

Anthropometry: An R Package for Analysis of Anthropometric Data

Guillermo Vinué

Department of Statistics and O.R., University of Valencia, Valencia, Spain.

Abstract

The development of new powerful 3D scanning techniques has enabled the generation of broad updated anthropometric databases which constitute high valued data to improve the ergonomic design of products adapted to the user population. Accordingly, Ergonomics and Anthropometry are two fields more and more quantitative, so advanced statistical methodologies and modern software tools are demanded to get full benefit from anthropometric data.

This paper presents a new R package, called **Anthropometry**, that brings together some statistical methodologies concerning clustering, statistical shape analysis, statistical archetypal analysis and the statistical concept of data depth, especially developed to deal with anthropometric data. They are proposed with the goal of providing effective solutions to some common anthropometric problems, such as clothing design or workstations design (focusing on the particular case of aircraft cockpits). The utility of the package is shown by analyzing the anthropometric data obtained from a survey of the Spanish female population performed in 2006 and from the 1967 United States Air Force Survey.

Keywords: R, anthropometric data, clustering, statistical shape analysis, archetypal analysis, data depth.

1. Introduction

Ergonomics is the science that investigates the interactions between human beings and the elements of a system. The application of ergonomic knowledge in multiple areas such as clothing and footwear design or both working and household environments is required to achieve the best possible match between the product and its users. To that end, it is fundamental to know the anthropometric dimensions of the targeted population. Anthropometry refers to the study of the measurements and dimensions of the human body and it is considered a very important branch of Ergonomics because its significant influence on the ergonomic design of products ([Pheasant 2003](#)).

A major issue when developing new patterns and products that fit well the target population is the lack of up-to-date anthropometric data. Improvements in health care, nutrition and living conditions and the transition to a sedentary life style have changed the body dimensions of people over recent decades. The anthropometric databases must therefore be updated regularly. Traditionally, human physical characteristics and measurements have been manually taken by using rudimentary methods like calipers, rulers or measuring tapes ([Simmons and Istook 2003](#); [Lu and Wang 2008](#); [Shu, Wuhrer, and Xi 2011](#)). These procedures are simple

(user-friendly), non-invasive and no particularly expensive. However, measuring by hand a statistically useful sampling of thousands of people is time-consuming and error-prone: the set of measurements obtained and therefore the shape information is usually imprecise and inaccurate.

In recent years, the development of new three-dimensional (3D) body scanner measurement systems has represented a huge step forward in the way of collecting and updating anthropometric data. This technology provides highly detailed, accurate and reproducible anthropometric data from which 3D shape images of the people being measured can be obtained (Istook and Hwang 2001; Lerch, MacGillivray, and Domina 2007; Wang, Wu, Lin, Yang, and Lu 2007; D'Apuzzo 2009). The great potential of 3D body scanning techniques constitute a true breakthrough in realistically characterizing people and they have made it possible to conduct new anthropometric large-scale size surveys in different countries (for instance, in the USA, the UK, France, Germany and Australia). In this context, the Spanish Ministry of Health sponsored a 3D anthropometric study of the Spanish female population in 2006 (Alemany, González, Náchter, Soriano, Arnáiz, and Heras 2010). A sample of 10415 Spanish females from 12 to 70 years old, randomly selected from the official Postcode Address File, was measured. Associated software provided by the scanner manufacturers made a triangulation deriving the 3D spatial location of a large number of points on the body surface. A 3D binary image of the trunk of each woman (white pixel if it belongs to the body, otherwise black) is produced from the collection of points located on the surface of each woman scanned, as explained in Ibáñez, Simó, Domingo, Durá, Ayala, Alemany, Vinué, and Solves (2012a). The two main goals of this study, which was conducted by the Biomechanics Institute of Valencia, were as follows: firstly, to characterize the morphology of females in Spain in order to develop a standard sizing system for the garment industry and secondly, to encourage an image of healthy beauty in society by means of mannequins that are representative of the population. In order to tackle both objectives, Statistics plays an essential role.

In every methodological and practical anthropometric problem, body size variability within the user population is characterized by means of a limited number of anthropometric cases. This is what is called a *user-centered design process*. An anthropometric case represents the set of body measurements the product evaluator plan to accommodate in design (HFES 300 Committee 2004). A case may be a particular human being or a combination of measurements. Depending on the features and needs of the product being designed, three types of cases can be distinguished: central, boundary and distributed. If the product being designed is a one-size product (one-size to accommodate people within a predetermined portion of the population) such as working environments design, the cases are selected on an accommodation boundary. However, if we focus on a multiple-size product (n sizes to fit n groups of people within a predetermined portion of the population), being clothing design the most apparent example, central cases are selected. The statistical methodologies that we have developed seek to define central and boundary cases to tackle the clothing sizing system design problem and the workplaces design problem (focusing on the particular case of an aircraft cockpit).

Clothing sizing systems divide a population into homogeneous subgroups based on some key anthropometric dimensions (size groups), in such a way that all individuals in a size group can wear the same garment (Ashdown 2007; Chung, Lin, and Wang 2007). An efficient and optimal sizing system must accommodate as large a percentage of the population as possible, in as few sizes as possible, that better describes the shape variability of the population. In addition, the garment fit for accommodated individuals must be as good as possible. Each

clothing size is defined from a person who is near the center for the dimensions considered in the analysis. This central individual, which is considered as the size representative (the size prototype), becomes the basic pattern from which the clothing line in the same size is designed. Once a particular garment has been designed, fashion designers and clothing manufacturers hire fit models to test and assess the size specifications of their clothing before the production phase. Fit models have the appropriate body dimensions selected by each company to define the proportional relationships needed to achieve the fit the company has determined (Ashdown 2005; Workman and Lentz 2000; Workman 1991). Fit models are usually people with central measurements in each body dimension. The definition of an efficient sizing system depend to a large extent on the accuracy and representativeness of the fit models.

Clustering is the statistical tool that classifies a set of individuals in groups (clusters), in such a way that subjects in the same cluster are more similar (in some sense) to each other than to those in other clusters (Kaufman, L. and Rousseeuw, P. J. 1990). In addition, clusters are represented by means of a representative central observation. Therefore, clustering comes up naturally as a useful statistical approach to try to define an efficient sizing system and to elicit prototypes and fit models. Specifically, five of the methodologies that we have developed are based on different clustering methods. Four of them are aimed at segmenting the population into optimal size groups. The first one, hereafter referred to as *trimowa*, has been published in Ibáñez, Vinué, Alemany, Simó, Epifanio, Domingo, and Ayala (2012b). It is based on using a especial distance function that mathematically captures the idea of garment fit. The second and third ones (called *CCbiclustAnthropo* and *TDDclust* belong to a paper in progress (Vinué and Ibáñez 2014). The *CCbiclustAnthropo* methodology adapts to the field of Anthropometry a particular clustering algorithm proposed for the analysis of gene expression data. Besides, *TDDclust* uses the statistical concept of data depth (Liu, Parelius, and Singh 1999) to group observations according to the most central (deep) one in each cluster. As mentioned, traditional sizing systems are based on using a suitable set of key body dimensions, so clustering must be carried out in the Euclidean space. In the three previous procedures, we have always worked in this way. Instead, in the fourth and last one, hereinafter called as *kmeansProcrustes*, a clustering procedure is developed for grouping women according to their body shape, represented by a configuration matrix of anatomical markers (landmarks). To that end, the statistical shape analysis (Dryden and Mardia 1998) will be fundamental. This contribution has been submitted for publication (Vinué, Simó, and Alemany 2014c). The fifth clustering proposal is presented with the goal of identifying accurate fit models. It is based on another clustering method originally developed for biological data analysis. This method, called *hipamAnthropom*, has been published in Vinué, León, Alemany, and Ayala (2014b). From well-defined fit models and prototypes, representative and precise mannequins of the population can be made.

A sizing system is intended only to cover what is called “standard” population, leaving out the individuals who might be considered outliers respect to a set of measurements. In this case, outliers are called disaccommodated individuals. Clothing industries usually design garments for the standard sizes in order to optimize market share. The four aforementioned methods concerned with apparel sizing system design (*trimowa*, *CCbiclustAnthropo*, *TDDclust* and *kmeansProcrustes*) take into account this fact. In addition, because *hipamAnthropom* is based on hierarchical features, it is able to discover and return true outliers.

