SUDAN UNIVERSITY OF SCIENCE AND TECHNOLOGY

COLLEGE OF GRADUATE STUDIES

Application of Data Mining Techniques for Establishing Standard Sizing System for Sudanese Army Officers Garments

تطبيق تقنيات تنقيب البيانات في تأسيس نظام قياس معياري لملبوسات ضباط القوات المسلحة السودانية

A PhD. THESIS SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN TEXTILE ENGINEERING

By:

Elawad Fathelrahman Elfaki Elnour

Supervisor:

Prof. Dr. Salah Eldin Abdel Latief Mohamed

Co – Supervisor:

Dr. Mohamed Elhafiz Mustafa Musa

July – 2015

Approval Page

Name of Candidate: Elanad Fathelrahman Elfaki

Thesis title: Application of Data Mining Techniques for Establishing standard System for Sudanese Army officer Government

Approved by:

1. External Examiner

Name: Prof. Fadlelmoula Abdalla Idris

Signature: F-moulaf Date: 28/7/2015

2. Internal Examiner

Name: Dr. Ahmed Ibrahim Ahmed

Signature: Date: 28/7/15

3. Supervisor

Name: Mohamed Elhafiz Mustafa Muse

Signature: Date: 28/7/15

Sudan University of Science and Technology
College of Graduate Studies

Declaration

I, the signing here-under, declare that I'm the sole author of the Ph.D. thesis entitled Application of Data Mining Techniques for Establishing Standard Sizing System for Sudanese Army Officers Garments

which is an original intellectual work. Willingly, I assign the copy-right of this work to the College of Graduate Studies (CGS), Sudan University of Science & Technology (SUST). Accordingly, SUST has all the rights to publish this work for scientific purposes.

Candidate's name: ELawad Fathelrahman ELfaki ELnour

Candidate's signature: .... Date: 12/8/2015

إقرار

أنا الموقع أدناه أُقر بأننى المؤلف الوحيد لرسالة الدكتوراه المعنونة تطبيق تقنيات تنقيب البيانات في تأسيس نظام قياس معياري لملبوسات ضباط القوات المسلحه السودانيه

وهى منتج فكري أصيل . وباختياري أعطى حقوق طبع ونشر هذا العمل لكلية الدراسات العليا جامعه السودان للعلوم والتكنولوجيا ،عليه يحق للجامعه نشر هذا العمل للأغراض العلمية .

اسم الدارس : العوض فتح الرحمن الفكي النور

توقيع الدارس : .... التاريخ : ٢٠١٥/٨/١٢

استهلال

قال تعالى:

{فَتَعالَى الله الْملك الْحق ولا تَعجلْ بِالْقُرآنِ
مِن قَبل أَن يقْضى إِلَيك وحيه وقُل رب زِدني علْماً}

صدق ا العظيم

سورة طه الآية (114)

Dedication

This thesis is dedicated

To
The soul of my loving mother,
The soul of my father,
My wife Mona, my son engineer Zaid, my daughter Mnasik,
My brothers, sisters, and extended family,

My teachers,
My colleagues, and friends,
My students and
Everyone who gave me his support and advice to this research directly or indirectly

Acknowledgement

Thanks and gratefulness for Allah the most gracious, the most merciful. I would like to express my sincere appreciation to my chief supervisor Prof. Dr. Salah Eldin Abdel lateef Mohamed for his wonderful support, excellent supervision and an endless guidance for helping me to take the first step in this research and during the entire period of my PhD studies and writing of this thesis. I would also like to extend my sincere gratitude to my co-supervisor Dr. Mohammed Elhafiz M. Musa for his professional assistance and remarkable insights, advice and encouragement for my research work. Special thanks to Miss Nariman S. A/Elkreem for her valuable assistance in applying the WEKA program to extract and analyze knowledge from database. Thanks also to Dr. Eltyeb O. Ahmed for his assistance in applying the SPSS program. The most special thanks go to my family for providing the moral support in this challenging period and especially to my wife Mona for her patient through the long periods I spent studying. Thanks go also to Elsir H. Ibrahim clothing design department for his help. Thanks also extended to my colleagues at the Department of Textile Engineering, College of Engineering and Industries technology for their fruitful discussion and support. Many people have contributed constructively to this work. I would like to thank them all. I owe special thanks to Sur Military Clothing factory staff especially Dr. Asim O. Mahgoub for providing the entire basic database. Apperceptions go to Dr. Ishag A. Ibrahim, Mrs Wisal A Ahmed and everybody who contributed to this research directly or indirectly. Last but not least, I thank the Sudan government represented in the Sudan University of Science and Technology for the partial finance for this research.

ABSTRACT

The purpose of this research was to use data mining techniques in establishing standard sizing system for Sudanese army officers garment based on an anthropometric body measurements variables. It also looked into the role of classification of the whole body types (upper and lower), due to its importance for garments manufacturers. From Sudanese army officers' database, a data was selected for establishing sizing system for producing army officers uniform (Poshirt) and to propose a new standard sizing system. The anthropometric data was collected from Sur military clothing factory in Sudan. The data set collected for 841 army officers was used .For each individual (13) anthropometric variables were involved. In this research data mining methods (WEKA and SPSS) were used for clustering and establishing sizing system. The K-means algorithm was implemented to determine the final cluster classification in two stages in the first stage, the (WEKA) clustering method was applied on the original database to categories the size figures. The (SPSS) method, as quantitative approach, was used in the second stage in order to generate descriptive statistics for the raw data following a step by step procedure. Cluster analysis using chest and waist as a control anthropometric variables revealed eight distinct clusters namely; XS, S, M, L, XL, XXL, XXXL and XXXXL respectively. Size codes and upper and lower size limits are generated. The new proposed sizing system SUD profile was compared with profile of three national standards sizing charts USA, EUR, and SUR. The results revealed that the proposed sizing system follow approximately the same profile as the three national standards sizes charts chosen.

المستخلص

تهدف هذه الدراسة لإستعمال تقنيات تنقيب البيانات فى تأسيس نظام قيـاس معيـاري لصناعة ملبوسات ضباط القوات المسلحة السودانيه يقوم على قياس المتغيرات لجسم الانـسان مايعرف بالانثروبومتريه (Anthropometric). كما ينظر ايضا الى أهميه قواعـد تـصنيف أنواع الأجسام وتأسيس نظام للقياس المعياري لزي كامل يغطي اعلى و اسفل الجـسم وذلـك بالنظر لأهميةهذا الأمر لمصنعى الملابس. تم اختيار بيانات لـضباط مـن القـوات المـسلحه السودانيه وذلك لتأسيس نظام معيارى لانتاج الزى الخاص بهذه القوات يعرف ب البوشـيرت (Poshirt)يتكون من قميص و بنطلون وتم إستعمالها لتأسيس المقترح الجديـد لنظـام القيـاس المعياري. جمعت البيانات الانثروبومتريه من مصنع سور للملبوسات العـسكريه بـالخرطوم بحرى – جمهورية السودان, لعدد ثمانمائه و واحد واربعون ضابطا (841), تم اخذ ثلاثه عشرة (13) متغيرا للبيانات الانثروبومتريه لكل فرد من الضباط. فى هـذه الدراسـة, تـم اسـتخدام طريقتين من طرق تنقيب البيانات وهما الويكـا (WEKA) والبرنـامج الإحـصائي للعلـوم الاجتماعية (SPSS) وذلك لتجميع وتأسيس نظام القيـاس المعيـارى المقتـرح. تـم تطبيـق خوارزميه (K-means) لتحديد تجميع التصنيف النهائي للمجموعة وذلـك علـى مـرحلتين . المرحله الاولى تم تطبيق تقنيه الويكا على البيانات الأصليه بغرض تقسيم مجموعـات القيـاس المختلفة .أما في المرحله الثانيه تم إستخدام البرنامج الإحصائي للعلـوم الاجتماعيـه كمـدخل للحصول علي بيانات إحصائيه وصفيه وذلك باتباع طريقه خطـوة بخطـوة (Step by Step) وبإستخدام قياسات الصدر و الوسط لضبط المتغيرات الانثروبومتريه, أدى ذلك لظهور ثمانيـة مجموعات مميزة للقياسات وهى فائق الصغر, صغير, وسط, كبير, فائق الكبـر, فـائق فـائق الكبر, فائق فائق فائق الكبر وفائق فائق فائق فائق الكبر على التوالى.وتم تحديـد رموزالقيـاس وايجادالحدود العلويه والسفليه للقياس. كما وأجريت مقارنه النسق للمقترح الجديد لنظام القيـاس المعياري مع انظمه قياس معياريه وطنيه فى الولايات المتحدة و اوروبا والسودان (مصنع سور للملبوسات العسكريه). وأظهرت النتائج المتحصله ان المقترح الجديد لنظام القيـاس المعيـاري يتبع تقريبا ذات النسق الذي تتبعه انظمه القياس الوطنيه الثلاثه التي تم اختيارها.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Abbreviations	Full Meaning of Abbreviations
D.M	Data Mining
OLAP	Online Analytical Processing
SQL	Structured Query Language
KDD	Knowledge Discovery from Database
D.T	Decision Tree
CART	Classification and Regression Trees
CHAID	Chi Square Automatic Interaction Detection
ID3	Iterative Dichotomizer 3
PCA	Principle Component Analysis
DNA	Deoxyribonucleic Acid
SOM	Self-Organizing Maps
DGA	Dissolved Gas Analysis
CRM	Customer Relationship Management
DSS	Decision Support Systems
CAD	Computer Aided Design
CAM	Computer-Aided Manufacturing
IT	Information Technology
CRPM	Center for research and policy making
k-nn	k-nearst neighber classification
WEKA	Waikato environment of knowledge
SPSS	Statistical package for social sciences
GNU	General Public license
STD	Standard Deviation
ANOVA	Analysis of Variance
BSI	British Standard Institution
SUD	Sudan Sizes Code
SUR	Sur Military Clothing Factory
EUR	EUR Sizes and Measurements for Men
USA	USA Sizes and Measurements for Men
FFIT	Female figure identification technique

CHAPTER ONE

INTRODUCTION

CHAPTER ONE

Introduction

1.1 Background:

In the last twenty years there was an extraordinary expansion of computer accessible data about all kinds of human activities. The availability of these large volumes of data, and our limited capabilities to process them effectively, creates a strong need for new methodologies for extracting useful, task – oriented methodologies for deriving plausible knowledge from small and directly relevant data. However, in many practical areas only limited data may be available, e.g., fraud detection, terrorism presentation computer intrusion detection, early cancer diagnosis.

In order to automatically generate useful knowledge from a variety of data, and presented it in human oriented forms, a powerful tools is strongly needed. Researchers have been exploring ideas and methods in different areas as efforts to satisfy this need. Such areas include; data mining, text mining, machine learning, statistical data analysis, data visualization, and pattern recognition.

According to (P.Lyman et al, 2008) survey the world produces between 1 and 2 Exabyte's of unique information per year, which roughly 250 megabytes for every man, women, and child on earth. Due to the wide range of computer devices nowadays, our ability to produce and collect data has impressively increased. As a result, huge amounts of data are collected from different application domains like business, science, telecommunication and health care systems. Also, the World Wide Web overwhelms us with information. A part from their huge quantity, modern data is also characterized by low level of abstraction, high degree of diversity and complexity. Furthermore, data are not only produced in a centralized but also in a distributed way, which imposes new challenges regarding their management. Due to these reasons, it is impossible for humans to thoroughly investigate these data collections through a manual process. Therefore, the compulsory need for data mining emerged.

1.2 Towards a Definition of Data Mining:

To understand the term 'data mining' it is useful to look at the literal translation of the word: to mine means to extract. The verb usually refers to mining operations that extract from the earth its hidden, precious recourses. The link of this word with data suggests an – in depth search to find additional information which previously were un-noticed in the mass of data available. From the viewpoint of scientific research, data mining is relatively new discipline that has developed mainly from studies carried out in other disciplines such as computing, marketing and statistics.

The term data mining came into popular use in late sixties of the last century. It is currently used in a variety of contexts. It has mostly been used by statisticians, data analysts and the management information systems (MIS) communities. It has also gained popularity in the database field. Since the concept of data mining was born changes in the definition of data mining have occurred reflecting a changing emphasis in the field of database. At present there appears to be numerous definitions which have achieved at least a modicum of acceptance. Various definitions of data mining are reviewed and it is apparent from the review that there is some ambiguity as to the definition of data mining.

Listed below are some famous definitions of data mining (DM):

(1) Data mining (DM) is, "the nontrivial extraction of information that resides implicitly in the data. This information was previously ignored or unknown and may be useful for some process. In other words, prepares data mining, data probes and explores to remove the hidden information in them".

(2) "Data mining is the process of selection, exploration, and modeling of large quantities of data to discover regularities or relations that are at first unknown with the aim of obtaining clear and useful results for the owner of the database.

(3) Data mining is not just about the use of a computer algorithm or a statistical techniques; "it is a process of business intelligence that can be used together with what is provided by information technology to support company decisions".

(4) "Data mining helps to find trends in data that human would never dream of looking for".

(5) "Data mining is the use of an appropriate set of technologies to exploit patterns of information from massive customer focused databases. However, data mining is not a single technology. Rather, it is a collection of tools that are used to extract information from data. Data mining is not just a technology but it is also a process. It cannot be fully automated as individuals must be active in the process to be sure that the information that is extracted is accurate".

(6) Fayyad et al, (1996) define, data mining as, "the process of extracting valid, previously unknown comprehensible information from large databases in order to improve and optimized business decisions. This definition seems to me more complete definition of data mining.

1.3 Definition of Some Terms Related to Data Mining:

1.3.1 Data:

Data are any facts, numbers, or text that can be processed by a computer. Nowadays, organizations are accumulating vast and growing amounts of data in different formats and different databases. This includes:

a) Operational or transactional data such as, sales, inventory, payroll, and accounting.

b) No operational data, such as industry sales, forecast data, and macro economic data.

c) Meta data – data about the data itself, such as logical database design or data dictionary definitions.

1.3.2 Information:

We are in an age often referred to as the information age, because information leads to power and success, and thanks to sophisticated technology such as computers, satellites internal, etc. We have been collecting a myriad of data from simple numerical measurements and text documents to more complex information such as spatial data, multimedia, channels, and hypertext document. The patterns, association, or relationship among all this data can provide information. For example, analysis of retail point of sale transaction data can yield information on which products were sold and when.

1.3.3 Knowledge:

Across a wide variety of frills, data are being collected and accumulated at a dramatic pace. There is an urgent need for a new generation of computational theories and tools to assist humans extracting. Useful information (knowledge) forms the rapidly grouping volumes of digital data. Information can be converted into knowledge about historical patterns and future trends. For example, summary information on retail supermarket sales can be analyzed in high of promotional efforts to provide knowledge of consumer buying behavior. Thus, a manufacturer or retailer could determine which items are most susceptible to promotional efforts.

1.3.4 Data warehouses:

A related field evolving from databases is data warehousing, which refers to the popular business trend of collecting and cleaning transactional data to make them available for online analysis and decision support. Data warehousing is defined as a process of centralized data management and retrieval.

Data warehousing represents an ideal vision of maintaining a central repository of all organizational data. Centralization of data is needed to maximize user access and analysis. Dramatic technological advances are making this vision a reality for many companies. Furthermore dramatic advances in data analysis software allow users to access this data freely.

1.3.5 Data cleaning:

As organizations are forced to think about a unified logical view of the wide variety of data and databases processed, the issues of mapping data addressed to a single naming convention, uniformly representing and handling missing data, noise and errors when possible.

1.3.6 Data access:

Uniform and well – defined methods must be created for accessing the data and providing access paths to data that were historically difficult to get to (for example, stored offline).

1.3.7 Online Analytical Processing (OLAP):

A popular approach proposed for analysis of data warehouses is called online analytical processing (OLAP), named for a set of principle (Codd, 1993). The (OLAP) Tools focus on providing multidimensional data analysis, which is superior to structured query language (SQL) in computing summaries and breakdowns along many dimensions. OLAP tools are targeted toward simplifying and supporting interactive data analysis.

1.4 Data Mining Verse Knowledge Discovery:

Data mining, uses of algorithms in order to extract the information and patterns derived by the knowledge discovery in database (KDD) process. On the other hand, (Fayyad et al, 1996) stated that the term knowledge discovery in databases (KDD) is the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data. Here, data are a set of facts (for example, cases in a database), and pattern is an expression in some language describing a subset of the data or a model applicable to the subset. Hence, in our usage here extracting a pattern also designates fitting a model to data the term process implies that KDD comprises many steps, that involve data preparation, search for patterns, knowledge evaluation, and refinement, all repeated in multiple iterations. By Nontrivial, means that some search or inference is involved; that is, it is not a straight forward computation of predefined quantities like computing the average value for a set of numbers.

The discovered patterns should be valid on new data with some degree of certainty. Patterns should also be novel and potentially useful (at least to the system and preferably to the user) leading to some benefit to the user or task. Furthermore, the patterns should be understandable, if not immediately then after some post processing.

The distinction between the (KDD) process and the data – mining step (within the process) is that, (KDD) refers to the overall process of discovering useful knowledge from data. On the other hand data mining (DM) is the application of specific algorithms for extracting patterns from data.

The knowledge discovery in databases process comprises a few steps leading from raw data collections to some form of new knowledge. The iterative process consists of the following steps.

1) Data Cleaning: also known as data cleansing, is a phase in which noisy and irrelevant data are removed from the collection.
2) Data Integration: at this stage, multiple data sources, often heterogeneous, may be combined in a common source.
3) Data Selection: at this step, the data relevant to the analysis is decided on and retrieved from the data collection.
4) Data transformation: also known as data consolidation, it is a phase in which the selected data is transformed into forms appropriate for the mining procedure.
5) Data mining: it is the crucial step in which the clever techniques are applied to extract patterns potentially useful. Therefore, (D.M) is considered by some researchers to be just one step in a large process known as knowledge discovery in database (KDD).
6) Pattern evaluation: in this step, strictly interesting patterns representing knowledge are identified based on given measures.
7) Knowledge representation: is the final phase in which the discovered knowledge is visually presented to the user.

8) Visualization: the knowledge discovered is visually presented to the user by visualization techniques in order to help users to understand and interpret the data mining results.

The KDD is an iterative process. Once the discovered knowledge is presented to the user, the evaluation measures can be enhanced, the mining can be further refined, new data can be selected or further transformed , or new sources can be integrated, in order to get different more appropriate results. Figure (1.1) shows the additional steps in the (KDD) process.

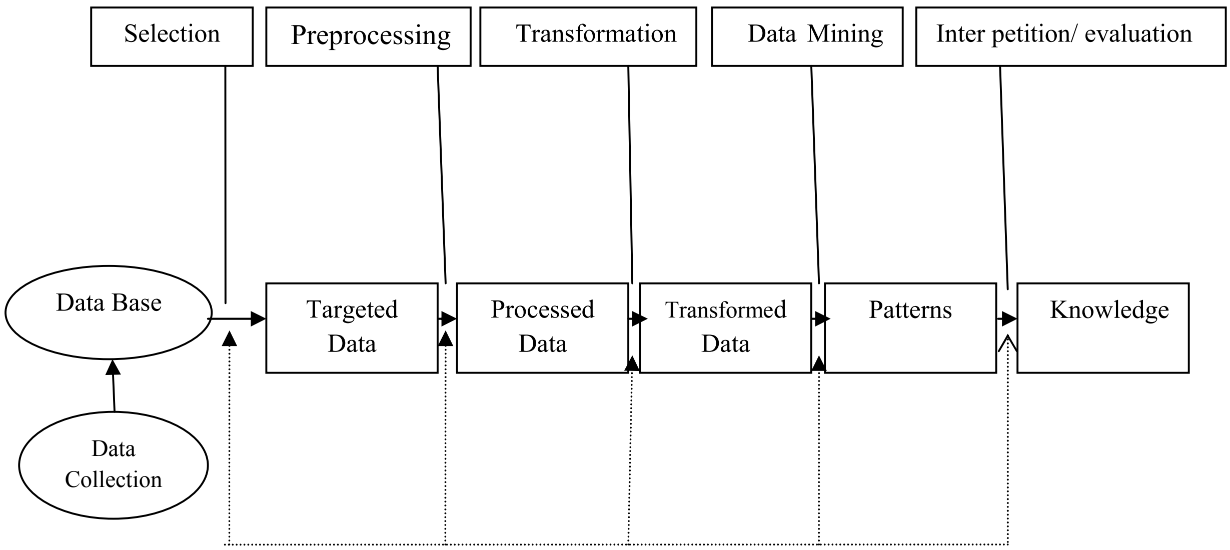


Figure (1.1): Steps of Knowledge Discovery on Database (KDD)

1.5 Data Mining Model and Tasks:

The kinds of pattern that can be discovered depend upon the data mining tasks employed. The two types of data mining tasks are:

1) Descriptive data mining tasks that describe the general properties of the existing data. Descriptive data mining describes a dataset in concise way and presents interesting characteristics of the data without having any predefined target.
2) Predictive data mining tasks that attempt to do predictions based on inference on available data. The term predictive data mining is usually applied to identify data mining projects with the goal to identify a statistical or neural network model or set of models that can be used to predict some response of interest. A predictive model is made up of a number of predictors, which are variable factors that are likely to influence future behavior or results. Figure (1.2) shows these two types of data mining tasks.

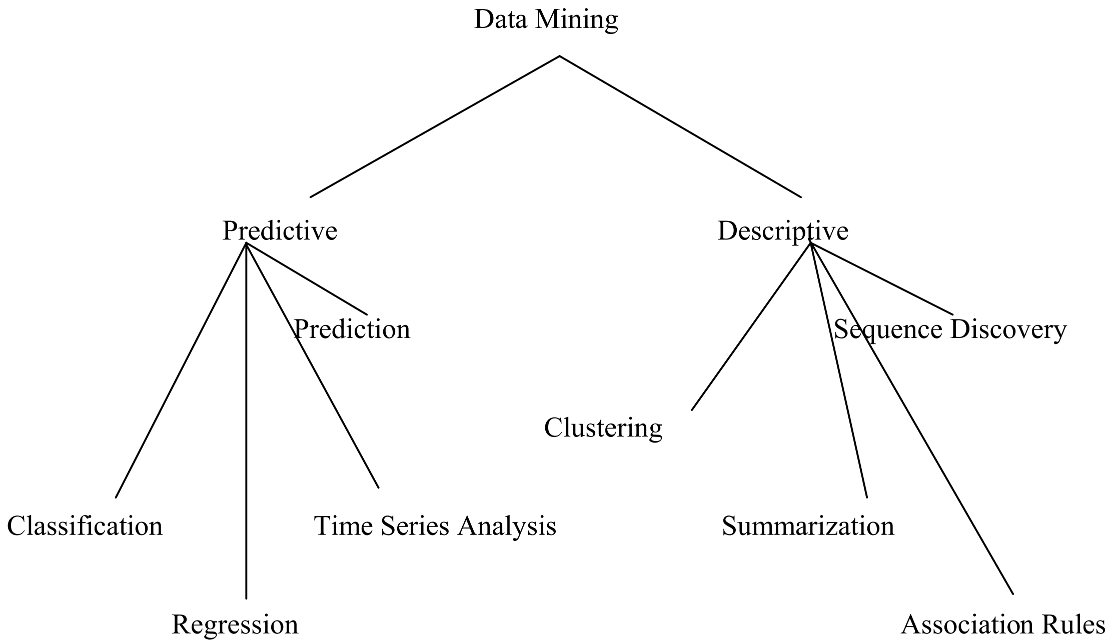


Figure (1.2): Data Mining Model and Tasks

1.6 Data Mining Functionalities:

As shown in figure (1.3), the data mining functionalities and the variety of knowledge they discover can be presented rather briefly as follows:

1) Characterization: Data characterization is a summarization of general features of objects in a target class, and produces what is called characteristic rules. The data relevant to a user – specified class are normally retrieved by a data base query and run through a summarization module to extract the essence of the data at different levels of abstractions. It is noted that with a data cube containing summarization of data, simple "OLAP" operations fit the purpose of data characterization.
2) Discrimination: Data discrimination produces what are called discriminate rules and is basically the comparison of the general features of objects between two classes referred to as the target class and the contrasting class. The techniques used for data discrimination are very similar to those used for data characterization with the exception that data discrimination results include comparative measures.
3) Association Analysis: Association analysis is the discovery of what are commonly called association rules. It studies the frequency of items occurring together in transactional databases, and based on a threshold called support, identifies the frequent item tests. Another threshold, confidence, which is the conditional probability than an item appears in a transaction when another item appears, is used to pinpoint association rules. Association analysis is commonly used for market basket analysis. The discovered association rules are of the form:

P→Q [s, c], where P and Q are conjunctions of attribute value ─ pairs, and (for support) is the probability that P and Q appear together in a transaction, and c (for confidence) is the conditional Probability that Q appears in a transaction when P is present. Association, patterns where one event is connected to another event such as purchasing a pencil, purchasing a paper.

4) Classification: classification (Supervised classification) analysis is the organization of data in given classes. The classification uses given class labels in order to classify the objects in the data collection. Classification approaches normally use a training set where all objects are already associated with known class labels. The classification algorithm learns from the training set and builds a model. The model is used to classify new objects. For example, wafer – cleaning processes (that include parameters set at various levels) may be classified according to the quality of the cleaning process outcome; thus the outcome of new cleaning processes can be identified. The classification analysis generates a model that could be used to either accept or reject credit request in the future. Classification leads to identification of new patterns such as coincidences between duct tape purchases and plastic sheeting purchases.

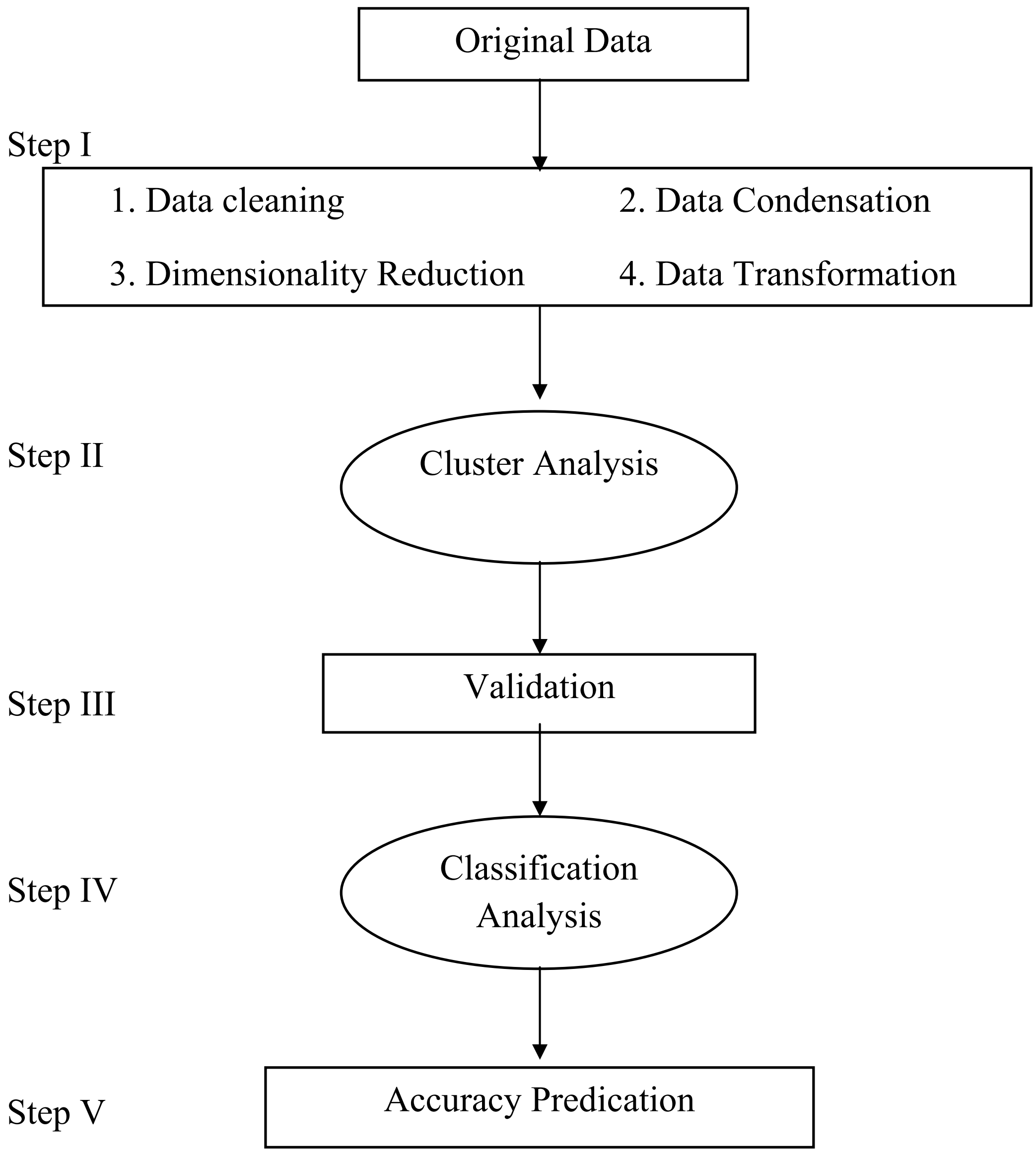


Figure (1.3): Data Mining Functionalities

5) Clustering: it focuses on partitioning a set of data items with similar characteristics together. Clustering is similar to classification, but, in clustering, class labels are unknown and it is up to the clustering algorithm to discover acceptable classes. Clustering is also called unsupervised classification, because the classification is not dictated by given class labels. There are many clustering approaches all based on the principle of maximizing the similarity between object in a same class (intra – class similarity) and (clustering finding and visually documenting groups of previously unknown facts, such as geographical location and brand preferences) minimizing the similarity between objects of different classes (inters – class similarity).
6) Prediction: Prediction has attracted considerable attention given the potential implications of successful forecasting in a business context. The first type of perdition can either try to predict some unavailable data values or pending trends, or predict a class exploration and modeling of large quantities of data to discover regularities or relations that are at first unknown with the aim of obtaining clear and useful results for the owner of the database" label for some data. The second is tied to classification. Therefore, a classification model is built based on the attribute value of the object and the attribute values of the classes. Prediction is however more often referred to the forecasting of missing numerical values or increase/ decrease trends in time related data. The major idea is to use large number of past values to consider probable future values. For example the prediction that, "people who join an athletic club may take exercise classes".
7) Outlier Analysis: Outliers are data elements that cannot grouped in a given class or cluster. Also known as exception or surprises, they are

often very important to identify. While outliers can be considered noise and discarded in some applications, they can reveal important knowledge in other domains, and thus can be very significant and their analysis is valuable.

8) Sequential Pattern Analysis: is concerned with the discovery of causal relationship among temporally oriented events or pattern where one event leads to another event, such as the birth of a child and purchasing diapers.

9) Model Visualization: it focuses on the decision makers attempt to convey the discovered knowledge in an understandable manner. Some examples are histograms, scatter plots, and outlier's identification.

10) Evaluation analysis: it pertains to the study of time related data those changes in time. Evaluation analysis models evolutionary trends in data, which consent to characterizing, comparing, classifying or clustering of time related data.

11) Deviation analysis: it considers differences between measured values and expected values, and attempts to find the cause of deviations from the anticipated values. Deviation analysis is used to detect deviation over time, deviation from the mean, and deviation between and observed value and a reference value as applied, for instance in quality control.

1.7 Data Mining and Computer Science:

Data mining in some sense flow from a conjunction of both computer science and statistical frameworks. The development of database theory and the structured query language (SQL) among information systems specialists allowed for logical queries to database and subsequent exploitation of knowledge in the database. However, being limited to logical queries, and the desire to exploit more numerical and even statistically oriented queries led to the early development of data mining.

The exploitation of large scale commercial databases was a natural application of supercomputer technology since early nineties of the last century. So there was both an academic pull and a commercial push to develop data mining in the context of computer science. The analysis of large to massive data set sizes that data mining deal with caused that analysis has to be automated, so that interactive approaches and approaches that exploit very complex algorithms are prohibited in a data mining frame work.

Specific data mining techniques consist of existing statistical models for rational statistical analysis such as: clustering, linear regression, neural networks, Bayesian networks, visualization, and tree based models decision trees (DT). Many of the techniques that are applied to data mine are from statistical methods that have been around for years. However, the main difference in today's data mining technology as compared to traditional methods is that a high speed computer is implemented to determine trends and relationships of large amount of data in a very short time. Without computer and optimized software can old methods of statistical analysis of the same data may take a very long time to come up with the same result, if they ever find the same relationship.

1.8 Methods and Algorithms Used in the Field of Data Mining (DM)

Data mining methods resemble the core of the data mining process, and can have different goals depending on the intended outcome of the overall data mining process. Berry et al, (1997) reported that a variety of methods are available to enable these goals. Different data mining methods serve different purposes, each offering its own advantages and disadvantages. The most commonly used methods of data mining come from artificial intelligence and statistics. These methods are algorithms that are more or less complex than when applied to data set to achieve results. Data mining methods can be classified into two categories, conventional and modern methods.

Data mining methods are allowed to sort through the data to research information to find easily predictive trends and relationships, while at the same time they are making a sweep through the data to find hidden patterns all in the same step. This data mining techniques find hidden trends with minimal time and effort.

1.8.1 Conventional methods:

These methods include, primarily, conventional statistical methods such as regression analysis, numerical taxonomy, and multidimensional scale.

1.8.2 Statistical methods:

Statistical methods governed most of the applications that are used data mine for many years compared with the other types of conventional methods. The introduction of the high – speed computer is the main difference in today's data mining techniques as opposed to conventional methods of statistics. The relationships and trends in large amount of data obtained when high – speed computer is used in a minimal time and effort. In order to come up with the same data answer, if they ever find the same relationship using traditional statistical analysis may take a very long time, compared with the application of computer and optimized software.

The statistical approach to data mining tends to be the most widely used basis for practical data mining applications given the typical presence of uncertainty in real – world data cementing processes. Statistical model is a symbolic expression in the form of equality or equation that is used in all experimental design and the regression to indicate the various factors that modify the response variable.

1.8.3 Numerical taxonomy:

Taxonomy is the practice and science of classification. Originally taxonomy only referred to the classifying of organisms or a particular classification of organisms. In a wider more general sense, it may refer to a classification of things or concepts, as well as to the principles underlying such a classification. Mathematically, a hierarchical taxonomy is a tree structure of classifications for a given set of objects.

