

# 基于 SPSS 数据分析的服装号型设计

刘睿智, 赵守香\*

(北京工商大学 计算机与信息工程学院, 北京 100048)

**摘要:**我国幅员辽阔,人口众多,不同地区人口体型不同,服装号型结构存在差异,服装生产时需要根据目标客户体型特点建立合理的号型结构以满足顾客的需求。以陕西地区男子体型为研究对象,利用 SPSS 软件进行大数据分析,结合回归分析、相关性分析、因子分析、聚类分析等方法,快速准确地建立上衣号型结构。该设计方案可以为服装设计和生产提供参考与借鉴。

**关键词:** SPSS 应用;数据分析;服装;号型

**中图分类号:** TS 941.79 **文献标志码:** A **文章编号:** 2096-1928(2019)06-0504-06

## Garment Size Design Based on SPSS Data Analysis

LIU Ruizhi, ZHAO Shouxiang\*

(School of Computer and Information Engineering, Beijing Technology and Business University, Beijing 100048, China)

**Abstract:** China is a vast country with a large population. Different regions have different body shapes and clothing size structures. It is necessary to establish a reasonable size structure to meet the needs of customers according to the body shape characteristics of the target customers. Taking the design of men's shirt size structure in Shaanxi Province as an example, the large data analysis was carried out by SPSS software, and the shirt size structure was established quickly and accurately by combining regression analysis model, correlation analysis, factor analysis, cluster analysis and so on. It is hoped that the design scheme can be used for reference by garment manufacturers.

**Key words:** SPSS application, data analysis, clothing, size

中国是最大的服装生产国、销售国和出口国,世界上每3件服装就有1件来自中国。服装是人们生活中的必需品,关系到其生活质量。服装号型对于服装销售至关重要,服装号型对消费者各种身材和体型覆盖率越大则越有利于产品的销售。虽然我国出台了相关的号型标准,但都是推荐性标准,生产企业在此基础上建立了自己的号型结构体系,并且不同地区、不同年龄的消费者身材体型不尽相同,需要生产企业深入调研设计号型标准。SPSS 软件是一款强大的数据分析软件,利用它可以简明、快速、准确地确定号型分类及设置。文中以陕西地区男式上衣为研究对象,具体讨论如何利用 SPSS 进行大数据分析,精确设计服装号型。

## 1 样本选取与数据采集

样本量的确定是数据统计的基础。采集数据量过大会费时费力,数据量过小容易产生较大误差,科学合理的样本量对统计分析预测至关重要。采用简单随机抽样方法,按下面公式进行计算:

$$N = \frac{Z^2 S^2}{d^2}$$

式中: $N$  为样本容量; $Z$  为置信水平下的统计量; $S$  为总体标准差; $d$  为允许误差。成年人体各控制部位尺寸可接受的误差和标准差见表1<sup>[1]</sup>。95%置信水平下 $Z$ 统计量为1.96,选取最大的 $S/d$ 值6.70,经计算 $N$ 为172。随机抽取了172名成年男子并测量获取到了他们的体型信息。

收稿日期:2019-02-04; 修订日期:2019-06-20。

基金项目:北京市科技计划项目(Z161100001616004)。

作者简介:刘睿智(1995—),男,硕士研究生。

\*通信作者:赵守香(1966—),女,副教授,硕士生导师。主要研究方向为重要产品追溯、客户关系管理、供应链管理。

Email:zhao\_sx4384\_cn@sina.com

表 1 人体各控制部位的数值  
Tab.1 Values of various control parts of the human body

项 目	d/cm	S/cm	S/d
身 高	1.00	6.20	6.20
胸 围	1.50	5.50	3.67
腰 围	1.00	6.70	6.70
臀 围	1.50	5.20	3.47
前腰节长	0.35	2.30	6.57
后腰节长	0.35	2.20	6.29

