|  |
| --- |
| МИНОБРНАУКИ РОССИИ |
| Федеральное государственное бюджетное образовательное учреждение высшего образования **«МИРЭА − Российский технологический университет»**  **РТУ МИРЭА** |

**Институт информационных технологий (ИИТ)**

**Кафедра прикладной математики (ПМ)**

**ОТЧЕТ ПО ПРАКТИЧЕСКОЙ РАБОТЕ**

по дисциплине «Прикладные задачи математической статистики»

**Практическое задание № 4**

|  |  |  |  |
| --- | --- | --- | --- |
| Студент группы ИНБО-01-17 | *ИНБО-05-19, Грузилова В.Д.* | (подпись) | |
| Преподаватель | *Буданцев А.В.* | (подпись) | |
| Отчет представлен | «\_\_\_»\_\_\_\_\_\_\_\_2021 г. | |  | |

Москва 2021 г.

Оглавление

[Задание 3](#_Toc87475102)

[Выполнение работы 4](#_Toc87475103)

[Файл 1 4](#_Toc87475104)

[Файл 2 5](#_Toc87475105)

[Файл 3 6](#_Toc87475106)

# Задание

Для выполнения практического задания необходимо:

1. Открыть папку, соответствующую своей группе.

2. Открыть папку с вариантом, совпадающим с вашим номером в списке.

В папке 3 файла с данными.

1-ый файл содержит 2 ряда данных. Необходимо построить линейную регрессию между ними (1-ый столбец содержит факторную переменную, 2-ой результирующую), оценить значимость полученных коэффициентов прямой и построить доверительные интервалы для них, оценить адекватность модели.

2-ой файл содержит 4 ряда данных. Первый ряд (столбец) содержит количественную факторную переменную, следующие два - качественную факторную переменную, последний - результирующую переменную.

Необходимо с помощью теста Чоу обосновать необходимость деления выборки по одной из качественных факторных переменных. Произвести разбиение и построить две линейных регрессии.

3-ий файл содержит 2 ряда данных. Необходимо двумя способами (тест Спирмена и тест Гольдфельда-Квандта) определить, присутствует ли в данных гетероскедастичность, построить линейную регрессию, оценить значимость полученных коэффициентов и адекватность модели.

Все расчеты проводить для уровня значимости (альфа) = 0.05.

# Выполнение работы

## Файл 1

**Формулы**

Регрессионная модель: φ = a\*x + b

Коэффициент корреляции:

Параметры регрессионной прямой: ,

F-критерий адекватности модели:

, ,

– дисперсия данных относительно регрессионной прямой, – общая выборочная дисперсия.

Проверка на значимость параметров:

Стандартные ошибки параметров: ,

Доверительные интервалы для параметров: ,

точность интервала , для b аналогично.

Критическое значение t-статистики: , для b аналогично.

**Выполнение задания**

Рассчитанный коэффициент корреляции: r = -0.98.

Параметры модели: a = -2.21, b = -2.02.

**Оценка адекватности модели:**

Критерий Фишера: Fрасч = 0.035. Критическое значение: Fкрит = 1.392.

Fрасч < Fкрит → модель признана адекватной с заданным уровнем значимости 𝛼 = 0.05.

**Доверительные интервалы параметров:**

-2.29543 a -2.12692

-3.37191 b -0.65999

**Оценка значимости параметров:**

Критические значения t-статистики:

Ta = -51.96, Tb = -2.94. Табличное значение: Tтабл = 1.98.

|Ta| > Tтабл и |Tb| > Tтабл → параметры статистически значимы.

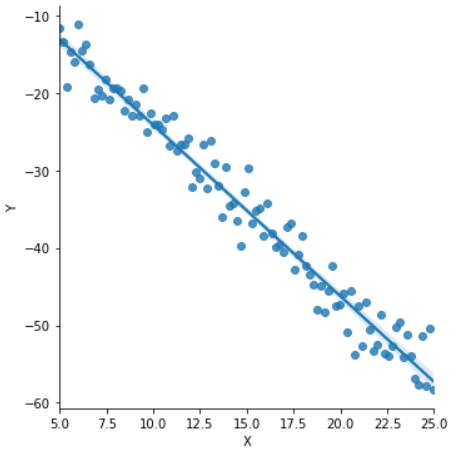


Рисунок 1. График линейной регрессии

## Файл 2

**Формулы**

Критерий для теста Чоу:

Сумма квадратов остатков:

RSS1 – сумма квадратов остатков для первой подвыборки, RSS2 – для второй. K – количество параметров модели.

