

$$\text{优势指数} = \frac{F_1}{P_1 + P_2/2}$$

237×抗7、15860×M₃17三组合与优势指数较小的东46×RW64A^{III}、36×94比较,前者杂种优势虽高,但就其实际拱土能力看,出苗指数为14.1、13.5和13.8,仍显著的小于后者16.3和15.3。分析其原因时,我们看到优势指数较高的三组合其亲本自交系平均出苗指数均较差,为7.1、9.1和9.6,而优势指

数较低的东46×RW64A^{III}、36×94其亲本自交系平均出苗指数却为13.6和13.2。为此,选育拱土能力强的玉米杂交种应以选择亲本自交系为基础,才易获得成功。在试验结果中看到,76~237这样的自交系在拱土能力方面,可能具有特殊的配合力,以及杂交优势等问题,需要进一步研究和探讨。

因子分析在春小麦品种与性状分类中的实际应用*

韩 龙 珠

(农牧渔业部哈尔滨土地管理干部学校)

通常在杂交育种实践中,常常希望获得具有综合优良性状的品种,但由于目标经济性状多为数量性状,且又存在复杂的性状间的表型、基因型的相关通路,难以研究分析。在后代选择中考虑多个性状时,往往顾此失彼。不难看出,这些问题用比较少量的有代表性的因子来说明由多个性状所提供的信息,并由此期望得到科学的单纯性或描述的经济性,显然将有助于问题的解决。为此,本文的目的在于试图利用因子分析,设法找出若干个数目较少独立的新的变量(它是一些原变量的线性组合),这对了解现象的内在原因有所帮助,在考虑多个性状时,显示出很大的优越性。

关于作物数量性状的多元统计技术应用于作物育种的杂种优势利用的研究,始于七十年代。Bhatt G.M. (1973)曾以此进行自花授粉作物亲本选配的研究;国内的刘来福、毛盛贤等(1979)首先将主成分分析(Principal Components Analysis)应用于冬小麦杂交育种的亲本选配研究上,继后,杨德、

代君锡(1982)提出了典范分析(Canonical Analysis)分别测定弱冬性小麦和水稻数量性状的后代选择与杂交亲本选配研究上。至于用因子分析(Factor Analysis)对样品或因素作出分类,国内尚未见报道,所以我们在遗传距离的聚类分析和模糊聚类分析的基础上,提出了Q型和R型因子分析,用因子荷载向量进行分类得到较好的分类效果。

现将分析结果简述如下:

一、 试验材料及方法

1. 供试材料

供试代表品种,分抗旱、耐湿、喜肥水及早熟4个生态类别,共59个材料。主要是黑龙江省近几十年不同时期引种、系统选择和杂交育成的推广品种⁽⁵⁾。试验采用随机区组,重复4次,每区调查10株,属本文研究的调查项目有:(1)小区产量(2)株高(3)穗

* 本文原始资料由黑龙江省农科院育种所褚延刚同志提供,黑龙江省电子计算机室段文同志帮助电算核对,在此一并致以谢意。

长(4)单株粒重(5)单株穗数(6)单株粒数(7)主穗粒数(8)主穗粒重(9)结实小穗数(10)千穗重等。品名与代号请参见黑龙江农业科学1983(2)。

2. 统计方法

1) 标准化数据矩阵 Z

对于原始资料进行标准化,用公式

$$Z_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma_{sj}}$$

计算得, Z_{ij} 为第 i 个品种第 j 个性状的值,其中 σ_{sj} 为第 j 个性状的表型标准差。

2) 求分析对象的相关阵或相似阵。

相关阵:一般生统遗传学文献介绍的方法均可采用,也可参照刘来福等^[4]的方法求出基因型相关系数。

相似阵:常用的相似系数夹角余弦计算,其公式为:

$$\cos \theta_{ij} = \frac{\sum_{k=1}^n x_{ik} x_{jk}}{\sqrt{\left(\sum_{k=1}^n x_{ik}^2\right) \left(\sum_{k=1}^n x_{jk}^2\right)}}$$

