A NOVEL APPROACH IN FORMULATING A SIZE CHART FOR FEMALE PANTS

Chandra Padmini Vithanage

118041A

Thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Textile and Clothing Technology

Department of Textile and Clothing Technology

University of Moratuwa

Sri Lanka

April 2015

DECLARATION

I declare that this is my own work and this thesis does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant to University of Moratuwa the non-exclusive right to reproduce and distribute my thesis, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature/ Date:

Declaration.

The above candidate has carried out research for the PhD thesis under my supervision.

Signature of the supervisor: ***UOM Verified Signature*** Date: 08/07/2015

ACKNOWLEDGEMENT

When I look back on my research journey which started as an M.Phil candidate and upgraded to a PhD later, I am completely humbled by the number of incredible people who helped me to bring it up to that level. I owe a great debt of gratitude to my research supervisors, Dr. T.S.S.Jayawardane, Dr. C.D.Tilakaratne and Eng. S.N. Niles for their profound support and guidance which led me to finish my research and thesis on time. I am deeply grateful to the members of the progress review committee for their valuable comments to make this research work a success.

Specially, I would like to place on record my gratitude to the following people who facilitated to capture data in their organizations; Lieutenant General J. Jayasooriya, Sri Lanka Army, Wing Commander Harsha Jayathilaka, Sri Lanka Air force, Major General W.P.P. Fernando, Captain A.S.L. Gamage and Dr. (Mrs.) Indira Dharmapriya, Sri Lanka Navy. My special gratitude to Mr. Samantha Seneviratne-Director, MAS Active (Pvt) Ltd and Management of Linea Intimo (Pvt) Ltd, Biyagama.

I would also like to express my gratitude to Dr. U.S.W. Gunasekara, Research Coordinator of the Department of Textile & Clothing Technology. I am also indebted to my colleagues in the Department who squeeze time from their busy time schedules to help me finish my research and thesis on time. Further, I would like to extend my gratitude to Technical Officers of the Department of Textile and Clothing Technology, Ms. Nirosha Samaraweera, Ms. Indu Bogoda and Ms. H.S.K. Nilmini for their support in sample making.

I am gratefully acknowledge the University grant commission for the research grant "Financial assistance by UGC for higher studies - 2013" given to me to carry out my research work.

Last but not least, I am so grateful to my loving family, my dear husband and three kids who continue to support and inspire me in my work, and whose presence in my life I dearly cherish.

ABSTRACT

Each country understands that they need their own size charts representing their population because researchers have found that human body shapes, proportions and measurements change significantly due to the geographical and demographical differences. Even though, many countries have developed their own size charts, ready-to-wear apparel industry still faces the problem of poor fit of apparels. Reasons for this fit problems may be due to several factors such as limitations of existing size chart development approaches, lack of up-to-date anthropometric data of the relevant population, vast body shape differences among the population, and restrictions in mass production systems. In this research, one of the above problems; issues in existing size chart development approaches, was studied comprehensively in order to identify drawbacks of the size chart development approaches. Statistical approach which uses descriptive statistics, k-means clustering combined with factor analysis and classification and regression decision tree method were widely used popular size chart development approaches. With the current lower body anthropometric data of Sri Lankan females of age 20-40 years, limitations of the above approaches were investigated. Through this explorative analysis, limitations of current approaches and potential improvements for a better approach were discerned. Thereby a novel approach for development of size charts was formulated. The proposed approach is capable of handling linearly inseparable data with high dimensionality without variable reduction. Further, number of clusters can be objectively determined and the transformation function could be optimized by tuning the parameters of it.

Kernel based learning is one of the latest data mining approaches in pattern recognition. A kernel based clustering technique called "global kernel k-means clustering technique", was adopted to cluster lower body anthropometric data in the proposed method. Selection of proper kernel function and tuning of kernel parameters are crucial in successful data clustering. For determining the number of clusters objectively, kernel based Dunn's index, which is one of the cluster validation technique, was successfully instrumented in the said novel approach. Thereby the lower body anthropometric dataset of females was successfully clustered through the proposed novel approach taking all variables into account. It was also proved that the developed size chart could successfully eliminate the fitting problems of Sri Lankan female pants. The size chart was validated through a well accepted index called Aggregate Loss of Fit index on theoretical ground and the live model fitting of fabricated pants according to the size chart through a standard feedback questionnaire. The complete approach in developing size charts could be of interest to other clustering applications in many fields also.

Keywords

Anthropometry, Clustering, Kernel based clustering, Cluster Validation, Development of size charts

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ALF	: Aggregate Loss of Fit factor
CART	: Classification And Regression Tree
GSM	: Grams per Square Meter
KDI	: Kernel-based Dunn's Index
MRI	: Magnetic Resonance Imaging
RTW	: Ready-To-Wear
WHR	: Waist-to-Hip-Ratio

List of Appendices

CHAPTER 1
INTRODUCTION

1.1 Background

One of the greatest challenges that apparel companies face today is providing satisfactory fit to its target consumers. According to Ashdown, Locker and Rucker (2007), there are two main reasons for this poor fit garments; one is lack of up to date anthropometric data and the other is non-availability of details on the issues related to fit of garments. Fit has been defined in the Oxford dictionary as "the ability to be the right shape and size". Further, according to Fan, Yu, and Hunter (2004), fit is directly related to the anatomy of the human body and well fitted garment conforms to the human body with adequate ease of movement. In the custom-made tailoring method of garment manufacturing, consumers achieve a good fit in their garments because the direct interaction between the consumer and the manufacturer facilitates of having exact measurements of the consumer as well as his/ her fit preferences. When this system is changed into a mass production system which is the manufacturing of garments in large quantities, fitting problems could be arisen in ready-to-wear garments.

Garments are manufactured in a mass scale using predefined size charts. The purpose of these size charts is to divide varied body measurements of a population into homogeneous subgroups. Size charts provide information for garment pattern production, pattern grading, production planning, marketing purposes and inventory planning. However, each consumer cannot achieve a good level of fit in ready-to-wear garments due to the existence of vast differences in body shape and proportion among the population, lack of up to date anthropometric data, drawbacks in existing size chart development approaches, restrictions that come across in mass production process and lack of required information in garment size labels (Winks, 1997; Pechoux & Ghosh, 2002).

A main issue of the usage of size chart is the relationship between the size charts and body dimensions and shapes which is not constant because of the changes that occur in the human population. Further, human biologists use the term "secular trend" to describe alterations in the human beings' body measurements which occur with the time. Over a period of century, biosocial changes have been occurring in the population which has led to increase in the rate of growth of children and also increase in adult stature (Iseri, 2008). Hence, body measurements should be updated regularly in order to provide current information on sizes and their distribution (Pechoux & Ghosh,2002). Having their own sizing systems representing their population is very important for every country to provide a good estimation of fit and size of ready-to-wear clothing (Ashdown 2007;Simmons, Istook & Devarajan,2004). There are several other factors that affect the variation of body measurements and body shapes among the population.

According to Pechoux and Ghosh (2002), individual variability, gender variability, race variability, generational variability, changing life styles, racial mixes and demographics are several factors that affect the changing of body measurements. Other than problems in sizing systems, there are some other reasons for fit problems of garments such as knowledge on one's body measurements, psychological factors regarding size and fit and under-evaluating garment sizes on label (Pechoux & Ghosh, 2002).

In developing size charts, different data analysis approaches have been used ranging from simple statistical methods to advanced data mining techniques. When these approaches are examined carefully, it can be seen that some shortcomings are there in the way of application. Hence, it will lead to a development of poor size chart which will not represent their target consumer resulting problems in garment fit.

Another reason for ready-to-wear apparel to have fit problems is the restrictions that appear in mass production system in apparel industry. In mass production system where the manufacturer and the consumer are not linked together, the manufacturer creates value to the product and it is up to the consumer to accept or reject it (Zilber & Nohara, 2009). Although this mass production systems enable low cost garment manufacturing, it badly affect to customer satisfaction due to poor fit. Hence, the concept, "mass customization" which was defined as "producing goods and services to meet individual customer's needs with near mass production efficiency", (Tseng & Jiao, 2001)was introduced in order to achieve customer satisfaction. Therefore, in mass customization, consumers are incorporated into the product development process at the beginning and hence, products follow the specifications of the consumer. When mass customization is targeted, availability of advanced technology such as computer aided design and manufacturing (CAD/CAM), 3-D body scanning facilities, flexible manufacturing systems are vital. Therefore, this mass customization concept was not popularized among majority of apparel manufacturers due to the huge investment on above mentioned technologies and other practical constraints such as receiving consumer's exact requirements and logistics matters. However, Levi Strauss & Co have succeeded the above constrains in providing custom fitted jeans (Zipkin, 2001).

In mass production system, standard size charts act as a communication tool among manufacturers, retailers and consumers. Therefore, the need for knowledge and understanding of the sizing systems of other nations becomes increasingly important to apparel companies which are conducting their businesses at international level (Jongsuk, 1993). According to McCulloch, Paal and Ashdown (1998), an effective and economical sizing system must satisfy several objectives, such as accommodating large percentage of population, providing good fit and maintaining a few sizes.

1.2 Identification to the problem

A Kurt Salmon Associates' study (2000) reported that 50percent of women cannot find a good fit in apparel and fit problems are the reason for 50percent of catalog returns (DesMarteau,2000). Furthermore, it was stated that the most dissatisfied apparel consumers were women (Alexander, Connell & Presley, 2005; Delong, Ashdown, Butterfield & Turnbladh, 1993). According to Otieno, Harrow and Greenwood (2005), from the women sample that they had analyzed, 54.7 percent had difficulty in finding well-fitting clothes. Well built women (size 16 plus) have greater difficulty in finding satisfactory fit. According to Bickle, Kotsiopulos, Dallas, & Eckman (1995), " Average or well-proportioned women were typically more satisfied with their own body shape and the fit of ready-to-wear clothing than less well proportioned and shorter persons".

LaBat and De Long (1990) found that women mostly dissatisfied with their lower body fit at the waist, hip and thighs and when compare with other apparel products, they dissatisfied mostly with pants. Anderson, *et al.* (2001) explored that 62 percent of the female survey group reported having fit problems in pant length while nearly 50 percent of the sample had fit problems at the waist, thigh and hip.

Dissatisfaction with fit is one of the most frequently stated problems with female ready- made garment purchases as consumer demand for good fit which enhance their feminine look (Otieno, *et al.*, 2005). Moreover, female consumers are dissatisfied with the search process needed to identify their correct sizes which give better fit (Roach, 1996). As a result, consumers have had difficulties finding garments which fit and retailers have thereby lost sales (Jongsuk, 1993). The view of good fit from the consumer, establishes a strong relationship between the consumer and the manufacturer (Anderson, *et al.*1999). While this dissatisfaction with fit of garments is a worldwide concern, in Sri Lankan context also, this might be a concern due to the unavailability of their own size chart.

A questionnaire survey was done by the researcher in 2012 among 112 university female students to find problems related to fit of female ready-to-wear apparel in Sri Lanka. In this survey, questions were asked about the range of garments such as pants, skirts, blouses and frocks. It was found that majority had problems with pants. Females who tend to select pants according to their waist complained about fit problems at pant hip area and the length. It was revealed that 66 percent of females return the pants due to fit problems and 74 percent of females had to fit-on three or more pants to select their best fitted one. Due to these facts, the rate of return of female pants as well as customer dissatisfaction is high in ready-to-wear garments in Sri Lanka.

Therefore, it was confirmed that females have fit problems in ready-to-wear garments, especially in pants. Further, it was proved that Sri Lankan females also have fit problems of pants. As mentioned in section 1.1, there were several reasons for poor fit of ready-to-wear garments. Among those reasons, the problems of lack of up to date anthropometric data and drawbacks in existing size chart development approaches were selected in solving fit problems of the female pants in Sri Lanka through this research. Therefore, it was decided to collect lower body anthropometric data from a sample of Sri Lankan females of aged 20-40 years. After investigating the existing approaches, the required features of a novel approach which can avoid the problems in existing approaches could be identified. Therefore, through this process, an effective approach in formulating size chart for female pants could be developed.

1.3 Objectives of the research

The main objective of this research is to propose a novel approach in formulating size charts for female pants in order to reduce fit problems. The specific objectives pertaining to the research can be summarized as follows:

- To apply widely used existing size chart development approaches to lower body anthropometric data of Sri Lankan females and explore the drawbacks of the approaches
- To explore a novel statistical approach to formulate size charts for female pants
- To develop size charts for pants for Sri Lankan females using new approach
- To validate the size chart in terms of better fit;
    - ✓ Using "aggregate loss of fit factor" with Sri Lankan females' anthropometric data
    - ✓ By a live fit assessment session followed by a questionnaire survey

1.4 Significance of the Study

In ready-to-wear apparel industry, size charts play a vital role. Hence, the size charts should be perfect in representing the target consumer. This research identifies the shortcomings of existing approaches in formulating size charts and develops a new statistical approach to create size charts for female pants with a better fit. Since the secondary data for lower body anthropometric data of Sri Lankan females were not available, it was decided to collect primary data which is more accurate.

The new approach proposed for developing size charts for pants for Sri Lankan females could most probably be generalized for size chart for female pants globally. However, generalizability of the new approach in formulating size charts for other garments need to be explored. The developed size charts can be utilized by apparel manufacturers who target Sri Lankan consumer, improving the customer satisfaction. Further, the anthropometric database of the lower body of female could be used for further studies.

1.5 Scope of the Study

According to Croney (1980), an average of 20 years is required for human beings to develop to full maturity. Based on that fact, lower limit of the age range was selected as 20 years. It was stated that estrogens cause higher levels of fat to be stored in a female body. It also affects body fat distribution, causing fat to be stored in the buttocks, thighs and hip in women, but generally not around her waist (Yegyan, 2010). When women reach menopause, the estrogen produced by ovaries declines and fat migrates from their buttocks, hips and thighs to their waists, later fat is stored at the abdomen (Harvard women's health watch,2006) causing to change the body measurements and shape. At around 40 years of age, human beings begin to shrink in stature that the shrinkage accelerates with age and that women shrink more than men (Iseri, 2008). Furthermore, it is very rare to find local elderly females who wear pants except in the Colombo area. These facts caused for the selection of upper limit of age range as 40 years. A size chart for elder females, who are above 40 years, should be done separately in order to get more effective size chart.

A convenience sample, in which female personnel from military forces were also included, was selected and permission for collecting data from military forces and other institutions was taken. Further, participants were informed about the ethical clearance of the collected data and personnel consent was taken. Females in many professions were included such as female personnel of Army, Air force and Navy, students of higher education institutions, working females in private sector and non-working females. This sample also covered females from different social backgrounds within the age limit. According to Abeysekara and Shahnavaze (1998), there were no significant differences in the measurements of human body representing all districts of Sri Lanka for both female and male populations, which indicated homogeneity in the country. However, the sample was selected such that all nine provinces of Sri Lanka were represented in different percentages according to the availability of subjects.

The size of the convenience sample was decided based on 95 percent confidence level (0.05 alpha level) and 3 percent of margin of error. In general, an alpha level of 0.05 is acceptable for most research (Bartlett, Kotrlik & Higgins, 2001). According to Krejcie and Morgan (1970), 3 percent margin of error is acceptable for continuous data. According to Sri Lankan Census and Statistical Department records (2001), Sri Lankan female population, whose age is between 20- 40 years, was 3.4 million. Accordingly, the sample size was calculated based on the above values and the theoretical sample size yields as 1068.

1.6 Structure of the Thesis

Chapter 1 presents a general introduction about the thesis with problem statement and objectives of the study. Further, it discussed the scope and significance of the study.

Chapter 2 focuses on anthropometry, past anthropometric surveys, anatomical structures of human body, female body shapes and its effect on garment fit. Further, it discusses different size chart development methods available and validation techniques of size charts.

Chapter 3 describes the theory behind kernel-based clustering method which is one of the latest techniques in data mining field. Under this, the researcher has used global kernel k-means clustering method in developing size charts for female pants and it explains in detail.

Chapter 4 presents the systematic approach in data collection, initial categorization of data, selection of proper kernel function and parameter tuning. It also explains the size chart development process and validation techniques.

Chapter 5 explores the validity of the widely used existing methods in developing size charts by applying lower body anthropometric data of Sri Lankan females. It details the drawbacks of the existing methods. Further, it describes the results of the research with new size chart for female pants. It also presents the results of live fit assessment session.

Chapter 6 dedicated to the conclusion which analyses the implications and contribution of the research, and perspective to be considered in future research work.

This research was supported by the University Grant Commission and the research was conducted at Department of Textiles and Clothing Technology, University of Moratuwa and Department of Statistics, University of Colombo.

CHAPTER 2

LITERATURE REVIEW I: ANTHROPOMETRY AND APPROACHES FOR DEVELOPING SIZE CHARTS

2.1 Introduction

In the plethora of literature, only few researches found to be on the area of the development of size charts. With the popularization of e-commerce, on-line shopping becomes an integral part of the life for the persons with computer literacy. As the computer literacy took an exponential growth across the globe, proportionately online shopping becomes highly significant. In this context, importance of perfect size charts is highly recognized globally.

Each country understands that they need their own size charts exactly representing their population since researchers have found that human body shapes, proportions and measurements change significantly with the geographical and demographical differences. When catering for the international apparel market, knowledge on different size charts is vital. With the development of computer software and improved data mining technique, author attempts to develop an advanced technique to formulate size charts with better precision.

The literature survey was carried out to study the relevant theories and underlying definitions of anthropometry and anatomical structure of the female body. Further, it reviewed the research findings of size charts development approaches.

2.2 What Is Anthropometry?

"Anthropometry is the science of measurement and the art of application that establishes the physical geometry, mass properties, and strength capabilities of the human body" (Roebuck, 1995). The meaning of anthropometry in Greek is *anthropos* "man" and *metrikos* "measure". Therefore "measurement of man" refers to the measurement of the human individual (Roebuck, 1995).

Anthropometry deals with body measurements, such as measurements of body size, shape, strength, and working capacity (Pheasant, 1996). Anthropometry is vital for criminology, medical practice, clothing design, industrial design, architecture and it also plays an important role in ergonomics (Pheasant, 1996).

The subject of anthropometry can be divided into two main sub categories, namely *static* and *dynamic* anthropometry.

i). Static anthropometry

It deals mainly with the physical structure of the body. It consists of measurements of the distance of bones between joint centres including some soft tissue measures in contour dimensions (Chakrabarti, 1997).

ii). Dynamic anthropometry

Measurements are taken when the body is in motion or engaged in a physical activity. "It includes reach, clearance and volumetric data. Reach signifies the extent that limbs can get to while clearance is the space allowed for a certain part of the body or the whole body itself" (Chakrabarti, 1997).

2.2.1 Anthropometric surveys

History of anthropometric surveys dates back to the era of the first World war in 1914.Countries like USA and UK, have done anthropometric surveys for military groups in order to design their uniforms and military equipments. Appendix 2.1 shows the anthropometric surveys done in USA and several other countries from its inception.

Meanwhile, very limited anthropometric surveys can be found in Sri Lanka except those done in the medical field. Abeysekera and Shahnavaze,(1998) have done an anthropometric survey on Sri Lankan workers, both male and female, and compared body sizes between developing countries with those of developed countries. Their survey covered people from every district in Sri Lanka aged between 21- 51 years, including every race in the country.

2.3 Anatomical Structure of the Female Body

As explained in section 1.2, fit problems of ready-to-wear garments arise due to huge variations in human body shapes existing in a population. Therefore, it is important to have a general awareness about the anatomical structure of the human body which directly related to above mentioned body shape variations.

The human body is composed of about 35 percent water and 65 percent solids in terms of its total body weight. According to the proportions of fat, muscles and bone content, which form the solid part, different figure types are formed. The formation of final figure types are based on the genetic reasons or acquired due to food habits, physical exercise, occupation, working habits and the influences of geographical locations (Chakrabarti, 1997).When anatomical structure of human body is discussed, it is vital to observe the anatomy of lower body of females because in this research, different female lower body shapes need to be considered.

The lower limb is subdivided by the hip joint, knee joint, and ankle joint into the regions; buttocks, thigh, leg and foot. The pelvis is the section between the legs and the torso that connects the spine (backbone) to the thigh bones (Ellis,1997).The pelvis contains a large compound bone structure at the base of the spine, which is connected with the legs. In adults, it is mainly constructed of two hip bones, one on the right and one on the left of the body. The two hip bones consist of 3 sections, the Ilium, Ischium and Pubis and these sections are fused together as shown in Fig.2.1 (Martini & Bartholomew, 2000). Along with the hip bones is the sacrum, the upper-middle part of the pelvis, which connects the spine (backbone) to the pelvis. Its major functions are to bear the weight of the upper body when sitting and standing, swings from side to side by a rotator movement during walking, provide attachments for muscles in the hand and foot and provide bony support for the birth canal (Ellis,1997).

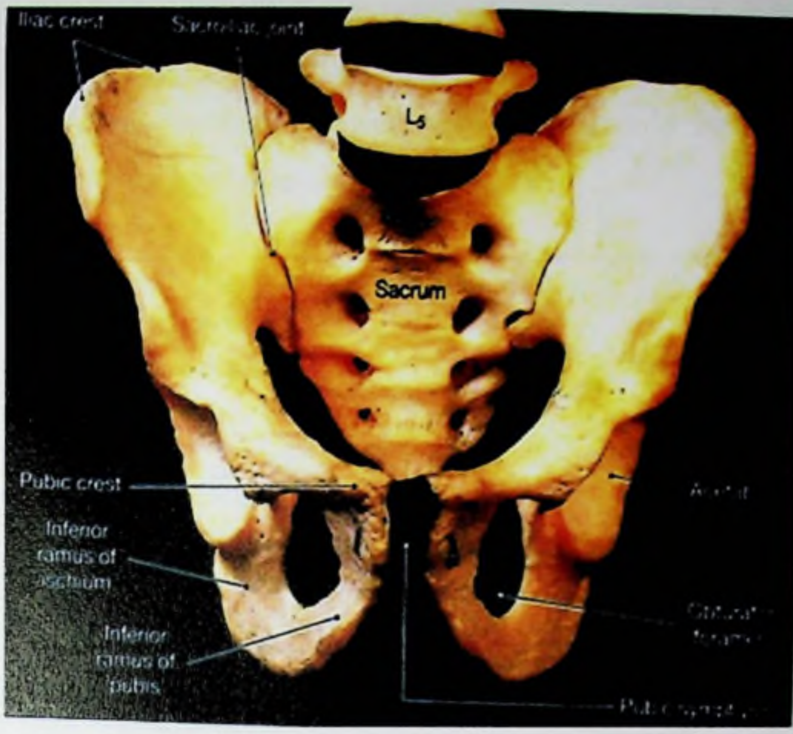


Fig.2.1 The bony pelvis

2.3.1 Variations of the pelvic shape

According to Ellis (1997), there are different female pelvic shapes which directly relate with different lower body shapes of females. These lower body shapes have strong impact on the fit of female pants. Therefore, it is important to have an understanding on those different pelvic shapes in developing size charts of pants.

According to Ellis (1997), the female pelvic shape can be subdivided as follows:

(i) The normal and its variants;
  - Gynaecoid
  - Android
  - Platypelloid
  - Anthropoid

(ii) Symmetrically contracted pelvis

(iii) The Rachitic flat pelvis

(iv) The Asymmetrical.

Fig 2.2 (Ellis, 1997) shows the different pelvic shapes given above. It is certain that the shapes of female lower body are determined by these pelvic shapes. Therefore, in developing size chart for female lower body, there should be a technique to differentiate these lower body shapes.

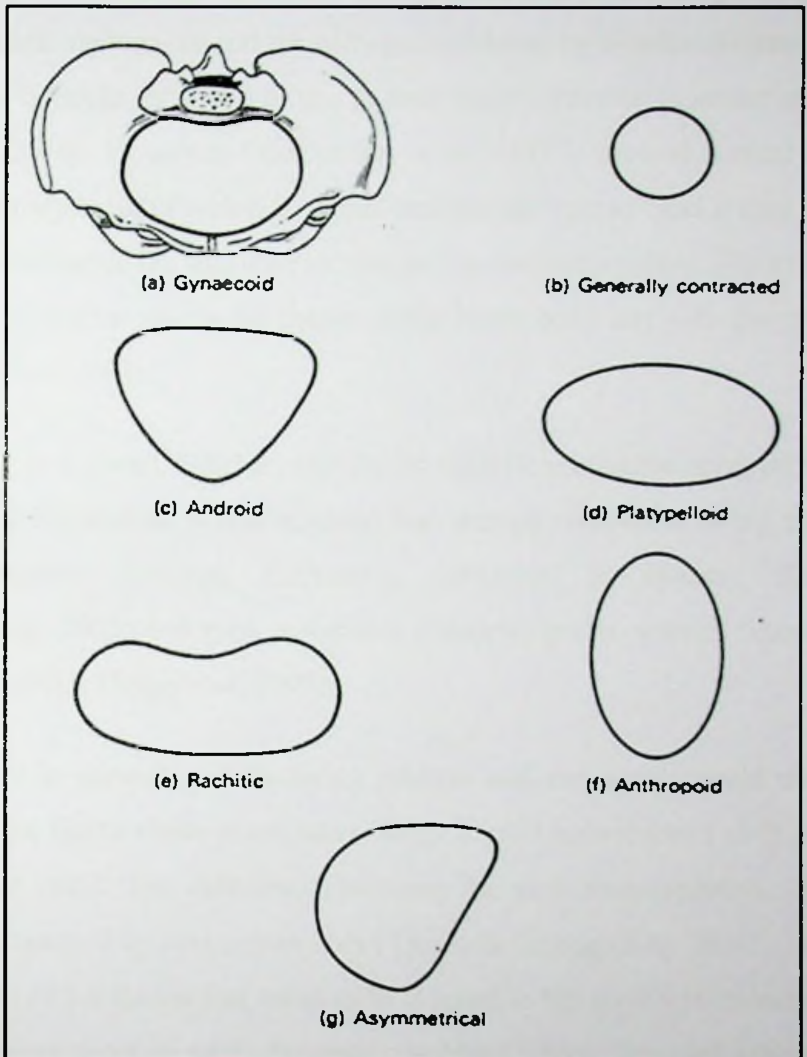


Fig. 2.2 Variations of females' pelvic shapes

2.3.2 Waist-to-hip ratio in classifying lower body shapes

When body shapes are considered, waist-to-hip ratio (WHR) is a significant factor in classifying lower body shapes. Further, this WHR of females is an important factor in several aspects such as attractiveness, healthiness, fertility, and cognitive ability of her child (Singh, 1993).

Due to testosterone and estrogen levels and their effect on fat distribution in the body, the waist to hip ratio varies. High estrogen levels results in low waist-to-hip ratios while high testosterone cause to high waist-to-hip ratios (Sing,1993). Estrogens affects body fat distribution, causing fat to be stored in the buttocks, thighs and hip in women, but generally not around her waist (Sugiyama,2005).When women reach menopause and the estrogen produced by ovaries declines, fat migrates from their buttocks, hips and thighs to their waists, later fat is stored at the abdomen (Yegyan, 2010). However, Frienlander, *et al.,* (1977) showed a steady longitudinal increase in hip breadth which is called 'middle age spread' and it may not be totally due to gathering of fat, but involve changes in the bony pelvis. Therefore, it is clear that fat distribution affects the shapes of the lower body and with the age, the pattern of change may differ.

According to Lidwell, Holden, and Butler (2003), preferable level of WHR ratio is 0.67 to 0.8 for women. It was revealed that women with waist to hip ratio of 0.7 are more attractive (Dixson, Grimshaw, Linklater, & Dixson, 2010; Streeter &McBurney, 2003) and men in western countries prefer women whose WHR is in between 0.6-0.7 (Sugiyama, 2005).

The people in apparel manufacturing process and consumers could understand and visualize the figure shape much more easily if ratio between two girth measurements were given rather than differences between the girth measurements, which was the most used method by past researchers (Gupta & Ghangadhar, 2004) . For example, a WHR ratio of 1.0 means that waist girth is equal to hip girth while similar rules apply to other figure types as well. According to Hsu (2009), "the girth ratio approach can provide a more reasonable and convenient identification of figure types to facilitate apparel manufacturing".

2.3.3 Variations of leg length

When female lower body is considered, the leg length and leg shape variations are also important in achieving correct fit of pant (Veblen, 2012). It is reported that adult leg length was positively related with parental height, birth weight, and weight at 4 years. (Wadsworth, *et al.*, 2001) According to the questionnaire survey done by the researcher, it was noticed that the length of a pant is also a strong factor to consider in providing pants which satisfy the customer. Hence, size charts for lower body should be formulated providing multiple choices for pant leg length. Anatomical structure of a human leg is shown in appendix 2.1.

2.3.4 Findings on human body proportions

Knowledge of human body proportions is very important in different fields such as medical practice, clothing design, industrial design, architecture and ergonomics. There are several findings and theories regarding body proportions such as "Vitruvian Man" and "Eight heads theory". Marcus Vitruvius Pollio, a Roman author, architect and engineer during the 1st century BC, explained in his book titled "De Architectura", human body proportions defining his "Vitruvian Man" that different proportions exist in the human body (Pollio, 1914/2006) and later Leonardo da Vinci converted it to a drawing as shown in Appendix 2.3. Pollio, (1914), revealed that the foot is one-seventh of the height of a man while from below the foot to below the knee is a quarter of the height of a man. There are several other important relationships among body measurements as revealed by him as shown in Appendix 2.2.

The Eight head theory explain that the human figure can be divided in to eight approximately equal parts (heads) and those parting lines goes through some particular body parts that can be identified straightforward (Appendix 2.3) (http://www.b- u.ac.in/sde_book/fashion_design.pdf). Making use of "Eight heads theory", a number of relative length measures can be obtained, for a grown up proportionate human body. For example, fore-arm length (i.e. armpit to wrist bone) equals to one-fourth of the total height, which is two heads. Several other examples on eight heads theory are given in Appendix 2.3. These are useful theories which could be used in apparel pattern production.

2.3.5 Relationship of consumer body shape with apparel fit

The satisfaction with clothing fit will be higher if the body shape of the wearer can be considered in manufacturing process. Properly fit garment will satisfy the consumer regardless of the price or style of the garment (Laitala, Klepp & Hauge, 2011). However, Lee, *et al.* (2007) explained that, it was not easy to define the shape of the human body because the human body has a very complex structure and each individual has unique body characteristics. Fit problems of apparel upsurge further because manufacturers expect the human body to match with the ideal figure, and produce accordingly (Strydom & Klerk, 2010). Therefore, dissatisfaction with fit is still one of the major complaints expressed by apparel consumers. Providing more consumers with better fit will benefit the retailer, the manufacturer, as well as the consumer (Strydom & Klerk, 2010).

