

**Школа лингвистики, 2020-21 уч. год****Теория вероятностей****Коэффициент корреляции и регрессия (25.05.2021)***И. В. Щуров, Д. А. Филимонов, Р. Я. Будылин*

**Задача 1.** По данным выборкам вычислить коэффициент корреляции Пирсона, Спирмена и Кендала. Что можно сказать о связи между переменными  $\xi$  и  $\eta$ ?

- (a) (1; 0), (2; 3), (3; 2)
- (b) (1; 1), (2; 10), (3; 100)

**Задача 2.** 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 3.** По данным выборкам посчитать коэффициенты для простой линейной модели регрессии переменной  $y$  на  $x$ .

- (a) (1; 0), (2; 3), (3; 2)
- (b) (1; 1), (2; 10), (3; 100)
- (c) (1; 3), (2; 2), (3; 1)

**Задача 4.** 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 5.** Для выяснения зависимости смертности от уровня загрязнения воздуха было проведено следующее исследование. Бралась 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.