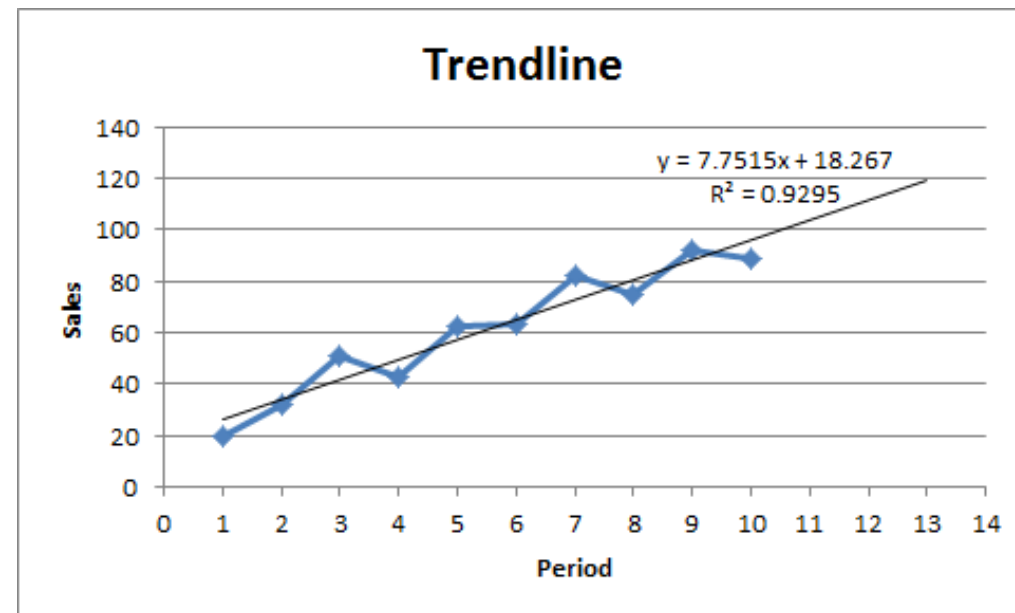


Регрессионный анализ

Лекция 8

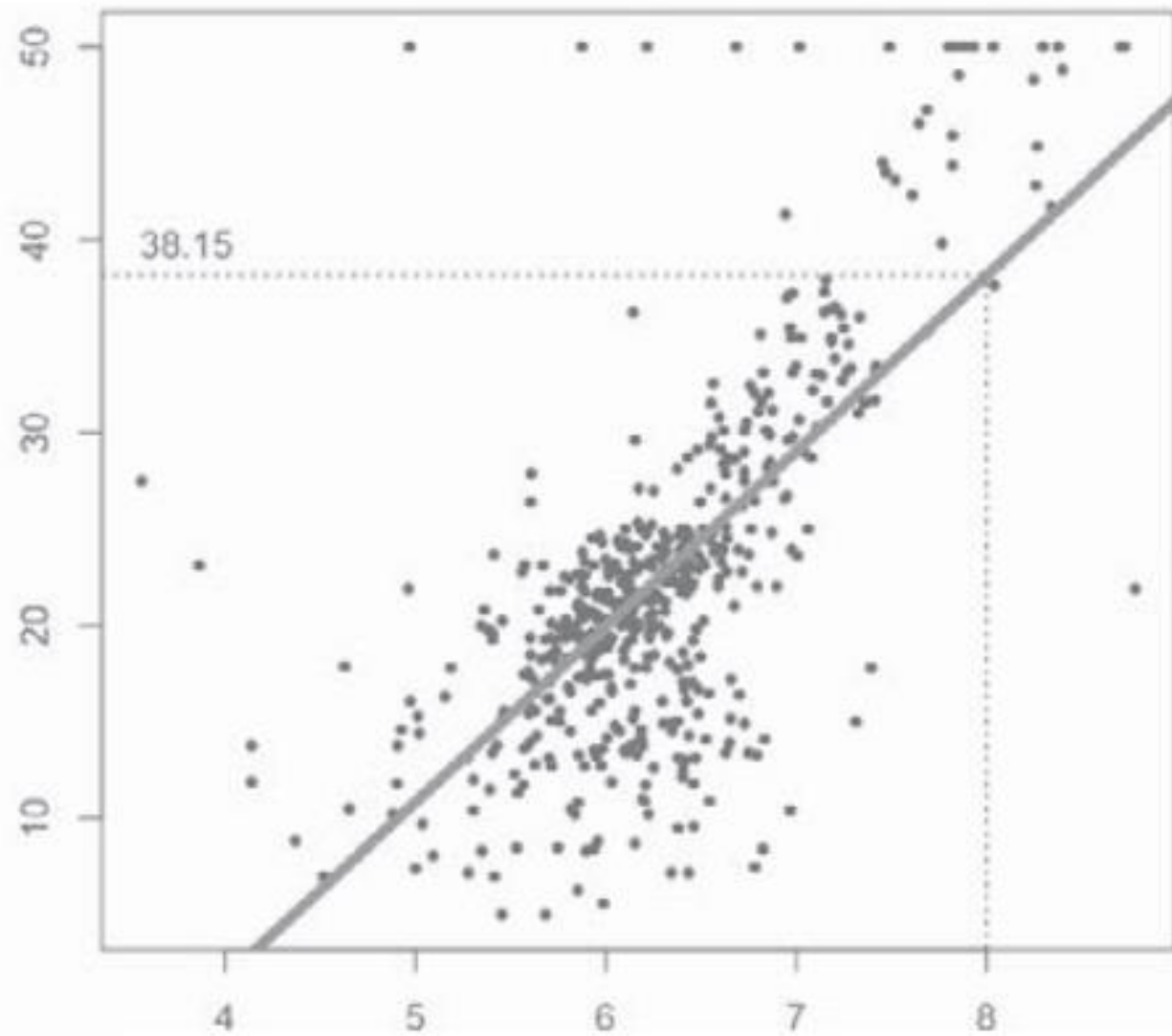
Выведение линии тренда



Пример: предсказание цен на дома

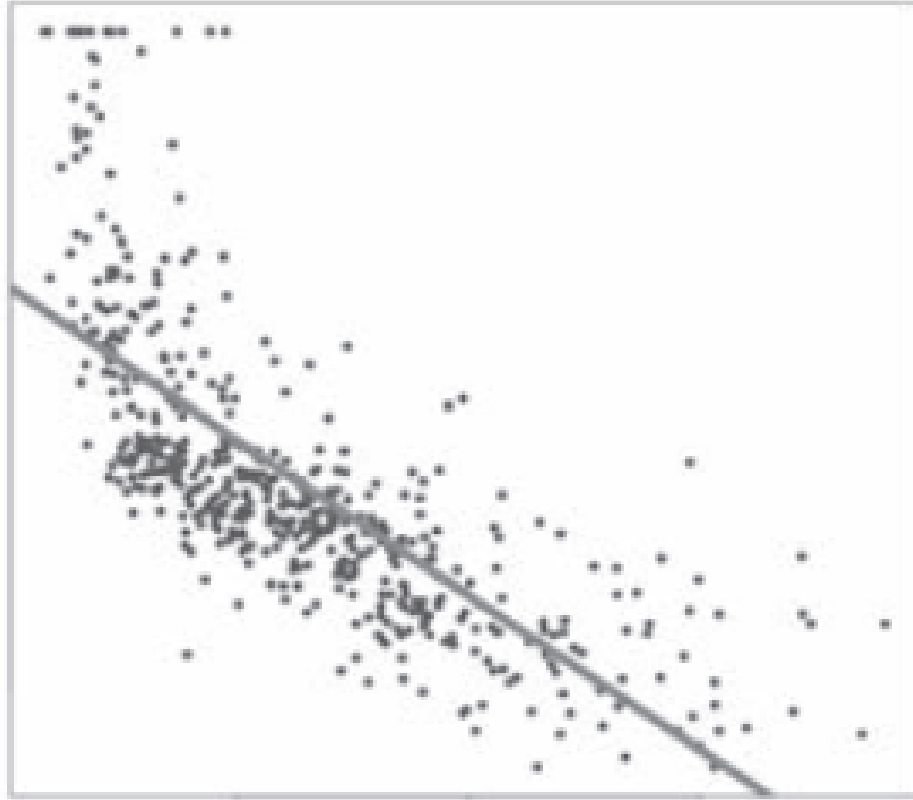
- Мы использовали данные за 1970-е годы о ценах на дома в Бостоне.
- Предварительный анализ показывает, что двумя сильнейшими предикторами цен на дома являются число комнат в доме и доля соседей с низким доходом

Средние цены на дома
в тыс. долл.