However, as explained above, garment fit is a very complex subject. Therefore, knowledge on different body shapes of consumers is vital in developing size charts which provide accurate fit of apparel.

2.4 Identification of Control (Key) Dimensions

Winks (1997) explained that control dimensions are "the body dimensions on which a sizing system is built, which are fundamental to the definition of body size, and which are used to assign a suitably sized garment to a wearer". Further, a control dimension must be a good predictor of other body dimensions related to a certain garment type and it must be one that can be measured accurately by consumers in deciding their garment sizes (Chun-Yoon, 1996). According to Strydom and Klerk (2010), key dimensions are used, for choosing fit models for fit testing and for communicating size description of garments to the consumer. Therefore, identification of proper control dimensions is very crucial in foundation step of developing size charts.

Key dimensions are identified by experts in the apparel sector, most probably based on their experience and knowledge on the subject. Moreover, Pearson correlation coefficients can be used to detect the dimension's correlation with other dimensions and most highly correlated (> 0.70) dimension with others can be decided as a key dimension (Beazley, 1998). In addition, from factor analysis, the dimensions with high factor loadings (> 0.75) in final factors can be considered as key dimensions(Chung, Lin, & Wang, 2007,Hsu, 2009).Further, Ashdown (1998) has selected four dimensions; hip, waist, crotch height(inseam height) and crotch length, as key variables to create her sizing system for female pants. However, apparel brands, such as Nike, Banana Republic, Old Navy, have used waist, hip and inseam height as key measurements for their female pants.

2.5 Evolution of Size Charts

Standard size charts are very important since it is the communication tool among manufacturers, retailers and consumers. McCulloch, *et al.* (1998) explained that an effective and economical size chart must satisfy several objectives:

- increase the accommodation of the population
- reduce the number of sizes in the system
- improve overall fit in the accommodated individuals

The structure of size charts used for the ready-to-wear apparel industry today was originated from proportional drafting systems developed by tailors in the past (Ashdown, 1998).In that particular method, individual body measurements were used to make the custom-fitted garments. With time, tailors identified various relationships that exist among different body measurements and these relationships were used to develop pattern drafting systems. Hence, large inventory of ready-to-wear garments were made available which could be sold to any number of similarly sized individuals. This process led to change the custom- fitted garment production system into a mass production system.

The first scientific size charts related to body measurements were published by the British Standards Institution (BSI), in 1953 (Laitala, *et al.* 2011). After that, in 1958, the USA developed the CS 215-58 standard sizing system on the experience of manufacturers and measurements of 10,042 women from an anthropometric survey done in 1941 (Ashdown, 1998). In 1970, the PS 42-70 standard size chart of the USA was developed incorporating military anthropometric data. However, the standard sizes from PS 42-70 no longer represented the current population because of the changes in body measurements with generations. Hence, the American Society of Testing and Material (ASTM) developed a new standard, D5585-94 which was cross–checked with US Army and Navy anthropometric databases. International sizing system for clothing was started in 1969 and the first international standard for clothing size designations (ISO 3635) was published in 1977 (Laitala, *et al.*, 2011).

2.6 Review on Different Approaches Used for Developing Size Charts

A detailed review was undertaken on various approaches used in developing size charts reported in literature. It was revealed that methods ranging from simple statistical methods to complex data mining methods have been used for the formation of size charts. Table 2.1 shows the summary of historical research work carried out on size chart development.

2.6.1 The approach based on descriptive statistics

In the past, standard size charts were developed in countries such as UK and USA using a statistical approach and descriptive statistics of the anthropometric database such as mean, standard deviation, range, maximum, and minimum were used for the process. It was found that the measurement range was evenly divided considering a constant size interval between sizes in getting different sizes (Haggar, 1990). Beazley (1997) has analyzed the past size charts and revealed that the size intervals that they maintained for bust, waist and hip were 4 cm and hence she recommended the same. Furthermore, it was noted that three body height categories were followed as short, regular and tall. However, Laitala, *et al.* (2011) revealed that past size charts were organized in a way that the vertical measurements increased along with the increment of girth measurements. However, Pearson correlation coefficients (Table 5.11) showed poor correlation between vertical and girth measurements and hence increasing vertical measurements along with the increment of girth measurements makes problems.

It was found that above statistical approach was improved by some researchers (Beazley, 1999; Guptha & Gangadhar,2004; Mpampa, Azariadis, & Sapidis, 2010) by incorporating drop values (hip girth–bust girth) to get different body shapes.

Table 2.1 Past Research on size chart development

Author/s	Year	Country	Sample details	Method used
Croney, J. E.	1977		317 female students,18 variables	Correlation coefficient & Factor analysis
Deonier,C.J.S.; DeLong,M.R. Martin,F.B. ; Krohn,K.R.	1985	USA	1217 US army women 17-35 years	Principal Component Sizing System
Tryfos,P.	1986	USA	4025 Air Force personnel	Integer programming to optimize number of sizes
Beazley,A.	1999	UK	100 college female students	Correlation coefficient ,Descriptive statistics
McCulloch,C.,Paal,B.&Ashdown,S.P	1998	USA	2208 US Army Women	Non-Linear Optimization techniques
Otieno,R. ;Fairhurst,C	2000	Kenya	618 Female children, 2-6 years, 33 variables	Correlation coefficient and bivariate classification
Sizirovicza,L.Ujevic.D.&Drenovac,M.	2002	Croatia	4268 men, 18-22 years ,50 variables	Discriminant analysis
Gupta,D.&Gangadhar,B.	2004	India	2095 adult females	Principal component analysis(PCA),Univariate Analysis
Hsu,C.&Wang,M.	2005	Taiwan	610 Army soldiers,265 variables	Factor analysis and decision tree (CART)
Gupta,D,Garg,N.Arora,K.,Priyadarshini,N.	2006	India	1900 females	linear programming approach
Salusso,C.J ; Borkowski,J.; Reich,N; Goldsberry,E.	2006	USA	55^{+} women	Principal component Sizing system
Chung ,M., Lin,H., and Wang,M.	2007	Taiwan	7800 children ,6-18 years ,36 variables	Factor analysis and k-means cluster analysis
Lin,H. ;Hsu,C.; Wang ,M.;;Lin,Y.	2007	Taiwan	610 Army soldiers, 265 variables	Factor analysis and decision tree CART
Zakaria,et.al.	2008	Malaysia	660 Boys, 7-12 years	Principle Component analysis, K- means clustering
Hsu,C.	2009	Taiwan	986 females, 52 variables	Factor analysis and two-stage clustering
Ariadurai,A, Nilusha,T. and Dissanayake,M.	2009	Sri Lanka	160 children 5-12 years	Bivariate classification using body types (height and girth)
Hsu,C.	2009	Taiwan	755 females,45-64years,44 variables	Factor analysis ,girth ratio,Descriptive Statistics
Doustaneh,A.Gorji,M.andVarsei,M.	2010	Iran	670 men	Self- organization method (SOM)
Bagherzadeh,R,, Latifi,M.&Faramarzi,A.R.	2010	Iran	1050 males ,16-22 years	Factor analysis, cluster analysis and decision tree technique.
Mpampa,M.L, Azariadis,P.N.andSapidis	2010	Greece	12,810 men	Descriptive Statistical analysis
Esfundarani,M.S. ; Shahrabi,J.	2012	Iran	males	PCA , non-hierarchical clustering
Jeyasingh,M.M. and Appavoo,K.	2012	India	Males ,25-66 years	Clustering technique

Under those different body shapes, size charts were developed considering constant size interval between sizes while maintaining constant drop values. However, Hsu (2009) explained that "although women's figure types were often classified based on the girth differences, these differences are not wholly reasonable or convenient for determining figure types". For example, a difference of 14–18 cm between bust and waist girths will yield a figure type but it is impossible to visualize the shape of intended customers because this can be due to larger bust girth or smaller waist girth (Hsu,2009). Therefore, a method which can assess the body shapes accurately should be incurred in developing size charts.

The statistical approach, as explained above, was used to develop a size chart using lower body anthropometric dataset of Sri Lankan females. Further, resultant size chart has been used to explain the reasons why this method is not appropriate in developing size charts.

The poor correlation between vertical and girth measurements found to be one reason to become the above technique not a success. Further, due to the negligence on body shape differences in developing size charts, this method do not represent the population well. When the developed size chart was validated statistically using aggregate loss of fit factor, it seemed that the result is beyond the acceptance level. Therefore, this statistical approach is not suitable for size chart development and it is discussed in detail under section 5.3.1.

2.6.2 Cluster analysis combined with factor analysis

Researchers have used cluster analysis methods combined with factor analysis in developing size charts (Hsu, 2009; Shahrabi & Esfundarani, 2010; Chung, *et al.*, 2007). Factor analysis was used to reduce the number of variables to a few new variables called factors representing original variables. From cluster analysis, new variables were clustered into several groups which were used to develop size charts.

The theory behind factor analysis and clustering approach were briefly explained and how past researchers applied it in the development of size charts was also described below.

2.6.2.1 Factor analysis

When the number of variables are high in cluster analysis, calculating the particular distance of a point from the centroid is a problem due to high dimensional spaces which is called "curse of dimensionality" (Sharma, 1996). Hence, factor analysis was used to reduce the number of variables and introduce new set of variables called factors, representing the original ones. The most popular method to determine the number of factors was the Eigen value greater than one rule and the scree plot (Sharma, 1996). Factor rotation, an orthogonal or an oblique rotation, was done to achieve a simpler factor structure that can be meaningfully interpreted while orthogonal rotation cause rotated factors to be orthogonal to each other (Sharma, 1996).Using factor loadings and researcher's knowledge about the variables, new factors were identified, representing all original variables. Each factor was named according to the set of body dimensions that belongs to it. However, it was found that variable reduction will result information loss since new factors cannot represent all original variables exactly.

2.6.2.2 Clustering technique

According to Izenman (2008), cluster analysis is known as "data segmentation or class discovery. The methodology consists of various algorithms to organize a given data set into homogenous subgroups or clusters". It was found that many clustering algorithms were available which were classified as hierarchical and partitioning methods. These methods follow the same basic concept of maximizing between-group variances while minimizing within-group variances. Other than the above mentioned methods, several other methods such as density-based, model-based, and grid-based methods were available (Rokach & Maimon, 2010). The two methods, namely hierarchical and partitioning, that were used for clustering of anthropometric data in past researches are explained below.

(i) Hierarchical methods

This method works by grouping data objects into a hierarchy or "tree" of clusters and it is useful for data summarization and visualization. A hierarchical clustering method can be either agglomerative or divisive. The agglomerative hierarchical clustering method uses a bottom-up strategy where each object form its own cluster and iteratively merges clusters into large clusters. The divisive hierarchical clustering method employs a top-down strategy and it place all objects in one cluster and then divides into several smaller sub clusters (Han, Kamber & Pei,2012).

Past researchers used this method to gain an understanding on possible number of clusters in their anthropometric data set. However, the exact number of clusters available in the dataset cannot be found from this method.

(ii) Partitioning method

K-means and k-medoids clustering methods are examples for this partitioning method of clustering. The *k*-means algorithms starts by assigning items to one of k pre-determined clusters and then computing the k cluster centres (centroids) (Han, *et al.*2012). The centre of each cluster is calculated as the mean of all the instances belonging to that cluster. However, the *k*-means algorithm is sensitive to initial seed selection resulting in a local minima and it is best applied for linearly separable clusters (Han, *et al.*, 2012).

It was found that the combination of factor analysis and two-step clustering method (hierarchical method and k-means algorithm) was used by past researchers (Hsu, 2009; Shahrabi & Esfundarani,2010; Chung, *et al.*, 2007) for clustering their anthropometric datasets. From factor analysis, a few factors such as girth, length and height factors were identified depending on their set of anthropometric variables. The dataset was clustered based on the factor scores that assigned to each case, using two-step clustering method. From hierarchical clustering, using "dendogram", the possible number of clusters (k) was assumed and these cluster numbers were used in *k*-means clustering to segment the factors. Past researchers used these clusters to develop size charts for the relevant population.

This approach was applied for the lower body anthropometric dataset of Sri Lankan females and from the results, it was noted that variable reduction is less effective in developing size charts and could not be applied to anthropometric data due to non-linearity of data. Furthermore, the problems incurred in *k*-means clustering algorithm such as trapping in local minima and finding only linearly separable clusters, affect the final clusters adversely. Therefore, this approach was not appropriate in developing effective size charts and it is discussed in detail under section 5.3.2.

2.6.3 Classification and regression (CART) decision tree approach

According to Han, *et al.* (2012), a decision tree is "a flow chart- like tree structure where each internal node denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node holds a class label". The tree nodes provide mean values for the target variable depending on the predictor variables. Hence, selection of predictor variables is very crucial in the regression tree.

Lin, Hsu, Wang, & Lin, (2008) used the CART decision tree method to develop a standard sizing system for soldiers. Here, a new target variable (body mass index) was created using anthropometric variables (weight (Kg)/ height^2(m^2)). The level of the tree was subjectively selected to classify the dataset and classification was done based on chest girth which was a predictor variable in his analysis. Further, the decision tree was completely dependent on the selection of predictor variables. Although key variables were selected based on the results of factor analysis, other variables may also affect the decision tree. The researchers selected height as the second predictor variable while it has already been used in creating BMI which was their target variable. Hence, the selection of a predictor variable (height) which was used to generate the target variable (BMI) might lead to wrong decisions. Further, the decision tree method is suitable for classifying datasets where previous knowledge of class labels is available. Therefore, the use of the CART decision tree in classifying an anthropometric dataset which does not have prior class labels, by subjectively selecting a tree level may result in a final solution that deviates significantly from the actual.

This approach was applied to the lower body anthropometric dataset of Sri Lankan females and the resultant decision tree was analyzed. It was found that there was no existing theory underpinning the selection of a tree level and classifying the human body based on particular measurements. Further, selection of a few predictor variables among a large number of variables was also a problem. Therefore from the results obtained it was noted CART decision tree approach is not a suitable approach in developing size charts. The result of this approach is discussed further under section 5.3.3.

2.6.4 Other approaches suggested for size chart development

McCulloch, *et al.*, (1998) introduced a novel approach for the construction of apparel sizing systems using "Nonlinear optimization techniques" to derive the sizing system. The core of the approach was to fix the number of sizes and optimize the quality of the fit.

Tryfos (1986) used the "Integer Linear Programming" approach to get optimal sizes formulating "minimum aggregate discomfort or loss" for any number of control measurements which calculated the distance between the person's measurements and allocated size. The probability of purchase depended on the distance between the sizing system of a garment and the real size of an individual. In order to optimize the number of sizes so as to minimize the distance, an integer programming approach was applied.

Doustaneh, Gorji and Varsei (2010) suggested "Self Organizing Maps (SOM)" which is an unsupervised learning approach to establish a non-linear sizing system. They selected height and bust girth as important measurements for size chart development and height was grouped into three categories. Under those height categories, bust girth was subdivided using SOM. Validation was done comparing with the statistical approach and it showed that the SOM method was much better than statistical method. However, according to Rokach and Maimon, (2005), "it is sensitive to the initial selection of weight vector, as well as to its different parameters, such as the learning rate and neighborhood radius".
The problems encountered with the SOM method are that artificial neural networks are more sensitive to data and when adding or changing a case of the system, the outcome will change dramatically. Even when repeating the process with the same dataset, the same outcome could not be achieved. Hence, generalizing the size chart which is developed from SOM is difficult.

The above mentioned approaches have shortcomings in the way of application such as selection of key variables, reduction of variables, expectation of linear relationships between variables and expectation of normal distribution of variables in the process of developing size charts. However, anthropometric data do not follow exact normal distribution and is not linearly related each other. Only two or three key variables were selected for analysis in the above existing approaches and this variable selection may be a wrong decision which leads to poor size charts. Since these key variables' impact on the developing process of size chart is critical, selecting a few key variables is a very serious task especially in case where the changes in dimensions are significantly varies. Therefore, without considering only a few variables, if all relevant variables can be considered in the process, an effective size chart could be expected. Hence, an approach which can be worked with high dimensional non-linear data is considered in this research to develop an effective size chart.

2.7 Validation of Size Charts

After developing the size chart, it should be validated to optimize the fit of apparels. One possible method is the use of "Aggregate loss of fit" factor (AFL) (Gupta & Gangadhar, 2004) and the data set that was reserved for validation will be used for calculation.

2.7.1 Aggregate loss as a measure of goodness of fit

The aggregate loss of fit will measure the difference between the wearer's body measurements and the assigned size for the wearer. This measure explains how well the size chart represent the relevant population (Ashdown, 1998).The "aggregate loss" index expresses the average Euclidean distance between the dimensions of individuals and their allocated garment size.

It was calculated using following equation:

$$ALF = \frac{\sum_{i=1}^{n} \sqrt{(a_{i1} - b_{i1})^2 + (a_{i2} - b_{i2})^2 + (a_{i3} - b_{i3})^2}}{n}$$

(Gupta & Gangadhar, 2004)

Where a_1, a_2 ,& a_3 - assigned values for the key variables of individual

b_1, b_2, & b_3 - actual values of the key variables of individual

n - the number of members in each clusters.

According to Gupta & Gangadhar (2004), for the developed size chart to be a better one, the maximum deviation of assigned measurements from acual body measurement could be considered as ±1 inch. Hence, according to the above formula, for three key measurements, aggregate loss of fit is √3, i.e. 1.73 inch (4.4cm). In general, it is equal to $(N)1/2$ where N is the number of key measurements used.

2.8 Chapter summary

This chapter has explained different definitions, topics related to body sizes and available size chart development approaches that were widely used by past researchers. However, the problem of garment fit still prevails in the ready-to-wear garment industry. As mentioned in Chapter 1, these fit problems are due to many different reasons. The size chart development approaches have their own drawbacks and they were identified as one of the reasons. Therefore, a review of the existing widely used methods was done and their drawbacks were identified. The shortcomings and limitations of the application of those methods in developing size charts were not adequately investigated in the literature and clearly revealed the inapplicability of these techniques for highly non-linear set of anthropometric data. Several other approaches (e.g. Integer programming) which were rarely used by the researchers, were also discussed and in this study, these methods were not considered.

CHAPTER 3

LITERATURE REVIEW II: KERNEL- BASED CLUSTERING APPROACH

3.1 Introduction

Chapter 2 has explored the potential adoptability of widely used techniques found in literature in developing body size charts using lower body anthropometric data of Sri Lankan females. It has been revealed that the application of these approaches have serious shortcomings in developing body size charts and this makes completely out in adopting these techniques for female lower body anthropometric dataset. Therefore, the essentiality of a novel approach to develop size charts addressing the fit problems incurred in ready-to-wear apparel industry was endorsed. A novel approach should possess the capabilities of handling non-linear, high dimensional data, when clustering anthropometric data. Literature showed that kernel-based clustering approach has the required potential and therefore, further research was carried out.

Taylor and Cristianini (2004) identified three revolutions in automated algorithms for pattern analysis: detecting algorithm for linear relations in the 1960s, introduction of back propagation multilayer neural networks and decision trees in the mid 1980s, and kernel-based learning methods which enabled the analysis of non linear relations efficiently in the mid 1990s. The kernel-based learning facilitates different approaches of pattern recognition such as classification, correlations, rankings, clustering, principal component analysis on different types of data such as vectors, strings, images, text documents and so on (Taylor & Cristianini, 2004). The problems of local minima and over fitting which appears in neural network and decision trees are also avoided in this kernel-based learning techniques. Pattern analysis algorithms possess important properties such as computational efficiency, robustness and statistical stability and kernel-based learning techniques fulfill these requirements (Taylor & Cristianini, 2004).

According to Taylor and Cristianini (2004), "any kernel methods solution comprises two parts: a module that performs the mapping into the embedding or feature space and a learning algorithm designed to discover linear patterns in that space". The main features of the process of kernel methods can be summarized as follows:

(i) Data items are embedded into a vector space called the feature space.

(ii) Linear relations are sought among the images of the data items in the feature space.

(iii) The algorithms are implemented in such a way that the coordinates of the embedded points are not needed, only their pair wise inner products.

The pair wise inner products can be computed efficiently directly from the original data items using a kernel function (Taylor & Cristianini, 2004).

In this approach, input data need to be mapped into high dimensional feature space through a mapping function to find a linear separation there. However, mapping input data into high dimensional data (e.g.; $\{x_1, x_2\}$ $z = \{x_1^2, \sqrt{2}x_1x_2, x_2^2\}$; which means changing the presentation of variables) may be very difficult with large datasets and computationally expensive. Hence, without explicitly mapping data to high dimensional feature space through a mapping function, kernel function can be used to implicitly fulfill it which is called a "kernel trick". Rai, (2011) explained the kernel transformation as follows;

$$\begin{aligned} k(x, z) &= (x^T z)^2 \\ &= (x_1 z_1 + x_2 z_2)^2 \\ &= x_1^2 z_1^2 + x_2^2 z_2^2 + 2x_1 x_2 z_1 z_2 \\ &= (x_1^2, \sqrt{2}x_1x_2, x_2^2)^T (z_1^2, \sqrt{2}z_1z_2, z_2^2) \\ &= \Phi(x)^T \Phi(z) \end{aligned}$$

Where $x = \{x_1, x_2\}$ and $z = \{z_1, z_2\}$; $\Phi(x) = (x_1^2, \sqrt{2}x_1x_2, x_2^2)$;

$\Phi(z) = (z_1^2, \sqrt{2}z_1z_2, z_2^2)$

Through this kernel function input data transform to a kernel matrix which is a n×n, (where n is the number of cases), symmetric, and positive semi definite matrix (where all eigen values are non-negative, $\lambda \geq 0$). This is also called as a " Gram Matrix" which can be defined as follows;

Let $V = \vec{v}_1,\ldots,\vec{v}_n$,a set of input vectors, then Gram Matrix K is defines as:

$$K = \begin{pmatrix} \langle\emptyset(\vec{v}_1).\emptyset(\vec{v}_1)\rangle \cdots \langle\emptyset(\vec{v}_1).\emptyset(\vec{v}_n)\rangle \\ \langle\emptyset(\vec{v}_1).\emptyset(\vec{v}_n)\rangle \ \ddots \qquad \vdots \\ \vdots \\ \langle\emptyset(\vec{v}_n).\emptyset(\vec{v}_1)\rangle \cdots \langle\emptyset(\vec{v}_n).\emptyset(\vec{v}_n)\rangle \end{pmatrix}$$

As explained by Hofmann, kernel functions are symmetric and follow Cauchy-Schwarz Inequality property as given below:

Symmetry $K(\mathrm{x},\mathrm{z}) = K(z,x)$

Cauchy-Schwarz Inequality

$$K(x,z)^2 = K(x,x)K(z,z)$$

Further, Khardon, (2008), mentioned following properties of kernel function.

if k_1 and k_2 are kernels,

- $k(x,z) = k_1(x,z) + k_2(x,z)$
- $k(x,z) = ak_1(x,z)$ where $a > 0$
- $k(x,z) = f(x).f(z)$ for any function f on x
- $k(x,z) = k_1(x,z).k_2(x,z)$
- $k(x,z) = \frac{k_1(x,z)}{\sqrt{k_1(x,x)}\sqrt{k_1(z,z)}}$

Although, kernel based pattern recognition approaches were used in other fields such as image processing, text classification (Lodhi, Saunders, Taylor, Cristianini, & Watkins, 2002 ; Methasate & Theeramunkong, 2007), object recognition (Wang, Xiong, Jiang, & Ling, 2012), gene expression profile analysis (Liu, Chen, & Bensmail, 2005), DNA and protein analysis (Zien,*et.al* 2000), it is still novel to the area of size chart development.

From the literature, a new kernel based clustering approach called "Global Kernel *K*- means clustering" (Tzortzis & Likas, 2009) which was used for MRI segmentation, was identified. It was a combination of global *k*-means clustering approach (Likas,Vlassisb & Verbeekb, 2003) and kernel *k*-means clustering approach and it avoids the problems encounter with *k*-means clustering such as trap in local minima, inability of handling linearly inseparable data and difficulty in managing large number of variables. This approach has been tested using three different artificial datasets and MRI images and compared them with kernel *k*-means clustering. They have proved that global kernel *k*-means clustering is better than kernel *k*-means clustering in terms of clustering error (Tzortzis & Likas, 2009).

According to Pochet, *et.al,* (2007), kernel *k*-means clustering approach is capable of handling high dimensional data while classical *k*-means need to follow some variable reduction method due to curse of dimensionality as explained in section 2.6.2.1. This is a very strong reason in using kernel *k*-means for clustering anthropometric data because it has more variables which *k*-means cannot handle efficiently.

3.2 *K*- means Clustering Approach

Among the clustering algorithms, *k*-means clustering is most popular in use due to lesser time complexity (which is *O(n)*,where n is the number of cases) (Halkidi, Batistakis, & Vazirgiannis, 2001). In this clustering algorithm, the number of clusters needs to be finalized in advance and the objects are assigned to the nearest initial cluster centres which are arbitrary selected. These centres are moved towards final locations at each step of iteration in such a way to minimize the sum of squared distance from its objects to the centre of the clusters. It iterates several times until the convergence occurs in a way that the intra-cluster squared distances is minimized while inter-cluster distance is maximized. This algorithm results in clusters which are approximately equal in size.

This algorithm has two major problems: the resultant clusters depend on the initial positions of the cluster centres which may trap in local optimal solutions and it can only find linearly separable clusters in input data space (Tzortzis & Likas, 2009). Since anthropometric data are not linearly separable due to complex combinations among variables (Appendix 4.1) *k*-means approach might not work well. According to Han, Kamber and Pei,(2012), for high dimensional data, the traditional distance measures can be ineffective, and hence finding clusters using *k*-means approach can be unreliable.

Outline of *k*-means algorithm

Input: Number of desired clusters *K* ,Data objects D= {d1, d2...dn}

Output: A set of *K* clusters

Steps:

1. Randomly elevate K data objects (as initial centres) from data set D.
2. Repeat;
3. Calculate the distance of each data object di (1 <= i<=n) from all *k* clusters C j (1 <= j<=k) and then assign data object di to the nearest cluster.
4. For each cluster j (1 <= j<=k)
5. Recalculate the cluster centre until no change in the centre of clusters.

(Na, Xumin, & Yong, 2010)

Matlab Syntax of *k*-means clustering algorithm is shown in Appendix 3.1 .

3.3 Global *K*- means Clustering Approach

This algorithm was proposed by Likas, Vlassisb, and Verbeekb, (2003) to solve the problem of initialization in *k*-means clustering approach. They have explained how global *k*-means clustering algorithm avoids local search procedure in *k*-means clustering algorithm and minimized clustering error. According to them, instead of selecting initial cluster centres arbitrarily, global *k*-means works in an incremental way of finding new cluster centres. That means, it starts with one cluster (k=1) and find the cluster centre using *k*-means algorithm and it is the optimal position of the cluster centre. Then it moves to two clusters (k=2) and the second best cluster centre is obtained after several iterations of k-means algorithm. By following this procedure, through sequentially adding clusters, optimal positions of the cluster centres can be achieved using k-means algorithm (Likas,2003). According to Likas *et. al.* (2003), instead of randomly selected initial values for all cluster centres, this incremental approach (adding one new cluster centre at each step) works better resulting globally optimal solutions with respect to the clustering error.

The Global *k*-means clustering method was modified by several other researchers such as "an efficient global k-means clustering algorithm" (Xie, Jiang, Xie & Gao, 2011), "modified global k-means algorithm" which is for clustering gene expression data sets (Bagirov &Mardaneh, 2006).

Outline of global *k*-means algorithm

Steps:

1. (Initialization) Compute the centroid x^1 of the set A:

$$X^1 = \frac{1}{m}\sum_{i=1}^{m} a^i, a^i \epsilon A, \qquad i = 1, \ldots\ldots m$$

   and set k = 1.

2. Set k = k + 1 and consider the centres $x^1, x^2, \ldots, x^{k-1}$ from the previous iteration.
3. Each point of A is the starting point for the k^{th} cluster centre, to obtain m initial solutions with k points $(x^1, \ldots, x^{k-1}, a)$; k-means algorithm is applied to each of them; keep the best k-partition obtained and its centres $x^1, x^2, \ldots, x^k$.
4. (Stopping criterion) If k = q then stops, otherwise go to Step 2.

   (Agrawal & Gupta,2013)

3.4 Kernel *K*-means Clustering Approach

According to Chitta, Jin, Havens, and Jain, (2011), "Kernel *k*-means is a nonlinear extension of the classical k-means algorithm. It replaces the Euclidean distance function employed in the *k*-means algorithm with a non linear kernel distance function". According to Zhang and Rudnicky (2002), "the incorporation of kernel functions enables the *k*-means algorithm to explore the inherent data pattern in the new space". Kim, Lee, Lee and Lee (2005) and Zhang and Rudnicky (2002) showed that kernel *k*-means algorithms perform better and significantly more accurate than conventional *k*-means algorithms in unsupervised classification.

According to Cristianini, Taylor, and Saunders, (2007), different kernel functions are available and selection of relevant kernel function and its parameters depends heavily on the type of data itself. The widely used kernel functions are as follows:

(i) Polynomial kernel of degree p:

$K(X, Y) = (X \cdot Y + c)^{p},$ where X and Y are vectors in the input space and p is a positive integer and $c \geq 0$ is a free parameter.

(ii) Gaussian (Radial Basis Function) kernel:

$K(X, Y) = exp\ (-\|X-Y\|^2/2\sigma^2)$

Where σ is called Gaussian kernel parameter which is a real number ($\mathbb{R}$).

(iii) Sigmoidal kernel :

$K(X, Y) = \tanh(a(X \cdot Y) + b)$, where $a, b \in \mathbb{R}$

According to Cristianini, *et.al*, (2007), Gaussian (RBF) kernels are the most widely used kernels in different subject areas. If X and Y are two feature vectors, $\|X-Y\|^2$ in RBF kernel can be considered as the squared Euclidean distance between the two feature vectors and it range from 0 to 1 (when $X = Y$).

