

Практика 10.

Задания по темам:

Вычисление линии регрессии при линейной корреляции.

Коэффициент корреляции r-Спирмена.

Обработка данных на компьютере (MS Excel, Statistica).

Блок 1. Вычисление линии регрессии при линейной корреляции

Задание 1 Параметры линейной регрессии

Для зависимости y от x заданы статистики:

$$\bar{x} = 10, \quad \bar{y} = 20, \quad s_x = 2, \quad s_y = 5, \quad r = 0.8.$$

- (a) Найдите коэффициент наклона $b = r \cdot \frac{s_y}{s_x}$.
- (b) Найдите свободный член $a = \bar{y} - b\bar{x}$.
- (c) Запишите уравнение линии регрессии $\hat{y} = a + bx$.
- (d) Предскажите $\hat{y}$ при $x = 12$ и $x = 15$.

Задание 2 Вычисление регрессии по исходным данным

Даны пары (x, y) :

$$(1, 3), (2, 4), (3, 6), (4, 7), (5, 9).$$

- (a) Вычислите средние $\bar{x}$, $\bar{y}$, а также $\sum (x_i - \bar{x})^2$, $\sum (y_i - \bar{y})^2$ и $\sum (x_i - \bar{x})(y_i - \bar{y})$.
- (b) Найдите коэффициент корреляции r .
- (c) Найдите коэффициенты регрессии b и a .
- (d) Постройте диаграмму рассеивания (схематично) и нарисуйте линию регрессии.

Задание 3 Связь между коэффициентами регрессии

Покажите, что коэффициент наклона b можно выразить как $b = \frac{\text{Cov}(x, y)}{\text{Var}(x)}$.
Используя данные задания 2, проверьте это равенство.

Задание 4 Остатки и их свойства

Для регрессии из задания 2:

- (a) Вычислите предсказанные значения $\hat{y}_i$ для каждого x_i .
- (b) Найдите остатки $e_i = y_i - \hat{y}_i$.
- (c) Убедитесь, что сумма остатков равна нулю (с точностью до округлений).
- (d) Чему равна ковариация между $\hat{y}_i$ и e_i ? (Качественно.)

Блок 2. Коэффициент корреляции r-Спирмена

Задание 5 Расчёт ранговой корреляции вручную

Два эксперта оценили 7 проектов по 10-балльной шкале:

Проект	1	2	3	4	5	6	7
Эксперт А	8	7	9	6	5	8	7
Эксперт В	7	6	8	5	4	9	6

- (a) Проранжируйте оценки каждого эксперта (назначая ранги от 1 до 7; при совпадении – средний ранг).
- (b) Вычислите разности рангов d_i и сумму квадратов разностей $\sum d_i^2$.
- (c) Найдите коэффициент Спирмена: $r_s = 1 - \frac{6 \sum d_i^2}{n(n^2-1)}$.
- (d) Интерпретируйте полученное значение: насколько согласованы мнения экспертов?

Задание 6 Свойства r_s

Ответьте на вопросы:

- (a) В каких пределах изменяется r_s ?
- (b) При каком условии $r_s = 1$? $r_s = -1$?
- (c) Если между рангами нет линейной монотонной связи, но есть нелинейная монотонная, то что покажет r_s ? (Например, $y = x^2$ на положительных x .)
- (d) Почему r_s более устойчив к выбросам, чем r Пирсона?

Задание 7 Проверка значимости r_s

Для $n = 12$ вычислен $r_s = 0.52$. Проверьте гипотезу $H_0: \rho_s = 0$ при $\alpha = 0.05$. Критические значения можно взять из таблицы: для $n = 12$ и $\alpha = 0.05$ (двусторонняя) $r_s \approx 0.587$ или можно использовать t-статистику $t = r_s \sqrt{\frac{n-2}{1-r_s^2}}$. Сделайте вывод.

Задание 8 Сравнение r_s и r Пирсона на одних данных

Для данных задания 2 (Блок 1) вычислите r_s (проранжируйте x и y). Сравните его с ранее найденным r Пирсона. Почему они близки? Когда r и r_s могут сильно различаться?

Блок 3. Обработка данных на компьютере (MS Excel, Statistica)

Задание 9 Линейная регрессия в MS Excel

Опишите пошагово (со скриншотами или подробно в тексте), как в MS Excel:

- (a) Построить точечную диаграмму (диаграмму рассеивания) для данных из задания 2 (Блок 1).
- (b) Добавить линию тренда (линейную) и вывести на график уравнение регрессии и R^2 .
- (c) Вычислить коэффициенты регрессии с помощью функции ЛИНЕЙН.
- (d) Найти коэффициент корреляции через функцию КОРРЕЛ.