Unlike clothing design, where representative cases correspond to central individuals, in designing a one-size product, such as working environments or the passenger compartment of

any vehicle including aircraft cockpits, the most common approach is to search for boundary cases. In these situations, the variability of human shape is described by extreme individuals, which are those that have the smallest or largest values (or extreme combinations) in the dimensions considered in the study. These design problems fall into a more general category: the accommodation problem. The supposition is that the accommodation of boundaries will facilitate the accommodation of interior points (with less-extreme dimensions) (Bertilsson, Högberg, and Hanson 2012; Parkinson, Reed, Kokkolaras, and Papalambros 2006; HFES 300 Committee 2004). For instance, a garage entrance must be designed for a maximum case, while for reaching things such as a brake pedal, the individual minimum must be obtained. In order to tackle the accommodation problem, two methodological contributions based on the statistical archetypal analysis are put forward. An archetype in Statistics is an extreme observation that is obtained as a convex combination of other subjects of the sample (Cutler and Breiman 1994). The first of these methodologies was published in Epifanio, Vinué, and Alemany (2013), whereas the second one has been submitted for publication (Vinué, Epifanio, and Alemany 2014a).

As far as we know, there is currently no reference in the literature related to Anthropometry or Ergonomics that provides the programming of the algorithms proposed. In addition, to the best of our knowledge, with the exception of modern human modelling tools like Jack and Ramsis, which are two of the most widely used tools by a broad range of industries (Blanchonette 2010), there are neither general software applications nor statistical packages available from the Internet, to tackle the definition of an efficient sizing system or the accommodation problem. In this context, a new R package (R Development Core Team 2013) called **Anthropometry** is introduced in this paper, which brings together all the abovementioned methodologies. All of them were applied to the anthropometric study of the Spanish female population and to the 1967 United States Air Force (USAF) survey. **Anthropometry** includes several data files related to both anthropometric databases. All the statistical methodologies, anthropometric databases and this R package were announced in the author's PhD thesis (Vinué 2014), which will soon be freely available on two Spanish institutional open archives.

The outline of the paper is as follows: Section 2 describes all the data files included in **Anthropometry**. Section 3 gives a brief exposition of each statistical technique suggested and Section 4 presents how they are implemented in this package. In Section 5, some examples of their application are shown. Finally, concluding remarks are in Section 6.

2. Data

2.1. Spanish anthropometric survey

The Spanish National Institute of Consumer Affairs (INC according to its Spanish acronym) of the Spanish Ministry of Health and Consumer Affairs commissioned a 3D anthropometric study of the Spanish female population in 2006, after signing a commitment with the main Spanish companies in the apparel industry. The Spanish National Research Council (CSIC in Spanish) planned and developed the design of experiments, the Complutense University of Madrid was responsible for providing advice on Anthropometry and the study itself was conducted by the Biomechanics Institute of Valencia (Alemany *et al.* 2010). The target sample was made up of 10415 women grouped into 10 age groups ranging from 12 to 70 years,

randomly chosen from the official Postcode Address File.

As illustrative data of the whole Spanish survey, **Anthropometry** contains a database called `dataDemo`, made up of a sample of 600 Spanish women and their measurements for five anthropometric variables: bust, chest, waist and hip circumferences and neck to ground length. These variables are chosen because three main reasons: recommendations of experts, they are commonly used in the literature and they appear in the European standard to sizing system. Size designation of clothes. Part 2: Primary and secondary dimensions (European Committee for Standardization 2002).

This data set will be used by *trimowa*, *TDDclust* and *hipamAnthropom*. As said, the women shape is represented by a set of landmarks, specifically 66 points. A data file called `landmarks` contains the configuration matrix of landmarks for each of the 600 women. The *kmeansProcrustes* methodology will need this data file.

As also noted above, a 3D binary image of each woman's trunk is available. Hence, the dissimilarity between trunk forms can be computed and a distance matrix between women can be built. The distance matrix used in Vinué *et al.* (2014a) is included in **Anthropometry** and it is called `cMDSwomen`.

2.2. USAF survey

This database saves the information provided by the 1967 United States Air Force (USAF) Survey. It can be downloaded from <http://www.dtic.mil/dtic/>. This survey was conducted during 1967 by the Anthropology Branch of the Aerospace Medical Research Laboratory (Ohio). A sample of 2420 subjects of the Air Force personnel, between 21 and 50 years of age, was measured at 17 Air Force bases across the United States of America. A total of 202 variables were collected. The dataset associated with the USAF survey is available in `dataUSAF`. In the methodologies related to the archetypal analysis, six anthropometric variables from the total of 202 will be selected. They are the same selected in Zehner, Meindl, and Hudson (1993) and are called cockpit dimensions because they are critical in order to design an aircraft cockpit.

2.3. Geometric figures

In the *kmeansProcrustes* approach, a numerical simulation with controlled data is performed to show the utility of our methodology. The controlled data are two geometric figures, a cube and a parallelepiped, made up of 8 and 34 landmarks. These configurations are saved in four files called `cube8`, `cube34`, `parallelepiped8` and `parallelepiped34`, respectively.

3. Statistical methodologies

In Section 3.1, the *trimowa*, *CCbiclustAnthropo*, *TDDclust* and *hipamAnthropom* are described. Section 3.2 focuses on the *kmeansProcrustes* methodology. Section 3.3 provides an exposition of the methodologies based on the archetypal analysis.

As a practical guidance, the way of proceeding of the clustering-based approaches has been the following: the data matrix was segmented using a primary control dimension (bust circumference in the case of *trimowa*, *hipamAnthropom*, *kmeansProcrustes* and *TDDclust*, and waist circumference in the case of *CCbiclustAnthropo*, according to the classes suggested in

the European standard to sizing system. Size designation of clothes. Part 3: Measurements and intervals (European Committee for Standardization 2005)). Then, a further segmentation using other secondary control anthropometric variables is carried out. In this way, the first segmentation provides a first easy input to choose the size, while the resulting clusters (subgroups) for each bust (or waist) and other anthropometric measurements optimize sizing. Regarding the methodologies using the archetypal analysis, the steps are as follows: first, depending on the problem, to standardize or not the data. Then, selecting an accommodation subsample for obtaining the archetypal individuals as the third and last step.

3.1. Anthropometric dimensions based clustering

Trimowa

The aim of a sizing system is to divide a varied population into groups using some key body dimensions (Ashdown 2007; Chung *et al.* 2007). Three types of approaches can be distinguished for creating a sizing system: traditional step-wise sizing, multivariate methods and optimization methods. Traditional methods are not useful because they use bivariate distributions to define a sizing chart and they do not consider the variability of other relevant anthropometric dimensions. Recently, more sophisticated statistical methods have been developed, especially using Principal Component Analysis (PCA) and clustering (Gupta and Gangadhar 2004; Hsu 2009b; Luximon, Zhang, Luximon, and Xiao 2012; Hsu 2009a; Chung *et al.* 2007; Zheng, Yu, and Fan 2007; Bagherzadeh, Latifi, and Faramarzi 2010). Peter Tryfos was the first one that suggested an optimization method (Tryfos 1986). Later on, McCulloch *et al.* (McCulloch, Paal, and Ashdown 1998) modified the Tryfos' approach.

The first clustering methodology proposed, called *trimowa*, is closed to that one developed in McCulloch *et al.* (1998). However, there are two main differences. First, when searching for k prototypes, a more statistical approach is assumed. To be quite specific, a trimmed version of the Partitioning Around Medoids (or k -medoids) clustering algorithm is used. The trimming procedure allows us to remove outlier observations (García-Escudero, Gordaliza, Matrán, and Mayo-Iscar 2008; García-Escudero, Gordaliza, and Matrán 2003). Second, the dissimilarity measure defined in McCulloch *et al.* (1998) is modified by using an OWA (Ordered Weighted Average) operator to consider the user morphology. Our approach allows us the derivation of more realistic prototypes (medoids) because they correspond to real women of the database and the selection of individual discommodities. In addition, the use of OWA operators has resulted in a more realistic dissimilarity measure between individuals and prototypes. This approach was published in Ibáñez *et al.* (2012b) and it is implemented in the `trimowa` function.

We learned from this situation that there is an ongoing search for advanced statistical approaches that can deliver practical solution solutions to the definition of central people and optimal size groups. Consequently, we have come across in the literature two different statistical strategies and we have aimed to discuss their potential usefulness in the definition of an efficient clothing sizing system. These approaches are based on Biclustering and Data Depth and will be summarized in the next two sections.

CCbiclustAnthropo

In the analysis of gene expression data, conventional clustering are limited to find local expression patterns. Gene data are organized in a data matrix where rows correspond to genes

and columns to experimental samples (conditions). The goal is to find submatrices, i.e., subgroups of genes and subgroups of conditions, where the genes exhibit high correlation for every condition (Madeira and Oliveira 2004). Biclustering is a novel clustering approach developed with this goal in mind. This technique consists in simultaneous partitioning of the set of rows and the set of columns into subsets.

