



酶促反应速度的回归分析

寇秀芬

(中国科学院微生物研究所, 北京)

在酶学研究中, 酶促反应速度是一个重要参数。在探讨酶的结构与功能的关系和研究酶作用机理中, 以及寻找酶促反应的最适条件等方面, 都需要酶促反应速度提供实验证据。通过对大量的实验数据的分析和处理, 可真实地反映出事物的客观规律。

(一) 回归分析的提出

在酶学研究中, 经常测定酶的反应速度来确定和比较酶的活力大小或了解催化反应合适的底物。测定酶活时, 目前常采用稀释原酶, 然后测定稀释后的酶活力, 所得结果乘以稀释倍数即为原酶活力。在这里实验条件是合理的, 保证了酶反应在一级反应范围之内。但是, 事实上有许多酶的原酶活力和稀释倍数之间不成比例关系。如纤维素酶其活力和稀释倍数的对数成直线关系, 其方程为 $A = A_0 + 2 \log r$ 。 A_0 为纤维素原酶活力, A 为稀释酶活力, r 为稀释倍数。显然, 用上述方法处理实验数据不能正确地反映纤维素酶活力的大小。

同样, 我们经常用的也是最直观的方法就是作图处理和分析实验结果。当我们做直线的时候, 平面上两点可以确定一条直线, 平面上的直线可以有无数条, 对实验获得一些离散点直观做图往往有误差, 有时候, 对同一组散点做图, 可以得到不同斜率的直线。

鉴于上述情况, 只有运用数学分析的方法, 也就是通常讲的运用回归分析法才能更深刻地反映客观规律。

(二) 回归直线的求法

1. 酶反应速度公式和回归直线方程: 酶促反应速度与底物浓度密切相关, 它可以用最大反应速度 V_m 、底物浓度 $[S]$ 和平衡常数 K_m 表示。

$$v = \frac{V_m [S]}{K_m + [S]} \quad (1.1)$$

反应速度的倒数

$$\frac{1}{v} = \frac{K_m}{V_m} \cdot \frac{1}{[S]} + \frac{1}{V_m} \quad (1.2)$$

在方程 (1.2) 中, 令

$$y = \frac{1}{v}, \quad x = \frac{1}{[S]},$$
$$a = \frac{1}{V_m}, \quad b = \frac{K_m}{V_m}.$$

得直线方程:

$$y = a + bx \quad (1.3)$$

这个方程就是回归直线方程的模拟方程。对一个具体的酶来讲, 其 V_m , K_m 都是常数, 所以它的回归直线方程很容易确定。

2. 确定回归直线的原则: 回归直线是一切直线中最接近实验点的直线, 用这条直线代表 x 与 y 的关系, 其误差比任何其他直线都小。因此, 它是一条较为合理的直线。如果用 (x_i, y_i) ($i = 1, 2, 3, \dots, N$) 表示 N 组测定数据, 那么, 任意一条直线的方程可写成如下形式:

$$y_i^* = a + bx_i \quad (2.1)$$

这条直线与客观存在的代表 x 与 y 关系的直线 ($y_i = a + bx_i$) 上每一个已知实验点 (x_i, y_i) 都有一个误差 δ_i ,

$$\delta_i = y_i - y_i^* = y_i - a - bx_i \quad (2.2)$$

而 N 个观测点所引起的误差构成了实验的总误差。显然这个总误差不能用这些 δ_i 的代数和来表示, 因为这些误差有正有负, 由于正负抵消不能代表真正的误差。用误差的绝对值之和

$$\sum_{i=1}^N |\delta_i| = |\delta_1| + |\delta_2| + |\delta_3| + \dots + |\delta_N|$$

来表示,但是这会给数学处理带来麻烦。因此采用每个误差的平方和 Q 表示总的误差,即:

$$Q = \sum_{i=1}^N \delta_i^2 = \sum_{i=1}^N (y_i - a - bx_i)^2 \quad (2.3)$$

回归直线就是在所有直线中误差平方和 Q 最小的一条直线。也就是说,回归直线的系数 b 及常数项 a 使 Q 达极小值。

根据数学分析中的极值原理,要使 Q 达到极小值,只需在(2.3)式中分别对 a, b 求偏微商,令它们等于0,于是 a, b 满足:

$$\frac{\partial Q}{\partial a} = -2 \sum_{i=1}^N (y_i - a - bx_i) = 0 \quad (2.4)$$

$$\frac{\partial Q}{\partial b} = -2 \sum_{i=1}^N (y_i - a - bx_i) \cdot x_i = 0 \quad (2.5)$$

