

部队装备新旧状态的大数据分析

王玉琢, 张建军

(海军工程大学 理学院应用数学系, 武汉 430033)

摘要:对已获得的模拟大数据进行了清洗和关联的预处理工作。以高程、经度、纬度、装备类型、部队种类5个方面作为影响因素,建立了3种方差分析模型,分别讨论了每种因素对装备新、堪、待、废4种状态的影响情况,以及各种因素联合起来产生的联合效应。通过计算相关系数,将每种因素与装备状态的关联程度进行量化比较。所得结论和提出的建议对部队的装备保障优化具有一定的意义。

关键词:数据预处理;单因素方差分析;无交互作用双因素方差分析;有交互作用双因素方差分析

本文引用格式:王玉琢,张建军.部队装备新旧状态的大数据分析[J].兵器装备工程学报,2017(8):124-130.

Citation format:WANG Yuzhuo, ZHANG Jianjun. Big Data Analysis About the States of the Military Equipment[J]. Journal of Ordnance Equipment Engineering, 2017(8):124-130.

中图分类号:O212.1

文献标识码:A

文章编号:2096-2304(2017)08-0124-07

Big Data Analysis About the States of the Military Equipment

WANG Yuzhuo, ZHANG Jianjun

(Department of Applied Mathematics of School of Science,
Naval University of Engineering, Wuhan 430033, China)

Abstract: By cleaning and correlating the big data about the states of the military equipment, the post-processing data were obtained and five possible factors that affect the states of equipment were found. For the five factors, some fit models consist of the one-way ANOVA model, the no interaction two-way ANOVA model and the two-way ANOVA model with interaction were chosen to analyzed different kinds of impacts neither from each factor respectively nor from the interactions of several factors. By calculating the correlation coefficient, the correlation degree between the factors and the states of the equipment were measured and compared. The conclusions obtained and suggestions proposed form this paper had some certain significance for improving the guarantee ability.

Key words: pre-process data; one-way ANOVA model; no interaction two-way ANOVA model; two-way ANOVA model with interaction

近年来,随着互联网技术的飞速发展,军队信息化也得到加快,各种军事资源数据的采集、储存与利用,为部队的现代化建设和科学化管理提供了有力的支持与理论依据。但是面对海量信息,如何结合部队实际,真实高效地利用这些数据资源做好部队的装备保障与维护工作,是当前亟待分析解决的一个重要问题^[1]。

装备的磨损消耗情况,是保障维护工作中常关心的问

题,它除了与装备本身的耐用性、维护保养情况、使用频率这些因素直接相关外,还受到如气候条件这类因素的间接影响^[2]。然而,出于保密等原因有些数据无法获得,即使这些数据是重要影响因子样本。为了便于研究,本研究基于已获得的模拟数据进行讨论,这些数据或许不是重要影响因子样本,但所用方法具有可复制性,结论具有参考性,可供有关部门借鉴。

收稿日期:2017-04-28;修回日期:2017-05-29

作者简介:王玉琢(1984—),女,硕士,讲师,主要从事概率论与数理统计研究。

1 数据采集

通过添加随机扰动项对原始数据实现模糊处理,本文获取了部分部队的地理位置及装备状态的模拟大数据。在大数据背景下,文中所用的检验统计量依大数定律仍可以反映出总体特征。部分数据如表1、表2所示。

表1中包含9439支部队的经度、纬度、高程3种地理位置信息。其中部队层次码采用多层变长的数字结构,两位一个层次,前1~2位为大单位名称编码,第3~4位为下一级单位名称编码,用来体现单位之间的上下级关系。如400009与400101属于同一级大单位,65020904与6502090203同属一个大单位,前者比后者高一级别。

表2中共有163178条装备状态数据,包含装备当前的新旧状况、购买日期、现有数量3类信息。其中装备层次码3位为一个层次,用来反映装备的类属关系和层次关系。如

001001可能表示电脑配件大类中的液晶显示屏,001001001001可能表示该大类下的鼠标。装备新旧状况被划分为新、堪、待、废4种,分别对应内码01、02、03、04,用表3解释。

