

Множественный Регрессионный Анализ

(МРА)

В МРА изменения фактора Y
определяются изменениями
нескольких факторов X_2, X_3, \dots, X_k .

$$Y = \beta_1 + \beta_2 * X_2 + \beta_3 * X_3 + \dots + \beta_k * X_k + u \quad (*)$$

(*) - модель множественной линейной регрессии (МЛР).

$\beta_1, \beta_2, \dots, \beta_k$ - подлежащие оценке параметры.

u - случайный член, случайная ошибка.

Уравнение линейного МРА
предполагает, что изменения
фактора Y складываются из:

- систематических, неслучайных изменений, строго функционально зависящих от изменений независимых факторов X_2, \dots, X_k ;
- случайных изменений, определяемых распределением случайного фактора u .

Модель (*). \rightarrow

Выборка реальных данных:

$(Y_i, X_{2i}, \dots, X_{ki}), i=1, \dots, n. \rightarrow$

МНК. \rightarrow

Оценка модели (выборочное уравнение):

$$\hat{Y} = b_1 + b_2 * X_2 + \dots + b_k * X_k. (**)$$

b_j - оценка β_j , $j=1,\dots,k$.

Чтобы оценить коэффициенты регрессии по МНК было возможно, должно быть:

- 1) $n \geq k$ (размер выборки не меньше числа параметров модели).
- 2) Между независимыми переменными X_2, X_3, \dots, X_k нет строгой линейной зависимости.

Интерпретация Коэффициентов Множественной Регрессии

В уравнении

$$Y = \beta_1 + \beta_2 * X_2 + \dots + \beta_k * X_k + u$$

любой коэффициент наклона β_j при X_j
($j = 2, \dots, n$) дает оценку влияния X_j на
 Y в случае **неизменности влияния**
на Y всех остальных X ,
включенных в уравнение.

ПРИМЕРЫ.

- 1) Y – затраты на питание в млрд. \$,
 X – располагаемый личный доход в млрд. \$,
 p – индекс относительных цен в пунктах.

Данные для США за 1959-1983 гг. $n = 25$.

$$\hat{Y} = 116,7 + 0,112 * X - 0,739 * p; \quad R^2 = 0,99$$

(9,6) (0,003) (0,114)

Интерпретация коэффициентов наклона:

- при каждом увеличении располагаемого личного дохода на 1 млрд. \$ (**при сохранении постоянных цен**) расходы на питание увеличатся в среднем на 112 млн.\$.

$$\hat{Y} = 116,7 + 0,112 * X - 0,739 * p; \quad R^2 = 0,99$$

(9,6) (0,003) (0,114)

- на каждую единицу увеличения индекса цен (**при сохранении постоянных доходов**) эти расходы уменьшатся в среднем на 739 млн. \$.

2) Y_t – потребление говядины на душу населения в фунтах в год t ;

p_t – цена 1 фунта говядины в центах в год t ;

X_t – среднедушевой доход в тысячах \$ в год t .

Данные для США за 1960-1987 гг. $n = 28$.

$$\hat{Y}_t = 37,53 - 0,88 \cdot P_t + 11,9 \cdot X_t$$

(15,2) (0,16) (1,76)

Интерпретация коэффициентов наклона:

- при каждом увеличении среднедушевого дохода на 1000\$ (**при сохранении постоянных цен на говядину**) спрос на говядину увеличивается в среднем на 11,9 фунтов на душу населения;
- на каждый добавочный цент к цене говядины (**при сохранении постоянной величины среднедушевого дохода**) спрос на говядину уменьшается в среднем на 0,88 фунтов на душу населения.

Замечание.

Факторы, НЕ включенные в уравнение регрессии, НЕ «сохраняются постоянными» при оценке влияния на Y включенных в уравнение факторов.

Например, в примере 2 спрос на говядину увеличивается в среднем на 11,9 фунтов при увеличении среднедушевого дохода на 1000\$ при сохранении постоянных цен на говядину, но не при сохранении постоянных цен на заменяющие говядину товары.

Аналогично при интерпретации коэффициента 0,88.

Интерпретация свободного члена - все то же, что и в парной линейной регрессии.