Задание по Теме 4

1. Dataset “Kuiper.xls” (переменные Price, Mileage, Cylinder, Liter, Cruise):
   1. Провести корреляционный анализ имеющихся данных.
   2. Построить базовую модель линейной регрессии.
   3. Вывести результаты анализа базовой модели.
   4. Записать уравнение линейной регрессии.
   5. Проверить значимость каждого отдельного коэффициента с помощью T-test.
   6. Проверить значимость построенного уравнения регрессии с помощью F-test.
   7. Построить график рассеяния и уравнения регрессии.
   8. Построить доверительные интервалы для коэффициентов регрессии.
   9. В случае подозрения на наличие выбросов, проверить так называемые важные наблюдения, которые значительно влияют на построение модели.
   10. Используя функцию Step или StepAIC, постараться улучшить модель.
   11. В случае получения в предыдущем пункте модели, отличной от базовой, повторить пп. c-i для новой модели.
   12. Построить графики: scatterplot, "Residuals vs Fitted", "Normal Q-Q" , "Residuals vs Leverage" и дать интерпретации.
   13. Проверить модель на наличие выбросов.
   14. Проверить модель на гетероскедастичность.
   15. Проверить остатки модели на автокорреляцию.
   16. Проверить остатки модели на нормальность распределения.
   17. Проверить модель на мультиколлинеарность данных.
   18. Попробовать применить трансформацию Box-Cox зависимой переменной.
   19. В случае получения новой модели в предыдущем пункте проанализировать новую модель.
2. Dataset “cigarettes.txt” (переменные по столбцам: carbon monoxide, tar, nicotine, weight):
   1. Построить модель линейной регрессии с зависимой переменной y=carbon monoxide, и независимыми переменными: x1=tar, x2=nicotine, x3=weight.
   2. Провести анализ как в предыдущем задании по пп. c-s.

**Примечание: можно и даже будет лучше, если Вы возьмете свои данные, их много в интернете, и проделаете регрессионный анализ на них.**