

基于R语言的等厚干涉实验数据统计分析

李峰 武红*

南京邮电大学理学院

摘要: 大学物理实验在处理数据的过程中经常需要用到统计学的方法, R软件作为一款专业的统计软件在大学物理实验中的应用还比较少, 本文以等厚干涉实验为例, 介绍R软件如何处理等厚干涉现象中出现的线性回归问题。为我们展示了R软件在处理大学物理实验中的统计学问题的优势。为我们有针对性改进大学物理实验教学方法提供客观的评价依据。

关键词: 等厚干涉; R语言; 线性回归; 相关系数

【DOI】 10.12252/j.issn.2096-6288.2022.03.008

引言

作图法是大学物理实验中常见的数据处理方法之一。学生可以通过坐标纸手工作图定性地分析出两个变量之间的关系。然而, 对于数据量较大且精度要求较高的实验, 作图法难以直观准确地反映变量之间的统计学关系, 这是物理实验中数据处理的一项难题。R语言软件是一款专业的统计学软件^[1, 2], 由贝尔实验室统计学家Robert Gentleman和Ross Ihaka创建, 可以更好的对实验数据进行统计分析, 向我们展示更多的数据的统计学性质^[3]。例如: 一元线性回归分析, 多元线性回归分析, 非线性回归分析问题等^[4]。

本文结合等厚干涉实验^[5-7], 介绍如何采用R语言来对物理实验中的数据进行统计学分析。此外, 通过R语言编程, 可以实现对全班学生实验结果的一键分析^[8, 9]。这既可以实时高效地对大量实验数据进行结果分析, 对实验误差进行归类, 也为后续进一步指导学生实验提供客观的评分标准, 是利用电子化信息手段改进实验教学效果的一种新尝试。

一、实验原理

在等厚干涉测量透镜曲率半径的实验中, 根据等厚干涉公式, 有 $D_m^2 - D_n^2 = 4\lambda R(m - n)$ 。其中, m 和 n 为等厚干涉的暗环级数, D_m 和 D_n 为第 m 和 n 级的暗环直径。即 $D = |X_{左} - X_{右}|$ 。本实验中我们选取钠光灯为光源, 入射单色光的波长 $\lambda = 589.3 \text{ nm}$ 。R为待测透镜的曲率半径。

为了对数据进行线性回归处理, 我们设 $Y = D_m^2 - D_n^2$, $X = m - n$, 则公式(1)可以表示成线性方程的形式 $Y = \beta_0 + \beta_1 X$, 即可得 $\beta_1 = 4\lambda R$ 。由于实验误差 ε 无法消除, 因此第 i 次实验的测量变量 Y_i 由下面的一元线性回归方程来表示 $Y_i = \beta_0 + \beta_1 X_i + \varepsilon$ 。其中, β_0 , β_1 是回归系数, ε 是在测量过程中产生的误差。因此, 可得球

面的曲率半径R的估计值为 $R = \frac{\beta_1}{4\lambda}$ 。由最小二乘法可得, 参数 β_1 的估计量为:

$$\beta_1 = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \quad (1)$$

此外, 已知测量值的相关系数 r 为

$$r = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \cdot \sqrt{n \sum_{i=1}^n y_i^2 - (\sum_{i=1}^n y_i)^2}} \quad (2)$$

因此, 测量的不确定度可以表示为

$$u_r(R) = \frac{u(R)}{R} = \sqrt{\frac{1}{r^2} - 1} \quad (3)$$

最终, 我们将对实验数据的显著性 p 值, 线性回归系数 β_0 、 β_1 , 样本决定系数 R^2 , 相关系数 r , 和不确定度 U_r 等方面对实验数据进行综合分析, 为学生的实验结果提供客观的统计学评分标准。

二、实例分析

一个小班共10位同学完成了本次利用等厚干涉原理测量透镜的曲率半径的实验。为了方便讨论, 我们将对前两位同学的实验数据进行比较, 见表1。首先, 我们对实验数据的关系的显著性 p 值进行检验。一般认为 $p < 0.05$ 表示两个变量之间的关系具有显著性, 相反 $p > 0.05$ 则认为两个变量之间的关系不显著。检验结果表明, 第一组实验数据的 $p = 8.0 \times 10^{-11}$, 第二组实验数据的 $p = 3.3 \times 10^{-6}$ 。P值分析结果显示两组数据都具有较好的显著性, 且第一组的实验数据的显著性更高。

