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.

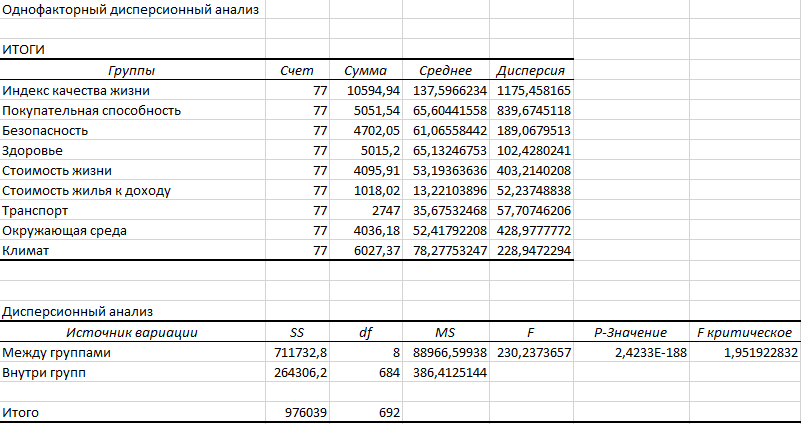


Рисунок 1 – Результаты однофакторного дисперсионного анализа

Сравнение средних значений показывает, что Индекс качества жизни позволяет добиться лучшего результата, а Индекс стоимости жилья к доходу наименее эффективен.

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

2. Выполним дисперсионный анализ средствами языка R.

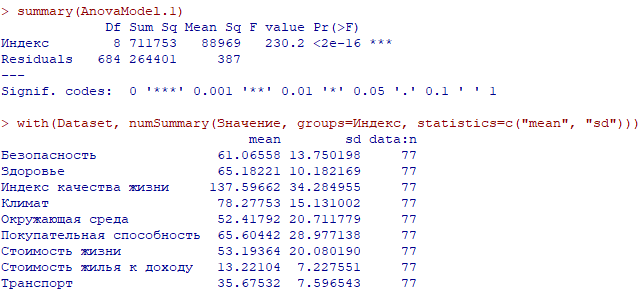


Рисунок 2 – Результаты дисперсионного анализа в среде Rcmdr

В столбце F value представлено рассчитанное по имеющимся данным значение F-критерия, он равен 230.2. 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 не высокая, равна 2^-16. 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 Индекс качества жизни позволяет добиться лучшего результата, а Индекс стоимости жилья к доходу наименее эффективен.

3. Построим диаграмму, отображающую средние значения и их доверительные интервалы для каждой группы.

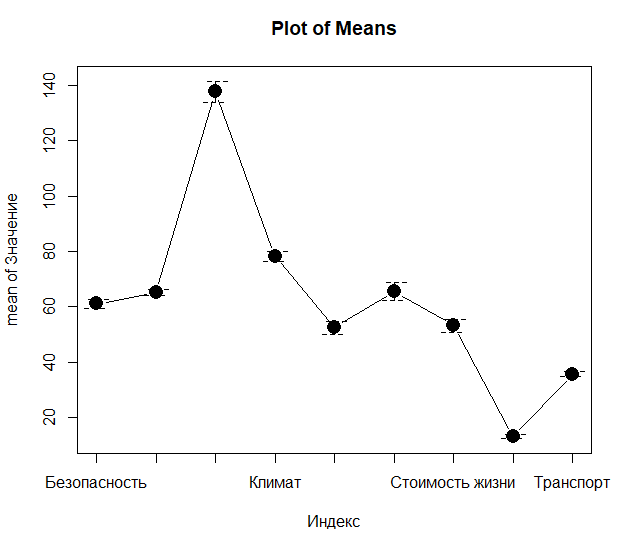


Рисунок 3 – Диаграмма, отображающая средние значения

и их доверительные интервалы

Данная диаграмма визуально подтверждает, что Индекс качества жизни позволяет добиться лучшего результата, а Индекс стоимости жилья к доходу наименее эффективен.

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 были получены следующие результаты:

* индекс качества жизни позволяет добиться лучшего результата, а Индекс стоимости жилья к доходу наименее эффективен;
* F > F-критическое, следовательно, отвергнута гипотеза Н0 и принята гипотеза Н1 и с вероятностью ошибки α = 0 можно утверждать, что влияние рассматриваемого фактора на результативный признак существенно.