Supplementary Material for

Looking for representative fit models for apparel sizing

G. Vinué[(1)], T. León[(1)], S. Alemany[(2)] and G. Ayala[(1)]

(1) Department of Statistics and O.R., Avda. Vicent Andrés Estellés, 1. 46100-Burjasot. University of Valencia, Valencia, Spain.
(2) Biomechanics Institute of Valencia, Universidad Politécnica de Valencia, Valencia, Spain.

Abstract

This paper is concerned with the generation of optimal fit models for use in apparel design. Representative fit models or prototypes are important for defining a meaningful sizing system. However, there is no agreement among apparel manufacturers and each one has their own prototypes and size charts i.e. there is a lack of standard sizes in garments from different apparel manufacturers.

We propose two algorithms based on a new hierarchical partitioning around medoids clustering method originally developed for gene expression data. We are concerned with a different application; therefore, the dissimilarity between the objects has to be different and must be designed to deal with anthropometric features. Furthermore, one of the algorithms incorporates a different rule to split the clusters, which, in our case, provides better results. Our procedures not only make it possible to obtain optimal prototypes, but also to detect outliers. These outliers should be removed before defining prototypes so that the companies' market share can be optimized.

All the analyses are performed using the anthropometric database obtained from a survey of the Spanish female population.

Keywords: HIPAM, Hierarchical tree, Partitioning around medoids, Fit models, Mean Split Silhouette, INCA statistic.

Email address: `guillermo.vinue@uv.es`, `teresa.leon@uv.es`,`sandra.alemany@ibv.upv.es`, `guillermo.ayala@uv.es` (G. Vinué[(1)], T. León[(1)], S. Alemany[(2)] and G. Ayala[(1)])

Preprint submitted to Elsevier *January 10, 2014*

4. Application of HIPAM to apparel sizing

4.2. Experimental results

4.2.1. Database segmented by bust circumference

First of all, some different scatterplots that represent the medoids and outliers returned by $HIPAM_{MO}$ and $HIPAM_{IMO}$ in every bust class are given in figures 1, 2, 3 and 4:

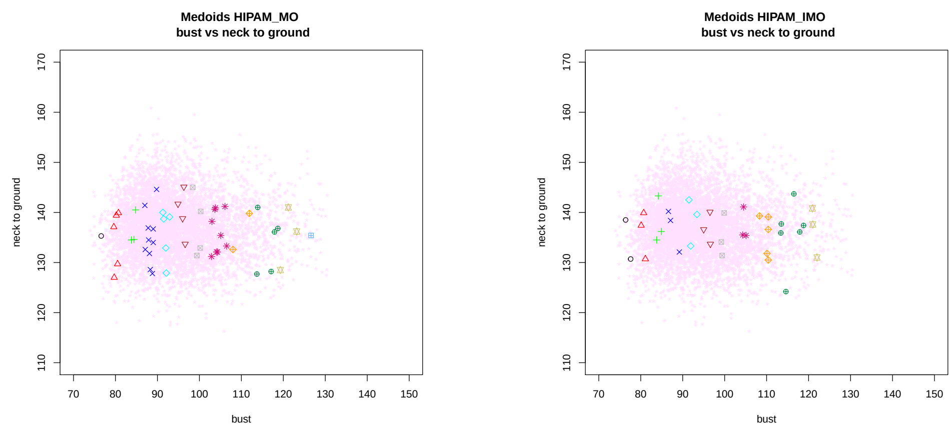

Figure 1. Bust vs neck to ground in the medoids obtained using $HIPAM_{MO}$ (left) and $HIPAM_{IMO}$ (right).

