

中图分类号: S532 文献标识码: B 文章编号: 1672-3635(2015)02-0120-05

基于因子分析的马铃薯淀粉质量指标的关联性研究

陈必琴*

(甘肃省定西市农产品质量安全管理站, 甘肃 定西 743000)

摘要: 为了提高定西市马铃薯淀粉质量, 促进淀粉加工企业向品牌化健康发展, 研究应用2007~2013年定西市连续7年15家马铃薯淀粉企业供检的105份淀粉质量检测数据, 建立SPSS数据录入系统。通过相关分析、因子分析和聚类分析研究了各项质量指标间的关联性, 利用马铃薯淀粉各变量间的相关系数矩阵和公因子方差, 以K-均值聚类法按提取的3个公因子对企业进行聚类。结果表明, 各项质量指标间存在着一定的相关性, 部分质量指标间存在着显著的关联性, 只需提取3个公因子就足以反映10个指标的信息。因此, 可根据最终聚类结果对马铃薯淀粉加工企业进行分类指导, 为定西市马铃薯淀粉生产的可持续发展提供决策参考。

关键词: 因子分析; 马铃薯淀粉; 质量指标; 关联

Correlation Research of Potato Starch Quality Indexes Based on Factor Analysis

CHEN Biqin*

(Dingxi Agricultural Products Quality and Safety Supervision and Management Station, Dingxi, Gansu 743000, China)

Abstract: One hundred and five data of ten starch quality indexes in 15 potato starch enterprises of Dingxi City in the period of 2007 to 2013 were collected and analyzed using SPSS software in order to improve the quality of potato starch and promote the healthy development of starch processing enterprises. The correlation between each quality index was studied through correlation analysis, factor analysis and cluster analysis, and cluster analysis was made using the correlation coefficient matrix and the common factor variance, and K-means clustering method based on the three extracted common factors of enterprises cluster. The results showed that correlation between each quality index existed to some extent, with some correlation significant, and the three extracted common factors was enough to reflect the information of ten indexes. Therefore, some suggestions would be proposed for potato starch processing enterprises according to the final clustering results, and decision-making reference provided for the sustainable development of potato starch production in Dingxi City.

Key Words: factor analysis; potato starch; quality index; correlation

马铃薯是甘肃省定西市的第一大支柱产业。多年来, 定西市围绕打造“马铃薯薯都”这一宏伟目标, 在马铃薯脱毒种薯普及、标准化生产、基地建设、马铃薯精深加工方面做了大量工作, 在带动农民增收、农业增效方面取得了明显成绩。马铃薯淀粉生产是定西马铃薯精深加工的基础, 全市先后有二十多家马铃薯加工企业从事淀粉生产, 年加工鲜薯50万t左右, 年产马铃薯淀粉7~

10万t, 有力地消化了全市马铃薯鲜薯, 发展壮大了加工企业, 发挥了龙头加工企业的良好带头作用。为提高马铃薯淀粉产品质量, 保障食品安全, 促进全市精淀粉加工企业向品牌化健康发展, 定西市农产品质量安全管理站自2004年开始就对全市重点淀粉生产企业产品进行检测服务和质量跟踪, 并通过对检测数据的分析, 为政府制定产业发展规划提供了科学依据。因此,

收稿日期: 2014-11-11

作者简介: 陈必琴(1978-), 女, 农艺师, 主要从事农产品质量安全监测检验工作。

*通信作者(Corresponding author): 陈必琴, E-mail: chenbq2008y@126.com。

在对 15 家企业连续 7 年的淀粉质量检测结果的研究基础上,借助因子分析和聚类分析,研究马铃薯淀粉质量各项指标的关联性及其影响淀粉质量指标的潜在因素,为进一步提高全市马铃薯淀粉质量,促进马铃薯淀粉向品牌化发展有着十分重要的现实意义。

1 材料与方法

1.1 分析材料

分析材料来源于定西市农产品质量安全监督管理站,从 2007~2013 年对 15 家淀粉加工企业供检的每年每家 3 份 7 年 315 份马铃薯淀粉质量,按照《马铃薯淀粉》GB/T 8884-2007^[1]标准方法检测的数据,分析采用的是将每年的 3 份数据进行平均而得到的 105 份数据。其中每份样品涉及的质量指标主要有白度、粘度、细度、水分、斑点、灰分、二氧化硫、电导率、菌落总数和霉酵母总数 10 项。