表1 部队地理位置信息

| 序号 | 部队层次码 | 经度/(°) | 纬度/(°) | 高程/m |
|-------|------------|--------|--------|-------|
| 1 | 400009 | 99.15 | 28.1 | 607.2 |
| 2 | 400101 | 97.32 | 29.36 | 672 |
| | | | | |
| 3078 | 65020904 | 121.1 | 36.26 | 50.4 |
| 3679 | 7103000003 | 119.05 | 49.41 | 177.6 |
| | | | | |
| 9439 | 99050102 | 107.28 | 35.16 | 1452 |

表2 部队装备状态信息表

| 序号 | 装备层次码 | 部队层次码 | 装备状况内码 | 购买日期 | 数量/个 |
|--------|--------------------|----------|--------|------------|-------|
| 1 | 001001001001 | 400105 | 2 | 2010/9/30 | 243 |
| 2 | 001001001001 | 400105 | 4 | 2011/9/30 | 192 |
| | | | | | |
| 47250 | 002003003 | 57040104 | 3 | 2014/10/18 | 3 |
| | | | | | |
| 163177 | 001005000000000000 | 99050200 | 1 | 2013/9/30 | 9 |
| 163178 | 001005000000000000 | 99050200 | 2 | 2014/9/30 | 10 |

表3 装备内码

| 新旧状况 | 装备状态内码 | 说明 |
|------|--------|------------|
| 新品 | 01 | 装备第一年为新品 |
| 堪用 | 02 | 新品从第二年转为堪用 |
| 待修 | 03 | 等待维修 |
| 待报废 | 04 | 等待报废 |

2 数据预处理

基于该数据,初步判断影响装备新旧状况的可能因素有5种:部队所在的经度、纬度、高程、部队类型、装备类型。

2.1 数据清洗

1) 清除空间上无用的数据。

考虑到我国军事单位大部分位于国境范围内,所以可将境外的部队地理信息视为无效数据进行剔除。中国国境范围为最北端漠河 53°33'N,最南端曾母暗沙 03°58'N,最东端黑瞎子岛 135°05'E,最西端帕米尔高原 73°49'E^[3] 以此作为

筛选依据,从表1中去掉324条数据,剩下9115条部队地理位置信息。

2) 清除时间上无用的数据。

表2中5年前数据量占比为0.15%,时间久远比重低,参考性不大。剔除掉此类数据247条后剩下162931条装备状态信息。

3) 清除信息不完整的数据。

通过比较发现,表2中有些部队层次码在表1中没有出现,即有些部队缺少地理位置信息,表1中有些部队层次码在表2中没出现,即有些部队缺少装备状态信息,这些数据对后续分析是无效的,需要删除。运用MATLAB软件进行两层循环遍历后,删除的不完整信息接近30%,缩减了后续数据分析的工作量。

2.2 数据关联

用EXCEL中的VLOOKUP函数,将处理后的两个表以“部队层次码”为桥梁整合到一个表中,最终该表包含87741条信息7个标题栏:部队层次码、装备层次码、装备状态内码、现有数量、经度、纬度、高程。

3 统计描述

3.1 装备状态的分类统计

装备的新旧状态为一个分类型变量,经统计4种新旧状态的可用装备总数比重依次为3.88%、94.73%、0.8%、0.52%,可见九成以上的装备均为堪用品,废品的比重最低。

3.2 各影响因素与装备状态之间的关联分析

3.2.1 地理位置与装备状态的关联分析

地理位置为一个三维连续型变量,首先对高程根据我国地貌特征划分为0~200 m、200~2 000 m和2 000~5 800 m 3类^[3],代表平原、丘陵、高原3种地貌。然后在每个海拔段上对经度、纬度采用K均值法聚类^[4],实现方位分区。综合考虑分区后各区拥有的单位数量,最终将低海拔、中海拔、高海拔再细分为7个区、4个区、3个区,总共聚为14类。

