

多元统计分析报告

工业工程 52 唐静雯 2015010846

目录

- 多元统计分析报告1
- 一、 数据总体分析1
 - 1. 数据详情1
 - 2. 数据可视化2
- 二、 初步回归分析3
 - 1. 人们喜爱程度3
 - 2. 总体风味强度4
 - 3. 人们喜爱程度&整体风味强度5
- 三、 主成分分析6
 - 1. 主成分个数选择6
 - 2. 主成分分析6
 - 3. 主成解释释7
 - 4. 效果验证8
 - 5. 用主成分进行线性回归9
- 四、 因子分析 11
 - 1. 选取因子个数 11
 - 2. 因子分析 12
 - 3. 因子旋转及解释 13
- 五、 典型相关分析 13
 - 1. 典型相关分析 13
 - 2. 解释 15
 - 3. 有效性检验 15
- 六、 总结 16

一、 数据总体分析

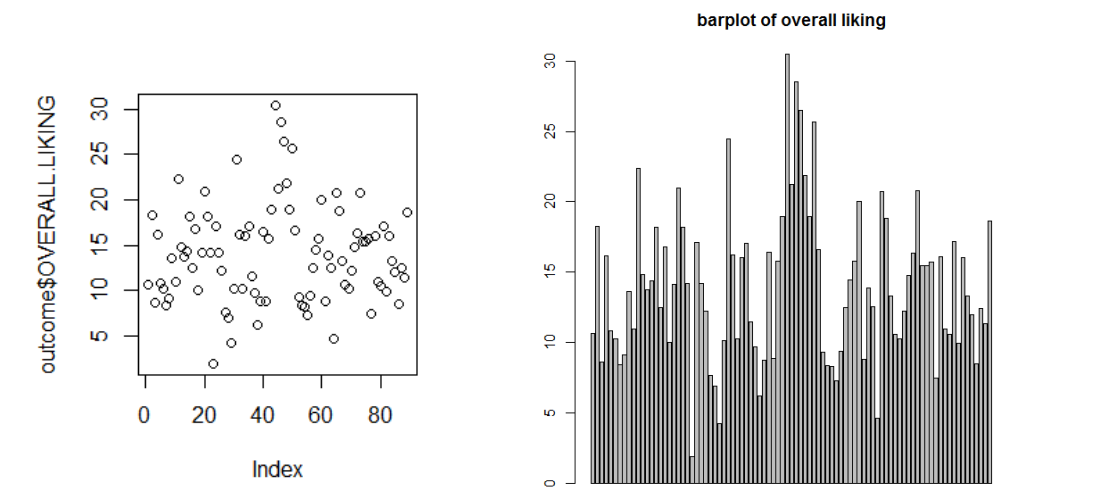
1. 数据详情

该数据由两部分组成。一个是对于 89 个番茄品种的 8 个维度的口味评价，包括两个总的指标：人们的喜爱程度，和总体风味强度，以及 6 个小的方面：质感，甜度，酸度，咸味，苦味，鲜味。第二部分是对于这 89 个番茄品种中的化学成分的检测和统计，研究对象包括 68 中化学物质。

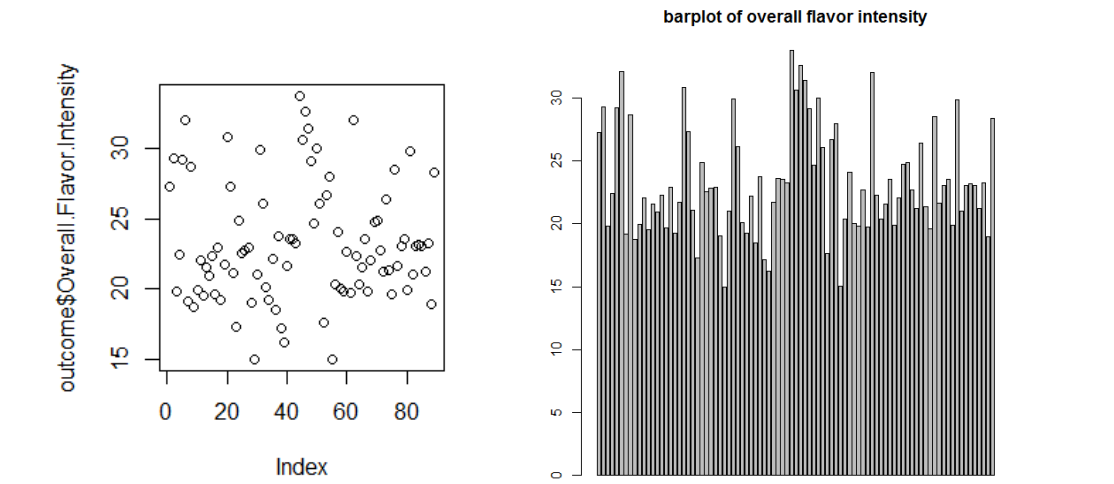
该数据的数据背景是与传统品种相比，现代商业番茄的风味质量下降，导致了人们对之喜爱程度的下降。由于任何食物的风味都是味觉和嗅觉的加和，为了解决这一问题，对番茄的化

学成分和风味的关系的研究就显得有实际意义。因此，该数据缘起于一项基于番茄的基因和化学成分的研究。

2. 数据可视化



从总体喜爱程度的散点图中可以看出，喜爱程度由 0-30 的数表示，分布较为均匀。从柱状图可以看出，对于这 89 种番茄品种人们的喜爱程度的分布接近于均匀分布。



从总体风味强度的散点图可以看出，风味强度由 15-30 的数表示，分布较为均匀。从柱状图可以看出，对于这 89 种番茄品种的风味强度评价接近于均匀分布。

总体来看，有关番茄风味的数据的散点图矩阵：



从图中可以看出，总体风味和各个分变量之间有着较为明显的相关性，以及人们喜爱程度和质地也呈现出一定的相关性。下面进行进一步的分析。

二、 初步回归分析

本部分对于味道数据进行原变量的回归分析，以初步掌握变量之间的关系。

1. 人们喜爱程度

```
> likefit.step$anova
Stepwise Model Path
Analysis of Deviance Table

Initial Model:
outcome$OVERALL.LIKING ~ 1

Final Model:
outcome$OVERALL.LIKING ~ TEXTURE + Sweetness + Sourness + Umami +
  Bitter + Salty
```