For $\sigma = 1$, the expansion of RBF kernel function is as follows;

$$\exp\left(-\frac{1}{2}\|X-Y\|^2\right) = \sum_{j=0}^{\infty} \frac{(X^T Y)^j}{j!} \exp\left(-\frac{1}{2}\|X\|^2\right) \exp\left(-\frac{1}{2}\|Y\|^2\right)$$

Therefore, the feature space of the RBF kernel has an infinite number of dimensions by mapping every point to an infinite dimensional space (Amnon, 2009).

According to Lin and Lin (2003), sigmoid kernel function behaves like RBF kernel for a range of parameter values and behavior is unknown for some parameter values. Furthermore, he recommended the RBF kernel for general users and hence, using sigmoid kernel was avoided. In high dimensional feature space the data can be separated linearly which in turn separate input data nonlinearly identifying the non-linear structures in input space (Sarma, Viswanath, & Reddy, 2011). In kernel *k*-means clustering algorithm, the kernel matrix is pre computed using a kernel function and stored. Here, the time and space requirements are $O\ (n^2)$ where n is the size of the sample. This is a drawback of kernel *k*-means clustering method when handling very large data sets and another drawback is the need of prior knowledge about the number of clusters (k) (Kim, *et al.*, 2005; Chitta, *et al.*,2011).

According to Chitta, *et.al*, (2011), kernel k-means clustering method substitute the Euclidean distance function $d^2(\varkappa_a, \varkappa_b) = \|\varkappa_a - \varkappa_b\|^2$ employed in the k-means algorithm with a non-linear kernel distance function defined as;

$d^2(\varkappa_a, \varkappa_b) = k(\varkappa_a, \varkappa_a) + k(\varkappa_b, \varkappa_b) - 2k(\varkappa_a, \varkappa_b)$, where $\varkappa_a \in \mathbb{R}$ and $\varkappa_b \in \mathbb{R}$ are two data points and k (.,.) is the kernel function.

Algorithm of kernel k-means clustering approach could be outlined as follows (Tzortzis & Likas,2009).

Outline of kernel *k*-means algorithm

Input: Kernel matrix K, Number of cluster k, Initial clusters $C_1,\dots C_k$

Output: Final clusters $C_1,\dots C_k$, Clustering error E

1. For each point x_n and every cluster C_i compute
   $\| \phi(x_n) - m_i \|^2$ using (3)
2. Find $c^*(x_n) = \text{argmin}_i (\| \phi(x_n) - m_i \|^2)$
3. Update clusters as $C_i = \{ x_n \mid c^*(x_n) = i \}$
4. If not converged go to step 1 otherwise stop and return final clusters
   $C_1,\dots C_k$ and E calculated using (2).

Refer Appendix 3.2 for the Matlab syntax for kernel k-means clustering method (Tzortzis & Likas, 2009).

3.5 Global Kernel *K*- means Clustering Approach

This algorithm combines the advantages of both global k-means and kernel k-means, and therefore it avoids both limitations; local minima problem and linearly separable clusters of k-means clustering algorithm and produces a final partition which is independent of the cluster initialization (Tzortzis & Likas, 2009). This approach identifies linear clusters in high dimensional feature space and it becomes nonlinear clusters in input data space. However, its computational complexity and time complexity are high.

It was proved that global kernel k-means clustering method is better in terms of clustering error as compared to kernel k-means method by using artificial datasets and MRI images (Tzortzis & Likas, 2009). Outline of the global kernel k-means clustering algorithm can be given as follows.

Outline of global kernel k-means algorithm (Tzortzis & Likas, 2009)

Input: Kernel matrix K, Total number of clusters M

Output: Final clustering of the points$C_1, C_2,\ldots,C_M$

There is no need to solve for one cluster as the solution is trivial and optimal.

$C_1^* = X$

1. Solve all k- clustering problems for $k = 2$ to M
2. For each such problem run kernel k-means N times for $n = 1$ to N with input $(K,\ k,\ C_1^*,\ldots,\ C_{k-1}^*, C_k^* = \{\ x_n\ \})$ and output $(C_1^n,\ldots,\ C_k^n, E_k^n)$
3. Find $E_k^* = min_n\ (E_k^n)$ and set $C_1^*,\ldots, C_k^*$ to the partitioning corresponding to E_k^*(this is the solution with k clusters).
4. Set $C_1 = C_1^*,\ \ldots,\ C_{M,} = C_M^*$ as output of the algorithm.

In the kernel-based data analysis process, input data was first embedded to a high dimensional feature space via a kernel function. Therefore, selection of a suitable kernel function is very important because it heavily depends on the dataset. In addition, tuning of kernel parameters is also very critical because it gives vital impact on the result.

3.5.1 Tuning of kernel parameter

Kernel parameter tuning is very important because the result heavily depends on it. In literature, there are several methods found for sigma (σ) tuning and in this research, two methods were considered and tested.

(i) Using kernel-based Dunn's index

According to Pochet,*et.al* (2007), cluster optimization techniques such as Dunn's Index, Davis-Bouldin (DB) Index, BIC index could be used for this tuning of parameters. Since, data analysis was done with kernel based algorithm in this research, a cluster optimization technique combined with kernel distances need to be used. As explained under section 3.6.1, Dunn's index was selected and kernel- based Dunn's Index (KDI) was developed in order to tune the kernel parameter and validate the clusters. In KDI, kernel distance between data points in clusters were taken instead of Euclidean distance (Appendix 5.6). In tuning the sigma parameter, different sigma values could be used to create kernel matrices and clustered it using global kernel k-means clustering approach. For each clustered data set, KDI could be calculated and the highest value could be selected. The sigma value that relevant to the highest KDI is the optimum value for sigma.

(ii) Using variance of the kernel matrix

According to Evangelista, Embrechts and Szymanski (2007) and Yunzhang (2011), highest variance of the kernel matrix which is resulted by a sigma value is the optimum value for sigma. Therefore, different sigma values can be used to create kernel matrices and then checked for its variances. The sigma value which gives the highest variance of the kernel matrix is the optimum value.

3.6 Validation of Clusters

There are two types of cluster validation methods; External cluster validation and internal cluster validation. For external validation, previous knowledge (class label) about dataset is required while internal validation is based on the information incorporated to the dataset (Rendón, Abundez, Arizmendi, & Quiroz, 2011). However, most data sets, especially anthropometric datasets, do not have prior class label to check with the cluster yielded and thus making external validity criterion non-applicable. Hence, number of clusters can be optimized using internal cluster validity indices which measure intra-cluster cohesion that is to measure how closely objects are related in a cluster and how significant the inter-cluster separation which is a measure of distinct or well-separated a cluster from other clusters (Rendón,*et al.*,2011). Further, according to Rendón,*et al.*, (2011), internal indices are more precise in real group number determination than external indices. These methods can be used not only to optimize the number of clusters but also to tuning the kernel parameters (Pochet, *et.al,* 2007). There are several internal cluster validity indices such as Dunn Index, Davis-Bouldin (DB) Index, BIC index, and Silhouette Index which are used in the input data space (Rendón, *et al.*, 2011). However, extended versions are needed in order to use with the feature space.

In kernel *k*-means clustering, depending on the kernel functions, tuning of kernel parameters is required. For example, in RBF kernel function, kernel parameter σ can be optimized. For this purpose, an internal cluster validity method can be used while it is used to optimize the number of clusters as well (Pochet, *et.al* ,2007).

3.6.1 Kernel- based cluster validity index

Fa, Nandi, and Jamous (2012) explained clearly how existing cluster validity indices are extended to kernel-based cluster validity indices. They have extended conventional indices; Dunn's index, the I-index, the Calinski Harabasz (CH) index and the geometrical index (GI), which calculate the ratio of the intra-cluster compactness to the inter-cluster separation in the input space.

The Dunn's index (Dunn, 1973) measures the ratio between the smallest inter-cluster distance and the largest intra-cluster distance and hence maximum value of Dunn index is corresponding to the optimal number of clusters. According to Bezdek and Pal (1998), although Dunn's index is sensitive to outlier points badly, it provides strong foundation for cluster validity indices for different types of clusters. Hence, several generalizations were suggested to reform this index (Pal & Biswas,1997; Bezdek & Pal, 1998).

Dunn's index is defined by Fa, Nandi, and Jamous, (2012), and it can be mathematically express as follows:

$$DI(K) = \min_{1 \le i \le K} \left\{ \min_{1 \le j \le K} \left\{ \frac{\delta(C_i, C_j)}{max_{1 \le k \le K} \{\Delta(C_k)\}} \right\} \right\}$$

Dunn's Index was extended to kernel Dunn's Index in the following way (Fa, Nandi, and Jamous, 2012);

$$kDI(K) = \min_{1 \le i \le K} \left\{ \min_{1 \le j \le K} \left\{ \frac{\delta^K(C_i, C_j)}{max_{1 \le k \le K} \{\Delta^K(C_k)\}} \right\} \right\}$$

Where $\Delta^{\kappa}(C_{\kappa})$ is the largest intra-cluster separation of cluster *k* in the feature space, $\delta^{K}(C_i, C_j)$ is the minimum of kernel based distance between cluster *i* and cluster *j* in the feature space.

Phillips and Venkatasubramanian, (2011) defined the kernel distance between two points *p and q* as follows;

$$\begin{aligned} D_K^2(\{p\},\{q\}) &= K(p,p) + K(q,q) - 2K(p,q) \\ &= 2(1 - K(p,q)) \end{aligned}$$

Here, the expression $1 - K(p,q)$ acts as a squared distance between the two points and hence has the following characteristics:

- If the two points are identical, this distance is zero
- $\mathcal{D}_k(p,q) = \mathcal{D}_k(q,p)$
- (iii)As the points become more distant, the self similarities remain the same while the cross similarity decreases, when $\mathcal{D}_k$ increases.

3.7 Chapter Summary

In this chapter on kernel- based clustering method, the global kernel k-means clustering method, which was developed by Tzortzis and Likas, (2009) and successfully deployed for clustering artificial datasets as well as for MRI segmentation, was discussed. The author of the thesis selected global *k*-means algorithm for clustering the anthropometric data based on the following rationale.

- kernel- based learning is one of the latest techniques found in data mining and it has a greater potential in clustering data successfully
- this method is free of problems in conventional k-means clustering technique
- it combines the advantages of both kernel k-means and global k-means clustering
- it is suitable for handling large number of dimensions without reducing the number of variables
- it constitutes linearly separable clusters in high dimensional feature space
- this method has not be tried out in the development of size charts

Kernel-based Dunn's index was selected as a cluster validation technique, which also can be used to optimize the number of clusters, as well as to tune the parameters of RBF function used for transformation. Application of the kernel- based Dunn's index is also a new approach to the subject of size chart development. Although the concept of kernel- based Dunn's index was formulated in 1973 its application is very rare in clustering data found in the literature.

The next chapter will be devoted to the methodology of the research.

CHAPTER 4
RESEARCH METHODOLOGY

4.1 Introduction

In chapter 3, kernel- based clustering approaches were reviewed and the global kernel k-means clustering approach was discussed in detail. Since this approach is capable of handling high dimensional data without variable reduction, it is highly appropriate in clustering anthropometric data. Further, this approach can find linearly separable clusters in feature space which is not linearly separable in input space. Since anthropometric variables have complex combinations among populations which cannot separate linearly, the global kernel k-means clustering approach will be successful in clustering this dataset. While, with kernel-based Dunn's index, optimizing the number of clusters as well as the tuning the parameters of transformation function are possible, cluster validation techniques could refine the potential of the scope of applicability.

This chapter is presented under three main sections: data collection, methodology for analyzing existing approaches and development of a new approach. Under data collection, sample selection criteria and procedure of anthropometric data collection are explained in detail. In the second section, the methods of applying existing size chart development approaches to Sri Lankan female lower body anthropometric dataset are explained. In the third section, the methodology for a novel approach with size chart validation methods is described.

4.2 Process of Anthropometric Data Collection

4.2.1 Pilot study

A pilot study is always helpful as a preliminary analysis before going to the larger research design. Therefore, a pilot study was carried out in order to understand the following:

- the anthropometric data collection process and its limitations
- ethical issues when working with females
- most convenient method for recording

- time duration for the process
- training the inexperienced assistants to collect data accurately

The pilot study of anthropometric data collection was started in January 2012. The lower body measurements of female students (university and several institutions of higher studies) of age between 20-30 years were collected. The detail procedure on data collection is explained under section 5.2.3. For the pilot study, a sample of 80 females was selected and measured. Therefore, the data collection process could be planned efficiently through the experience gained from this pilot study.

4.2.2 Sample selection

In Sri Lanka, where people are strongly influenced by social and religious norms, getting consent for measuring the lower body of females is difficult. Therefore, selection of a random sample is practically impossible. Hence, convenience sampling technique was selected for collecting anthropometric data. However, to make the selected sample to be more representative of the population, females from each province of the country, different races and different professions were included. Furthermore, according to Abeysekera and Shahnavaze (1998), there were no significant differences in the measurements of human body for both female and male populations, which indicated a homogeneous distribution of physical structures of people in Sri Lanka, but the findings do not bear any specific emphasis to lower body measurement. Therefore, selection of convenience sampling might not affect the final result adversely.

According to the Department of Census and Statistics records in 2001, the female population in the age group of 20 – 40 years was 3.4 million . According to Bartlett, Kotrlik, and Higgins (2001), the confidence interval of 95 % (alpha level of 0.05) is acceptable in determining a sample size in educational research studies while acceptable margin of error for continuous data is 3 %. Therefore, for 95% confidence level and 3% margin of error, the sample size estimation is found to be 1068. However, a total of 1430 females were measured in order to reserve a sample for validating the results.

4.2.3 Selection of age limit

The selected age limit of the sample for the study was 20- 40 years. According to Croney (1980), an average of 20 years is required for human beings to develop to full maturity in physical structure, which the rational for selecting the above lower limit of the age. Furthermore, after secondary education in Sri Lanka, the young population enters their twenties and they start their higher education or career life and they are the most potential customers in fashion apparel products. According to the questionnaire survey done in 2011 among Sri Lankan females to find problems related with ready-to-wear apparels, it was revealed that female pants has a significant place in young generation's apparel choices. Therefore, taking the above facts into consideration, age 20 was selected as the lower limit of the age range.

Female's menopause occurs in midlife, during their late 40s or early 50s and it causes to the changes to the body measurements and shape due to the changes in fat accumulation patterns as a result of the changes in body hormones. In addition, at around 40 years of age, human beings begin to shrink in stature and women shrink more than men (Iseri, 2008). Therefore, females above 40 years need to be considered separately in developing size charts. Furthermore, it is seemed that the current adult female population in Sri Lanka, except in Colombo suburbs, do not prefer pants, which may be due to social, cultural bounds, though this trend will change with the time. Therefore, it was found difficult in getting lower body anthropometric data from adult females. Hence, these facts caused for the selection of 40 years of age as the upper limit.

4.2.4 Method of anthropometric data collection

First, thirteen dimensions of the lower body of females' were identified for the development of size chart for pants, and these dimensions were measured by using a tape measure as explained in section 4.2.4.2. The definitions of body dimensions and the procedures of taking body dimensions were followed according to Beazley (1996) who followed the ISO8559:1989 protocol (quoted in parentheses in section 4.2.4.2.)

4.2.4.1 Preparation for measuring

All the participants in the sample were always informed fully about the purpose and the procedure of the measuring process in order to obtain their consent before taking measurements. A separate room was arranged to protect their privacy and two female assistants who work as a measurer and a recorder were there with the participant. The participant was asked to wear a pair of leggings (stretchable pant) before taking measurements. The participant was standing up barefoot with her hair tied up. The measurements were taken from the right side of the participant. A fabric tape was fastened around the natural waist line and landmarks, (adhesive small arrow stickers), were positioned on the participant as follows (Fig. 4.1):

(i) Hip level: The widest hip girth level is located. Landmarks were positioned at the centre back, right side and left side.

(ii) Knee level: The subject bends her right knee slightly to define the crease line of the tibial knee joint. (refer appendix 2.2)

(iii) Ankle level: The lower edge of the tibial bone is located on the right ankle of the subject.

(iv) Mid Thigh level: the mid level in between crotch level and knee level.

(v) Mid calf level: the mid level in between knee level and ankle level.

Fig. 4.1 Positions of the landmarks on lower body

4.2.4.2 Procedure of females' lower body measuring

(Measurement Definitions; Beazley,1996)

A trained female assistant was employed to measure the subjects while an assistant recorded those readings in a record sheet. Additionally, other demographical information such as age, native province, profession of the subject was also noted. Thirteen measurements that were taken from the lower body of the subject were as follows;

(i) Waist girth (ISO8559 2.1.11)

This is the circumference of the waist at the natural waist level. The measuring tape was held firmly around the waist level and the reading was recorded. Appendix 4.1 (i) illustrate the typical way of obtaining the waist girth measurement.

(ii) Hip girth (ISO8559 2.1.12)

This is the maximum circumference of the hips at the level of the maximum posterior protrusion of the buttocks. The measuring tape measure was held firmly over the hip level landmarks and the reading was recorded [Appendix 4.1 (ii)].

(iii) Thigh girth (ISO8559 2.2.18)

The measurement was taken horizontally around the thickest part of the right upper thigh just below the crotch level [Appendix 4.1 (iii)].

(iv) Mid thigh girth

The measurement was taken horizontally over the mid thigh landmarks.

[Appendix 4.1 (iv)].

(v) Knee girth

The measurement was taken horizontally over the knee landmarks.

[Appendix 4.1 (v)].

(vi) Mid calf girth

The measurement was taken horizontally over the mid calf landmarks.

[Appendix 4.1 (vi)].

(vii) Ankle girth (ISO8559 2.1.24)

The measurement was taken horizontally over the ankle landmarks.

[Appendix 4.1 (vii)].

(viii) Inside leg from ground (Inseam) (ISO8559 2.2.27)

The measurement was taken from the crotch level of the subject to the ground level.

[Appendix 4.1 (x)].

(ix) Side waist to ground (out seam) (ISO85592.2.25)

The measurement was taken from the lower edge of the waist tape vertically down to the ground level from the right side [Appendix 4.1 (xi)].

(x) Crotch length (ISO8559 2.2.19)

The measurement was taken from the centre back waist level between the legs to the centre front waist level [Appendix 4.1 (ix)].

(xi)Centre front waist to hip

The measurement was taken vertically between the lower edge of the waist tape to the centre front hip landmark [Appendix 4.1 (viii)].

(xii) Knee height from ground

The measurement was taken vertically from the knee landmark to the ground.

[Appendix 4.1 (xii)].

(xiii) Ankle height from ground

The measurement was taken vertically from the ankle landmark to the ground.

[Appendix 4.1 (xiii)]

4.3 Methodology for Analyzing Existing Approaches

This section explains the methodology for analyzing the existing approaches: the approach based on descriptive statistics, *k*-means clustering algorithm combined with factor analysis and classification and regression decision tree approach, by using lower body anthropometric data of a sample of 1068 Sri Lankan females. Through this exploration, it is intended to recognize the drawbacks pertaining in the above approaches in developing size charts. Analysis was carried out using SPSS version 16 software.

4.3.1 The approach based on descriptive statistics

The approach was studied and the procedure was applied to the lower body anthropometric dataset of Sri Lankan females. This dataset consists of thirteen dimensions; waist girth, hip girth, thigh girth, mid thigh girth, knee girth, mid calf girth, ankle girth, outside leg length (outseam), inside leg length (inseam), crotch length, knee height, waist to hip height and ankle height.

Out of the thirteen dimensions, key dimensions (control variables) need to be selected in order to develop size charts. It was decided based on the correlation between dimensions (Table 5.11). The dimensions which had strong correlation of 0.7 or over with many other dimensions were selected as key dimensions.

Afterwards, standard size intervals for waist and hip were taken as 4cm as explained by Beazley (1997). Then, three inseam categories were identified as short (65-71 cm), medium (71.1-77 cm) and tall (77.1-83 cm) for the lower body to be on par with the height categories in existing size charts. Descriptive statistics of the dataset under different height categories was obtained. Using minimum, maximum, range and standard size intervals, key dimensions were divided into several segments and secondary dimensions (other dimensions in the dataset except key dimensions) were distributed among above sizes empirically determining the size intervals.

4.3.1.1 Validation of the size chart

Aggregate loss of fit factor was calculated using a validation sample of 362 anthropometric dataset as explained in section 2.7.1 and three key measurements (waist, hip and inside leg length) were considered for the calculation.

4.3.2 *K*-means clustering approach with factor analysis

This analysis method was applied on the lower body anthropometric dataset of Sri Lankan females. Before applying factor analysis, the Kaiser-Meyer-Olkin measure of sampling adequacy (KMO) which is a statistic that indicates the proportion of variance in the variables that might be caused by underlying factors, was checked. Further, Bartlett's Test of Sphericity which is to see whether the correlation matrix deviates significantly from the identity matrix, was checked. From the table of "Total variance explained" and Scree plot number of factors were decided based on the Eigen value greater than one criterion. Further, from Varimax rotated solution, which results in orthogonal coordinates maximizing the sum of the variances of the squared loadings, number of factors can be confirmed. The sample has been clustered based on factor scores using *k*-means algorithm.

Hierarchical clustering method was used to identify the possible number of clusters in the dataset. This number of clusters were used in the *k*-means algorithm to cluster the factor scores and check the behavior of variables inside the clusters.

4.3.3 Classification and regression (CART) decision tree approach

The CART decision tree approach was applied to the anthropometric dataset of Sri Lankan females in order to investigate the possible potential and the drawbacks of the method in developing a size chart. Since this dataset consists only of lower body measurements, waist to hip ratio (waist / hip) has been used as the generated target variable for explaining the decision tree approach. Waist girth and thigh girth are selected as predictor variables to be on par with past research. The resulted decision tree was analyzed in order to develop a size chart.

4.4 Development of New Approach

The new approach, global kernel *k*-means clustering, was applied to the lower body anthropometric dataset of Sri Lankan females. The Fig. 4.2 shows the flow chart of the data mining process of the research and the sequence of the sub-processes in a nutshell. It illustrates the data preparation procedure, data analysis procedure and result evaluation procedure in detail.

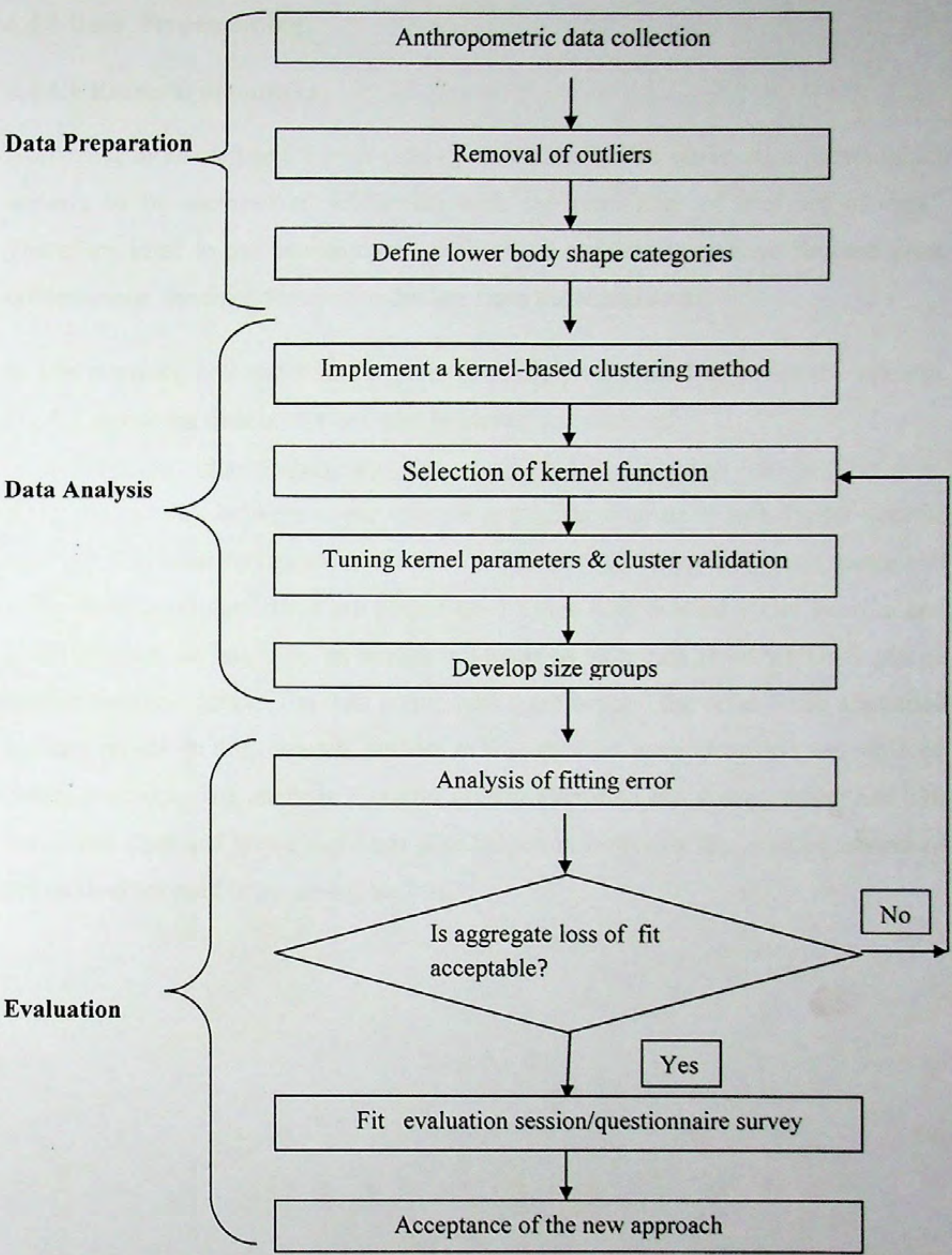


Fig.4.2 Flow chart of data analysis process

4.4.1 Data Preprocessing

4.4.1.1 Removal of outliers

According to Barnett and Lewis (1994), an outlier is "an observation (few) which appears to be inconsistent (different) with the remainder of that set of data". Therefore, prior to the data analysis, outliers are required to remove because those outliers cause the final outcome to deviate from the actual result.

In this research, box and whisker plots (box plot) were used to detect the outliers. Fig.4.3 shows the details of a box plot in identifying outliers.
(https://nelsontouchconsulting.wordpress.com/2011/01/17/deeper-into-box-plots)
Here, the distance between upper quartile and lower quartile is called inter-quartile range (IQR). Inner fences are decided on 1.5 times IQR beyond upper quartile and lower quartile. Outer fences are placed on 3 times IQR beyond upper quartile and lower quartile. In box plot, an outlier is identified as a data point which is placed outside the inner fence. The data points which are beyond the outer fence are called extreme points. In this research, outliers in the variables were identified and replaced before preceding the analysis. Cooklin (1995) identified hip sizes greater than 110 cm as out sizes and hence that basis also helped in justifying the appropriateness of the method adopted in removing outliers.

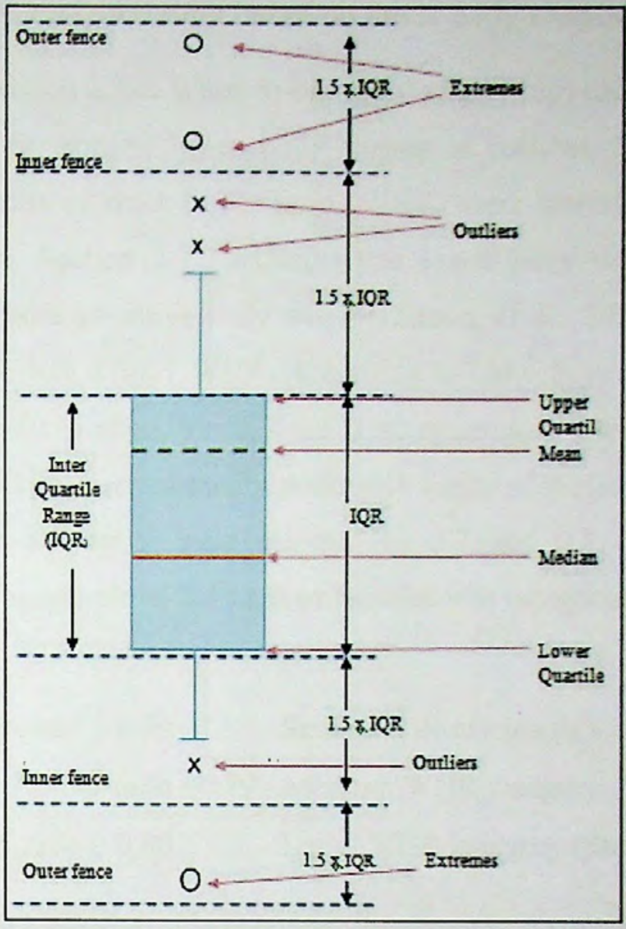


Fig. 4.3 Details of a box and whisker plot

4.4.1.2 Identifying key measurements of the lower body of females

As explained in section 2.4, key measurements of the lower body of females were identified based on following aspects:

- the variables (measurements) which showed high Pearson correlation coefficients with other variables
- the variables which showed higher factor loadings of final factors resulted in factor analysis
- evidence from literature
- experience and knowledge of the experts in the apparel field

4.4.1.3 Initial data categorization based on lower body shapes

As mentioned in section 2.3.2, Waist-to-hip ratio (waist / hip) can be considered as a standard measure for sorting lower body shapes of females. Hence, referring to literature, three limits of waist-to-hip ratio (WHR) were identified to segment the lower body shapes. Section 2.3.2 explains that lower body with WHR of 0.7 is considered as the most attractive body shape (Dixson, *et al.*, 2010) while European men prefer females with 0.6-0.7 WHR. According to Lidwell, *et al.* (2003), 0.67-0.8 WHR is preferable for women. Since, there is no systematic literature regarding Sri Lankan females' WHR, internationally preferable limits of WHR were considered in this research. Cut off limits were selected as 0.7 and 0.8 considering above-mentioned facts. The sample of Sri Lankan females was categorized in the following way:

- waist-to-hip-ratio ≤ 0.70 - Small WHR category (Curvy lower body)
- $0.71 \leq$ waist-to-hip-ratio ≤ 0.79 - Medium WHR category
- waist-to-hip-ratio ≥ 0.80 - Large WHR category (Straight lower body)

Here, the small WHR category encompassed females who have small waist and large hip resulting in a curvy lower body. The females whose difference between waist and hip is less, resulting in a straight lower body, formed the large WHR category, while the medium WHR category included females who have average waist and hip measurements. After initially dividing the dataset into the above three categories, further analysis was carried out separately on these three datasets.