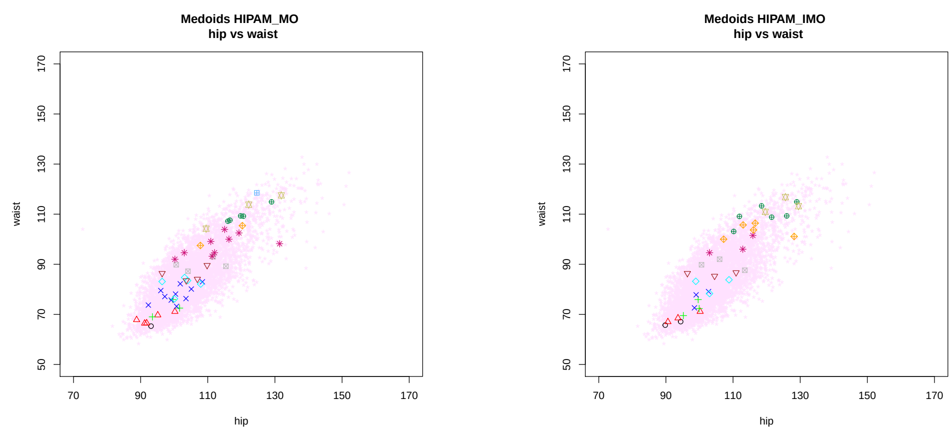

Figure 2. Hip vs waist in the medoids obtained using $HIPAM_{MO}$ (left) and $HIPAM_{IMO}$ (right).

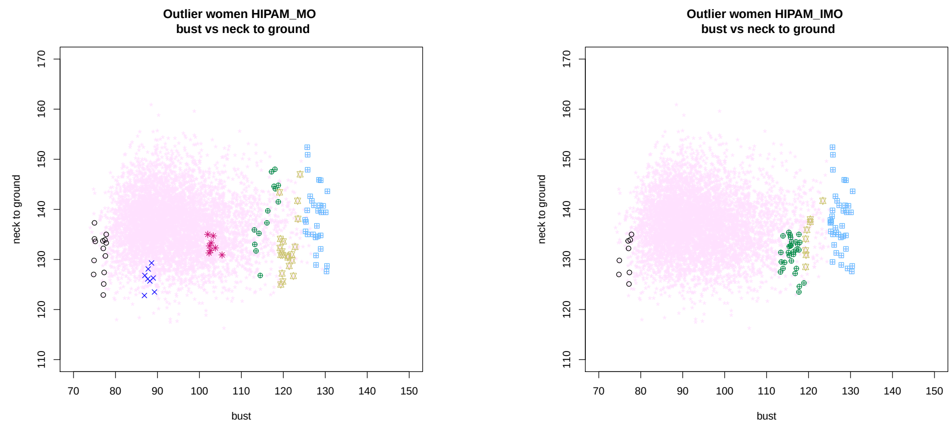

Figure 3. Bust vs neck to ground in the outliers obtained using $HIPAM_{MO}$ (left) and $HIPAM_{IMO}$ (right).

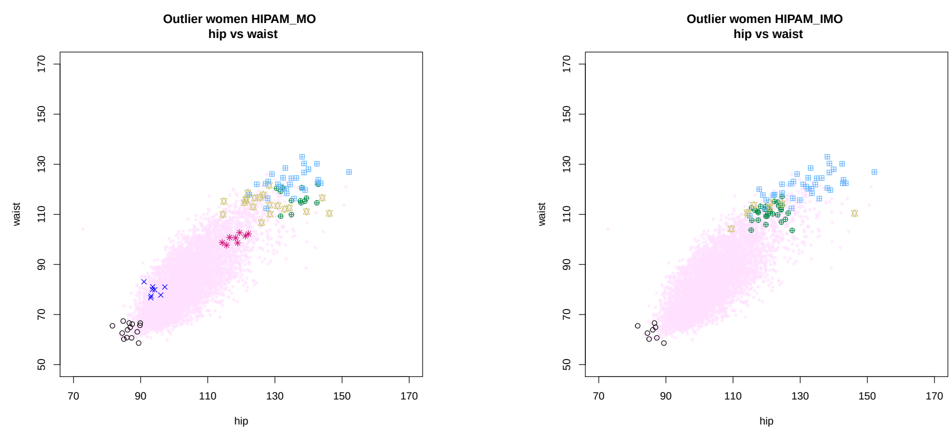

Figure 4. Hip vs waist in the outliers obtained using $HIPAM_{MO}$ (left) and $HIPAM_{IMO}$ (right).

