

Школа лингвистики, 2023-24 уч. год**Теория вероятностей****Коэффициент корреляции и регрессия (21.05.2024)***И. В. Щуров, Д. А. Филимонов, Р. Я. Будылин*

Задача 1. По данным выборкам вычислить коэффициенты корреляции Пирсона, Спирмена и Кендала. Что можно сказать о связи между переменными ξ и η ?

(а) (0.7; 0), (1.1; 4), (1.2; 2)

(б) (1; 1), (2; 10), (3; 100)

Задача 2. Пусть дана выборка (1.1; 5), (0.3; -3), (1.1; 1), (0.9; -1), (0.7; 0), (1.1; 4), (0.9; 2), (0.7; -2), (0.9; 3), (1.1; -4). Выпишите выборку из рангов для дальнейшего подсчёта коэффициента корреляции Спирмена (сам коэффициент считать не нужно).

Задача 3. Социологи взяли средние уровни образованности и дохода по штатам США, и корреляция между ними составила 0.7. С другой стороны, если измерять эти две величины индивидуально, то коэффициент корреляции оказывается равным приблизительно 0.5. Как это можно объяснить?

Задача 4. По данным выборкам посчитать коэффициенты для простой линейной модели регрессии переменной y на x .

(а) (1; 0), (2; 3), (3; 2)

(б) (1; 1), (2; 10), (3; 100)

(в) (1; 3), (2; 2), (3; 1)

Задача 5. У детей измерили уровень IQ, потом занимались с ними по специальной программе. После этого уровень IQ был измерен ещё раз. Тест показал, что программа не оказала сильного влияния на рост IQ. Но также оказалось, что дети с высоким IQ в среднем понизили свой уровень, а дети с из группы с более низким IQ его повысили. Чем можно объяснить такое усреднение?

Задача 6. Для выяснения зависимости смертности от уровня загрязнения воздуха было проведено следующее исследование. Брались данные о числе смертей на 100 тыс. человек за день и об уровне загрязнения в разных городах в течении года. Затем все эти точки были нанесены на график. И получилось, что точки сгруппированы вокруг прямой с отрицательным наклоном, то есть чем выше уровень загрязнения, тем ... ниже смертность. Попробуйте объяснить, почему так вышло.