2 数据预处理

在数据采集过程中难免会出现偏差,造成数据的失真,因此需要对这些数据进行奇异值的检查和处理,剔除一些异常数据以保证结果准确。在进行统计分析时,许多分析方法要求数据符合某种概率分布,大多数方法要求符合正态分布,因此还需要进行正态分布检验。

检查是否存在奇异值主要有两种方法,分别是茎叶图和箱图,具体如图 1 所示。茎叶图可以看出奇异值的数量,箱图可以找到奇异值的位置,文中将两种方法组合运用,以增强数据的准确性。在茎叶图中,显示了整体身高变量的频数、茎和叶<sup>[2]</sup>。茎表示数值的整体部分,叶表示数值的小数部分,“≥186”表示身高变量数据中存在一个奇异值。从箱图中可以看出,107 号是奇异值,该奇异值是由于身高过高导致的,因此将 107 号样本数据剔除。按照该方法对其他变量进行检验,将样本中 4 个奇异值剔除,最终样本数量为 168 个。

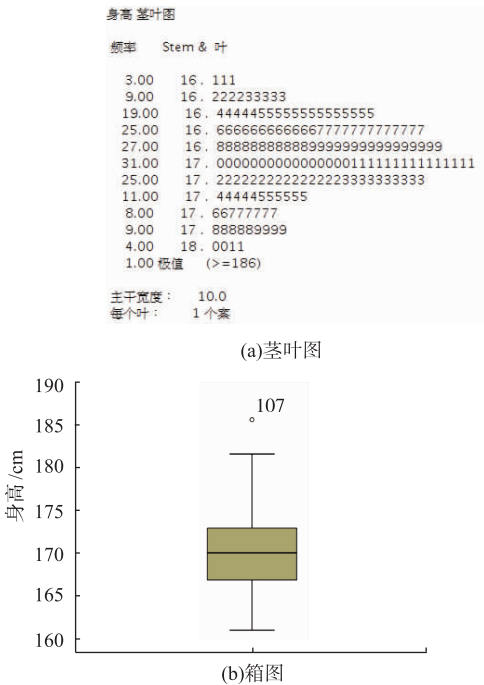


图 1 检查奇异值的两种方法  
Fig.1 Two ways to cheek singular values

正态分布检验常用的方法是 P-P 概率图和 Q-Q 概率图,其中 P-P 概率图简单直观、便于判断,因此文中采用该方法对身高变量进行正态分布检验,检验结果如图 2 所示。由图 2 可以看出,样本数据大致聚集在一条直线上,可以认定身高变量基本符合正态分布,且各点是无规则的即样本是随机的。同理可得,其他变量也基本遵循正态分布。