1.8.4 Limitations of conventional method:

While it is useful for many applications, these techniques have inherent limitations. For example, a statistical analysis can determine distributions, covariance's and correlations among variables in data, but it is not able to characterize these dependencies at an abstract, conceptual level as humans can, and produce a causal explanation why these dependencies exist. While a statistical data analysis can determine the central tendencies and variance of given factors, it cannot produce a qualitative description of the regularities, nor can it determine a dependence on factors not explicitly provided in the data.

On the other hand, a numerical technique can create a classification of entities, and specify a numerical similarity among the entities assembled into the same or different categories, but it cannot alone build qualitative descriptions of the classes created and present a conceptual justification for including entities into a given category.

1.8.5 Modern methods:

This includes methods such as, classification and regression trees, association rules, and Bayesian nets. Popular classification (or decision) trees or forests can represent a relationship between the input and output variables, but their representation power is very modest. These methods may thus produce a very complex tree even for a conceptually simple relationship. Similarly, association rules, which are popular in data mining, have a limited representation power. Bayesian nets are very attraction for many applications, but typically rely on human input as to their structure, and can automatically determine only relatively simple interrelationships among attributes or concepts.

The above methods typically create patterns that use only attributes that are present in the data. They do not by themselves draw upon background domain knowledge in order to automatically generate additional relevant attributes, nor do they determine attributes, changing relevance to different data analysis problems. In cases where the goal is to address such tasks as those listed above, data analysis system has to be equipped with a substantial amount of background knowledge, and be able to conduct symbolic reasoning involving that knowledge and the input data. Some of these modern methods are listed below;

1) Artificial neural networks**.** They are a paradigm of learning and automatic processing inspired by the way does the nervous system of animals. It is a system of interconnected neurons in a network that collaborates to produce stimulus. In other words artificial neural networks are non – linear predictive models that learn through training

and resemble biological neural networks in structure. Some examples of neural networks are:

1) The perceptron.
2) The multivariate perceptron.
3) The self – organizing maps.

2) Decision trees. A decision tree is a predictive model used in the field of artificial intelligence, given a database of these diagrams are constructed logical constructions similar to forecasting systems based on rules that serve to represent and categorize a series of conditions that occur in succession, to solve a problem. Decision tree is a tree – shaped structures that represent sets of decisions. These decisions generate rules for the classification of a dataset. These decisions include classification and Regression Trees (CART) and Chi Square Automatic Interaction Detector (CHAID).

CART and CHAID are decision tree techniques used for classification of a dataset (http, 2011). These trees provide a set of rules that can be applied to a new (unclassified dataset to predict which records will have a given outcome. CART segments a dataset by creating 2–way split while CHAID segments using Chi square tests to create multi– way splits. CART typically requires less data preparation than CHAID.

There are two main types of decision trees used in data mining.

a) Classification tree analysis is when the predicted outcome is the class to which the data belongs.

b) Regression tree analysis is when the predicted outcome can be considered a real number.

The term Classification and Regression Tree (CART) analysis is an umbrella term used to refer to both of the above procedures, first introduced explained that trees used for classification have some similarities – but also some differences, such as the procedure used to determine where to split (http, 2010).

3) The Chi Squared Automatic International Detector (CHAID):

This name drives from the basic algorithm that is used to construct (non – binary) trees, which for classification problems (when the dependent variable is categorical in nature) relies on the Chi– square test to determine the best next split at each step; for regression – type problems (continuous dependent variable) .

The program will actually compute F-tests. The CHAID originally proposed will "build" non – binary trees (i.e. trees where more than two branches can attack to a single root or node).

4) Basic tree – building algorithm of CHAID of exhaustive CHAID:

Specifically the algorithm proceeds as follows:

a) Preparing predictors: The first step is to create categorical predictors out of any continuous predictors by dividing the respective continuous distributions into a number of categories with an approximately equal number of observations. For categorical predictor, the categories (classes) are naturally defined.

b) Merging categories: The next step is to cycle through the predictors to determine for each predictor the pair of (predictor) categories that is least significantly different with respect to the dependent variable; classification problems (where the dependent variable is categorical as well), it will compute a Chi – square test (Person Chi – square); for regression problems (where the dependent variable is continuous), F-tests. If the respective test for a given pair of predictor categories is not statistically significant as defined by an alpha – to – merge value, then it will merge the respective predictor categorists and repeat this step (i.e. respective pair of predictor categories is significant (less than the Respective alpha – to – merge value) it will compute p- value for the set of categories for the respective predictor.

c) Selecting the spilt variable: The next step is to choose the spilt the predictor variable with the smallest adjusted p-value, i.e., the predictor variable that will yield the most significant split; if the smallest adjusted p-value for any predictor is greater than some alpha – to – split value, then no further splits will be performed, and the respective node is a terminal node. Continue this process until no further splits can be performed, (given the alpha – to merge and alpha – to – split value).

1.8.6 CHAID and exhaustive CHAID algorithms:

The basic CHAID algorithm is modified known as Exhaustive CHAID which performs a more through merging and testing of predictor variables, and hence requires more computing time. Specifically, the merging of categories continuous (without reference to any alpha – to – merge value) until only two categorists remain for each predictor. The algorithm then proceeds as described above in the split variable step, and selects among the predictors the one that yields the most significant split. For large datasets, and with many continuous predictor variables, this motivation of the simpler CHAID algorithm requires significant computing time.

A well-known tree-growing algorithm for generating decision trees based on univariate splits is ID3 with an extended version called C4.5. Greedy search methods, which involve growing and pruning decision tree structures, are typically employed in these algorithms to explore the exponential space of possible models.

These trees construction algorithm is differ in several aspects, for instance:

a) C4.5 test are allow two or more outcomes but while classification and regression trees CART tests are always binary.

b) C4.5 prunes trees using a single – pass algorithm derived from binomial confidence limits; CART estimated by cross- validation.

c) C4.5 used information – based criteria; where as CART used the Gini diversity index to rank tests, when some of the case's value is unknown.

d) C4.5 apportions the case probabilistically among the outcomes; but CART looks for surrogate tests that approximate the outcomes.

1.8.7 The k-means algorithm:

It is a simple iterative method to partition a given dataset into user specified number of clusters. The algorithm operates on a set of d-dimensional vectors;

$D = \{ x_i{}^{[} i = 1, \ldots. N]\}$ where x_i € K^d denote the ith data point.

The algorithm is initialized by picking k points in K^d as the initial k cluster representative or "centroids". Despite k- means algorithm drawbacks; it remains the most widely used partitioned clustering algorithm in practice. The algorithm is simple, easily understandable and reasonably scalable, and can be easily modified to deal with streaming data.

1.8.8 kNN: k –nearest neighbor classification:

It is used to find a group of k objects in the training set that are the closest to the test object, and base the assignment of a label on the predominance of a particular class in this neighborhood.

The three key elements of this algorithm are:

a) A set of labeled objects, e.g., a set of stored records.

b) A distance or similarity metric to compute distance between objects.

c) The value of k, the number of nearest neighbors.

To classify unlabeled object , the distance of this object to the labeled objects is computed , its k- nearest neighbors are identified, and the class labels of the nearest neighbors are then used to determine the class label of the object.

1.8.9 ID3 basic algorithm:

The basic idea of ID3 algorithm is to construct the decision tree by employing top- down, greedy search through the given sets to test each attribute at every tree node. Initial definition of ID3 is restricted in dealing with discrete sets of values. It handles symbolic attribute effectively. However that extend it sphere to continuous – valued attributes (numeric attribute) to fit the real world scenario. This is achieved by definition of new discrete valued attributes that partition the continuous-valued attributes into symbolic attribute again .The decision tree learning algorithm ID3 examined by implementing it by using Java programming.

Firstly, ID3 implemented to deal with target function that has discrete output values. Then ID3 extended to deal with real – valued output, such as numeric data and discrete outcome rather than simply Boolean value.

The Java applet provided at least section offers a simulation of decision –tree learning algorithm in various situations.

1.8.10 Genetic algorithms:

There are numerical optimization techniques, in which that variable or variables that are to be optimized together with the study variables constitute a reportable sequent. Those configurations of the variables of analysis to obtain best values for the response variable correspond to the segments with greater reproductive capacity. It can also introduce random elements for the modification of the variables (mutations). It was stated that (http, 2011), after a certain number of iterations, the population will consist of good dilutions to the optimization problem, because the bad solutions have been discarded, iteration after iteration. In (http, 2010) mentioned that genetic algorithms, known as optimization techniques that use process such as genetic combination, mutation, and natural selection in a design based on the concepts of natural evolution.

1.8.11 linear regression:

It is widely used to build relationships between data. Quick and effective but insufficient in multidimensional spaces where they can relate more than two variables.

1.8.12 Data visualization:

It deals with the visual interpretation of complex relationships in multidimensional data. Graphics tools are used to illustrate data relationships.

1.8.13 Rotational forest:

In which every decision tree is trained by first applying principal component analysis (PCA) on a random subset of the input features.

1.8.14 Machine learning method:

In effort to satisfy the growing need for new data analysis tools that can overcome the above limitations of traditional methods, researchers have turned to ideas and methods developed in symbolic machine learning. The field of machine learning is a natural source of ideas for this purpose, because the essence of research in this field is to develop computational models for acquiring knowledge from facts and background knowledge. The above and related efforts led to the emergence of a research area concerned with logical data analysis, and the development of methods for data mining and knowledge discovery.

A natural step in this progression appears to be the development of systems that closely integrate databases with inductive learning and data mining capabilities. Such systems would be able to, for example, automatically all upon a decision rule generator, regression analysis, and conceptual cluster or attribute generation operator depending on the state of data analysis. The machine learning method is connected to computer science and artificial intelligence. It is concerned with finding relations and regularities in data that can be translated into general truths.

The aim of machine learning is the reproduction of the data – generating process, allowing analysts to generalize from the observed data to new, unobserved cases. Rosenblatt (1962) introduced the first machine learning model, called the perception, following on from this; neural networks developed in the second half of the 1980s. In the 1990s statisticians began showing interest in machine learning methods as well, which led to important development in methodology. Machine learning methods were used for elaborates and specific marketing campaigns. The term knowledge discovery in databases (KDD) was coined to describe all those methods that aimed to find relations and regularity among the observed data.

1.8.15 Data mining algorithms requirements:

a) Discover patterns in very large databases, rather than simply verify that a pattern exists;
b) Have a completeness property that guarantees that all patterns of the certain types have been discovered;

c) Have high performance and near – linear scaling on very large (multiple gigabytes) real- life databases.

1.8.16 Thesis Outlines:

This thesis consists of five chapters. Chapter one gives a brief introduction of data mining techniques, definition, tasks, functionalities and data mining process. Chapter 2, presents the literature review and discusses the applications of where data mining has been used are reviewed. The techniques and methods that are present in data mining software packages are discussed. It also reviews the applications of data mining in textiles spinning, weaving, knitting quality control and garment manufacturing and it covers the research problem and objectives. Chapter three shows the material and methods and sheds some light on the methods that were chosen for the analysis and how are run. The process used to mine the data, which will be followed for this research, were also explained. In chapter four the actual effort of mining the data will be shown and then discussed. The analysis of database using the chosen software packages the (WEKA and SPSS) was explained. In this chapter the proposed standard sizing systems introduced. A comparison of the proposed standard sizing system with national standard sizing system for men such as USA, EUR and SUR was explained. Finally conclusions of the research were stated in chapter five, along with recommendations and suggestions for future studies.

CHAPTER TWO

LITERATURE REVIEW

CHAPTER TWO

Literature Review

2.1 Examples to Illustrate the Use of Data Mining:

Data mining is used for a variety of purposes in both the private and public sector in business, science and engineering, and manufacturing etc., these will be discussed rather briefly as follows;

2.1.1 In business:

Data mining can contribute significantly in the application of corporate governance based on the relationship with the customer. The classic example of application of data mining deals with the detection of shopping behavior in supermarkets. A common example is detection of patterns of flight. In many industries, including banking telecommunications, etc. there is an understandable interest in detecting early customers who may be planning to terminate their contracts to possibly switch to the competition. Data mining helps determine which customers are more likely to unsubscribe from studying their behavior patterns and comparing them with samples of customers, who actually dropped out in the past. A similar case is the transaction of money laundering or fraud in the use of credit cards or mobile phone services and even in the list of taxpayers with the tax.

Data mining can also be useful for the departments of human resources to identifying the characteristics of their most successful employees. The information obtained can assist in the recruitment of staff, focusing on the efforts of its employees and the results obtained by them. In addition the assistance offered by strategic management application in an enterprise result in obtaining benefits at the corporate level, such as improving the profit margin or share objectives and improving operational decisions such as development plans of production or labor management. It is also a fashionable area in the analysis of visitors behavior, especially when potential customers on a web-Internet. In http (2011) declare that, data mining has been cited as methods by which the unit able Danger U.S. Army had identified the leader of the terrorist attacks of September 11, 2001, Mohammed Atta, and three other hijackers "11-S" as potential members of a cell of Al Qaeda operating in the U.S. more than a year before the attack.

2.1.2 In Science and engineering:

In recent years, data mining is being used extensively in various areas related to science and engineering. Some examples of these applications in these fields are:

a) Genetics: When studying of genetic health, the main objectives is to understand the relationship mapping between the parties and individual variation in the sequences of Deoxyribonucleic Acid(DNA) and human variability in susceptibility to disease. In plainer terms, the question is how changes in DNA sequence affect an individual's risk of developing common diseases (such as cancer). This is very important because it helps to improve the diagnosis, prevention, and treatment of disease. The data mining used to perform this task is known as, "multifactor dimensionality reduction".

b) Electrical engineering: In the field of electrical engineering, data mining techniques have widely used to monitor the conditions of the facilities of high voltage. The purpose of this monitoring is to obtain valuable information about the state of the isolation of the equipment. To monitor the vibration or the analysis of load changes in transformers used techniques to data grouping (clustering), such as: Self – Organizing Maps (SOM).These maps are used to detect abnormal conditions and to estimate the nature of these anomalies.

c) Gas analysis: Data mining techniques were also used for dissolved gas analysis (DGA) in electrical transformers. The self- Organizing Maps (SOM) was used to analyze data and identify trends that might be overlooked using conventional techniques (DGA).

d) In manufacturing: The first applications of artificial intelligence in engineering in general and in manufacturing in particular were developed in the late 1980s. Kusiak (2006) presented a comprehensive overview of data mining application in manufacturing especially in the areas of production processes, control, maintenance, Customer Relationship Management (CRM), decision support systems (DSS), quality improvement, fault detection, and engineering design, more details about the application of D.M in manufacturing will be considered later.

e) Application of data mining in spinning manufacturing: In the spinning manufacturing industry, most of the articles that have been written shows that data mining being used for forecasting, production planning, promotion and distribution, However, for the most parts this was only done at the retail level. Very few articles have been found at the manufacturing of the textile industry. Two examples are in the production of mono – filament nylon fibers, and production of natural fiber versus man – made nylon fibers. In these cases the companies used classification and regression trees to help determine spin breaks during production and a different production process was used in the case of a natural fiber versus man – made nylon fibers respectively. The analysis utilized data that was available from both on line and offline measurement and testing.

f) Data mining for Design and Manufacturing: Tremendous amounts of data leading valuable information can be collected and stored in databases at various stages of design and production. The data may be related to designs, products, machines, materials, processes, inventories, sales, and marketing and performance data. Furthermore it may include patterns, trends, associations, and dependencies. For example, understanding the data and the quantitative relationships among product design, product geometry and materials, manufacturing process, equipment capabilities, and related activities could be considered as strategic information that leads to power and success. Extracting, organizing, and analyzing such useful information could be utilized to improve and optimize company planning and operations.

Today, more information can be available from business transactions, scientific data, satellite pictures, texts reports, military intelligence, various stages of production and design. The huge collections of data that generated during daily operations, containing hundreds of attributes that needed to be simultaneously considered when accurately model the systems, behavior. Confronted with the abundance of data has impeded the ability to extract useful knowledge, and making it impractical to manually analyze for valuable decision-making information. This complexity calls for new needs to help us to make better managerial choices. These needs such as new techniques and tools known as knowledge discovery in databases (KDD) that can intelligently and (semi) automatically turn low – level data into high – level and useful knowledge.

g) Data mining in product design and development: Data mining is primarily used in business retail. Application to design and manufacturing are still under utilized and infrequently used on a large scale. The application of (DM) in various organizations reveals interest awareness among manufacturing companies across many industry sectors regarding the potential of data mining for changing business performance. For instance, data mining techniques have been used by Texas Instrument (fault diagnosis), Caterpillar (effluent control and warrant claims analysis), Ford (harshness, noise, and vibration analysis), Boeing (Post – Flight Diagnostics), and Kodak (data visualization). Still, the application of data mining to design and manufacturing is not broadly integrated within companies. Decision makers are hampered from fully exploiting data mining techniques by complexity of broad – based integration.

A product development process is the sequence of activities that a manufacturing company employs in order to turn opportunities and ideas into successful products.

Figure (2.1) shows the product development stage. Braha, et al. (1998) reported that, product development goes through several stages. Starting with identifying customer needs and ending with production and then delivery to market.

The nature of the product development process can be viewed as a sequential process. The design process evolves from concept through realization. For instance a part cannot be assembled until the components are machined; the components cannot be machined without a dimensioned part model; the part model cannot be dimensional without a set of relevant requirements.

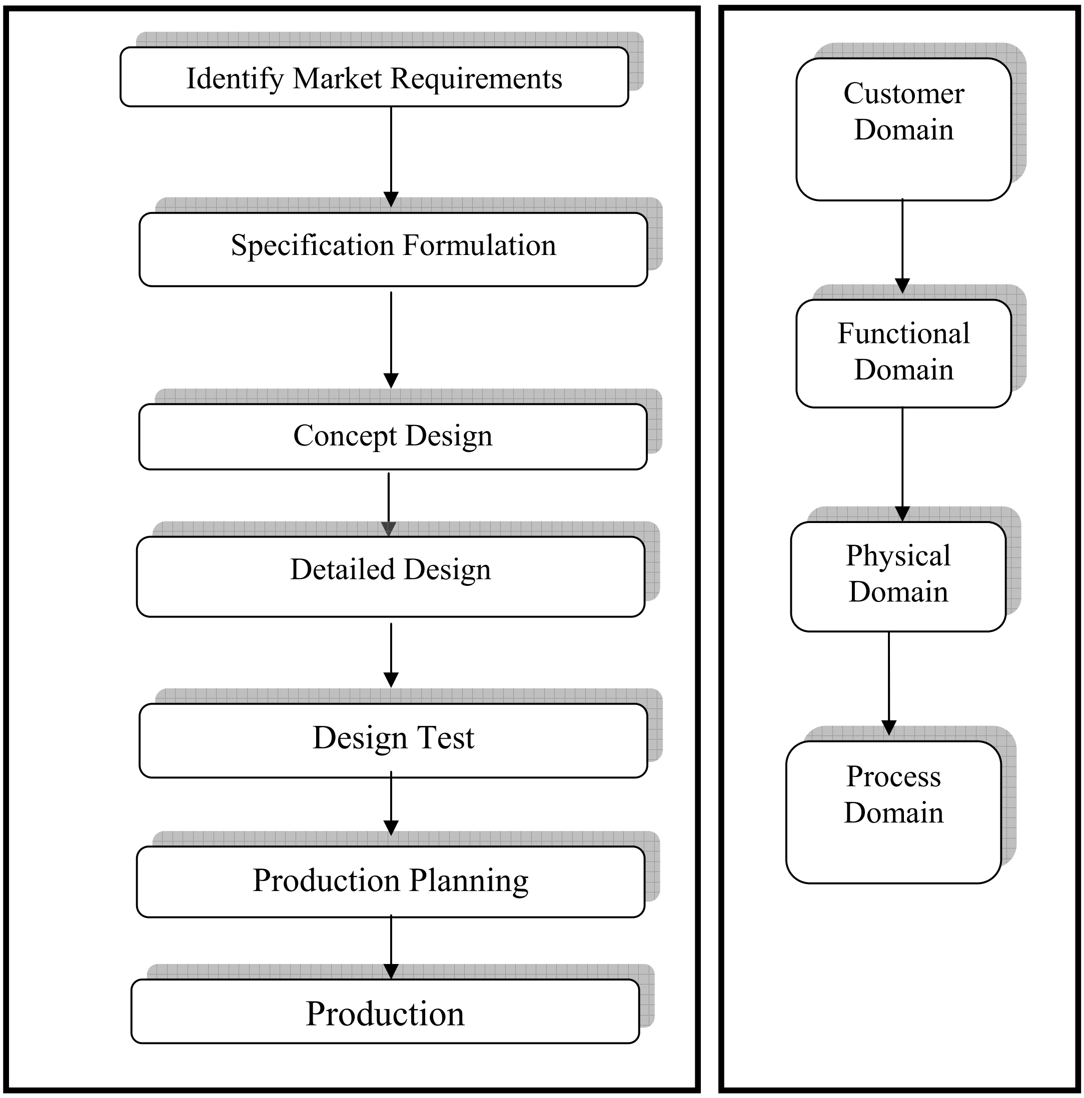


Figure (2.1): Traditional View of Product Design and Development

Also Braha, (1998) stated that, product development process, the product development iteration and categorizes can occur either between stages (inter -stage iteration) or within a stage (intra-stage iteration). See figure (2.2).

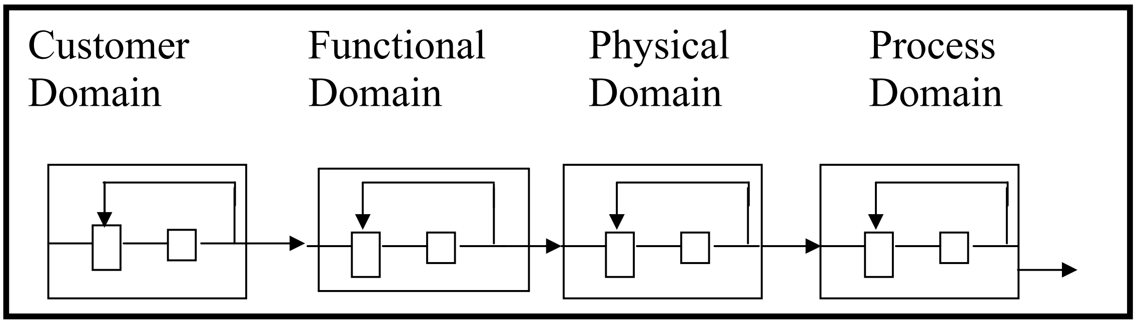


Figure (2.2): Combining Sequential and Iterative Design

In reality, these sequentially orders to most design processes may be affected. And not follow the same sequence. Product development cannot simultaneously consider every relevant aspect of any given design, because the new information, ideas, and technologies became available during the progresses of the product development process. The limitation of the design systems hinders the requirements to reach the optimum design. However the external environment reactions affect the model in any design system. All these points clearly indicate that there is an inherent, iterative nature to the product development process.

As shown in Figure (2.2) product development iteration is categorized as occurring either between stages (inter – stage iteration) or within a stage (intara – stage iteration). Product development still flows sequentially from initial concept through realization. Each process stage providing the data and requirements for the subsequent stage. Within each process stage, however, the designer iteratively creates a design that needs the given requirements.

This model largely represents the current state – of – the – art in Computer Aided Design (CAD), and Computer-Aided Manufacturing systems (CAM). There are numerous software modules to assist the designer during inter – stage design iteration, in order to help identify customer needs and to analyze a current design. However the tools are generally not well integrated at the inter – stage level. Dan Braha (2002) stated, it has been recognized that minimizing the inter – stage and inter – stage iterations tends to reduce product development lead time and cost.

2.2 Utilization of Data Mining During the Product Design and Development Process:

Data mining has the potential of becoming one of the key components in achieving the above mentioned goals. During the product design and development process data mining can be used in order to determine relationships among "internal' factors at each stage and "external" factors at consecutive and previous stages. The following examples illustrate how data mining can be utilized.

2.2.1 Data mining and extraction of patterns from customer needs:

Data mining can be used to extract patterns from customer needs, to learn interrelationships between customer needs and design specifications, and to group products based on functional similarity for the purpose of benchmarking modular design, and mass customization.

2.2.2 Data mining and the concept design selection:

At the concept design stage, data mining can support concept selection by dynamic indexing and retrieval of design information in knowledge bases (e.g., patents and benchmarking products). Clustering of design cases for design reuse, extracting design knowledge for supporting knowledge – based systems, extracting guidelines and rules for design – for X (manufactureality, assembly, economics, environment), and exploring interactively conceptual designs by visualizing relationships in large product development databases.

2.2.3 Data mining and system – level design:

During system level design data mining can aid in extracting the relationships between product architecture, product portfolio, and customer needs data.

2.2.4 Data mining and industrial design:

In industrial design, information about the complex relationships can be extracted with data mining and used in redesign. These include:

a) Relationships between tangible product features (such as color and ergonomics factors).

b) Intangible aspects of product that relate to the user (such as aesthetics, comfort, enthusiasm, and feelings).

2.2.5 Data Mining and detailed design stage:

At the detailed design stage, data mining can support material selection and cost evaluation systems.

2.2.6 Data mining and testing the design:

When testing the design, product characteristics can be extracted from prototypes. This may be used for determining the best design practices (e.g., design for reuse).

2.2.7 Data mining and product development planning:

During product development planning, data mining can be beneficial for certain activities such as the predication of product development span time and cost, effectiveness of cross – functional teams, and exploration of tradeoffs between overlapping activities and coordination costs. Data mining may also be used for identifying dependencies among design tasks that can be used to develop an effective product development plan.

2.2.8 Data mining and organizational level:

Data mining can be seen as a supportive vehicle for organizational learning, e.g., based on past projects, the factors (and their interdependencies) that affect a project's success/ failure may be identified. This may include the intricate interaction between the project and the company, market, and macro environment.

The improvement of product design and development can be achieved by the utilization of data mining to understand the relationship among internal and external factors facilitates the inter – stage and intra – stage iterations.

2.3 Examples of Successful Application of Data Mining in Manufacturing:

As with any new technology there are always doubts about whether or not it's suitable for any particular circumstances. While no technology will answer every business problem; data mining can be useful for certain things .Data mining is successful at solving problems with these characteristics:

a) Have vast amounts of data available,

b) The data consists of many variables,

c) The data is complex, multivariate, and non-linear,

The following examples show where data mining can be used successfully:

1) Fault diagnosis such as predicting assembly errors and defects, which may be used to improve the performance of the manufacturing quality control activity.

2) Preventive machine maintenance, which is concerned with deciding the point in time and type of maintenance of tools and instruments. For instance, cutting tool state may be classified and used for tool condition monitoring.
3) Manufacturing knowledge acquisition by examine relevant data, which implicitly contains most of the required expert knowledge. The extracted knowledge rules can then be incorporated by expert systems for decision support such as fuzzy controllers, diagnosis systems, and intelligent scheduling systems.
4) Operational manufacturing control such as intelligent scheduling systems that learn the effect of local dynamic behavior on global outcomes, and use the extracted knowledge to generate control policies.
5) Learning in the context of robotics (e.g., navigation and exploration, mapping, feature recognition, and extracting knowledge from numerical and graphical sensor data.
6) Quality and process control, which is concerned with monitoring standards; taking measurements; and taking corrective action in case deviation from the norm is detected and/ or discernible patterns of data overtime are present. Extracted knowledge may include classification to predetermined types of deviation from the norm, and causal relationships among temporally oriented events.
7) Adaptive human – machine interface for machine operation.
8) Summarization and abstraction of large and high – dimensional manufacturing data.
9) Enabling supply and delivery for casing, e.g., by classifying types of suppliers involved in transportation and distribution of the product.

2.4 Data Mining Processes in Design and Manufacturing Environments:

In order to successfully implement data mining in design and manufacturing processes the following issues should be considered. See figure (2.3)

1) Data cleaning: also known as data cleansing, it is a phase in which noise data and irrelevant data are removed from the collection.
2) Data integration: at this stage, multiple data sources, often heterogeneous, may be combined in a common source.
3) Data selection: at this step, the data relevant to the analysis is decided on and retrieved form the data collection.
4) Data transformation: also known as data consolidation, it is a phase in which the selected data is transformed into forms appropriate for the mining procedure.

It is common practice to combine some of these issues together. For instance, data cleaning and data integration can be performed together as a pre-processing phase to generate a data warehouse. Data selection and data transformation can also be combined where the consolidation of the data is the result of the selection, or, as for the case of data warehouses, the selection is done on transformed data.

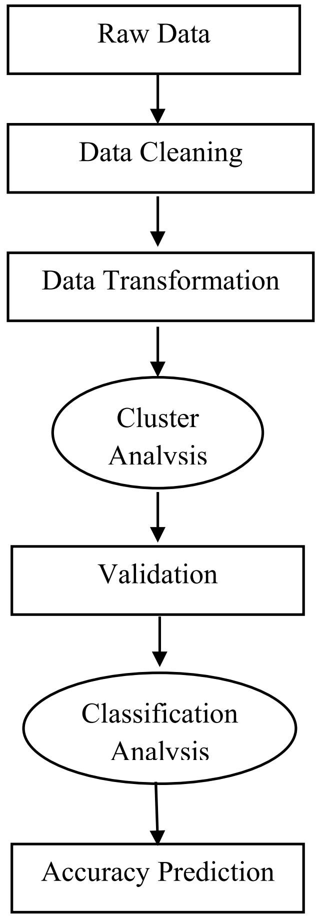


Figure (2.3): Data Mining Process

2.5 Enabling Technology for Data Mining:

Braha (1998) suggested that, the following enabling or supportive technologies are valuable:

1) Data Warehousing: is considered as one of the most important supportive technologies and it is defined as the process of integrating legacy operational systems (storing data related to product, process, assembly, inventory, purchasing, etc.) within a company to provide

centralized data management and retrieval for design support purposes. Thus the preprocessing, including cleaning and transforming data with intent of analysis and discovery can be facilitated by a data warehouse.

2) Report Generators: which are used to present the extracted patterns in a user – readable way, are another type of supportive technology. If discovered knowledge is further used by various computers (e.g., industrial robots), it is imperative that computers be able to interpret the output.

3) Computationally massive data mining operations: can be enabled through parallel – computing plat forms and distributed computation. The parallelism aspect is deployed proactively and systematically throughout the manufacturing environment, and is used for continuous tasks such as preventive machine maintenance and real-time monitoring of the overall manufacturing processes.

2.6 Overview of Industrial Data Mining Application:

Lee (1993) stated that, the use of data mining techniques in industry and manufacturing began in early 1990s. Several domains of data mining applications in manufacturing engineering from early 1990s to 2005 were presented, (Harding et al, 2006). These studies were carried out to determine production process, fault detection, maintenance, decision support, product quality improvement, and quality control. Semiconductor industry had enormous influence on data mining appliances. Bertion et a l, (1999), also reported that, data mining technique has been successively applied to wafer manufacturing.

Other manufacturing industries such as, the computer component manufacturing, the quality control automated data mining system, automated construction of compact and interpretable models from highly noisy data sets, product quality improvement in ultra – precision manufacturing industry, clothing manufacturing to generate useful patterns and rules for standard size charts, carpet manufacturing byproducts to improve robustness of large scale bioprocesses, prediction and diction of significant meteorological phenomena and in medicine. Table (2.1) shows the major stages of knowledge database discovery process in manufacturing.

The reviewed literature showed that, there is a rapid growth in the application of data mining techniques in the industry and manufacturing sector. The necessary data can be collected during the normal manufacturing processes and therefore there is no need to introduce dedicated processes.

Furthermore, despite this rapid growth, there is still slow adoption of this technology in some industries for several reasons including; difficulties in determining the type of data mining function to perform in any particular knowledge area and question of choice the most appropriate data mining technique regarding too many possibilities.

Table (2.1): Major Stages of Knowledge Database Discovery process in manufacturing

No.	Stage Name	Description
1.	Understanding the manufacturing domain	Stands the relevant prior knowledge related to manufacturing application and targeted goal.
2.	Collecting the targeted data	This stage includes the collecting raw data, selecting the data sets and focusing on the set of variables affecting the problem partly defined in the first stage.
3.	Data cleaning, pre-processing and transformation	This incorporates the pre-processing of data such as noise removal, and both replacement of missing values and data cleaning. Data are consolidated into forms appropriate for mining.
4.	Data integration	This step includes multiple manufacturing heterogeneous data sources integration.
5.	Choosing the functions of data mining	Depending on the problem defined (stage 1) various data mining functions (clustering, classification, prediction, association, regression, summarization etc.) need to be performed to derive the model.
6.	Choosing the appropriate data mining algorithm	The selection of technique is important to perform the desired function for finding patterns in the data.
7.	Data mining	Includes searching for patterns of interest in a particular representational form or a set of such representations.
8.	Interpretation and Visualization	Tasks that include the interpretation and visualization of patterns to derive novel knowledge.
9.	Implementation of Discovered knowledge	Incorporation of discovered knowledge into the manufacturing domain performance system. The feedback is received and the knowledge can be modified further based on the achieved feedback.
10.	Knowledge storage, reuse and integration into the manufacturing system	This includes the storage of discovered knowledge for future reuse and possible integration in to the manufacturing system.

Source: (Choudhary A. et al, 2009)

2.7 Application of Data Mining Techniques in Spinning, Weaving and Knitting Industries:

While data mining has been widely used in many fields, little research has been done on its applications in textile manufacturing. In the textile industry, most of the work that have been conducted presents the application of data mining for production planning, forecasting, promotion, distribution and more recently stylistics: "Mining data from Fashion", most of these researches done at the retail level. Few researches have been found applying data mining techniques at the manufacturing processing of the textile industry. The following examples illustrate where data mining techniques being used in the field of spinning, weaving and knitting manufacturing. Industries

1) Menezes, J. (1999) used data mining techniques in the production of mono-filament nylon fibers. The data mining method used by the company were classification and regression trees in order to help determining spin breaks during production. The analysis utilized data that was available from both on and off line measurement and testing.
2) Schertel, Stacey (2002) understood the possible uses of data mining in the textile industry; specifically a spinning mill was studied. Results showed that there was a similarity between this research and the one mentioned above. Despite the fact that there is a slight difference in the production processes for natural fiber versus man – made fibers (nylon). Data was collected from a spinning mill operation and then cleansed and merged to create a data warehouse. In this warehouse the different elements and formats that are needed were listed for each process in the production of a cotton fiber. A simple data mining process was used, due to successes and failures that were experienced during the research

a new data mining process model were created. This model has six major steps, which contains a total of 28 specific activities that may be included in the data mining process model. The proposed model describes how data mining can be implemented in a manufacturing setting.

3) Waldernar et al, (2011) stated that in carpet manufacturing, "most of the latest publications from 2010 concern a new approach based on particle swarm optimization algorithm for clustering problems description (Duran et al, 2010) or knowledge induction form data to detect and isolate machine breakdowns in carpet manufacturing.

