Impact of Life Style Factors on Atherosclerosis: A Modelling Based Study

Xi He

Doctor of Philosophy

Aston University

September 2017

To my parents

Aston University

Impact of Life Style Factors on Atherosclerosis: A Modelling Based Study

Xi He

Doctor of Philosophy

September 2017

SUMMARY OF THE THESIS

Atherosclerosis is a low density cholesterol promoted medical condition in which the walls of the artery thicken due to fatty acid, cholesterol deposition (plaque). Such medical aggravations are known to escalate coronary and cardio-vascular heart diseases (CHD & CVD). This thesis models the time dynamical evolution of atherosclerosis, and in turn coronary heart disease (CHD), as a function of natural ageing and affectation due to life-style parameters like alcohol consumption, cheese consumption, smoking habit, high blood pressure, cereal-fruit-vegtable consumption. Principally based on data modelling (13 European countries, including the UK, have been analysed), followed by a continuum model based prediction, the thesis probabilistically estimates how a change in life style factors could help in controlling CHD/atherosclerosis.

The thesis is structured within three major sections. First, real data from open access databases (WHO & FAO) were analysed using standard statistical tools to establish dependence of CHD rates on the aforementioned lifestyle and ageing parameters. Two major conclusions could be drawn: a) linear dependence of all life style parameters on time, in the post-statin era; b) CHD death rate analysis demarcated the importance of statin usage in medical optimisation of life style factors.

Second, joint variation of (many, if not all) available parameters, including their inter-dependence, was analysed using machine learning based data visualization tools, like Principal Component Analysis (PCA) and NeuroScale (NSC). Two-fold conclusions were drawn: a) low dimensional clustering of high dimensional data established the interdependence of certain parameters; b) a key outcome of this research is the quantification of the moderating influence of the *healthy* lifestyle factors (fruit/vegetable and cereal consumption) on the *negative indicators* (systolic blood pressure, smoking, alcohol and cheesy food). This result is expected to lead to a major life saving tool for medical personnel in advising patients on what to eat, how much to eat, and what not to eat.

Combining information from the two sections above, a time varying model was developed that could predict how the population biology data based conclusions could be probabilistically projected to make future predictions of patient behaviour and concerned life expectations related to CHD deaths. This work is presently ongoing.

Keywords: Atherosclerosis, Life-style, Data Visualization, Machine Learning

PUBLICATIONS

1. *CHD Risk Minimisation through Lifestyle Control: A Population Biology Perspective* by **Xi He**, B. Matam, S. Bellary and A. K. Chattopadhyay; submitted to **Nature**.
2. *CHD Life expectancy augmentation through Lifestyle Control: an European overview* by **Xi He**, S. Bellary and A. K. Chattopadhyay; manuscript under preparation.
3. *Life Style Impact on Atherosclerosis and CHD: A Continuum Model Based Study* by **Xi He** and A. K. Chattopadhyay; work in progress.

ACKNOWLEDGEMENTS

First of all, I would like to express my sincere gratitude to my supervisor Dr. Amit K Chattopadhyay who has patiently advised me and continued to support me during my study. I am also thankful to him for insightful guidance in improving my work and endless knowledge sharing. Nothing in this short paragraph can express my gratitude.

I am thankful to Dr Srikanth Bellary for his inputs on biology and medical science, especially on the role of atherosclerosis in CHD.

Moreover, I am grateful to all members of the NCRG group, present and past, in particular Dr. Shahzad Mumtaz, Dr. Michel F. Randrianandrasana, Dr. Diar Nasiev and Dr. Rajeswari Matam who helped me in understanding the detailed nuances of mathematical programming, Dr Sotos Generalis who I always enjoyed with his stories since my undergraduate. Hearty thanks to Alex Brulo for advice related to cluster usage. I am also thankful to Sandra Mosley, Kanchan Patel and Susan Doughty for administrative support.

My completion of this study could not have been accomplished without the support of my colleagues and friends. I regret that I cannot mention them all by name here. I am very thankful to my friends Dr. Tianyu Qiu, Dr. Xueting Wang, Dr. Zhongyuan Sun, Dr. Chunhui Li, Lei He, Yingying Cai, Erika He for their patience accompanied with endless support. Many thanks are due to Dr. Lizi Harrison, Dr. Jordan Raykov, Shabnam Bibi, Gagan Aggarwal for the enjoyable time we had.

A very special thanks to Dr. Tom Davenport, for lots of very enjoyable teaching time, leisure time with him and his family.

I am grateful to my family: my husband, Dr. Mingchao Zhang who I met during my study years, for his love and support, my parents, grandparents and our extended family members for their unconditional love and care in my life.

Finally, I appreciate my precious little girl Jessenia Zixi Zhang who decided to join my life, thereby enriching and changing the meaning of my life since 2016.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

CAD Coronary Artery Disease

CHD Coronary Heart Disease

CVD Cardiovascular Disease

GTM Generative Topographic Mapping

GPLVM Gaussian Process Latent Variable Model

HDL High Density Lipoprotein

LDL Low Density Lipoprotein

NSC NeuroScale

ODE Ordinary Differential Equations

PDF Probability Density Function

PCA Principal Component Analysis

MeEU Mediterranean European Countries

ScEU Scandinavian European Countries

WeEU Western European Countries

1

INTRODUCTION

1.1 THE BIOLOGICAL PROBLEM

Cardiovascular (CVD) and Coronary Heart (CHD) diseases were responsible for nearly half (47%) of all deaths in Europe and 28% in the UK (Wilkins et al., 2017), according to 2017 WHO statistics. Approximately 70% of these numbers in either geographical domain is attributed to life style factors, like high blood pressure, smoking, cheesy (fat) food consumption, alcohol intake and lack of exercise, and a further 10% fluctuation is attributed to gender dependence. While it is also largely believed that ethnicity plays a key role in CVD and CHD related deaths, precise numerical estimation is yet to be confirmed. Based on extensive data analysis spanning 13 European countries, including the UK, we analyse the post statin (a cholesterol constraining drug) CVD statistics to predict timeline growth/decay rates of individual factors and then follow that up with detailed data mining studies to characterise and probabilistically predict the life style factor dependence of CVD.

