Отчёт к лабораторной работе

по дисциплине  
«Интеллектуальный анализ данных»

выполнил   
студент гр. ИС/б-18-1-з Демиденко А. А.  
зачётная книжка № 481483  
принял Шумейко И. П.

Лабораторной работа № 3  
«Задача дисперсионного анализа. Однофакторный дисперсионный анализ»

## Цель работы

* приобрести практические навыки в проведении дисперсионного анализа по экспериментальным данным
* исследовать возможности языка R для проведения дисперсионного анализа

## Ход работы

1. Создадим файл с исходными данными (см. рисунок 1):

  
Рисунок 1 – Данные для анализа, вариант 28

С помощью «Пакета анализа» программы Excel произведём однофакторный дисперсионный анализ данных, для заданной однофакторной таблицы зависимости урожайности пшеницы (т/га) от количества осадков (мм) за год (см. рисунок 2):

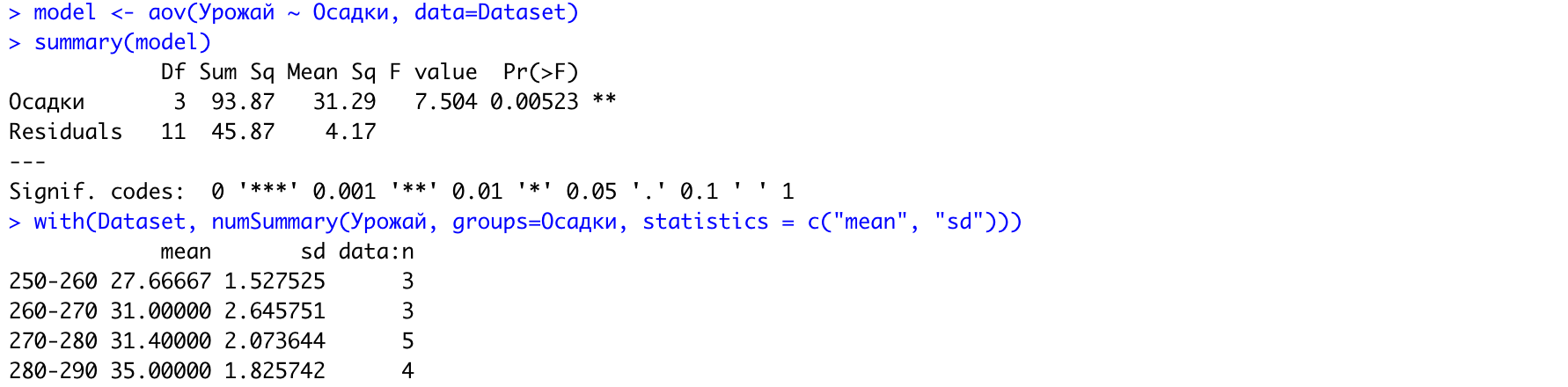
  
Рисунок 2 – Результирующая таблица выполненного дисперсионного анализа

Сформулируем выводы полученного анализа.

Исходя из таблицы «ИТОГИ», сравнение средних значений показывает, что при осадках 280-290 мм урожайность достигает лучшего результата, а при осадках 250-260 мм – наихудшего.

Проанализируем таблицу «Дисперсионный анализ». Сравнение F и Fкритическое показывает, что F > Fкритическое, следовательно отвергнута гипотеза Н0 и принята гипотеза Н1, и с вероятностью ошибки α = 0,05 можно утверждать, что влияние рассматриваемого фактора на результативный признак существенно.

Далее проведём дисперсионный анализ средствами языка R пакета Rcmdr.

