

基于最小二乘法的线性回归方程推导与应用分析

□陈雨彤 江苏省梅村高级中学

【摘要】 在日常的生产生活以及学习过程中，数据处理往往是发现和认识事件规律的关键核心，如经济数据分析、机器使用时间预测等，而事实上这些数据有存在内在的关联性，需要借助数学工具达到发掘规律的目的。本文正是基于对日常工作学习中遇到的一类线性关系，文章详细地利用最小二乘法来估计参数并推导出线性回归方程原理进行了阐述，并借助线性回归方程对学习中的—些物理实验画图问题进行了实例分析，以此拓展线性回归方法的应用。

【关键词】 最小二乘法 线性回归方程 参数估计 物理实验

一、引言

在高中物理实验题中，存在一些数据对，其中的自变量和因变量基本呈线性关系，题目中往往要求画出一条和数据基本符合的函数图像，但这样画出的函数图像是否是最符合数据之间内在的关系呢，却很难得到确切的验证，或者说只是一个近似结果。但是，通过了解到函数拟合中的线性回归方程，发现在一定的准则前提下根据数据可以拟合一个最优的函数图像，因此，如果需要准确绘制图像时，往往使用线性回归方程，但是对于给出的线性回归方程是如何得来的，一开始并没有深入了解。

本文正是基于对线性回归方程的浓厚兴趣，通过调研资料，对线性回归方程的学习和掌握，对经典的基于最小二乘法推导出了线性回归方程中参数的计算公式，并应用此公式计算一些物理实验题中与给出数据最为相符的函数关系，绘制出较为准确的函数图像，以便得到最合理的解答。

二、线性回归方程的原理

2.1 线性回归模型的建立

顾名思义，线性回归方程是建立因变量与自变量之间的线性关系，为了根据给出数据建立准确的计算函数关系，对

于一般意义下的一元线性关系，可以建立简单的一元线性函数模型，来描述因变量 y 是如何随自变量 x 的变化而变化的，即建立式 (1) 的线性关系式。

$$y = \alpha x + \beta \quad (1)$$

其中 α, β 都为实数，该线性模型本质上就是高中所学的直线方程， α 为直线方程的斜率， β 为直线方程的截距。其实对式 (1) 中的参数 α, β 进行估计方法有很多，如广义最小二乘法、稳健回归等。为了论述方便，本文将详细论述基于最小二乘法来对线性模型中的参数 α, β 估计方法。

2.2 基于最小二乘法的参数估计

最小二乘法是一种数学优化技术，它通过最小化误差的平方和寻找数据的最佳函数匹配。利用最小二乘法可以简便地求得未知的数据，并使得这些求得的数据与实际数据之间误差的平方和为最小，即建立如式 (2) 所述的目标函数。

$$W = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2)$$

其中 y_i 是实际的观测数据， $\hat{y}_i$ 是根据式 (1) 所定义的线性模型计算得到的预测值，于是，当在目标最小时，可以通过数学的方法确定线性模型中的参数 α, β 表达式。相应的推导过程如下描述：

未受到实质性伤害的前提下，对同人创作保持宽容与鼓励的态度，但自身著作权一经受侵犯，必然有法律帮助其维权和赔偿。

四、结论

保护著作权的目的主要是保护著作权人的经济利益不受侵害，在笔者看来，著作权法是对原作者经济利益的保护和利益的平衡。而有一些同人作品的出现，打破了平衡，损害了原作应有的经济利益，于是便出现了有关同人作品著作权问题的辨析和对其侵权行为的讨论。从法律角度来说，我国并未有一条法律条例对同人进行过清晰的定位并规定其创

作时需注意和创作会可享有的权利。因此，谨防同人创作侵权问题是同人作者与原作者都应自觉注重之事。我们虽对侵权行为深恶痛绝，但并不意味着要限制同人作品的创作。把握同人影响界限，平衡好双方利益是我们利用著作权的途径。在保障激励同人创作的时候，也不能纵容其任意发展最终造成对原作者的经济利益伤害。在维护原作者的著作权是，尊重同人作品的著作权，在其未侵权情况下，不得干涉同人的创作。为有效缓解同人作品侵权的事情出现和鼓励正当的同人创作，需要同人作者与原作者的共同努力来构建同人圈得一片乐土。