	Step	Df	Deviance	Resid. Df	Resid. Dev	AIC
1				88	2562.5333	301.0503
2	+ TEXTURE	1	1677.27998	87	885.2533	208.4541
3	+ Sweetness	1	438.19019	86	447.0631	149.6516
4	+ Sourness	1	59.37808	85	387.6850	138.9686
5	+ Umami	1	35.11850	84	352.5665	132.5177
6	+ Bitter	1	54.39668	83	298.1699	119.6034
7	+ Salty	1	10.44581	82	287.7241	118.4295

运用逐步分析模型得到下面的最终模型。

```
> likefit = lm(outcome$OVERALL.LIKING ~ TEXTURE + Sweetness + Sourness + Umami +  
+           Bitter + Salty, data = outcome)  
> summary(likefit)
```

```
Call:  
lm(formula = outcome$OVERALL.LIKING ~ TEXTURE + Sweetness + Sourness +  
    Umami + Bitter + Salty, data = outcome)
```

```
Residuals:  
    Min       1Q   Median       3Q      Max  
-3.7609 -1.1760 -0.1889  1.0806  4.5241
```

```
Coefficients:  
              Estimate Std. Error t value Pr(>|t|)  
(Intercept) -2.96590     1.42273  -2.085  0.0402 *  
TEXTURE      0.51716     0.04729  10.935 < 2e-16 ***  
Sweetness    0.42417     0.07366   5.759 1.43e-07 ***  
Sourness    -0.25692     0.12181  -2.109  0.0380 *  
Umami        0.83805     0.19997   4.191 6.97e-05 ***  
Bitter      -0.91158     0.21624  -4.216 6.37e-05 ***  
Salty        0.43855     0.25417   1.725  0.0882 .  
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 1.873 on 82 degrees of freedom  
Multiple R-squared:  0.8877,    Adjusted R-squared:  0.8795  
F-statistic: 108.1 on 6 and 82 DF,  p-value: < 2.2e-16
```

可以看出人们对于一种番茄的喜爱程度,有着正面影响的因素由高到低分别是 鲜味,质地,咸味,甜度;而有着负面影响的因素由大到小是苦味和酸味。和我们日常生活的感受认知是一致的。

2. 总体风味强度

```
> flavorfit.step$anova  
Stepwise Model Path  
Analysis of Deviance Table
```

```
Initial Model:  
outcome$OVERALL.LIKING ~ 1
```

```
Final Model:  
outcome$OVERALL.LIKING ~ TEXTURE + Sweetness + Sourness + Umami +  
    Bitter + Salty
```

	Step	Df	Deviance	Resid. Df	Resid. Dev	AIC
1				88	2562.5333	301.0503
2	+ TEXTURE	1	1677.27998	87	885.2533	208.4541
3	+ Sweetness	1	438.19019	86	447.0631	149.6516
4	+ Sourness	1	59.37808	85	387.6850	138.9686
5	+ Umami	1	35.11850	84	352.5665	132.5177
6	+ Bitter	1	54.39668	83	298.1699	119.6034
7	+ Salty	1	10.44581	82	287.7241	118.4295

得到下面的最终模型：

```
> flavorfit = lm(outcome$Overall.Flavor.Intensity ~ TEXTURE + Sweetness + Sourness + Umami +
+               Bitter + Salty,data = outcome)
> summary(flavorfit)

Call:
lm(formula = outcome$Overall.Flavor.Intensity ~ TEXTURE + Sweetness +
    Sourness + Umami + Bitter + Salty, data = outcome)

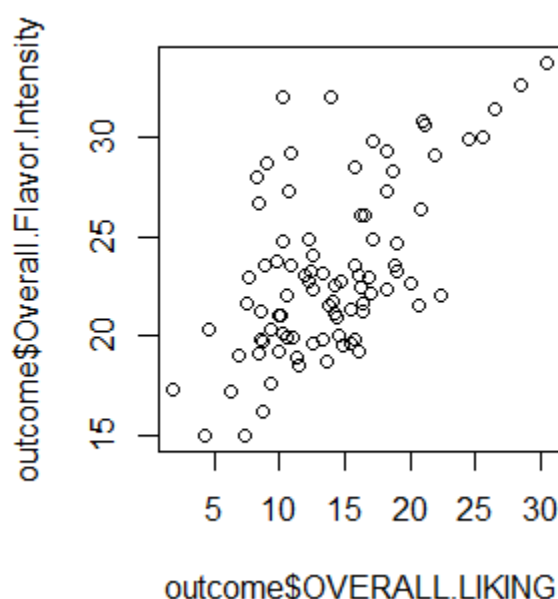
Residuals:
    Min       1Q   Median       3Q      Max
-3.0020 -0.8331 -0.1043  0.7889  4.8309

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.88026    1.12540  -0.782  0.43636
TEXTURE      0.06642    0.03741   1.775  0.07954 .
Sweetness    0.50438    0.05826   8.657 3.47e-13 ***
Sourness     0.31044    0.09635   3.222  0.00183 **
Umami        0.44619    0.15818   2.821  0.00601 **
Bitter       -0.04745    0.17105  -0.277  0.78217
Salty        0.60451    0.20105   3.007  0.00350 **
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.482 on 82 degrees of freedom
Multiple R-squared:  0.8861,    Adjusted R-squared:  0.8777
F-statistic: 106.3 on 6 and 82 DF,  p-value: < 2.2e-16
```

可以看出一种番茄的总体风味强度，和咸味，甜度，鲜味关系最为密切，和酸味也有一定的正相关关系。而苦味的效果并不是很显著。

3. 人们喜爱程度&整体风味强度



```
> cor(outcome$OVERALL.LIKING,outcome$Overall.Flavor.Intensity)
[1] 0.6024601
```

二者相关系数为 0.6。呈现一定程度正相关。从上面的分析可以猜测二者的差异（即整体风味强度高并不意味着人们就一定喜爱）的原因可能是：

- 1) 质地对二者的影响不一样，对于人们喜爱程度影响较大（0.52）但是对于整体风味强度没有太大关系（0.06）。
- 2) 苦味对二者的影响不一样，对于人们喜爱程度影响较大（-0.91）但是对于整体风味强度没有太大关系（0.04）。