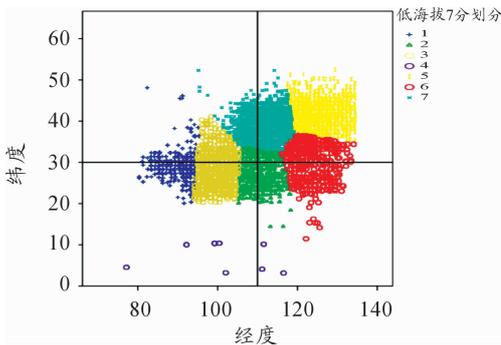


图1 低海拔7区划分

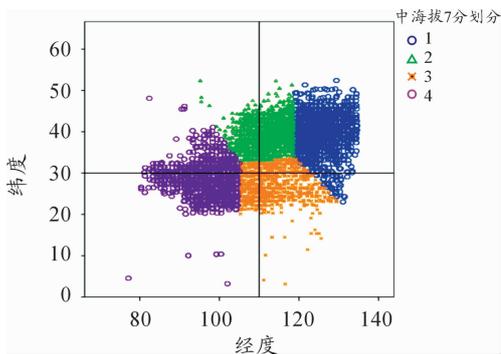


图2 中海拔4区划分

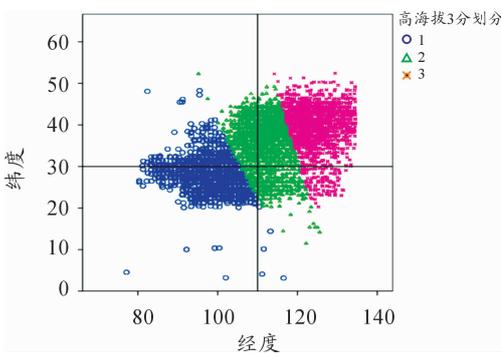


图3 高海拔3区划分

统计14个区域上各新旧状态的装备数量,可以通过折线图依次观察不同区域上的新品率、堪用率、待修率、报废率的波动情况。为了节省篇幅,这里通过一张百分比堆积柱形图(图4)反映变量间的比重关系,可见不同地理位置上的新品率、堪用率波动不大,待修率、报废率差异明显。

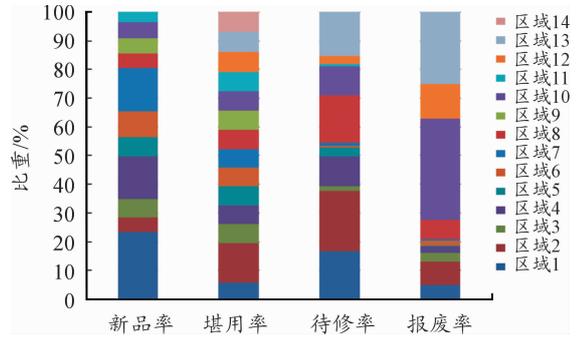


图4 不同区域下新旧装备数量比重

3.2.2 装备类型与装备状态的关联分析

经EXCEL分类汇总,表2中共有275种装备,分为001与002两大类,且98%为001型。由于装备类型众多,选择数量最多的前5种装备为代表进行研究,这5种装备型号为001001001001,001001001002,001001001004,001001001006,001001004001,简记为A~E型。

统计这5种装备各新旧状态可用数量,绘制百分比堆积柱形图(图5),发现5种装备堪用率基本相当,其他几种差异都比较大,其中B、E型装备的废品率比其他几种要高许多。因此装备类型对装备状态存在影响。

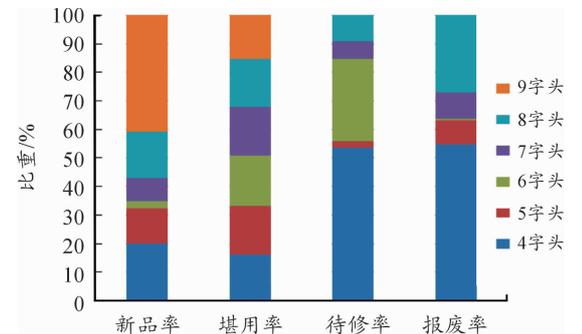


图5 6类部队各新旧装备数量的比重关系

3.2.3 部队类型与装备状态的关联分析

部队层次码的首位数从4~9一共6个整数,代表着部队隶属的最高类别,可按首位数字将部队类型划分为6类。

统计6类部队下各种新旧状态的装备数量,绘制百分比堆积柱形图(图6)发现,6类部队的堪用率基本均衡,4字头部队装备的待修率和废品率比其他部队高出许多。因此部队类型对装备状态存在影响。

