

文章编号: 1005-0906(2005)03-0021-04

主成分分析在甜玉米育种中的应用

董海合,李凤华,张旭,楼辰军,钱芳,杨兆顺

(天津市农科院农作物研究所,天津 300112)

摘要: 利用主成分分析方法对 36 个普通甜玉米(su)品种进行了分析。结果表明:前 5 个主成分值可以用来表述全部性状的信息。第一主成分为品种的生育期因子,第二主成分为糖分因子,第三主成分为赖氨酸因子,第四主成分为口感因子,第五主成分为鲜果穗产量和外观品质因子。36 个甜玉米品种可以分成 5 类,根据各类的特点可对品种进行集中选择。12 个性状可以分成 6 类,同一类性状可以通过其中的一个进行相关选择。

关键词: 主成分分析;甜玉米;育种;品种分析

中图分类号: S513.03

文献标识码: A

The Application of Principal Components Analysis Method in Sweet Corn Breeding

DONG Hai-he, LI Feng-hua, ZHANG Xu, et al.

(Tianjin Agriculture Academy Crop Research Institute, Tianjin 300112, China)

Abstract: Utilize the method of principal components analysis to analyse the 36 varieties of ordinary corn (su). The result showed: The value of the first 5 pieces of principal components can indicate the information of all properties. The first principal component is a factor of breeding time of the variety; The second principal component is a sugar factor; The third principal component is a lysine factor; The fourth principal component is a gustation factor; The fifth principal component is yield of the fresh ear of corn and quality factor of the appearance. The 36 varieties of sweet corn can be divided into 5 kinds, and can be selectd according to characteristic of the varieties. 12 properties can be divided into 6 kinds, The properties of the same type can be selected relevantly through the properties of one of them.

Key words: Principal Components analysis; Sweet corn; Breeding; Variety analysis

甜玉米营养丰富,富含糖分、必需氨基酸和维生素。甜玉米作为一种菜果兼用食品,深受消费者的欢迎。种植甜玉米对提高农业生产的经济效益、调整种植业结构具有重要意义。近年来甜玉米的需求量逐年增加,对甜玉米的品质育种提出了更高的要求,生产上急需含糖量、氨基酸、维生素等营养成分均衡、综合性状好的甜玉米品种。在甜玉米育种中,由于品质性状、农艺性状大多为数量性状,并且性状数量较多,单从个别性状进行选择,往往出现顾此失彼现象,给综合选择带来一定的困难。引入主成分分析方法,可以将多指标问题化为较少的新的指标问题,可获得许多重要信息,有利于提高多目标育种的亲本选配效果,提高育种水平。

1 材料与方法

采用 36 个普通甜玉米(su)杂交种,田间排列为随机区组设计,3 行区,3 次重复。行长 6 m,行距 0.6 m,株距 0.3 m(种植密度为 5.5 万株/hm²)。

取样在授粉后 22 天进行。每区在中行套袋 10 株自交,取其中的 5 穗,采后立即剥皮处理。然后将每区典型的 3 穗考种,并分别在穗的中部取等量的子粒,混合后迅速用烘干箱在 105℃下 10 min 杀死,之后在 60℃恒温下干燥 48 h 测干重。糖分测定用蒽酮比色法,氨基酸测定用日立 835-50 型氨基酸自动分析仪测定。

主成分分析选用鲜穗长、鲜穗粗、鲜穗重、总可溶糖、蔗糖、其它糖(其它糖=总糖-蔗糖,下同)、干物质、赖氨酸、株高、穗位高、生育日数、出苗至吐丝日数共 12 个性状,原始数据进行标准化处理,以消除量纲的影响。

收稿日期: 2004-12-21

基金项目: 天津市自然科学基金项目(023804411)

作者简介: 董海合(1962-),男,吉林省德惠县人,博士,长期从事玉米育种工作。Tel:022-27796376(O) 13820548292

2 结果与分析

2.1 样本的相关矩阵及相关分析

从表 1 可以看出,穗长(X_1)与蔗糖(X_5)显著正相关,穗粗(X_2)与鲜穗重(X_3)极显著正相关,穗粗(X_2)与总糖(X_4)、其它糖(X_6)显著负相关,穗粗(X_2)与穗位高(X_{10})、生育日数(X_{11})显著正相关,穗粗(X_2)与出苗至吐丝日数(X_{12})极显著正相关,穗重(X_3)与其它糖(X_6)显

