

Практика 9.

Задания по темам:

Линейные и ранговые корреляции.
Коэффициент корреляции r-Пирсона.
Коэффициент детерминации.
Линия регрессии.

Блок 1. Коэффициент корреляции r-Пирсона

Задание 1 Вычисление r-Пирсона вручную

Даны две переменные x и y :

$x : 2, 4, 5, 6, 8$

$y : 3, 5, 6, 7, 9$

- (a) Вычислите средние $\bar{x}$, $\bar{y}$.
- (b) Найдите ковариацию $\text{Cov}(x, y) = \frac{1}{n-1} \sum (x_i - \bar{x})(y_i - \bar{y})$.
- (c) Найдите стандартные отклонения s_x и s_y .
- (d) Вычислите $r = \frac{\text{Cov}(x, y)}{s_x s_y}$.
- (e) Интерпретируйте знак и величину r .

Задание 2 Свойства коэффициента Пирсона

Ответьте на вопросы:

- (a) В каких пределах изменяется r ?
- (b) Что означает $r = 1$? $r = -1$? $r = 0$?
- (c) Докажите, что r не меняется при линейных преобразованиях: $x' = a + bx$, $y' = c + dy$ (при $b > 0$, $d > 0$).

Задание 3 Проверка значимости r

По выборке объема $n = 20$ получен $r = 0.35$. Проверьте гипотезу $H_0 : \rho = 0$ при $\alpha = 0.05$. Используйте формулу $t = r \sqrt{\frac{n-2}{1-r^2}}$. Сравните с критическим значением $t_{0.025, 18}$. Сделайте вывод.

Блок 2. Коэффициент детерминации R^2

Задание 4 Смысл R^2

Что показывает коэффициент детерминации R^2 ? Запишите формулу через

сумму квадратов: $R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$, где $SS_{tot} = \sum (y_i - \bar{y})^2$, $SS_{res} = \sum (y_i - \hat{y}_i)^2$.
Чему равен R^2 при $r = 0.6$? $r = 0.9$? Что означает $R^2 = 0.81$?

Задание 5 Расчёт R^2 по данным

Для данных из задания 1:

- (а) Используя найденный r , вычислите R^2 .
- (б) Интерпретируйте: какая доля дисперсии y объясняется линейной связью с x ?

Задание 6 Заблуждения о R^2

Верны ли утверждения?

- (а) Чем больше R^2 , тем лучше модель подходит для предсказания.
- (б) $R^2 = 0$ означает полное отсутствие связи между x и y .
- (с) R^2 может быть отрицательным.

Блок 3. Линия регрессии

Задание 7 Уравнение прямой регрессии

Для данных задания 1 найдите коэффициенты регрессии y на x :

- (а) Наклон $b = r \cdot \frac{s_y}{s_x}$.
- (б) Свободный член $a = \bar{y} - b\bar{x}$.
- (с) Запишите уравнение $\hat{y} = a + bx$.
- (д) Предскажите $\hat{y}$ для $x = 7$.

Задание 8 Геометрический смысл

Нарисуйте схематично диаграмму рассеивания (по данным задания 1) и проведите линию регрессии. Отметьте на графике точку $(\bar{x}, \bar{y})$. Почему линия регрессии всегда проходит через эту точку?

Задание 9 Метод наименьших квадратов (МНК)

Объясните, почему МНК минимизирует сумму квадратов остатков $\sum (y_i - \hat{y}_i)^2$. Что произойдёт с линией регрессии, если вместо минимизации квадратов минимизировать сумму модулей остатков? (Качественно.)

Задание 10 Предсказание и экстраполяция

Для выборки, отражающей зависимость роста ребёнка от возраста (от 0 до 12 лет), построена регрессионная прямая. Можно ли использовать её для предсказания роста взрослого человека в 25 лет? Почему?

Блок 4. Ранговые корреляции (Спирмен, Кендалл)

Задание 11 Расчёт ранговой корреляции Спирмена

Даны оценки двух судей (баллы) для 5 участников:

Участник	A	B	C	D	E
Судья 1	5	4	3	2	1
Судья 2	3	4	1	5	2

- (a) Проранжируйте оценки каждого судьи (назначая ранги от 1 до 5).
- (b) Вычислите разности рангов d_i и $\sum d_i^2$.
- (c) Найдите $\rho_s = 1 - \frac{6 \sum d_i^2}{n(n^2-1)}$.
- (d) Интерпретируйте: насколько согласованы мнения судей?