参考文献

- [1] 曹洵，虚拟社区的动漫迷文化实践模式研究，青年研究，2011年第4期。
- [2] 孙山，同人作品传播中的《著作权法》限制，科技与出版，2017年第12期。
- [3] 张楚主编，知识产权（第二版），高等教育出版社，2010年出版。
- [4] 胡婧，论非演绎性同人作品与原作品的知识产权“对碰”，法政学刊，2018年第1期。
- [5] 陈秋英，同人作品著作权侵权风险评析——以虚拟角色商品化权为视角的考量，浙江万里学院学报，2018年第1期。
- [6] 徐微，探究同人作品的现状及发展前景，赤子（下旬），2016年第9期。

$$\begin{aligned}
 W &= \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - \alpha x_i - \beta)^2 \\
 &= (y_1 - \alpha x_1 - \beta)^2 + (y_2 - \alpha x_2 - \beta)^2 + \dots + (y_n - \alpha x_n - \beta)^2 \\
 &= y_1^2 - 2y_1(\alpha x_1 + \beta) + (\alpha x_1 + \beta)^2 + y_2^2 - 2y_2(\alpha x_2 + \beta) + (\alpha x_2 + \beta)^2 + \dots \\
 &\quad + y_n^2 - 2y_n(\alpha x_n + \beta) + (\alpha x_n + \beta)^2 \\
 &= y_1^2 - 2\alpha x_1 y_1 - 2\beta y_1 + \alpha^2 x_1^2 + 2\alpha\beta x_1 + \beta^2 + y_2^2 - 2\alpha x_2 y_2 - 2\beta y_2 + \alpha^2 x_2^2 \\
 &\quad + 2\alpha\beta x_2 + \beta^2 + \dots + y_n^2 - 2\alpha x_n y_n - 2\beta y_n + \alpha^2 x_n^2 + 2\alpha\beta x_n + \beta^2 \\
 &= (y_1^2 + y_2^2 + \dots + y_n^2) - 2\alpha(x_1 y_1 + x_2 y_2 + \dots + x_n y_n) - 2\beta(y_1 + y_2 + \dots + y_n) \\
 &\quad + \alpha^2(x_1^2 + x_2^2 + \dots + x_n^2) + 2\alpha\beta(x_1 + x_2 + \dots + x_n) + n\beta^2
 \end{aligned} \tag{3}$$

随后, 可以利用平均数对其进行进一步化简, 相应的化简基础如 (4)-(8):

$$y_1^2 + y_2^2 + \dots + y_n^2 = n\bar{y}^2 \tag{4}$$

$$2\alpha(x_1 y_1 + x_2 y_2 + \dots + x_n y_n) = 2n\alpha\bar{x}\bar{y} \tag{5}$$

$$2\beta(y_1 + y_2 + \dots + y_n) = 2n\beta\bar{y} \tag{6}$$

$$\alpha^2(x_1^2 + x_2^2 + \dots + x_n^2) = n\alpha^2\bar{x}^2 \tag{7}$$

$$2\alpha\beta(x_1 + x_2 + \dots + x_n) = 2n\alpha\beta\bar{x} \tag{8}$$

因此, 式 (3) 可化简为:

$$W = n\bar{y}^2 - 2n\alpha\bar{x}\bar{y} - 2n\beta\bar{y} + n\alpha^2\bar{x}^2 + 2n\alpha\beta\bar{x} + n\beta^2 \tag{9}$$

为了使 W 的值取得最小, 可以对式 (9) 中的 W 分别 α, β 求偏导数, 并令偏导数都为 0, 如式 (10),(11) 所示。

$$\frac{\partial W}{\partial \alpha} = -2n\bar{x}\bar{y} + 2n\alpha\bar{x}^2 + 2n\beta\bar{x} = 0 \tag{10}$$

$$\frac{\partial W}{\partial \beta} = -2n\bar{y} + 2n\alpha\bar{x} + 2n\beta = 0 \tag{11}$$

进一步化简, 消掉 $2n$ 可得到关于 α, β 方程组 (12)(13)。

$$-\bar{x}\bar{y} + \alpha\bar{x}^2 + \beta\bar{x} = 0 \tag{12}$$

$$-\bar{y} + \alpha\bar{x} + \beta = 0 \tag{13}$$

最后可以解得 α 和 β 的求解公式 (14)(15) 所示:

$$\alpha = \frac{\bar{x}\bar{y} - \bar{x}\bar{y}}{(\bar{x})^2 - \bar{x}^2} \tag{14}$$

$$\beta = \bar{y} - \alpha\bar{x} \tag{15}$$

在式 (14)(15) 的帮助下, 可以借助观测数据求得的 α 和 β 的值, 便可以得到基于给出数据最为相符的线性回归函数关系, 并根据计算得出的函数关系可绘制出较为准确的函数图像, 用作后续的规律推演。

