

Математика и статистика, часть 2**Коэффициент корреляции Пирсона (28.05.2020)**

А. А. Макаров, А. А. Тамбовцева, Н. А. Василёнок

Задача 1. По результатам исследования черт личности на примере мультфильма «Чип и Дейл», проведенного среди студентов 1 курса ОП «Политология», выяснилось, что выборочная ковариация между веселостью в процентах от Дейла x_i и ворчливостью в процентах от Чипа y_i составила -169 .

- (а) Рассчитайте выборочный коэффициент корреляции Пирсона, если известно, что в исследовании приняли участие 73 человека, выборочное стандартное отклонение веселости составило $s_x = 26.2$, а выборочная дисперсия ворчливости $s_y = 26.6$. Проинтерпретируйте полученное значение.
- (б) Проверьте гипотезу о равенстве истинного коэффициента корреляции нулю на 5% уровне значимости против двухсторонней альтернативы через расчет p -value. Сделайте содержательный вывод.

Задача 2. По 27 парам наблюдений двух количественных признаков вычислили коэффициент корреляции Пирсона. Он оказался равен $R = 0.4$. Можно ли считать, что истинная корреляция между этими признаками отлична от нуля? Проверьте гипотезу о равенстве истинного коэффициента корреляции нулю на 5% уровне значимости против двухсторонней альтернативы через построение критической области. Сделайте содержательный вывод.

Задача 3. По 38 парам наблюдений двух количественных признаков вычислили коэффициент корреляции Пирсона. Он оказался равен $R = -0.3$.

- (а) Можно ли считать, что истинная корреляция между этими признаками отлична от нуля? Проверьте гипотезу о равенстве истинного коэффициента корреляции нулю на 5% уровне значимости. Сделайте содержательный вывод.
- (б) Сколько респондентов должно участвовать в исследовании, чтобы указанный коэффициент корреляции значимо отличался от нуля?

Губерния	% крепостных	ВРП на д.н.
Олонецкая	4	56
Оренбургская	7	56
Пензенская	46	48
Харьковская	30	66
Бессарабская	1	78
Гродненская	40	56
Могилевская	65	52

Задача 4. В таблице выше приведены данные экономического развития 7 случайно отобранных губерний европейской части Российской империи. Для каждой губернии

известен процент крепостных крестьян от всего населения губернии в 1856 г. и ВРП на душу населения в 1897 г.¹

- (a) Постройте диаграмму рассеяния для приведенных данных, откладывая значения процента крепостных на оси OX .
- (b) Перенесите имеющиеся данные в R, создав вектора `serfs` и `grpps`. Рассчитайте средние значения для каждого показателя.
- (c) Постройте доверительный интервал для среднего процента крепостных от населения губернии в помощь функции `MeanCI()`. Как вы думаете, почему доверительный интервал получился таким?
- (d) Рассчитайте коэффициент корреляции Пирсона с помощью функции `cor.test()`. Чему равна вероятность $P(|T| \geq t_{\text{набл}})$? Сделайте вывод о значимости коэффициента корреляции.

¹Источник: Markevich, Andrei, A Regional Perspective on the Economic Development of the late Russian Empire. Available at SSRN: <https://ssrn.com/abstract=2555273>