



R在烟草及其加工制品研发中的应用

上海烟草集团有限责任公司
余苓

2011-11-12



1

简介

2

多元数据的直观分析

3

多元数据的相关分析

4

多元数据的降维分析



烟草数据特征

感官评吸、辅材设计、工艺参数、理化检测和市场分析

品牌	地域	...	类别1		...	类别2		...
			化合物1	化合物2		化合物3	化合物4	
中华								
红双喜								
...								

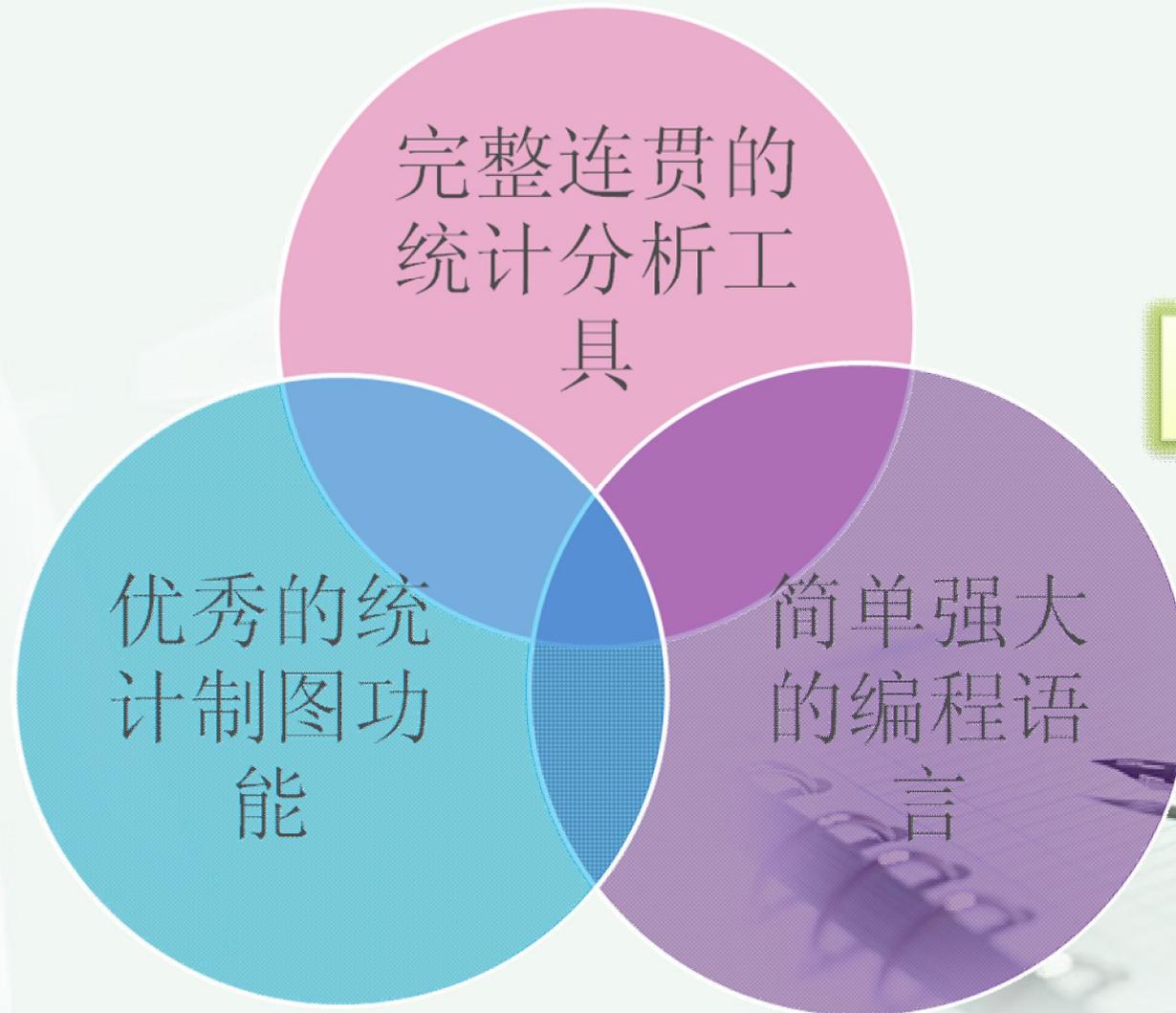
多指标

多类别

多元数据



R是一套完整的数据处理、计算和制图软件系统



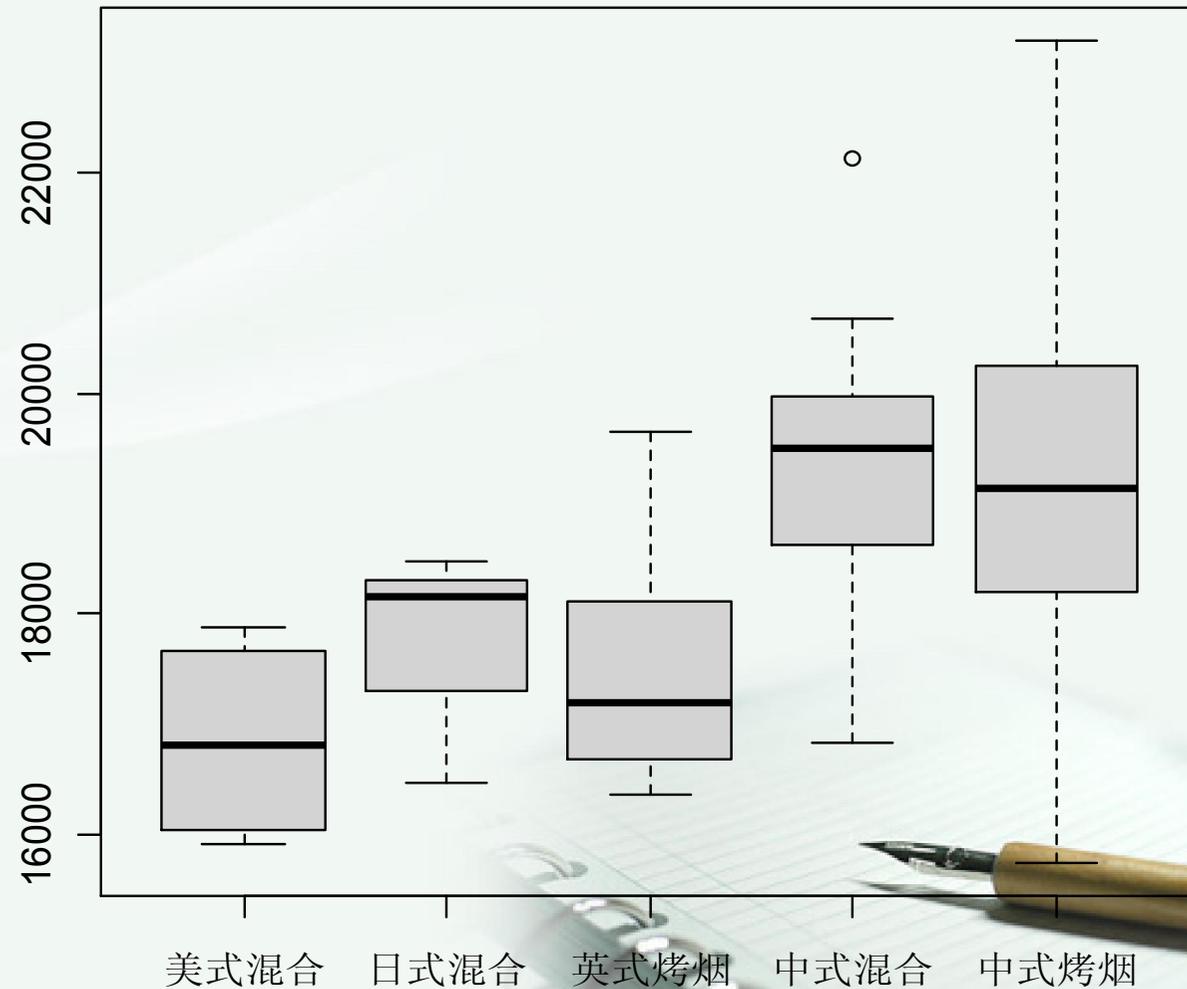
开源



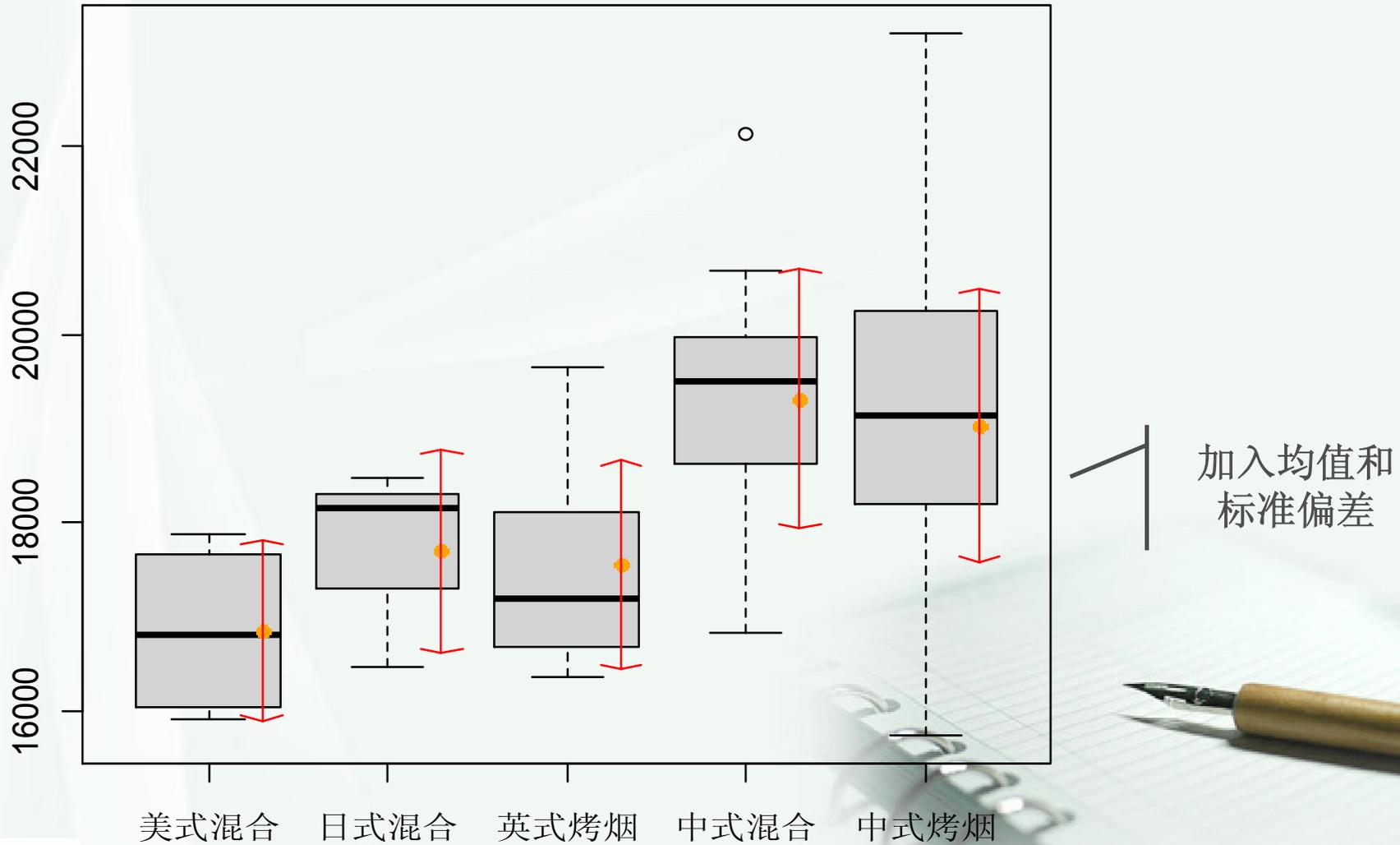
