Уровень значимости для всех гипотез задается 0.01. Зеленым обозначено условие для первого варианта, желтым для второго.

1) Проверить влияет ли Страна производитель (Origin) (или тип Кузова (Type)) на расход бензина в городе (MPG\_city) (или на трассе MPG\_highway) c заданным уровнем значимости.

2) Преобразовать категориальные переменные так, чтобы не было «неразличимых» групп (используя график diffogram, попарный t-test и шаг данных для объединения «неразличимых» групп).

3) Добавить предиктор тип Кузова (Type) (или Страна производитель (Origin) соответсвенно), чтобы понять улучшается ли модель (по RMSE). Объединения неразличимых групп не проводить. Если предиктор Origin\*Type не нужен, то исключить из модели. Построить финальную модель.

4) Построить регрессионную модель (с помощью GLMSELECT) зависимости расхода бензина на трассе (MPG\_highway) (или в городе (MPG\_city) от числовых переменных Length Weight Wheelbase Horsepower Invoice EngineSize Cylinders, выбрав наилучшую модель по AIC (SBC), причем варианты моделей должны формироваться прямым (обратным) пошаговым методом с критерием F (SELECTION=SL и уровнем значимости 0.01).

5) Найти наилучшую (по R2) регрессионную модель с двумя перменными (перебором всех вариантов с помощью процедуры REG). Построить 3D график зависимости отклика от этих переменных с равномерной сеткой 20 на 20 точек (сетку в наборе данных сгенерировать самостоятельно).

6) Сформировать pdf отчет, содержащий: ANOVA Таблицу и таблицу статистик (fit statistics) шага 3, трассу коэфициентов шага 4 и 3D график шага 5.