4.4.2 Data analysis procedure

As explained in section 3.5, Global kernel *k*-means clustering approach was selected for analyzing the lower body anthropometric data of Sri Lankan females (Matlab code in appendix 4.4). For analysis, Matlab version7.7 software and SPSS version 16 were used. Two kernel functions, polynomial and Gaussian (RBF) kernel functions were tested with different parameters.

As explained in section 3.1, input data transform to a kernel matrix which is a symmetric, (n×n) square kernel matrix (where n is the number of females in the sample) through the selected kernel function. This kernel matrix is the input for global kernel k-means clustering algorithm.

4.4.2.1 Global kernel k-means algorithm with polynomial kernel function

From the most popular kernel functions, polynomial kernel function which is given below, was first tested with different "degrees". This approach was carried out to investigate whether a noticeable linear relationship exists even after categorizing the dataset into three segments.

Polynomial kernel of degree p can be written as:

$K(X, Y) = (X \cdot Y + c)^p$, where p is a positive integer and $c \in \mathbb{R}$.

The default value for polynomial degree was two, and hence, initially, d=2 was selected to create the kernel matrix. Afterwards, different "degree" values such as 3 and 4 were used to create kernel matrices. For d< 1, it was found that kernel matrix is not a valid one as it is not positive semi-definite (negative eigen values in the kernel matrix). Gamma (c) value was taken as one because it's impact on the kernel matrix is very small when overseeing the resulted matrix. Afterwards, this kernel matrices were sent through the clustering process using global kernel k-means clustering algorithm and this process was repeated for different cluster numbers such as 2, 3, and 4 (Matlab code used was given in Appendix 5.4). However, this approach could not yield an acceptable level of success as explained under section 6.3.1. The failure compels to move to the Radial Basis Function (RBF) as the kernel function.

4.4.2.2 Global kernel k-means clustering algorithm with Gaussian (RBF) kernel function

In the second step Gaussian kernel function, which was given below, was selected as the kernel function for transformation.

$K(X, Y) = exp(-\|X-Y\|^2 / 2\sigma^2)$, where $\sigma \in \mathbb{R}$ is called Gaussian kernel parameter.

4.4.2.3 Tuning of kernel parameter

(i) Using kernel-based Dunn's index

In tuning the sigma parameter, different sigma values were used to create kernel matrices and clustered it using global kernel *k*-means clustering approach. For each clustered data set, KDI was calculated and the highest value was selected. The sigma value that relevant to the highest KDI was the optimum value for sigma.

(ii) Using variance of the kernel matrix

Different sigma values were used to create kernel matrices and then checked for its variances (refer Appendix 4.5 for the Matlab code). The sigma value which gave the highest variance of the kernel matrix, was selected as the optimum value.

Then, the kernel matrix which was produced with optimum sigma value (Refer Appendix 4.3), was used to proceed with the clustering process using Global kernel *k*-means clustering approach (Appendix 4.4). Since prior knowledge of the number of clusters was not available, the clustering process was repeated for several numbers of clusters. The best number of clusters was recognized through cluster validation. Here, kernel- based Dunn's index (Appendix 4.7) was used as the cluster validation approach as explained in section 3.5.1. Two dimensional scatter plots were also used to view the behavior of the variables. This procedure was repeated for all three categories of datasets.

4.4.3 Development of size charts

The data sets which were clustered were separated according to cluster number to develop size charts. Three different size charts need to be developed for three categories of lower body shapes that were identified as small WHR (curvy lower body), medium WHR and large WHR (straight lower body). First, in the dataset which was separated according to the cluster number, range, minimum and maximum values of key variables (waist, hip and inseam) were found for deciding the size intervals and number of sizes within that relevant cluster. Size intervals were selected considering the acceptable tolerances of key variables for the consumer. Further, reducing the size interval may result unnecessary number of sizes making the size chart complicate. Therefore, size intervals for waist and hip measurements were selected as 4 cm which may result the maximum tolerance of ± 2 cm which is acceptable. Further, Beazley (1997) also recommended 4 cm size intervals for waist and hip measurements analyzing past international size standards. In addition, Beazley (1997) explained that larger sizes require a larger interval of 6 cm for the waist and hip measurements because beyond a certain range, the garments become disproportional.

In deciding the inseam length interval, Table 4.3 shows inseam categorization of female pants in different apparel brands (http://www.nike.com/us/en_us/c/size-fit-guide,http://bananarepublic.gap.eu,http://oldnavy.gap.com, https://www.ae.com/sizechart/). It shows that in UK, USA, and Europe, females have long legs and their shortest inseam height is around 29.5 inches. From the data set of Sri Lankan females, it was found that majority have average inseam length of 74 cm (29.5 inches) and they are shorter than UK,USA regular height females. It was also confirmed by Abeysekera and Shahnavaze (1998) that in developing countries, including Sri Lanka, people are shorter than that of developed counties. Accordingly, three inseam groups were identified from the data set of Sri Lankan females based on minimum, maximum inseam length and international standards as follows; short(68 cm/26.5 inches), Regular (74cm/29 inches) and Tall (80cm/31.5 inches) maintaining 6 cm size interval.

The identified three key measurements (waist, hip, inseam) were separated into size groups maintaining above mentioned size intervals, and mean values of the secondary variables (thigh, mid thigh, knee, mid calf, ankle, outseam, crotch length, hip height, knee height and ankle height) relevant to that size group were also added to complete the size chart.

Table 4.1 Inseam categorization of female pants in different apparel brands

Inseam	**Nike**	**American Eagle outfitters**	**Old Navy**	**Banana Republic**
Short	29.5 inches	30 inches	30 inches	33 inches
Regular	31.5inches	32.5 inches	32 inches	33 inches
Long	33.5inches	34 inches	34 inches	33 inches

4.4.4 Size chart validation

Two methods were used to validate the developed size charts: using statistics and by live fit assessment sessions followed by a questionnaire survey.

4.4.4.1 Validation by statistics

Size chart validation was done by calculating "aggregate loss of fit" factor as explained in section 2.7.1, which measures the difference between assigned size and actual body measurements. For this purpose, "validation sample" of anthropometric data which consists of 362 females was used.

4.4.4.2 Validation by live fit assessment sessions

The second method was "live fit assessment sessions" which access the fit of pants which were produced according to the developed size chart. For the fit session, twenty participants were arbitrarily selected and their waist, hip and inseam measurements were taken after getting their consent for the process. Among them, participants who belong to different sizes representing three shape categories of new size chart were selected. Accordingly, twelve participants were selected and the rest, who have similar measurements with selected participants were informed and removed. Replacement could not be done due to time and cost constraints. The fabric selected for samples was medium weighted calico (1×1 plain weave muslin of 156.4 gsm) as a common practice (Veblen, 2012), and its neutral color clearly shows the issues of the pants. Then, required ease allowances were decided based on Hagger (1990) and patterns for the pants were produced. Accordingly, samples were stitched and fitted-on by the participants. Then, photographs of front, back and side views of the lower body of the participants were taken. A pre-tested questionnaire of Bickle, *et al.* (1995) was used to get the feedback regarding the fit of the pants (appendix 5.7). The survey result was statistically analyzed to identify the causable reasons amidst the possible coincident due to randomness.

4.5 Chapter Summary

In this chapter, the research methodology was explained in detail. Under the data collection section, the anthropometric data collection procedure, which includes sample selection, measuring points of the body and method of body measuring was described clearly. The methodology of applying existing size chart development approaches to lower body anthropometric dataset of Sri Lankan females was also explained. Subsequently, methodology of applying the new approach was discussed. Data preprocessing procedure to the removal of outliers, initial data grouping based on lower body shapes and identifying of key variables were explained. Then, the method of data analysis, global kernel k-means clustering approach which is a new approach for anthropometric data clustering were presented. The result validation procedure which comprises of statistical method and live fit assessment method, was explained in the latter part of the chapter. The next chapter presents the results and discussion of the research.

CHAPTER 5

RESULTS AND DISCUSSION

5.1 Introduction

The methodology of the research which explained the data analysis process in detail was covered in Chapter 4. This chapter explains the results of the research under three sections. Under the first section, results of data collection process are presented. Then, the results of existing size chart development approaches are discussed and in the final section, results of the new approach are presented.

The method of "Global kernel *k*-means algorithm" which was used in MRI segmentation in literature, was used successfully in clustering of lower body anthropometric data of Sri Lankan females with added sophistications to optimize the number of clusters and to tune parameters of kernel transformation functions. Further, selected clusters were validated with a set of separate data which was not used for the development of the size charts.

5.2 Results of Data Collection

The selected sample consisted of females from different professions, different social backgrounds and different races within the age limit representing all the provinces in Sri Lanka. Table 5.1 shows the provincial distribution of the sample. However, getting consent for measuring the lower body from Tamil and Muslim females was found to be difficult, resulted in poor representation from Northern and Eastern provinces to the sample.

Fig.5.1 shows the pie chart of the distribution of the sample based on the provinces in Sri Lanka. It shows that the Western province has the highest representation, while the Northern and Eastern provinces have less representation.

Table 5.1 Provincial distribution of the sample

province	Northern	North Central	North Western	Central	Eastern	Western	Sabaragamuwa	Uva	Southern	Total
No of women	22	108	114	192	21	501	162	144	166	1430
%	1.54	7.55	7.97	13.43	1.47	35.03	11.33	10.07	11.61	100%

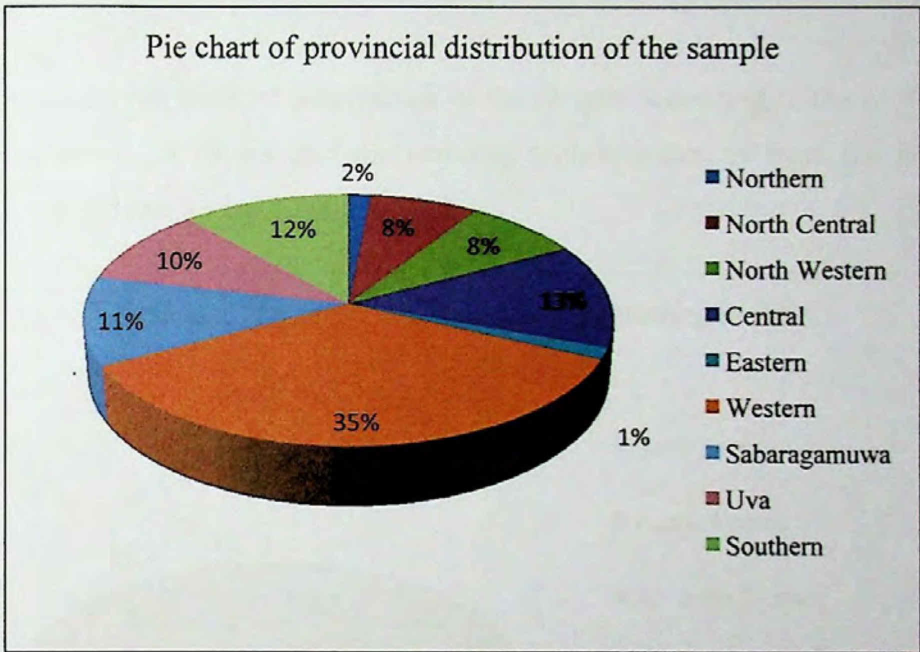


Fig.5.1 Provincial distribution of the sample

Table 5.2 shows the distribution of the sample based on professions of the participants. The sample consists of female personnel from Army, Navy, Air force, University students, students of higher education institutions, working females in different private companies, self employed females and non-employed females of aged 20-40 years, representing nine provinces in Sri Lanka. Addition of Navy, Army and Air force females made the sample more balanced.

Table 5.2 Distribution of the sample according to the profession

Profession	Navy women	Army women	Air force women	University students	Students in higher education institutions	Working females in private sector	Self-employed females	Non- employed females	Total
No .of Women	100	88	178	247	206	462	46	103	1430
Percentage	6.99	6.15	12.45	17.27	14.41	32.31	3.22	7.2	100%

Fig. 5.2 shows the pie chart of distribution of the sample according to the profession of the participants. It shows that the majority representation is from the females working in the private sector.

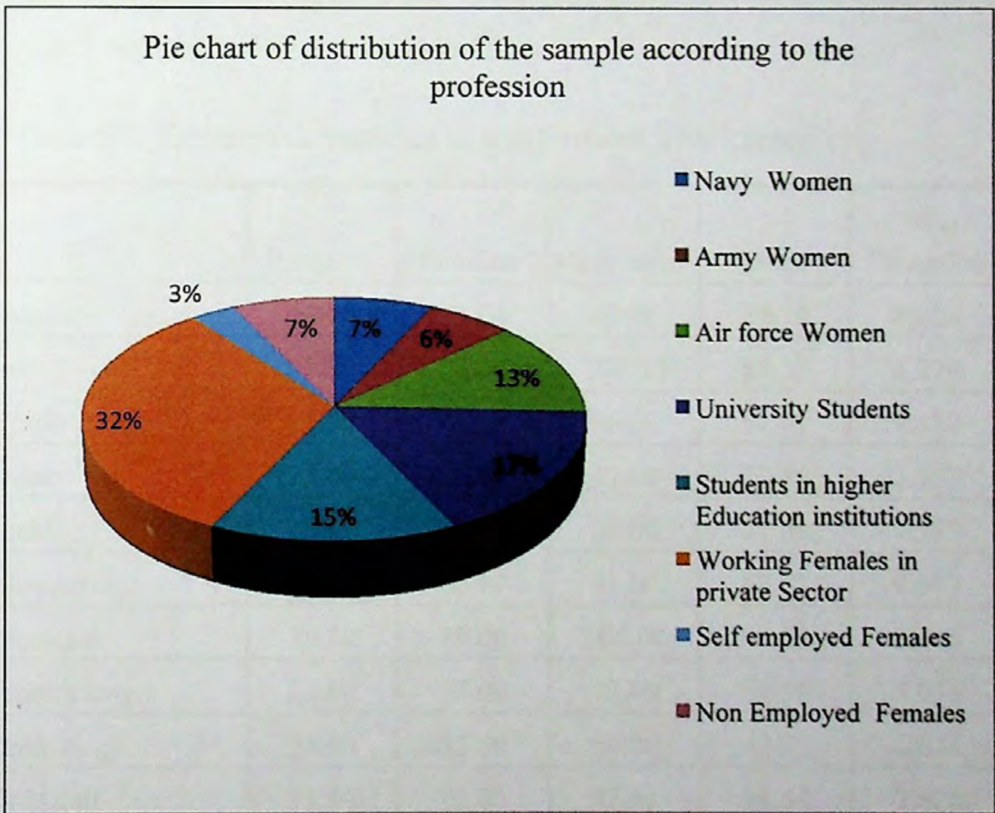


Fig.5.2 Distribution of the sample according to the profession

5.3 Analysis of Existing Size Chart Development Approaches

As explained in section 4.3, three existing size chart development approaches were applied to the lower body anthropometric dataset of Sri Lankan females, and the results are discussed below.

5.3.1 The approach using descriptive statistics

When applying this approach to the Sri Lankan dataset, key measurements were initially identified. Considering the Pearson correlation coefficients between measurements (Table 5.11), measurements with high factor loadings in factor analysis (Table 5.10) and literature, waist girth, hip girth and inseam height were selected as key measurements. The process of selection of key measurements is discussed in detail under section 5.4.2. Afterwards, descriptive statistics were checked and result of short inseam height category is shown in table 5.3. As explained in section 4.3.1, using these descriptive statistics and size intervals, a size chart was developed and shown in table 5.4.

Table 5.3 Descriptive Statistics of short inseam height category

	Range	Minimum	Maximum	Mean	Std. Deviation
waist	36.00	53.00	89.00	67.74	8.108
hip	32.00	75.00	107.00	88.32	6.274
thigh	25.50	40.50	66.00	51.41	4.889
knee	15.00	27.00	42.00	32.85	2.522
ankle	7.00	18.00	25.00	21.09	1.337
inseam	9.00	65.00	71.00	68.73	2.147
outseam	19.00	89.00	108.00	96.08	3.640
crotch length	25.00	54.00	79.00	65.59	5.036
mid thigh	20.50	33.50	54.00	41.97	4.041
mid calf	15.50	22.00	37.50	28.64	2.828
hip height	6.00	14.00	20.00	16.93	1.361
knee height	13.50	38.00	51.50	43.46	2.266
ankle height	3.00	5.00	8.00	6.34	0.664

Table 5.[illegible] Size chart for short inseam height category (inseam = 65-71 cm)

	Sizes									
Description	**S1**	**S2**	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**	**S9**	**S10**
Waist	54.0	58.0	62.0	66.0	70.0	74.0	78.0	82.0	86.0	90.0
Hip	74.0	78.0	82.0	86.0	90.0	94.0	98.0	102.0	106.0	110.0
Thigh	40.0	43.0	46.0	49.0	52.0	55.0	58.0	61.0	64.0	67.0
Mid thigh	30.0	32.5	35.0	37.5	40.0	42.5	45.0	47.5	50	52.5
Knee	27.0	28.5	30.0	31.5	33.0	34.5	36.0	37.5	39.0	41.0
Mid calf	22.0	23.5	25.0	26.5	28.0	29.5	31.0	32.5	34.0	35.5
Ankle	18.0	18.8	19.6	20.4	21.2	22.0	22.8	23.6	24.4	25.2
Out seam	90.0	92.0	94.0	96.0	98.0	100.0	102.0	104.0	106.0	108.0
Crotch length	55.0	57.5	60.0	62.5	65.0	67.5	70.0	72.5	75.0	77.5
Hip height	14.0	14.8	15.6	16.4	17.2	18.0	18.8	19.6	20.4	21.2
Knee height	38.0	39.5	41.0	42.5	44.0	45.5	47.0	48.5	50.0	51.5
Ankle height	5.2	5.5	5.8	6.1	6.4	6.7	7.0	7.3	7.7	8.0

Aggregate loss of fit was calculated and this value for the developed size chart was 7.69 cm which is much higher than the theoretical value.

According to the above calculation, it is clear that the developed size chart will not provide a better fit for pant to the consumer. Other than the above-mentioned reason, there are several drawbacks in this statistical method. Although there are great differences in human body shapes, proportions and measurements, it was not addressed in this method. For example, the anatomical structure of the female pelvic shows several shape variations (Fig.2.2). Hence, there should be an opportunity for multiple choices of hip measurements for a single waist measurement in a size chart in providing a correct fit to the consumer. Therefore, these different body shapes should be taken into consideration in order to represent the population correctly.

Furthermore, height variables (e.g. outseam, inseam, knee height, hip height) and girth variables (e.g. waist, hip, thigh) show poor correlation with each other (Table 5.11) and also, the girth variables and length variables are not linearly connected as shown in Fig.5.3. Hence, increasing length measurements along with girth measurements leads to poor results.

Further, increasing all the measurements from small values to higher values throughout the size chart is not correct because the measurements do not follow perfect linear relationships (Fig. 5.4). Therefore, due to these reasons, size charts developed with the statistical technique based on descriptive statistics would not be successful in solving fit problems of ready-to-wear garments.

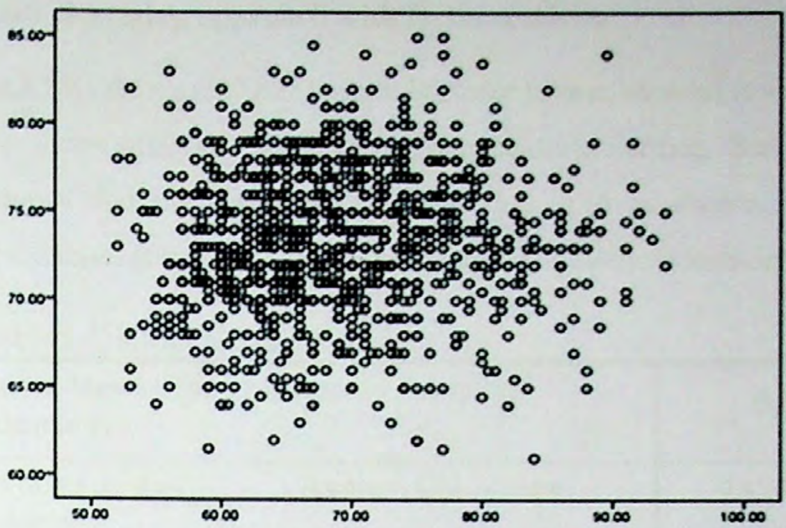


Fig. 5.3 Scatter Plot of inseam vs waist

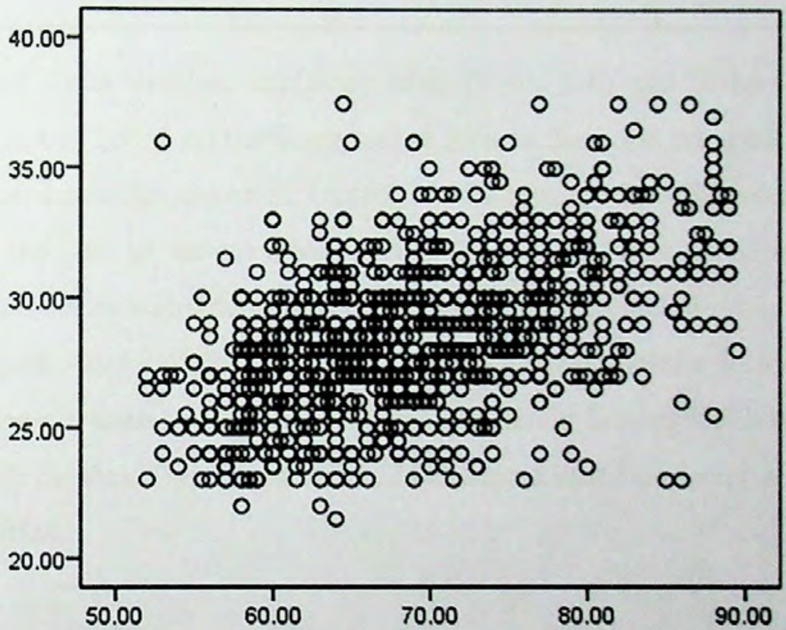


Fig. 5.4 Scatter plot of mid-calf vs waist

Hence, size charts developed using statistical approach do not represent the population well and resulting in customer dissatisfaction in ready-to wear garments.

5.3.2 *K*-means clustering approach with factor analysis

The resultant KMO measure, 0.875, which is closer to one, showed it was acceptable to perform a factor analysis (IBM knowledge centre). Further, Bartlett's Test of Sphericity shows that the significance is less than 0.05 as shown in Table 5.5. Therefore, the dataset is suitable for applying the factor analysis technique.

Table 5.5 KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.875
Bartlett's Test of Sphericity	Approx. Chi-Square	9.429E3
	df	78
	Significance	0.000

The generated Total variance explained table (Table 5.6) and Scree plot (Fig 5.5) show three factors based on the Eigen value greater than one criterion. Further, the Varimax rotated solution shown in Table 5.7, which results in orthogonal coordinates maximizing the sum of the variances of the squared loadings, confirmed the three factors. These factors were identified as girth factor (which includes waist, hip, thigh, mid thigh, knee, mid-calf, ankle), length factor (which include inside leg length, outside leg length, knee length, ankle length) and height factor (which include crotch length and hip height). The sample has been clustered based on factor scores using *k*-means algorithm.

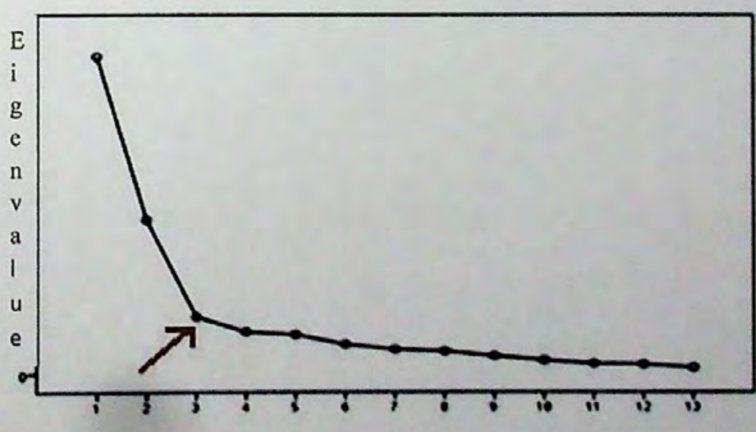


Fig.5.5 Scree plot (Eigen value vs Component number)

The Scree plot in Fig.5.5 plots resultant eigen values for 13 components. It shows that from component number three, the graph has a sharp elbow meaning that there are three potential factors. Therefore, three factors were selected and the analysis continued.

Table 5.6 Table of total variance explained

	Initial Eigen values		Extraction sums of squared loadings				Rotation sums of squared loadings		
component	Total	%of variance	Cumulative %	Total	%of variance	Cumulative %	Total	%of variance	Cumulative %
1	5.589	42.990	42.990	5.589	42.990	42.990	4.906	37.735	37.735.
2	2.708	20.833	63.824	2.708	20.833	63.824	2.576	19.819	57.554
3	1.000	7.695	71.519	1.000	7.695	71.519	1.815	13.965	71.519
4	0.732	5.634	77.153						
5	0.684	5.259	82.412						
6	0.509	3.919	86.331						
7	0.422	3.245	89.576						
8	0.386	2.967	92.542						
9	0.305	2.345	94.838						
10	0.230	1.772	96.610						
11	0.171	1.315	97.975						
12	0.160	1.230	99.206						
13	0.103	0.794	100.00						

Table 5.7 Rotated component matrix

	Component		
	1	2	3
waist	.870	.084	-.106
hip	.910	.112	.187
thigh	.896	.043	.150
knee	.829	.196	.125
ankle	.512	.310	.244
inseam	.009	.885	.087
outseam	.057	.758	.503
crotch length	.400	-.004	.740
mid thigh	.911	.016	.085
mid calf	.740	-.113	.255
hip height	.034	.355	.779
knee height	.133	.847	.037
ankle height	.096	.454	.428

Fig.5.6 shows the three-dimensional scatter plot of factor scores for three clusters which showed three clusters clearly. Although new factors were created representing the variables, those original variables were needed in developing size charts. Hence, its behavior inside these clusters should be studied. Scatter plots of original variables (e.g. waist girth Vs hip girth) showed that the points were scattered highly and the clusters were invisible (Fig.5.7). Moreover, mean, minimum and maximum were studied in selected key variables (waist, hip & inside leg length) in those clusters and it was verified that those variables overlapped among clusters (Table 5.8).

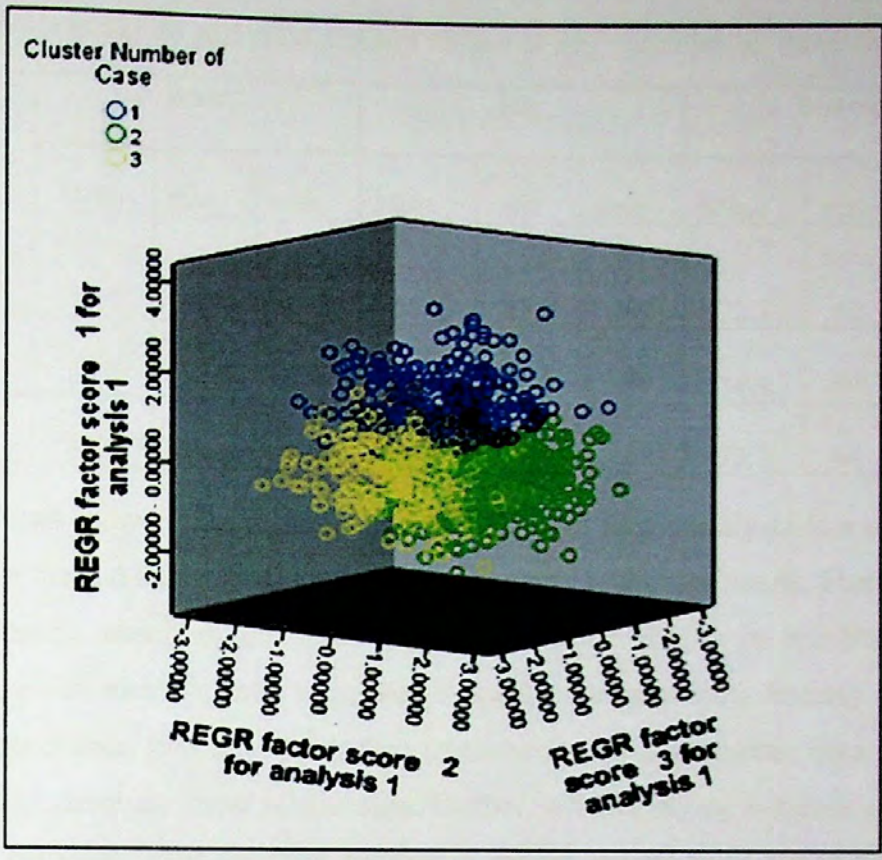


Fig.5.6 Three-dimensional scatter plot of factor score clusters

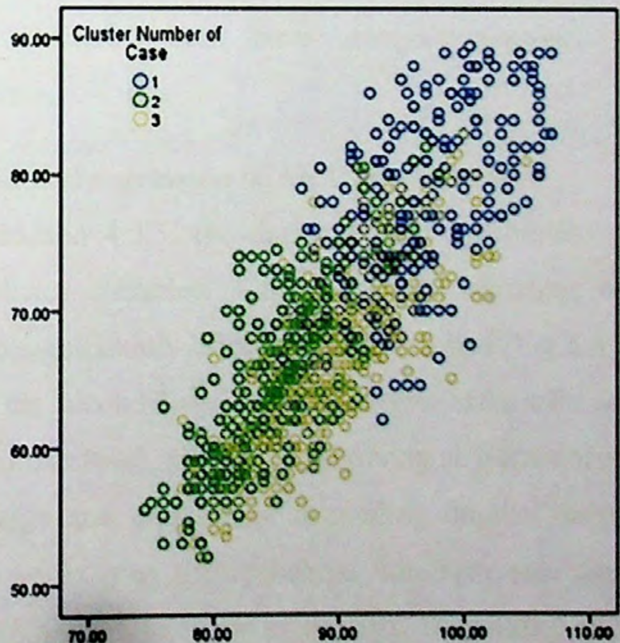


Fig. 5.7 Scatter plot of hip Vs waist in 3 clusters

Table 5.8 Mean values and measurement ranges of key variables in three clusters

Cluster No	Waist			Hip			Inseam		
	Mean	min.	max.	Mean	min	max	Mean	min	max
1	77.1	62.0	89.5	95.8	82.0	107	72.5	62	85
2	64.9	52.0	82.0	85.2	74.5	98	74.3	64	83
3	66.0	52.0	83.0	89.0	77.5	105	73.3	64	85

This analysis showed that variable reduction through factor analysis is meaningless because at the end the original variables must appear in the size charts. Further, these anthropometric variables cannot represent from another due to its non-linearity. In addition, the k- means cluster technique is suitable for clustering linearly separable data and therefore, it is not appropriate in clustering anthropometric data which do not show an adequate linear relationship. Further, when applying *k*-means clustering, variable reduction must be done because it cannot handle high dimensional data. However, variable reduction will lose some information and also the generated factors are meaningless in the subject of size chart development. Due to these reasons, *k*-means clustering with factor analysis approach is not suitable in developing size charts.