Figure 1 can be plotted using the World Health Organization (WHO, 2017) data. This plot clearly shows that Sri Lanka and Japan seem to have the lowest death rates whereas Scotland and Northern Ireland are the worst offenders. In Europe, CVD/CHD problem assumes even greater importance.

Data also suggest that men are more likely to die of CVD than women (Lerner and Kannel, 1986). Studies show that the ratio of coronary disease related mortality in men compared to women vary between a wide range from 2.5:1.0 to 4.5:1.0 in countries with different rates of coronary artery disease (CAD) (Kalin and Zumoff, 1990).

Though this plot shows aggregate data from the past (30 years span), the problem in Europe is more acute than in other parts of the world, as is evidenced in Figure 1. In 2013, the Global Burden of Disease (GBD) study reported that there are 17.3 mil-

Figure 1: Age-standardized coronary disease death rates by sex in 1987 from 52 countries (Kalin and Zumoff, 1990)

lion CVD caused death globally, in which, over 4 million (almost a quarter) of CVD caused death occurs across the continent.

Figure 1 shows the wide differences in CHD death rates between countries, along with gender related variations of the same (Gerhard-Herman, 2002).

Generally, it has been predicted that life style parameters play key roles in high CVD/CHD mortality rates in Europe. The problem becomes more complicated when we note gender variation even within the same country. This has been clearly demonstrated in a recent work where the authors observed such gender differences in CHD suggesting that the probability of CHD in female populations are substantially lower than those for the male sectors, both data taken from the same country specific database (Isles et al., 1992). As our later analysis will show, this prediction is correct; but in the process, we also evaluate the exact mortality rates for each gender group that would enable us to make probabilistic predictions of atherosclerotic CHD deaths for some future point in time.

1.1.1 *Atherosclerosis and Coronary Heart Disease (CHD)*

Atherosclerosis is caused by the deposition of excessive fatty substances (mostly cholesterol[1]) in the interior walls of the arteries in the body(Russell and Cohn, 2012), include arteries in the heart, brain, arms, legs etc. As a result, different serious diseases may develop based on which arteries are affected knowns as CVD, and which atherosclerosis occurs when plaque builds up in the coronary arteries and these arteries supply oxygen-rich blood to the heart, this results in hardening and narrowing of the coronary arteries (Awojoyogbe et al., 2011; Nhs.uk, 2013), often leading to constriction, followed by haemorrhage. It is a potentially serious condition can lead to serious problems, including CHD (but CHD is not imperative of a lack of atherosclerosis), heart attack[3], stroke[4], or even death (Ross, 1993).

Figure 2: Formation of atherosclerosis Source: (Phillips-fit, 2013)

Investigations on atherosclerosis date back to ancient times. "Vessels in the elderly restrict the transit of blood through thickening of the tunics." was the first description of atherosclerosis by Leonardo da Vinci (1452-1519) (Slijkhuis, Mali and Appelman, 2009). Several hundred years later, atherosclerosis (and the resulting cardiovascular diseases) has become one of the most prevalent causes of mortality in the UK and all over the 'developed' world (George and Johnson, 2010). It has replaced infectious

1 **Cholesterol**:($C_{27}H_{46}O$)is a fatty substance known as a lipid[2] and is vital for the normal functioning of the body. It is an essential structural component of animal cell membranes and is required to build and maintain proper membrane permeability and fluidity. It is mainly made by the liver but can also be found in some foods we eat.

3 **Heart attack** is a serious medical emergency in which the supply of blood to the heart is suddenly blocked, usually by a blood clot. Lack of blood to the heart can seriously damage the heart muscle.

4 **stroke** is a serious medical condition that occurs when the blood supply to part of the brain is cut off. The brain needs the oxygen and nutrients provided by blood to function properly. If the supply of blood is restricted or stopped, brain cells begin to die.

disease as the leading cause of death in the developed world (Thompson et al., 2013), accounting for one in three deaths around the globe.

This problem is the most pronounced in high-income countries (HICs) (Wilson and O'Donnell, 2018). Each year atherosclerosis (and the resulting cardiovascular diseases) causes over 4 million deaths here that account for 47% of all deaths in Europe (52% for women and 42% for men) (Nichols et al., 2012).Although rates of death from atherosclerosis have fallen to 24% from 28% over the last three decades (López-Candales, 2002), Göran and Hansson have predicted it will also be the leading cause of death within the next 7 years in most developing countries and in eastern Europe (Hansson, 2005). In other words, this is a global problem, not restricted to the precincts of a few countries only. The most recent analysis shows, in 2015 CHD affected 110 million people and resulted in 8.9 million deaths (Haidong et al., 2016; Vos et al., 2016). Which holds 15.9% all causes of death globally (Vos et al., 2016). Therefore, In this study, the investigations are concentrate on CHD.

1.1.2 *Risk Factors*

Studies show that the exact causes of atherosclerosis are known only qualitatively. Not much is known is terms of numbers. For example, although it is well acknowledged that smoking increases the probability of atherosclerotic CHD (Wilhelmsen, 1988), nothing precise is known as to what would be the change in the CHD mortality rate if the smoking habit changes by, say, r%. The problem becomes more pronounced technically when we come to understand that not one but multiple such life style factors are liable to change simultaneously. As an example, a heavy drinker is often found to be a smoker as well (Bien and Burge, 1990). So if his/her drinking increases by a certain fraction, his/her smoking too may increase (or decrease). Literal data do not depict a quantitative picture of such simultaneous variation of affecting factors which is where mathematical analysis is unavoidable. Ours is an effort in this direction.

In 1973, nine probable risk factors leading to atherosclerosis were analysed based on predictions from a multiple logistic model (Lee, 1986; Murray, 2002), which included a high ratio of cholesterol, smoking, hypertension, etc. as the major affecting

factors (Wilhelmsen, Wedel and Tibblin, 1973). Some other well known risk factors (negative indicators) of atherosclerosis include family history, lack of exercise, high fat diet, diabetes. Due to the lack of any quantifiable data on some negative indicators, in this thesis, we only studying on 4 negative indicators: Alcohol Consumption, Cheese Consumption, Smoking Habit, High Blood Pressure. And accompanied with two positive indicators: Cereal consumption, Fruit and Vegetable Consumption.

