Министерство образования и науки Российской Федерации

Севастопольский государственный университет

Институт информационных технологий

Кафедра ИС

# ОТЧЁТ

по лабораторной работе №3

ЗАДАЧА ДИСПЕРСИОННОГО АНАЛИЗА. МЕТОДЫ ДИСПЕРСИОННОГО

АНАЛИЗА. ОДНОФАКТОРНЫЙ ДИСПЕРСИОННЫЙ АНАЛИЗ

Выполнил:

ст. гр. ИС/б-21-2-о

Мовенко К. М.

Проверила:

Сырых О. А.

Севастополь

2024

# ЦЕЛЬ РАБОТЫ

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

# ЗАДАНИЕ

* 1. Создать файл с исходными данными о результатах исследования влияния цены за единицу продукции на объем продаж (шт.) в месяц (таблица 1) в MS Excel;
  2. Запустить “Пакет анализа”. Провести однофакторный дисперсионный анализ входных данных;
  3. Провести однофакторный дисперсионный анализ тех же данных в среде R Commander. Построить диаграмму, отображающую средние значения и их доверительные интервалы для каждой группы;
  4. Выполнить пункты 2.2-2.3 для выборок из собственных экспериментальных данных, рассмотренных в ЛР №2.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Номер наблюдения | Цена за единицу продукции (руб.) | | | |
| 1000-1100 | 1100-1200 | 1200-1300 | 1300-1500 |
| 1 | 310 | 311 | 308 | 299 |
| 2 | 314 | 309 | 307 | 287 |
| 3 | 311 | 305 | 300 | 301 |
| 4 |  | 307 |  | 300 |

Таблица 1 – Входные данные (Вариант − 6)

# АНАЛИЗ ДАННЫХ ПО ВАРИАНТУ

## Дисперсионный анализ в MS Excel

Данные были перенесены в табличный процессор MS Excel (рисунок 1).

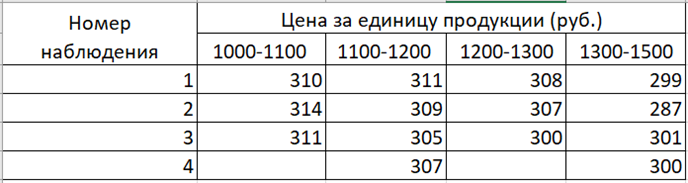


Рисунок 1 – Входные данные

В качестве фактора был выбран параметр «Цена на единицу продукции», в качестве результата – объём продаж за месяц. Фактор разбит на четыре группы. С помощью анализа данных Excel был проведён однофакторный дисперсионный анализ приведённых выборок при (рисунок 2).

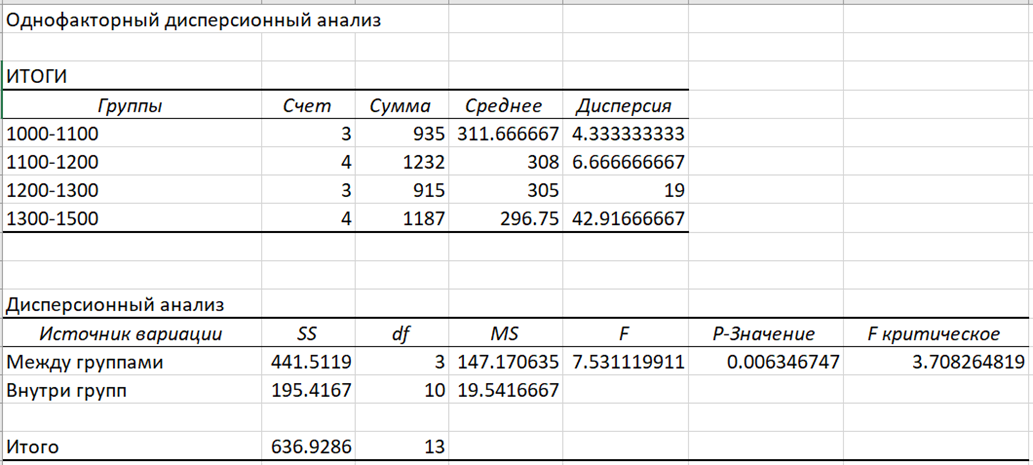


Рисунок 2 − Результаты однофакторного дисперсионного анализа по варианту

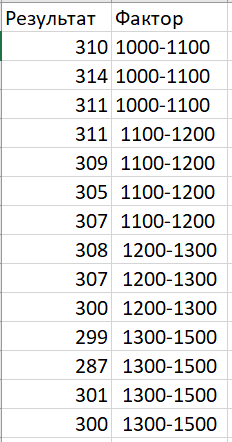
Средние значения по выборкам имеют небольшое различие, прослеживается тенденция на снижение объёма продаж при увеличении стоимости. Лучший результат получен при цене 1000-1100, худший при 1300-1500.

