

doi: 10.7690/bgzd.2016.08.012

## 基于影响因素主成分分析的航材消耗预测模型

张 梁<sup>1</sup>, 崔崇立<sup>2</sup>, 贵徐伟<sup>3</sup>, 辛 昱<sup>1</sup>, 曹 亮<sup>4</sup>(1. 空军勤务学院学员一大队二队, 江苏 徐州 221000; 2. 空军勤务学院航材管理系, 江苏 徐州 221000;  
3. 中国人民解放军 94906 部队, 江苏 苏州 215157; 4. 中国人民解放军 94783 部队, 浙江 湖州 313111)

**摘要:** 为了确定航材消耗预测各影响因素的影响程度, 提高航材消耗规律分析及其预测的准确性, 提出一种基于影响因素主成分分析与广义线性回归的航材消耗预测模型。对航材消耗影响因素进行定性、定量分析, 通过主成分分析筛选变量, 对影响因素数据进行降维, 得到最佳的影响因子集合; 然后通过广义线性回归构建航材消耗方程, 并验证其拟合优度, 最后进行消耗预测。分析结果表明: 该模型降低了影响因子相关性, 减小了回归分析难度, 提升了回归分析的精度, 提升了航材消耗预测模型的准确度。

**关键词:** 航材消耗; 主成分分析; 广义线性回归

**中图分类号:** TP391.92 **文献标志码:** A

## Prediction Model of the Aviation Materials Consumption Based on PCA of Influencing Factors

Zhang Liang<sup>1</sup>, Cui Chongli<sup>2</sup>, Ben Xuwei<sup>3</sup>, Xin Yu<sup>1</sup>, Cao Liang<sup>4</sup>(1. No. 2 Team, No. 1 Cadet Brigade, Air Force Logistics College, Xuzhou 221000, China;  
2. Department of Air Material Management, Air Force Logistics College, Xuzhou 221000, China;  
3. No. 94906 Unit of PLA, Suzhou 215157, China; 4. No. 94783 Unit of PLA, Huzhou 313111, China)

**Abstract:** To define the multi-factor affecting degree of consuming prediction of aviation materials and improve the accuracy of consumption analysis and prediction, we proposed a method for predicting the consumption of aviation materials based on principal component analysis (PCA) of influencing factors and generalized liner regression. Firstly, quantitatively and qualitatively analysis the influence factors of aviation materials consumption. Then, screen out the main factors by PCA and make it dimensionality reduction to get the best collection of variables. Finally, use the generalized linear regression to build the air material consumption equation, and verify its goodness-of-fit, and finally make consumption forecast. The results show that the model reduces the impact factor correlation, and the difficulty of regression analysis, and improves the precision of regression analysis, and the accuracy of the prediction model of aviation material consumption.

**Keywords:** consuming of aviation materials; PCA; generalized liner regression

### 0 引言

航空器材的消耗预测是空军航材保障工作中的重要环节。该项工作直接影响空军航空兵部队作战训练任务的完成和保障工作的军事、经济效益<sup>[1-2]</sup>。进行航材消耗预测的关键在于准确掌握航材历史消耗信息和恰当选择分析方法, 以确定影响因素的影响程度, 从而为航材保障实践活动提供指导。基于此, 笔者提出一种基于影响因素主成分分析与广义线性回归的航材消耗预测模型, 采用定性、定量分析的方法, 分析环境条件和保障任务等诸多航材消耗影响因素, 通过主成分分析确定影响主因子, 减小影响因素相关性对航材消耗规律线性回归分析的影响, 从而减小广义线性回归建模的工作量, 提高航材消耗规律分析的准确性。

### 1 模型介绍

#### 1.1 主成分分析

在多数实际问题中, 数据的不同指标(属)之间有一定的相关性。主成分分析是利用降维的思想, 将多个变量转化为少数几个综合变量(即主成分), 其中每个主成分都是原始变量的线性组合, 各主成分之间互不相关, 使这些主成分能够反映原始变量的绝大部分信息, 且所含的信息互不重叠<sup>[3]</sup>。计算步骤如下:

