Министерство науки и высшего образования РФ

Севастопольский государственный университет

Кафедра информационных систем

ОТЧЕТ

по лабораторной работе №3

Задача дисперсионного анализа. Методы дисперсионного анализа. Однофакторный дисперсионный анализ

по дисциплине «Интеллектуальный анализ данных»

Выполнил:

Студент группы ИС/б 17-2-о

Черняев Н.Г.

Проверил:

Сырых О.А.

г. Севастополь 2020

1.Цель работы

Приобрести практические навыки в проведении дисперсионного анализа по экспериментальным данным. Исследовать возможности языка R для проведения дисперсионного анализа.

2.Постановка задачи

* Создать набор данных согласно варианту;
* Провести однофакторный дисперсионный анализ в среде Rcmdr;
* Построить диаграмму, отображающую средние значения и их доверительные интервалы для каждой группы;
* По результатам дисперсионного анализа сформулировать выводы;

3.Ход работы

1. Выполним дисперсионный анализ в Excel для своего варианта.

В качестве фактора была выбрана переменная “Расходы на рекламу”, исследуется ее влияние на переменную “Объем выручки”. Фактор разбит на следующие группы:

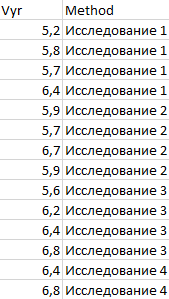


Рисунок 3.1 – Разбиение на группы

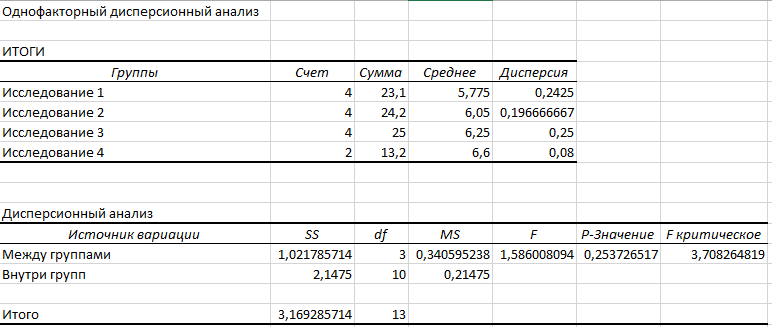


Рисунок 3.2 – Результаты однофакторного дисперсионного анализа по варианту

Сравнение средних значений показывает, что Исследование 4 позволяет добиться лучшего результата, а Исследование 1 наименее эффективно.

Сравнение F и F-критическое показывает, что F < F-критическое, следовательно, отвергнута гипотеза Н1 и принята гипотеза Н0 и с вероятностью ошибки α = 0,2537 можно утверждать, что влияние рассматриваемого фактора на результативный признак не существенно.

Выполним дисперсионный анализ средствами языка R для своего варианта.

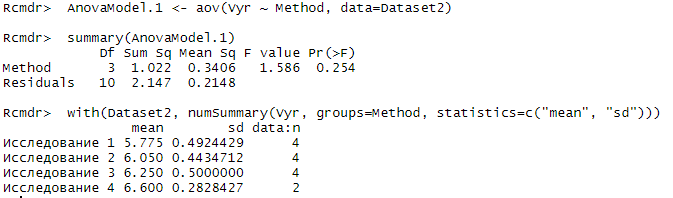


Рисунок 3.3 – Результаты дисперсионного анализа в среде Rcmdr по варианту

В столбце F value представлено рассчитанное по имеющимся данным значение F-критерия, он равен 1.586. В столбце Pr (>F) представлена вероятность получить F-значение, равное или превышающее то значение, которое в действительности рассчитали по имеющимся выборочным данным. Как видно, эта вероятность равна 0.254 и превышает 5%-ный уровень значимости, в связи с чем мы заключаем, что нулевая гипотеза верна. Таким образом, можно утверждать, что экспериментальные условия оказали не существенное влияние на результативный признак.

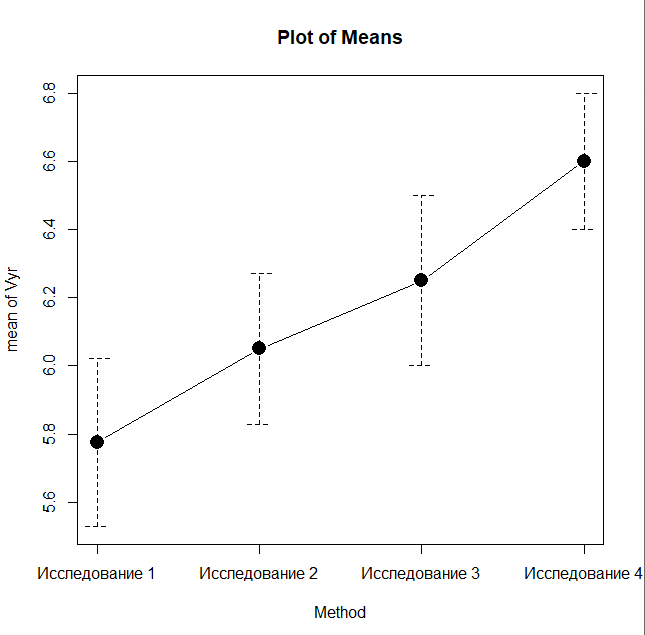


Рисунок 3.4 – Диаграмма, отображающая средние значения и их доверительные интервалы