3.3 小结

本节对研究对象(装备新旧状态)和影响因素进行了初步的描述性统计,一定程度上反映了总体内在的规律和关联,为后续工作提供了思路。但要给出更具有统计学意义的

结论,则需要利用模型和统计量做进一步的统计推断。

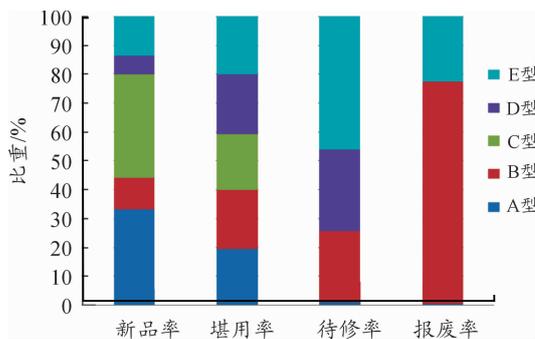


图6 A-E型装备中新旧装备数量的比重关系

4 建立模型及数据分析

4.1 高程因子对装备状态的影响分析

由于高程因子与其他因素之间相对独立,因此采用单因素方差分析判别法。此时将“高程”作为检验对象,不考虑其余因素产生的影响。

4.1.1 单因素方差分析判别模型^[5]

在单因素试验中,首先假定每个总体都服从正态分布,方差相同,观测值相互独立。 A 表示影响因素,具有 k 个水平 A_1, A_2, \dots, A_k ,是因素 A 的具体表现。每个 $x_{ij} (i=1, 2, \dots, k; j=1, 2, \dots, r)$ 表示第 i 个水平下的第 j 个观测值,来自正态总体 $N(\mu_i, \sigma_i^2)$ 。

1) 提出原假设与备择假设

$H_0: \mu_1 = \mu_2 = \dots = \mu_i = \dots = \mu_k$, 自变量对因变量没有显著影响;

著影响;

$H_1: \mu_i (i=1, 2, \dots, k)$ 不完全相等,自变量对因变量有显著影响。

2) 构造检验统计量

定义总误差平方和 $SST = \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2$,水平项误差平方和 $SSA = \sum_{i=1}^k \sum_{j=1}^{n_i} (\bar{x}_i - \bar{x})^2 = \sum_{i=1}^k n_i (\bar{x}_i - \bar{x})^2$,误差项平方和 $SSE = \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$,其中 $SST = SSE + SSA$ 。组间均方 $MSA = \frac{SSA}{k-1}$,组内均方 $MSE = \frac{SSE}{n-k}$,则统计量 $F = \frac{MSA}{MSE} \sim F(k-1, n-k)$ 。

3) 进行F检验

对给定的显著水平 α 计算统计量 F ,判定 $F > F_\alpha(k-1, n-k)$ 是否成立,若成立拒绝 H_0 ,即认为各水平均值间有显著差异,否则接受 H_0 。

4.1.2 建模与求解

通常情况下,新品与废品之间人们往往更关心废品快速报废的原因,因此本文列出以废品数为样本的分析过程,以新品数、堪用品数、待修品数为样本的分析只列出结果供比较。

1) 划分因素水平。根据上节对高程的聚类,将高程划分为3个海拔段,代表 A_1, A_2, A_3 三种水平。

2) 统计每个水平下的样本值。以废品为样本,经筛选低海拔下有40种报废装备,中海拔下有72种,高海拔下有21种,每种装备的当前数量如表4所示。

表4 各海拔下报废品数量统计

| | 1 | 2 | ... | 20 | 21 | 22 | 23 | ... | 39 | 40 | 41 | 42 | ... | 71 | 72 |
|--------------|----|----|-----|----|----|----|----|-----|----|----|----|----|-----|----|----|
| 低海拔(A_1) | 4 | 2 | ... | 4 | 4 | 9 | 3 | ... | 2 | 2 | | | | | |
| 中海拔(A_2) | 10 | 6 | ... | 4 | 3 | 3 | 2 | ... | 26 | 5 | 16 | 2 | ... | 36 | 18 |
| 高海拔(A_3) | 12 | 18 | ... | 4 | 3 | | | | | | | | | | |

3) 计算F统计量,做显著性检验。经计算 $F = 4.492663 > F_\alpha = 3.065839$,故拒绝原假设,表明不同海拔下的废品数量之间有显著差异。