1.1.2.1 *Alcohol Consumption*

"Alcohol has a bi-form relationship with CHD"- while controlled consumption (no more than 10g) every day appears to protect the cardiovascular system (Gunzerath, Zakhari and Warren, 2004; Møller, Anderson and Moloney, 2010) by raising the levels of HDL cholesterol (Klatsky, 1999), heavy drinking at a higher frequency increases the risk of CHD (Anderson and Baumberg, 2006).

However, some studies (Han et al., 2013; Ikehara et al., 2013) seem to conclude very differently in connection with CHD lowering due to low alcohol consumption as has been claimed in the studies above. A meta-analysis study provides strong evidence that whatever be the volume of alcohol consumed, there are two major enzymes in alcohol that may be associated with increasing the risk of coronary artery disease (CAD) (Han et al., 2013). The Japan Public Health centre-based prospective study examined a sample consisting of 47,100 women aged 40-69 years during 1990 to 2009. Results seem to indicate that there was no association between alcohol consumption and risk of CHD (Ikehara et al., 2013). Once again, this may point to the fact that occurrences of atherosclerosis, as well as CHD, are country specific with separate behavioural issues for the country concerned.

1.1.2.2 *Cheese Consumption (High Fat Diet)*

Another major contributing factor to the cause of CHD is the high saturated-fat diet usage. Statistics show that the risk of having a CHD can be reduced by 14% if the saturated fat content in the diet is reduced in diets (Hooper et al., 2012). Another finding suggests that consuming polyunsaturated fats instead of saturated fats reduces the risk of CHD as the HDL concentration increases (Mozaffarian, Micha and Wallace, 2010).

Although most of the studies indicate High-fat diet as being a key risk factor for CHD growth, not all studies conclude the same (De Oliveira Otto et al., 2012). A multi-ethnic study of atherosclerosis investigated the links between saturated fat consumption from a variety of food and the eventual occurrence rate of CHD. Data collected from 2000-2010 could not necessarily conclude about any specific link between saturated fat intake and CHD death rate, it may depend on specific fatty acids present in some food, in addition to saturated fats (De Oliveira Otto et al., 2012). Our analysis based on cheese consumption data seems to give credence to this observation, as opposed to popular dictums.

Cheese consumption has a very complex connection with CHD. It can be classified as dairy food, same as milk, yoghurt, has been connected with the reduced risk of CHD. A large number of studies show that dairy consumption does not associate with increased CHD incidence or even death. In other words, while small volumes of cheese may aid in the coagulation process, thereby acting as a CHD deterrent, high volumes of cheese could constrict the blood vessels through atherosclerotic plaque growth.

Currently, there is no known study clearly establishing a quantitative relationship between well-defined High-fat food and the risk of CHD, a lateral reason for our consideration of cheese as a key parameter in our analysis.

1.1.2.3 *Smoking Habit*

Tobacco is known to cause at least 20% yearly deaths in England alone that further combines with 14% casualty from heart disease.

The carbon dioxide emanated from cigarettes thickens the blood which reduces effective oxygenated blood flow to the heart. This constrains the myocardium's ability to use oxygen to generate more adenosine triphosphate, leading to a higher risk of developing several chronic disorders, which includes atherosclerosis. Heavy cigarette smoking is agreed to be a major hazard associated with health; it contributes significantly to CHD deaths.

Comparing between non-smokers who live with smokers and smokers living with non-smokers, meta-analyses have shown that the risk of coronary heart disease is around 30% greater (Glantz and Parmley, 1991, 1995; Kritz, Schmid and Sinzinger,

1995; Steenland, 1992; Wells, 1994) for the first group as a result of so-called 'passive smoking'. Many studies have reported that passive cigarette smoking increases the risk of CHD (He et al., 1999; Law, Morris and Wald, 1997) and it is a major preventable risk factor for CHD (Parish et al., 2000).

1.1.2.4 *High Blood Pressure*

High blood pressure is often referred to as *t*he silent killer. Most people with high blood pressure (also known as hypertension) do not have any high blood pressure symptoms since the effects occur inside the body.

High blood pressure can damage the artery walls, making arteries stiffer and narrower, that in turn leads to insufficient blood flow into the heart the heart muscles. This can cause cardiovascular disease, or even death.

Isolated systolic hypertension[5] has been investigated as an important negative indicators increasing the death rate of CHD for both women and men, especially in 45-64 years' age group (Antikainen, Jousilahti and Tuomilehto, 1998). Also, research shows that not only diastolic hypertension but also isolated instances of systolic hypertension are responsible for atherosclerosis (Kannel, Dawber and McGee, 1980).

1.1.3 *"Good Men" at Work: Cereal, Fruit and Vegetable*

Cereals, fruits and vegetables provide a significant part of life nutrition, they are universally advocated as health foods, which supply vitamins and minerals to the diet, and also include a diverse group of plant foods contains various nutrients, dietary fibre. Studies have shown that fibre intake lowers CVD (Kim and Je, 2016; Liu et al., 2002). There has been some statistic based suggestions that whole grain, fruits and vegetables intake may prevent CHD, which associated with reducing risk of CHD and CVD (Aune et al., 2016; Dauchet et al., 2006).

Some minor opposite voice is standing out with an over 11-7 follow-up study of the relations of cereal, fruits and vegetables intakes with the risk of total mortality and the incidence of coronary artery disease (CAD) and ischaemic stroke in the African

5 **Isolated systolic hypertension**: A type of hypertension (high blood pressure): The diastolic blood pressure (DBP) under a normal range (< 90 mm Hg), but the systolic blood pressure is greater than a normal range (> 160 mm Hg).

American and white men and women (Steffen et al., 2003), the study shows there did inversely associate with total mortality and the incidence of CAD, but not on the risk of ischaemic stroke.

