Парная регрессия

Термин «регрессия»

- Означает «отступление, возврат к чему-либо»
- Положительная зависимость между ростом отца и ростом его взрослого сына (у высоких отцов высокие сыновья и наоборот);
- Тенденция: у высоких отцов сыновья ниже отцов, хотя сами и высокие, а у низких отцов сыновья выше отцов, хотя сами низкие ростом – возврат к среднему состоянию.

$$y = \hat{f}(x) \quad (1)$$

$$y = \hat{f}(x_1, x_2, \ldots, x_k) \quad (2)$$

$$y_j = \hat{y}_{x_j} + \varepsilon_j$$

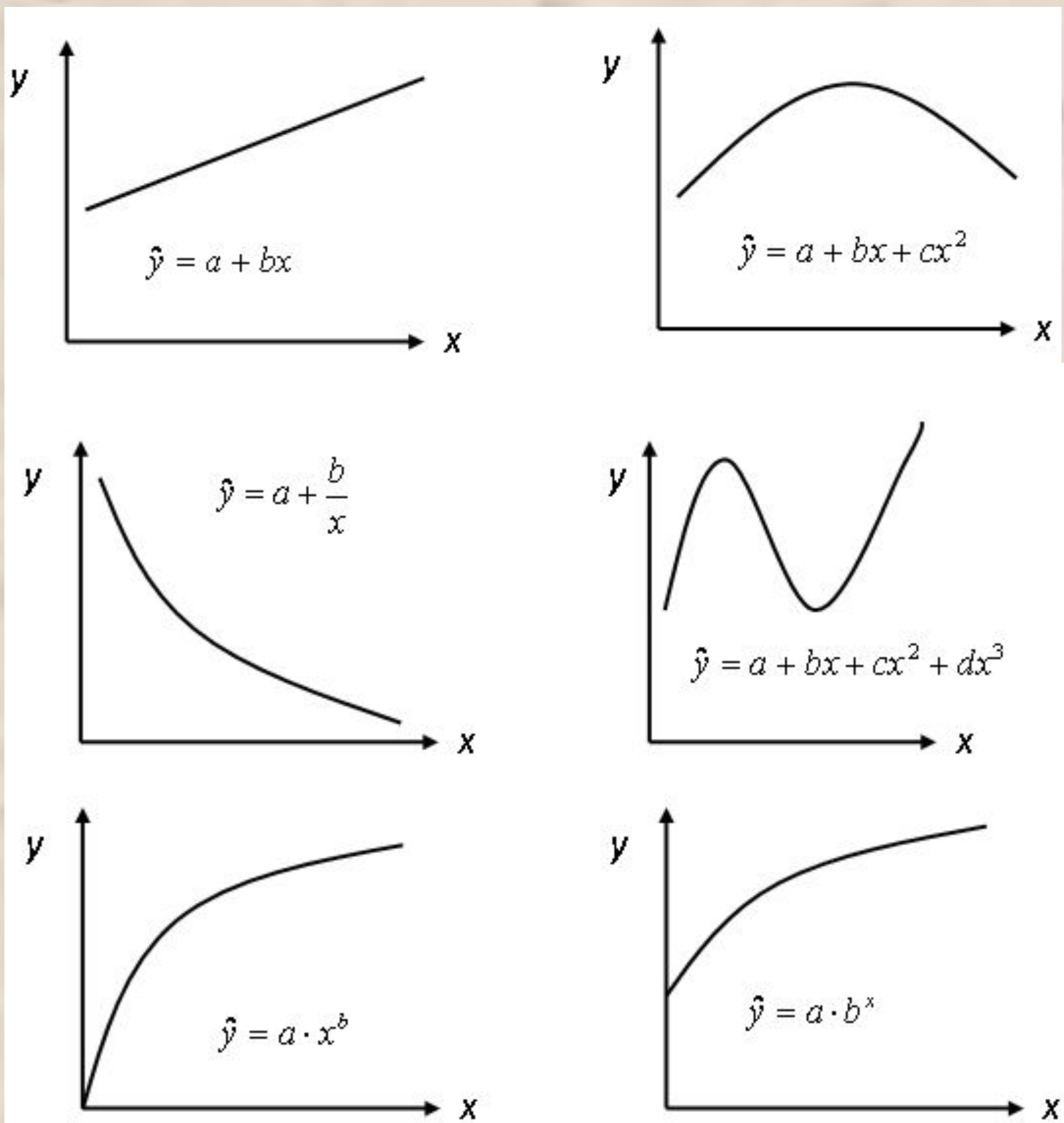
Источники ошибок эконометрической модели:

- спецификация
- выборочный характер исходных данных
- ошибки измерения

Методы выбора математической функции *f(x)*

- графический
- аналитический
- экспериментальный

Графический метод



Другие типы кривых:

$$\hat{y} = \frac{1}{a + bx + cx^2}$$

$$\hat{y} = \frac{a}{1 + be^{-cx}}$$

$$\hat{y} = \frac{1}{a + b \cdot x}$$

$$\hat{y} = a + b \cdot \lg x$$

$$\lg \hat{y} = a + bx + cx^2$$

$$\hat{y} = a + bx + \frac{c}{x}$$

Аналитический метод

Пример: зависимость потребления электроэнергии *y* от объема выпускаемой продукции *x*:

$$\hat{y} = a + bx$$

Удельный расход электроэнергии на единицу продукции *z=y/x*:

$$\hat{z} = b + \frac{a}{x}$$

Экспериментальный метод

$$D_{ост} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2$$

Подгонка функции f(x) под результаты наблюдений

$$f\left(x, \beta_1, \beta_2, ..., \beta_k\right)$$

$$f\left(x, a, b\right) = a + bx$$

Меры отклонения функции от набора наблюдений

1) $$Q=\sum_{i=1}^{n}\left(y_i-f\left(x_i,\beta_1,\dots,\beta_p\right)\right)^2$$

2) $$Q=\sum_{i=1}^{n}\left|y_i-f\left(x_i,\beta_1,\dots,\beta_p\right)\right|$$

3) $$Q=\sum_{i=1}^{n}g\left(y_i-f\left(x_i,\beta_1,\dots,\beta_p\right)\right)$$

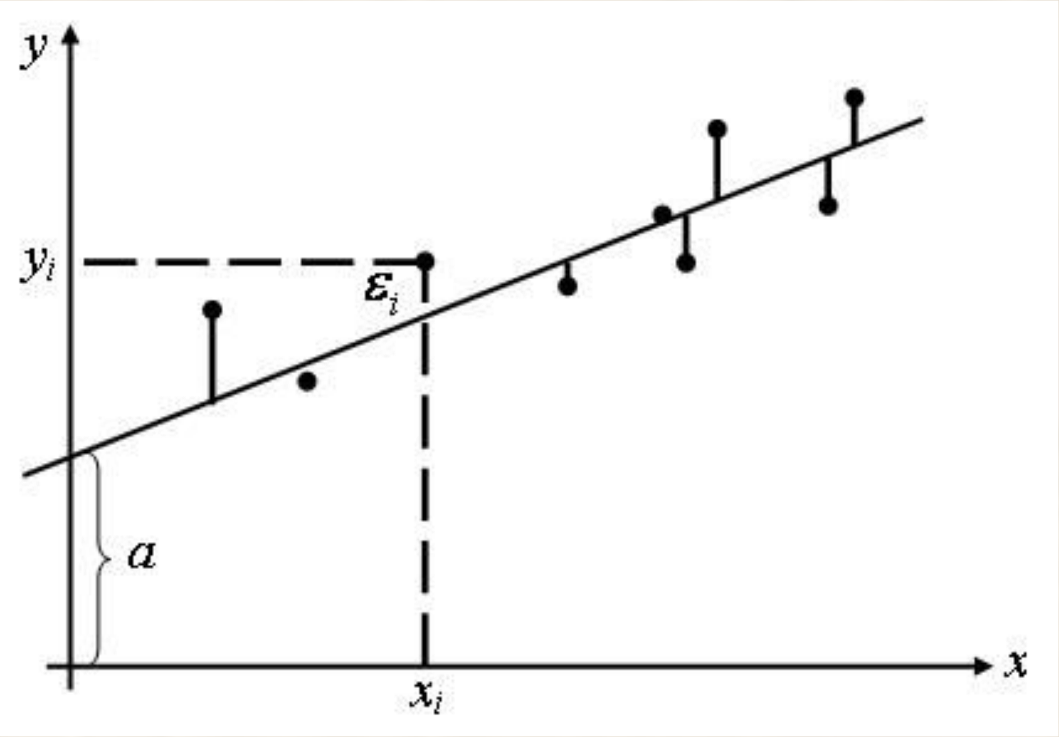
Функция Хубера: $$g(x)=\begin{cases}x^2, & |x|<c,\\ 2cx-c^2, & x\ge c,\\ -2cx-c^2, & x\le -c.\end{cases}$$

