

doi: 10.11823/j.issn.1674-5795.2022.02.05

一元线性回归分析方法在压力传感器数据拟合中的应用

李久龙, 李玲, 刘瑞敏, 孙旭, 薛炜, 许腾, 杨懿*

(北京航天试验技术研究所, 北京 100074)

摘要: 针对压力传感器在工程测量中需要根据计量检定数据获取压力传感器输出值和所加载标准压力值之间关系的问题, 采用正态检验和方差分析的方法分析了压力传感器输出电压值和标准压力值之间的特征关系。根据相关系数值的计算结果, 建立一元线性回归模型。分析了一元线性回归模型的拟合优度、模型线性关系、回归系数和模型的残差。结果表明: 一元线性回归分析方法能够应用于计量检定工作中的压力传感器检定数据拟合分析, 这对计量检定人员研究其他类似检定数据的拟合方法具有重要的参考价值。

关键词: 压力传感器; 检定数据; 一元线性回归

中图分类号: TB9; V438+.4 **文献标识码:** A **文章编号:** 1674-5795(2022)02-0040-10

Application of univariate linear regression analysis in data fitting of pressure sensor

LI Jiulong, LI Ling, LIU Ruimin, SUN Xu, XUE Wei, XU Teng, YANG Yi*

(Beijing Institute of Aerospace Test Technology, Beijing 100074, China)

Abstract: Aiming at the necessity to obtain the relationship between the output value of the pressure sensor and the loaded standard pressure value according to the metrological verification data when a pressure sensor is used in engineering measurement, the characteristic relationship between the output voltage of the pressure sensor and the standard pressure is analyzed by using the methods of normal test and variance analysis. According to the calculation results of correlation coefficient, a univariate linear regression model is established. The goodness of fit, linear relationship, regression coefficient and residual error of the univariate linear regression model are analyzed. The results show that the univariate linear regression analysis method can be applied to the fitting analysis of pressure sensor verification data in metrological verification, which has important reference value for metrological verification personnel to study other similar verification data fitting methods.

Key words: pressure sensor; verification data; univariate linear regression

0 引言

在工程实践中, 传感器测量技术是获取测量数据的重要手段, 其测量方式大致可分为线性测量和非线性测量两种^[1]。对测量数据进行分析,

多采用一元、多元线性回归分析等方法建立测量数据与物理量值之间的关系^[2-4]。

压力传感器测量技术多采用线性测量的方式。线性测量具有操作简单、测量准确度高等特点, 因而在工程中应用非常广泛。在进行参数测量前,

收稿日期: 2022-01-07; 修回日期: 2022-04-12

基金项目: 国家重点研发计划(2017YFB0603702); 国家自然科学基金(51976178)

引用格式: 李久龙, 李玲, 刘瑞敏, 等. 一元线性回归分析方法在压力传感器数据拟合中的应用 [J]. 计测技术, 2022, 42 (2): 40-49.

Citation: LI J L, LI L, LIU R M, et al. Application of univariate linear regression analysis in data fitting of pressure sensor [J]. Metrology and measurement technology, 2022, 42 (2): 40-49.



需要将压力传感器送计量检定机构进行检定,以获取输入压力值与输出电压值之间对应的函数关系。

在计量检定中,通常根据压力传感器的量程设定多个压力参考点,通过标准活塞压力计输出对应压力点的标准压力值获取对应的电压/电流信号,再通过数学方法拟合压力值(MPa)与电压值/电流(mV/mA)之间的数学关系。在工程应用中,通常采用线性回归分析方法拟合因变量与自变量之间的函数关系。多位研究人员对线性回归方法在生产实践中的应用进行了深入研究。

郑锦秀^[5]分析了电力系统测试中温度对大电流分流器测量的影响,采用一元线性回归方程拟合了大电流分流器阻值和温度的线性回归方程,通过显著性检验、相关系数等方法验证了回归方程的合理性。

许莉莉^[6]针对一元线性回归方法中自变量在进行正交变换过程中会影响运算结果的缺陷,借鉴动力学中转动惯量的对称性质,提出了以转动惯量为基础的线性回归分析方法。通过仿真实验和回归实例证明了该方法比传统一元线性回归准确度更高、稳定性更好。但是该方法在应用过程中需要获取对称观测值,即获取平面上任意样本点相对于平面内直线 $y=kx+b$ 的转动惯量,不适用于压力传感器输出数据的分析。

宫凯勋等人^[7]针对传统压阻式压力传感器仿真设计中,压敏电阻应力值的垂直变化以及非线性影响设计精度的问题,提出了多元线性回归法对 Z 坐标下的 X 、 Y 坐标与应力值之间的关系式进行多次拟合,获取压敏电阻的整体平均应力值,有效提高了传感器的设计精度。

王惠文等人^[8]针对传统线性回归方法中采用局部信息替代全部信息,无法充分有效利用原始数据信息的缺陷,将符号数据表的全部信息作为分析基础,建立了正态分布型数据矩阵模型和数字特征。通过仿真实验,验证了该方法的优越性。但是该方法在使用过程中仅仅针对正态分布型数据,在推广应用中具有一定的局限性。

