



спецификация



Множественная регрессия

L/O/G/O

www.themegallery.com



Множественная регрессия

- Уравнение множественной регрессии в натуральном масштабе:

$$\hat{y} = a + b_1x_1 + b_2x_2 + \dots + b_px_p$$

- Где Y – зависимая переменная;
- x_1, x_2, \dots, x_p – независимые переменные;
 a и b_1, b_2, \dots, b_p – параметры (коэффициенты) модели

Напоминание:

- Y, x_1, x_2, \dots, x_p – изучаемые показатели или явления;
- a, b_1, b_2, \dots, b_p – числа, характеризующие связь между y и x , рассчитываются по формулам или в столбце «Коэффициенты» пакета анализа «Регрессия» в Excel



Множественная регрессия

- Регрессионная модель в стандартизованном масштабе :

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

– Где $t_y, t_{x_1}, t_{x_2}, \dots, t_{x_p}$ – стандартизованные переменные; для которых среднее значение равно нулю, а среднее квадратическое отклонение равно единице:

$$t_y = \frac{y - \bar{y}}{\sigma_y}; \quad t_{x_j} = \frac{x_j - \bar{x}_j}{\sigma_{x_j}}, \quad j = \overline{1, n}$$

β_j – стандартизованные коэффициенты регрессии, или β – коэффициенты

➔ Расчет:

$$\begin{cases} \beta_1 + \beta_2 r_{x_2 x_1} + \beta_3 r_{x_3 x_1} + \beta_p r_{x_p x_1} = r_{yx_1} \\ \beta_1 r_{x_1 x_2} + \beta_2 + \beta_3 r_{x_3 x_2} + \beta_p r_{x_p x_2} = r_{yx_2} \\ \boxtimes \quad \quad \quad \boxtimes \quad \quad \quad \boxtimes \quad \quad \quad \boxtimes \quad \quad \quad \boxtimes \\ \beta_1 r_{x_1 x_p} + \beta_2 r_{x_2 x_p} + \beta_3 r_{x_3 x_p} + \beta_p = r_{yx_p} \end{cases}$$

Частный случай: наличие 2х факторов x_1 и x_2

$$\begin{cases} \beta_1 + \beta_2 r_{x_1 x_2} = r_{yx_1} \\ \beta_1 r_{x_1 x_2} + \beta_2 = r_{yx_2} \end{cases}$$

$r_{x_1 x_2}$ r_{yx_1} r_{yx_2} - коэффициенты корреляции

Взаимосвязь уравнений в стандартизованном и натуральном масштабах:

В парной зависимости стандартизованный коэффициент регрессии есть линейный коэффициент корреляции r

$$a = \bar{y} - b_1 \bar{x}_1 - b_2 \bar{x}_2 - \dots - b_p \bar{x}_p$$

$$\bar{\varepsilon}_j = b_j \frac{\bar{x}_j}{\bar{y}}$$

$$b_j = \beta_j \frac{\sigma_y}{\sigma_{x_j}}$$



Интерпретация

коэффициентов:

- В модели множественной регрессии в натуральном и стандартизированном масштабах, а также по эластичности:

$$b_1, b_2 \dots b_p$$

показывают на сколько **единиц** изменится y при изменении x_i на 1 **единицу**, при неизменности прочих факторов

$$\beta_1, \beta_2 \dots \beta_p$$

на сколько значений **с.к.о.** изменится в среднем y , если соответствующий фактор x_j изменится на одну **с.к.о.** при неизменном среднем уровне других факторов

$$\varepsilon_1, \varepsilon_2 \dots \varepsilon_p$$

Эластичность показывает на сколько % в среднем изменится y при изменении x_i на 1%



Частная корреляция

Коэффициенты частной корреляции

Частные коэффициенты корреляции характеризуют тесноту связи между результатом и соответствующим фактором при устранении влияния других факторов, включенных в уравнение регрессии.

Задача состоит в том, чтобы:

найти «чистую» корреляцию между двумя переменными, исключив (линейное) влияние других факторов.