Оценка параметров линейной регрессии:

$$\hat{y}_x = a + bx \quad (3)$$

$$y = a + bx + \varepsilon$$

$$\sum_i \left(y_i - \hat{y}_{x_i}\right)^2 \to \min, \qquad \left(\sum_i \varepsilon_i^{\,2} \to \min\right) \quad (4)$$



$$\Sigma\varepsilon_i^2 = S = \Sigma(y - \hat{y}_x)^2 = \Sigma(y - a - bx)^2$$

$$\begin{cases} \dfrac{\partial s}{\partial a} = -2\sum y + 2na + 2b\sum x = 0; \\ \dfrac{\partial s}{\partial b} = -2\sum y \cdot x + 2a\sum x + 2b\sum x^2 = 0 \end{cases}$$

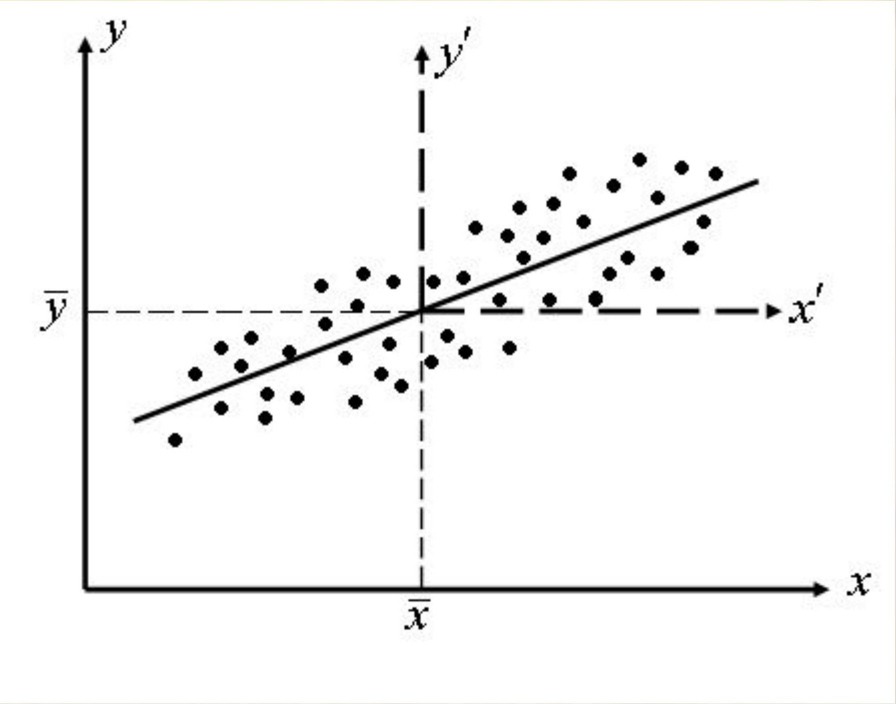
$$\begin{cases} n \cdot a + b\sum x = \sum y, \\ a\sum x + b\sum x^2 = \sum yx \end{cases}$$