Так же при увеличении цен заметен рост дисперсии выборки. При меньшем диапазоне (1000-1100) дисперсия минимальна (4.33), при 1300-1500 сильно возрастает (42.9), что также, вероятно, связано с большей величиной диапазона.

Поскольку , нулевая гипотеза может быть отвергнута и можно с вероятностью ошибки 0.05 принять гипотезу, что влияние фактора (цена продукции) на результирующий признак (объём продаж) существенно.

## Дисперсионный анализ средствами языка R

Данные в файле Excel были преобразованы в форму, доступную для считывания языком R (рисунок 3). В первом столбце значения объёма выручки (результирующего признака), во втором – соответствующие выборки (значения фактора).

  
Рисунок 3 – Преобразование данных

В среде R Commander был проведён однофакторный дисперсионный анализ указанных данных (рисунок 4).

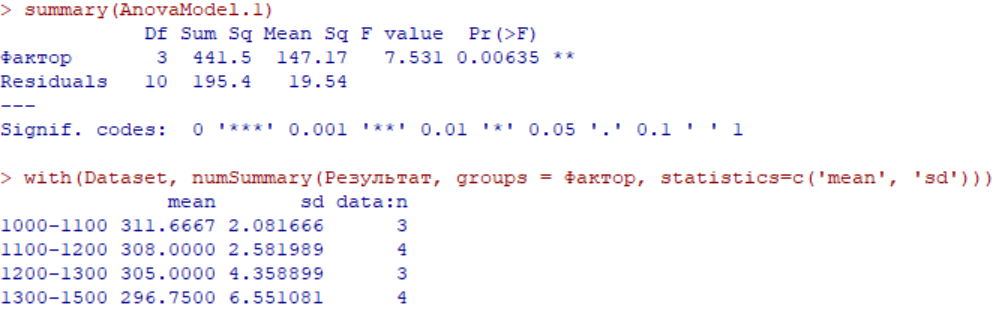


Рисунок 4 – Однофакторный дисперсионный анализ в R

Рассчитанное значение F-критерия равно 7.531. Вероятность получить равна 0.006 (0.6%), что сильно ниже выбранного уровня значимости (5%). Это позволяет отвергнуть нулевую гипотезу.

Была построена диаграмма, отображающая средние значения и их доверительные интервалы для каждой группы (рисунок 5). Сравнение средних указывает, что при цене 1000-1100 объём продаж наиболее высок, а при цене 1300-1500 наиболее низок.