刘军等人^[9]采用最小二乘法对不同温度下超声波热量的时差和流量数据进行拟合。通过软件仿真得到不同温度下的时差、温度与流量的关系。通过对比分析试验,证明了最小二乘法拟合方法

较传统的拟合方法具有拟合精度稳定、误差小的特点。

王海峰等人^[10]分析了冲击波超压测量数据处理流程,采用自适应滤波的方法消除测量数据中的噪声,以阶跃波形为准则完成测量数据的有效性检验,根据不同距离传感器数据拟合多项式曲线,数据有效性验证证明了该数据处理方法实现了冲击波超压测量数据的自动化处理,提高了数据处理的效率和准确度。

赵彦凯等人^[11]综合比较了普通最小二乘法和整体最小二乘法在电子测压器动态校准过程中数据拟合中的差异,建立了两种方法的回归模型,推导了整体最小二乘法的演算过程,从模型、参数估计和精度评定等方面对比分析了两种方法的优缺点,并指明了两种方法在电子测压器动态校准过程中数据拟合中的应用范围。

许浩源等人^[12]采用微压力传感器测量声表面波的压力数据,利用最小二乘法建立数学模型,分析、拟合所测频率值与对应的载荷力之间的关系。通过构建BP神经网络模型,训练已采集的数据样本,预测压力传感器的输入与输出之间的关系。遗传算法对BP神经网络优化结果显示该方法能有效减少误差。

结合压力传感器测量技术的特点、输出数据的特征和工程应用的实际情况,本文分析了压力传感器检定数据的特征,采用一元线性回归分析方法拟合压力值(MPa)与电压值(mV)之间的数学关系,并对拟合的结果进行分析。该方法对计量检定人员研究检定数据拟合方法具有重要的参考价值。

1 压力传感器检定系统和检定数据结构

1.1 检定系统

本研究的数据来源于国防三级计量标准——0.02级活塞压力计计量检定装置。压力传感器检定系统由0.02级活塞压力计(含砝码)、精密直流稳压电源、数字多用表、计算机等设备组成^[13]。

压力传感器检定系统组成如图1所示。

系统组成设备的功能:活塞压力计和专用砝码为被检压力传感器加载标准压力。精密直流稳压电源为被检压力传感器提供激励电压。数字多用表采集压力传感器输出的毫伏级电压信号,其

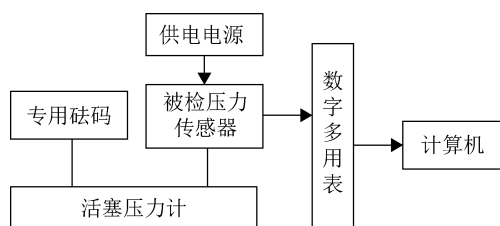


图1 压力传感器检定系统组成图

Fig.1 Composition diagram of pressure sensor verification system

输出端通过 RS-232 串口与采集计算机实现通讯。计算机完成数据的采集、存储、分析和处理。

1.2 检定数据结构

选取 1 支 YB-1 型号的压力传感器, 其量程为 12 MPa, 编号为 H0608, 最大综合基本误差为 0.1%。依照相应检定规程^[14]进行检定。检定过程分为三个进-回程。活塞压力计从 0, 2, 4, 6, 8, 10, 12 MPa 七个点加载标准压力称之为进程, 从 12, 10, 8, 6, 4, 2, 0 MPa 的卸压过程称之为回程。整个检定过程分为进程 1、回程 1、进程 2、回程 2、进程 3 和回程 3, 检定数据如表 1 所示。

表1 H0608 压力传感器检定数据表

Tab.1 H0608 pressure sensor verification data sheet

加载 压力值/MPa	输出电压值/mV					
	进程 1	回程 1	进程 2	回程 2	进程 3	回程 3
0.0000	0.1618	0.1697	0.1696	0.1707	0.1623	0.1713
2.0000	1.8262	1.8350	1.8364	1.8374	1.8268	1.8381
4.0000	3.4890	3.5014	3.5030	3.5043	3.4895	3.5048
6.0000	5.1596	5.1684	5.1696	5.1696	5.1599	5.1699
8.0000	6.8316	6.8360	6.8393	6.8390	6.8322	6.8395
10.0000	8.4994	8.5062	8.5045	8.5130	8.4999	8.5136
12.0000	10.1651	10.1660	10.1684	10.1710	10.1657	10.1716

2 检定数据分析

2.1 传感器输出电压值与标准压力值的特征分析

根据表 1 的检定数据, 以进程 1 的数据为例进行画图分析, 压力值与电压值散点图如图 2 所示。观察标准压力值与电压值之间对应关系可知, 二者基本呈线性关系。

