

对应分析在小麦品种与性状分类的研究初探

韩龙珠

(黑龙江省土地管理干部学校)

提 要

本研究以黑龙江省近百年不同时期推广的 59 个代表品种为材料, 所调查的 10 个数量性状应用对应分析, 按品种与性状的聚合自然趋势, 将全部性状和品种划分为 IV 区, 并作图和成因解释, 找出了相应性状突出的品种。从对应分析结果表明, 影响产量的特征系数为单株粒数和千粒重, 同时受两个主因子轴的控制。

有关作物品种与性状分类的研究, 将能增进对作物品种资源的正确认识。

1973 年, Bhatt. G.M 首先曾以白花授粉作物为例, 采用多元分析方法进行亲本分类, 刘来福、杨德等[3, 4, 5] 提出了主成份分析、典范分析、因子分析和模糊聚类等方法。这些工作, 对作物品种资源数量分类和育种工作中性状选择具有指导作用。

这项工作, 国内开展稍晚, 近年来日益增多, 但应用对应分析对品种与性状进行分类及分析还未见报道。

本研究以黑龙江省近百年不同时期推广的 59 个代表品种为材料, 对 10 个数量性状作了初步分析。

材 料 与 方 法

(一) 供试材料

供试代表品种为黑龙江省近百年不同时期

推广的 4 个生态型(抗旱、耐湿、喜肥水及早熟) 59 个品种(表 1)。试验采用随机区组法设计, 重复 4 次, 株行距为 $(50+20) \times 5$ 厘米, 行长 1 米, 单粒点播, 前茬为大豆间地。收获前, 每区调查 10 株, 调查项目有: 小区产量(1)、株高(2)、穗长(3)、单株粒重(4)、单株穗数(5)、单株粒数(6)、主穗粒数(7)、主穗粒重(8)、结实小穗数(9)、千粒重(10)等 10 项指标。

(二) 计算方法

1. 59 个品种资源所调查的原始资料, 可排成矩阵 $X = [x_{ij}]$, 其中 x_{ij} 为第 i 个品种的第 j 个性状的观测值。

首先将 X 按行、列分别求和。

2. 计算矩阵

$$Z = [Z_{ij}] \quad \text{其中 } Z_{ij} = \frac{x_{ij} - \frac{x_{i.} \cdot x_{.j}}{T}}{\sqrt{x_{i.} \cdot x_{.j}}}$$

$$A = [a_{ij}]_{(10 \times 10)} = Z'Z$$

3. 因子分析

I. R—型

i) 计算矩阵 $A = Z'Z$ 的特征值: $\lambda_1 \geq \lambda_2 \geq$

$\dots \geq \lambda_p$ 。按其累计百分比 $\frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^p \lambda_i} \geq 95\%$, 取其前 k 个特征值 $\lambda_1, \lambda_2, \dots, \lambda_k$ 。并计算它们相应的特征向量 u_1, u_2, \dots, u_k 。

注: 本文承蒙安徽农学院刘垂瑜教授审阅, 黑龙江省农科院育种所小麦研究室提供原始资料, 特此一并致谢。

表 1

不同生态类型品种及代号

代号	品种名	代号	品种名	代号	品种名	代号	品种名	代号	品种名
1	勇捷	13	松花江1号	25	龙麦1号	37	克繁	49	克早5号
2	滨南	14	松花江2号	26	龙麦2号	38	克茂	50	克早6号
3	兰寿	15	松花江3号	27	新曙光1号	39	克早1号	51	克早7号
4	克华	16	松花江7号	28	新曙光3号	40	克珍	52	克69-701
5	南风	17	东农101	29	龙麦4号	41	克群	53	克早8号
6	北大青芒	18	东农104	30	龙麦5号	42	克全	54	黑春1号
7	甘肃96	19	东农106	31	龙麦6号	43	克丰1号	55	克丰2号
8	麦粒多	20	合春7号	32	龙麦7号	44	克游2号	56	辽春4号
9	合作2号	21	合春11号	33	龙麦9号	45	克游2号	57	垦149
10	合作4号	22	合春12号	34	克强	46	克游3号	58	沈68-71
11	合作6号	23	北新2号	35	克壮	47	克早2号	59	他诺瑞
12	合作7号	24	北新4号	36	克健	48	钢107		