4) Research indicates that in many industries including textiles, Information Technology (IT) was not being used efficiently or effectively. Georg et al (2001) reported in a project aimed to enhance decision effectiveness in textile manufacturing using the data to decision model as a basis, thereby designing new and efficient information stage.

5) More recently, (Prof.Tygi et al, 2011) used data mining tools and techniques to manage the textile quality control data for strategic decision making. It was found that by introducing these techniques into textile production processes it is possible to achieve a substantial increase in productivity and quality of work.

2.8 Historical Background for Standard Garment–Sizing Systems:

Garment sizing systems were originally based on those developed by tailors in the late 18th century. Before that all garments were hand- made to order. Tailors measured the body dimensions of each customer, and then drew and cut patterns for each garment; for specific customer after many original patterns had been accumulated the tailors discovered correlations between bodily dimensions, regardless of the individual differences. Tailors gradually developed these patterns into a system of garment storage, which could be used to make clothes for people with similar figures.

Emanuel et al, (1959) established a set of procedures for formulating a standard sizes for figure types. Accordingly, people of all figure types were first classified into one of four body weight groups, which are subdivided by height. Therefore people were divided into eight categories based on similar sizing systems with classification based on two or three sizing variables for male and female as well.

The sizing variables for male garment are height, chest girth and waist girth. The sizing variable most commonly used by female garments are height, bust girth, and hip girth.

McCulloch et al, (1998) introduced criteria by which the employed (used) systems could be evaluated. They justified that sizing systems should:

a) Cover the greatest number of people and,
b) Require the fewest number of sizes.

At times these criteria conflict with each other. Depending on circumstances, one may take priority over another. Since the late 1800s, anthropologists used tape measures and calipers which are still being utilized for measuring the human body, (Detong et al, 1993). During that period most clothing was custom – made by tailors. Various measuring methods were developed by professional dressmakers and craftsmen. Their techniques for measuring and fitting their clients were unique.

In the 1920s; the demand for the mass production of garments created the need for a standard sizing system. The mail order in 1930s led to frequent returns of defective garments. Devarajan et al, (2002) stated that, "the sizing system for women's apparel developed as a result of a large anthropometric survey of 10042 women.

The Joint Clothing Council was the first publisher of the classical terminology and methods of body measurement for the clothing field, which leads to standard reference for body measurements. Body measurements, were divided into four groups these are: stature, segment length, body breadth, and circumference. Beazley (1996) suggested a procedure by undertaking a size survey using International Organization for Standardization (ISO) 8559(E) (1989) which include a natural sequence of body measurement comprising three types of data: horizontal, vertical, and others. Beazley (1997) reported that; in Japanese Body Size Data (1992-1994), the definition, equipment, methods and procedures of body measurements were discussed.

2.9 Image Processing and Modeling Methods for Measuring Human Body:

The methods of measurements mentioned in the introduction above were criticized on that, they are time consuming and often not accurate. Therefore, many researchers all over the world have directed their efforts towards obtaining more reliable measurements such as 3D profile of the human body using various techniques.

The utilization of image processing and modeling methods for the digitization of the human body considered as the third technology. In this technique, 3D measurements are not performed, but 3D information is generated and extracted from 2D images.

Nicola (2007) proved that, "technologies used commercially for the digital measurement of the human body can be divided into five different groups:

a) Laser scanning,

b) Projection of white light patterns,

c) Combination modeling and image processing,

d) Digital manual measurement,

e) Technologies based on other active sensors.

Since early twentieth century research directed towards body image. The optical devices for non – contact measurements used as scanning technologies for all body. Before the establishment of 3D methods, different types of 2D photographic techniques, such as silhouette, were commonly used to present a complicated body profile. During the 1980s 3D body scanning technologies have grown rapidly and they can be grouped into the following three categories:

a) Structure light,

b) Laser infrared,

c) And photogrammetric.

In Sudan Elabid, Amel E. A (2009) carried out a research in order to develop a technique that could be used to measure the human body without any direct contact using image technique. The results obtained showed that this technique reduces the time needed for measurement and increases the accuracy of measurement Manual and image technique measurements were taken for (male and female). The measuring technique procedure was as follows:

Firstly, the images were fed to the computer and processing them using "Adobe Photoshop" program. Secondly, measurement calculations were transferred to Microsoft Excel program and a number of equations were developed for size measurement. Statistical calculations including the mean, standard deviations, and correlation coefficients were carried out for image and manual techniques.

The results revealed that the image technique had taken a period of 0.96 minutes, compared with 3.8 minutes for every subject which indicates that measurement by image technique needs a shorten time. Regarding the size measurement values for both techniques, the results obtained showed that there is no difference.

2.10 Application of Data Mining Methods in Garments Industry:

Data mining methods have been used in many fields, but little research has been done on its application to sizing systems for the manufacturing of garments. Norsaadah et al, (2008) established a sizing system for the manufacture of garment using decision tree based data mining to determine the pants size of army soldier's uniform. Firstly, it was defined the subject and constructed a large anthropometric database. Secondly, the data was prepared and analyzed, performed factor analyses, and then extracted important sizing variables. Thirdly, the data was used and classifies significant patterns in the body shapes of soldiers. The justification of the implementation of decision tree techniques –to establish sizing systems was based on.

a) It allows for a wider coverage of body shapes with a fewer number of sizes,
b) It Generates regular sizing patterns and rules, and
c) It provides manufacturers with reference points to facilitate production.

It was stated that the newly developed sizing system can provide garment manufacturers with size specifications, design development, pattern grading, and market analysis. In addition to that it can minimize the inventory costs of the production plans due to mismatches and make it more realistic.

Hsu (2008) conducted an empirical study in apparel industry in order to support manufacturing decision for production management as well as marketing with various customers' needs.

Hsu (2009) also discussed a two – stage cluster approach based framework was considered, it generates useful patterns and rules for standard size charts. More recently (Bagherzadeh et al, 2010) introduced a study for developing sizing systems by data mining techniques using anthropometric data. A three – stage data mining procedure were employed to develop sizing system for lower body figure type of Iranian male. The approach includes three phases:

a) Factor analysis employed to mine through the variable and to determine the main effective factors.
b) A two – step cluster analysis used included both hierarchical and nonhierarchical clustering cases sorted to the cluster according to the factor analysis results.
c) The decision tree analysis employed to extract the most significant classification rules of the body type based on the results of cluster and factor analysis. The results showed that three body type and sizing system developed, have a good fit performance.

Hai et al, (2008) studied the application of data mining techniques for developing a sizing system for army soldiers' uniform in Taiwan. The study aimed to establish systems for determine the sizes of garments for army personnel using the data mining techniques. An anthropometric database for the purpose of simplifying the entire process was constructed. It was found that the newly developed sizing systems can be adopted to better accurately predict the requirements of different sizes of uniforms, generate a practical production planning procedure, and minimize unnecessary inventory costs resulting from sizing mismatches.

The classification and regression tree (CART) technique was used to explore and to identify significant pattern based on a large amount of anthropometric data of Taiwan military personnel. In Malaysia; (Norsaadah et al, 2008) used data mining technique to explore anthropometric data for the development of sizing system .An anthropometric survey of girls aged between 7 and 12 years old were conducted. The whole data was analyzed using descriptive analysis of average, mean and standard deviation. The data obtained was further explored using the factor analysis method. Principal component analysis technique (PCA) was done to reduce the variables to similar factor components. The decision tree was used to validate the cluster groups obtained. These segmented groups were then converted into size tables.

Table (2.2): illustrates some examples of the application of data mining techniques in developing sizing systems for clothing industry in different countries.

Table (2.2): Application of data mining techniques in developing sizing system for garments in different countries.

Country	Taiwan	Iran	Ghana	Malay	India
Targeted size	Size System for Army Soldiers	Size system for pants\suit Male:16-22	Size system for Ghanaian Women 16-35	Size system for female 7-12	Size system for shirt for men 25-66 years
Data Mining Techniques Applied	Classification Decision-tree	Quadratic average of difference Aggregate loss goodness of fit sizing Hierarchical &Clustering Classification Decision tree (CART)	K-means Algorithm	Decision Tree Clustering	Error Improvement Regression Classification K-nearest neighbor 2(knn)+Random tree
Algorithms	Factor Loading Clustering Body Mass Index(BMI)	K-means finally Multivariate Analysis cluster SPSS(Equal-Variance Maximum like LiHood (EML) factor analysis	Step by Step Procedure	Principle Component Analysis(PCA) Factor Analysis exploration SPSS	X – means clustering missing data by multivariance techniques
Selected Techniques	Decision Trees	Three Stages	SPSS	Two Stage	Two stage
Year	2008	2010	2012	2010	2012

Source: Researcher design

From this background the data mining methods used in the field of sizing system in clothing industry can be summarized as follows:

1) Neural net work.
2) Cluster analysis.
3) The decision tree approach.
4) Two stage cluster analysis.
5) Three stage data mining procedure.
6) Two stage based data mining procedure include cluster analysis and classification algorithms.

2.11 Limitations of Data Mining Methods Used in Sizing System for Garments:

In recent years data mining methods have been widely used in area of science and engineering. The application domain is quite broad covers area such as bionic- formtics, genetics, medicine, education, electrical power engineering, marketing, production, human resources management, risk prediction, biomedical , technology and health insurance.

In the clothing a cluster which is typically grouped by the similarity of its members' body shape can be considered as a size category or a figure type. The drawback of these methods is that it requires one to pre – assigning the number of clusters to initialize the algorithm and it is usually subjectively determined by experts. Also the use of two stages clustering, will not define the rule of the classification for categorizing the figure type.

2.12 Research Problem Statement:

1) The sizing systems used by clothing manufacturers for men's wear vary widely in size ranges and measurement charts. Most often these

measurements are very different from the standards recommended by the International Standards Organization (ISO).

Many problems caused by an inappropriate garment sizing for manufacturers, retailers and customers as well. This will result in financial loss because of missing of adequate data and current sizing specifications.

2) Inventory planning and evaluations of market potential depend largely on rules of thumb or ideal sizes that have an uncertain relationship to real consumer size specifications. This will lead to high production and inventory planning, markdowns, loose of marketing opportunities and high levels detective products

Pechoux et al, (2002) claimed that, people are different, having their very own mix of measurements and proportions. If they all followed a normal distribution, with fixed mean and standard deviation, statistical models used for processing anthropometric data and converting it into size measurement charts would be right on target.

3) A considerable amount of current sizing systems are based on statistical models derived from outdated or inappropriate anthropometric data.

Kurt, S.A (1996) study, revealed that half of all American consumers are unable to find ready – made garment that fit them properly". Furthermore J. ABEND (1993) claimed that an estimated 70-80% of garments on the rack did not correspond to the reported size, forcing consumers to content with a wide variation in fit and mail order houses to experience a 30% of defective garments due to poor fit.

2.13 Summary:

One of the basic necessities in our life is clothing. Therefore, it is essential to wear clothes that is comfortable and has a good fit.

Review of the literature clearly showed that in Sudan there has been no studies carried out in the field of sizing and fit system using data mining techniques. However a study was carried out in order to develop a technique that could be used for measuring human body without direct contact.

The author experience in the field of manufacturing of ready to wear clothes in Sudan clearly indicates that there is a problem of size and fit facing Sudanese people when they purchase readymade garments.

The problem of this research based on the assumption that the current sizing system for (poshirt) uniform which consists of jacket and trouser used by Sudanese army officers were needed to be manufactured according to the standard garments sizing systems. Therefore, the problem of this research was to establish standard sizing system for Sudanese army officers garments (poshirt) by application of data mining technique.

2.14 Objectives of the Current Research:

The main objectives of this research were to;

1- Establish a sizing system that could be used to determining the size of garments for Sudanese army officers (poshirt) uniform by applying data mining methods.
2- Explore the body types of Sudanese army officers (male) based on the anthropometric variables involving all body dimensions.

3- Groping Sudanese army officers (men) into homogenous body size group using cluster and visualization techniques.
4- Identify the key dimensions for upper and lower or whole body sizing system.

2.15 The Steps to Investigate the Research Problem:

The steps that would be followed to investigate the research problem are given in figure (2.4)

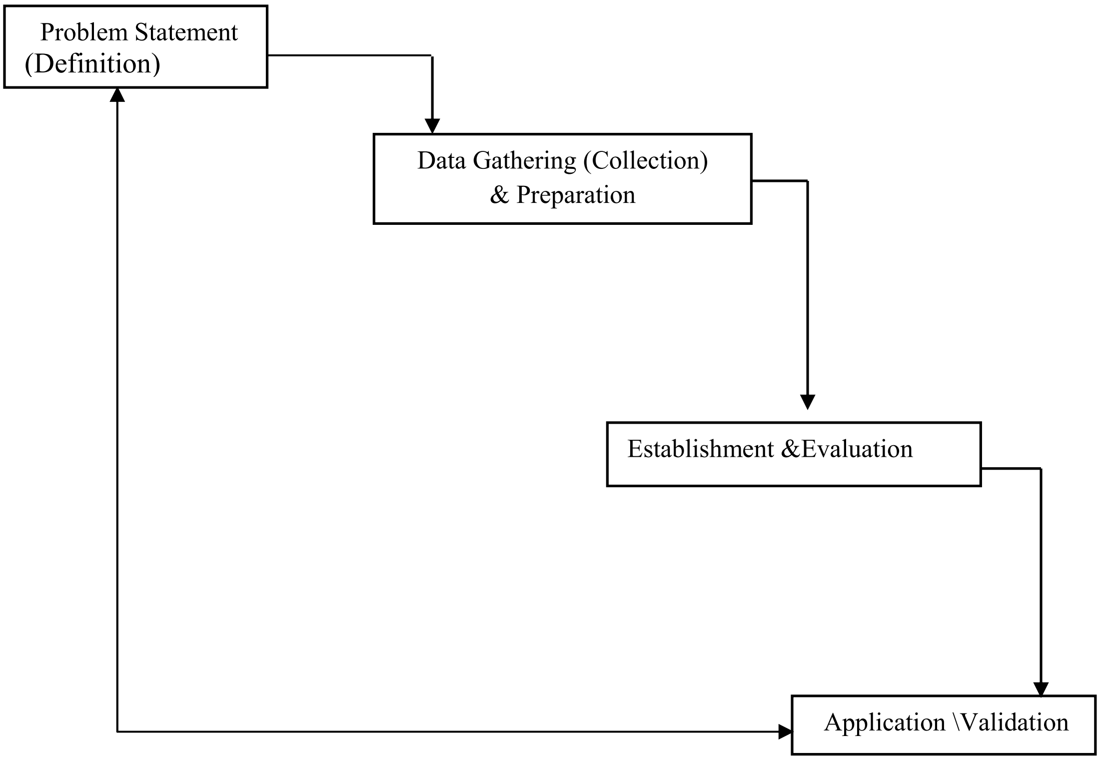


Figure (2.4): Flow Chart of the Investigation of the Research Problem

CHAPTER THREE

MATERIAL AND METHODS

CHAPTER THREE

Material and Methods

3.1 Material:

In this research, the anthropometric database for army officers was obtained from (Sur Military Clothing Factory) located in Khartoum North (Industrial Area), in the Sudan used. The based dataset collected has a total of (841) record. These records have existing measurements authorized by an anthropometric expert using a measuring tape. The measurements for the anthropometric variables were in inches with decimals. Therefore, they were converted into integer centimeters in order to ease the comparison with the commonly used international garment sizing standard units. Further, they were preprocessed i.e. they were examined and purified to omit the outliers in order to increase the efficiency and ensure the accuracy of its analysis during the processing and transformation. For each record (individual) 13 anthropometric variables were employed resulting in a total of 10933 variables.

The measurements of the anthropometric variables were taken from army officers with age ranging from 16 to 60 years. The bodily dimensions commonly adopted by the garment manufacturing sector were first considered. Domains experts were consulted to identify anthropometric variables that are necessary for producing poshirt (jacket + trouser) uniform.

Based on the experience and the advice of the experts the selected variables for the jacket were collar, chest, waist, length, across shoulder, sleeve and sleeve and cuff. For the trouser variables were; waist, hip circumference, thigh circumference, knee girth, foot, and trouser length. These measurements of anthropometric data followed the ISO 85591/1989 body measurement standard. The thirteen dimensions for (poshirt) measured by using a measuring tape.

As mentioned before most anthropometric databases descriptions of body size based on measurements such as segment length, body breadths sutured and circumferences were obtained over the body surface with standard tools such as measuring tape and calipers.

3.2 Methods:

One of the difficulties when deciding to choose a data mining system is to determine which method (technique) is appropriate to establish a sizing system for clothing. This is because there are many methods (techniques) that could be used. Some of the criteria that are important in determining the methods (techniques) to be used are determined by trial and error; depending on the type and objectives of the research and the method or techniques available. Bearing in mind the strengths and weaknesses of each method.

3.2.1 Determining the software package (technique):

The choice of the software was not an easy job. From the review of earlier research, it was observed that there are several different mining software packages that had their own strengths and weakness. Based on this information it was decided to select two software packages that were suitable for establishing sizing system. The chosen packages were:

3.2.2 WEKA 3.6.9:

In this research Waikato Environment of Knowledge (WEKA) was used. WEKA is software written in Java and runs on almost any plat form. It is a data mining system developed at the University of Waikato in New

Zealand. WEKA is free software available under the GNU (General Public License). The WEKA work bench contains collection of visualization tools and algorithms for data analysis and predictive modeling together with graphical user interfaces for easy access to this functionality.

The reasons for choosing these techniques were; WEKA 3.6.9 is more advanced version which has been implemented in Java with latest Windows7 operating in Intel Core ZQuad@2.83 GHz and 2 GB memory, with a fairly simple microprocessor as software allows for the data to be imported into their software directly from Excel. Martin et al (2012).

This is an advantage because; all the data collected from anywhere can be downloaded into Excel, which is a well known and widely used program. This data set can be applied and identified systematic clusters in bodily dimensions. Based on these clusters the representative figure types can be classified for establishing the required sizing systems.

In addition to that the following points support the reasons why WEKA technique was selected;

1) Portability since it is fully implement in the Java programming language and thus runs on almost any modern computing platform.
2) A comprehensive collection of data preprocessing and modeling techniques.
3) It is easily useable by people who are not data mining specialists, and also due its graphical user interfaces.

It provides flexible facilities for scripting experiments.

4) It has kept up-to-date, with new algorithms being added.
5) It provides many different algorithms for data mining and machine learning.

These are the key features responsible for WEKA' s success.

Sunita et al, (2011) added that all of WEKA's techniques are predicted on the assumption that the data is available as a single flat file or relation, where each data points described by a fixed number of attributes (normally, numeric or nominal attributes, but some other attribute are also used).WEKA is not capable of multi-relational data mining, but there is separate software for converting a collection of linked database tables into a single table that is suitable for processing using WEKA.

3.2.2.1 Launching WEKA Explorer:

In this research the following steps were followed through the analysis of the problem using WEKA Explorer, preprocessing, classification, clustering, association, attribute selection, and visualization. Figure (3.1) shows these steps when using "cluster" tab at the top of WEKA Explorer window. WEKA can be launched from c:\ program files directly, from the desktop selecting WEKA 3.4, shortcut 2 KB icon or from the windows task bar "start" → programs → WEKA 3.4. When WEKA GUI chooser, window appears on screen, one of the four options button of the window can be selected, R. Kirkby, (2002).

WEKA 3.6.9 could be launched by following the same steps mentioned for WEKA 3.4. When WEKA GUI chooser' windows appears on screen and one of the following steps could be selected. The steps one; (preprocessed, classify, cluster, associate, select attribute, and visualization) as shown in figure (3.1).

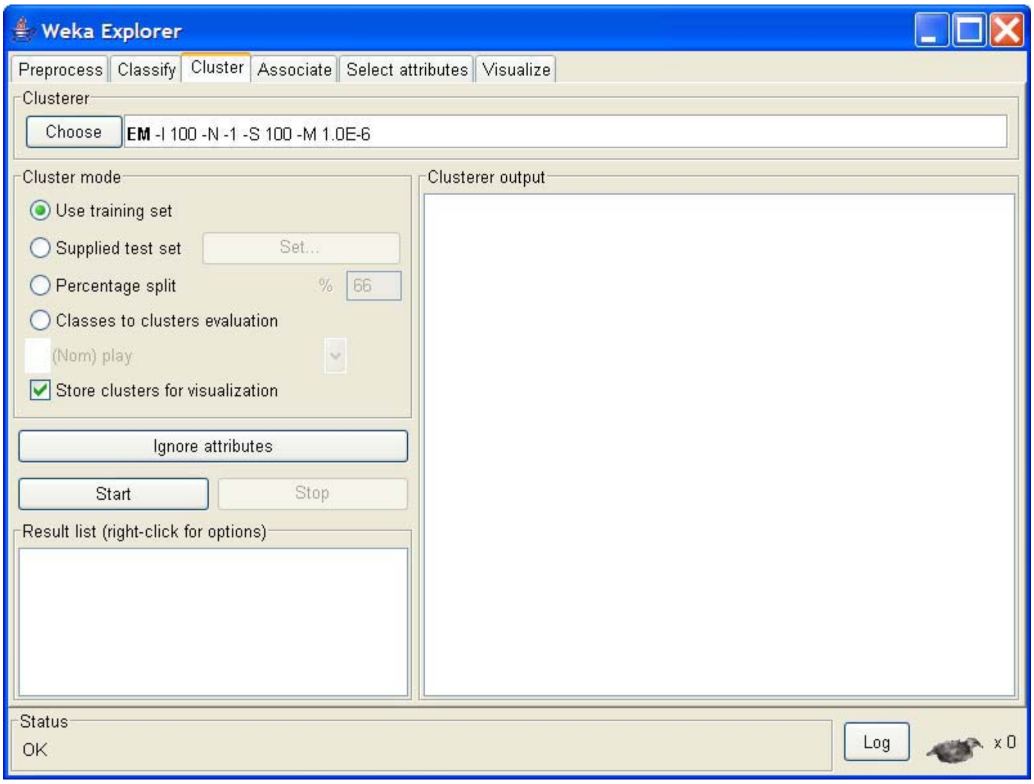


Figure (3.1): Using 'Cluster' tab at the top of WEKA Explorer window

3.2.2.2 WEKA Cluster Analysis:

Clustering is the portioning of a dataset into subset (clusters), so that the data in each subset (ideally) share some common trait. Cluster analysis is a powerful technique to divide heterogeneous data into groups. Cluster analysis was used as an exploratory data analysis tool for classification. In the clothing cluster which is typically grouped by the similarity of its member's body shape can be considered as a size category or a figure type. The simple K-means method was implemented to determine the final cluster categorization. Simple K-means algorithm was used as the clustering approach. This is a prototype–based clustering technique that attempts to find a user–specified number of clusters (Pang, 2006). "K is a certain number of clusters fixed as a priori K centroids are defined for each cluster. Each sample is then assigned to the closest centroid, and each collection of samples that assigned to a centroid is a cluster".

The dataset in this research has different characteristics, such as the number of instances, the number of attributes, the number of classes, the number of records, and the percentage of class's occurrences.

In this research the database was designed in (size.csv) file system to store the collected data. The data was formed according to the required format and structures. Further the database was converted to (size.csv) format to be processed in WEKA. The text file describes a list of instances that sharing a set of attributes. After processing the (size.csv) file in WEKA a list of all attributes, statistics, and other parameters can be utilized as shown in figure 3.2.

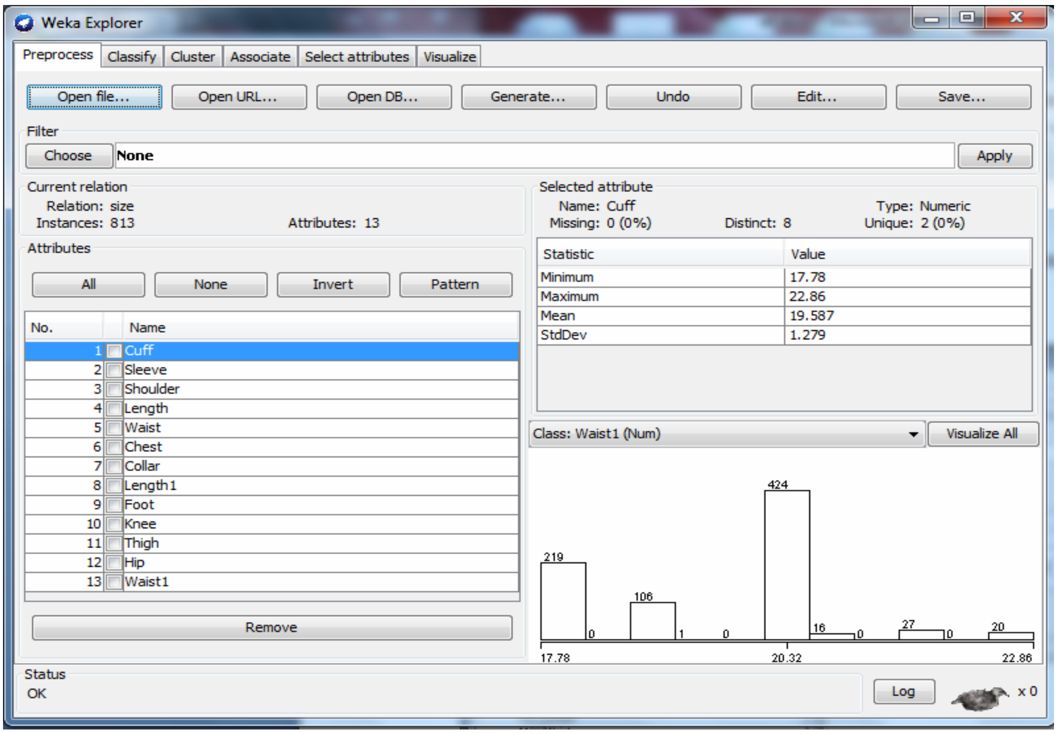


Figure (3.2): Processed (size.csv) file in WEKA

In this research, an anthropometric database collected from Sur Military Clothing Factory was used. Anthropometric data of 841 army officers were obtained. A total of 813 number of instances were processed, with (13) different attribute. The attributes for the jacket were, cuff, sleeve, shoulder, length, waist, chest and collar. For the trouser the attributes were length1, foot, knee, thigh, hip, and waist1. For the trouser, the term length1 and waist1 were used in order to distinct them from those of the Jacket.

Because WEKA does not work with numbers, these (813) instances for (jacket + trouser) were categorized in numerical groups. Based on Sur Military Clothing Factory Poshirt (U4) size specification all the (813) instances were grouped in (7) groups; named XS, S, M, L, XL, 2XL and 3XL respectively. The distribution of the proposed groups was based on waist and length attributes.

3.2.2.3 Analysis of Data Processed in WEKA

As mentioned before, the processed data in WEKA can be analyzed by using different data mining techniques such as, classification, clustering, association route mining, visualization algorithms etc. Figure (3.3) shows the processed attributes visualized into a 2 dimensional graphical representation.

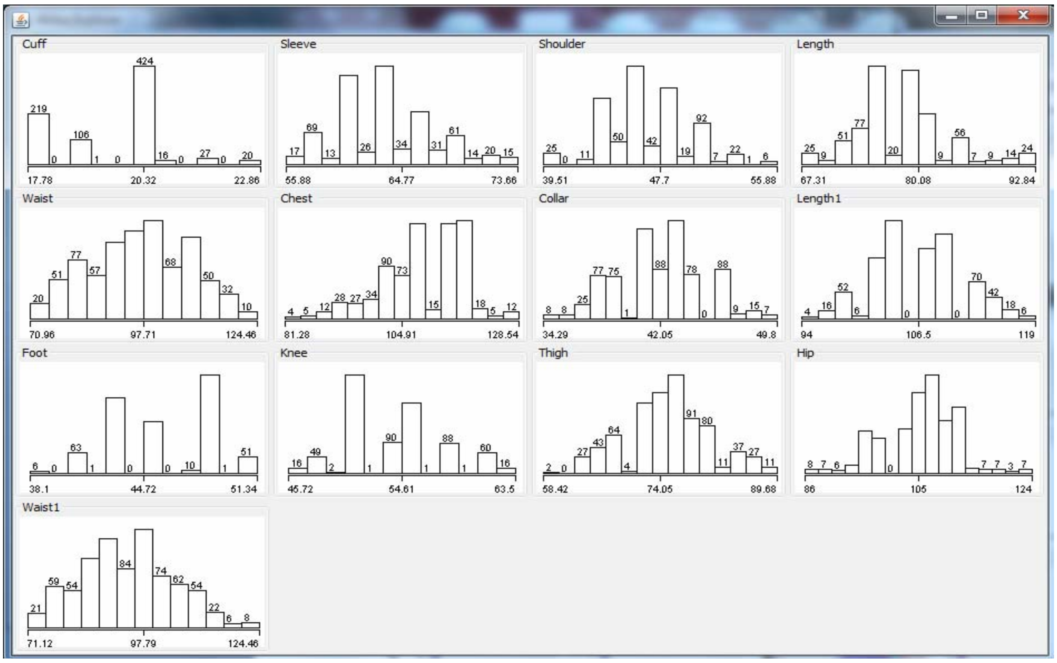


Figure (3.3): Graphical visualization of processed attributes

As shown in figure (3.4) there are many algorithms. From these algorithms, the cluster scheme "Simple K Means" was selected. The implementations of K-means only allow the selected numerical values for attributes.

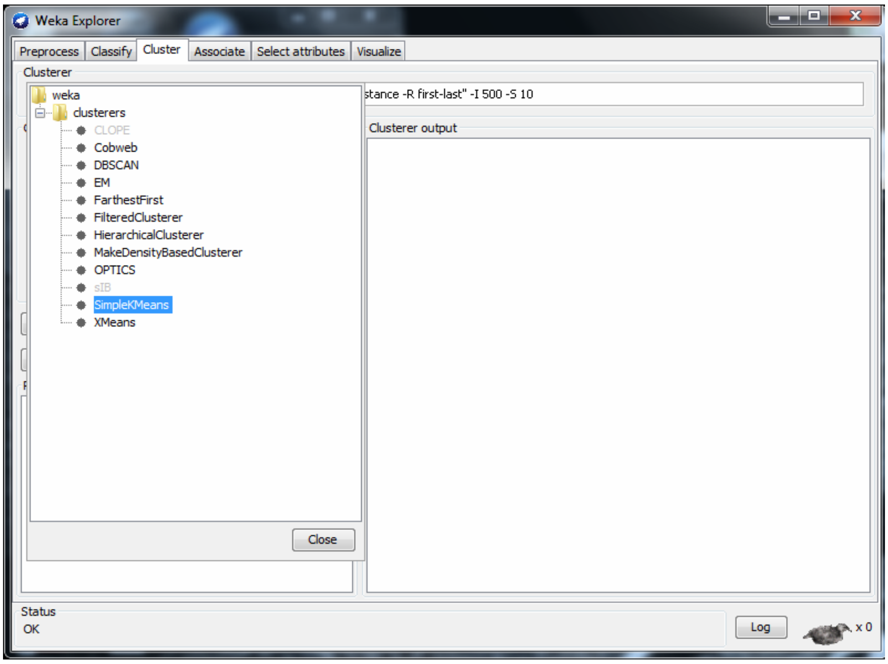


Figure (3.4): WEKA explorer

3.2.3 The Statistical Package for Social Sciences (SPSS) Version 18.0

This package was the second method selected because it is a simple clustering method and shows optimal results. However all variables must be independent and normally distributed. Clusters were identified according to the different body types. It was employed for anthropometric data analysis. It reduces the large samples in same groups contains similar number. The co-efficient correlation in determining the relationships between the dimensions can be obtained. Finally the WEKA and the (SPSS) methods could be used to establish new sizing systems.

CHAPTER FOUR

RESULTS AND DISCUSSION

CHAPTER FOUR

Results and Discussion

4.1 Cluster Output:

The cluster values were set in (num Clusters) box to use the clustering scheme: Simple K-Means with 7, 8 and 9 clusters respectively. These clusters include the attributes from 0 – 6, 0 – 7, and 0 – 8 respectively.

- First set the value in (num Clusters box) to 7 instead of default (2), 8 instead of 7 and finally 9 instead of 8. Therefore, the clustering scheme used was Simple K-Means with 7 clusters as an example. When set to run mode the following information will appear on the dialogue box.
- The relation name "size".
- Number of instances in the relation is 813.
- Number of attributes in the relation is 13.

The 13 attributes used in clustering are shown in figure (4.1).

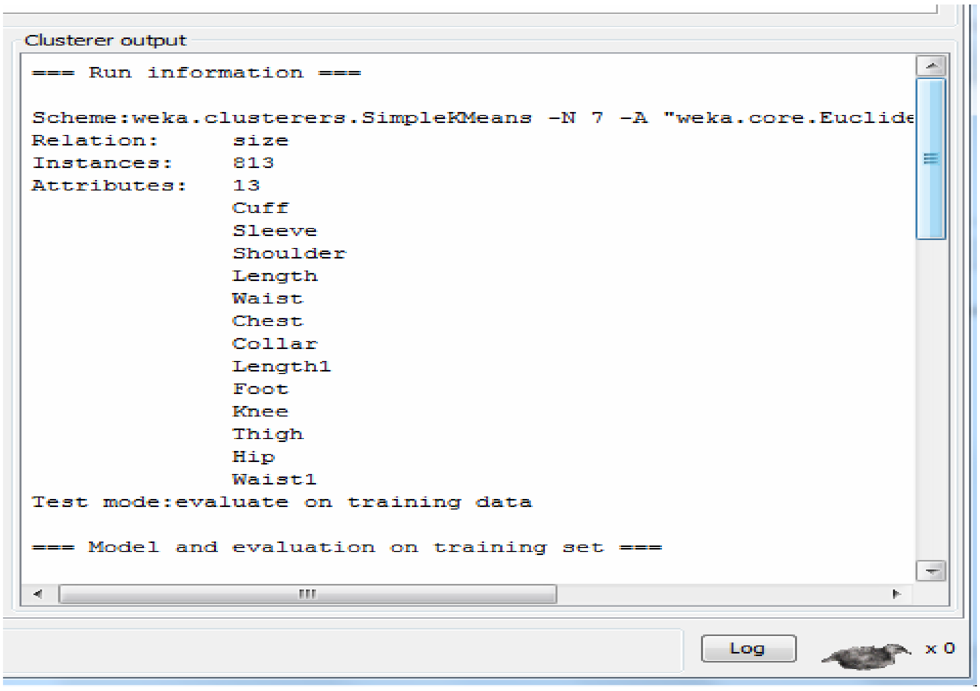


Figure (4.1): The output from the run mode

4.2 WEKA Results and Discussion:

The distribution of classes in clusters 7, 8 and 9 in the actual training data for classifier evaluation on the occurrences and the percentage of size categories using Pie chart are shown in figures (4.2, 4.3 and 4.4). Figures (4.5, 4.6 and 4.7) represent the results of clustering through visualization for cluster 7, 8 and 9 respectively.