4.1.3 关系强度的度量

在单因素方差分析中, $R = \sqrt{\frac{SSA}{SST}}$ 称为相关系数,用来测量两个变量之间的关系强度^[5-6]。定义高程因子与4种装备新旧状态的相关系数为 $R_{(01)}^{GC}, R_{(02)}^{GC}, R_{(03)}^{GC}, R_{(04)}^{GC}$,由公式即可得到结果。

4.1.4 结果分析

以新品数、堪用品数、待修品数为样本做方差分析的结果如表5所示。

表5 高程对装备状态的影响情况统计

| 高程 | 装备状态 | | | |
|---------------|--------|--------|--------|----------|
| | 新(01) | 堪(02) | 待(03) | 废(04) |
| 有无显著影响 | 没有 | 没有 | 没有 | 有 |
| 相关系数 R^{GC} | 0.0585 | 0.0161 | 0.0699 | 0.254165 |

可以看到,只有废品数受到高程的显著影响,其他3种受高程的影响并不显著;相关系数显示 $R_{(04)}^{GC} > R_{(03)}^{GC} > R_{(01)}^{GC} > R_{(02)}^{GC}$,且 $R_{(04)}^{GC}$ 远大于后三者,表明高程与装备的报废之间有着较强关系。

出现这样的结果主要是因为高海拔地区气象的特殊性。

高海拔地区大气气压低,空气密度小,会导致设备的散热能力下降;高原地区昼夜温差大,使设备的密封结构容易破裂,外壳容易变形破裂;此外,海拔到达5 000 m以上时,太阳的辐射强度比低海拔时高1.25倍,热辐射会对设备起加热作用,降低有机绝缘材料的性能等等,这些原因都会加快设备报废^[7]。建议相关部门查找出设备快速报废的原因,有针对性的改进设备的设计,提高高海拔地区设备的使用寿命^[8]。

4.2 经度、纬度因子对装备状态的影响分析

由于经度、纬度两者联合起来能描述部队的方位,两因素又相互独立,因此采用无交互作用的双因素方差分析法进行数据挖掘更有价值。

4.2.1 无交互作用的双因素方差分析法判定模型^[5]

在该模型中,一个因素在行位置,有 k 个水平;另一个在列位置,有 r 个水平,行列因素的每一个水平都搭配成一组。若每个组合下对应着一个观察值 x_{ij} ($i=1,2,\dots,k$; $j=1,2,\dots,r$),则 x_{ij} 看作是从行因素和列因素组合成的 $k \times r$ 个总体中抽取的样本容量为1的独立随机样本,其中这 $k \times r$ 个总体每一个都服从正态分布,且有相同的方差。

1) 提出原假设及备择假设

对行因素的假设 $H_{0A}: \mu_1 = \mu_2 = \dots = \mu_i = \dots = \mu_k, H_{1A}: \mu_i$ ($i=1,2,\dots,k$)不全相等;

对列因素的假设 $H_{0B}: \mu_1 = \mu_2 = \dots = \mu_j = \dots = \mu_r, H_{1B}: \mu_j$ ($j=1,2,\dots,r$)不全相等。

2) 构造行因素和列因素的统计量

由于需要做两因素检验,因此定义行因素均方 $MSR =$

$$\frac{SSR}{k-1} = \frac{\sum_{i=1}^k \sum_{j=1}^r (\bar{x}_{i.} - \bar{x})^2}{k-1}, \text{列因素均方 } MSC = \frac{SSC}{r-1} =$$

$$\frac{\sum_{j=1}^r \sum_{i=1}^k (\bar{x}_{.j} - \bar{x})^2}{r-1}, \text{随机误差项均方 } MSE =$$

$$\frac{SSE}{(k-1) \times (r-1)}. \text{则构造行统计量 } F_R = \frac{MSR}{MSE} \sim F(k-1,$$

$$(k-1) \cdot (r-1)), \text{列统计量 } F_C = \frac{MSC}{MSE} \sim F(r-1,$$

3) 对给定的显著水平 α 做 F 检验

若 $F_R > F_\alpha$,拒绝原假设 H_0 ,所检验行因素对观测值有显著影响;