3) 求相关阵或相似阵的特征值及其特征向量。本文采用 jacob 方法计算。

4) 确定公共因子数目之后,利用公式 $a_i = \alpha_i / \sqrt{\lambda_i}$ 求出在第 i 个公共因子上的荷载向量。

5) 仅取头 2 个公共因子上的荷载向量,作成两个公共因子为轴的平面图,利用图上点子的聚集状况进行分类。

6) 计算 59 个变量的公共性为:

$$h_i^2 = \sum_{j=1}^m \alpha_{ij}^2$$

类内最大者可作为代表品种。

二、结果分析与讨论

(一) Q 型因子分析

1. 对黑龙江省近几十年不同时期推广的 59 个品种,10 个数量性状的原始数据进行规格化。

2. 首先计算 59 个品种之间的相似系数。由此可得到一个 59×59 的相似阵。

$$Q_{59 \times 59} = \begin{pmatrix} 1 & 0.42 & 0.88 & 0.60 \cdots & -0.40 \\ 0.42 & 1 & 0.76 & 0.68 \cdots & 0.34 \\ 0.88 & 0.76 & 1 & 0.76 \cdots & -0.10 \\ 0.60 & 0.68 & 0.76 & 1 & \cdots & -0.32 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -0.40 & 0.34 & -0.10 & -0.32 \cdots & 1 \end{pmatrix}$$

求 Q 阵的特征值为: $\lambda_1 = 23.300$, $\lambda_2 = 15.856$, $\lambda_3 = 7.801$, $\lambda_4 = 5.402 \cdots$, $\lambda_{59} = -2.148$ 。并前三个特征向量 V_1, V_2, V_3 列在表 2 中。

由于头 3 个主要因子的相对方差贡献率

$$P = \frac{\sum_{i=1}^3 \lambda_i}{\sum_{i=1}^{59} \lambda_i} = \frac{46.957}{59} = 80\%$$

故取头 3 个公共因子,计算 59 个品种在头 3 个公共因子上的荷载向量为:

$$\alpha_i = \sqrt{\lambda_1} \cdot V_1 = \sqrt{23.300} \begin{pmatrix} -0.095 \\ -0.156 \\ \vdots \\ -0.117 \end{pmatrix} = \begin{pmatrix} -0.460 \\ -0.753 \\ \vdots \\ -0.566 \end{pmatrix}$$

详细值列在表 1 中。

表 1

品种代号	V_1	V_2	V_3	a_1	a_2	a_3	h	品种代号	V_1	V_2	V_3	a_1	a_2	a_3	h
1	-0.095	-0.185	0.062	-0.460	-0.735	0.174	0.884	6	0.071	-0.193	-0.091	0.341	-0.767	-0.255	0.877
2	-0.156	-0.089	-0.191	-0.753	-0.354	-0.533	0.989	7	-0.201	0.015	-0.032	-0.973	0.060	-0.228	1.001
3	-0.143	-0.180	-0.021	-0.693	-0.716	-0.060	0.998	8	-0.175	-0.095	-0.132	-0.843	-0.021	-0.370	0.921
4	-0.053	-0.200	-0.117	-0.254	-0.798	-0.326	0.899	9	-0.053	0.096	-0.145	-0.255	0.384	-0.405	0.614
5	0.024	-0.022	-0.242	0.115	-0.327	-0.677	0.761	10	-0.112	-0.112	-0.166	-0.543	-0.471	-0.462	0.855