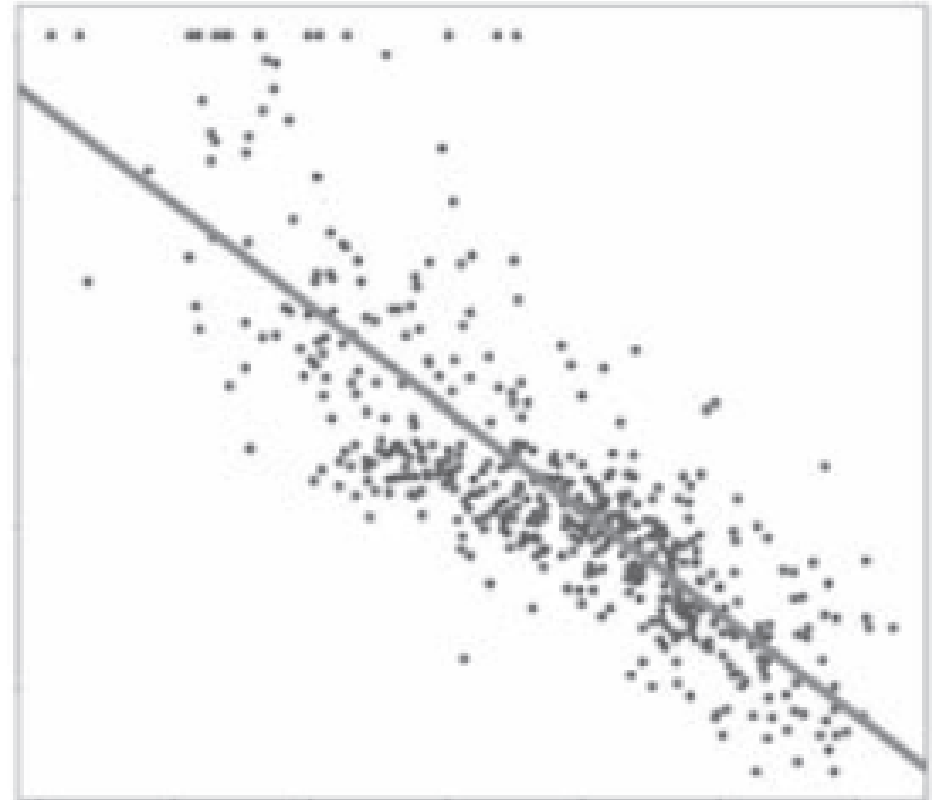


Число комнат

Средние цены на дома