Let us now compare the outliers obtained with both $HIPAM_{MO}$ and $HIPAM_{IMO}$ algorithms with those obtained by a quite common method in the apparel sizing literature: confidence ellipses (see for instance [7]). This method consists in determining a confidence curve: an ellipse containing a specified percentage of the data, assuming a bivariate normal distribution for height and weight. Under bivariate normality, the percentage of elements falling inside the ellipse should closely agree with the fixed confidence level. Data falling outside the ellipse are considered as abnormal.

We have obtained confidence ellipses at a 99% confidence level for each bust segment by means of the R function *dataEllipse* of the *car* package [4]. As an example, figure 5 shows the outliers returned for the bust segment $[78, 82[$ cm.

The labels of the outliers mark which women are identified as outliers. Height is measured in mm. and weight in kg. We note that with this kind of procedure, a woman is already considered as an outlier if she deviates substantially from the mean of only one of those variables. In addition, figure 6 shows the scatter plot of bust circumference against hip jointly with the outliers obtained from the normality ellipse in each bust class. As more relevant results, we appreciate that there are no outliers in the segments $[74, 78[$, $[119, 125[$ and $[125, 131[$ cms, which are the most extreme segments. However, both HIPAM algorithms find the outliers precisely in these segments because they contain the most unusual bust circumferences. In addition, it can be seen in figure 6 that there are a lot of women that are coded as outliers according to the normality ellipses built in each segment. Nevertheless, they are not actually outliers because they are clearly located over the point cloud.

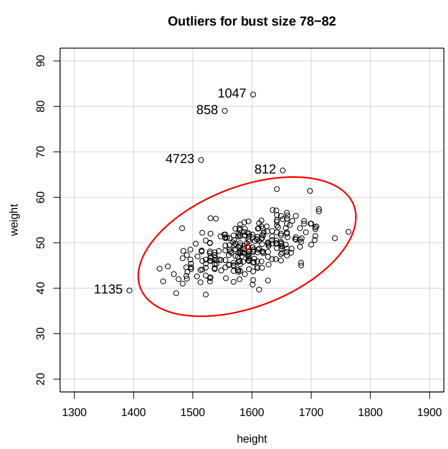

Figure 5. Outlier women returned by the normality ellipse for the segment [78,82[cm.

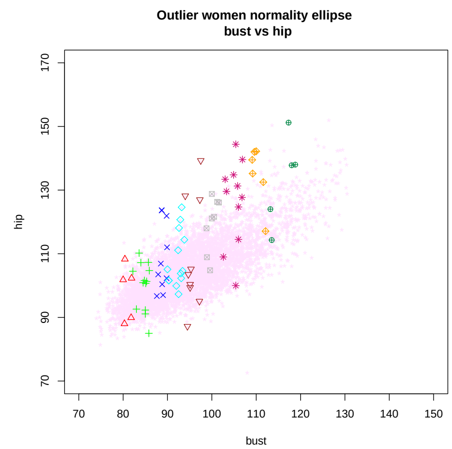

Figure 6. Bust vs hip in the outliers obtained using the normality ellipse.

