

*Desarrollo de metodologías estadísticas aplicadas
a datos antropométricos y orientadas al diseño
ergonómico del producto*



Guillermo Vinué Visús
Facultad de Matemáticas
Universidad de Valencia
www.uv.es/vivigui/



Defensa de la Tesis Doctoral

Dirigida por:

M^a Teresa León Mendoza Amelia Simó Vidal Irene Epifanio López

Asociada al proyecto de investigación "Aplicación de técnicas
morfométricas al diseño y evaluación funcional de indumentaria", dirigido
por Guillermo Ayala Gallego

Programa de Doctorado en Estadística y Optimización

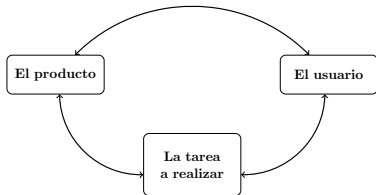
Mayo de 2014, Burjassot, España

Esquema de la presentación

- 1 **Introducción**
 - Ergonomía, Antropometría y toma de medidas
- 2 **Estudio antropométrico de la población femenina en España**
 - Puntos anatómicos: landmarks
- 3 **Casos antropométricos**
- 4 **Objetivos**
- 5 **Metodologías estadísticas**
 - Clustering
 - Análisis estadístico de formas
 - Profundidad estadística
 - Análisis estadístico de arquetipos y arquetipoides
 - Paquete de R Anthropometry
- 6 **Conclusiones**
- 7 **Bibliografía básica**

Ergonomía, Antropometría y toma de medidas I

- La **Ergonomía** es el proceso de adaptar el producto al usuario.
- Ejemplos: diseño de ropa y calzado, diseño de entornos laborales.
- Para desarrollar un producto se debe tener en cuenta las medidas antropométricas de las personas a las que va dirigido.
- Se basa en el *principio del diseño centrado en el usuario*.
- El objetivo es lograr el mejor ajuste posible entre el producto y sus usuarios, teniendo en cuenta la tarea que van a realizar.



Ergonomía, Antropometría y toma de medidas II

- La **Antropometría** estudia las medidas y dimensiones del cuerpo humano.
- Se le considera una rama muy importante de la Ergonomía.
- Las medidas corporales han sido habitualmente tomadas utilizando procedimientos rudimentarios como reglas o cintas métricas.

* Ventajas:

- Simples.
- No invasivos.
- Baratos.



* Desventajas:

- Medidas poco precisas e inexactas.
- Proceso largo y laborioso.
- Modernos sistemas de escaneo 3D permiten obtener datos antropométricos de gran calidad, exactos y reproducibles.
- Nuevos estudios antropométricos en diferentes países (p. ej., España).

Estudio antropom. de la población femenina en España

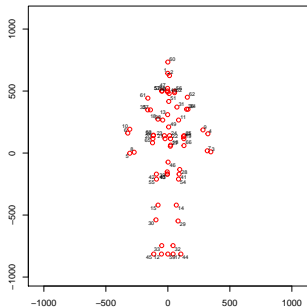
- Fue promovido por el Ministerio de Sanidad y Consumo en 2006 y realizado por el Instituto de Biomecánica de Valencia (IBV).



- Objetivos:
 - * Caracterizar la forma de la actual población de mujeres españolas.
 - * Fomentar una imagen de belleza saludable.
- Base de datos: 10415 mujeres españolas elegidas aleatoriamente:
 - * Entre 12 y 70 años de edad.
 - * 95 medidas antropométricas.
 - * 66 **puntos anatómicos** para representar su forma.
 - * Encuesta sociológica.

Puntos anatómicos: landmarks

- La forma de todas las mujeres de la base de datos está representada por unos marcadores anatómicos llamados landmarks.
- Landmark: punto (x, y, z) de correspondencia en cada individuo que coincide entre poblaciones y dentro de cada una de ellas.
- Se llama configuración al conjunto de landmarks $\rightarrow X \in \mathcal{M}_{66 \times 3}(\mathbb{R})$

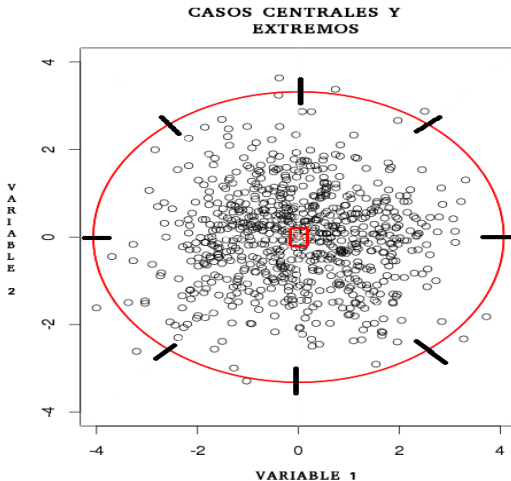


Landmark	Descripción
1. Parte trasera de la cabeza	Punto más prominente de la cabeza en el plano sagital
2. Frente (de la cabeza)	Glabela (punto más prominente de la frente)
3. Muñeca del antebrazo izquierdo	Perímetro máximo del antebrazo izquierdo
4. Perímetro del antebrazo izquierdo	Perímetro máximo del antebrazo izquierdo justo debajo del codo
5. Muñeca del antebrazo derecho	Perímetro máximo del antebrazo derecho
.....
66. Cresta ilíaca izquierda	Marcador físico en la izquierda de la cresta ilíaca

Casos antropométricos I

- La variabilidad antropométrica de la población se caracteriza mediante unos pocos casos.
- Un **caso antropométrico** representa las medidas corporales que el diseñador quiere acomodar en el producto.
- Según las propiedades del producto, se distinguen tres tipos de casos:
 - * **Centrales.**
 - * **Extremos (*boundaries*).**
 - * **Distribuidos.**
- Si queremos n tallas para acomodar a n grupos dentro de una población dado un % de acomodación, buscaremos casos centrales → Ej.: diseño de ropa.
- Si queremos una talla para acomodar a toda una población dado un % de acomodación, buscaremos casos extremos → Ej.: diseño de entornos de trabajo.

Casos antropométricos II



Objetivos

- **Objetivos en el diseño de ropa:**
 - * Desarrollar metodologías estadísticas para obtener prototipos y modelos de ajuste (casos centrales) que permitan definir un sistema de tallaje eficiente y fabricar maniqués representativos.
 - * Agrupar personas de acuerdo a sus dimensiones antropométricas y a su forma corporal representada por landmarks.
- **Objetivos en el diseño de la cabina de una aeronave:**
 - * Desarrollar metodologías estadísticas para identificar pilotos con características extremas que permitan la acomodación de todos los demás.
 - * Aumentar la seguridad, eficacia y comodidad de las acciones a realizar por parte de los pilotos.

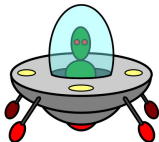
Metodologías estadísticas

- **Sistemas de tallaje y diseño de ropa (casos centrales)**



- * Clustering.
- * Análisis estadístico de formas.
- * Profundidad estadística.

- **Problema de acomodación en la cabina de una aeronave (casos extremos)**



- * Análisis estadístico de arquetipos y arquetipoides.

- **Paquete de R**



Sistemas de tallaje y diseño de ropa (casos centrales)



Sistemas de tallaje de ropa

Definiciones

- Un sistema de tallaje de ropa divide a la población en grupos homogéneos utilizando un conjunto de variables antropométricas control, de manera que las personas de un mismo grupo puedan llevar la misma prenda.
- Un sistema de tallaje **óptimo y eficiente** debe acomodar el mayor porcentaje posible de la población utilizando el mínimo número posible de tallas que mejor describan la variabilidad corporal. Además, el ajuste de las prendas debe ser el mejor posible.
- Cada talla de ropa se define a partir de una persona **central**, llamada **prototipo** (representante de la talla).
- Una vez se ha diseñado una prenda, se utilizan **modelos de ajuste** (*fit models*) para evaluar su ajuste y comodidad (también **casos centrales**).
- Los sistemas de tallaje acomodan a la población estándar, eliminando las personas atípicas (*outliers*).

Sistemas de tallaje de ropa

Enfoque matemático-estadístico

- El clustering es la herramienta estadística que particiona un conjunto de datos en grupos homogéneos (clusters).
- Cada cluster se representa por una observación central.
- Proponemos tres métodos basados específicamente en clustering:
 - * *trimowa*: prototipos & ropa parte superior (variable control: busto).
 - * *biclustAnthropom*: prototipos & ropa parte inferior (cintura).
 - * *hipamAnthropom*: modelos de ajuste & ropa parte superior (busto).

Busto	74-78	78-82	82-86	86-90	90-94	94-98
Cintura	58-62	62-66	66-70	70-74	74-78	78-82
Busto	98-102	102-107	107-113	113-119	119-125	125-131
Cintura	82-86	86-91	91-97	97-103	103-109	109-115

Tallas de busto y cintura según la Normativa Europea UNE 13402-3-2004.

Diseño de ropa: casos centrales

Clustering: *trimowa*, *biclustAnthropom* y *hipamAnthropom*

1^{er} trabajo: *trimowa*

Antecedentes y propuesta

- Enfoque tradicional: tablas de tallaje bidimensionales.
 - * No consideran la variabilidad de otras dimensiones importantes.
- Métodos matemáticos y de optimización.
- Trabajo de **McCulloch et al. (1998)**:
 - * Define un sistema de tallaje eficiente mediante un algoritmo de optimización.
 - * Define k prototipos, uno por talla.
- **Nuestra propuesta**:
 - * Definir un sistema de tallaje eficiente mediante clustering.
 - * Definir k prototipos, uno por talla, más realistas.
 - * Eliminar una proporción de los datos.
 - * Combinar el desajuste en cada dimensión antropométrica mediante operadores OWA.

