Министерство науки и высшего образования РФ

Федеральное государственное бюджетное образовательное учреждение

высшего образования

«Томский государственный университет систем управления

и радиоэлектроники» (ТУСУР)

Кафедра автоматизированной обработки информации (АОИ)

**ОДНОФАКТОРНЫЙ ДИСПЕРСИОННЫЙ АНАЛИЗ**

Отчет по лабораторной работе № 4

По дисциплине «Теория вероятностей и

математическая статистика»

Вариант №2

Выполнил:

Студент гр. 429-3

Бабец А.А.

Принял:

Доцент каф. Математики

Лугина Н.Э.

Томск 2021

# Постановка задачи

Разработчики игр проводят исследование: «Влияние класса персонажа на скорость прохождения подземелий». Присутствуют шесть игроков, которые проходят одно и тоже подземелье (на персонаже подходящего уровня и с подходящей экипировкой) на четырёх последовательных этапах игры, после 10, 25, 50 и 100 часов, проведённых в игре соответственно. В таблицу выведено время прохождения подземелья в минутах.

Фактор F – класс персонажа. Наблюдение – время прохождения подземелья.

Н0: класс персонажа не влияет существенно на время прохождения подземелья; Df = Deps.

Н1: класс персонажа влияет на скорость прохождения подземелья; Df > Deps.

Полученные данные оценки скорости прохождения представлены в таблице:

Таблица 1.1 - Данные

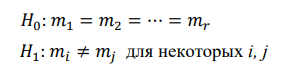
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Этап | Игрок | 1 | 2 | 3 | 4 | 5 | 6 |
| 1-ый этап | | 53 | 52 | 45 | 43 | 48 | 54 |
| 2-ой этап | | 49 | 55 | 38 | 47 | 46 | 49 |
| 3-ий этап | | 48 | 50 | 51 | 54 | 52 | 46 |
| 4-ый этап | | 44 | 47 | 46 | 49 | 50 | 42 |

Кол-во факторов r = 4, кол-во наблюдений k = 6, общее кол-во наблюдений n = 24, α примем за 0,01.

# Классическая схема

## Теория

В классической схеме однофакторного дисперсионного анализа на некую СВ Х действует некий фактор F имеющий r уровней. С помощью дисперсионного анализа можно дать ответ на вопрос: «Влияет ли данный фактор на эту СВ?». На каждом уровне производится ряд наблюдений исследуемого признака, рассматриваемых как независимые выборочные значения из генеральных совокупностей Х1, Х2, …, Хr, распределённых по нормальному закону с одинаковыми, хоть неизвестными дисперсиями и мат. ожиданиями. При этом предполагается, что ошибки наблюдений распределены нормально с нулевым математическим ожиданием и одинаковыми дисперсиями. Задача дисперсионного анализа формулируется как задача о равенстве всех математических ожиданий



Основная идея дисперсионного анализа состоит в переходе от задачи 1 к эквивалентной ей задаче сравнения дисперсий: «факторной» дисперсии, оценивающей разброс, вносимый в результате воздействия фактора F и «остаточной» дисперсии, оценивающей разброс, возникший в результате случайных причин.

Основные требования к экспериментальным данным:

1. результаты наблюдений на разных уровнях рассматриваются как независимые выборочные значения из генеральных совокупностей, распределенных по нормальному закону
2. с одинаковыми 𝜎1 = 𝜎2 = ⋯ = 𝜎𝑟, хотя и неизвестными дисперсиями и математическими ожиданиями 𝑚1, 𝑚2, …, 𝑚𝑟.
3. математическое ожидание возмущения 𝜀𝑖𝑗 для любого i равно нулю: 𝑀[𝜀𝑖𝑗] = 0; возмущения 𝜀𝑖𝑗 взаимно независимы; дисперсия возмущения 𝜀𝑖𝑗 (или переменной 𝑥𝑖𝑗) постоянна для любых i, j, то есть 𝐷[𝜀𝑖𝑗] = 𝜎2; возмущение 𝜀𝑖𝑗 (или переменной 𝑥𝑖𝑗) имеет нормальный закон распределения 𝜀𝑖𝑗~𝑁(0, 𝜎𝜀 ).