##### 1) 变量标准化。

协方差矩阵的计算根据自变量不同分为 2 种情况: 当各变量的度量单位相同且数值差异不大时, 可直接计算变量的协方差矩阵; 反之, 则需要将原

收稿日期: 2016-04-15; 修回日期: 2016-05-12

作者简介: 张 梁(1992—), 男, 四川人, 在读硕士, 从事航材保障决策与信息技术研究。

始变量做标准化处理，减小或者消除变量之间的方差差异，然后计算标准化变量的协方差矩阵，以此保证主成分分析的结果不会过分倾向于方差大的变量。利用各个变量的标准差归一化进行标准化转换，改变变量的均值和标准差，使变量变换到以 0 为中心的区域，计算方法为

$$X = \frac{x_i - \bar{x}}{\sigma} \quad (1)$$

其中： $X$  为变换后的值； $X_i$  为待变换的值； $\bar{x}$ 、 $\sigma$  分别为变量的均值和标准差。

2) 计算相关系数矩阵

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ r_{n1} & r_{n2} & \cdots & r_{nn} \end{bmatrix}$$

$r_{ij}(i, j=1, 2, 3, \dots, n)$  为原变量  $X_i$  与  $X_j$  的相关系数。

$R$  为对称矩阵( $r_{ij}=r_{ji}$ )，其计算公式为

$$r_{ij} = \frac{\sum_{k=1}^n (X_{ki} - \bar{X}_i)(X_{kj} - \bar{X}_j)}{\sqrt{\sum_{k=1}^n (X_{ki} - \bar{X}_i)^2 \sum_{k=1}^n (X_{kj} - \bar{X}_j)^2}}$$

3) 计算特征值与特征向量。

解特征方程  $|\lambda I - R| = 0$ ，通常采用雅比克法求出特征值  $\lambda_i (i=1, 2, \dots, n)$ ，并使其按大小顺序排列，即  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$ ；求出对应于特征值  $\lambda_i$  的特征向量  $e_i (i=1, 2, 3, \dots, n)$ ，并要求  $\|e_i\| = 1$ ，即  $\sum_{j=1}^n e_{ij}^2 = 1$ ，

其中  $e_{ij}$  表示向量  $e_i$  第  $j$  个分量。

4) 计算主成分贡献率及累计贡献率。

$$\text{贡献率: } \lambda_i / \sum_{k=1}^n \lambda_k \quad (i=1, 2, \dots, n)。$$

$$\text{累计贡献率: } \sum_{k=1}^i \lambda_k / \sum_{k=1}^n \lambda_k \quad (i=1, 2, \dots, n)。$$

选取主成分的原则：特征值大于 1 且累计贡献率达到 80%~95%，特征值  $\lambda_1, \lambda_2, \dots, \lambda_m$  所对应的是第 1, 第 2,  $\dots$ , 第  $m (m \leq n)$  个主成分。

5) 计算主成分载荷。

$$t_{ij} = \sqrt{\lambda_i} e_{ij} \quad (i, j=1, 2, \dots, n)。$$

主成分表达式为： $C_k = t_{1k}X_1 + t_{2k}X_2 + \dots + t_{nk}X_n$ ，称系数  $t_{ij}$  为第  $k$  个成分  $C_k$  在第  $i$  个变量  $X_i$  上的载荷，载荷的绝对值越大，主成分涵盖该变量的信息越多，越可以代表该变量。同时，线性相关的变量会在同一个成分中同时具有较大的载荷<sup>[4]</sup>。

## 1.2 航材消耗预测模型

由于航材消耗的影响因素众多，且不同系统的航材消耗影响因素也有差异，在进行航材消耗预测时不能一概而论，应当针对不同系统中的航材，有针对性地筛选影响其消耗的主要因素，从而提高预测的准确性。综上所述，笔者将主成分分析应用到回归分析建模中，分析航材消耗的主要因素，为回归分析建模提供线性相关性低且数据可靠的变量。如图 1 所示，建模步骤为：