**Выполнение задания**

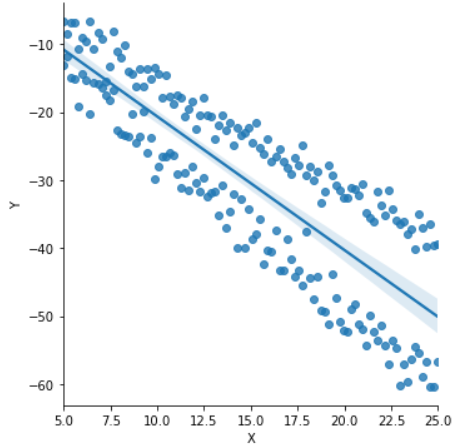


Рисунок . График линейной регрессии без разбиения на подвыборки

Разбиение по **первой** качественной факторной переменной:

F = 1270.08

Fкрит = 3.087

F > Fкрит → подвыборки не однородны.

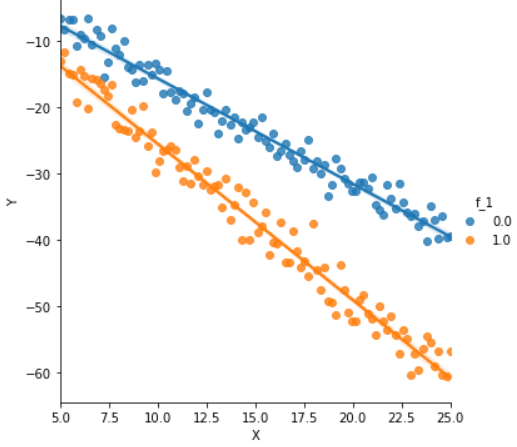


Рисунок . График линейных регрессий с разделением по первой качественной переменной

Разбиение по **второй** качественной факторной переменной:

F = 0.54

Fкрит = 3.087

F > Fкрит → подвыборки однородны.

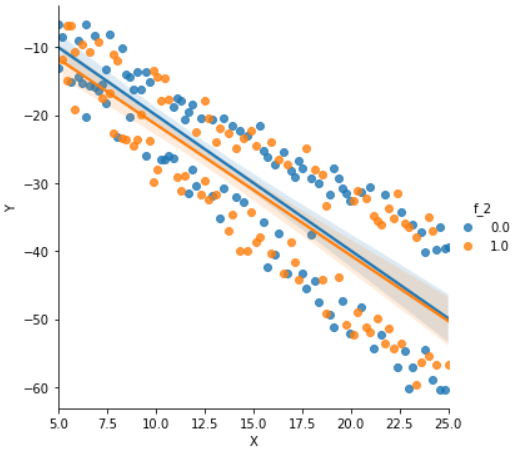


Рисунок . График линейных регрессий с разделением по второй качественной переменной

## Файл 3

**Формулы**

Критерий теста Спирмена:

, где Di – разность между рангом X и ε в i-том наблюдении.

Критерий теста Гольдфельда-Квандта: , где

**Выполнение задания**

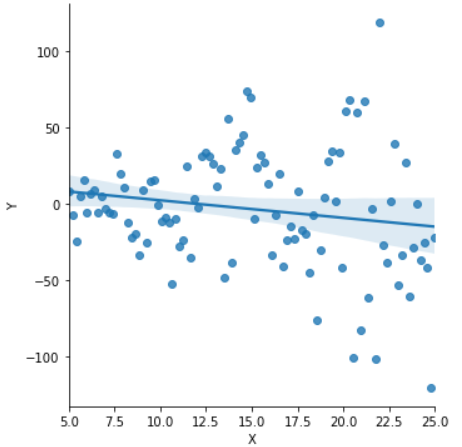


Рисунок . График линейной регрессии

**Оценка адекватности модели:**

Критерий Фишера: Fрасч = 0.98. Критическое значение: Fкрит = 1.392.

Fрасч < Fкрит → модель адекватна с заданным уровнем значимости 𝛼 = 0.05.

**Оценка значимости параметров:**

Критические значения t-статистики:

Ta = -1.72, Tb = 1.27. Табличное значение: Tтабл = 1.98.

|Ta| < Tтабл и |Tb| < Tтабл, поэтому параметры статистически не значимы.

**Проверка на гетероскедатичность остатков:**

Критерий Спирмана: = 4.45. Табличное значение tтабл = 1.98.

> tтабл → есть линейная корреляция между остатками и X.

Критерий Гольдфельда-Квандта: F = 0.15.

Табличное значение Fтабл = 1.69.

F > Fтабл → дисперсии в двух подвыборках (начало и конец отсортированной изначальной выборки) не равны.

# Выводы

Файл 1: регрессионная модель адекватна, оба параметра статистически значимы, для них рассчитаны доверительные интервалы.

Файл 2: с помощью теста Чоу обоснована необходимость деления выборки по первой качественной факторной переменной.

Файл 3: и по тесту Спирмена, и по тесту Гольдфельда-Квандта доказана гетероскедатичность остатков, модель адекватна, параметры статистически не значимы.