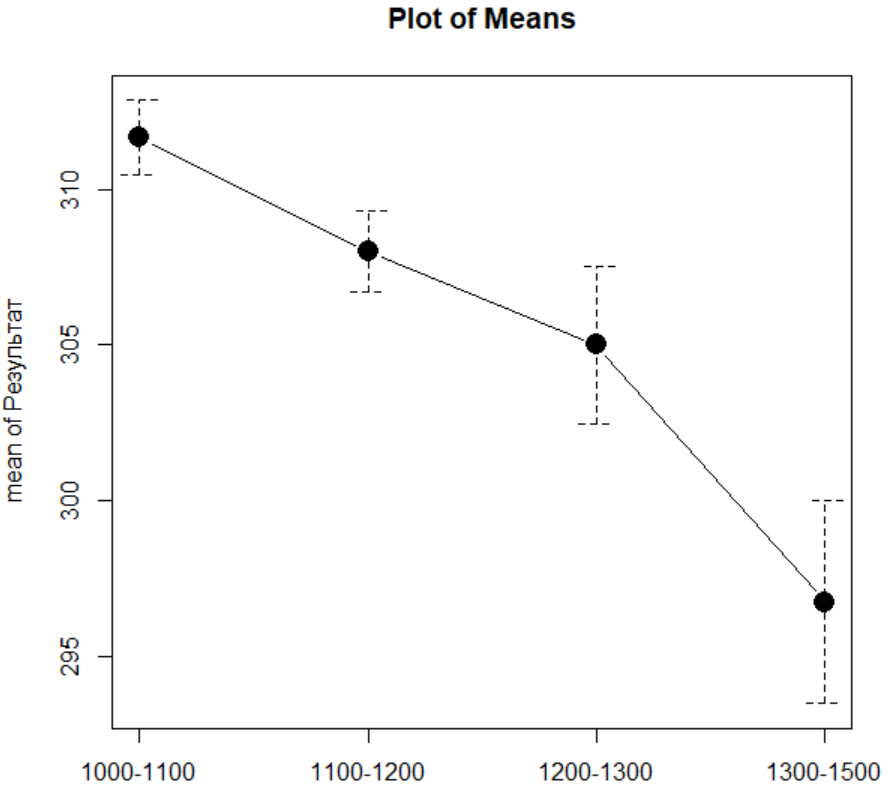


Рисунок 5 – Диаграмма средних значений

# АНАЛИЗ СОБСТВЕННЫХ ДАННЫХ

## Дисперсионный анализ в MS Excel

Были рассмотрены данные об индексе качества уровня жизни из ЛР №2. В качестве фактора был взят коэффициент климата, в качестве результирующего признака – индекс качества жизни. Генеральная совокупность была разбита на 4 выборки по диапазону климата (рисунок 6).

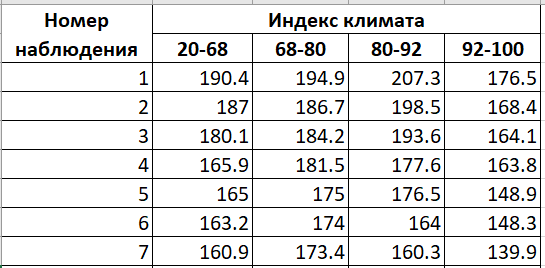


Рисунок 6 – Выборки по диапазону климата

В среде Excel был проведён однофакторный дисперсионный анализ полученных данных при α=0.05 (рисунок 7).

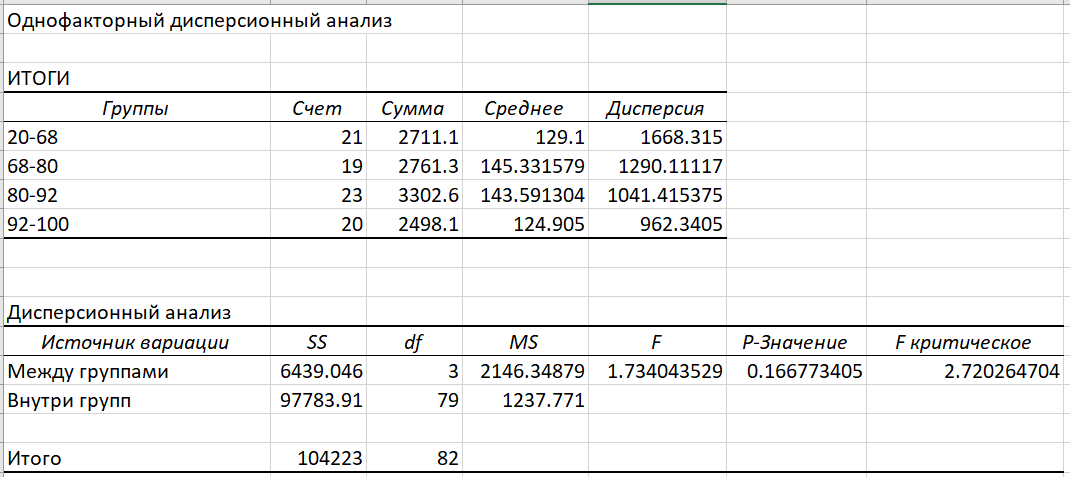


Рисунок 7 – Результаты однофакторного дисперсионного анализа собственных данных

Средние значения по выборкам имеют существенное различие, однако чёткой тенденции не прослеживается. Лучшие средние значения качества жизни находятся в средних диапазонах 68-80 и 80-92, в то время как при крайних значениях климата 20-68 и 92-100 они сильно ниже.

При увеличении климата заметен рост дисперсии выборки. Это связано с тем, что подавляющее число результатов принадлежит выборкам в диапазоне 80-100, группирование неравномерное по диапазонам климата.

Поскольку (1.73<2.72), нулевая гипотеза не может быть отвергнута. Оснований утверждать, что индекс качества жизни существенно зависит от климата, нет.

## Дисперсионный анализ средствами языка R

Данные были приведены в вид, доступный для анализа средствами языка R (рисунок 8). Затем в среде R Commander был проведён их однофакторный дисперсионный анализ (рисунок 9).