1) 利用主成分分析筛选出影响航材消耗的主要因子，将主因子作为回归分析建模的输入；

2) 选择同类型的航材实际消耗数据，按照主成分分析产生的主因子表达式，计算出输入的自变量取值，从而构造出此类航材消耗规律的回归函数，最后将当前的样本数据作为回归函数的输入，来预测出消耗值。

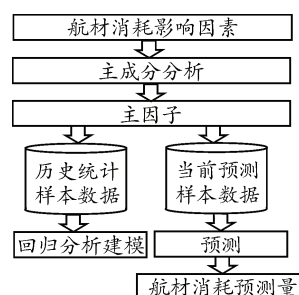


图 1 航材消耗预测模型结构

## 2 航材消耗影响因素主成分分析

### 2.1 定性分析

通常，影响航材消耗的因素主要包括航材自身性能、保障任务和环境条件 3 类<sup>[2]</sup>。笔者结合航材保障工作实际，初步确定影响航材消耗的因素<sup>[1-2,5]</sup>如下：

1)  $X_1$  为单机年均飞行小时数，编制单位当年实际飞行总小时数与在编飞机总数之比；

2)  $X_2$  为单机年均起落数，编制单位当年实际飞行总起落数与在编飞机总数之比；

3)  $X_3$  为编制单位飞机总数，编制单位现有飞机总数；

4)  $X_4$  为飞行训练当量，各飞行训练科目的训练量折算值；

5)  $X_5$  为年异常温度系数，规定航材保管库房温度 5~30℃ 为正常温度，把日均温度不在规定值的视为异常温度日，把当年异常温度日与正常的天数之比作为该因素的值；

6)  $X_6$  为年异常湿度系数, 规定航材保管库房相对湿度 45%~70%为正常值, 把相对湿度不在规定值的视为异常湿度日, 把当年异常湿度日与正常的天数之比作为该因素的值;

7)  $X_7$  为人员素质, 人员包括航材管理人员、机务人员和飞行人员。统计每一类人员的中高级人员数, 算出每一类中高级人员对应类别总人数的比例, 然后乘以相应类别人员的权重, 进而求和确定该因素的综合量化值, 通过综合量化值来衡量各类人员素质对库存备件消耗的影响情况。

根据专家建议初步认定: 初级飞行人员是指尚未完成飞行改装训练计划的飞行人员, 将其他飞行人员视为中高级人员; 初级维修人员是指独立保障经验未满 4 a 的战士或独立保障经验未满 2 a 的干部, 将其余维修人员视为中高级人员; 初级库存管理者是指独立保障经验未满 2 a 的战士或独立保障经验未满 1 a 的干部, 将其余库存管理者视为中高级人员<sup>[5]</sup>。其中需要量化处理的是人员素质和飞行训练当量。

## 2.2 量化处理

在实际保障过程中, 人员素质影响因素的 3 个子因素: 飞行人员素质、维护人员素质、航材管理人员素质对航材消耗影响程度是不一样的。此外, 不同的飞行科目对航材使用的频率和程度也不一样; 因此, 笔者采用层次分析法对各子因素进行综合评价, 求解出影响因素的综合量化值。

### 2.2.1 人员素质影响因素量化处理

#### 1) 建立判断矩阵。

采用 9 分位比率排定各子因素的相对优劣顺序, 以此构造出子因素的判断矩阵。根据专家建议, 对各因素给出如下的判断矩阵:

$$A = \begin{bmatrix} 1 & 1/3 & 1/2 \\ 3 & 1 & 2 \\ 2 & 1/2 & 1 \end{bmatrix}。$$

#### 2) 层次排序。

在此排序中, 只有一个层次; 因此, 单排序则为总排序。笔者用几何平均法计算判断矩阵权重:

计算判断矩阵  $A$  各行元素乘积:

$$M_1 = 1 \times 1/3 \times 1/2 = 1/6;$$

$$M_2 = 3 \times 1 \times 2 = 6;$$

$$M_3 = 2 \times 1/2 \times 1 = 1。$$

计算  $M_i$  的 3 次方根:

$$\omega_1 = 0.5503, \quad \omega_2 = 1.8171, \quad \omega_3 = 1。$$

归一化处理:

$$\omega_1 = 0.1634, \quad \omega_2 = 0.5396, \quad \omega_3 = 0.2970。$$

#### 3) 一致性检验。

最大特征根:

$$\lambda = \sum_{i=1}^n \frac{(A\omega)_i}{n\omega_i} = 3.0093 \quad (n=3)。$$

一致性指标:

$$CI = \frac{\lambda - n}{n - 1} = 0.0047 \quad (n=3)。$$

经查表得到平均随机一致性指标  $RI=0.58$ 。

随机一致性指标:

$$CR = \frac{CI}{RI} = 0.0080 < 0.10。$$

一致性检验说明判断矩阵一致性较好, 符合建模要求。

设某部飞行人员总数为  $f$ , 中高级飞行员人数为  $f_1$ ; 维修人员数为  $w$ , 中高级维修人员数为  $w_1$ ; 航材管理人员为  $h$ , 中高级航材管理人员数为  $h_1$ 。

人员素质综合量化值方程为:

$$X_7 = \frac{f_1}{f} \omega_1 + \frac{w_1}{w} \omega_2 + \frac{h_1}{h} \omega_3。$$

### 2.2.2 飞行训练当量影响因素量化处理

#### 1) 建立判断矩阵

$$B = \begin{bmatrix} 1 & 1/2 & 1/3 & 1/5 & 1/7 \\ 2 & 1 & 1/2 & 1/3 & 1/5 \\ 3 & 2 & 1 & 1/2 & 1/3 \\ 5 & 3 & 2 & 1 & 1/2 \\ 7 & 5 & 3 & 2 & 1 \end{bmatrix}。$$

#### 2) 层次排序与一致性检验。

由于与人员素质计算方法一样, 在此省略具体的计算过程。

$$\omega_1 = 0.0523, \quad \omega_2 = 0.0887, \quad \omega_3 = 0.1525,$$

$$\omega_4 = 0.2622, \quad \omega_5 = 0.4443。$$

一致性检验符合建模要求,  $A_i$  为第  $i$  个科目年飞行小时数, 飞行训练当量影响因素量化公式为:

$$X_4 = \omega_1 A_1 + \omega_2 A_2 + \omega_3 A_3 + \omega_4 A_4 + \omega_5 A_5。$$

## 2.3 航材消耗影响因素分析<sup>[6]</sup>

选取收集到的某航空兵部队在 2005 年至 2011 年航材消耗影响因素相关数据, 经式 (1) 标准化后, 如表 1 所示。

表 1 标准化后的影响因素数据

年份	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$
2005	-1.467	-1.362	-0.751 0	-1.430 0	-2.226 0	-0.492 0	-1.412 0
2006	-0.843	-1.004	1.351 7	-0.727 0	0.699 0	1.166 7	-0.947 0
2007	-0.499	-0.396	-0.750 0	-0.624 0	0.498 0	-1.325 0	-0.343 0
2008	0.081	-0.033	-0.750 0	-0.037 0	0.419 0	1.055 6	-0.237 0
2009	0.452	0.570	0.300 4	0.491 0	0.207 6	-0.523 0	0.870 0
2010	0.940	0.811	1.351 7	0.958 4	0.237 6	0.849 2	0.914 2
2011	1.336	1.415	-0.751 0	1.369 0	0.164 0	-0.730 0	1.155 0

笔者借助 SPSS 软件完成对航材消耗影响因素的主成分分析，根据主成分分析结果(如表 2)，选取累计贡献率 90%以上，提出 2 个影响因素主成分。

表 2 影响因素的解释总方差

影响因素	主成分	
	1	2
$X_1$	0.988	-0.086
$X_2$	0.776	-0.199
$X_3$	0.411	0.254
$X_4$	0.888	-0.052
$X_5$	0.097	0.733
$X_6$	0.037	0.885
$X_7$	0.976	-0.145

主成分系数如表 3 所示，从表中可以看出： $X_1, X_2, X_3, X_4, X_7$  在第一主成分上载荷较大， $X_5, X_6$  在第二主成分上载荷较大；因此，将 7 影响因素降为 2 个，即保障任务和环境条件。