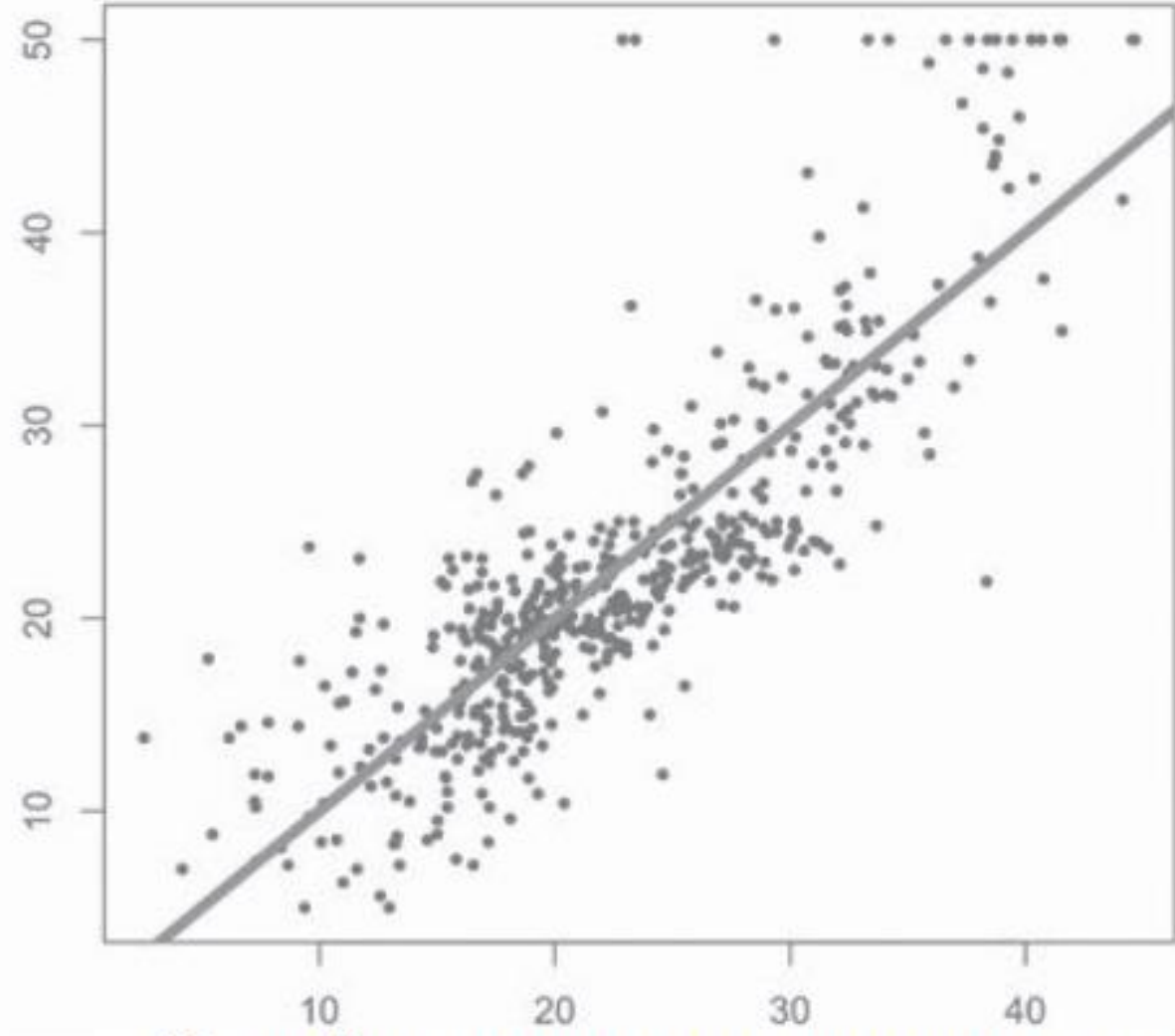


а) исходный



б) измененный

Средние цены на дома
в тыс. долл.