品种代号	V_1	V_2	V_3	a_1	a_2	a_3	h	品种代号	V_1	V_2	V_3	a_1	a_2	a_3	h
11	-0.182	-0.041	0.030	-0.878	-0.162	-0.085	0.897	36	-0.132	-0.093	0.159	-0.635	-0.371	0.445	0.860
12	-0.188	-0.106	-0.015	-0.906	-0.422	-0.043	1.000	37	-0.178	-0.048	-0.088	-0.861	-0.189	-0.246	0.915
13	-0.112	-0.142	0.190	-0.541	-0.564	0.530	0.944	38	0.025	-0.149	0.202	0.119	-0.592	0.565	0.827
14	-0.149	0.028	0.178	-0.717	0.110	0.498	0.880	39	0.186	-0.055	0.122	0.600	-0.218	0.340	0.937
15	-0.184	0.061	0.030	-0.888	0.243	0.084	0.925	40	0.147	-0.106	-0.142	0.707	-0.423	-0.397	0.915
16	-0.164	0.150	-0.033	-0.794	0.598	-0.092	0.998	41	0.181	-0.044	0.047	0.873	-0.177	0.131	0.900
17	-0.188	-0.049	0.100	-0.905	-0.196	0.279	0.967	42	0.164	-0.089	0.174	0.791	-0.354	0.486	0.994
18	-0.059	0.147	-0.174	-0.287	0.586	-0.486	0.814	43	0.155	0.017	0.160	0.748	0.068	0.447	0.874
19	-0.049	0.109	-0.089	-0.235	0.432	-0.254	0.554	44	0.119	0.057	0.010	0.574	0.229	0.029	0.619
20	0.139	0.113	-0.127	0.671	0.452	-0.356	0.894	45	0.153	-0.113	0.078	0.737	-0.452	0.217	0.891
21	-0.081	0.097	0.009	-0.391	0.388	0.024	0.550	46	0.143	0.160	0.081	0.690	0.638	0.226	0.967
22	0.012	0.204	0.034	0.057	0.813	0.095	0.821	47	0.181	-0.026	-0.130	0.878	-0.103	-0.362	0.953
23	-0.039	-0.183	-0.096	-0.189	-0.727	-0.288	0.798	48	0.037	0.188	-0.150	0.178	0.749	-0.420	0.877
24	0.133	0.170	-0.006	0.640	0.678	-0.018	0.933	49	0.178	0.003	-0.171	0.860	0.013	-0.478	0.984
25	0.155	-0.087	-0.113	0.740	-0.267	-0.314	0.855	50	0.186	-0.090	0.027	0.896	-0.357	0.077	0.968
26	-0.050	-0.001	-0.211	-0.242	-0.005	-0.588	0.636	51	0.186	-0.099	0.065	0.897	-0.394	0.182	0.997
27	-0.074	0.227	-0.050	-0.359	0.905	-0.140	0.984	52	0.153	-0.047	0.230	0.740	-0.188	0.642	0.998
28	-0.132	0.184	0.039	-0.637	0.724	0.108	0.978	53	0.057	0.206	0.019	0.276	0.822	0.053	0.889
29	0.172	0.058	-0.127	0.830	0.233	-0.353	0.932	54	0.074	0.176	-0.120	0.357	0.701	-0.336	0.855
30	0.028	0.198	-0.046	0.135	0.787	-0.127	0.809	55	0.147	0.055	-0.012	0.709	0.218	-0.035	0.743
31	0.073	0.221	0.032	0.354	0.880	0.090	0.953	56	-0.119	0.137	0.207	-0.573	0.547	0.580	0.932
32	0.117	-0.122	-0.223	0.565	-0.485	-0.622	0.970	57	-0.131	0.166	0.013	-0.634	0.663	0.037	0.918
33	-0.030	0.009	0.314	-0.145	0.036	0.877	0.890	58	-0.038	0.224	0.113	-0.183	0.891	0.317	0.963
34	-0.087	-0.171	0.125	-0.422	-0.682	0.348	0.874	59	-0.117	0.155	0.002	-0.586	0.617	0.005	0.837
35	-0.087	-0.133	-0.197	-0.420	-0.530	-0.551	0.872								

3. 以头两个公共因子为轴, 把各品种在这两因子上的荷载分别点在图上, 如图 1。

根据点子聚集情况, 可分为 7 类:
 第一类是 20、24、25、29、32、39、40、41、42、43、44、45、46、47、49、50、51、52、55 等 19 个品种。从公共性上面 (见表 2), 这一类中以 42、51、52, 即克全、克早 7 号、克 69-701 为最大, 可作为代表性品种。
 第二类是 22、30、31、48、53、54 等