1.2 数据处理

试验数据采用 Excel 和 SPSS 19.0^[2]软件进行统计分析和处理。采用主成分分析法提取公因子进行因子分析得出各因子得分,从而计算出各因子的综

合得分并排序,然后根据综合得分对各个企业的指标进行 R 型(变量)聚类。

2 结果与分析

2.1 各指标相关分析

将 7 年 15 家企业 10 个检测项目的 105 份检测数据在 Excel 表中构建成有序数据矩阵,将新矩阵导入 SPSS 19.0 统计分析软件进行均值和标准差变换及双变量相关分析,得到白度(X_1)、粘度(X_2)、细度(X_3)、水分(X_4)、斑点(X_5)、灰分(X_6)、二氧化硫(X_7)、电导率(X_8)、菌落总数(X_9)和霉酵母数(X_{10})等 10 种相关分析的相关系数矩阵(表 1)。

从表 1 可以看出, X_1 与其他因子的关联系数依次为: $G(1,2)=0.116$, $G(1,3)=0.286$, $G(1,4)=-0.135$, $G(1,5)=-0.689$, $G(1,6)=-0.277$, $G(1,7)=-0.018$, $G(1,8)=-0.365$, $G(1,9)=0.047$, $G(1,10)=0.109$ 。关联度绝对值排列为 $X_5 > X_8 > X_3 > X_6 > X_4 > X_2 > X_{10} > X_9 > X_7$,表明白度与其余 9 项主要质量指标的关联度各不相同,其中白度与斑点、灰分、电导率呈极显著负相关性,而与细度呈极显著正相关。其余各项指标的关系以此类推。

表 1 马铃薯淀粉各检项之间的相关系数矩阵

Table 1 Matrix of correlation coefficient between each inspection item of potato starch

相关性 Pearson	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
X_1	1									
X_2	0.116	1								
X_3	0.286**	-0.088	1							
X_4	-0.135	0.047	-0.159	1						
X_5	-0.689**	-0.120	-0.297**	0.184	1					
X_6	-0.277**	0.034	-0.234*	-0.111	0.221*	1				
X_7	-0.018	-0.531**	0.164	-0.093	-0.061	-0.148	1			
X_8	-0.365**	-0.557**	-0.009	-0.026	0.297**	0.033	0.480**	1		
X_9	0.047	-0.555**	0.167	-0.143	-0.106	-0.117	0.660**	0.519**	1	
X_{10}	0.109	-0.544**	0.186	-0.135	-0.167	-0.151	0.671**	0.481**	0.772**	1

注: *表示在 0.05 水平(双侧)上相关显著, **表示在 0.01 水平(双侧)上相关极显著, $N=105$ 。 X_1 -白度, X_2 -粘度, X_3 -细度, X_4 -水分, X_5 -斑点, X_6 -灰分, X_7 -SO₂ Oxygen Saturation, X_8 -电导率, X_9 -菌落总数, X_{10} -霉酵母数。

Note: * and ** denote significance at 0.05 and 0.01 levels of probability (two sides), $N=105$. X_1 -whiteness, X_2 -viscosity, X_3 -fineness, X_4 -water content, X_5 -spot, X_6 -ash, X_7 -SO₂ Oxygen Saturation, X_8 -conductivity, X_9 -total number of colony, and X_{10} -number of mold yeast.