In a traditional row cluster, each row is defined using all the columns of the data matrix. Something analogue would occur with a column cluster. However, regarding biclustering, each row in a bicluster is defined using only a subset of columns and vice versa. Therefore, clustering derives a global model but biclustering defines a local one. This interesting property made us think that maybe biclustering could be useful to derive efficient size groups, since they would be only defined for the most relevant anthropometric dimensions that describe a body in the detail necessary to design a garment that fit well.

Recently, a large number of biclustering methods have been developed. Some of them are implemented in different sources, including R. Up to now, the most complete R package for biclustering is **biclust** (Kaiser and Leisch 2008; Kaiser, Santamaria, Khamiakova, Sill, Theron, Quintales, and Leisch 2011). The utility of the approaches included in **biclust** to deal with anthropometric data was investigated in Vinué (2012). Among the conclusions reached, the most important was concerned with the possibility to consider the Cheng & Church biclustering algorithm (Cheng and Church 2000) (from now on, CC) as a potential statistical approach to be used for defining size groups. Specifically, in Vinué (2012) an algorithm to find size groups (biclusters) and disaccommodated women with CC was set out. This methodology is called *CCbiclustAnthropo* and it is implemented in the `CCbiclustAnthropo` function.

Designing lower body garments depends not only on the waist circumference (the principal dimension in this case), but also on other secondary control dimensions (for upper body garments the bust circumference is usually needed only). Biclustering produces subgroups of objects that are similar in one subgroup of variables and different in the remaining variables. Therefore, it seems more interesting to use a biclustering algorithm with a set of lower body dimensions. For that purpose, all the body variables related to the lower body part included in the Spanish anthropometric survey were chosen (there were 36). As promising results, an efficient partition into different biclusters was obtained. All individuals in the same bicluster can wear a garment designed for the specific body dimensions (waist and other variables) which were the most relevant for defining the group. Each group is represented by the median woman. The CC algorithm is nonexhaustive, i.e, some rows (and columns) do not belong to any bicluster. This property can be used to fix a proportion of no accommodated sample.

The major interest of this approach was descriptive and exploratory and the important point to note here is that `CCbiclustAnthropo` cannot be used with `dataDemo`, since this data file does not contain variables related to the lower body part in addition to waist and hip. However, this function is included in the package in the hope that it could be helpful or useful for other researchers. All theoretical and practical details are given in Vinué (2014) and Vinué (2012).

TDDclust

The statistical concept of data depth is other general framework for descriptive and inferential analysis of numerical data in a certain number of dimensions. In essence, the notion of data depth is a generalization of the median to data in higher-dimensions. A depth function

measures the degree of centrality of a point regarding a probability distribution or a data set. Highest depth values correspond to central points and lowest depth values correspond to tail points (Liu *et al.* 1999; Zuo and Serfling 2000). Therefore, the depth paradigm is other very interesting strategy to identify central prototypes.

The development of clustering and classification methods using data depth measures has received increasing attention over the last few years (Dutta and Ghosh 2012; Lange, Mosler, and Mozharovskiy 2012; López and Romo 2010; Ding, Dang, Peng, and Wilkins 2007). The most relevant contribution to this field has been made by Rebecka Jörnsten in Jörnsten (2004) (see also Jörnsten, Vardi, and Zhang (2002) and Pan, Jörnsten, and Hart (2004)). She introduced two clustering and classification methods (*DDclust* and *DDclass*, respectively) based on L_1 data depth (see Vardi and Zhang (2000)). The *DDclust* method is proposed to solve the problem of minimizing the sum of L_1 -distances from the observations to the nearest cluster representatives. The L_1 data depth is the amount of probability mass needed at a point z to make z the multivariate L_1 -median (a robust representative) of the data cluster.

An extension of *DDclust* by incorporating a trimmed procedure is introduced, aimed at segmenting the data into efficient size groups, using central (the deepest) people. In what follows, this methodology will be called *TDDclust* and it can be used within **Anthropometry** by using a function with the same name. All details about *TDDclust* are described in Vinué (2014).

hipamAnthropom

Representative fit models are important for defining a meaningful sizing system. However, there is no agreement among apparel manufacturers and almost every company employs a different fit model. Companies try to improve the quality of garment fit by scanning their fit models and deriving dress forms from the scans (Ashdown 2007; Song and Ashdown 2010). A fit model's measurements correspond to the commercial specifications established by each company to achieve the company's fit (Loker, Ashdown, and Schoenfelder 2005; Workman and Lentz 2000; Workman 1991). Beyond merely wearing the garment for inspection, a fit model provides objective feedback about fit, movement or comfort of a garment in place of the consumer.

In order to provide new insights about this problem, the *hipamAnthropom* methodology is proposed. It consists of two classification algorithms based on the Hierarchical Partitioning Around Medoids (HIPAM) clustering method presented in Wit and McClure (2004), modified to deal with anthropometric data. This procedure was published in Vinué *et al.* (2014b). The dissimilarity measure defined in McCulloch *et al.* (1998) and a different method to obtain a classification tree (Irigoién and Arenas 2008) were incorporated. One algorithm was called *HIPAM_{MO}* and the other one, *HIPAM_{IMO}*. The outputs of both of them include a set of central representative subjects or medoids taken from the original data set which constitute our fit models. In addition, they can detect outliers. This methodology is available in the *hipamAnthropom* function.

3.2. Statistical shape analysis: kmeansProcrustes

The clustering methodologies explained in Section 3.1 use a set of control anthropometric variables as the basis for a different type of sizing system in which people are grouped in a size group based on a full range of measurements. Consequently, clustering is done in the Euclidean space. The shape of the women recruited into the Spanish anthropometric survey is

represented by a set of correspondence points called landmarks. Taking advantage of this fact, we have adapted the k -means clustering algorithm to the field of statistical shape analysis, to define size groups of women according to their body shapes (Vinué *et al.* 2014c). The representative of each size group is the mean woman.

The k -means method is based on the fact that the sample mean is the value that minimizes the Euclidean distance from each point, to the centroid of the cluster to which it belongs. In order to adapt k -means to shape analysis, it is sufficient merely to replace the sample mean and the Euclidean distance by the Procrustes mean and the Procrustes distance, which are basic concepts of the shape analysis (Dryden and Mardia 1998). Thus, k -means can be used to cluster objects based on landmarks. Several attempts have been made in this regard, each one adapting a different version of the k -means. Amaral *et al.* in Amaral, Dore, Lessa, and Stosic (2010) adapted the Hartigan-Wong (H-W) version of k -means, while V. Georgescu in Georgescu (2009) used an algorithm in some aspects similar to the Lloyd version of k -means.

The main difference between Lloyd and H-W is that H-W proceeds point by point. Therefore, it requires the mean to be updated many more times than Lloyd. Unlike the sample mean in the Euclidean space, the calculus of the Procrustes mean needs much more time. Consequently, H-W should have a high computational burden in the shape space, losing efficiency, especially with a large sample size. In order to confirm empirically these hypothesis, we adapted both H-W and original Lloyd versions of k -means to the field of shape analysis and we demonstrated, by means of a simulation study, that the original Lloyd version of k -means is more efficient to clustering shapes than the H-W version (Vinué *et al.* 2014c).

The function to use the Lloyd version of k -means adapted to shape analysis (what we called *kmeansProcrustes*) is `LloydShapes`. In addition, the function to use the Hartigan-Wong version of k -means adapted to shape analysis is `HartiganShapes`. In addition, a trimmed version of *kmeansProcrustes* can be executed with `trimmedLloydShapes`.

3.3. Archetypal Analysis

In ergonomic related problems, where the goal is to create more efficient people-machine interfaces, a small set of extreme cases (boundary cases), called human models, is searched. Designing for extreme individuals is appropriate where some limiting factor can define either a minimum or maximum value which will accommodate the population. The basic principle is that accommodating boundary cases will be sufficient to accommodate the whole population.

For too long, the classical solution to select this small group of boundary models was based on the use of percentiles. However, percentiles are a kind of univariate descriptive statistic, so they are suitable only for univariate accommodation and should not be used in designs that involve two or more dimensions. In addition, they are not additive (Zehner *et al.* 1993; Robinette and McConville 1981; Moroney and Smith 1972). Today, the alternative commonly used for the multivariate accommodation problem is based on PCA (Friess and Bradtmiller 2003; Hudson, Zehner, and Meindl 1998; Robinson, Robinette, and Zehner 1992; Bittner, Glenn, Harris, Iavecchia, and Wherry 1987). However, it is known that the PCA approach presents some drawbacks (Friess 2005). In Epifanio *et al.* (2013), a different statistical approach for determining multivariate limits was put forward: the archetypal analysis (Cutler and Breiman 1994), and its advantages regarding to PCA were demonstrated. The function that allows us to reproduce the results discussed in Epifanio *et al.* (2013) is `archetypesBoundary`.

