2.1

Средний балл мальчиков будет равен 4.16026667-0.18740000\*1, что равно 3.97286667

Средний балл девочек будет равен 4.16026667-0.18740000\*0, что равно 4.16026667

Средний балл учеников будет равен (3.97286667+4.16026667)/2, что равно 4,06656667

2.2

Давайте заменим is\_male на is\_female, тогда наши коэффициенты будут 3.97286667 и 0.18740000, то есть уравнение будет выглядеть как 3.97286667+0.18740000\*is\_female

2.3

Полученная регрессия не будет лучше объяснять оценки потому что is\_male и is\_female это две противоположные бинарные величины, то есть is\_male = (1-is\_female) и is\_female = (1-is\_male), то есть по значению любой переменной мы можем определить точное значение другой переменной, поэтому смысла добавлять две переменные нет.

Одно из возможных уравнений для двух переменных будет выглядеть так:

4.06656667 - 0.0937\*is\_male + 0.0937\*is\_female, если его расписать как

4.06656667 - 0.0937\*is\_male + 0.0937\*(1-is\_male), а это уже наше уравнение из условия

4.16026667-0.18740000\*is\_male

2.4

Если просто удвоить изначальный набор исследований, то качество регрессии станет хуже, мы переобучимся, а именно по сути для каждой изначальной точки(до раздвоения) величина квадрата ошибки удвоится, поэтому новая регрессия относительно старой будет более приближена к дальним значениям(на который изначально величина квадрата ошибки была большой) то есть коэффициенты будут более чувствительны к выбросам.