

# Обработка экспериментальных данных

## Лекция 6.

### Регрессионный и корреляционный анализ.

### Нелинейная зависимость

Преподаватель: Аникеева Александра Евгеньевна

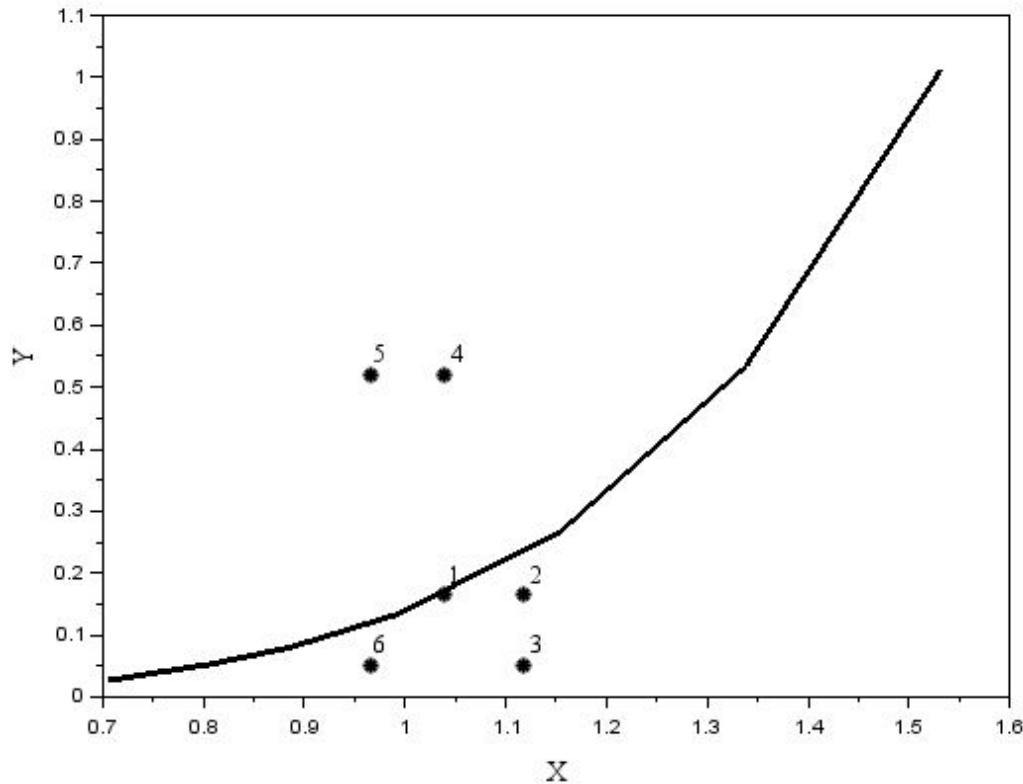
## Метод средних точек

№	Формула	Средние точки	
1	$y = ax^b$	$\sqrt{x_1 x_n}$	$\sqrt{y_1 y_n}$
2	$y = ab^x$	$\frac{x_1 + x_n}{2}$	$\sqrt{y_1 y_n}$
3	$y = \frac{1}{a + bx}$	$\frac{x_1 + x_n}{2}$	$\frac{2y_1 y_n}{y_1 + y_n}$
4	$y = a + b \lg x$	$\sqrt{x_1 x_n}$	$\frac{y_1 + y_n}{2}$
5	$y = a + \frac{b}{x}$	$\frac{2x_1 x_n}{x_1 + x_n}$	$\frac{y_1 + y_n}{2}$
6	$y = \frac{ax}{b + x}$	$\frac{2x_1 x_n}{x_1 + x_n}$	$\frac{2y_1 y_n}{y_1 + y_n}$

$T, K$	270,5	280,8	288,6	299,2	315,4	333,8	353,2
$p*10^{-5}, Pa$	0,0267	0,0533	0,0800	0,1333	0,2667	0,5333	1,0133

$$x = \frac{T - 200}{100}$$

$T, K$	270,5	280,8	288,6	299,2	315,4	333,8	353,2
$p*10^{-5}, Pa$	0,0267	0,0533	0,0800	0,1333	0,2667	0,5333	1,0133