$$b=\frac{n\sum yx-\left(\sum y\right)\left(\sum x\right)}{n\sum x^2-\left(\sum x\right)^2} \quad (7)$$

$$a=\frac{1}{n}\sum y-\frac{b}{n}\sum x \quad (8)$$

$$b=\frac{\overline{yx}-\bar{y}\cdot\bar{x}}{\overline{x^2}-\bar{x}^2}=\frac{cov(x,y)}{\sigma_x^2} \quad (9)$$

$$y' = y - \bar{y}, \quad x' = x - \bar{x} \quad \Rightarrow \quad y' = bx' \quad (10)$$



$$r = b\frac{\sigma_x}{\sigma_y} \quad (11) \qquad -1 \le r \le 1$$

Предпосылки МНК

1^0. Математическое ожидание случайного отклонения равно нулю для всех наблюдений: $M\left(\varepsilon_i\right)=0, \forall i$

.

2^0. Дисперсия случайных отклонений постоянна:

$$D(\varepsilon_i)=D(\varepsilon_j)=\sigma^2, \forall i, j$$

3^0. Случайные отклонения ε_i и ε_j являются независимыми друг от друга для $i \neq j$

$$\operatorname{cov}(\varepsilon_i, \varepsilon_j)=\begin{cases} 0, & i \neq j \\ \sigma^2, & i=j. \end{cases}$$

4^0. Случайное отклонение должно быть независимо от объясняющих переменных.

Проверка адекватности

$$\sum\left(y-\bar{y}\right)^2=\sum\left(\hat{y}-\bar{y}\right)^2+\sum\left(y-\hat{y}\right)^2 \qquad (13)$$

TSS RSS ESS

$$\sum\left(\hat{y}-\bar{y}\right)^2\to 0 \quad\Rightarrow\quad r^2\to 0$$

$$\sum\left(y-\hat{y}\right)^2\to 0 \quad\Rightarrow\quad r^2\to 1$$

$$df_{общ} = n - 1$$

$$df_{фактор} = 1$$

$$df_{остат} = n - 2$$

$$df_{\text{общ.}} = df_{\text{факт.}} + df_{\text{остат.}} \quad (14)$$

$$D_{общ.} = \frac{\sum\left(y - \bar{y}\right)^2}{n - 1} \quad (15)$$

$$D_{факт.} = \frac{\sum\left(\hat{y}_x - \bar{y}\right)^2}{1} \quad (16)$$

$$D_{ост.} = \frac{\sum\left(y - \hat{y}_x\right)^2}{n - 2} \quad (17)$$

$$H_0 : D_{факт.} = D_{ост.}$$

$$F = \frac{D_{факт.}}{D_{ост.}} \qquad (18)$$

$$H_1 : D_{факт.} > D_{ост.}$$

$$F < F_{табл}(\alpha;1;n-2) \Rightarrow H_0 \quad не \quad отклоняется$$

$$F \geq F_{табл}(\alpha;1;n-2) \Rightarrow H_0 \quad отклоняется$$

$$в \quad пользу \quad H_1 \quad (модель\ адекватна)$$

$$F = \frac{r^2}{1 - r^2} \cdot (n - 2) \quad (19)$$

$$m_b = \sqrt{\frac{\sum (y - \hat{y}_x)^2 / (n - 2)}{\sum (x - \bar{x})^2}} = \sqrt{\frac{S^2}{\sum (x - \bar{x})^2}} \quad (20)$$