Рисунок 8 – Преобразование данных

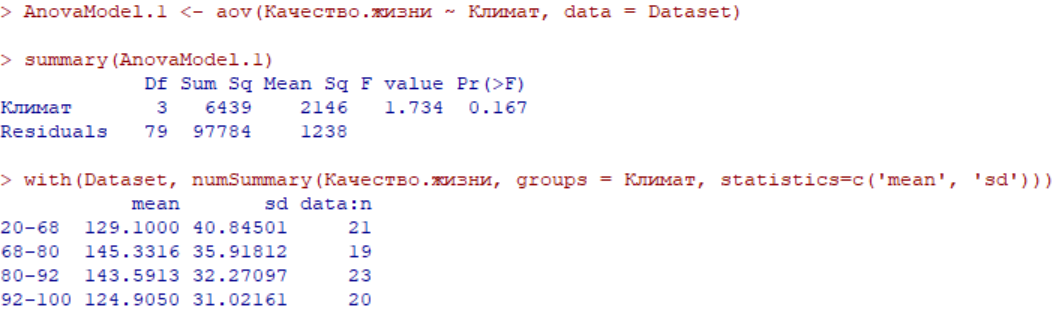


Рисунок 9 − Однофакторный дисперсионный анализ собственных данных в R

Рассчитанное значение F-критерия равно 1.73. Вероятность получить равна 0.167 (16%), что сильно выше выбранного уровня значимости (5%). Нулевая гипотеза не может быть отвергнута, результаты предыдущего анализа подтвердились.

Была построена диаграмма, отображающая средние значения и их доверительные интервалы для каждой группы (рисунок 10). Наилучший средний уровень качества жизни получен в диапазоне климата 68-80

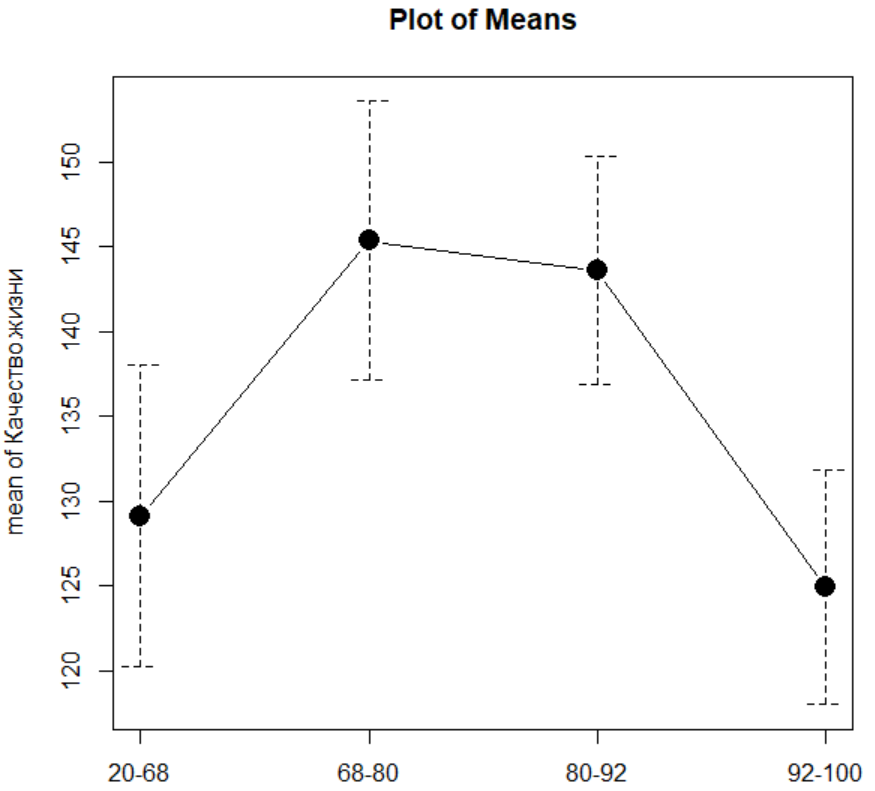


Рисунок 10 – График средних для собственных данных

# ВЫВОД

В ходе работы был проведён однофакторный дисперсионный анализ данных по варианту и собственных экспериментальных данных. В качестве инструмента был использован табличный процессор Excel и среда R Commander.

В задании по варианту в качестве фактора была рассмотрена цена за единицу товара, в качестве результирующего признака – объём продаж. Была опровергнута нулевая гипотеза (), лучшее значение получено в диапазоне цен 1000-1100.

При анализе собственных данных в качестве фактора был рассмотрен индекс климата, в качестве результирующего признака – индекс качества жизни. Нулевая гипотеза не была опровергнута, существенная связь не подтвердилась. Лучшие средние значения качества жизни получены в средних диапазонах: 68-80 и 80-92.