

颜色读数辨识物质浓度

符号说明

符号	含义
Y	物质浓度
ppm	mg/L
B	蓝色颜色值
G	绿色颜色值
R	红色颜色值
H	色调
S	饱和度

问题一

1. 数据预处理

对 Data1 的数据进行分离，同时修正错误数据，以组胺的数据为例：

组胺	B	G	R	H	S
0	68	110	121	23	111
100	37	66	110	12	169
50	46	87	117	16	155
25	62	99	120	19	122
12.5	66	102	118	20	112
0	65	110	120	24	115
100	35	64	109	11	172
50	46	87	118	16	153
25	60	99	120	19	126
12.5	64	101	118	20	115

最终得到组胺、溴酸钾、工业碱、硫酸铝钾、奶中尿素五种物质的浓度—颜色读数表。

2. 多元线性回归

假设物质浓度与五种颜色读数存在一定的函数关系：

$$Y = f(B, G, R, H, S)$$

对每种物质分别建立多元线性回归模型，源代码见附录。

$$Y = \beta_1 B + \beta_2 G + \beta_3 R + \beta_4 H + \beta_5 S + \beta_0$$

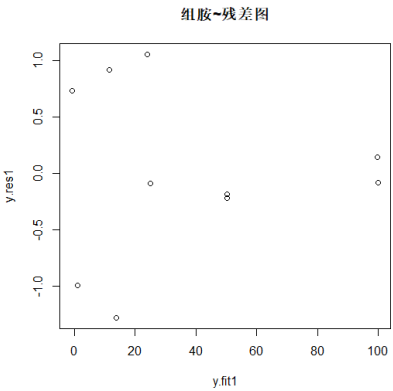
得到如下结果：

```
Call:
lm(formula = 组胺 ~ B + G + R + H + S, data = data1)

Residuals:
    1      2      3      4      5      6      7      8      9     10 
-0.99313 -0.08324 -0.18405 -0.08707  0.92020  0.73337  0.14180 -0.22235  1.05651 -1.28202 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -212.7650    84.1714  -2.528  0.064819 .
B              2.8548     0.9089   3.141  0.034819 *
G             -4.4873     0.4470 -10.039  0.000554 ***
R              2.3213     0.6533   3.553  0.023733 *
H              4.5932     0.8663   5.302  0.006078 **
S              1.1415     0.4298   2.656  0.056641 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.145 on 4 degrees of freedom
Multiple R-squared:  0.9996,    Adjusted R-squared:  0.9991 
F-statistic: 1904 on 5 and 4 DF,  p-value: 7.71e-07
```



$$Y(\text{组胺}) = 2.8548B - 4.4873G + 2.3213R + 4.5932H + 1.1415S - 212.765$$

Call:
lm(formula = 溴酸钾 ~ B + G + R + H + S, data = data2)

Residuals:

	1	2	3	4	5	6	7	8
	0.47852	-2.50015	0.20607	-4.26064	-1.33663	0.05506	3.84708	0.52627
	9	10						
	-1.61096	4.59538						

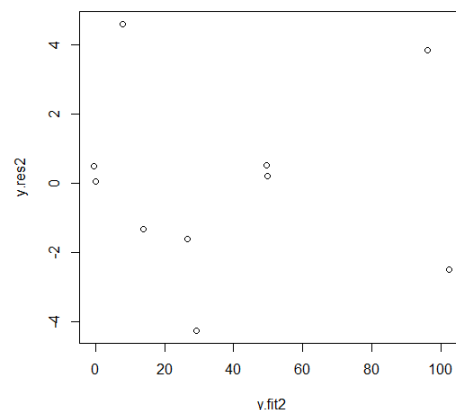
Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1449.592	372.648	3.890	0.0177 *
B	-12.588	6.346	-1.983	0.1183
G	4.868	2.702	1.802	0.1460
R	-1.277	2.747	-0.465	0.6661
H	-7.094	2.427	-2.923	0.0431 *
S	-6.347	3.586	-1.770	0.1514

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.039 on 4 degrees of freedom
Multiple R-squared: 0.9948, Adjusted R-squared: 0.9883
F-statistic: 152.4 on 5 and 4 DF, p-value: 0.0001186