Accordingly, from a scientific perspective, the promotion of cereal, fruit and vegetable consumption could be a preferable strategy to decrease the burden of CHD in Western countries (Boeing et al., 2012). Public Health England (PHE) has launched the refreshed UK's healthy eating model plate (PHE, 2017):

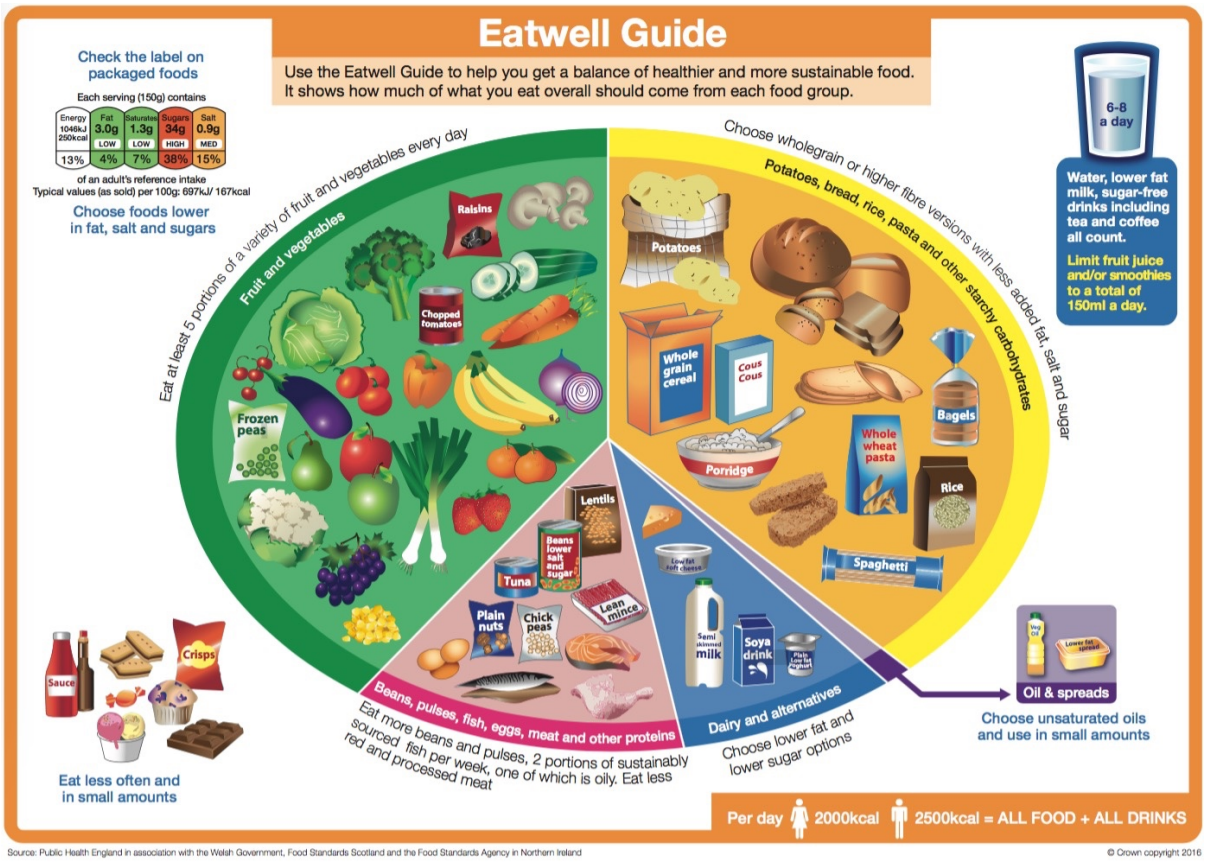


Figure 3: Eatwell Guide, Source: (PHE, 2017)

To mitigate the risk of heart disease, stroke and obesity, PHE recommends:

- Eat at least 5 portions of the variety of fruit and vegetables every day.
- Consume meals based on potatoes, bread, rice, pasta or other starchy carbohydrates; choosing wholegrain versions where possible.

A key component of this research is to precisely quantify what percentage volume change in fruit-vegetable and/or cereal intake could lower BP, and in turn smoking, etc. eventually

lowering the CHD afflicted health situation. Our analysis here relays a formula that connects all such affecting factors in precise numerical terms.

1.2 COUNTRIES STUDIED

Due to the high CHD death rate in European countries, our research focuses on 13 European countries, including the UK. Bearing in mind the traditional food and lifestyle habits of the different European nations, we have grouped the selected countries combining their geographical location with life and dietary requirements (Berglee, 2012; King, Proudfoot and Smith, 2014; Rossi, 2015):

- United Kingdom (UK)
- Mediterranean European Countries (MeEU) Block - France, Greece, Italy and Spain
- Scandinavian European Countries (ScEU) Block - Denmark, Finland, Iceland, Norway and Sweden
- Western European Countries (WeEU) Block - Germany, Netherlands, Switzerland

1.3 CREATION OF DATABASES

This section further classifies the grouping structure mentioned above:

1.3.1 *Data Source and characteristic*

The dataset consists of 13 European countries, including the UK, where each country's data consists of features such as alcohol consumption, cheese consumption, smoking habit, systolic blood pressure, cereal consumption, fruits and vegetables consumption, smoking. For features such as CHD death rate, smoking habit and systolic blood pressure, we have gender specific details.[6]. It is also important to note

6 **Missing Data** Smoking and Systolic blood pressure have missing values for certain years that we approximate using extrapolation from our linear model detailed below

that the absolute value of the variables have widely different ranges that we later renormalise (details below).

CHD death rate: Calculated by: SDR[7] of ischaemic heart disease/SDR of all causes, including all ages, per 100,000, males and females separately.

Source of data: WHO Global Health Observatory Data Repository, available from `http://apps.who.int/ghodata/`;

collected year: 1970-2014

Alcohol Consumption: Recorded adult (15+ years) per capita (APC)[8] consumption of pure alcohol. In order to make the conversion into litres of pure alcohol, the alcohol content of beer, wine and spirits is considered to be 5%, 12% and 40% respectively. Alcohol consumption here includes all alcoholic drinks.