由此, 得因子载荷矩阵:

$$F = \begin{pmatrix} u_{11} \sqrt{\lambda_1} & u_{12} \sqrt{\lambda_2} & \cdots & u_{1k} \sqrt{\lambda_k} \\ \cdots & \cdots & \cdots & \cdots \\ u_{p1} \sqrt{\lambda_1} & u_{p2} \sqrt{\lambda_2} & \cdots & u_{pk} \sqrt{\lambda_k} \end{pmatrix}$$

ii) 在两两因子轴平面上作性状点图。

II、Q—型

i) 对 I 中所得 k 个特征值 $\lambda_1 \geq \lambda_2 \cdots \geq \lambda_k$

计算其对应于矩阵 $Z'Z'$ 的单位特征向量 $V_1 = Zu_1, V_2 = Zu_2, \cdots, V_k = Zu_k$ 。由此, 得 Q—型因子载荷矩阵:

$$G = \begin{pmatrix} V_{11} \sqrt{\lambda_1} & V_{12} \sqrt{\lambda_2} & \cdots & V_{1k} \sqrt{\lambda_k} \\ \cdots & \cdots & \cdots & \cdots \\ V_{N1} \sqrt{\lambda_1} & V_{N2} \sqrt{\lambda_2} & \cdots & V_{Nk} \sqrt{\lambda_k} \end{pmatrix}$$

ii) 在与 R—型相应的因子平面上作品种点图。

4. 计算品种与性状的绝对贡献和相对贡献

$$\text{绝对贡献: } C_{\alpha_k}(i) = \frac{P_i \cdot g_{ik}^2}{\lambda_k}$$

$$C_{\alpha_k}(j) = \frac{P_{.j} \cdot f_{jk}^2}{\lambda_k}$$

$$\text{相对贡献: } C_{T_k}(i) = \frac{g_{ik}^2}{d^2(i, G)}$$

$$C_{T_k}(j) = \frac{f_{jk}^2}{d^2(j, G)}$$

$$\text{其中 } d^2(i, G) = \sum_{i=1}^n g_{ik}^2$$

结果与分析

(一) 方差分析

根据 59 个品种主要性状的考种结果进行方差估计, 上述 10 个性状的品种间方差大于机误方差, F 值极显著, 误差率 ($m\%$) 均小于 1, 说明这 10 个性状品种间差异主要是品种本身决定的, 故可作进一步分析。

(二) 计算矩阵和特征根

上述 59 个品种 10 项指标的数据变换, 求出其矩阵 $A = Z'Z$, 结果给出一个 10 阶的方差—协方差矩阵。用 Jacobi 法进行相似变换, 将上述 A 阵求出特征根和特征向量, 使累计率达到 95% 入选的特征根及相应的特征向量见表 2。

表 2

特 征 根 λ_i	0.0050	0.0040	分 量
累 计 %	56.47	92.61	来 源
特 征 向 量 u_k	-0.3132 0.7729 0.1160 -0.1059 0.1131 -0.4554 -0.0981 -0.0226 0.1456 0.1677	-0.5286 0.1577 -0.0802 -0.1542 0.1653 0.4774 -0.0675 -0.1148 -0.0003 -0.6263	小 区 产 量 株 高 穗 长 单 株 粒 重 单 株 穗 数 单 株 粒 数 主 穗 粒 数 主 穗 粒 重 结 实 小 穗 数 千 粒 重

(三) 因子分析

R—型因子载荷矩阵,如表3取前二个特征值 $\lambda_1=0.005$ 、 $\lambda_2=0.004$ 及相应的特征向量 u_1 、 u_2 计算因子载荷阵;Q—型因子载荷

