

基于 SPSS 回归分析的颜色与物质浓度辨识模型

陈 溥 杨 颢

(柳州铁道职业技术学院, 广西 柳州 545616)

摘 要 本文根据 2017 年高教社杯全国大学生数学建模竞赛 C 题中提供的物质在不同浓度下的颜色读数采样数据, 运用 SPSS 软件对数据进行多元线性回归分析, 得到了各种物质颜色读数与物质浓度的数学模型。对所得模型进行误差分析发现, 只有组胺的模型达到精度要求, 其它四种物质的模型误差均偏大, 且模型失效。对失效的模型, 考虑更优的建模方法, 计算四种物质数据中五个影响指标两两之间的相关性系数, 选取两两之间相关性系数最大的三对指标进行重新组合, 得到一组新的数据, 之后进行重新建模, 最终得到了各物质颜色读数与浓度优化后的非线性数学模型。

关键词 SPSS 回归分析 数学模型

中图分类号: TP311.54

文献标识码: A

文章编号: 1007-0745(2021)06-0024-04

1 前言

世界是多姿多彩、五彩缤纷、光鲜亮丽的, 我们在日常生活中经常会看见五颜六色的事物, 这时便会有朋友好奇颜色为什么会有这么多, 即使是相同的颜色, 又为什么会有深浅, 大多数人都会说是浓度不同所导致的, 的确深浅在我们看来, 就是浓淡, 亦是物质的浓度, 那么如何根据颜色的深浅来确定物质的浓度呢, 这使得研究颜色读数与浓度之间的关系就变得越来越重要。

比色法是目前常用的一种检测物质浓度的方法, 即把待测物质制备成溶液后滴在特定的白色试纸表面, 等其充分反应以后获得一张有颜色的试纸, 再把该颜色试纸与一个标准比色卡进行对比, 就可以确定待测物质的浓度档位了。由于每个人对颜色的敏感差异和观测误差, 使得这一方法在精度上受到很大影响。随着照相技术和颜色分辨率的提高, 希望建立颜色读数和物质浓度的数量关系, 即只要输入照片中的颜色读数就能够获得待测物质的浓度。

2 模型的建立与求解

首先将附件中的采样数据导入 SPSS, 其次在 SPSS 软件中对各组数据进行多元线性回归分析, 得到各物质颜色读数与物质浓度的模型如下: (模型中的符号说明见下表)

x_1	蓝色颜色值 (B)	y_1	组胺浓度值
x_2	绿色颜色值 (G)	y_2	溴酸钾浓度值
x_3	红色颜色值 (R)	y_3	工业碱浓度值
x_4	色调 (H)	y_4	硫酸铝钾浓度值
x_5	饱和度 (S)	y_5	奶中尿素浓度值

组胺的模型:

$$y_1 = -212.765 + 2.855x_1 - 4.487x_2 + 2.321x_3 + 4.593x_4 + 1.142x_5$$

溴酸钾的模型:

$$y_2 = -859.900 - 1.356x_1 + 2.074x_2 - 5.180x_3 + 10.237x_4$$

工业碱的模型:

$$y_3 = 261.065 + 0.164x_1 - 1.398x_2 - 0.314x_3 - 0.131x_4 - 0.880x_5$$

硫酸铝钾的模型:

$$y_4 = 17.537 + 0.053x_1 - 0.018x_2 - 0.102x_3 - 0.134x_4 - 0.019x_5$$

奶中尿素的模型:

$$y_5 = 12221.204 + 280.116x_1 + 495.160x_2 - 811.290x_3 - 365.931x_4 + 251.102x_5$$

3 模型的误差分析

根据以上所得各物质的颜色读数与浓度的模型, 将采样数据中各颜色读数代入模型, 求得所对应的各物质的浓度模型值 y_i , 之后与所对应的浓度采样值 Y_i 进行比较, 代入下面平均相对误差的计算公式:

$$MRE = \frac{1}{n} \sum_{i=1}^n \left| \frac{Y_i - y_i}{Y_i} \right| \times 100\%$$

