Лекция № 6
множественная регрессия и корреляция.
(продолжение)

- При исключении из регрессии p+1 фактора коэффициент детерминации должен уменьшаться, а остаточная дисперсия возрастать;

$$R^2_{p+1} \geq R^2_p \quad \text{и} \quad S^2_{p+1} \leq S^2_p.$$

Выбор формы уравнения регрессии.

- Как и в парной зависимости, возможны разные виды уравнений множественной регрессии: линейные и нелинейные.
- наиболее широко используются линейная и степенная функции .

- В линейной множественной регрессии

$$\hat{y}_x = a + b_1 \cdot x_1 + b_2 \cdot x_2 + ... + b_p \cdot x_p$$

параметры при переменной x называются **коэффициентами «чистой» регрессии.** Они характеризуют среднее изменение результата с изменением соответствующего фактора на единицу при неизменном значении других факторов, закрепленном на среднем уровне.

- Возможен и иной подход к построению уравнения множественной регрессии ,когда на основе матрицы коэффициентов корреляции строится уравнение регрессии в стандартизованном виде:

$$t_y = \beta_1 \cdot t_{x_1} + \beta_2 \cdot t_{x_2} + \square + b_p \cdot t_{x_p}$$

- Где $t_y, y_{x_1}, \ldots, t_{x_p}$ -стандартизованные переменные

$$t_{x_i} = \frac{x_i - \bar{x}_i}{\sigma_{x_i}}, \qquad t_y = \frac{y - \bar{y}}{\sigma_y}$$

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_i -стандартизованные коэффициенты регрессии.

- Стандартизованные коэффициенты регрессии показывают, на сколько % изменится в среднем результат, если соответствующий фактор x_i изменится на 1 % при **неизменном** среднем **уровне других факторов**.

- Стандартизованные коэффициенты регрессии β_i **сравнимы** между собой. Сравнивая их друг с другом, можно ранжировать факторы по силе их воздействия на результат.
- В этом основное стандартизованных коэффициентов регрессии в отличие от коэффициентов “чистой ” регрессии, которые несравнимы между собой.

- коэффициенты “чистой ” регрессии связанны со стандартизованными коэффициентами регрессии следующими формулами :

$$b_i = \beta_i \frac{\sigma_y}{\sigma_{x_i}}$$

- Это позволяет от уравнения регрессии в стандартизованном виде переходить к уравнению регрессии в естественном виде.

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

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

- 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.

- уравнение регрессии в стандартизованном виде выглядит так

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

- Вывод:

- 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 виде .

- Рассмотренный смысл стандартизованных коэффициентов регрессии позволяет их использовать при отсеве факторов –из модели исключаются факторы **с наименьшим** значением β_j .

Индекс множественной корреляции

- Оценивает тесноту совместного влияния факторов на результат

(0 ; 1)

Значение $R_{yx_1 x_2 \ldots x_i \ldots x_p}$ должно быть больше или равно максимальному парному индексу корреляции

ЧАСТНАЯ КОРРЕЛЯЦИЯ

- Частные коэффициенты (или индексы) корреляции характеризуют **тесноту связи** между результатом и соответствующим фактором при **неизменном уровне других факторов**, включенных в уравнение регрессии
- **(-1;1)**

- частные показатели корреляции широко используются при
- 1) решении проблемы отбора факторов.
- 2) ранжировании факторов, участвующих в множественной регрессии – для линейных связей (при нелинейной взаимосвязи исследуемых признаков эту функцию выполняют частные индексы детерминации).

- **Порядок** частного коэффициента корреляции определяется количеством факторов, влияние которых исключается.

- Например, $r_{yx_1 \cdot x_2}$ — коэффициент частной корреляции первого порядка.

$$r_{yx_2 \cdot x_1 x_3 x_4} -$$

- Коэффициенты частной корреляции более высоких порядков можно определить через коэффициенты частной корреляции более низких порядков по рекуррентной формуле

$$r_{yx_i \cdot x_1 x_2 \ldots x_p} = \frac{r_{yx_i \cdot x_1 x_2 \ldots x_{p-1}} - r_{yx_p \cdot x_1 x_2 \ldots x_{p-1}} \cdot r_{x_i x_p \cdot x_1 x_2 \ldots x_{p-1}}}{\sqrt{(1 - r^2_{yx_p \cdot x_1 x_2 \ldots x_{p-1}}) \cdot (1 - r^2_{x_i x_p \cdot x_1 x_2 \ldots x_{p-1}})}}.$$

При двух факторах и *i*= 1 данная формула примет вид:

$$r_{yx_1 \cdot x_2} = \frac{r_{yx_1} - r_{yx_2} \cdot r_{x_1 x_2}}{\sqrt{(1 - r_{yx_2}^2) \cdot (1 - r_{x_1 x_2}^2)}}.$$

- Соответственно при $i = 2$ и двух факторах частный коэффициент корреляции y с фактором x_2 можно определить по формуле