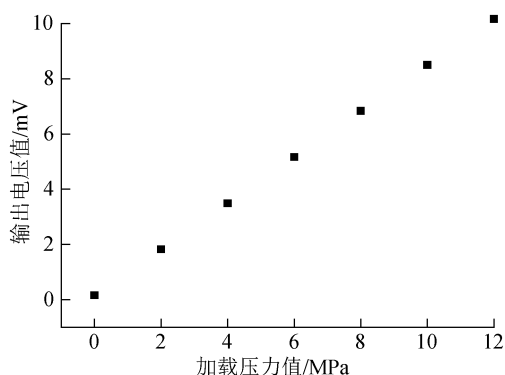


图2 进程 1 传感器压力值与输出电压值数据散点图

Fig.2 Scatter diagram of pressure and output voltage data of sensor in process 1

2.2 正态性检验

假定压力传感器电压输出的观测值来自正态分布总体的简单随机样本, 对所观测值进行正态性检验, 绘制变量的正态概率图。以进程 1 的数据为例, 运用 Origin 软件对检定数据进行正态检验。检验结果如图 3 所示。

采用相同的方法对其他进-回程的数据进行正态检验分析可知, 3 个进-回程压力传感器电压输出值均服从正态分布。

2.3 方差分析

由于压力传感器的电压输出值仅与所加载标准压力值相关, 故采用单因素方差分析法对数据的方差进行分析。单因素方差分析只考虑一个因素对观测数据的影响效应。

其基本原理^[15]为: 在水平 A_j ($j = 1, 2, \dots, 6$) 下, 进行 j 次独立试验。假定各个水平 A_j ($j = 1, 2, \dots, 6$) 下的样本来自具有相同方差 σ^2 , 均值分别为 μ_j ($j = 1, 2, \dots, 6$)。 μ_j 和 σ^2 未知, 且在不同水平下样本之间相互独立。设每次采集的数据为

Normality Test (2022/2/23 09:27:40)

NormalityTest

Shapiro-Wilk

	DF	Statistic	p-value	Decision at level(5%)
A	7	0.978	0.94929	Can't reject normality
B	7	0.97785	0.94846	Can't reject normality

A: At the 0.05 level, the data was significantly drawn from a normally distributed population.

B: At the 0.05 level, the data was significantly drawn from a normally distributed population.

Lilliefors

	DF	Statistic	p-value	Decision at level(5%)
A	7	0.10844	0.2	Can't reject normality
B	7	0.10859	0.2	Can't reject normality

A: At the 0.05 level, the data was significantly drawn from a normally distributed population.

B: At the 0.05 level, the data was significantly drawn from a normally distributed population.

Kolmogorov-Smirnov

	DF	Statistic	p-value	Decision at level(5%)
A	7	0.10844	1	Can't reject normality
B	7	0.10859	1	Can't reject normality

A: At the 0.05 level, the data was significantly drawn from a normally distributed population.

B: At the 0.05 level, the data was significantly drawn from a normally distributed population.

图3 进程1的正态性检验图

Fig.3 Normality test chart of process 1

X_{ij} , $X_{ij} \sim N(\mu_j, \sigma^2)$, 即有 $(X_{ij} - \mu_j) \sim N(0, \sigma^2)$ 。将 $(X_{ij} - \mu_j)$ 看作随机误差, 记 $X_{ij} - \mu_j = \varepsilon_{ij}$, 则可建立单因素试验方差分析的数学模型。

1) 单因素试验方差分析的数学模型

单因素试验方差分析的数学模型条件: ① $X_{ij} = \mu_j + \varepsilon_{ij}$, $i = 1, 2, \dots, n$, $j = 1, 2, \dots, s$; ② $\varepsilon_{ij} \sim N(0, \sigma^2)$, 各 ε_{ij} 独立; ③ $\sum_{j=1}^s n_j \delta_j = 0$ 。

假设所检验的数据符合数学模型的条件①~③, 则需要检验 s 个总体 $N(\mu_1, \sigma^2), \dots, N(\mu_s, \sigma^2)$ 的均值是否相等, 即检验假设 ($\alpha =$ 设定值, α 为观测值的附加效应, 即误差): $H_0: \mu_1 = \mu_2 = \dots = \mu_s$, $H_1: \mu_1, \mu_2, \dots, \mu_s$ 不全相等。

设 μ_j 的加权平均值 $\frac{1}{n} \sum_{j=1}^s n_j \mu_j = \mu$, $\delta_j = \mu_j - \mu$, $j = 1, 2, \dots, s$ 。其中, δ_j 为水平 A_j 下的总体评价与总平均的差异。此时有 $n_1 \delta_1 + n_2 \delta_2 + \dots + n_s \delta_s = 0$ 。

在实际运用中可按照表2中的方差来源进行方差分析。

S_A (A_j 水平下样本均值与数据总平均的差异), S_E (误差的平方和), S_T (总偏差的平方和) 的具体计算定义如式(1)~式(4)所示。

$$\bar{X} = \frac{1}{n} \sum_{j=1}^s \sum_{i=1}^{n_j} X_{ij} \quad (1)$$

表2 单因素试验方差法分析表

Tab.2 Variance analysis table of one-way test

方差来源	平方和	自由度
因素 A	S_A	$s - 1$
误差	S_E	$n - s$
总和	S_T	$n - 1$

$$S_T = \sum_{j=1}^s \sum_{i=1}^{n_j} X_{ij}^2 - n \bar{X}^2 \quad (2)$$

$$S_A = \sum_{j=1}^s n_j \bar{X}_j^2 - n \bar{X}^2 \quad (3)$$

$$S_E = S_T - S_A \quad (4)$$

式中: S_A 为 A_j 水平下样本均值与数据总平均的差异; S_E 为误差的平方和; S_T 为总偏差的平方和; X_{ij} 为每次采集的数据; j 为试验次数; s 为检验总体个数;

