Учреждение образования

«Белорусский государственный университет информатики

и радиоэлектроники»

Инженерно-экономический факультет

Кафедра экономической информатики

Средства и технологии анализа и разработки информационных систем

Лабораторная работа № 7

«Применение инструментов Data Mining»

Выполнил: студент группы 772302, Шмыров И.Д.

Проверил: Кунцевич А.А.

Минск 2020

**Линейные регрессии наименьших квадратов**

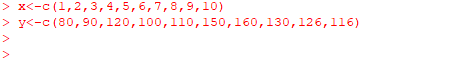


Рисунок 1 – указание данных

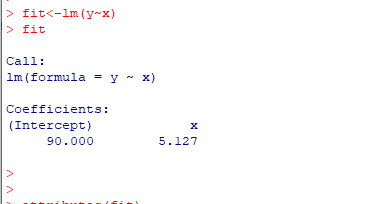


Рисунок 2 – установление зависимости x от y

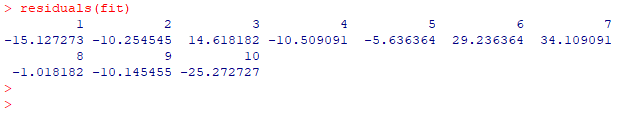


Рисунок 3 – вычисление остатков

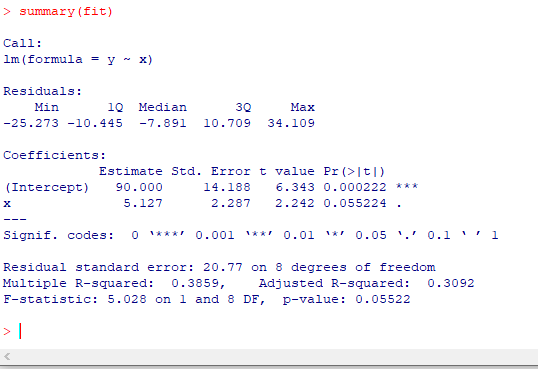


Рисунок 4 – итоговая модель

Так как остатки не распределены нормально и probability f-статистики >0.05 – итоговая модель не адекватна и не значима.

**README**

1. **Что такое data mining?**

Data mining – это процесс обнаружения в больших массивах данных, информационных системах, базах данных ранее неизвестных и не изученных данных доступных для изучения и анализа, которые могут облегчить принятие решений в различных сферах деятельности компании, человека и т.д.

1. **МНК.**

Метод наименьших квадратов – математический метод для решения задач в статистике и эконометрике, основанный на минимизации суммы квадратов отклонений для нахождения коэффициентов линейной зависимости.

1. **Что такое остатки.**

Остатки - это разности между полученными значениями и предсказанными в регрессионной модели. Чем меньше остатки, тем лучше регрессионная модель.

1. **F-статистика.**

Это критерий, тестовая статистика которого при выполнении нулевой гипотезы имеет распределение Фишера, применяется для проверки равенства дисперсий двух статистических выборок.

1. **Probability.**

Probability – это статистический критерий который отображает степень возможности наступления некоторого события.