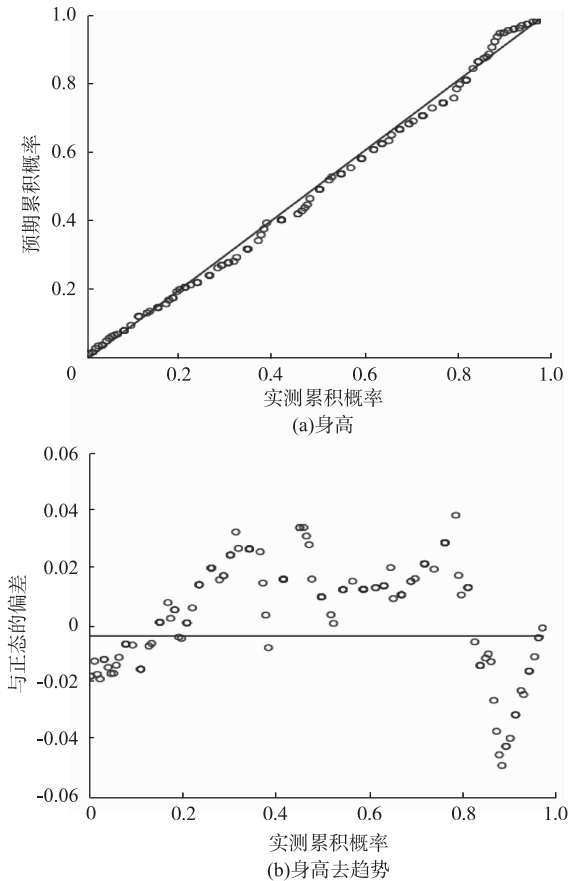


图 2 身高及身高去趋势的正态 P-P 分布  
Fig.2 Height and height trending normal P-P diagram

3 数据的基本分析

3.1 描述性分析

文中选取最大值、最小值、平均值及标准差等具有代表性的指标进行分析,分析结果见表 2。

表 2 变量统计性描述  
Tab.2 Statistical description of variables

项 目	个案数	最小值	最大值	平均值	方 差
身 高	168	161.0	181.0	170.020	19.568
体质量	168	50.2	74.0	62.405	23.215
胸 围	168	79.0	104.2	91.108	25.781
颈 围	168	35.2	44.3	39.635	3.266
总肩宽	168	39.1	53.7	46.317	7.762
后背长	168	38.5	49.7	43.655	5.492
全臂长	168	52.5	60.0	55.598	2.730
有效个案数 (成列)	168				

由表2可以看出,陕西地区成年男子身高、体质量、胸围方差较大,表示个体在这些变量上存在着很大差异,其他变量差异性较小。

### 3.2 相关性分析

相关性分析是研究变量之间关系密切程度常用的方法之一。变量间的相关程度可以用相关系数衡量。相关系数用  $r$  表示,正态分布的等间隔测度变量  $x$  与  $y$  间的相关系数采用 Pearson 积矩相关公式计算:

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

式中: $\bar{x}$  和  $\bar{y}$  为变量  $x$  和  $y$  的平均值; $x_i$  和  $y_i$  为  $x$  和  $y$

的第  $i$  个观测值; $r$  的取值范围在  $-1 \sim 1$  之间。若  $r < 0$  表示变量之间存在负相关, $r > 0$  表示变量之间存在正相关; $|r|$  越接近 1 代表相关程度越高,越接近于 0 相关度越低。 $|r| \geq 0.8$ , 则为高度相关; $0.5 \leq |r| < 0.8$ , 则为中度相关; $0.3 \leq |r| < 0.5$ , 则为低度相关; $|r| < 0.3$ , 则为不相关。

表3为各变量间相关系数矩阵。由表3可以得出:所有变量均呈正相关关系。其中,身高与体质量、全臂长之间存在中度相关或高度相关;胸围与肩宽、后背长、颈围存在中度相关或高度相关;其他变量之间也存在中度相关及低度相关。相关性分析是进行回归分析、建立回归模型的重要依据,相关性越强,线性回归模型效果越好。

表3 变量间相关系数矩阵

Tab.3 Matrix of correlation coefficients between variables

	身高	体质量	胸围	颈围	总肩宽	后背长	全臂长
身高	1.000	0.868	0.539	0.500	0.359	0.477	0.790
体质量	0.868	1.000	0.509	0.504	0.338	0.457	0.717
胸围	0.539	0.509	1.000	0.798	0.861	0.850	0.340
颈围	0.500	0.504	0.798	1.000	0.721	0.793	0.411
总肩宽	0.359	0.338	0.861	0.721	1.000	0.801	0.229
后背长	0.477	0.457	0.850	0.793	0.801	1.000	0.399
全臂长	0.790	0.717	0.340	0.411	0.229	0.399	1.000