Source of data: WHO Global Health Observatory Data Repository, available from `http://www.who.int/substance_abuse/publications/global_alcohol_report/en/index.html`;

collected year: 1970-2015

Cheese Consumption: is the total supply amount of cheese per capita per year in kilogram.

Source of data: FAO Statistics Division, Food Supply Sheets, available from `http://www.fao.org/faostat/en`;

collected year: 1970-2013

Smoking Habit: was measured using the standard questionnaire during a health interview of a representative sample of the population aged 15 years and above. Many countries are carrying out such health interview surveys on a more or less regular basis. However, most of the data are collected from multiple sources by the Tobacco or Health unit at WHO/EURO, males and females separately.

Source of data: WHO Global Health Observatory Data Repository, available from

7 **SDR:** This is the age-standardised death rate calculated using the direct method, i.e. represents what the crude rate would have been if the population had the same age distribution as the standard European population.

8 **Recorded APC:** This is defined as the recorded amount of alcohol consumed per adult (15+ years) over a calendar year in a country, in litres of pure alcohol. The indicator only takes into account the consumption which is recorded from production, import, export, and sales data often via taxation.

http://apps.who.int/ghodata/;

collected year: 1980-2015

Systolic Blood Pressure: This is the mean systolic blood pressure trends, age-standardised (mmHg), males and females recorded separately.

Source of data: WHO Global Health Observatory Data Repository, available from *http://apps.who.int/gho/data/view.main.12467EST?lang=en*;

collected year: 1980-2015

Cereal Consumption: This is the average amount of cereal consumed per person per year in kilograms, excluding the carbohydrate content in beer.

Source of data: FAO Statistics Division, Food Balance Sheets, available from *http://www.fao.org/faostat/en*;

collected year: 1970-2013

Fruit and Vegetable Consumption: This is the average amount of fruits and vegetables consumed per person per year in kilograms, excluding the fruit content in wines.

Source of data: FAO Statistics Division, Food Balance Sheets, available from *http://www.fao.org/faostat/en*.

collected year: 1970-2013

1.3.2 *Data Usage*

The recorded data span multiple year ranges that have not always been consistently recorded. It is important to standardise each database, including a general time span that, for practical purposes, in agreement with most available data bases for all the different variables used, we consider from 1970 to 2013. This provides a broad range spanning ca 43 years for our population biology based data modelling.

1.3.2.1 *Statin Usage*

The plot showed in Figure 4 describes the CHD death rate versus time from 1970 to 2013 in the UK, the collected data belong to two clearly separate regimes. This regime differentiation is based on the usage of a group of popular drugs, called

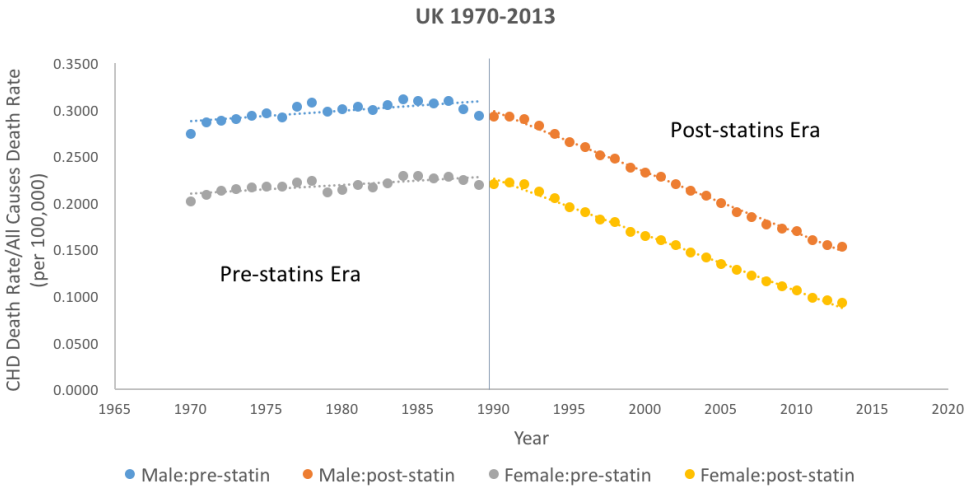


Figure 4: Plot of CHD death rate of UK versus time, 1970-2013

'statins' (Endo, 2010), that reduce the level of low-density lipoprotein (LDL) cholesterol content in the body and thereby decreases the probability of atherosclerosis, and hence that of CHD as well (Endo, 1992). Statins were first introduced into the market in September 1987. By 1989, they have been developed and launched internationally, and made available to health science practitioners worldwide. Across the range of all 13 countries studied, the statins launched year are vary, but by browsing the data, we find that in the pre-statin era, with increasing year number, the CHD death rate increased with time while the gradient reversed in the post-statin phase. We then separated the data into two eras: *Pre-statin and Post-statin*. Since the entire emphasis of our work is in calibration-prediction of atherosclerosis for future years, the focal regime of our interest is the post-statin era. With this (Post-statin era) in mind, the choice of raw database is stipulated from 1990 to 2013 which are tabled in Appendix A.

1.3.2.2 *Data Normalisation*

As different factors (variables) is each dataset have different ranges of value, we therefore transform all variables on the similar ranges by applying a linear transformation (Standard Score, or more commonly referred to as Z-score transformation). For this purpose, we consider each variable as an independent variable.

First we compute mean (μ) and standard deviation (σ) as

$$\mu_i = \frac{1}{N}\sum_{n=1}^{N} x_i \tag{1}$$

$$\sigma_i = \sqrt{\frac{1}{N-1}\sum_{n=1}^{N}(x_i^n - \mu_i)^2}, \tag{2}$$

where $n = 1, \cdots, N$ are the indexes of the data points, and μ_i and σ_i are the mean and standard deviation of i-th variables respectively. Finally the values can be scaled as

$$\tilde{x}_i^n = \frac{x_i^n - \mu_i}{\sigma_i}, \tag{3}$$

where $\tilde{x}_i^n$ represents the scaled value of the ith variable for the nth points. Examples of the UK raw and normalised scale features as probability density functions (PDFs) are given in Figure 5 and Figure 6 respectively.