Trimowa

Métodos: trimmed k -medoids (o trimmed PAM)

- Notación:
 - * $x_i = (x_{i1}, \dots, x_{ip})$: p medidas del individuo i -ésimo ($i = 1, \dots, n$).

- Método de clustering que utilizamos: **trimmed k -medoids (o trimmed PAM)**:

$$\{x_{i_1}^*, \dots, x_{i_k}^*\} = \underset{\mathbf{Y}, x_{i_1}, \dots, x_{i_k}}{\operatorname{argmin}} \frac{1}{\lceil n(1 - \alpha) \rceil} \sum_{x_j \in \mathbf{Y}} \inf_{1 \leq j \leq k} d(x_i, x_{i_j})$$

donde \mathbf{Y} se refiere a los subconjuntos de x_1, \dots, x_n que contienen $\lceil n(1 - \alpha) \rceil$ puntos y $\lceil \cdot \rceil$ denota la parte entera de un número cualquiera.

- Propiedades:
 - * Los medoides $x_{i_j}^*$ ($j = 1, \dots, k$) son individuos del banco de datos.
 - * Elimina una proporción α de los datos (procedimiento *trimmed*).
 - * Es más robusto frente a los outliers.

Trimowa

Métodos: disimilaridad utilizada I

- Una función distancia permite expresar la idea del ajuste de una prenda.
- $d_i(x_i, y_i)$ mide el desajuste entre $x = (x_1, \dots, x_p)$ e $y = (y_1, \dots, y_p)$ en la variable i ($i = 1, \dots, p$), (y es el prototipo).
 - * **Nota:** x_i denota aquí la medida del individuo x en la variable i .
- Para definir la función distancia utilizada, nos basamos en las ideas de McCulloch et al.:
 - 1 El ajuste se predice mejor mediante diferencias proporcionales entre x_i e $y_i \rightarrow$ log-transformación.
 - 2 Hay un intervalo donde no hay diferencia entre x_i e y_i .
 - 3 La función distancia no es simétrica \rightarrow **disimilaridad**.
 - 4 Las diferencias en ciertas dimensiones influyen más sobre el ajuste que otras.

Trimowa

Métodos: disimilaridad utilizada II

- Elegimos la función propuesta por McCulloch et al.:

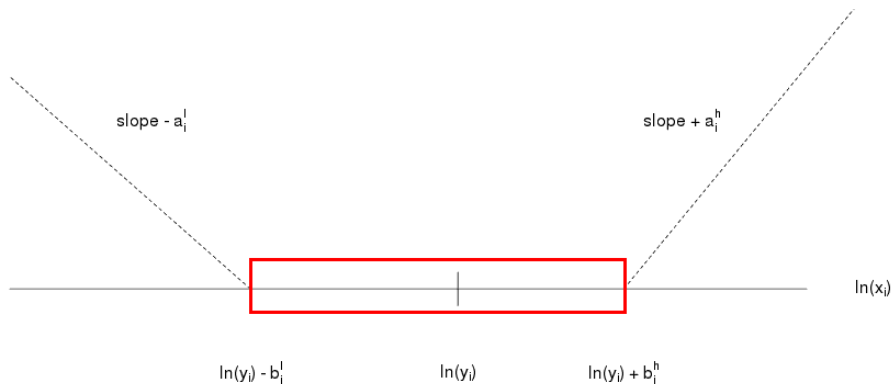
$$d_i(x_i, y_i) = \begin{cases} a_i^l(\ln(y_i) - b_i^l - \ln(x_i)) & \text{if } \ln(x_i) < \ln(y_i) - b_i^l \\ 0 & \text{if } \ln(y_i) - b_i^l \leq \ln(x_i) \leq \ln(y_i) + b_i^h \\ a_i^h(\ln(x_i) - b_i^h - \ln(y_i)) & \text{if } \ln(x_i) > \ln(y_i) + b_i^h \end{cases}$$

donde a_i^l , b_i^l , a_i^h y b_i^h son constantes para cada dimensión i :

- b_i : rango en el cual el ajuste es perfecto.
- a_i : índice en el que el ajuste se deteriora fuera del rango.

Trimowa

Métodos: disimilaridad utilizada III



Trimowa

Métodos: disimilaridad utilizada IV

- **Disimilaridad global de McCulloch et al.:** $d(x, y) = \sum_{i=1}^p (d_i(x_i, y_i))^2$
- Estamos *agregando* las diferentes disimilaridades $d_i(x_i, y_i)$.
- Incorporamos en $d(x, y)$ el operador OWA (Ordered Weighted Average).
- Operador OWA de dimensión n : $f : \mathbb{R}^n \rightarrow \mathbb{R}$ con:
 - * $W = (w_1, \dots, w_n)$ t.q. $\sum_{j=1}^n w_j = 1$.
 - * $f(a_1, \dots, a_n) = \sum_{j=1}^n w_j b_j$.
 - * b_j es el j -elemento más grande de los a_1, \dots, a_n agregados.
- **Disimilaridad global de trimowa:** $d(x, y) = \sum_{i=1}^p w_i (d_i(x_i, y_i))^2$
- La mayor discrepancia recibe el mayor peso, la segunda mayor discrepancia recibe el segundo mayor peso, etc. para las p medidas.
- Se tiene en cuenta al usuario dando mayor peso a las medidas en las que el desajuste es más importante.

Trimowa

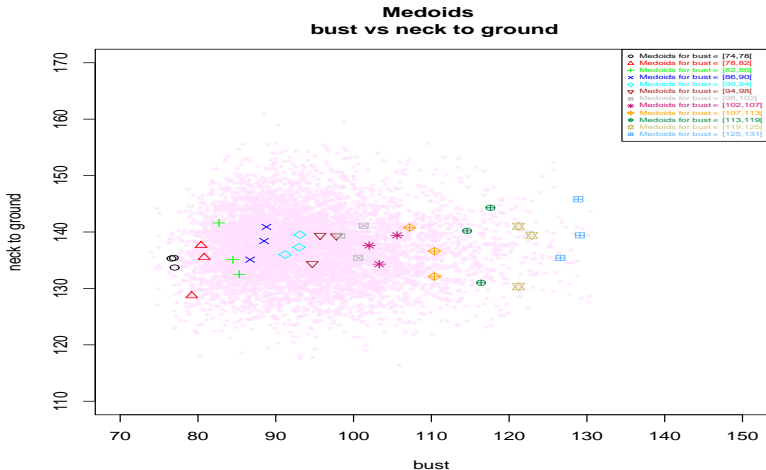
Datos y procedimiento

- Selección de 6013 mujeres:
 - * No embarazadas.
 - * No lactantes.
 - * Sin cirugía estética.
 - * Entre 20 y 65 años.
- Variable control primaria: busto.
- Variables control secundarias: pecho, cadera, cintura y nuca-tierra.
- Procedimiento:
 - 1 Los datos se segmentan en 12 tallas de busto según la Norma UNE.
 - 2 En cada talla, se aplica el *trimmed k-medoids* (o *trimmed PAM*) con $k = 3$.

Busto	74-78	78-82	82-86	86-90	90-94	94-98
Busto	98-102	102-107	107-113	113-119	119-125	125-131

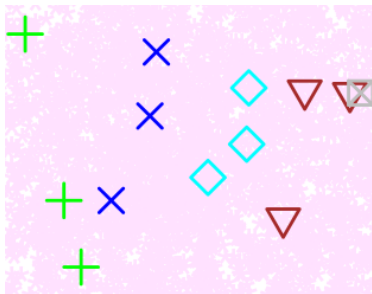
Trimowa

Resultados I: medoides busto vs nuca-tierra



Trimowa

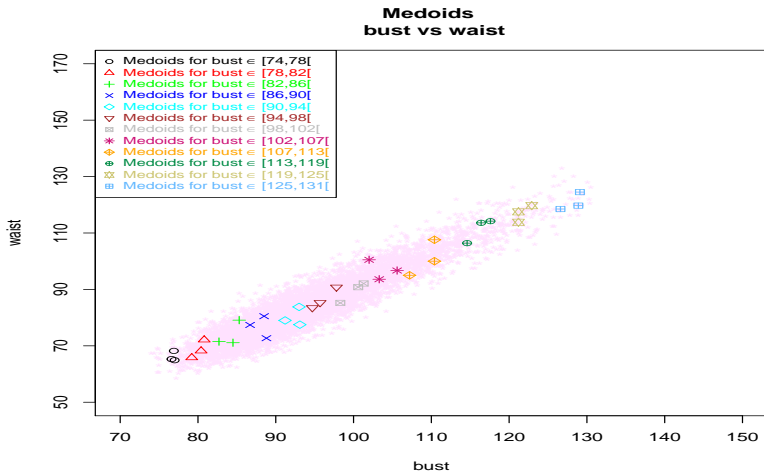
Resultados II: medoides busto vs nuca-tierra



- Rango [82,86[(cruces verdes): tres tallas (los medoides están distribuidos).
- Rango [94,98[(triángulos marrones invertidos): sería más conveniente dos tallas (dos medoides tienen una longitud nuca-tierra similar).

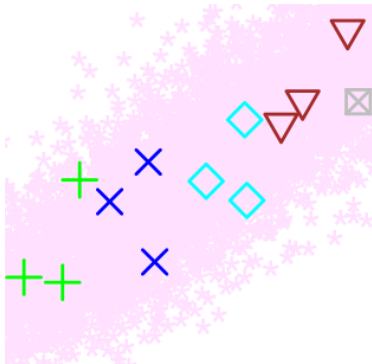
Trimowa

Resultados III: medoides busto vs cintura



Trimowa

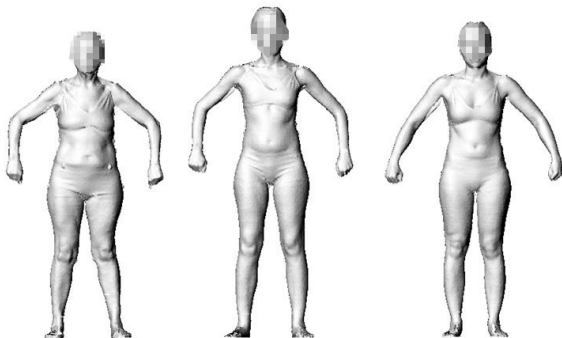
Resultados IV: medoides busto vs cintura



- Rango [82,86[(cruces verdes): dos tallas (dos medoides tienen una cintura similar).
- Rango [94,98[(triángulos marrones invertidos): tres tallas.