表 3 主成分系数

成份	合计	方差的/%	累积/%
1	4.259	60.875	60.875
2	1.769	25.282	86.157
3	0.562	8.038	94.195
4	0.390	5.577	99.772
5	0.015	0.210	99.982
6	0.001	0.018	100.000
7	-1.0E-013	-2.716E-15	100.000

根据特征向量矩阵(略)得到主成分  $C_1, C_2$ ，其表达式为：

$$C_1 = -0.479X_1 - 0.472X_2 - 0.102X_3 - 0.478X_4 - 0.290X_5 - 0.017X_6 - 0.473X_7; \quad (2)$$

$$C_2 = 0.064X_1 + 0.149X_2 - 0.643X_3 + 0.039X_4 - 0.327X_5 - 0.663X_6 + 0.109X_7. \quad (3)$$

### 3 航材消耗的广义回归分析

#### 3.1 广义回归分析建模

通过上一章的影响因素主成分分析，将原来的 7 个影响因素转化为 2 个主因子。通过已求得的主成分表达式，结合式 (1) 标准化后的航空兵某部某件航材消耗的历史数据，将原影响因素数据转化为主成分数据。标准化后的数据样本如表 4 所示。

表 4 标准化后的数据样本

$C_1$	$C_2$	消耗数
3.427 8	0.065 7	2.000 0
1.313 0	-0.469 6	3.000 0
0.842 1	0.080 8	5.000 0
0.045 0	-1.347 5	6.000 0
-1.214 9	0.690 4	6.000 0
-1.946 1	0.545 8	5.000 0
-2.471 4	0.424 5	7.000 0

利用 SPSS 分别将主因子  $C_1, C_2$  与消耗数进行曲线估计得到拟合图，如图 2、图 3 所示。

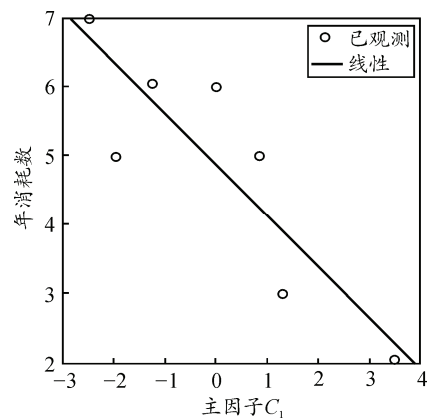


图 2 主因子  $C_1$  与消耗量的拟合散点

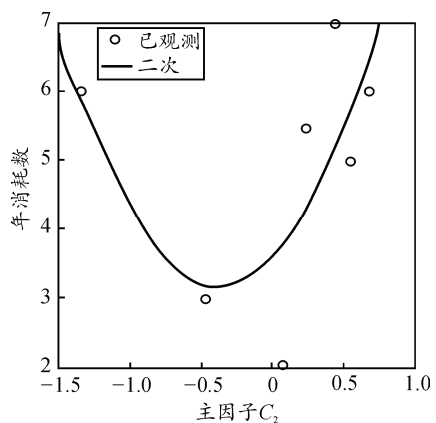


图 3 主因子  $C_2$  与消耗量的拟合散点

从图中可以看出，主成分  $C_1$  与消耗数大体呈线性相关， $C_2$  与消耗数呈二次项相关；因此，建立自变量  $C_1, C_2$  与因变量  $Y$  的二阶多项式回归模型：

$$y = k_1 C_1 + k_2 C_2 + k_3 C_2^2 + k_4$$

直接对  $C_1$ 、 $C_2$  进行回归分析比较困难。笔者先对  $C_1$  和  $C_2$  进行变量转换，使航材的消耗量满足多元线性关系，再运用多元线性回归模型对航材消耗量进行预测。令