若 $F_C > F_\alpha$,拒绝原假设 H_0 ,所检验列因素对观测值有显著影响。

4.2.2 建模与求解

1) 确定因素水平。将中国境内 $73^\circ \sim 135^\circ$ 的经度范围以 5° 为步长划分为12个水平, $3^\circ \sim 53^\circ$ 的纬度范围以 10° 为步长划分为5水平;

2) 统计样本矩阵。以废品为样本,依次筛选出落在第 i ($i=1,2,\dots,5$)个行水平、第 j ($j=1,2,\dots,12$)个列水平下的所有废品,统计它们的可用数总和作为 x_{ij} ,得到样本矩阵

$$X_{5 \times 12} = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 136 & 64 & \dots & 1464 \\ \dots & \dots & \dots & \dots & \dots \\ 122 & 4025 & 4943 & \dots & 2448 \end{pmatrix}_{5 \times 12}$$

例如矩阵中的 x_{32} 表示纬度落在 $23^\circ \sim 33^\circ$ 、经度落在 $78^\circ \sim 83^\circ$ 的所有废品数量总和为136个。

3) 计算 F_R, F_C 两个统计量,分别对行因素、列因素做检验。

4.2.3 关系强度的度量

SSR, SSC 两个指标可以分别度量纬度、经度对因变量产生的影响。然而在双因素判别法中,人们更关心两个因素联合起来的影响效应。定义 $R^2 = \frac{\text{联合效应}}{\text{总效应}} = \frac{SSR + SSC}{SST}$,其平方根 R 就反映了双因子与因变量之间的关系强度,称为多重相关系数^[5]。

4.2.4 结果分析

1) 经度对装备的4种新旧状态都有显著影响,而纬度只对新品有影响。这是因为我国地域呈雄鸡形状,经度跨越范围大,纬度变化范围小,因此经度带来的影响也应更大一些。

2) 4个多重相关系数 R 均大于0.5,说明两因素产生了较强的联合影响效应,验证了将两因素联合考虑的合理性,也表明不论装备是新还是旧,与地理位置均存在很大联系。

表6 经度、纬度对装备状态的影响情况统计

| | 装备状态 | | | |
|------------|-----------|-----------|-----------|-----------|
| | 新(01) | 堪(02) | 待(03) | 废(04) |
| 行因素(纬度) | | | | |
| 有无显著影响 | 有 | 没有 | 没有 | 没有 |
| 列因素(经度) | | | | |
| 有无显著影响 | 有 | 有 | 有 | 有 |
| 多重相关系数 R | 0.767 078 | 0.800 165 | 0.647 306 | 0.631 286 |

4.3 装备类型、部队类型对装备状态的影响程度分析

在日常生活中,根据长期的使用经验,人们对熟悉的产品存在着习惯性使用行为和习惯性购买行为,因此某种部队可能会对某型装备产生特殊“偏好”,这种偏好会对该型装备的磨损产生影响,即这两个因素并不是独立的产生作用,而是会对因变量产生新的附加效应,因此本节采用有交互作用的双因素方差分析法。

4.3.1 有交互作用的双因素方差分析模型^[6]

在该模型中,每个 i, j 组合水平下需要进行 $k(k \geq 2)$ 次重复独立试验才能进行分析。设 x_{ij} 为对应于行因素的第 i 个水平和列因素的第 j 个水平的第 $l(l \leq k)$ 行的观测值; $\bar{x}_{i.}$ 为行因素的第 i 个水平的样本均值; $\bar{x}_{.j}$ 为列因素的第 j 个水平的样本均值; \bar{x}_{ij} 为对应于行因素的第 i 个水平和列因素的第 j 个水平组合的样本均值; \bar{x} 为全部 n 个观察值的总均值。

定义总平方和 $SST = \sum_{i=1}^k \sum_{j=1}^r \sum_{t=1}^m (x_{ij} - \bar{x})^2$, 行变量平方和 $SSR = rm \sum_{i=1}^k (\bar{x}_i - \bar{x})^2$, 列变量平方和 $SSC = km \sum_{j=1}^r (\bar{x}_j - \bar{x})^2$, 交互作用平方和 $SSRC = m \sum_{i=1}^k \sum_{j=1}^r (\bar{x}_{ij} - \bar{x}_i - \bar{x}_j + \bar{x})^2$, 误差项平方和 $SSE = \sum_{i=1}^k \sum_{j=1}^r \sum_{t=1}^m (x_{ijt} - \bar{x}_{ij})^2$. 各项均方依次为 $MSR = \frac{SSR}{k-1}, MSC = \frac{SSC}{r-1}, MSE = \frac{SSE}{rk(t-1)}, MSRC = \frac{SSRC}{(k-1)(r-1)-1}$.

