

**Лекция № 5**  
**множественная регрессия и**  
**корреляция.**

**Множественная регрессия широко  
используется в решении**

**проблем спроса,  
доходности акций,  
изучение функции издержек  
производства,  
в макроэкономических расчетах.**

- **Основная ЦЕЛЬ множественной регрессии – построить модель с большим числом факторов, определив при этом влияние каждого из них в отдельности, а также совокупное их воздействие на моделируемый показатель.**



# например

- Современная потребительская функция чаще всего рассматривается как модель вида

$$C = f(y, P, M, Z),$$

- $C$  – потребление;
- $y$  – доход;
- $P$  – цена, индекс стоимости жизни;
- $M$  – наличные деньги;
- $Z$  – ликвидные активы;

- **Построение уравнения множественной регрессии начинается с решения вопроса о спецификации модели.**

## Условия включения факторов при построении множественной регрессии.

- 1. Они должны быть количественно измеримы. Если необходимо включить модель качественный фактор, не имеющий количественного измерения, то ему нужно придать количественную определенность.

- например,
- в модели урожайности качество почвы задается в виде баллов;
- в модели стоимости объектов недвижимости учитывается место нахождения недвижимости: районы могут быть пронумерованы.

- **2. Факторы не должны быть интеркоррелированы.**



- Если между факторами существует высокая корреляция, то нельзя определить их изолированное влияние на результативный показатель и параметры уравнения регрессии оказываются *неинтерпретируемыми*.

- Так, в уравнении

$$y = a + b_1 x_1 + b_2 \cdot x_2$$

предполагается, что факторы  $x_1$  и  $x_2$  независимы друг от друга, т.е.  $r_{x_1 x_2} = 0$ . Тогда можно говорить, что параметр  $b_1$  измеряет силу влияния фактора  $x_1$  на результат  $y$  при неизменном значении фактора  $x_2$ . Если же  $r_{x_1 x_2} \neq 0$ , то с изменением фактора  $x_1$  фактор  $x_2$  не может оставаться неизменным. Отсюда и нельзя интерпретировать  $b_1$  и  $b_2$  как показатели раздельного влияния  $x_1$  и  $x_2$  на  $y$ .

$$x_1 \quad x_2$$

## Пример.

- Рассмотрим регрессию себестоимости: единицы продукции (руб.,  $y$ ) от заработной платы работника (руб.,  $x$ ) и производительности его труда (единиц в час,  $z$ ):

$$y = 22600 - 5 \cdot x - 10 \cdot z$$

- $r_{xz} = 0,95$

# **Отбор факторов при построении множественной регрессии.**

- отбор факторов обычно осуществляется в две стадии
- на первой подбираются факторы исходя из сущности проблемы;
- на второй – на основе матрицы показателей корреляции определяют существенность включения в уравнение регрессии каждого из факторов.

- Коэффициенты интеркорреляции – коэфф. корреляции между объясняющими переменными.
- Считается, что две переменные явно *коллинеарны*, т.е. находятся между собой в линейной зависимости, если  $r_{x_i x_j} > 0,7$ .

Поэтому одним из условий построения уравнения множественной регрессии является независимость действия факторов .

- Если факторы явно коллинеарны, то они дублируют друг друга и один из них рекомендуется исключить из регрессии.

- Предпочтение отдается не фактору, более тесно связанному с результатом, а тому фактору, который при достаточной тесной связи с результатом имеет наименьшую тесноту связи с другими факторами.



- Пусть, например, при изучении зависимости матрица парных коэффициентов корреляции оказалась следующей:

	$y$	$x$	$z$	$v$
$y$	1			
$x$	0,8	1		
$z$	0,7	0,8	1	
$v$	0,6	0,5	0,2	1

- Очевидно, что факторы  $x$  и  $z$  дублируют друг друга. В анализ целесообразно включить фактор  $z$ , а не  $x$ , хотя корреляция  $z$  с результатом  $y$  слабее, чем корреляция фактора  $x$  с  $y$  ( $r_{yz} < r_{yx}$ ), но зато слабее, чем межфакторная корреляция  $r_{zv} < r_{xv}$ . Поэтому в данном случае в уравнении множественной регрессии включаются факторы  $z, v$ .

# пример

	<i>y</i>	<i>x</i>	<i>z</i>	<i>v</i>
<i>y</i>	1			
<i>x</i>	0,3	1		
<i>z</i>	0,7	0,75	1	
<i>v</i>	0,6	0,5	0,8	1

- По величине парных коэффициентов корреляции обнаруживается лишь явная коллинеарность факторов. Наибольшие трудности возникают при наличии *мультиколлинеарности* факторов, когда более чем два фактора связаны между собой линейной зависимостью.

- Для оценки мультиколлинеарности факторов может использоваться определитель матрицы парных коэффициентов корреляции между факторами.
- Если бы факторы не коррелировали между собой, то матрица парных коэффициентов корреляции была бы единичной матрицей т.е.

$$\text{Det}|R| = \begin{vmatrix} r_{x_1x_1} & r_{x_2x_1} & r_{x_3x_1} \\ r_{x_1x_2} & r_{x_2x_2} & r_{x_3x_2} \\ r_{x_1x_3} & r_{x_2x_3} & r_{x_3x_3} \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = 1,$$

- Если же, наоборот, между факторами существует полная линейная зависимость и все коэффициенты корреляции равны единице, то определитель такой матрицы равен нулю:

$$\text{Det}|R| = \begin{vmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{vmatrix} = 0.$$

- Таким образом, чем ближе к нулю определитель матрицы межфакторной корреляции, тем сильнее мультиколлинеарность факторов и ненадежнее результаты множественной регрессии.



- Через коэффициенты множественной детерминации можно найти переменные, ответственные за мультиколлинеарность факторов.

- Сравнивая между собой коэффициенты множественной детерминации факторов

$$R^2_{x_1|x_2, x_3 \dots x_p} ; R^2_{x_2|x_1 x_3 \dots x_p} ; \square$$

- оставляем в уравнении факторы с минимальной величиной коэффициента множественной детерминации.

- При дополнительном включении в регрессию  $p+1$  фактора коэффициент детерминации должен возрасти, а остаточная дисперсия уменьшиться;

$$R_{p+1}^2 \geq R_p^2 \quad \text{и} \quad S_{p+1}^2 \leq S_p^2.$$

- Если же этого не происходит и данные показатели практически мало отличаются друг от друга, то включаемый в анализ фактор не улучшает модель и практически является лишним фактором.

- Так, если для регрессии, включающих пять факторов, коэффициент детерминации составил 0,857 и включение шестого фактора дало коэффициент детерминации 0,858, то вряд ли целесообразно дополнительно включать в модель этот фактор.