溴酸钾~残差图



$$Y(\text{溴酸钾}) = -12.588B + 4.868G - 1.277R - 7.094H - 6.347S + 1449.592$$

Call:
lm(formula = 工业碱 ~ B + G + R + H + S, data = data3)

Residuals:

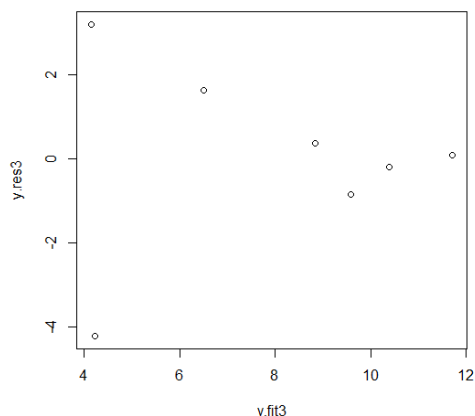
	1	2	3	4	5	6	7
	3.19319	1.63005	-0.84568	0.36213	-0.20690	0.08484	-4.21762

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	261.0649	633.1744	0.412	0.751
B	0.1642	0.6282	0.261	0.837
G	-1.3982	3.0907	-0.452	0.730
R	-0.3136	0.7679	-0.408	0.753
H	-0.1306	1.5097	-0.086	0.945
S	-0.8799	1.7826	-0.494	0.708

Residual standard error: 5.616 on 1 degrees of freedom
Multiple R-squared: 0.6314, Adjusted R-squared: -1.212
F-statistic: 0.3426 on 5 and 1 DF, p-value: 0.8518

工业碱~残差图



$$Y(\text{工业碱}) = 0.1642B - 1.3982G - 0.3136R - 0.1306H - 0.8799S + 261.0649$$

Call:
lm(formula = 硫酸铝钾 ~ B + G + R + H + S, data = data4)

Residuals:

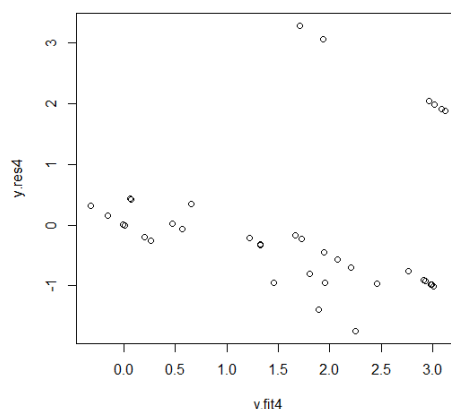
	Min	1Q	Median	3Q	Max
	-1.7490	-0.9134	-0.2271	0.3213	3.2899

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	17.53733	20.13486	0.871	0.390
B	0.05343	0.07840	0.681	0.501
G	-0.01783	0.13538	-0.132	0.896
R	-0.10208	0.31344	-0.326	0.747
H	-0.13388	0.09284	-1.442	0.159
S	-0.01935	0.15131	-0.128	0.899

Residual standard error: 1.285 on 31 degrees of freedom
Multiple R-squared: 0.4687, Adjusted R-squared: 0.383
F-statistic: 5.469 on 5 and 31 DF, p-value: 0.001013

硫酸铝钾~残差图



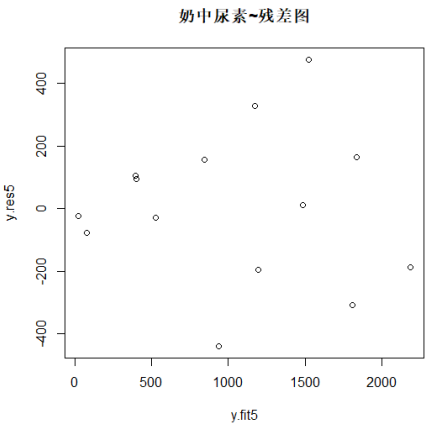
$$Y(\text{硫酸铝钾}) = 0.05343B - 0.01782G - 0.10208R - 0.13388H - 0.01935S + 17.53733$$

```
Call:
lm(formula = 奶中尿素 ~ B + G + R + H + S, data = data5)

Residuals:
    Min       1Q   Median       3Q      Max
-438.88 -132.89  -24.03   130.75   476.24

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 10432.6    10482.3   0.995  0.3456
B             233.2      272.5   0.856  0.4142
G             550.1      176.5   3.117  0.0124 *
R            -801.9      305.9  -2.621  0.0278 *
H            -378.3      119.2  -3.172  0.0113 *
S             222.5      153.7   1.448  0.1816
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 294.4 on 9 degrees of freedom
Multiple R-squared:  0.8991,    Adjusted R-squared:  0.843
F-statistic: 16.04 on 5 and 9 DF,  p-value: 0.0002994
```