$$t_b = \frac{b}{m_b} \quad (21)$$

$$H_0 : b = 0 \qquad H_1 : b \neq 0$$

$$\left|t_b\right| \geq t_{табл}(\alpha; n-2), \quad H_0 \text{ отклоняется,}$$

параметр статистически значим

$$\left|t_b\right| < t_{табл}(\alpha; n-2), \quad H_0 \textit{ не отклоняется}$$

$$t_b^2 = F \qquad (22)$$

$$b - m_b \cdot t_{табл} \leq b^* \leq b + m_b \cdot t_{табл} \qquad (23)$$

$$t_a = \frac{a}{m_a} \quad (25)$$

$$t_r = \frac{r}{\sqrt{1-r^2}}\sqrt{n-2} \qquad (26)$$

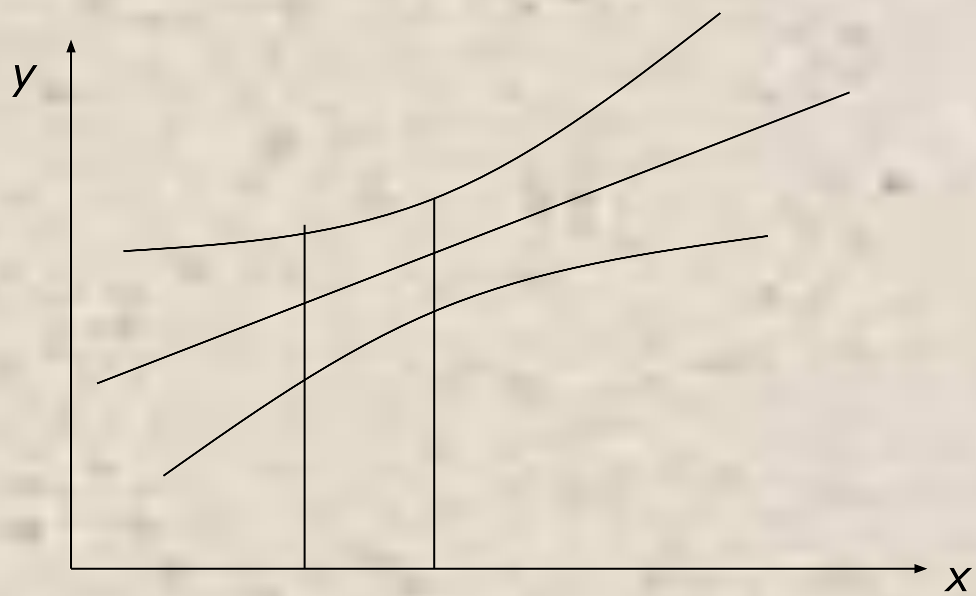
$$t_r^2 = F$$

$$z = \frac{1}{2}\ln\frac{1+r}{1-r} \qquad (27) \qquad m_z = \frac{1}{\sqrt{n-3}} \qquad (28)$$

Прогнозирование по парной линейной модели

$$\hat{y} - m_{\hat{y}} \cdot t_{табл} \le y^* \le \hat{y} + m_{\hat{y}} \cdot t_{табл} \qquad (30)$$

$$m_{y_{i(x_p)}} = S\sqrt{\frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum (x - \bar{x})^2}} \qquad (31)$$



$$m_{y_{i(x_p)}} = S\sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum (x - \bar{x})^2}} \qquad (33)$$

Нелинейные модели регрессии

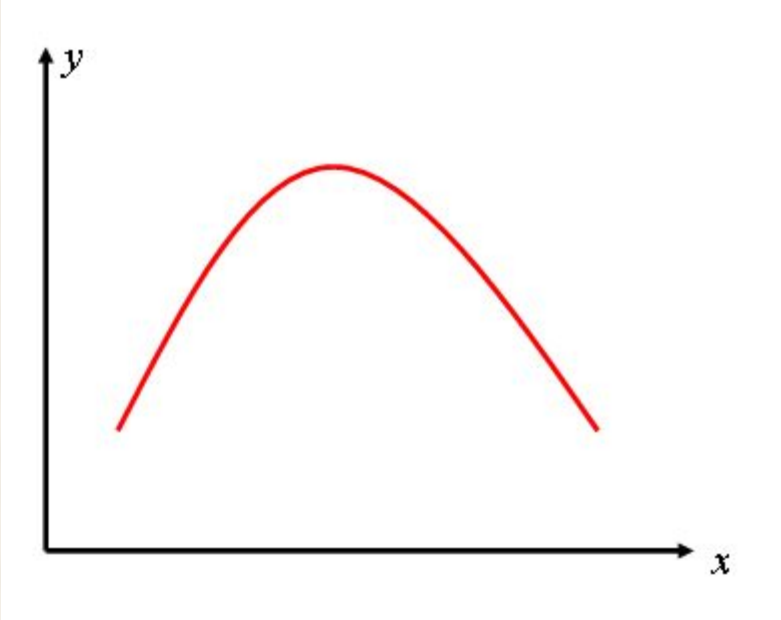
$$y = a_0 + a_1 x + a_2 x^2 + \ldots + a_k x^k + \varepsilon \quad (34)$$

