

ОПРЕДЕЛЕНИЕ КОЭФФИЦИЕНТА КОРРЕЛЯЦИИ ВЕЛИЧИН ПРИ МАЛОМ КОЛИЧЕСТВЕ НАБЛЮДЕНИЙ

Захаров И.П., Сафарян Г.Г., Сергиенко М.П.

Харьковский национальный университет радиоэлектроники
(61166, Харьков, пр. Ленина, 14, кафедра метрологии и измерительной
техники, тел. (057) 702-1331)

E-mail: newzip@ukr.net; факс (057) 702-1013

There were analyzed an exactness of accidental data's correlation coefficient determination by parametric and nonparametric methods. The correction coefficients for these methods optimization were received. The distribution law does not have an influence on the determination correlation coefficient result. It was demonstrated by the example of data with normal, uniform, arc sine distributions and its combinations.

В процессе решения различных теоретических и прикладных задач, включающих математическую обработку полученных данных, часто возникает вопрос определения и учета возможной корреляции между этими данными, поскольку в противном случае существует риск существенно завысить либо занижить (в зависимости от характера корреляционной связи) характеристики точности результатов эксперимента. Особенно эта проблема актуальна для тех видов измерений, где получение большого количества наблюдений технически и/или экономически затруднено.

Обычно для нахождения корреляции между двумя выборками используют параметрические (коэффициент корреляции R и выборочный коэффициент корреляции r) и непараметрические (коэффициенты ранговой корреляции ρ Спирмена и τ Кендалла) методы [1]. Более распространены на практике параметрические методы, однако они чувствительны к наличию в выборках грубых погрешностей и промахов, которые затруднительно устранить при малом объеме выборки, поскольку при уменьшении объема данных точность результатов заметно снижается. Основными преимуществами непараметрических методов расчета коэффициента корреляции является независимость от закона распределения исследуемых величин и нечувствительность к возможным выбросам в исходных данных. При этом коэффициент ранговой корреляции ρ Спирмена является аналогом выборочного коэффициента корреляции r и эти критерии сравнимы по мощности, в то же время коэффициент ранговой корреляции τ Кендалла отличается повышенной сложностью расчета.

Рассмотрим случай, когда исходные величины распределены по нормальному закону, т.е. когда коэффициенты R и r идентичны. Исследования показали, что при расчете коэффициента корреляции при малом числе наблюдений для коррекции результата целесообразнее применение выра-

жения $r^* = r \left[1 + \frac{1-r^2}{2(n-1)} \right]$ вместо выражения $r^* = r \left[1 + \frac{1-r^2}{2(n-3)} \right]$, приведенного в [2].

Использование коэффициента R с учетом эффективных оценок математического ожидания и среднего квадратического отклонения (СКО) для разных законов распределения не является адекватным. При этом, как показали проведенные исследования, СКО коэффициента корреляции может возрасти до 3. Коэффициент выборочной корреляции r с достаточной точностью может быть использован при разных законах распределения исходных величин. Его СКО при этом не изменяется.

Для расчета коэффициентов ранговой корреляции ρ Спирмена и τ Кендалла необходимо вводить поправки, показанные на рис. 2.

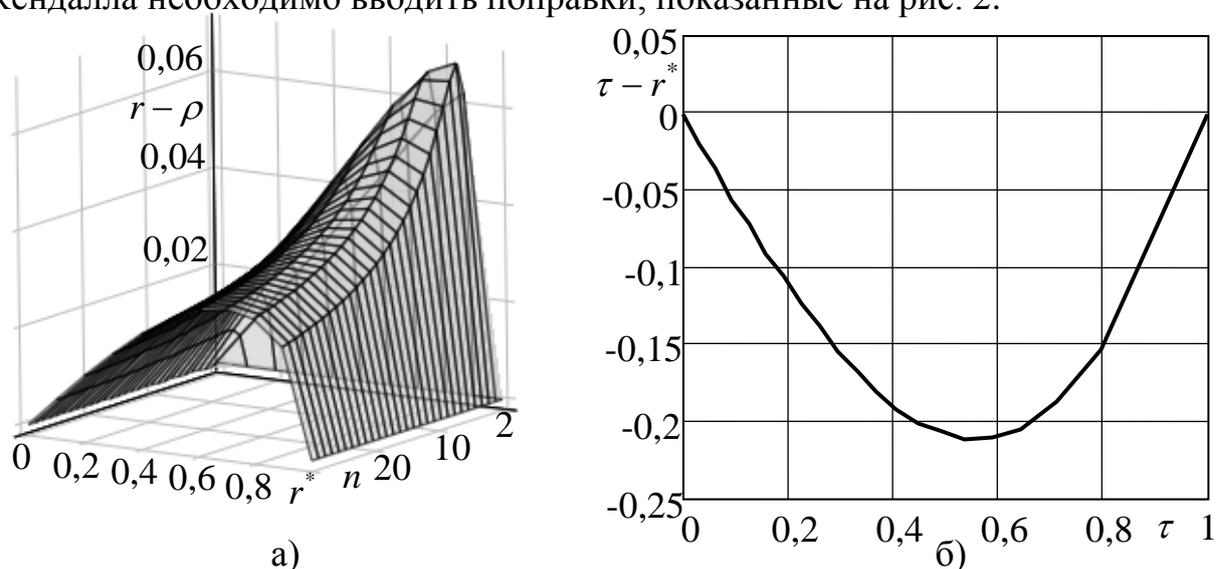


Рис.1. Поправки к коэффициентам ранговой корреляции ρ Спирмена (а) и τ Кендалла (б)

Преимуществом определения непараметрического коэффициента с введением соответствующей поправки является возможность его применения при полном отсутствии информации о законе распределения исходных величин и наличии грубых погрешностей и промахов.

Таким образом, исследования показали, что оптимальным является использование выборочного коэффициента корреляции r с учетом поправки. Если же в исходных данных присутствуют грубые погрешности и промахи, следует воспользоваться коэффициентом ранговой корреляции ρ Спирмена или τ Кендалла с введением соответствующей поправки.

Список литературы:

1. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. – М.: Наука, 1983. – 416 с.
2. Степнов М.Н. Статистические методы обработки результатов механических испытаний: Справочник. – М.: Машиностроение, 1985. – 232 с.