

Задание №2 по курсу
«Пакеты прикладных программ для статистической обработки и анализа данных»

Задание выполнил:
студент 520 группы Ковальчук Александр
Вариант 2

Часть 1

Для проверки влияния типа кузова на расход бензина на трассе использовался дисперсионный анализ (процедура `proc glm`). Гипотеза H_0 состоит в равенстве средних по различным типам кузова. В результате проверки гипотезы p -value оказалось меньше <0.0001 . При уровне значимости 0.01 гипотеза H_0 отвергается. Значит, тип кузова влияет на расход бензина на трассе.

Часть 2

Исходя из графика `diffogram` и попарного `ttest` я сделал вывод о том, что SUV и Truck и Sedan и Wagon можно объединить. В результаты группы стали различимы.

Часть 3

После добавления предиктора `Origin` исходя из значения `RMSE` модель улучшилась (было: 4.163423 , стало: 4.051007). После удаления предиктора `Type*Origin` модель также улучшилась (`RMSE` = 4.043384).

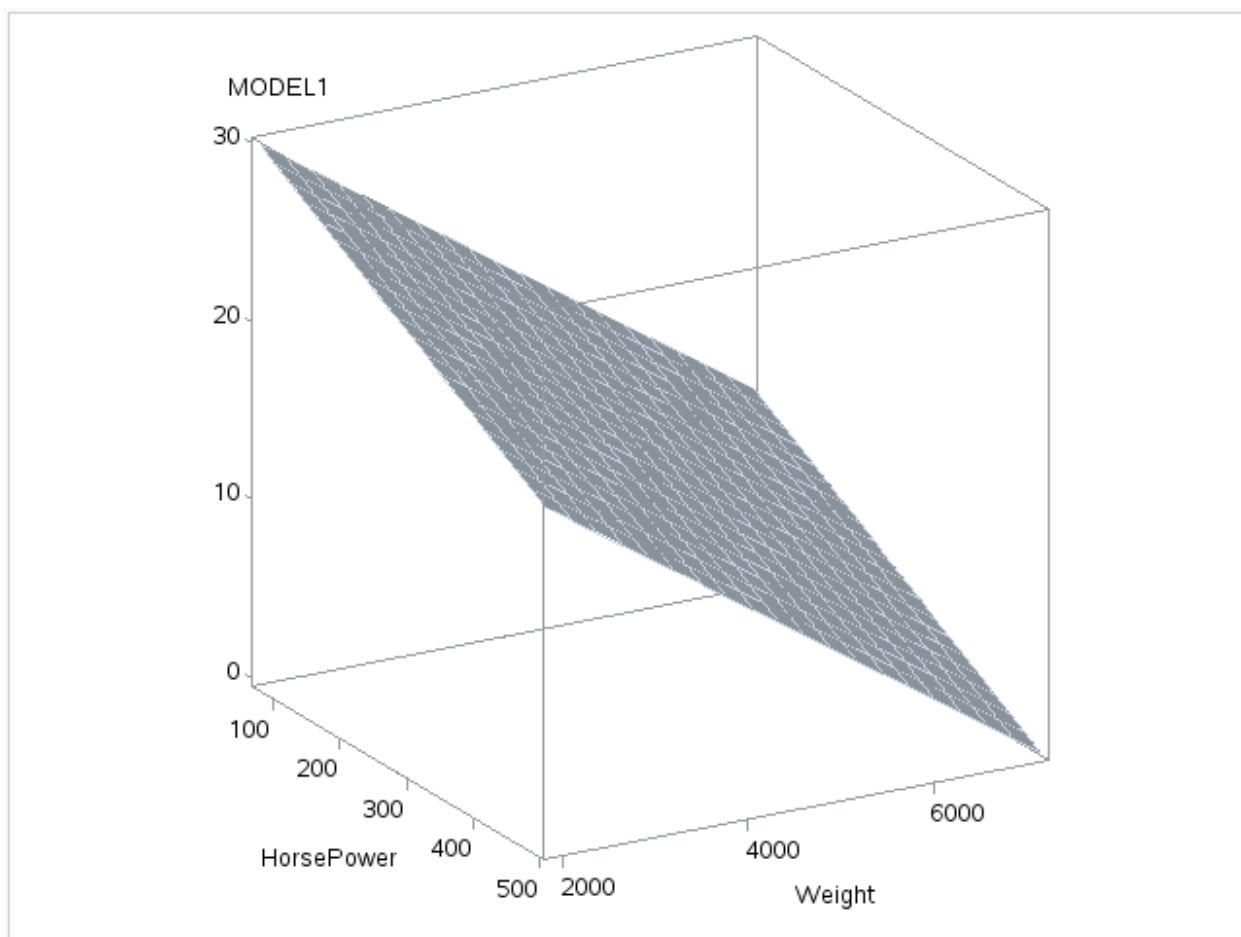
Часть 4

В результате работы `GLMSelect` с критерием `SBC` пошаговым обратным методом с критерием `F` были выбраны переменные `Weight`, `Horsepower`, `Invoice` и константа `Intercept`. При этом отклик описывается следующим образом:

$$Y = 38.732816 - 0.003461 * \text{Weight} - 0.035113 * \text{Horsepower} + 0.000043653 * \text{Invoice}$$

Часть 5

В результате работы процедуры `REG` моделью в наибольшим значением `R-Sqaure` оказалась модель от переменных `Weight` и `Horsepower`. 3D график представлен ниже.



Часть 6

PDF отчет содержится в файле report.pdf