I.

	Средний балл
Мальчики	3.97286667
Девочки	4.16026667
Все ученики	p_male * 3.97286667 + (1 - p_male) * 4.16026667

Для того, чтобы узнать средний балл для всех учеников, надо узнать примерно или точно долю, которую составляют мальчики в классе p_male (или, что аналогично, долю. которую составляют девочки). После этого можно будет посчитать взвешенное среднее из средних, которое очевидно равно среднему.
На самом деле, при изменении переменной is_male переменная grade может принять только два значения: 4.16026667 (какое-то значение для всех девочек) или 4.16026667-0.18740000=3.97286667 (какое-то значение для всех мальчиков). Очевидно, что один из корректных способов минимизировать сумму квадратов такой: выбрать два значения: одно минимизирующее сумму квадратов ошибок для мальчиков (значение x), второе - минимизирующее сумму квадратов ошибок для девочек (значение y), и построить модель: grade = y + (x-y)*is_male. То есть значение, которое принимает grade при is_male=0, минимизирует сумму квадратов ошибок такого значения как оценки среднего балла для всех девочек, а значение при is_male=1 минимизирует сумму квадратов ошибок такого значения как оценки среднего балла для всех мальчиков. Тогда значение при is_male=0, равное 4.16026667, получается при минимизации суммы квадратов, и, как известно, равно среднему. Аналогично, значение при is_male=1, равное 3.97286667, равно среднему. При этом средний балл по всем ученикам определить скорее нельзя, так как неизвестно, какой процент класса составляют девочки, а какой - мальчики. Нельзя предположить, что и те, и те составляют примерно 50%, потому что известно, что, например, в классах с углублённым изучением математики обычно меньше девочек.
II.
Коэффициенты будут такими: grade= 3.97286667+ 0.18740000*is_female. Заметим, что множество возможных значений регрессии для разных полов не меняется: всё ещё можно выбрать одно произвольное значение для мальчиков, другое произвольное для девочек и по ним построить модель, их дающую. Тогда значения grade для мальчиков и для девочек должны сохраниться. Линейная функция задаётся своими значениями в двух точках, и у нас в обоих случаях две точки: is_male = 0 или 1 и is_female = 0 или 1. Заметим, что is_male = 1-is_female. Тогда регрессия от is_female должна давать при каждом значении is_female то же, что и регрессия от is_male при is_male = 1 - is_female. Тогда в регрессии от is_female уже зафиксированы (из оптимальности приведённой в условии регресии от is_male) значения при is_female=1, is_male=0 и is_female=0, is_male=1. Получаем, что линейная функция при замене is_female = (1 - is_male), is_male = (1 - is_female) должна быть той же, что и функция в оптимальной регрессии от is_male. Получаем:
grade = a+b*is_female =a+b*(1-is_male) =(тождественно равно)4.16026667-0.18740000*is_male, где a, b - некоторые числа. (a+b) - b*is_male = 4.16026667-0.18740000*is_male. (a+b) = 4.16026667, b= 0.18740000, a = 4.16026667-0.18740000=3.97286667.
III.
Нет. Параметр is_female линейно зависит от is_male: is_female = 1-is_male. Заменим в новой модели is_female на 1 - is_male и приведём подобные. Тогда мы получим линейную регрессию от is_male: если была модель a+b*is_female+c*is_male, то станет модель a + b*(1-is_male)+c*is_male = a+b + (c-b)*is_male. Тогда для любых значений, которые может выдавать новая модель на области определения, существует модель линейной регрессии от is_male, выдающая те же значения. Тогда при минимизации суммы квадрата ошибок более хорошую оценку, чем лучшая модель линейной регрессии от is_male, новая модель дать не может.

IV. Коэффициенты не изменятся, качество не изменится. Как мы определили, для любой пары значений для is_male=0 и is_male=1 существует модель линейной регрессии от is_male, выдающая эти значения. При удваивании набора наблюдений оптимальное значение для is_male=0, которое нужно взять, останется тем же, так как всё ещё нужно будет взять среднее (или можно заметить, что для любого значения, которое мы попробуем подставить, сумма квадратов ошибок просто возрастёт в два раза). Аналогично, оптимальное значение для is_male=1 останется тем же. Тогда и модель останется той же. При сохранении коэффициентов качество измениться не может.