Archetypes computed by archetypal analysis are a convex combination of the sampled indi-

viduals, but they are not necessarily real observations. In some problems, it is crucial that the archetypes are real subjects, observations of the sample, and not fictitious. To that end, we have proposed a new archetypal concept: the archetypoid, which correspond to specific individuals and each observation of the data set can be represented as a mixture of these archetypoids (Vinué *et al.* 2014a). We have developed an efficient computational algorithm based on PAM to compute archetypoids (called archetypoid algorithm), we have analyzed some of their theoretical properties, we have explained how they can be obtained when only dissimilarities between observations are known (features are unavailable) and we have demonstrated some of their advantages regarding classical archetypes. The `stepArchetypoids` function calls the `archetypoids` function to run the archetypoid algorithm repeatedly.

4. The Anthropometry R package

First, the package must be loaded into R:

```
library(Anthropometry)
```

Next we briefly describe the R functions to execute each one of the methodologies introduced in Section 3.

4.1. Antropometric dimensions based clustering

A key element of two of the aforementioned clustering methodologies, *trimowa* and *hipamAnthropom*, is the dissimilarity function used. It is based on the dissimilarity measure proposed in McCulloch *et al.* (1998). The code to compute the dissimilarities with **Anthropometry** is written in C and it is exported from the NAMESPACE file.

Both clustering methodologies incorporate the calculus of the dissimilarity matrix within their main functions, `trimowa` and `hipamAnthropom`, respectively.

trimowa function

```
trimowa(x,w,K,alpha,niter,Ksteps,ahVect=c(23,28,20,25,25))
```

Its arguments are the following:

- `x`: Data frame. In our approach, this is each one of the subframes originated after segmenting the whole anthropometric Spanish survey in twelve bust segments, according to the European standard to sizing system. Size designation of clothes. Part 3: Measurements and intervals. Each row corresponds to an observation, and each column corresponds to a variable. All variables are numeric.
- `w`: The aggregation weights of the OWA operators. They are computed with the `WeightsMixtureUB`.
- `K`: Number of clusters.
- `alpha`: Proportion of trimmed sample.

- `niter`: Number of random initializations.
- `Ksteps`: Steps per initialization.
- `ahVect`: Constants that define the *ah* slopes of the distance function in `GetDistMatrix`. Given the five variables considered, this vector is `c(23,28,20,25,25)`. This vector would be other according to the variables considered.

TDDclust function

```
TDDclust(x,K,lambda,Th,A,T0,alpha,lplot,Trimm,data1)
```

Its arguments are the following:

- `x`: Data frame. Each row corresponds to an observation, and each column corresponds to a variable. All variables must be numeric.
- `K`: Number of clusters.
- `lambda`: Tuning parameter that controls the influence the data depth has over the clustering, see [Jörnsten \(2004\)](#).
- `Th`: Threshold for observations to be relocated, usually set to 0.
- `A`: Number of iterations.
- `T0`: Simulated annealing parameter. It is the current temperature in the simulated annealing procedure.
- `alpha`: Simulated annealing parameter. It is the decay rate, default 0.9.
- `lplot`: Tracking convergence, default 0.
- `Trimm`: Proportion of no accommodated sample.
- `data1`: The same data frame as `x`, used to incorporate the trimmed observations to the rest of them for the next iteration.

hipamAnthropom function

```
hipamAnthropom(x,asw.tol=0,maxsplit=5,local.const=NULL,orness=0.7,
               type,ahVect=c(23,28,20,25,25),...)
```

Its arguments are the following:

- `x`: Data frame. In our approach, this is each one of the subframes originated after segmenting the whole anthropometric Spanish survey in twelve bust segments, according to the European standard to sizing system. Size designation of clothes. Part 3: Measurements and intervals. Each row corresponds to an observation, and each column corresponds to a variable. All variables are numeric.

- `asw.tol`: If this value is given, a tolerance or penalty can be introduced (`asw.tol > 0` or `asw.tol < 0`, respectively) in the branch splitting procedure. Default value (0) is maintained. See (Wit and McClure 2004, p.154) for more details.
- `maxsplit`: The maximum number of clusters that any cluster can be divided when searching for the best clustering.
- `local.const`: If this value is given (meaningful values are those between -1 and 1), a proposed partition is accepted only if the associated `asw` is greater than this constant. Default option for this argument is maintained, that is to say, this value is ignored. See (Wit and McClure 2004, p.154) for more details.
- `orness`: Quantity to measure the degree to which the aggregation is like a min or max operation. See `WeightsMixtureUB` and `GetDistMatrix`.
- `type`: Type of HIPAM algorithm to be used. The possible options are 'MO' (for $HIPAM_{MO}$) and 'IMO' (for $HIPAM_{IMO}$).
- `ahVect`: Constants that define the *ah* slopes of the distance function in `GetDistMatrix`. Given the five variables considered, this vector is `c(23,28,20,25,25)`. This vector would be other according to the variables considered.
- ...: Other arguments that may be supplied to the internal functions of the HIPAM algorithms.

4.2. Statistical shape analysis

LloydShapes function

```
LloydShapes(dg,K,Nsteps=10,niter=10,stopCr=0.0001,simul,print)
```

Its arguments are the following:

- `dg`: Array with the 3D landmarks of the sample objects. Each row corresponds to an observation, and each column corresponds to a dimension (x,y,z).
- `K`: Number of clusters.
- `Nsteps`: Number of steps per initialization. Default value is 10.
- `niter`: Number of random initializations. Default value is 10.
- `stopCr`: Relative stopping criteria. Default value is 0.0001.
- `simul`: A logical value. If TRUE, this function is used for a simulation study.
- `print`: Logical value. If TRUE, some messages associated with the running process are displayed.

HartiganShapes function

```
HartiganShapes(dg,K,Nsteps=10,niter=10,stopCr=0.0001,simul,initLl,
               initials,print)
```

Its arguments are the following:

- `dg`: Array with the 3D landmarks of the sample objects. Each row corresponds to an observation, and each column corresponds to a dimension (x,y,z).
- `K`: Number of clusters.
- `Nsteps`: Number of steps per initialization. Default value is 10.
- `niter`: Number of random initializations. Default value is 10.
- `stopCr`: Relative stopping criteria. Default value is 0.0001.
- `simul`: A logical value. If `TRUE`, this function is used for a simulation study.
- `initLl`: Logical value. If `TRUE`, see next argument *initials*. If `FALSE`, they are new random initial values.
- `initials`: If *initLl=TRUE*, they are the same random initial values used in each iteration of `LloydShapes`. If *initLl=FALSE* this argument must be passed simply as an empty vector.
- `print`: Logical value. If `TRUE`, some messages associated with the running process are displayed.

trimmedLloydShapes function

```
trimmedLloydShapes(dg,n,alpha,K,Nsteps=10,niter=10,
                   stopCr=0.0001,print)
```

Its arguments are the following:

- `dg`: Array with the 3D landmarks of the sample objects. Each row corresponds to an observation, and each column corresponds to a dimension (x,y,z).
- `n`: Number of individuals.
- `alpha`: Proportion of trimmed sample.
- `K`: Number of clusters.
- `Nsteps`: Number of steps per initialization. Default value is 10.

- `niter`: Number of random initializations. Default value is 10.
- `stopCr`: Relative stopping criteria. Default value is 0.0001.
- `print`: Logical value. If TRUE, some messages associated with the running process are displayed.

4.3. Archetypal analysis

`archetypesBoundary` function

```
archetypesBoundary(data, numArchet, verbose, nrep)
```

- `data`: USAF 1967 database (see `dataUSAF`). Each row corresponds to an observation, and each column corresponds to a variable. All variables are numeric.
- `numArchet`: Number of archetypes.
- `verbose`: Logical value. If TRUE, some details of the execution progress are shown (this is the same argument as that of the `stepArchetypes` function of the **archetypes** R package (Eugster and Leisch 2009)).
- `nrep`: For each archetype run `archetypes` `nrep` times (this is the same argument as that of the `stepArchetypes` function of **archetypes**).