In addition to the previous comparison, we have done a survey of the R packages which are implemented to detect multivariate outliers and next we will compare their results with those obtained with our proposal. As far as we know, the multivariate outlier detection methods implemented as R packages are based on robust methods. An example is the *mvoutlier* package [5], which is the implementation in R of the methodology developed in ref. [3]. Ref. [3] proposes to discover outliers by using the mcd estimator to make the mahalanobis distance robust, which is very sensitive to the presence of outliers. We are using a particular distance adapted to our data (see Section *2.1. The dissimilarity measure* in our original paper). Therefore, in order to correctly compare the outliers returned by the HIPAM algorithms with the outliers calculated by the *mvoutlier* package, the Mahalanobis distance should be replaced by our dissimilarity measure. However, given the implementation of this R package, this is not possible, which indicates to us that this kind of methods implemented are not flexible and cannot be used for all types of data. Hence, any HIPAM algorithm can be considered as a very interesting method to detect outliers regardless of the type of data. The usual plot to represent outliers with the *mvoutlier* R package is shown in figure 7. In particular, this represents the women for the bust class [78,82[cms, where the outliers are marked with red circles. The *mvoutliers* package does not return the same results each time its methodology is applied, although they are very similar. On the contrary, the HIPAM algorithm always identifies the same clustering results. This property of the HIPAM algorithm is called

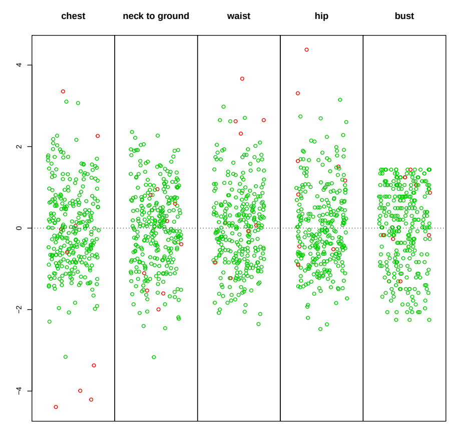

Figure 7. Outlier women returned by the *mvoutlier* package of R for the segment [78,82[cm.

reproducibility and ensures that the returned clusters really exist and are not merely the result of a certain random solution. Figure 8 shows the outliers calculated with the *mvoutliers* package for every bust segment. As can be seen, there are a lot of points that are clearly not atypical data. Therefore, after this analysis, we can state that both this R package and the application of the procedure explained in ref. [7], are overestimating the number of outliers in each bust class with respect to both $HIPAM_{MO}$ and $HIPAM_{IMO}$ algorithms, which are based on hierarchical features to discover outliers and return true outliers.

Next, we add some extra sections that do not appear in our original paper. On the one hand, we carry out an anthropometric analysis of the outliers obtained by $HIPAM_{MO}$ and $HIPAM_{IMO}$. On the other hand, we check whether any HIPAM algorithm can be considered as an alternative to the current European Normative to sizing system [1] for defining an apparel sizing system.

4.2.4. Anthropometric analysis of the outliers

Table 1 displays the answers for the outlier women provided by the $HIPAM_{MO}$ and $HIPAM_{IMO}$ algorithms, to the question about whether they have had problems in finding their correct size.

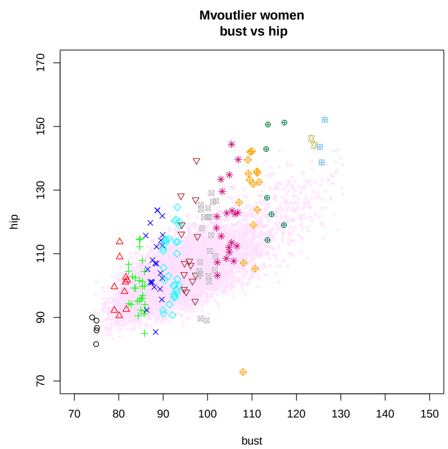

Figure 8. Bust vs hip in the outliers obtained using the R package *mvoutlier*.

Algorithm \ Answer	Never	Almost never	Yes, sometimes	Yes, always
$HIPAM_{MO}$	16	7	27	42
$HIPAM_{IMO}$	14	11	23	34

Table 1. Outlier women problems in finding their correct size.