Trimowa

Resultados V: prototipos talla de busto [82,86[

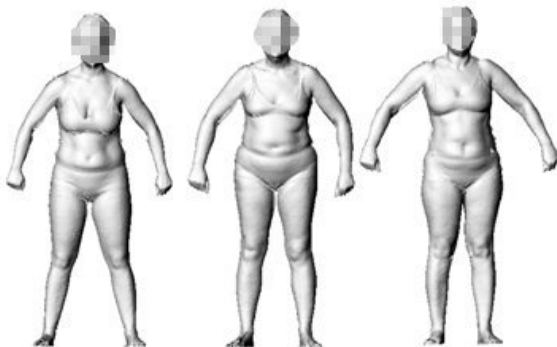


Izquierda a derecha:

- CANDE021
- SEVI132
- LLEID074

Trimowa

Resultados VI: prototipos talla de busto [94,98[



Izquierda a derecha:

- SILLE034
- JAEN075
- CANDE068

Trimowa Conclusiones

- Para diseñar una prenda de ropa se debe tener en cuenta el aspecto multivariante del problema.
- *Trimowa* permite, de manera simultánea:
 - * asignar (acomodar) cada individuo a una talla.
 - * seleccionar prototipos.
 - * descartar personas atípicas (no acomodadas).
- Los prototipos corresponden con personas reales.
- La disimilaridad utilizada describe de modo más realista las diferencias entre los individuos y los prototipos.



Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa



Apparel sizing using trimmed PAM and OWA operators

M.V. Ibáñez^{a,*}, G. Vinué^b, S. Alemany^c, A. Simó^a, I. Epifanio^a, J. Domingo^d, G. Ayala^b

^a Department of Mathematics, Universitat Jaume I, Castellón, Spain

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

^c Biomechanics Institute of Valencia, Universidad Politécnica de Valencia, Valencia, Spain

^d Department of Informatics, University of Valencia, Valencia, Spain

ARTICLE INFO

Keywords:

Anthropometric data
Sizing systems
Trimmed k-medoids
OWA operators

ABSTRACT

This paper is concerned with apparel sizing system design. One of the most important issues in the apparel development process is to define a sizing system that provides a good fit to the majority of the population. A sizing system classifies a specific population into homogeneous subgroups based on some key body dimensions. Standard sizing systems range linearly from very small to very large. However, anthropometric measures do not grow linearly with size, so they can not accommodate all body types. It is important to determine each class in the sizing system based on a real prototype that is as representative as possible of each class. In this paper we propose a methodology to develop an efficient apparel sizing system based on clustering techniques jointly with OWA operators. Our approach is a natural extension and improvement of the methodology proposed by McCulloch, Paal, and Ashdown (1998), and we apply it to the anthropometric database obtained from an anthropometric survey of the Spanish female population, performed during 2006.

© 2012 Elsevier Ltd. All rights reserved.

2º trabajo: *biclustAnthropom*

Antecedentes y propuesta I

- **Clustering convencional:** agrupa por observaciones (filas) utilizando todas las variables (columnas) y viceversa.
- En algunos contextos, un subconjunto de observaciones muestra un comportamiento (patrón) similar sólo en un subconjunto de variables.
- El clustering no es capaz de detectar estos patrones → **biclustering**.
- **Biclustering:** encuentra subgrupos de filas y columnas (biclusters), lo más parecidos entre sí y lo más diferentes al resto.
- En un bicluster:
 - * las filas se agrupan si se expresan igual en un subconjunto de columnas.
 - * las columnas se agrupan si incluyen filas que se expresan igual.

BiclustAnthropom

Antecedentes y propuesta II

- Por ejemplo, la fila r_3 se define por las columnas c_1 y c_2 , la fila r_4 por c_2 y c_3 , mientras que las columnas c_1 , c_2 y c_3 se definen por las filas r_1 , r_2 y r_5 .

	c_1	c_2	c_3	c_4	c_5	...
r_1	7	8	6	4	3	...
r_2	7	8	6	1	8	...
r_3	7	7	6	2	9	...
r_4	4	8	8	5	2	...
r_5	7	8	6	5	1	...
...

	c_1	c_2	c_3
r_1	7	8	6
r_2	7	8	6
r_3	7	7	6
r_4	4	8	8
r_5	7	8	6

BiclustAnthropom

Antecedentes y propuesta III

- Matriz $A = (X, Y)$, X filas, Y columnas.
 - * Bicluster: submatriz (I, J) de A con $I \subseteq X$ y $J \subseteq Y$ en la que a_{ij} representa la relación entre la fila i y la columna j .
- Nuestra propuesta:
 - * Utilizar el biclustering para encontrar el subconjunto de variables antropométricas que sean las más relevantes para diseñar una talla.
- En mi trabajo fin de Máster (dirigido por Guillermo Ayala):
 - * Revisamos algunos métodos biclustering.
 - * Analizamos su posible utilidad en la definición de tallas eficientes.
 - * Concluimos que el método de Cheng & Church (CC) ofrecía los resultados más interesantes.
- Proponemos un algoritmo utilizando CC para encontrar biclusters e individuos no acomodados, de manera simultánea.
- Software estadístico utilizado: paquete de R biclust.



BiclustAnthropom

Método de Cheng & Church (CC): definición

- $a_{iJ} = \frac{1}{|J|} \sum_{j \in J} a_{ij}$, $a_{Ij} = \frac{1}{|I|} \sum_{i \in I} a_{ij}$, $a_{IJ} = \frac{1}{|I||J|} \sum_{i \in I, j \in J} a_{ij}$
- Bicluster: filas y columnas con un índice de similaridad alto.
- Para estudiar la similaridad, CC utiliza el residuo cuadrado medio, H .
- CC busca el bicluster más grande con $H < \delta$, $\delta \geq 0$
- δ -bicluster perfecto: $a_{ij} = a_{iJ} + a_{Ij} - a_{IJ}$
- Por la presencia de ruido, los biclusters no son perfectos.
- Concepto de residuo: $r(a_{ij}) = a_{ij} - a_{iJ} - a_{Ij} + a_{IJ}$
- Residuo cuadrado medio: $H(I, J) = \frac{1}{|I||J|} \sum_{i \in I, j \in J} (r(a_{ij}))^2$
- Bicluster: submatriz tal que $H(I, J) < \delta$

BiclustAnthropom

Algoritmo propuesto

Para cada talla defino un objeto **nc**, un objeto **delta** y un objeto **disac**.

nc es el número propuesto de biclusters a encontrar en cada talla.

delta es el valor δ del algoritmo Cheng & Church. Inicialmente igual a 1.

disac es el número de mujeres que se van a quedar sin agrupar. Inicialmente igual al número de mujeres de la talla correspondiente.

La proporción de mujeres no acomodadas que queremos en cada talla la fijamos al 1%.

mientras $\text{disac} > \text{ceiling}(0.01 * \text{número de mujeres de la talla correspondiente})$

`biclust(datosTalla, method=BCCC(), delta=delta, alpha=1.5, number=nc)`

$\text{disac} = \text{número de mujeres que se han quedado sin agrupar.}$

$\text{delta} = \text{delta} + 1$

fin mientras

BiclustAnthropom

Datos y procedimiento

- Las prendas de ropa inferior dependen no sólo del contorno de cintura, sino de otras dimensiones.
- El biclustering encuentra grupos para un subconjunto de dimensiones.
- Se propone utilizar el biclustering para sistemas de tallaje de ropa inferior.
- Mismo conjunto de 6013 mujeres que en *trimowa*.
- Variable control primaria: cintura.
- Variables control secundarias: todas las de la parte inferior del cuerpo.
- Procedimiento:
 - 1 Los datos se segmentan en 12 tallas de cintura (Norma UNE).
 - 2 En cada talla, se aplica el *biclustAnthropom*.

Cintura	58-62	62-66	66-70	70-74	74-78	78-82
Cintura	82-86	86-91	91-97	97-103	103-109	109-115

BiclustAnthropom

Resultados I: comentarios generales

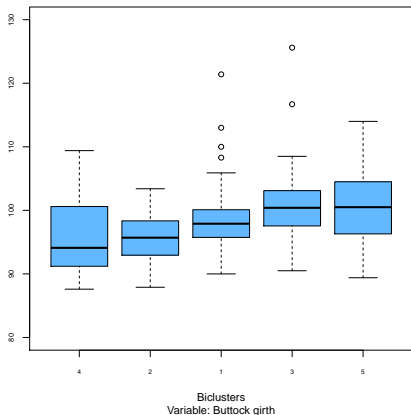
- Biclusters no definidos para el total de variables.
- Biclusters no solapados.
- Pocas mujeres desacomodadas.
- δ pequeños.
- Resultados reproducibles.