矩阵,取前两个特征值 λ_1 、 λ_2 和单位特征向量 $V_1=Zu_1$ 、 $V_2=Zu_2$,得Q—型两个主因子的载荷表(表4)。

表 3

主 因 子 分量序号	F ₁	F ₂
1	-0.0221	-0.0334
2	0.0547	0.0100
3	0.0082	-0.0051
4	-0.0075	-0.0098
5	0.0080	0.0105
6	-0.0322	0.0302
7	-0.0069	-0.0043
8	-0.0016	-0.0073
9	0.0103	-0.00002
10	0.0119	-0.0396
方差贡献	0.0050	0.0040
累计方差贡献%	56.47	92.61

由表3、表4可作出各品种及性状在因子平面 F_1-F_2 、 G_1-G_2 上的投影图(图1)。

(四) 品种与性状的绝对贡献和相对贡献

在某些特殊的情况下,两个很接近的投影可能代表两个相隔很远的点,为了防止出

现这种情况我们可以计算每一个体对于某一因子的贡献,绝对贡献和相对贡献,如表5。绝对贡献 Ca_k 就表示性状对于因子 k 的出现所作出的贡献,相对贡献 Cr_k 表示因子 k 对于性状的离差的贡献。

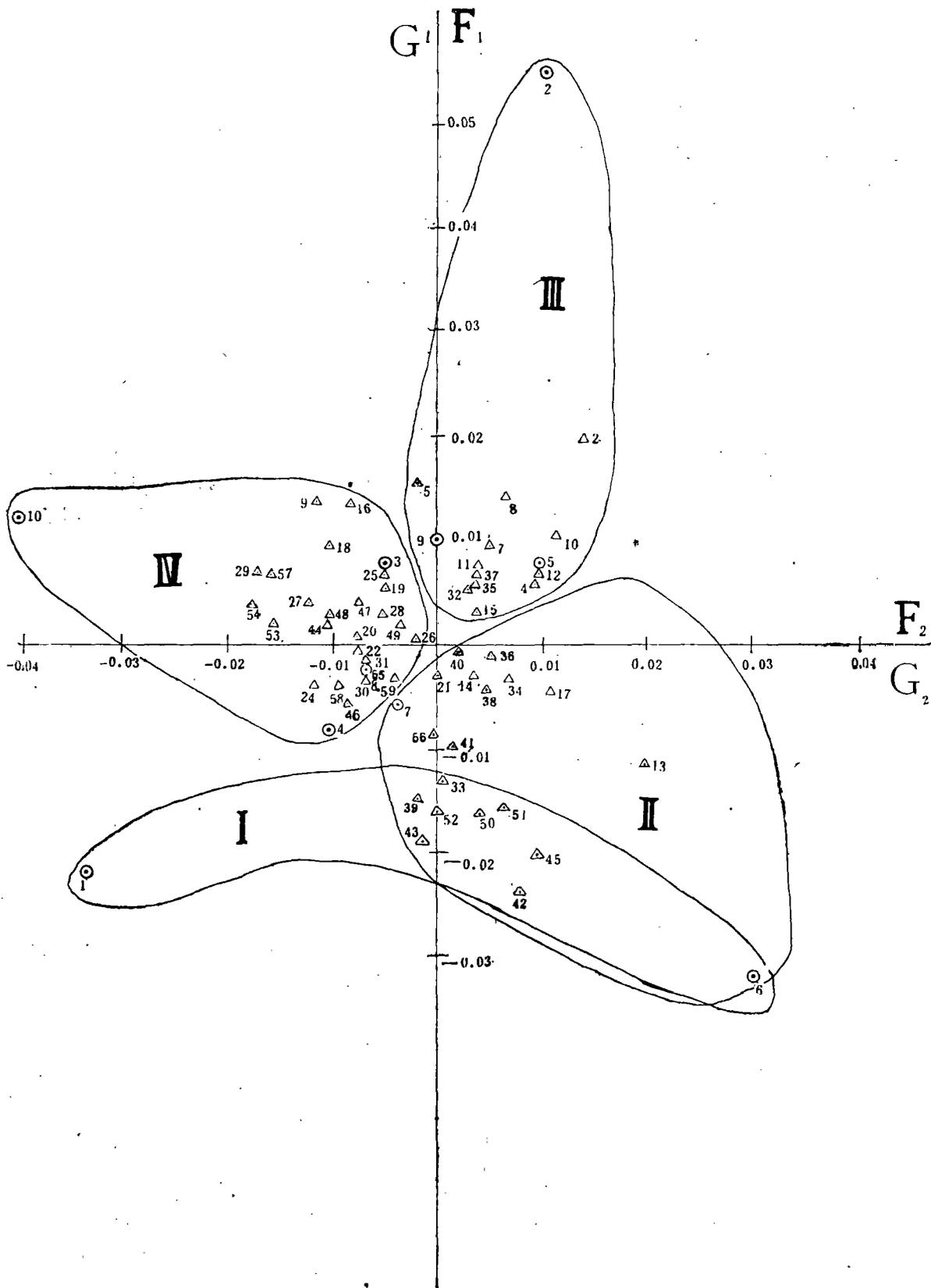


图 1. 59 个品种对应分析主因子载荷平面聚点图

表 4

59 个品种的因子载荷

品种代号	G_1	G_2	品种代号	G_1	G_2	品种代号	G_1	G_2
1	0.0023	0.0164	21	-0.0033	0.0007	41	-0.0097	0.0015
2	0.0206	0.0123	22	-0.0006	-0.0088	42	-0.0230	0.0052
3	0.0057	0.0180	23	0.0009	0.0130	43	-0.0186	-0.0012
4	0.0069	0.0093	24	-0.0034	-0.0120	44	0.0025	-0.0085
5	0.0175	-0.0007	25	0.0065	-0.0048	45	-0.0189	0.0072
6	0.0013	0.0130	26	0.0003	-0.0018	46	-0.0055	-0.0081
7	0.0089	0.0019	27	0.0055	-0.0104	47	0.0047	-0.0073
8	0.0162	0.0059	28	0.0028	-0.0047	48	0.0041	-0.0090
