомодально (имеет два или более «пика»).

Ассиметрия – это свойство распределения выборки, которое характеризует несимметричность распределения СВ. На практике симметричные распределения встречаются редко и чтобы выявить и оценить степень асимметрии, вводят следующую меру:

,

Эксцесс равен:



  Кривая распределения может быть островершинной, плосковершинной, средне вершинной. Эти четыре момента составляют набор особенностей распределения при анализе данных. Для нормального распределения А=0, Е=0.

**Лекция 4.**

***Корреляционный анализ: основные понятия. Различные коэффициенты корреляции. Выбор коэффициента корреляции в зависимости от измерительных шкал***

**Корреляционный анализ,** совокупность основанных на математической теории корреляции методов обнаружения корреляционной зависимости между двумя случайными признаками или факторами. К. а. экспериментальных данных заключает в себе следующие основные практические приёмы: 1) построение корреляционного поля и составление корреляционной таблицы; 2) выборочных коэффициентов корреляции или корреляционного отношения; 3) проверка статистической гипотезы значимости связи. Дальнейшее исследование заключается в установлении конкретного вида зависимости между величинами Зависимость между тремя и большим числом случайных признаков или факторов изучается методами многомерного К. а. (вычисление частных и множественных коэффициентов корреляции и корреляционных отношений).

  Корреляционное поле и корреляционная таблица являются вспомогательными средствами при анализе выборочных данных. При нанесении на координатную плоскость выборочных точек получают корреляционное поле. По характеру расположения точек поля можно составить предварительное мнение о форме зависимости случайных величин (например, о том, что одна величина в среднем возрастает или убывает при возрастании другой). Для численной обработки результаты обычно группируют и представляют в форме корреляционной таблицы. В каждой клетке корреляционной таблицы приводятся численности *гц;* тех пар *(х, у),* компоненты которых попадают в соответствующие интервалы группировки по каждой переменной.

 Предполагая длины интервалов группировки (по каждому из переменных) равными между собой, выбирают центры *xi* (соответственно *yj*) этих интервалов и числа *nij* в качестве основы для расчётов.

 Коэффициент корреляции и корреляционное отношение дают более точную информацию о характере и силе связи, чем картина корреляционного поля

 **Корреляционная зависимость** двух признаков - это **зависимость в среднем.**

Основным инструментом выяснения корреляционной зависимости является коэффициент корреляции.

Значение коэффициента корреляции лежит между -1 и +1. Если наблюдается тенденция возрастания одной величины при росте другой, то говорят о положительной коррелированности величин, если наблюдается тенденция увеличения одной величины при уменьшении другой, то говорят об отрицательной коррелированности величин. Если значение коэффициента корреляции равно 0, то корреляционная зависимость отсутствует.

Оценку коэффициента корреляции дает **выборочный коэффициент корреляции**. По его значению можно проверить, равен истинный коэффициент корреляции 0, или нет (проверка значимости коэффициента корреляции).

Выбор коэффициента корреляции определяется типом измерительных шкал для обоих признаков:

|  |  |  |
| --- | --- | --- |
| Измерительная шкала для1-го признака | Измерительная шкала для2-го признака | Коэффициент корреляции |
| Количественная (интервальная или отношений) | Количественная (интервальная или отношений) | Пирсона |
| Количественная или порядковая | Количественная или порядковая | Спирмена |
| Качественная с двумя значениями | Количественная или порядковая | Точечно- бисериальный |
| Качественная с двумя значениями | Качественная с двумя значениями | Ассоциации (четырехпольный)  |

Замечание: коэффициент корреляции Пирсона исследует только линейную корреляционную зависимость и используется только в случае, когда оба признака имеют нормальное распределение.

**Формулы для вычисления выборочных коэффициентов корреляции:**

1. *Коэффициент корреляции Пирсона:*

 **

где *и -* средние выборочные для X и Y.

 * и -* средне-квадратические отклонения для X и Y.

 *N -* объем выборки.

1. *Коэффициент корреляции Спирмена:*

 **

где - квадрат разности рангов значений признаков X и Y для i-того испытуемого;  *N -* объем выборки.