Схема решения:

1. формулируем основную и альтернативную гипотезы (сформулированы выше);
2. задаём уровень значимости α;
3. вычисляем ;
4. находим границу критической области , критическая область правосторонняя ;
5. делаем вывод о принятии или отвержении гипотезы.

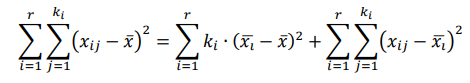
## Решение

Сначала находим средние значения по уровням фактора и по всем данным (используем функцию Microsoft EXCEL СРЗНАЧ).

Таблица 2.2.1 – Средние значения

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Номер уровня** | | **Номер наблюдения** | | | | | | **Среднее по уровням** |
| 1 | 2 | 3 | 4 | 5 | 6 |
| **Уровни фактора F** | 1 | 53 | 52 | 45 | 43 | 48 | 54 | 49,16666667 |
| 2 | 49 | 55 | 38 | 47 | 46 | 49 | 47,33333333 |
| 3 | 48 | 50 | 51 | 54 | 52 | 46 | 50,16666667 |
| 4 | 44 | 47 | 46 | 49 | 50 | 42 | 46,33333333 |
| **Среднее значение по всей таблице** | | | | | | | | 48,25 |

Для этого сумма квадратов отклонений наблюдаемых значений от их общего среднего по всей таблице 𝑥̅разлагается на две части в виде основного дисперсионного тождества.



В этом тождестве левая часть обозначается 𝑄2общ и служит для оценки общего разброса в наблюдаемых данных; первое слагаемое в правой части – 𝑄2𝐹 – для оценки «факторного» разброса; второе слагаемое в правой части – 𝑄2𝜀 – для оценки «случайного» разброса в данных.



Таким образом,

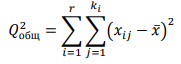
𝑄2общ – общая или полная сумма квадратов отклонений;

𝑄2𝐹 – сумма квадратов отклонений групповых средних от общей средней, или межгрупповая (факторная) сумма квадратов отклонений;

𝑄2𝜀 – сумма квадратов отклонений наблюдений от групповых средних, или внутригрупповая (остаточная)сумма квадратов отклонений.

В основном дисперсионном тождестве заключена основная идея дисперсионного анализа. Если поделить обе части равенство на число наблюдений, то получим рассмотренное ранее в теории вероятностей правило сложения дисперсий. В дисперсионном анализе анализируются не сами суммы квадратов отклонений, а так называемые средние квадраты, являющиеся несмещенными оценками соответствующих дисперсий, которые получаются делением сумм квадратов на соответствующее число степеней свободы.

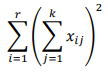
Находим дисперсии:

1. Общая дисперсия 
2. Межгрупповая дисперсия
3. Внутригрупповая дисперсия 

Упростим формулы выше:

1. 
2. 
3. 

Для упрощения вычислений по этим формулам находим следующие суммы:

1. 
2. 
3. 

Вычисление данных сумм на рисунке 2.2.1

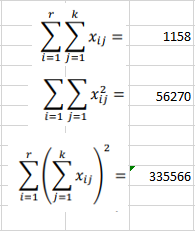


Рисунок 2.2.1 – Вычисление сумм

По этим суммам вычисляем компоненты дисперсии

Таблица 2.2.2 – Компоненты дисперсии

|  |  |
| --- | --- |
| Компоненты дисперсии | Сумма квадратов |
|
| Общая(T) | 396,5 |
| Межгрупповая(F) | 54,16666667 |
| Внутригрупповая(R) | 342,3333333 |