$$y = a + bx + cx^2 + \varepsilon \quad (35)$$

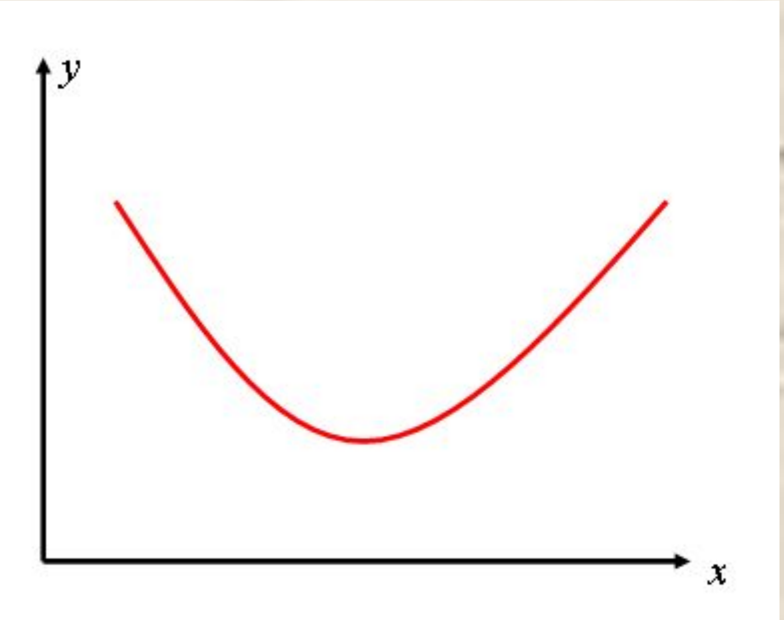
$$\begin{cases} a \cdot n & + b\sum x & + c\sum x^2 & = \sum y, \\ a\sum x & + b\sum x^2 & + c\sum x^3 & = \sum yx, \\ a\sum x^2 & + b\sum x^3 & + c\sum x^4 & = \sum yx^2. \end{cases} \quad (36)$$