$\bar{X}_j$ 为 A_j 水平下的样本均值, $\bar{X}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} x_{ij}$ 。

2) 效应检验

在实际的检定过程中, 每一个进-回程分别对应 0~12 MPa 量程内 7 个采样点, 总计形成 $6 \times 7 = 42$ 个数据。受篇幅所限, 选取 2 MPa 压力点的数据作为效应检验对象。在标准活塞压力计加载 2 MPa 压力时, 对应每一个进-回程采样点, 在较短时间内连续快速采集 3 次数据, 详细数据如表 3 所示。

表3 2 MPa压力点压力传感器连续采集3次数据表
Tab.3 Data sheet of 3 times consecutive acquisition
by pressure sensor at 2 MPa

采集 次数	输出电压值/mV					
	进程1	回程1	进程2	回程2	进程3	回程3
1	1.8262	1.8350	1.8364	1.8374	1.8268	1.8381
2	1.8260	1.8351	1.8362	1.8371	1.8267	1.8384
3	1.8261	1.8349	1.8365	1.8375	1.8271	1.8379

假设加载压力值对电压输出的影响效应分别为 μ_1 (进程1), μ_2 (回程1), μ_3 (进程2), μ_4 (回程2), μ_5 (进程3), μ_6 (回程3)。数据符合数学模型的条件, 检验假设 ($\alpha = 0.05$)^[16]:

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6$ (进-回程对电压输出值没有显著影响);

$H_1: \mu_1, \mu_2, \mu_3, \mu_4, \mu_5, \mu_6$ 不全相等 (进-回程对电压输出值有显著影响)。

根据式 (1) ~ 式 (5) 对表4中数据进行计算, 设 X_{ij} 为每次采集的电压值。 $s=6, j=3, n = s \times j=18$ 。计算结果见表4。

表4 2 MPa压力点电压输出数据方差计算表
Tab.4 Variance calculation of output voltage data at
2 MPa

进/回程	输出电压值/mV			样本总 和/MPa	样本均 值/MPa
	第1次 加载	第2次 加载	第3次 加载		
进程1	1.8262	1.8260	1.8261	5.4783	1.8261
回程1	1.8350	1.8351	1.8349	5.5050	1.8350
进程2	1.8364	1.8362	1.8365	5.5091	1.8364
回程2	1.8374	1.8371	1.8375	5.5120	1.8373
进程3	1.8268	1.8267	1.8271	5.4806	1.8269
回程3	1.8381	1.8384	1.8379	5.5144	1.8381

根据式 (2) ~ 式 (4) 可得 $S_T = 0.00043584$; $S_A = 0.00043545$; $S_E = S_T - S_A = 0.00000039$ 。 S_T, S_A, S_E 的自由度依次为 17, 5, 12, 方差分析见表5。

$F_{0.05}(5, 12) = 3.11 < 2702.8$, 故在水平 0.05 下拒绝 H_0 , 认为在加载 2 MPa 标准压力下, 在不同进-回程中, 压力传感器的输出电压有显著的差异。

表5 2 MPa压力传感器输出电压方差分析表
Tab.5 Variance analysis of output voltage of pressure
sensor at 2 MPa

方差来源	平方和	自由度	均方	F比
因素	0.00043545	5	0.000087091	2702.8
误差	0.00000039	12	0.000000032	
总和	0.00043584	17	/	/

注: 根据所提出的假设检验, 设定相应的拒绝域, 即 F 比, 其形式为 $F = \frac{S_A(s-1)}{S_E(n-s)} \geq k, k$ 由预先给定的显著水平 α 确定。

3 一元线性回归分析

一元线性回归分析主要用于研究单个因变量和自变量之间的线性关系, 其重点是考察一个特定因变量, 将自变量看作影响这一变量的因素, 通过建立适当的数学模型将变量之间的关系合理、准确地表达出来, 进而通过自变量的取值来预测因变量的取值。因此, 在实际应用中, 需要针对选定的数据, 分析因变量与自变量之间的关系, 通过建立回归模型, 并分析该模型的误差, 以实现因变量与自变量之间关系的准确描述。

由图2的分析可知, 观测点基本分布在一条直线周围, 压力值增加, 其电压输出值随之增加, 电压值与压力值呈线性相关关系, 因此可以对其进行一元线性回归分析。

3.1 相关系数计算

选取表1中回程1的数据计算相关系数 r 。计算结果为

$$r = \frac{\sum (y - \bar{y})(x - \bar{x})}{\sqrt{\sum (y - \bar{y})^2 \sum (x - \bar{x})^2}} = 1$$