2.2 因子分析

2.2.1 应用 KMO 检验和 Bartlett 球形检验做因子分析的适应性检验

KMO 是用来检验变量间的偏相关系数的大

小,其值越接近 1,说明变量间相关性越强,因子分析的效果越好; Bartlett 球形检验是用来判断相关阵是否为单位阵,如果为单位阵,则各变量独立,因子分析无效。

表2 KMO和Bartlett的检验
Table 2 KMO and Bartlett tests

检验方法 Test method	检验值 Test value
取样足够度的 Kaiser-Meyer-Olkin 度量	0.805
Bartlett 的球形度检验近似卡方	389.840
df	45
Sig.	0.000

该资料的 KMO 统计量为 0.805，Bartlett 检验：
C2 = 389.840，sig. = 0.000。2 种方法说明各变量间有
较强的相关性，适合进行因子分析。检验结果见表2。

表3 各公因子总方差解释
Table 3 Total variance of each common factor explained

公因子 Common factor	特征值 Characteristic value	占总方差(%) Percentage of total variance	累积(%) Cumulation
1	3.382	33.823	33.823
2	2.285	22.854	56.677
3	1.112	11.125	67.801

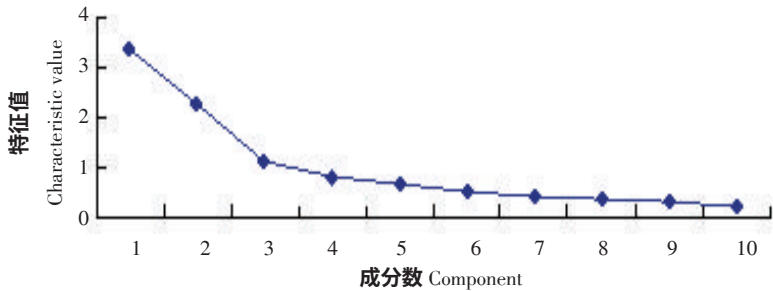


图1 特征根散点图
Figure 1 Eigenvalue scatter diagram

2.2.3 因子旋转

在此基础上，计算出10个变量对应于3个公因子的因子载荷矩阵，并对其进行最大方差正交旋转，产生旋转因子载荷矩阵，旋转的目的是使复杂的矩阵变得简洁。选取对各公因子贡献值较大(选

2.2.2 公因子的提取和判识

以特征根大于1为公因子提取原则，对公因子进行提取。共提取出3个公因子(包括7项指标)，累计贡献率为67.801%，即3个公因子已经包括了原来10个指标的绝大部分信息，损失的信息量较少，具有较高的代表性。旋转后的方差贡献率说明，公因子 F1、F2 和 F3 可以解释原始信息的能力分别是33.823%、22.854%和11.125%，结果见表3。

碎石图显示(图1)，前3个公因子的折线坡度较陡，而至第4个公因子以后就逐渐趋于平缓，该图从另一个侧面说明了取前3个因子为宜。

取相关系数绝对值>0.5)的变量。这些公因子是原来变量的线性组合，能够承载原来10个变量的大部分信息。结果见表4。

2.2.4 因子命名

由表4可知，主因子 F1 在 SO₂、菌落总数、霉

表4 旋转后对各个公因子作用较大的变量
Table 4 Variables of each common factor of larger role after rotation

公因子 Common factor	载荷变量 Load variable
1	粘度(-0.777)、SO ₂ (0.823)、电导率(0.746)、菌落总数(0.861)、霉菌菌数(0.849)
2	白度(0.856)、细度(0.549)、斑点(-0.854)
3	水分(0.811)、灰分(-0.667)

酵母数这3个指标上的因子载荷值最大,均超过82%,该因子反映淀粉在卫生指标方面的信息,这与GB/T 8884-2007规定的卫生指标是一致的,可命名为卫生指标因子;主因子 F_2 在白度和斑点这2个指标上的因子载荷最大,绝对值均超过85%,该因子反映精淀粉在感官方面的信息,可命名为感官要求因子;主因子 F_3 在水分指标上的因子载荷最大,超过80%,因此可命名为理化指标因子。

2.2.5 因子综合得分

通过因子分析,各因子的得分已保存在数据文件中,综合得分为 $F_{\text{综}} = 0.33823F_1 + 0.22854F_2 + 0.11125F_3$,由此可求得每家企业的综合得分及其排名,见表5(仅取前20名)。