Resultados para la talla [74, 78[
Filas y columnas de los datos: 809 10					
Mujeres desacomodadas: 4					
Valor de δ : 3					
	BC 1	BC 2	BC 3	BC 4	BC 5
Número de filas:	388	119	136	93	69
Número de columnas:	10	10	10	8	8

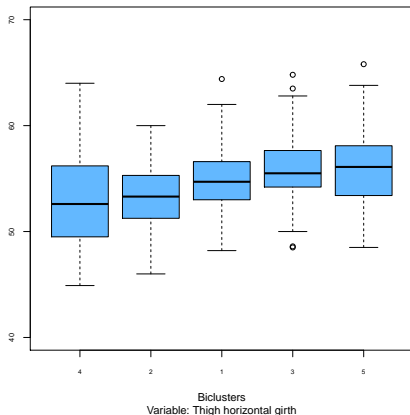
BiclustAnthropom

Resultados II: gráficos para la talla de cintura [74, 78[

Boxplots for each bicluster



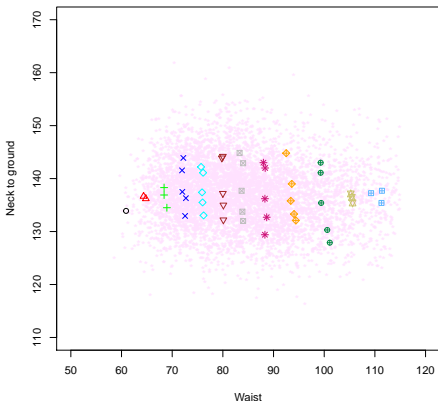
Boxplots for each bicluster



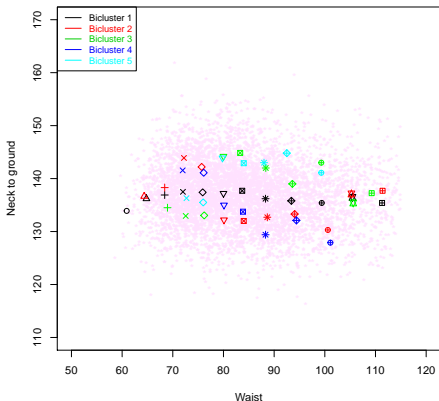
BiclustAnthropom

Resultados III: gráficos para todas las tallas

Biclustering



Biclustering



BiclustAnthropom

Conclusiones

- Para cada talla de cintura, se obtienen grupos de mujeres definidos para un número específico (y no pre-determinado) de variables corporales.
- Las prendas asociadas a esos grupos deberían diseñarse utilizando sólo esas variables.
- Prototipos: mediana de cada bicluster.
- Desacomodados: individuos no agrupados.
- Grupos muy restrictivos y homogéneos.
- Selección automática de variables.
- Reproducibilidad: CC es un algoritmo determinista.

Data depth and biclustering applied to anthropometric data: exploring their utility in apparel design

Abstract

Anthropometry is a field rapidly becoming more quantitative and numerous statistical and computational studies have proliferated to tackle some common anthropometric problems. This paper aims to investigate the application of two different statistical methods to anthropometric data and to discuss their potential usefulness in the definition of an efficient sizing system oriented towards the clothing design. These approaches are Biclustering and Data Depth.

We present two methodologies: the first one is based on a well-known biclustering method. The second one is an extension of a clustering algorithm based on data depth previously published in the literature. Both strategies represent each size group by means of a central individual (the size prototype) and they are adapted to satisfy some very important requirements of the apparel design context.

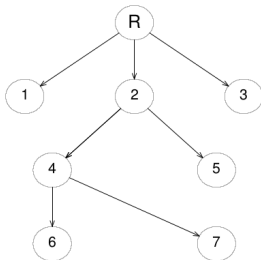
The anthropometric data used come from an anthropometric survey of the Spanish female population, performed during 2006. Results of this analysis show that biclustering, unlike classical clustering, finds groups not always defined for the same total number of body dimensions involved in the study. At the same time, it discovers which are the most relevant dimensions to build every size group. This feature is very useful for the apparel design. On the other hand, the data depth paradigm can be used to get prototypes of prefixed sizes. Furthermore, clustering based on data depth allows homogeneous size groups to be obtained. It shows better computational performance with small sample sizes.

Keywords: Apparel design, Sizing systems, Clustering, Biclustering, Data depth, Size prototypes

3^{er} trabajo: *hipamAnthropom*

Antecedentes y propuesta I

- El algoritmo **HIPAM** es un método de clustering jerárquico divisivo basado en PAM (Wit et al. (2004)).
- Originalmente creado para el análisis de datos de genes.
- Combina las ventajas de los métodos jerárquicos y de partición.



- Comienza con un grupo con todas las observaciones.
- En cada nivel del árbol, los clusters se dividen utilizando PAM.

HipamAnthropom

Antecedentes y propuesta II

- El número de grupos se decide maximizando la silueta media (asw).
- Utiliza distancias p -norma (p. ej., Manhattan, euclídea).
- La versión local del HIPAM divide los clusters hasta que se alcanza un nivel de homogeneidad (criterio de parada).
- Nuestra propuesta:
 - * Utilizar el HIPAM local con datos antropométricos para definir modelos de ajuste o de patronaje (*fit models*).
 - * Sustituir las dist. p -norma por la disimil. global de *trimowa*, $d_{trimowa}$.
 - * Incorporar el criterio INCA para dividir clusters y como criterio de parada (Irigoien et al. (2008)).
- Proponemos dos algoritmos:
 - * $HIPAM_{MO}$: HIPAM + asw + $d_{trimowa}$.
 - * $HIPAM_{IMO}$: HIPAM + INCA + $d_{trimowa}$.

HipamAnthropom

Datos y procedimiento

- Mismo conjunto de 6013 mujeres.
- Variable control primaria: busto.
- Variables control secundarias: pecho, cadera, cintura y nuca-tierra.
- Procedimiento:
 - 1 Los datos se segmentan en 12 tallas de busto (Norma UNE).
 - 2 En cada talla, se aplican $HIPAM_{MO}$ y $HIPAM_{IMO}$.

Busto	74-78	78-82	82-86	86-90	90-94	94-98
Busto	98-102	102-107	107-113	113-119	119-125	125-131

- Software estadístico utilizado: paquete de R *smida*.



HipamAnthropom

Resultados I: medoides

- Seleccionamos tallas con un número de mujeres > 250 .
- Ambos algoritmos devuelven un conjunto de medoides \rightarrow modelos de ajuste.

Medoides (modelos de ajuste)

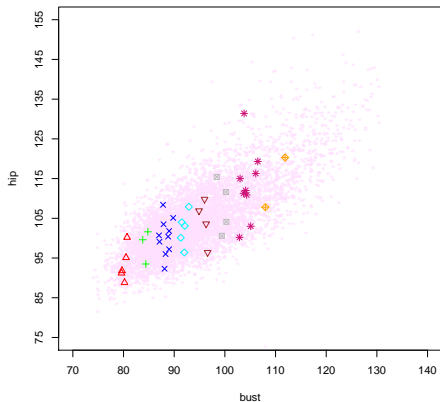
Talla de busto	[78, 82[[82, 86[[86, 90[[90, 94[
Núm. mujeres	287	732	1028	952
Núm. medoides $HIPAM_{MO}$	5	3	10	5
Núm. medoides $HIPAM_{IMO}$	3	3	3	3
Talla de busto	[94, 98[[98, 102[[102, 107[[107, 113[
Núm. mujeres	818	633	547	356
Núm. medoides $HIPAM_{MO}$	4	4	9	2
Núm. medoides $HIPAM_{IMO}$	3	3	3	5

- $HIPAM_{IMO}$ obtiene menos clusters y están mejor distribuidos.

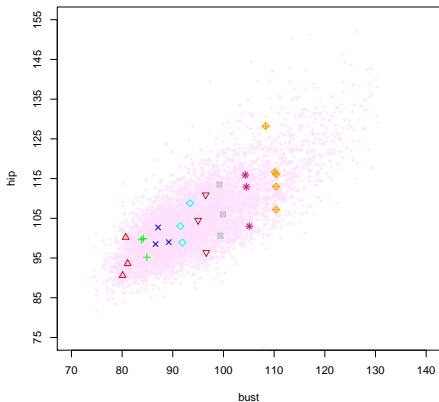
HipamAnthropom

Resultados II: medoides

Medoids HIPAM_MO
bust vs hip



Medoids HIPAM_IMO
bust vs hip



HipamAnthropom

Resultados III: outliers

- Dada su naturaleza jerárquica, HIPAM detecta outliers.
- Los outliers corresponden con las mujeres desacomodadas.

Outliers (mujeres no acomodadas)

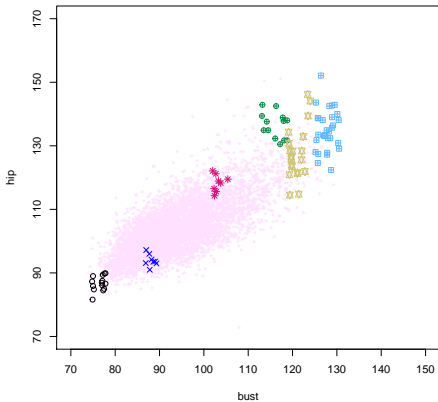
Algoritmo	Talla de busto	[74,78[[86,90[[102,107[[113,119[[119,125[[125,131[
	<i>HIPAM_{MO}</i>		30%	0.77%	1.5%	6.5%	23%
<i>HIPAM_{IMO}</i>		17%	—	—	14%	9%	100%

- *HIPAM_{IMO}* detecta outliers sólo en las tallas más pequeñas y grandes.

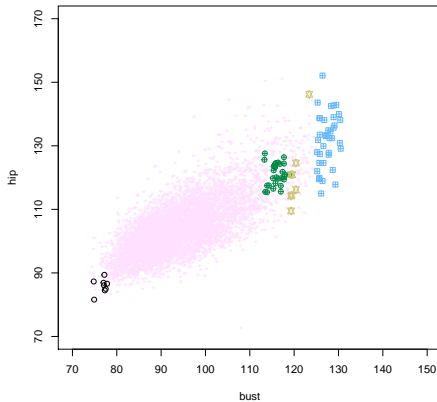
HipamAnthropom

Resultados IV: outliers

Outlier women HIPAM_MO
bust vs hip



Outlier women HIPAM_IMO
bust vs hip



HipamAnthropom

Conclusiones

- HIPAM es una herramienta estadística robusta para determinar modelos de ajuste representativos.
- Los modelos de ajuste corresponden con personas reales.
- HIPAM permite detectar de manera automática individuos atípicos.
- La elección entre $HIPAM_{MO}$ y $HIPAM_{IMO}$ depende del analista.
- Aportar nuevas ideas para definir modelos de ajuste:
 - No hay métodos en la literatura desarrollados con este objetivo.
 - Las compañías de ropa no explican cómo definen sus modelos.