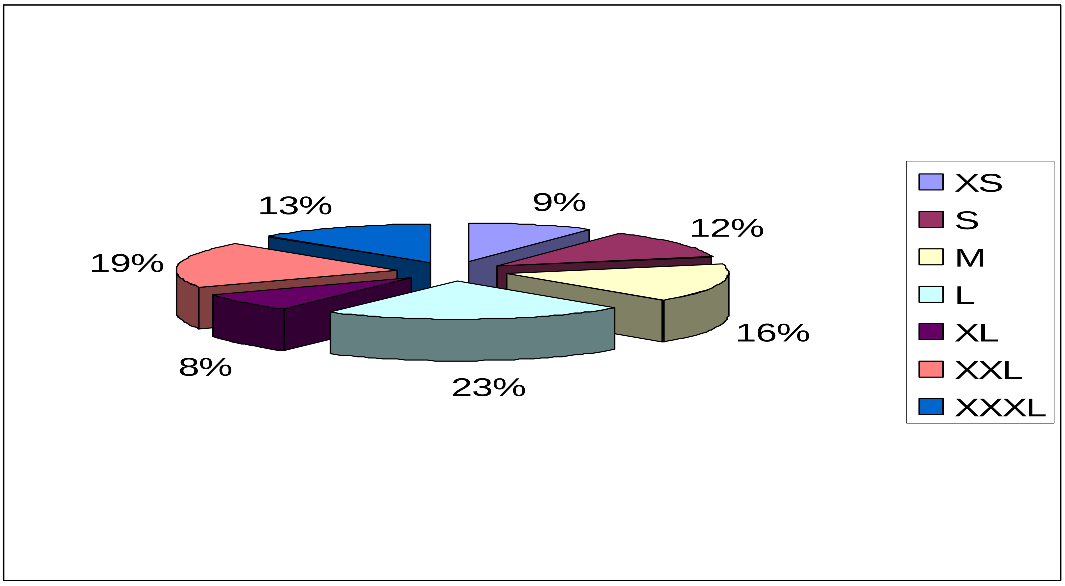


Figure (4.2) :the percentage of size categories for cluster 7

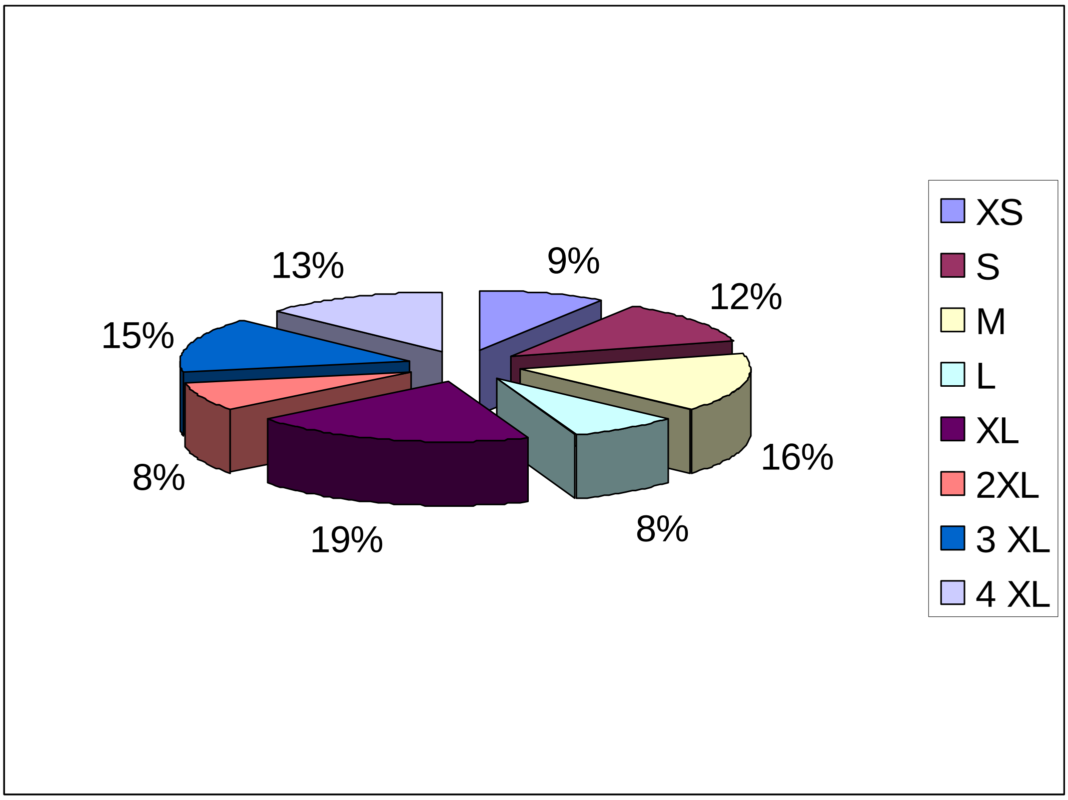


Figure (4.3): the percentage of size categories for clusters 8

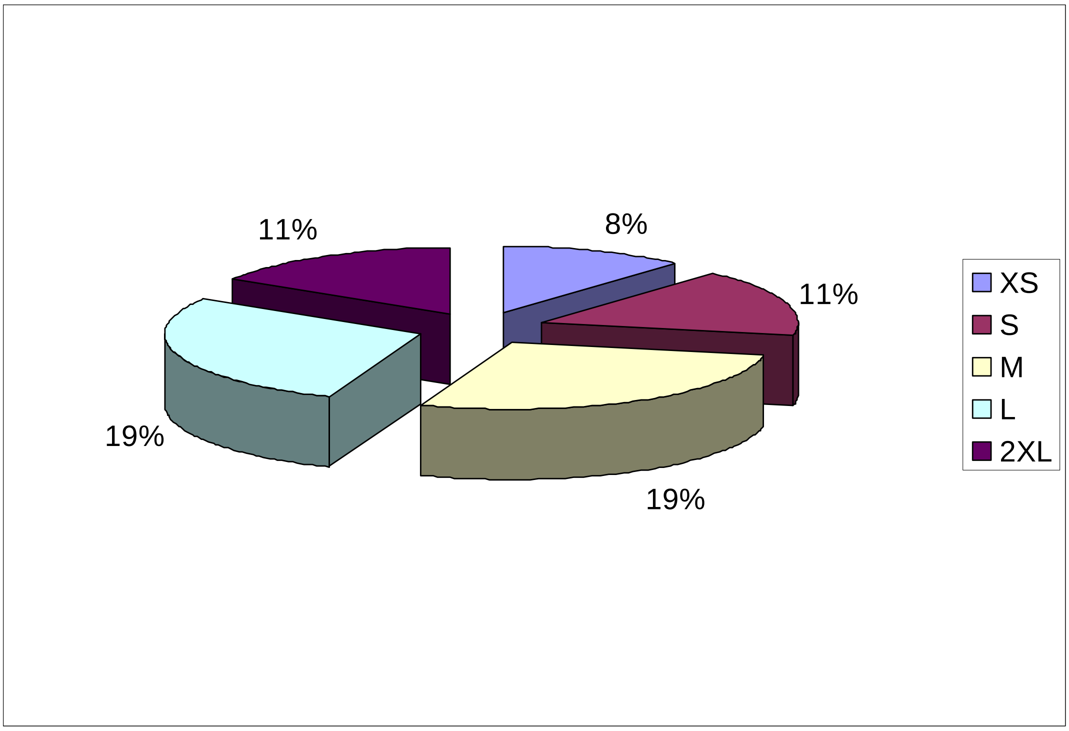


Figure (4.4): The Percentage of size categories for clusters 9

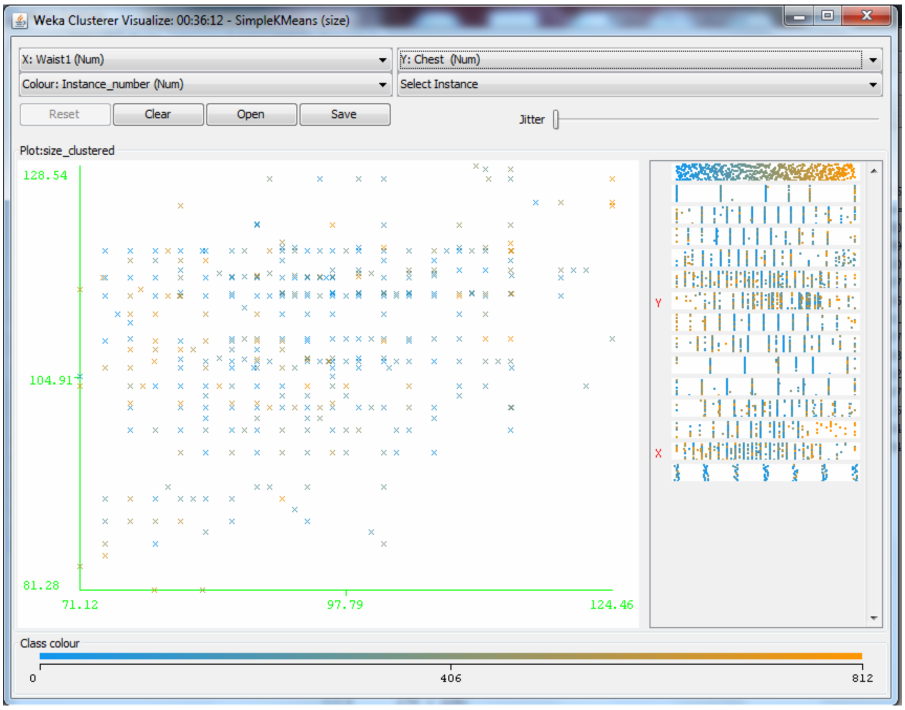


Figure (4.5): Visualization for cluster 7

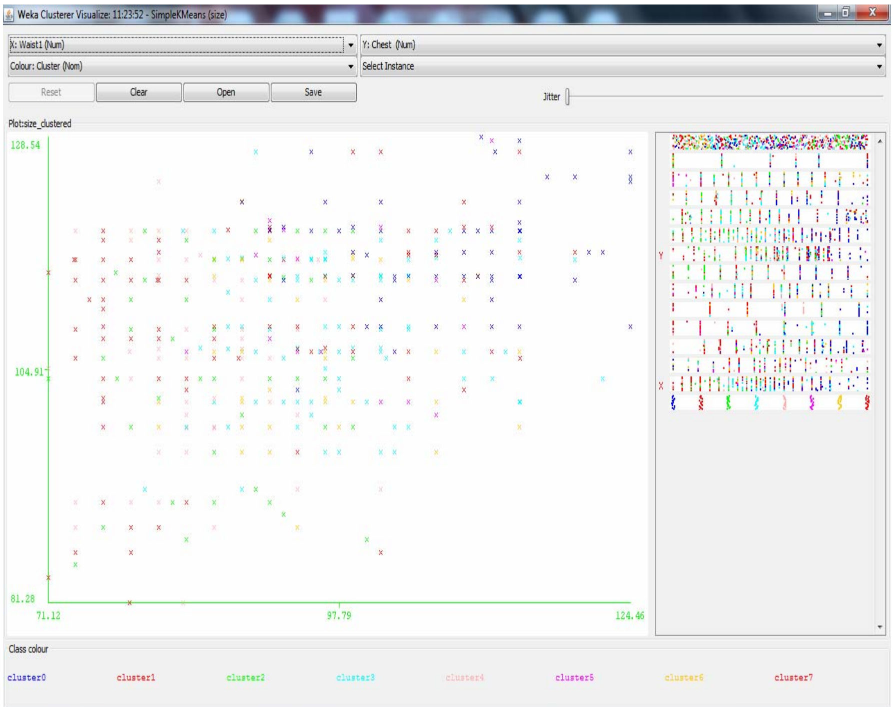


Figure (4.6): Visualization for cluster 8

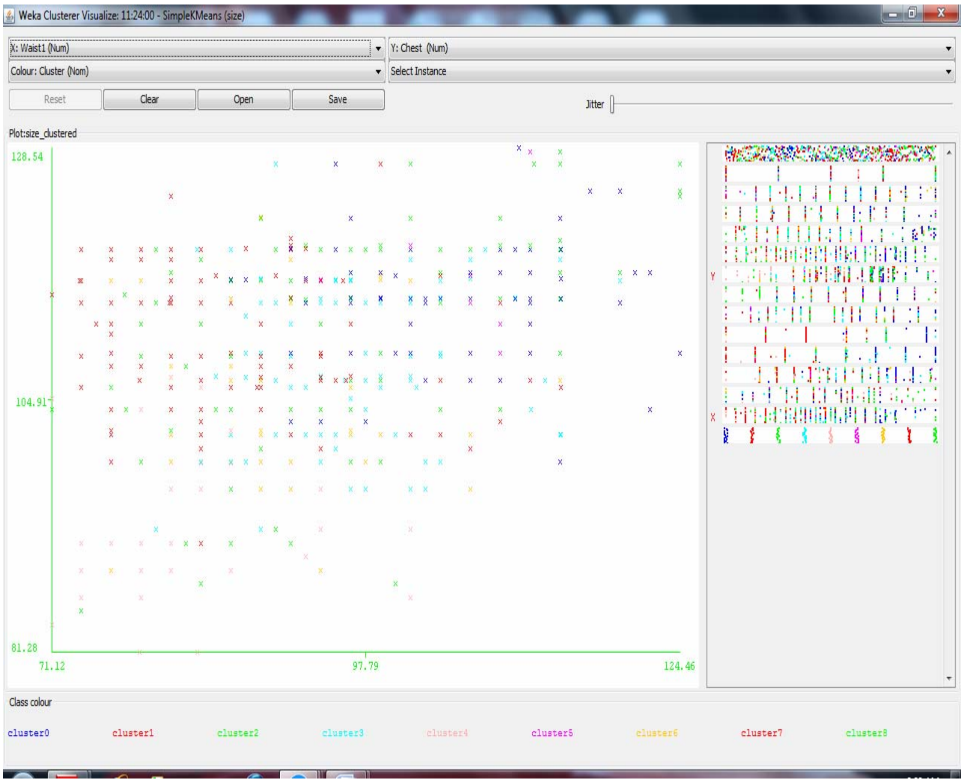


Figure (4.7): Visualization for cluster 9

4.3 Procedure for Establishing the New Sizing Systems:

The clustering model shows the centroid of each cluster and statistics on the number and percentage of instances assigned to different clusters. Cluster centroids are the mean vectors for each cluster; therefore, each dimension value and the centroid represent the mean value for that dimension in the cluster.

Therefore, centroids can be used to characterize the clusters. Table (4.1) shows the statistical analysis for the size ranges from the raw data.

Table (4.1) Statistical Analysis for the Size Ranges from the Raw Data:

Body Dimensions(cm)	Mean	Standard Error of Mean	Std. Deviation	Min	Max
Cuff	19.6	0.04	1.3	17.8	22.9
Sleeve	63.6	0.12	3.5	55.9	73.7
Shoulder	46.6	0.11	3	39.5	55.9
Length	78.2	0.07	4.8	67.3	92.8
Waist	96.5	0.41	11.6	71	124.5
Chest	109	0.3	8.5	81.3	128.5
Collar	41.9	0.1	2.9	34.3	49.8
Length 1	106.3	0.2	4.6	94	119
Foot	45.8	0.1	2.9	38.1	51.3
Knee	53.9	0.14	3.9	45.7	63.5
Thigh	75.2	0.2	5.7	58.4	89.7
Hip	105.2	0.21	6.0	86	124
Waist 1	94	0.37	10.7	71.1	124.5

n = 813 (all values are in centimeters) ___ jacket ___ trouser

4.3.1 Generated Clusters:

The new clusters generated using the WEKA was:

- Cluster 7 gives the size that represents these clusters and it shows the number of persons fall in these classes and the percentage from the total instances.
- Cluster 8 gives the size that represents these clusters and it shows the number of persons fall in these classes and the percentage from the total instances.

- Cluster 9 gives the size that represents these clusters and it shows the number of persons fall in these classes and the percentage from the total instances see appendix (A).

4.3.2 Selection of Appropriate Cluster:

In cluster 7 the percentage covered by the proposed sizing systems was 97%. However, the size figures 4XL and 5XL were not selected by the WEKA. This may be due to the fewer number of persons in these size figures. On the other hand, in cluster 8 the percentage covered by the proposed sizing system was 96.7 and there was only one size figure (5XL) with no classes that was not represented. This may be due to the fewer number of persons in this size figure. In cluster 9 the percentage covered by the proposed sizing system was 76.7, but there were four sizing figures XL, 3XL, 4XL and 5XL were not represented. This may be due to the fewer number of persons in these size figures.

As can be seen from the results, in the three clusters (7, 8, and 9) the size 5XL was omitted in all clusters. This may be attributed to the nature of the Sudanese male body shape.

Therefore, cluster 8 seems to be the best sizing system that represents the data collected from Sur factory. This is because it covered nearly 96.7% of the data and it includes 8 figure types. The new established sizing systems consists of the following sizing figures, XS, S, M, L, XL, 2XL, 3XL, and 4XL respectively.

4.4 The (SPSS) Results and Discussion:

The Statistical Package for the Social Sciences (SPSS) version 18.0 for windows was employed for an anthropometric data analysis. In this work data analysis were carried but by using K-means method to reduce the large samples in same groups contains similar number. The results were obtained for clusters 7 up to cluster 9. The iteration was 10 for all the clusters mentioned above. Descriptive statistics including Analysis of Variance (ANOVA) means square, standard error of mean, standard deviation, minimum and maximum values were calculated and utilized for the analysis and the determination of the correlations see table (4.1). The values were calculated in centimeters. All values of the standard deviation and others descriptive statistics are rounded to two decimals. Co-efficient correlation was used to determine the relationships between the body dimensions. The values used for the determination of the correlations between the dimensions and identifying the key parameters were based on BS 7231 (BSI, 1990). The standard specifies that, if correlation co-efficient is less than 0.5 then there is no relationship; if correlation co-efficient is between 0.6 – 0.76 then there is a moderate relationship; and if correlation co-efficient is more than 0.76 it shows a strong or high relationship.

The results of the cluster samples were analyzed statistically using (SPSS) version 18.0 package. The analysis results (mean and standard deviation) of the sample are tabulated in table (4.1).

4.5 Size Interval in a Size Chart:

Before the eleven figures types were classified it was necessary to select the size interval in order to get some flexibility for easy allowances during the seaming process. Kunick (1984) stated that, size interval is the division of sizes in a size chart. The BSEN 13402–3,(2004) recommended that an interval of 4 cm or 6 cm for both bust and waist1 and 4 cm or 5 cm for hip in order to have a flexible link between the bust, waist and hip. Beazley (1998) used 4 cm interval for the key dimensions (bust, waist and hip) for size 8 – 14 and 6 cm interval for size 16 to normalize the intervals.

Many British companies use 5 cm interval between all sizes, Aldrich,(2008). Kunick, (1984) proposed that 6 cm size interval is used by most countries. According to Winks, (1997), most countries applied 4 cm as the interval of chest girth. In this research, the intervals for the key anthropometric dimensions, chest, waist and hip were 4 cm. As mentioned before, from the total data (841), only (28) samples were excluded (outliers). Therefore, the percentage covered by the proposed sizing systems for the chest, waist and hip was 97%.

4.6 Establishment and Evaluation of the New Sizing System:

4.6.1 Establishing of the New Sizing System:

It would be a very tedious task using all 13 anthropometric variables to establish sizing systems. The development of the size chart was carried out by using values obtained from the statistical information of body dimensions. Winks (1997) states that, the mean value can be a convenient indication of obtaining central tendency. The mean values are the most widely used value for size steps. Table (4.1) shows the descriptive statistics for body dimension, and it is equivalent to the average size (mean) and also equivalent to size 12 of every size chart, Boakye S. et al,(2012). Eleven size steps approach was used to establish the new size chart as given in table (4.2). The eleven size steps used as a base for the determination of the outliers. The values that were less than the smallest size and those higher than the biggest size were eliminated and classified as outliers.

Table (4.2) Eleven Steps Size Ranges

	XXXXS M-5STD	XXXS M-4STD	XXS M-3STD	XS M-2STD	S M-1STD	M Mean	L Mean+1STD	XL M+2STD	XXL M+3STD	XXXL M+4STD	XXXXL M+5STD
Cuff	13.1	14.4	15.7	17	18.3	19.6	20.9	22.2	23.5	24.8	26.1
Sleeve	46.1	49.6	53.1	56.6	60.1	63.6	67.1	70.6	74.1	77.6	81.1
Shoulder	31.6	34.6	37.6	40.6	43.6	46.6	49.6	52.6	55.6	58.6	61.6
Length	54.2	59	63.8	68.6	73.4	78.2	83	87.8	92.6	97.4	102.2
Waist	38.7	50.3	61.9	73.5	85.1	96.7	108.3	119.9	131.5	143.1	154.7
Chest	66.5	75	83.5	92	100.5	109	117.5	126	134.5	143	151.5
Collar	18.9	30.3	33.2	36.1	39	41.9	44.8	47.7	50.6	53.5	64.9
Length 1	83.3	87.9	92.5	97.1	101.7	106.3	110.9	115.5	120.1	125	129.3
Foot	31.3	36	38.9	40	42.9	45.8	48.7	51.6	54.5	57.4	60.3
Knee	34.4	38.3	42.2	46.1	50	53.9	57.8	61.7	65.6	69.5	73.4
Thigh	46.7	52.6	58.2	63.8	69.5	75.2	80.9	86.6	92.2	97.8	103.7
Hip	75.2	81.2	87.2	93.2	99.2	105.2	111.2	117.2	123.2	129.2	135.2
Waist 1	40.5	51.3	61.9	72.6	83.3	94	104.7	115.4	126.1	136.8	147.5

n = 813 all values are in centimeters ___ jacket ___ trouser

As shown in table (4.2), in order to obtain eleven steps for eleven categories of body size, (1STD), (2STD), (3STD), (4STD) and (5STD) values were added to the mean and subtracted from the mean respectively. This was carried out in order to obtain five values that are higher and five values that are lower than the mean. According to Ashdown (1998) by subtracting (-1STD), (-2STD), (-3STD), (-4STD) and (-5STD) from the mean, the values obtained represents size 2, 4, 6, 8 and 10 respectively. When (1STD), (2STD), (3STD), (4STD) and (5STD) are added to the mean, the values obtained represent sizes 14, 16, 18, 20 and 22 respectively. The mean and standard deviation values were all rounded up to 0.1 decimals. Values above 0.15 were rounded up to 0.2 cm, and values below 0.15 have been reduced to 0.1 cm. This was done to make the comparison between SUR, EUR, and US size charts with the new established size charts more easy and understandable. To get the new established sizing charts from the eleven steps, the values of the 5XL figure size were omitted based on the results of the outlier from cluster 8, where there were no classes represented in this figure size. See table (4.2). Therefore, the new established sizing systems chart which consists of 8 figure size is given in table (4.3).

Table (4.3) the Proposed New Established Size System

	XS M-2STD	**S M-1STD**	**M Mean**	**L Mean+1STD**	**XL M+2STD**	**XXL M+3STD**	**XXXL M+4STD**	**XXXXL M+5STD**
Cuff	17	18.3	19.6	20.9	22.2	23.5	24.8	26.1
Sleeve	56.6	60.1	63.6	67.1	70.6	74.1	77.6	81.1
Shoulder	40.6	43.6	46.6	49.6	52.6	55.6	58.6	61.6
Length	68.6	73.4	78.2	83	87.8	92.6	97.4	102.2
Waist	73.5	85.1	96.7	108.3	119.9	131.5	143.1	154.7
Chest	92	100.5	109	117.5	126	134.5	143	151.5
Collar	36.1	39	41.9	44.8	47.7	50.6	53.5	64.9
Length 1	97.1	101.7	106.3	110.9	115.5	120.1	125	129.3
Foot	40	42.9	45.8	48.7	51.6	54.5	57.4	60.3
Knee	46.1	50	53.9	57.8	61.7	65.6	69.5	73.4
Thigh	63.8	69.5	75.2	80.9	86.6	92.2	97.8	103.7
Hip	93.2	99.2	105.2	111.2	117.2	123.2	129.2	135.2
Waist 1	72.6	83.3	94	104.7	115.4	126.1	136.8	147.5

n = 813 all values are in centimeters ___ jacket ___ trouser

4.6.2 Evaluation of the New Established Sizing Systems:

As stated by the International Standards Organization (ISO), and those others Organizations mentioned in the literature; “the use of control dimensions and size interval can effectively facilitate to recognize the parameters for developing sizing systems" Winks, (1997).

After the eleven figure sizes were classified by the WEKA and SPSS soft wares, the new established size system of the eight figures size were determined and the results are given in table (4.3). Figure (4.8) shows the relevant scatter plots of chest on the X-axis verse the waist on the Y-axis and the interval was 4 cm to demonstrate the distribution of all figures type. It has been reported that, Cooklin, (1992) the chest is the most important anthropometric variable in establishing sizing systems in the field of garment making. The waist is also an important variable for sizing male garments in many countries. Figure (4.9) illustrates the differences between the eight types for the new established sizing systems. The figure was plotted as a line graph to yield a better insight into the differences between the new established sizing systems. The eight figures types are exhibited by clear differences in chest and waist. The eight figure types also follow the order, SX, S, M, L, XL, XXL, XXXL and XXXXL. In addition, figure (4.9) shows the differences between the eight figures types mentioned above.

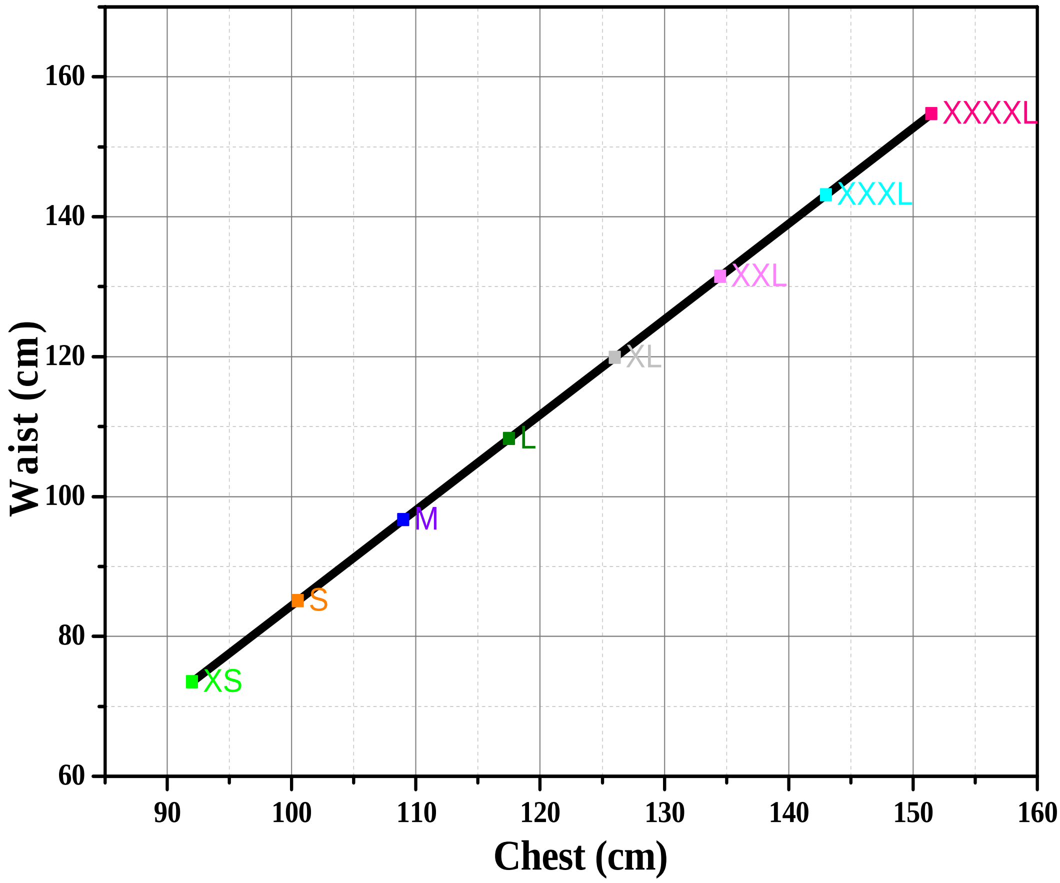


Figure (4.8): Scatter plot of chest verse waist for the proposed new established size system

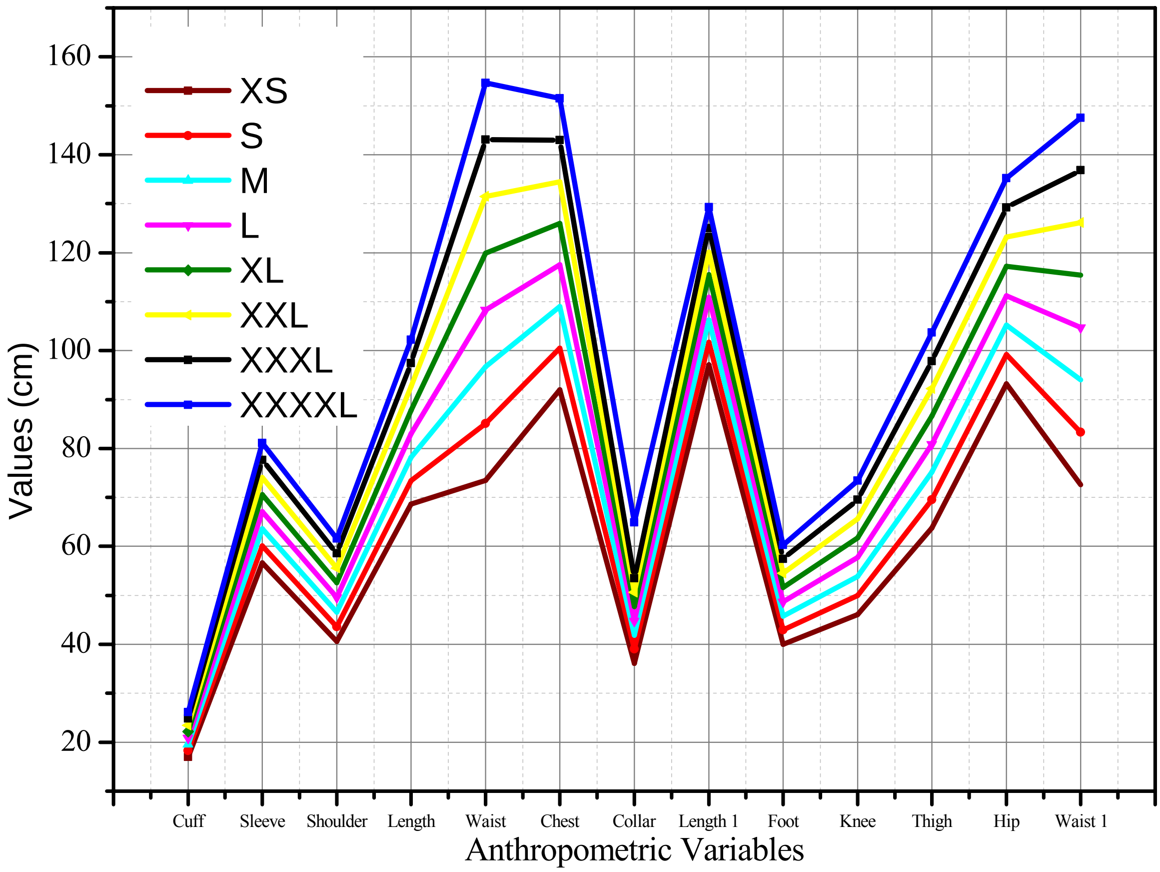


Figure (4.9): The distinct anthropometric variables between the proposed new established size systems

4.7 The Size Code:

The size codes are based on the numerical coding methods. In this research the size codes were determined after generating the eight size steps values from the figure sizes, as given in table (4.3).

The new size code was termed (SUD) abbreviated for Sudan, and then the code was added to establish the 8 size figures for Sudanese army officers (poshirt uniform). Therefore, the new size codes are, SUD XS,

SUD S, SUD M, SUD L, SUD XL, SUD XXL, SUD XXXL and SUD XXXXL. See table (4.4).

Table (4.4) Size codes for Sudanese officers (poshirt)

	SUD XS	SUD S	SUD M	SUD L	SUD XL	SUD XXL	SUD XXXL	SUD XXXXL
Cuff	17	18.3	19.6	20.9	22.2	23.5	24.8	26.1
Sleeve	56.6	60.1	63.6	67.1	70.6	74.1	77.6	81.1
Shoulder	40.6	43.6	46.6	49.6	52.6	55.6	58.6	61.6
Length	68.6	73.4	78.2	83	87.8	92.6	97.4	102.2
Waist	73.5	85.1	96.7	108.3	119.9	131.5	143.1	154.7
Chest	92	100.5	109	117.5	126	134.5	143	151.5
Collar	36.1	39	41.9	44.8	47.7	50.6	53.5	64.9
Length 1	97.1	101.7	106.3	110.9	115.5	120.1	125	129.3
Foot	40	42.9	45.8	48.7	51.6	54.5	57.4	60.3
Knee	46.1	50	53.9	57.8	61.7	65.6	69.5	73.4
Thigh	63.8	69.5	75.2	80.9	86.6	92.2	97.8	103.7
Hip	93.2	99.2	105.2	111.2	117.2	123.2	129.2	135.2
Waist 1	72.6	83.3	94	104.7	115.4	126.1	136.8	147.5

n = 813 all values are in centimeters ___ jacket ___ trouser

4.8 Lower and Upper Limits of Sizes

In order to establish the limits of each size and demonstrate the extent of coverage for inter size ranges one of the important step to determine the lower and upper limits of sizes. The value obtained for each size code is used as a midpoint and the lower and upper limit are determined from it. The half value of the standard deviation of each body dimension added or subtracted to the midpoint value. A value of 0.01 is subtracted from the figure obtained below the midpoint to create limits between the lower value of the next size and the upper value of the previous one. To avoid overlapping of figures with the next size value of 0.01 is subtracted from the upper limit making it less than the next value. Beazley, (1998) and other researchers used this procedure. In order to know what percentages of the populations are covered by each size we need to establish lower and upper limits, table (4.5) tabulates and presents the lower and upper limits of the body dimensions.