分布特征

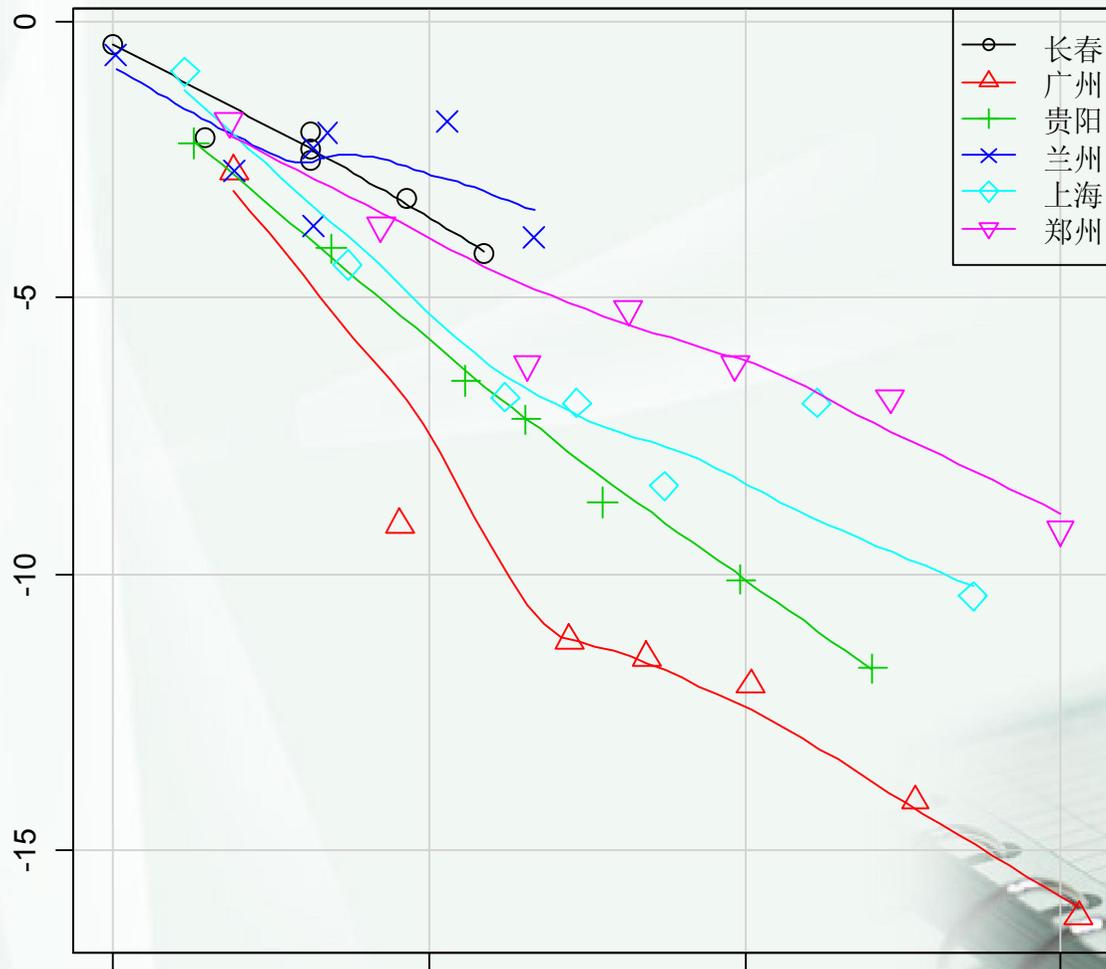
四分位数
异常值

不同类型卷
烟比较分析



R在烟草中的应用2011/11/12



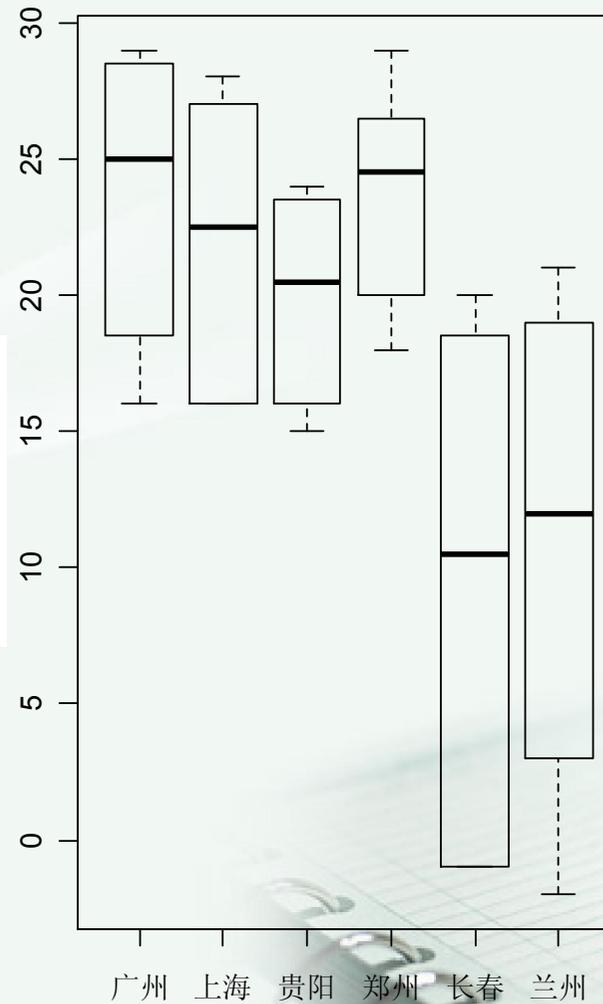
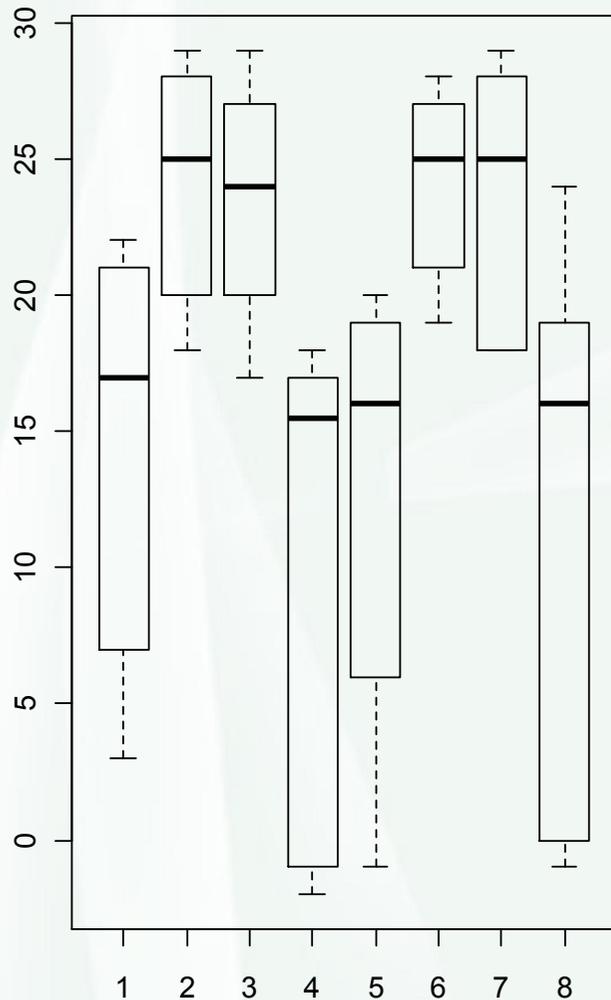