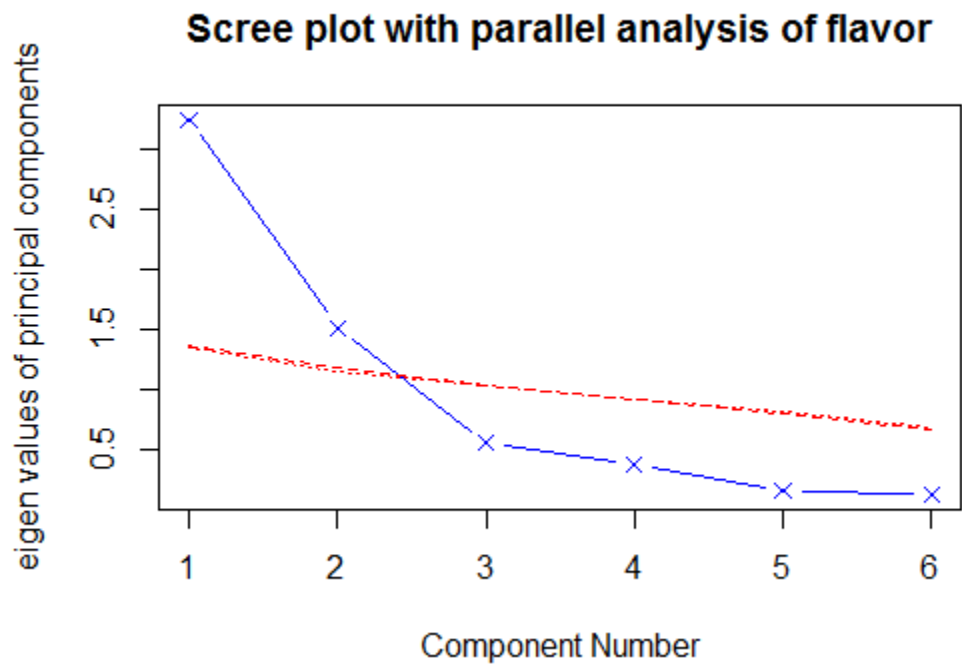
3) 酸度对二者的影响不一样，对于人们喜爱程度有着较大的负面影响 (-0.26) 但是对于整体风味强度有着正面影响 (0.31)。

三、主成分分析

本部分计划对风味数据部分进行主成分分析。

1. 主成分个数选择

运用 Cattell 碎石检验 (由线段和 x 符号组成)，根据 100 个随机矩阵推导出来的特征值均值 (虚线)，得出选取两个主成分的结论。



2. 主成分分析

```
> cor(outcome.sub)
```

	TEXTURE	Sweetness	Sourness	Salty	Bitter	Umami
TEXTURE	1.00000000	0.45687445	0.05801857	0.2320522	-0.15424787	0.3463417
Sweetness	0.45687445	1.00000000	0.19085141	0.5182630	0.08412435	0.6105188
Sourness	0.05801857	0.19085141	1.00000000	0.7677604	0.71384441	0.4528074
salty	0.23205217	0.51826300	0.76776036	1.0000000	0.66163945	0.7054950
Bitter	-0.15424787	0.08412435	0.71384441	0.6616395	1.00000000	0.5419926
Umami	0.34634167	0.61051878	0.45280739	0.7054950	0.54199261	1.0000000

首先是去除整体风味强度和人们喜爱程度这两个较为综合的指标之后的数据的相关系数矩阵。(从中可以注意到只有质地和苦味是呈负相关，其余的变量之间都是正相关。相关性最大的一对是咸味和酸度，达到了 0.77；其次是酸味和苦味，咸味和鲜味这两对，约 0.7。)

接下来用这个矩阵进行主成分分析。

```
> outcome.pc<-princomp(outcome.sub,cor=TRUE)
> summary(outcome.pc,loadings=TRUE)
Importance of components:
               Comp.1      Comp.2      Comp.3      Comp.4      Comp.5      Comp.6
Standard deviation  1.7980932  1.2306692  0.74943808  0.61659522  0.41221197  0.37489777
Proportion of Variance 0.5388565 0.2524245 0.09360957 0.06336494 0.02831979 0.02342472
Cumulative Proportion 0.5388565 0.7912810 0.88489055 0.94825549 0.97657528 1.00000000

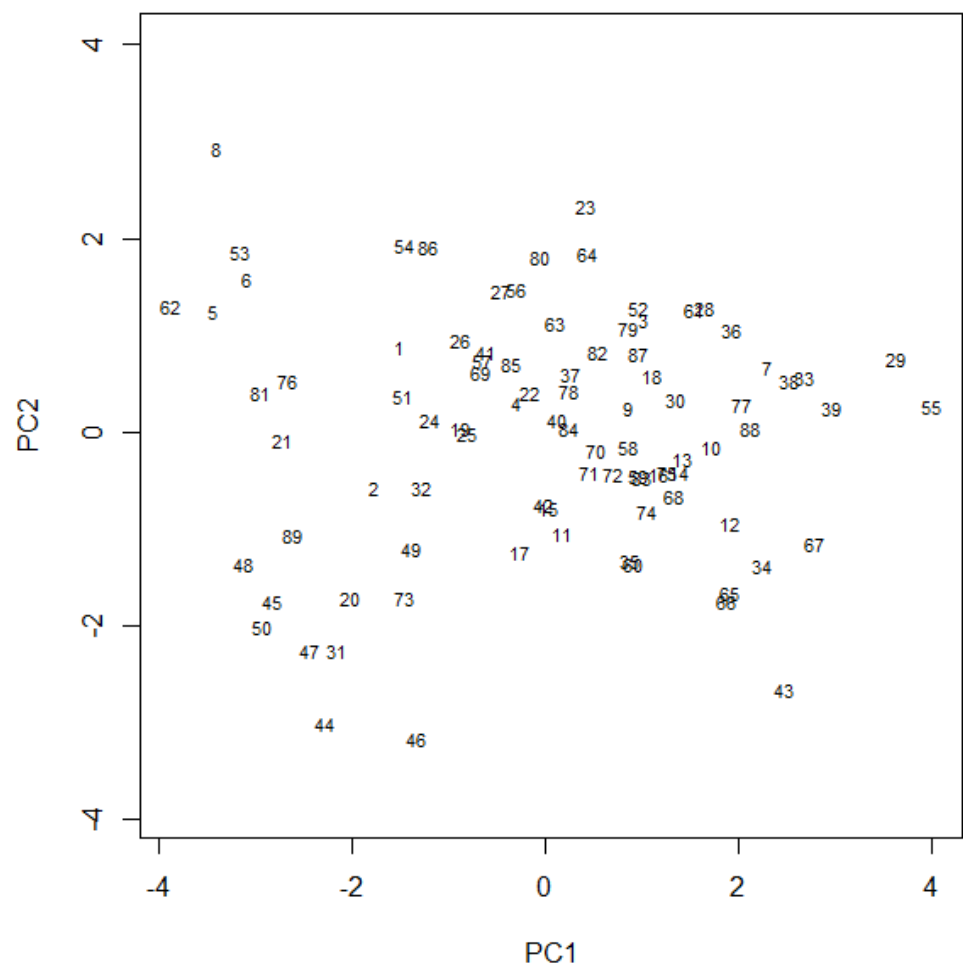
Loadings:
               Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6
TEXTURE      -0.180 -0.632  0.682 -0.244  0.200
Sweetness    -0.339 -0.502 -0.499  0.447  0.404  0.145
Sourness     -0.439  0.325  0.424  0.422 -0.114  0.575
Salty        -0.519          0.272 -0.340 -0.732
Bitter       -0.409  0.461      -0.391  0.670 -0.131
Umami        -0.473 -0.166 -0.320 -0.580 -0.467  0.303
```

