Министерство образования и науки Российской Федерации

Севастопольский государственный университет

Кафедра ИС

Отчет

По дисциплине: “Интеллектуальный анализ данных”

Лабораторная работа №3

“Задача дисперсионного анализа. Методы дисперсионного анализа.

Однофакторный дисперсионный анализ”

Выполнил:

ст.гр. ИС/б-32

Долженко И.А.

Проверила:

Сырых О.А.

Севастополь

2020

1 ЦЕЛЬ РАБОТЫ

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

2 ХОД РАБОТЫ

1. Выполним дисперсионный анализ в Excel для своего варианта (Вариант 6).

В качестве фактора была выбрана переменная “Цена на единицу продукции”, исследуется ее влияние на переменную “Объем продаж”. Фактор разбит на следующие группы:

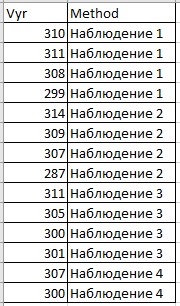


Рисунок 1 – Разбиение на группы

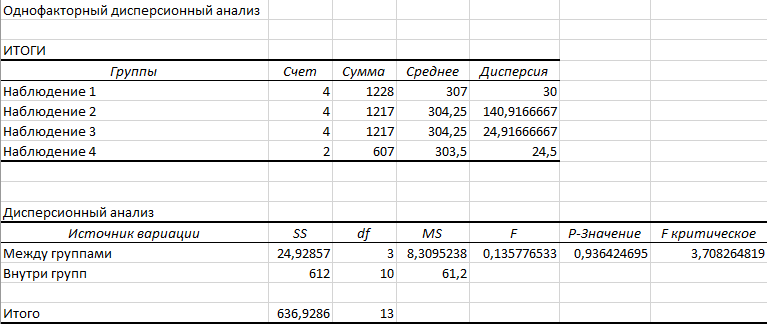


Рисунок 2 – Результаты однофакторного дисперсионного анализа по варианту

Сравнение средних значений показывает, что Наблюдение 1 позволяет добиться лучшего результата, а Наблюдение 4 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 = 0,9364 можно утверждать, что влияние рассматриваемого фактора на результативный признак не существенно.

2. Выполним дисперсионный анализ средствами языка R для своего варианта.

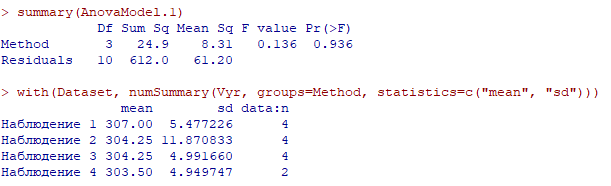


Рисунок 3 – Результаты дисперсионного анализа в среде Rcmdr по варианту

В столбце F value представлено рассчитанное по имеющимся данным значение F-критерия, он равен 0.136. 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 0.936 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.

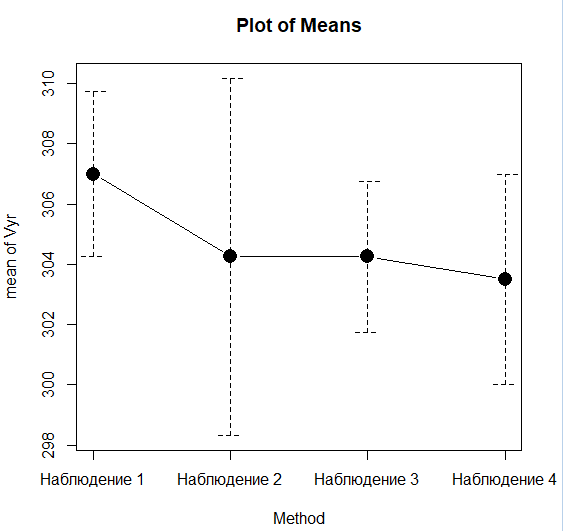


Рисунок 4 – Диаграмма, отображающая средние значения

и их доверительные интервалы

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

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

В качестве фактора была выбрана переменная “Индекс стоимости жизни”, исследуется ее влияние на переменную “Индекс покупательной способности”. Фактор разбит на следующие группы:

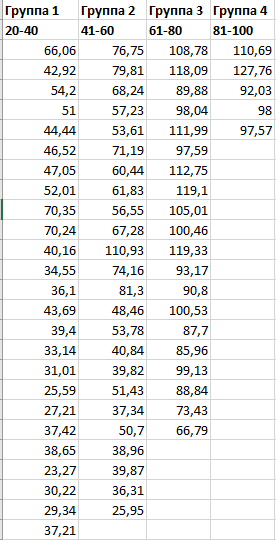


Рисунок 5 – Разбиение на группы

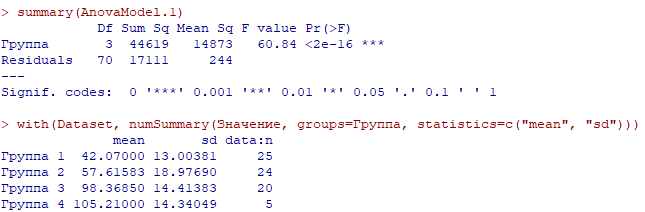


Рисунок 6 – Результаты дисперсионного анализа в среде Rcmdr

для своих данных

В столбце F value представлено рассчитанное по имеющимся данным значение F-критерия, он равен 60.84. 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 2e^-16 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:

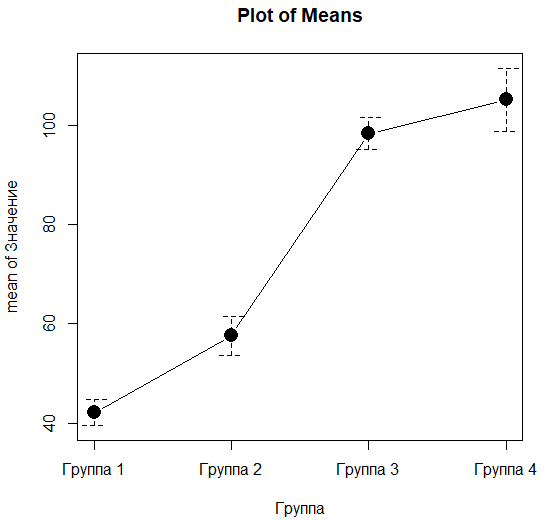


Рисунок 7 – Диаграмма, отображающая средние значения и интервалы

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 = 2e^-16 можно утверждать, что влияние рассматриваемого фактора на результативный признак существенно.