`archetypoids` function

```
archetypoids(i, data, huge=200, step, init, ArchObj, nearest, sequ, aux)
```

- `i`: Number of archetypoids.
- `data`: Data matrix. Each row corresponds to an observation and each column corresponds to an anthropometric variable. All variables are numeric.
- `huge`: This is a penalization added to solve the convex least squares problems regarding the minimization problem to estimate archetypoids, see Eugster and Leisch (2009). Default value is 200.
- `step`: Logical value. If TRUE, the archetypoid algorithm is executed repeatedly within `stepArchetypoids`. Therefore, this function requires the next argument `init` (but neither the `ArchObj` nor the `nearest` arguments) that specifies the initial vector of archetypoids, which has been already computed within `stepArchetypoids`. If FALSE, the archetypoid algorithm is executed once. In this case, the `ArchObj` and `nearest` arguments are required to compute the initial vector of archetypoids.

- `init`: Initial vector of archetypoids for the BUILD phase of the archetypoid algorithm. It is computed within `stepArchetypoids`. See next `nearest` argument to know how this vector is calculated.
- `ArchObj`: The list returned by the `stepArchetypesMod` function. This function is a slight modification of the original `stepArchetypes` function of `archetypes` to apply the archetype algorithm to raw data. The `stepArchetypes` function standardizes the data by default and this option is not always desired. This list is needed to compute the nearest individuals to archetypes. Required when `step=FALSE`.
- `nearest`: Initial vector of archetypoids for the BUILD phase of the archetypoid algorithm. Required when `step=FALSE`. This argument is a logical value: if TRUE (FALSE), the `nearest` (*which*) vector is calculated. Both vectors contain the nearest individuals to the archetypes returned by the `archetypes` function of `archetypes` (In [Vinué et al. \(2014a\)](#), archetypes are computed after running the archetype algorithm twenty times). The `nearest` vector is calculated by computing the Euclidean distance between the archetypes and the individuals and choosing the nearest. It is used in [Epifanio et al. \(2013\)](#). The *which* vector is calculated by identifying consecutively the individual with the maximum value of alpha for each archetype, until getting the number of archetypes defined. It is used in [Eugster \(2012\)](#).
- `sequ`: Logical value. It indicates whether a sequence of archetypoids (TRUE) or only a single number of them (FALSE) is computed. It is determined by the number of archetypes computed by means of `stepArchetypesMod`.
- `aux`: If `sequ=FALSE`, this value is equal to $i-1$ since for a single number of archetypoids, the list associated with the archetype object only has one element.

`stepArchetypoids` function

```
stepArchetypoids(i,nearest,data,ArchObj)
```

- `i`: Number of archetypoids.
- `nearest`: Initial vector of archetypoids for the BUILD phase of the archetypoid algorithm. This argument is a logical value: if TRUE (FALSE), the `nearest` (*which*) vector is calculated. Both vectors contain the nearest individuals to the archetypes returned by the `archetypes` function of `archetypes` (In [Vinué et al. \(2014a\)](#), archetypes are computed after running the archetype algorithm twenty times). The `nearest` vector is calculated by computing the Euclidean distance between the archetypes and the individuals and choosing the nearest. It is used in [Epifanio et al. \(2013\)](#). The *which* vector is calculated by identifying consecutively the individual with the maximum value of alpha for each archetype, until getting the number of archetypes defined. It is used in [Eugster \(2012\)](#).
- `data`: Data matrix. Each row corresponds to an observation and each column corresponds to an anthropometric variable. All variables are numeric.

- ArchObj: The list returned by the `stepArchetypesMod` function. This function is a slight modification of the original `stepArchetypes` function of **archetypes** to apply the archetype algorithm to raw data. The `stepArchetypes` function standardizes the data by default and this option is not always desired. This list is needed to compute the nearest individuals to archetypes.

stepArchetypesMod function

```
stepArchetypesMod(data,k,nrep=3,verbose=TRUE)
```

- data: Data to obtain archetypes.
- k: Number of archetypes to compute, from 1 to k .
- nrep: For each k , run `archetypes` $nrep$ times.
- verbose: If TRUE, the progress during execution is shown.

5. Examples

5.1. Anthropometric dimensions based clustering

The following code executes the `trimowa` methodology. A similar code was used to obtain the results described in [Ibáñez *et al.* \(2012b\)](#). In the next example, the number of random initializations is 10, with three steps per initialization. Figure 1 shows the scatter plots of bust circumference against neck to ground with the three medoids obtained for each bust class without (left) and with (right) the prototypes defined by the European standard. Note that any other number of clusters could be also obtained.

```
#Data loading:
dataDef <- dataDemo
num.variables <- dim(dataDef)[2]
bust <- dataDef$bust

#Aggregation weights calculation:
orness <- 0.7
w <- WeightsMixtureUB(orness,num.variables)

#Bust classes definition according to the European standard:
bustCirc_4 <- seq(74,102,4)
bustCirc_6 <- seq(107,131,6)
bustCirc <- c(bustCirc_4,bustCirc_6)

#Trimowa parameters:
```

```

nsizes <- length(bustCirc)
K <- 3 ; alpha <- 0.01 ; niter <- 10 ; Ksteps <- 7
ahVect <- c(23,28,20,25,25)

#For reproducing results, seed for randomness:
set.seed(2014)
res_trimowa <- list()
for (i in 1 : (nsizes-1)){
data = dataDef[(bust >= bustCirc[i]) & (bust < bustCirc[i + 1]), ]
res_trimowa[[i]] <- trimowa(data,w,K,alpha,niter,Ksteps,ahVect=ahVect)
}

#Medoids provided for each bust class:
medoids <- list()
for (i in 1 : (nsizes-1)){
medoids[[i]] <- res_trimowa[[i]]$meds
}

#Plotting arguments:
bustVariable <- "bust"
xlim <- c(70,150)
color <- c("black", "red", "green", "blue", "cyan", "brown", "gray",
           "deeppink3", "orange", "springgreen4", "khaki3", "steelblue1")

variable <- "necktoground"
ylim = c(110,160) #see range(dataDef[,variable])
title <- "Medoids \n bust vs neck to ground"

#Figure 1 (left and right):
plotMedoids(dataDef,medoids,nsizes,bustVariable,variable,color,
            xlim,ylim,title,FALSE)
plotMedoids(dataDef,medoids,nsizes,bustVariable,variable,color,
            xlim,ylim,title,TRUE)

```

In addition, the following sentences illustrate how to use the `hipamAnthropom` methodology. Specifically, some results returned by the `HIPAMIMO` algorithm are shown. Figure 2 displays the medoids (left) and the outlier women (right) corresponding to each bust size.

We emphasize that `dataDemo` is divided into twelve bust segments here as well. Thus, each segment has a small sample size. This might explain that this algorithm (and also `HIPAMMO`) does not find big homogeneous clusters and therefore, identifies a lot of women as outliers in each class for this case. One of the features of the HIPAM algorithm is that it is a very sensitive algorithm to identify outliers. A broad discussion, analysis and thoughts of the anthropometric meaning of these outliers is given in [Vinué *et al.* \(2014b\)](#) (including its supplementary material).


```

#hipamBigGroups is a function of Anthropometry that returns the medoids
#of the clusters with more than 2 elements:
list.meds <- lapply(1:(nsizes - 1), FUN = hipamBigGroups, hip)
#outlierHipam is a function of Anthropometry that returns the individuals
#of the clusters with 1 or 2 elements (outliers):
list_outl1_2 <- sapply(1 : (nsizes - 1), FUN = outlierHipam, hip)

#Plotting arguments:
bustVariable <- "bust"
xlim <- c(70,150)
color <- c("black", "red", "green", "blue", "cyan", "brown", "gray",
           "deeppink3", "orange", "springgreen4", "khaki3", "steelblue1")

variable <- "hip"
ylim <- c(80,160)
title <- "Medoids HIPAM_IMO \n bust vs hip"
title_outl <- "Outlier women HIPAM_IMO \n bust vs hip"

#Figure 2 (left and right):
plotMedoids(dataDef,list.meds,nsizes,bustVariable,variable,color,
            xlim,ylim,title,FALSE)
plotTrimmOutl(dataDef,list_outl1_2,nsizes,bustVariable,variable,color,
              xlim,ylim,title_outl)

