Задача 1

Условие

Набор данных: swiss

Объясняемая перменная: Catholic

Регрессоры: Agriculture, Examination

1. Оцените среднее значение, дисперсию и СКО переменных, указанных во втором и третьем столбце.
2. Постройте зависимости вида y = a + bx, где y – объясняемая переменная, x – регрессор (для каждого варианта по две зависимости).
3. Оцените, насколько «хороша» модель по коэффициенту детерминации R2?
4. Оцените, есть ли взаимосвязь между объясняемой переменной и объясняющей переменной (по значению p-статистики, «количеству звездочек» у регрессора в модели).

Решение

1. Для оценки среднего значения используем функцию mean(), дисперсии – var(), СКО – sd(). Полученный результат:
   * Среднее значение *Catholic* = 41.143829787234
   * Среднее значение *Agriculture* = 50.6595744680851
   * Среднее значение *Examination =* 16.4893617021277
   * Дисперсия *Catholic =* 1739.29453718779
   * Дисперсия *Agriculture =* 515.799417206291
   * Дисперсия *Examination =* 63.6466234967623
   * СКО *Catholic =* 41.7048502837235
   * СКО *Agriculture =* 22.7112178714901
   * СКО *Examination* = 7.97788339703973
2. Для построение линейной зависимости используем функцию lm(). Полученный результат:
   * y = 3.8313 + 0.7365x *(Catholic ~ Agriculture)*
   * y = 90.514 – 2.994x *(Catholic ~ Examination)*
3. Для просмотря R2 используем функцию summary(). Мы видим, что коэффицент детерминации для первой модели *(Catholic ~ Agriculture)* равен 0.1609, то есть первая модель объясняет 16% колебаний перменной *Catholic*, это очень плохой показатель. Коэффицент детерминации для второй модели *(Catholic ~ Exmination)* равен 0.328, то есть вторая модель объясняет 33% колебаний перменной *Catholiс*, это тоже очень плохой показатель.
4. P-статистика тоже доступна при использовании функции summary(). P-значения первой модели имеют 0 и 2 звездочки, что говорит о крайне слабой взаимосвязи между параметрами *(Catholic ~ Agriculture)*. P-значения второй модели имеют 3 звездочки каждое, что, наоборот, говорит о невероятно сильной связи между параметрами *(Catholic ~ Examination)*.

Вывод

В первой модели *(Catholic ~ Agriculture)* прчинно-следственная связь между поведением перменных крайне слабая, к тому же она нелинейна. Вторая модель *(Catholic ~ Examination)* показывает очень сильную причинно-следственную связь между поведением переменных, но она нелинейна и/или требует дополнительных регрессоров.

Приложение 1

Код решения задания:

