ОБОЗРЕНИЕ
ПРИКЛАДНОЙ И ПРОМЫШЛЕННОЙ
Том 21 МАТЕМАТИКИ Выпуск 5
2014

А. Н. Тырсин, О. В. Ворфоломеева (Челябинск, ЮУрГУ).
Об асимптотической симметрии корреляционного отношения.

Коэффициент парной корреляции Пирсона является одним из распространенных измерителей тесноты корреляционной связи между случайными величинами. Однако его использование ограничено линейным видом корреляционной зависимости. При нелинейных формах зависимостей используют индекс корреляции, который для случая двух переменных обычно называют корреляционным отношением [1]. Корреляционное отношение $R_{Y/X}$ равно

$$R_{Y/X} = \sqrt{R^2_{Y/X}},$$

где $R^2_{Y/X}$ — коэффициент детерминации, равный доле дисперсии результативной переменной Y, объясненной вариацией факторной переменной X.

Справедлива следующая лемма.

Лемма. *Пусть имеем две непрерывные случайные величины X и Y. Тогда теоретические значения корреляционных отношений обладают симметрией*

$$R_{Y/X} = R_{X/Y}. \tag{1}$$

Доказательство. В [2] доказано, что дифференциальная энтропия $H(\mathbf{Y})$ системы непрерывных случайных величин $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_m)$ равна

$$H(\mathbf{Y}) = \sum_{k=1}^{m} H(Y_k) + \frac{1}{2}\sum_{k=2}^{m} \ln(1 - R^2_{Y_k/Y_1\ldots Y_{k-1}}). \tag{2}$$

Согласно (2) для случая двух случайных величин имеем

$$H(X,Y) = H(X) + H(Y) + \frac{1}{2}\ln(1 - R^2_{Y/X}),$$

$$H(Y,X) = H(Y) + H(X) + \frac{1}{2}\ln(1 - R^2_{X/Y}).$$

Очевидно, что $H(X,Y) = H(Y,X)$. Поэтому $\ln(1 - R^2_{Y/X}) = \ln(1 - R^2_{X/Y})$. Отсюда следует, что $R^2_{Y/X} = R^2_{X/Y}$, а значит $R_{Y/X} = R_{X/Y}$.

В ряде источников, например в [3, 4], указывается, что между корреляционными отношениями $R_{X/Y}$ и $R_{Y/X}$ нет какой-либо простой зависимости. Однако, на самом деле, согласно лемме, они равны. Это справедливо для теоретических значений корреляционных отношений. Выборочные оценки $\widehat{R}_{X/Y}$ и $\widehat{R}_{Y/X}$, в отличие от парных линейных коэффициентов корреляции $\widehat{r}_{XY}$ и $\widehat{r}_{YX}$, будут различными. Это вызвано также тем, что для нелинейных корреляционных зависимостей на ограниченных выборках сглаживание и группирование данных не инвариантно к выбору независимой переменной. Но с увеличением объема выборки оценки $\widehat{R}_{X/Y}$ и $\widehat{R}_{Y/X}$ сходятся по вероятности к теоретическому значению $R_{X/Y} = R_{Y/X}$. Как показало статистическое моделирование, оценки $\widehat{R}_{X/Y}$ и $\widehat{R}_{Y/X}$ уже при объеме выборки, равном 100 наблюдениям, оказываются довольно близкими. Разница, как правило, не превышает 0,03.

Соотношение (1) показывает, что при оценивании тесноты нелинейной корреляционной связи между переменными их порядок не существенен.

СПИСОК ЛИТЕРАТУРЫ

1. *Ферстер Э.*, *Ренц Б.* Методы корреляционного и регрессионного анализа. М.: Финансы и статистика, 1983, 302 с.
2. *Тырсин А. Н.*, *Ворфоломеева О. В.* Исследование динамики многомерных стохастических систем на основе энтропийного моделирования. — Информатика и ее применения, 2013, т. 7, в. 4, с. 3–10.
3. *Айвазян С. А.*, *Енюков И. С.*, *Мешалкин Л. Д.* Прикладная статистика: Исследование зависимостей. М.: Финансы и статистика, 1985, 487 с.
4. *Кендалл М.*, *Стьюарт А.* Статистические выводы и связи. М.: Наука, 1973, 900 с.