Связь с коэффициентом детерминации R²

$$r_{yx_1 \cdot x_2}^2 = \frac{R^2 - r_{yx_2}^2}{1 - r_{yx_2}^2} \quad \text{где } r_{yx_2} \text{ - обычный коэффициент корреляции}$$

В коэффициенте частной корреляции через точку указываются факторы, влияние которых устраняется



Расчет по рекуррентной формуле:

Влияние парной корреляции на коэффициент детерминации

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

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

$$R^2 = \beta_1 r_{yx_1} + \beta_2 r_{yx_2}$$

$$r_{yx_1 \cdot x_2}^2 = \frac{R^2 - r_{yx_2}^2}{1 - r_{yx_2}^2}$$

$$r_{yx_2 \cdot x_1}^2 = \frac{R^2 - r_{yx_1}^2}{1 - r_{yx_1}^2}$$



Тест на обоснованность исключения новых k факторов из модели

Гипотезы:

$$H_0 : R_1^2 = R_2^2 \quad H_1 : R_1^2 > R_2^2$$

Наблюдаемое и критическое значение

$$F_{\text{набл}} = \frac{R_1^2 - R_2^2}{1 - R_1^2} \cdot \frac{n - p - 1}{k} \quad F_{\text{кр}} = F(\alpha; k; n - p - 1)$$

Вывод:

$F_{\text{набл}} < F_{\text{кр}} = H_0$ (исключение **обоснованно**)

$F_{\text{набл}} > F_{\text{кр}}$ то H_1 (исключение **не обоснованно**)

R_1 – коэффициент детерминации до исключения;
 R_2 – коэффициент детерминации после исключения;
 n – объем выборки;
 p – количество независимых факторов до исключения;
 k – количество исключаемых факторов



Тест на обоснованность включения новых k факторов в модель

Гипотезы:

$$H_0 : R_2^2 = R_1^2 \quad H_1 : R_2^2 > R_1^2$$

Наблюдаемое и критическое значение

$$F_{\text{набл}} = \frac{R_2^2 - R_1^2}{1 - R_2^2} \cdot \frac{n - p - 1}{k} \quad F_{\text{кр}} = F(\alpha; k; n - p - 1)$$

Вывод:

$F_{\text{набл}} < F_{\text{кр}} = H_0$ (включение **не обоснованно**)

$F_{\text{набл}} > F_{\text{кр}}$ то H_1 (включение **обоснованно**)

R_1 – коэффициент детерминации до включения;
 R_2 – коэффициент детерминации после включения;
 n – объем выборки;
 p – количество независимых факторов после включения;
 k – количество включаемых факторов



Тест Чоу на наличие структурных сдвигов:

Гипотезы:

$$H_0 : s_0 = s_1 + s_2; \quad H_1 : s_0 > s_1 + s_2$$

Наблюдаемое и критическое значение

$$F_{\text{набл}} = \frac{s_0 - (s_1 + s_2)}{s_1 + s_2} \cdot \frac{n - 2p - 2}{p + 1} \quad F_{\text{кр}} = F(\alpha; p + 1; n - 2p - 2)$$

Вывод:

$F_{\text{набл}} < F_{\text{кр}} = H_0$ (структурных сдвигов **нет**)
 $F_{\text{набл}} > F_{\text{кр}}$ то H_1 (структурные сдвиги **есть**)

$$s_0 = \sum_{i=1}^n e_i^2$$

$$s_1 = \sum_{i=1}^k e_i^2$$

$$s_2 = \sum_{i=n-k+1}^n e_i^2$$

s_0 – сумма квадратов остатков всей выборки;

s_1 – сумма квадратов остатков первой подвыборки;

s_2 – сумма квадратов остатков второй подвыборки;

n – объем выборки;

p – количество независимых факторов в модели



Тест Спирмена на наличие гетероскедастичности:

Гипотезы:

$$H_0 : r_{x,e} = 0$$

$$H_1 : r_{x,e} \neq 0$$



Наблюдаемое и критическое значение

$$t_{\text{набл}} = \frac{r_{x,e} \sqrt{n-2}}{\sqrt{1-r_{x,e}^2}}$$

$$t_{\text{крит}}(\alpha; n-2)$$

Вывод:

$$t_{\text{набл}} < t_{\text{кр}} = H_0 \text{ (гомоскедастичность)}$$

$$t_{\text{набл}} > t_{\text{кр}} \text{ то } H_1 \text{ (гетероскедастичность)}$$

$$r_{x,e} = 1 - 6 \cdot \frac{\sum d_i^2}{n(n^2 - 1)}$$

$r_{x,e}$ – коэффициент ранговой корреляции Спирмена;
 d – разность рангов x_i и модулей остатков $|e_i|$;