As can be seen in table 1, most of the outlier women answer that they sometimes or always have had problems in finding their correct size. In other words, there are few outlier women that claim never to have had problems in finding their correct size. Following the suggestions of the experts, there are several possible hypotheses to explain the reasons why they have had problems in finding their size:

1. The most probable reason might be that they are overweight. They fall outside the usual sizing range. Manufacturers usually offer the most common sizes, that is to say, those that will be sent in greater quantities to reduce unsold stock. Following this trend, extremely thin women will also have problems in finding their size, but these cases will be a minority as the population tends to overweight.
2. Another explanation could be that this kind of woman, despite not being overweight, are women with a less frequent morphotype, with a narrow waist or wide hip. In this case, the women would find their

clothes size, but items would not fit correctly and they would need to get the clothes altered.

3. Finally, these women could fall between sizes, the difference usually being around 5 cms bust or waist circumference depending on whether upper or lower body garments. It is possible that these 5 cms for certain garments are being considered a big difference and therefore they would fall between sizes. In a real situation, since each manufacturer has his own sizing systems, when a woman does not fit one brand, it is probable that she will find her size in another store.

We evaluate these hypotheses for the outliers returned by $HIPAM_{MO}$ and $HIPAM_{IMO}$:

- HYPOTHESIS 1: With respect to the 92 outliers obtained by the $HIPAM_{MO}$ algorithm, 70 of them are overweight ($BMI > 25$), 12 are underweight ($BMI < 18.5$) and 10 are normal weight ($18.5 \leq BMI < 24.99$). With respect to the 82 outliers obtained by $HIPAM_{IMO}$, 74 are overweight, 7 are underweight and there is only one woman who is normal weight (BMI is the Body Mass Index). Therefore, with both algorithms we see, that the majority of the outliers are overweight, while underweight women are the minority. Indeed, the outlier population tends to be overweight.

- HYPOTHESIS 2: We analyse the overall body shape of the outlier women with normal weight with both algorithms. To that end, we use the *drop value* [6, 2]. To identify the morphotypes, a parameter called drop value is used, which is defined as the difference between a woman's hip circumference and bust circumference. Drop values allow us to identify different relationships between key anthropometric dimensions that determine body shape. The population can be classified into the following categories, as can be read in ref. [6]:

  - Triangular or pear-shaped (bust much smaller than hip).
  - Inverted triangle (bust much bigger than hip).
  - Rectangular (bust equal to hip).
  - The other categories lie in between these.

Table 2 details hip and bust circumferences and consequently the drop value of the 10 outlier women that according to hypothesis 1, are normal weight with $HIPAM_{MO}$. With $HIPAM_{IMO}$, only one woman was normal weight and in addition, she is one of the 10 obtained by $HIPAM_{MO}$.

Woman Code	BARNA066	ELGOI023	MELGA066	CAMB037	ABAD071	MALAG131	CAMAR142	SILLE148	MADY165	MADY229
Hip measurement	89.4	87.5	96.0	94.2	93.6	91.0	93.1	97.2	93.5	93.0
Bust measurement	77.2	77.1	87.7	88.2	89.0	87.8	86.9	87.0	88.6	89.3
Drop value	12.2	10.4	8.3	6.0	4.6	3.2	6.2	10.2	4.9	3.7

Table 2. Drop values for the outlier women with a normal weight for $HIPAM_{MO}$.

According to the hip and bust measurements and therefore to the drop values shown in table 2, BARNA066, ELGOI023, MELGA066, CAMB037, CAMAR142 and SILLE148 can be considered pear shaped. The other women are rectangle shaped because they have very similar hip and bust measurements. The only woman with a normal weight obtained with $HIPAM_{IMO}$ is BARNA066.

- HYPOTHESIS 3: We analyse the four women with a rectangular shape (ABAD071, MALAG131, MADY165 and MADY229).

  Table 3 presents bust and waist circumferences for the outlier women with a rectangular shape. Table 4 summarizes the first seven bust and waist segments defined by the European Normative to sizing system [1].

Woman Code	ABAD071	MALAG131	MADY165	MADY229
Bust measurement	89.0	87.8	88.6	89.3
Waist measure ment	81.1	83.1	80.0	76.7

Table 3. Bust and waist measurements for the outlier women with a rectangular shape.