著负相关,穗重(X_3)与出苗至吐丝日数(X_{12})极显著正相关,总糖(X_4)与蔗糖(X_5)、其它糖(X_6)呈极显著正相关,蔗糖(X_5)与其它糖(X_6)显著正相关,蔗糖(X_5)与干物质(X_7)呈极显著负相关,其它糖(X_6)与出苗至吐丝日数(X_{12})显著负相关。此外,株高(X_9)与穗位高(X_{10})、株高(X_9)与生育日数(X_{11})、穗位高(X_{10})与生育日数(X_{11})、穗位高(X_{10})与出苗至吐丝日数(X_{12})、生育日数(X_{11})与出苗至吐丝日数(X_{12})呈极显著正相关。

表 1 36 个甜玉米品种的样本相关矩阵

变量	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}
X_1	1											
X_2	-0.031 2	1										
X_3	0.264 9	0.871 1**	1									
X_4	0.252 5	-0.337 7*	-0.301 9	1								
X_5	0.331 3*	-0.184 7	-0.119 8	0.785 1**	1							
X_6	0.113 4	-0.356 9*	-0.355 3*	0.870 9**	0.379 3*	1						
X_7	-0.098 1	0.112 3	0.069 3	-0.159 8	-0.430 7**	0.103 7	1					
X_8	0.011 1	-0.249 9	-0.305 2	0.146 2	-0.027 8	0.239 6	-0.249 6	1				
X_9	0.092 5	0.280 1	0.182 1	0.299 2	0.230 0	0.263 9	-0.279 7	0.163 1	1			
X_{10}	0.095 4	0.369 6*	0.317 4	0.189 3	0.201 9	0.122 8	-0.218 9	0.098 7	0.757 0**	1		
X_{11}	0.057 6	0.360 7*	0.325 2	0.107 5	0.255 5	-0.041 6	-0.280 8	-0.152 5	0.536 5**	0.674 8**	1	
X_{12}	0.173 4	0.516 1**	0.529 3**	-0.213 6	0.060 7	-0.367 0*	-0.134 7	-0.268 5	0.317 1	0.515 5**	0.724 1**	1

注: $r_{0.05}=0.328\ 8$, $r_{0.01}=0.423\ 3$; ** 表示 0.01 水平显著, * 表示 0.05 水平显著。

2.2 样本的主成分表达式及其实际意义分析

由表 2 可以看出,前 5 个主成分特征根的累积贡献率已达到 85.51%,超过 85%,因此只要求出前 5 个主成分即可。如果样本主成分用 $y_1, y_2, \cdots, y_{12}$ 表示,则由表 3 可知,前 5 个主成分的表达式分别为:

$$0.231\ 6X_{10}+0.098\ 1X_{11}-0.039\ 4X_{12}$$
$$y_4=-0.264\ 5X_1+0.230\ 5X_2+0.082\ 4X_3+0.136\ 4X_4-0.308\ 2X_5+0.448\ 9X_6+0.643\ 3X_7+0.022\ 1X_8+0.256\ 1X_9+0.213\ 7X_{10}-0.048\ 9X_{11}-0.164\ 6X_{12}$$
$$y_5=0.605\ 1X_1+0.153\ 6X_2+0.315\ 0X_3-0.079\ 9X_4-0.162\ 9X_5+0.008\ 9X_6-0.035\ 8X_7+0.585\ 9X_8+0.080\ 8X_9+0.019\ 9X_{10}-0.322\ 2X_{11}-0.150\ 8X_{12}$$

表 2 样本相关阵的特征根、贡献率及累积贡献率			
序号	特征根	贡献率(%)	累积贡献率(%)
1	3.610 3	30.09	30.09
2	3.140 9	26.17	56.26
3	1.345 2	11.21	67.47
4	1.197 4	9.98	77.45
5	0.967 4	8.06	85.51
6	0.643 0	5.36	90.87
7	0.411 6	3.43	94.30
8	0.254 8	2.12	96.42
9	0.201 2	1.68	98.10
10	0.163 6	1.36	99.46
11	0.064 6	0.54	100.00
12	0.000 0	0.00	100.00