Introducción
Estudio antropométrico de la población femenina en España
Casos antropométricos
Objetivos
Metodologías estadísticas
Conclusiones
Bibliografía básica

Clustering
Análisis estadístico de formas
Profundidad estadística
Análisis estadístico de arquetipos y arquetipoides
Paquete de R Anthropometry



Stradivarius

Modelo de Patronaje
Sallent (Manresa) Barcelona - España

Para Stradivarius

Funciones

Seleccionamos para nuestra central, ubicada en Sallent (Manresa), una modelo de patronaje de prenda de mujer.

Trabajarás junto con nuestro equipo de diseño, compras y comercial, probándote la ropa que después estará en todas nuestras tiendas.

Imprescindible disponer de las medidas del patrón base de corte, que deben coincidir con la talla 38 de pantalón y una talla M de camiseta. Las medidas requeridas son las siguientes: Altura 170-178 cm, busto: 90-95 cm, cintura 68-70 cm, cadera alta 81-84 cm, cadera 97-99 cm, muslo 57-59 cm.



ELSEVIER

Contents lists available at ScienceDirect

Decision Support Systems

journal homepage: www.elsevier.com/locate/dss



Looking for representative fit models for apparel sizing

G. Vinué^{a,*}, T. León^a, S. Alemany^b, G. Ayala^a

^a Department of Statistics and O.R., Avda. Vicent Andrés Estellés, 1, 46100-Burjassot, University of Valencia, Valencia, Spain

^b Biomechanics Institute of Valencia, Universidad Politécnica de Valencia, Valencia, Spain

ARTICLE INFO

Article history:

Received 15 February 2013

Received in revised form 23 July 2013

Accepted 23 July 2013

Available online 1 August 2013

Keywords:

HIPAM

Hierarchical tree

Partitioning around medoids

Fit models

Mean split silhouette

INCA statistic

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.

© 2013 Elsevier B.V. All rights reserved.

Diseño de ropa: casos centrales

Análisis estadístico de formas

kmeansProcrustes

Antecedentes y propuesta

- **Algoritmo k -medias:** minimiza $\sum_{i=1}^k \sum_{j \in C_i} d_E(x_j, \bar{x}_{C_i})^2$, siendo \bar{x}_{C_i} la **media muestral (centroide)** de cada grupo C_i y d_E la **dist. euclídea**.
- **Nuestra propuesta:** integrar la **media de Procrustes** y la **distancia de Procrustes** en el k -medias.
- Varios intentos (**Amaral et al. (2010)**, **Georgescu (2009)**):
 - * Amaral et al. → versión **Hartigan-Wong (H-W)** del k -medias.
 - * Georgescu → algoritmo similar a la versión **Lloyd** del k -medias.
- **Objetivos:**
 - 1 Adaptar el k -medias (Lloyd y H-W) al espacio de la forma para agrupar individuos de acuerdo a su forma dada por landmarks.
 - 2 Mostrar que Lloyd es mejor que H-W en el espacio de la forma.
 - 3 Analizar sus resultados en términos de tallaje.

kmeansProcrustes

Definiciones

- **Pre-forma de un objeto:** es lo que resulta tras eliminar los efectos de la posición y escala.
- **Espacio de la pre-forma:** conjunto de todas las pre-formas.
- **Forma de un objeto:** es lo que resulta tras eliminar los efectos de la posición, escala y rotación.
- **Espacio de la forma Σ_3^{66} :** conjunto de todas las formas.
- **Distancia completa de Procrustes, $d_F(X_1, X_2)$:** raíz cuadrada de la suma de las distancias al cuadrado entre los mismos landmarks en dos configuraciones superpuestas de manera óptima.
- **Distancia de Procrustes, ρ :** distancia mínima entre dos pre-formas en el espacio de la pre-forma $\Rightarrow d_F = \sin(\rho)$.
- **Media de Procrustes:** es la forma que tiene la mínima suma de las distancias Procrustes al cuadrado respecto a todas las configuraciones de la muestra.

kmeansProcrustes

k-medias en el espacio de la forma

- Lo aplicamos para agrupar X_1, \dots, X_n configuraciones.
 - (i) Dado $Z = ([Z_1], \dots, [Z_k]), [Z_i] \in \Sigma_3^{66}$, minimizamos con respecto a $\mathcal{C} = (C_1, \dots, C_k)$ asignando cada forma $[X_j]$ al grupo cuyo centroide tiene la menor distancia Procrustes a dicha forma.
 - (ii) Dado \mathcal{C} , minimizamos con respecto a Z , tomando $Z = ([\widehat{\mu}_1], \dots, [\widehat{\mu}_k])$, siendo $[\widehat{\mu}_i]$ la media Procrustes de las formas de C_i .
- Los pasos (i) y (ii) se repiten hasta que el algoritmo converge.
- Utilizamos la distancia de Procrustes, ρ , por una razón computacional.
- Adaptamos las versiones Lloyd y H-W.
- H-W calcula la media punto a punto. El cálculo de la media de Procrustes supone más carga computacional que la media muestral.

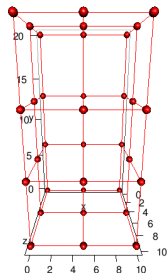
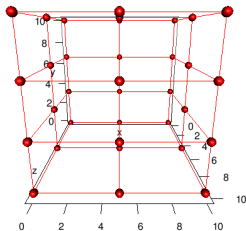
kmeansProcrustes

Estudio de simulación I

- **Hipótesis:** Lloyd tiene un menor coste computacional que H-W en el espacio de la forma (especialmente con muestras grandes).
- Representamos un cubo y un paralelepípedo con $l = 8, 34$ landmarks.
- Simulamos n_1 cubos para un primer cluster (1) y n_2 paralelepípedos para otro segundo cluster (2).
- El cluster 1 (resp. cluster 2) se define por una normal multivariante cuyo vector de medias $3l$ -dimensional se representa por el cubo (resp. paralelep.) y una matriz de var.-covar. $l \times l$ $\Sigma_1 = \sigma_1 I_{3l}$ (resp. $\Sigma_2 = \sigma_2 I_{3l}$), $l = 8, 34$.

kmeansProcrustes

Estudio de simulación II



kmeansProcrustes

Estudio de simulación III

- Aplicamos Lloyd y H-W a una combinación de los dos clusters para valores de n_1 , n_2 , σ_1 y σ_2 : $n_1 + n_2 = 50, 500, 900$ y $\sigma_1 = \sigma_2 = 0.1, 3, 6$.
- Índice de asignación: proporc. de observaciones agrupadas correctamente.
- Tiempo computacional: tiempo transcurrido para obtener resultados.

l=34 landmarks

n_1	n_2	σ_1	σ_2	Algoritmo Lloyd				Algoritmo Hartigan-Wong			
				Índice de asign.		Tiempo comput.		Índice de asign.		Tiempo comput.	
				\bar{x}	sd	\bar{x}	sd	\bar{x}	sd	\bar{x}	sd
25	25	0.1	0.1	1	0	20s.	3,47s.	1	0	23,59s.	17,71s.
25	25	3	3	0.924	0.05	39,61s.	15,15s.	0.96	0	46s.	19,66s.
25	25	6	6	0.674	0.078	24,78s.	5,79s.	0.712	0.11	43,22s.	9,56s.
250	250	0.1	0.1	1	0	3m.12s.	26,16s.	1	0	22m.08s.	19m.50s.
250	250	3	3	0.9866	0.0013	5m.	38,27s.	0.9863	0.00076	57m.43s.	15m.
250	250	6	6	0.86	0.017	8m.30s.	1m.	0.87	0.009	1h.18m.	11m.
450	450	0.1	0.1	0.988	0.0008	9m.	42s.	0.988	0	2h.45m.	50m.
450	450	3	3	0.989	0.0009	7m.27s.	1m.10s.	0.989	0	2h.29m.	55m.
450	450	6	6	0.903	0.0032	15m.	2m.	0.886	0.004	3h.30m.	1h.10m.

kmeansProcrustes

Datos y procedimiento

- Mismo conjunto de 6013 mujeres.
- Variables control primarias: busto y altura.
- Procedimiento:
 - * Los datos se segmentan por tallas de busto y altura.
 - * En cada talla, se aplica el *kmeansProcrustes* (Lloyd) con $k = 3$.

