# 主成分分析在猪选种选配中的应用

王先明 刘益钱 金访中 (杭州市种猪试验场 311115)

【摘要】 用杭州种猪试验场 1999 年春产 347 头长白后备种猪性能测定记录,取平均日增重($X_1$)、达 90kg 日龄($X_2$)、6 月龄体重($X_3$)、体长($X_4$)、体高($X_5$)、腿臀围($X_6$)、背膘厚度($X_7$)共 7 个性状作主成分分析。结果表明:主成分作为综合性状可用于选种,一个主成分可构成一个被选择的综合性状,也可用两个主成分构成一个选择指数,后备猪在主成分上的得分作为选择依据。个体在各个主成分上的得分能很好地反映出该个体在各综合性状上的优劣,因此能对每头后备猪作出较全面的评价,可作为制订选配计划的部分依据。

关键词:猪,选种选配,主成分分析。

## 前言

主成分分析法可将多个性状,根据它们间的相关性,找出能反应主要信息的少数几个彼此间独立的综合性状,并显示出更稳定的统计性质。它主要用于简化数据结构,寻找综合因子,样品排序及分类等方面。近年来在家畜育种工作中的应用也陆续有些报道[1,2,3],但都局限在对多个测定性状的综合分析,找出一些综合因子即综合性状,并进行种种描述,未对样品进行得分分析和排序,也未阐明它在育种工作中的具体应用方法。本研究将后备猪性能测定所得的 7 个性状,用主成分分析法综合成几个综合性状后,再计算出每头后备猪在各个主成分(即综合性状)上的得分多少,并进行排序,探讨在猪选种和选配工作中应用的可能性和具体应用方法。

## 1 材料和方法

### 1.1 材料

本研究所用资料取自本场承担的国家攻关项目《中国瘦肉猪新品系 $S_{IV}$系的持续选育与提高》,1999 年春产 347 头后备猪在性能测定中生长发育性能测定记录,分析性状有:2 月龄至 90kg 体重期间的平均日增重($X_1$)、达 90kg 日龄($X_2$)、6 月龄体重($X_3$)、体长($X_4$)、体高($X_5$)、腿臀围($X_6$)、90kg 体重时用 B 超测定的背膘厚度($X_7$)。

### 1.2 分析方法[4]

1.2.1 主成分贡献率:

$$\lambda_i / \sum_{k=1}^{h}\lambda_K (i = 1,2 \cdots\cdots, k; k = 1,2 \cdots\cdots, h) \quad \cdots\cdots (1)$$

式中 $\lambda_i$ 为第 i 个主成分的特征根。

1.2.2 样本主成分得分

$$Z_i = a_{kj} \cdot X_{ij}' \quad \cdots\cdots (2)$$

式中:$a_{kj}$为第 K 个主成分特征根相对应的特征向量;

$X_{ij}$ '为样品原始数据的标准化结果,其公式为:

$$X_{ij} \rightarrow X'_{ij} = \frac{X_{ij} - \bar{X}_j}{S_j}$$

全部数据用 SAS 软件进行 r 型主成分分析。

## 2 结果与分析

### 2.1 各主成分的特征根、贡献率、特征向量

表 1 主成分特征根、贡献率、特征向量

| 主成分 | | $Z_1$ | $Z_2$ | $Z_3$ | $Z_4$ | $Z_5$ | $Z_6$ | $Z_7$ |
|---|---|---|---|---|---|---|---|---|
| 特征根 | | 2.9719 | 1.3030 | 1.0304 | 0.8510 | 0.7014 | 0.1012 | 0.0409 |
| 贡献率(%) | | 0.4246 | 0.1861 | 0.1472 | 0.1215 | 0.1002 | 0.0145 | 0.0059 |
| 累计贡献率(%) | | 0.4246 | 0.6107 | 0.7579 | 0.8795 | 0.9797 | 0.9942 | 1.0000 |
| 贡献率 | $X_1$ | 0.5529 | -0.0577 | 0.0445 | -0.1344 | -0.0597 | 0.8142 | 0.0674 |
| | $X_2$ | -0.5630 | 0.0524 | -0.0705 | 0.1265 | 0.0399 | 0.3533 | 0.7299 |
| | $X_3$ | 0.5657 | 0.0022 | 0.0443 | -0.0697 | -0.0600 | -0.4584 | 0.6777 |
| | $X_4$ | 0.0165 | 0.6951 | 0.0721 | 0.2187 | -0.6791 | 0.0238 | -0.0423 |
| | $X_5$ | 0.0791 | 0.6607 | 0.2596 | -0.0258 | 0.6988 | 0.0266 | -0.0078 |
| | $X_6$ | 0.2250 | -0.0059 | -0.5239 | 0.7948 | 0.2028 | 0.0245 | -0.0373 |
| | $X_7$ | 0.0085 | -0.2745 | 0.8025 | 0.5301 | -0.0226 | 0.0165 | 0.0050 |

特征根的大小表示该主成分所包含的信息多寡,而特征向量则表示某个主成分中,各个性状所起作用的大小,又称系数载荷[2],每个主成分都可由它的特征向量构成主成分方程,如第 1 主成分方程为:

$$Z_1 = 0.5529X_1 - 0.5630X_2 + 0.5657X_3 + 0.0165X_4 + 0.0791X_5 + 0.225X_6 + 0.0085X_7$$

从表 1 中可以看出,第 1 主成分的特征根最大,贡献率达 42.46%,表明它所含的信息量占总信息量的 42%,这个主成分中的特征向量以平均日增重($X_1$)和 6 月体重($X_3$)为最大,这两个性状对 $Z_1$ 所起的作用最大,它们的共同特点是都表示猪的生长速度,因此可称它为生长速度综合因子或生长速度综合性状。在 $Z_1$ 里达 90kg 体重日龄($X_2$)也有很大的特征向量但呈现负值,表明它与 $Z_1$ 有负的相关关系。第 2 主成分 $Z_2$ 贡献率为 18.61%,是第 2 个信息含量最多的主成分,它对应的特征向量以体长($X_4$)和体高($X_5$)为最大,这两个

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

性状可勾画出猪的体形轮廓,故可称它为体形综合因子。第 3 主成分 $Z_3$ 的贡献率为 14.72%,起作用最大的性状是背膘厚度($X_7$),故可称为背膘综合因子,其中腿臀围($X_6$)所起的作用也不小,但呈现为负值,表明背膘越薄则腿臀围越大。其他主成分因贡献率少或与前面主成分的含义相似可不予解释。本研究的目的是在选种和选配上的应用,不追求总贡献率的大小,根据育种工作的要求选取前面三个主成分作进一步分析。

## 2.2 样本的主成分得分

将每头后备猪的实测值经标准化后代入公式(2),即得各头猪在各主成分上的得分,选取部分有代表性的个体分值列于表 2:

表 2 部分样本的主成分得分

| 耳号 | 代码 | $Z_1$ | $Z_2$ | $Z_3$ |
|---|---|---|---|---|
| 990778 | 1 | 5.5537 | -0.3290 | -1.4321 |
| 991320 | 2 | 4.2955 | -0.5208 | 0.8281 |
| 991328 | 3 | 3.6545 | 0.3778 | -0.7944 |
| 991226 | 4 | 3.3194 | 4.1980 | -0.9069 |
| 991322 | 5 | 2.8713 | 1.1495 | -1.4779 |
| 990448 | 6 | 1.1775 | 1.6675 | 1.9909 |
| 990747 | 7 | 0.9504 | 0.3117 | 0.4159 |
| 990296 | 8 | 0.5551 | 0.0657 | 1.1919 |
| 990792 | 9 | 0.0010 | -0.3586 | 0.5230 |
| 990228 | 10 | -0.2468 | -0.4589 | 1.3840 |
| 990090 | 11 | -1.7534 | 0.9455 | -0.1249 |
| 990932 | 12 | -2.4096 | -0.0084 | -1.3005 |
| 990523 | 13 | -3.3857 | -5.5648 | -0.4688 |
| 990544 | 14 | -4.2552 | -1.0743 | -0.8004 |
| 991503 | 15 | -5.3999 | 1.1575 | -0.7214 |