Задание 10 Ранговая корреляция Спирмена в MS Excel

В стандартном «Пакете анализа» Excel нет прямой функции для r_s . Предложите два способа получить r_s :

- а) С помощью функций РАНГ.СР (или РАНГ.СРВ) и КОРРЕЛ.
- б) С помощью надстройки «Анализ данных» → «Корреляция» (но там только Пирсон). Поэтому опишите ручной расчёт с ранжированием.

Приведите последовательность действий для данных из задания 5.

Задание 11 Работа в Statistica

В пакете Statistica (предположительно) есть модули для корреляционного и регрессионного анализа.

- (a) Как в Statistica вычислить коэффициент корреляции Пирсона и Спирмена? (Назовите меню: «Статистика» → «Основные статистики и таблицы» → «Корреляционные матрицы» и т.д.)
- (b) Как построить линию регрессии и получить её уравнение? (Например, модуль «Множественная регрессия».)
- (c) В чём преимущества Statistica перед Excel при выполнении непараметрических корреляций и регрессионной диагностики?

Задание 12 Сравнение вывода двух пакетов

Возьмите данные из комплексного задания (ниже). Выполните (мысленно или на реальном компьютере) вычисление регрессии и корреляции Пирсона/Спирмена в Excel и в Statistica. Должны ли результаты совпадать? Почему могут быть незначительные расхождения (например, округление, способы ранжирования)?

Комплексное задание

Задание 13 Анализ реальных данных: зависимость дохода от возраста

Исследователь собрал данные о 8 случайно выбранных людях:

Возраст (x), лет	22	25	30	35	40	45	50	55
Доход (y), тыс. руб.	30	35	40	48	55	58	60	62

- (a) Постройте диаграмму рассеивания (от руки или в Excel). Оцените визуально связь (знак, форма, теснота).
- (b) Вычислите вручную (или с помощью калькулятора) коэффициенты линейной регрессии y на x : найдите b , a и запишите уравнение.
- (c) Найдите коэффициент корреляции Пирсона r .
- (d) Найдите коэффициент детерминации R^2 и интерпретируйте его.
- (e) Вычислите ранговую корреляцию Спирмена r_s .
- (f) Сравните r и r_s . Объясните возможное различие.
- (g) Используя уравнение регрессии, предскажите доход для возраста 28 лет и 65 лет. Оцените надёжность предсказаний.

Задание 14 Обработка в Excel и Statistica (практическое задание)

Для данных из задания 12 выполните следующее (опишите действия, даже если нет доступа к ПО, дайте мысленную инструкцию):

- (a) В Excel: постройте точечную диаграмму, добавьте линию тренда, выведите R^2 и уравнение. Сравните с ручными расчётами.
- (b) В Excel: вычислите r_s через ранжирование и функцию КОРРЕЛ. Сравните с ручным значением.
- (c) В Statistica (или в мысленном режиме): постройте график регрессии с доверительными интервалами, получите матрицу корреляций (Пирсона и Спирмена).
- (d) Интерпретируйте результаты: можно ли утверждать, что между возрастом и доходом существует сильная положительная линейная связь? Является ли эта связь строго линейной на всём диапазоне возрастов?

Задание 15 Анализ малой выборки с выбросом

Возьмём данные из задания 12, но добавим 9-го человека: возраст 20 лет, доход 100 тыс. руб. (аномально высокий для 20 лет).

- (a) Как изменится визуальная картина диаграммы рассеивания?
- (b) Пересчитайте (приблизённо или с помощью идеи) коэффициент Пирсона r и коэффициент Спирмена r_s для расширенной выборки. Какой из них изменится сильнее? Почему?
- (c) Как изменится линия регрессии? Станет ли она более пологой или крутой? Обоснуйте.

Задание 16 Создание отчёта на основе компьютерной обработки

Представьте, что вы выполнили расчёты в Excel для данных задания 12. Оформите краткий отчёт, содержащий:

- Диаграмму рассеивания с линией тренда.
- Таблицу с основными статистиками: средние, стандартные отклонения, r , r_s , R^2 , коэффициенты регрессии.
- Уравнение регрессии и его интерпретацию.
- Вывод о характере и силе связи (с использованием как линейной, так и ранговой корреляции).