计算得到各物质模型的平均相对误差如下表:

物质	组胺	溴酸钾	工业碱	硫酸铝钾	奶中尿素
MRE	3%	15%	14%	66%	19%

通过表格中的数据可以看出, 只有组胺模型的平均相对误差小于 10%, 在可以接受的范围内, 另外四种物质模型的平均相对误差较大, 超出了可接受范围, 模型失效。^[1]

4 模型的改进

对于四种物质的失效模型, 考虑更优的建模方法, 运用 SPSS 软件计算四种物质采样数据中五个影响指标 (B、G、R、H、S) 两两之间的相关性系数, 得到的相关性系数

★基金项目: 柳州铁道职业技术学院校级项目, 项目编号: 2021-JGB01。

表1 溴酸钾

颜色	B	G	R	H	S
B		0.898	0.073	0.846	1.000
G	0.898		0.456	0.871	0.889
R	0.073	0.456		0.224	0.053
H	0.846	0.871	0.224		0.844
S	1.000	0.889	0.053	0.844	

表2 工业碱

颜色	B	G	R	H	S
B		0.08	0.22	0.671	0.946
G	0.08		0.886	0.265	0.15
R	0.22	0.886		0.12	0.043
H	0.671	0.265	0.12		0.652
S	0.946	0.15	0.043	0.652	

表3 硫酸铝钾

颜色	B	G	R	H	S
B		0.837	0.958	0.930	0.970
G	0.837		0.921	0.814	0.899
R	0.958	0.921		0.945	0.998
H	0.930	0.814	0.945		0.951
S	0.970	0.899	0.998	0.951	

表4 奶中尿素

颜色	B	G	R	H	S
B		0.834	0.862	0.710	0.832
G	0.834		0.874	0.973	1.000
R	0.862	0.874		0.839	0.88
H	0.710	0.973	0.839		0.972
S	0.832	1.000	0.88	0.972	

表5 溴酸钾

浓度	BG	BS	GS	RR	SS
0	18189	3483	3807	21025	729
100	931	1687	32053	21025	58081
50	7980	8700	19285	19881	21025
25	9384	9177	18088	21025	17689
12.5	11815	9010	14734	21025	11236
0	18048	3584	3948	20736	784
100	931	1694	32186	21025	58564
50	7581	8607	20083	19881	22801
25	9590	9240	18084	21316	17424
12.5	12006	8874	14076	21316	10404

表6 工业碱

浓度	BB	GH	RR	GS	HS
7.34	23409	15120	17424	4900	3780
8.14	22801	14768	17689	4118	3016
8.74	24964	15120	16129	6552	6240
9.19	25921	11220	13924	10200	15840
10.18	16129	3087	14161	4431	31017
11.8	8836	888	8281	1422	35076
0	23104	14910	17424	4544	3360

表7 硫酸化钾

浓度	BS	RS	HS	GG	HH
0	5104	4576	3344	15876	5776
0	5130	4680	3330	15876	5476
0	4720	4200	3120	15625	6084
0	4746	4326	3066	15376	5329
0	4446	4056	2925	15376	5625
0	5085	4680	3240	15876	5184
0.5	25752	8178	17400	12544	10000
0.5	26700	7832	17800	12321	10000
0.5	16974	8733	12054	13924	9604
0.5	16592	8540	11956	13924	9604
0.5	18224	8576	13132	13689	9604
0.5	18360	8640	13095	13924	9409
0.5	22218	8050	15939	12321	9801
1	25628	8256	17028	13456	9801
1	25650	8379	17100	13225	10000
1	23373	8745	15741	14161	9801
1	21605	9280	14500	14161	10000
1	21840	8424	15444	12769	9801
1	21920	8160	15840	12321	9801
1.5	27540	7920	18180	12769	10201
1.5	28152	7728	18400	12769	10000
1.5	26163	8550	17271	13225	10201
1.5	26928	8272	17600	13225	10000
1.5	25384	8684	16700	13456	10000
1.5	26163	8379	17100	13456	10000
2	31044	6766	20298	11236	10404
2	31752	7252	20188	11449	10609
2	30590	7600	19380	12100	10404
2	31622	7372	19788	12321	10404
2	31482	6930	20394	10816	10609
2	31284	6930	20394	11025	10609
5	30690	6732	19998	11449	10201
5	30888	6732	19998	11664	10201
5	26448	8352	17400	13456	10000
5	25368	8568	16800	13225	10000
5	30646	6567	20298	11025	10404
5	30732	6895	20094	11025	10404