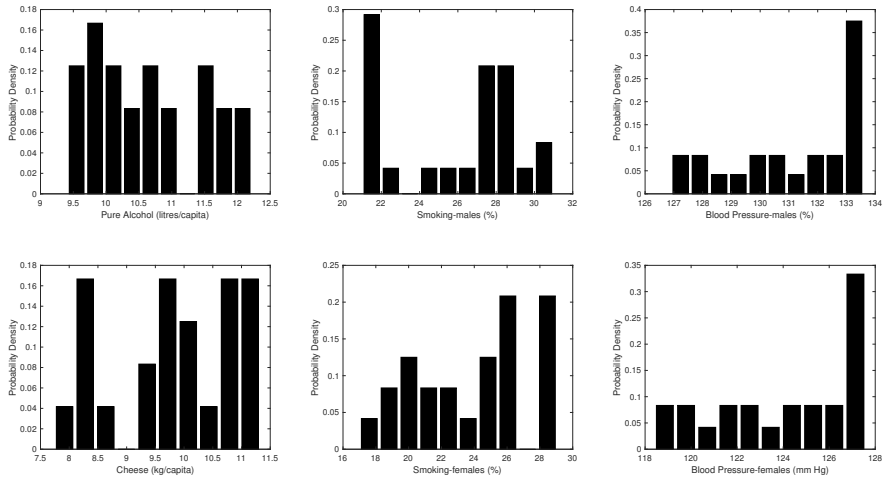


Figure 5: PDF plots of the raw data features for UK.

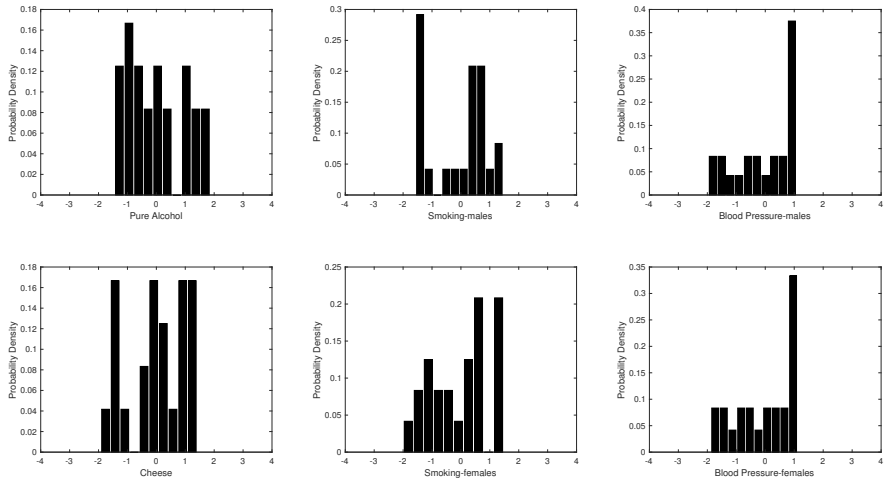


Figure 6: PDF plots of the normalised data features for UK.

Histograms are data frequency charts that mathematically represent the probability density functions of such data. For example, in the first plot of Figure 5, the quantity of alcohol consumed is subdivided into 10 groups (horizontal axis), the number of year which falls into each group is counted, and the counts are converted to percentage (vertical axis). In this histogram (Figure 5), the first bar tells us that 12.5% of a total of 24 years have consumed alcohol between 9.5 to 9.75 Litres/capita, the second bar shows 16.67% of 24 years consume between 9.75 to 10 Litres/capita, which is the peak point of the graph, and so on.

The PDFs play an important role in the investigation of each risk factor in the following study. The relevant histograms for the other 12 European countries are shown in Appendix B respectively.

1.4 OBJECTIVES

This strongly interdisciplinary research aims to build a model of functional interrelationship between 6 life-style parameters, which are alcohol consumption, cheese consumption, smoking habit, cereals, and fruit-veg consumptions. Our objective is to develop a regime of statistical analysis based (deterministic and stochastic) data and mathematical modelling that would be able to predict Coronary (CHD) and Cardiovascular disease (CVD) inflicted death rate probabilities with percentage changes in life style habits, on a person to person basis.

Although this is a key medical problem of the modern era, not enough has been done in connection with prognosis directed theoretical analysis. In the longer run, we hope to bridge this gap by providing a black box tool to medical personnel that would enable them to make fast, probabilistic predictions of patient mortality based on present measures of life style conditions.