从结果中可以看出，解释的方差比重，第一主成分 0.54，第二主成分 0.25，然而到了第三主成分就到了 0.09，因此前两个主成分已经足够解释 79%的方差，符合之前对选取主成分个数的判断。

3. 主成分解释

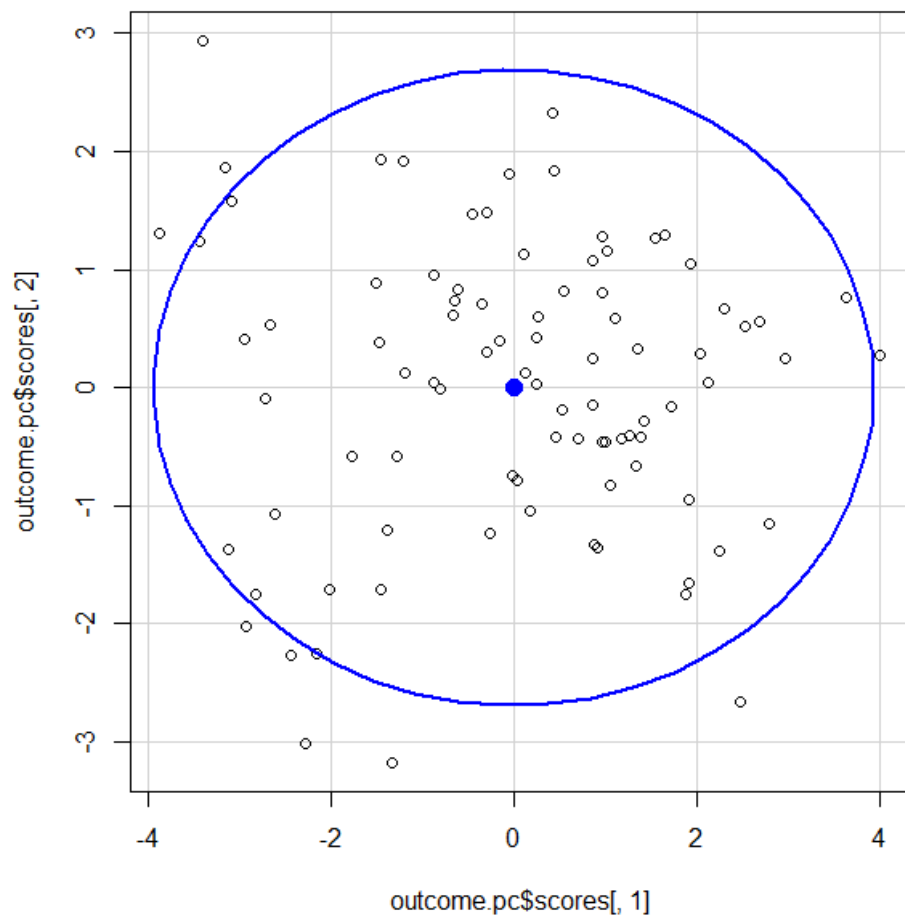
第一个主成分是个味道的负系数组合，且系数都处于 0.2-0.5 范围内，因此可以将第一个主成分看作是一项综合性指标，可以看做是一系列特征的强度综合的负值，即“寡淡度”。第二个主成分中，根据前面的回归分析，“不太受人们喜爱的”特征的系数均为正值，而“受人喜爱”的特征的系数为负值或零，意味着第二个主成分代表着该番茄种类和人类口味偏好相违背的程度。

4. 效果验证



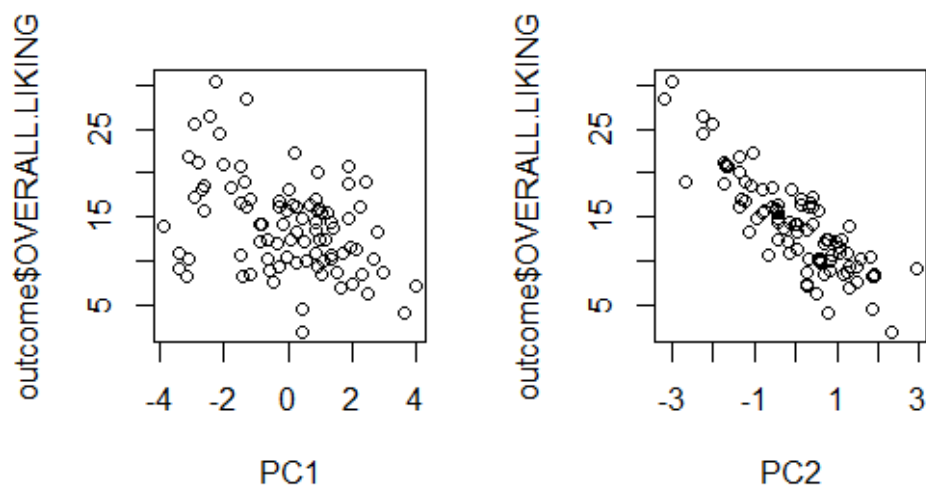
画出前

两个主成分的载荷（具体数值见 R markdown）的散点图。图上数字表示的是番茄的种类。可以看出基本是椭圆的分布，下面的带 90%置信区间的椭圆的图更加说明了这一点。因此该主成分分析是较为有效的。



5. 用主成分进行线性回归

1) 人们喜爱程度



```
> summary(lm(outcome$OVERALL.LIKING~outcome.pc$scores[,1]+outcome.pc$scores[,2]))
```

Call:

```
lm(formula = outcome$OVERALL.LIKING ~ outcome.pc$scores[, 1] +  
    outcome.pc$scores[, 2])
```

Residuals:

Min	1Q	Median	3Q	Max
-4.3014	-1.4643	-0.2584	1.2898	4.7066

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	13.9932	0.1980	70.66	<2e-16 ***
outcome.pc\$scores[, 1]	-1.1841	0.1101	-10.75	<2e-16 ***
outcome.pc\$scores[, 2]	-3.7135	0.1609	-23.08	<2e-16 ***