相关系数 $r = 1$ 表明加载压力值与输出电压值之间是完全正线性相关关系。

3.2 一元线性回归模型

设压力值 y 关于电压值 x 的回归函数为 $\mu(x)$, 利用样本来估计 $\mu(x)$ 。假设对于电压值 x (在某个区间内) 的每一个值都有

$$y \sim N(a + bx, \sigma^2) \quad (5)$$

式中: a, b, σ^2 为不依赖于 x 的未知参数。记 $\varepsilon = y - (a + bx)$, 对压力值 y 作正态假设, 其一元回归模型为 $y = a + bx + \varepsilon, \varepsilon \sim N(0, \sigma^2)$ 。

根据正规方程组^[17]

$$na + (\sum_{i=1}^n x_i)b = \sum_{i=1}^n y_i \quad (6)$$

$$(\sum_{i=1}^n x_i)a + (\sum_{i=1}^n x_i)b = \sum_{i=1}^n x_i y_i \quad (7)$$

有唯一解, 求解得到 a, b 的最大似然估计值 $\hat{a}, \hat{b}$ 分别为

$$\hat{b} = \frac{n \sum_{i=1}^n x_i y_i - (\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (8)$$

$$\hat{a} = \frac{1}{n} \sum_{i=1}^n y_i - \hat{b} \frac{1}{n} \sum_{i=1}^n x_i = \bar{y} - \hat{b} \bar{x} \quad (9)$$

式中: $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$; $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ 。

对于给定的 x , 取 $\hat{a} + \hat{b}x$ 作为回归函数 $\mu(x) = a + bx$ 的估计, 即 $\mu(x) = \hat{a} + \hat{b}x$ 。 $\hat{a}, \hat{b}$ 由式 (10) ~ 式 (14) 计算。

$$S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - \frac{1}{n} (\sum_{i=1}^n x_i)^2 \quad (10)$$

$$S_{yy} = \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n y_i^2 - \frac{1}{n} (\sum_{i=1}^n y_i)^2 \quad (11)$$

$$S_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n x_i y_i - \frac{1}{n} (\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i) \quad (12)$$

$$\hat{a} = \frac{1}{n} \sum_{i=1}^n y_i - \hat{b} \frac{1}{n} \sum_{i=1}^n x_i \quad (13)$$

$$\hat{b} = \frac{S_{xy}}{S_{xx}} \quad (14)$$

以进程 1 的 7 个采样点数据为例进行分析计算, 详细数据见表 6。

表 6 压力传感器进程 1 检定数据表

Tab.6 Verification data of pressure sensor in process 1

采样点数	加载压力值 y_i/MPa	输出电压值 x_i/mV
1	0.0000	0.1618
2	2.0000	1.8262
3	4.0000	3.4890
4	6.0000	5.1596
5	8.0000	6.8316
6	10.0000	8.4994
7	12.0000	10.1651

根据式 (10) ~ 式 (14) 计算, 结果为

$$S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - \frac{1}{n} (\sum_{i=1}^n x_i)^2 = 77.885309$$

$$S_{xy} = \sum_{i=1}^n x_i y_i - \frac{1}{n} (\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i) = 93.397800$$

$$S_{yy} = \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n y_i^2 - \frac{1}{n} (\sum_{i=1}^n y_i)^2 = 112$$

故得到

$$\hat{b} = \frac{S_{xy}}{S_{xx}} = 1.199171$$

$$\hat{a} = -0.189898$$

得到回归直线方程为

$$\hat{y} = -0.189898 + 1.199171x \quad (15)$$

按照同样的方法分别计算回程 1、进程 2、回程 2、进程 3 和回程 3 的回归直线方程, 截距 a 和斜率 b 的估计值 $\hat{a}, \hat{b}$ 的计算结果如表 7 所示。

对 6 个不同的 $\hat{a}, \hat{b}$ 值取平均, 得到所有进程数据拟合的一元线性回归方程为

$$\hat{y} = -0.199255 + 1.199545x \quad (16)$$

表 7 压力传感器检定数据拟合回归系数表

Tab.7 Fitting regression coefficient of pressure sensor verification data

拟合回归系数	系数值						
	进程 1	回程 1	进程 2	回程 2	进程 3	回程 3	平均值
$\widehat{a}$	-0.189898	-0.202848	-0.203839	-0.203890	-0.190488	-0.204568	-0.199255
$\widehat{b}$	1.199171	1.200019	1.199942	1.199482	1.199166	1.199487	1.199545

3.3 一元线性回归模型的拟合优度计算

回归直线 $\hat{y} = \hat{a} + \hat{b}x$ 在一定程度上可以描述变量 x 与 y 的关系, 将回归直线与观测点的接近程

度称为拟合优度, 将决定系数作为回归方差拟合优度的衡量标准。

在统计学中, 一个具体观测值的波动称之为

变差。变差大小可用实际观测值 y 与其均值 $\bar{y}$ 之差来表示。 n 次观测值的总变差(SST)大小可用观测值的离差,即 $(y - \bar{y})$ 的平方和来表示^[13]。即