5.3.3 Classification and regression (CART) decision tree

As explained in section 4.3.3, the decision tree is completely dependent on the selection of predictor variables. Although a few variables were selected, other variables may also significantly affect the decision tree. Fig.5.8 shows the resultant decision tree and the second level of the tree was arbitrarily selected without any theory behind it. At that level, different lower body size categories were identified as small, medium, large and extra large depending on the waist girth (Table 5.9). However, there is no existing theory behind this body size classification which is based on waist girth and the selection of cut-offs. Depending on the selection of tree level and cut-off of the variable, the size chart will be completely different. Further, selecting waist girth as one of the predictor variable leads to wrong decisions because waist girth is directly connected to the target variable (waist to hip ratio).

Table 5.9 Size categories resulted from second level of the CART decision tree

Size Category	Small	Medium	Large	Extra Large
Waist	< 60.25	60.25- 68.75	68.75- 77.75	> 77.75

Due to the above-mentioned problems in application, the CART decision tree method was found to be ineffective for body classification in developing size charts.

Therefore, this research proves that existing approaches in developing size charts have drawbacks which resulting poor size charts. Therefore, the necessity of a novel approach which can solve above problems and successfully cluster non-linear high dimensional data was confirmed.

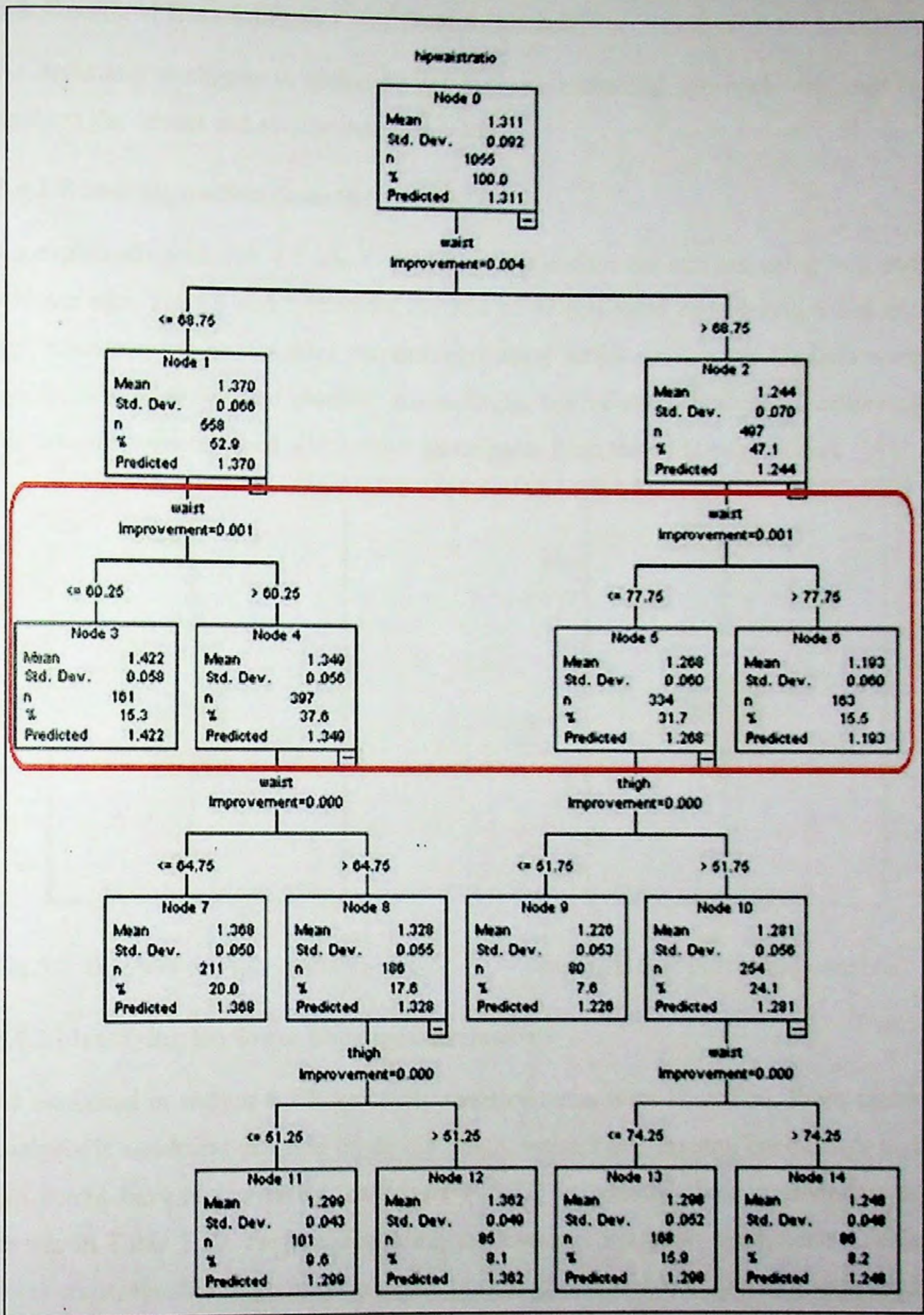


Fig.5. 8 CART Decision Tree

5.4 Results of Data Analysis Using New Approach

As explained in chapter 4, global kernel *k*-means clustering approach was used to analysis the dataset and results were presented.

5.4.1 Removing outliers from the dataset

As explained in section 4.4.1.1, variables were checked for outliers using box and whisker plot. Fig.5.9 and 5.10 show the box plots generated considering waist and hip measurements as variables respectively, using SPSS version 16. Outliers were shown here with the case number. Accordingly, twelve subjects were identified as outliers and were replaced with twelve participants from the additional dataset.

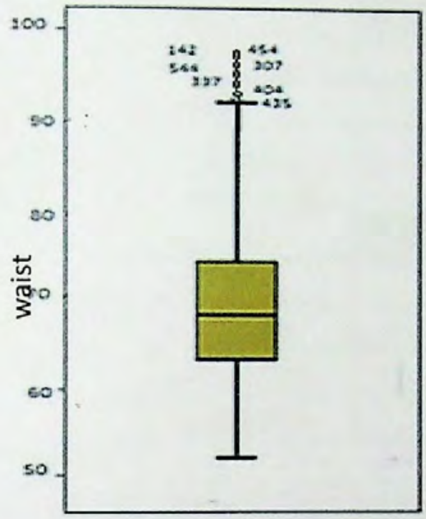


Fig.5.9 Box plot of waist variable

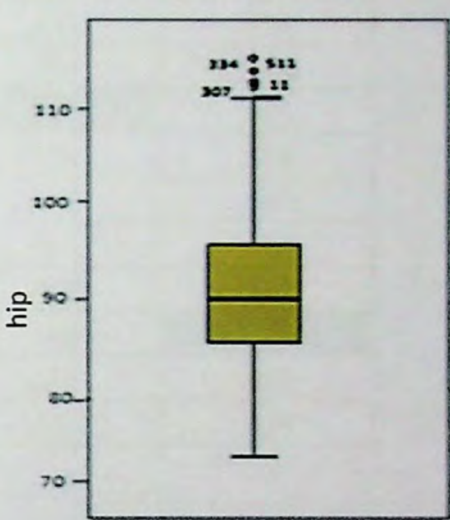


Fig.5.10 Box plot of hip variable

5.4.2 Identifying key lower body measurements

As explained in section 4.3.3, key body measurements were identified. From factor analysis, it was found that mid thigh, hip, thigh, waist, knee, inseam, knee height and hip height have higher factor loadings (> 0.75) in resulted three components as shown in Table 5.10. Further, according to Pearson correlation coefficients (Table 5.11) waist, hip and thigh measurements have higher correlation (> 0.70) with each other and also with mid thigh and knee measurements. Further, Ashdown (1998) has selected four dimensions; hip, waist, crotch height (inseam height) and crotch length, as key variables to create her sizing system for female pants. However, apparel brands such as Nike, Banana Republic and Old Navy have used waist, hip and inseam height as key measurements for their female pants. Literature also shows that waist, hip and inseam length are found to be key measurements in pant development. Therefore, in developing a size chart for pants, waist measurement, hip measurement and inseam height were considered as key dimensions.

Table 5.10 Factor loadings of rotated component matrix resulted from factor analysis

Measurements	Component1	Component2	Component3
waist	.869	.080	-.102
hip	.909	.108	.191
thigh	.895	.039	.151
knee	.829	.197	.124
Ankle	.514	.309	.246
inseam	.008	.887	.085
Outseam	.054	.758	.504
Crotch length	.394	-.009	.740
Mid thigh	.911	.016	.084
Mid calf	.740	-.107	.247
Hip height	.035	.348	.779
Knee height	.130	.848	.041
Ankleheight	.099	.446	.429

Table 5.11 Pearson product moment correlation between variables

	1	2	3	4	5	6	7	8	9	10	11	12	13
1 waist	1												
2.hip	0.819	1											
3.thigh	0.737	0.858	1										
4.mid thigh	0.733	0.811	0.831	1									
5.knee	0.636	0.750	0.721	0.747	1								
6.midcalf	0.503	0.622	0.617	0.675	0.616	1							
7.ankle	0.399	0.490	0.427	0.420	0.496	0.415	1						
8.inseam	0.071	0.127	0.068	0.063	0.159	0.050	0.239	1					
9.outseam	0.047	0.230	0.155	0.121	0.240	0.103	0.352	0.717	1				
10.crotch length	0.301	0.513	0.451	0.411	0.378	0.401	0.291	0.108	0.434	1			
11.hip height	0.039	0.227	0.184	0.107	0.210	0.148	0.323	0.363	0.586	0.407	1		
12.knee height	0.151	0.203	0.150	0.110	0.272	0.076	0.276	0.612	0.640	0.161	0.319	1	
13ankle height	0.092	0.227	0.185	0.148	0.253	0.109	0.226	0.359	0.423	0.209	0.415	0.330	1

5.4.3 Initial categorization of the dataset

First, the dataset was divided into three categories based on Waist-to-Hip Ratio (WHR) as described in section 4.3.4. These three categories represent three different lower body shapes existing in the sample. Table 5.12 shows these three categories with the number of subjects in each category.

Table 5.12 Division of the dataset according to WHR

Waist-to-Hip-Ratio	Lower body shape category	No. of subjects
WHR$\leq$ 0.70	Small WHR (curvy lower body)	124
0.71 $\leq$ WHR $\leq$0.79	Medium WHR(medium lower body)	629
WHR$\geq$ 0.80	Large WHR (straight lower body)	315

5.4.4 Data analysis using "Global kernel *k*-means clustering" method

The three sets of data were analyzed separately using different kernel functions as explained in section 3.4. Here, polynomial kernel function and Gaussian (RBF) kernel functions were used because these were the most popular kernel functions in prior art of anthropometric studies.

5.4.5 Global kernel *k*-means algorithm with polynomial kernel function

The polynomial kernel function of degree p, $K(X, Y) = (X \cdot Y + c)^p$, is used to transform the input data space to feature space, where p is a positive integer and $c \in \mathbb{R}$.

The degree of kernel function, p, was first selected as two and c as one. Medium WHR dataset was first mapped to high dimensional feature space through this polynomial function which result a kernel matrix of size (629$\times$ 629). Then, this kernel matrix was input to global kernel *k*-means clustering algorithm and number of clusters was given as two. This process segmented the dataset into two clusters.

Fig. 5.11 and 5.12 scatter plots showed the behavior of waist vs hip measurements and waist vs thigh measurements respectively in two clusters. Here, the points were scattered among the two clusters making it impossible in finding homogenous subgroups. Therefore, it was understood that medium WHR dataset was not successfully clustered for the above parameters. It was experienced that when the degree was increased to three and four, the resulted clusters became heavily scattered. When a decimal number was selected for the degree, (0 <p< 1), the resultant kernel matrix was not a positive semi-definite one because negative eigen values were available in the matrix. Therefore, polynomial kernel function was not suited in clustering the medium WHR dataset. Therefore, this analysis with polynomial kernel function was not continued for the remaining datasets.

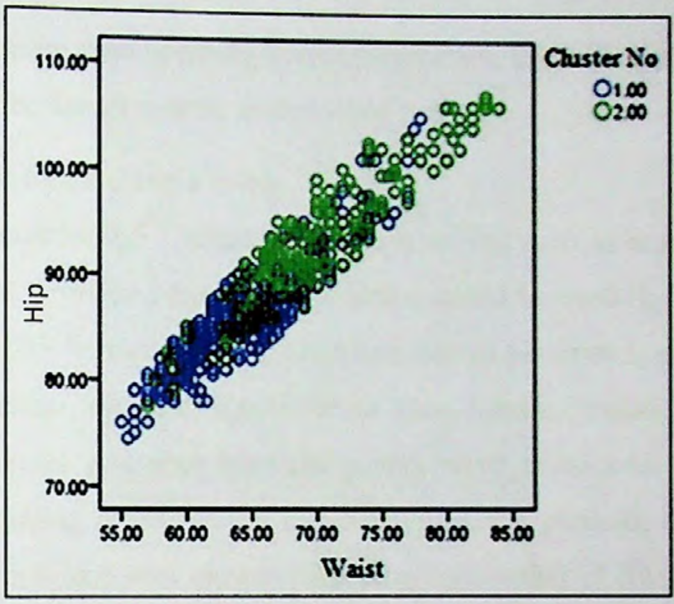


Fig. 5.11 Scatter plot of hip vs waist measurement

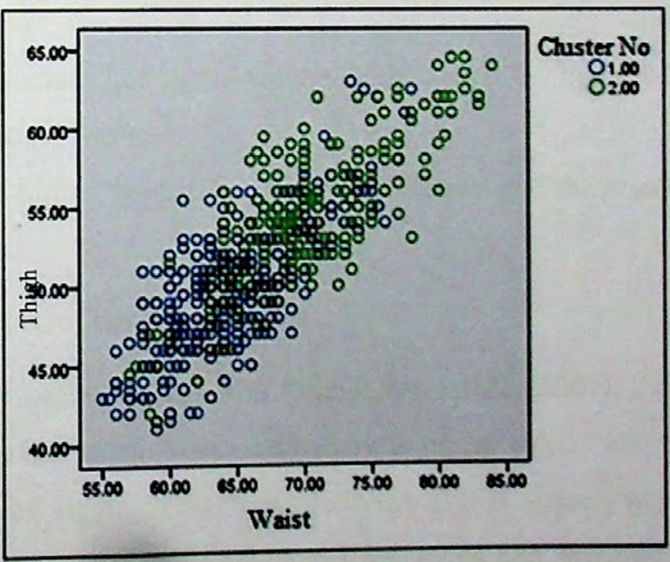


Fig. 5.12 Scatter plot of thigh vs waist measurement

5.4.6 Global Kernel *K*-means algorithm with Gaussian (RBF) kernel function

Since polynomial kernel function was not successful in clustering the dataset as explained in section 5.4.5, Gaussian (RBF) kernel function, was used next.

Gaussian (RBF) kernel function is given below;

$K(X, Y)=exp(-\|X-Y\|^2/2\sigma^2)$, Where $\sigma \in \mathbb{R}$ is called Gaussian kernel parameter.

5.4.6.1 Tuning kernel parameters

Two techniques were used in tuning kernel parameters; kernel-based Dunn's index and variance of the kernel matrix, as described below.

(i). Using kernel-based Dunn's index

As explained in section 3.5.1, cluster validation indices such as kernel-based Dunn's index which was developed from Dunn's index, could be used in optimizing kernel parameters. In RBF kernel function, Gaussian kernel parameter, sigma, need to be optimized. Therefore, different sigma values were used in creating kernel matrices and then the kernel distances between points were calculated (Matlab code in Appendix 4.6). Using global kernel k-means clustering method, the data in above different kernel matrices were clustered keeping the number of clusters constant (e.g. number of clusters=2). Then, the kernel-based Dunn's indices (Matlab code in Appendix4.7) were calculated for above clustered datasets which created using different sigma values. The sigma value which caused the highest KDI was selected as optimum value for sigma.

This technique was first applied in large WHR dataset and the results were explained in section 5.4.7.

(ii) Using variance of the kernel matrix

According to Yunzhang (2011), and Evangelista, *et al.* (2007), the highest variance of the kernel matrix which was resulted by a sigma value, was considered as the optimum value of sigma. Therefore, different sigma values were used to create kernel matrices and then checked for its variances and selected the sigma value which gave the highest variance.

This technique was first applied in large WHR dataset and the results were explained in section 5.4.7. Further, the three datasets were analyzed separately because kernel parameter and number of clusters were depending on the dataset.

5.4.7 Analysis of large WHR dataset (straight lower body)

Descriptive statistics of all the variables of the dataset were checked to have a general knowledge on the variables. Here, range, maximum, minimum, mean, standard deviation (SD) and coefficient of variation (CV= mean/ SD) were obtained for key variables and tabulated in Table 5.13. Coefficient of variation showed that dispersion is higher in waist measurement than hip measurement in this dataset.

Table 5.13 Descriptive Statistics of large WHR dataset

Variable	Range	Min	Max	Mean	SD	CV
waist	27.5	62.0	89.5	77.17	5.99	0.077621
hip	31.0	76.0	107.0	92.54	6.09	0.065809
inseam	23.0	62.0	85.0	73.57	4.04	0.054914

5.4.7.1 Tuning of kernel parameter

As explained in section 5.4.6.1, two techniques were used in tuning kernel parameter. The results were discussed below.

(i) Using kernel-based Dunn's index

Table 5.14 shows different values of sigma and resultant kernel-based Dunn's indices for two clusters. Accordingly, the KDI increased when decreasing sigma value. Sigma value of two gave the highest KDI value of one. Since, the KDI is inter cluster distance divided by intra cluster distance, the meaning of KDI equals 1 is that inter cluster distance equals to intra cluster distance. This means all points belongs to one cluster which was clearly shown by hip versus waist scatter plot in Fig. 5.13. This shows that when decreasing the sigma, the cross similarity values of the points of the resulted kernel matrix is reducing. Since, the kernel distance is given by 2{1-k(p,q)}; where k(p,q) is the cross similarity of points, the values of the distance matrix is increasing. Due to higher values in distance matrix, KDI gradually increased. Therefore, tuning of sigma value using cluster validation index was not successful in this research.

Table 5.14 KDI for different sigma values of large WHR dataset

Sigma	Kernel-based Dunn's index
8.0	0.1918
7.0	0.2427
6.0	0.3521
5.0	0.3750
4.0	0.6114
3.0	0.9945
2.0	1.0000

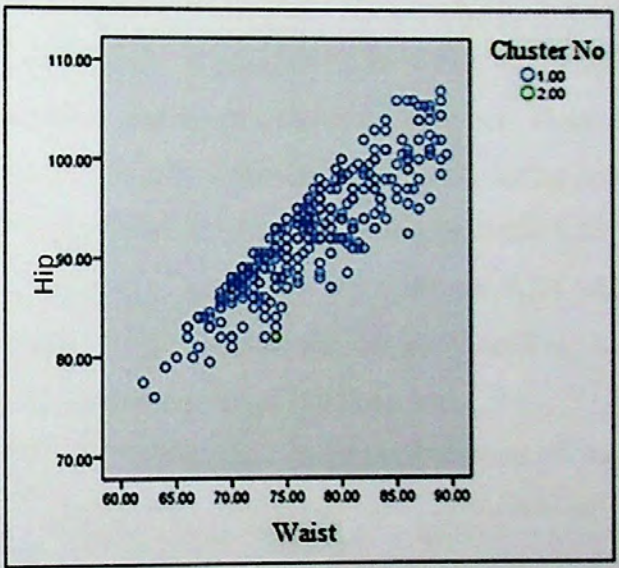


Fig. 5.13 Scatter plot of hip vs waist measurement of large WHR dataset for $\sigma = 2$

(ii) Using variance of the kernel matrix

As explained above, kernel-based Dunn's index cannot be used in this research as a parameter tuning method. Therefore, the second method, highest variance of kernel matrix method was used and Table 5.15 shows different sigma values and resultant variances of the kernel matrices. Accordingly, highest kernel matrix variance was resulted when sigma value is equal to 13.

Table 5.15 Matrix variances for different sigma values of large WHR dataset

Sigma	Matrix Variance
10.0	0.0454
12.0	0.0522
12.5	0.0524
13.0	0.0530
13.5	0.0529
14.0	0.0525

Therefore, optimal sigma value was taken as 13 and relevant kernel matrix was used to proceed the analysis. For this kernel matrix, kernel distance matrix was produced. Then, the next step was to find the optimal number of clusters which could be done using KDI. Here, global kernel *k*-means clustering method was used and the data set was clustered for different cluster numbers. Afterwards, KDI was calculated and the number of clusters which gave the highest KDI value was selected as the optimal cluster number. Therefore, different numbers such as two, three, four and five were used as number of clusters in global kernel *k*-means clustering approach and KDI for each step was calculated and the results were shown in Table 5.16.

Accordingly, cluster number of three gave the optimum KDI which means the best cluster number for large WHR dataset was three. Therefore, further analysis was done considering three clusters for large WHR dataset.

Table 5.16 KDI for different number of clusters of large WHR dataset

Cluster number	KDI
2	0.0636
3	0.0727
4	0.0523
5	0.0453

Two-dimensional scatter plots of the variables were also produced to see the behavior of the variables within three clusters. Scatter plots were generated against waist measurement (e.g. hip vs waist in Fig.5.14) and it showed obvious three clusters clearly with some overlaps. These cluster overlaps were obvious because anthropometric data are not linearly separable due to its natural complex blends.

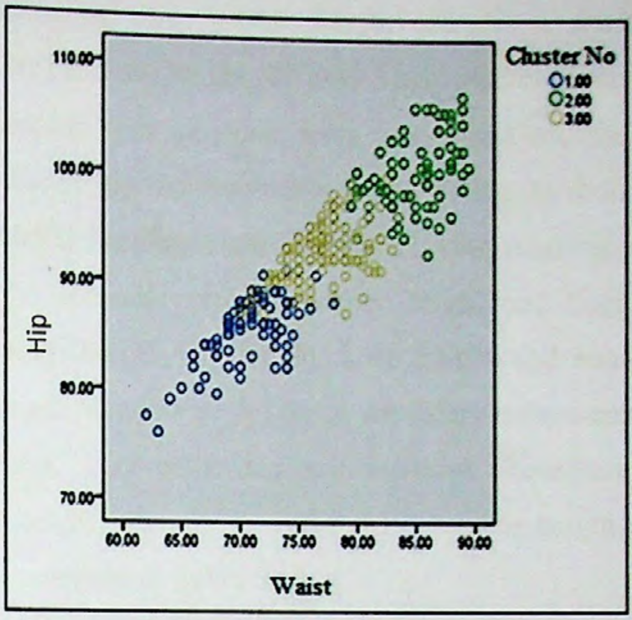


Fig.5.14 Scatter plot of hip Vs waist measurement of large WHR category

Scatter plot of inseam vs waist measurement in Fig. 5.15 shows that inseam was not separated within three clusters. Therefore, for all three clusters, same inseam height categories; short, regular, and tall, were used.

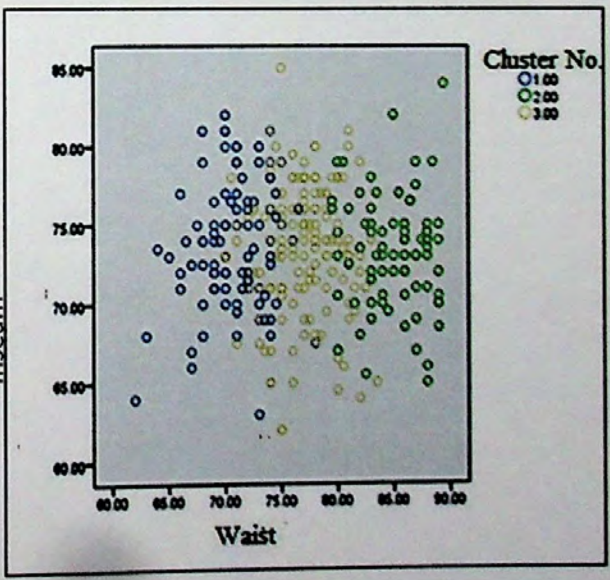


Fig.5.15 Scatter plot of inseam vs waist measurement of large WHR category

As explained in section 4.4.3, size charts were developed from the clustered data. Accordingly, for large WHR dataset (straight lower body), three separate size charts were developed following three inseam height categories; short height (68 cm), regular height (74 cm) and tall height (80 cm). Then, available waist groups for each height category, keeping 4cm intervals, were recognized and the dataset separated accordingly. Afterwards, hip measurements which belong to each waist group were identified and tabulated keeping 4 cm interval. Having filtering waist and hip into different groups, the secondary measurements (thigh, mid thigh, knee, mid calf, ankle, outseam, crotch length, hip height, knee height and ankle height) in each group were considered. Mean values of those secondary measurements which belong to a particular group, were calculated and included. Resultant size charts were presented as short height category in Table 5.17, regular height category in Table 5.18 and tall height category in Table 5.19.

In these size charts, multiple hip sizes were available for some waist groups (e.g. in Table 5.17; for 74cm waist, 86 cm & 90 cm hips). Further, these size charts clearly showed that height variables were not increasing along with girth variables which represent the actual situation of anthropometric data. Further, other girth variables also do not follow any consistent increment with waist and hip variables confirming the fact that anthropometric data were not linearly connected

Table 5.17 Size chart for large WHR (straight lower body) - Short height

Inseam	Short (65cm-71cm)									
Waist	62.0	66.0	70.0	74.0		78.0		82.0		88.0
Hip	78.0	82.0	86.0	86.0	90.0	90.0	94.0	94.0	98.0	104.0
Thigh	45.0	46.6	49.8	50.5	54.8	53.5	54.8	54.1	58.1	59.9
Mid thigh	37.5	38.7	41.9	40.6	44.7	45.1	44.5	45.4	47.1	49.0
Knee	29.6	30.2	32.2	32.3	34.0	34.5	32.6	34.0	36.6	35.9
Mid calf	26.3	25.8	28.1	27.1	30.7	30.2	29.9	29.6	32.1	32.1
Ankle	19.5	20.2	21.9	20.4	21.5	21.4	20.6	21.3	22.3	22.2
Out seam	92.7	93.7	95.0	94.2	95.0	95.3	94.1	94.3	96.1	94.6
Crotch length	57.0	61.5	64.0	63.5	64.5	65.3	67.1	66.5	68.9	68.2
Hip height	15.8	15.5	17.0	16.1	16.5	16.9	16.1	16.1	17.3	16.2
Knee height	41.3	41.4	41.2	41.0	41.9	41.7	40.2	41.8	40.3	41.1
Ankle height	5.5	6.1	6.0	6.2	6.2	6.1	6.3	6.0	6.7	6.6
% from short height category only	3.41	5.68	10.23	11.36	10.23	14.77	21.59	6.82	6.82	9.09
% from large WHR category	0.95	1.59	2.86	3.17	2.86	4.13	6.03	1.90	1.90	2.54

Table 5.18 Size chart for large WHR(straight lower body)- Regular height

Inseam	Regular (>71cm- 77cm)								
Waist	66.0	70.0	74.0		78.0		82.0		88.0
Hip	82.0	86.0	86.0	90.0	94.0	98.0	94.0	98.0	104.0
Thigh	48.0	50.0	50.6	53.4	55.6	56.5	55.8	57.2	61.1
Mid thigh	39.4	40.6	41.0	42.5	44.3	44.3	43.8	46.8	49.0
Knee	31.4	32.1	32.0	33.3	34.3	34.8	33.6	35.0	37.0
Mid calf	25.4	27.2	27.3	28.0	28.7	29.7	29.4	30.5	32.1
Ankle	20.5	21.4	21.0	21.7	21.9	22.5	21.6	22.8	22.8
Out seam	98.0	98.3	98.6	100.7	100.2	102.0	99.6	102.0	99.8
Crotch length	61. 2	63.5	62.6	65.3	65.8	68.3	65.5	68.1	70.1
Hip height	17.8	17.0	16.7	17.5	17.5	19.1	17.4	18.9	17.6
Knee height	44.3	44.4	44.5	45.0	44.9	44.9	44.7	45.3	45.3
Ankle height	6.1	6.3	6.2	6.4	6.6	6.5	6.4	6.6	6.5
% from Regular height category only	5.20	15.61	6.94	15.61	17.34	4.62	5.78	12.14	16.76
% from large WHR category	2.86	8.57	3.81	8.57	9.52	2.54	3.17	6.67	9.21

Table 5.19 size chart for large WHR (straight lower body)- Tall height

Inseam		Tall (>77cm - 83cm)				
Waist	70.0	74.0		78.0	82.0	88.0
Hip	86.0	86.0	90.0	94.0	98.0	104.0
Thigh	47.8	48.7	52.9	54.8	56.2	59.4
Mid thigh	40.1	40.0	44.0	44.7	45.3	47.6
Knee	32.4	32.3	34.2	34.9	36.2	37.2
Mid calf	26.7	27.4	28.4	29.4	30.2	32.2
Ankle	21.7	20.7	22.5	22.1	22.3	22.8
Out seam	103.6	104.0	104.5	103.9	104.5	104.4
Crotch length	61.5	62.6	64.2	65.5	66.8	69.0
Hip height	18.1	18.0	18.1	17.8	18.2	18.5
Knee height	47.1	47.1	47.1	46.6	47.4	47.7
Ankle height	6.5	6.6	7.2	7.3	7.1	7.1
% from tall height category only	18.37	16.33	20.40	26.53	10.20	8.16
% from large WHR category	2.85	2.54	3.17	4.12	1.59	1.27

5.4.8 Analysis of medium WHR dataset (medium lower body)

Descriptive statistics of key variables of the medium WHR dataset were tabled to have a general image of the dataset. The Coefficient of Variation (CV) of variables showed that the variables in this dataset were more dispersed than large WHR dataset (Table 5.13). Inseam variable follows same value range as large WHR category.