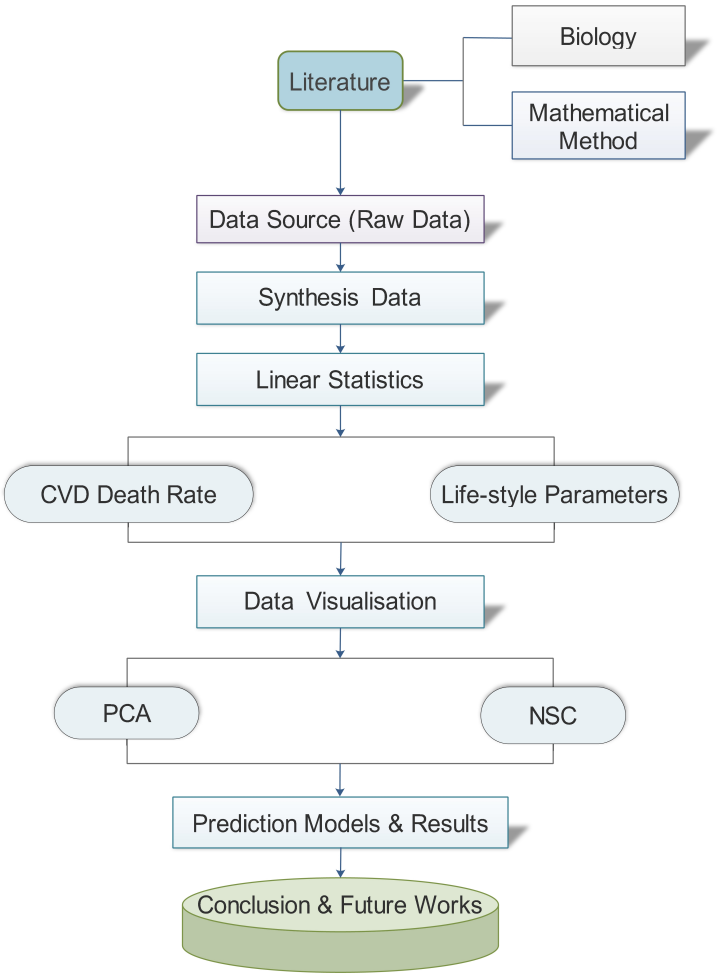
The work presented in this thesis analyses the interaction between alcohol consumption, cheese consumption, smoking habit, high blood pressure, cereal-fruit-veg consumption with ischaemic Heart Disease death rate for 13 European countries, including the UK. CVD death data collected over 40 years for all 13 countries were least square fitted[9], clearly showing linear trends across both genders, over all years and non-dimensionalised[10] against total deaths. Six life-style factors clearly showed linearly evolving (decaying or growing, depending on country and gender) trends with time. These data were then mined using dimensional regularisation techniques, selectively choosing PCA and NSC, as the respective representatives of linear and nonlinear visualization tools. Target outcomes:

1. Rank affecting factors in their order of importance and

2. Predict how fluctuations in individual parameters, and collectively, will probabilistically affect the CVD rate in the future.

In other words, results from this research are expected to serve as a guidance tool for medical personnel in advising how a life style change will probabilistically lower the chance of cardiovascular death rates.

A key challenge here is to construct effective models that can identify the nature of these relationships between the interactions and then numerically estimate the strength of such correlations. This research will lay down a mechanism to use past data based on predicting future life-style affectation of CHD/CVD death rates.



1.5 THESIS OUTLINE

The flowchart below consummate the thesis in five complementary chapters; following are individual chapter summaries:

CHAPTER 1: Introduction; This chapter reviews the basic biology of atherosclerosis and Coronary Heart Disease (CHD), along with summarises current studies involving the negative indicators considered in our study that are expected to affect them: alcohol consumption, cheese consumption, smoking and high blood pressure; also cereal consumption, fruit and vegetables consumption serve as

9 **Least square fitting:** A mathematical method to construct a best-fitting curve to a given set of data points on the basis of the residuals of the data points

10 **non-dimensionalisation:** 'Scaling' or 'Normalising', means the removal of units from physical quantities by certain substituted variables, the non-dimensionalisation method used in this thesis explained in subsubsection 1.3.2.2.

the positive indicators. A later part of our analysis will also focus on the multivarying inter-relation between blood pressure and the other three factors.

Followed by the biology understanding, countries selection and grouping are introduced with the data source and the selection of the data. We have collated data over 13 European countries, the country to be subdivided into four separate sectors - Mediterranean countries (Italy, Spain, France, Greece), Scandinavian countries (Denmark, Norway, Iceland, Finland, Sweden), Western Countries (Germany, Netherlands, Switzerland) and the United Kingdom. Focusing more on the UK data, due to closer association with the NHS. The subdivision is based on traditional CHD inflictions food and other habits. The setting up of the database used in this thesis are explained.

CHAPTER 2: We use real data of each country to produce a linear regression on each of the parameter, including 4 risk parameters and 2 potentially positive parameters ('parameters' or 'variables' are often interchanged in our description in conformity with the biological literature; the definitions become a lot more precise in the mathematical modelling part). Our data analysis confirms a linear regression between CHD rates of the different countries (both pre- and post-statin eras) with the time concerned. Results from all countries have been presented.

CHAPTER 3: In this chapter, we review the data mining methods used in this thesis: Principal Component Analysis (PCA), NeuroScale (NSC), we also review the other two projection algorithms: Generative Topographic Mapping (GTM), Gaussian Process Latent Variable Model (GPLVM) for a comparison with the systems we have chosen. We then train our data using PCA and NSC. In passing, we also emphasise the reason of restricting our dimensional visualization choice only to PCA and NSC.

CHAPTER 4: We demonstrate experimental results for synthetic datasets in order to show their effectiveness. We then rank all affecting negative indicators in an ascending order of importance, including the tolerance estimation of this analysis (expressed as standard deviation) by using of PCA, NSC and SPSS.

In our results, we rank each of the risk parameters and develop a prediction model.

CHAPTER 5: We develop a time varying model that could predict how all these data based conclusions could be probabilistically projected to make future predictions of patient behaviour and concerned life expectations related to CHD deaths. This work is presently ongoing.

CHAPTER 6: This chapter summarises the outcomes from each chapter and also discusses possible future extensions of this work, emphasising the inclusion of two key missing elements, that of 'ethnicity' and 'diabetes', along with the present set of 6 parameters/variables.

APPENDIX A: Data sets we have used in this thesis for all 13 countries.

APPENDIX B: Histogram of Probability Density Functions for 12 European countries except UK.

APPENDIX C: The plots and linear regression results of 6 life-style parameters for 12 European countries except UK.

APPENDIX D: PCA visualization for 12 European countries except UK.

2

STATISTICAL DATA MODELLING

This chapter represents statistical analysis of available databases, leading to linear regression relations connecting the CHD rates of each of the countries considered against their time of evolution, analysis are detailed in section 2.2. Similar linear relationships are also found to be true for all 6 life-style parameters (alcohol consumptions, cheese consumption, smoking habit, systolic blood pressure, cereal consumption quantity and fruit and vegetables) varying with time, details of UK showed in section 2.3 and details of the other 12 countries can be found in Appendix D.

We have generically used least square fitting mechanism for both linear regressions mentioned above.

