Лекция 12

Статистический анализ зависимостей между гидрологическими переменными

Множественная линейная корреляция. Оценка точности уравнения множественной линейной корреляции

(Ахметов С.К.)

Множественная линейная корреляция

На практике возможны случаи, когда ***СВ*** ***Y*** зависит сразу от ряда ***СВ*** $X_1, X_2 \ldots X_n$.

Уравнение регрессии в этом случае будет иметь вид

$$(y - y_{ср.}) = a_1 (x_1 - x_{1,ср.}) + a_2 (x_2 - x_{2,ср.}) + \ldots a_m (x_m - x_{m,\ ср.})$$

$a_{1,}\ a_2 \ldots a_m$ ***- коэффициенты регрессии;***

$y_{ср.},\ x_{1,ср.},\ x_{2,ср\ldots}.\ x_{m,\ ср.}$ ***– средние значения соответственно предиктанта и предикторов;***

m ***– число предикторов***

Т.о., задача сводиться к определению значений $a_{1,}\ a_2 \ldots a_m$***. При небольшом числе предикторов, задачу можно решить методом Крамера. В этом случае нужно рассчитать главный определитель*** D

Расчет главного определителя

$$D = \begin{vmatrix} 1 & r_{0,1} & r_{0,2} & r_{0,3} & \cdots & r_{0,j} & \cdots & r_{0,m} \\ r_{1,0} & 1 & r_{1,2} & r_{1,3} & \cdots & r_{1,j} & \cdots & r_{1,m} \\ r_{2,0} & r_{2,1} & 1 & r_{2,3} & \cdots & r_{2,j} & \cdots & r_{2,m} \\ r_{3,0} & r_{3,1} & r_{3,2} & 1 & \cdots & r_{3,j} & \cdots & r_{3,m} \\ \vdots & \vdots & \vdots & \vdots & & \vdots & & \vdots \\ r_{i,0} & r_{i,1} & r_{i,2} & r_{i,3} & \cdots & r_{i,j} & \cdots & r_{i,m} \\ \vdots & \vdots & \vdots & \vdots & & \vdots & & \vdots \\ r_{m,0} & r_{m,1} & r_{m,2} & r_{m,3} & \cdots & r_{m,j} & \cdots & 1 \end{vmatrix}$$

$r_{i,j}$ – парные коэффициенты корреляции.

Например, $r_{0,3}$ – коэффициент корреляции между предиктантом и третьим предиктором, $r_{2,1}$ – коэффициент корреляции между вторым и первым предикторами. При этом $r_{i,j} = r_{j,i}$, а на главной диагонали $r_{0,0} = r_{1,1} = r_{2,2} = \ldots\ldots r_{m,m} = 1.$

Расчет коэффициентов регрессии

$$a_j = -(-1)^j \frac{\sigma_y}{\sigma_{x,j}} \frac{D_{0,j}}{D_{0,0}}$$

σ_y – среднеквадратическое отклонение предиктанта

$\sigma_{x,j}$ – среднеквадратическое отклонение j – того предиктора

$D_{0,0}$ – определитель, получаемой из главного определителя системы путем вычеркивания из него нулевой строки и нулевого столбца

$D_{0,j}$ - определитель, получаемой из главного определителя системы путем вычеркивания из него нулевой строки и j - того столбца

Правила расчета определителя

$$\begin{vmatrix} a & c \\ b & d \end{vmatrix} = ad - bc$$

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = a_1b_2c_3 + a_3b_1c_2 + a_2b_3c_1 - a_3b_2c_1 - a_1b_3c_2 - a_2b_1c_3$$

Если число переменных равно трем

$$D = \begin{vmatrix} 1 & r_{0,1} & r_{0,2} \\ r_{1,0} & 1 & r_{1,2} \\ r_{2,0} & r_{2,1} & 1 \end{vmatrix}$$

$$D_{0,0} = \begin{vmatrix} 1 & r_{1,2} \\ r_{2,1} & 1 \end{vmatrix} = 1 - r_{2,1} r_{1,2} = 1 - r_{1,2}^2$$

$$D_{0,1} = \begin{vmatrix} r_{1,0} & r_{1,2} \\ r_{2,0} & 1 \end{vmatrix} = r_{1,0} - r_{2,0} r_{1,2}$$

$$D_{0,2} = \begin{vmatrix} r_{1,0} & 1 \\ r_{2,0} & r_{2,1} \end{vmatrix} = r_{1,0} r_{2,1} - r_{2,0}$$

$$a_1 = + \frac{\sigma_y}{\sigma_{x,1}} \frac{(r_{1,0} - r_{2,0} r_{1,2})}{1 - r_{1,2}^2}$$

$$a_2 = - \frac{\sigma_y}{\sigma_{x,2}} \frac{(r_{1,0} r_{2,1} - r_{2,0})}{1 - r_{1,2}^2}$$

Оценка точности уравнения множественной линейной регрессии

- **Сводный коэффициент корреляции можно вычислить двумя способами:**

1. По формуле

$$R = \sqrt{1 - \frac{D}{D_{0,0}}}$$

2. Как парный коэффициент корреляции между фактическими значениями предиктанта и значениями предиктанта, полученными по уравнению множественной линейной регрессии при тех же значениях аргументов.

В отличие от парного коэффициента корреляции, который может меняться от -1 до +1, коэффициент множественной корреляции меняется от 0 до +1.

Оценка точности уравнения множественной линейной регрессии (2)

- ***Стандартная ошибка коэффициента множественной корреляции определяется по формуле***

$$\sigma_R = \frac{1 - R^2}{\sqrt{n - m - 1}}$$

n – длина анализируемых рядов

m - число предикторов

- ***Стандартную ошибку уравнения множественной линейной регрессии можно оценить по формуле***

$$\sigma_{y(X)} \approx \sigma_y^* \sqrt{1 - R^2}$$

- ***Стандартная ошибка j –того коэффициента регрессии можно определить по формуле***

$$\sigma_{a_j} = \frac{\sigma_y^*}{\sigma_{x_j}^*} \sqrt{\frac{(1 - R^2)\Delta_{j,j}}{(n - m - 1)D_{0,0}}}$$

$\Delta_{j,j}$ - минор определителя $D_{0,0}$

Условия приемлемости уравнения регрессии

1. $R \geq 0{,}7$

2. $(R/\sigma_R) \geq 2$

3. $(a_j/\sigma_{a,j}) \geq 2$

$n \geq 10$ ***(при одном предикторе)***

$n \geq 25\text{-}30$ ***(при двух предикторах)***

$n \geq 50\text{-}60$ ***(при четырех предикторах)***

Третье условие считается выполненным, если оно выполняется для каждого коэффициента регрессии в отдельности

СПАСИБО ЗА ВНИМАНИЕ!