Lowess

评吸数据
地域特性



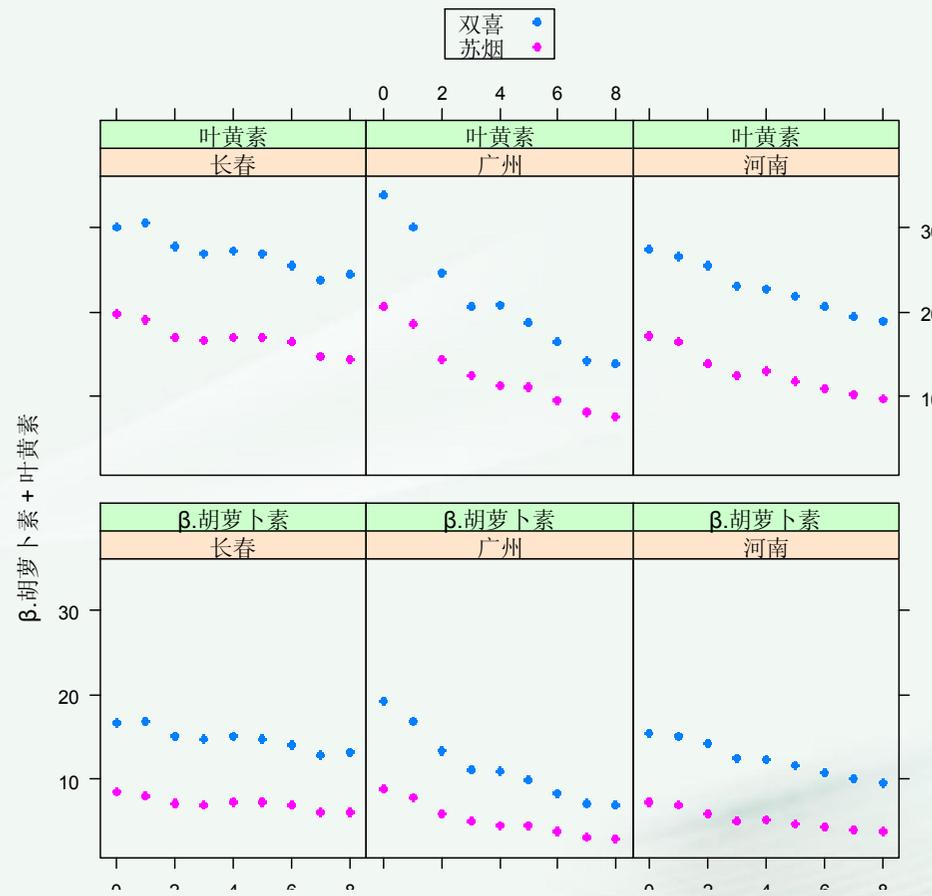
R在烟草中的应用2011/11/12



生产温湿度地域和季节特性



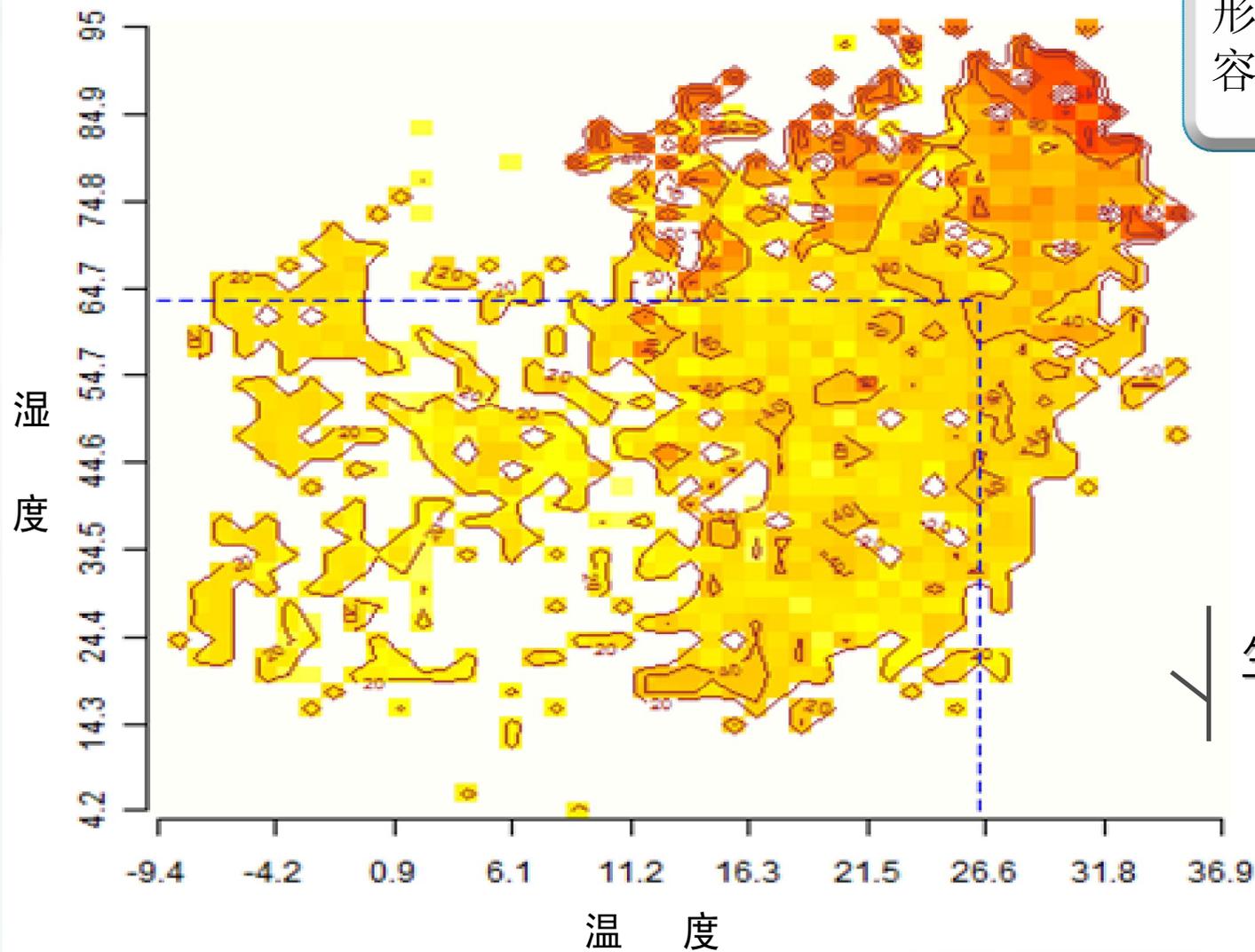
R在烟草中的应用2011/11/12



化学数据地域特性



Lattice包



形象展示数据，
容易找到规律

生产环境分析



两种工艺处理前后指标脸谱图

工艺1-前 AVG



工艺1-后 AVG



工艺2-前 AVG



工艺2-后 AVG



每个脸部特征代表
一个质量指标

活泼生动，
好玩而又不
失规律性

祥瑞祥和
吉祥如意



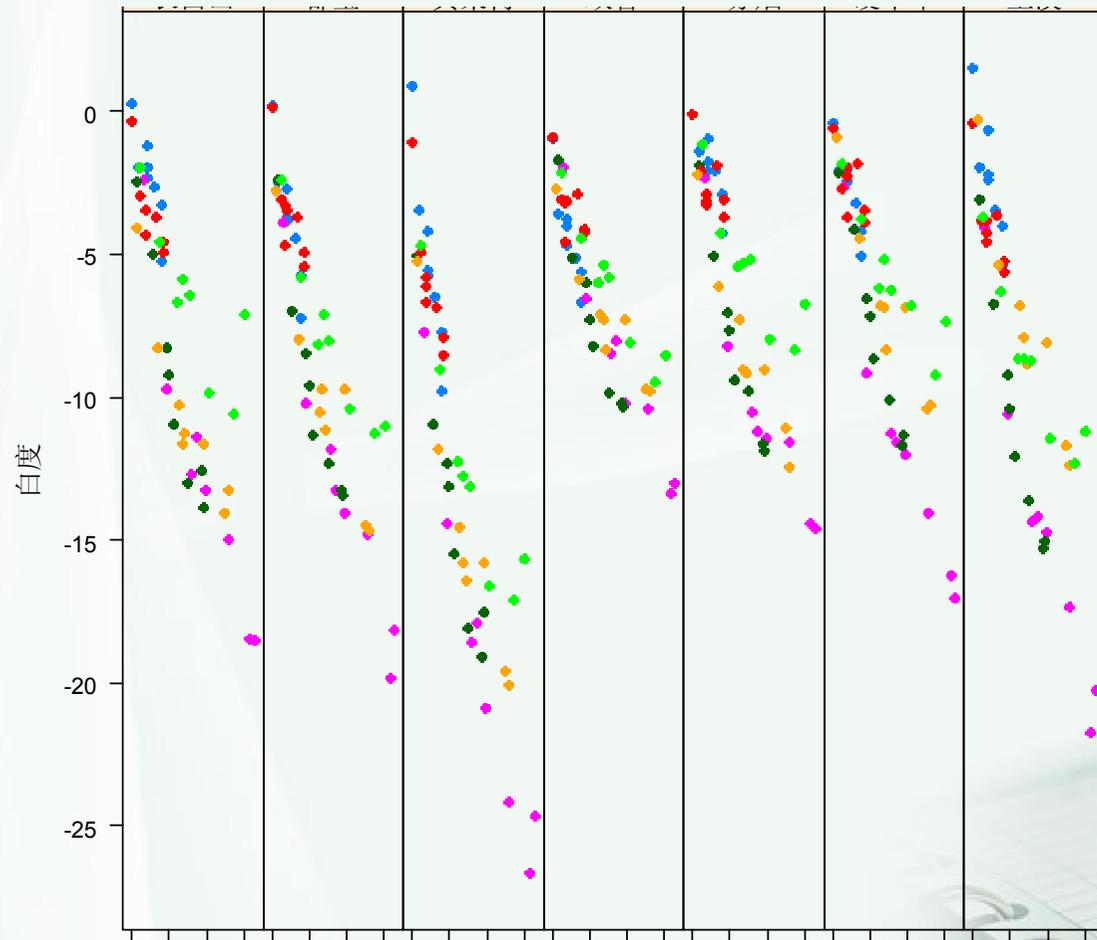
工艺1处理前后化合物的变化率和显著性

64 ***	65 ***	66 ***	67 ***					
55 **	56 ***	57 **	58 ***	59 **	60 ***	61	62 *	63 ***
