

## ВОПРОСЫ НА ПОНИМАНИЕ

Перед тем как приступить к решению домашних задач, попробуйте ответить на следующие вопросы. Это простые вопросы на понимание. Ответы на них включать в домашнюю работу не нужно.

1. Что можно сказать про случайные величины  $X$  и  $Y$ , если  $\text{Corr}(X, Y) = 1$ ?
2. Чем отличаются коэффициенты корреляции Пирсона и Спирмена? В каких случаях лучше пользоваться коэффициентом корреляции Спирмена?
3. Что такое коэффициент детерминации? Что происходит с коэффициентом детерминации, при добавлении признаков в регрессионную модель? А при удалении?
4. В каких предположениях работают критерии Стьюдента и Фишера о значимости коэффициентов регрессии?
5. Какую гипотезу проверяет критерий Фишера ( $F$ -тест), когда мы вызываем `summary` при построении линейной регрессии в Python или R?

## КОВАРИАЦИЯ И КОРРЕЛЯЦИЯ. ЛИНЕЙНАЯ РЕГРЕССИЯ.

Следующие задания — практические. В них необходимо попытаться применить все возможные инструменты, которые мы изучали, чтобы ответить на заданный в задаче вопрос. Не забывайте об одном из самых важных инструментов — визуализации. Ваши решения необходимо сопроводить краткими комментариями и выводами, которые Вы сделали на основе анализа.

**Упражнение 1** (15 баллов). В файле `lifeline.xls` содержатся 50 пар наблюдений из исследования докторов Л. Матера и М. Уилсона. В нем рассматривались следующие переменные:  $X$  — длина «линии жизни» на левой руке в сантиметрах (с точностью до 0.15 см) и  $Y$  — продолжительность жизни человека (округленная до ближайшего целого года). Изучите корреляцию  $X$  и  $Y$ . Верно ли, что  $X$  и  $Y$  связаны линейной регрессионной зависимостью?

**Упражнение 2** (35 баллов). В файле `homes.csv` содержатся данные о цене продажи 50 домов. В файле содержатся следующие признаки:

- 1) `Sell` — цена продажи;
- 2) `List` — запрашиваемая цена;
- 3) `Living` — жилая площадь;
- 4) `Rooms` — количество комнат;
- 5) `Beds` — количество спален;
- 6) `Baths` — количество ванных комнат;
- 7) `Age` — возраст жилого помещения;
- 8) `Acres` — площадь жилого помещения;
- 9) `Taxes` — налоги, которые должен платить владелец жилого помещения.

Что можно сказать о связи первых двух признаков? Постройте (хорошую) линейную регрессию для признака `Sell` с и без признака `List`. Не забудьте о преобразовании признаков, если они будут нужны.