Если четырехпольная таблица для двух качественных признаков, каждый из которых имеет ровно 2 значения, имеет вид

|  |  |
| --- | --- |
| a | b |
| c | d |

то коэффициент ассоциации вычисляется по формуле:

 

коэффициент контингенции вычисляется по формуле:



**Лекция 5.**

***Теоретические основы проверки статистических гипотез. Нулеваяя и альтернативная гипотезы. Критерии для проверки гипотез. Понятие статистики критерия, уровня значимости, критической области.***

*Понятие нулевой и альтернативной гипотезы*

Поскольку статистика как метод исследования имеет дело с данными, в которых интересующие исследователя закономерности искажены различными случайными факторами, большинство статистических вычислений сопровождается проверкой некоторых предположений или гипотез об источнике этих данных.

Статистическая гипотеза – это предположение о свойствах случайных величин или событий, которое мы хотим проверить по имеющимся данным. Примеры статистических гипотез в педагогических исследованиях:

Нулевая гипотеза – это основное проверяемое предположение, которое обычно формулируется как отсутствие различий, отсутствие влияние фактора, отсутствие эффекта, равенство нулю значений выборочных характеристик и т.п. Примером нулевой гипотезы в педагогике является утверждение о том, что различие в результатах выполнения двумя группами учащихся одной и той же контрольной работы вызвано лишь случайными причинами.

Другое проверяемое предположение (не всегда строго противоположное или обратное первому) называется конкурирующей или альтернативной гипотезой. Так, для упомянутого выше примера гипотезы Н0 в педагогике одна из возможных альтернатив Н1 будет определена как: уровни выполнения работы в двух группах учащихся различны и это различие определяется влиянием неслучайных факторов, например, тех или других методов обучения.

Выдвинутая гипотеза может быть правильной или неправильной, поэтому возникает необходимость проверить ее. Так как проверку производят статистическими методами, то данная проверка называется статистической.

При проверке статистических гипотез возможны ошибки (ошибочные суждения) двух видов:

— можно отвергнуть нулевую гипотезу, когда она на самом деле верна (так называемая ошибка первого рода);

— можно принять нулевую гипотезу, когда она на самом деле не верна (так называемая ошибка второго рода).

Ошибка, состоящая в принятии нулевой гипотезы, когда она ложна, качественно отличается от ошибки, состоящей в отвержении гипотезы, когда она истинна. Эта разница очень существенна вследствие того, что различна значимость этих ошибок.

Допустимая вероятность ошибки первого рода (Ркр) может быть равна 5% или 1% (0.05 или 0.01).

Уровень значимости – это вероятность ошибки первого рода при принятии решения (вероятность ошибочного отклонения нулевой гипотезы).

Альтернативные гипотезы принимаются тогда и только тогда, когда опровергается нулевая гипотеза. Это бывает в случаях, когда различия, скажем, в средних арифметических экспериментальной и контрольной групп настолько значимы (статистически достоверны), что риск ошибки отвергнуть нулевую гипотезу и принять альтернативную не превышает одного из трех принятых уровней значимости статистического вывода:

первый уровень — 5% (р=5%); где допускается риск ошибки в выводе в пяти случаях из ста теоретически возможных таких же экспериментов при строго случайном отборе испытуемых для каждого экспе­римента;

второй уровень — 1%, т. е. соответственно допускается риск ошибиться только в одном случае из ста;

третий уровень — 0,1%, т. е. допускается риск ошибиться только в одном случае из тысячи.

Последний уровень значимости предъявляет очень высокие требования к обоснованию достоверности результатов эксперимента и потому редко используется. В педагогических исследованиях, не нуждающихся в очень высоком уровне достоверности, представляется разумным принять 5% уровень значимости.

Статистика критерия (Т) — некоторая функция от исходных данных, по значению которой проверяется нулевая гипотеза. Чаще всего статистика критерия является числовой функцией, но она может быть и любой другой функцией, например, многомерной функцией.

Всякое правило, на основе которого отклоняется или принимается нулевая гипотеза называется критерием для проверки данной гипотезы. Статистический критерий (критерий) – это случайная величина, которая служит для проверки статистических гипотез.

Критическая область – совокупность значений критерия, при котором нулевую гипотезу отвергают. Область принятия нулевой гипотезы (область допустимых значений) – совокупность значений критерия, при котором нулевую гипотезу принимают. При справедливости нулевой гипотезы вероятность того, что статистика критерия попадает в область принятия нулевой гипотезы должна быть равна 1-Ркр.