$$Y(\text{奶中尿素}) = 233.2B + 550.1G - 801.9R - 378.3H + 222.5S + 10432.6$$

3. 优劣准则：
- (1) 显著性水平 P 值<0.05，且越接近于 0 越好；
 - (2) 拟合系数 R^2 越接近 1 越好；
 - (3) 均方误差 MSE 越小越好。

统计各物质相关数据：

物质	R^2	P 值	MSE
组胺	0.9996	7.71×10^{-7}	0.5249
溴酸钾	0.9948	1.18×10^{-4}	6.5254
工业碱	0.6314	8.52×10^{-1}	4.5054
硫酸铝钾	0.4687	1.01×10^{-3}	1.3833
奶中尿素	0.8991	2.99×10^{-4}	52019

其中工业碱对应的 P 值>0.05，多元线性回归模型不成立，因此数据最差。
最终得到优劣排序结果：

组胺>溴酸钾>奶中尿素>硫酸铝钾>工业碱

问题二

1. 多元线性回归

基于问题一，建立多元线性回归模型，源代码见附录，之后查看残差值：

```
> residuals(lm.sol)
      1      2      3      4      5      6
-3.2151502 -2.4350955  8.8837424  8.8837424  2.8124271 -14.1766876
      7      8      9     10     11     12
-14.3615696 -15.0156709  5.9264686 16.0885535 21.3614123 10.3603442
     13     14     15     16     17     18
-28.3890574 -32.8229329 -38.5580767 -5.7694966 -11.0423555 -0.0412873
     19     20     21     22     23     24
 10.6425456  5.5615032 15.7235880 11.2089108 20.5978754 11.6711957
     25
16.1050712
```

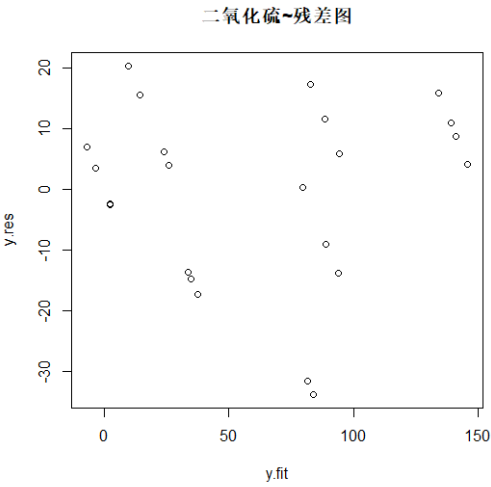
其中第 15 组数据残差值的绝对值最大，将其剔除后得到新的多元线性回归模型：

```
Call:
lm(formula = 二氧化硫 ~ B + G + R + H + S, data = data)

Residuals:
    Min       1Q   Median       3Q      Max
-33.724 -10.126   4.125   9.331  20.326

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2910.630   1096.463   2.655  0.016133 *
B              4.796     3.453   1.389  0.181735
G            -21.156     4.568  -4.631  0.000207 ***
R              3.587     5.292   0.678  0.506476
H            -10.532     2.980  -3.534  0.002372 **
S             -6.751     5.439  -1.241  0.230431
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 16.45 on 18 degrees of freedom
Multiple R-squared:  0.925,    Adjusted R-squared:  0.9042
F-statistic: 44.42 on 5 and 18 DF,  p-value: 1.662e-09
```



