Задание по Теме 4

1. Dataset “Doctor.xlsx” (переменные Y=исход беременности (зависимая переменная), x1, x2, x3, x4, x5 (независимые переменные, факторы)=показатели исследуемой женщины):

* Построить описательную статистику с помощью функции summary.
* При необходимости создать категориальную переменную (функция factor).
* Построить базовую модель логистической регрессии с максимально возможным количеством предикторов с использованием функции glm.
* Записать уравнение бинарной регрессии, используя оценки коэффициентов.
* Протестировать значимость коэффициентов регрессии в отдельности.
* Проверить значимость регрессии в целом по критерию Вальда и максимального правдоподобия.
* Построить доверительные интервалы для коэффициентов регрессии (функции confint и confint.default).
* Провести сравнительный анализ логит и пробит моделей.
* Построить таблицу сопряженности с пороговой вероятностью 0.5 (функция confusionMatrix).
* Посчитать специфичность и чувствительность модели (функции sensitivity и specificity).
* Найти оптимальное пороговое значение вероятности предсказания. Построить таблицу сопряженности для этой вероятности, посчитать специфичность и чувствительность модели (функция optimalCutoff).
* Если выборка большого объема, то можно предварительно разбить выборку на две части: тренировочная и тестовая выборки (функция sample\_frac).
* Построить ROC кривую, интерпретировать результаты (функция plotROC).
* Попробовать улучшить логит или пробит модель с использованием коэффициента AIC (функция stepAIC).