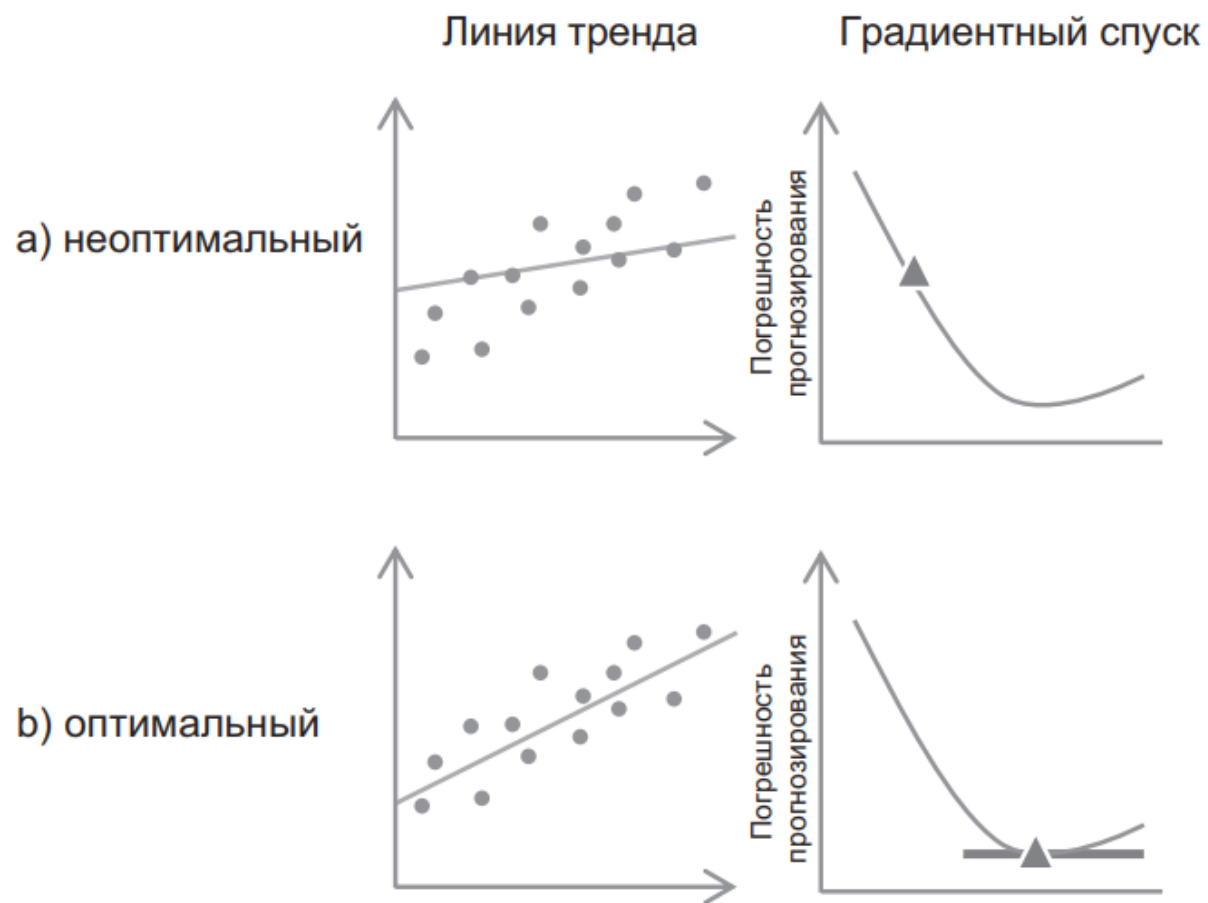


Скомбинированные предикторы

Таблица 1. Средняя прогностическая ошибка при использовании трех разных линий тренда

	Погрешность прогнозирования (в тыс. долл.)