46	47 ***	48 ***	49 **	50 ***	51 ***	52 **	53 ***	54 ***
37	38 ***	39	40 *	41 ***	42 ***	43 ***	44	45 ***
28 **	29 *	30 **	31 **	32	33 ***	34 **	35 ***	36
19 ***	20 ***	21 *	22 **	23 **	24 **	25 ***	26 *	27 *
10 ***	11 ***	12 ***	13	14 *	15 **	16 ***	17 *	18
1 ***	2 **	3 *	4 **	5 ***	6 ***	7 **	8 ***	9 ***

颜色 浅—深
变化率 小—大

星号个数代表变化的
显著性





不同品牌分析

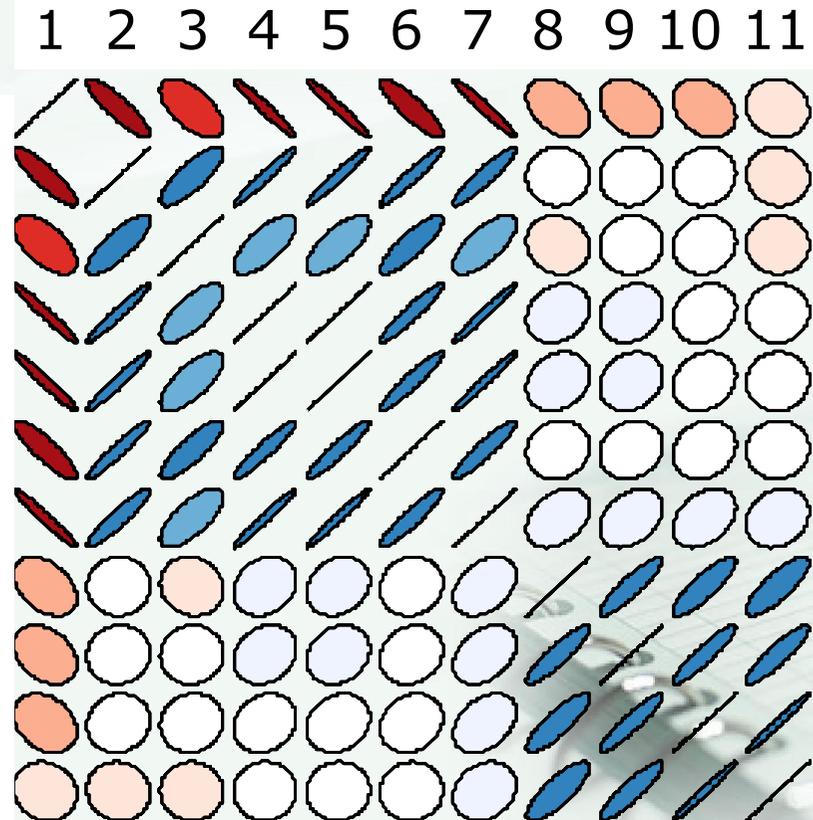


分析变量之间的关系

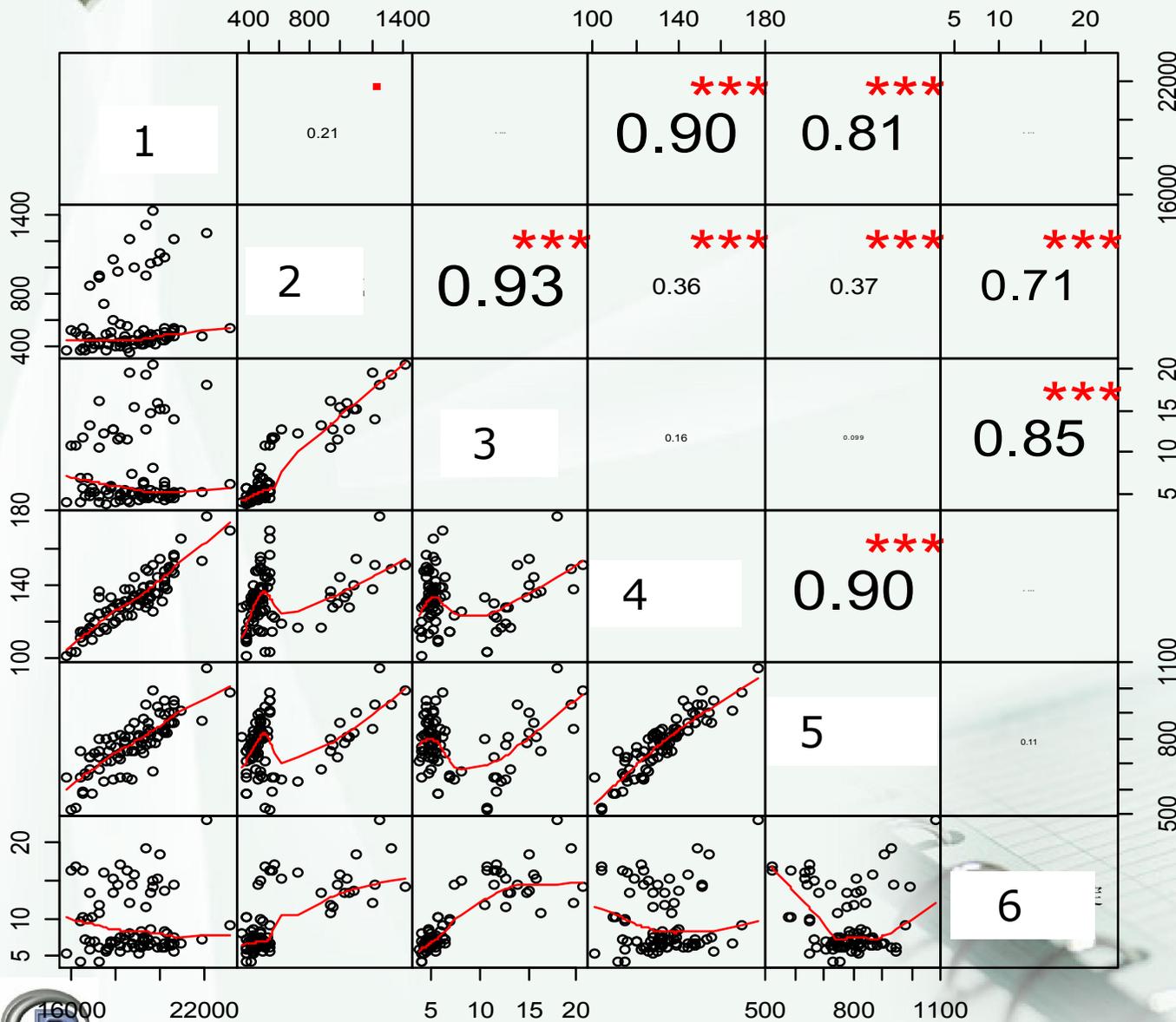
向右倾斜代表为正;
向左倾斜代表为负;