### 3.3 因子分析

因子分析的目的是从众多变量中挑选出一个或几个具有代表性的变量<sup>[3]</sup>,因此因子分析的前提条件是变量之间存在较强的相关关系。在因子分析前需要检验变量间是否满足较强相关关系,常用的检验方法有 KMO 检验和 Bartlett 球形度检验。KMO 检验统计量用于比较变量间简单相关系数和偏相关系数的指标,计算公式为

$$KMO = \frac{\sum_i \sum_{j \neq i} r_{ij}^2}{\sum_i \sum_{j \neq i} r_{ij}^2 + \sum_i \sum_{j \neq i} p_{ij}^2}$$

式中: $r_{ij}$  为变量  $x_i$  和其他变量  $x_j$  间的简单相关系数; $p_{ij}$  为变量  $x_i$  和其他变量  $x_j$  间在控制剩余变量下的偏相关系数。KMO 取值范围在  $0 \sim 1$  之间,当所有变量间简单相关系数平方和大于偏相关系数平方和时,KMO 接近于 1,变量间的相关性强,适合进行因子分析。文中运用 KMO 和 Bartlett 球形度检验对样本数据进行测试,具体结果见表4。Bartlett 球形度检验的统计量根据相关系数矩阵的行列式计算得到,若变量观测值较大且对应的  $p$  值小于给定的显著性水平  $\alpha$ ,认为原有变量适合进行因子分析。由表4可以看出,KMO 值为 0.832,较接近于 1,说明适合进行因子分析;同时,Bartlett 球形度检验中

显著性接近于 0,显著性水平  $\alpha$  为 0.05,0 小于显著性水平  $\alpha$ ,适合进行因子分析。因此,两种检验方法均证明变量间适合进行因子分析。

表4 KMO 和 Bartlett 检验结果

Tab.4 KMO and Bartlett test results

KMO 检验	Bartlett 球形度检验		
	近似卡方	自由度	显著性
0.832	1 121.998	21	0.000

总方差解释见表5。表5中初始特征值列反映了因子分析初始解的情况,第1个因子的特征值为 4.531,解释原有 7 个变量总方差的 64.732%,累计方差贡献率 64.732%;第2个因子的特征值为 1.509,解释原有 7 个变量总方差的 21.560%,累计方差贡献率 86.283%,其余数据含义类似,初始解中提取了 7 个因子,原有变量总方差均被解释,累计方差贡献率 100%。提取载荷平方和列描述了因子解的情况,由于指定提取 2 个因子,它们共解释了原有变量总方差的 86.283%,整体上,原有变量信息丢失较少,因子分析效果较理想;旋转载荷平方和列描述了最终因子解情况,因子旋转后累计方差贡献率没有发生变化,但重新分配了各个因子解释原有变量的方差,使得因子更易于解释。

表 5 总方差解释  
Tab.5 Total variance interpretation

成 分	初始特征值			提取载荷平方和			旋转载荷平方和		
	总 计	方差百分比	累积/%	总 计	方差百分比	累积/%	总 计	方差百分比	累积/%
1	4.531	64.723	64.723	4.531	64.723	64.723	3.363	48.040	48.040
2	1.509	21.560	86.283	1.509	21.560	86.283	2.677	38.243	86.283
3	0.315	4.501	90.785						
4	0.261	3.733	94.517						
5	0.168	2.399	96.916						
6	0.129	1.837	98.753						
7	0.087	1.247	100.000						