Table 5.20 Descriptive statistics of medium WHR category

variable	Range	Min	Max	Mean	SD	CV
waist	29.0	55.5	84.0	66.52	5.57	0.083734
hip	32.5	74.5	107	88.71	6.34	0.071469
inseam	21.0	64.0	85.0	73.37	4.22	0.057517

5.4.8.1 Optimizing sigma value

In order to get the optimum value for the sigma parameter, kernel matrices were created using RBF kernel function with different sigma values and then, variances of the matrices were calculated. Table 5.21 shows different sigma values of the RBF kernel function and resultant variances.

Table 5.21 Sigma values and resultant kernel matrix variance of medium WHR dataset

Sigma value	Matrix variance
12.0	0.0576
12.5	0.0582
13.0	0.0583
13.5	0.0581
14.0	0.0577

The optimal sigma value was thirteen for medium WHR dataset, as highlighted in Table 5.21.Therefore, the kernel matrix created with sigma parameter of 13 was used for Global kernel k-means clustering approach in order to cluster medium WHR dataset.

5.4.8.2 Optimization of Number of clusters

Since number of clusters was not priorly known, several cluster numbers were used for clustering. Then, KDI values relevant to different number of clusters were checked and the highest value was selected.

Table 5.22 KDI values for different cluster numbers for medium WHR category

Cluster No.	KDI
2	0.0339
3	0.0466
4	0.0415
5	0.0298

According to the Table 5.22, optimal number of clusters for medium WHR category was three. Therefore, dataset was divided into three clusters and further analysis was done. Scatter plots were also created to see the behavior of the variables within the respective clusters and it showed that first and second clusters have more overlaps of points (e.g. hip Vs waist in Fig.5.16). This was due to the different combinations of variables available within the sample.

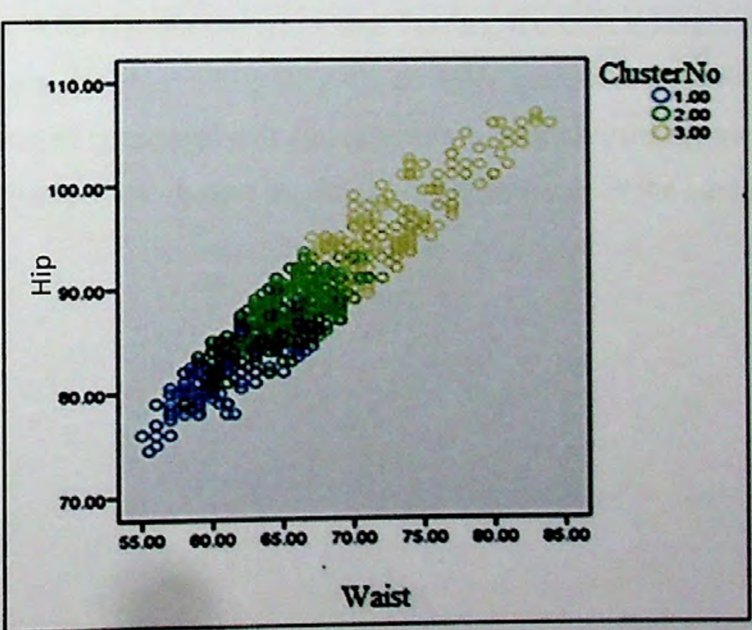


Fig.5.16 Scatter plot of hip Vs waist of medium WHR category

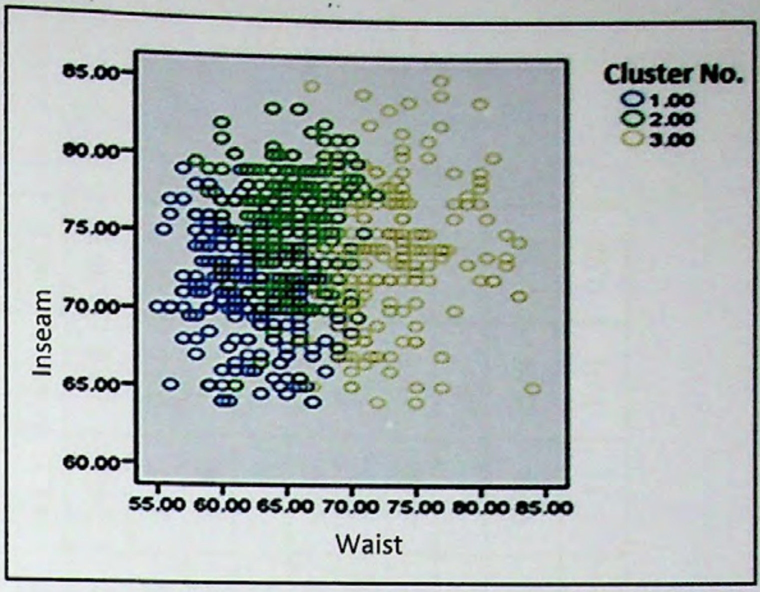


Fig.5.17 Scatter plot of inseam Vs waist of medium WHR category

Fig.5.17 shows that the same inseam categories could be used for all three clusters as done in the large WHR category. Therefore, three size charts were developed under three height categories as short, regular and tall. The resultant size charts were presented in Table 5.23 the short height category, Table 5.24 the regular height category and Table 5.25 the tall height category. The common features in these size charts are multiple choices of hip and very small variation in knee height, hip height and ankle height throughout the size chart. Further, the crotch length increased with hip and out seam variables while all girth variables increased across the size charts. Therefore, it can be concluded that global kernel *k*-means clustering method has been successfully clustered the dataset by identifying differences of the variables clearly.

Table 5.23 Size chart for medium WHR dataset - Short height

Inseam	Short (65cm-71cm)									
Waist	58.0		62.0		66.0		70.0		74.0	78.0
Hip	78.0	82.0	82.0	86.0	86.0	90.0	90.0	94.0	94.0	102.0
Thigh	44.5	46.7	47.5	49.1	49.4	51.7	52.7	56.0	55.5	61.0
Midthigh	36.4	38.3	38.9	40.1	40.5	41.2	44.8	45.2	46.3	49.1
Knee	30.2	30.7	30.6	31.6	31.7	32.6	33.5	34.3	34.8	36.3
Mid calf	26.0	26.5	26.8	27.6	27.5	28.7	29.6	30.7	30.1	32.0
Ankle	19.7	20.6	20.5	21.2	20.8	21.7	21.5	21.8	22.1	22.7
Outseam	94.7	95.1	94.7	95.3	94.9	95.1	95.3	97.1	95.8	95.9
Crotchlength	60.4	62.5	62.3	64.2	63.5	65.8	65.7	67.4	68.7	71.5
Hipheight	16.5	16.5	16.7	16.6	16.5	16.8	17.0	17.3	16.6	17.4
Knee height	40.9	40.8	40.3	41.5	40.9	40.6	40.5	41.4	40.5	40.7
Ankle height	6.1	6.0	6.1	6.6	6.5	6.6	6.3	6.4	6.7	7.0
% from short height category only	7.48	9.35	12.62	16.36	16.82	10.75	8.41	9.81	5.61	2.80
% from medium WHR category	2.54	3.18	4.29	5.56	5.72	3.66	2.86	3.34	1.91	0.95

Table 5.24 Size chart for medium WHR dataset - Regular height

Inseam	Regular (71.1 cm-77cm)											
Waist	58.0		62.0		66.0		70.0		74.0		78.0	82.0
Hip	78.0	82.0	82.0	86.0	86.0	90.0	90.0	94.0	94.0	98.0	102.0	106.0
Thigh	44.6	47.0	46.7	49.7	48.4	52.4	52.2	54.7	55.0	57.9	58.7	62.6
Midthigh	35.7	38.9	38.6	40.3	39.5	42.2	42.3	44.7	44.6	47.9	48.1	51.3
Knee	30.2	31.3	31.4	32.3	31.6	33.2	33.3	34.7	35.4	36.0	37.2	37.9
Mid calf	24.7	26.4	26.3	27.5	26.5	28.6	28.1	30.2	30.3	30.5	32.0	34.7
Ankle	19.8	20.6	20.6	20.8	20.9	21.5	21.6	22.0	21.6	22.0	23.3	23.6
Outseam	98.2	99.0	97.8	102.6	101.5	101.8	102.2	100.9	102.0	101.8	102.2	102.3
Crotchlength	61.0	62.7	62.4	66.4	66.3	67.4	67.3	68.3	67.6	69.8	71.6	71.5
Hipheight	17.2	17.2	16.7	18.2	18.0	18.4	18.1	18.5	17.9	17.9	18.7	19.1
Knee height	44.6	44.3	43.8	45.4	44.6	45.3	44.4	44.7	45.0	44.3	45.1	43.8
Ankle height	6.4	6.2	6.2	6.6	6.6	6.6	6.6	6.7	6.6	6.6	6.8	7.0
% from regular height only	5.79	10.29	10.29	18.01	12.54	9.32	4.82	11.25	5.14	6.43	3.54	2.57
% from medium WHR category	2.86	5.08	5.08	8.90	6.20	4.61	2.38	5.56	2.54	3.18	1.75	1.27

Table 5.25 Size chart for medium WHR dataset - Tall height

Inseam	Tall (77.1cm-83cm)							
Waist	62.0		66.0		70.0		74.0	
Hip	82.0	86.0	86.0	90.0	90.0	94.0	94.0	98.0
Thigh	47.5	50.1	50.7	51.7	52.5	55.3	54.6	60.7
Midthigh	38.4	40.6	41.0	42.5	42.5	45.7	46.0	49.8
Knee	32.1	32.8	33.3	33.5	33.7	35.3	35.4	36.7
Mid calf	26.2	26.8	27.6	28.1	28.1	30.2	30.7	31.7
Ankle	21.1	21.5	22.3	22.1	22.1	22.1	22.4	22.5
Outseam	104.6	105.4	104.6	105.8	104.5	106.7	106.1	107.5
Crotchlength	64.0	65.4	65.1	66.3	65.7	68.6	68.5	70.7
Hipheight	17.4	18.6	18.0	18.4	18.3	19.1	18.5	19.3
Knee height	45.8	46.4	46.3	46.7	46.8	47.1	46.6	47.1
Ankle height	7.0	6.9	7.1	7.0	6.9	7.1	7.1	7.5
% from tall height category only	11.54	14.42	16.35	15.38	17.31	11.54	9.62	3.85
% from medium WHR category	1.91	2.38	2.70	2.54	2.86	1.91	1.58	0.64

5.4.9 Analysis of small WHR dataset (curvy lower body)

As explained in previous sections, small WHR category was also analysed following global kernel *k*-means cluster method.This is the smallest dataset (124 subjects only) which comprises curvy lower body shapes.Eventhough the young females were also included in the sample, curvy lower body shapes (WHR ≤ 0.7) were rare among them. Descriptive statistics of small WHR dataset (Table 5.26) showed that the coefficient of variation (cv) of key variables were smaller than the other two datasets meaning that the dispersion of variables fromits mean were small. Same inseam height categories which were used in other two datasets could be used because inseam height maintain same value range.

Table 5.26 Descriptive statistics of key variables in small WHR dataset

variable	Range	Min	Max	Mean	SD	CV
waist	19.0	52.0	71.0	59.43	4.13	0.069494
hip	26.0	76.0	102.0	86.89	5.64	0.064910
inseam	19.0	64.0	83.0	73.75	3.93	0.053288

5.4.9.1 Optimizing sigma value

As described in section 5.4.8.1, optimum sigma value was obtained by checking variance of kernel matrices resulted by different sigma values. According to the Table 5.27, optimum sigma value for small WHR dataset was 11.

Table 5.27 Kernel matrix variances for different values of sigma of small WHR category

Sigma	Matrix Variance
10.0	0.0602
10.5	0.0604
[redacted]	[redacted]
11.5	0.0607
12.0	0.0603

5.4.9.2 Optimization of Number of Clusters

Using sigma as 11, kernel matrix was produced for small WHR category. Then, for different cluster numbers, the dataset was clustered using global kernel k-means clustering. For those differently clustered small WHR dataset, optimal KDI value was selected (Table 5.28).

Table 5.28 KDI for different number of clusters of small WHR category

Cluster No.	KDI
2	0.1040
3	0.1014
4	0.0961

According to the Table 5.28, optimum cluster number was two. Limiting the number of clusters to two might be due to the less dispersion of the variables in the input space comparing with other two datasets. Therefore, the dataset was clustered into two using global kernel *k*-means clustering approach. Scatter plot of hip vs waist in Fig.5.18 shows the two clusters. Since the dataset was non-linearly separated by the above approach, some points were overlapped between the clusters which was acceptable.

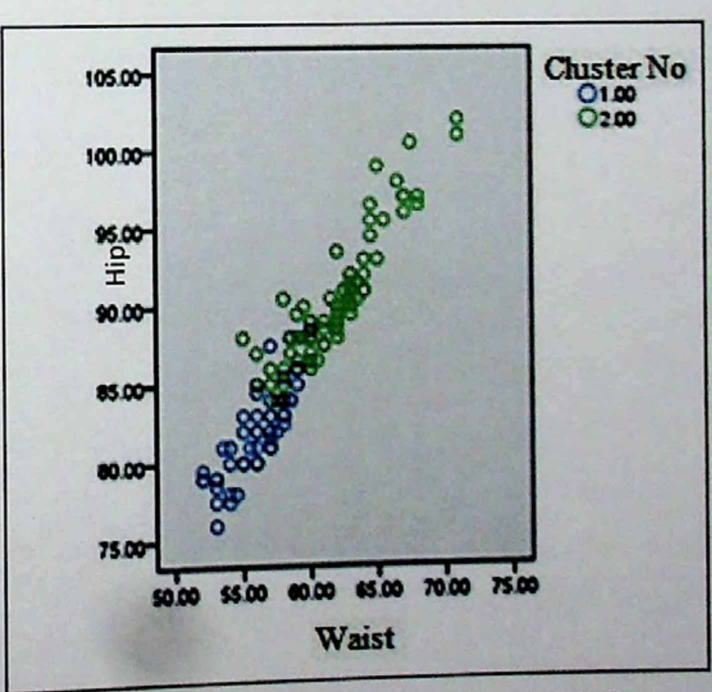


Fig.5.18 Scatter plot of hip vs waist of small WHR category

From these clustered data, size charts were developed under three height categories; short height category (Table 5.29), regular height category (Table 5.30) and tall height category (Table 5.31). Under this small WHR category, 16 sizes were resulted for all height groups. Smaller waist and larger hip where the difference between waist and hip was greater than 24 cm were included in this category.

Table 5.29 Size chart for small WHR dataset (curvy lower body)-Short height

Inseam	Short (65cm-71cm)			
Waist	54.0		58.0	
Hip	78.0	82.0	86.0	90.0
Thigh	44.0	45.6	48.4	48.8
Midthigh	37.4	38.4	40.3	40.6
Knee	30.0	30.6	31.8	32.7
Mid calf	26.0	27.2	27.8	28.5
Ankle	20.0	20.5	21.0	21.8
Outseam	94.5	95.4	97.5	97.0
Crotchlength	60.2	62.0	63.2	65.5
Hipheight	17.1	17.1	18.5	18.7
Knee height	43.4	42.6	43.1	43.3
Ankle height	5.8	6.3	6.6	6.8
% from short height category only	28.57	25	35.71	10.71
%from small WHR category	6.45	5.65	8.06	2.42

Table 5.30 Size chart for small WHR dataset (curvy lower body)- Regular height

Inseam	Regular (71.1cm-77cm)						
Waist	54.0		58.0		62.0	66.0	70.0
Hip	78.0	82.0	86.0	90.0	90.0	98.0	102.0
Thigh	43.8	45.3	47.8	50.7	51.3	57.2	62.5
Midthigh	36.1	37.5	39.6	41.8	41.5	45.2	48.5
Knee	30.4	30.8	32.3	32.6	33.0	35.0	35.0
Mid calf	25.1	25.8	28.3	29.0	28.3	30.6	30.0
Ankle	20.6	21.1	21.5	21.4	21.5	22.2	24.0
Outseam	99.7	100.0	102.2	103.1	102.0	101.5	104.5
Crotchlength	60.6	63.8	66.5	68.2	68.0	70.1	74.3
Hipheight	17.4	17.1	18.6	18.7	18.6	18.1	18.5
Knee height	44.9	45.6	45.6	44.6	45.1	44.0	44.5
Ankle height	6.5	6.9	6.8	6.9	7.3	6.6	7.2
%from regular height category	9.46	5.41	32.43	5.41	29.73	13.51	4.05
% from small WHR category	5.65	3.23	19.35	3.23	17.74	8.06	2.42

Table 5.31 Size chart for small WHR dataset (curvy lower body) - Tall height

Inseam	Tall (77.1cm-83cm)				
Waist	54.0	58.0	62.0	66.0	70.0
Hip	78.0	86.0	90.0	94.0	102.0
Thigh	44.3	47.8	49.4	53.6	62.5
Midthigh	36.4	38.4	40.2	44.5	49.5
Knee	29.5	31.2	33.4	35.1	35.0
Mid calf	25.7	27.2	27.7	29.3	30.0
Ankle	20.4	20.7	22.1	22.4	24.0
Outseam	104.8	105.4	105.6	107.4	107.0
Crotchlength	62.3	66.0	67.8	69.7	75.0
Hipheight	17.7	18.1	19.4	19.8	20.5
Knee height	47.7	47.8	47.6	47.9	47.5
Ankle height	7.0	6.8	7.4	7.2	7.2
%from tall height category only	13.63	27.27	27.27	22.73	9.10
%from small WHR category	2.42	4.84	4.84	4.03	1.61

5.4.10 Size chart validation

5.4.10.1 Validation using statistics

As explained in section 4.4.4, developed size charts were validated using aggregate loss of fit factor which was to verify how close these sizes to actual body measurements of wearers. For this purpose, anthropometric data sample which consist of 362 females were used. Aggregate loss of fit factor (ALF) was calculated for developed size charts of three body shape categories.

Table 5.32 shows the resulted values for aggregate loss of fit factor for three different size chart categories under three height categories. As mentioned in section 2.8.1, theoretically acceptable value of aggregate loss of fit factor is 4.4 cm. However, the resulted size chart has less aggregate loss of fit factor except in large WHR–tall height category which is 4.82 cm.

Table 5.32 Aggregate loss of fit factor for size charts

Size chart category		ALF
Small WHR (curvy lower body)	Short height	3.30
	Regular height	3.15
	Tall height	3.37
Medium WHR	Short height	2.95
	Regular height	2.49
	Tall height	2.70
Large WHR (straight lower body)	Short height	3.31
	Regular height	2.77
	Tall height	4.82

Descriptive statistics of large WHR category of the validation sample is shown in Table 5.33. It shows that the maximum values of waist and hip measurements are beyond the limits of the largest size of the size chart shown in Table 5.13. Therefore, when assigning a size to them, the difference between assigned value and actual value is higher which affect the AFL factor. Therefore, the value of AFL factor in large WHR -Tall height category is slightly higher than the theoretical value of 4.4cm.

Table 5.33 Descriptive statistics of large WHR category of validation sample

	Minimum	Maximum	Mean	Std. Deviation
waist	66.00	91.50	78.4407	6.48640
hip	78.50	108.00	93.7966	6.40153
inseam	63.50	85.00	75.2966	3.74065

For three shape categories, regular height sub-category achieved minimum value for AFL factor as shown in Table 5.30. When inseam variable of the validation sample was plotted as shown in the histogram in Fig.5.19, it showed that the majority represent regular height category (71cm < inseam < 77cm).

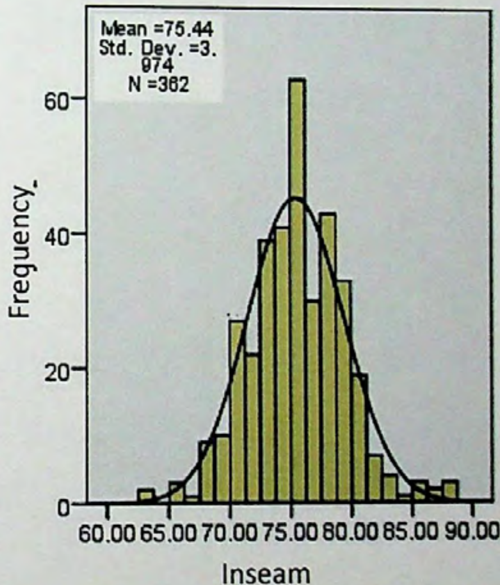


Fig. 5.19 Histogram of inseam in validation sample

Therefore, with the increase of the number of subjects (n) who belong to a particular size, the corresponding value of AFL decrease. Hence, this was the reason for the minimum AFL values in regular height category.

5.4.10.2 Live fit assessment in validating size charts

Live fit assessment session was held followed by a questionnaire survey in order to assess the fit and appearance of the pants as explained in section 4.4.4.2. Samples were produced using new size chart and fitted-on by the selected participants. Photographs of the sample pants were taken on front, back and side views. After that, a questionnaire which was to get the feedback regarding the fit of the pant, was filled by the participants (Appendix 4.8).

Fig. 5.20,(i) ,(ii), and (iii), show front views of the fitted-on samples which belong to three body shape categories; small WHR, medium WHR and large WHR. Accordingly, the appearances of the front figures are satisfactory and it follows the body curves accurately.

(i) Small WHR category

(ii) Medium WHR category

(iii) Large WHR category

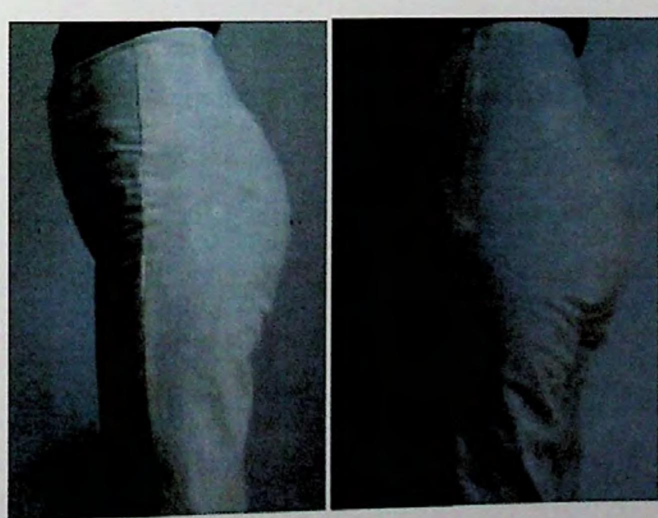
Fig.5.20 Front view of participants in three body shape categories

Pant fit may be troublesome due to the shape and depth of the pelvic structure and the size and shape of the waist, stomach, buttock, and upper thigh. Further, bowed legs and other leg variations can add poor appearance to the pant wearer (Veblen,2012). This poor appearance could be seen in back upper thigh area of the pants developed for fit sessions. As shown in Fig.5.21 (i) and (ii), the images of back view of two participants in the same small WHR category showed different appearances. Although, these two participants were in the same small WHR category, their shapes and depth of back muscle were different. Therefore, the first participant in Fig.5.21 (i) showed satisfactory appearance for back view and the second participant's (Fig.5.21 (ii)) back view showed wrinkle appearance. In this regard, the researcher has identified that the back rise curve need to be scooped for the second participant in order to provide room for her back muscle to sit in. Therefore, by identifying sub categories of body shapes available within the main categories, required changes of the pant patterns could be done as a provision to customize the product.

(i) First participant (ii)Second participant

Fig.5.21 Back view of participants in small WHR category

Fig 5.22 (i) and (ii) shows that the two participants were in the same body shape category according to WHR, however, differ in prominence of buttocks. According to Watkins,(1995), for same circumferential proportions, different width/ depth proportions and different angles may prevail which make body shape categorization more complex. Further, according to Petrova and Ashdown (2008), a more sophisticated combination of circumferential/arc proportions, width/depth proportions, and body heights/lengths proportions need to be considered in defining body shapes. Therefore, as explained above, adding subcategories which need only pattern modifications could improve the customer satisfaction.



(i) First participant (ii) Second participant

Fig.5.22 Side view of participants in small WHR Category

Photographs of twelve samples were added in Appendix 5.1. Participants actual key body measurements were included with assigned measurements for comparison.

5.4.10.3 Results of the questionnaire survey

Questionnaire survey was accompanied with fit assessment session in order to have feedback from the participants regarding the fit of the sample pants. The questionnaire
which include nine statements regarding the fit of the pant followed "five likert system" with 1 - Strongly disagree,2 – Disagree, 3- Neither agree nor disagree,4 – agree and 5 - Strongly agree as shown in Appendix 4.8.

The participants have mentioned that the pants they buy have fit problems in waist, hip, thigh area and also in pant length. However, they have agreed with the fit of the sample pants in all nine areas included in the questionnaire while a few responses of "neither agree nor disagree" for some questions were noted. No one disagreed with the fit of the pants. When analyzing the feedback of the questionnaire, "agree" and "strongly agree" responses were considered as "agree" and " neither agree nor disagree" was considered separately. The percentages were calculated and plotted on a column chart as shown in Fig.5.23.
The column chart showed that more than 80 percent agreed with the fit of the pants they fitted-on while question number 1, 5 and 8 achieved 100 percent agreement. The areas that achieved 100 percent fit agreement were waist, leg and pant length. The question number 4, "Crotch is loose so that buttock and thigh are not defined", scored minimum of 83 percent since in two samples, crotch was not loose resulting the buttock to be defined. However, according to the analysis of questionnaire feedback, it can be concluded that the fit of the samples are satisfactory which in turn confirm the accuracy of the new size chart.

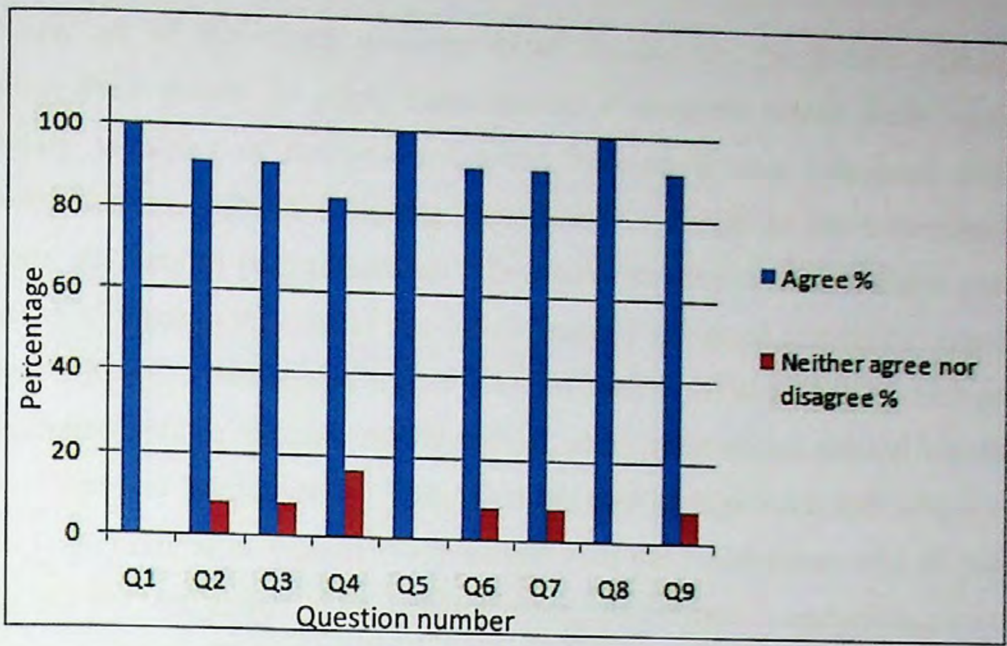


Fig. 5.23 Column chart of the percentages of "Agree" and " Neither agree nor disagree" of the questionnaire

5.5 Discussion

The method used in clustering anthropometric datasets in this research was the global kernel *k*-means clustering method. Although this method was successfully used in other fields such as MRI segmentation, this is the first attempt to employ it in the size chart development field. It was further improved by combining the kernel parameter tuning method and cluster validation method. Here, the cluster validation method was further modified by introducing kernel distance to the conventional Dunn's index. Furthermore, it avoided trial and error method used in cluster validation by introducing this consistent method. The most important fact is that this approach can handle high dimensional data without variable reduction which is a requirement for anthropometric data clustering. Further, this method facilitates non-linear separation of clusters which is also an important prerequisite in clustering of anthropometric data. Therefore, all drawbacks incurred in existing size chart development methods were overcome by this method. In addition, the ability to handle large datasets is also an advantage though the time complexity is high in processing. Further, high capacity for storage will be needed and it could be achieved with the rapid development of technology.

This new set of size-charts provides more choices for the female customer in selecting their proper fit pant. Furthermore, it presents actual mean values of secondary variables of the relevant group because it was calculated from the clustered dataset, without assigning measurements based on pre-determined size intervals, as done in past researches. Therefore, the resultant size charts precisely represent Sri Lankan females of age 20-40 years. This set of size charts consists of 70 sizes which is considerably higher than the traditional size charts which provide less selection options for customer. However, some international apparel brands (e.g. Land's End and Eddie Bauer) found that the cost of high inventory which occurs due to large number of sizes is compensated with the lower return rate of garments due to bad fit (McCulloch, *et al.* 1998). Further, on-demand manufacturing system of apparels which was developed by Levi-Strauss company can gain the benefit of this type of size charts because it avoids keeping large inventories (McCulloch, *et al.* 1998).

When implementing this new sizing system, it is necessary to develop a detailed size label which clearly illustrates three key measurements of the pant (waist, hip and inseam). In existing size label of pants in local apparel retail shops, only waist measurement is appeared because there are no multiple choices except waist. The customer should be aware of her waist, hip and inseam length measurements for which a shop assistance can help. With the development of technology, the selection of correct size will not be a complex issue any further. Validation process of resulted size charts which was done statistically and through live fit assessment session, proved that the new size chart was strongly representing the lower body of Sri Lankan females of aged 20-40 years. Furthermore, the results of the questionnaire survey verified that the participants were satisfied with the fit of the sample pant they fitted-on. Therefore, it was confirmed that the new approach; global kernel *k*-means clustering approach, with a suitable kernel function and optimized kernel parameters would provide better solution in clustering of anthropometric data.

