Множественная регрессия

На любой экономический показатель как правило оказывает влияние не один, а несколько факторов. Например, спрос на некоторое благо определяется не только ценой данного блага, но и ценами на замещающие и дополняющие блага, доходом потребителей и многими другими факторами. В этом случае вместо парной регрессии $M\left(Y \middle| x\right) = f(x)$ рассматривается *множественная регрессия*

$$M\left(Y \middle| x_1, x_2, \dots, x_k\right) = f(x_1, x_2, \dots, x_k)$$

Задача оценки статистической взаимосвязи переменных Y и $X^{(1)}, X^{(2)}, \dots, X^{(k)}$ формулируется аналогично случаю парной регрессии. *Уравнение множественной регрессии* может быть представлено в виде

$$Y = f(\alpha, \mathrm{X}) + \varepsilon$$

где $\mathrm{X} = \left(X^{(1)}, X^{(2)}, \dots, X^{(k)}\right)$ есть вектор независимых объясняющих переменных (регрессоры), α — вектор *параметров* (подлежащих определению), ε — *случайная ошибка (отклонение)* , Y — *зависимая (объясняемая) переменная.*

Наиболее простой моделью множественной регрессии является линейная модель множественной регрессии аддитивного вида.

Теоретическое линейное уравнение регрессии имеет вид:

$$Y = \alpha_0 + \alpha_1 X^{(1)} + \alpha_2 X^{(2)} + \ldots + \alpha_k X^{(k)} = \alpha_0 + \sum_{j=1}^{k} \alpha_j X^{(j)}$$

или для индивидуальных наблюдений

$$y_i = \alpha_0 + \alpha_1 x_i^{(1)} + \alpha_2 x_i^{(2)} + \ldots + \alpha_k x_i^{(k)}$$

Здесь $\alpha = (\alpha_0, \alpha_1, \alpha_2, \ldots, \alpha_k)$ — вектор размерности *(m* + 1) неизвестных параметров α_j, $(j = \overline{1,k})$, называется j -ым теоретическим коэффициентом регрессии (частичным коэффициентом регрессии). Он характеризует чувствительность величины Y к изменению $X^{(j)}$ регрессора, который отражает влияние на условное математическое ожидание зависимой переменной Y объясняющей переменной $X^{(j)}$, при условии, что все другие объясняющие переменные модели остаются постоянными. α_0 — *свободный член,* определяющий значение Y в случае, когда все объясняющие переменные равны нулю.

Методом оценки параметров уравнения множественной линейной регрессии является *метод наименьших квадратов* (МНК), суть которого состоит в минимизации суммы квадратов отклонений наблюдаемых значений зависимой переменной Y от ее значений $\tilde{Y}$, получаемых по уравнению регрессии. Определим предпосылки МНК, позволяющие его применить

Предпосылки МНК

1. *Математическое ожидание случайного отклонения* ε, *равно нулю для всех наблюдений:*

$$\forall i: \ M(\varepsilon_i) = 0$$

1. *Гомоскедастичность (постоянство дисперсии отклонений).* Дисперсия случайных отклонений e_i постоянная:

$$\forall (i, j): \ D(e_i) = D(e_j) = \sigma^2$$

1. *Отсутствие автокорреляции.*
   Случайные отклонения ε_i и ε_j являются независимыми друг от друга

$$\operatorname{cov}(\varepsilon_i, \varepsilon_j) = \begin{cases} 0, & i \neq j \\ \sigma^2, & i = j \end{cases}$$

Предпосылки МНК (продолжение)

4. *Случайное отклонение должно быть независимо от объясняющих переменных:*

$$\forall (i, j):\ \operatorname{cov}(\varepsilon_i, x_i^{(j)}) = 0$$

4. *Модель является линейной относительно параметров.* Для случая множественной линейной регрессии существенными являются еще две предпосылки.
5. *Отсутствие мультиколлинеарности.*

Между объясняющими переменными отсутствует строгая (сильная) линейная зависимость.

7. *Ошибки* ε_i *имеют нормальное распределение* $\varepsilon_i \sim N(0,\sigma)$.

Теоретическое уравнение регрессии оценивается *эмпирическим уравнением регрессии,* которое имеет вид:

$$Y = a_0 + a_1 X^{(1)} + a_2 X^{(2)} + \ldots + a_k X^{(k)} + e = a_0 + \sum_{j=1}^{k} a_j X^{(j)} + e$$

Здесь $a_0, a_1, a_2, \ldots, a_k$ — оценки теоретических значений $\alpha_0, \alpha_1, \alpha_2, \ldots, \alpha_k$ коэффициентов регрессии *(эмпирические коэффициенты регрессии);* e — оценка отклонения ε. Очевидно, для индивидуальных наблюдений имеем:

$$y_i = a_0 + a_1 x_i^{(1)} + a_2 x_i^{(2)} + \ldots + a_k x_i^{(k)} + e_i$$