УДК 303.723:311.16

ПРОВЕРКА ЗНАЧИМОСТИ ДЛЯ КОЭФФИЦИЕНТА КОРРЕЛЯЦИИ

Савельева А.С., Черненко Н.А.

ФГБОУ ВО «Ставропольский государственный аграрный университет», Ставрополь, e-mail: saveleva25.73@mail.ru

Исследователя нередко интересует, как связаны между собой две или большее количество переменных в одной или нескольких изучаемых выборках. Например, может ли рост влиять на вес человека или может ли давление влиять на качество продукции? Такого рода зависимость между переменными величинами называется корреляционной, или корреляцией. Задача корреляционного анализа сводится к установлению направления (положительное или отрицательное) и формы (линейная, нелинейная) связи между варьирующими признаками, измерению ее тесноты, и, наконец, к проверке уровня значимости полученных коэффициентов корреляции. В статье рассматривается применение корреляционного анализа с расчетом коэффициентов корреляции Стьюдента. Изложенный материал дает общие сведения об оценке степени тесноты взаимосвязи между переменными.

Ключевые слова: корреляционный анализ, коэффициенты корреляции, выборка, переменные, случайные колебания, генеральная совокупность

TEST OF SIGNIFICANCE FOR CORRELATION COEFFICIENT

Saveleva A.S., Chernenko N.A.

Stavropol State Agrarian University, Stavropol, e-mail: saveleva25.73@mail.ru

The researcher is often interested in how two or more variables are related to one or more of the samples studied. For example, can growth influence the weight of a person, or can pressure affect product quality? This kind of relationship between variables is called correlation, or correlation. The problem of correlation analysis is reduced to establishing the direction (positive or negative) and the form (linear, nonlinear) of the relationship between the varying signs, measuring its tightness, and, finally, to checking the significance level of the correlation coefficients obtained. The article discusses the use of correlation analysis with calculation of correlation coefficients t-test. The article gives an overview of the assessment of the degree of closeness of the relationship between the variables.

Keywords: correlation analysis, correlation coefficients, sampling, variables, random vibration, a general population

Как неоднократно отмечалось, для статистического вывода о наличии или отсутствии корреляционной связи между исследуемыми переменными необходимо произвести проверку значимости выборочного коэффициента корреляции. В связи с тем что надежность статистических характеристик, в том числе и коэффициента корреляции, зависит от объема выборки, может сложиться такая ситуация, когда величина коэффициента корреляции будет целиком обусловлена случайными колебаниями в выборке, на основании которой он вычислен. При существенной связи между переменными коэффициент корреляции должен значимо отличаться от нуля. Если корреляционная связь между исследуемыми переменными отсутствует, то коэффициент корреляции генеральной совокупности р равен нулю [9].

Как всякая статистическая характеристика, выборочный коэффициент корреляции является случайной величиной, т.е. его значения случайно рассеиваются вокруг одноименного параметра генеральной совокупности (истинного значения коэффициента корреляции) [1].

При отсутствии корреляционной связи между переменными у и х коэффициент

корре ляции в генеральной совокупности равен нулю. Но из-за случайного характера рассеяния принципиально возможны ситуации, когда некоторые коэффициенты корреляции, вычисленные по выборкам из этой совокупности, будут отличны от нуля [5].

Процедура проверки значимости начинается с формулировки нулевой гипотезы H_0 . В общем виде она заключается в том, что между параметром выборки и параметром гене ральной совокупности нет какихлибо существенных различий. Альтернативная гипотеза H_1 состоит в том, что между этими параметрами имеются существенные различия. Например, при проверке наличия корреляции в генеральной совокупности нулевая гипотеза заключается в том, что истинный коэффициент корреляции равен нулю H_0 : p = 0. Если в результате проверки окажется, что нулевая гипотеза не приемлема, то выборочный коэффициент корреляции $r_{x,y}$ значимо отличается от нуля (нулевая гипотеза отвергается и принимается альтернативная H_1) [7, 10].