Table (4.5) Lower and Upper Limit of Size Code

	SUDXS M-2STD	**SUDS Mean-1STD**	**SUDM Mean**	**SUDL Mean+ 1STD**	**SUD XL M+ 2STD**	**SUD XXL M+ 3STD**	**SUD XXXL M+ 4STD**	**SUD XXXXL M+ 5STD**
Cuff	16.35	17.65	18.95	20.25	21.55	22.85	24.15	25.45
	17	**18.3**	**19.6**	**20.9**	**22.2**	**23.5**	**24.8**	**26.1**
	17.64	18.94	20.24	21.54	22.84	24.14	25.44	26.74
Sleeve	54.85	58.35	61.85	65.35	68.85	72.35	75.85	79.35
	56.6	**60.1**	**63.6**	**67.1**	**70.6**	**74.1**	**77.6**	**81.1**
	58.34	61.84	65.34	68.84	72.34	75.84	79.34	82.40
Shoulder	39.10	42.10	45.10	48.10	51.10	54.10	57.10	61.10
	40.6	**43.6**	**46.6**	**49.6**	**52.6**	**55.6**	**58.6**	**61.6**
	42.09	45.09	48.09	51.09	54.09	57.09	60.09	63.09
Length	66.20	71.00	75.80	80.60	85.40	91.20	95.00	99.80
	68.6	**73.4**	**78.2**	**83**	**87.8**	**92.6**	**97.4**	**102.2**
	70.99	75.79	80.59	85.39	91.19	94.99	99.79	104.59
Waist	67.70	79.30	90.90	102.50	114.10	125.70	137.30	148.90
	73.5	**85.1**	**96.7**	**108.3**	**119.9**	**131.5**	**143.1**	**154.7**
	79.29	90.89	102.49	114.09	125.69	137.29	148.89	160.49
Chest	87.75	96.25	104.75	113.25	121.75	130.25	138.75	147.25
	92	**100.5**	**109**	**117.5**	**126**	**134.5**	**143**	**151.5**
	96.24	104.74	113.24	121.74	130.24	138.74	147.24	155.74
Collar	34.65	37.55	40.45	43.35	46.25	49.15	52.05	54.95
	36.1	**39**	**41.9**	**44.8**	**47.7**	**50.6**	**53.5**	**64.9**
	37.54	40.44	43.34	46.24	49.14	52.04	54.94	67.69
Length 1	74.50	99.40	104.00	108.60	113.20	117.80	124.40	127.80
	97.1	**101.7**	**106.3**	**110.9**	**115.5**	**120.1**	**125**	**129.3**
	99.39	103.99	108.59	113.19	117.79	122.39	127.79	131.59
Foot	40.35	41.45	44.35	47.25	50.15	53.05	55.95	58.85
	40	**42.9**	**45.8**	**48.7**	**51.6**	**54.5**	**57.4**	**60.3**
	41.44	44.34	47.24	50.14	53.04	55.94	58.84	61.74
Knee	44.15	48.05	51.95	55.85	59.75	63.75	67.55	71.45
	46.1	**50**	**53.9**	**57.8**	**61.7**	**65.6**	**69.5**	**73.4**
	48.04	51.94	55.84	59.74	63.74	67.54	71.44	75.34
Thigh	60.95	66.65	72.35	78.05	83.75	89.35	95.95	100.65
	63.8	**69.5**	**75.2**	**80.9**	**86.6**	**92.2**	**97.8**	**103.7**
	66.64	72.34	78.04	83.74	89.44	95.04	100.64	106.54
Hip	90.20	96.20	102.20	108.20	114.20	120.20	126.20	132.20
	93.2	**99.2**	**105.2**	**111.2**	**117.2**	**123.2**	**129.2**	**135.2**
	96.19	102.19	108.19	114.19	120.19	126.19	132.19	138.19
Waist 1	67.25	77.95	88.65	99.35	110.05	127.75	131.45	142.15
	72.6	**83.3**	**94**	**104.7**	**115.4**	**126.1**	**136.8**	**147.5**
	77.94	88.64	99.34	110.04	127.74	131.44	142.14	152.84

n = 813 all values are in centimeters ___ jacket ___ trouser

4.9 Application of the New System for Establishment of Garment Measurements:

The new established size systems need to be checked to find out its accuracy. In order to verify the size chart, garments measurements should be established for the preparation of the patterns and subsequently the garment for trials. For establishing the garment measurement, ease allowance was added to each body dimension on the established size chart. By using measurements information from the new established size system, patterns were constructed manually. The basic blocks constructed were grade in all figure sizes indicated, i.e. SUD XS, SUD S, SUD M, SUD L, SUD XL, SUD 2XL, SUD 3XL and SUD 4XL. As mentioned earlier, the basic block for the base size M was decreased four steps down and increased four steps up to obtain the other sizes.

4.10 Validation of the New Established Size System:

Because of the restrictions given to Sudanese army uniforms, private communication was made with Sur Military Clothing Factory manager to find out the possibility of preparing a marker and mini-marker for poshirt uniform (jacket + trouser) for only one size (SUD XL) based on the established sizing systems and comparing it with (SUR XL). A Marker is a thin paper on which all patterns pieces for all sizes for a particular style of garments are drawn before cutting as shown on figures 4.10 and figure 4.11 below in a min marker form. The results revealed that the proposed new standard sizing SUD XL in a marker and min-marker will be acceptable compared with SUR XL marker and mini-marker.

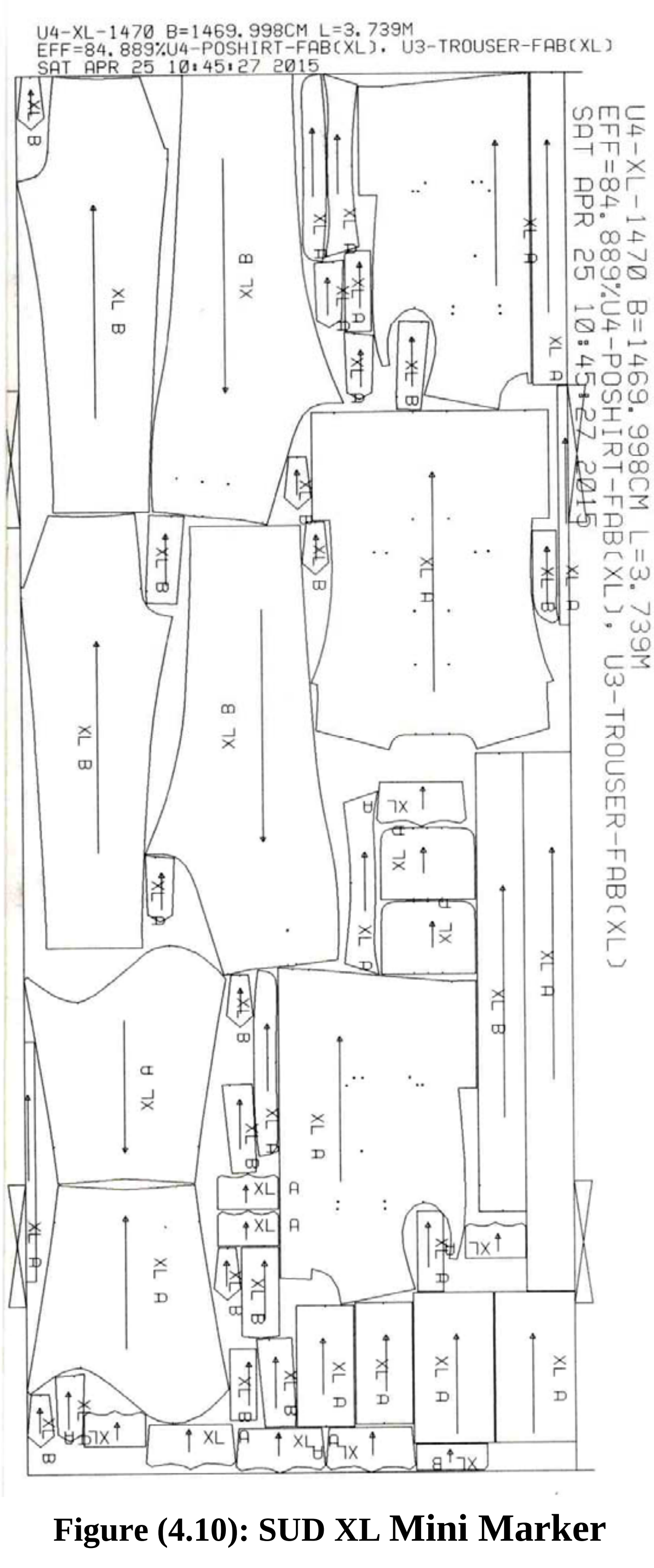


Figure (4.10): SUD XL Mini Marker

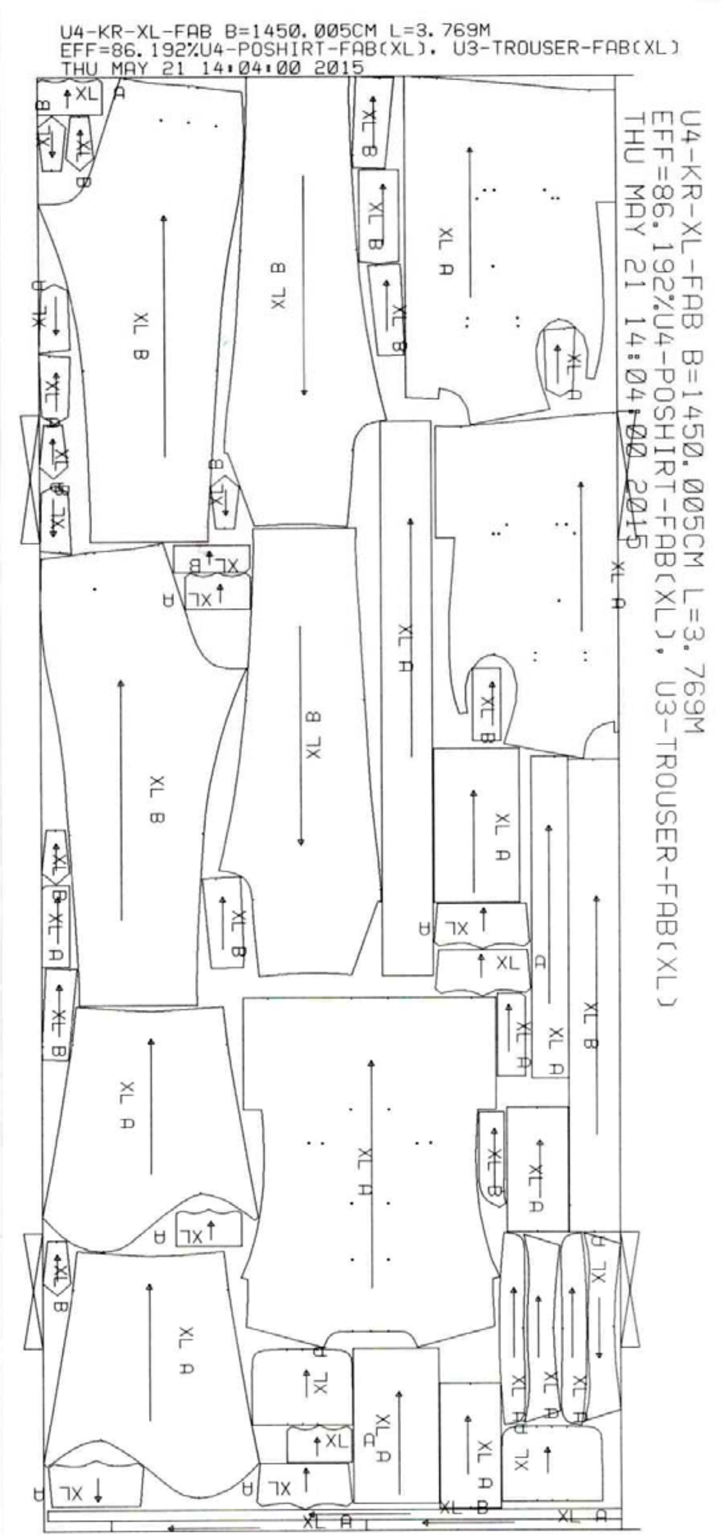


Figure (4.11): SUR XL Mini Marker

4.11 A Comparison between the New Established Sizing Systems SUD and SUR, USA, EUR Sizing Systems Specifications:

In order to compare the new established sizing systems with others sizing systems, five figure types; SX, S, M, L and XL, and five variables; chest, waist, hip, collar and sleeves were chosen from the eight figure types, see table (4.6). As can be seen from tables (4.6, 4.7, and 4.8) and figures (4.12, 4.14, 4.16 and 4.18), the new size system (SUD) follows the size chart for the three sizing systems (SUR, USA and EUR). Therefore, the scatter plot of chest verse waist for the new size system will follow the same trend that was plotted in figure (4.14). For the new established size figure (SUD) figure (4.14) plots a distribution graph of chest on the X-axis verse waist on the Y-axis to demonstrate the distribution of all five figure types.

Figure (4.13) compared with figures (4.15, 4.17 and 4.19) for (SUR, USA and EUR) respectively. It can be seen that in figure (4.20) for both the new established sizing systems (SUD and SUR) the plot of chest verse waist fall in straight line. However, when compared (SUD,SUR,USA and EUR) as shown in figure (21) in the USA size systems there was a deviation in the line for the (M and L) figure types. Also there was a deviation in the line for the (S and M) figure types in the case of the EUR size systems.

Figures 4.13, 4.15, 4.17 and 4.19 show the differences between the five figure types for the new established size systems (SUD, SUR, USA and EUR) respectively. All the sizing systems follow the same profile, but there was slight differences between the new established systems SUD and the others sizing systems as shown in figure (22).

This may be due to:

a) The national standards (SUR + USA + EUR) deal essentially with the size designation of clothing and are not directly concerned with the sizing systems as such. In other words, the establishment of a size designation system that indicates the body size of a person that a garment is intended to fit. Therefore, the size designation system is based on body and not garment measurements. Choice of garments measurements is normally left to the garment designer and manufacture, who are concerned with style, cut and other fashion elements, and he must make due allowance for garments normally worn beneath a specific garment.

b) The new established sizing systems deals with clothing designer and manufactures where the measurements are taken manually and for every individual. Therefore, the measurements may not be accurate. The accuracy of measurements depends on the skill of the person who takes the measurements. Also this difference may be due to the nature of the Sudanese body size compared with other nationalities.

Table (4.6): Five Steps Size Ranges Obtained from the New Established Sizing Systems

Size Body Dimensions (cm)	**XS Mean -2STD**	**S Mean -1STD**	**M Mean**	**L Mean +1STD**	**XL Mean +2STD**
Chest	92	100.5	109	117.5	126
Waist	73.5	85.1	96.7	108.3	119.9
Hip	93.2	99.2	105.2	111.2	117.2
Collar	36.1	39	41.9	44.8	47.7
Sleeve	56.6	60.1	63.6	67.1	70.6

n = 813   all values are in centimeters

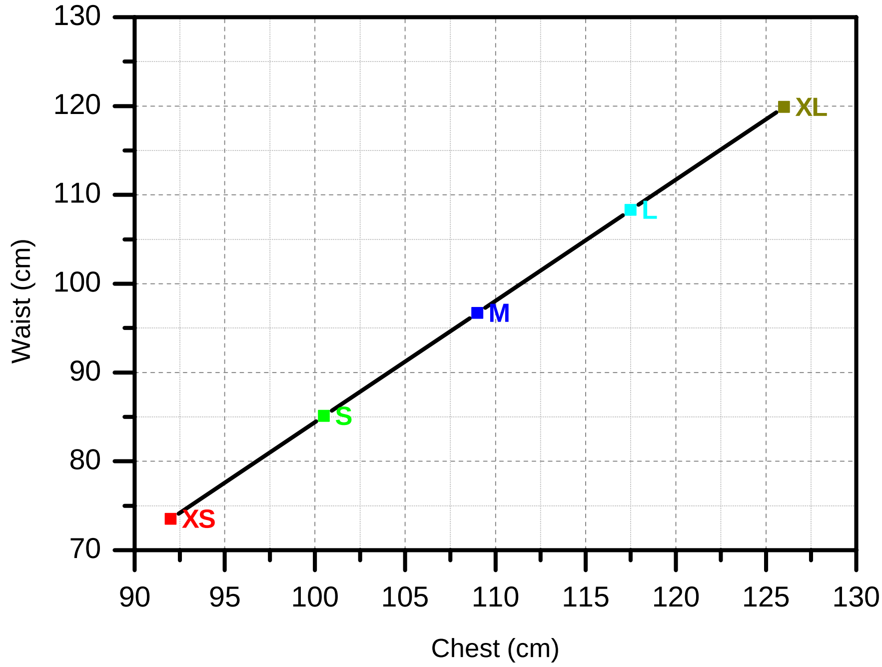


Figure (4.12): Scatter plot of chest verse waist for 5 figure types for the new established sizing systems

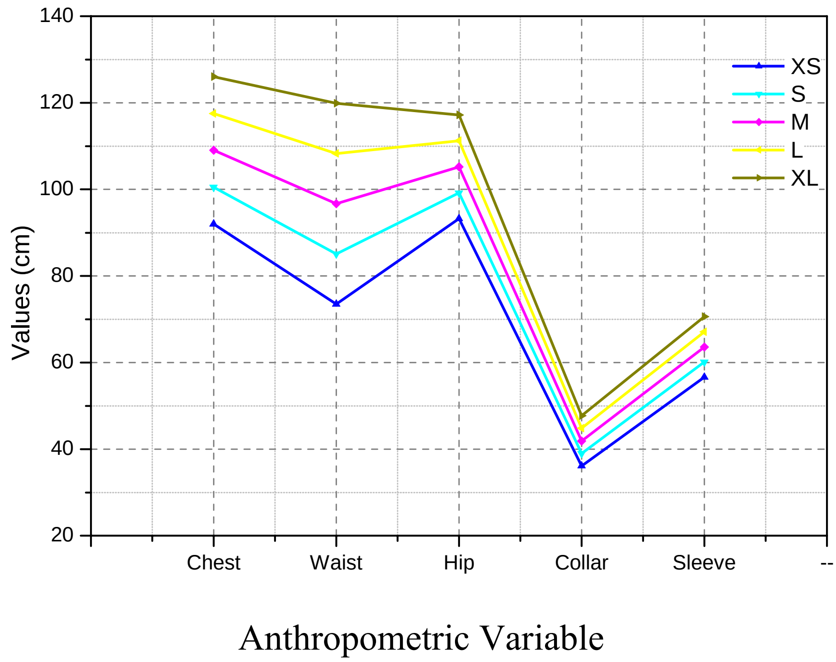


Figure (4.13): New established sizing system 5 figure types and corresponding anthropometric variables

☐

Table (4.7): SUR Military Clothing Factory Poshirt (U4) Size Specification

Size	XS (cm)	S (cm)	M (cm)	L (cm)	XL (cm)
Chest	104.1	106.7	109.2	111.8	114.3
Waist	84	89	94	99	104
Hip	99.8	104.8	109.8	114.8	119.8
Collar	36.8	38.1	39.4	40.6	41.9
Sleeve	61	62.2	63.5	64.8	66

All values are in centimeters

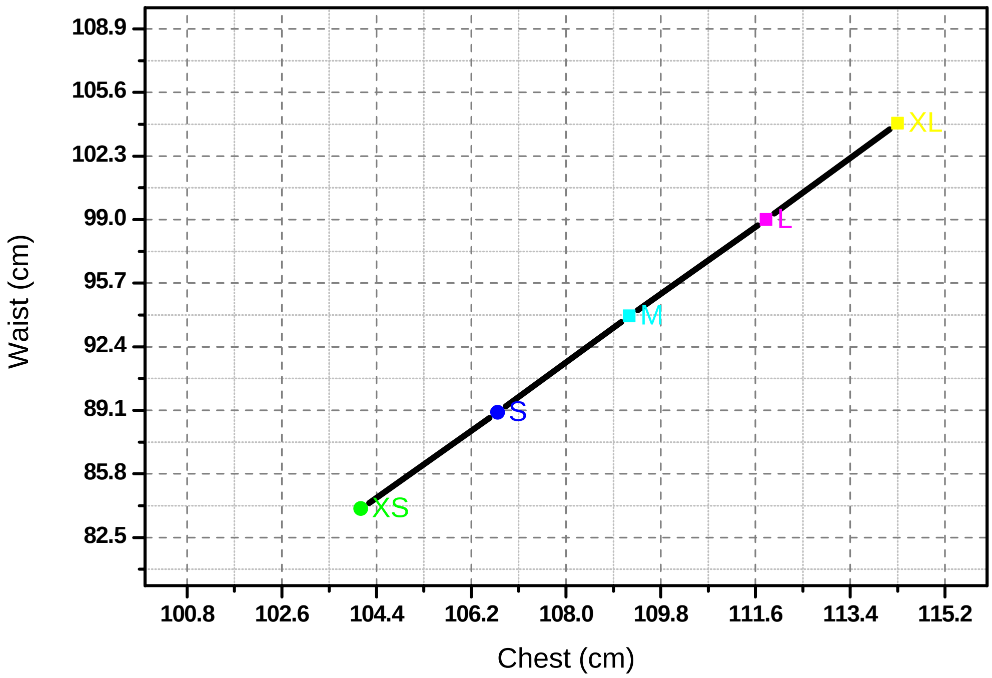


Figure (4.14): Scatter plot of chest verse waist for 5 figure types for SUR size systems

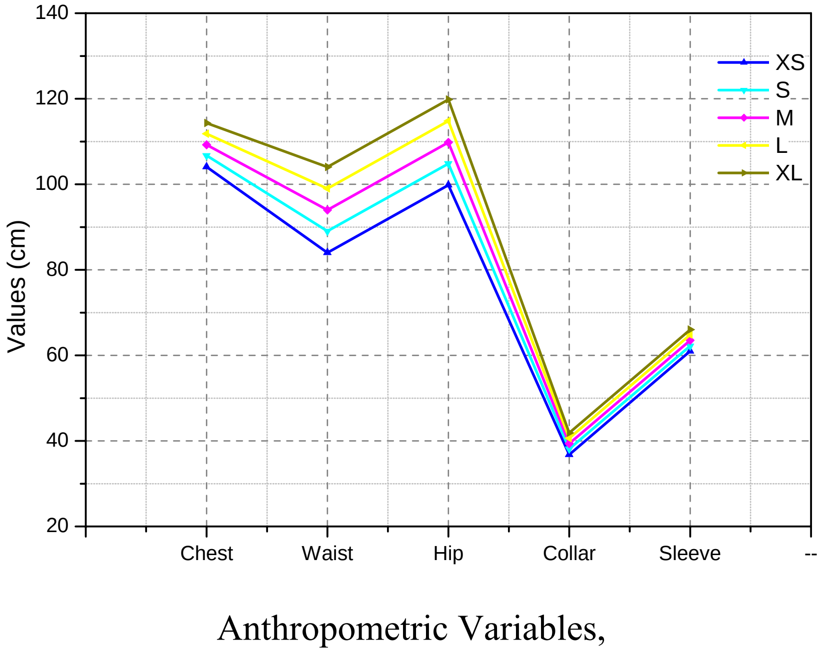


Figure (4.15): SUR 5 figure types and corresponding anthropometric variables

Table (4.8): USA size charts specification for (men)

Size	XS (cm)	S (cm)	M (cm)	L (cm)	XL (cm)
Chest	96.5	101.6	106.7	111.8	116.8
Waist	81.3	86.4	91.4	94	106.7
Hip	88.9	104.1	109.4	114.3	119.4
Neckband (collar)	38.1	39.4	40.6	41.9	43.2
Sleeve	83.8	83.8	86.4	86.4	88.9

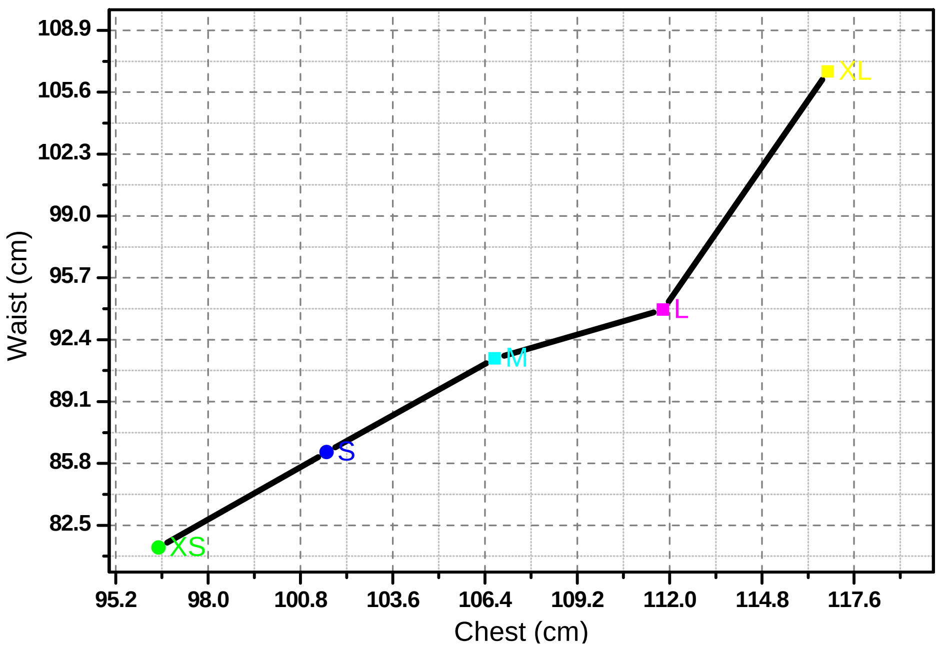


Figure (4.16): Scatter plot of chest verse waist for 5 figure types for USA size systems

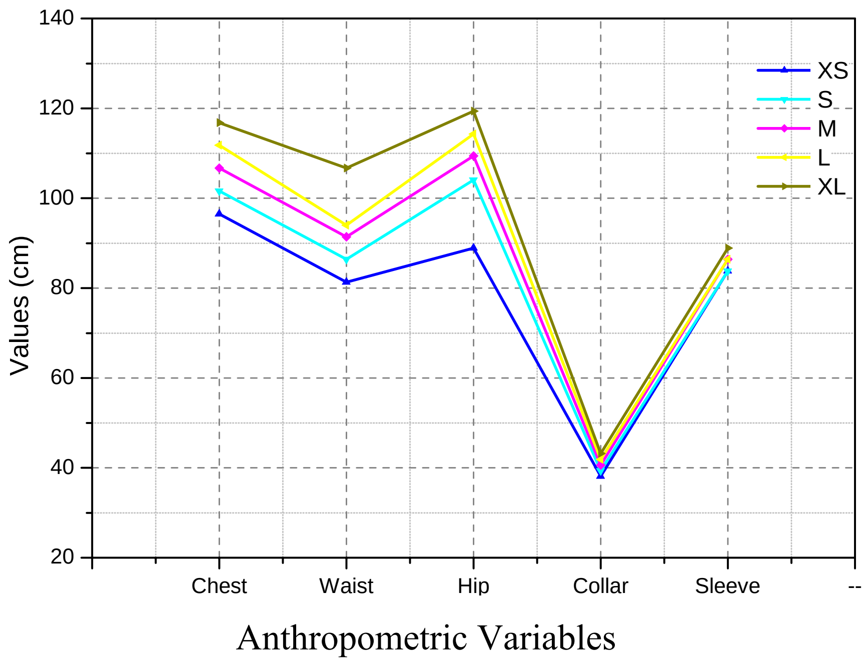


Figure (4.17): USA 5 figure types and corresponding anthropometric variables

Table (4.9): EUR size charts specification for (men)

Size	XS (cm)	S (cm)	M (cm)	L (cm)	XL (cm)
Chest	92	97	107	112	117
Waist	81	87	92	99	107
Hip	99	104	109	114	119
Neckband Collar	38	39	40	42	43
Sleeve	84	84	87	87	89

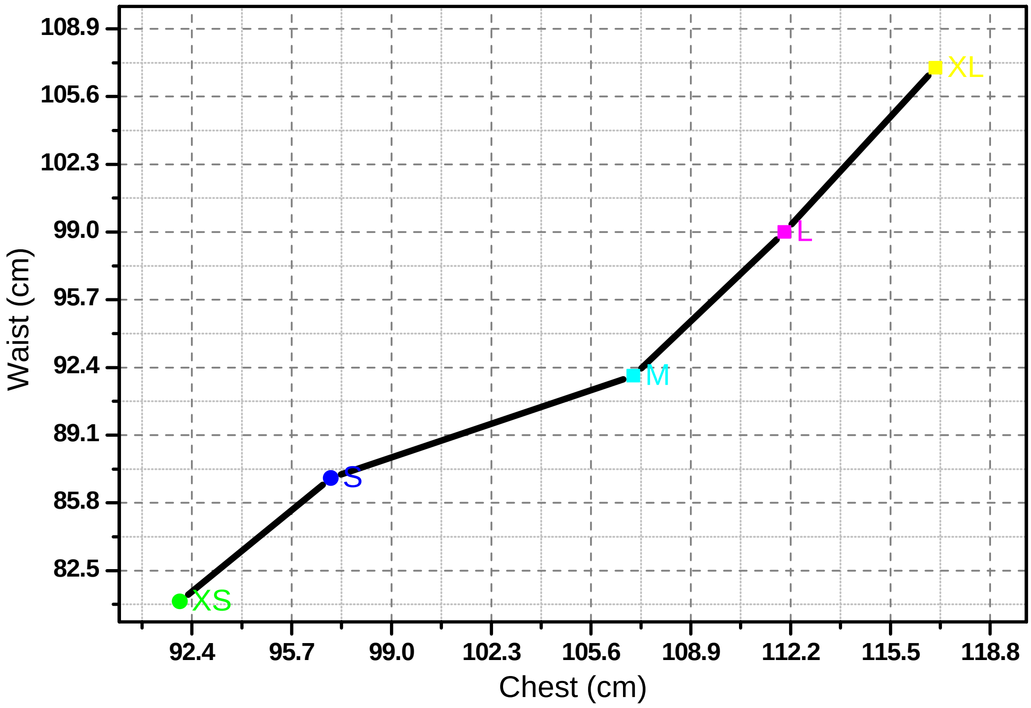


Figure (4.18): Scatter plot of chest verse waist for 5 figure types for EUR size systems

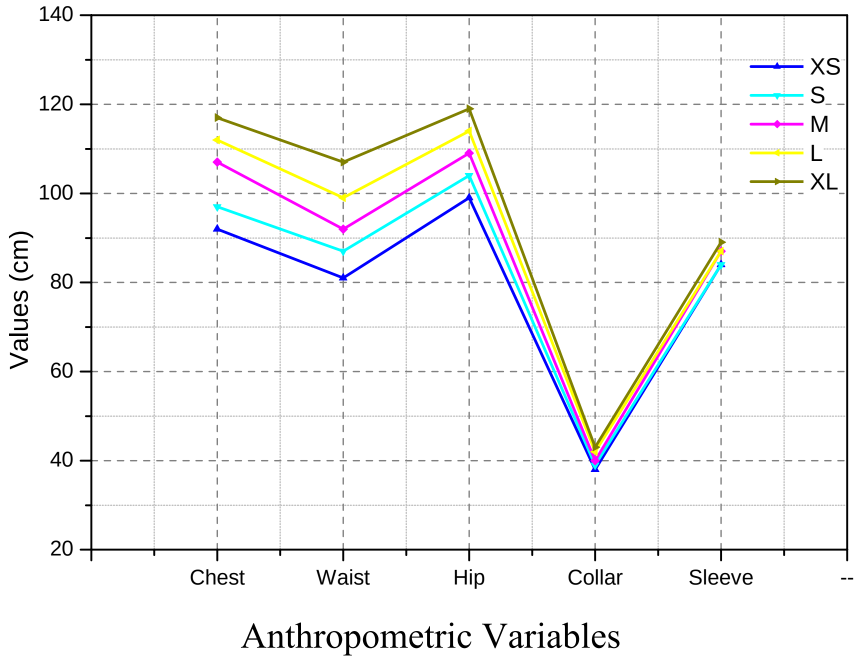


Figure (4.19): EUR 5 figure types and corresponding anthropometric variables

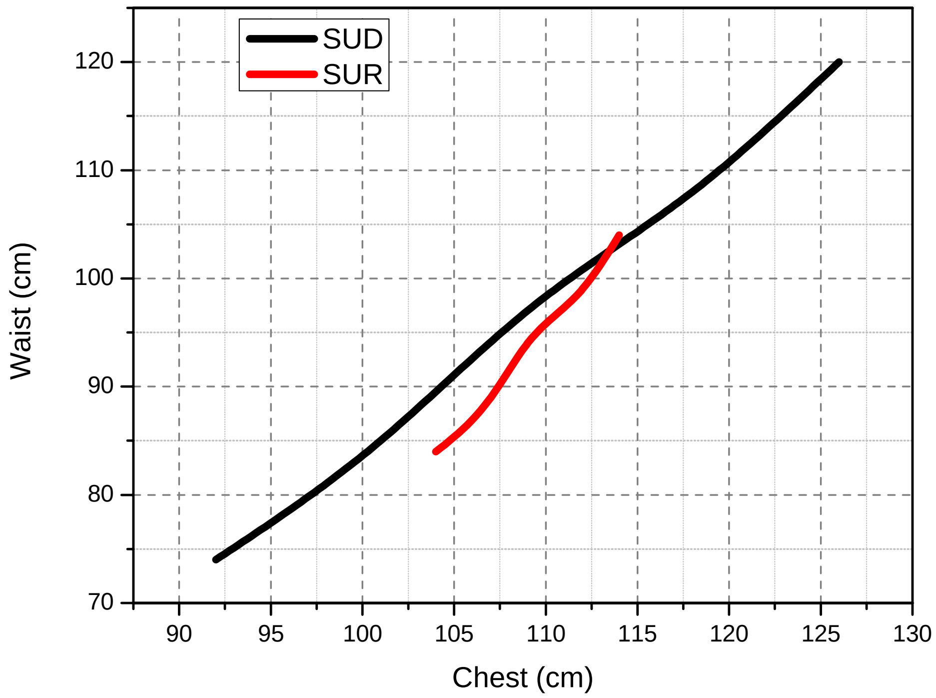


Figure (4.20): The New SUD and SUR scatter plot of chest verse waist for 5 figure types