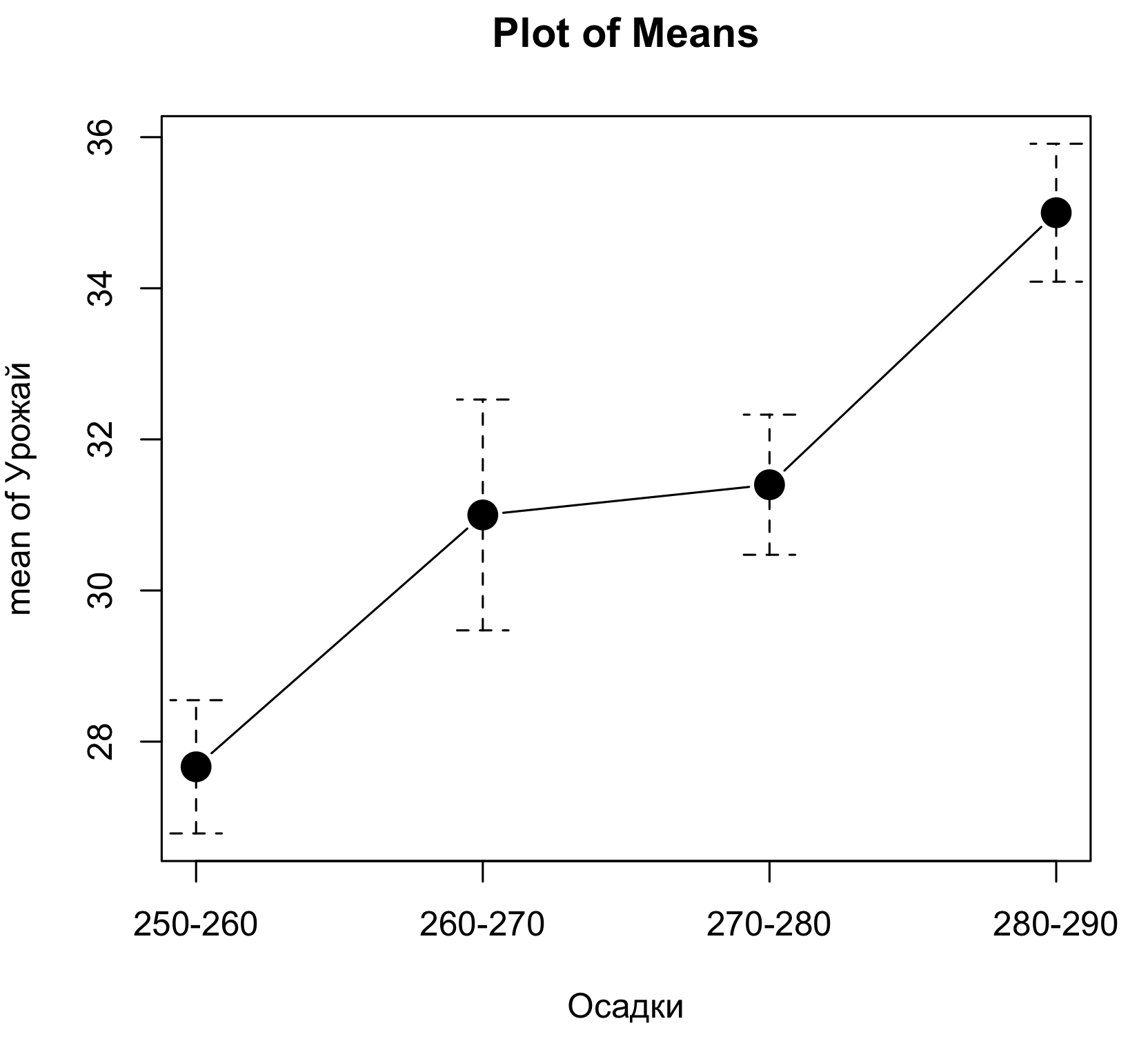


В столбце F value представлено рассчитанное по имеющимся данным значение F-критерия, равное 7,504. В столбце Pr(>F) представлена вероятность получить F-значение, равное или превышающее то значение, которое в действительности рассчитали по имеющимся выборочным данным (при условии, что нулевая гипотеза верна). Как видно, эта вероятность равна 0,0052 (0,52%), что ниже 5%-ного уровня значимости. В связи с этим мы заключаем, что нулевая гипотеза неверна. Таким образом, экспериментальные условия оказали существенное влияние на результативный признак.

Сравнение средних значений показывает, что осадки 280-290 мм позволяют добиться лучшей урожайности, а уровень осадков 250-260 мм – худшей.

Построим диаграмму, отображающую средние значения и их доверительные интервалы для каждой группы (см. рисунок 3).



  
Рисунок 3 – График средних

Полученный график визуально подтверждает выводы о зависимости осадков и урожая.

1. Ответим на контрольные вопросы.

* Факторы в дисперсионном анализе.

В основе дисперсионного анализа лежит предположение о том, что одни переменные могут рассматриваться как причины (факторы, независимые переменные): 𝑓1, ... 𝑓𝑘, а другие как следствия (зависимые переменные).

* Цель применения дисперсионного анализа.

Применяется для исследования влияния одной или нескольких качественных переменных (факторов) на одну зависимую количественную переменную (отклик).

* Различия между однофакторным и многофакторным дисперсионным анализом.

Однофакторный дисперсионный анализ используется в тех случаях, когда есть в распоряжении три или более независимые выборки, полученные из одной генеральной совокупности путем изменения какого-либо независимого фактора, для которого по каким-либо причинам нет количественных измерений.

* Три основные математические допущения дисперсионного анализа.

1. распределение зависимой переменной для каждой группы фактора соответствует нормальному закону;
2. дисперсии выборок, соответствующих разным градациям фактора, равны между собой;
3. выборки, соответствующие градациям фактора, должны быть независимы. Независимыми называются выборки, в которых объекты исследования набирались независимо друг от друга, то есть вероятность отбора любого испытуемого одной выборки не зависит от отбора любого из испытуемых другой выборки.

* Внутригрупповая дисперсия в дисперсионном анализе.

Градации, соответствующие независимым выборкам объектов, называются межгрупповыми, а градации, соответствующие зависимым выборкам, — внутригрупповыми.

* Статистическая значимость F-отношения.

Если расчетное значение критерия Фишера будет меньше, чем табличное значение нет оснований считать, что независимый фактор оказывает влияние на разброс средних значений, в противном случае, независимый фактор оказывает существенное влияние на разброс средних значений

* Функции для выполнения дисперсионного анализа на языке R.

Функция `aov`.

**Выводы**

В ходе данной лабораторной работы были приобретены практические навыки в проведении дисперсионного анализа по экспериментальным данным, а также исследованы возможности языка R и программы Excel для проведения дисперсионного анализа.