$$x = -\frac{b}{2a}$$

$b > 0, \quad c < 0:$



$b < 0, \quad c > 0:$



$$y = a + \frac{b}{x} + \varepsilon \qquad (37)$$

$$z = 1/x \quad \Rightarrow \quad \hat{y} = a + b \cdot z \qquad (38)$$

$$y = a + b \cdot \ln x + \varepsilon \qquad (39)$$

$$z = \ln x \quad \Rightarrow \quad \hat{y} = a + b \cdot z$$

$$y = e^{a+bx} \cdot \varepsilon \quad (40)$$

$$\Rightarrow \ln y = a + bx + \ln \varepsilon \quad (43) \Rightarrow$$

$$Y = a + bx + E \quad (44)$$

$$y = a \cdot e^{bx} \cdot \varepsilon \quad (41)$$

$$\Rightarrow \ln y = \ln a + bx + \ln \varepsilon \quad (45) \Rightarrow$$

$$Y = A + bx + E \quad (46)$$

$$y = a \cdot b^{x} \cdot \varepsilon \quad (42)$$

$$\Rightarrow \ln y = \ln a + x \ln b + \ln \varepsilon \Rightarrow$$

$$Y = A + Bx + E \quad (47)$$

$$y = a \cdot x^b \cdot \varepsilon \quad (48)$$

$$\Rightarrow \ln y = \ln a + b \ln x + \ln \varepsilon \Rightarrow$$

$$Y = A + bX + E \quad (49)$$

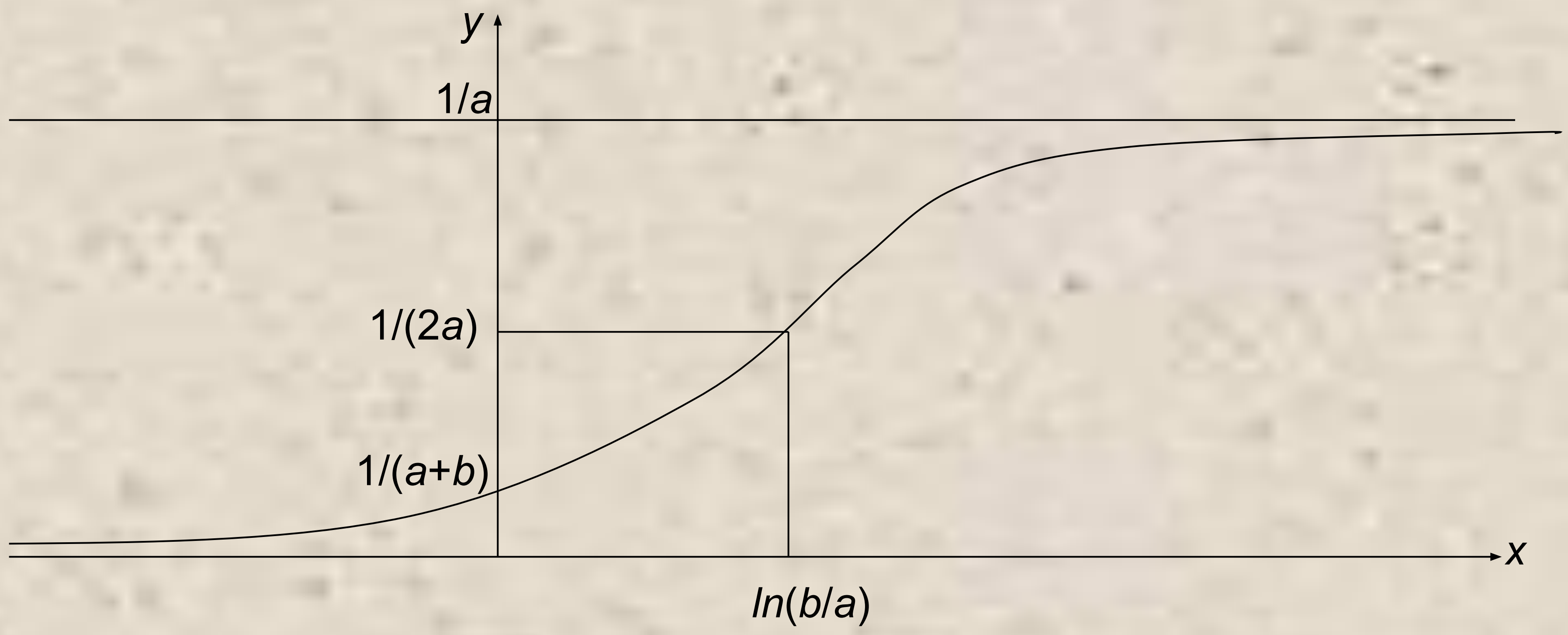
$$y = \frac{1}{a + bx + \varepsilon} \quad (50)$$

$$\Rightarrow \quad u = 1/y \quad \Rightarrow$$

$$u = a + bx + \varepsilon \quad (51)$$

$$y = \frac{1}{a + be^{-x} + \varepsilon} \quad (52) \Rightarrow \quad u = 1/y, \quad z = e^{-x} \quad \Rightarrow$$

$$u = a + bz + \varepsilon$$



$$R = \sqrt{1 - \frac{\sum \left(y - \hat{y}_x\right)^2}{\sum \left(y - \overline{y}\right)^2}} \quad (54)$$