5.6 Chapter Summary

This chapter analyzed the existing size chart development approaches and proved that these approaches have shortcomings resulting size charts which do not represent the population well. Further, this chapter discussed the results of the novel approach. First, the data set was initially categorized into three groups namely small WHR, medium WHR and large WHR, considering lower body shapes. Further analysis of the above categories was done separately. Two kernel functions, polynomial and Gaussian (RBF) kernel functions were tested and the latter one was selected. Two methods were used in tuning of kernel parameter, σ, and highest matrix variance method was applied. In optimizing number of clusters, kernel-based Dunn's index was used. The clustered datasets were used in developing size charts under three different height categories. Resultant size charts were validated statistically and through live fit assessment session.

CHAPTER 6

CONCLUSIONS

Achieving an accurate fit is a major problem in ready-to-wear apparel industry globally. There are several reasons for this poor fit such as, problems in size chart development methods, vast body shape differences existing within the population, changing of body measurements with time, limitations in mass production systems and the consciousness of clothing fit problems among consumers. The problem of apparel fit is particularly a main concern in female garments. Intense fitting problems are common with female pants. Therefore, the main objective of this research was to explore a better method in formulating size charts to address these critical fitting problems in female pants, because existing methods such as the *k*-means clustering, statistical method with descriptive statistics and classification and regression decision tree method have intrinsic drawbacks in developing size charts.

The purpose of an apparel size chart is to divide a varied anthropometric dataset in a population into homogeneous subgroups. Along with the large number of relevant body dimensions, the body proportions and shapes vary by a great extent. Classifying a population into homogenous body sizes is therefore a highly complex problem. Further, this body dimensions and shapes vary depending on different facts such as ethnicity, life style changes, racial mixes, and generational variability. This made the complexity of the problem much more serious. Therefore, establishing its own sizing system based on up-to-date anthropometric data of the population is a must for any country to eliminate fit problems (Ashdown, 1998). It was understood that selection of a proper method to segment the dataset is also very critical in developing an accurate size chart.

In this research, lower body anthropometric data of 1430 Sri Lankan females, aged 20-40 years, were collected covering all provinces and different professions including female personnel in three military forces. Thirteen lower body measurements, including height measurements and girth measurements, were collected. Existing methods were applied on the data and it was proved that those methods have drawbacks in the manner of application.

Kernel-based learning has been identified as the latest data mining technique. It has been used in different pattern recognition problems such as classification, clustering, correlation, ranking, and principal component analysis with different types of data such as vectors, strings and text. In kernel- based learning, non-linear data in the input space is mapped to a high dimensional feature space where linear patterns could be found depending on the kernel transformation function selected. Therefore, in clustering problems of lower body female anthropometric data, where variables are not linearly related in input space, linearly separable clusters could be found in this feature space. Further, it was evident that the reduction of variables in anthropometric data leads to serious errors in existing methods, and any technique with dimensional reduction may be less suitable in clustering anthropometric data. Therefore, the ability to handle high dimensional data without any dimensional reduction which is a key feature of kernel based clustering, is very important in anthropometric data clustering.

6.1 General Conclusions

The global kernel k-means clustering approach which has been used for MRI segmentation (Tzortzis & Likas,2009), was used to cluster anthropometric data in this research. The applied method was a combination of global *k*-means and kernel *k*-means clustering techniques. The development of size charts with this clustering approach is novel in the field of size chart development and it could achieve better fitting due to transformation of non-linear merged data space into linear separable feature space and its ability to cluster without reduction of dimension. Hence it provided a better solution to the problems in clustering approaches used for size chart development. *K*-means clustering has problems such as it traps in local minima in finding centroids, needs variable reduction, finds only linearly separable clusters, and the number of clusters need to be pre-determined. All the above problems, except the last one, could be solved by this method as discussed in chapter 3.

For the problem of finding optimum number of clusters, kernel-based Dunn's index was implemented as a cluster validation method. Applying a cluster validation method in size chart development field was a new approach while applying kernel-based cluster validation method was an improved approach. It successfully optimized the number of clusters avoiding a trial and error method.

The main important feature in this size chart was the ability of multiple choices of hip measurement and inseam length for a particular waist measurement. This would provide the customer several choices for achieving a better fit in their pants. These size charts also provide percentage of coverage for each size which might beneficial for the manufacturer as well as the retailer in inventory management.

In addition, size chart validation which was done statistically and using live fit assessment followed by a questionnaire survey showed that the size charts are clearly representing the Sri Lankan females aged 20-40 years, which will lead to a reduction of fit problems in ready-to-wear pant market. Even though, the inventory was slightly higher in this system than conventional system, it would compensate by the higher level of customer satisfaction resulting higher sales, lower returns and higher level of brand loyalty.

Therefore, the global kernel *k*-means clustering approach combined with kernel based Dunn's index provides a complete solution for the problem of anthropometric data clustering in the subject of size charts development.

6.2 Limitations of the Research

There were limitations in selecting samples for the anthropometric data collection. Therefore, the selected sampling method was a convenience sampling method instead of a random sampling method. Convenience sampling technique was used in similar studies also despite its poor generalizability due to less representativeness of population by the sample. However to improve the representativeness of the sample all provinces and different professions were covered and thereby tried to minimize the limitations of generalizability.

The selected age limit was 20-40 years based on biological aspects. Hence, females above 40 years and below 20 years of age need to be considered separately. Further, getting anthropometric data is very time consuming and costly. Therefore, the size of the sample was limited to 1430 subjects.

In this research, only widely recognized existing methods were explored, although there are several other methods, such as integer programming, have been suggested in academia.

6.3 Future Possibilities for Exploration

This research was focused on fit problems of female pants. The researched technique of developing size charts could be extended further with the use of full body measurements when developing size charts for different types of garments. The findings will be beneficial for Sri Lankan apparel manufactures as well as retailers. Further, this research was restricted to a limited age group. It could be broadened to represent the majority of the population. The process would have been more efficient and accurate if a 3-D body scanner was used in measuring human body.

In this research it was revealed that, different lower body shapes need to be identified for achieving a perfect fit. Although, the researcher could realize three lower body shapes, further development need to be done incorporating width/ depth and arc proportions in defining varied lower body shapes.

Further, the global kernel *k*-means clustering approach used in clustering anthropometric data might be used for other clustering problems successfully by selecting proper kernel function and its tuned parameters.

REFERENCES

Abeysekera, J.D.A. &Shahnavaze, H. (1998). Body size data of Sri Lankan workers & Comparison with other populations in the world: Its impact on the use of imported goods. *Journal of National Science Council, Sri Lanka*, 16(1), 67-86.

Alexander, M., Connell, L.J., & Presley, A. (2005). Clothing fit preferences of young female adult consumers. *International Journal of Clothing Science and Technology*, Vol. 17(1), pp.52 - 64.

Amnon,S. (2009). "Introduction to Machine Learning: Class Notes 67577" arXiv:0904.3664v1 [cs.LG].[Accessed on 06.05.2014]

Anderson, L. J. ,Brannon, E.L., Ulrich, P.V., Marshall,T., Staples, N. ,Grasso,M. *et al.* (1997). Discovering the process of mass customization; A paradigm shift for competitive manufacturing .*National Textile Centre Research Briefs*,I95-A19

Anderson, L. J. ,Brannon, E.L., Ulrich, P.V., Presley,A.B., Woronka,D., Grasso,M. *et al.*(1999). Understanding Fitting Preferences of Female Consumers: Development an Expert System to Enhance Accurate Sizing Selection. *National Textile Centre Annual Report* , I98-A8/1- 9.

Ariadurai,A.,Nilusha,T., Dissanayake,M. (2009). An Anthropometric Study on Sri Lankan School Children for developing clothing sizes. *Journal of Social Sciences* 19(1) ,pp51-59.

Ashdown, S.P.(1998) "An investigation of the structure of sizing systems: A comparison of three multidimensional optimized sizing systems generated from anthropometric data with the ASTM standard D5585-94", *International Journal of Clothing Science and Technology*, Vol. 10 (5), pp.324 – 341

Ashdown, S.P., Locker, S.,Ruker, M., (2007). Improved Apparel sizing: Fit and Anthropometric 3D scan data. *National Textile Centre Research Briefs*: S04-CR01

Bagherzadeh, R., Latifi, M., & Faramarzi, A.R. (2010). Employing a Three-Stage Data Mining Procedure to Develop Sizing System. *World Applied Sciences Journal* 8 (8), pp.923-929.

Barnett, V., & Lewis, T. (1994). Outliers in Statistical Data(3rd ed.) Chichester: John Wiley & Sons.

Bagirov, A.M., &Mardaneh, K. (2006).Modified global k-means algorithm for clustering in gene expression datasets. *Conferences in Research and practice in Information Technology*, Vol. 73. Hobart: Australian Computer Society, Inc.

Bartlett, J., Kotrlik, J., & Higgins, C. (2001).Organizational Research: Determining Appropriate Sample size in Survey Research. *Information Technology, Learning and Performance Journal,* Vol 19(1), pp.43-50.

Beazley, A. (1997). Size and Fit: Procedures in undertaking a survey of body measurements. *Journal of Fashion Marketing and Management*, Vol.2 (1), pp.55-85.

Beazley, A. (1998). Size and fit: Formulation of body measurement tables and sizing systems; part2. *Journal of Fashion Marketing and Management* Vol. 2 (3), pp.260-284.

Beazley, A. (1999). Size and Fit: The development of size charts for clothing; part 3. *Journal of Fashion Marketing and Management*; Vol. 3, (1), pp.66 -84.

Belle, G. (2002). Statistical Rules of Thumb. New York: John Wiley & Sons Inc.

Bezdek, J. C., & Pal, N.R.(1998). Some new indexes of cluster validity. *IEEE Transactions on Systems, Man, and Cybernetics,* Part B 28(3), pp.301-315.

Bickle, M.C., Kotsiopulos, A., Dallas, M.J., & Eckman, M. (1995). Fit of Women's Jeans: An Exploratory Study Using Disconfirmation Paradigm. *Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior* Vol.8, pp.208-213.

Boguslawska, M. (2010). Analysis of the contemporary problem of garment sizes. TEXSCI. Czech Republic.

Bye, E., LaBat, K.L., & DeLong, M.R. (2006). Analysis of body Measurement systems for Apparel, *Clothing & Textile Research Journal* Vol.24, pp.66-79.

Carrier, S. F. (2009). What is your target market 's size and shape. *Advances in Production Engineering & Management*, Vol.4 , pp.5- 18.

Chakrabarti, D. (1997). Indian Anthropometric Dimensions for Ergonomic Design Practice. Ahmedabad: National Institute of Design.

Chitta, R., Jin, R., Havens, T. C., and Jain, A. K.(2011) ,Approximate Kernel k-means: solution to Large Scale Kernel Clustering. KDD, San Diego, CA,

Chung, M., Lin, H., & Wang, M. (2007). The development of sizing systems for Taiwanese elementary and high school students. *International Journal of Industrial Ergonomics*Vol.37(8), pp.707- 716.

Chun-Yoon, J., & Jasper, C.R. (1996). Key Dimensions of Women's Ready-to-Wear Apparel: Developing a Consumer Size- Labeling System. *Clothing and Textile Research Journal*, Vol.14(1), pp.89-95.

Cooklin, G. (1995). Master Patterns and Grading for women's outsizes. Oxford: Blackwell Science.

Cole, T. (2003). The Secular Trend in Human Physical Growth: A Biological View. *Economics and Human Biology*, Vol.1(2), pp. 161-168 .

Cristianini, N., Taylor, J.S., & Saunders, C. (2007). Kernel Methods: A Paradigm for Pattern Analysis. In Valls,G., Alvarez, J.,& Ramon, M., Kernel Methods in Bioengineering, Signal and Image Processing (pp.1-41). Hershey: Idea Group Publishing.

Croney, J. (1980).Anthropometry for Designers. London: Batsford Academic & Educational Limited.

Deas Marteau, K. (2000).CAD: Let the fit revolution begin. *Bobbin* 42(2), pp.42-56

DeLong, M.R., Ashdown, S.P., Butterfield, L., & Turnbladh, K. (1993). Data specification needed for apparel production using computers. *Clothing & Textiles Research Journal* Vol. 11(3), pp.1-7.

Dixson, B. J., Grimshaw, G. M., Linklater, W. L., & Dixson, A. F. (2010). Watching the hourglass: Eye tracking reveals men's appreciation of the female form. *Human Nature*, Vol.21(4), pp.355-370.

Donelson, S.M., & Gordon, C.C. (1996). 1995 Mached Anthropometric database of U.S. Marine Corps Personnel: Summary Statistics. Massachusetts: GEO Centres, Inc.

Doustaneh, A.H., Gorji, M., &Varsei, M. (2010).Using Self Organization Method to Establish Nonlinear Sizing System. *World Applied Sciences Journal,* Vol.9 (12), pp.1359-1364.

Dunn, J.C. (1973). A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters. *Journal of Cybernetics* Vol. 3(3), pp.32-57.

Ellis, H. (1997). Clinical Anatomy: A revision and applied anatomy for clinical students . UK: Blackwell Science Ltd.

Evangelista, P., Embrechts, M., & Szymanski, B. (2007). Some Properties of the Gaussian Kernel for One class Learning. *International Conference on Artificial Neural Networks* (pp.269-278). Berlin: Springer.

Fa, R., Nandi, A.K., & Jamous, B.A. (2012).Development and Evaluation of Kernel-Based Clustering Validity.*20th European Signal Processing Conference* (pp.634-638). Bucharest: EURASIP.

Fan, J., Yu, W., & Hunter, L. (2004). Clothing Appearance and Fit: Science and Technology. Cambridge: Woodhead Publishing Ltd.

Faust, M., Carrier, S., & Baptist, P. (2006). Variations in Canadian women's ready-to-wear standard sizes. *Journal of Fashion Marketing and Management*, Vol.10(1) , pp.71 - 83.

Frienlander, J.S., Costa, P.T., Bosse, R., Ellis, E., Rhoads, J.G. and Stoudt, H.W.(1977), Longitudinal physique changes among healthy white veterans at Boston. *Human Biology,* Vol.49, pp.541-558

Gunnell, D. J., Smith, G.D., Frankel, S.J, Kemp, M., & Peters, T. (1998). Socio-economic and dietary influences on leg length and trunk length in childhood: a reanalysis of the Carnegie (Boyd Orr) survey of diet and health in prewar Britain (1937–39). *Pediatric and Perinatal Epidemiology,* Vol. 12(S1), pp.96-113.

Gupta, D., & Gangadhar, B.R. (2004).A Statistical Model for Developing Body Size Charts for Garments. *International Journal of Clothing Science and Technology*, Vol.16 (5), pp.458-469.

Gupta, S. (1993).Research methodology and statistical Techniques. New Delhi: Deep & Deep Publications.

Haggar, A. (1990). Pattern Cutting for Lingerie, Beachwear & Leisurewear. Oxford: Blackwell Publishing.

Halkidi, M., Batistakis, Y., & Vazirgiannis, M. (2001).On Clustering Validation Techniques. *Journal of Intelligent Information Systems*, Vol.17(2/3), pp.107–145.

Han, J., Kamber, M., & Pei, J. (2012).Data mining Concepts and Techniques. Waltham: Morgan Kaufmann publishers.

Hashim, Y.A. (2010). Determining sufficiency of sample size in management survey research activities. *International Journal of Organizational Management& Entrepreneurship Development*,Vol.6, pp.119-130.

Hofmann, M., Kernels and the Kernel Trick "Support Vector Machines" http://www.cogsys.wiai.uni-bamberg.de/teaching/ss06/hs_svm/slides/SVM_and_Kernels.pdf

Hsu, C. , (2007). Employing Data Mining to Identify the Significant Rules for Classifying Body Types. *WSEAS International Conference on Computer Engineering and Applications*, (pp. 29-34). Gold Coast.

Hsu, C. (2009a). Data mining to improve industrial standards and enhance production and marketing: An empirical study in apparel industry. *Expert systems with Applications*, Vol.36(3) , pp4185-4191.

Hsu, C. (2009b). Developing Accurate Industrial Standards to Facilitate Production in Apparel Manufacturing Based on Anthropometric Data. Human Factors and Ergonomics in Manufacturing, Vol.19(3), pp.199–211.

Hsu, C., & Wang, M. (2004). Using Decision tree based data mining to establish a sizing system for the manufacture of garments. *International journal of Advanced Manufacturing Technology,* Vol.(26), pp.669-674.

Iseri, A. (2008). Determining the Anthropometric characteristics of Turkish population. Istanbul: (Faith University).

Izenman, A.J. (2008). Modern Multivariate Statistical Techniques. NewYork: Springer Science+ Business Media.

Jongsuk, C.Y. , Jasper, C.R. (1993). Garment-sizing Systems: An International Comparison. *International Journal of Clothing Science and Technology*,Vol.5(5), pp.28-37.

Jongsuk, C. Y. ,Jasper, C.R. (1996). Key dimensions of Women's Ready-to-wear Apparel: Developing a Consumer Size Labeling System. *Clothing and Textile research Journal* Vol. 14 (1) , pp 89-95.

Khardon, R.(2008) Computational Learning Theory http://www.cs.tufts.edu/~roni/Teaching/CLT2008S/ [Accessed on 22.01.2014]

Kim, D.W., Lee, K.Y., Lee, D., & Lee, K.H. (2005).Evaluation of the performance of clustering algorithms in kernel-induced feature space. *Pattern Recognition* Vol.38, pp.607 – 611.

Krejcie, R.V., and Morgan, D.W. (1970). Determining Sample Size for Research Activities. Educational and Psychological Measurement Vol.30(3), pp.607-610

LaBat, K.L., and DeLong, M.R. (1990). Body cathexis and satisfaction with fit of apparel. *Clothing Textile Research Journal* Vol.8(2), pp.43-48.

Laitala, K., Klepp, I.G., &Hauge, B. (2011). Materialized Ideals: Sizes and Beauty. *Culture Unbound* Vol.3, pp.19-41.

Lassek, W.D., & Gaulin, S.J.C. (2008). Waist-to-hip ratio and cognitive ability: Isgluteofemoral fat a privileged store of neuro developmental resources? *Evolution and Human Behavior* Vol.29, pp.26-34.

Lee, J.Y., Istook, C.L., Nam, Y.J., & Park, S.M. (2007). Comparison of body shape between USA and Korean women. *International Journal of Clothing Science and Technology* Vol. 19(5), pp.374-391.

Lidwell, W., Holden, K., & Butler, J. (2003).Universal Principles of Design. Massachusetts: Rockport Publishers.

Likas, A., Vlassisb, N., & Verbeekb, J. (2003).The global k-means clustering algorithm. *The Journal of Pattern Recognition Society* ,Vol.36, pp.451-461.

Lin, H., Hsu, C., Wang, M.& Lin, Y. (2007). An application of data mining technique in Engineering to facilitate production management of garments. *11th WSEAS international Conference on Computers*. Greece: Agios Nikolaos.

Lin, H., Hsu,C., Wang, M. & Lin,Y. (2008). An Application of Data Mining Technique in Developing Sizing System for Army Soldiers in Taiwan. *WSEAS Transactions on Computers*, Vol.7(4), pp.245-252.

Lin, H.T., & Lin, C.J. (2003). A Study on Sigmoid Kernels for SVM and the Training of non-PSD Kernels by SMO-type Methods. CiteSeer X.

Liu, Z., Chen, D., & Bensmail, H. (2005). Gene Expression Data Classification With Kernel Principal Component Analysis. *Journal of Biomedicine and Biotechnology* Vol. 2005(2), pp.55-159.

Lodhi, H., Saunders, C., Taylor, J.S., Cristianini, N., & Watkins, C. (2002). Text Classification using String Kernels. *The Journal of Machine Learning Research* Vol.2, pp.419-444.

Loker, S., Ashdown, S.P., & Schoenfelder, K. (2005). Size- specific Analysis of Body Scan Data to Improve Apparel Fit. *Journal of Textile and Apparel, Technology and Management*, Vol. 4 (3), pp.1-7 .

Mampa, M.L., Azariadis, P.N., & Sapidis, N.S. (2010).A new methodology for the development of sizing systems for the mass customization of garments. *International Journal of Clothing Science and Technology*, Vol. 22 (1), pp.49 - 68.

Martini, F.H., & Bartholomew, E.F. (2000).Essentials of Anatomy & Physiology. New Jersey: Prentice-Hall Inc.

McCulloch, C.E., Paal, B., & Ashdown, S.P (1998). An Optimization approach to apparel sizing. *Journal of the Operational Research Society*,Vol.49, pp.492-499.

Merry, C.V. (2005). Mind - Primary Cause of Human Evolution . Victoria: Trafford Publishing.

Methasate, I., & Theeramunkong, T. (2007).Experiments on Kernel Tree Support Vector Machines for Text Categorization. In Z. Zhou, & Q. Yang, Advances in Knowledge Discovery and Data Mining Vol. 4426 (pp.720-727). Berlin: Springer-Verlag.

Molarius, A. (1999). Waist and hip circumferences, and waist-hip ratio in 19 populations of the WHO MONICA project. International Journal of Obesity(23) , pp.116-125.

Otieno, R., Harrow, C., & Greenwood, G.L. (2005).The unhappy shopper, a retail experience: exploring fashion, fit and affordability. *International Journal of Retail & Distribution Management*, Vol. 33(4), pp.298 - 309.

Pal, N.R., & Biswas, J. (1997). Cluster Validation using graph theoretic concepts. Pattern Recognition, Vol.30(6), pp.847-857

Pechoux, B., & Ghosh, T. (2002). Apparel sizing & Fit. Manchester: The Textile Institute.

Petrova, A., & Ashdown, S.P. (2008). Three - dimensional Body Scan Data Analysis:Body Size and Shape Dependence of Ease Values for Pants' Fit. *Clothing and Textiles Research Journal,* Vol.26(3), pp.227-252.

Pisut, G., & Connell, L. (2007). Fit preferences of female consumers in the USA. Journal of Fashion Marketing and Management, 11(3), pp.366-379.

Pheasant, S. (1996). Bodyspace : anthropometry, ergonomics and design of work. London: Taylor& Francis.

Phillips, J., & Venkatasubramanian, S. (2011). A Gentle Introduction to the kernel Distance. Computing Research Repository-CORR Vol. abs/1103.1 .

Pochet, N.L.M., Ojeda, F., De Smet, F., De Bie, T., Suykens, J., & De Moor, B. (2007).Kernel Clustering for Knowledge Discovery in Clinical Microarray Data. In G. Valls, J. Alvarez, & M.

Ramon, Kernel Methods In Bioengineering, Signal and Image Processing (pp.64-92). Hershey: Idea Group Publishing

Rai, P. (2011). Kernel Methods and Nonlinear Classification .http://www.cs.utah.edu/~piyush/teaching/15-9-print.pdf. [Accessed on 12.03.2014]

Ramos, J. (2012). Dunn's Index
http://www.mathworks.com/matlabcentral/fileexchange/27859-dunns-index/content/dunns.m [Accessed on 15.05.2014]

Rendón, E., Abundez, I., Arizmendi, A., & Quiroz, E. (2011).Internal versus External cluster validation indexes. *International Journal of Computers and Communications* Vol. 5(1), pp.27-34.

Roach, M. (1996). The Numbers Game. *Vogue* , 94.

Roebuck, J.A. (1995). *Anthropometric Methods: Designing to fit the Human Body.* Santa Monica: HFES Publications Department.

Rokach, L., & Maimon, O.(Eds.) (2010). *Data Mining and Knowledge Discovery Handbook.* Springer US.

Sarma, T.H., Viswanath, P., & Reddy, B.E. (2011). Fast Approximate Kernel k-means Clustering Method For Large Data sets. Recent Advances in Intelligent Computational Systems (RAICS), (pp. 545 - 550). IEEE.

Shahrabi, J. ,Esfandarani, M.S., (2010). Development of a new Sizing System Based on Data Mining Approaches.*7th International Conference- TEXSCI* . Liberec, Czech Republic

Sharma, S. (1996).Applied Multivariate Techniques. Newyork: John Wiley Sons, Inc.

Simmons, K., Istook, C.L., & Devarajan, P. (2004). Female Figure Identification Technique(FFIT) for Apparel; Development of Shape Sorting Software. *Journal of Textile and Apparel, Technology and Management*, Vol.4(1), pp.1-15.

Song, H.K., & Ashdown, S.P (2012). Development of Automated Custom- Made Pants Driven by Body Shape. *Clothing and Textile research Journal* Vol.30(4), pp.315-329.

Streeter, S.A., & McBurney, D.N. (2003).Waist–hip ratio and attractiveness, New evidences and a critique of "a critical test". *Evolution and Human Behavior* Vol.24, pp.88-98.

Strydom, M., & Klerk, H.M. (2010). Key to good fit: Body measurement problems specific to key dimensions. *Journal of Family Ecology and Consumer Sciences*, Vol.38, pp.74- 83.

Sugiyama, L.S. (2005). Physical Attractiveness in Adaptationist Perspective. Handbook of Evolutionary Psychology (pp.292-343). Hoboken: John Wiley.

Singh, D. (1993). Adaptive significance of female physical attractiveness: Role of waist-to-hip ratio. *Journal of Personality and Social Psychology, 65(2),* pp.293–307

Taylor, J.S., & Cristianini, N. (2004).Kernel Methods for Pattern Analysis. NewYork: Cambridge University Press.

Tzortzis, G.F, & Likas, A.C. (2009).The Global Kernel K-Means Algorithm for Clustering in Feature Space. IEEE Transactions on Neural Networks, Vol.20(7), pp.1181-1194.

Taylor, P. (1990). Computers in the Fashion Industry. Oxford: Heinemann Professional Publishing Ltd.

Tryfos, P. (1986, October). An Integer Programming Approach to the Apparel Sizing Problem. The Journal of the Operational Research Society, Vol.37(10), pp.1001-1006.

Tseng, M., & Jiao, J. (2001). Mass Customization: Handbook of Industrial Engineering, Technology and Operation Management. New York: Wiley.

Veblen, S. (2012). The Complete Photo Guide to Perfect Fitting. Minneapolis: Creative Publishing International, Inc.

Wadsworth, M.E., Hardy, R.J., Paul, A., Marshall, S., & Cole, T. (2001).Leg and trunk length at 43 years in relation to childhood health, diet and family circumstances; evidence from the 1946 national birth cohort. *International Journal of Epidemiology* Vol. 31(2), pp383-390.

Wang, B., Xiong, H., Jiang, X., & Ling, F. (2012).Semi-Supervised Object Recognition Using Structure Kernel. *19th IEEE International Conference on Image Processing (ICIP)*, (pp.2157 - 2160). IEEE.

Winks, J.M. (1997). Clothing Sizes: International Standardization. Manchester: The Textile Institute.

Wardle, J. W. (1996). Body fat distribution in South Asian women and children. *International Journal of obesity & Related Metabolic Disorders*,Vol.20(3), pp.267-271.

Xie, J., Jiang, S., Xie, W., & Gao, X. (2011). An Efficient Global K-means Clustering Algorithm. *Journal of Computers,* Vol.6(2), pp.271-279.

Yunzhang, Z. (2011). *CSCI1430 :Project 4 Face detection with a sliding window.* from cs.brown.edu: http://cs.brown.edu/courses/csci1430/2011/results/proj4/yunzhang / [Accessed on 15.09.2014]

Yoon, J., & Jasper, C.R. (1996). Key Dimensions of Women's Ready-to-Wear Apparel: Developing a Consumer Size-Labeling System. *Clothing and Textiles Research Journal January* vol. 14(1), pp.89-95 .

Zien, A., Rätsch, G., Mika, S., Schölkopf, B., Lengauer, T., & Müller, K. (2000). Engineering support vector machine kernels that recognize translation initiation sites in DNA. *Bioinformatics*, vol.16, pp.799–807.

Zakaria, N. ,Taib, J.S.,Tan,Y.Wah,Y. (2008).Creating new sizing system for children's wear using anthropometric data. *5th International Cyberspace Conference on Ergonomics* .

Zilber, S.N., & Nohara, J.J. (2009).Mass Customization and Strategic Benefits :A Case Study in Brazil. *The Electronic Journal on Information Systems in Developing Countries* Vol.36(5), pp.1- 26 .

Zipkin,P. (2001). The Limits of Mass Customization.*Magazine* http://sloanreview.mit.edu/article/the-limits-of-mass-customization/ (Accessed on 08.05.2014)

Zhang, R., & Rudnicky, A. (2002).A large scale clustering scheme for kernel k-means. *The 16th International Conference on Pattern Recognition* (pp. 289-292). IEEE.