2.3 线性拟合的评价方法

由于 2.2 计算得到的函数关系只是基于最小二乘法原理, 即基于偏差平方和最小的原则下得到, 是一个近似的线性关系。因此在实际中需要通过一定的标准检验并判断计算得出的函数关系实际值存在多大的偏差, 偏差是否过大, 该线性模型是否合理等。

为了解决这个问题, 在线性回归方程的评价中引入皮尔逊相关系数 R , 并用 R^2 来判断计算出的函数关系实际值之间的偏差大小。定义如式 (16) 所示。

$$R^2 = \frac{SSR}{SST} \text{ 或 } R^2 = 1 - \frac{SSE}{SST} \tag{16}$$

其中, SSR 表示回归线可以解释的偏差, SSE 表示回归线不能解释的偏差, 即其他因素影响造成的偏差, SST 表示总偏差, 其计算公式如下 (17)(18)(19) 所示:

$$SST = \sum_{i=1}^n (y_i - \bar{y})^2 \tag{17}$$

$$SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \tag{18}$$

$$SSR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \tag{19}$$

根据给出数据算得 R^2 后, 即可知道函数关系与实际之间的偏差大小, R^2 取值在 0, 1 之间, 越接近 1, 说明拟合程度越好, 通常在得到线性回归方程后, 都需要对所得到的线性关系进行科学合理的评价。

三、线性回归方程在物理实验中的应用分析

3.1 应用背景介绍

在高中物理电学实验题目中, 经常遇到给出两组数据, 寻找其内在关系, 如伏安法测电阻、电源电动势及内阻的实验中, 本质上都是线性关系。本文以干电池的电动势与内阻实验为例进行线性回归原理的应用阐述与分析。

题目: 某实验小组利用图 1 所示电路测量一节干电池的电动势和内阻, 所得实验数据如表 1 所示, 请在给出的直角坐标系上画出 $U-I$ 图像。

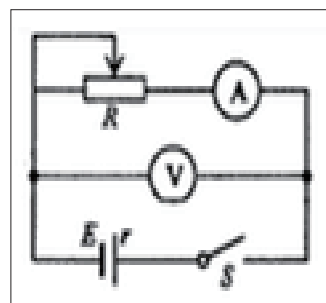


图 1 实验原理电路图

表 1 实验所测数据

U/V	1.96	1.86	1.80	1.84	1.64	1.56
I/A	0.05	0.15	0.25	0.35	0.45	0.55

在没有掌握线性回归方程之前, 遇到此类题目往往会比较随意, 但通过对线性回归的理解与掌握, 下面将进行严格的数学应用计算分析。

3.2 应用过程分析

在这道题目中电流 I 相当于自变量 x , 电压 U 相当于因变量 y , 可以根据论文第 2 部分所述内容所推导出的线性回归公式计算出与数据相符的函数图像的斜率和截距。

首先, 设 $U = \alpha I + \beta$, 根据给出数据, 可以计算出 U 和 I 的平均值, 如 (20)(21)(22)(23) 所示:

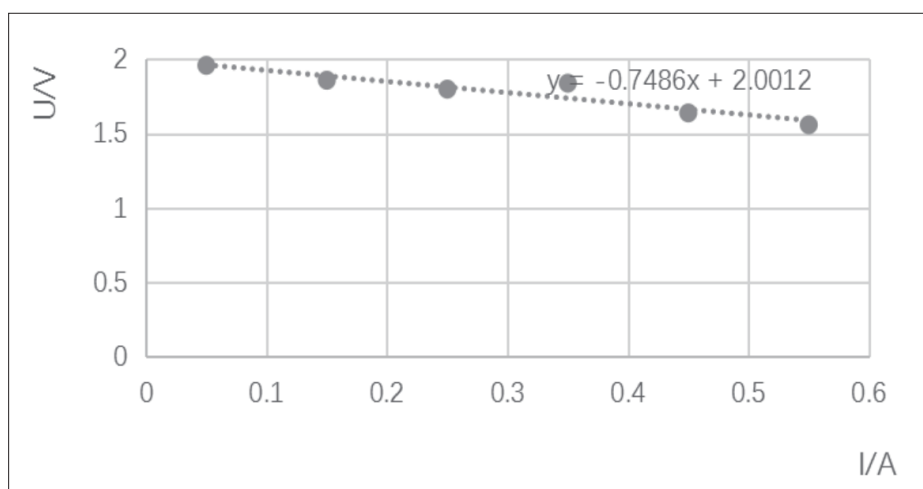


图2 基于 Excel 拟合的线性关系图

$$\bar{U} = \frac{U_1 + U_2 + \dots + U_6}{6} = \frac{1.96 + 1.86 + 1.80 + 1.84 + 1.64 + 1.56}{6} \approx 1.777 \quad (20)$$