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

Residual standard error: 1.868 on 86 degrees of freedom

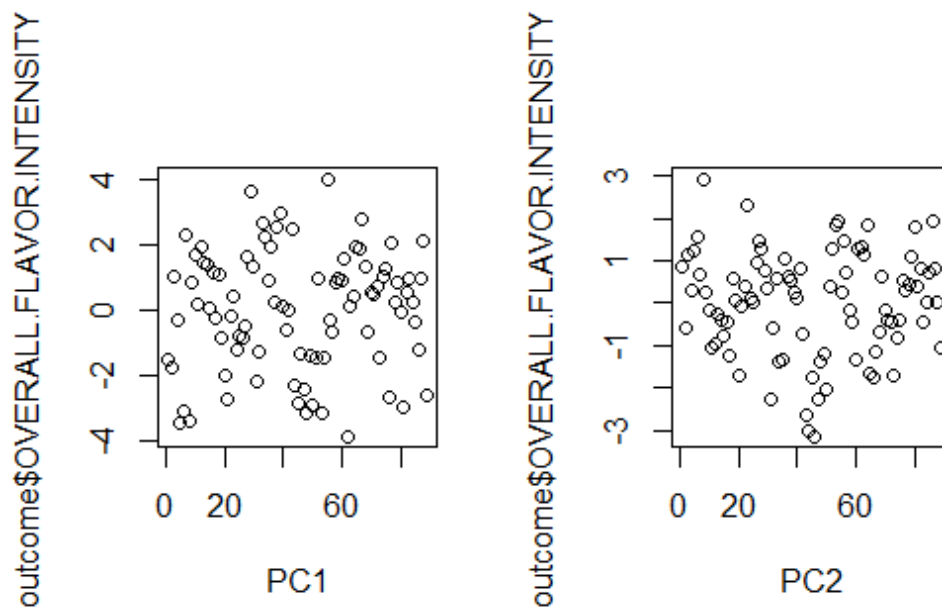
Multiple R-squared: 0.8828, Adjusted R-squared: 0.8801

F-statistic: 324 on 2 and 86 DF, p-value: < 2.2e-16

可以看出人们喜爱程度和第一第二主成分都显著负相关，因此和之前主成分的解释一致。值得注意的是和第二主成分的负相关性更强一些，因为第二主成分相较于第一主成分更符合人们喜好的反面。

和之前的回归的 Adjusted R-squared (0.8795) 相比，Adjusted R-squared 增大到 0.8801。意味着模型的解释性更强一些。

2) 整体风味强度



```
> summary(lm(outcome$Overall.Flavor.Intensity~outcome.pc$scores[,1]+outcome.pc$scores[,2]))  
Call:  
lm(formula = outcome$Overall.Flavor.Intensity ~ outcome.pc$scores[,  
  1] + outcome.pc$scores[, 2])  
Residuals:  
    Min       1Q   Median       3Q      Max   
-3.2641 -1.2291 -0.0520  0.8797  4.2067  
Coefficients:  
                Estimate Std. Error t value Pr(>|t|)      
(Intercept)      23.27065    0.17662  131.759 < 2e-16 ***  
outcome.pc$scores[, 1] -2.03594    0.09822  -20.728 < 2e-16 ***  
outcome.pc$scores[, 2] -1.05063    0.14351   -7.321 1.22e-10 ***  
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
Residual standard error: 1.666 on 86 degrees of freedom  
Multiple R-squared:  0.8489,    Adjusted R-squared:  0.8454  
F-statistic: 241.6 on 2 and 86 DF,  p-value: < 2.2e-16
```

可以看出人们喜爱程度和第一第二主成分都显著负相关，因此和之前主成分的解释一致。值得注意的是和第一主成分的负相关性更强一些，因为第一主成分按照之前的解释是“寡淡程度”，和整体风味强度更为密切。

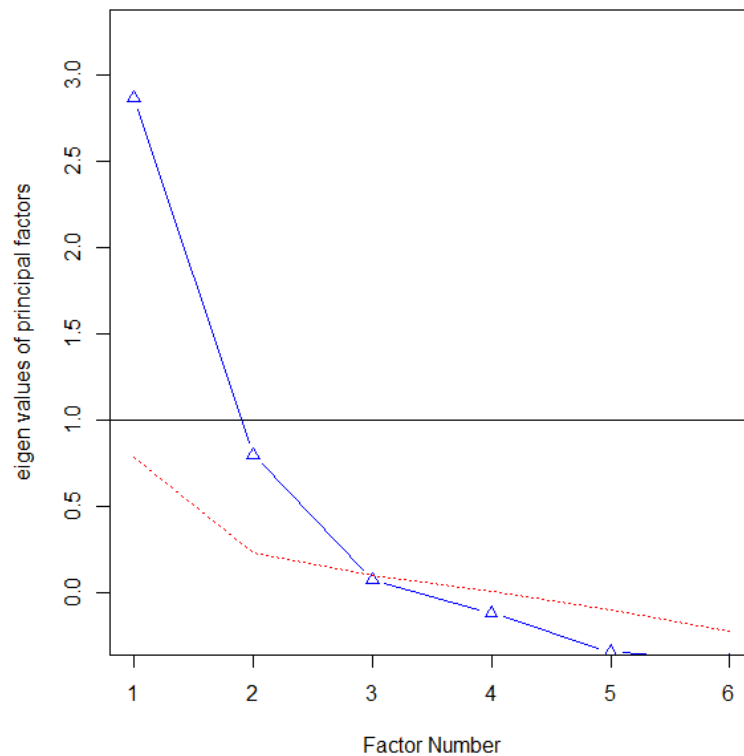
四、 因子分析

本部分运用因子分析对味道数据集进行分析。

1. 选取因子个数

运用 Cattell 碎石检验（由线段和三角符号组成），根据 100 个随机矩阵推导出来的特征值均值（虚线），得出选取两个因子的结论。（因为将特征值与 0 比较而不是 1）

Scree plots with parallel analysis of factor analysis of outcome



2. 因子分析

采用最大似然法进行因子分析，得到以下结果：

```
> fa = fa(r = correlations, n.factors = 2, rotate = "none", fm = "ml")
> fa
Factor Analysis using method = ml
Call: fa(r = correlations, n.factors = 2, rotate = "none", fm = "ml")
Standardized loadings (pattern matrix) based upon correlation matrix
```

	ML1	ML2	h2	u2	com
TEXTURE	0.25	0.51	0.32	0.68	1.4
Sweetness	0.57	0.68	0.79	0.21	1.9
Sourness	0.79	-0.37	0.75	0.25	1.4
Salty	0.94	-0.02	0.89	0.11	1.0
Bitter	0.71	-0.46	0.71	0.29	1.7
Umami	0.77	0.24	0.65	0.35	1.2

```
SS loadings          ML1 ML2
Proportion var       0.50 0.19
Cumulative var       0.50 0.68
Proportion Explained 0.73 0.27
Cumulative Proportion 0.73 1.00
```

Mean item complexity = 1.5
Test of the hypothesis that 2 factors are sufficient.