由方程(2.4)得

$$a = \frac{1}{N} \sum_{i=1}^N y_i - \frac{1}{N} b \sum_{i=1}^N x_i$$

由方程(2.5)得

$$b = \frac{\sum_{i=1}^N y_i \cdot x_i - \frac{1}{N} \left(\sum_{i=1}^N y_i \right) \left(\sum_{i=1}^N x_i \right)}{\sum_{i=1}^N x_i^2 - \frac{1}{N} \left(\sum_{i=1}^N x_i \right)^2}$$

在计算 a, b 的公式中,所有的量都是从实验中得出,因此, $y = a + bx$ 直线方程可以确定。

3. 具体计算格式: 根据计算 a, b 的公式,求回归直线的具体计算步骤是列表进行。

例: 用回归直线求多孔玻璃固定化葡萄糖淀粉酶对麦芽糖底物的 K_m 值。根据方程(1.2)和(1.3)之间的关系。

(1) 实验记录了底物浓度的倒数 x 值和酶反应速度的倒数 y 值,列于表1中,对每一组 x 与 y 分别计算 x^2, y^2 及 xy 的值。

(2) 将表1中每列 $N = 6$ 个数值相加其和写在表最后一行。

(3) 计算 x 的离差平方和

$$L_{xx} = \sum_{i=1}^N x_i^2 - \frac{1}{N} \left(\sum_{i=1}^N x_i \right)^2,$$

y 的离差平方和

$$L_{yy} = \sum_{i=1}^N y_i^2 - \frac{1}{N} \left(\sum_{i=1}^N y_i \right)^2,$$

以及 x 与 y 的离差乘积之和

$$L_{xy} = \sum_{i=1}^N x_i y_i - \frac{1}{N} \left(\sum_{i=1}^N x_i \right) \left(\sum_{i=1}^N y_i \right)$$

列于表2。

我们可以获得多孔玻璃固定化葡萄糖淀粉酶对麦芽糖底物的回归直线方程:

$$y = 0.036 + 9.3 \times 10^{-5} x$$

根据这一方程获得一条直线与 $-x$ 相交于380处。于是得 $K_m = -\frac{1}{-380} = 2.6 \times 10^{-3} \text{g/ml}$ (图1)。

根据这一方程同样可以计算 K_m 值,当 $y = 0$ 时:

$$x = \frac{-0.036}{9.3 \times 10^{-5}} = -387.1$$

$$K_m = -\frac{1}{-387.1} = 2.58 \times 10^{-3} (\text{g/ml})$$

多孔玻璃固定化葡萄糖淀粉酶对麦芽糖底物的 K_m 值为 $2.6 \times 10^{-3} \text{g/ml}$ 。

4. 相关系数: 前面介绍了从观测数据计算回归直线方程的方法。应当指出,对一堆杂乱无章的散点,用最小二乘法配一条直线是毫无意义的,只有当 x, y 两个变量成线性或大致成线性关系时才适宜配回归直线。我们用一个量性的指标来描述两个变量线性关系的程度,这个指标叫相关系数,通常用 r 表示:

$$r = \frac{L_{xy}}{\sqrt{L_{xx} \cdot L_{yy}}}$$

$$= \frac{\sum xy - \frac{1}{N} (\sum x)(\sum y)}{\sqrt{\left[\sum x^2 - \frac{1}{N} (\sum x)^2 \right] \cdot \left[\sum y^2 - \frac{1}{N} (\sum y)^2 \right]}}$$

$|r|$ 愈大,愈接近1, x, y 之间线性关系愈密切。只有当 $|r|$ 大到一定程度才可用回归直线来近似地表示 x 与 y 之间的关系。

另外, $|r|$ 的大小可以衡量一个实验的质量。如果已知 x, y 的关系是线性关系,而计算 r 的值较小,也就是说 x, y 之间线性关系不好。

表 1 回归直线方程计算步骤 (I)

编 号	x	y	x^2	y^2	xy
1	400.0	0.230	160,000.0	0.0549	26.8
2	666.6	0.169	444,355.6	0.01103	67.0
3	1000.0	0.151	1,000,000.0	0.01769	133.0
4	1333.3	0.133	1,777,688.9	0.02592	214.7
5	1538.0	0.105	2,365,444.0	0.02856	259.9
6	2000.0	0.067	4,000,000.0	0.05290	460.0
Σ	6937.9	0.865	9,747,488.5	0.140585	1161.4