$$Z_1=C_1, Z_2=C_2, Z_3=C_2^2,$$

得到  $y = k_1 Z_1 + k_2 Z_2 + k_3 Z_3 + k_4$ 。

### 3.2 模型有效性分析

建模结果分析如表 5 所示。

表 5 建模结果分析

模型		系数		t	Sig
		B	标准误差		
1	常量	4.857 0	0.366 0	13.257 0	0.000
	$Z_1$	-0.744 0	0.192 0	-3.882 0	0.012
2	常量	4.542 0	0.435 0	10.432 0	0.000
	$Z_1$	-0.710 0	0.185 0	-3.828 0	0.019
	$Z_3$	0.735 0	0.604 0	1.217 0	0.020
3	常量	4.529 0	0.726 0	6.237 0	0.008
	$Z_1$	-0.704 0	0.317 0	-2.219 0	0.113
	$Z_3$	0.763 0	1.411 0	0.541 0	0.626
	$Z_2$	0.032 0	1.3500	0.023 0	0.983

从表 5 可以看出：根据 SPSS 回归分析结果，当  $Z_2$  进入回归分析时，sig(p 值)显著增大<sup>[7]</sup>。因此，剔除  $Z_2$ ，得到回归函数：

$$y = -0.71C_1 + 0.735C_2^2 + 4.542$$

复相关系数和判定系数如表 6 所示。从表中可以看出：被解释变量与解释变量直接的复相关系数为 0.905，说明拟合效果非常好。此外，调整系数好也较为接近于 1；因此，可以认为拟合优度较高，被解释变量可以被模型解释的部分较多。

表 6 复相关系数和判定系数

模型	R	R 方	调整 R 方	标准估计的误差
2	0.905	0.818 2	0.727 3	0.925 8

### 3.3 实例验证分析

笔者选取该单位同项器材 2012—2015 年消耗数据与影响数据，经标准化处理后，分别代入经过主成分分析的消耗预测回归方程与未经过影响因素主成分分析的预测回归方程，计算得到预测误差结果，如图 4 所示。

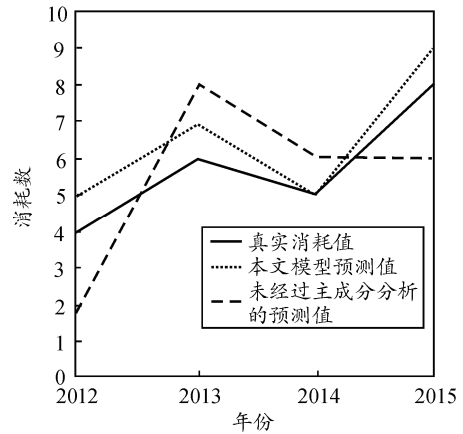


图 4 实例结果分析

从结果可以看出：影响因素主成分分析提高了航材消耗预测的精度，同时也简便了建立回归模型的难度与工作量；因此，笔者研究的基于影响因素主成分分析的航材消耗预测模型具有较高的实用性和可操作性。

## 4 结论

针对航材消耗的影响因素众多的问题，笔者定量分析了相关影响因素，利用主成分分析筛选出航材消耗的主要影响因子，从而降低了影响因子相关性，减小了回归分析难度，提升了回归分析的精度和航材消耗预测模型的准确度。

### 参考文献：

- [1] 何亚群, 谢瑞龙, 荣军. 航材供应学[M]. 长沙: 国防科技出版社, 2001: 29-35.
- [2] 韩兴才. 航材管理工程[M]. 北京: 蓝天出版社, 2003: 15-142.
- [3] 高鹏, 邢国平, 孙德翔, 等. 灰色关联支持向量机在备件库存中的应用[J]. 电光与控制, 2012(3): 100-105.
- [4] 张作刚, 胡新涛, 刘望. 主成分聚类分析在航材分类中的应用[J]. 兵工自动化, 2012, 31(11): 25-28.
- [5] 徐廷学, 陈洪光, 韩建立, 等. 改进加权分析方法的新机型航材消耗预测[J]. 四川兵工学报, 2012(7): 55-57.
- [6] 李蒙. 基于影响因素主成分分析的采购数量预测模型[D]. 大连: 大连理工大学, 2011.
- [7] 蔡信, 李波, 汪宏华, 等. 基于神经网络模型的动力电池 SOC 估计研究[J]. 机电工程, 2015, 32(1): 128-132.