The degrees of freedom for the null model are 15 and the objective function was 3.69
The degrees of freedom for the model are 4 and the objective function was 0.3

The root mean square of the residuals (RMSR) is 0.05
The df corrected root mean square of the residuals is 0.09

Fit based upon off diagonal values = 0.99
Measures of factor score adequacy

	ML1	ML2
Correlation of (regression) scores with factors	0.97	0.89
Multiple R square of scores with factors	0.94	0.80
Minimum correlation of possible factor scores	0.88	0.60

可以看到两个因子解释了 68%的方差。

第一个因子仍然是各个指标的综合，可以看成是整体口味的因子；第二个因子则是番茄的甜鲜方面的因子。

此时的因子的意义不是特别的明显（其实是之前主成分分析的负值），所以可以尝试使用因子旋转看是否有助于因子的解释。

3. 因子旋转及解释

采用正交旋转法提取因子。

```
> fa.varimax = fa(correlations,nfactors = 2,rotate = "varimax",fm = "ml")
> fa.varimax
Factor Analysis using method = ml
Call: fa(r = correlations, nfactors = 2, rotate = "varimax", fm = "ml")
Standardized loadings (pattern matrix) based upon correlation matrix
```

	ML1	ML2	h2	u2	com
TEXTURE	-0.03	0.57	0.32	0.68	1.0
Sweetness	0.16	0.87	0.79	0.21	1.1
Sourness	0.87	0.06	0.75	0.25	1.0
Salty	0.83	0.44	0.89	0.11	1.5
Bitter	0.84	-0.06	0.71	0.29	1.0
Umami	0.56	0.58	0.65	0.35	2.0

	ML1	ML2
SS loadings	2.49	1.61
Proportion Var	0.42	0.27
Cumulative Var	0.42	0.68
Proportion Explained	0.61	0.39
Cumulative Proportion	0.61	1.00


```
Mean item complexity = 1.3
Test of the hypothesis that 2 factors are sufficient.

The degrees of freedom for the null model are 15 and the objective function was 3.69
The degrees of freedom for the model are 4 and the objective function was 0.3

The root mean square of the residuals (RMSR) is 0.05
The df corrected root mean square of the residuals is 0.09

Fit based upon off diagonal values = 0.99
Measures of factor score adequacy
```

	ML1	ML2
Correlation of (regression) scores with factors	0.95	0.91
Multiple R square of scores with factors	0.91	0.83
Minimum correlation of possible factor scores	0.81	0.66

此时结果很明显可以看出，整个番茄的口味的数据(6个方面)可以看做有两个因子在主导，第一个因子是酸苦口感方面的因子，第二个是甜鲜方面的因子。

五、 典型相关分析

1. 典型相关分析

CCA 得到的 A 矩阵：

```
> ca$cor
[1] 0.9847824 0.9735600 0.9576218 0.9127072 0.8829802 0.8084332
> ca$xccoef
      [,1]      [,2]      [,3]      [,4]      [,5]      [,6]
TEXTURE 0.04678406 -0.098317226 0.04028193 0.014078735 0.01707418 5.776635e-02
Sweetness 0.04836981 0.081952606 0.02939660 -0.028002074 0.12337881 7.051985e-05
Sourness 0.10724433 -0.029540985 -0.07444456 -0.030838482 0.02342441 -1.364288e-01
Salty -0.03970483 0.050547802 -0.04532449 -0.050022077 -0.13409635 1.660876e-01
Bitter 0.01058637 0.002165150 0.04019064 0.141525361 0.10139220 7.857270e-02
Umami -0.02107454 -0.004652506 0.07477872 0.009230914 -0.10467119 -1.343120e-01
```

