Обработка экспериментальных данных

Лекция 6. Регрессионный и корреляционный анализ. Нелинейная зависимость

Преподаватель: Аникеева Александра Евгеньевна

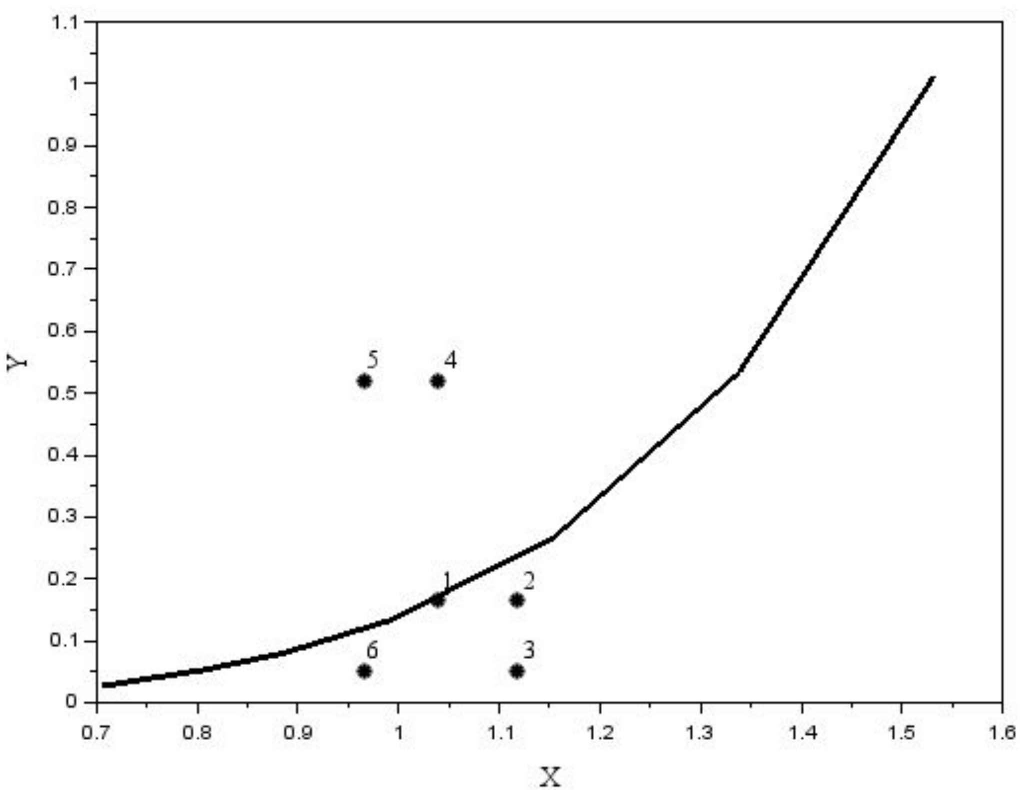
Метод средних точек

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

T, K	270,5	280,8	288,6	299,2	315,4	333,8	353,2
*p*10-5, Па*	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, Па*	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 2кривой –(1,1185; 0,1644844)
Средняя точка для 3кривой –(1,1185; 0,0520291)
Средняя точка для 4кривой –(1,0392594; 0,52)
Средняя точка для 5кривой –(0,9656325; 0,52)
Средняя точка для 6кривой –(0,9656325; 0,0520291)