Busto	Altura 1	Altura 2
	≤ 162	$[162, 174[$
[74, 82[240	97
[82, 90[1052	694
[90, 98[1079	671
[98, 106[772	311
[106, 118[446	170

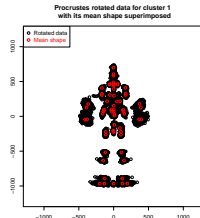
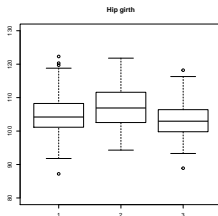
- Software estadístico utilizado: paquete de R shapes.



kmeansProcrustes

Resultados

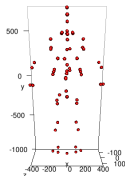
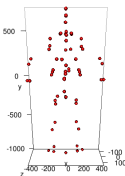
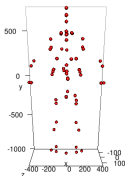
Busto $\in [90, 98[$; Altura $\in [162, 174[$		
671 mujeres		
Cluster 1	Cluster 2	Cluster 3
299	184	188



Mean shape cluster 1

Mean shape cluster 2

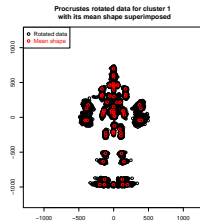
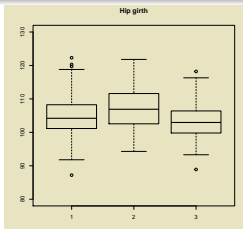
Mean shape cluster 3



kmeansProcrustes

Resultados

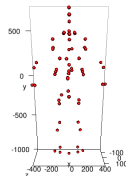
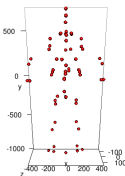
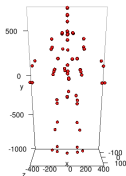
Busto $\in [90, 98[$; Altura $\in [162, 174[$		
671 mujeres		
Cluster 1	Cluster 2	Cluster 3
299	184	188



Mean shape cluster 1

Mean shape cluster 2

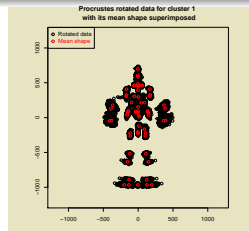
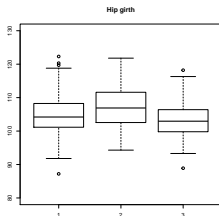
Mean shape cluster 3



kmeansProcrustes

Resultados

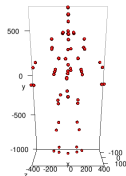
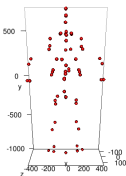
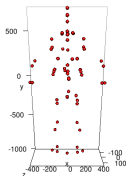
Busto $\in [90, 98[$; Altura $\in [162, 174[$		
671 mujeres		
Cluster 1	Cluster 2	Cluster 3
299	184	188



Mean shape cluster 1

Mean shape cluster 2

Mean shape cluster 3



kmeansProcrustes

Conclusiones

- Algoritmo k -medias (Lloyd y H-W) adaptado al espacio de forma para agrupar individuos de acuerdo a su forma representada por landmarks.
- Lloyd (*kmeansProcrustes*) es más eficiente en el espacio de forma que H-W.
- La metodología *kmeansProcrustes* permite:
 - Encontrar una partición en tallas óptimas.
 - Definir prototipos a partir de los centroides de cada grupo.
- Versión trimmed de *kmeansProcrustes* para descartar individuos atípicos (explicado en la tesis).

k-means algorithm for clustering human body shapes oriented to apparel design

G. Vinné^a, A. Simó^{b,*}, S. Alemany^c

^aDepartment of Statistics and O.R., University of Valencia, Valencia, Spain.

^bDepartment of Mathematics-IMAC, University Jaume I. Castellón, Spain.

^cBiomechanics Institute of Valencia, Universidad Politécnica de Valencia, Valencia, Spain.

Abstract

Since the basic foundation of k -means algorithm is to use 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, it arises in a natural way the idea of integrating Procrustes distance and Procrustes mean into k -means algorithm to adapt it to the shape analysis context. There has been several attempts in that sense, each one adapting a different version of k -means algorithm. In this paper we present a study comparing the performance of adapting Hartigan-Wong and Lloyds versions of the k -means algorithm to the field of statistical shape analysis. Both algorithms are applied to a 3D data base obtained from an anthropometric survey of the Spanish female population conducted in this country in 2006. Thus, our paper is concerned with apparel sizing system design.



SEGUNDA REVISIÓN

Diseño de ropa: casos centrales

Profundidad estadística

Profundidad estadística

Antecedentes y propuesta

- Los medoides y centroides son individuos centrales y buscamos grupos a partir de ellos.
- **Nuestra propuesta:** a partir de grupos ya definidos, buscar el individuo más centrado → medidas de centralidad o **profundidad**.
- Profundidad estadística: ordenación de datos multivariantes (Liu et al. (1999)). “De dentro hacia fuera”:
 - * Los puntos más cercanos al centro tienen un valor de profundidad más alto y a medida que nos alejamos del centro, la profundidad de los puntos disminuye.
- Dada una distrib. de probabilidad F en \mathbb{R}^d , $d \geq 1$, una función de profundidad mide cómo de profundo (central) es un cierto punto $x \in \mathbb{R}^d$ respecto a F o respecto a una nube de puntos $\{x_1, \dots, x_n\}$.
- Han sido propuestas muchas funciones de profundidad.

Profundidad estadística

1^{er} trabajo: prototipos de tallas prefijadas. Datos y procedimiento

- Software estadístico utilizado: paquete de R `depth`:
 - * Tukey.
 - * Liu.
 - * Oja.
- Mismo conjunto de 6013 mujeres.
- Variable control primaria: busto.
- Variables control secundarias: cintura y nuca-tierra.
- Procedimiento:
 - * Elegimos la talla de busto $[86, 90[$ (1028 mujeres).
 - * Aplicamos las profundidades de Tukey y Oja (`depth`), Mahalanobis y convex hull peeling (programadas) y L_1 (Jörnsten (2004)).

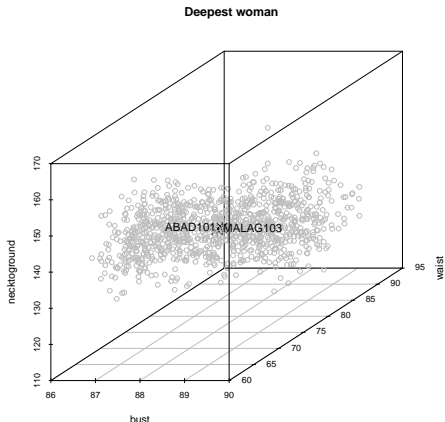


Profundidad estadística

1^{er} trabajo: prototipos de tallas prefijadas. Resultados I

- ABAD101 y MALAG103 son las mujeres más profundas (prototipos).

Mujeres	Dimensión	Nuca-tierra	Cintura	Busto
ABAD101		137.4	76.6	88
MALAG103		137.2	76.4	87.9



Profundidad estadística

1^{er} trabajo: prototipos de tallas prefijadas. Resultados II



ABAD101



MALAG103

Profundidad estadística

2º trabajo: clustering y profundidad: *TDDclust*. Teoría

- R. Jörnsten presentó un método clustering basado en la profundidad L_1 del punto y respecto al cluster k , $D(y|k) \rightarrow$ *DDclust* (Jörnsten (2004)).
- *DDclust* devuelve la partición $I_1^K = \{I(1), \dots, I(K)\}$ que maximiza:

$$C(I_1^K) = \frac{1}{N} \sum_{k=1}^K \sum_{i \in I(k)} (1 - \lambda) sil_i + \lambda (D_i^w - D_i^b)$$

- * sil es la silueta de x_i y $\lambda \in [0, 1]$ controla la influencia de la prof. sobre el clustering.
- * $D_i^w = D(x_i|k)$: prof. de $x_i : i \in I(k)$ resp. a su cluster $I(k)$.
- * $D_i^b = D(x_i|l)$: prof. de $x_i : i \in I(k)$ resp. a su cluster más cercano $I(l)$.
- Región de profundidad α -trimmed: $DD_F^\alpha = \{x : DD_F(x) \geq \alpha\}$.
- **Extensión de *DDclust***: trimmed *DDclust* \rightarrow *TDDclust*.
 - 1 Definir regiones trimmed.
 - 2 Aplicar *DDclust* sobre las observaciones restantes.

Profundidad estadística

2º trabajo: clustering y profundidad: *TDDclust*. Datos y procedimiento

- Mismo conjunto de 6013 mujeres.
- Variable control primaria: busto.
- Variables control secundarias: cintura y nuca-tierra.
- Procedimiento:
 - 1 Los datos se segmentan en 12 tallas de busto (Norma UNE).
 - 2 En cada talla, se aplica *TDDclust*.

Busto	74-78	78-82	82-86	86-90	90-94	94-98
Busto	98-102	102-107	107-113	113-119	119-125	125-131

- Software estadístico utilizado: funciones de R de *DDclust*.



Profundidad estadística

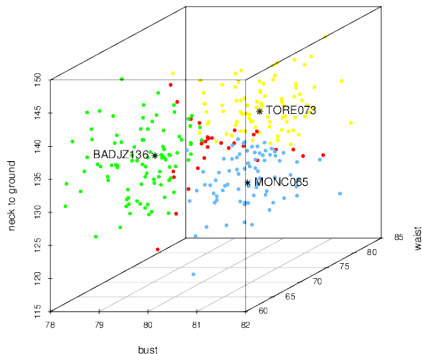
2º trabajo: clustering y profundidad: *TDDclus*. Resultados I

Busto $\in [78,82[$		
287 mujeres		
Cluster 1	Cluster 2	Cluster 3
93	88	77

Mujer	Dimensión		
	Nuca-tierra	Cintura	Busto
BADJZ136	135.6	66.7	79.4
TORE073	140.1	71.6	81
MONC065	130.7	68.5	81.1

- Clusters muy bien separados.
- Mujeres trimmed localizadas en las fronteras de los clusters.

Red points are trimmed observations
 Black stars are the deepest women of each cluster
 Green, yellow and blue points correspond to cluster 1,2 and 3



Profundidad estadística

2º trabajo: clustering y profundidad: *TDDclust*. Resultados II



BADJZ136



TORE073



MONCO65

Profundidad estadística

Conclusiones

- Procedimiento simple y efectivo para identificar prototipos.
- TDDclust, método de clustering basado en profundidad:
 - * Permite definir tallas eficientes sin individuos atípicos (en la misma línea que *trimowa*).
- Gran coste computacional cuando el tamaño muestral es grande.
- Podría ser utilizado con tallas *especiales* (grandes y pequeñas).

