ՀԱՅԿԱԿԱՆ ՍՍՀ ԳԻՏՈՒԹՅՈՒՆՆԵՐԻ ԱԿԱԴԵՄԻԱՅԻ ԶԵԿՈՒՅՑՆԵՐ
ДОКЛАДЫ АКАДЕМИИ НАУК АРМЯНСКОЙ ССР

XLVIII 1969 5

УДК 519.210

МАТЕМАТИКА

Ф. А. Овсепян

О влиянии коррелированности аргументов на регрессионную функцию

(Представлено чл.-корр. АН Армянской ССР Р. А. Александряном 26/II 1969)

Исследуется регрессионная зависимость одного из составляющих трехмерного случайного вектора от двух других при изменении коррелированности последних.

В дальнейшем всюду: r_{ij} — коэффициент корреляции составляющих X_i и X_j исследуемого вектора; $M(X_i/X_j)$ и $M(X_i/X_j, X_k)$ — соответственно парная и множественная регрессии, представляющие собой линейные функции относительно своих аргументов*; $DX_{i\cdot j}$ и $DX_{i\cdot jk}$ — дисперсия X_i относительно парной и множественной регрессии соответственно.

Известно [1], что $DX_{1.2}$ и $DX_{1.23}$ характеризуют точность представления X_1 соответствующими регрессионными зависимостями, причем имеют место следующие неравенства

$$\begin{gathered} DX_{1.23} \leqslant DX_{1.2}, \\ DX_{1.23} \leqslant DX_{1.3}, \end{gathered} \tag{1}$$

т. е представление X_1 множественной регрессией точнее. Если величины X_2 и X_3 находятся в линейной функциональной зависимости, то точность представления X_1 множественной регрессией такая же, как и по парным. Казалось бы, что точность представления X_1 регрессией $M(X_1/X_2, X_3)$ должна увеличиваться по мере уменьшения коррелированности X_2 и X_3. Однако это не так.

Справедлива следующая

Теорема. *Если случайная величина X_1 коррелирована хотя бы с одной из величин X_2 и X_3, то существует множественная регрессия $M(X_1/X_2, X_3)$ такая, что $DX_{1.23}$ в этом случае меньше дис-*

* Необходимо указать, что это требование не ограничивает общности последующих результатов, поскольку соответствующей заменой нелинейное уравнение можно привести к линейному виду.

персии $\widetilde{D}X_{1.23}$ величины X_1 относительно множественной регрессии при некоррелированных X_2 и X_3.

Доказательство. Без ограничения общности выводов будем предполагать, что составляющие исследуемого вектора центрированы и нормированы. Дисперсия X_1 относительно множественной регрессии в этом случае равна

$$DX_{1.23} = 1 - R^2_{1.23}, \tag{2}$$

где

$$R^2_{1.23} = \frac{r^2_{12} + r^2_{13} - 2r_{12}r_{13}r_{23}}{1 - r^2_{23}}.$$

Утверждение теоремы иными словами можно сформулировать следующим образом: если хотя бы одно из значений r_{12} и r_{13} не равно нулю, то существует r_{23} такое, что $DX_{1.23}$ в этом случае меньше $\widetilde{D}X_{1.23}$, где $\widetilde{D}X_{1.23}$ определяется по (2) при условии $r_{23} = 0$. Укажем эти значения r_{23}.

Введем следующие обозначения

$$\alpha = \sqrt{(1-r^2_{12})(1-r^2_{13})}, \tag{3}$$

$$\beta = \frac{2r_{12}\, r_{13}}{r^2_{12} + r^2_{13}}. \tag{4}$$

Случай I. $r_{12} \neq r_{13}$ и их знаки совпадают

$$\beta < r_{23} \leqslant r_{12}r_{13} + \alpha, \tag{5}$$

$$r_{12}r_{13} - \alpha \leqslant r_{23} < 0. \tag{6}$$

Случай II. $r_{12} = r_{13}$

$$2r^2_{12} - 1 \leqslant r_{23} < 0. \tag{7}$$

Случай III. $|r_{12}| \neq |r_{13}|$ и имеют разные знаки

$$r_{12}r_{13} - \alpha \leqslant r_{23} < \beta, \tag{8}$$

$$0 < r_{23} \leqslant r_{12}r_{13} + \alpha. \tag{9}$$

Случай IV. $r_{12} = -r_{13}$

$$0 < r_{23} \leqslant 1 - 2r^2_{12}. \tag{10}$$

Случай V. $r_{12} = 0$

$$0 < |r_{23}| \leqslant \sqrt{1 - r^2_{13}}. \tag{11}$$

Теорема доказана.

Более того, справедливо следующее утверждение: если хотя бы один из коэффициентов r_{12} и r_{13} не равен нулю, то существует множественная регрессия $M(X_1/X_2, X_3)$, абсолютно точно представляющая случайную величину X_1, т. е. $DX_{1.23}=0$. Коэффициенты соответствующих регрессий определяются при значениях r_{23}, на которых в (5) ÷ (11) достигается равенство.

Рассмотрим частный случай, когда исследуемый вектор распределен по нормальному закону. Используя понятия теории информации, доказанную теорему можно интерпретировать следующим образом: о величине X_1 иногда можно получить значительно больше шенноновской информации, когда величины X_2 и X_3 зависимы, чем в случае их независимости друг от друга. Более того, при некоторой зависимости между X_2 и X_3 количество информации об X_1 равно бесконечности, т. е. по наблюдениям за X_2 и X_3 можно точно указать значение случайной величины X_1.

При другой постановке задачи аналогичные результаты получены в ([2]), но только для случая нормального распределения.

Вычислительный центр
Академии наук Армянской ССР

Ց. Ա. ՀՈՎՍԵՓՅԱՆ

Արգումենտների կոռելյացիոն կապի ազդեցությունը ռեգրեսային ֆունկցիայի վրա

Ուսումնասիրվում է պատահական եռաչափ վեկտորի բաղադրիչներից մեկի ռեգրեսային կապակցությունը՝ մնացածների նկատմամբ, երբ վերջիններիս կոռելյացիոն կապը փոփոխվում է։ r_{ij}-ով նշանակված է հետազոտվող եռաչափ վեկտորի X_i և X_j բաղադրիչների կոռելյացիայի գործակիցը։ Ապացուցվում է հետևյալը. եթե r_{12} և r_{13} կոռելյացիայի գործակիցներից որևէ մեկը զրոյի հավասար չէ, ապա գոյություն ունի այնպիսի կոռելյացիայի գործակից r_{23}, որի դեպքում բազմաթիվ ռեգրեսիան ճշգրիտ է ներկայացնում X_1 պատահական մեծությունը, քան թե չկոռելյացված արգումենտներով ռեգրեսիան։ Կարելի է ապացուցել, որ ինչպես էլ լինեն r_{12} և r_{13} կոռելյացիայի գործակիցները, գոյություն ունի այնպիսի բազմակի ռեգրեսիա, որի օգնությամբ X_1 պատահական մեծության արժեքները որոշվում է անսխալ։

ЛИТЕРАТУРА—ԳՐԱԿԱՆՈՒԹՅՈՒՆ

1 *Т. Андерсон.* Введение в многомерный статистический анализ, М., 1963.
2 *А. Г. Француз*, Известия АН СССР, Техническая кибернетика, № 4, 1967.