Тест Голдфелда – Квандта на наличие гетероскедастичности :

Гипотезы:

$$H_0 : s_1 = s_3; \quad H_1 : s_3 > s_1$$



Наблюдаемое и критическое значение

$$F_{\text{набл}} = \frac{s_3}{s_1} \quad F_{\text{кр}} = F(\alpha; k - p - 1; k - p - 1)$$

Вывод:

$F_{\text{набл}} < F_{\text{кр}} = H_0$ (гомоскедастичность)
 $F_{\text{набл}} > F_{\text{кр}}$ то H_1 (гетероскедастичность)

$$s_1 = \sum_{i=1}^k e_i^2$$

$$s_3 = \sum_{i=n-2k+1}^n e_i^2$$

s_1 – сумма квадратов остатков первой подвыборки;

s_2 – сумма квадратов остатков второй подвыборки;

k – объем подвыборки;

p – количество независимых факторов в модели

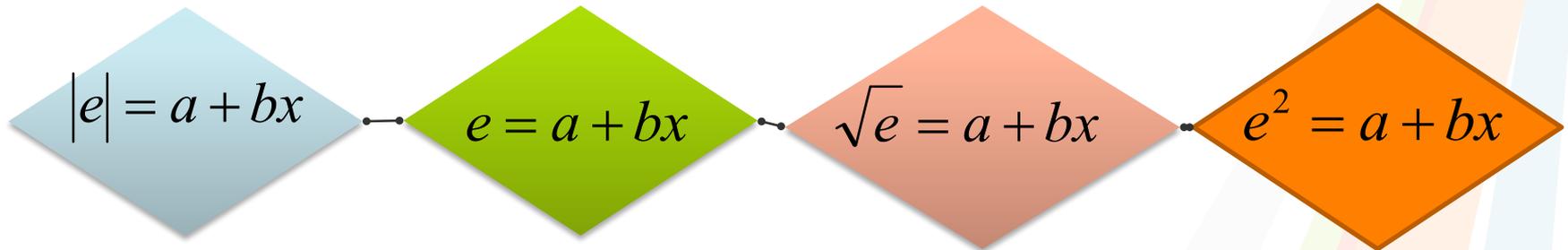


Тест Глейзера на гетероскедастичность

Тест основан на проверке статистической значимости коэффициентов регрессии моделей зависимости остатков от x

$H_0: b=0$ $p\text{-значение} > \alpha$

$H_1: b \neq 0$ $p\text{-значение} < \alpha$



- Если хоть в одной из представленных моделей коэффициент регрессии статистически значим ($p\text{-значение} < \alpha$), то существует гетероскедастичность



Корректировка гетероскедастичности Метод взвешенных наименьших квадратов

• Предпосылка:

Пересчитываются коэффициенты модели линейной регрессии если известны дисперсии остатков для каждого наблюдения σ_i^2

Ввод новых
переменных

$$y_i^* = \frac{y_i}{\sigma_i}; \quad x_i^* = \frac{x_i}{\sigma_i}; \quad z_i = \frac{1}{\sigma_i}$$

Оценка
параметров
регрессии

$$y^* = az + bx^*$$

*свободный член равен нулю (константа-ноль)

Возврат к
исходной
модели

$$y = a + bx$$

*модель гомоскедастична



Корректировка гетероскедастичности Обобщенный метод наименьших квадратов

• Предпосылка:

Пересчитываются коэффициенты модели линейной регрессии, дисперсии остатков для каждого наблюдения не известны

Ввод новых
переменных

$$y_i^* = \frac{y_i}{\sqrt{x_i}} \quad z_i^* = \frac{1}{\sqrt{x_i}} \quad x_i^* = \frac{x_i}{\sqrt{x_i}}$$

Оценка
параметров
регрессии

$$y^* = az^* + bx^*$$

*свободный член равен нулю (константа-ноль)

Возврат к
исходной
модели

$$y = a + bx$$

*модель гомоскедастична



Тест Дарбина – Уотсона на наличие автокорреляции :

$$DW = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2}$$

положительная АКЛЛ

отрицательная АКЛЛ

Зона неопр.

Зона неопр.