6 个品种。从公共性方面 31, 龙麦 6 号为代表性品种。

第三类是 5、6、38。6 号北安大青芒为公共性上最大, 作为代表性品种。

第四类是 26、33。33 号龙麦 9 号为代表性品种。

第五类是 9、18、19、21、27、58 等 6 个品种。从公共性方面 27 号新曙光一号为代表性的品种。

第六类是 16、28、56、57、59 等 5 个

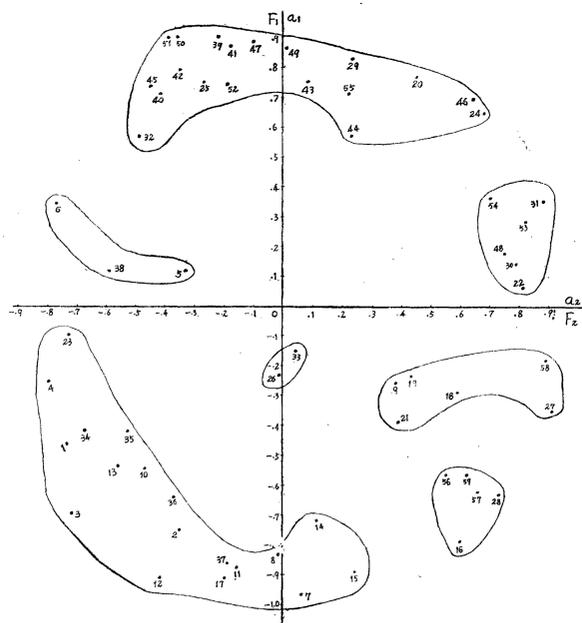


图1 Q型因子分析

品种。从公共性上,以16号松花江7号为代表性的品种。

合作7号为代表性的品种。

第七类是1、2、3、4、7、8、10、11、12、13、14、15、17、23、24、35、36、37等18个品种。从公共性上7、12,甘肃96、

本文结合黑龙江省春小麦品种推广实践与育成过程,对上述分类结果作如下讨论:

(1) 从7个类中的大群体看,60年代初以前推广的品种归一类,60年代后期到70

年代推广品种归一类,并从第一主成分考虑,后一类品种大多数 a_1 值在0.5以上。这充分说明,我省春小麦育种水平从60年代后期开始逐渐提高,对品种的抗病、灾力上,在生育后期的耐湿性、产量稳定性上有明显改进。

(2) 在各类中的代表性品种看,都是各不同时期表现突出,有一定面积的、生产上成为主体品种,其中合作7号、松花江7号、克全、克早7号、克69~701等品种一度生产上起了很大作用,有些仍为当前一些地区的主体品种。

(3) 从品种的系谱考虑,第四类的两个品种恰好是天兰冰草系统;第五、六类是Minn2761及其衍生品种系统,它们尽管细分为两类,但图上这两类距离很近,在第六类品种中外品种成分较多,第一类主要是“克字号”品种,其姐妹系同为一群,这一类是

在生产进一步发展的情况下涌现出一批以抗旱型为主的新类群,多数表现出丰产性状突出、抗逆、抗病性强、适于机械化栽培的新品种。

综上所述,通过因子荷载向量进行分类是符合本地区生产实际情况,比主成分分析进了一步。

(二) R型因子分析

R型因子分析是对性状进行分类,计算技术上与Q型一样,只不过是求样品间相似性矩阵改为变量间的相关矩阵。

1. 在作物育种实践中,对性状间遗传变异的相关程度作出估价,然而估算遗传协方差和遗传相关系数是研究相关遗传效应的基础,为此,我们用方差-协方差分析模式^[2]估算出遗传相关系数。由10个数量性状得到 $R_{10 \times 10}$ 的遗传相关阵:

$$R_{10 \times 10} = \begin{pmatrix} | & 0.0782 & 0.4581 & 0.9390 & -0.2336 & 0.6610 & 0.7275 & 0.8881 & 0.5399 & 0.4938 & \text{产量} \\ 0.0782 & | & 0.7223 & 0.6935 & 0.4516 & 0.3524 & 0.2701 & 0.1319 & 0.6175 & -0.3706 & \text{株高} \\ 0.4581 & 0.7223 & | & 0.7509 & -0.0313 & 0.4671 & 0.6215 & 0.6122 & 0.7607 & -0.0003 & \text{穗长} \\ 0.9390 & 0.6935 & 0.7509 & | & -0.2261 & 0.6658 & 0.7302 & 0.8940 & 0.5429 & 0.4879 & \text{单株粒重} \\ -0.2336 & 0.4516 & -0.0313 & -0.2261 & | & 0.2563 & -0.3318 & -0.4985 & -0.0214 & -0.6095 & \text{单株穗数} \\ 0.6610 & 0.3524 & 0.4671 & 0.6658 & 0.2563 & | & 0.7763 & 0.5381 & 0.6539 & -0.3196 & \text{单穗粒数} \\ 0.7275 & 0.2701 & 0.6215 & 0.7302 & -0.3318 & 0.7763 & | & 0.8268 & 0.7954 & 0.0111 & \text{主穗粒数} \\ 0.8881 & 0.1319 & 0.6122 & 0.8940 & -0.4985 & 0.5381 & 0.8268 & | & 0.6122 & 0.4808 & \text{主穗粒重} \\ 0.5399 & 0.6175 & 0.7607 & 0.5429 & -0.0214 & 0.6539 & 0.7954 & 0.6122 & | & -0.1690 & \text{穗、尖小粒数} \\ 0.4938 & -0.3706 & -0.0003 & 0.4879 & -0.6095 & -0.3196 & 0.0111 & 0.4808 & -0.1690 & | & \text{千粒重} \end{pmatrix}$$