$$Y(\text{二氧化硫}) = 4.796B - 21.156G + 3.587R - 10.532H - 6.751S + 2910.630$$

物质	R^2	P 值	MSE
二氧化硫	0.925	1.66×10^{-9}	202.9887

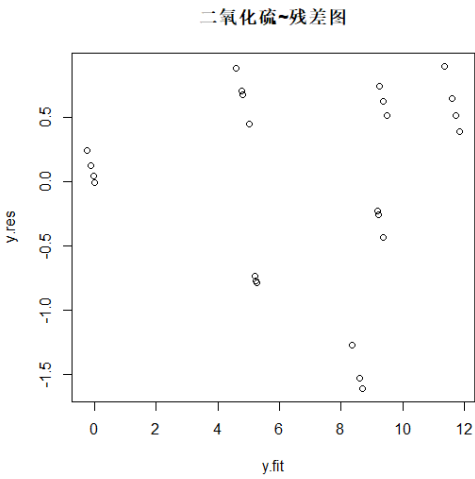
为了优化模型，并保证数据的完整性，选择建立新的模型。
2. 多元二次回归模型

```
Call:
lm(formula = sqrt(二氧化硫) ~ B + G + R + H + S, data = data)

Residuals:
    Min       1Q   Median       3Q      Max
-1.6122 -0.4332  0.2427  0.6250  0.8945

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  113.7581    56.2118   2.024  0.05729 .
B              0.2044     0.1768   1.156  0.26194
G            -0.6829     0.2329  -2.933  0.00854 **
R            -0.1450     0.2645  -0.548  0.58997
H            -0.3175     0.1528  -2.078  0.05149 .
S            -0.1318     0.2762  -0.477  0.63876
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8436 on 19 degrees of freedom
Multiple R-squared:  0.9675,    Adjusted R-squared:  0.9589
F-statistic: 113 on 5 and 19 DF,  p-value: 1.871e-13
```



$$\sqrt{Y(\text{二氧化硫})} = 0.2044B - 0.6829G - 0.1450R - 0.3175H - 0.1318S + 113.7581$$

物质	R^2	P 值	MSE
二氧化硫	0.9675	1.87×10^{-13}	0.5409

新的模型拟合效果更好，而且保证了数据的完整性，拟合误差也很小，可以作为最终的回归模型。

问题三

对于工业碱，其最初的多元线性回归模型已不成立，因此不考虑数据量和颜色维度对其回归模型的影响。下面分析其他物质的数据量和颜色维度对回归模型的影响，源代码见附录。

1. 数据量

回归分析模型对样本数据量具有很强的依赖性，样本太少无法保证模型的精确度和准确性。对于每种物质，逐一减少最后的样本数据，直到拟合的多元回归模型不成立，统计相关数据，结果如下：

(1) 组胺

R^2	P 值	MSE
0.9996	7.71×10^{-7}	0.5409
0.9998	7.18×10^{-6}	0.2160
0.9999	1.78×10^{-4}	0.1030
0.9999	9.10×10^{-3}	0.0471

(2) 溴酸钾

R^2	P 值	MSE
0.9948	1.19×10^{-4}	6.5254
0.9985	1.93×10^{-4}	1.9400
0.9992	1.79×10^{-3}	1.0354
0.9999	1.73×10^{-2}	0.1710

(3) 硫酸铝钾 (由于数据量较大, 依次减少 5 组最后的样本数据)

R^2	P 值	MSE
0.4687	1.01×10^{-3}	1.3833
0.6407	3.67×10^{-5}	0.3466
0.8430	8.64×10^{-8}	0.0610
0.7909	5.68×10^{-5}	0.0531
0.9427	1.81×10^{-6}	0.0082
0.9965	2.71×10^{-7}	0.0002
0.9999	1.66×10^{-2}	0.0000

(4) 奶中尿素

R^2	P 值	MSE
0.8991	2.99×10^{-4}	52019.00
0.8875	1.25×10^{-3}	52934.85
0.8809	3.91×10^{-3}	56726.07
0.8918	7.33×10^{-3}	55622.05
0.9025	1.45×10^{-2}	53600.50
0.8874	4.93×10^{-2}	58821.83

(5) 二氧化硫 (依次减少 4 组最后的样本数据)

R^2	P 值	MSE
0.9675	1.87×10^{-13}	0.5409
0.9725	3.66×10^{-11}	0.3373
0.9965	4.12×10^{-13}	0.0339
0.9963	2.42×10^{-8}	0.0262
0.9996	2.59×10^{-5}	0.0022

可以发现,随着数据量的减少,组胺、溴酸钾、奶中尿素和二氧化硫的拟合系数 R^2 和显著性水平 P 值一直在增加,说明其数据量的减少可以让拟合效果更好。硫酸铝钾的拟合系数也在一直增加,但显著性水平 P 值却呈先减后增的趋势,说明适合的数据量可以拟合出更好的回归模型。