- *индекс* корреляции (для нелинейных моделей)

$$(0 \le R \le 1)$$

$$F = \frac{R^2}{1 - R^2} \cdot \frac{n - m - 1}{m} \quad (55)$$

$$F_{табл} = F\left(\alpha; m; n - m - 1\right)$$

$$H_0 : R^2 = r^2, \quad H_1 : R^2 > r^2$$

$$t = \frac{R^2 - r^2}{m_{\left|R^2 - r^2\right|}}$$

$$m_{\left|R^2 - r^2\right|} = 2 \cdot \sqrt{\frac{\left(R^2 - r^2\right) - \left(R^2 - r^2\right)^2 \cdot \left(2 - \left(R^2 - r^2\right)\right)}{n}}$$

$$t > t_{tab}\left(\alpha; n - m - 1\right)$$

Коэффициент эластичности *y* по *x*:

$$Э = y'\frac{x}{y}$$

Вид уравнения регрессии	Коэффициент эластичности
$y = a + b \cdot x + \varepsilon$	$\dfrac{b \cdot x}{a + bx}$
$y = a + bx + cx^2 + \varepsilon$	$\dfrac{(b + 2cx) \cdot x}{a + bx + cx^2}$
$y = a + \dfrac{b}{x} + \varepsilon$	$\dfrac{-b}{ax + b}$
$y = a \cdot b^x \cdot \varepsilon$	$x \ln b$
$y = a \cdot x^b$	b
$y = a + b \ln x + \varepsilon$	$\dfrac{b}{a + b \ln x}$
$y = \dfrac{1}{a + bx + \varepsilon}$	$\dfrac{-bx}{a + bx}$

Оценивание параметров регрессии по методу максимального правдоподобия

$$y_i = a + bx_i + \varepsilon_i \qquad (3')$$

$$\varepsilon_i \sim N\left(0, \sigma^2\right) \qquad (58)$$

$$y_i \sim N\left(a + bx_i, \sigma^2\right) \qquad (59)$$

$$\left(x_i, y_i\right), \quad i = \overline{1,n}$$

$$f\left(y_i\right)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{\left(y_i-\hat{y}_i\right)^2}{2\sigma^2}} \quad (60)$$

$$L\left(y_1,\ldots,y_n,a,b,\sigma^2\right)=p\left(y_1,\ldots,y_n \middle| x_1,\ldots,x_n,a,b,\sigma^2\right)=$$

$$\prod_{i=1}^{n}p\left(y_i\right)=\prod_{i=1}^{n}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{\left(y_i-\hat{y}_i\right)^2}{2\sigma^2}}=\left(\frac{1}{\sqrt{2\pi}}\right)^n\sigma^{-n}e^{-\frac{\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2}{2\sigma^2}} \quad (61)$$

$$\ln L\left(y_1,\ldots,y_n,a,b,\sigma^2\right)=-\frac{n}{2}\ln\left(2\pi\right)-\frac{n}{2}\ln\left(\sigma^2\right)-$$

$$-\frac{1}{2\sigma^2}\sum\left(y_i-a-bx_i\right)^2 \quad (62)$$

$$\frac{\partial \ln L}{\partial a} = \frac{1}{\sigma^2} \sum \left(y_i - a - bx_i \right) = 0 \quad (63)$$

$$\frac{\partial \ln L}{\partial b} = \frac{1}{\sigma^2} \sum x_i \left(y_i - a - bx_i \right) = 0 \quad (64)$$

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{n}{2} \frac{1}{\sigma^2} + \frac{1}{2\sigma^4} \sum \left(y_i - a - bx_i \right)^2 = 0 \quad (65)$$

$$\begin{cases} b_{ММП} = \dfrac{\sum\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)}{\sum\left(x_i - \bar{x}\right)^2}; \\ a_{ММП} = \bar{y} - b_{ММП}\bar{x}; \\ \sigma^2_{ММП} = \dfrac{1}{n}\sum\left(y_i - \hat{y}_i\right)^2 \end{cases} \qquad (66)$$

$$псевдо - R^2 = 1 - \frac{\ln L}{\ln L_0} \quad (67)$$

LR-статистика:

$$2\ln\frac{L}{L_0} = 2(\ln L - \ln L_0) \quad (68)$$