Data depth and biclustering applied to anthropometric data: exploring their utility in apparel design

Abstract

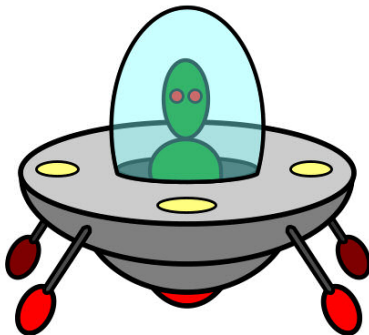
Anthropometry is a field rapidly becoming more quantitative and numerous statistical and computational studies have proliferated to tackle some common anthropometric problems. This paper aims to investigate the application of two different statistical methods to anthropometric data and to discuss their potential usefulness in the definition of an efficient sizing system oriented towards the clothing design. These approaches are Biclustering and Data Depth.

We present two methodologies: the first one is based on a well-known biclustering method. The second one is an extension of a clustering algorithm based on data depth previously published in the literature. Both strategies represent each size group by means of a central individual (the size prototype) and they are adapted to satisfy some very important requirements of the apparel design context.

The anthropometric data used come from an anthropometric survey of the Spanish female population, performed during 2006. Results of this analysis show that biclustering, unlike classical clustering, finds groups not always defined for the same total number of body dimensions involved in the study. At the same time, it discovers which are the most relevant dimensions to build every size group. This feature is very useful for the apparel design. On the other hand, the data depth paradigm can be used to get prototypes of prefixed sizes. Furthermore, clustering based on data depth allows homogeneous size groups to be obtained. It shows better computational performance with small sample sizes.

Keywords: Apparel design, Sizing systems, Clustering, Biclustering, Data depth, Size prototypes

Problema de acomodación en la cabina de una aeronave (casos extremos)



Análisis de arquetipos

Antecedentes y propuesta

- Problema de acomodación: el ajuste de los extremos permite el ajuste de los demás.
- Ejemplos: entrada de una garaje (máximo), pedales de coche (mínimo).
- Tradicionalmente, se han utilizado tres procedimientos estadísticos:
 - * **Análisis de percentiles** → Perspectiva univariante.
 - * **Regresión** → Proporciona valores medios para las predicciones.
 - * **PCA** → Parte de la variabilidad se elimina. Cuanta más variab. se considere, más casos se necesitan definir. El PCA no tiene por objetivo calcular individuos extremos.
- **Nuestra propuesta:** análisis estadístico de arquetipos (AA) (Cutler et al. (1994)).

Análisis de arquetipos

¿Qué es un arquetipo?: en nuestro lenguaje cotidiano...

Arquetipo de anciano vencido

JAVIER MEMBA
El debut en la gran pantalla de Joan Dalmau fue tardío. Aun así, su actividad cinematográfica se prolongó a lo largo de más de 50 títulos. Ahora bien, tuvieron que pasar varios años para que impusiera su arquetipo de anciano vencido en un pasado, casi siempre concerniente a la Guerra Ci-

vil. Tras dar vida a algunos de estos personajes a las órdenes de Montxo Armendáriz, le consagró Miralles, el posible salvador de Rafael Sánchez Mazas (Ramón Fontserè) en *Soldados de Salamina* (2003), la aplaudida adaptación de David Trueba de la novela homónima de Javier Cercas. Aquella creación fue merced a

en menor medida también en la pequeña pantalla, cuando Dalmau se estrenó en el cine por primera vez a las órdenes de Gonzalo Herralde en *Últimas tardes con Teresa* (1984). Ya en 1987, Jaime de Arminán le confiaba el personaje del vicealmirante Céspedes en su comedia *Mi amor*

Los hermanos Pastor nos proponen una visión arquetípica del fin del mundo basada en

Análisis de arquetipos

¿Y en Estadística?

meaningful subsets of the population. An archetype is defined as a centrally located subject in each cluster that is the true representative of the cluster [2]. The use of cluster analysis to identify such archetypes would go a long way to increase our understanding of the data. For

Y:

this goal, the clustering approach should be able to represent each cluster by its archetype [2].

Fuente: Paquet, E. *Exploring Anthropometric Data Through Cluster Analysis*. Published in Digital Human Modeling for Design and Engineering (DHM). June 15-17, 2004. Oakland University, Rochester, Michigan, USA. NRC 46564.

- ¡Un arquetipo no es un prototipo!

Análisis de arquetipos

Definición estadística de arquetipo

- Sea \mathbf{X} una matriz $n \times m$ y k el número de arquetipos.
- El objetivo del AA es obtener las matrices α y β ($n \times k$) que minimicen:

$$RSS = \sum_{i=1}^n \left\| \mathbf{x}_i - \sum_{j=1}^k \alpha_{ij} \mathbf{z}_j \right\|^2 = \sum_{i=1}^n \left\| \mathbf{x}_i - \sum_{j=1}^k \alpha_{ij} \sum_{l=1}^n \beta_{jl} \mathbf{x}_l \right\|^2$$

bajo las restricciones:

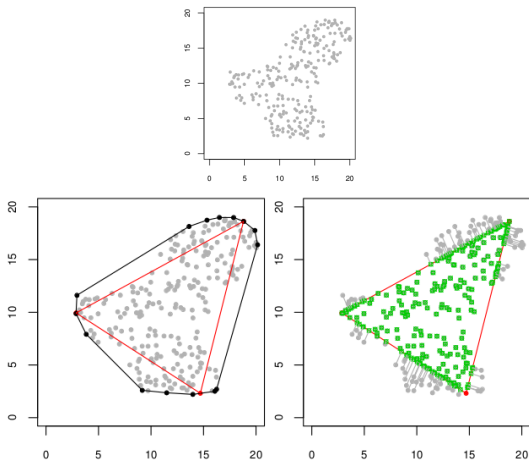
- 1) $\sum_{j=1}^k \alpha_{ij} = 1$ con $\alpha_{ij} \geq 0$ y $i = 1, \dots, n \implies \hat{\mathbf{x}}_i = \sum_{j=1}^k \alpha_{ij} \mathbf{z}_j$
- 2) $\sum_{l=1}^n \beta_{jl} = 1$ con $\beta_{jl} \geq 0$ y $j = 1, \dots, k$

- **ARQUETIPO**: vector extremo de la matriz de datos que se obtiene como una mixtura de los demás: $\mathbf{z}_j = \sum_{l=1}^n \beta_{jl} \mathbf{x}_l$ (Cutler et al. (1994)).
- Los arquetipos se pueden calcular con el paquete de R archetypes.



Análisis de arquetipos

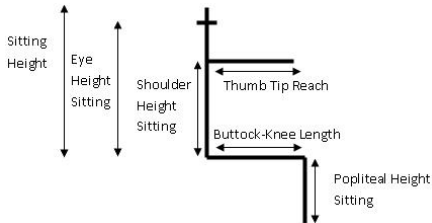
¿Cómo se calculan los arquetipos?



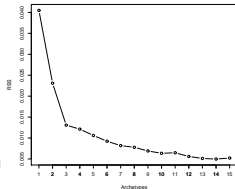
Análisis de arquetipos

1^{er} trabajo: AA vs PCA

- Utilizamos una base de datos de pilotos de aviones (encuesta USAF).
- Seleccionamos las 6 variables más importantes para diseñar una cabina.



- Conclusiones más relevantes:
 - * El objetivo del AA es obtener individuos extremos.
 - * El AA alcanza el nivel de acomodación exacto.
 - * Los arquetipos no se pueden calcular con el PCA.
 - * El número de arquetipos puede ser determinado de manera subjetiva (opinión del diseñador) u objetiva (criterio del codo).





Contents lists available at SciVerse ScienceDirect

Computers & Industrial Engineering

journal homepage: www.elsevier.com/locate/caie



Archetypal analysis: Contributions for estimating boundary cases in multivariate accommodation problem



I. Epifanio^{a,*}, G. Vinué^b, S. Alemany^c

^a Dept. de Matemàtiques, Universitat Jaume I, Castelló, Spain

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

^c Biomechanics Institute of Valencia, Universidad Politècnica de Valencia, Valencia, Spain

ARTICLE INFO

Article history:

Received 3 August 2012

Received in revised form 6 December 2012

Accepted 7 December 2012

Available online 29 December 2012

Keywords:

Representative case

Archetype

Percentile

Anthropometry

Representative human model generation

ABSTRACT

The use of archetypal analysis is proposed in order to determine a set of representative cases that entail a certain percentage of the population, in the accommodation problem. A well-known anthropometric database has been used in order to compare our methodology with the common used PCA-approach, showing the advantages of our methodology: the level of accommodation is reached unlike the PCA approach, no more adjustments are necessary, the user can decide the number of archetypes to consider or leave the selection by a criterion. Unlike PCA, the objective of the archetypal analysis is obtaining extreme individuals, so it is the appropriate statistical technique for solving this type of problem. Archetypes cannot be obtained with PCA even if we consider all the components, as we show in the application.

© 2012 Elsevier Ltd. All rights reserved.

Análisis de arquetipoides

Definición I

- Los arquetipos no siempre son observaciones de la muestra:

$$\mathbf{z}_j = \sum_{l=1}^n \beta_{jl} \mathbf{x}_l \quad \text{con} \quad \sum_{l=1}^n \beta_{jl} = 1 \quad \text{y} \quad \boxed{\beta_{jl} \geq 0}$$

- En algunas situaciones, es fundamental que sí que lo sean.
- La solución hasta ahora es calcular los individuos más cercanos a los arquetipos. Se puede hacer de dos maneras:
 - 1 **nearest**: Aquellos cuya d_E a los arquetipos es la mínima.
 - 2 **which**: Aquellos que tienen el mayor α para cada arquetipo.
- Sin embargo, estos individuos pueden ser “artificiales”: “*no economist in our sample fits this archetype to 100%*” (Seiler et al. (2013)).