因此,五种物质的多元回归模型的数据量应当适中,最好保持在 10 组左右。

2. 颜色维度

颜色维度分为五维(RGBHS)、三维(RGB)和二维(HS),对于每种物质,分别对比分析结果。

(1) 组胺

维度	R^2	P 值	MSE
RGBHS	0.9996	7.71×10^{-7}	0.5249
RGB	0.9995	2.39×10^{-7}	5.9844
HS	0.9714	3.96×10^{-6}	35.7600

(2) 溴酸钾

维度	R^2	P 值	MSE
RGBHS	0.9948	1.19×10^{-4}	6.5255
RGB	0.9408	4.44×10^{-4}	74.0269
HS	0.9478	3.25×10^{-5}	65.2286

(3) 硫酸铝钾

维度	R^2	P 值	MSE
RGBHS	0.4687	1.01×10^{-3}	1.3833
RGB	0.4091	5.29×10^{-4}	1.5383
HS	0.4576	3.04×10^{-5}	1.4120

(4) 奶中尿素

维度	R^2	P 值	MSE
RGBHS	0.8991	2.99×10^{-4}	52019
RGB	0.7800	6.15×10^{-4}	113428.2
HS	0.7526	2.29×10^{-4}	127558.7

(5) 二氧化硫

维度	R^2	P 值	MSE
RGBHS	0.9675	1.87×10^{-13}	0.5409
RGB	0.9599	7.99×10^{-15}	0.6668
HS	0.9454	1.29×10^{-14}	0.9078

可以发现，随着维度的下降和改变，模型的通用性降低，而且颜色维度的大小比数据量的多少对回归模型的影响更大。

3. 相关性分析

为了进一步分析颜色维度对每种物质的影响，针对每种物质，分别分析五种颜色读数与物质浓度之间的相关性，得到如下表的分析结果，显著性水平为 0.05，分析结果均保留三位有效数字。

物质	分析结果	B	G	R	H	S
组胺	相关系数	-0.972	-0.997	-0.931	-0.978	0.963
	p 值	0.000	0.000	0.000	0.000	0.000
溴酸钾	相关系数	-0.956	-0.868	-0.163	0.696	0.953
	p 值	0.000	0.001	0.653	0.025	0.000
工业碱	相关系数	-0.491	-0.664	-0.624	0.708	0.658
	p 值	0.264	0.104	0.134	0.075	0.108
硫酸铝钾	相关系数	0.606	-0.619	-0.623	0.503	0.613
	p 值	0.000	0.000	0.000	0.002	0.000
奶中尿素	相关系数	-0.88	0.003	-0.14	0.521	0.866
	p 值	0.000	0.992	0.618	0.047	0.000
二氧化硫	相关系数	-0.844	-0.867	0.696	-0.150	0.830
	p 值	0.000	0.000	0.000	0.476	0.000

可以得到以下结论：

- (1) 组胺的浓度与 B、G、R、H、S 有显著的线性相关，程度极强；
- (2) 溴酸钾的浓度与 B、G、H、S 有显著的线性相关，与 R 的相关程度极低弱；
- (3) 工业碱的浓度与 B、G、R、H、S 有显著的线性相关，程度适中；
- (4) 硫酸铝钾的浓度与 B、G、R、H、S 有显著的线性相关，程度较强；
- (5) 奶中尿素与 B、H、S 有显著的线性相关，与 G、R 的相关程度极弱；
- (5) 二氧化硫与 B、G、R、S 有显著的线性相关，与 H 的相关程度极弱。

可知组胺、工业碱和硫酸铝钾的多元回归模型不变，仍为问题一的结论。利用以上结论，改变相应物质的颜色维度，建立新的多元回归模型。

(1) 溴酸钾

$$Y(\text{溴酸钾}) = -14.923B + 4.923G - 6.523H - 7.686S + 1581.191$$

物质	R^2	P 值	MSE
溴酸钾	0.9945	7.83×10^{-6}	6.8782

残差值如下：

1	2	3	4	5	6
0.65153299	-3.19541606	-0.11133412	-4.32732770	-0.34590636	-0.06280642
7	8	9	10		
4.49020489	1.23425800	-2.01363481	3.68042958		

