**Федеральное агентство по образованию Российской Федерации**

**Государственное образовательное учреждение**

**высшего профессионального образования**

**«КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»**

**Кафедра прикладной математики**

Преподаватель,

д.т.н. А.А. Халафян

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**ОТЧЕТ**

**СТАТИСТИЧЕСКИЙ АНАЛИЗ АВТОМОБИЛЕЙ**

Работу выполнил студент 4 курса

факультета компьютерных технологий и прикладной математики  
спец. 01.03.02 – Прикладная информатика и информатика

Краснодар 2023

СОДЕРЖАНИЕ

[**1 Исходные данные 3**](file:///K:\Халафян\primer%20otcheta.docx#_Toc233138293)

[**2 Графический анализ 5**](file:///K:\Халафян\primer%20otcheta.docx#_Toc233138294)

[**2.1 2D Graphs 5**](file:///K:\Халафян\primer%20otcheta.docx#_Toc233138295)

[**2.1.1 2D Histogramms 5**](file:///K:\Халафян\primer%20otcheta.docx#_Toc233138296)

[**2.1.2 2D Scatterplots 6**](file:///K:\Халафян\primer%20otcheta.docx#_Toc233138297)

[**2.1.3 2D Box Plots 9**](file:///K:\Халафян\primer%20otcheta.docx#_Toc233138298)

[**2.2 Средство «закрашивание» 10**](file:///K:\Халафян\primer%20otcheta.docx#_Toc233138299)

[**2.3 3D SequentialGraphs 12**](file:///K:\Халафян\primer%20otcheta.docx#_Toc233138300)

[**2.3.1 Raw Data Plots 12**](file:///K:\Халафян\primer%20otcheta.docx#_Toc233138301)

[**2.3.2 Bivariate Histograms 13**](file:///K:\Халафян\primer%20otcheta.docx#_Toc233138302)

[**2.3.3 Box Plots 13**](file:///K:\Халафян\primer%20otcheta.docx#_Toc233138303)

[**3 Основные статистики 15**](file:///K:\Халафян\primer%20otcheta.docx#_Toc233138304)

[**3.1 Описательные статистики 15**](file:///K:\Халафян\primer%20otcheta.docx#_Toc233138305)

[**3.2 Корреляционная матрица 15**](file:///K:\Халафян\primer%20otcheta.docx#_Toc233138306)

[**3.3 Критерий Стьюдента сравнения средних 16**](file:///K:\Халафян\primer%20otcheta.docx#_Toc233138307)

[**3.3.1 t-test, independent, by groups 16**](file:///K:\Халафян\primer%20otcheta.docx#_Toc233138308)

[**3.3.2 t-test, independent, by variables 16**](file:///K:\Халафян\primer%20otcheta.docx#_Toc233138309)

[**3.3.3 t-test, dependent samples 16**](file:///K:\Халафян\primer%20otcheta.docx#_Toc233138310)

[**3.3.4 t-test, single samples 17**](file:///K:\Халафян\primer%20otcheta.docx#_Toc233138311)

[**3.4 Группировка и однофакторная ANOVA 17**](file:///K:\Халафян\primer%20otcheta.docx#_Toc233138312)

[**4 Частотный анализ 19**](file:///K:\Халафян\primer%20otcheta.docx#_Toc233138313)

[**4.1 Таблицы частот 19**](file:///K:\Халафян\primer%20otcheta.docx#_Toc233138314)

[**4.2 Таблицы кросстабуляции 21**](file:///K:\Халафян\primer%20otcheta.docx#_Toc233138315)

[**5 Корреляционный анализ 23**](file:///K:\Халафян\primer%20otcheta.docx#_Toc233138316)

[**6 Дисперсионный анализ 24**](file:///K:\Халафян\primer%20otcheta.docx#_Toc233138317)

[**7 Линейное многомерное моделирование взаимосвязей 27**](file:///K:\Халафян\primer%20otcheta.docx#_Toc233138318)

[**8 Нелинейное многомерное моделирование взаимосвязей 29**](file:///K:\Халафян\primer%20otcheta.docx#_Toc233138319)

[**8.1 Fixed Nonlinear Regression 29**](file:///K:\Халафян\primer%20otcheta.docx#_Toc233138320)

[**8.2 Логит регрессия 31**](file:///K:\Халафян\primer%20otcheta.docx#_Toc233138321)

[**8.3 Пробит регрессия 32**](file:///K:\Халафян\primer%20otcheta.docx#_Toc233138322)

[**8.4 Экспоненциальная регрессия 33**](file:///K:\Халафян\primer%20otcheta.docx#_Toc233138323)

[**8.5 Кусочно-линейная регрессия 34**](file:///K:\Халафян\primer%20otcheta.docx#_Toc233138324)

[**8.6 Определенная пользователем регрессия 35**](file:///K:\Халафян\primer%20otcheta.docx#_Toc233138325)

[**9 Канонический анализ 36**](file:///K:\Халафян\primer%20otcheta.docx#_Toc233138326)

[**10 Дискриминантный анализ 41**](file:///K:\Халафян\primer%20otcheta.docx#_Toc233138327)

[**11 Классификационный анализ без обучения 48**](file:///K:\Халафян\primer%20otcheta.docx#_Toc233138328)

[**11.1 Кластерный анализ 48**](file:///K:\Халафян\primer%20otcheta.docx#_Toc233138329)