Análisis de arquetipoides

Definición II

- Se propone un nuevo concepto: el **ARQUETIPOIDE**:

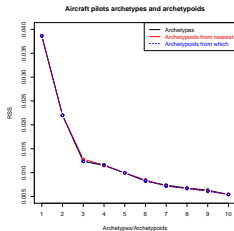
$$z_j = \sum_{l=1}^n \beta_{jl} x_l \quad \text{con} \quad \sum_{l=1}^n \beta_{jl} = 1 \quad \text{y} \quad \boxed{\beta_{jl} \in \{0, 1\}}$$

- Un arquetipoide es un individuo de la base de datos.
- Propiedades más importantes:
 - * Los arquetipoides no tienen por qué coincidir con los individuos *nearest* o *which* → Caso de estudio en la tesis: jugadores de la NBA.
 - * Los arquetipoides siempre existen, incluso en el caso de que no se disponga de variables → Caso de estudio en la tesis: diseño de ropa.
- Un nuevo algoritmo, inspirado por PAM, permite calcular arquetipoides:
 - 1 El vector inicial de arquetipoides es *nearest* o *which*.
 - 2 Este vector se intenta mejorar en sucesivas iteraciones.



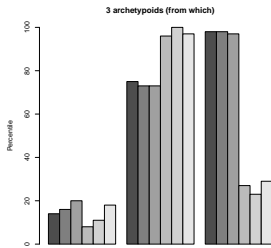
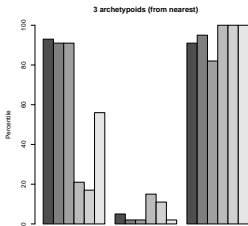
Análisis de arquetipoides

2º trabajo: búsqueda de arquetipoides



● Caso de estudio: encuesta USAF.

	RSS
3 arquetipos	0.012380
3 arquetipoides (which) (1632,1822,52)	0.012385
3 arquetipoides (nearest) (2177,2240,1691)	0.0128
which (1421,314,1691)	0.0195
nearest (511,314,1691)	0.0182



Archetypoids: a new approach to define representative archetypal data

G. Vinué⁽¹⁾, I. Epifanio^{(2)(*)}, S. Alemany⁽³⁾

(1) Department of Statistics and O.R., University of Valencia, Valencia, Spain. (2) Dept. Matemàtiques. Universitat Jaume I, Castelló, Spain () Corresponding author.*

(3) Biomechanics Institute of Valencia, Valencia, Spain.



SEGUNDA REVISIÓN

Abstract

This paper introduces a new concept: archetypoids. Archetypoid analysis represents each observation in a dataset as a mixture of actual observations in the dataset, which are pure type or archetypoids. Unlike archetype analysis, archetypoids are real observations, not a mixture of observations. This is relevant when existing archetypal observations are needed, rather than fictitious ones. An algorithm is also proposed to find them and some of their theoretical properties are introduced. We also show how they can be obtained when only dissimilarities between observations are known (features are unavailable). Archetypoid analysis is illustrated in two design problems and several examples, comparing them with the archetypes, the nearest observations to them and other unsupervised methods.

Keywords: Archetype, Convex hull, Unsupervised learning, Extremal point, Non-negative matrix factorization.

Paquete de R Anthropometry

- Un nuevo paquete de R, llamado **Anthropometry**, reúne todos los algoritmos programados.
- Incluye la base de datos de la encuesta USAF y una muestra de 600 mujeres del estudio antropométrico.
- Contiene un documento explicativo (viñeta).
- Disponible en CRAN y www.uv.es/vivigui/software.



Journal of Statistical Software

MMMMMM YYYY, Volume VV, Issue II. <http://www.jstatsoft.org/>

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**, which is available on the Comprehensive R Archive Network. It 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.

Conclusiones I

- Los sistemas de escaneo corporal 3D han permitido disminuir el coste y duración del procedimiento de toma de medidas.
- Además, obtienen datos muy precisos y actualizados de la población.
- Estas nuevas bases de datos constituyen una información muy valiosa para...
 - * Mejorar y optimizar el diseño de ropa y de entornos laborales.
 - * Entender la forma corporal de la población.
 - * Reducir el ciclo del proceso de diseño.
- Para sacar el máximo partido de ellas, se deben desarrollar métodos estadísticos rigurosos y herramientas software de fácil acceso y utilización.

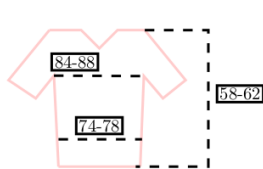
Conclusiones II

- En esta tesis doctoral, se han presentado las siguientes metodologías:
 - * Basadas exclusivamente en clustering (1):
 - *trimowa*.
 - *biclustAnthropom*.
 - *hipamAnthropom*.
 - * Basadas en clustering y el análisis de formas (2): *kmeansProcrustes*.
 - * Basadas en clustering y profundidad estadística (3): *TDDclust*.
 - * Basadas en el análisis de arquetipos y arquetipoides (4).
- Se ha presentado un nuevo paquete de R que permite utilizar y modificar todos los algoritmos asociados a cada propuesta.

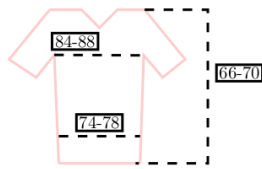
Conclusiones III

- (1), (2) y (3) permiten definir prototipos y modelos de ajuste, a partir de los que se pueden fabricar sistemas de tallaje y maniqués.
- (1), (2) y (3) han seguido el mismo esquema:
 - * Los datos se han segmentado en grupos a partir de una primera dimensión control (busto o cintura).
 - * Cada grupo se vuelve a dividir en subgrupos utilizando otras dimensiones control secundarias.
- La primera segmentación permite elegir la talla (grupo) fácilmente.
- Los subgrupos resultantes optimizan el tallaje.

Conclusiones IV



Busto:	84-88 cm.
Cintura:	74-78 cm.
Nuca-cadera:	58-62 cm.



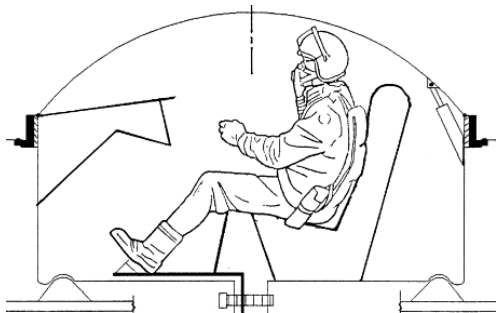
Busto:	84-88 cm.
Cintura:	74-78 cm.
Nuca-cadera:	66-70 cm.



Busto:	84-88 cm.
Cintura:	70-74 cm.
Nuca-cadera:	62-66 cm.

Conclusiones V

- (4) permite identificar casos extremos (arquetipoides). Se utilizan en el diseño de lugares de trabajo.
- Acomodar a los extremos permite acomodar a los demás.
- El análisis de arquetipos y arquetipoides representa una mejora respecto a los enfoques utilizados hasta ahora.



Bibliografía básica I



Alemany, S., González, J. C., 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 International Conference on 3D Body scanning Technologies. Lugano, Switzerland, pp. 1–10.



Amaral, G. J. A., Dore, L. H., Lessa, R. P., Stosic, B., 2010. *k*-Means Algorithm in Statistical Shape Analysis. Communications in Statistics - Simulation and Computation 39 (5), 1016–1026.



Ashdown, S. P., 2007. Sizing in clothing: Developing effective sizing systems for ready-to-wear clothing. Woodhead Publishing in Textiles.



Claude, J., 2008. Morphometrics with R. Use R! Springer.



Cutler, A., Breiman, L., November 1994. Archetypal Analysis. Technometrics 36 (4), 338–347.



Dryden, I. E., Mardia, K. V., 1998. Statistical Shape Analysis. John Wiley & Sons.



Dryden, I. L., 2012. *shapes* package. R Foundation for Statistical Computing, Vienna, Austria, contributed package.



Eugster, M. J. A., Leisch, F., April 2009. From Spider-Man to Hero - Archetypal Analysis in R. Journal of Statistical Software 30 (8), 1–23.

Bibliografía básica II



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.



Genest, M., Masse, J.-C., Plante, J.-F., 2012. depth: Depth functions tools for multivariate analysis. R package version 2.0-0.



HFES 300 Committee, 2004. Guidelines for Using Anthropometric Data in Product Design. Human Factors and Ergonomics Society.



Jörnsten, R., 2004. Clustering and classification based on the L_1 data depth. Journal of Multivariate Analysis 90, 67–89.



Kaiser, S., Leisch, F., 2008. A Toolbox for Bicluster Analysis in R. Tech. rep., 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.



Kaufman, L. and Rousseeuw, P. J., 1990. Finding Groups in Data: An Introduction to Cluster Analysis. John Wiley, New York.

Bibliografía básica III



León, T., Zuccarello, P., Ayala, G., de Ves, E., Domingo, J., 2007. Applying logistic regression to relevance feedback in image retrieval systems. *Pattern Recognition* 40, 2621–2632.



Liu, R. Y., Parelius, J. M., Singh, K., 1999. Multivariate analysis by data depth: Descriptive statistics, graphics and inference. *The Annals of Statistics* 27 (3), 783–858.



Madeira, S. C., Oliveira, A. L., 2004. Biclustering Algorithms for Biological Data Analysis: A Survey. *IEEE Transactions on Computational Biology and Bioinformatics* 1, 24–45.



McCulloch, C. E., Paal, B., Ashdown, S. P., 1998. An optimization approach to apparel sizing. *Journal of the Operational Research Society* 49, 492–499.



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.



Wit, E., McClure, J., 2004. *Statistics for Microarrays: Design, Analysis and Inference*. John Wiley & Sons, Ltd.



Wit, E., McClure, J., 2006. *Statistics for Microarrays: Inference, Design and Analysis*. R package version 0.1.

MUCHAS GRACIAS POR LA ATENCIÓN