Задание 12 Сравнение Пирсона и Спирмена

Для каких данных (приведите пример) r Пирсона будет близок к 1, а ρ_s Спирмена — к 0? (Подсказка: немонотонная зависимость.) Нарисуйте пример диаграммы рассеивания.

Задание 13 Коэффициент корреляции Кендалла τ (ознакомительно)

В чём отличие τ Кендалла от ρ_s Спирмена? Какой из них более устойчив к малым объёмам выборки?

Блок 5. Обработка данных на компьютере (MS Excel)

Задание 14 Вычисление r и построение диаграммы в Excel

Опишите пошагово, как для данных из задания 1:

- (a) Ввести данные в два столбца.
- (b) Построить точечную диаграмму (диаграмму рассеивания).
- (c) Добавить линию тренда (линейную) и отобразить на графике уравнение регрессии и R^2 .
- (d) С помощью функции КОРРЕЛ вычислить коэффициент корреляции.
- (e) Сравнить полученные значения с ручными расчётами.

Задание 15 Использование функции ЛИНЕЙН

В Excel есть функция ЛИНЕЙН для расчёта параметров регрессии. Что она возвращает? Как извлечь наклон, свободный член и стандартные ошибки? (Краткая справка.)

Задание 16 Ранговые корреляции в Excel

В стандартном Excel нет прямой функции для ρ_s Спирмена. Предложите способ вычисления ρ_s с помощью функций РАНГ.СП и КОРРЕЛ. Напишите последовательность действий.

Комплексные задания

Задание 17 Полный регрессионно-корреляционный анализ

Ниже приведены данные о расходе топлива (л/100 км) в зависимости от скорости (км/ч) для 8 автомобилей:

Скорость	30	40	50	60	70	80	90	100
Расход	9.2	8.0	7.4	6.8	6.7	7.0	7.8	9.5

- (a) Постройте диаграмму рассеивания (можно от руки или в Excel). Какой характер связи (линейный, нелинейный) вы наблюдаете?

- (b) Вычислите коэффициент корреляции Пирсона.
- (c) Найдите коэффициент детерминации и интерпретируйте его.
- (d) Постройте линию линейной регрессии $\hat{y} = a + bx$.
- (e) Оцените расход при скорости 55 км/ч и при 120 км/ч. Насколько надёжна вторая оценка?
- (f) Рассчитайте ранговую корреляцию Спирмена. Сравните с r .

Задание 18 Сравнение двух корреляций

Исследователь изучил две пары переменных: Пара А: $x = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10$; $y = 1, 2, 3, 4, 5, 6, 7, 8, 9, 100$ (выброс в последней точке). Пара В: x те же, $y = 10, 9, 8, 7, 6, 5, 4, 3, 2, 1$ (строгая отрицательная линейная).

- (a) Для каждой пары вычислите r Пирсона (можно приближённо или качественно).
- (b) Для каждой пары вычислите ρ_s Спирмена.
- (c) Какой коэффициент лучше характеризует связь в паре А? Почему?
- (d) В паре В что покажут r и ρ_s ? Сделайте вывод о робастности.

Задание 19 Применение в психологии

Психолог предположил, что между уровнем тревожности (баллы по тесту) и успеваемостью (баллы за экзамен) существует отрицательная корреляция. Он обследовал 12 студентов и получил $r = -0.48$, $p = 0.03$.

- (a) Какой процент дисперсии успеваемости объясняется тревожностью? Как это вычислить?
- (b) Постройте уравнение линейной регрессии, если известно: $\bar{x} = 35$, $\bar{y} = 70$, $s_x = 8$, $s_y = 12$.
- (c) Какой успеваемости следует ожидать при тревожности 40 баллов? При тревожности 50? Стоит ли доверять последнему предсказанию?
- (d) Какие ещё факторы могут влиять на успеваемость, не учтённые в модели?

Задание 20 Расчёты в Excel (пошаговая инструкция)

Создайте в MS Excel таблицу с данными из комплексного задания 13 (расход топлива). Выполните:

- (a) Построение точечной диаграммы, добавление линии тренда с выводом R^2 и уравнения.
- (b) Вычисление коэффициента корреляции через КОРРЕЛ.
- (c) Расчёт коэффициента Спирмена через ранжирование и КОРРЕЛ.
- (d) Подготовьте краткий отчёт (на основе вывода Excel) с интерпретацией: какой вид связи, насколько она сильна, какая доля вариации объяснена, каковы коэффициенты регрессии.