*Общие принципы проверки статистических гипотез*

Процедура проверки нулевой гипотезы в общем случае включает следующие этапы:

1. задается допустимая вероятность ошибки первого рода (Ркр=0,05)

2. выбирается статистика критерия (Т)

3. ищется область допустимых значений

4. по исходным данным вычисляется значение статистики Т

5. если Т (статистика критерия) принадлежит области принятия нулевой гипотезы, то нулевая гипотеза принимается (корректнее говоря, делается заключение, что исходные данные не противоречат нулевой гипотезе), а в противном случае нулевая гипотеза отвергается и принимается альтернативная гипотеза.[1] Это основной принцип проверки всех статистических гипотез.

Обычно первые три этапа выполняют профессиональные математики, а последние два – пользователи для своих частных данных.

В современных статистических пакетах на ЭВМ используются не стандартные уровни значимости, а уровни, подсчитываемые непосредственно в процессе работы с соответствующим статистическим методом. Эти уровни, обозначенные буквой P, могут иметь различное числовое выражение в интервале от 0 до 1, например, 0,7 0,23 0,012. Понятно, что в первых двух случаях полученные уровни значимости слишком велики и говорить о том, что результат значим нельзя. В последнем случае результаты значимы на уровне 12 тысячных. Это достоверный результат.

При проверке статистических гипотез с помощью статистических пакетов, программа выводит на экран вычисленное значение уровня значимости Р и подсказку о возможности принятия или неприятия нулевой гипотезы.

Если вычисленное значение Р превосходит выбранный уровень Ркр,

то принимается нулевая гипотеза, а в противном случае — альтернативная гипотеза. Чем меньше вычисленное значение Р, тем более исходные данные противоречат нулевой гипотезе.

Число степеней свободы у какого-либо параметра определяют как число опы­тов, по которым рассчитан данный параметр, минус количество одинаковых значений, найденных по этим опытам независимо друг от друга.

Величина Ф называется мощностью критерия и представляет собой вероятность отклонения неверной нулевой гипотезы, то есть вероятность правильного решения. Мощность критерия – вероятность попадания критерия в критическую область при условии, что справедлива альтернативная гипотеза. Чем больше Ф, тем вероятность ошибки 2-го рода меньше.

*Проверяем значимость коэффициента корреляции*. Для этого вычисляем значение величины

t выб=

Эта величина имеет распределение Стьюдента с числом степеней свободы n-2

Коэффициент корреляции значим, если

 | tвыб | > t Крит (α/2, n- 2)

В таблице приведены критические значения коэффициентов корреляции для уровня значимости α=0.05. Объем выбрки N= десятки\*10+ единицы. Если абсолютная величина выборочного коэффициента корреляции > критического значения, то коэффициент корреляции значим ( есть корреляционная зависимость).

|  |  |
| --- | --- |
|  | единицы |
| десятки |  | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 0 |  |  |  |  |  | 0.94 | 0.85 | 0.78 | 0.72 | 0.68 |
| 1 | 0.64 | 0.61 | 0.58 | 0.56 | 0.54 | 0.52 | 0.50 | 0.48 | 0.47 | 0.46 |
| 2 | 0.45 | 0.44 | 0.43 | 0.42 | 0.41 | 0.40 | 0.39 | 0.38 | 0.38 | 0.37 |
| 3 | 0.36 | 0.36 | 0.36 | 0.34 | 0.34 | 0.33 | 0.33 | 0.33 | 0.32 | 0.32 |
| 4 | 0.31 |  |  |  |  |  |  |  |  |  |

**Лекция 6.**

***Непараметрические критерии: критерий знаков, критерий Вилкоксона, критерий Манна-Уитни, критерий Фридмана.***

 Следующей задачей статистического анализа, решаемой после определения основных (выборочных) характеристик и анализа одной выборки, является совместный анализ нескольких выборок. Важнейшим вопросом, возникающем при анализе двух выборок, является вопрос о наличии различий между выборками. Обычно для этого проводят проверку статистических гипотез о принадлежности обеих выборок одной генеральной совокупности или о равенстве средних.

Если вид распределения или функция распределения выборки нам заданы, то в этом случае задача оценки различий двух групп независимых наблюдений может решаться с использованием **параметрических** **критериев** статистики: либо кри­терия Стьюдента (t), если сравнение выборок ведется по сред­ним значениям (X и У), либо с использованием критерия Фишера (F), если сравнение выборок ведется по их дисперсиям.

Использование параметрических критериев статистики без предварительной про­верки вида распределения может привести к определенным ошибкам в ходе проверки рабочей гипотезы.

Для преодоления указанных трудностей в практике психологических исследований следует использовать **непараметрические** **критерии** **статистики**, такие, как критерий знаков, критерий Вилкоксона, критерий Манна-Уитни.