При проверке значимости исследователь устанавливает уровень значимости α, который дает определенную практическую уверенность в том, что ошибочные заключения будут сделаны только в очень

редких случаях. Уровень значимости выражает вероятность того, что нулевая гипотеза H_0 отвергается в то время, когда она в действительности верна. Ясно, что имеет смысл выбирать эту вероятность как можно меньшей [2].

Пусть известно распределение выборочной характеристики, являющейся несмещенной оценкой параметра генеральной совокупности. Выбранному уровню значимости α соответствуют под кривой этого распределения заштрихованные площади (см. рис. 1). Незаштрихованная площадь под кривой распределения определяет вероятность

$$P=1-a$$
.

Границы отрезков на оси абсцисс под заштрихованными площадями называют критическими значениями, а сами отрезки образуют критическую область, или область отклонения гипотезы [3].

Когда же надо убедиться в том, что одна величина в среднем строго больше или мень ше другой, используется односторонняя критическая область. Если распределение выборочной характеристики симметрично, то уровень значимости двусторонней критической области равен α , а односторонней

 $-\frac{a}{2}$ (см. рис. 1). Далее мы лишь укажем критерии значимости для различных процедур, не останавливаясь на их построении [6].

Проверяя значимость коэффициента парной корреляции, устанавливают наличие или отсутствие корреляционной связи между исследуемыми явлениями. При отсутствии связи коэффициент корреляции генеральной совокупности равен нулю (p=0). Процедура проверки начинается с формулировки нулевой и альтернативной гипотез [8]:

 H_0 : различие между выборочным коэффициентом корреляции r и $\rho = 0$ незначимо,

 H_1 : различие между r и p=0 значимо, и следовательно, между переменными у и х имеется существенная связь. Из альтернативной гипотезы следует, что нужно воспользоваться двусторонней критической областью.

Выборочный коэффициент корреляции при определенных предпосылках связан со случайной величиной t, подчиняющейся распределению Стьюдента с f=n-2 степенями свободы [4].

Вычисленная по результатам выборки статистика

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}\tag{1}$$

сравнивается с критическим значением, определяемым по таблице распределения Стьюдента при заданном уровне значимости α и f=n-2 степенях свободы. Правило применения критерия заключается в следующем: если $|t| > t_{f,a}$, то нулевая гипотеза на уровне значимости α отвергает-

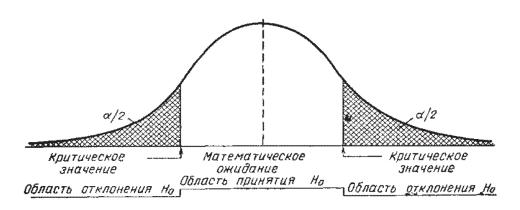


Рис. 1 Проверка нулевой гипотезы H_0

ся, т. е. связь между переменными значима; если

$$|t| > t_{f,a}$$

то нулевая гипотеза на уровне значимости α принимается. Отклонение значения r от $\rho = 0$ можно приписать случайной вариации. Данные выборки характеризуют рассматриваемую гипотезу как весьма возможную и правдоподобную, т. е. гипотеза об отсутствии связи не вызывает возражений.

Рассмотрим на конкретном примере проверку гипотезы о значимости выборочного коэффициента корреляции.

Для 15 пар порядковых переменных (x, y) был вычислен коэффициент корреляции Спирмена, который оказался равным $P_s = 0,64$. При уровне значимости $\alpha = 5\%$ проверить гипотезу H_0 о равенстве генерального коэффициента ранговой корреляции нулю $Ps_r = 0$ при конкурирующей гипотезе $Ps_r \neq 0$.

