Guide of Basics

内容概要： 数学建模算法

创建时间： 2022/4/7 13:41 **更新时间：** 2022/4/17 15:27

作者： **TwinkelStar**

最小二乘法拟合

least square fit

1、操作系统相关环境

1）硬件环境：

- 电脑

2）软件环境：

- Python3.7(向下兼容 Python3)(程序设计语言)
- Numpy1.19.5(兼容大部分版本)(科学计算库)

3）操作系统(2 选 1)：

- Windows7
- Windows10
- Windows11

2、最小二乘法

我们知道，用作图法求出直线的斜率 a 和截据 b，可以确定这条直线所对应的经验公式，但用作图法拟合直线时，由于作图连线有较大的随意性，尤其在测量数据比较分散时，对同一组测量数据，不同的人去处理，所得结果有差异，因此是一种粗略的数据处理方法，求

出的 a 和 b 误差较大。用最小二乘法拟合直线处理数据时,任何人去处理同一组数据，只要处理过程没有错误，得到的斜率 a 和截据 b 是唯一的。

最小二乘法就是将一组符合 Y=a+bX 关系的测量数据，用计算的方法求出最佳的 a 和 b。显然，关键是如何求出最佳的 a 和 b。

1）求回归方程

设直线方程的表达式为：

$$y = a + bx \tag{1}$$

要根据测量数据求出最佳的 a 和 b。对满足线性关系的一组等精度测量数据（x_i，y_i），假定自变量 x_i 的误差可以忽略，则在同一 x_i 下，测量点 y_i 和直线上的点 $a+bx_i$ 的偏差 d_i 如下：

$$\begin{aligned} d_1 &= y_1 - a - bx_1 \\ d_2 &= y_2 - a - bx_2 \\ &\vdots \\ d_n &= y_n - a - bx_n \end{aligned} \tag{2}$$

显然最好测量点都在直线上（即 $d_0=d_2=\cdots\cdots=d_n=0$），求出的 a 和 b 是最理想的，但测量点不可能都在直线上，这样只有考虑 d_1、d_2、……、d_n 为最小，也就是考虑 $d_1+d_2+\cdots\cdots+d_n$ 为最小，但因 d_1、d_2、……、d_n 有正有负，加起来可能相互抵消，因此不可取；而$|d_1|+|d_2|+\cdots\cdots+|d_n|$又不好解方程，因而不可行。现在采取一种等效方法：当 $d_{12}+d_{22}+\cdots\cdots+d_{n2}$ 对 a 和 b 为最小时，d_1、d_2、……、d_n 也为最小。取（$d_1^2+d_2^2+\cdots\cdots+d_n^2$）为最小值，求 a 和 b 的方法叫最小二乘法。

令：

$$D=\sum_{i=1}^{n}d_i^2=\sum_{i=1}^{n}[y_i-a-b_i]^2 \tag{3}$$

D 对 a 和 b 分别求一阶偏导数为：

$$\frac{\partial D}{\partial a}=-2[\sum_{i=1}^{n}y_i-na-b\sum_{i=1}^{n}x_i] \tag{4}$$

$$\frac{\partial D}{\partial b}=-2[\sum_{i=1}^{n}x_iy_i-a\sum_{i=1}^{n}x_i-b\sum_{i=1}^{n}x_i{}^2] \tag{5}$$

再求二阶偏导数为：

$$\frac{\partial^2 D}{\partial a^2}=2n \tag{6}$$

$$\frac{\partial^2 D}{\partial b^2}=2\sum_{i=1}^{n}x_i^2 \tag{7}$$

显然：

$$\frac{\partial^2 D}{\partial a^2}=2n\geq 0 \tag{8}$$

$$\frac{\partial^2 D}{\partial b^2}=2\sum_{i=1}^{n}x_i^2\geq 0 \tag{9}$$

满足最小值条件，令一阶偏导数为零：

$$\sum_{i=1}^{n}y_i-na-b\sum_{i=1}^{n}x_i=0 \tag{10}$$

$$\sum_{i=1}^{n}x_iy_i-a\sum_{i=1}^{n}x_i-b\sum_{i=1}^{n}x_i{}^2=0 \tag{11}$$

引入平均值：

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i\ ;\quad \bar{y}=\frac{1}{n}\sum_{i=1}^{n}y_i\quad \overline{x^2}=\frac{1}{n}\sum_{i=1}^{n}x_i^2\ ;\quad \overline{xy}=\frac{1}{n}\sum_{i=1}^{n}x_iy_i\ ; \tag{12}$$