**Непараметрические критерии статистики** *-* свободны от допущения о законе распределения выборок и базируются на предположении о независимости наблюдений.

 Сравнивая на глазок (по процентным соотношениям) результаты до и после какого-либо воздействия, исследователь приходит к заключению, что если наблюдаются различия, то имеет место различие в сравниваемых выборках. Подобный подход категорически неприемлем, так как для процентов нельзя определить уровень достоверности в различиях. Проценты, взятые сами по себе, не дают возможности делать статистически достоверные выводы. Чтобы доказать эффективность какого-либо воздействия, необходимо выявить статистически значимую тенденцию в смещении (сдвиге) показателей. Для решения подобных задач исследователь может использовать ряд критериев различия. Ниже будет рассмотрены непараметрические критерии: критерий знаков и критерий хи-квадрат.

***Критерий знаков (G-критерий)***

Критерий предназначен для срав­нения состояния некоторого свойства у членов двух **зави­симых** **выборок** на основе измерений, сделанных по шка­ле не ниже ранговой.

Имеется две серии наблюдений над случайными переменными X и У, полученные при рассмотрении двух **зависимых выборок**. На их основе составлено N пар вида (хi, уi), где хi, уi — результаты двукратного измерения одного и того же свойства у одного и того же объекта.

Элементы каждой пары хi, уi сравниваются между собой по величине, и паре присваивается знак **«+»**, ес­ли хi < уi , знак **«—»**, если хi > уi и **«0»**, если хi = уi.

**Нулевая гипотеза** формулируются следующим обра­зом: в состоянии изучаемого свойства нет значимых различий при первичном и вторичном измерениях. Альтернативная гипотеза: законы распределения величин X и У различны, т. е. состояния изучаемого свойства существенно раз­личны в одной и той же совокупности при первичном и вторичном измерениях этого свойства.

**Ста­тистика критерия** (Т) определяется следую­щим образом:

допустим, что из N пар (х, у,) нашлось несколько пар, в которых значения хi  и уi равны. Такие пары обозначаются знаком «0» и при подсчете значения ве­личины Т не учитываются. Предположим, что за вы­четом из числа N числа пар, обозначенных знаком «0», осталось всего n пар. Среди оставшихся n пар подсчита­ем число пар, обозначенных знаком «-», т.е, пары, в которых xi<yi. Значение величины Т и равно чис­лу пар со знаком минус.

Нулевая гипотеза принимается на уровне значимости 0,05, если наблю­даемое значение T<n-ta, где значение n-ta определя­ется из статистических таблиц для критерия знаков

**Лекция 7.**

**Однофакторный дисперсионный анализ**

 Дисперсионный анализ, предложенный Р. Фишером, является статистическим методом, предназначенным для выявления влияния ряда отдельных факторов на результаты экспериментов.

В основе дисперсионного анализа лежит предположение о том, что одни переменные могут рассматриваться как причины (факторы, независимые переменные), а другие как следствия (зависимые переменные). Независимые переменные называют иногда регулируемыми факторами именно потому, что в эксперименте исследователь имеет возможность варьировать ими и анализировать получающийся результат.

Сущность дисперсионного анализа заключается в расчлене­нии общей дисперсии изучаемого признака на отдельные компо­ненты, обусловленные влиянием конкретных факторов, и про­верке гипотез о значимости влияния этих факторов на исследуе­мый признак. Сравнивая компоненты дисперсии друг с другом посредством F — критерия Фишера, можно определить, какая доля общей вариативности результативного признака обусловле­на действием регулируемых факторов.

Исходным материалом для дисперсионного анализа служат данные исследования трех и более выборок, которые могут быть как **равными**, так и **неравными** по численности, как **связными**, так и **несвязными**. По количеству выявляемых регулируемых фак­торов дисперсионный анализ может быть **однофакторным** (при этом изучается влияние одного фактора на результаты экспери­мента), **двухфакторным** (при изучении влияния двух факторов) и **многофакторным** (позволяет оценить не только влияние каждого из факторов в отдельности, но и их взаимодействие)**.**

Дисперсионный анализ относится к группе параметрических методов и поэтому его следует применять только тогда, когда доказано, что распределение является нормальным. (Суходольский Г.В., 1972; Шеффе Г., 1980).