[**11.1.1 Метод к-средних 48**](file:///K:\Халафян\primer%20otcheta.docx#_Toc233138330)

[**11.1.2 Двухвходовая кластеризация 50**](file:///K:\Халафян\primer%20otcheta.docx#_Toc233138331)

[**11.1.3 Древовидная кластеризация 52**](file:///K:\Халафян\primer%20otcheta.docx#_Toc233138332)

[**11.2 Деревья классификации 53**](file:///K:\Халафян\primer%20otcheta.docx#_Toc233138333)

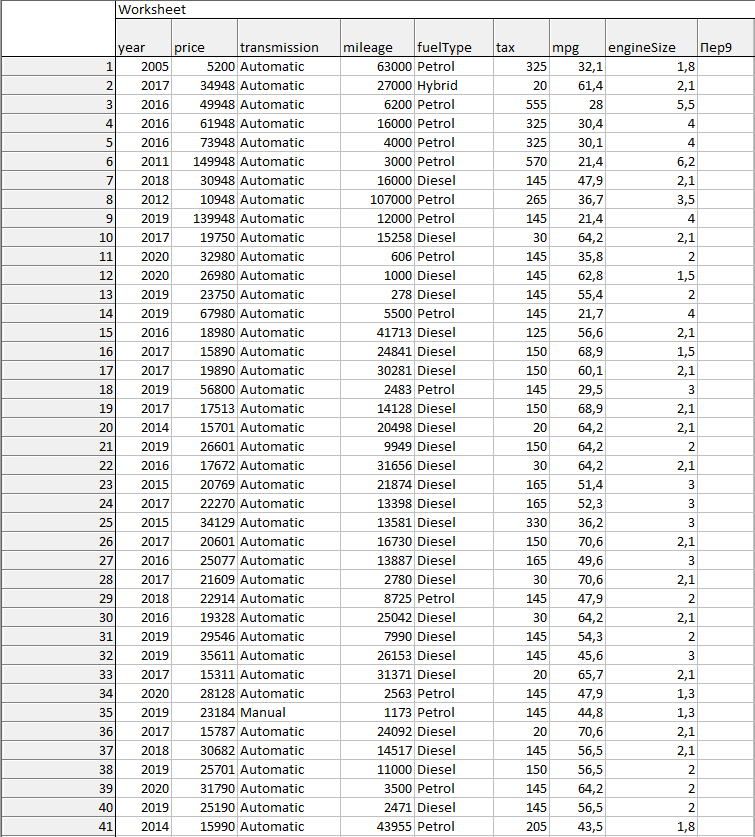
[**12 Методы редукции данных 57**](file:///K:\Халафян\primer%20otcheta.docx#_Toc233138334)

[**12.1 Факторный анализ 57**](file:///K:\Халафян\primer%20otcheta.docx#_Toc233138335)

[**12.2 Метод анализ главных компонент и классификация 61**](file:///K:\Халафян\primer%20otcheta.docx#_Toc233138336)

[**13 Многомерное шкалирование 65**](file:///K:\Халафян\primer%20otcheta.docx#_Toc233138337)

**1 Исходные данные**



В исходной таблице представлены данные по популярным автомобилям.

Показатели:

1. Year – год выпуска автомобиля.
2. Price – цена автомобиля при выпуске из завода.
3. Transmission – коробка передач.
4. Mileage – пробег.
5. FuelType – тип топлива.
6. Tax – налог на машину.
7. Mpg – ЛС.
8. EngineSize – размер двигателя.

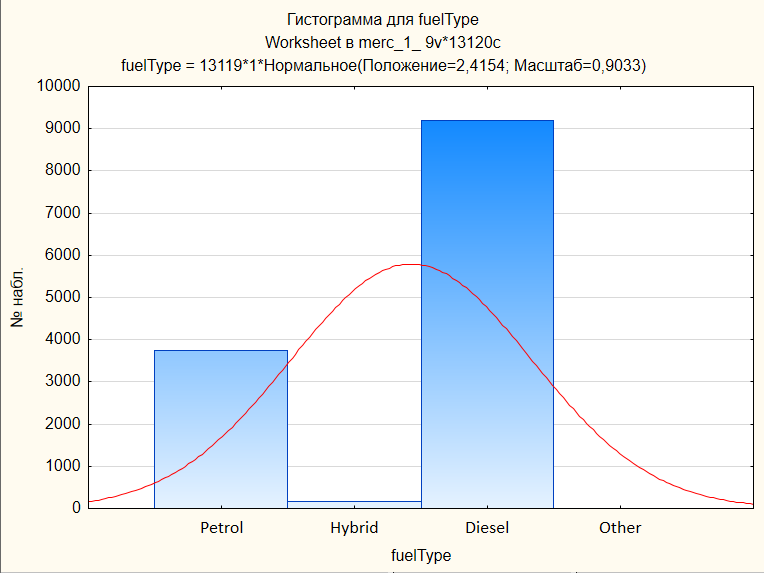
**2 Графический анализ**

**2.1 2D Graphs**

**2.1.1 2D Histogramms**

2D Histogramms являются графическими представлениями распределения частот выбранных переменных.

2D Histogramms Regular (простые) – столбчатая диаграмма распределения частот.



Данная 2D Histogramms Regular построена используя данные «fuelType»». Из гистограммы видно, что чаще всего встречаются автомобили, использующие бензин (9150 из 13000).

2D Histogramms Multiple (составные) – изображают распределение частот для нескольких переменных на одном графике.