下面是 B 矩阵。

```
> ca$ycoef[,1:6]
      [,1]      [,2]      [,3]      [,4]      [,5]      [,6]
soluble.solds 0.0227271984 -8.815088e-02 0.125728807 -0.130578574 -0.152083445 1.243363e-02
citric 0.0447304866 -1.611054e-02 -0.024503258 0.078973756 0.020468680 1.093534e-02
malic.acid -0.0091504367 3.743771e-02 -0.005452605 0.057496883 0.010562521 -3.906087e-03
glucose -0.0714306125 7.008748e-02 -0.118247556 -0.040688427 0.147044450 2.767158e-02
fructose 0.0009381404 1.322696e-01 0.036838362 0.060985010 0.007436612 -8.196063e-02
glutamic.acid 0.0340606490 -4.597299e-02 0.007357593 0.031737597 0.022035994 8.731186e-03
X1.nitro.2.phenylethane 0.0079877847 -6.306656e-02 0.001616796 0.051622435 0.090376807 2.595641e-02
X1.nitro.3.methylbutane -0.0463076275 9.181356e-04 0.035332663 0.152187399 0.130178907 -5.962903e-02
X1.octen.3.one 0.0430211178 -1.406709e-02 0.002051577 -0.080785757 -0.008035650 -8.918305e-02
X1.pentanol -0.0773828072 7.360016e-02 0.031420001 0.024517118 -0.112830320 -1.514850e-02
X1.penten.3.ol -0.0240452341 4.310626e-02 -0.054788890 0.084234902 0.102249721 -8.685613e-02
X1.penten.3.one -0.0094333230 -2.188554e-02 -0.088998703 0.098231715 0.106293931 7.047488e-02
X2.5.dimethyl.4.hydroxy.3.2H..furanone 0.0975271779 -2.265765e-01 -0.012488043 0.049873798 0.071596598 1.364108e-01
X2.butylacetate -0.0112923237 -5.236957e-03 0.045061762 -0.001391411 0.048398377 -4.208883e-02
X2.ethylfuran 0.0079005416 -2.172009e-02 0.002671468 -0.020138874 0.045148205 1.194582e-01
X2.isobutylthiazole 0.0742988905 -9.039790e-02 0.002832051 -0.085533885 -0.041714661 5.641802e-03
X2.methyl.1.butanol 0.0267165577 -6.32867e-02 0.012950066 0.068302715 0.008820413 1.269727e-01
X2.methyl.2.butenal -0.0540367472 1.064314e-01 0.037747376 0.016079040 -0.142967728 -1.342409e-01
X2.methylbutyl.acetate -0.0837786426 4.391960e-02 -0.052012544 -0.253323915 0.075110752 -6.178294e-02
X2.methylbutyraldehyde 0.0213077704 -5.079939e-02 0.024032868 0.067554096 0.082791198 -1.970187e-02
X2.phenyl.ethanol -0.0048325703 1.449321e-02 0.036917905 0.068118599 0.071516990 -1.476584e-01
X3.methyl.1.butanol 0.1720841261 -1.527045e-01 0.051878327 0.041964801 0.034918003 -1.410511e-01
X3.methyl.1.pentanol 0.0293206341 5.617653e-02 0.027048995 0.065250589 -0.076995120 -5.801514e-02
X3.methyl.2.butenal -0.0138235017 -1.045323e-02 -0.008525990 0.044934476 0.116796667 6.399212e-06
X3.pentanone 0.0640459027 2.675079e-02 0.056845020 -0.035129270 -0.123224950 -7.538895e-02
X4.carene -0.0236699217 9.109343e-02 -0.012797611 0.045130189 0.004949603 -3.182263e-02
X6.methyl.5.hepten.2.ol 0.0108942051 -3.490098e-02 0.002955933 0.008838378 -0.067759152 1.358635e-01
X6.methyl.5.hepten.2.one -0.0074048718 2.423975e-02 0.151264373 0.130567292 0.008379537 -6.814183e-02
b.cyclocitral -0.0169037565 1.759939e-02 -0.049570310 -0.089958730 -0.010295962 -7.816802e-02
b.damascenone -0.0041161140 -1.743116e-02 -0.018796157 0.022210740 0.010345734 1.273623e-02
benzaldehyde 0.0285046485 -8.265314e-02 0.036683187 0.041807317 0.006699711 2.197063e-02
benzothiazole 0.0242855786 6.627738e-02 -0.007721603 -0.081040835 -0.009516969 -7.724421e-02
benzyl.alcohol -0.0264037072 -9.630979e-04 -0.017729119 -0.026046143 0.009961123 3.518818e-02
benzyl.cyanide 0.0124859139 -1.179314e-02 -0.022969253 -0.033322922 -0.077615954 8.552388e-03
b.ionone 0.0093134914 -2.785388e-03 0.003615352 0.073167765 0.080740038 -3.576953e-02
butyl.acetate 0.0408810010 3.513055e-02 0.008589687 -0.141176830 -0.213631944 2.676728e-02
cis.2.penten.1.ol -0.0165936443 -1.315173e-01 -0.031880282 -0.083579112 0.010431780 1.216088e-01
cis.3.hexen.1.ol -0.0192296094 -3.259732e-02 -0.045956655 -0.233036196 -0.094662340 1.358428e-01
cis.3.hexenal 0.0206766182 -3.183407e-02 0.046191243 0.028677234 0.053746118 2.932437e-02
cis.3.hexenyl.acetate -0.0432433786 2.777997e-02 -0.039773401 0.346630104 0.222397931 -4.692456e-02
cis.4.decenol 0.0153079643 4.251382e-02 0.042994321 -0.120663331 -0.006873370 6.247454e-02
eugenol -0.0105427657 3.157444e-02 -0.001103365 0.032426364 -0.009587350 -2.152802e-02
geranial -0.0065304060 8.599869e-02 -0.122225799 0.025086105 -0.010295962 -7.816802e-02
geranylacetone -0.0188749855 3.059856e-02 -0.052713863 -0.026992328 0.002168229 -4.900958e-03
guaiacol 0.0077948440 -3.374245e-03 0.043343301 -0.053292237 -0.065605605 8.896516e-03
heptaldehyde -0.0093448920 3.874868e-02 -0.031476825 0.051715720 0.025139195 -9.385257e-03
hexanal 0.0307915478 -3.382932e-02 0.022612180 -0.148197157 -0.065636684 -9.304605e-02
hexyl.acetate -0.0024882122 -9.624392e-02 0.038394099 -0.095763868 -0.095174854 3.180683e-02
hexyl.alcohol 0.0524322619 -1.002890e-01 0.037755529 0.189892518 0.111837066 6.111390e-03
isobutyl.acetate 0.0717034443 6.932486e-03 -0.017177937 0.232769258 0.028467617 2.325164e-02
isopentyl.acetate -0.1111951339 1.015584e-02 0.113746957 -0.041297038 -0.169446635 2.033432e-01
isovaleraldehyde -0.0202584286 1.381945e-01 -0.127739987 -0.036212027 -0.031059218 -9.59275e-02
isovaleric.acid -0.0022939498 -1.340653e-02 0.022318305 -0.068843627 -0.012611383 3.677451e-02
isovaleronitrile -0.0318526352 -6.666100e-03 -0.017792184 -0.122829879 -0.054186323 1.566077e-01
methional -0.0466114553 5.326905e-02 -0.019150229 0.021307688 0.049693937 -4.693824e-02
methylsalicylate -0.0150673661 1.430450e-02 -0.030417241 0.043941741 -0.053808432 3.108967e-02
neral 0.0011550903 -1.637691e-02 0.011006614 -0.012247648 0.001778018 -9.521368e-03
nonyl.aldehyde -0.0038412597 -7.259197e-03 0.009102368 -0.014401133 0.008529236 -2.494402e-02
p.anisaldehyde 0.0357190451 6.984250e-03 -0.026933118 0.022284388 -0.014952616 -5.201571e-02
phenylacetaldehyde 0.0040114105 -1.865389e-02 0.006536981 -0.140984879 -0.147266720 1.093673e-01
prenyl.acetate 0.0357411541 -1.244444e-05 0.048263542 -0.059880499 -0.061240268 -4.650595e-02
propyl.acetate -0.0452897524 -2.404864e-02 -0.071552650 0.021150247 0.202686956 2.031037e-02
salicylaldehyde -0.0170574713 -6.333382e-02 -0.011499948 -0.022227854 0.014380436 3.544088e-02
trans.trans.2.4.decadienal 0.0053332328 8.280302e-02 -0.033128123 -0.061339406 0.012651903 -1.715827e-01
trans.2.heptenal -0.0967938776 1.735863e-01 -0.028099044 0.020330030 -0.075319517 6.856745e-02
trans.2.hexenal -0.0401985746 5.809912e-02 0.036535798 -0.041348108 -0.046819033 -4.426770e-02
trans.2.pentenal 0.0191658502 2.931733e-02 0.118767880 -0.051607236 -0.109880636 -7.914100e-02
trans.3.hexen.1.ol -0.0183994364 1.036051e-01 -0.068375384 -0.016205806 0.026183875 -6.160188e-02
```

```
> corcoef_test <- corcoef.test(r=ca$cor,n=20,p=3,q=3)
号7)
> corcoef_test
[1] 6
```