颜色只是为了表现
形象,无实际意义。

1
2
3
4
5
6
7
8
9
10
11



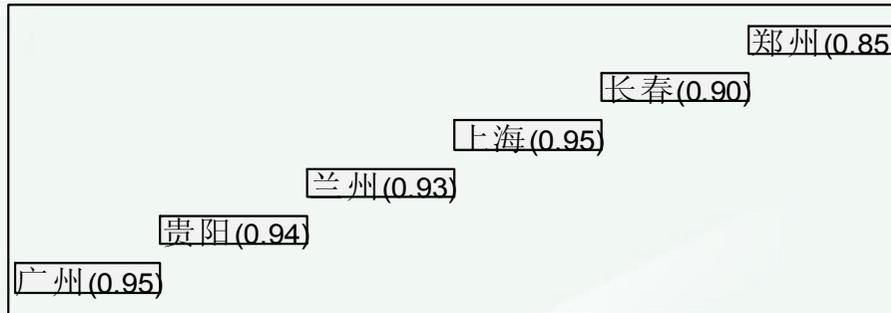
R在烟草中的应用2011/11/12



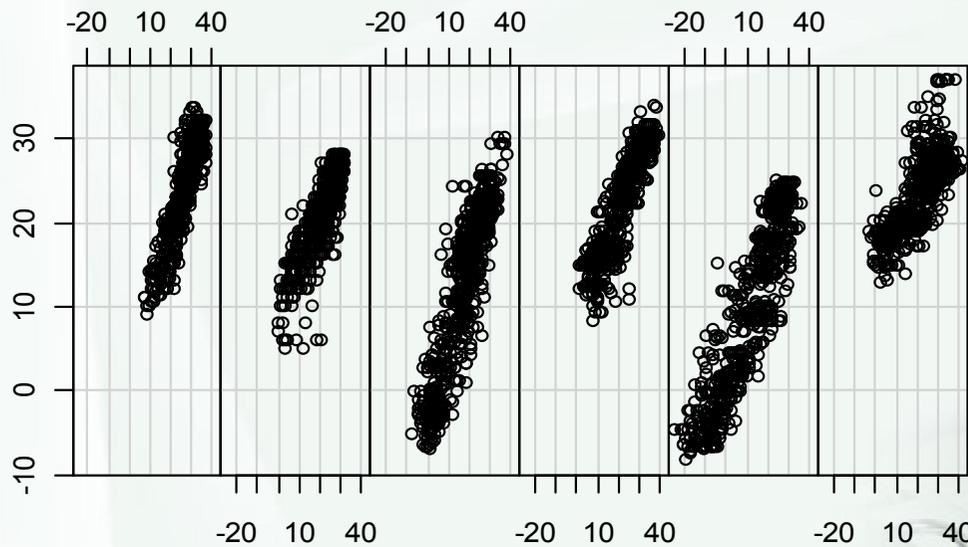
多
信
息
量

散点图+拟合曲线
 相关系数(数字展示+大小代表)
 显著性水平(星号)