表1 两位同学的等厚干涉测量数据

环的级数		5	10	15	20	25	30
同学1	环的位置/mm ($X_{左}$)	26.978	27.519	27.999	28.391	28.759	29.091
	环的位置/mm ($X_{右}$)	23.382	22.709	22.212	21.796	21.431	21.092
	环的直径/mm $D= X_{左}-X_{右} $	3.596	4.810	5.787	6.595	7.328	7.999
	$ D_m^2 - D_n^2 /\text{mm}^2$	12.931	23.136	33.489	43.494	53.700	63.984
同学2	环的位置/mm ($X_{左}$)	18.883	19.471	19.971	20.371	20.741	21.084
	环的位置/mm ($X_{右}$)	14.965	14.367	13.882	13.467	13.112	12.763
	环的直径/mm $D= X_{左}-X_{右} $	3.918	5.104	6.089	6.904	7.629	8.321
	$ D_m^2 - D_n^2 /\text{mm}^2$	15.351	26.051	37.076	47.665	58.202	69.239

其次，基于R语言对两组数据分别进行线性回归分析。图1 (ab) 分别给出了两组实验数据的散点图和估计方程的曲线。回归分析结果显示，两组数据的估计方程分别为 $y_i = 2.760 + 2.040x_i$ 和 $y_i = 1.442 + 2.143x_i$ 。分别利用公式 $\hat{R} = \frac{\hat{\beta}_1}{4\lambda}$ 计算，可得两组实验测量的曲率半径分别为865.334mm和909.270mm。后者比前者大了约5.0%。然而，考虑到镜片个体的差异性，我们无法从散点图和回归方程中判断哪位同学的实验结果更优。

我们借助统计分析R软件对数据的决定系数 R^2 值进行拟合程度分析。决定系数 R^2 值的大小决定了自变量和因变量两者间相关的密切程度。决定系数越大，则自变量对因变量的解释程度越高，观察点在回归直线附近越密集。结果表明两组数据的样本决定系数 R^2 值分别为0.9999和0.9963。因此从统计角度来看，第一组实验数据的拟合效果更好。

接着，我们对两组实验进行相关系数 r 分析。一般认为 $r < 0.3$ 为弱相关， $0.3 \leq r \leq 0.7$ 为中相关， $r > 0.7$ 为