Решение:

$$H_0: Ps_r = 0$$
$$H_1: Ps_r \neq 0$$

Постановка задачи:

Найдем наблюдаемое значение критерия:

$$t = \frac{Ps\sqrt{n-2}}{\sqrt{1-Ps^2}} \to t_{\text{Ha6},n} =$$
$$= \frac{0.64\sqrt{15-2}}{\sqrt{1-0.64^2}} \approx 3.00.$$

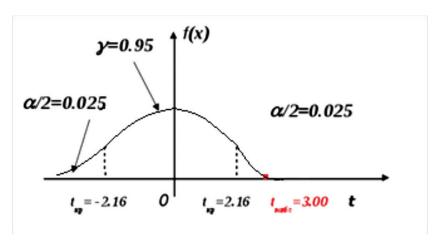
Найдем значения границ двусторонней критической области из условия, что при малых объемах выборок критерий t распределен (при справедливости нулевой гипотезы) по закону распределения Стьюдента с числом степеней свободы k=15-2=13.

На основе таблицы «Критические точки распределения Стьюдента» найдем границы двусторонней критической области на основе

$$a = 0.05 \text{ M} k = 13 \rightarrow t_{\text{KP}} = 2.16.$$

Полученные результаты покажем графически (рис. 2).

Поскольку наблюдаемое значение критерия попало в критическую область, то следует отклонить нулевую гипотезу в пользу альтернативной гипотезы. Это означает, что коэффициент ранговой корреляции значим: между двумя порядковыми случайными величинами X и Y в генеральной совокупности имеется значимая связь, которая говорит о наличии связи этих двух случайных величин.



Puc. 2

Список литературы

- 1. Бондаренко В.А., Донец З.Г., Цыплакова О.Н. Теория игр и финансовые рынки // Финансово-экономические и учетно-аналитические проблемы развития региона: материалы Ежегодной 78-й научно-практической конференции. 2014. С. 231–236.
- 2. Бондаренко В.А., Мамаев И.И., Сахнюк П.А., Сахнюк Т.И. Решение задачи планирования посевов с использованием теории игр // Экономические, инновационные и информационные проблемы развития региона: материалы Международной научно-практической конференции. 2014. С. 56–62.
- 3. Бондаренко Д.В., Бражнев С.М., Литвин Д.Б., Варнавский А.А. Метод повышения точности измерения векторных величин // Наука Парк. 2013. № 6 (16). С. 66–69.
- 4. Гулай Т.А., Литвин Д.Б., Попова С.В., Мелешко С.В. Прогнозирование в регрессионном анализе при построении статистических моделей экономических задач с помощью программы Microsoft excel // Экономика и предпринимательство. 2017. № 8–2 (85–2). С. 688–692.
- 5. Долгополова А.Ф., Гулай Т.А., Литвин Д.Б. Финансовая математика в инвестиционном проектировании (учебное пособие) // Международный журнал прикладных и фундаментальных исследований. -2014. -№ 8-2. -C. 178-179.
- 6. Долгополова А.Ф., Гулай Т.А., Литвин Д.Б. Совершенствование экономических механизмов для решения

- проблем экологической безопасности // Информационные системы и технологии как фактор развития экономики региона. II Международная научно-практическая конференция, 2013. С. 68–71.
- 7. Литвин Д.Б. Дифференциальное исчисление в исследовании производственных моделей. // В сборнике: экономические приоритеты и информационный механизм устойчивого развития регионов России: Сборник научных статей Всероссийской научно-практической конференции, 2017. С. 102–105.
- 8. Литвин Д.Б., Гулай Т.А., Жукова В.А., Мамаев И.И. Модель экономического роста с распределенным запаздыванием в инвестиционной сфере. // Вестник АПК Ставрополья. -2017. № 2 (26). C. 225–228.
- 9. Литвин Д.Б., Шепеть И.П. Моделирование роста производства с учетом инвестиций и выбытием фондов. // Социально-экономические и информационные проблемы устойчивого развития региона: Международная научно-практическая конференция. 2015. С. 114–116.
- 10. Литвин Д.Б., Шепеть И.П., Бондарев В.Г., Литвина Е.Д. Применение дифференциального исчисления функций нескольких переменных к разработке алгоритма определения координат объекта. // Финансово-экономические и учетно-аналитические проблемы развития региона: Материалы Ежегодной 78-й научно-практической конференции, 2014. С. 242–246.