2. 解释

U1 主要是酸度，对应着 V1 则是 X3.methyl.1.butanol (0.172084126) 和 isopentyl.acetate (-0.1111951339) 这两种化学物质。即番茄中的 X3.methyl.1.butanol 物质为酸度的主导物质，量越大，酸度越高；isopentyl.acetate 物质质量越大，越不酸。

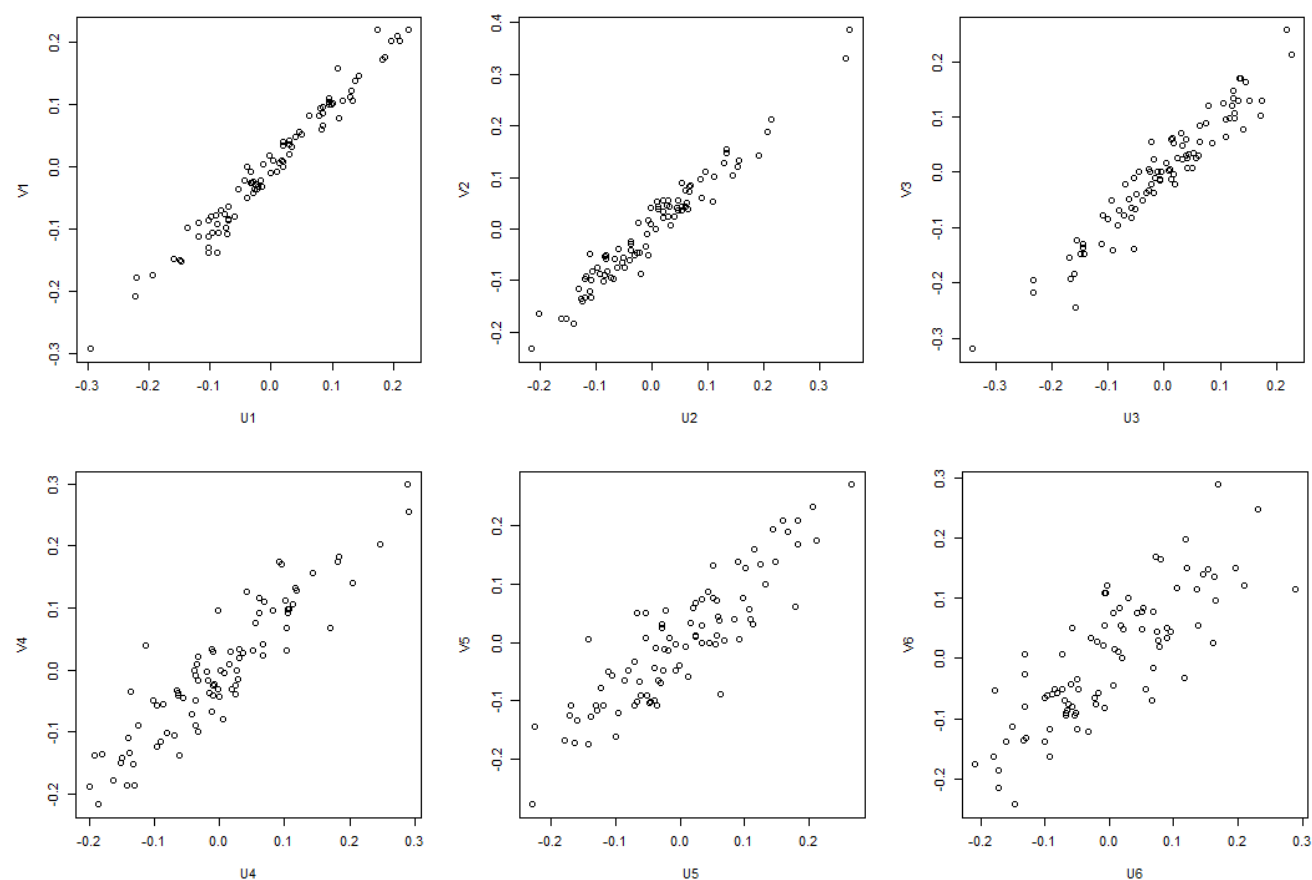
U2 主要是甜度和质地不好，对应着 V2 则是对应着 X2.5.dimethyl.4.hydroxy.3.2H..furanone (-2.265765e-01)，即番茄中对甜度和质地影响较大的物质是 X2.5.dimethyl.4.hydroxy.3.2H..furanone，该物质质量越大，甜度越高，质地越不好。

U3 主要是鲜味和不酸的程度，对应着的 V3 则是 Soluble.solds (0.1257288) X6.methyl.5.hepten.2.one (0.1512644) geranial (-0.1222258) isovaleraldehyde (-0.1277400) 四种主要化学物质的影响。

U4 主要是苦味，根据 V4 得到主要影响番茄苦味的化学物质是 X2.methylbutyl.acetate (-0.2533239)，cis.3.hexen.1.ol (-0.2330362)，cis.3.hexenyl.acetate (0.2327693) 和 isobutyl.acetate (0.3466301)。

3. 有效性检验

根据典型相关系数的显著性检验，选择前 6 对相关系数对，U 和 V 值见 Rmarkdown 文件。画出前六对的分布图如下，可以看出相关性十分明显。



六、 总结

本样例分析在对数据进行了初步简单和回归分析后,由于化学成分数据组的属性个数太多导致样本量不足以支撑做主成分分析或者是因子分析,因此主成分分析和因子分析仅涉及味道数据组,典型相关分析涉及两个数据组之间的联系。得到的结论分别是:

- 1) 番茄的 6 种评价数据(质地,甜度,酸度,鲜味,咸味,苦味)最终可以简化成两个综合指标即味道的寡淡程度(整体风味强度)和人们胃口厌恶程度(人们喜爱程度)。
- 2) 番茄的 6 种评价数据最终可以有两个因子来解释,一个是酸苦口感因子,一个是甜鲜口感因子。
- 3) 番茄的味道属性分别受对应的化学物质的控制(具体见典型相关分析的解释部分)。