Средняя точка для 1 кривой –  
 (1,0392594; 0,1644844)  
 Средняя точка для 2 кривой –  
 (1,1185; 0,1644844)  
 Средняя точка для 3 кривой –  
 (1,1185; 0,0520291)  
 Средняя точка для 4 кривой –  
 (1,0392594; 0,52)  
 Средняя точка для 5 кривой –  
 (0,9656325; 0,52)  
 Средняя точка для 6 кривой –  
 (0,9656325; 0,0520291)

$$y = ax^b$$

# Расчет значений коэффициентов уравнения регрессии

$$\begin{cases} \sum_{i=1}^n g_i = nc_0 + c_1 \sum_{i=1}^n t_i \\ \sum_{i=1}^n g_i t_i = c_0 \sum_{i=1}^n t_i + c_1 \sum_{i=1}^n t_i^2 \end{cases}$$

$$\begin{cases} \sum_{i=1}^n \ln y_i = nc_0 + c_1 \sum_{i=1}^n \ln x_i \\ \sum_{i=1}^n \ln y_i \ln x_i = c_0 \sum_{i=1}^n \ln x_i + c_1 \sum_{i=1}^n \ln^2 x_i \end{cases}$$

$$\begin{cases} -13,0329 = 7c_0 + 0,1692c_1 \\ 1,8467 = 0,1692c_0 + 0,4696c_1 \end{cases}$$

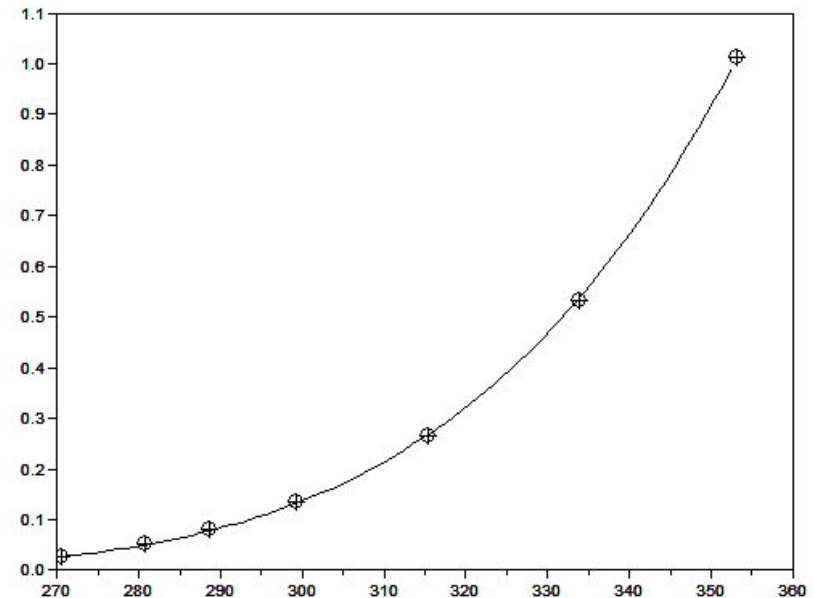
$$\ln y = \ln ax^b = \ln a + b \ln x$$

$$g = \ln y, \quad c_0 = \ln a, \quad c_1 = b, \\ t = \ln x$$

$$p = 0,1389 \cdot \left( \frac{T - 200}{100} \right)^{4,6395}$$

$$c_0 = -1,9740; \quad c_1 = 4,6395$$

$$b = c_1 = 4,6395; \quad a = e^{c_0} = 0,1389$$



## Коэффициент детерминации

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - f(x_i))^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

$T$	$y$	$p = 0,1389 \cdot \left(\frac{T - 200}{100}\right)^{4,6395}$	$(y-p)^2$	$(y-ycp)^2$
270,5	0,0267	0,0274	5,5E-07	0,0752
280,8	0,0533	0,0517	2,7E-06	0,0613
288,6	0,08	0,0792	6,1E-07	0,0488
299,2	0,1333	0,1338	2,7E-07	0,0281
315,4	0,2667	0,2700	1,1E-05	0,0012
333,8	0,5333	0,5363	8,9E-06	0,0540
353,2	1,0133	1,0051	6,7E-05	0,5075
		Сумма	9,1E-05	0,7761
$ycp$	0,300943		$R^2$	<b>0,9999</b>

# Проверка адекватности модели по критерию Фишера-Снедекора

Проверим гипотезу об адекватности полученной сглаживающей кривой исходным данным по критерию Фишера при уровне значимости  $\alpha=0,05$ .

$$F_{\text{выб}} = \frac{R^2(n-2)}{1-R^2} = \frac{0,9999 \cdot (7-2)}{1-0,9999} = 55013,3$$

Здесь  $R^2$  – коэффициент детерминации,  $n=7$ . По числу степеней свободы  $k_1=1$  и  $k_2=n-2=5$  найдем критическое значение  $F_{кр} = 6,607891$ .