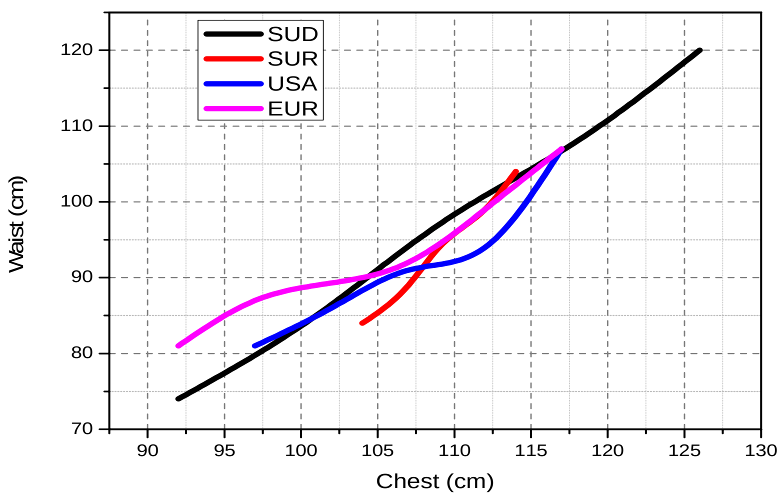
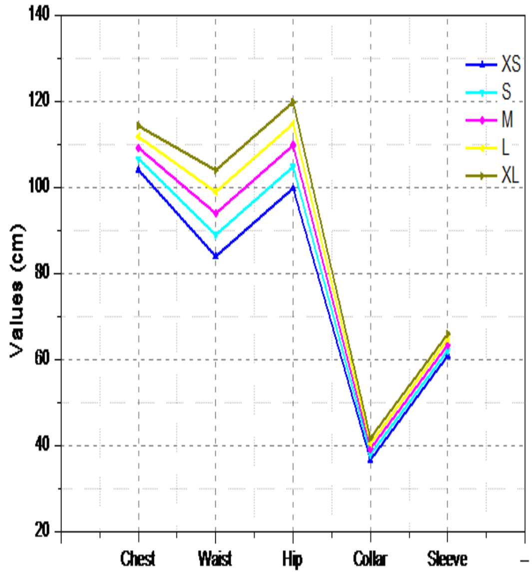
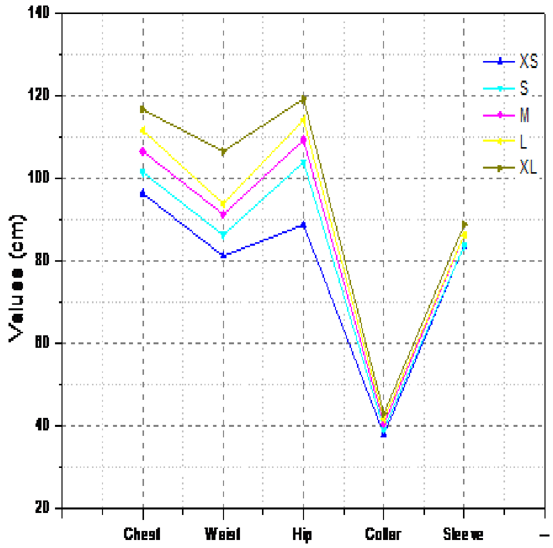


Figure (4.21): The sum of New SUD, SUR, EUR and USA scatter plot of chest verse waist for 5 figure types



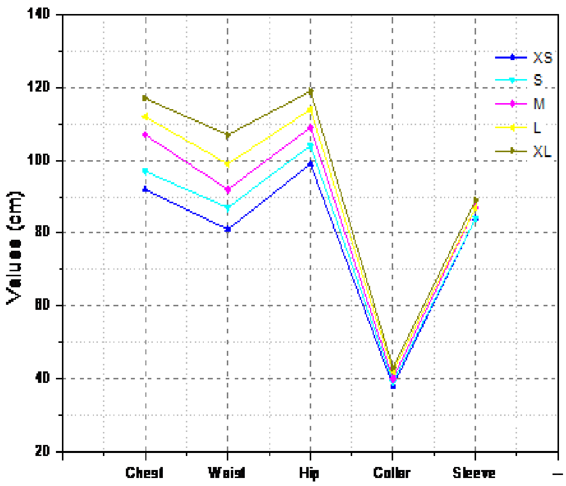
Anthropometric Variables

SUR 5 figure types and corresponding anthropometric variables



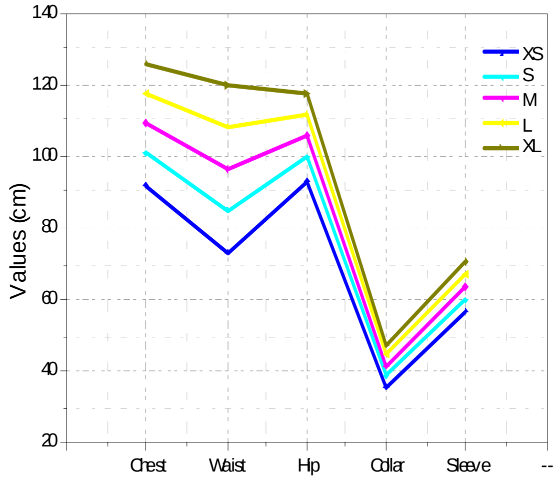
Anthropometric Variables

USA 5 figure types and corresponding anthropometric variables



Anthropometric Variables

EUR 5 figure types and corresponding anthropometric variables



Anthropometric Variables

The new SUD established sizing systems 5 figure types and corresponding anthropometric variables

Figure (5.22): The sum of new SUD, SUR, EUR and USA 5 figure types and corresponding anthropometric variables

CHAPTER FIVE

CONCLUSION AND RECOMMENDATIONS

CHAPTER FIVE

Conclusion and Recommendations

5.1 Conclusion:

In this study, data mining methods (WEKA and SPSS) were applied in order to establish sizing systems for Sudanese army officers uniform (poshirt). The study used database obtained from Sur Military Clothing Factory in Sudan. Anthropometric data for (841) officers was used. For each individual (13) anthropometric variables were involved resulting in a total of 10933 variables. Large amounts of data were analyzed by applying the (WEKA and SPSS) methods, and identify systematic patterns in bodily dimensions. Based on these patterns, the representative figure types of Sudanese army officers were clustered and classified, and then standard sizing systems was established. The army officers' database was selected for establishing sizing systems because of the urgent need for accurate sizing systems for producing army officer's uniforms (poshirt).

The WEKA and SPSS methods were used for clustering and establishing sizing system by implementing simple K-means algorithm to determine the final cluster classification.

Cluster analysis using chest and waist as a control anthropometric variables reveal eight distinct clusters. Each cluster showed distinct difference between clusters but similar within each cluster.

The study also showed that within the army officers aged 16-60 years there were eight types of body shapes namely; XS, S, M, L, XL, XXL, XXXL and XXXXL respectively.

The size codes were based on numerical coding method abbreviated from Sudan (SUD) and added to the sizes obtained. Therefore the size codes obtained are, SUD XS, SUD S, SUD M, SUD L, SUD XL, SUD XXL, SUD XXXL and SUD XXXXL respectively.

The value obtained for each size code is used as the midway point for determining the lower and upper limit which helps in establishing the limit of each size in order to demonstrate the extent of coverage for inter size ranges.

The total applicability of the new established sizing system was 96.7% which is very high coverage.

Scatter plots of chest on the X-axis; verses waist on the Y-axis, using an interval of 4 cm demonstrates the distribution of all figure types of the established sizing systems.

A line graph was plotted and it shows significant differences in chest and waist among the eight figure types of the new proposed established sizing systems. This was appear very clearly when New SUD compared with national SUR, USA, and EUR size systems as shown on (Figure: 4.21)

The establishment of garment measurement for the new proposed established sizing systems was obtained by adding ease allowance to each size figure type.

The established sizing systems were compared with three standard national sizing charts such as, USA, EUR and SUR. The results revealed that the new established sizing systems follow approximately the same profile as was shown in (Figure 4.22) the plotted graphs for chest and waist anthropometric variables except the variation on the waist of the new established SUD.

With clustering technique it was possible to discover the differences in body shapes that exited among army officers poshirt uniform in Sudan.

Therefore it is important to consider their body shapes differences in garments production.

From the reviewed literature it was clear that the rule of classification of the whole body types (upper and lower) has been investigated in a few works before. Therefore the present research looked into this topic. This was done because it helps the garment manufacturer to better understand the customer's body shape characteristics and therefore planning for production accordingly.

Moreover, this work provide an effective procedure of extracting of the significant rules to categorize human body type for many applications, such as clothing industry, physiology, medical treatment sports etc.

The percentage of army officers who fall in a certain figure type and sizes can serve as a good reference to indicate the quantity of garments to be produced for specific market. Thus a realistic plan for producing male army officer’s uniforms can be established.

Manufacturing of clothing can produce different types of garments with pre-determined quantities, and their judgments could base on the new proposed sizing systems.

This research contributes largely to knowledge of size chart by providing a detailed procedure involved in establishing standard sizing system based on anthropometric variables and will serve as the basis for other future research in garments industry in Sudan.

5.2 Recommendations for Further Work:

- This study aimed to establish a sizing system for army officer’s uniform (poshirt).
- Further anthropometric studies needed to cover army soldiers and other civil population.
- Others anthropometric variables such as body height and body weight could be used for establishing sizing system.
- It would be of large interesting if different methodologies such as body scanning technology, and the developed software for Female Figure Identification Technique (FFIT) for apparel could be used for establishing sizing systems.
- The established sizing systems should make garments standards accessible to all garments manufactures and ensure that all garments items producers are based on these standards.
- Further studies are recommended in order to establish a Sudanese sizing standards for Sudanese garments industry.

REFERENCES

REFERENCES

(1) P .Lyman and H. R. Varian How Much Information? http://www2.sims.berkeley.edu/ research/ projects/how – much-info/, 2003 (valid as of May 2008)

(2) Codd, E.F. (1993). Providing OLAP (On-line Analytical Processing) to User-Analysis an IT Mandate. E. F. Codd and Associates.

(3) Fayyad, U. M., Piatetsky – Shapiro, G., and Smyth, P., R. (Eds.), Advances in knowledge Discovery and Data Mining. Cambridge, M. A: AAA1 Press/MIT Press, 1996.

(4) Berry, M. J., and Linoff, G., Data Mining Techniques. New York: John Wiley and Sons,(1997).

(5) http://www.anderson.ucla.edu/faculty/jason.frand/teacher/ technologies/pala…11/3/2011

(6) http:// translate.googleusercontent.com/translate_c?h1=en & lang pair=es%7c… 27/04/2010 Data Mining – Wikipedia, the free encyclopedia.

(7) G.Shakhnarovish, T. Darrell and P. Indyk, 2005, "Nearest Neighbor Methods in Learning and Vision", MIT Press, Informatics. Vol. 37 No.6, December 2004, p.p. 461-470.

(8) Kusiak, A., Srinivas, S., Shahbas, M., and Harding, J., A., November 2006, Vol. 128/969. Data Mining in Manufacturing a Review. J. of Manu. Sc. and Engineering.

(9) Lee, M. H., 1993. Knowledge Based Factory, Artif. Intell. Eng. 8, [[. 109-125.

(10) Braha D., and Maimon, O., A Mathematical Theory of Design: Foundation, Algorithms and Applications. Boston, MA: Kluwer Academic Publisher, 1998.

(11) Dan Brah (ED), Data Mining for Design and Manufacturing, Spring, 2002, 544 p., Hardcover, ISBN: 1-4020-0034-0.

(12) Schertel, Stacey Lee. Data Mining and its Potential Use in Textiles: A Spinning Mill, PhD thesis. North Carolina State University, 2002.

(13) Menezes,J. “Database makers deepen mining capabilities.” Computing Canada, 07|30|99, Vol. 25, Issue 29 p 19.

(14) Waldemer Wojcik and Konral Gramszek, Knowledge – Oriented Applications in Data Mining, D. M. Industrial Allication. Page (435-442). In Tech, January, 20

(15) Hai- F., Lin., Chin – H. H., Mao – J. J. W., and Yu-C. Lin, An Application of Data Mining Technique in Developing System for Army Soldiers in Taiwan, Issue 4, Volume 7, April 2008, ISSN: 1109-2750.

(16) Norsaadah Z., Jamil S.M., Nasir T., Yong Y.T. and Yap B. W. (2008). Using Data Mining Technique to Explore Anthropometric Data Towards the Development of Sizing System.

(17) Amal Elkhair Abbas Elabid, Development of Human Body Size Measurement for Clothing Using Image Analysis. MSc thesis unpublished, College of Textile, Gezera Univesity, Sudan, January, (2009).

(18) Pechoux B.Le and Ghosh T.K (2002). Apparel Sizing and Fit: Textile Progress. Vol. 32 (1), pp. 11-19.

(19) Kurt Salmon Associates. Consumer Pulse. KSA Annual Consumer Pluses. Survey Results, 1996, p. 10.

(20) J. Abend, 1993. Our Fits over Fits. Bobbin, 34, No. 11, 78-79.

(21) ISO 8559, 1989. Garment Construction and Anthropometric Surveys – Body Dimensions, International Organization for Standardization.

(22) Emanuel. I., M. Alexander, E. Churchill and B. Truett, 1959. A height-weight sizing system for flight clothing. Wright Air Development Center, WADC Technical Report 56-365, Wright-Patterson Air Force Base, Ohio.

(23) SPSS, 1999, CRISP-DM 1.0 Step- By- Step Data Mining Guide.

(24) D. Hand, H. Mannila, P. Smyth, Principle of data Mining. The MTT Press, Massachusetts. London. England. 2001

(25) Liang Xun, 2006, Data Mining: Algorithms and Application. Beijing University Press: p.p. 22-42

(26) George Hodge, William Oxenham, Michelle Jones, Yatin Karpe, Sedef Uncu and Stacey Schertel, National Textile Center Report: November 2001,(199-210). Information Engineering: Textile Industry ,s value adding Key to Effective Decision – Making.

(27) M. Martin Jeyasingh, Kumaravel, I.J. Information Technology and Computer Science 2012, 6, 12-17 Published online June 2012 in MECS (http://www.mecs-press.org/)

(28) R. Kirkby, WEKA Explorer User Guide for Version 3-3-4, University of Weikato, 2002.

(29) Jamal et al 2010, Development of a New Sizing System Based on Data Mining Approaches, 7th International Conference. TEXSCI 2010, September 6 – 8, Liberec, CZECH REPUBLIC.

(30) Pang–Ning Tan, Introduction to data Mining, Addison–Wesley Companion Book Site, 2006.

(31) Adu-Boakye S., Power, Jess, Wallace, T. and Chen, Z. (2012). Development of a sizing system for Ghanaian Women for the Production of Ready-to-Wear Clothing. In: The 88th Textile Institute World Conference 2012, 15th – 17th May 2012, Selangor, Malaysia.

(32) Ashdown, P.S. (1998) an investigation of the structure of sizing systems: 'A comparison of three multidimensional optimized sizing system generated from anthropometric data with ASTM standard D55855 – 94'. Journal of Clothing Science and Technology. Vol. 10, No. 5, pp. 324 – 341.

(33) Kunick, P. (1984). Modern sizing and pattern making for women's and children's garments. London: Phillip Kunick Publication.

(34) British Standard Institution (2004). BS En 13402-3. Size designation of clothes. Measurements and Intervals. British Standard Institution.

(35) Beazley, A. (1997) 'Size and Fit: Procedures in undertaking a survey of body measurements'. Journal for Fashion Marketing and Measurement. Vol. 2, No. 1, pp 55 – 85.

(36) Winks, J. M., Clothing Sizes: International Standardization, Redwood, U.K. 1997.

(37) Cooklin, G., Pattern Grading for Men's Clothes, Blackwell, Oxford. 1992.

APPENDICES

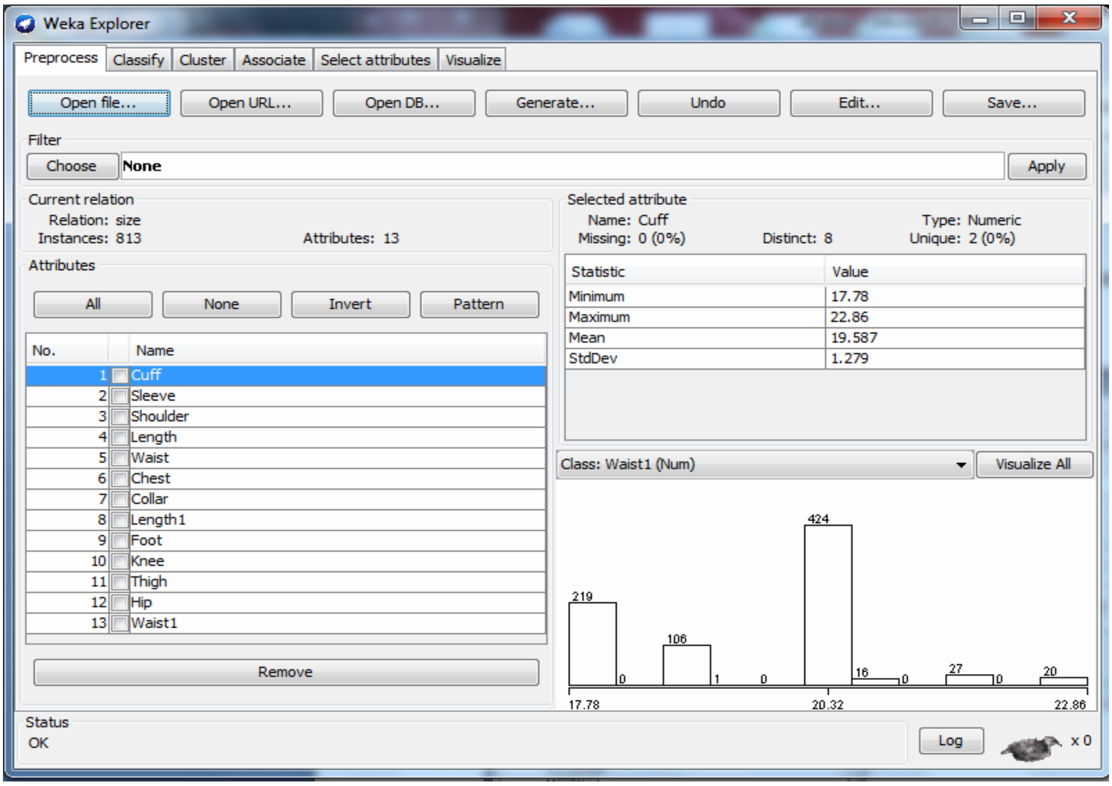
APPENDICES

Appendix (A)

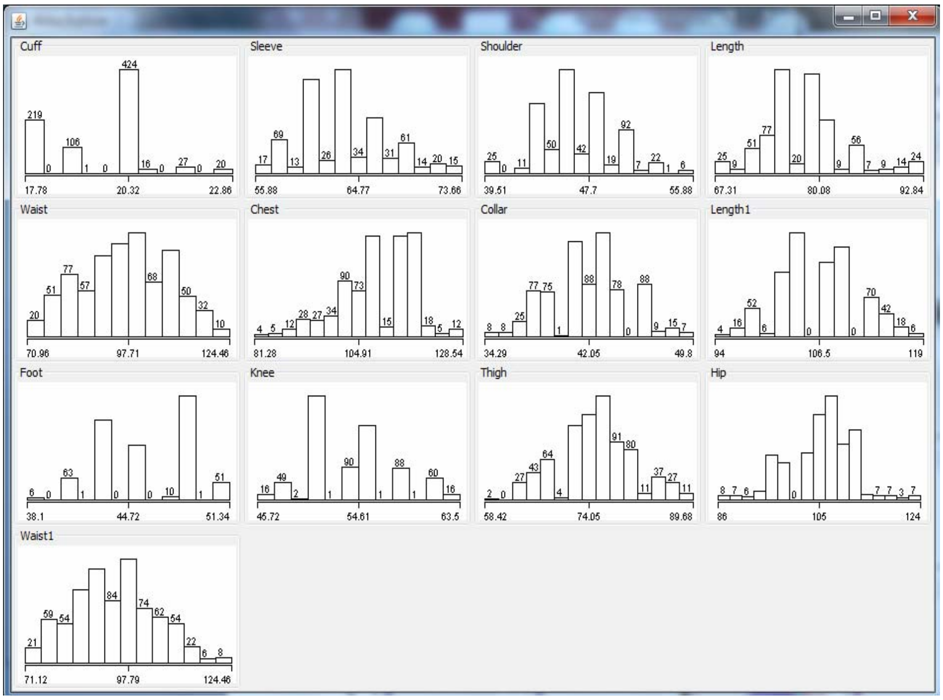
<u>WEKA</u>
<u>Result and discussion</u>

Clustering All the Data

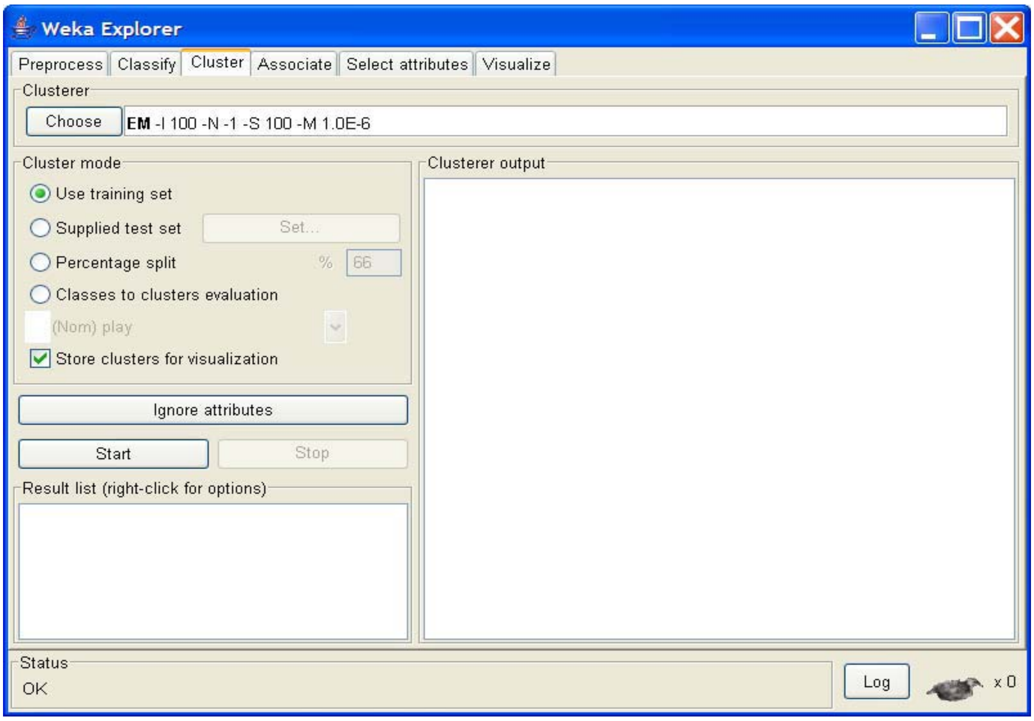
Use Sizing data that is contained in “size.csv” file and analyze it with k-means clustering scheme.



Visualization all sizing data

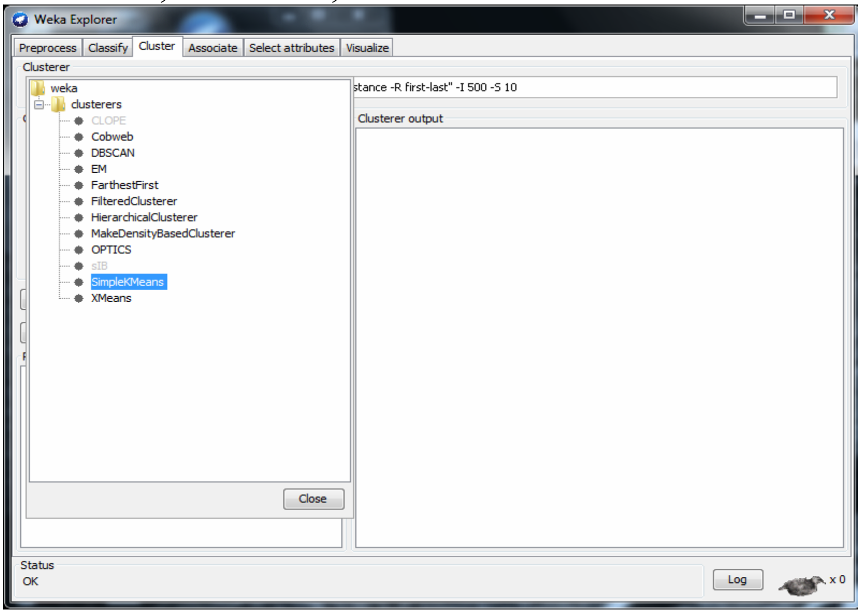


Using 'Cluster' tab at the top of WEKA Explorer window

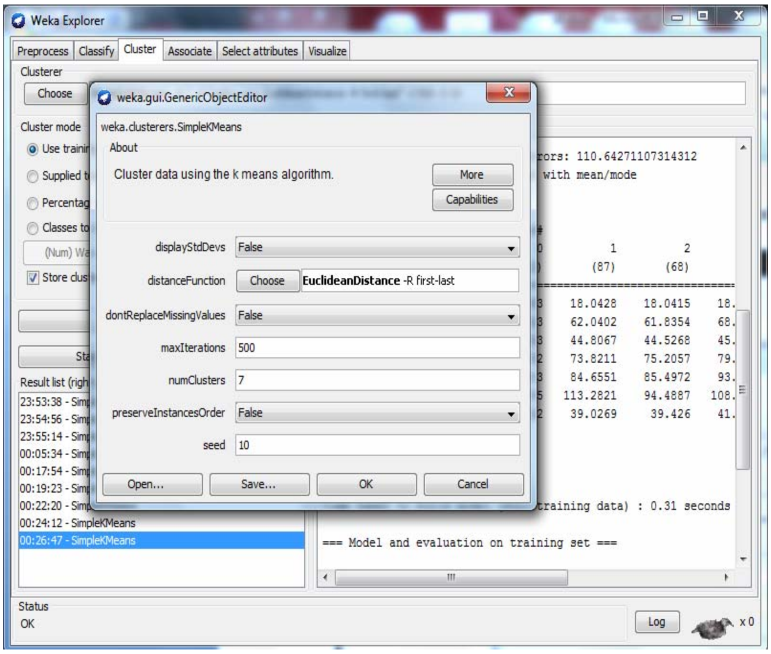


Choosing Clustering Scheme

Select the cluster scheme 'Simple K Means'. Some implementations of K-means only allow numerical values for attributes; therefore, we do not need to use a filter.



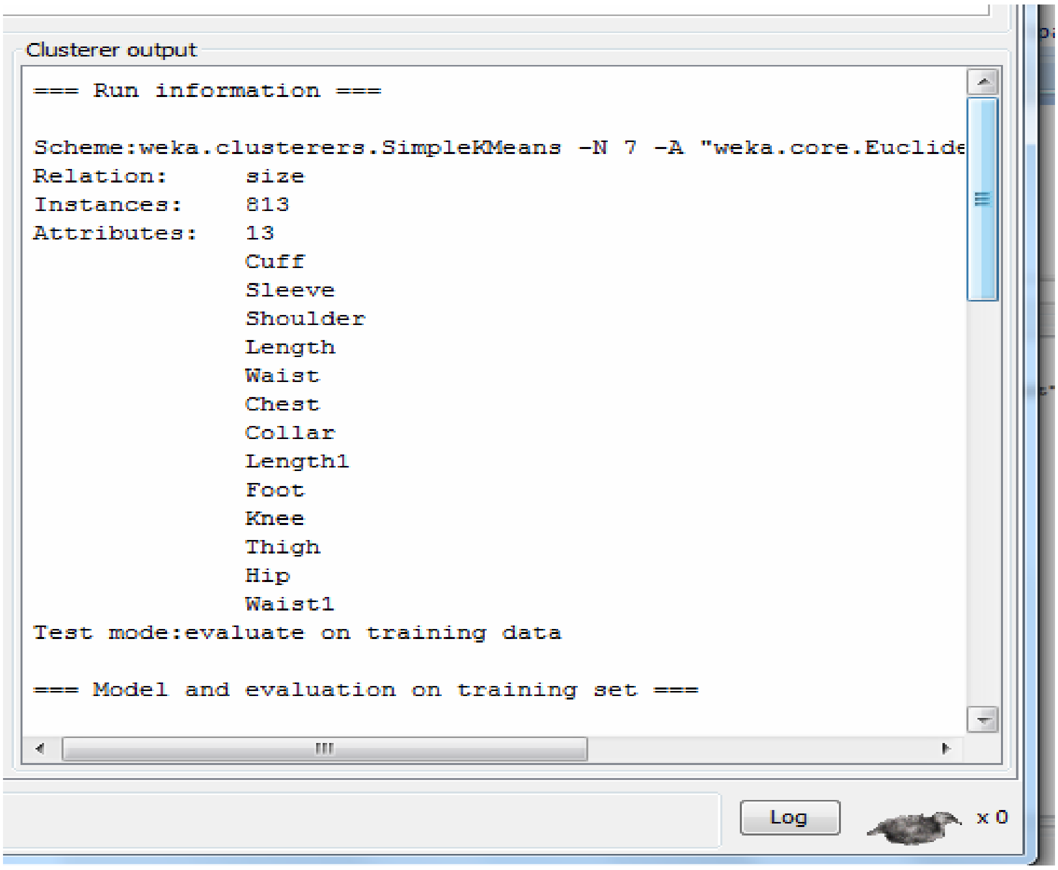
Set the value in "numClusters" box to 7 (instead of default 2)



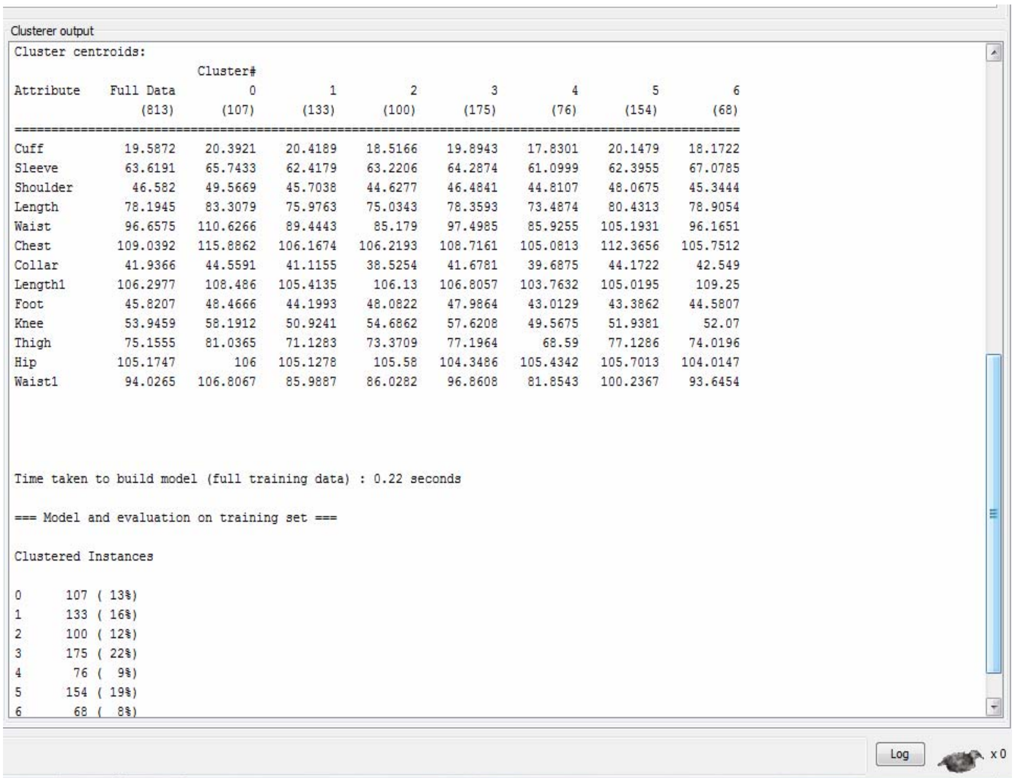
Results of training and testing

Run Information' gives you the following information:

- The clustering scheme used: Simple K- Means with 7 clusters
- The relation name "size"
- Number of instances in the relation 813
- Number of attributes in the relation 13
- List of attributes used in clustering



Analyzing Results



Clusterer output

Cluster centroids:

Attribute	Full Data (813)	Cluster# 0 (107)	1 (133)	2 (100)	3 (175)	4 (76)	5 (154)	6 (68)
Cuff	19.5872	20.3921	20.4189	18.5166	19.8943	17.8301	20.1479	18.1722
Sleeve	63.6191	65.7433	62.4179	63.2206	64.2874	61.0999	62.3955	67.0785
Shoulder	46.582	49.5669	45.7038	44.6277	46.4841	44.8107	48.0675	45.3444
Length	78.1945	83.3079	75.9763	75.0343	78.3593	73.4874	80.4313	78.9054
Waist	96.6575	110.6266	89.4443	85.179	97.4985	85.9255	105.1931	96.1651
Chest	109.0392	115.8862	106.1674	106.2193	108.7161	105.0813	112.3656	105.7512
Collar	41.9366	44.5591	41.1155	38.5254	41.6781	39.6875	44.1722	42.549
Length1	106.2977	108.486	105.4135	106.13	106.8057	103.7632	105.0195	109.25
Foot	45.8207	48.4666	44.1993	48.0822	47.9864	43.0129	43.3862	44.5807
Knee	53.9459	58.1912	50.9241	54.6862	57.6208	49.5675	51.9381	52.07
Thigh	75.1555	81.0365	71.1283	73.3709	77.1964	68.59	77.1286	74.0196
Hip	105.1747	106	105.1278	105.58	104.3486	105.4342	105.7013	104.0147
Waist1	94.0265	106.8067	85.9887	86.0282	96.8608	81.8543	100.2367	93.6454

Time taken to build model (full training data) : 0.22 seconds

=== Model and evaluation on training set ===

Clustered Instances

0 107 (13%)
1 133 (16%)
2 100 (12%)
3 175 (22%)
4 76 (9%)
5 154 (19%)
6 68 (8%)

The clustering model shows the centroid of each cluster and statistics on the number and percentage of instances assigned to different clusters. Cluster centroids are the mean vectors for each cluster; so, each dimension value and the centroid represent the mean value for that dimension in the cluster.

Thus, centroids can be used to characterize the clusters. WEKA generated clusters are:

Cluster 0 shows that the size that represents this cluster is Cluster 0 <--

3XL which represent about 107 people that have about 13% from data.

Cluster 1 show that the size that represents this cluster is Cluster 1 <-- M

which represent about 133 people that have about 16% from data.