***Однофакторный дисперсионный анализ для несвязанных выборок***

Изучается действие только одной переменной (фактора) на исследуемый признак. Исследователя интересует вопрос, как изменяется определенный признак в разных условиях действия переменной (фактора). Например, как изменяется время решения задачи при разных условиях мотивации испытуемых (низкой, средней, высокой мотивации) или при разных способах предъявления задачи (устно, письменно или в виде текста с графиками и иллюстрациями), в разных условиях работы с задачей (в одиночестве, в комнате с преподавателем, в классе). В первом случае фактором является мотивация, во втором – степень наглядности, в третьем – фактор публичности.[[1]](#footnote-1)

В данном варианте метода влиянию каждой из градаций подвергаются разные выборки испытуемых. Градаций фактора должно быть не менее **трех**.

**Пример 1.** Три различные группы из шести испытуемых получили списки из десяти слов. Первой группе слова предъявлялись с низкой скоростью -1 слово в 5 секунд, второй группе со средней скоростью - 1 слово в 2 секунды, и третьей группе с большой скоростью - 1 слово в секунду. Было предсказано, что показатели воспроизведения будут зависеть от скорости предъявления слов. Результаты представлены в табл. 1.

Таблица 1. Количество воспроизведенных слов (по J. Greene, M D'Olivera, 1989, p. 99)

|  |  |  |  |
| --- | --- | --- | --- |
| № испытуемого  | Группа 1 низкая скорость  | Группа 2 средняя скорость  | Группа 3 высокая скорость  |
| 1  | 8  | 7  | 4  |
| 2  | 7  | 8  | 5  |
| 3  | 9  | 5  | 3  |
| 4  | 5  | 4  | 6  |
| 5  | 6  | 6  | 2  |
| 6  | 8  | 7  | 4  |
| суммы  | 43  | 37  | 24  |
| средние  | 7,17  | 6,17  | 4,00  |
| Общая сумма  | 104  |    |    |

Дисперсионный однофакторный анализ позволяет проверить гипотезы:

H0: различия в объеме воспроизведения слов *между* группами являются не более выраженными, чем случайные различия *внутри* каждой группы

H1: Различия в объеме воспроизведения слов *между* группами являются более выраженными, чем случайные различия *внутри* каждой группы.

*Последовательность операций в однофакторном дисперсионном анализе для несвязанных выборок:*

1. подсчитаем **SSфакт** - вариативность признака, обусловленную действи­ем исследуемого фактора. Часто встречающееся обозначе­ние SS - сокращение от "суммы квадратов" (sum of squares). Это со­кращение чаще всего используется в переводных источниках (см., на­пример: Гласс Дж., Стенли Дж., 1976).

, (1)

где Тс – сумма индивидуальных значений по каждому из условий. Для нашего примера 43, 37, 24 (см. табл. 1);

с – количество условий (градаций) фактора (=3);

n – количество испытуемых в каждой группе (=6);

N – общее количество индивидуальных значений (=18);

- квадрат общей суммы индивидуальных значений (=1042=10816)

Отметим разницу между , в которой все индивидуальные значения сначала возводятся в квадрат, а потом суммируются, и , где индивидуальные значения сначала суммируются для получения об­щей суммы, а потом уже эта сумма возводится в квадрат.

По формуле (1) рассчитав фактическую вариативность признака, получаем:



2. подсчитаем **SSобщ** – общую вариативность признака:

 (2)

3. подсчитаем случайную (остаточную) величину **SSсл**, обусловленную неучтенными факторами:

 (3)

4.      **число степеней свободы** равно:

=3-1=2 (4)

 

5.      **«средний квадрат»** или **математическое ожидание суммы квадратов,** усредненная величина соответствующих сумм квадратов SS равна:

**** (5)

****

6.      значение **статистики критерия** **Fэмп** рассчитаем по формуле:

 (6)

Для нашего примера имеем*: Fэмп=15,72/2,11=7,45*

7.      определим **Fкрит** по статистическим таблицам **Приложения 3** для df1=k1=2 и df2=k2=15 табличное значение статистики равно 3,68

8. если **Fэмп<** Fкрит, то нулевая гипотеза принимается, в противном случае принимается альтернативная гипотеза. Для нашего примера **Fэмп** > Fкрит (7.45>3.68), следовательно принимается альтернативная гипотеза.

**Вывод:** различия в объеме воспроизведения слов между группами являются более выраженными, чем случайные различия внутри каждой группы (р<0,05). Т.о. скорость предъявления слов влияет на объем их воспроизведения.

**Разработчик : Н.М. Саукова, к.п.н., доцент кафедры информационных технологий в образовании МПГУ**

1. [↑](#footnote-ref-1)