2. 求R阵的特征值为: $\lambda_1 = 5.3809$, $\lambda_2 = 2.5848$, $\lambda_3 = 0.9393$, $\lambda_4 = 0.7530$, $\lambda_5 = 0.2658$, $\lambda_6 = 0.1337$, $\lambda_7 = 0.0430$, $\lambda_8 = 0.0169$,

$\lambda_9 = 0.0029$, $\lambda_{10} = -0.1208$,由于前三个主要因子的相对方差贡献为87%,故取前三个公共因子上的荷载向量(见表2)。

表 2

性状代号	V_1	V_2	V_3	a_1	a_2	a_3	h	性状代号	V_1	V_2	V_3	a_1	a_2	a_3	h
1	0.377	-0.184	-0.243	0.874	-0.296	-0.236	0.952	6	0.328	0.244	-0.534	0.760	0.392	-0.518	0.999
2	0.165	0.465	0.487	0.383	0.748	0.472	0.964	7	0.391	0.018	-0.146	0.907	0.029	-0.141	0.908
3	0.342	0.186	0.484	0.793	0.299	0.479	0.969	8	0.393	-0.213	0.859	0.312	-0.342	0.038	0.975
4	0.396	-0.159	-0.086	0.918	-0.254	-0.089	0.956	9	0.353	0.239	0.135	0.818	0.384	0.131	0.913
5	-0.096	0.486	-0.266	-0.222	0.781	-0.258	0.852	10	0.093	-0.541	0.253	0.215	-0.869	0.245	0.828

3. 也以前两个公共因子为轴,把10个数量性状的荷载向量分别点在如图2,然后根据点聚集情况分为四类:

第一类是1, 3, 4, 6, 7, 8, 9等7个性状,从公共性上6号单株粒数为最大,可作为这一类的代表性性状。

第二类是2, 株高性状。

第三类是10, 千粒重性状。

第四类是5, 单株穗数性状。

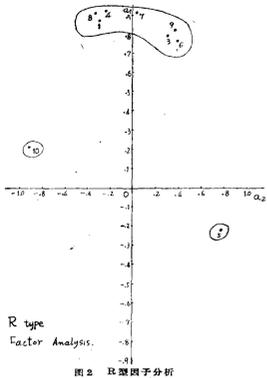


图2 R型因子分析

对于求得各类性状中的代表性性状可组成: 粒数因子、粒重因子、穗数因子和株高因子。这与我们在遗传相关系数的聚类分析相符。

4. 再从上述因子荷载方差极大旋转,可得到如表3的性状与因子关系。

经方差极大旋转后,因子荷载方差由0.2329增加到0.3586。从旋转后的因子荷载矩阵来看,主因子 F_1 上小区产量和主穗粒数荷载分别为0.97、530.9477明显表示了粒数

表3

变 量	F_1	F_2	F_3
1	0.9537		
2		0.9530	
3		0.8806	
4	0.9121		
5		0.9757	
6	0.8695		
7	0.9477		
8	0.8404		
9			0.7186
10		-0.9690	

在影响产量的直接关系,故定为粒数因子。在主因子 F_2 中,正向考虑穗数荷载向量最大,故定为穗数因子。主因子 F_3 中,主要起作用的变量是株高,所以定为株高因子。

总的看来,影响产量的主要特征性状是:粒数、穗数和株高因子,如果在主因子 F_1 上负向考虑,性状组成是粒数、粒重和株高因子。

参 考 文 献

- (1) 张尧庭 方开泰 1982 多元统计分析引论 科学出版社。
- (2) 黄嘉佑 1982 气象统计预报实用教材 北京大学地球物理系。
- (3) 刘来福 1979 遗传 1(5):44~48。
- (4) 王学仁 1982 地质数据的多变量统计分析 科学出版社。
- (5) 陈洪文等 1983 黑龙江农业科学(2):8~12。