Число комнат	4,4
Влияние окружения	3,9
Число комнат и влияние окружения	3,7

Градиентный спуск



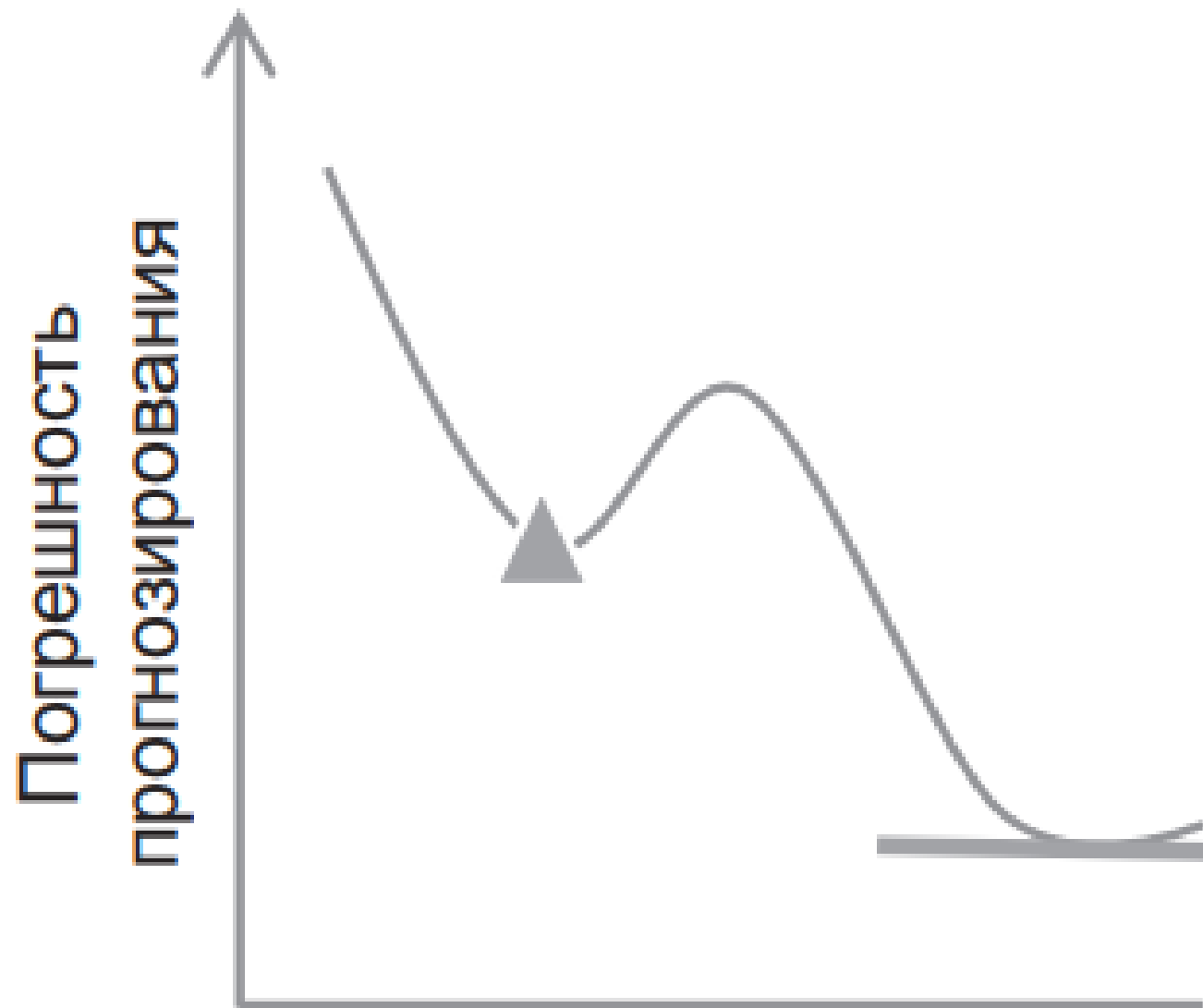
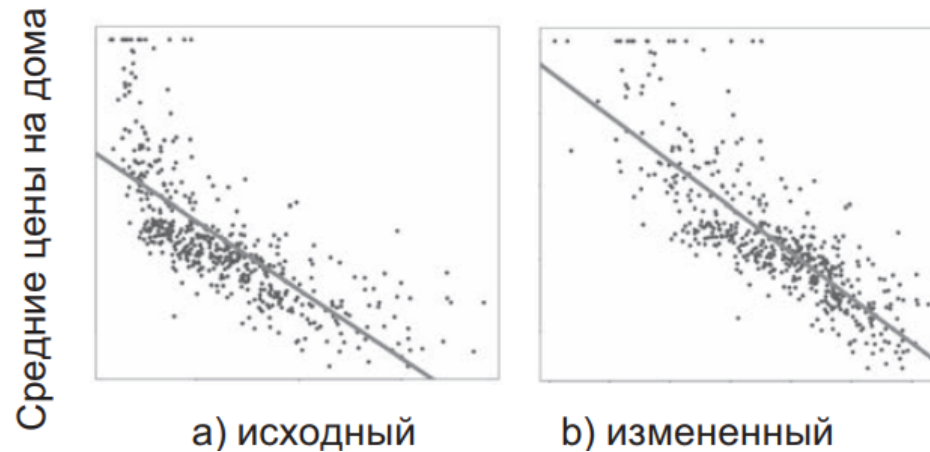