注:提取方法为主成分分析法。

图 3 为因子分析碎石图。

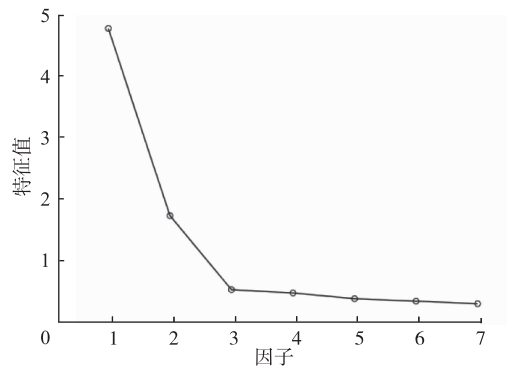


图 3 因子分析碎石图  
Fig.3 Gravel map

由图 3 可以看出,第 1 个因子的特征值很高,对解释原有变量的贡献最大;第 3 个因子以后的特征值都较小,对解释原有变量的贡献很小,因此提取两个因子是合适的。

旋转后的成分矩阵见表 6。由表 6 可以看出,第 1 个因子主要解释了总肩宽、胸围、后背长和颈围这几个变量,身高、全臂长和体质量主要由第 2 个因子解释。

4 聚类分析

聚类分析是将样本数据按照一定的规则进行分类,同类样本具有较大的相似度,不同类样本之间存在着较大差异。K-means 均值聚类方法具有思想简单、易于理解、容易实现、处理速度快、占用内存少等优点,适合大样本聚类,分类效果较好<sup>[4]</sup>,因此文中采用此方法对体型数据进行聚类分析。聚类分析的核心步骤是:确定聚类数目;确定初始类

中心点;确定样本点到中心点的欧氏距离,完成一次迭代过程;重新确定中心点并重复以上步骤,直至迭代终止。

表 6 旋转后的成分矩阵  
Tab.6 Component matrix after rotation

项 目	成 分	
	1	2
总肩宽	0.930	0.089
胸 围	0.913	0.277
后背长	0.891	0.268
颈 围	0.833	0.323
身 高	0.286	0.910
全臂长	0.137	0.897
体质量	0.275	0.885

注:提取方法为主成分分析法;旋转方法为凯撒正态化最大方差法;旋转在 3 次迭代后已收敛。

4.1 聚类变量的确定

通过体型的特征指标确定聚类变量,特征指标包括身体的不同部位,即身高、体质量、胸围等<sup>[5]</sup>。通过因子分析可知,在两个因子中都存在着具有代表性的变量,通过计算因子中相关指数,找出相关指数最大的变量作为控制变量即聚类变量,相关指数计算公式为

$$R_i^2 = \frac{\sum r_{ij}^2}{m - 1}$$

式中: $i = 1, 2, \dots, m$ ,  $m$  为所在类的指标个数; $r_{ij}$  为相关系数,  $i \neq j, j = 1, 2, \dots, m$ 。计算结果见表 7。

表 7 因子相关指数  
Tab.7 Related index

因 子	因子 1				因子 2		
	总肩宽	胸 围	后背长	颈 围	身 高	全臂长	体质量
相关指数	0.634	0.700	0.669	0.595	0.689	0.569	0.634

4.2 聚类结果分析

将身高和胸围作为聚类变量,按照 K-means 均值聚类方法对样本数据进行分类。考虑到服装生产的实际,服装号型设置不宜过多,因此 3~5 类最为合理。经过比较分析,发现聚类数为 3 时分类结果最清晰,效果最好,最终聚类中心见表 8。由表 8 可以看出,不同类别个案数及所占比例中间多、两边较少,表明聚类结果合理。胸围的聚类中心大致为 86,93,99,身高为 166,171,177,参考国家号型标准同时兼顾分析结果,将相邻体型之间的胸围差设置为 4,身高差设置为 5,结果清晰均匀。

表 8 最终聚类中心

Tab. 8 Final cluster center

项 目	聚 类		
	1	2	3
胸 围	85.9	92.6	98.8
身 高	166.4	170.6	177.1
个案数	57	89	22
比 例	34	53	13

4.3 中间体数值的确定

将不同类别样本数据分离出来,分别计算不同变量的平均值,作为中间体数值,具体结果见表 9。中间体具有一定的代表性,反映了不同分类样本的体型特点:M 表示体型样本身高较低,胸围较小;N 表示体型样本频率最大,表明此体型人数最多,体型中等;P 为体型样本身高较高,身材健壮<sup>[6]</sup>。为便于表示体型,将表 8 中 1 体型用 M 表示,2 体型用 N 表示,3 体型用 P 表示。如果按照表 9 中的数值进行服装号型的推算,必然会带来许多不便,因此需要对其进行圆整,圆整后结果见表 10。