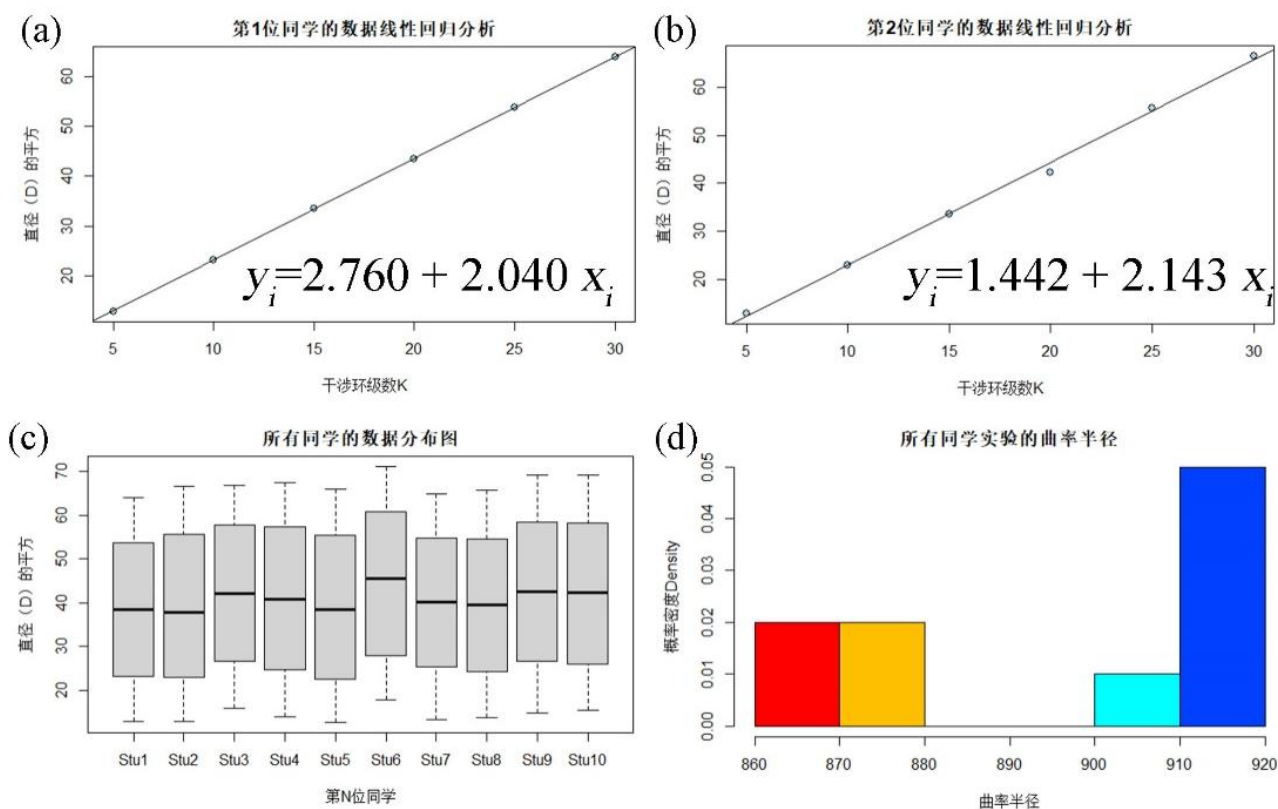


图1 (a, b) 第一和二号学生的实验数据线性回归分析。(c) 全班同学的原始实验数据分布图。(d) 全班同学的实验曲率半径结果分布图。

强相关。利用公式5分析数据，结果表明第一组实验的相关系数 $r_1=0.9999$ ，测量的不确定度 $u_{r_1}(R)=0.0019$ ，第二组实验的相关系数 $r_2=0.9985$ ，测量的不确定度 $u_{r_2}(R)=0.0273$ 。从测量的不确定度来看第二组实验的不确定度更大些，这与从统计上的样本决定系数的角度得到的结论是一致的。

采用同样的统计方法，我们将小班中所有同学的实验数据进行统计学分析。如图1(c)所示，整体来看10位同学的数据分布较为合理，其等厚干涉条纹直径 D^2 的中间值在 40mm^2 上下浮动。其中，第6位同学的数值明显偏大，说明其未能很好掌握本次实验，需教师帮助该生找到数据偏大的原因。最后，对本次实验所有同学测量的曲率半径结果进行统计学分析，如图1(d)所示。统计发现，同学们计算得到的曲率半径分布在两个区间内：800~880mm和900~920mm。导致这一差异的主要原因来自两个部分：一是不同平凸玻璃由于打磨加工导致的客观差异；二是在实验前是否调整了平凸玻璃与平面玻璃之间的间距使牛顿环中间正好形成一个暗斑。由于玻璃挤压，因此调整后的平凸玻璃的曲率半径会比不调整的情况更大。通过R语言辅助，我们可以快速筛选出第1、3、7、8个同学未按照要求在实验前正确调整牛顿环中间暗斑位置。

```
###本实验的R语言代码：
library(readxl)
Interference=read_excel("Equal-thickness-
interference.xlsx")
lambda=589.3 #钠光灯，单色光波长，单位nm
####处理第1个同学的数据，其他同学以此类推
plot(Interference$Stu1~
Interference$circle, main="第1位同学的数据线性
回归分析", xlab="干涉环级数K", ylab="直径
(D)的平方", bg="lightblue", col="black",
cex=1.1, pch=21, frame=TRUE)
fit=lm(Interference$Stu1~
Interference$circle, data=Interference)
abline(fit) #线性回归拟合
summary(fit)
Ur[10]=(1/r[10]^2-1)/(length(YY)-
2))^0.5 #计算不确定度
Radius=Fit[2,,1]/lambda*1000000 #计算曲
```

率半径

结语

本文以等厚干涉为例，介绍了R语言在大学物理实验中的具体用法和分析手段。R语言的优势是能够批量处理全班同学的实验数据结果，即可以高效快速获得大批量实验数据的统计学性质，也可以提供了实验结果优劣的客观评判标准，是进行信息化教学设计的一种有益尝试，有望与虚拟仿真实验技术相结合，进一步提高大学物理实验的教学质量。

参考文献

- [1] Bali, R., Sarkar, D. R语言机器学习：实用案例分析[M]. 北京：机械工业出版社，2017.
 - [2] 徐珉久. R语言与数据分析实战[M]. 北京：人民邮电出版社，2017.
 - [3] 薛毅, 陈立萍. 统计建模与R软件[M]. 清华大学出版社，2015.
 - [4] 何晓群, 刘文卿. 应用回归分析[M]. 北京：中国人民大学出版社，2019.
 - [5] 藤坚, 徐军华. 牛顿环实验数据处理的最佳方法[J]. 西安邮电学院学报, 2016, 11, (2): 129-130.
 - [6] 马天成, 李全彬. 基于Python的牛顿环实验数据拟合及分析方法[J]. 大学物理实验, 2019, 32, (3): 112-115.
 - [7] 张庆春, 赵瑛, 刘艳, 等. 基于R软件的牛顿环实验数据的回归分析[J]. 大学物理实验, 2021, 34, (6): 110-112.
 - [8] 顾辉, 杨青, 蒋成功. 关联规则在成绩分析中的研究及应用[J]. 计算机应用, 2015, 35, (198): 149-151.
 - [9] 汤银才. R语言与统计分析[M]. 北京：高等教育出版社，2009.
- 作者简介：李峰（1986.04.25-），男，汉族，江苏省南京人，博士研究生，副教授，研究方向：物理学。
- 武红（1986.05.12-），女，汉族，江苏省南京市，博士研究生，副教授，研究方向：光学。
- 基金项目：南京邮电大学教学改革研究项目（重点），项目号JG00721JX14；大学物理情境教学法的研究与应用，项目号JG00719JX31。