1) 提出原假设与备择假设

① 对行因素提出假设; ② 对列因素提出假设; ③ 对交互作用项提出假设:

$$H_{0AB}: \mu_1 = \mu_2 = \dots = \mu_j = \dots = \mu_r,$$

$$H_{1AB}: \mu_j (j=1, 2, \dots, r) \text{ 不全相等}$$

其中,对行列因素的假设与无交互作用方差分析法相同,这里不再陈述。交互作用项若通过了假设检验,表明两因素的交互作用对样本产生了显著影响。

2) 构造检验统计量

行因素统计量

$$F_A = \frac{MSR}{MSE} \sim F((k-1), rk(t-1))$$

列因素统计量

$$F_B = \frac{MSC}{MSE} \sim F((r-1), rk(t-1))$$

交互作用项的检验统计量

$$F_{A \times B} = \frac{MSRC}{MSE} \sim F((k-1)(r-1)-1, rk(t-1))$$

3) 对给定的显著性水平 α , 分别计算统计量 $F_A, F_B, F_{A \times B}$, 若大于 F_α 拒绝原假设, 小于 F_α 则接受原假设。

4.3.2 建模与求解

1) 确定因素水平。根据前文的分析,将部队类型按首位数划分为6个水平;对于装备层次码,它以3位为一级来反映层次关系,经统计它的位数只有5种类型,因此按位数的长短将装备类型划分为5个水平。

2) 统计样本矩阵。以废品为样本,在每个 i, j 组合水平下 ($i=1, 2, \dots, 5; j=1, 2, \dots, 12$) 都取定 $k=5$, 以等距抽样的方式统计出该组合水平下各种废品数量的总和作为样本值 x_{ij} , 得到样本如表7所示。

表7 不同装备类型、部队类型下废品总数统计

| 废品数量/个 | 部队类型 | | | | | |
|--------|--------|--------|--------|--------|--------|--------|
| | 4 字头部队 | 5 字头部队 | 6 字头部队 | 7 字头部队 | 8 字头部队 | 9 字头部队 |
| 6 位码 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | | | | | |
| 12 位码 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 3949 | 2127 | 311 | 1431 | 5870 | 8354 |
| 15 位码 | | | | | | |
| | 287 | 565 | 190 | 403 | 695 | 589 |
| 18 位码 | | | | | | |
| | 1316 | 3423 | 16 | 619 | 1992 | 1915 |
| 21 位码 | | | | | | |
| | 43 | 13 | 0 | 30 | 47 | 747 |
| 21 位码 | | | | | | |
| | 110 | 15 | 0 | 43 | 23 | 208 |
| 21 位码 | | | | | | |
| | 15 | 0 | 7 | 0 | 32 | 199 |
| 21 位码 | | | | | | |
| | 15 | 0 | 7 | 0 | 45 | 166 |

3) 依模型计算统计量,做假设检验。

4) 计算交互作用相关系数^[5] $R_{交互} = \sqrt{\frac{SSRC}{SST}}$ 。

4.3.3 结果分析

可以看到,第一,部队类型、装备类型对于装备各新旧状态均有着显著影响。这是因为不同数字头的部队职责划分不同,对于装备的使用频率也不尽相同,自然磨损程度会在存在差异。而不同类型的装备在使用中的磨损程度也是不一

样的,比如液晶显示屏显然要比鼠标键盘这类低值易耗品耐用得多。第二,交互作用均通过了显著性检验,且相关系数 $R_{交互}$ 均在0.4左右,说明两个因素对装备状态产生了较明显的附加效应,其中消费习惯可能就是产生该效应的原因之一。事实上,消费习惯现象是客观普遍存在的,在做统计分析时不可忽视这种现象,简单地将各因素视作相互独立加以处理,会导致统计结果与客观事实产生偏差、甚至出现相违背的情况。

