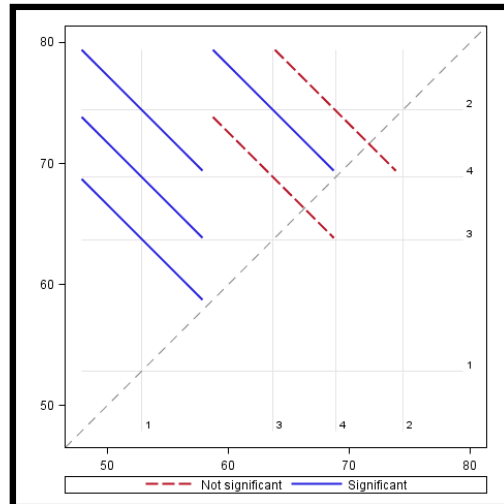


Уровень значимости для всех гипотез задается 0.01. Зеленым обозначено условие для первого варианта, желтым для второго. Реализовать R код, делающий все последовательности шагов:

- 1) Проверить влияет ли Страна производитель (Origin) (или тип Кузова (Type)) на расход бензина в городе (MPG_city) (или на трассе MPG_highway) с заданным уровнем значимости.
- 2) Преобразовать категориальные переменные так, чтобы не было «неразличимых» групп (используя график diffogram, попарный t-test и преобразование данных для объединения «неразличимых» групп). Написать соответствующий код.
- 3) Реализовать код (готовые сторонние реализации diffogram использовать нельзя) для «кастомизированной» визуализации попарных сравнений в виде diffogram как в примере ниже (можно использовать любые графические пакеты):



- 4) Добавить предиктор тип Кузова (Type) (или Страна производитель Origin соответственно), чтобы понять улучшается ли модель. Объединения «неразличимых» групп не проводить.
- 5) Проверить нужен ли эффект взаимодействия Origin*Type и если не нужен, то исключить из модели. Для проверки построить соответствующий график и применить один из формальных методов (на основе расчета коэффициентов модели или критерия Фишера), текстом обосновать почему решили оставить или удалить эффект взаимодействия. Построить финальную модель. Визуализировать diffogram'ы для данной модели.
- 6) С помощью оператора Contrast из ghlт или emmeans или aov проверить гипотезу, что средний расход бензина MPG_city (MPG_highway) не отличается в группе европейских и азиатских седанов от группы американских траков. Проверить, что полученные результаты совпадают с проверкой «в лоб» - изменить соответствующим образом исходную выборку (оставить только модели из двух групп европейских + азиатских седанов vs американские траки) и применить подходящую процедуру ANOVA или t-test.
- 7) Показать (добавив комментарии), что выполняются или не выполняются требования применимости полученной в пункте 6 модели, с точки зрения нормальности распределения целевой переменной и равенства дисперсий в группах.
- 8) Реализовать проверку гипотезы из пункта 6 с использованием непараметрической ANOVA.
- 9) Сформировать pdf отчет, содержащий ТОЛЬКО таблицы ANOVA со всех шагов для всех моделей, значения P-values больше заданного уровня значимости (0.01) подсвечивать в отчете красным или (отмечать «!»).