Bust Circumference	76	80	84	88	92	96	100	...
Range	74-78	78-82	82-86	86-90	90-94	94-98	98-102	...
Waist Circumference	60	64	68	72	76	80	84	...
Range	58-62	62-66	66-70	70-74	74-78	78-82	82-86	...

Table 4. First seven bust and waist segments defined by the European Normative to sizing system [1].

As we see in table 3, ABAD071 and MADY229 may be considered women between two bust sizes (88 and 92) when comparing their bust measurement with the sizes detailed in table 4. On the other hand, for MALAG131 and MADY165 is not so easy to state that they are located between two sizes for their bust or waist dimensions.

Summarizing, we have carried out an exploratory anthropometric analysis of the outlier women obtained by both HIPAM algorithms. In particular, we have checked that most of the outlier women identified with both algorithms who state that they always have problems in finding their size, verify the hypotheses that can explain the reasons for their problems.

4.2.5. HIPAM as an alternative to current sizing system Normatives

The core of our approach in the original paper has been to use both $HIPAM_{MO}$ and $HIPAM_{IMO}$ within twelve particular segments, defined taking into account bust circumference values according to the sizes described in the European Normative to sizing system [1]. We would like to check now which clustering results any HIPAM algorithm obtains when applied it to the whole data set without presegmenting in any group. We want to examine whether the clusters returned might be considered a good approximation of the sizes defined by the European Normative to sizing system. After applying both algorithms, $HIPAM_{MO}$ returned ten clusters, while $HIPAM_{IMO}$ returned three. So we concluded in this case that $HIPAM_{MO}$ could be more useful to that objective. Figure 9 shows the clustering results obtained by $HIPAM_{MO}$ when applying it to the whole data set without presegmenting. Table 5 details the key anthropometric measurements of the cluster medoids obtained by $HIPAM_{MO}$ ordered by bust circumference in an increasing order. Table 6 shows the basic descriptives for bust dimension for each one of these ten clusters calculated by $HIPAM_{MO}$. On the other hand, table 7 shows the basic descriptives for bust dimension for each one of the twelve sizes defined by the European Normative to sizing system. When comparing both tables, we realize that the clusters obtained by $HIPAM_{MO}$ when applying it to the whole data base are somewhat different with respect to the sizes defined by the current Normative. However, we also appreciate that these clusters returned by $HIPAM_{MO}$ are quite different to each other, except for clusters labelled as 9 and 10, which are very similar. In spite of the fact that the bust measurement of the medoid in cluster 10 is greater than the corresponding one for the medoid in cluster 9 (see table 5), all the descriptives in cluster 9 are greater than those in cluster 10, (see table 6). As a conclusion, although we have seen that $HIPAM_{MO}$ applied to the whole database, does not approximate well to the sizes defined by the current European Normative to sizing system, we have proved that any HIPAM algorithm does serve to segment a huge data base in different groups according to the most relevant anthropometric dimensions.

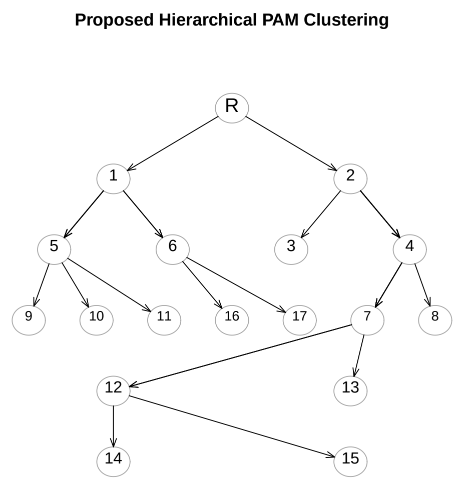

Figure 9. Clustering results returned by $HIPAM_{MO}$ applied to the whole data base.