从第一主成分(Y_1)看,出苗至吐丝日数(X_{12})、生育日数(X_{11})的系数较大,鲜穗粗(X_2)、鲜穗重(X_3)、株高(X_9)、穗位高(X_{10})的系数也比较大,因此当品种的 12 个指标值代入第一主成分表达式,所得的第一主成分值较大时,必然是出苗至吐丝日数,生育日数较长,鲜果穗较粗,鲜果穗较重,株高和穗位高相对较高。因此,我们可以认为第一主成分为品种的生育期因子。第二主成分(Y_2)中,总可溶糖(X_4)、蔗糖(X_5)、其它糖(X_6)的系数较大,株高(X_9)、穗位高(X_{10})的系数相对也较大,说明当第二主成分值较大时,总可溶糖、蔗糖、其它糖的含量较高,株高、穗位高相对较高。因此,我们可以将第二主成分定为糖分因子。第三主成分(Y_3)中,赖氨酸(X_8)的系数 0.529 8 最大;其次为株高(X_9)和穗位高(X_{10}),鲜穗长(X_1)的系数最小为负值,鲜穗粗(X_3)、总可溶糖(X_4)、蔗糖(X_5)、干物质(X_7)的系数也为相对较小的负值,说明当第三主成分值较大时,赖氨酸含量较高,总可溶糖、蔗糖、干物质等含量相对较小。因此,第三主成分我们可以认为是赖氨酸

$$y_1=0.090\ 9X_1+0.403\ 2X_2+0.397\ 2X_3-0.091\ 4X_4+0.038\ 4X_5-0.166\ 7X_6-0.102\ 1X_7-0.132\ 1X_8+0.302X_9+0.302\ 8X_{10}+0.412\ 9X_{11}+0.443\ 9X_{12}$$
$$y_2=0.167\ 7X_1-0.198\ 2X_2-0.182\ 1X_3+0.513\ 8X_4+0.435\ 7X_5+0.421\ 5X_6-0.234\ 8X_7+0.180\ 8X_8+0.303\ 0X_9+0.240\ 9X_{10}+0.170\ 0X_{11}-0.038\ 0X_{12}$$
$$y_3=-0.475\ 0X_1-0.091\ 6X_2-0.277\ 9X_3-0.249\ 6X_4-0.294\ 0X_5-0.140\ 0X_6-0.321\ 4X_7+0.529\ 8X_8+0.267\ 6X_9+$$

因子。第四主成分(Y_4)中,干物质(X_7)的系数最大为 0.643 3;其次为其它糖(X_6)、蔗糖(X_5)的系数最小,说明当第四主成分值较大时,干物质和其它糖含量必然较高,蔗糖含量相对较低,表明在甜玉米子粒中干物质含量与蔗糖的转化有直接关系。由于干物质含量决定甜玉米子粒的内含物多少,而内含物是适口性的一个重要指标。因此,我们可以称第四主成分为干物质即内含物或口感因子。第五主成分(Y_5)中,鲜

穗长(X_1)的系数最大为 0.605 1;其次为赖氨酸(X_8)、鲜穗重(X_3)和鲜穗粗(X_2),生育日数(X_{11})的系数最小为负值,说明当第五主成分较大时,穗长和穗粗一定较大,穗较重,此外赖氨酸含量也较高。由于穗长和穗粗决定鲜果穗的产量和外观。因此,我们可以认为第五主成分是穗部性状因子,是鲜果穗产量和外观品质的度量。

表 3 样本相关阵的特征向量

项 目	q_1	q_2	q_3	q_4	q_5	q_6	q_7	q_8	q_9	q_{10}	q_{11}	q_{12}
X_1	0.090 9	0.167 7	-0.475 0	-0.264 5	0.605 1	0.420 0	-0.215 8	0.073 6	0.025 3	-0.124 0	0.239 7	-0.000 4
X_2	0.403 2	-0.198 2	-0.091 6	0.230 5	0.153 6	-0.478 7	0.174 9	-0.015 1	-0.136 1	-0.057 3	0.656 7	0.001 0
X_3	0.397 2	-0.182 1	-0.277 9	0.082 4	0.315 0	-0.335 9	0.089 7	0.172 5	0.174 6	0.027 2	-0.666 3	-0.000 3
X_4	-0.091 4	0.513 8	-0.249 6	0.136 4	-0.079 9	-0.152 9	0.167 9	0.032 6	-0.021 8	0.094 7	-0.005 7	0.760 4
X_5	0.038 4	0.435 7	-0.294 0	-0.308 2	-0.162 9	-0.286 1	0.175 1	-0.520 1	-0.117 7	-0.179 5	-0.093 5	-0.403 7
X_6	-0.166 7	0.421 5	-0.140 0	0.448 9	0.008 9	-0.002 3	0.113 4	0.461 5	0.062 1	0.283 3	0.067 5	-0.508 7
X_7	-0.102 1	-0.234 8	-0.321 4	0.643 3	-0.035 8	0.365 4	0.174 2	-0.396 7	-0.140 0	-0.251 0	-0.110 4	0.000 2
X_8	-0.132 1	0.180 8	0.529 8	0.022 1	0.585 9	0.039 1	0.527 6	-0.120 4	-0.110 6	0.109 5	-0.082 6	-0.000 2
X_9	0.302 8	0.303 0	0.267 6	0.256 1	0.080 8	-0.031 2	-0.567 4	-0.031 7	-0.556 6	-0.111 1	-0.163 3	-0.000 9
X_{10}	0.383 0	0.240 9	0.231 6	0.213 7	0.019 9	0.146 3	-0.159 4	-0.401 2	0.665 2	0.195 2	0.084 9	0.000 3
X_{11}	0.412 9	0.170 0	0.098 1	-0.048 9	-0.322 2	0.266 1	0.267 8	0.379 9	0.108 5	-0.621 4	0.003 7	0.000 5
X_{12}	0.443 9	-0.038 0	-0.039 4	-0.164 6	-0.150 8	0.384 9	0.339 5	-0.043 3	-0.369 1	0.589 6	-0.032 4	-0.000 3