Напомним определение. Число степеней свободы определяется как общее число наблюдений минус число связывающих их уравнений. Поэтому для среднего квадрата 𝑆2𝐹, являющегося несмещенной оценкой межгрупповой дисперсии, число степеней свободы 𝑟 − 1, так как при его расчете имеется r групповых средних, связанных между собой одним (1) уравнением (формула для расчета групповых средних 𝑥̅𝑖 = 1 𝑘𝑖 ∑ 𝑘𝑖 𝑗=1 𝑥𝑖𝑗). А для среднего квадрата 𝑆2𝑅, являющегося несмещенной оценкой внутригрупповой дисперсии, число степеней свободы 𝑛 − 𝑟, так как при расчете имеется n наблюдений связаны между собой уравнениями .

Таким образом, могут быть найдены несмещенные оценки межгрупповой и внутригрупповой дисперсий: .

В данном случае оценки равны 10,8 и 17,1 соответственно.

Находим Fнабл по формуле (с помощью функции excel F.ОБР.ПХ):



Получилось значение равное 0,63.

Находим границу критической области по параметрам α, r-1, n-r



Получаем значение равное 4,1.

Вывод: так как Fнабл < Fкрит, то нет оснований отвергать основную гипотезу, следовательно класс персонажа не влияет существенно на скорость прохождения подземелья.

# Критерий Крускала-Уоллиса

## Теория

Классическая схема однофакторного дисперсионного анализа требует

* нормальности распределения исследуемого признака;
* нормальности распределения ошибок наблюдений.

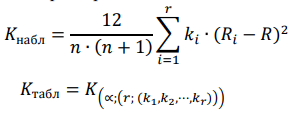
Если эти условия не выполняются, то для решения задачи о равенстве средних более предпочтительными являются непараметрические методы.

Критерий Крускала – Уоллиса позволяет проверить гипотезу о равенстве средних с минимальными требованиями к выборочным данным.

Предполагается, что ошибки наблюдений 𝜀𝑖𝑗

* независимы;
* имеют непрерывное распределение.

Алгоритм проверки основной гипотезы H0 о равенстве средних

1. Выборочные значения 𝑥𝑖𝑗 заменяют их рангами 𝑥𝑖𝑗′. Напомним, что ранг – это число, соответствующее порядковому номеру наблюдаемого значения в выборке.
2. Для каждого уровня вычисляется средний ранг  и сравнивается с общим средним рангом, который в предположении справедливости гипотезы H0, равен .
3. Средний ранг сравниваем с общим средним рангом 
4. Для сравнения используется статистика Крускала–Уоллиса, критические точки которой приведены в таблице.   
   

Критическая область в данной задаче тоже является правосторонней.

## Решение

Сначала вычислим ранги с помощью функции Microsoft EXCEL РАНГ.СР, также находим общий ранг, средние ранги и разности среднего ранга уровня и общего ранга (разности находим для упрощения дальнейших вычислений) (таблица 3.2.1).

Таблица 3.2.1 – Ранги

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Уровни** | **Ранги** | | | | | | **Средний ранг** | **(Ri-R)** |
| **1** | **2** | **3** | **4** | **5** | **6** |
| **1** | 21 | 19,5 | 5 | 3 | 11,5 | 22,5 | 13,75 | 1,25 |
| **2** | 14 | 24 | 1 | 9,5 | 7 | 14 | 11,58333333 | -0,916666667 |
| **3** | 11,5 | 16,5 | 18 | 22,5 | 19,5 | 7 | 15,83333333 | 3,333333333 |
| **4** | 4 | 9,5 | 7 | 14 | 16,5 | 2 | 8,833333333 | -3,666666667 |

Общий ранг равен 12,5.

Kкрит вычисляем с помощью функции excel ХИ2.ОБР.ПХ.

Kнабл и Kкрит равны 3,36 и 11,34 соответственно.

Вывод: так как Kнабл < Kкрит, то нет оснований отвергать основную гипотезу, следовательно класс персонажа не влияет существенно на скорость прохождения подземелья.