```

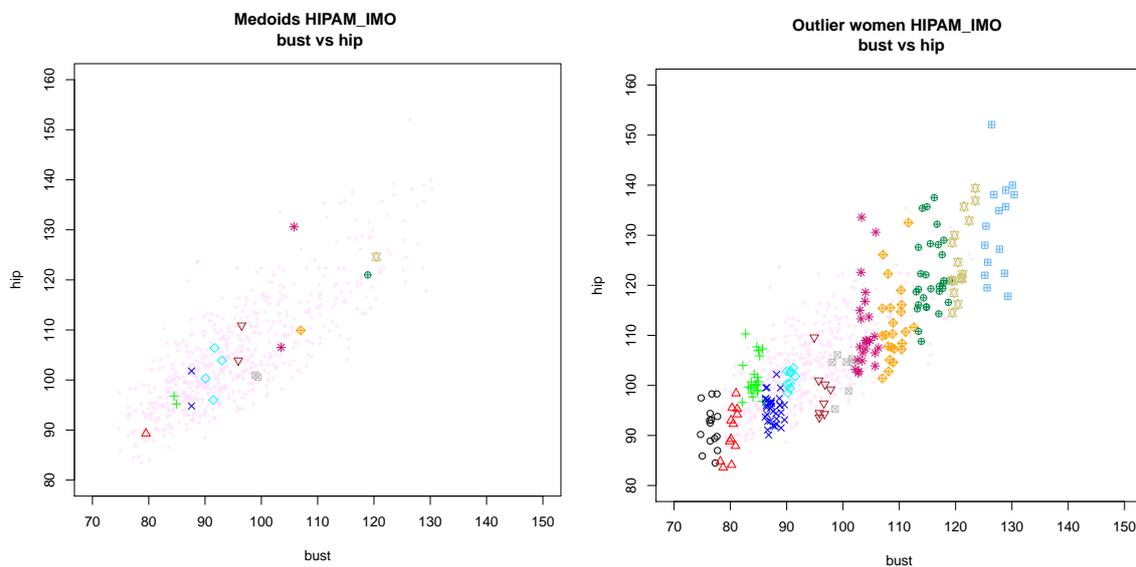


Figure 2: Bust vs hip with the medoids (left) and with the outliers (right) obtained using $HIPAM_{IMO}$.

Next, a basic example of the TDD_{clus} methodology is shown.

```

#In the interests of simplicity of the computation involved
#only a small sample (the first 50 individuals) is selected:
dataDef <- dataDemo[1:50,c(2,3,5)] #Neck to ground, waist and bust variables.
data1 <- dataDemo[1:50,c(2,3,5)]

#TDDclust parameters:
K=3      ; lambda=0.5      ; Th=0
A=5      ; T0=0            ; alpha=.9
lplot=0  ; percTrimm=0.1

Dout <- TDDclust(x=dataDef,K=K,lambda=lambda,Th=Th,A=A,T0=T0,alpha=alpha,
                lplot=lplot,Trimm=percTrimm,data1=data1)

#Clustering results:
table(Dout$NN[1,])
#Final value of the optimal partition:
Dout$Cost
#Iteration in which the optimal partition was found:
Dout$klBest
#Trimmed observations:
Dout$indivTrimmed

```

5.2. Statistical shape analysis

Next, the results obtained with the trimmed *kmeansProcrustes* by using five iterations, five number of steps per initialization and a relative stopping criteria of 0.0001 are presented. Figure 3 displays the boxplots for neck to ground measurement for the three clusters calculated (left) and the projection on the plane xy of the registered points and mean shape for cluster 1 (right).

```

landmarks1 <- na.exclude(landmarks)
num.points <- (dim(landmarks1)[2]) / 3
#In the interests of simplicity of the computation involved
#only a small sample (the first 50 individuals) is selected:
landmarks2 <- landmarks1[1:50,]
n <- dim(landmarks2)[1]

#Array with the 3D landmarks of the sample objects:
dg <- array(0,dim = c(num.points,3,n))
for(k in 1:n){
  for(l in 1:3){
    dg[,l,k] <- as.matrix(as.vector(landmarks2[k,][seq(1,dim(landmarks2)[2]+
      (l-1),by=3)]),ncol=1,byrow=T)
  }
}

```

```

#kmeansProcrustes parameters:
K <- 3 ; alpha <- 0.01 ; Nsteps <- 5 ; niter <- 5 ; stopCr <- 0.0001
#For reproducing results, seed for randomness:
set.seed(2013)
res <- trimmedLloydShapes(dg,n,alpha,K,Nsteps,niter,stopCr,TRUE)

#Numerical and graphical results:
asig <- res$asig #table(asig) shows the clustering results.
copt <- res$copt #optimal centers.

#To identify the trimmed individuals of the optimal iteration:
iter_opt <- res$trimmsIter[length(res$trimmsIter)]
trimm <- res$trimmWomen[[iter_opt]][[res$betterNstep]]

#Generalised Procrustes analysis into each cluster:
out_proc <- list()
for(h in 1 : K){
  out_proc[[h]] = shapes::procGPA(dg[, , asig == h], distances = T,
                                pcaoutput = T)
}

data <- dataDemo[1:50,]
data <- data[-trimm,]
#Figure 3 (left):
boxplot(data$necktoground ~ as.factor(asig), main = "Neck to ground")
#Figure 3 (right):
plotshapes(out_proc[[1]]$rotated)
points(copt[, , 1], col = 2)
legend("topleft", c("Registrated data", "Mean shape"), pch = 1,
       col = 1:2, text.col = 1:2)
title("Procrustes registrated data for cluster 1 \n
      with its mean shape superimposed", sub = "Plane xy")

```

5.3. Archetypal analysis

Three archetypoids for `dataUSAF` are calculated with the following code. Figure 4 shows the percentiles of three archetypoids, beginning with *nearest* (left) and with *which* (right).

```

#Cockpit design problem:
#In the interests of simplicity of the computation involved
#only a small sample (the first 50 individuals) is selected:
m <- dataUSAF[1:50,]
#Variable selection (cockpit dimensions):
sel <- c(48,40,39,33,34,36)
#Changing to inches:
mpulg <- m[,sel] / (10 * 2.54)

```

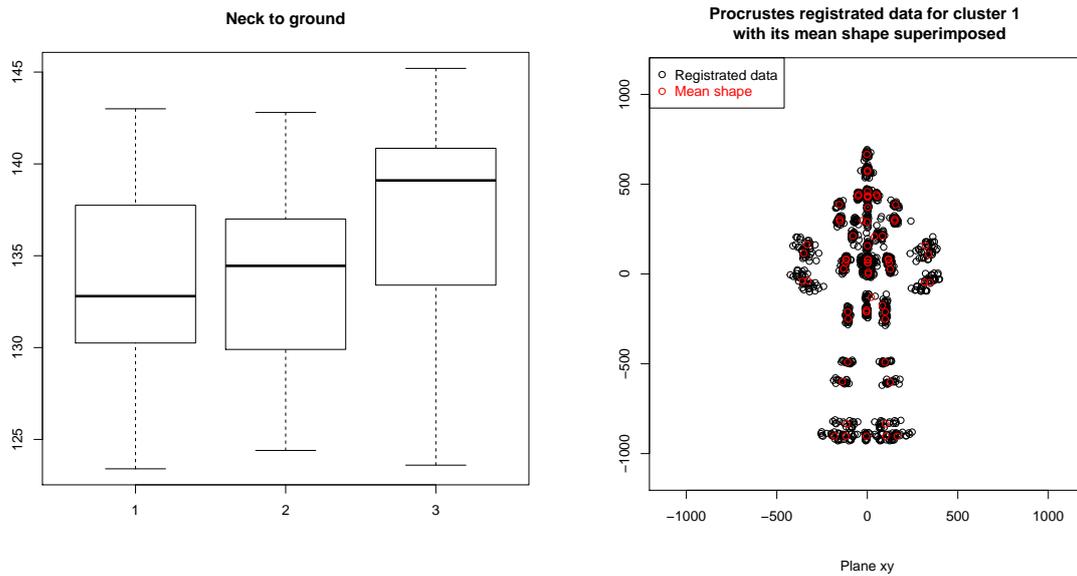


Figure 3: Boxplots for the neck to ground measurement for three clusters (left) and projection on the plane xy of the registered points and mean shape for cluster 1 (right). Results provided by trimmed *kmeansProcrustes*.

```
#Data preprocessing:
preproc <- accommodation(mpulg,TRUE,0.95,TRUE)

#For reproducing results, seed for randomness:
set.seed(2010)
#Run archetype algorithm repeatedly from 1 to numArch archetypes:
numArch <- 10 ; nrep <- 20
lass <- stepArchetypesMod(data=preproc$data,k=1:numArch,verbose=FALSE,
                          nrep=nrep)
screepplot(lass)

i <- 3 #number of archetypoids to compute.
res <- archetypoids(i,preproc$data,huge=200,step=FALSE,ArchObj=lass,
                  nearest=TRUE,sequ=TRUE)
res_which <- archetypoids(i,preproc$data,huge=200,step=FALSE,ArchObj=lass,
                        nearest=FALSE,sequ=TRUE)

aux <- res$archet
aux_wh <- res_which$archet
#In this case, the nearest and which archetypoids match
#(although the nearest and which archetypes do not),
#so it's enough to represent a single percentiles plot:
```

```

percs <- list()
#In case the nearest and which archetypoids don't match:
#percs_wh <- list()
for(j in 1:length(aux)){
  percs[[j]] <- sapply(1:dim(preproc$data)[2], compPerc, aux[j], preproc$data, 0)
  #percs_wh[[j]] <- sapply(1:dim(preproc$data)[2], compPerc, aux_wh[j],
    #preproc$data, 0)
}
m <- matrix(unlist(percs), nrow=6, ncol=length(percs), byrow=F)
#m1 <- matrix(unlist(percs_wh), nrow=6, ncol=length(percs_wh), byrow=F)

#Figure 4:
barplot(m, beside=TRUE,
        main = paste(i, " archetypoids", sep = ""),
        ylim=c(0,100), ylab="Percentile")