在改变颜色维度后，溴酸钾的回归模型通用性上升。

(2) 奶中尿素

$$Y(\text{奶中尿素}) = -86.74B - 49.64H + 17.85S + 11155.96$$

物质	R^2	P 值	MSE
奶中尿素	0.7876	5.09×10^{-4}	109510.5

残差值如下：

```
1      2      3      4      5      6      7
-339.92162 101.22747 -361.82070 248.07494 531.09161 -45.40784 242.91591
8      9      10     11     12     13     14
-669.08753 -119.23461 306.56564 197.53847 -311.81974 149.67641 -328.41175
15
398.61334
```

在改变颜色维度后，奶中尿素的回归模型通用性下降。

(3) 二氧化硫

其多元二次回归模型的拟合效果很好，不需要降低颜色维度。

可见，更改颜色维度后回归模型的通用性不一定变好，如果综合考虑模型的数据量和颜色维度相关性，则可以得到更好的多元回归模型。

附录

1. 主程序

```
1. #####问题一#####
2.
3. source('D:/桌面/大作业/程序/regression.r') #多元线性回归分析
4. for(i in 1:5)
5. {
6.   print(switch(i,'组胺','溴酸钾','工业碱','硫酸铝钾','奶中尿素'))
7.   regression(i)
8. }
9.
10. #####问题二#####
11.
12. data=read.csv('D:/桌面/大作业/二氧化硫.csv',encoding='utf-8')
13. data=data[-15,] #删除异常值
14. lm.sol=lm(二氧化硫~B+G+R+H+S,data)
15. summary(lm.sol)
16. y.res=residuals(lm.sol)
17. y.fit=predict(lm.sol)
18. mse=mean(y.res^2);mse
19. plot(y.res~y.fit,main=bquote('二氧化硫~残差图'))
20.
21. lm.new=update(lm.sol,sqrt(.)~.) #多元二次回归分析
22. summary(lm.new)
23. y.res=residuals(lm.new)
```



```

24. y.fit=predict(lm.new)
25. mse=mean(y.res^2);mse
26. plot(y.res~y.fit,main=bquote('二氧化硫~残差图'))
27.
28. #####问题三#####
29.
30. source('D:/桌面/大作业/程序/num.r') #数据量
31. length=c(10,10,7,37,15,25)
32. for(i in 1:6)
33. {
34.     print(switch(i,'组胺','溴酸钾','工业碱','硫酸铝钾','奶中尿素','二氧化
        硫'))
35.     for(j in length[i]:1)
36.     {
37.         res=num(i,j);print(res)
38.         if(res$'回归模型'==FALSE) break
39.     }
40. }
41.
42. source('D:/桌面/大作业/程序/dim.r') #颜色维度
43. for(i in 1:6)
44. {
45.     print(switch(i,'组胺','溴酸钾','工业碱','硫酸铝钾','奶中尿素','二氧化
        硫'))
46.     for(j in 1:3)
47.     {
48.         print(dim(i,j))
49.     }
50. }
51.
52. source('D:/桌面/大作业/程序/correlation.r') #相关性分析
53. library(Hmisc) #相关性分析程序包
54. library(openxlsx) #表格数据写入包
55. for(i in 1:6)
56. {
57.     rp=paste0('rp',i)
58.     assign(rp,correlation(i))
59. }
60. sheets=list('组胺'=rp1,'溴酸钾'=rp2,'工业碱'=rp3,'硫酸铝钾'=rp4,'奶中尿
        素'=rp5,'二氧化硫'=rp6)
61. write.xlsx(sheets,'D:/桌面/大作业/结果.xlsx',rowNames=T)
62.
63. data=read.csv('D:/桌面/大作业/溴酸钾.csv',encoding='utf-8')
64. lm.sol=lm(溴酸钾~B+G+H+S,data)

```

```

65.
66. data=read.csv('D:/桌面/大作业/奶中尿素.csv',encoding='utf-8')
67. lm.sol=lm(奶中尿素~B+H+S,data)
68.
69. summary(lm.sol)
70. y.res=residuals(lm.sol);y.res
71. mse=mean(y.res^2);mse