表 2 中 $Z_1$ 是按个体得分大小顺次排列的,分值由正到负,在零左右的个体,表示在该主成分所代表的主要性状上的表型值处在群体平均数上下,如 990792 号猪,在 $Z_1$ 上得分为 0.0010,其 6 月龄体重($X_3$)实测值为 100.3kg,群体平均数为 101.0kg。得分高表示个体在该主成分所代表的主要性状上有很高的实测值,如 990778 号猪第 1 主成分的得分 5.5537 为全群最高,6 月龄体重实测值为 130.6kg,也是全群最高。前面已经提到第 1 主成分称为生长速度综合因子,得分高的个体生长速度也快;得分低的个体生长速度就慢。若按个体在第 1 主成分上得分高低进行选择,就是对生长速度性状的选择,这样一个主成分就可作为一个可用于选择的综合性状。

由表 2 还可发现,每头后备猪都有在各个主成分上的得分,从而可以看出它们在各个性状上的优劣。如 990778 号猪它的生长速度虽居全群之冠,但它的第 2 主成分是负值,表明它的体形在全群平均数之下,第 3 主成分得分也是负值,表明它的背膘较薄而腿臀围较大瘦肉率较高。这样我们可对每头猪作出较全面而正确的评价,既有利于选种更有利于选配计划的制订。

## 2.3 样本在平面坐标上的图形

样本在主成分上的得分就是在该主成分轴上的坐标,任意两个主成分都可构成一个平面坐标系,本研究中部分个体在 $Z_1$ 和 $Z_3$ 主成分构成的平面坐标上的分布如图 1 所示:

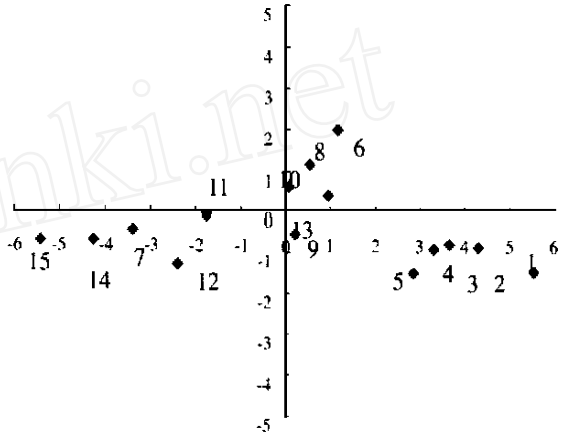


图 1 样本在 $Z_1$ 和 $Z_3$ 平面上的分布

图 1 清楚地表明,散落在不同象限上的个体具有不同的特性,在第 I象限上的个体生长速度快但背膘较厚,在第 II象限上个体正好相反。在第 IV象限上的个体生长速度快而且背膘也较薄是很好的种用对象,因此利用主成分构成的坐标系可作为对两个综合性状同时进行选择的一种方法。

## 3 讨论

3.1 主成分分析能提供一些综合性状可作为选择性状。根据主成分分析的原理,每个主成分都是由原性状间的相关关系而重新组成的一组线性组合[4],它包含着所有原始性状所提供的信息,又突出一些相关密切的主要性状,形成一个新的线性综合性或称综合因子。很明显性状是由若干个普通性状合成的,英国数量遗传学家 K 马瑟称它为"超性状"。这种综合性状在生物界中普遍存在,如畜禽的"体形"就是一个综合性状,它至少由体长、体高两个普通性状合成。综合性状一般无法测定也难有具体指标。但有些综合性状却都是育种的重要性状,象体形很小的一些地方猪种在改良中就要增大它的体形,这时若将一些测定性状资料

