

Пример. Свинец в донных отложениях р.Миасс на территории г. Челябинска

Было отобрано 10 проб донных отложений и атомно-абсорбционным спектрофотометрическим методом определено содержание свинца. Результаты, выраженные в мг/кг сухого вещества обозначены x_i и представлены в первой колонке таблицы П 1-1. Для дальнейшего анализа потребовалось нормализовать эти значения.

Этап I. Подготовительные расчеты

Выбирается диапазон значений параметра λ и шаг его изменения. Обычно достаточно интервала $[-2; 2]$, но иногда требуется его увеличение в какую-то одну сторону. В таблице ПТ1-1 выбран диапазон $[-2; 2]$ с шагом 0,5.

Для каждого значения x_i при каждом значении λ из выбранного интервала рассчитывается функция y и ее квадрат, согласно системе:

$$y = \begin{cases} \frac{x^{\lambda} - 1}{\lambda}, & \lambda \neq 0 \\ \ln x, & \lambda = 0 \end{cases}$$

Так, для $x_1=10$ и $\lambda=-2$ имеем: $y = \frac{x^{\lambda} - 1}{\lambda} = \frac{10^{-2} - 1}{-2} = \frac{0,01 - 1}{-2} = \frac{-0,99}{-2} = 0,495$; $y^2 = 0,245025$.

Все расчеты сводятся в таблицу и рассчитываются суммы значений y_i и y_i^2 . Обратите внимание, что для $\lambda=0$ расчет ведется по другой формуле – находится логарифм x_i . Результат:

Таблица ПТ1-1.

		λ						
№ п/п	x_i	-2	-1	-0,5	0	0,5	1	2
Значения y								
1.	10,0	0,4950000	0,9000000	1,3675445	2,3025851	4,3245553	9,00	49,500
2.	6,7	0,4888617	0,8507463	1,2273326	1,9021075	3,1768716	5,70	21,945
3.	6,7	0,4888617	0,8507463	1,2273326	1,9021075	3,1768716	5,70	21,945
4.	50,0	0,4998000	0,9800000	1,7171573	3,9120230	12,1421356	49,00	1249,500
5.	34,0	0,4995675	0,9705882	1,6570028	3,5263605	9,6619038	33,00	577,500
6.	66,7	0,4998876	0,9850075	1,7551122	4,2002050	14,3340136	65,70	2223,945
7.	131,4	0,4999710	0,9923896	1,8255254	4,8782461	20,9259678	130,40	8632,480
8.	154,9	0,4999792	0,9935442	1,8393043	5,0427797	22,8917657	153,90	11996,505
9.	94,4	0,4999439	0,9894068	1,7941533	4,5475411	17,4319325	93,40	4455,180
10.	115,7	0,4999626	0,9913570	1,8140641	4,7510006	19,5127869	114,70	6692,745
Σ		4,9718352	9,5037859	16,2245291	36,9649561	127,5788044	660,50	35921,245
Значения y^2								
1.	10,0	0,2450250	0,8100000	1,8701779	5,3018981	18,7017787	81,00	2450,250000
2.	6,7	0,2389857	0,7237692	1,5063453	3,6180130	10,0925134	32,49	481,583025
3.	6,7	0,2389857	0,7237692	1,5063453	3,6180130	10,0925134	32,49	481,583025
4.	50,0	0,2498000	0,9604000	2,9486292	15,3039240	147,4314575	2401,00	1561250,250000
5.	34,0	0,2495677	0,9420415	2,7456584	12,4352185	93,3523848	1089,00	333506,250000
6.	66,7	0,2498876	0,9702398	3,0804190	17,6417216	205,4639456	4316,49	4945931,363025
7.	131,4	0,2499710	0,9848372	3,3325428	23,7972851	437,8961288	17004,16	74519710,950400
8.	154,9	0,2499792	0,9871301	3,3830403	25,4296276	524,0329372	23685,21	143916132,215025
9.	94,4	0,2499439	0,9789258	3,2189859	20,6801298	303,8722701	8723,56	19848628,832400
10.	115,7	0,2499627	0,9827886	3,2908285	22,5720070	380,7488524	13156,09	44792835,635025
Σ		2,4721085	9,0639014	26,8829726	150,3978377	2131,6847819	70521,49	289921408,911925

Этап II. Расчет функции максимального лог-правдоподобия

а). Расчет дисперсии преобразованных величин по формуле: $s_{TR}^2 = \frac{\sum y^2 - \frac{(\sum y)^2}{n}}{n-1}$.

Для $\lambda = -2$ имеем:

$$s_{TR}^2 = \frac{2,4721085 - \frac{(4,9718352)^2}{10}}{10-1} = \frac{0,000193974}{9} = 2,15527 \cdot 10^{-5}; \ln s_{TR}^2 = -10,74500892.$$

б). Расчет функции лог-правдоподобия L по формуле: $L = -\frac{n}{2} \ln s_{TR}^2 + (l-1) \frac{n}{n} \sum \ln x$,

где: L – значение функции лог-правдоподобия; v – число степеней свободы, равное $n-1$; n – объем выборки; s_{TR}^2 – дисперсия преобразованных данных, x – исходные значения в наборе данных.