生产环境：温度日值



气象数据：最高温度日值

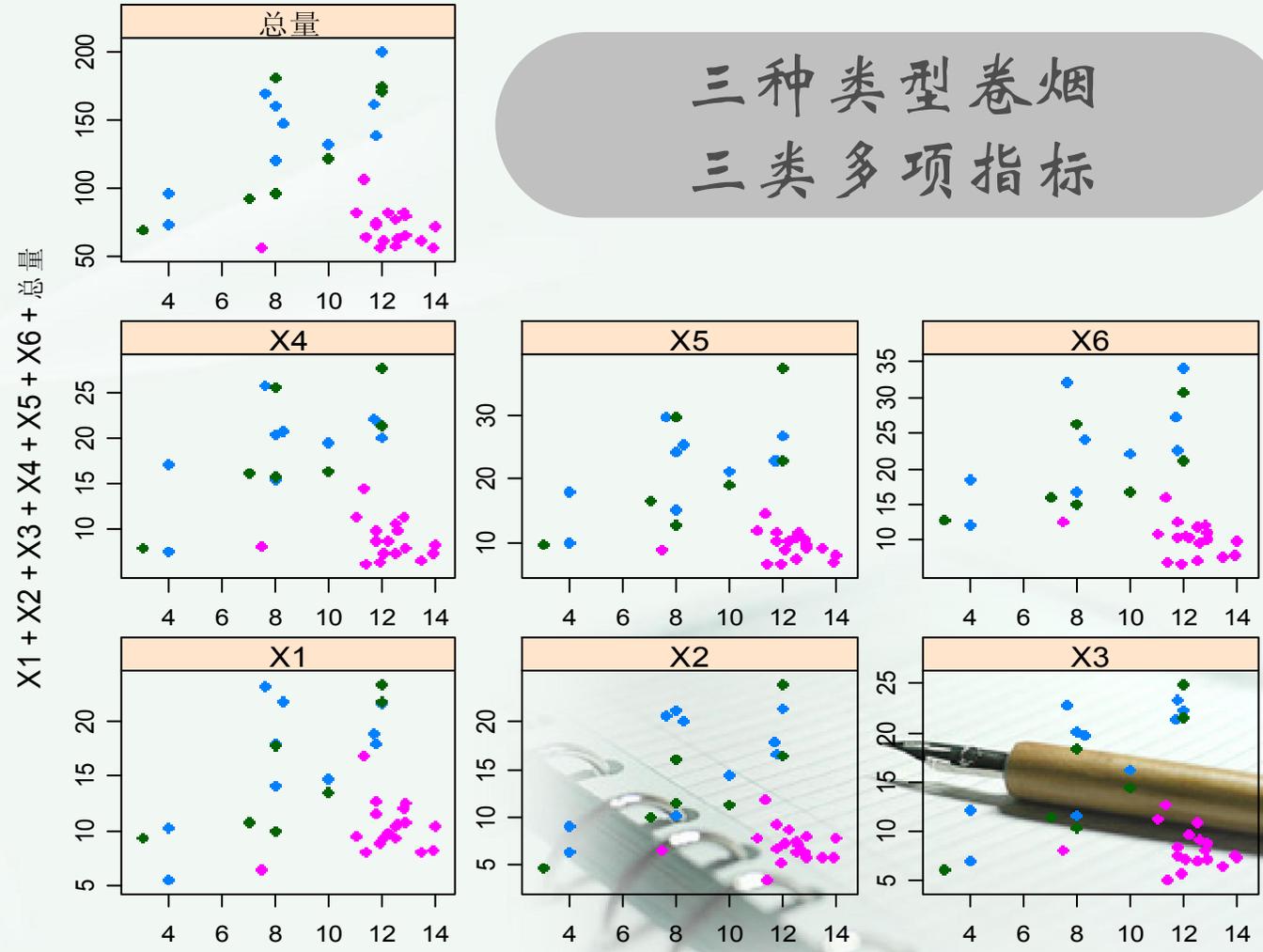
生产温度与气象温度相关性



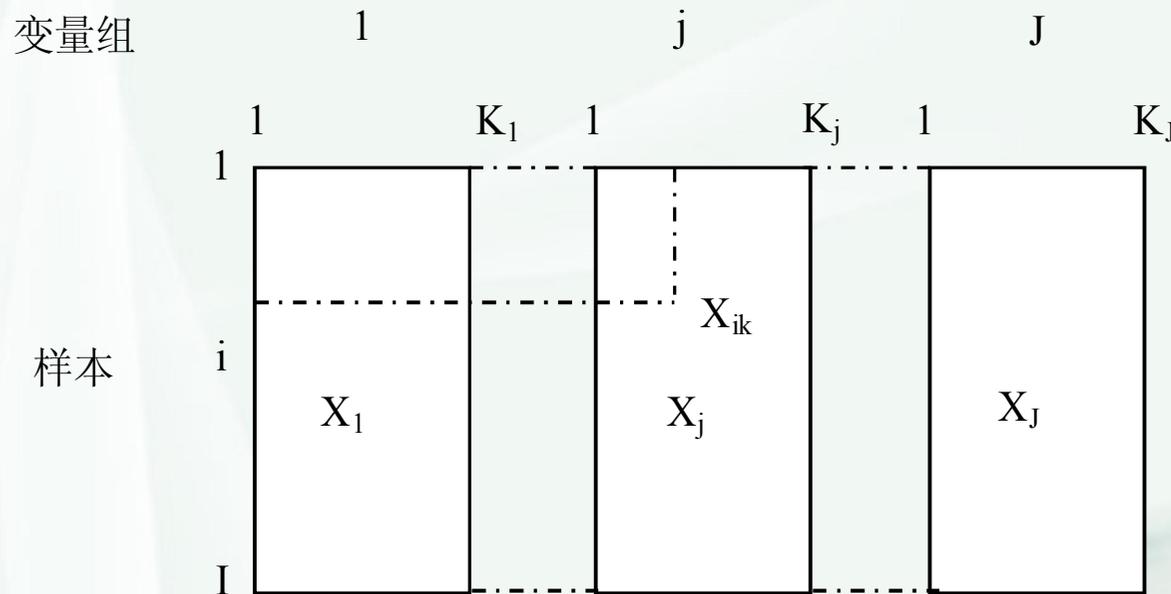
多元数据的降维



三种类型卷烟
三类多项指标



MFA—multiple factor analysis

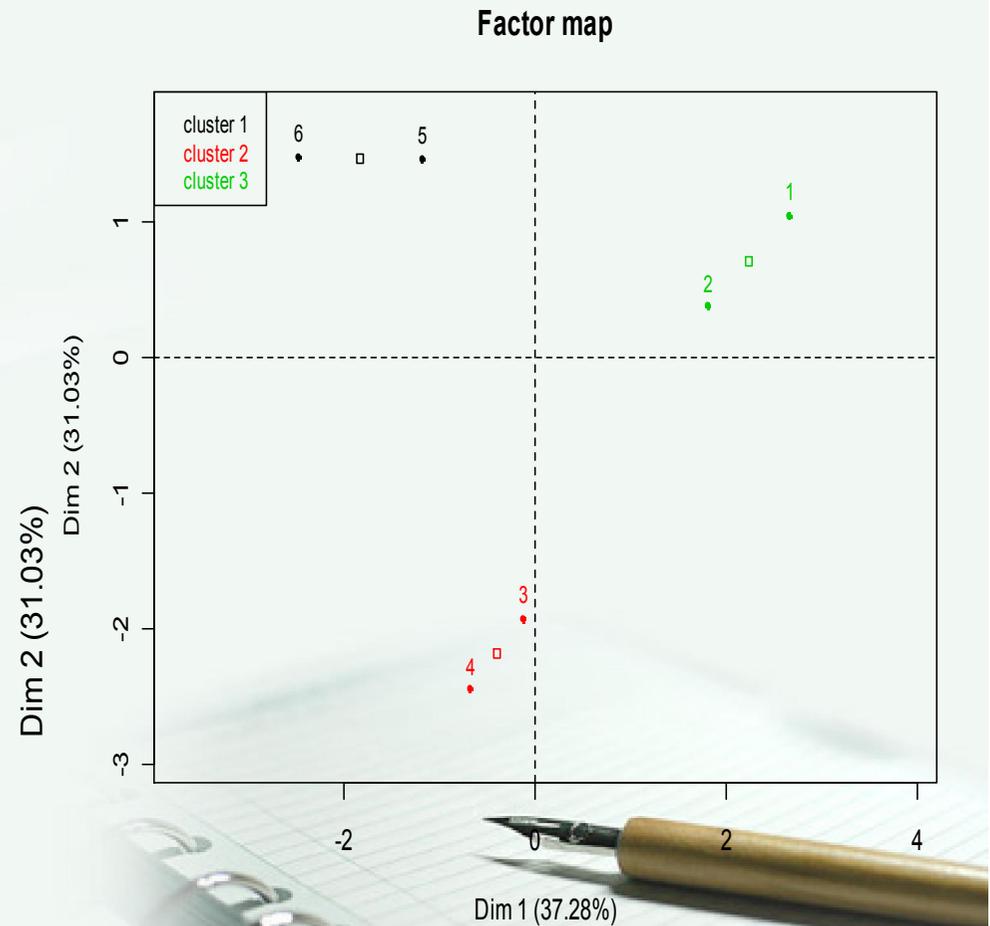
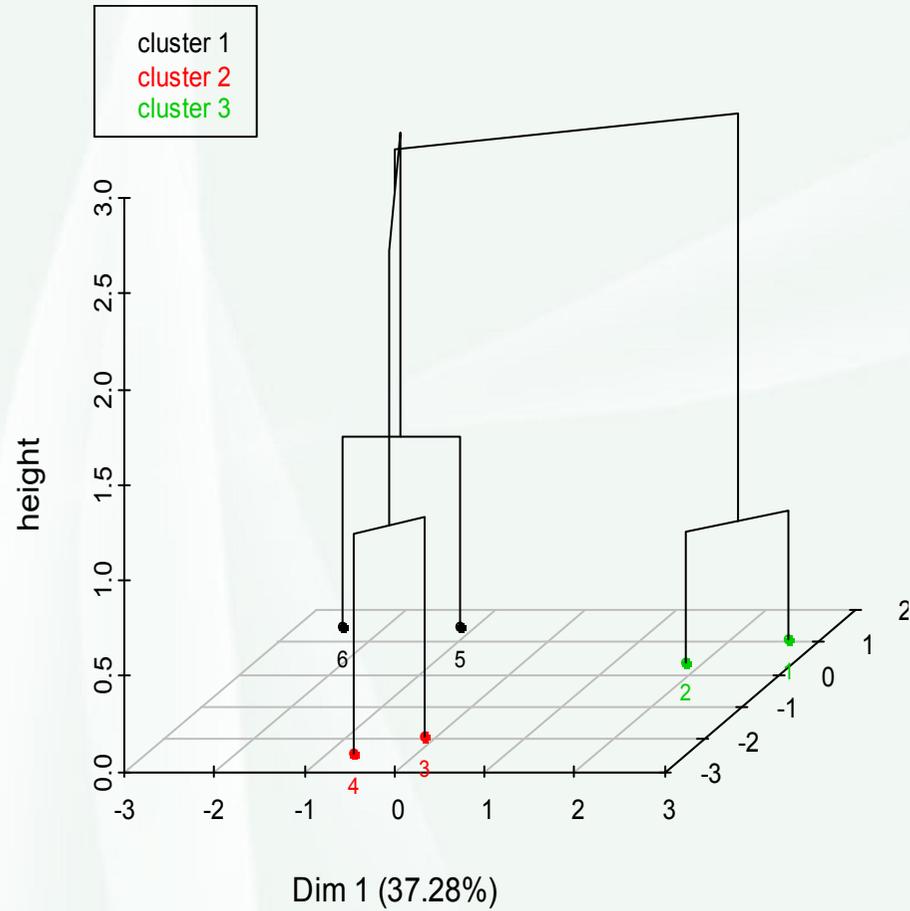