表 11 线性回归方程

Tab. 11 Linear regression equation

项 目	M	N	P
体质量	$0.766H - 68.444$	$0.990H - 106.081$	$1.002H - 107.800$
颈 围	$0.292B + 13.030$	$0.227B + 19.104$	$0.423B - 0.183$
总肩宽	$0.464B + 4.087$	$0.635B - 11.741$	$0.455B + 4.625$
后背长	$0.354B + 11.065$	$0.394B + 7.864$	$0.479B - 0.960$
全臂长	$0.329H - 0.187$	$0.366H - 6.634$	$0.412H - 15.710$

首先设置身高和胸围的档差分别是 5 和 4,将其代入不同部位的线性回归方程可以得到相应的档差。为便于生产,将 3 类体型不同档差进行统一化,得到了最终确定的档差,具体结果见表 12。由

表 9 平均值

Tab. 9 Average value

项 目	M	N	P
身 高	166.417	170.587	177.064
体质量	59.063	62.773	69.573
胸 围	85.895	92.552	98.773
颈 围	38.074	40.153	41.586
总肩宽	43.970	47.013	49.586
后背长	41.505	44.368	46.341
全臂长	54.602	55.845	57.177

表 10 圆整后平均值

Tab. 10 Rounded average

项 目	M	N	P
身 高	166	171	177
体质量	59	63	69
胸 围	86	93	99
颈 围	38	40	42
总肩宽	44	47	50
后背长	42	44	46
全臂长	55	56	57

4.4 各部位档差的设置

档差的设置是为了便于号型的制定。档差设置时,不仅要考虑顾客的舒适性,还要便于企业生产,档差设置太大不能满足众多的消费者需求,档差设置太小则不利于生产者批量化生产。文中利用回归方程设置各部位档差。胸围在第 1 个因子中具有代表性,因此其他变量用胸围来描述,同理第 2 个因子用身高表示。在不同体型样本中,分别建立身高与体质量、全臂长的线性关系,胸围与颈围、总肩宽、后背长之间的线性关系<sup>[7]</sup>。身高用  $H$  表示,胸围用  $B$  表示,线性关系见表 11。

表 12 可以看出,颈围和总肩宽档差大于国家标准中规定的数值(颈围档差为 1,总肩宽为 1.2),这充分体现了陕西男子的体型特征,身材中等,肩宽颈粗的特点,这与陕西男子的体型特征相一致。

表 12 档差设置  
Tab. 12 Gear setting

项 目	M	N	P
体质量	3.830(5)	4.950(5)	5.010(5)
颈 围	1.168(2)	0.908(2)	1.692(2)
总肩宽	1.856(3)	2.540(3)	1.820(3)
后背长	1.416(2)	1.576(2)	1.916(2)
全臂长	1.645(2)	1.830(2)	2.060(2)

注:括号外为计算值,括号内为采用值。

#### 4.5 确定不同号型具体数值

在计算不同号型具体数值时,首先要统计分析不同体型各部位的最大值和最小值,然后结合中间体、档差、最大值、最小值和线性方程综合考虑数值的设置。以体型 N 为例,首先统计各部位最大值和最小值,具体见表 13;再按照中间体及档差设置数值;最后根据最值和线性方程进行调整修正,最终号型划分结果见表 14<sup>[8]</sup>。同理,按照此步骤对 M 和 P 体型部位数值进行设置。

表 13 N 体型变量统计描述  
Tab. 13 Statistical description of N body

项 目	个案数	最小值	最大值
身 高	89	163.3	179.0
体质量	89	56.0	74.0
胸 围	89	86.2	98.5
颈 围	89	37.2	43.6
总肩宽	89	42.8	52.7
后背长	89	40.8	48.3
全臂长	89	53.0	60.0