Так как  $F_{\text{выб}} > F_{кр}$ , делаем вывод о том, что полученное уравнение экспоненциальной регрессии:

$$p = 0,1389 \cdot \left( \frac{T - 200}{100} \right)^{4.6395}$$

статистически значимо описывает результаты эксперимента.

# Сгруппированные данные

$y_i \backslash x_j$	5	10	15	20	25	$p_i = \sum_{j=1}^5 n_{ij}$	$p_i y_i$
10	0	0	0	1	4	5	50
11	0	3	6	4	1	14	154
12	1	3	2	0	1	7	84
13	3	0	1	0	0	4	52
$w_j$	4	6	9	5	6	<b>30</b>	<b>340</b>
$w_j x_j$	20	60	135	100	150	<b>465</b>	сумма
$w_j x_j^2$	100	600	2025	2000	3750	<b>8475</b>	
$\sum_{i=1}^4 n_{ij} y_i x_j$	255	690	1545	1080	1575	<b>5145</b>	

$$y = a + \frac{b}{x}$$

$$\left\{ \begin{array}{l} \sum_{i=1}^m p_i y_i = na + b \sum_{j=1}^k \frac{w_j}{x_j} \\ \sum_{i=1}^m \sum_{j=1}^k n_{ij} \frac{y_i}{x_j} = a \sum_{j=1}^k \frac{w_j}{x_j} + b \sum_{j=1}^k \frac{w_j}{x_j^2} \end{array} \right.$$

## Сгруппированные данные

$y_i \backslash \frac{1}{x_j}$	0,2	0,1	$\frac{1}{15}$	0,05	0,04	$P_i$	$P_i y_i$
10	0	0	0	1	4	<b>5</b>	50
11	0	3	6	4	1	<b>14</b>	154
12	1	3	2	0	1	<b>7</b>	84
13	3	0	1	0	0	<b>4</b>	52
$w_j$	<b>4</b>	<b>6</b>	<b>9</b>	<b>5</b>	<b>6</b>	<b>30</b>	340
$\frac{w_j}{x_j}$	0,8	0,6	0,6	0,25	0,24	<b>2,49</b>	$\Sigma$
$\frac{w_j}{x_j^2}$	0,16	0,06	0,04	$\frac{1}{80}$	0,0096	<b>0,2821</b>	
$\sum_{i=1}^4 n_{ij} \frac{y_i}{x_j}$	10,2	6,9	6,8667	2,7	2,52	<b>29,1867</b>	

$$\left\{ \begin{aligned} \sum_{i=1}^m p_i y_i &= na + b \sum_{j=1}^k \frac{w_j}{x_j} \\ \sum_{i=1}^m \sum_{j=1}^k n_{ij} \frac{y_i}{x_j} &= a \sum_{j=1}^k \frac{w_j}{x_j} + b \sum_{j=1}^k \frac{w_j}{x_j^2} \end{aligned} \right.$$

$$340 = 30a + 2,49b$$

$$29,1867 = 2,49a + 0,2821b$$



## Коэффициент детерминации

$$f(x) = 10,269653 + \frac{12,81542723}{x}$$

$$R^2 = \frac{\sum_{j=1}^k (f(x_j) - \bar{y})^2 w_j}{\sum_{i=1}^m (y_i - \bar{y})^2 p_i}$$

$$\sum_{i=1}^m (y_i - \bar{y})^2 p_i = 24,6667$$

$$R^2 = \frac{12,3883}{24,6667} = 0,50223$$

$$\bar{y} = \frac{\sum_{i=1}^4 m_i y_i}{n} = \frac{340}{30} = 11,3333$$

$x_j$	5	10	15	20	25
$f(x_j)$	12,8327	11,5512	11,1240	10,9104	10,7823
$w_j$	4	6	9	5	6
$(f(x_j) - \bar{y})^2 w_j$	8,9929	0,2848	0,3943	0,8943	1,8220
$\sum_{j=1}^5 (f(x_j) - \bar{y})^2 w_j$	<b>12,3883</b>				

Уравнение регрессии

$$f(x) = 10,269653 + \frac{12,81542723}{x}$$

объясняет примерно 50% всей вариации зависимой величины  $Y$ .

