Задача 2

1. Формула:

|  |  |
| --- | --- |
|  | 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. |

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

1. Добавление обеих переменных никак не поможет улучшить точность формулы, потому что значение переменной обратно пропорционально , а именно их зависимость можно описать следующей формулой: (предполагается, что аналитик может точно определить принадлежность каждого ученика ровно одному из полов).
2. 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 своём классе, а если он уже это сделал, то единственное, что поможет лучше предсказывать средний балл — добавление других факторов к регрессии.