2.3 聚类分析

根据3个因子得分函数,将SPSS自动计算所得的15家企业的3个因子得分作为新变量,对15家企业淀

表5 用因子分析法得到的综合评价

Table 5 Comprehensive evaluation using factor analysis

企业名称 Enterprise	卫生指标因子 F_1 Health indicator factor F_1	感官要求因子 F_2 Sensory requirement factor F_2	理化指标因子 F_3 Physicochemical index factor F_3	综合得分 $F_{\text{综}}$ Comprehensive score F_c	综合得分排名 Comprehensive score ranking
安陇 13 Anlong 13	3.371 45	0.769 39	0.195 52	1.34	1
陇清 10 Longqing 10	2.039 58	1.261 77	0.130 55	0.99	2
安蓝 13 Anlan 13	2.194 96	-0.615 14	2.905 92	0.93	3
安薯 07 Anshu 07	2.331 90	0.536 23	-0.113 00	0.90	4
安陇 10 Anlong 10	1.976 20	0.651 04	-0.067 26	0.81	5
安陇 09 Anlong 09	1.663 48	0.388 13	0.774 49	0.74	6
陇清 09 Longqing 09	1.387 26	0.614 73	0.695 45	0.69	7
陇凯 10 Longkai 10	1.507 77	0.481 35	0.247 65	0.65	8
安博 13 Anbo 13	0.992 84	1.387 74	-0.571 87	0.59	9
陇清 11 Longqing 11	1.515 53	0.241 08	0.179 33	0.59	10
安超 09 Anchao 09	1.365 35	0.737 26	-0.437 44	0.58	11
陇祥 08 Longxiang 08	1.081 52	1.642 44	-1.447 48	0.58	12
陇祥 09 Longxiang 09	0.788 46	0.929 71	0.670 38	0.55	13
陇祥 07 Longxiang 07	1.735 12	-0.198 28	0.010 91	0.54	14
陇清 07 Longqing 07	1.218 72	0.562 85	-0.466 97	0.49	15
安陇 12 Anlong 12	0.315 08	1.213 57	0.937 64	0.49	16
安超 13 Anchao 13	0.436 31	1.535 33	-0.172 20	0.48	17
安超 12 Anchao 12	1.459 19	0.089 03	-0.400 15	0.47	18
安陇 08 Anlong 08	1.331 57	-0.337 02	0.739 27	0.46	19
陇祥 13 Longxiang 13	1.462 22	0.900 47	-2.225 54	0.45	20

注:企业名称这一列如‘安陇 13’表示‘安陇’这个企业2013年的淀粉样品信息。下同。

Note: In column of enterprise, Anlong 13 denotes starch sample information of enterprise Anlong in 2013. The same below.

表6 最终聚类中心

Table 6 Final cluster center

公因子 Common factor	聚类 Clustering		
	1	2	3
F_1	1.261 40	-0.636 59	-0.015 52
F_2	0.459 73	0.293 09	-1.580 96
F_3	-0.295 50	0.102 27	0.144 22

粉质量指标进行K-means聚类分析,将15家企业7年的质量指标分成3类,聚类结果见表6和表7。

根据最终聚类中心情况(表6),根据分类的结果对淀粉企业进行指导,第一类企业在卫生指标方面有较强优势,而在理化指标方面较为薄弱;第二类企业在感官要求方面较强,而在卫生指标方面有待加强;第三类企业在理化指标方面稍有优势,而在感官要求方面有明显的劣势。

表7 每个聚类中的案例数及成员
Table 7 Case numbers and members in each cluster

聚类 Clustering	案例数 Case number	成员 Member
1	29	安陇13、陇清10、安蓝13、安薯07、安陇10、安陇09、陇清09、陇凯10、安博13、陇清11、安超09、陇祥08、陇祥09、陇祥07、陇清07、安超13、安超12、安陇08、陇祥13、陇凯08、陇祥10、安蓝12、陇凯13、安超07、安超08、渭海07、陇清08、渭海08、陇祥11
2	57	安陇12、安蓝11、临兴09、临腾12、安薯12、临盛12、安蓝10、安薯09、临盛13、安薯10、陇凯11、陇清13、临盛09、临兴08、陇祥12、临盛11、陇凯09、临腾11、临通11、临通12、安陇11、安博12、陇清12、安陇07、临通07、安宏09、临通10、临兴12、渭海10、临兴11、安博08、临盛10、安博11、安薯13、临腾13、安宏11、临兴10、安薯11、渭晶13、安博07、渭海13、临通13、安宏12、临兴13、安宏10、临腾09、陇凯12、安宏13、安博10、渭晶12、临腾10、安薯08、临兴07、临盛07、临腾07、临腾08、渭海11
3	19	安超11、安超10、安蓝07、渭晶08、安宏07、临盛08、安宏08、陇凯07、渭海09、渭海12、安蓝09、安蓝08、渭晶09、临通08、临通09、渭晶11、渭晶10、渭晶07、安博09

