№2

Пункт1: Если построить график, то у нас всего на нем будет 2 столбца с точками, так как переменная бинарная, но тогда если у нас линейная регрессия, то в каждом из этих столбцов нам нужно найти такую точку, что сумма квадратов отклонений будет минимальной, и именно эти точки будут значениями регрессии в точках 0 и 1, именно эти точки будут средними (следует из определения дисперсии), тогда получается что **для девочек это 4.16026667** (при х=0). Для **мальчиков 3.97286667** (при х=1). А общее мы не можем определить, так как не знаем отношение количества девочек к количеству мальчиков.

Пункт2: Найдем 2 точки через которые проходит регрессия, а именно точки при х = 0 и 1. Это будут 4.16026667 и 3.97286667 соответственно. Тогда так как у нас бинарная переменная, то у графика вид 2 столбцов(при х=0 и х=1), тогда заметим, что если мы меняем 0 на 1 и наоборот, то эти столбцы поменяются местами (так как значение у мальчиков и девочек прыи котором сумма квадратов ошибок наименьшая не изменится), тогда нам нужно найти функцию, проходящую не через точки (0;4.16026667) и (1; 3.97286667), а через точки (1; 4.16026667) и (0; 3.97286667), у нее вид Y=A+BX. Тогда А = 3.97286667 (значение при 0), а В = (Y(1)-A)/X=(Y(1)-A)/1=Y(1)-A=0.18740000. Следовательно, мы получили регрессию

**grade= 3.97286667+0.18740000\*is\_female**

Пункт3: пусть у нас есть новая регрессия вида grade = a1+a2\*ismale+a3\*isfemale. Но тогда заметим, что так как у нас бинарные переменные, то переменную isfemale мы можем поменять на переменную 1-ismale, таким образом мы снова получим изначальную регрессию, так как нам не важны значения в точках кроме 0 и 1. Но тогда эта регрессия будет ничем не лучше, так как по суте она будет той же самой что и в первом пункте.

Пункт4: Коэффициенты не изменятся, так как оптимальные значения оценок для мальчиков и девочек не изменятся (с минимумом квадратов ошибок), так как мы просто продублируем все старые измерения, следовательно, для новых измерений эти значения так же оптимальные, следовательно, так как по 2 оптимальным точкам среди мальчиков и девочек (других точек нет, следовательно нам нужно учитывать только квадратные отклонения в 1 и 0) мы однозначно строим линейную функцию, следовательно, она не изменится и так как она не изменится, ее качество тоже не изменится.