

## Задача 2

- I. Формула:  $grade = 4.16026667 - 0.18740000 * is\_male$

	Средний балл
Мальчики	3.97286667
Девочки	4.16026667
Все ученики	4.06656667, при условии, что мальчиков и девочек одинаковое количество, если это не так — необходимо знать соотношение мальчиков к девочкам для вычисления.

- II.  $grade = 3.97286667 + 0.18740000 * is\_female$

Тогда значение для мальчиков будет свободным членом формулы, а когда переменная  $is\_female = 1$ , надо добавлять разность между средними баллами, чтобы средний балл увеличивался для девочек (по отношению к среднему баллу мальчиков, но сами значения не изменятся).

- III. Добавление обеих переменных никак не поможет улучшить точность формулы, потому что значение переменной  $is\_male$  обратно пропорционально  $is\_female$ , а именно их зависимость можно описать следующей формулой:  $is\_male = 1 - is\_female$  (предполагается, что аналитик может точно определить принадлежность каждого ученика ровно одному из полов).

- IV. Никак не изменятся ни коэффициенты, ни качество регрессии. При добавлении абсолютно таких же данных, он получит такую же регрессию, а точность не увеличится, так как фактически будет использовано ровно столько же информации для получения формулы. Ему стоит опросить всех участников в своем классе, а если он уже это сделал, то единственное, что поможет лучше предсказывать средний балл — добавление других факторов к регрессии.