3 讨 论

相关分析表明, 各项指标之间存在着一定的相关性, 部分指标之间存在着显著的相关性; 因子分析表明, 基于变量之间的高相关性, 对各企业淀粉的10个质量指标只需提取3个公因子已足以反映这些指标的信息。而根据各因子综合得分排名来看, 安陇2013年的淀粉质量综合指标明显优于别的企业, 排在首位。最终聚类分析表明, 影响各企业各年度淀粉质量指标的因素各有差别, 如影响第一类企业质量指标的综合因素是理化指标, 影响第二类企业质量的因素是卫生指标, 影响第三类企业淀粉质量指标的是感官方面的因素。因此, 可根据各个企业的具体情况进行分类指导, 对改进马铃薯淀粉加工流程^[3]和提高马铃薯淀粉质量整体上升意义深远, 作用重大。

建议企业引进配套先进的加工设备^[4], 从硬件上保障产品质量提高, 使产品的质量指标从一个达标到关联的指标都达标甚至达优级。例如, 水质是马铃薯淀粉制备中对淀粉粘度的主要影响因素^[5], 如果增加去离子水生产设备, 可以使淀粉粘度明显提高, 从而使与粘度有关联性的SO₂、电导率、菌落总数及霉酵菌数也相应提高; 其次, 建议企业在选择加工原料时, 尽量选择芽眼浅的马铃薯, 这样能减少泥沙带入量, 从而使灰分指标降低^[6]; 另外, 引进先进淀粉烘干设备, 可提高淀

粉的白度和粘度等指标。马铃薯淀粉质量与各企业的加工工艺^[7,8]、加工环境关系密切。要保证在生产前各种设备清洁无污染, 避免杂物等残留产生细菌而影响马铃薯淀粉质量。马铃薯淀粉加工期短和生产人员的不稳定性, 一定程度上影响到加工技术的提高, 进而影响马铃薯淀粉质量的稳定和提高。因此, 要不断通过组织技术培训, 严格生产工艺, 使工序操作工和精制工序人员技术娴熟, 以实现淀粉各检项达优级标准。

[参 考 文 献]

- [1] 顾正彪, 周庆锋, 吕春林, 等. 马铃薯淀粉 GB/T 8884-2007 [S]. 北京: 中国标准出版社, 2007.
- [2] 邓维斌, 唐兴艳, 胡大权, 等. SPSS19 (中文版) 统计分析实用教程 [M]. 北京: 电子工业出版社, 2012.
- [3] 李贵春, 王江, 郑丽莉. 马铃薯精淀粉加工考察报告 [J]. 马铃薯杂志, 1997, 11(1): 60-61.
- [4] 张琦琦, 孙鑫淼, 石瑛, 等. 马铃薯淀粉制备条件对淀粉主要品质特性的影响 [J]. 中国马铃薯, 2009, 23(1): 15-18.
- [5] 李清萍. 影响定西市马铃薯精淀粉品质的因素分析 [J]. 农产品加工, 2009, 10: 55-56.
- [6] 孟凡玲, 梁志刚. 浅析影响马铃薯淀粉七项主要指标的因素 [J]. 科技创新与应用, 2014, 4: 65.
- [7] 祖泓舟. 马铃薯淀粉加工工艺原理及设备初探 [J]. 中外食品工业 (下半月), 2013, 10: 14-15.
- [8] 周庆锋. 中国马铃薯淀粉产业发展报告 [J]. 中国食品添加剂, 2006, 21: 62-64.