

Практика 8.

Задания по темам:

Понятие корреляции, диаграмма рассеивания.

Простейшие разновидности корреляции.

Коэффициент корреляции.

Величина корреляции и сила связи.

Блок 1. Понятие корреляции. Диаграмма рассеивания

Задание 1 Построение диаграммы рассеивания вручную

Даны пары значений (x, y) :

$(1, 2), (2, 3), (3, 5), (4, 4), (5, 6), (6, 7), (7, 8), (8, 9), (9, 10), (10, 11)$

- (a) Нанесите точки на координатную плоскость (сделайте рисунок от руки).
- (b) Какой вид связи (положительная, отрицательная, отсутствует) вы наблюдаете?
- (c) Предположите, будет ли коэффициент корреляции близок к $+1$, -1 или 0 .

Задание 2 Выброс и корреляция

Представьте диаграмму рассеивания из 9 точек, образующих плотное облако с положительной корреляцией, и одну точку-выброс, расположенную далеко справа внизу.

- (a) Как выброс повлияет на значение коэффициента корреляции (увеличит, уменьшит, не изменит)?
- (b) Покажите схематически, как изменится «облако» после удаления выброса. Какой вывод о чувствительности корреляции к выбросам можно сделать?

Блок 2. Простейшие разновидности корреляции

Задание 3 Классификация по знаку и форме

Заполните таблицу, приведя по одному реальному примеру из жизни или науки для каждого типа корреляции:

Тип корреляции	Пример
Положительная линейная	
Отрицательная линейная	
Нулевая (отсутствие связи)	
Нелинейная (не монотонная)	

Задание 4 Монотонная vs немонотонная связь

Какая связь называется монотонной? Приведите пример немонотонной зависимости (например, параболической): когда при увеличении x сначала y растёт, а затем убывает. Можно ли для такой связи использовать коэффициент корреляции Пирсона? Почему?

Блок 3. Коэффициент корреляции

Коэффициент корреляции Пирсона (r)

Задание 5 Формула и вычисление вручную

Даны две переменные: $x = 1, 2, 3, 4, 5$; $y = 2, 4, 5, 4, 5$.

- Вычислите средние $\bar{x}$, $\bar{y}$ и произведения отклонений $\sum (x_i - \bar{x})(y_i - \bar{y})$.
- Вычислите суммы квадратов отклонений $\sum (x_i - \bar{x})^2$ и $\sum (y_i - \bar{y})^2$.
- Найдите коэффициент корреляции Пирсона: $r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$.
- Интерпретируйте полученное значение r .

Задание 6 Свойства r

Выберите верные утверждения:

- r изменяется от 0 до 1.
- r безразмерен.
- $r = 0$ означает полное отсутствие какой-либо связи.
- $r = -0.9$ указывает на сильную отрицательную линейную связь.
- При перестановке x и y значение r не меняется.
- На значение r не влияют линейные преобразования шкал ($x' = a + bx$, $y' = c + dy$).

Задание 7 Проверка значимости r

По выборке объёма $n = 15$ вычислен коэффициент корреляции $r = 0.45$. Проверьте гипотезу $H_0 : \rho = 0$ против $H_1 : \rho \neq 0$ при $\alpha = 0.05$. Используйте формулу $t = r\sqrt{\frac{n-2}{1-r^2}}$ и сравните с критическим значением $t_{0.025,13}$. Какой вывод?

Ранговый коэффициент корреляции Спирмена (ρ_s)

Задание 8 Расчёт ранговой корреляции

Даны ранги R_x и R_y для 5 объектов:

Объект	1	2	3	4	5
R_x	1	2	3	4	5
R_y	3	4	1	2	5

- (a) Вычислите разности $d_i = R_{x_i} - R_{y_i}$ и их квадраты.
- (b) Примените формулу Спирмена: $\rho_s = 1 - \frac{6 \sum d_i^2}{n(n^2-1)}$.
- (c) Что показывает полученное значение?
- (d) В каких случаях предпочтительнее использовать ρ_s вместо r Пирсона?

Задание 9 Связь между Пирсоном и Спирменом

При каких условиях ρ_s приблизительно равен r ? Если данные имеют выбросы, что произойдёт с r и ρ_s ? Какой из коэффициентов более робастен?

Блок 4. Величина корреляции и сила связи

Задание 10 Шкала Чеддока для оценки силы связи

По шкале Чеддока (или аналогичной) дайте качественную оценку силы линейной корреляции для следующих значений $|r|$: 0.1, 0.3, 0.5, 0.7, 0.9, 0.99. Заполните таблицу:

$ r $	Сила связи (по Чеддоку)
0.0 – 0.1	_____
0.1 – 0.3	_____
0.3 – 0.5	_____
0.5 – 0.7	_____
0.7 – 0.9	_____
0.9 – 0.99	_____

Задание 11 Коэффициент детерминации

Что такое коэффициент детерминации R^2 ? Чему он равен для $r = 0.6$ и $r = 0.8$? Как интерпретировать $R^2 = 0.64$?

Задание 12 Статистическая значимость vs практическая значимость

При очень большом объёме выборки ($n = 1000$) коэффициент корреляции $r = 0.08$ оказался статистически значимым ($p < 0.001$). Можно ли утверждать, что связь сильная? Объясните различие между статистической значимостью и величиной эффекта.

Комплексные задания

Задание 13 Анализ реальной зависимости

Исследователь измерил рост (см) и вес (кг) у 10 случайных прохожих:

Рост	165	170	168	172	175	180	178	182	169	174
Вес	55	65	60	68	70	85	80	90	62	72

- (a) Постройте диаграмму рассеивания (на миллиметровке или схематично).
- (b) Вычислите коэффициент корреляции Пирсона (вручную или с помощью калькулятора, представив вычисления).
- (c) Оцените силу связи по шкале Чеддока. Проверьте значимость при $\alpha = 0.05$.
- (d) Как изменится вывод, если добавить человека ростом 190 см и весом 50 кг (выброс)? Сделайте предположение.

Задание 14 Сравнение корреляций Пирсона и Спирмена

По данным предыдущего задания вычислите ранговую корреляцию Спирмена (предварительно проранжировав рост и вес).

- (a) Сравните значения r и ρ_s . Одинаковы ли они? Если нет, то за счёт чего различие?
- (b) Какая корреляция (линейная или монотонная) лучше описывает зависимость роста и веса в этой выборке?
- (c) Какой коэффициент вы бы использовали, если бы распределение веса было сильно асимметричным? Почему?

Задание 15 Интерпретация коэффициента в контексте

Прочитайте выдержку из отчёта: «Коэффициент корреляции между количеством часов, потраченных на подготовку к экзамену, и итоговым баллом составил $r = 0.75$ ($p < 0.01$)».

- (a) Можно ли утверждать, что увеличение часов подготовки на 1 час приводит к увеличению балла на 0.75? Поясните.
- (b) Что означает $p < 0.01$ в этом контексте?
- (c) Какая доля изменчивости баллов объясняется количеством часов подготовки?
- (d) Может ли существовать третья переменная, влияющая на обе переменные? Приведите пример.

Задание 16 Построение диаграммы рассеивания в Excel

Опишите пошагово, как построить диаграмму рассеивания (точечную диаграмму) в MS Excel для данных из задания 12. Укажите, как добавить линию тренда и отобразить на ней коэффициент R^2 . Что покажет R^2 на этой диаграмме? Как в Excel вычислить коэффициент корреляции (функция КОРРЕЛ)?