表 2 回归直线方程计算步骤 (II)

$$\begin{aligned}
 N=6 \quad \frac{1}{N} \Sigma x &= 1156.3, & \frac{1}{N} \Sigma y &= 0.1442 \\
 \frac{1}{N} \Sigma x^2 &= 1624581.4, & \frac{1}{N} \Sigma y^2 &= 0.023431 \\
 \frac{1}{N} (\Sigma xy) &= 8022409.4, & \frac{1}{N} (\Sigma y)^2 &= 0.003294 \\
 (\Sigma x)(\Sigma y)/N &= 1000.2
 \end{aligned}$$

$$\begin{aligned}
 b &= \frac{1161.4 - 1000.2}{9747488.5 - 8022409.4} = 9.3 \times 10^{-5} \\
 a &= 0.1442 - 9.3 \times 10^{-5} \times 1156.3 = 0.036
 \end{aligned}$$

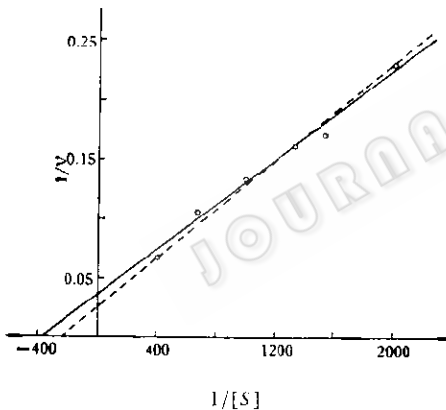


图 1 麦芽糖浓度与反应速度的关系
——回归直线 ---直观做图直线

说明实验误差较大,应该分析结果,找出原因。

(三) 回归直线和直观做图的比较

求得一条回归直线并不是我们的唯一目的。在求得一个回归方程后,它的效果如何,方程揭示的规律性,以及根据 x 变量预报 y 的取值其精度如何,都是我们所关心的。

前面谈到相关系数,它的绝对值愈大,则回归效果愈好。计算多孔玻璃固定化葡萄糖淀粉酶对麦芽糖底物的回归直线相关系数为 0.98,

实验质量是理想的。这条直线揭示其 K_m 值为 $2.6 \times 10^{-3} \text{g/ml}$ 。直观做图求得其 K_m 为 $3.8 \times 10^{-3} \text{g/ml}$ (图 1)。

二者相差千分之一。这个差异是由于做图的斜率不同引起的。直观做图是使直线尽量多地通过某些实验点,而其他实验点考虑的因素就小了些。回归直线是通过所有实验点的重心位置,这条直线是所有直线中误差最小的一条直线。所以说它可以更深刻地反映实验本身的客观情况。

还应该指出,回归分析法与直观做图法一样,实验观测点愈多,分析效果愈好。

(四) 回归分析解决的问题

回归分析是处理变量相关关系的数学统计方法。工业生产和科学研究中的许多问题可以用这种方法得到解决。

E. Galas 等人通过数学运算得到纤维素酶原酶活力和稀释酶活力的直线方程:

$$A = A_0(1 + \log r)$$

他认为其斜率是一个常数,与作用底物无关,与酶中各组份无关。该方法不仅适合于纤

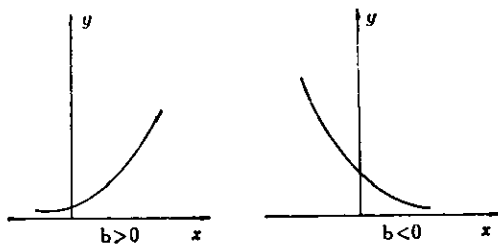


图2 指数曲线 $y = de^{bx}$

纤维素酶和半纤维素酶,还适用一切DNS测定还原糖的方法来表示酶活力的其他酶。这样就可以把不同文献或不同实验室在不同稀释倍数的情况下测定的酶活力换算成原酶活力进行比较。

应当指出,回归分析不只是解决两个变量成线性关系的问题。实际工作中我们常遇到一些曲线。如细菌培养,每一时刻生物生成总量

y 与培养时间 x 之间有指数关系(图2)。

$$y = de^{bx}$$

对此,我们化曲线为直线,令

$$y' = \ln y, \quad a = \ln d,$$

则有直线方程 $y' = a + bx$ 。

对于实验所得一批数据配回归曲线时,如果曲线不易从理论上或经验确定,可以根据实验观测点分布的形状与已知函数图形进行比较,确定一种与之相近的曲线,再化曲线为直线的方法予以解决。

对于不能化为常规曲线的问题,从理论上讲可以用一个多项式来模拟,即:

$$y = \sum_{i=0}^N a_i x^i$$

虽然计算较复杂,但可以解决问题。