$$\bar{I} = \frac{I_1 + I_2 + \dots + I_6}{6} = \frac{0.05 + 0.15 + 0.25 + 0.35 + 0.45 + 0.55}{6} = 0.3 \quad (21)$$

$$\overline{UI} = \frac{U_1 I_1 + U_2 I_2 + \dots + U_6 I_6}{6} \approx 0.511 \quad (22)$$

$$\bar{I}^2 = \frac{I_1^2 + I_2^2 + \dots + I_6^2}{6} \approx 0.119 \quad (23)$$

由此根据式 (14)(15), 可以根据推导出的线性回归方程计算出 α 及 β 的值:

$$\alpha = \frac{\overline{UI} - \bar{U}\bar{I}}{(\bar{I})^2 - \bar{I}^2} = \frac{1.777 \times 0.3 - 0.511}{0.09 - 0.119} \approx -0.762 \quad (24)$$

$$\beta = \bar{U} - \alpha \bar{I} = 1.777 + 0.762 \times 0.3 = 2.0056 \quad (25)$$

因此, 得到的线性函数关系式如下:

$$U = -0.762I + 2.0056 \quad (26)$$

最后, 通过 Excel 拟合线性关系得出的函数关系和图像 (如图 2 所示) 与计算出的关系式基本一致, 可以说明推导出的线性关系参数公式是合理的。

接下来, 按照 2.2 所述的线性拟合评价方法对上述所得到的线性回归方程进行科学评价, 需要计算得出的函数关系与实际值的误差大小, 首先按照 (17)(18) 计算出相应的 SSR 及 SST, 得出如下数值:

$$SST = \sum_{i=1}^n (U_i - \bar{U})^2 = (U_1 - \bar{U})^2 + (U_2 - \bar{U})^2 + \dots + (U_6 - \bar{U})^2 = 0.110734$$

$$SSR = \sum_{i=1}^n (\hat{U}_i - \bar{U})^2 = (\hat{U}_1 - \bar{U})^2 + (\hat{U}_2 - \bar{U})^2 + \dots + (\hat{U}_6 - \bar{U})^2 = 0.1016127$$

所以, $R^2 = \frac{SSR}{SST} = \frac{0.1016127}{0.110734} \approx 0.9176$, 与 1 十分接近, 因此误差较小, 可见线性回归方程的应用可以提高计算的准确度,

但是在实际考试中需要进行权衡, 毕竟进行线性回归计算相对耗时, 且容易出错, 如果采用现成的计算机程序便可快速得到答案。

3.3 应用过程总结

通过对 3.2 所述这道题目的线性回归应用分析, 可以发现推导得出的线性关系方程与实际情况符合得很好, 因此这样的线性关系方程是可取的较为正确的。实际上, 现实生活中有很多类似的关系都可以通过线性回归进行预测, 如在数学领域可以用于寻找两个变量的线性相关性, 在经济领域用于趋势判断, 如 GDP、股票、商品价格等, 总之, 线性回归方程在国民生产生活中应用十分广泛, 具有十分重要的应用价值。

四、结论

经过利用最小二乘法推导出线性关系方程, 并利用得出的公式, 可以得到给出实际数据较为准确相符的函数关系式, 并可以根据函数关系式绘制出比较精确的函数图像, 而根据绘制出的函数图像可以尝试预测实际数据没有涵盖的其他可能数据, 并且可以根据引入的系数判断计算得出的函数关系与实际值之间的误差大小, 从而得出准确的判断。这样的线性关系方程, 既可以在物理实验画图题目中使用, 在日常生产生活中, 若有存在线性关系趋势的数据时, 同时又需要分析或预测时, 同样可以用到本文所述的基于最小二乘法线性回归方程。

除此之外, 线性回归方程还只是函数拟合中最为简单的一种, 有的数据之间并不存在线性关系, 可能有二次关系、三角关系等, 需要充分利用数据理论和计算机技术充分挖掘内在的数学关系对实际问题的分析与解决是十分重要的。

参考文献

- [1] 李玉毛, 何涛, 刘冬. 一元线性回归方法的理论及其应用 [J]. 赤峰学院学报 (自然版), 2017, 33(15):1-2.
- [2] 梁博尧. 浅析最小二乘法及其在数学建模中的应用 [J]. 中国新通信, 2018(3).
- [3] 陈明. 一元线性回归模型预测图书借阅量 [J]. 大学教育, 2016(5):111-112.
- [4] 程 sir. 一元线性回归的细节. <https://www.jianshu.com/p/fcd220697182#>, 2016.01.27.
- [5] 张春琦. 线性回归方程 [M]. 《新高考 (高三数学)》, 2006 (11):32-35.