表8 奶中尿酸

浓度	BH	BS	GR	BB	GG
0	2950	4366	18904	13924	18496
500	3159	4797	19043	13689	18769
1000	3024	5832	18768	11664	18496
1500	2860	5720	18904	12100	18496
2000	3024	6480	19880	11664	19600
0	3120	3960	18768	14400	18496
5	3094	4760	19880	14161	19600
500	2997	6105	19738	12321	19321
1500	2782	6206	18904	11449	18496
2000	2940	6090	18632	11025	18496
0	2500	3375	18900	15625	18225
500	2850	5016	18492	12996	17956
1000	3024	4704	17688	12544	17424
1500	2730	6300	18492	11025	17956
2000	2782	6099	18630	11449	18225

表9

物质	溴酸钾	工业碱	硫酸铝钾	奶中尿素
MRE	3%	9.58%	9%	7%

结果如上表1~4。

在各种物质中,选取五个影响指标两两之间相关性系数最大的三对指标进行两两组合(对应的指标数据两两之间做乘法运算,得到新的组合数据),两两之间相关性系数最小的指标不进行组合(对应的指标数据取各自的平方,得到新的数据),从而各物质均得到一组新的五个组合指标,并且得到其所对应的数据如表5~8。

将数据导入SPSS软件,并对各物质新组合的数据进行多元线性回归分析,得到了各物质下颜色读数与物质浓度的模型如下:

溴酸钾的模型:

$$y_2 = 413.977 - 0.015x_1x_2 - 0.09x_1x_3 - 0.05x_3^2 - 0.03x_5^2$$

工业碱的模型:

$$y_3 = -35.338 + 0.02x_1 + 0.01x_2x_4 - 0.01x_3^2 - 0.03x_2x_5 + 0.01x_4x_5$$

硫酸化钾的模型:

$$y_4 = -12.578 - 0.00033x_1x_5 - 0.001x_3x_5 + 0.001x_4x_5 + 0.001x_2^2 + 0.00036x_4^2$$

奶中尿酸的模型:

$$y_5 = 7569.571 - 3.404x_1x_4 + 2.019x_1x_5 - 5.845x_2x_3 + 1.015x_1^2 + 4.867x_2^2$$

根据以上所得的非线性数学模型,将采样数据中各颜色读数代入上述模型,求得所对应的各物质的浓度模型值 y_i ,之后与所对应的浓度采样值 Y_i 进行比较,代入下面平均相对误差的计算公式:

$$MRE = \frac{1}{n} \sum_{i=1}^n \left| \frac{Y_i - y_i}{Y_i} \right| \times 100\%$$

计算得到以上四种物质模型的平均相对误差如上表9所示。

从表中可以看出,四种物质模型的误差均小于10%,由此可见,改进后的非线性模型与之前由多元线性回归分析所得的模型对比,有了很大的优化。^[2]

5 结语

本文运用SPSS软件对采样数据进行了多元线性回归分析,建立了颜色读数和物质浓度的数学模型,并对所建立的模型进行了误差分析和改进,得到了优化后的非线性数学模型,根据颜色读数来预测物质浓度,在实际应用中有一定的参考价值。

参考文献:

- [1] 张文彤.SPSS统计分析基础教程[M].北京:高等教育出版社,2017.
- [2] 武松,潘发明.SPSS统计分析大全[M].北京:清华大学出版社,2014.