则：

$$\bar{y}-a-b\bar{x}=0 \tag{13}$$

$$\overline{xy}-a\bar{x}-b\overline{x^2}=0 \tag{14}$$

解得：

$$a=\bar{y}-b\bar{x} \tag{15}$$

$$b=\frac{\overline{xy}-\bar{x}\bar{y}}{\overline{x^2}-\bar{x}^2} \tag{16}$$

将 a、b 值带入线性方程$y=a+bx$，即得到回归直线方程。

2）y、a、b 的标准差

在最小二乘法中，假定自变量误差可以忽略不计，是为了方便推导回归方程。操作中函数的误差大于自变量的误差即可认为满足假定。实际上两者均是变量，都有误差，从而导致结果 y、a、b 的标准差(n≥6)如下：

$$\sigma_y=\sqrt{\frac{\sum_{i=1}^{n}d_i^2}{n-2}}=\sqrt{\frac{\sum_{i=1}^{n}(y_i-bx_i-a)^2}{n-2}} \tag{17}$$

3）相关系数

相关系数是衡量一组测量数据 x_i、y_i 线性相关程度的参量，其定义为：

$$r=\frac{\overline{xy}-\bar{x}\bar{y}}{\sqrt{(\overline{x^2}-\bar{x}^2)(\overline{y^2}-\bar{y}^2)}} \tag{17}$$

r 值在 0<|r|≤1 中。 |r|越接近于 1，x 、y 之间线性好；r 为正，直线斜率为正，称为正相关；r 为负，直线斜率为负，称为负相关。|r|接近于 0，则测量数据点分散或 xi、yi 之间为非线性。不论测量数据好坏都能求出 a 和 b，所以我们必须有一种判断测量数据好坏的方法，用来判断什么样的测量数据不宜拟合，判断的方法是|r|<r_0时，测量数据是非线性的．r_0称为相关系数的起码值，与测量次数 n 有关，如下表 1 所示：

表 1.r 实验数据

table1.r experimental data

n	r_0	n	r_0	n	r_0
3	1.000	9	0.798	15	0.641
4	0.990	10	0.765	16	0.623
5	0.959	11	0.735	17	0.606
6	0.917	12	0.708	18	0.590
7	0.874	13	0.684	19	0.575
8	0.834	14	0.661	20	0.561

在进行一元线性回归之前应先求出 r 值，再与 r_0 比较，若|r|> r_0，则 x 和 y 具有线性关系，可求回归直线；否则反之。

3、例题

灵敏电流计的电流常数 Ki 和内阻 Rg 的测量公式为：

$$R_2 = \frac{R_s}{K_i R_1 d} U - R_g \qquad (17)$$

其中间处理过程如下，试用最小二乘法求出 K_i 和 R_g，并写出回归方程的表达式。

解：测量公式与线性方程表达式 y＝a+bx 比较：

$$y=R_2;\quad x=U;\quad b=\frac{R_s}{K_iR_1d};\quad a=-R_g;$$

数据处理如表 2 所示：

表 2.实验数据

table1.experimental data

i	1	2	3	4	5	6	7	8	平均值
$R_2(\Omega)$	400.0	350.0	300.0	250.0	200.0	150.0	100.0	50.0	225.0
U(V)	2.82	2.49	2.15	1.82	1.51	1.18	0.84	0.56	1.67125
R_2^2 ($10^4\Omega^2$)	16.00	12.25	9.000	6.250	4.000	2.250	1.000	0.250	6.375
U^2 (V^2)	7.95	6.20	4.62	3.31	2.28	1.39	0.71	0.31	3.34625
R_2U ($10^2\Omega V$)	11.3	8.72	6.45	4.55	3.02	1.77	0.84	0.28	4.615625

中间过程可多取位，通过计算可知：

$\bar{x}=1.67125$；$\bar{y}=225.0$；$\overline{x^2}=3.34625$；$\overline{y^2}=6.375\times10^4$；$\overline{xy}=461.5625$

相关系数 r：

$$r=\frac{\overline{xy}-\bar{x}\bar{y}}{\sqrt{(\overline{x^2}-\bar{x}^2)(\overline{y^2}-\bar{y}^2)}}=0.998$$

查表得知，当 n=8 时，r_0=0.834，两者比较 $r>r_0$，说明 x、y(即 U、R_2)之间线性相关，可以求回归直线：

$$b=\frac{\overline{xy}-\bar{x}\bar{y}}{\overline{x^2}-\bar{x}^2}=154.6192304$$

$$a=\bar{y}-b\bar{x}=-33.4$$

代换可得：

$$R_g=-a=33.4\Omega$$

$$\frac{R_s}{K_iR_id}=b=154.6192304$$

$$K_i=\frac{R_s}{bR_id}=3.7170\times10^{-9}A/mm$$

计算标准差为：

$\sigma_y = 2.64561902$； $\sigma_a = 2.300545589$； $\sigma_b = 1.257626418$

计算不确定度：

$\Delta R_g = \sigma_a = 2\Omega$； $\frac{\Delta K_i}{K} = \frac{\sigma_b}{b} = 0.81\%$； $\Delta K = 0.03\times10^{-9}A/mm$

电流计内阻：

$R_g = (33\pm2)\Omega$； $\frac{\Delta R_g}{R_g} = 6.1\%$

电流常数：

$K = (3.72\pm0.03)\times10^{-9}A/mm$； $\frac{\Delta K_i}{K} = 0.81\%$

求解出回归方程：

$R_2 = 155U - 33$