2.3 应用主成分值对品种进行分类

由品种的第一主成分为横坐标,第二主成分为纵坐标作成图 1。由图 1 可以看出,36 个品种可以分为 5 类:

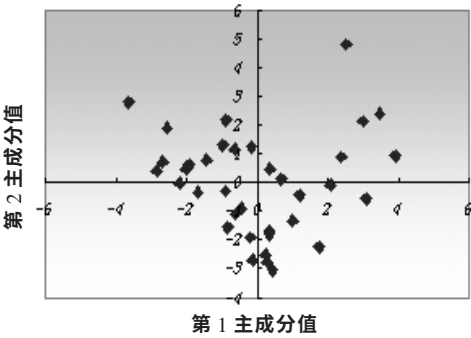


图 1 36 个甜玉米品种的分类

- I 类包括 1、2、4 号 3 个品种,为生育期较长,鲜穗较粗、重,含糖量较高,植株高大的品种类型;
- II 类包括 3、8、12、18、23、26 号 6 个品种,为糖分中等,鲜穗粗、重,植株适中的品种;
- III 类包括 10、15、16、17、21、22、24、26、27、30、32、33 号 12 个品种,生育期较短,株高适中,果穗中等,含糖量较低品种类型;
- IV 类包括 9、11、13、14、19、25、28、29、31、35、36 号 11 个品种,生育期较短,果穗较小,含糖量中等偏低,株高较矮的品种类型;
- V 类包括 7、20、34 号 3 个品种,含糖量较高,果穗中等偏小,株高较矮,生育期较短的品种类型。

由上述分类可知,首先选择好的品种类型,在同类品种中选择综合性状好的 1~2 个品种,即可达到优中选优的目的,不必按品种逐一考察。

将表 3 中第一主成分和第二主成分所对应的两个特征向量 q_1 和 q_2 的 12 对分量,分别作为指标变量 $X_1 \rightarrow X_{12}$ 的点坐标,并以 q_1 为横坐标, q_2 为纵坐标绘制二维平面散点图(图 2),则可以得到 12 个性状的分

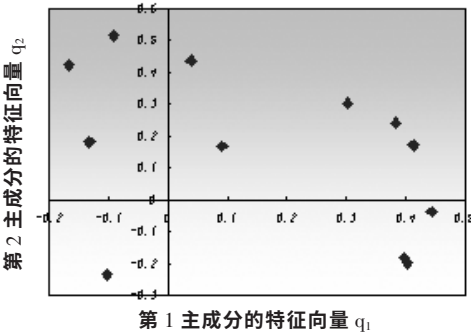


图 2 甜玉米 12 个性状的分类

- I 类: X_9 、 X_{10} 、 X_{11} 为一类,为株高和生育期性状;
- II 类: X_4 、 X_5 、 X_6 为一类,是有关糖分含量的一类性状;
- III 类: X_1 自成一类,为穗长性状;
- IV 类: X_8 自成一类,为赖氨酸性状;
- V 类: X_2 、 X_3 、 X_{12} 为一类,为穗部和营养生长(出苗至吐丝)性状;
- VI 类: X_7 自成一类,为干物质性状。