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

经主成分分析,就可形成一个赋有具体指标能说明体形的综合性状,并有一个测定指标,这个指标就是个体在这个主成分上的得分,按个体得分多少进行选择就能实现对体形这个综合性状的选择,达到化繁为简的目的。根据育种工作的需要还可以用两个主成分进行联合选择,这时既可用平面坐标系形式在图上进行选择,如本研究结果 3.2 小节所示,又可在两个主成分得分前添加权重系数,再合成一个主成分选择指数进行选择,根据我们试验检验结果,在相同权重下,主成分选择指数与无约束选择指数,在被选中个体的排序上是非常一致的。

3.2 样本在主成分上的得分可作为选择的依据。样本在主成分上的得分是用样本在各个观察性状上的实测值经标准化后,代入主成分方程计算出来的,其分值大小取决于方程中各性状的系数和实测的表型值的大小。一个样本(个体)在某个主成分中几个系数都很高的性状上表型值又都很大,那么样本在这个主成分上的得分就高,在这个主成分所代表的综合性状上表现就好,因此可以用样本的得分作为选择的依据。例如第 2 主成分又称体形综合性状,最大系数性状是 $X_4$ 和 $X_5$,凡在这两个性状上表型值大的个体在 $Z_2$ 上的得分就较高,体形也较高大,表 2 中的 991226 号后面猪的得分很高,它的 6 月龄体重($X_4$)实测为 119cm,体高($X_5$)为 61cm,都在平均数以上,这是一头体形较高大的猪。当然同时还要注意,分值的大小还受到系数较大又为负值性状的影响,这时若负值性状的表型值很大,会明显降低样本的得分,对得分处在选择临界线上的个体应特别注意。

3.3 主成分分析有利于加深对样品的全面认识。在主成分分析中,被选用的主成分都有较大的特征根,包含着较多的信息量,并能表示出某个综合性状。样本在主成分上的得分说明样本(个体)的某个综合性状在群体中所处的位置,每个样本都能获得在各个主成分上的得分,从而可使人们从多方面加深对样本的认识,进而可对每个样本作出较全面而客观的评价,这在选种和选配工作中是十分重要和必须的。这是主成分分析的一个很大优点,也是其他分析方法不可比拟的地方。通过本文 3.2 小节最后一段对 990778 号猪的分析,我们就能对它有一个较全面而又具体的认识。在选择中可对每头被选的猪进行评定和比较,以提高选择的效果。在选配中可根据被选留公母猪的特点,根据育种要求进行合理选配,定能加快选育进展。

3.4 本研究的主成分分析是用实测的表型值为原始材料,进行 r 型分析所获得的结果,对遗传力较高用表型选择有较好效果的性状,用主成分选择法是可行的,所获得的信息用于制订选配方案是十分有效的。作为一种选种方法还应能算出选择反应,选择进展等参数,这些都有待于深入研究,若从性状间的遗传相关开始进行主成分分析,可望能帮助解决前述问题。

**主要参考文献**


[1] 郭万正,徐子清. 猪活体测定性状的主成分分析. 湖北畜牧兽医,1997(2),11～13

[2] 项云. 主成分分析在金华猪瘦肉系选育中的应用. 湖北畜牧兽医,1995(4),11～13

[3] 张代坚,赵含章. 因子分析在肉质评定与选育中的应用. 中国畜牧杂志,1993,29(3),25～27

[4] 王天行,张泽. 多元生物统计学. 成都科学技术出版社,1992,126～131


本文蒙南京农业大学动物科技学院
杨茂成教授审阅指导,特此致谢!

国家科技部批准 教育部主管 华中农业大学主办
《养殖与饲料》邮发代号:38 - 381
独特的行业媒体 及时的市场信息 实用的专业技术 优良的读者服务
《养殖与饲料》告诉您:可家庭化经营 ——小富小康;可产业化发展 ——富甲一方
订阅《养殖与饲料》 科技致富指门道
全年定价:22.8 元 地址:武汉华中农业大学校内《养殖与饲料》杂志社 邮编:430070

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net