表8 装备类型、部队类型对装备状态的
影响情况统计

| | 装备状态 | | | |
|---------------------|-------|-------|-------|-------|
| | 新(01) | 堪(02) | 待(03) | 废(04) |
| 行因素(装备层次) 有无显著影响 | 有 | 有 | 有 | 有 |
| 列因素(部队层次) 有无显著影响 | 有 | 有 | 有 | 有 |
| 交互作用有无 显著影响 | 有 | 有 | 有 | 有 |
| 相关系数 $R_{交互}$ | 0.34 | 0.32 | 0.41 | 0.45 |

5 结论

本文基于已获得的模拟大数据,用不同的方差分析模型分别研究了5种因素对装备新旧状况的影响。建议相关部门若是在各项数据均完备的情况下,先采用主成分分析、因子分析法提取出主要影响因子,再研究它们对装备状态的影响情况。

(责任编辑 杨继森)

参考文献:

- [1] 尹承督,贾红丽,李卫东.大数据在部队信息资源管理中的应用[J].兵器装备工程学报,2016(8):118-121.
 - [2] 代恒,宋华文,王鹏.基于大数据的装备指挥决策问题研究[J].装备学院学报,2016(27):38-43.
 - [3] 李炳元,潘保田,韩嘉福.中国陆地基本地貌类型及其划分指标探讨[J].第四纪研究,2008(4):535-543.
 - [4] 蒋帅.K均值聚类算法研究[D].西安:陕西师范大学,2010.
 - [5] 贾俊平.统计学[M].2版.北京:清华大学出版社,2006.
 - [6] 戴金辉,代金辉.方差分析在跳水运动成绩管理中的应用[J].统计与决策,2016(22):80-82.
 - [7] 周琼方.高压电器设备的高原适应性要求[J].电器工业,2006(5):60-62.
 - [8] 胡连桃,魏东涛,罗广旭.高原环境下航空地面制氧设备的技术选择及改进措施[J].装备环境工程,2012(4):88-91.
- (责任编辑 杨继森)
-
- (上接第111页)
- [9] QIAN Z, WANG G, WANG J. UAV Route Planning Based on Dual Populations Genetic Ant Colony Algorithm [C]//International Conference on Information Engineering for Mechanics and Materials, 2015.
 - [10] DUAN Y, JI X, LI M, et al. Route Planning Method Design for UAV Under Radar ECM Scenario [C]//International Conference on Signal Processing IEEE, 2015:108-114.
 - [11] NIU S, ZHANG J, ZHANG F, et al. A Method of UAVs Route Optimization Based on the Structure of the Highway Network [J]. International Journal of Distributed Sensor Networks, 2015, 2015(1):1667-167.
- (责任编辑 杨继森)
-
- (上接第116页)
- [6] ALLEN R V. Automatic earthquake recognition and timing from single traces[J]. Bulletin of the Seismological Society of America, 1978, 68(5):1521-1532.
 - [7] AKAIKE H. Information Theory and an Extension of the Maximum Likelihood Principle[M]//Breakthroughs in Statistics. Springer New York, 1973:610-624.
 - [8] SARAGIOTIS C D, HADJILEONTIADIS L J, PANAS S M. PAI-S/K: A robust automatic seismic P phase arrival identification scheme [J]. Geoscience & Remote Sensing IEEE Transactions on, 2002, 40(6):1395-1404.
 - [9] BOSCHETTI F. A fractal-based algorithm for detecting first arrivals on seismic traces [J]. Geophysics, 1996, 61(4):1095-1102.
 - [10] 刘劲松,王赞,姚振兴.微地震信号到时自动拾取方法[J].地球物理学报,2013,56(5):1660-1666.
 - [11] 赵大鹏,刘希强,李红,等.峰度和AIC方法在区域地震事件和直达P波初动自动识别方面的应用[J].地震研究,2012,35(2):220-225.
 - [12] 叶根喜,姜福兴,杨淑华.时窗能量特征法拾取微地震波初始到时的可行性研究[J].地球物理学报,2008,51(5):1574-1581.
 - [13] 张唤兰,朱光明,王云宏.基于时窗能量比和AIC的两步法微震初至自动拾取[J].物探与化探,2013,37(2):269-273.
- (责任编辑 杨继森)