$$y = ax^b$$

Расчет значений коэффициентов уравнения регрессии

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

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

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

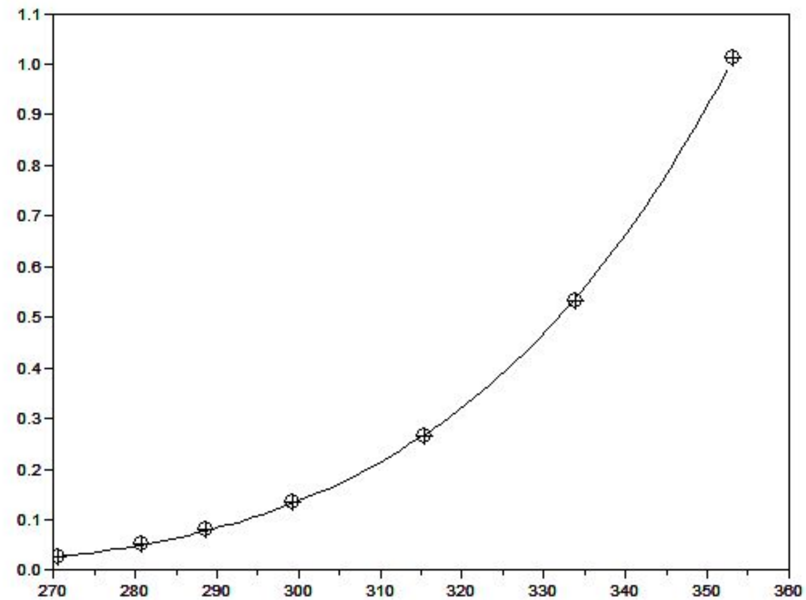
c_0=- 1,9740; c_1= 4,6395

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

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

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

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



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	$p = 0,1389 \cdot \left(\frac{T-200}{100}\right)^{4,6395}$	$(y\text{-}p)^2$	$(y\text{-}ycp)^2$
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	**0,9999**

Проверка адекватности модели по критерию Фишера-Снедекора

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

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

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 $F_{выб}$>$F_{кр}$, 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 \ 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	**30**	**340**
$w_j x_j$	20	60	135	100	150	**465**	сумма
$w_j x_j^2$	100	600	2025	2000	3750	**8475**	
$\sum_{i=1}^{4} n_{ij} y_i x_j$	255	690	1545	1080	1575	**5145**	

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

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

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	**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**	*30*	340
$\frac{w_j}{x_j}$	0,8	0,6	0,6	0,25	0,24	**2,49**	Σ
$\frac{w_j}{x_j^2}$	0,16	0,06	0,04	$\frac{1}{80}$	0,0096	**0,2821**	
$\sum_{i=1}^{4} n_{ij} \frac{y_i}{x_j}$	10,2	6,9	6,8667	2,7	2,52	**29,1867**	

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

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

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

Коэффициент детерминации

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

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

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

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

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

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
$(f(x_j - \overline{y})^2 w_j$	8,9929	0,2848	0,3943	0,8943	1,8220
$\sum_{j=1}^{5}(f(x_j - \overline{y})^2 w_j$	**12,3883**				

Уравнение регрессии

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

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

Проверка адекватности

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

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

$$Q_{адекв}=\sum_{j=1}^{5}\left(f(x_j)-\overline{y_j}\right)^2 w_j=(12{,}83274-12{,}75)^2\cdot 4+$$

$$Q_{повт}=10{,}77222$$

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

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

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

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

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

$f_{кр}(0{,}05,\ 3,\ 25)=2{,}991$

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

Определение силы криволинейной связи

Корреляционное отношение η

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

Межгрупповая дисперсия

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

Общая дисперсия

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

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

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

Значение	Характер связи	Значение	Характер связи
η = 0	Отсутствует	0,5 ≤ η < 0,7	Заметная
0 < η < 0,2	Очень слабая	0,7 ≤ η < 0,9	Сильная
0,2 ≤ η < 0,3	Слабая	0,9 ≤ η < 1	Весьма сильная
0,3 ≤ η < 0,5	Умеренная	η = 1	Функциональная

Теоретическое корреляционное отношение определяется по формуле:

$\eta = \sqrt{1 - \frac{\sigma_{ост}^2}{\sigma^2}} = \sqrt{\frac{\delta^2}{\sigma^2}}$, где δ^2 – дисперсия выровненных значений результативного признака, т.е. рассчитанных по уравнению регрессии;

σ^2 – дисперсия эмпирических (фактических) значений результативного признака;

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