9	0.0129	-0.0086	29	0.0079	-0.0137	49	0.0019	-0.0037
10	0.0109	0.0107	30	-0.0023	-0.0075	50	-0.0162	0.0034
11	0.0069	0.0032	31	-0.0018	-0.0069	51	-0.0157	0.0053
12	0.0073	0.0098	32	0.005	0.0024	52	-0.0148	0.0001
13	-0.0104	0.0181	33	-0.0126	0.0006	53	0.0029	-0.0134
14	-0.0038	0.0026	34	-0.0026	0.0061	54	0.0045	-0.0126
15	0.0032	0.0027	35	0.0059	0.0033	55	-0.0026	-0.0065
16	0.0132	-0.0069	36	-0.0013	0.0042	56	-0.0085	-0.0006
17	-0.0028	0.0106	37	0.0062	0.0033	57	0.0078	-0.0132
18	0.0106	-0.0076	38	-0.0042	0.0042	58	-0.0026	-0.0101
19	0.0058	-0.0048	39	-0.0149	-0.0012	59	-0.0025	-0.0043
20	0.0011	-0.0077	40	-0.0006	0.0019			

表 5

性状对主因子的绝对贡献和相对贡献

性 状	第一主因子		第二主因子	
	$O_{a_k}(j)$	$O_{r_k}(j)$	$O_{a_k}(j)$	$O_{r_k}(j)$
1	0.0121	0.2603	0.0344	0.5945
2	0.1299	0.9377	0.0054	0.0313
3	0.0003	0.0996	0.0001	0.0385
4	0.0001	0.2379	0.0003	0.4061
5	0.0002	0.1344	0.0004	0.2316
6	0.0837	0.5016	0.0920	0.4412
7	0.0009	0.0414	0.0004	0.0161
8	0.052	0.0103	0.00004	0.2139
9	0.0008	0.3280	0.034	0.051
10	0.0021	0.0760	0.0286	0.8420