权重PCA
实现多组变量
之间的共性和
差异的比较

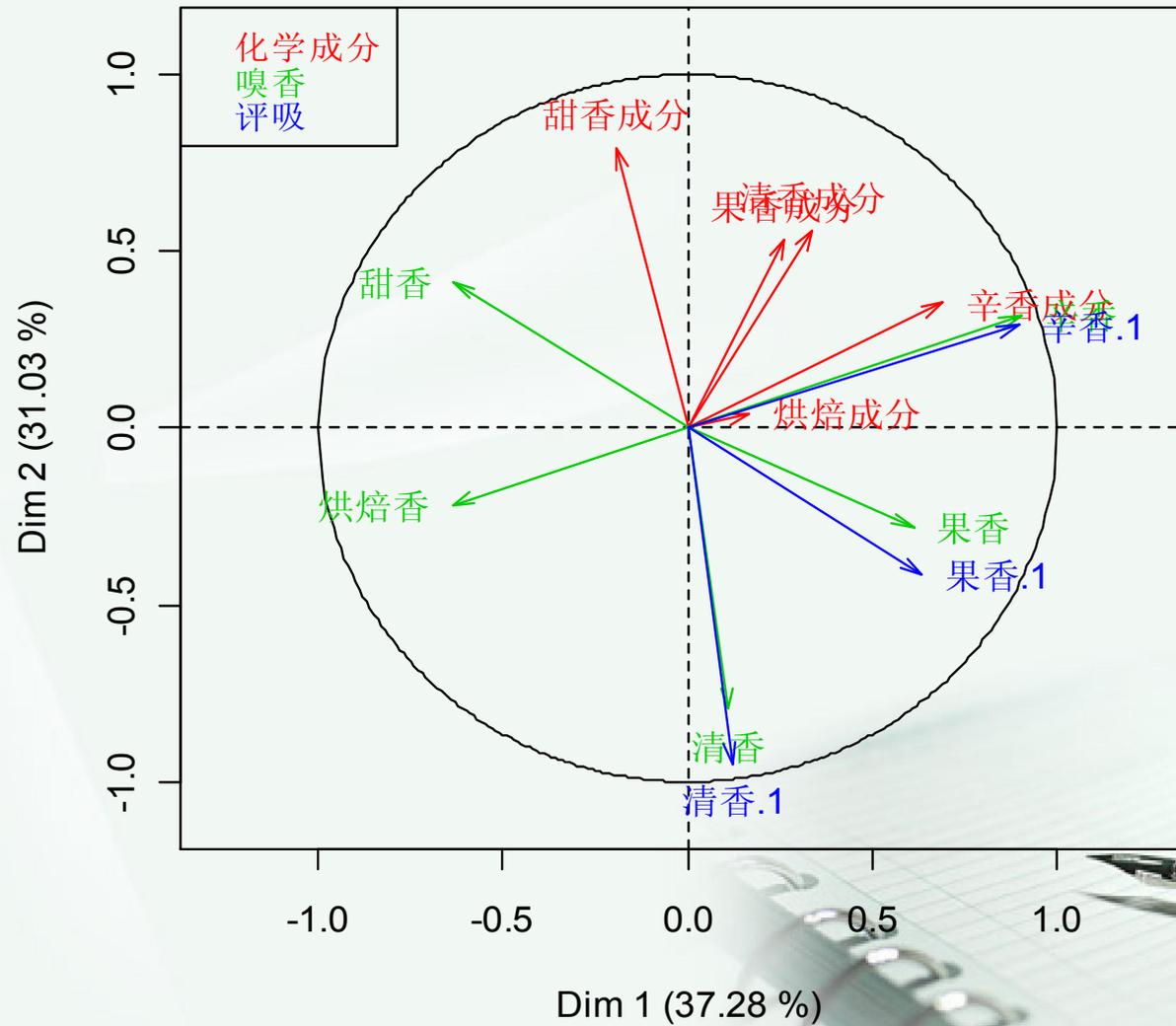
$$X = \left[\frac{1}{\sqrt{\lambda_{11}}} X_1 \cdot \cdot \cdot \frac{1}{\sqrt{\lambda_{j1}}} X_j \cdot \cdot \cdot \frac{1}{\sqrt{\lambda_{J1}}} X_J \right]$$



Hierarchical clustering on the factor map



R在烟草中的应用2011/11/12





Thank You !