表 14 N 体型下的号型设置  
Tab. 14 Number setting under N body type

项 目	165/60	170/65	175/70	180/75
身 高	165	170	175	180
体质量	60	65	70	75
胸 围	86	90	94	98
颈 围	38	40	42	44
总肩宽	43	46	49	52
后背长	42	44	46	48
全臂长	54	56	58	60

## 5 结 语

号型设置是否合理对于服装生产和销售至关重要。利用 SPSS 数据分析软件并根据随机样本数据信息可以快速、准确地构建服装号型体系,方便生产企业根据不同地域、不同年龄的目标客户建立合理的号型标准<sup>[9]</sup>。利用陕西地区男子身材体型数据信息,结合 SPSS 数据分析工具,划分了 N、M、P 3 种体型,并

在每种体型下设置了相关号型及不同号型的数据特征,为服装设计及生产提供参考与借鉴。

#### 参考文献:

- [1] 王俊峰. 数据挖掘技术在服装号型投产决策的应用研究[D]. 西安:西安工程大学,2015.
- [2] 何天虹,刘玉娜,赵星. 天津市女大学生体型分类及其原型修正[J]. 毛纺科技,2018,46(7):53-57.  
HE Tianhong,LIU Yuna,ZHAO Xing. Body classification and prototype amendment of female college students in Tianjin [J]. Wool Textile Journal,2018,46(7):53-57. (in Chinese)
- [3] 朱碧空,李月. 女装号型分类方法对比研究与优化[J]. 毛纺科技,2018,46(4):41-45.  
ZHU Bikong,LI Yue. Contrastive study and optimization of the female size standarads [J]. Wool Textile Journal,2018,46(4):41-45. (in Chinese)
- [4] 申亚楠,潘波. 北京市初中生的服装号型初探[J]. 北京服装学院学报(自然科学版),2018,38(1):1-10.  
SHEN Yanan,PAN Bo. Study on size designation of school uniform for junior high school students in Beijing area [J] Journal of Beijing Institute of Clothing Technology (Natural Science Edition),2018,38(1):1-10. (in Chinese)
- [5] 袁惠芬,王旭,齐雪良,等. 基于神经网络的批量定制服装号型分类研究[J]. 武汉纺织大学学报,2018,31(3):41-45.  
YUAN Huifen,WANG Xu,QI Xueliang,et al. Investigation on mass customization clothing shape classification by artificial neural network [J]. Journal of Wuhan Textile University,2018,31(3):41-45. (in Chinese)
- [6] 刘飞,王云仪,王竹君. 男裤号型表示方法的优化[J]. 服装学报,2016,1(05):461-465.  
LIU Fei,WANG Yunyi,WANG Zhujun. Optimization of size representation method of men's trousers [J]. Journal of Clothing Research,2016,1(5):461-465. (in Chinese)
- [7] 齐静,李毅,张欣. 我国西部地区青年男性体型描述与体型分类研究[J]. 纺织学报,2010,31(5):107-111.  
QI Jing,LI Yi,ZHANG Xin. Description and classification of young men's body type in the west China [J]. Journal of Textile Research,2010,31(5):107-111. (in Chinese)
- [8] 韩雨,潘波. 北京市初中生的体型分类及号型研究[J]. 北京服装学院学报(自然科学版),2015,35(3):40-46.  
HAN Yu,PAN Bo. Study on body shape classification and size of middle school students in Beijing [J]. Beijing Institute of Clothing Technology (Natural Science Edition),2015,35(3):40-46. (in Chinese)
- [9] 李艳梅. 改善我国服装号型标准的对策研究[J]. 上海纺织科技,2006,34(6):70-72.  
LI Yanmei. Study on improving the standards of apparel type and size in our country [J]. Shanghai Textile Science and Technology,2006,34(6):70-72. (in Chinese)

(责任编辑:邢宝妹)