Зеленым обозначено условие для первого варианта, желтым для второго. Вместе с заданием передается два макроса:

- prepare_data генерирует из набора данных cars случайным образом тренировочный и тестовый наборы (с сохранением распределения откликов). С тренировочным набором можно производить любые манипуляции (менять любые переменные, удалять записи и т.д.), в тестовом можно только дописывать входные переменные, например, X*X или f(X), причем по тем же правилам, что и в тренировочном наборе, менять переменные отклика в тестовом наборе нельзя.
- calc_mape считает на тестовом наборе оценку качества модели MAPE=усреднённая сумма по всем наблюдениям от модуля остатка, деленного на значением истинного отклика.

Задача состоит в том, чтобы построить регрессионную модель для прогнозирования расхода бензина на трассе (MPG_highway) (или в городе (MPG_city) от числовых переменных Length Weight Wheelbase Horsepower Invoice EngineSize Cylinders и категориальных переменных Origin и Type. Можно использовать процедуры, рассмотренные в рамках курса: GLM, GLMSELECT, REG, GLMMOD, LIN, LOESS, а также их комбинации, например, можно фильтровать переменные одной процедурой, а модель строить другой по отобранным переменным, или оценивать выбросы одной, потом их удалять из тренировочного набора, а строить модель другой процедурой по отфильтрованному набору. Для оценки качества разработанной модели использовать макрос calc_mape. Обратите внимание, что при разных запусках prepare_data генерируются разные тестовые и тренировочные наборы. Ваша модель должна работать так, чтобы с любыми сгенерированными наборами (не использовать предположение, что такое-то наблюдение есть в тестовом или в тренировочном наборе) на 10 любых запусках усредненное MAPE было меньше 0.065. Рекомендуется использовать техники:

- Преобразования категориальных переменных (группировка значений) с использованием дисперсионного анализа (не забывайте делать эти преобразования и в тестовом наборе)
- Отбор значимых переменных с помощью REG или GLMSELECT (тестовый набор в качестве валидационного использовать нельзя, но можно из тренировочного выделить часть для валидации или использовать кросс-валидацию)
- Преобразование входных переменных и добавление полиномиальных членов в уравнение регрессии (не забывайте делать эти преобразования и в тестовом наборе)
- Использовать нелинейные уравнения зависимости в том числе и делать свою «нейросеть» и обучать ее как нелинейную регрессию с помощью NLIN.
- Преобразование отклика или использование обобщенных линейных моделей с разными распределениями ошибки и функциями связи (не забывайте делать правильный пересчет отклика после прогноза).

Для получения требуемого качества модели обычно достаточно использовать 2х из перечисленных выше техник.

Построить 3D график зависимости отклика от всех пар отобранных переменных с равномерной сеткой 20 на 20 точек (сетку в наборе данных сгенерировать самостоятельно), значение остальных переменных, не вошедших в пару (если есть), при построении графика усреднять.