

Стандартизированные коэффициенты регрессии

Стандартизированные коэффициенты регрессии

- Позволяют сравнивать коэффициенты между собой

$$t_y = \beta_1 \cdot t_{x_1} + \beta_2 \cdot t_{x_2} + \dots + \beta_p \cdot t_{x_p} + \varepsilon,$$

где $t_y, t_{x_1}, \dots, t_{x_p}$ – стандартизованные переменные: $t_y = \frac{y - \bar{y}}{\sigma_y}$, $t_{x_i} = \frac{x_i - \bar{x}_i}{\sigma_{x_i}}$,

для которых среднее значение равно нулю: $\bar{T}_y = \bar{T}_{x_i} = 0$,
а среднее квадратическое отклонение равно единице:

$$\sigma_{t_y} = \sigma_{t_x} = 1;$$

β – стандартизованные коэффициенты регрессии.

Стандартизированные коэффициенты регрессии

Стандартизированные коэффициенты регрессии показывают, на сколько сигм изменится в среднем результат, если соответствующий фактор x_i изменится на одну сигму при неизменном среднем уровне других факторов. В силу того, что все переменные заданы как центрированные и нормированные, стандартизированные коэффициенты регрессии β_i сравнимы между собой. Сравнивая их друг с другом, можно ранжировать факторы по силе их воздействия на результат. В этом основное достоинство стандартизированных коэффициентов регрессии в отличие от коэффициентов «чистой» регрессии, которые несравнимы между собой.

Пример

(Эконометрика, под ред. Елисейевой)

Пример. Пусть функция издержек производства y (тыс. руб.) характеризуется уравнением вида

$$y = 200 + 1,2 \cdot x_1 + 1,1 \cdot x_2 + \varepsilon,$$

где x_1 — основные производственные фонды (тыс. руб.);
 x_2 — численность занятых в производстве (чел.).

Анализируя его, мы видим, что при той же занятости дополнительный рост стоимости основных производственных фондов на 1 тыс. руб. влечет за собой увеличение затрат в среднем на 1,2 тыс. руб., а увеличение численности занятых на одного человека способствует при той же технической оснащенности предприятий росту затрат в среднем на 1,1 тыс. руб. Однако это не означает, что фактор x_1 оказывает более сильное влияние на издержки производства по сравнению с фактором x_2 . Такое сравнение возможно, если обратиться к уравнению регрессии в стандартизованном масштабе. Предположим, оно выглядит так:

Пример (Эконометрика, под ред. Елисейевой)

$$t_y = 0,5 \cdot t_{x_1} + 0,8 \cdot t_{x_2}.$$

Это означает, что с ростом фактора x_1 на одну сигму при неизменной численности занятых затраты на продукцию увеличиваются в среднем на 0,5 сигмы. Так как $\beta_1 < \beta_2$ ($0,5 < 0,8$), то можно заключить, что большее влияние оказывает на производство продукции фактор x_2 , а не x_1 , как кажется из уравнения регрессии в натуральном масштабе.

Стандартизированные коэффициенты в Excel

Имеются данные по 10 предприятиям о прибыли (Y, млн.руб.), выработке продукции на 1 работника (X1, единиц) и доле продукции, производимой на экспорт (X2,%).

Найдем оценки параметров регрессионного уравнения

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

№ п/п	Y	X1	X2
1	2	11	3
2	1	10	2
3	3	12	4
4	8	18	10
5	7	15	11
6	5	13	6
7	4	13	5
8	6	15	7
9	7	16	10
10	7	17	12

Стандартизированные коэффициенты в R-studio

1. Создадим набор стандартизированных переменных. Для этого воспользуемся функцией `scale`.
2. `h = swiss`
`h_st = mutate_each(h, "scale")`
`glimpse(h_st)`
3. Построим регрессию со стандартизированными коэффициентами.
4. `model = (data=h_st, Fertility~Catholic+Agriculture+Examination)`
`summary(model_st)`

```
Coefficients:
      Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.729e-16  1.123e-01  0.000 1.000000
Catholic     1.416e-01  1.385e-01  1.022 0.312347
Agriculture -1.730e-01  1.561e-01 -1.108 0.273930
Examination -6.836e-01  1.745e-01 -3.918 0.000315 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```