Cluster 2 it shows that it represents the 2 <-- S which have about 100person

and represent about 12% from database.

Clusters 3 represent the cluster 3 <-- L and it has about 175person and about 22% from database.

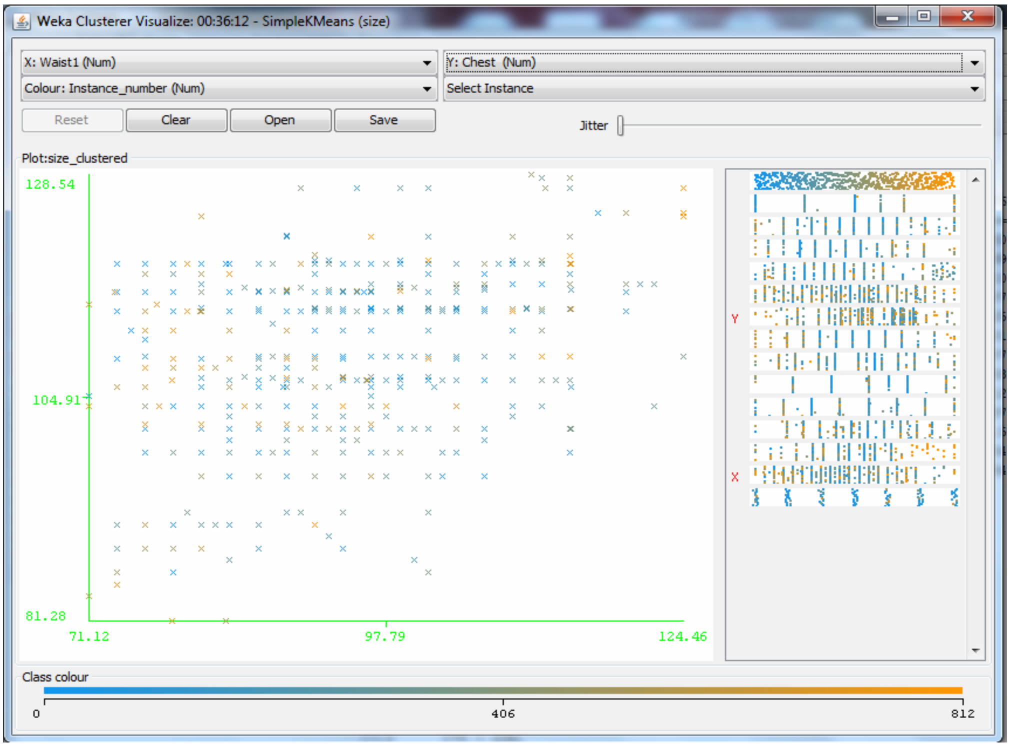
Clusters 4 represent the small size 4 <-- XS and it has about 76person and about 9% from database.

Cluster 5 shows that the size that represents this cluster is Cluster 5 <-- 2XL which represent about 154 people that have about 19% from data.

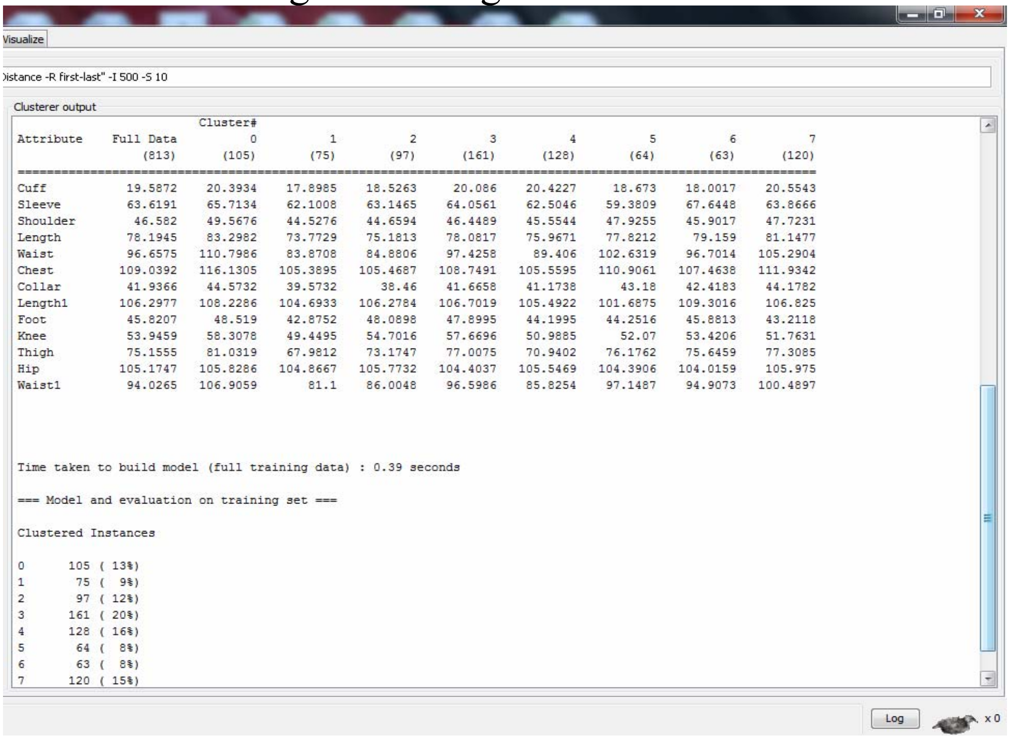
Cluster 6 it shows that it represents the 6 <-- XL which have about 68person and represent about 8% from database.

Visualization of Results

Representation the results of clustering through visualization



Results of training and testing with 8 clusters


Visualize

Distance -R first-last" -I 500 -S 10

Clusterer output

Attribute	Full Data (813)	Cluster# 0 (105)	1 (75)	2 (97)	3 (161)	4 (128)	5 (64)	6 (63)	7 (120)
Cuff	19.5872	20.3934	17.8985	18.5263	20.086	20.4227	18.673	18.0017	20.5543
Sleeve	63.6191	65.7134	62.1008	63.1465	64.0561	62.5046	59.3809	67.6448	63.8666
Shoulder	46.582	49.5676	44.5276	44.6594	46.4489	45.5544	47.9255	45.9017	47.7231
Length	78.1945	83.2982	73.7729	75.1813	78.0817	75.9671	77.8212	79.159	81.1477
Waist	96.6575	110.7986	83.8708	84.8806	97.4258	89.406	102.6319	96.7014	105.2904
Chest	109.0392	116.1305	105.3895	105.4687	108.7491	105.5595	110.9061	107.4638	111.9342
Collar	41.9366	44.5732	39.5732	38.46	41.6658	41.1738	43.18	42.4183	44.1782
Length1	106.2977	108.2286	104.6933	106.2784	106.7019	105.4922	101.6875	109.3016	106.825
Foot	45.8207	48.519	42.8752	48.0898	47.8995	44.1995	44.2516	45.8813	43.2118
Knee	53.9459	58.3078	49.4495	54.7016	57.6696	50.9885	52.07	53.4206	51.7631
Thigh	75.1555	81.0319	67.9812	73.1747	77.0075	70.9402	76.1762	75.6459	77.3085
Hip	105.1747	105.8286	104.8667	105.7732	104.4037	105.5469	104.3906	104.0159	105.975
Waist1	94.0265	106.9059	81.1	86.0048	96.5986	85.8254	97.1487	94.9073	100.4897

Time taken to build model (full training data) : 0.39 seconds

=== Model and evaluation on training set ===

Clustered Instances

```
0      105 ( 13%)
1       75 (  9%)
2       97 ( 12%)
3      161 ( 20%)
4      128 ( 16%)
5       64 (  8%)
6       63 (  8%)
7      120 ( 15%)
```

Log x 0

Analyzing Results

Cluster 0 represent the 0<-- 4XL is the new size that added to cover the clustered sizes in the database and it represent the 105 person that having a percentage of 13%.

Cluster 1 show represent the small size 1 <-- XS which represent about 75 people that have about 9% from data.

Cluster 2 it shows that it represents the 2 <-- S which have about 97person and represent about 12% from database.

Clusters 3 represent the cluster 3 <-- XL and it has about 161person and about 20% from database.

Clusters 4 shows that the size that represents this cluster is Cluster 4 <-- M and it has about 128person and about 16% from database.

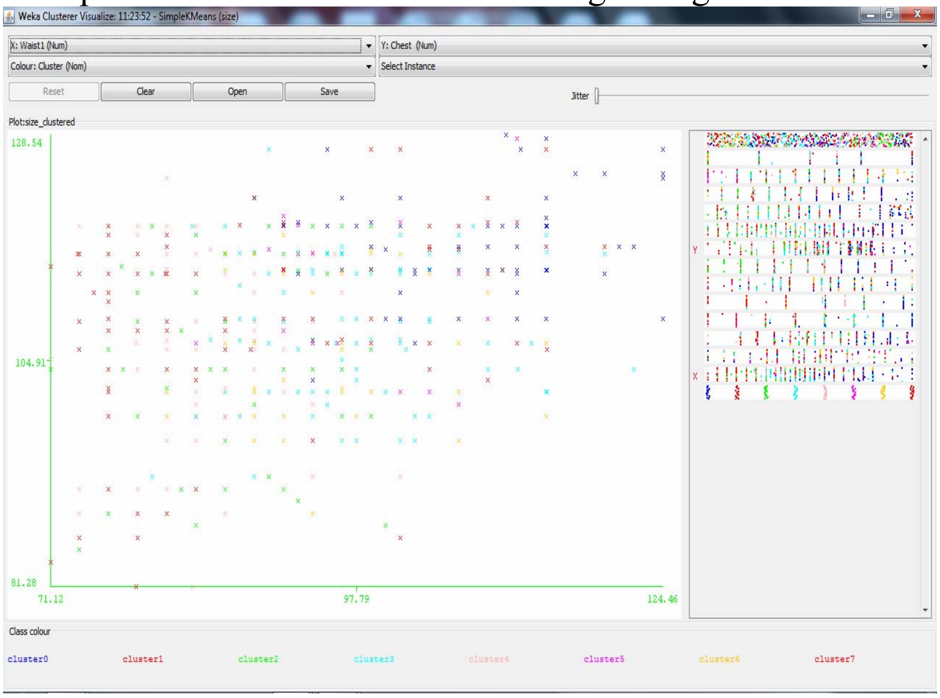
Cluster 5 shows that the size that represents this cluster is Cluster 5 <-- 2XL which represent about 64 people that have about 8% from data.

Cluster 6 it shows that it represents the 6 <-- L which have about 63person and represent about 8% from database.

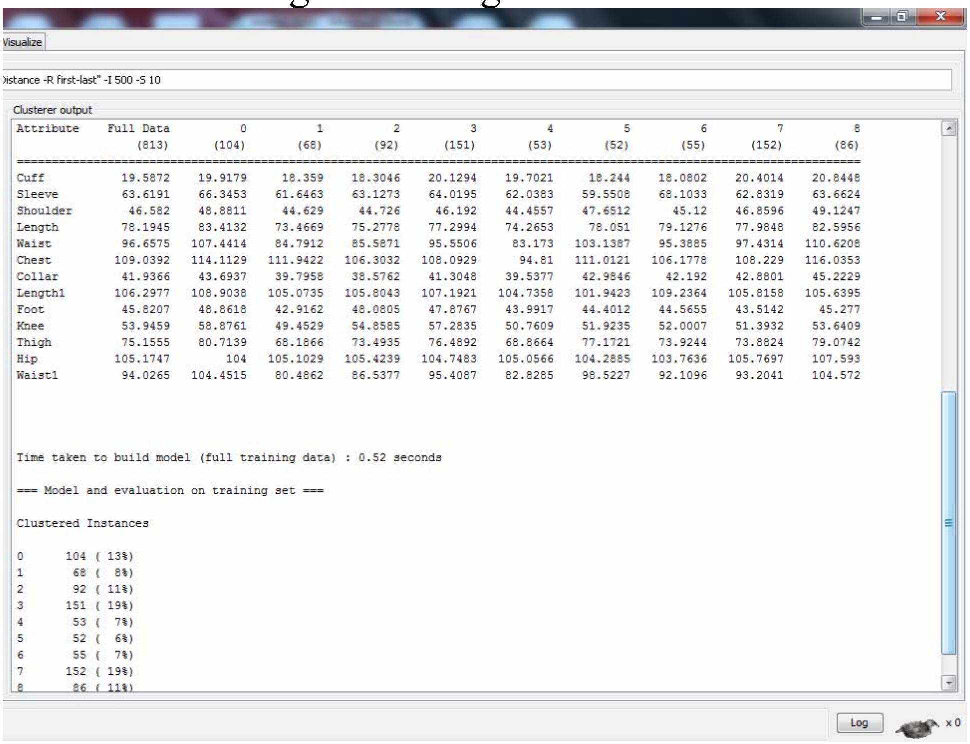
Cluster 7 show that the size that represents this cluster is Cluster 7 <-- 3XL which represent about 120 people that have about 15% from data.

Visualization of Results

Representation the results of clustering through visualization



Results of training and testing with 9 clusters



Visualize

Distance -R first-last" -I 500 -S 10

Clusterer output

Attribute	Full Data (813)	0 (104)	1 (68)	2 (92)	3 (151)	4 (53)	5 (52)	6 (55)	7 (152)	8 (86)
Cuff	19.5872	19.9179	18.359	18.3046	20.1294	19.7021	18.244	18.0802	20.4014	20.8448
Sleeve	63.6191	66.3453	61.6463	63.1273	64.0195	62.0383	59.5508	68.1033	62.8319	63.6624
Shoulder	46.582	48.8811	44.629	44.726	46.192	44.4557	47.6512	45.12	46.8596	49.1247
Length	78.1945	83.4132	73.4669	75.2778	77.2994	74.2653	78.051	79.1276	77.9848	82.5956
Waist	96.6575	107.4414	84.7912	85.5871	95.5506	83.173	103.1387	95.3885	97.4314	110.6208
Chest	109.0392	114.1129	111.9422	106.3032	108.0929	94.81	111.0121	106.1778	108.229	116.0353
Collar	41.9366	43.6937	39.7958	38.5762	41.3048	39.5377	42.9846	42.192	42.8801	45.2229
Length1	106.2977	108.9038	105.0735	105.8043	107.1921	104.7358	101.9423	109.2364	105.8158	105.6395
Foot	45.8207	48.8618	42.9162	48.0805	47.8767	43.9917	44.4012	44.5655	43.5142	45.277
Knee	53.9459	58.8761	49.4529	54.8585	57.2835	50.7609	51.9235	52.0007	51.3932	53.6409
Thigh	75.1555	80.7139	68.1866	73.4935	76.4892	68.8664	77.1721	73.9244	73.8824	79.0742
Hip	105.1747	104	105.1029	105.4239	104.7483	105.0566	104.2885	103.7636	105.7697	107.593
Waist1	94.0265	104.4515	80.4862	86.5377	95.4087	82.8285	98.5227	92.1096	93.2041	104.572

```
Time taken to build model (full training data) : 0.52 seconds

=== Model and evaluation on training set ===

Clustered Instances

0      104 ( 13%)
1       68 (  8%)
2       92 ( 11%)
3      151 ( 19%)
4       53 (  7%)
5       52 (  6%)
6       55 (  7%)
7      152 ( 19%)
8       86 ( 11%)
```

Log

x 0

Analyzing Results

Cluster 0 represent the 0<-- XL is the new size that added to cover the clustered sizes in the database and it represent the 104 person that having a percentage of 13%.

Cluster 1 show represent the small size 1 <-- XS which represent about 68 people that have about 8% from data.

Cluster 2 it shows that it represents the 2 <-- S which have about 92person and represent about 11% from database.

Clusters 3 represent the cluster 3 <-- L and it has about 151person and about 19% from database.

Clusters 4 shows that the size <-- No class because of the lower persons 53 in that size and the percentage to the database id 7%.

Cluster 5 shows that the size <-- No class because of the lower persons 52 in that size and the percentage to the database id 6%.

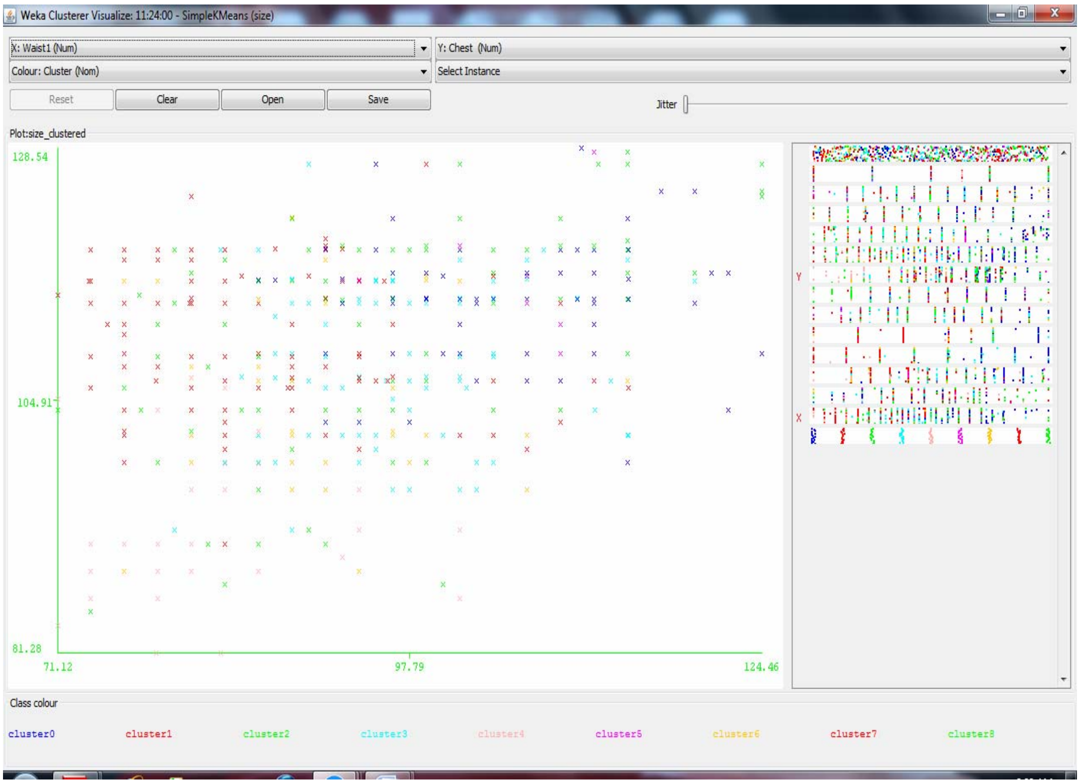
Cluster 6 shows that the size <-- No class because of the lower persons 55 in that size and the percentage to the database id 7%.

Cluster 7 show that the size that represents this cluster is Cluster 7 <-- M which represent about 152 people that have about 19% from data.

Cluster 8 represents the cluster 8 <-- 2XL and it has about 86person and about 11% from database.

Visualization of Results

Representation the results of clustering through visualization



Appendix (B)
SPSS Analysis of Results

7 clusters (size)

Table (16)

Initial cluster centers

	Cluster						
	1	2	3	4	5	6	7
CUFF	17.78	20.32	22.86	20.32	17.78	17.78	22.86
SLEEVE	60.96	60.32	63.50	73.66	63.50	60.96	66.04
SHOULDER	43.18	40.64	48.26	53.43	43.18	44.45	53.43
LENGTH	68.58	78.74	81.28	81.28	73.69	73.69	92.69
WAIST	106.68	99.06	83.82	102.00	71.12	71.12	124.46
CHEST	108.00	99.06	116.84	127.00	104.00	86.36	127.00
COLLAR	38.10	43.18	48.26	48.26	34.29	38.10	49.34
LENGTH1	97.00	119.00	104.00	109.00	107.00	99.00	112.00
FOOT	43.18	48.26	45.72	48.26	50.80	38.10	48.26
KNEE	48.26	60.96	50.80	58.42	55.88	50.80	55.88
THIGH	73.87	81.28	65.88	76.20	68.58	66.04	88.90
HIP	104.00	102.00	96.00	89.00	94.00	112.00	117.00
WAIST1	76.20	102.87	119.38	95.25	71.12	101.60	124.46

Table (17)

Iteration history

Iteration	Change in Cluster Centers						
	1	2	3	4	5	6	7
1	19.502	17.797	19.823	20.187	18.547	20.037	22.827
2	2.459	1.793	4.482	3.313	4.148	5.512	3.757
3	.664	1.046	2.451	1.220	2.244	2.153	1.266
4	.535	.515	.678	.685	.359	.552	.442
5	.199	.408	.791	.605	.176	.619	.275
6	.452	.405	.384	.450	.132	.920	.192
7	.320	.320	.526	.226	.000	.664	.286
8	.197	.130	.000	.330	.178	.338	.188
9	.149	.145	.000	.211	.000	.295	.162
10	.079	.123	.000	.168	.000	.000	.134

Table (18)
Number of cases in each center cluster

	Cluster						
	1	2	3	4	5	6	7
Valid	183.000	190.000	57.000	116.000	116.000	62.000	89.000

Table (19)
Final cluster centers

	Cluster						
	1	2	3	4	5	6	7
CUFF	19.51	19.70	19.48	20.00	18.94	19.05	20.26
SLEEVE	63.04	64.03	64.17	63.78	63.29	62.50	64.57
SHOULDER	45.97	46.99	45.69	47.89	44.95	44.64	49.30
LENGTH	76.58	79.39	77.44	81.35	74.53	75.06	82.29
WAIST	93.81	101.74	83.66	108.98	82.37	82.78	112.24
CHEST	106.21	108.11	110.61	113.85	112.56	91.44	117.24
COLLAR	41.42	42.66	41.43	43.92	39.20	39.23	44.64
LENGTH1	105.38	106.39	108.68	106.67	106.10	105.87	106.53
FOOT	45.34	45.84	46.98	45.99	45.14	45.18	47.13
KNEE	52.65	54.27	56.33	55.42	51.70	51.91	56.80
THIGH	73.18	76.71	78.28	77.47	69.61	70.25	81.52
HIP	105.27	106.67	102.93	100.66	105.05	105.23	109.22
WAIST1	88.24	96.90	104.28	100.30	80.18	84.02	110.09

Table (20)

Analysis of variance (ANOVA)

	Cluster		Error		F	Sig
	Mean Square	df	Mean Square	df		
CUFF	21.913	6	1.485	806	14.759	.000
SLEEVE	47.305	6	12.213	806	3.873	.001
SHOULDER	257.617	6	7.269	806	35.440	.000
LENGTH	931.708	6	16.749	806	55.629	.000
WAIST	15145.520	6	23.614	806	641.387	.000
CHEST	5179.357	6	33.612	806	154.094	.000
COLLAR	432.950	6	4.988	806	86.794	.000
LENGTH1	86.338	6	21.057	806	4.100	.000
FOOT	59.101	6	7.905	806	7.477	.000
KNEE	412.070	6	12.429	806	33.155	.000
THIGH	1836.042	6	18.564	806	98.904	.000
HIP	756.318	6	30.922	806	24.459	.000
WAIST1	11612.791	6	28.255	806	410.994	.000

8 clusters (size)

Table (16)

Initial cluster centers

	Cluster							
	1	2	3	4	5	6	7	8
CUFF	20.32	20.32	22.86	20.32	17.78	17.78	20.32	22.86
SLEEVE	63.50	66.04	58.42	73.66	63.50	60.96	59.69	66.04
SHOULDER	45.72	40.64	50.80	53.43	43.18	44.45	44.45	53.43
LENGTH	83.82	78.74	81.28	81.28	73.69	73.69	81.28	92.69
WAIST	119.38	91.44	116.84	102.00	71.12	71.12	81.28	124.46
CHEST	91.44	114.00	116.84	127.00	104.00	86.36	106.68	127.00
COLLAR	43.18	43.18	45.72	48.26	34.29	38.10	44.45	49.34
LENGTH1	114.00	112.00	104.00	109.00	107.00	99.00	98.00	112.00
FOOT	43.18	45.72	43.18	48.26	50.80	38.10	48.26	48.26
KNEE	50.80	50.80	50.80	58.42	55.88	50.80	60.96	55.88
THIGH	68.58	71.12	88.90	76.20	68.58	66.04	86.36	88.90
HIP	107.00	119.00	104.00	89.00	94.00	112.00	86.00	117.00
WAIST1	83.82	88.90	81.28	95.25	71.12	101.60	113.03	124.46

Table (17)

Iteration history

Iteration	Change in Cluster Centers							
	1	2	3	4	5	6	7	8
1	21.144	16.379	21.554	20.628	16.971	19.995	21.010	22.393
2	4.689	2.096	3.995	2.234	4.899	3.985	4.350	2.438
3	1.816	1.671	1.006	.848	3.950	3.734	2.457	.624
4	1.257	.778	.844	1.201	1.567	1.478	1.392	.514
5	1.087	.414	1.012	1.363	.325	.643	.864	.382
6	.515	.401	.662	.644	.233	.277	.845	.573
7	.459	.363	.353	.309	.268	.239	1.154	.326
8	.404	.358	.332	.586	.000	.426	.827	.281
9	.342	.331	.253	.376	.152	.546	.653	.401
10	.216	.205	.199	.252	.158	.502	.374	.000

Table (18)

Number of cases in each center cluster

	Cluster							
	1	2	3	4	5	6	7	8
Valid	142.000	125.000	151.000	79.000	103.000	90.000	56.000	67.000

Table (20)

Final cluster centers

	Cluster							
	1	2	3	4	5	6	7	8
CUFF	19.63	19.52	19.89	19.82	18.91	19.11	19.46	20.46
SLEEVE	63.10	63.57	63.63	64.41	63.14	62.98	64.38	64.81
SHOULDER	46.48	45.80	47.70	47.92	44.99	44.79	45.83	49.66
LENGTH	78.13	76.78	79.67	82.15	74.41	75.34	77.76	82.98
WAIST	98.68	93.77	105.94	107.30	81.69	84.99	83.75	113.76
CHEST	102.37	112.60	113.16	110.94	112.66	94.55	109.62	118.41
COLLAR	42.36	41.33	43.41	43.71	38.99	39.51	41.49	44.93
LENGTH1	105.26	106.17	106.47	107.10	105.84	106.12	108.50	106.49
FOOT	45.20	45.77	46.05	46.04	44.97	45.38	46.77	47.56
KNEE	53.62	52.90	54.64	55.69	51.61	51.99	56.43	57.09
THIGH	74.70	73.70	78.40	77.17	69.23	70.39	78.43	81.91
HIP	105.51	104.70	108.60	97.37	104.86	106.20	102.71	108.00
WAIST1	92.87	88.84	99.10	101.83	79.72	84.01	104.99	111.81

Table (20)

Analysis of variance (ANOVA)

	Cluster		Error		F	Sig
	Mean Square	Df	Mean Square	df		
CUFF	19.827	7	1.477	805	13.419	.000
SLEEVE	39.508	7	12.238	805	3.228	.002
SHOULDER	232.268	7	7.179	805	32.355	.000
LENGTH	795.714	7	16.795	805	47.379	.000
WAIST	12549.987	7	27.399	805	458.052	.000
CHEST	5270.981	7	26.423	805	199.487	.000
COLLAR	382.800	7	4.893	805	78.239	.000
LENGTH1	72.630	7	21.095	805	3.443	.001
FOOT	58.878	7	7.843	805	7.507	.000
KNEE	339.298	7	12.565	805	27.003	.000
THIGH	1644.963	7	17.968	805	91.552	.000
HIP	1086.663	7	27.149	805	40.027	.000
WAIST1	10039.336	7	27.547	805	364.446	.000

9 clusters (size)

Table (16)

Initial cluster centers

	Cluster								
	1	2	3	4	5	6	7	8	9
CUFF	17.78	22.86	22.86	19.05	20.32	17.78	20.32	17.78	20.32
SLEEVE	60.96	58.42	66.04	68.58	63.50	60.96	63.50	60.96	60.96
SHOULDER	48.26	50.80	53.43	50.80	46.99	44.45	45.72	55.88	43.18
LENGTH	71.12	81.28	83.82	92.71	80.01	73.69	83.82	83.82	71.12
WAIST	73.66	116.84	119.38	92.71	115.57	71.12	119.38	83.82	86.36
CHEST	83.82	116.84	124.46	102.87	106.68	86.36	91.44	128.08	118.00
COLLAR	34.29	45.72	49.80	39.37	45.72	38.10	43.18	48.26	38.10
LENGTH1	109.00	104.00	107.00	112.00	114.00	99.00	114.00	104.00	97.00
FOOT	45.72	43.18	48.26	48.26	48.26	38.10	43.18	40.64	43.18
KNEE	53.34	50.80	60.96	55.88	63.50	50.80	50.80	50.80	48.26
THIGH	63.50	88.90	68.58	81.28	81.28	66.04	68.58	76.20	63.50
HIP	102.00	104.00	122.00	112.00	86.00	112.00	107.00	99.00	112.00
WAIST1	71.12	81.28	124.46	91.44	109.22	101.60	83.82	111.76	81.28

Table (17)

Iteration history

Iteration	Change in Cluster Centers								
	1	2	3	4	5	6	7	8	9
1	15.912	21.954	21.325	18.212	19.275	19.469	20.071	18.423	15.477
2	1.718	2.574	3.145	1.198	1.607	4.881	3.635	4.934	1.664
3	.762	1.259	.743	.614	.505	2.503	2.764	1.193	1.328
4	.667	.744	.662	.808	.626	1.270	1.964	.000	.939
5	.410	.508	.345	.789	.466	.883	1.339	.366	.545
6	.618	.374	.606	1.003	.702	.400	1.571	.000	.625
7	.349	.410	.801	1.071	.870	.795	.659	.000	.295
8	.324	.318	.146	.736	.577	.517	.395	.000	.124
9	.000	.529	.000	.675	.759	.253	.422	.448	.000
10	.000	.353	.000	.451	.792	.359	.403	.653	.130

Table (18)

Number of cases in each center cluster

	Cluster								
	1	2	3	4	5	6	7	8	9
Valid	43.000	120.000	65.000	116.000	80.000	84.000	127.000	44.000	134.000

Table (19)

Final cluster centers

	Cluster								
	1	2	3	4	5	6	7	8	9
CUFF	18.68	19.92	20.41	19.78	19.94	19.35	19.41	19.61	19.12
SLEEVE	61.91	62.97	64.81	66.30	63.90	63.76	61.71	64.42	63.14
SHOULDER	44.69	47.88	49.81	46.58	47.80	45.08	46.52	45.98	44.93
LENGTH	74.08	79.59	83.05	80.59	80.96	76.84	76.44	78.25	74.67
WAIST	79.98	106.51	114.23	98.12	109.12	88.17	97.76	85.43	83.92
CHEST	92.00	112.85	118.42	109.54	111.61	98.55	105.33	112.42	113.56
COLLAR	38.72	43.60	44.77	42.12	44.01	40.43	42.05	42.15	39.48
LENGTH1	105.98	105.52	106.74	109.32	106.34	106.93	103.15	107.91	106.31
FOOT	45.48	45.77	47.46	46.59	45.97	45.66	44.66	46.89	45.27
KNEE	51.46	54.11	56.93	55.43	55.88	53.43	52.16	56.93	51.75
THIGH	68.97	77.98	81.78	76.84	77.87	72.58	73.39	79.68	70.12
HIP	104.63	109.43	108.49	103.74	98.71	106.61	105.02	103.41	104.85
WAIST1	78.79	98.81	111.44	94.65	102.87	90.84	89.56	107.99	82.03

Table (20)

Analysis of variance (ANOVA)

	Cluster		Error		F	Sig
	Mean Square	Df	Mean Square	df		
CUFF	18.081	8	1.472	804	12.283	.000
SLEEVE	203.990	8	10.567	804	19.304	.000
SHOULDER	215.137	8	7.069	804	30.433	.000
LENGTH	747.501	8	16.306	804	45.843	.000
WAIST	11232.587	8	24.932	804	450.534	.000
CHEST	4341.701	8	29.146	804	148.963	.000
COLLAR	331.292	8	4.935	804	67.129	.000
LENGTH1	319.445	8	18.575	804	17.197	.000
FOOT	64.161	8	7.727	804	8.303	.000
KNEE	358.665	8	11.966	804	29.974	.000
THIGH	1452.370	8	17.860	804	81.318	.000
HIP	851.274	8	28.173	804	30.216	.000
WAIST1	8751.365	8	27.910	804	313.557	.000

Measurement of size:

Table (62)

Measurements of size

	Cuff	SLEEVE	SHOULDER	LENGTH	WAIST	CHEST	COLLAR	LENGTH1	FOOT	KNEE	THIGH	HIP	WAIST1
Mean	19.5872	63.6191	46.5820	78.195	96.6575	109.0392	41.9366	106.2977	45.8207	53.9459	75.155	105.1747	94.0265
Std. Error of Mean	.04485	.12386	.10591	.17005	.40803	.29684	.10013	.16277	.10094	.13755	.19837	.21125	.37422
Std. Deviation	1.27893	3.53166	3.01978	4.84865	11.63409	8.46371	2.85491	4.64105	2.87802	3.92196	5.65627	6.02348	10.67030
Minimum	17.78	55.88	39.51	67.31	70.96	81.28	34.29	94	38.10	45.72	58.42	86.00	71.12
Maximum	22.86	73.66	55.88	92.84	124.46	128.54	49.80	119	51.34	63.50	89.68	124.0	124.46