Label medoid	chest	neck to ground	waist	hip	bust	cluster label
869	84.55	133.20	66.70	89.30	79.50	13
150	86.53	135.10	69.10	94.60	82.30	14
232	88.42	134.90	73.20	96.00	84.60	15
302	91.36	134.80	73.10	97.40	86.30	8
5503	98.68	140.90	86.50	105.70	94.00	3
1120	103.24	130.90	90.40	104.10	99.60	9
1506	102.35	141.60	88.70	114.20	100.00	10
3988	109.26	136.30	100.30	110.20	106.30	11
431	113.83	131.00	105.10	117.40	110.50	16
501	120.86	141.20	112.30	128.30	115.50	17

Table 5. Cluster medoids returned by $HIPAM_{MO}$ applied to the whole data base ordered by bust circumference value.

	13	14	15	8	3	9	10	11	16	17
Min.	74.00	74.90	80.00	79.10	81.00	92.50	88.50	97.70	97.40	103.30
1st Qu.	79.20	82.90	84.70	87.80	91.90	99.00	97.93	104.10	110.00	115.60
Median	80.80	84.60	86.50	89.80	94.50	100.40	99.80	106.40	113.60	120.40
Mean	80.96	84.51	86.32	89.38	93.84	100.50	99.60	106.70	113.50	119.90
3rd Qu.	82.98	86.50	88.20	91.40	96.30	102.20	102.00	109.00	117.70	125.80
Max.	86.60	89.60	90.60	95.10	108.00	108.50	106.70	117.00	126.10	130.50
Number women	350	574	396	1106	1389	510	350	575	366	111

Table 6. Basic descriptives for bust dimension for each one of the ten clusters returned by $HIPAM_{MO}$ applied to the whole data base.

	[74,78[	[78,82[	[82,86[	[86,90[	[90,94[	[94,98[	[98,102[	[102,107[	[107,113[	[113,119[	[119,125[	[125,131[
Min	74.00	78.00	82.00	86.00	90.00	94.00	98.00	102.00	107.00	113.00	119.00	125.20
1st Qu.	76.30	79.55	83.30	87.00	90.90	94.90	98.70	103.00	108.30	113.80	119.80	125.80
Median	77.00	80.50	84.20	88.00	91.80	95.80	99.70	104.20	109.60	115.40	121.20	127.30
Mean	76.64	80.39	84.13	87.97	91.89	95.88	99.74	104.20	109.70	115.60	121.40	127.40
3rd Qu.	77.60	81.30	85.10	89.00	92.80	96.90	100.80	105.40	111.10	117.00	122.70	128.90
Max.	77.90	81.90	85.90	89.90	93.90	97.90	101.90	106.90	112.90	118.90	124.90	130.50
Number women	47	287	732	1028	952	818	633	547	356	203	87	37

Table 7. Basic descriptives for bust dimension for the first twelve bust sizes defined by the European Normative to sizing system.

References

[1] European Committee for Standardization, 2002. European Standard EN 13402-2: Size system of clothing. Primary and secondary dimensions.

[2] Fan, J., Yu, W., Hunter, L., 2004. Clothing appearance and fit: Science and technology. Woodhead Publishing in Textiles.

[3] Filzmoser, P., Garrett, R. G., Reimann, C., 2005. Multivariate outlier detection in exploration geochemistry. Computers & Geosciences 31, 579–587.

[4] Fox, J., Weisberg, S., 2011. An R Companion to Applied Regression, 2nd Edition. Sage, Thousand Oaks CA.
URL `http://socserv.socsci.mcmaster.ca/jfox/Books/Companion`

[5] Gschwandtner, M., Filzmoser, P., 2011. mvoutlier: Multivariate outlier detection based on robust methods. R package version 1.8.
URL `http://CRAN.R-project.org/package=mvoutlier`

[6] Gupta, D., Gangadhar, B., 2004. A statistical model for developing body size charts for garments. International Journal of Clothing Science and Technology 16 (5), 458–469.

[7] Hsu, C.-H., 2009. Data mining to improve industrial standards and enhance production and marketing: An empirical study in apparel industry. Expert Systems with Applications 36, 4185–4191.