据知,各特征根的大小代表各综合指标方差的大小,各特征根累计的百分率代表各综合指标为总方差贡献的百分比。

由表 3 和图 1 可以看出,第一因子轴 (F_1) 方差贡献达到 56.47%,因而, F_1 是本地区 59 个品种 10 个数量性状所提供的生物信息中占主导地位的一个因子轴。在 F_1 因子载荷中所占比重最大的性状是株高(0.0547),

且位于最上端,其他特征系数,按其因子载荷排序为:单株粒数、小区产量、千粒重、结实小穗数、穗长、单株穗数、单株粒重、主穗粒数、主穗粒重。由表 5 得知,株高和单株粒数对第一因子的出现作出了主要贡献,从相对贡献来看,第一主因子对株高、单株粒数和结实小穗数有较大的离差贡献,反映了株高和单株粒数受第一主因子的影

响。

第二个主因子轴(F_2)的方差贡献也达到36.14%，在这一因子轴中因子载荷比较大的性状是千粒重，且位于最左端，其余性状中因子载荷较大者有小区产量和单株粒数。从性状对主因子的两个贡献来看，绝对贡献大的性状为单株粒数，说明它对第二主因子的出现作出一定贡献，而相对贡献大的性状为千粒重与小区产量，受 F_2 所支配。

综上所述， F_1 和 F_2 两个因子轴对数量性状的认识，第一主因子为长度和粒数因子，第二主因子为重量因子。从性状的图上分布来看，穗长与结实小穗穗数、单株穗数较接近；单株粒重与主穗粒数、主穗粒重较接近；株高、千粒重、小区产量、单株粒数散布在四个边缘处。

为了便于解释，在图1按性状和品种的聚合自然趋势，将全部性状和品种划分为I—IV。这样，就可根据每区所含性状的含义来对区内品种进行解释。

第一聚点区位于含单株粒数和小区产量的最下端，区内的品种有龙麦9号、克早一号、克全、克丰一号、克涝二号、克早6号、克早7号、克69—701等8个品种，在59个品种中产量较高的均落入这一聚点区。这些品种在第一主因子的出现作出了较大贡献。

第二聚点区位于含单株粒数和主穗粒数的右下端，属 F_1 、 F_2 同时受影响的混合区，该区内品种有22个粒数型品种，呈半园状散布，单株粒数在182—263粒，主穗粒数在36—63粒。该区内22个品种包括了第一聚点区的品种，我们也可把这两区合并为一区再分为两个亚区。这样，结合类内的品种在黑龙江省各历史时期所表现的贡献，反映了

粒数在提高产量中的意义及其作用。

第三聚点区位于右上端，属于株高、结实小穗数、单株穗数，受 F_1 所支配。区内品种有12个，株高范围在100—114厘米。

第四聚点区位于中左端，属于千粒重、穗长、主穗粒重和单株粒重，主要受 F_2 的影响，区内有25个粒重品种，千粒重范围在32—39克。

小 结

1. R-型因子分析表明，第一因子轴(F_1)的方差贡献达56.47%，在 F_1 因子载荷中所占比重最大的性状是株高和单株粒数因子；第二因子轴(F_2)的方差贡献为36.14%，这一因子轴中因子载荷所占比重较大的性状是千粒重和小区产量因子。

2. 对应分析结果表明，品种与性状的聚合自然趋势，将全部性状与品种划分为IV区，仅就影响产量的特征系数考虑，单株粒数和千粒重，同时受两个主因子轴的控制。

3. 本研究供试的59个品种在黑龙江省各历史时期所表现的贡献与对应分析结果是一致的，说明它提供的信息，也就是说，同一类型的品种将为邻近的性状所表征。

参 考 文 献

- [1] 王学仁：1982。地质数据的多变量统计分析 科学出版社
- [2] 张尧庭 方开泰：1982。多元统计分析引论 科学出版社
- [3] 刘来福：1979。遗传学报，6(3)：349—195
- [4] 杨德、戴君锡：1982。遗传学报，9(3)188—195
- [5] Bhatt G、M:1970、Aust. J. Agric. Res, 21: 1-7
- [6] 于崇文等：1980。数学地质的方法与应用 冶金工业出版社