Сумму $\sum \ln x$ удобно взять из таблицы ПТ1-1, где она находилась для y при $\lambda=0$.

Для $\lambda = -2$ имеем:

$$L = -\frac{9}{2}(-10,74500892) + (-2-1) \frac{9}{10} 36,9649561 = 48,35254014 - 99,80538147 \approx -51,453.$$

Аналогично находятся L для остальных значений λ . Результаты сведены в таблицу ПТ1-2.

Таблица ПТ1-2.

λ	-2	-1	-0,5	0	0,5	1	2
s_{TR}	$2,15527 \times 10^{-5}$	0,00352297	0,062159794	1,528559973	56,00551652	2988,385	17876424,96
$\ln s_{TR}$	-10,74500892	-5,6484501	-2,77804688	0,424326098	4,025450195	8,00248839	16,69899336
L	-51,453	-41,119	-37,401	-35,178	-34,749	-36,011	-41,877

Этап III. Нахождение параметра λ

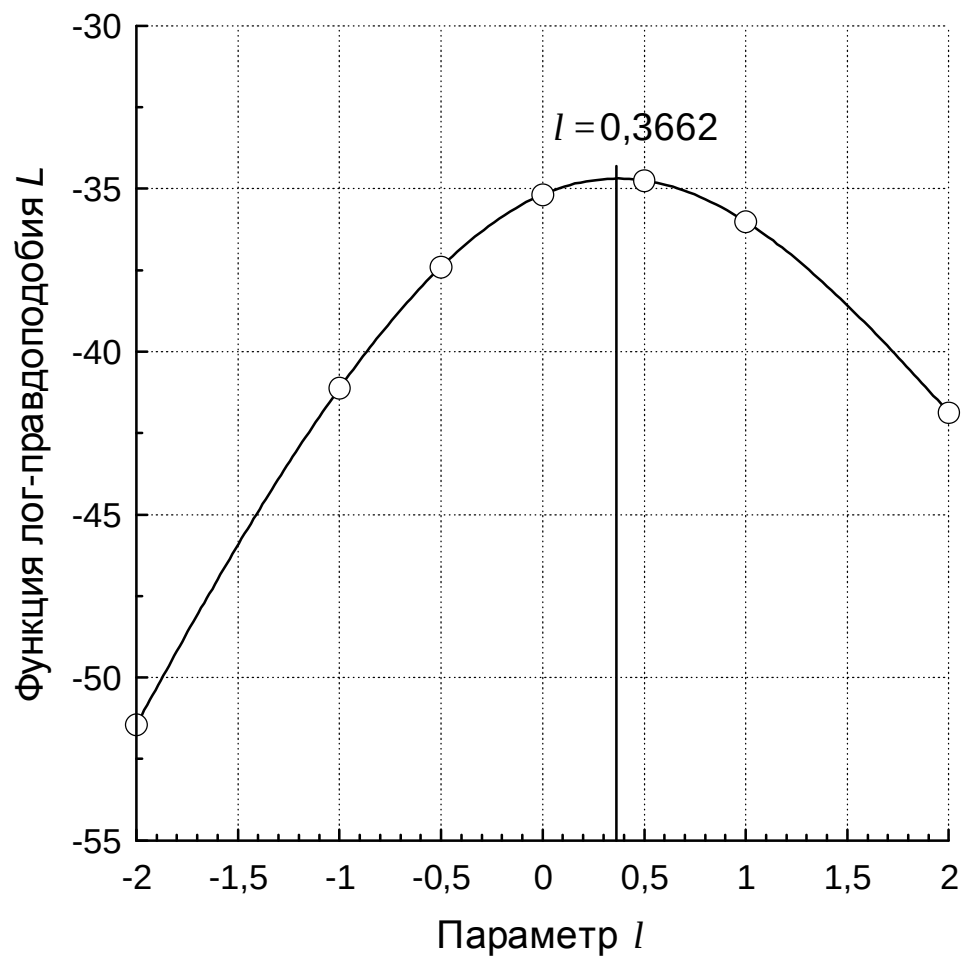
По данным таблицы ПТ1-2 строится график зависимости функции L от λ . Искомое значение λ соответствует максимуму функции лог-правдоподобия L . Наиболее рациональный алгоритм поиска λ – итерационный. Более простой, но неэкономичный алгоритм:

- (1) поиск всех значений L на заданном отрезке с очень малым шагом, например 0,001;
- (2) принятие наибольшего из найденных значений за максимум функции L , а соответствующего ему значения λ – за искомый показатель. Именно так работает, например, программа Rndom BC.

В нашем случае $\lambda=0,366$, а график изменения функции L представлен на рис. ПР1.

Этап IV. Расчет итоговых значений y_i

x_i	10,0	6,7	6,7	50,0	34,0	66,7	131,4	154,9	94,4	115,7
y_i	3,614	2,749	2,749	8,706	7,200	9,978	13,558	14,569	11,701	12,817



По источнику:

<http://www.dsa.unipr.it/soliani/soliani.html>

Глава 13. TRASFORMAZIONI DEI DATI CON TEST PER NORMALITA' E PER OUTLIER