```

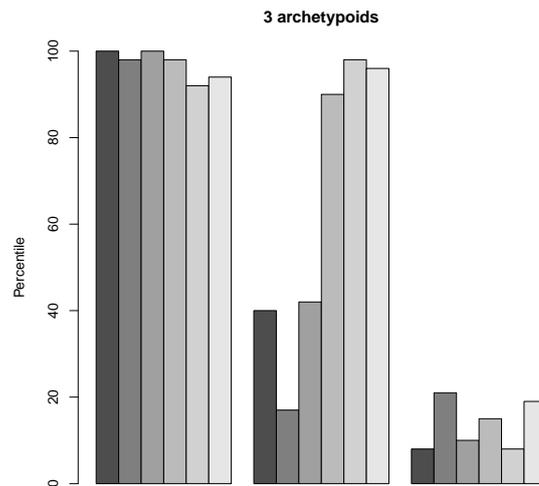


Figure 4: Percentiles of three archetypoids, beginning with *nearest* and with *which* for `dataUSAF`. In this case, the *nearest* and *which* archetypoids coincide.

6. Conclusions

New three-dimensional whole-body scanners have drastically reduced the cost and duration of the measurement process. These types of systems, in which the human body is digitally scanned and the resulting data converted into exact measurements, make it possible to obtain accurate, reproducible and updated anthropometric data. These databases constitute very valuable information to effectively design better-fitting clothing and workstations, to understand the body shape of the population and to reduce the design process cycle. Therefore,

rigorous statistical methodologies and software applications must be developed to make the most of them.

This paper introduces a new R package called **Anthropometry** that brings together different statistical methodologies concerning clustering, the statistical concept of data depth, the statistical shape analysis and the archetypal analysis, especially developed to deal with anthropometric data. They use the data obtained from a 3D anthropometric survey of the Spanish female population and from the USAF survey. Procedures related to clustering, data depth and shape analysis are aimed at defining optimal clothing size groups and central prototypes. The two approaches based on the archetypal analysis are useful for determining boundary human models which could be useful to improve industry practice in workspace design.

The **Anthropometry** R package is a positive contribution to tackle some statistical problems related to Ergonomics and Anthropometry. It provides a useful software tool to engineers and researchers in these fields so that they can analyze their anthropometric data in a comprehensive way.

Acknowledgments

The author gratefully acknowledges the many helpful suggestions of I. Epifanio and G. Ayala. The author would also like to thank the Biomechanics Institute of Valencia for providing us with the Spanish anthropometric data set and the Spanish Ministry of Health and Consumer Affairs for having commissioned and coordinated the “Anthropometric Study of the Female Population in Spain”. This paper has been partially supported by the following grants: TIN2009-14392-C02-01, TIN2009-14392-C02-02.

References

- Alemany S, González JC, Nácher B, Soriano C, Arnáiz C, Heras H (2010). “Anthropometric Survey of the Spanish Female Population Aimed at the Apparel Industry.” In *Proceedings of the 2010 Intl. Conference on 3D Body scanning Technologies*. Lugano, Switzerland.
- Amaral GJA, Dore LH, Lessa RP, Stosic B (2010). “k-Means Algorithm in Statistical Shape Analysis.” *Communications in Statistics - Simulation and Computation*, **39**(5), 1016–1026.
- Ashdown S & Loker S (2005). “Improved Apparel Sizing: Fit and Anthropometric 3D Scan Data.” *Technical report*, National Textile Center Annual Report.
- Ashdown SP (2007). *Sizing in clothing: Developing effective sizing systems for ready-to-wear clothing*. Woodhead Publishing in Textiles.
- Bagherzadeh R, Latifi M, Faramarzi AR (2010). “Employing a Three-Stage Data Mining Procedure to Develop Sizing System.” *World Applied Sciences Journal*, **8**(8), 923–929.
- Bertilsson E, Högberg D, Hanson L (2012). “Using experimental design to define boundary manikins.” *Work: A Journal of Prevention, Assessment and Rehabilitation*, **41**(Supplement 1), 4598–4605.

- Bittner AC, Glenn FA, Harris RM, Iavecchia HP, Wherry RJ (1987). "CADRE: A family of manikins for workstation design." In *Asfour, S.S. (ed.) Trends in Ergonomics/Human Factors IV. North Holland*, pp. 733–740.
- Blanchonette P (2010). "Jack Human Modelling Tool: A Review." *Technical Report DSTO-TR-2364*, Defence Science and Technology Organisation (Australia). Air Operations Division.
- Cheng Y, Church GM (2000). "Biclustering of expression data." *Proceedings of the Eighth International Conference on Intelligent Systems for Molecular Biology*, **8**, 93–103.
- Chung MJ, Lin HF, Wang MJJ (2007). "The development of sizing systems for Taiwanese elementary- and high-school students." *International Journal of Industrial Ergonomics*, **37**, 707–716.
- Cutler A, Breiman L (1994). "Archetypal Analysis." *Technometrics*, **36**(4), 338–347.
- D'Apuzzo N (2009). *Recent Advances in 3D full body scanning with applications to fashion and apparel*. Gruen, A., Kahmen, H. (eds.). Optical 3-D Measurement Techniques IX, Vienna, Austria.
- Ding Y, Dang X, Peng H, Wilkins D (2007). "Robust clustering in high dimensional data using statistical depths." *BMC Bioinformatics*, **8**(Suppl 7:S8), 1–16.
- Dryden IE, Mardia KV (1998). *Statistical Shape Analysis*. John Wiley & Sons.
- Dutta D, Ghosh A (2012). "On robust classification using projection depth." *Annals of the Institute of Statistical Mathematics*, **64**(3), 657–676.
- Epifanio I, Vinué G, Alemany S (2013). "Archetypal Analysis: Contributions for Estimating Boundary Cases in Multivariate Accommodation Problem." *Computers & Industrial Engineering*, **64**, 757–765.
- Eugster MJ, Leisch F (2009). "From Spider-Man to Hero - Archetypal Analysis in R." *Journal of Statistical Software*, **30**(8), 1–23.
- Eugster MJA (2012). "Performance Profiles Based on Archetypal Athletes." *International Journal of Performance Analysis in Sport*, **12**(1), 166–187.
- European Committee for Standardization (2002). "Size designation of clothes. Part 2: Primary and secondary dimensions."
- European Committee for Standardization (2005). "Size designation of clothes. Part 3: Measurements and intervals."
- Friess M (2005). "Multivariate Accommodation Models using Traditional and 3D Anthropometry." *Technical report*, SAE.
- Friess M, Bradtmiller B (2003). "3D Head Models for Protective Helmet Development." *Technical report*, SAE.
- García-Escudero LA, Gordaliza A, Matrán C (2003). "Trimming Tools in Exploratory Data Analysis." *Journal of Computational and Graphical Statistics*, **12**(2), 434–449.