# Проверка адекватности

$$S_{повт}^2 = \frac{Q_{повт}}{30 - 5}; \quad S_{адекв}^2 = \frac{Q_{адекв}}{5 - 2}$$

$x_j$	5	10	15	20	25	
$f(x_j)$	12,833	11,551	11,124	10,910	10,782	
$\bar{y}_j$	12,75	11,5	11,444	10,8	10,5	
$w_j$	4	6	9	5	6	
$(f(x_j) - \bar{y}_j)^2 w_j$	0,0274	0,0157	0,9241	0,0610	0,4781	<b>1,506</b>

$$Q_{адекв} = \sum_{j=1}^5 (f(x_j) - \bar{y}_j)^2 w_j = (12,83274 - 12,75)^2 \cdot 4 +$$

$$Q_{повт} = 10,77222$$

$$+ (11,5512 - 11,5)^2 \cdot 6 + (11,12401 - 11,44444)^2 \cdot 9 + (10,91042 - 10,8)^2 \cdot 5 +$$

$$+ (10,78227 - 10,5)^2 \cdot 6 = 1,506211$$

$$f_{выб} = \frac{0,502070352}{0,430889} = 1,165197$$

$$f_{кр}(0,05, 3, 25) = 2,991$$

$$S_{повт}^2 = \frac{10,77222}{25} = 0,430889;$$

$$S_{адекв}^2 = \frac{1,506211}{3} = 0,502070352.$$

Поскольку  $f_{выб} < f_{кр}$  делаем вывод об адекватности модели выборочным данным.

# Определение силы криволинейной связи

**Корреляционное отношение  $\eta$**

$$\eta_{yx} = \sqrt{\frac{\bar{S}_y^2}{S_y^2}}$$

**Межгрупповая дисперсия**

$$\bar{S}_y^2 = \frac{1}{n} \sum_{j=1}^k (\bar{y}_j - \bar{y})^2 w_j$$

**Общая дисперсия**

$$S_y^2 = \frac{1}{n} \sum_{i=1}^m (y_i - \bar{y})^2 p_i$$

$$S_y^2 = \frac{1}{n} \sum_{i=1}^m (y_i - \bar{y})^2 p_i = \frac{1}{30} \cdot 24,66667 = 0,82222 \quad \eta_{yx} = \sqrt{\frac{\bar{S}_y^2}{S_y^2}} = \sqrt{\frac{0,463148}{0,82222}} = 0,750525$$

$$\eta = \sqrt{1 - \frac{S_{yx}^2}{S_y^2}}$$

$$S_{yx}^2 = \frac{1}{n} \sum_{i=1}^m (y_i - y_{x_i})^2$$

Значение	Характер связи	Значение	Характер связи
$\eta = 0$	Отсутствует	$0,5 \leq \eta < 0,7$	Заметная
$0 < \eta < 0,2$	Очень слабая	$0,7 \leq \eta < 0,9$	Сильная
$0,2 \leq \eta < 0,3$	Слабая	$0,9 \leq \eta < 1$	Весьма сильная
$0,3 \leq \eta < 0,5$	Умеренная	$\eta = 1$	Функциональная

**Теоретическое корреляционное отношение** определяется по формуле:

$$\eta = \sqrt{1 - \frac{\sigma_{ост}^2}{\sigma^2}} = \sqrt{\frac{\delta^2}{\sigma^2}}, \quad \text{где } \delta^2 - \text{ дисперсия выровненных значений}$$

результативного признака, т.е. рассчитанных по уравнению регрессии;

$\sigma^2$  – дисперсия эмпирических (фактических) значений результативного признака;

$\sigma_{ост}^2$  – остаточная дисперсия.