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

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \quad (18)$$

总变差(SST)可分解为回归平方和(SSR)和与残差平方和(SSE)的加总。即

$$\text{总变差}(SST) = \text{回归平方和}(SSR) + \text{残差平方和}(SSE) \quad (19)$$

回归平方和(SSR)与总变差(SST)的比例称为决定系数,计算公式为

$$R^2 = \frac{SSR}{SST} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (20)$$

以回程2的数据为例,进行一元线性回归模型的拟合优度计算,数据计算结果为

$$R^2 = \frac{SSR}{SST} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = 0.999999$$

估计标准误差是残差的平方和的均方根,用 s_e 表示,计算公式为

$$s_e = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-2}} = \sqrt{\frac{SSE}{n-2}} = 0.001752$$

从估计标准误差 $s_e = 0.001752$ 可知,观测值与回归估计值之间的差异程度非常小。从一元线性回归方程的拟合优度值 $R^2 = 0.999999$ 可知,回归直线与各观测点几乎重合,该回归直线是拟合数据较为理想的拟合直线。

3.4 一元线性回归模型线性关系检验

在建立回归模型之前,假定压力传感器所测压力值和输出电压值之间是线性关系,但是该假定是否成立需要进行检验。上一节拟合了一元线性回归方程,求出了系数 $\hat{a}$, $\hat{b}$ 。为了检验 x, y 之间关系的显著性,采取数理统计方法进行验证。将回归平方(SSR)除以与残差平方和(SSE)的结果称为回归均方,记为 MSR , SSE 除以相应自由度的结果称为残差均方,记为 MSE ^[15]。提出假设

$$H_0: b = 0 \quad (x, y \text{ 之间的线性关系不显著})$$

$$H_1: b \neq 0 \quad (x, y \text{ 之间的线性关系显著})$$

$$F = \frac{SSR/1}{SSE/(n-2)} = \frac{MSR}{MSE} \sim F(1, n-2) \quad (21)$$

若假设 $H_0: b = 0$ (x, y 之间的线性关系不显著)成立, $\frac{MSR}{MSE}$ 的值应接近1。若原假设不成立,则 $\frac{MSR}{MSE}$ 的值趋近无穷大。

以回程2的数据为例,计算结果为 $SSR = 111.999985$, $SSE = 0.001752$, $F = \frac{SSR/1}{SSE/(n-2)} = 319559.556751$,远大于1。因此,假设 H_0 不成立,拒绝 H_0 ,即 x, y 之间的线性关系显著。

3.5 一元线性回归模型回归系数检验

回归系数检验的目的是检验自变量对因变量的影响是否显著。统计学中通常采用 t 检验法检验自变量对因变量的影响的显著性^[13]。在一元线性回归中,由于自变量只有一个,故回归系数的检验等价于线性关系的检验。检验假设为

$$H_0: b = 0 \quad (\text{自变量 } x \text{ 对因变量 } y \text{ 的影响不显著})$$

$$H_1: b \neq 0 \quad (\text{自变量 } x \text{ 对因变量 } y \text{ 的影响显著})$$

引入标准差估计量 $s_{\hat{b}}$

$$s_{\hat{b}} = \frac{s_e}{\sqrt{\sum_{i=1}^n x_i^2 - \frac{1}{n} (\sum_{i=1}^n x_i)^2}} \quad (22)$$

将回归系数标准化得到用于检验回归系数 $\hat{b}$ 的统计量 t 。若假设成立, $\hat{b} - b = \hat{b}$, $\hat{b} \sim N(b, \sigma^2/S_{xx})$, $\hat{\sigma}^2 = (S_{yy} - \hat{b}S_{xy})/(n-2)$ 。 S_{xx} , S_{xy} 的定义参照式(10)和式(12)。

当 H_0 为真,即 $b = 0$ 时

$$t = \frac{\hat{b}}{\hat{\sigma}} \sqrt{S_{xx}} \sim t(n-2) \quad (23)$$

设 α 为显著水平,得到 H_0 的拒绝域 $|t|$ 为

$$|t| = \frac{|\hat{b}|}{\hat{\sigma}} \sqrt{S_{xx}} \geq t_{\alpha/2}(n-2) \quad (24)$$

以回程2的数据为例,计算结果为

$$\hat{b} = 1.199482, S_{xx} = 77.844959, \hat{\sigma}^2 = 0.008696$$

查表得

$$t_{0.05/2}(n-2) = t_{0.05/2}(5) = 2.5706$$

$|t| = 113.4878 > 2.5706$,故拒绝 $H_0: b = 0$,认为

回归效果是显著的, 即自变量 x 对因变量 y 的影响显著。

同理, 可以运用相同方法对其他数据进行显著性检验分析。

3.6 一元线性回归模型的残差分析

在一元回归模型 $y = a + bx + \varepsilon$ 中, 假定 ε 是期望值为 0, 方差相等且服从正态分布的一个独立随机变量^[15]。但若此假定不成立, 则所做检验以及预测则均不准确。因此, 需要对 ε 进行残差分析。因变量的观测值 y_i 与根据估计的回归方程求出的预测值 $\hat{y}_i$ 之差即为残差。第 i 个观测值的残差数学定义为

$$e_i = y_i - \hat{y}_i \quad (25)$$

3.6.1 残差图检验

可通过残差图来分析检验误差项 ε 的假设是否成立。通过一元线性回归公式 (17) 计算表 1 中 7 个进、回程传感器输出电压值的残差, 数据的残差图如图 4 所示。