Данная диаграмма визуально подтверждает, что Исследование 4 позволяет добиться лучшего результата, а Исследование 1 наименее эффективно.

1. Выполним дисперсионный анализ средствами языка R для своих данных.

В качестве фактора была выбрана переменная “Индекс качества жизни”, исследуется ее влияние на переменную “Покупательная способность”. Фактор разбит на следующие группы:

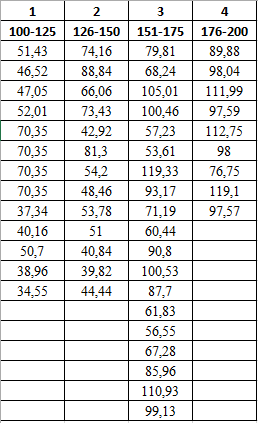


Рисунок 3.5 – Разбиение на группы

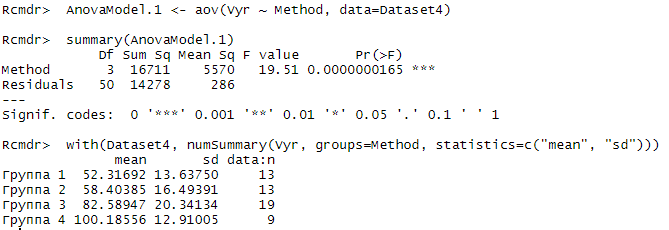


Рисунок 3.6 – Результаты дисперсионного анализа в среде Rcmdr для своих данных

В столбце F value представлено рассчитанное по имеющимся данным значение F-критерия, он равен 19.51. В столбце Pr (>F) представлена вероятность получить F-значение, равное или превышающее то значение, которое в действительности рассчитали по имеющимся выборочным данным. Как видно, эта вероятность равна 1,65\*10^-8 и не превышает 5%-ный уровень значимости, в связи с чем мы заключаем, что нулевая гипотеза не верна. Таким образом, можно утверждать, что экспериментальные условия оказали существенное влияние на результативный признак.

Сравнение средних значений показывает, что Группа 4 позволяет добиться лучшего результата, а Группа 1 наименее эффективна.

Построим диаграмму, которая визуально это подтверждает:

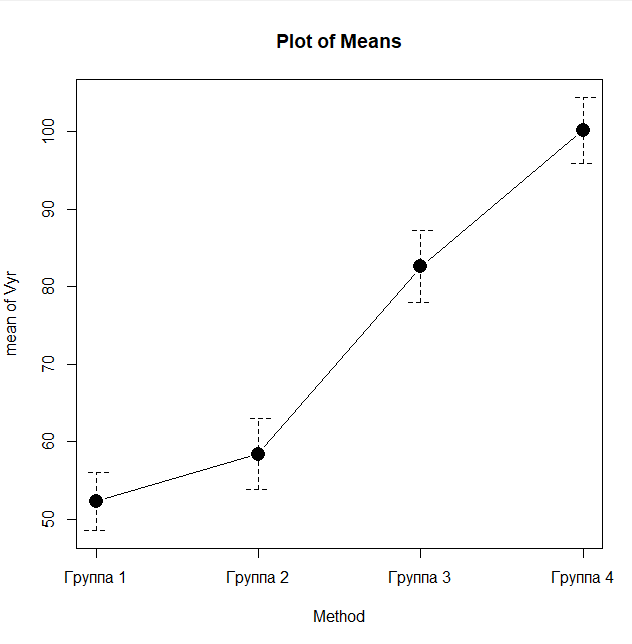


Рисунок 3.7 – Диаграмма, отображающая средние значения и интервалы

4.Вывод

В ходе выполнения лабораторной работы были приобретены практические навыки в проведении дисперсионного анализа по экспериментальным данным. Исследованы возможности языка R для проведения дисперсионного анализа.

В ходе анализа своих данных были получены следующие результаты:

* Группа 4 позволяет добиться лучшего результата, а Группа 1 наименее эффективна;
* F > F-критическое, следовательно, отвергнута гипотеза Н0 и принята гипотеза Н1 и с вероятностью ошибки α = 1,65\*10^-8 можно утверждать, что влияние рассматриваемого фактора на результативный признак существенно.