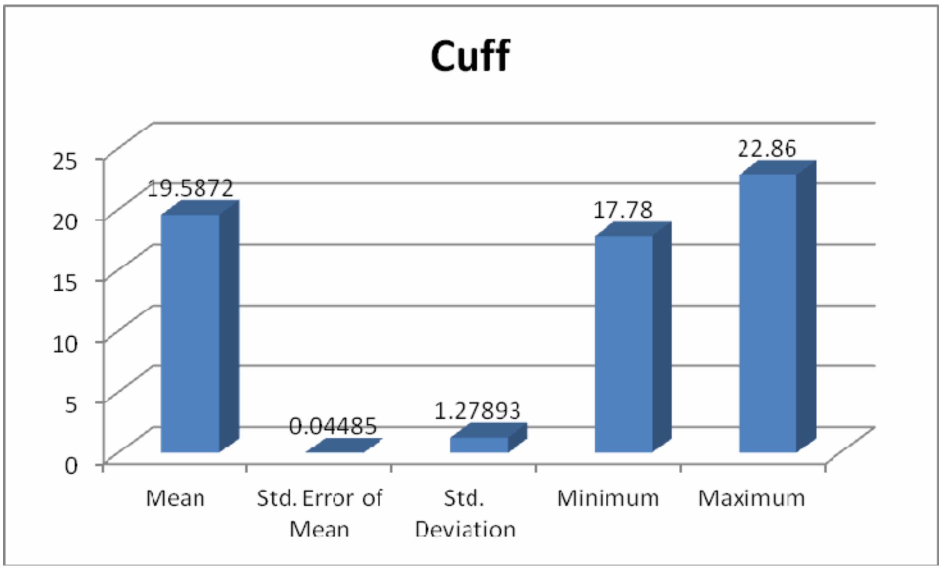


Figure 5.12: Visualization of cuff

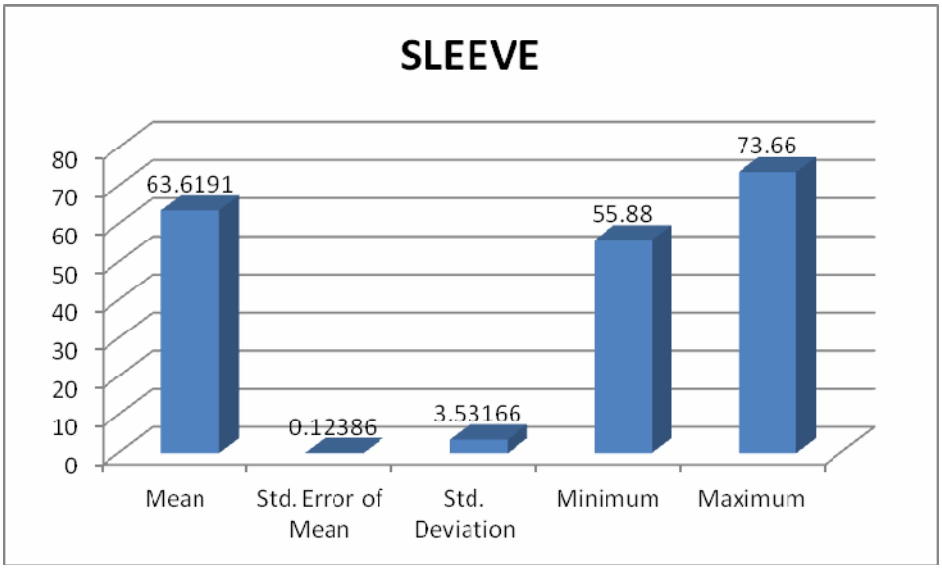


Figure 5.13: Visualization of sleeve

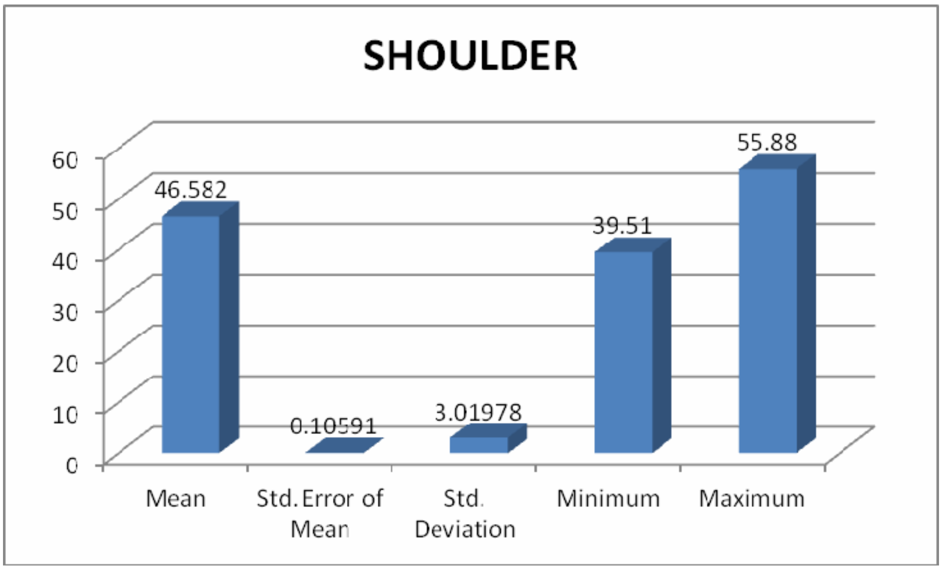


Figure 5.14: Visualization of shoulder

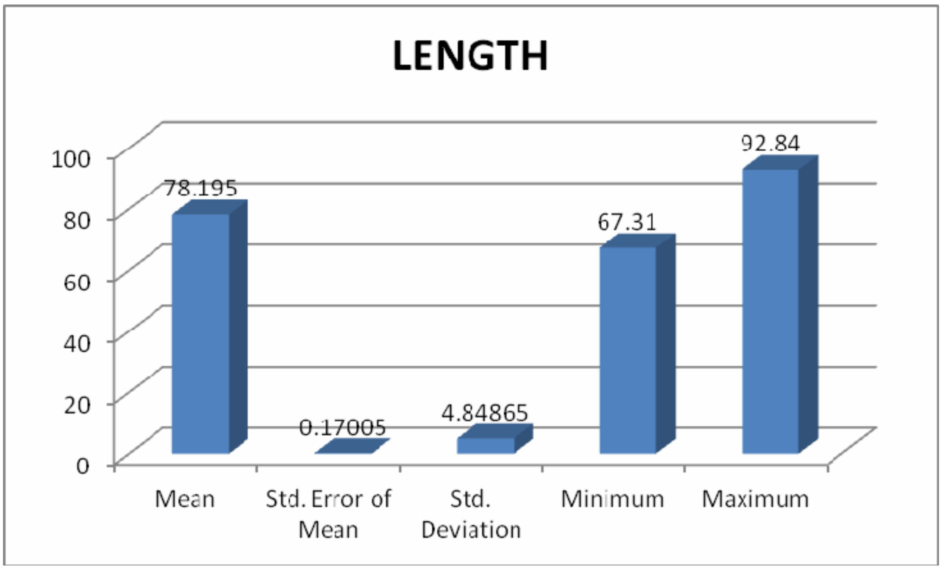


Figure 5.15: Visualization of length

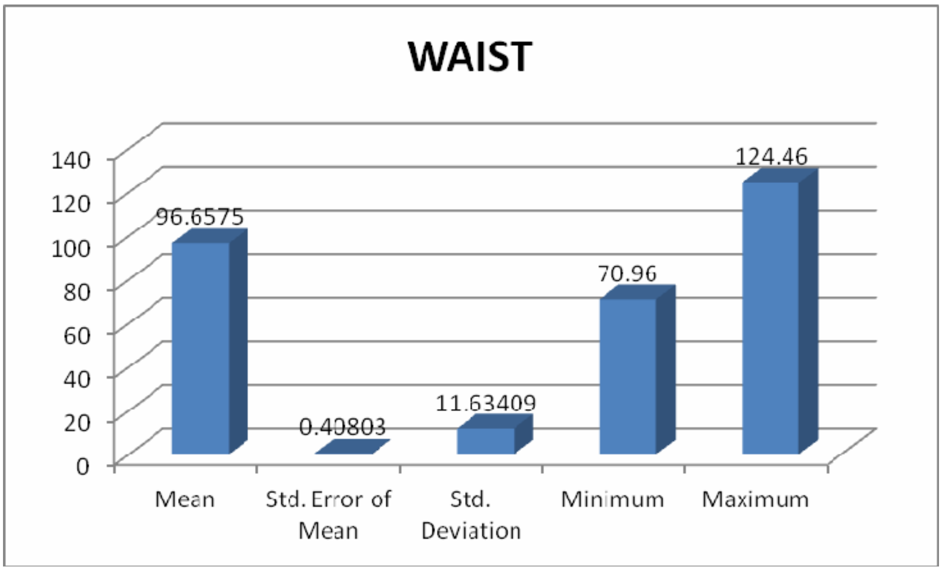


Figure 5.16: Visualization of waist

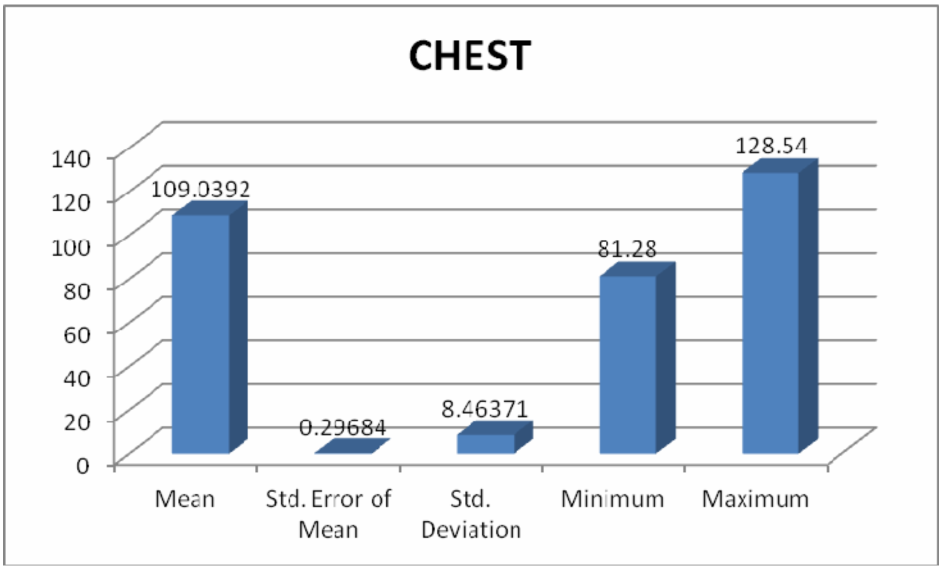


Figure 5.17: Visualization of chest

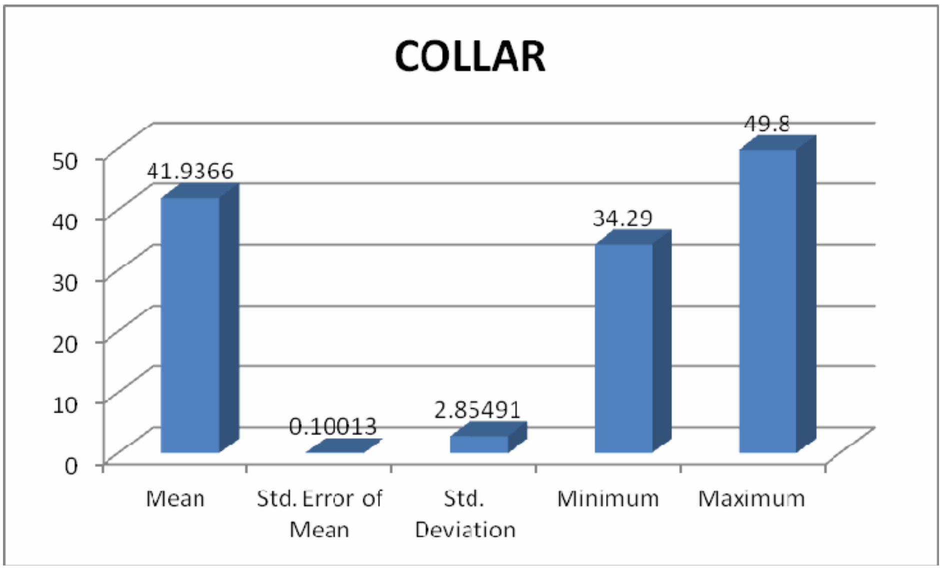


Figure 5.18: Visualization of collar

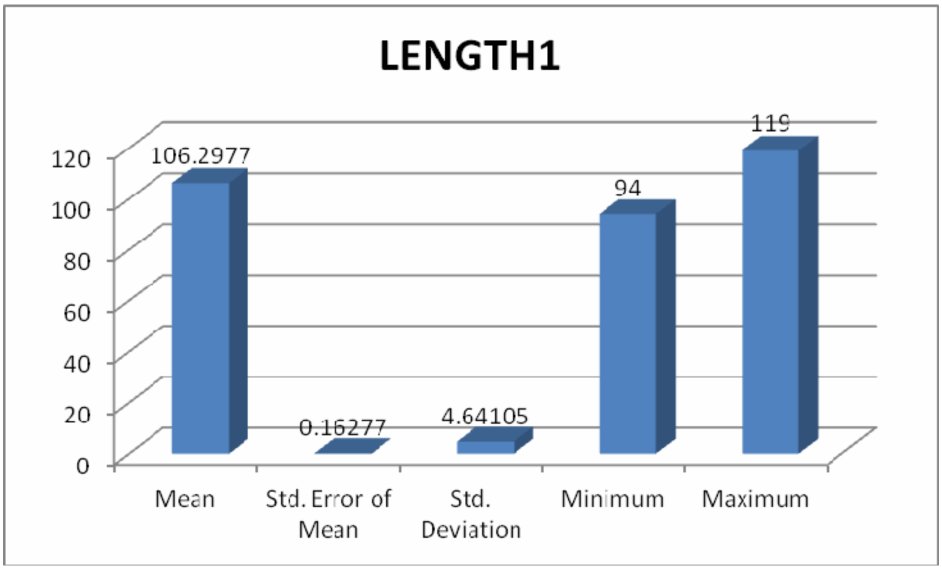


Figure 5.19: Visualization of length1

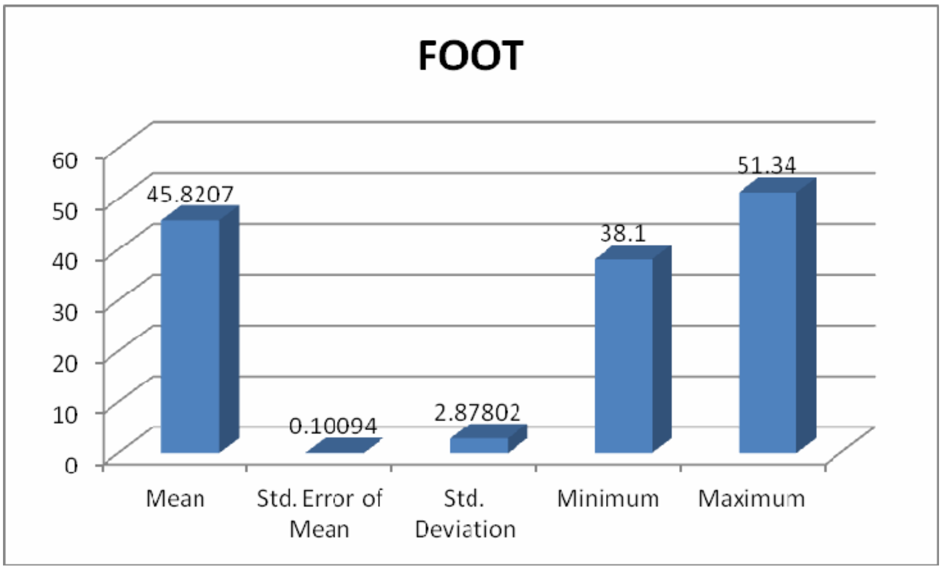


Figure 5.20: Visualization of foot

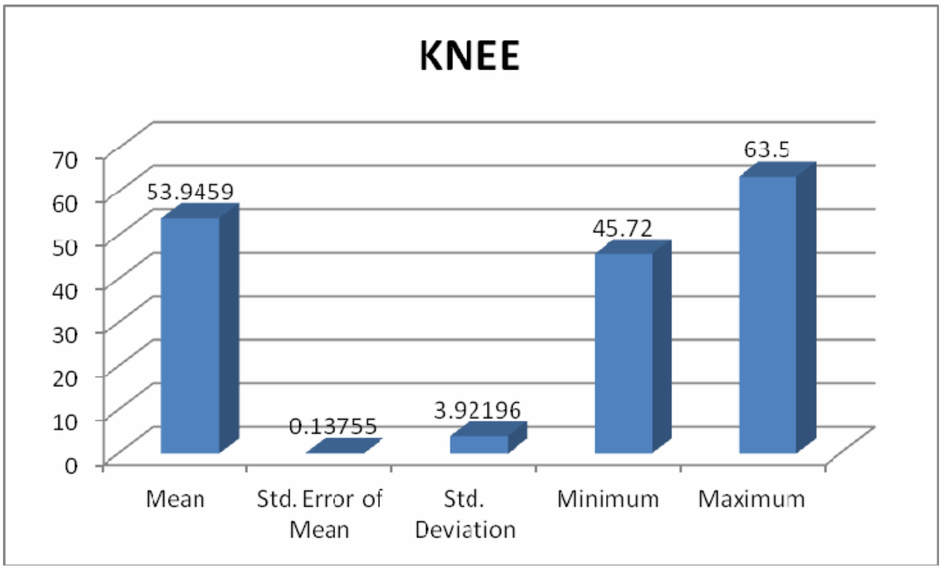


Figure 5.21: Visualization of knee

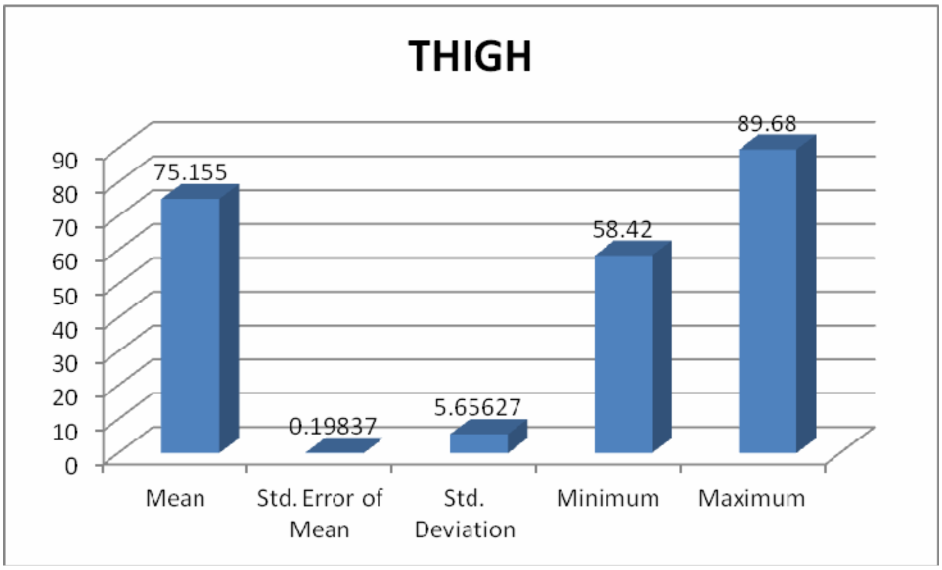


Figure 5.22: Visualization of thigh

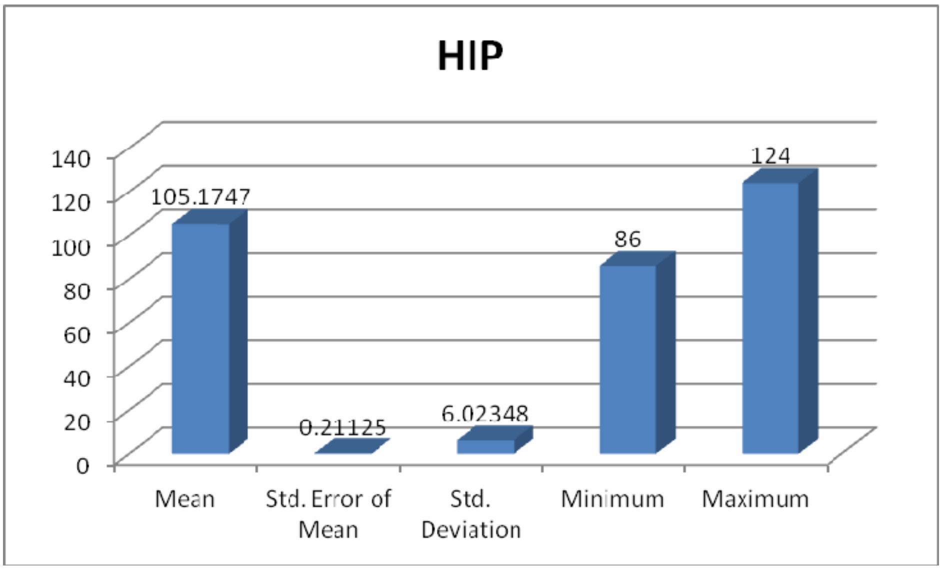


Figure 5.23: Visualization of hip

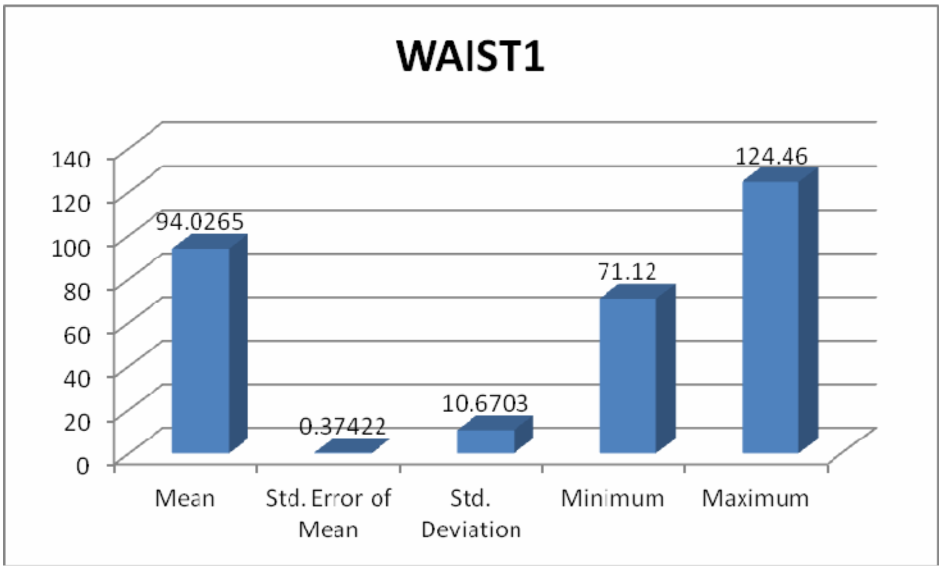


Figure 5.24: Visualization of waist1

Appendix (C)

SUR, USA, and ERU standard size charts

SUR MILITARY CLOTHING FACTORY
CAD CAM DESIGN DEPARTMENT

POSHIRT (U4) SIZE SPECIFICATION

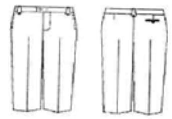
Date : 20, April 2013
Customer : Land Force

Size	Xsmall		Small		Medium *		Large		XLarge		2XLarge		3XLarge		4XLarge		5XLarge		Tol.
Conversion	Inch	Cm.	Inch	Cm.	Inch	Cm.	Inch	Cm.	Inch	Cm.	Inch	Cm.	Inch	Cm.	Inch	Cm.	Inch	Cm.	+/-
Collar Length	14 1/2	36.8	15	38.1	15 1/2	39.4	16	40.6	16 1/2	41.9	17	43.2	17 1/2	44.4	18	45.7	18 1/2	47	1/2
Chest	41	104.1	42	106.7	43	109.2	44	111.8	45	114.3	46	116.8	47	119.4	48	121.9	49	124.5	1/2
Front Length	30 5/8	77.8	31 1/8	79	31 5/8	80.3	32 1/8	81.6	32 5/8	82.8	33 1/8	84.1	33 5/8	85.4	34 1/8	86.7	34 5/8	87.9	1/2
Back Length	30 1/4	76.8	30 3/4	78.1	31 1/4	79.4	31 3/4	80.6	32 1/4	81.9	32 3/4	83.2	33 1/4	84.4	33 3/4	85.7	34 1/4	87	1/2
Across Shoulder	18 1/4	46.3	18 3/4	47.6	19 1/4	48.9	19 3/4	50.1	20 1/4	51.4	20 3/4	52.7	21 1/4	54	21 3/4	55.2	22 1/4	56.5	1/2
Sleeve Length	24	61	24 1/2	62.2	25	63.5	25 1/2	64.8	26	66	26 1/2	67.3	27	68.6	27 1/2	69.9	28	71.1	1/2

Prepared by: ____________ Noted by: ____________ Aprroved by: ____________

SUR MILITARY CLOTHING FACTORY
CAD CAM DESIGN DEPARTMENT

POSHIRT TROUSER (U4) SIZE SPECIFICATION



Date : 17, October 2011
Customer : Land Force

*

Reference	Xsmall		Small		Medium		Large		XLarge		2XLarge		3XLarge		4XLarge		5XLarge		TOL.
Size	Inch 33	CM. 84	Inch 35	CM. 89	Inch 37	CM. 94	Inch 39	CM. 99	Inch 41	CM. 104	Inch 43	CM. 109	Inch 45	CM. 114	Inch 47	CM. 119	Inch 49	CM. 124	+/-
Waist	33	84	35	89	37	94	39	99	41	104	43	109	45	114	47	119	49	124	
Hips	39 1/4	99.8	41 1/4	104.8	43 1/4	109.8	45 1/4	114.8	47 1/4	119.8	49 1/4	124.8	51 1/4	129.8	53 1/4	134.8	55 1/4	139.8	1/2
Outseam	40 1/2	102.9	41 3/4	106	43	109.2	45 1/4	114.9	45 1/2	115.6	45 3/4	116.2	47	119.4	47 1/4	120	47 1/2	120.6	1/2
Inseam	29	73.6	30	76.2	31	78.8	33	83.8	33	83.8	33	83.8	34	86.4	34	86.4	34	86.4	1/2

Sur/Spec2010/Sheet5/WTU3/GMayo'08

11-27-14 3:57 PM

Prepared by: ____________ **Noted by:** ____________ **Aprroved by:** ____________

Sizes and Measurements **Simplicity®**

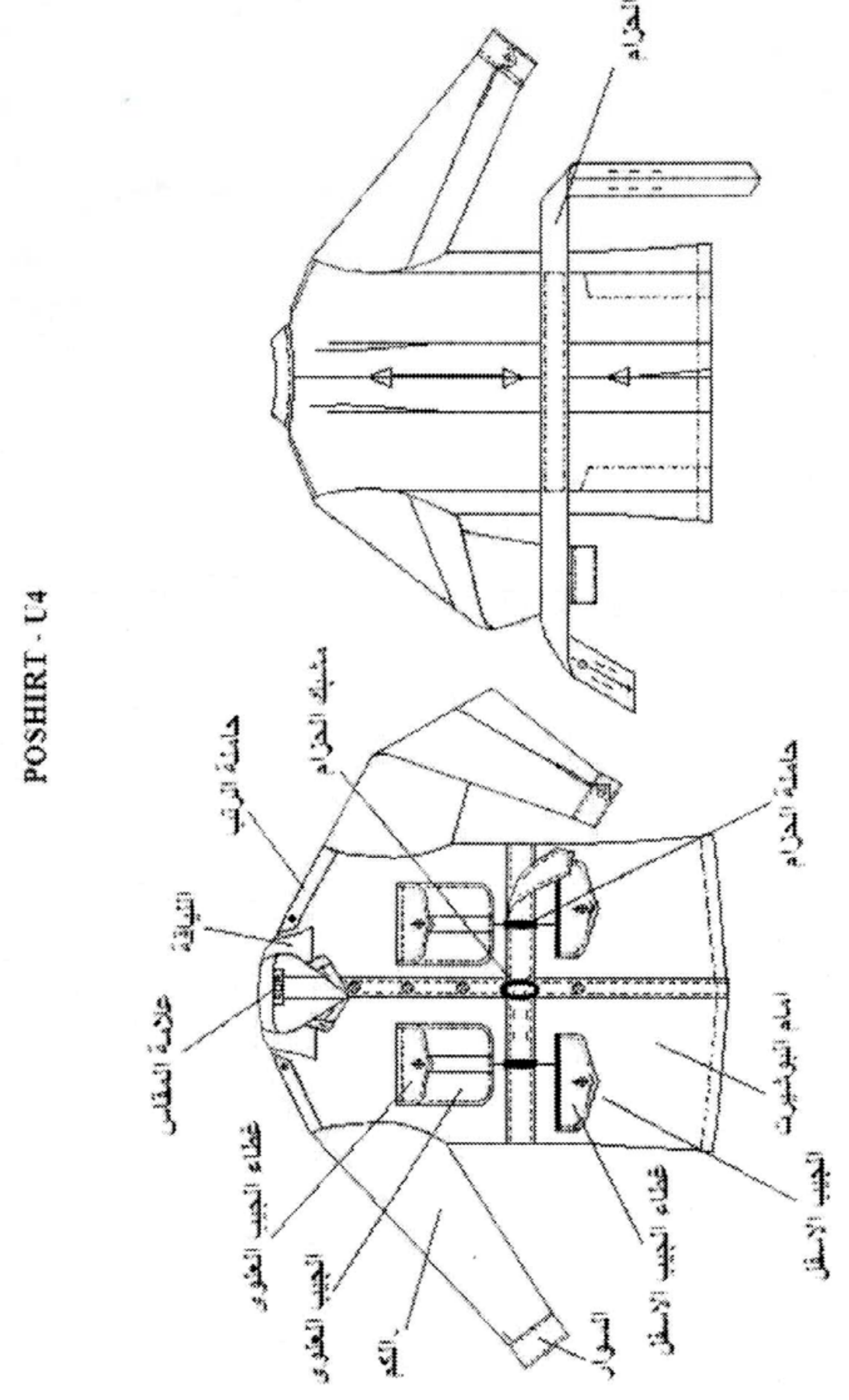
Men's For men of average build; about 5'10" without shoes.

Sizes US	**32**	**34**	**36**	**38**	**40**	**42**	**44**	**46**	**48**	**50**	**52**
Chest	32"	34"	36"	38"	40"	42"	44"	46"	48"	50"	52"
Waist	27"	28"	30"	32"	34"	36"	39"	42"	44"	46"	48"
Hip	34"	35"	37"	39"	41"	43"	45"	47"	49"	51"	53"
Neck Band	13½"	14"	14½"	15"	15½"	16"	16½"	17"	17½"	18"	18½"
Shirt Sleeve	31"	32"	32"	33"	33"	34"	34"	35"	35"	36"	36"
Sizes EUR	**42**	**44**	**46**	**48**	**50**	**52**	**54**	**56**	**58**	**60**	**62**
Chest	82cm	87cm	92cm	97cm	102cm	107cm	112cm	117cm	122cm	127cm	132cm
Waist	66cm	71cm	76cm	81cm	87cm	92cm	99cm	107cm	112cm	117cm	122cm
Hip	84cm	89cm	94cm	99cm	104cm	109cm	114cm	119cm	124cm	130cm	135cm
Neck Band	34,5cm	35,5cm	37cm	38cm	39,5cm	40,5cm	42cm	43cm	44,5cm	45,5cm	47cm
Shirt Sleeve	78,5cm	81cm	81cm	84cm	84cm	87cm	87cm	89cm	89cm	91,5cm	91,5cm
Sizes FR	**42**	**44**	**46**	**48**	**50**	**52**	**54**	**56**	**58**	**60**	**62**
Chest	82cm	87cm	92cm	97cm	102cm	107cm	112cm	117cm	122cm	127cm	132cm
Waist	66cm	71cm	76cm	81cm	87cm	92cm	99cm	107cm	112cm	117cm	122cm
Hip	84cm	89cm	94cm	99cm	104cm	109cm	114cm	119cm	124cm	130cm	135cm
Neck Band	34,5cm	35,5cm	37cm	38cm	39,5cm	40,5cm	42cm	43cm	44,5cm	45,5cm	47cm
Shirt Sleeve	78,5cm	81cm	81cm	84cm	84cm	87cm	87cm	89cm	89cm	91,5cm	91,5cm

www.**simplicity**.com

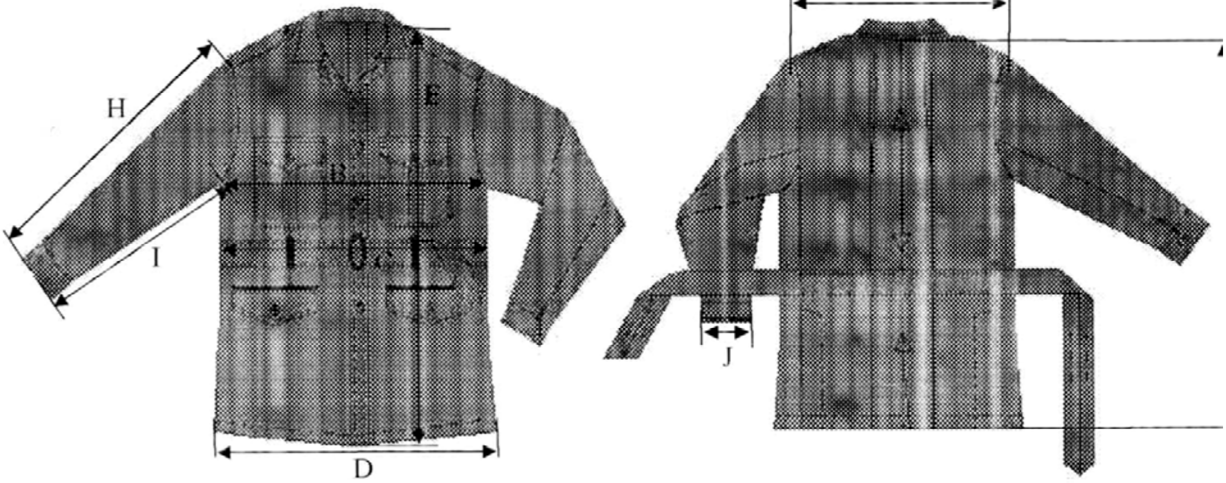
Appendix (D)

Poshirt measurements

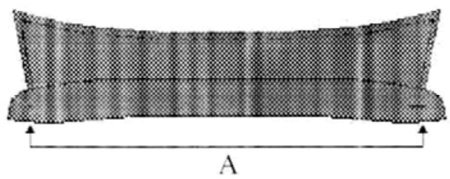


LOCATION OF BASIC MEASUREMENT

(POSHIRT U4)



WITH STAND COLLAR



A)طول الياقة - (من منتصف الزر إلى نهايه العروه).

B)نصف الصدر - (1 بوصه اسفل فتحه الكم).

C)نصف الخصر - (5 بوصات أسفل فتحه الكم).

D)نصف الدائرة - (من بدايه طرف الدائرة إلى نهايه الطرف الاخر).

E)الطول الامامي - (من بدايه الكتف إلى نهايه الدائرة).

F)الطول الخلفي- (من منتصف أعلى الظهر إلى منتصف اسفل الدائرة).

G)عرض الكتف - (من نهايه الكتف إلى نهايه الكتف الاخر).

H)طول الكم - (من بدايه خط حياكة الكم مع الكتف الى نهاية كفة الكم السفلية).

I) طول الكم الداخلي- (من بدايه خط حياكة الكم مع الكتف الى نهاية فتحه السوار).

J)كفه الكم - (من بدايه كفه الكم إلى نهايتها).