НЕТ АКЛЛ

0

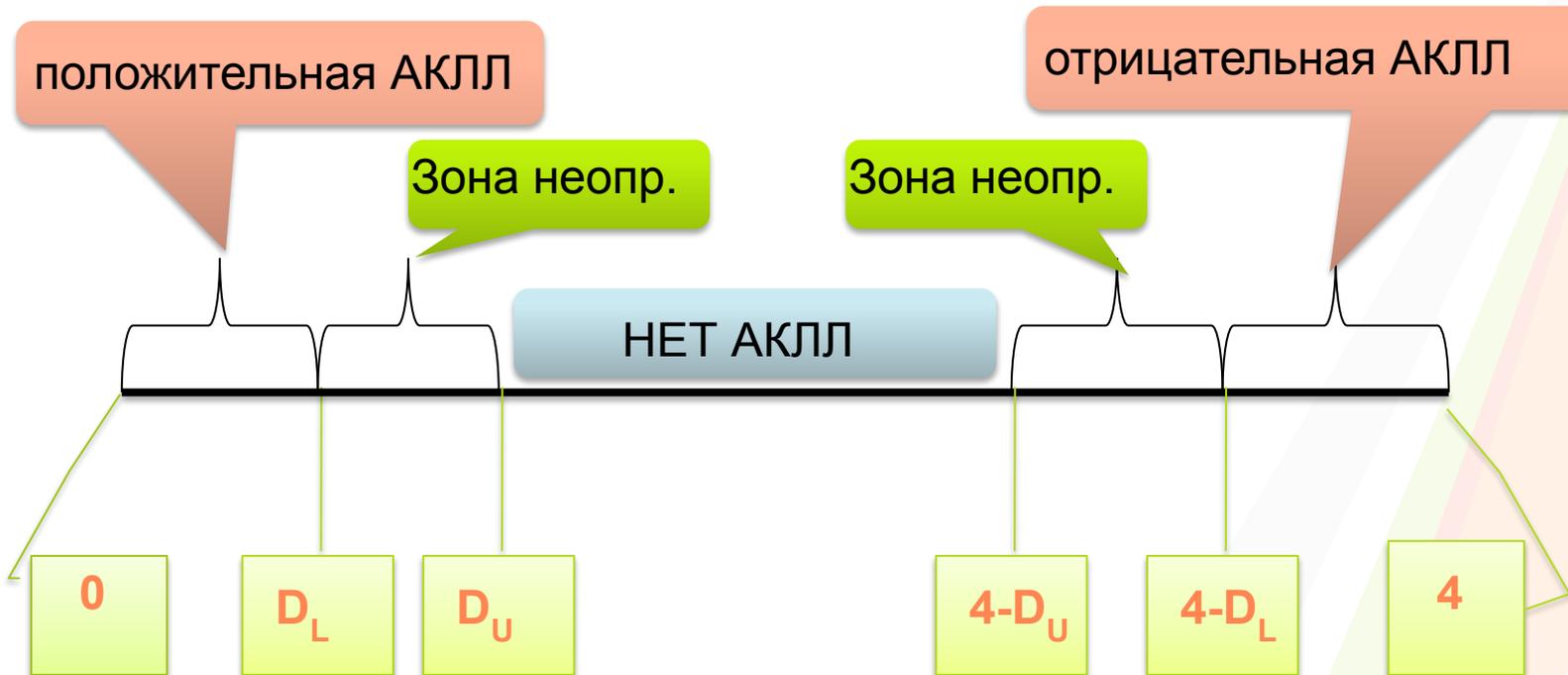
D_L

D_U

$4 - D_U$

$4 - D_L$

4





Корректировка автокорреляции

Авторегрессионная схема первого порядка $AR(1)$

• Предпосылка:

Применяется для пересчета коэффициентов модели, если автокорреляция вызвана внутренними свойствами ряда $\{e_t\}$

Определение ρ
и ввод новых
переменных

$$e_t = \rho e_{t-1}$$
$$y_t^* = y_t - \rho y_{t-1}; \quad x_t^* = x_t - \rho x_{t-1}$$

Оценка
параметров
регрессии

$$y_t^* = a^* + bx_t^*$$

Возврат к
исходной
модели

$$a = \frac{a^*}{1 - \rho} \quad y = a + bx$$