表 4 36 个甜玉米品种的前 5 个主成分值

品种代号	第 1 主成分值	第 2 主成分值	第 3 主成分值	第 4 主成分值	第 5 主成分值
1	3.541 9	2.413 5	1.426 9	-0.556 5	-0.918 9
2	3.910 9	0.916 7	-0.291 8	-0.152 9	0.682 2
3	2.370 0	0.880 9	2.533 5	-1.416 4	0.846 7
4	2.996 8	2.131 6	-1.452 5	-0.074 1	-0.976 9
5	2.515 2	4.801 3	-1.381 3	-0.497 3	-0.437 6
6	2.053 9	-0.094 4	0.865 0	2.708 3	1.843 1
7	-0.904 9	2.161 0	1.079 1	1.542 4	0.455 1
8	3.091 4	-0.569 9	-1.118 0	0.862 1	0.169 8
9	-0.177 0	1.249 7	-2.420 7	-0.500 7	-0.057 0
10	0.323 2	-1.703 6	1.169 0	-1.994 1	-0.890 6
11	-1.683 6	-0.343 9	-0.144 9	1.780 0	-1.956 6
12	0.353 2	0.458 7	1.027 9	-0.587 3	-2.320 8
13	-1.436 4	0.793 1	0.721 9	0.335 0	-0.978 1
14	-2.688 8	0.724 6	-0.135 7	-0.483 7	0.862 0
15	0.236 6	-2.530 7	1.405 0	0.578 4	0.253 8
16	1.747 1	-2.249 7	0.984 2	0.657 0	0.663 7
17	-0.642 6	-1.101 3	-1.375 9	-1.070 8	1.015 9
18	0.657 6	0.105 9	1.011 8	-1.670 5	1.136 0
19	-0.987 1	1.290 8	1.768 8	0.972 0	0.515 2
20	-2.556 9	1.886 0	0.056 5	-2.157 3	1.102 7
21	0.424 3	-3.044 8	0.277 9	0.062 5	-0.608 7
22	-0.846 9	-1.552 1	-0.283 1	-0.128 9	-0.607 3
23	1.184 2	-0.438 3	-0.998 3	0.098 1	-0.503 8
24	-0.214 5	-1.920 4	-0.397 3	-0.016 9	-1.384 4
25	-2.842 5	0.410 5	0.880 0	-1.046 4	0.190 4
26	0.269 6	-2.778 7	1.023 3	-0.220 3	0.313 6
27	0.985 1	-1.327 9	-0.753 5	-0.535 3	1.221 7
28	-2.026 5	0.468 6	-0.294 0	1.557 8	1.323 9
29	-1.930 1	0.622 4	-1.004 7	-0.711 4	0.040 0
30	-0.155 1	-2.699 1	-1.745 2	-0.230 5	0.007 0
31	-0.651 1	1.169 5	-0.681 8	2.273 4	-0.252 6
32	-0.467 2	-0.917 1	-1.613 7	0.161 8	0.964 2
33	0.326 2	-1.689 7	0.177 6	0.379 3	-1.499 5
34	-3.653 9	2.788 6	1.175 9	0.685 3	-0.423 8
35	-0.908 5	-0.293 9	-0.955 8	-0.165 3	0.910 4
36	-2.213 5	-0.017 7	-0.536 5	-0.437 1	-0.700 9

3 结 论

(1)前 5 个主成分值的贡献率达到 85%以上,其所表达的综合信息可以用来表述全部性状的信息。第一主成分为品种的生育期因子,第二主成分为糖分因子,第三主成分为赖氨酸因子,第四主成分为口感因子,第五主成分为鲜果穗产量和外观品质因子。在甜玉米育种中,可以尝试集中考察上述 5 个综合性状因子,以提高亲本的选配效果,提高育种水平。

(2)36 个甜玉米品种可以分成 5 类,根据各类的特点可以对品种进行相关选择,减少品种决选的工作量。

(3)12 个性状可以分成 6 类,同一类的性状可以通过其中的某一个进行选择,完全可以达到对其它

性状间接选择的效果。我们在甜玉米育种中,可以在同类性状中选择相对容易选择的性状进行相关选择,可以减少多性状选择的盲目性和工作量。

参考文献:

[1] 裴鑫德. 多元统计分析及其应用[M]. 北京农业大学出版社, 1991. 196-248 .

[2] 刘多森. 主成分分析在分辨土壤类型及风化成土过程上的应用[J]. 土壤学报, 1979, (2): 172-183 .

[3] 肖 红, 等. 利用因子分析方法研究冬小麦亲本性状及其品种分类[J]. 北京农业大学学报, 1991, (4): 69-73 .

[4] 孙卫永, 等. 玉米自交系性状的遗传相关分析和主成分分析[J]. 扬州大学学报(自然科学版), 2001, 4(1) .

[5] 吴高岭, 等. 数值分类法在玉米品种资源上的应用[J]. 玉米科学, 2000, 8(1): 23-27 .