Appendix 2.1Summary of anthropometric surveys

Anthropometric surveys done in USA

Anthropologists/ country	Year	Sample Details	Sample size	Published by	Year
Medical dept. of the US Army	1917	World War I soldiers	2,000,000	Davenport and Love	1921
Field,W. Dayton	WWII	US air forces personnel	-	Randall et.al	1946
Randall,F.E.	1946	Army Corps -Men -women	105,062 8,864	Newman and White Randall & Munro	1951 1949
Beer,W.J.	1948	Marine Corps men	2,000	-	-
US Army Quarter-master Corps	1949	Army men	7,272	-	-
US Air force	1950	Flying personnel	4,063	Hertzberg, Daniels & Churchill	1954
US Air force	1952	Air force male trainees	3 332	Daniels,Meyers& Churchill	1953
Us Navy	1964	Navy & Marine Corps	1 549	Gifford,Provost&Lazo	1965
Us Air force	1965	Air force trainees	2 632- 158mmts	Churchill, Kikta	1977
US Army, Navy & Marine	1965	Army personnel Navy Marine Corps	6 682 4 095 2 008	White & Churchill - White & Churchill	1971 1977
US Air force	1967	Flying personnel	2 420- 87mmts	Grunhofer&Kroh	1975
US Army	1968	Flight trainees	1,640	Schane, Littell& Moultrie	1969
US Army	1970	Army Aviators	1,482- 85mmts	Churchill et.al	1971
US Air force	1968	Air force women officers,	1905- 137mmts	Clauser et.al	1972
US Army	1976	Women officers, ,nurses	1331- 128mmts	Churchill et.al	1977
US Civilian					
Bureau of Home Economics	1939	US Civilian Women	15,000-59mmts	O'Brien and Shelton	1941
US Dept. of Health, &Education	1962	US Civilian	3,091-men 3,581 women	Stoudt et.al	1965;1970
US Dept.ofHealth,Education& Welfare	1971	Men &women	20,749	Abraham et al.	1976
Naval Electronics Lab	1974	Law enforcement officers	3000-23mmts	Martin et al.	1975

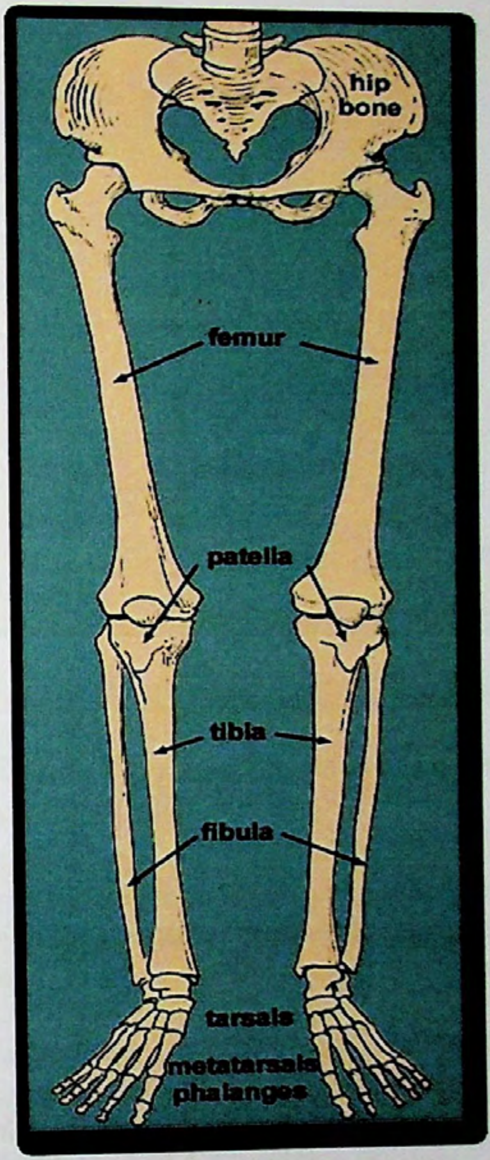
Anthropometric surveys done in other countries

Country	Year	Sample Details	Sample size	Published by	Year
Canada	1974	Forces personnel	565 - 32mmts	McCann et al.	1975
US Army Tropic test centre	1965-70	People in 18 south & Central American countries	1985	Dobbins &Kindick	1972
British Army	1975	Royal armoured Corps	1134 - 62mmts	Gooderson&Beebee	1976,77
Royal Aircraft Establishment	1970	Royal Air force air crew	2000 - 63mmts	Bolton et al.	1973
University of Paris	1967	Military personnel	2000 - 67mmts	Coblentz&Ignazi	1968
Germany	1967	Airforce personnel	1465	Grunhofer&Kroh	1975
Germany	1968	German dreftees	7144- 43mmts	Jurgens, Benthin&Lengsfeld	1972
Germany	1970	Armed Forces men	2643- 54mmts	Jurgens, Benthin&Lengsfeld	1973
NATO	1960	Turkey,Greece, Italy Forces	3356	Hertzberg et al.	1963
South Africa	1968	Armed Forces men	1445- 57mmts	Strydom et al.	1968
Australia	1968	Armed Forces men	3695- 28mmts	Maclean	1968
Australia	1969	Army men	3695- 11mmts	Australian Army	1970
New Zealand	1972	Aircrew	238- 62mmts	Toulson	1974
Thailand	1962	Army men	2950- 52mmts	White	1964
Vietnam	1963	Armed Forces men	2128	White	1964
Japan	1961	Airforce Pilots	239- 62mmts	Oshima et al.	1962
Japan	1972	Airforce personnel	1176- 108mmts	Japanese Air force	1972
Republic of Korea	1961	Airforce pilots	264- 132mmts	Kay	1961
South Korea	1965	Armed Forces men	3747- 59mmts	Hart,Rowland&Malina	1967
Iran	1968	Forces men	9414- 68mmts	Noorani& White	1971
Sri Lanka	1981/82	Men & women	724- 90 mmts	Abeysekera, &Shahnavaze,	1998

Source: White,R.M.(1978) Anthropometry and Human Engineering, yearbook of Physical Anthropology vol(21)

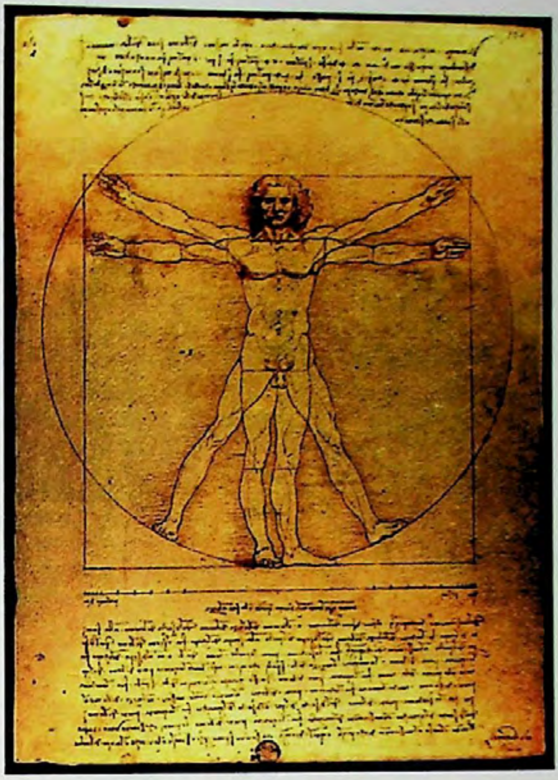
Appendix 2.2

Anatomical structure of human lower body



source: http://home.comcast.net/~wnor/llbones.htm [Accessed on 15.01.2012]

Appendix 2.3 Vitruvius man



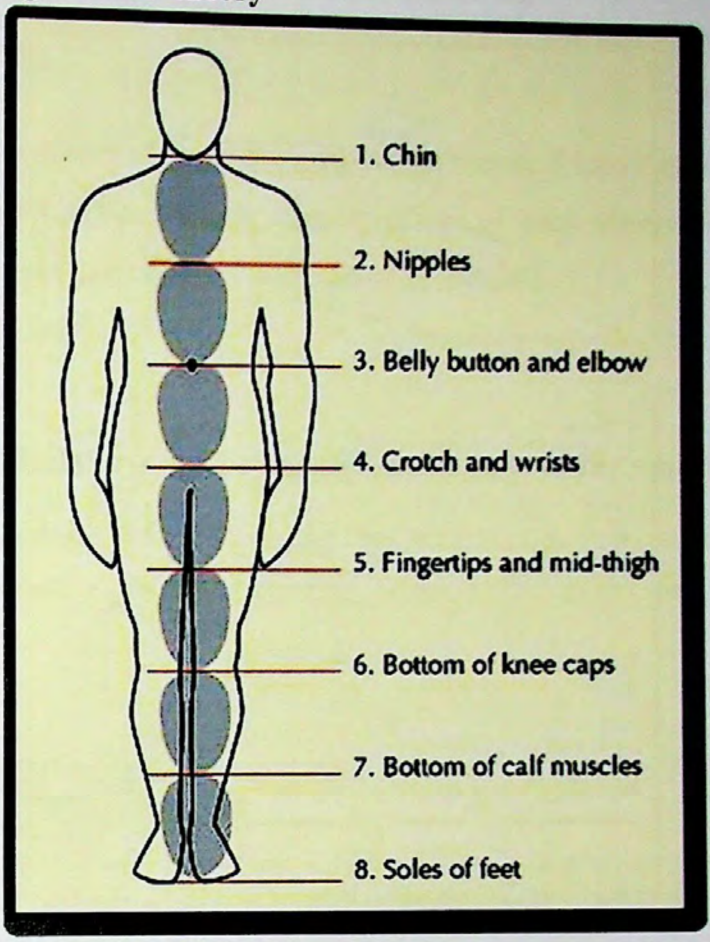
source:http://en.wikipedia.org/wiki/Vitruvian_man[Accessed on 10.08.2014]

Body proportions explained in Vitruvius, The ten books on Architecture (translated by Morgan, M.H) are as follows;

- the length of the outspread arms is equal to the height of a man
- from the hairline to the bottom of the chin is one-tenth of the height of a man
- from below the chin to the top of the head is one-eighth of the height of a man
- from above the chest to the top of the head is one-sixth of the height of a man
- from above the chest to the hairline is one-seventh of the height of a man.
- the maximum width of the shoulders is a quarter of the height of a man.
- from the breasts to the top of the head is a quarter of the height of a man.
- the distance from the elbow to the tip of the hand is a quarter of the height of a man.
- the distance from the elbow to the armpit is one-eighth of the height of a man.
- the length of the hand is one-tenth of the height of a man.
- the root of the penis is at half the height of a man.
- the foot is one-seventh of the height of a man.
- from below the knee to the root of the penis is a quarter of the height of a man.
- the distances from the below the chin to the nose and the eyebrows and the hairline are equal to the ears and to one-third of the face.

(Pollio, 2006)

Appendix 2.4 Eight heads theory



Source:http://mhsart1.wikispaces.com/Proportions[Accessed on 03.03.2012]

Nape to the level of armpit= one-eighth of the height (i.e. 1 head).

Natural waist (i.e. nape to waist) = one-fourth of the height (i.e.-2 heads).

Fore-arm (i.e. armpit to wrist bone) = one-fourth of the height (i.e. 2 heads).

Elbow to armpit = one-eighth of the height (i.e. 1 head).

Inside leg or leg measure = half the full height (i.e. 4 heads) less 5 to 6cm (2 to 2 ¼").

Slope of shoulder= one-sixth of the natural waist length.

Sleeve length (upto wrist) from shoulder =three-eighth of the height (i.e. 3 heads) less 2 to 4 cm 3/4 to 1 ½").

Both the arms extended = full height of the figure (i.e. 8 heads).

Source: http://www.b-u.ac.in/sde_book/fashion_design.pdf[Accessed on 20.03.2012]

Appendix 3.1 Matlab syntax for *k*-means clustering

idx = kmeans(X,k)

This partition the observations of the *n*-by-*p* data matrix X into k clusters, and returns an *n*-by-1 vector (idx) containing cluster indices of each observation. Rows of X correspond to points and columns correspond to variable.
(Source: Matlab help)

Appendix 3.2 Matlab syntax for kernel k-means clustering approach

```
function [Cluster_elem,Clustering_error,Centre_dist]=
Weighted_Kernel_K_Means(Cluster_elem,K,Dataset_Weights,Cluster
s,Display)


%Function Inputs
%===============

%Cluster_elem is an initial partitioning of the dataset. It is
a column vectorcontaining the cluster index of each point. The
clusters are indexed1,2,...,Clusters.

%K is the kernel matrix of the dataset. It must be a positive
definitesquare matrix (Gram matrix) in order to guarantee
algorithm convergence.
%If K is not positive definite the algorithms may still
converge though.

%Dataset_Weights is a column vector containing the weight of
each
%datapoint. It is used to run the weighted version of Kernel
K-Means. By setting allweights equal to 1 the non-weighted
version is run. The weights must be positive numbers.

%Clusters is the number of clusters.

%Display when set to 'details' prints information after each
iteration.
%Any other value results in no printing.
%Function Outputs
%================
%Cluster_elem is a column vector containing the final
partitioning of the dataset. The clusters are indexed
1,...,Clusters.
```

```
%Clustering_error is the value of the Kernel K-Means objective
function corresponding to the final partitioning..

%Centre_dist is a column vector containing the distance of
each point to its cluster centre.

%Converge if clustering error difference is less than e.
e=0.000001;

%Store the objective function value.
Clustering_error=0;

%Dataset size.
Data_num=size(K,1);

%Store the distance between points and their cluster centre.
Centre_dist=zeros(Data_num,1);

Iter=1;

while 1

%Keep the clustering error of the previous iteration.
Old_clustering_error=Clustering_error;
Clustering_error=0;

    Intra=zeros(Clusters,1);
Cluster_dist=zeros(Data_num,Clusters);

Fori=1:Clusters

%Find the dataset points that belong to cluster i and their
weights.
This_elem=find(Cluster_elem==i);
Dataset_Weights_This_elem=Dataset_Weights(This_elem);

%Calculate intra-cluster pairwise quantity for cluster i.
Intra(i)=(K(This_elem,This_elem)*Dataset_Weights_This_elem)'*D
ataset_Weights_This_elem;
Intra(i)=Intra(i)/sum(Dataset_Weights_This_elem)^2;

%Calculate point-cluster quantity between all points and
cluster i.
Cluster_dist(:,i)=K(:,This_elem)*Dataset_Weights_This_elem;

%Calculate the distance of all points to the centre of cluster
i.
Cluster_dist(:,i)= (-
2*Cluster_dist(:,i)/sum(Dataset_Weights_This_elem))+Intra(i)+d
iag(K);
```

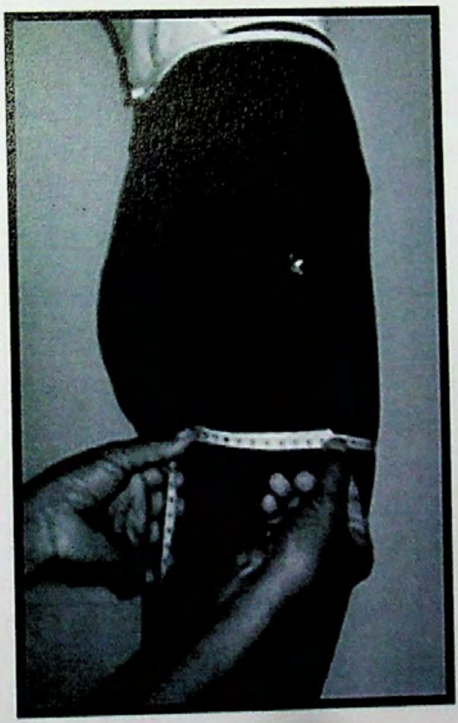
```
%Store the distance of cluster'si points to their cluster
centre.
Centre_dist(This_elem)=Cluster_dist(This_elem,i);

%Add the contribution of cluster i to the clustering error.

Clustering_error=Clustering_error+Dataset_Weights_This_elem'*C
luster_dist(This_elem,i);
end

%Update the assignment of points to clusters by placing each
point to the closest centre.
    [min_dist,Update_cluster_elem]=min(Cluster_dist,[],2);

ifstrcmp(Display,'details')
fprintf('\nKernel K-Means Iteration %d\n',Iter);
fprintf('Clustering error=%g.\n',Clustering_error);
end

%Check for convergence.
IfIter>1

Ifstrcmp(Display,'details')
fprintf('Clustering error reduced by
%g\n',Old_clustering_error-Clustering_error);
end

if abs(Old_clustering_error-Clustering_error) < e
break;
end
end

Cluster_elem=Update_cluster_elem;

%Drop empty clusters.
Count=0;
Fori=1:Clusters
if size(find(Update_cluster_elem==i),1)==0
tmp=find(Update_cluster_elem>i);
Cluster_elem(tmp)=Cluster_elem(tmp)-1;
Count=count+1;
warning('Droping empty cluster');
end
end
%Reduce the number of clusters if some were dropped.
    Clusters=Clusters-count;

Iter=Iter+1;
end

return
```

Source: (Tzortzis&Likas, 2009)

Appendix 4.1 Female lower body measuring procedure

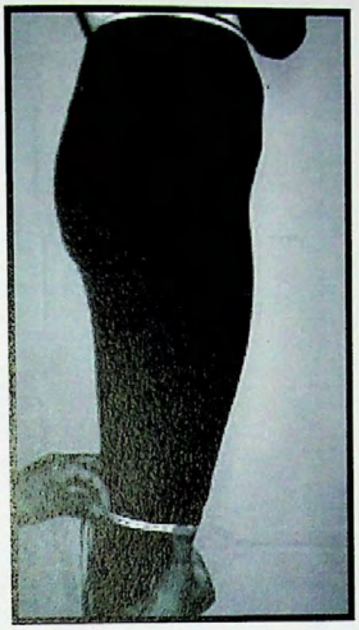
(i)Measuring of waist girth

(ii) Measuring of hip girth



(iii)Measuring of thigh girth

(iv) Measuring of mid-thigh



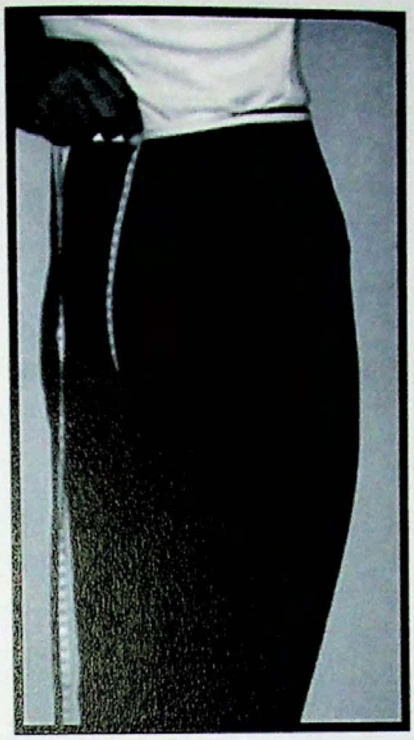
(v) Measuring of knee girth

(vi) Measuring of mid-calf girth

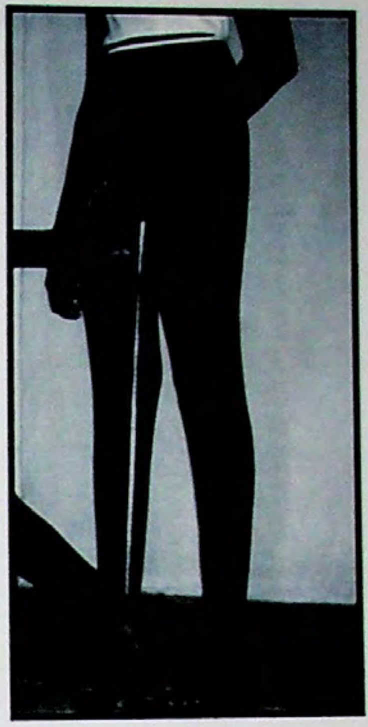
(vii) Measuring ankle girth

(viii)Measuring centre front waist

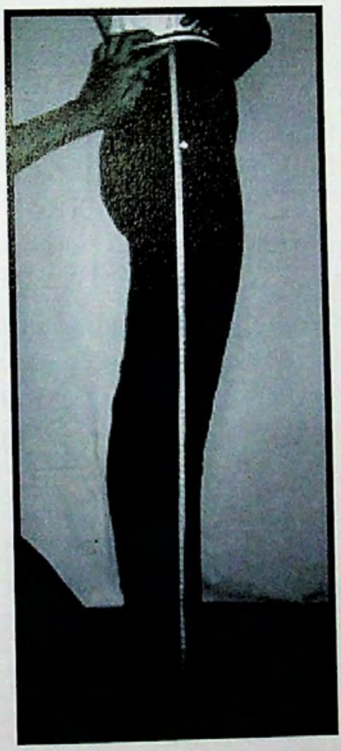
to hip



(ix)Measuring crotch length



(x) Measuring Inside leg



(xi)Measuring outseam

(xii)Measuring knee height

(xiii)Measuring ankle height

Appendix 4.2 Matlab syntax for Polynomial Kernel function

```
gamma   = 1;

delta  =2;
Temp    = zeros(13,1);
matrix1=zeros(629);
fori  =1:629
for j = 1:629
Temp   = dot(mediumWHR13(i,:)',mediumWHR13(j,:));
matrix1(i,j) = (Temp + gamma)^delta;
matrix1(j,i) = matrix1(i,j);
end
% matrix1 = zeros(629,629);
end


xlswrite('polynomialmedium-d2.xlsx',matrix1)
```

Appendix 4.3 Matlabsyntaxfor Gaussian Kernel function

```
Sigma = 8;

Temp= zeros(13,1);
matrix1=zeros(315);
fori =1:315
for j = i:315
        Temp = largeWHR13(i,:)-largeWHR13(j,:);
matrix1(i,j) = exp(-((norm(Temp))^2)/(2*sigma^2));
matrix1(j,i) = matrix1(i,j);
end
% matrix1 = zeros(315,315);
end


xlswrite('RBFlarge13-s8.xlsx',matrix1)
```

Appendix 4.4 Matlab syntax for Global Kernel K-means Clustering

```
function
[Cluster_elem,Clustering_error]=Weighted_Global_Kernel_K_Means
(K,Dataset_Weights,Total_clusters,Display)


%[Cluster_elem,Clustering_error]=Weighted_Global_Kernel_K_Mean
s(K,Dataset_Weights,Total_clusters,Display)

%Function Inputs
%===============

%K is the kernel matrix of the dataset. It must be a positive
definitesquare matrix (Gram matrix) in order to guarantee
algorithm convergence.If K is not positive definite the
algorithms may still converge though.

%Dataset_Weights is a column vector containing the weight of
each datapoint. It is used to run the weighted version of
Global Kernel K-Means. By setting all weights equal to 1 the
non-weighted version is run. The weights must be positive
numbers.

%Total_clusters is the number of clusters.

%Display when set to 'nutshell' prints information only about
Global Kernel K-Means and when set to 'details' also prints
information about each iteration of Kernel K-Means.
%Any other value results in no printing.

%Function Outputs
%===============

%Cluster_elem is a column vector containing the final
partitioning of the dataset. The clusters are indexed
1,...,Total_clusters.

%Clustering_error is the value of the Kernel K-Means objective
function corresponding to the final partitioning.

%Dataset size.
Data_num=size(K,1);

%Store the optimal clustering error when searching for
1,2,...,Total_clusters.
Best_error=inf(1,Total_clusters);
```

```
%Store the assignment of points to clusters corresponding to
the optimal solution for 1,2,...,Total_clusters.
Best_clusters=ones(Data_num,Total_clusters);

%Find one cluster solution.
[Best_clusters(:,1),Best_error(1)]=Weighted_Kernel_K_Means(Bes
t_clusters(:,1),K,Dataset_Weights,1,Display);

%Find 2,...,Total_Clusters solutions.
for m=2:Total_clusters

%Solve the m clustering problem by trying all datapoints as
possible initializations for the m-th cluster.
for n=1:Data_num

%Place the m-th cluster initially at point n.
%The other clusters are initiallized using the solution to the
m-1 clustering problem.
Cluster_elem=Best_clusters(:,m-1);
Cluster_elem(n)=m;

Ifstrcmp(Display,'details') || strcmp(Display,'nutshell')
fprintf('\n\nSearching for %d clusters.Intitially placing %dth
cluster at point %d',m,m,n);
end

%Find the solution with m clusters.

[Cluster_elem,Clustering_error]=Weighted_Kernel_K_Means(Cluste
r_elem,K,Dataset_Weights,m,Display);

ifstrcmp(Display,'details') || strcmp(Display,'nutshell')
fprintf('\nFinal Clustering error=%g\n',Clustering_error);
end

%Keep the best solution with m clusters i.e. the one with
lowest clustering error.
ifBest_error(m)>Clustering_error
Best_error(m)=Clustering_error;
Best_clusters(:,m)=Cluster_elem;
end
end

if size(unique(Best_clusters(:,m)),1)<m
error('Not able to find more than %d clusters\n',m-1);
end
end

%Keep as final solution of Global Kernel K-Means the one with
the lowest clustering error (usually its the one with
Total_clusters).
[Clustering_error,Clusters]=min(Best_error);
```

```
Cluster_elem=Best_clusters(:,Clusters);

ifstrcmp(Display,'details') || strcmp(Display,'nutshell')
fprintf('+++++++++++++++++++++++++++++++++++++++++++++++++++\
n');
fprintf('Best fit:%d clusters with Clustering
Error=%g\n',Clusters,Clustering_error);
fprintf('+++++++++++++++++++++++++++++++++++++++++++++++++++\
n');
end
```

(Tzortzis&Likas, 2009)

Appendix 4.5

Matlab Syntax for kernel matrix variance calculation

```
a=xlsread('kernelmatrixtrial14.xls');

V11=var(a(:));
```

Appendix 4.6

MatlabSyntax for kernel distance calculation

```
clearall;

a=xlsread('RBFlarge13-s13.xlsx');
[n,m]=size(a);
ForI =1:n
for j=1:m
b(i,j)=2*(1-a(i,j));
% b(i,j)=sqrt(2*(1-a(i,j)));
end
end
xlswrite('distM.xlsx',b);
```

Appendix 4.7

MatlabSyntax for Kernel –based Dunn's Index

```
function DI=dunns(clusters_number,distM,ind)
%Dunn's index for clustering compactness and separation
measurement
% dunns(clusters_number,distM,ind)
% clusters_number = Number of clusters
% distM = Dissimilarity matrix
% ind   = Indexes for each data point aka cluster to which
each data point belongs
I =clusters_number;
Denominator=[];

for i2=1:i
indi=find(ind==i2);
indj=find(ind~=i2);
    x=indi;
    y=indj;
temp=distM(x,y);
denominator=[denominator;temp(:)];
end

num=min(min(denominator));
neg_obs=zeros(size(distM,1),size(distM,2));

for ix=1:i
indxs=find(ind==ix);
neg_obs(indxs,indxs)=1;
end

dem = neg_obs.*distM;
dem = max(max(dem));

DI =num/dem;
end
```

(Ramos, 2012)

Appendix 4.8

Questionnaire survey

This Questionnaire Survey is to get the feedback of the sample pant that you have worn. The sample is made to a newly developed size chart for female pants. Please answer the questions below.

1.What is the waist size of your pant that you normally buy?................

3 Do you satisfy with the fit of pants you buy? Yes / No

4 If No, please mention the problems of the pant you buy.

....

....

....

Please mark your level of satisfaction with fit of the sample pant you wear in following areas using the scale mentioned below.

1 - Strongly disagree 2 – Disagree

3- Neither agree nor disagree

4 – agree 5 – Strongly agree

1. Pant fit comfortably at waist..............
2. Waist line lies at the natural waist line when I sit down...........
3. Waist line lies at the natural waist line when I stand................
4. Crotch is loose so that buttock and thigh are not defined................
5. Pant is easy to get on over my feet and legs....................
6. Pant is loose through hips and legs...............
7. Pant is easy to get on over my hips..................
8. Pant length is satisfactory.............................
9. Overall fit of the pant is satisfactory....................

Other comments..

Appendix 6.1 Images of live fit assessment session

Small WHR- Regular height category

Fig 1 Front View Back View Side View

	waist	hip	inseam
Actual measurement	63	90	77
signed measurement	62	90	74

Small WHR- Tall height category

Fig 2 Front View Back View Side View

	waist	hip	inseam
Actual measurement	64.5	95	80.5
Assigned measurement	66	94	80

Small WHR- Tall height category

Fig 3 Front View Back View Side View

	waist	hip	inseam
Actual measurement	72	103	79.5
Assigned measurement	70	102	80

Medium WHR- Regular height category

Medium WHR- Short height category

Fig 4 Front View Back View Side View

	waist	hip	inseam
Actual measurement	71	94.5	70
Assigned measurement	70	94	68

Medium WHR- Regular height category

Fig 5 Front View Back View Side View

	waist	hip	inseam
Actual measurement	70.5	92	77
Assigned measurement	70	94	74

Fig6 Front View Back View Side View

	waist	hip	inseam
Actual measurement	65	83.5	76
Assigned measurement	66	86	74

Medium WHR- Tall height category

Fig 7 Front View Back View Side View

	waist	hip	inseam
Actual measurement	74	98.5	82
Assigned measurement	74	98	80

Medium WHR- Tall height category

Fig.8 Front View Back View Side View

	waist	hip	inseam
Actual measurement	63	84.5	78
Assigned measurement	62	86	80

Medium WHR- Tall height category

Fig 9 Front View Back View Side View

	waist	hip	inseam
Actual measurement	72.5	97	78
Assigned measurement	74	98	80

Large WHR- Short height category

Fig 10 Front View Back View Side View

	waist	hip	inseam
Actual measurement	87	100	69.5
Assigned measurement	88	104	68

Large WHR- Regular height category

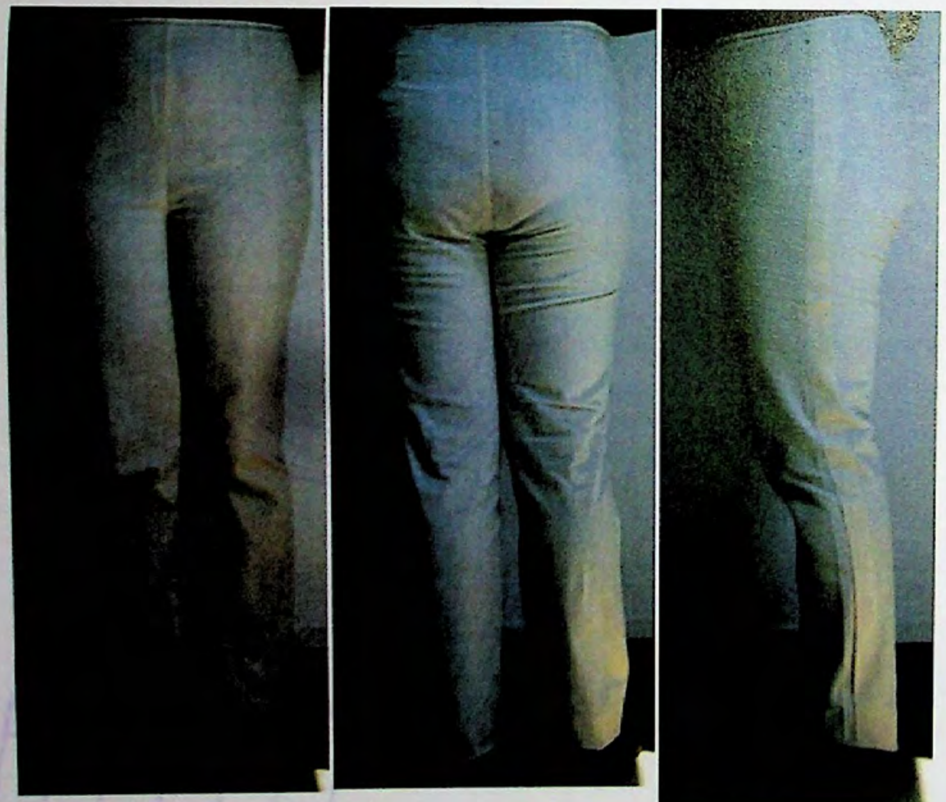


Fig 11 Front View Back View Side View

	waist	hip	inseam
Actual measurement	79	97	72.5
Assigned measurement	78	98	74

Large WHR- Tall height category

Fig 12 Front View Back View Side View

	waist	hip	inseam
Actual measurement	73	87	79
Assigned measurement	74	86	80