```

2. 多元回归分析

(1) 初始函数

```

1. regression=function(n){
2.     name=switch(n,'组胺','溴酸钾','工业碱','硫酸铝钾','奶中尿素')
3.     file=paste('D:/桌面/大作业/',name,'.csv',sep='')
4.     data=read.csv(file,encoding='utf-8')
5.     y=switch(n,data$组胺,data$溴酸钾,data$工业碱,data$硫酸铝钾,data$奶中
        尿素)
6.     lm.sol=lm(y~B+G+R+H+S,data)
7.     lm.sum=summary(lm.sol)
8.     y.res=residuals(lm.sol)
9.     y.fit=predict(lm.sol)
10.    R=lm.sum$r.squared
11.    f=lm.sum$fstatistic
12.    p=1-pf(f[1],f[2],f[3])
13.    mse=mean(y.res^2)
14.    if(is.nan(p) || p>0.05) answer=F else answer=T
15.    result=data.frame('拟合系数'=R,'P 值'=p,'均方误差'=mse,'回归模型
        '=answer)
16.    print(lm.sum);print(result)
17.    plot(y.res~y.fit,main=paste(name,'残差图',sep='~'))
18. }

```

(2) 数据量函数

```

1. num=function(n,m){
2.     name=switch(n,'组胺','溴酸钾','工业碱','硫酸铝钾','奶中尿素','二氧化
        硫')
3.     file=paste('D:/桌面/大作业/',name,'.csv',sep='')
4.     data=read.csv(file,encoding='utf-8')
5.     y=switch(n,data$组胺,data$溴酸钾,data$工业碱,data$硫酸铝钾,data$奶中
        尿素,data$二氧化硫)
6.     data=data[1:m,]
7.     y=y[1:m]
8.     lm.sol=lm(y~B+G+R+H+S,data)
9.     if(n==6) lm.sol=update(lm.sol,sqrt(.)~.)
10.    R=summary(lm.sol)$r.squared
11.    f=summary(lm.sol)$fstatistic
12.    p=1-pf(f[1],f[2],f[3])

```

```

13.     y.res=residuals(lm.sol)
14.     mse=mean(y.res^2)
15.     if(is.nan(p) || p>0.05) answer=F else answer=T
16.     result=data.frame('拟合系数'=R,'P 值'=p,'均方误差'=mse,'回归模型'
                          '=answer')
17.     return(result)
18. }

```

(3) 颜色维度函数

```

1. dim=function(n,k){
2.     name=switch(n,'组胺','溴酸钾','工业碱','硫酸铝钾','奶中尿素','二氧化
    硫')
3.     file=paste('D:/桌面/大作业/',name,'.csv',sep='')
4.     data=read.csv(file,encoding='utf-8')
5.     y=switch(n,data$组胺,data$溴酸钾,data$工业碱,data$硫酸铝钾,data$奶中
    尿素,data$二氧化硫)
6.     lm.sol=switch(k,
7.         lm(y~B+G+R+H+S,data),
8.         lm(y~B+G+R,data),
9.         lm(y~H+S,data))
10.    if(n==6) lm.sol=update(lm.sol,sqrt(.)~.)
11.    R=summary(lm.sol)$r.squared
12.    f=summary(lm.sol)$fstatistic
13.    p=1-pf(f[1],f[2],f[3])
14.    y.res=residuals(lm.sol)
15.    mse=mean(y.res^2)
16.    if(is.nan(p) || p>0.05) answer=F else answer=T
17.    result=data.frame('拟合系数'=R,'P 值'=p,'均方误差'=mse,'回归模型'
                        '=answer')
18.    return(result)
19. }

```

3. 相关性分析

```

1. correlation=function(n){
2.     name=switch(n,'组胺','溴酸钾','工业碱','硫酸铝钾','奶中尿素','二氧化
    硫')
3.     file=paste('D:/桌面/大作业/',name,'.csv',sep='')
4.     data=read.csv(file,encoding='utf-8')
5.     y=switch(n,data$组胺,data$溴酸钾,data$工业碱,data$硫酸铝钾,data$奶中
    尿素,data$二氧化硫)
6.     r=cor(y,data[,2:6],method='pearson')
7.     p=rcorr(as.matrix(data),type='pearson')$P[2:6,1]
8.     rp=rbind(data.frame(r),p)
9.     row.names(rp)=c('相关系数','p 值')
10.    return(rp)
11. }

```