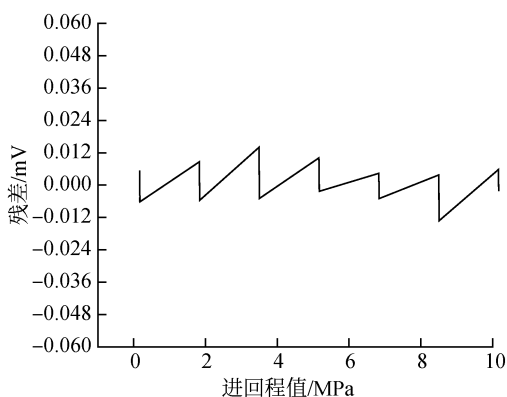


图4 检定数据残差图

Fig.4 Residual diagram of verification data

根据图 4 可知, 各残差基本位于一条水平带中间, 基本无固定模式, 表明在压力传感器压力值与电压输出值的一元线性回归中, 线性的假定以及对 ε 的假定是成立的。

3.6.2 正态性检验

通过标准化残差分析对 ε 正态假定进行检验。标准化残差用 z_e 表示, 第 i 个观测值的标准化残差定义^[13] 为

$$z_{e_i} = \frac{e_i}{s_e} = \frac{y_i - \hat{y}_i}{s_e} \quad (26)$$

若关于 ε 正态假定成立, 则标准化残差的分布也服从正态分布, 且大约有 95% 的标准化残差落在 $-2 \sim +2$ 区间。

计算表 1 中各数据点的标准化残差, 计算数据如表 8 所示。

表8 压力传感器检定数据标准化残差表

Tab.8 Standardized residual of pressure sensor verification data

采样 点数	输出电压值/mV					
	进程 1	回程 1	进程 2	回程 2	进程 3	回程 3
1	0.5342	-1.2009	-1.0957	-0.7751	0.5053	-0.7996
2	0.8937	-0.5325	-0.9391	-0.6740	0.8767	-0.7228
3	1.4515	-0.2319	-0.7198	-0.6067	1.4868	-0.6305
4	1.0422	-0.1320	-0.5004	-0.2692	1.0754	-0.2918
5	0.4593	-0.2327	-1.2540	-0.6239	0.4119	-0.6463
6	0.3971	-1.2028	-0.5953	-1.7552	0.3586	-1.6939
7	0.5954	1.3048	0.4714	-0.1854	0.5575	-0.2616

从表 8 中标准化残差的计算结果可知, 标准化残差均分布在 $-2 \sim +2$ 区间内, 表明关于 ε 服从正态分布的假定成立。

通过本章节的分析可知, 一元线性回归分析的主要步骤:

1) 确定变量之间的关系

可采用散点图, 计算、检验相关系数的方法确定自变量与因变量之间的关系。

2) 建立一元线性回归模型

一般采用最小二乘法确定回归方程及其系数。

3) 一元线性回归模型的拟合优度计算

一般选用决定系数作为拟合优度的重要统计量。

4) 一元线性回归模型线性关系和系数检验

采用数理统计 F 检验和 t 检验方法验证线性关系和系数的显著性。

5) 一元线性回归模型的残差分析

采用残差图检验和正态性检验的方法验证残差 ε 正态假定是否成立。

4 总结

从压力传感器的实际应用出发, 分析了压力传感器计量检定数据的特征, 采用一元线性回归

分析方法拟合压力传感器检定数据, 建立输出值(mV)和加载标准压力值(MPa)之间的数学函数关系。通过正态检验、方差分析、回归模型拟合优度分析、模型线性关系分析、回归系数和模型残差分析等方法验证了一元线性回归分析方法能够应用于压力传感器检定数据的拟合工作中。对业内计量检定人员拟合检定数据具有重要的参考价值。