Рис. 5. Как ближайшая яма может быть ошибочно принята за оптимальную точку (треугольник), хотя истинная оптимальная точка находится ниже ее (черта)

Коэффициенты регрессии

- После получения оптимального набора регрессионных предикторов их нужно интерпретировать.
- Вес регрессионных предикторов называется коэффициентом регрессии.
- Уравнение регрессии для примера с ценами на дома будет выглядеть примерно так:

цена = 2,7 (количество комнат) – 6,3(% соседей с низким доходом).



Коэффициенты корреляции

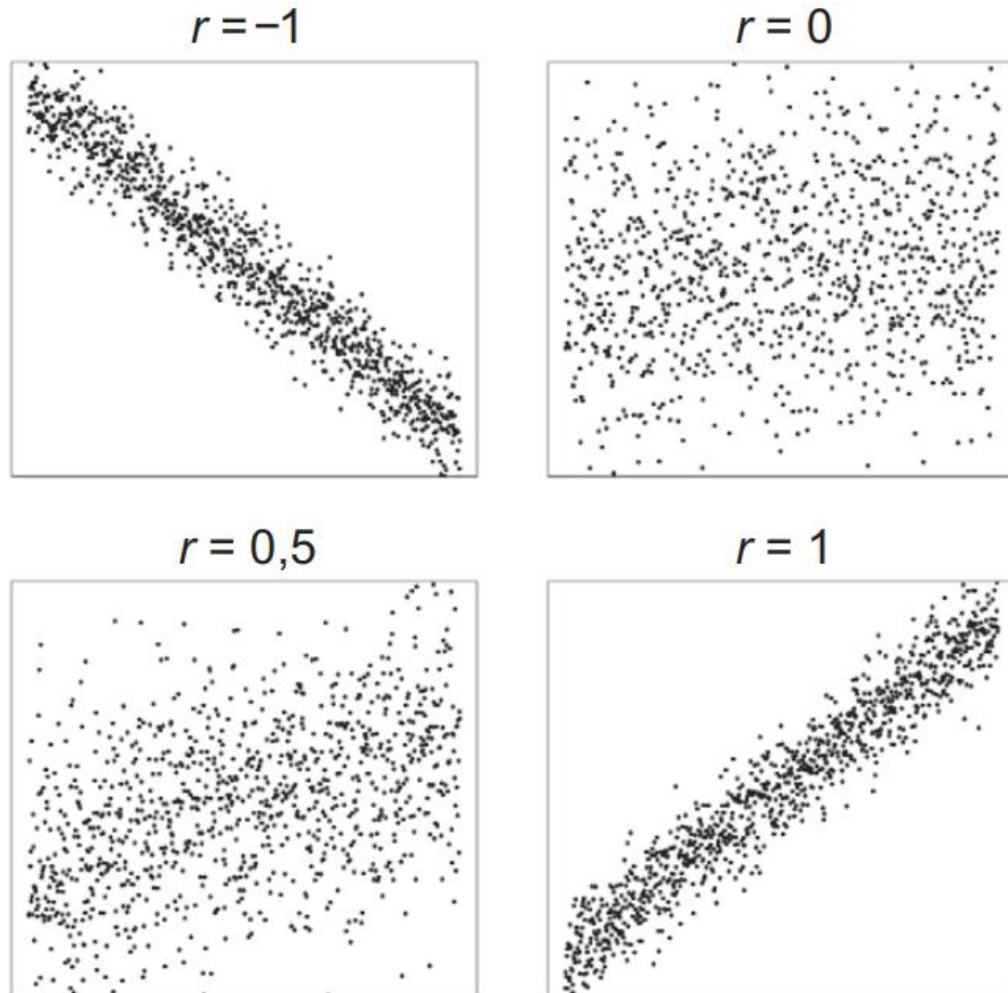
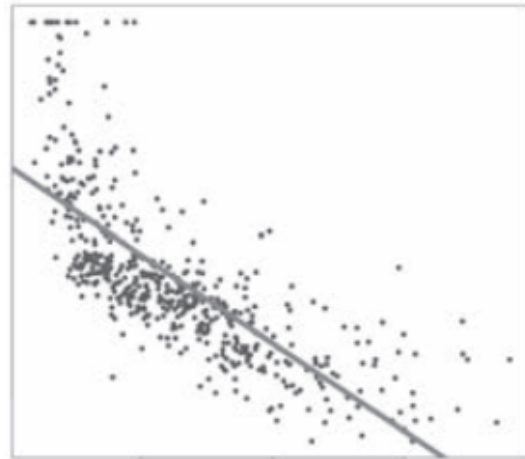


Рис. 6. Пример распределения данных в соответствии с различными коэффициентами корреляции

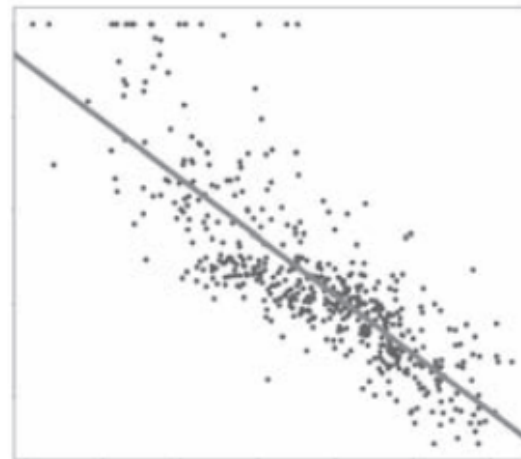
Ограничения

- Чувствительность к резко отклоняющимся значениям
- Искажение веса при корреляции предикторов
- Криволинейные тренды
- Корреляция не говорит о причинности

Средние цены на дома



а) исходный



б) измененный

Краткие итоги

- Регрессионный анализ находит линию наилучшего соответствия, тяготеющую к максимально возможному числу элементов данных.
- Линия тренда выводится на основании уравновешенной комбинации предикторов. Вес предиктора называется коэффициентом регрессии. Он показывает силу одного предиктора в присутствии других.
- Регрессионный анализ хорошо работает в условиях низкой корреляции между предикторами, отсутствия резко отклоняющихся значений и там, где линия тренда ожидается в виде прямой линии.