- García-Escudero LA, Gordaliza A, Matrán C, Mayo-Isacar A (2008). “A General Trimming Approach to Robust Cluster Analysis.” *The Annals of Statistics*, **36**, 1324–1345.
- Georgescu V (2009). “Clustering of Fuzzy Shapes by Integrating Procrustean Metrics and Full Mean Shape Estimation into k-Means Algorithm.” In *IFSA-EUSFLAT Conference*.
- Gupta D, Gangadhar BR (2004). “A statistical model for developing body size charts for garments.” *International Journal of Clothing Science and Technology*, **16**(5), 458–469.
- HFES 300 Committee (2004). *Guidelines for Using Anthropometric Data in Product Design*. Human Factors and Ergonomics Society.
- Hsu CH (2009a). “Data mining to improve industrial standards and enhance production and marketing: An empirical study in apparel industry.” *Expert Systems with Applications*, **36**, 4185–4191.
- Hsu CH (2009b). “Developing accurate industrial standards to facilitate production in apparel manufacturing based on anthropometric data.” *Human Factors and Ergonomics in Manufacturing*, **19**(3), 199–211. doi:10.1002/hfm.20148.
- Hudson JA, Zehner GF, Meindl RD (1998). “The USAF Multivariate Accommodation Method.” *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, **42**(10), 722–726.
- Ibáñez MV, Simó A, Domingo J, Durá E, Ayala G, Alemany S, Vinué G, Solves C (2012a). “A Statistical Approach to Build 3D Prototypes from a 3D Anthropometric Survey of the Spanish Female Population.” In *Proceedings of the 1st International Conference on Pattern Recognition Applications and Methods*, volume 1, pp. 370–374. Vilamoura, Algarve, Portugal.
- Ibáñez MV, Vinué G, Alemany S, Simó A, Epifanio I, Domingo J, Ayala G (2012b). “Apparel Sizing Using Trimmed PAM and OWA Operators.” *Expert Systems with Applications*, **39**, 10512–10520.
- Irigoien I, Arenas C (2008). “INCA: New Statistic for Estimating the Number of Clusters and Identifying Atypical Units.” *Statistics in Medicine*, **27**, 2948–2973.
- Istook CL, Hwang SJ (2001). “3D Body Scanning Systems with Application to the Apparel Industry.” *Journal of Fashion Marketing and Management*, **5**, 120–132.
- Jörnsten R (2004). “Clustering and classification based on the L_1 data depth.” *Journal of Multivariate Analysis*, **90**, 67–89.
- Jörnsten R, Vardi Y, Zhang C (2002). “A robust clustering method and visualization tool based on data depth.” In *Statistical data analysis based on the L_1 norm and related methods (Neuchâtel, 2002)*, *Stat. Ind. Technol.*, pp. 353–366.
- Kaiser S, Leisch F (2008). “A Toolbox for Bicluster Analysis in R.” *Technical report*, Department of Statistics (University of Munich).

- Kaiser S, Santamaria R, Khamiakova T, Sill M, Theron R, Quintales L, Leisch F (2011). *biclust: BiCluster Algorithms*. R package version 1.0.1, URL <http://CRAN.R-project.org/package=biclust>.
- Kaufman, L and Rousseeuw, P J (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley, New York.
- Lange T, Mosler K, Mozharovskiy P (2012). “Fast nonparametric classification based on data depth.” *Statistical Papers*, **5**, 1–22.
- Lerch T, MacGillivray M, Domina T (2007). “3D Laser Scanning: A Model of Multidisciplinary Research.” *Journal of Textile and Apparel, Technology and Management*, **5**, 1–22.
- Liu RY, Parelius JM, Singh K (1999). “Multivariate Analysis by Data Depth: Descriptive Statistics, Graphics and Inference.” *The Annals of Statistics*, **27**(3), 783–858.
- Loker S, Ashdown S, Schoenfelder K (2005). “Size-specific Analysis of Body Scan Data to Improve Apparel Fit.” *Journal of Textile and Apparel, Technology and Management*, **4**(3), 1–15.
- López A, Romo J (2010). “Simplicial similarity and its application to hierarchical clustering.” *Technical report*, Universidad Carlos III de Madrid. Departamento de Estadística. Working papers. Statistics and Econometrics.
- Lu JM, Wang MJJ (2008). “Automated Anthropometric Data Collection Using 3D Whole Body Scanners.” *Expert Systems with Applications*, **35**, 407–414.
- Luximon A, Zhang Y, Luximon Y, Xiao M (2012). “Sizing and grading for wearable products.” *Computer-Aided Design*, **44**, 77–84.
- Madeira SC, Oliveira AL (2004). “Biclustering Algorithms for Biological Data Analysis: A Survey.” *IEEE Transactions on Computational Biology and Bioinformatics*, **1**, 24–45.
- McCulloch CE, Paal B, Ashdown SP (1998). “An Optimization Approach to Apparel Sizing.” *Journal of the Operational Research Society*, **49**, 492–499.
- Moroney WF, Smith MJ (1972). “Empirical reduction in potential user population as the result of imposed multivariate anthropometric limits.” *Technical report*, Naval Aerospace Medical Research Laboratory.
- Pan JZ, Jörnsten R, Hart RP (2004). “Screening anti-inflammatory compounds in injured spinal cord with microarrays: a comparison of bioinformatics analysis approaches.” *Physiological Genomics*, **17**, 201–214.
- Parkinson MB, Reed MP, Kokkolaras M, Papalambros PY (2006). “Optimizing Truck Cab Layout for Driver Accommodation.” *Journal of Mechanical Design*, **129**(11), 1110–1117.
- Pheasant S (2003). *Bodyspace: Anthropometry, Ergonomics and the Design of Work*. Taylor & Francis, Ltd.
- R Development Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.

- Robinette KM, McConville JT (1981). “Alternative to Percentile Models.” *Technical report*, SAE.
- Robinson JC, Robinette KM, Zehner GF (1992). “User’s guide to the anthropometric database at the computerized anthropometric research and design (CARD) laboratory (U).” *Technical report*, Systems Research Laboratories Inc.
- Shu C, Wuhrer S, Xi P (2011). “Geometric and Statistical Methods for Processing 3D Anthropometric Data.” In *International Symposium on Digital Human Modeling*.
- Simmons KP, Istook CL (2003). “Body measurement techniques: Comparing 3D body-scanning and anthropometric methods for apparel applications.” *Journal of Fashion Marketing and Management*, **7**(3), 306–332.
- Song HK, Ashdown SP (2010). “An Exploratory Study of the Validity of Visual Fit Assessment From Three-Dimensional Scans.” *Clothing and Textiles Research Journal*, **28**(4), 263–278.
- Tryfos P (1986). “An Integer Programming Approach to the Apparel Sizing Problem.” *The Journal of the Operational Research Society*, **37**(10), 1001–1006.
- Vardi Y, Zhang CH (2000). “The multivariate L_1 -median and associated data depth.” *Proceedings of the National Academy of Sciences*, **97**, 1423–1426.
- Vinué G (2012). *Métodos biclustering aplicados a datos antropométricos: Exploración de su posible aplicación en el diseño de indumentaria*. Master’s thesis, School of Mathematics, University of Valencia (Spain). In Spanish.
- Vinué G (2014). *Development of statistical methodologies applied to anthropometric data oriented towards the ergonomic design of products*. Ph.D. thesis, Faculty of Mathematics, University of Valencia, Spain. To be defended in the next few months.
- Vinué G, Epifanio I, Alemany S (2014a). “Archetypoids: A new approach to define representative archetypal data.” Submitted.
- Vinué G, Ibáñez MV (2014). “Data depth and Biclustering applied to anthropometric data: Exploring their utility in apparel design.” In progress.
- Vinué G, León T, Alemany S, Ayala G (2014b). “Looking for representative fit models for apparel sizing.” *Decision Support Systems*, **57**, 22–33.
- Vinué G, Simó A, Alemany S (2014c). “The k -means algorithm for 3D shapes with an application to apparel design.” Submitted.
- Wang MJJ, Wu WY, Lin KC, Yang SN, Lu JM (2007). “Automated anthropometric data collection from three-dimensional digital human models.” *The International Journal of Advanced Manufacturing Technology*, **32**, 109–115.
- Wit E, McClure J (2004). *Statistics for Microarrays: Design, Analysis and Inference*. John Wiley & Sons, Ltd.
- Workman J (1991). “Body Measurement Specifications for Fit Models as a Factor in Clothing Size Variation.” *Clothing and Textiles Research Journal*, **10**(1), 31–36.

- Workman JE, Lentz ES (2000). "Measurement Specifications for Manufacturers' Prototype Bodies." *Clothing and Textiles Research Journal*, **18**(4), 251–259.
- Zehner GF, Meindl RS, Hudson JA (1993). "A Multivariate Anthropometric Method for Crew Station Design: Abridged." *Technical report*, Human Engineering Division, Armstrong Laboratory, Wright-Patterson Air Force Base, Ohio.
- Zheng R, Yu W, Fan J (2007). "Development of a new chinese bra sizing system based on breast anthropometric measurements." *International Journal of Industrial Ergonomics*, **37**, 697–705.
- Zuo Y, Serfling R (2000). "General Notions of Statistical Depth Function." *The Annals of Statistics*, **28**(2), 461–482.

Affiliation:

Guillermo Vinué
Department of Statistics and Operations Research
Faculty of Mathematics
University of Valencia
46100 Burjassot, Spain
E-mail: Guillermo.Vinue@uv.es
URL: <http://www.uv.es/vivigui>