参考文献

- [1] 胡斯, 王国富, 叶金才, 等. 基于非线性温度传感器的高精度测量方案设计[J]. 仪表技术与传感器, 2017, 5: 112-116.
HU S, WANG G F, YE J C, et al. Design of high precision measurement scheme based on nonlinear temperature sensor[J]. Instrument technique and sensor, 2017, 5: 112-116. (in Chinese)
- [2] 董素君, 王凯, 高红霞, 等. 板翅式换热器换热效能三元线性回归模型及其系数辨识[J]. 航空学报, 2012, 33(9): 1571-1577.
DONG S J, WANG K, GAO H X, et al. A ternary linear regression model and its coefficients identification of heat transfer efficiency for plate-fin heat exchanger[J]. Journal of aeronautics, 2012, 33(9): 1571-1577. (in Chinese)
- [3] 林天水, 陈佩树. 一元线性回归中异方差的处理[J]. 统计与决策, 2015, 21: 86-88.
LIN T S, CHEN P S. Treatment of heteroscedasticity in univariate linear regression[J]. Statistics and decision making, 2015, 21: 86-88. (in Chinese)
- [4] 段新胜, 林清龙, 毛汉川, 等. 岩土层原位导热系数的多元线性回归分析方法[J]. 太阳能学报, 2018, 39(2): 385-389.
DUAN X S, LIN Q L, MAO H C, et al. Multiple linear regression analysis method for in-situ thermal conductivities of rock and soil layer[J]. Acta energiae solaris sinica, 2018, 39(2): 385-389. (in Chinese)
- [5] 郑锦秀, 童欣. 一元线性回归方程在大电流分流器测量中的应用[J]. 计测技术, 2009, 29(5): 17-19.
ZHEN J X, TONG X. Application of unitary linear regression equation for measuring this great diffuent utensil[J]. Metrology and measurement technology, 2009, 29(5): 17-19. (in Chinese)
- [6] 许莉莉. 基于转动惯量的一元线性回归方法[J]. 统计与决策, 2015, 16(5): 33-35.
XU L L. Univariate linear regression method based on moment of inertia[J]. Statistics and decision making, 2015, 16(5): 33-35. (in Chinese)
- [7] 宫凯勋, 雷程, 董志超, 等. 基于最小二乘法数据拟合的微小压力传感器仿真设计[J]. 传感技术学报, 2021, 34(10): 1320-1325.
GONG K X, LEI C, DONG Z H, et al. Simulation design of tiny pressure sensor based on least squares data fitting[J]. Journal of sensing technology, 2021, 34(10): 1320-1325. (in Chinese)
- [8] 王惠文, 李楠. 基于全信息的正态分布型数据的线性回归分析[J]. 北京航空航天大学学报, 2012, 38(10): 1275-1279.
WANG H W, LI N. Linear regression analysis of normal distribution data based on full information[J]. Journal of Beijing university of aeronautics and astronautics, 2012, 38(10): 1275-1279. (in Chinese)
- [9] 刘军, 崔学伟, 田甜. 最小二乘法在流量计算中的应用[J]. 电子测量技术, 2018, 41(21): 35-41.
LIU J, CUI X W, TIAN T, et al. Application of least square method in flow calculation[J]. Electronic measurement technology, 2018, 41(21): 35-41. (in Chinese)
- [10] 王海峰, 禄晓飞, 涂国勇, 等. 冲击波超压测量数据处理方法研究[J]. 现代防御技术, 2020, 48(2): 69-74.
WANG H F, LU X F, TU G Y, et al. Study on data processing method for shock wave overpressure measurement[J]. Modern defence technology, 2020, 48(2): 69-74. (in Chinese)
- [11] 赵彦凯, 裴东兴, 石垒, 等. 嵌入式电子测压器应用环境下动态校准数据拟合方法研究[J]. 火炮发射与控制学报, 2021, 42(1): 78-83.
ZHAO Y K, PEI D X, SHI L, et al. Research on fitting method of dynamic calibration data in the application environment of internal electronic pressure gauge[J]. Journal of gun launch and control, 2021, 42(1): 78-83. (in Chinese)
- [12] 许浩源, 李媛媛. GA-BP神经网络对SAW压力传感器测量数据的拟合[J]. 电子测量与仪器学报, 2021, 35(4): 7-14.
XU H Y, LI Y Y. Fitting analysis of SAW micro pressure sensor measurement data by GA optimized BP neural net-

- work[J]. Journal of electronic measurement and instrumentation, 2021, 35(4): 7-14. (in Chinese)
- [13] 杨懿,陈文丽,贾志杰,等. 降低发动机试验稳态压力测量不确定度方法研究[J]. 计测技术, 2020, 40(5): 15-21.
- YANG Y, CHEN W L, JIA Z J, et al. Improvement of steady state pressure measurement method for engine test [J]. Metrology and measurement technology, 2020, 40 (5): 15-21. (in Chinese)
- [14] 压力传感器检定规程(静态):JJG 860-2015[S]. 北京:中国计量出版社, 2015.
- Verification regulation of pressure sensor (static) : JJG 860-2015 [S]. Beijing: China Metrology Press, 2015. (in Chinese)
- [15] 贾俊平. 统计学[M]. 北京:中国人民大学出版社, 2019.
- JIA J P. Statistics [M]. Beijing: China Renmin University Press, 2019. (in Chinese)
- [16] 杨建. 定量分析方法[M]. 北京:清华大学出版社, 2018.
- YANG J. Quantitative analysis method [M]. Beijing: Tsinghua University Press, 2018. (in Chinese)
- [17] 盛骤,谢式千,潘承毅. 概率论与数理统计[M]. 北京:高等教育出版社, 2008.
- SHENG C, XIE S Q, PAN C Y. Probability theory and mathematical statistics [M]. Beijing: Higher Education Press, 2008. (in Chinese)

(本文编辑:朱俊真)

第一作者:李久龙(1983—),男,山西朔州人,高级工程师,主要从事大型航天技术应用产业项目运营与管理等方面的研究。

通讯作者:杨懿(1984—),男,湖南岳阳人,高级工程师,主要从事航天技术应用产业项目运营与管理等方面的研究。