2.1 LINEAR LEAST SQUARE

Linear least square fitting (Plackett, 1950) is the most common method of linear regression and approach a best straight line fitting solution for a set of points. Suppose we have a set of (x, y) points: $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, and looking for a best-fitted line: $y = \beta_1 + \beta_2 x$ of this dataset. In the other words, we supposed to find out the β_1 and β_2 best solve the following equations system (Lawson and Hanson, 1995)

$$\begin{cases} \beta_1 + x_1\beta_2 = y_1 \\ \beta_1 + x_2\beta_2 = y_2 \\ \qquad \vdots \\ \beta_1 + x_n\beta_2 = y_n \end{cases} \tag{4}$$

The residual, means the error at each of the point between the linear fit, which is the difference beside the equation system above. The least square proceeds by finding

the sum of the squares of the deviations of this set of data as small as possible, which is to find the minimum of the function

$$R(\beta_1, \beta_2) = [y_1 - (\beta_1 + x_1\beta_2)]^2 + [y_2 - (\beta_1 + x_2\beta_2)]^2 + \cdots + [y_n - (\beta_1 + x_n\beta_2)]^2 \tag{5}$$

The condition for R to be a minimum is

$$\begin{cases} \frac{\partial R}{\partial \beta_1} = a\beta_1 + b\beta_2 + c = 0 \\ \frac{\partial R}{\partial \beta_2} = d\beta_1 + e\beta_2 + f = 0 \end{cases} \tag{6}$$

Where a, b, c, d, e, f are all constant calculated by deviating of Equation 5. By solving this two equations with two unknowns system, we have $\beta_1 = \frac{bf-ce}{ae-bd}; \beta_2 = \frac{af-cd}{bd-ae}$, the equation of best-fitted line is

$$y = \frac{bf - ce}{ae - bd} + \frac{af - cd}{bd - ae}x \tag{7}$$

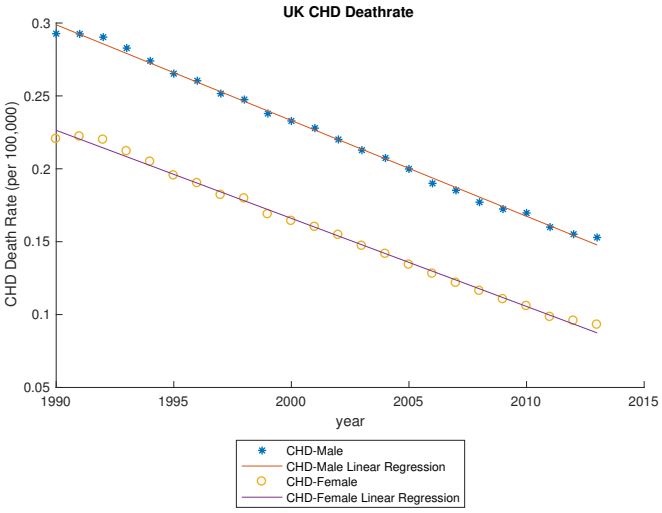
2.2 LINEAR REGRESSION OF CHD DEATH RATE

In section 2.1 the method of statistical analysis is briefed. The purpose of this section is to generate the linear regression relations connection the CHD rates of each of the countries considered against their time using Linear least square method.

2.2.1 *United Kingdom*

Figure 7 shows the scatter plot of raw real database of UK CHD death rate from 1990 to 2013, and a linear fitted equation generated. The table in Figure 7 shows the regression statistics, the number we most interested in is:

Mutiple R: Known as Pearson's Correlation Coefficient, R is a measure of the strength of the linear association between two variables. It takes a value between +1 (a perfect positive correlation between the two variables, such that an increase in one of them is matched by a set amount of increase in the other) and -1 (a perfect



UK-male Regression Statistics	
Multiple R	0.998490152
R Square	0.996982584
Adjusted R Square	0.996845429
Standard Error	0.002611695
Observations	24

UK-female Regression Statistics	
Multiple R	0.99805026
R Square	0.996104322
Adjusted R Square	0.995927246
Standard Error	0.002731414
Observations	24

Figure 7: UK Linear plot and regression fit of CHD deathrate

negative correlation, such that an increase in one of the variables is always matched by a set amount of decrease in the other). $R = 0$ means that the two variables are not correlated at all. In this case, both UK CHD death rate of male and female all have R larger than 0.998, which means there is a nearly perfect correlation between UK CHD death rate and time either male or female.

Table 1: ANOVA test statistic of UK CHD death rate

UK-male	*df*	*SS*	*MS*	*F*	*Significance F - ρ*
Regression	1	0.049582	0.049582	7269.007	3.18E-29
Residual	22	0.00015	6.82E-06		
Total	23	0.049732			

UK-female	*df*	*SS*	*MS*	*F*	*Significance F - ρ*
Regression	1	0.041968	0.041968	5625.284	5.28E-28
Residual	22	0.000164	7.46E-06		
Total	23	0.042132			

Table 1 shows the results of Analysis of Variance (ANOVA). 'F' is the value of test statistic, for evaluating if the value of F is large enough that it is unlikely to have occurred by chance, two sets of degrees of freedom need to be taken in to account, they are 'df1' and 'df2'. For our purposes, *'Significance F - ρ'* is the one of the most interesting value. It shows the probability of getting a value of F as large as our obtained one, merely by chance. If *'Significance F - ρ'* is smaller than

0.05, we can conclude that our value of 'F' is large enough that it is unlikely to have occurred by merely chance, which means our regression line is fitting better to the data than a model just based on using the mean of the values for the predicted variable. In this case, $F(1, 22) = 7269.007, \rho = 3.18 \times 10^{-29}$ for male and $F(1, 22) = 5625.284, \rho = 5.28 \times 10^{-28}$ for female, both ρ-values are largely smaller than 0.05, so we can conclude that in this model, the regression line is significantly better at predicting the CHD death rate from time scale than if merely use the mean CHD death rate each time.

Based on the above analysis, we extrapolate the UK CHD death rate data to obtain the following linear formulae:

$$y_{CHD_{male}} = -0.00657\ t + 13.36542 \tag{8a}$$

$$y_{CHD_{female}} = -0.00604\ t + 12.24803 \tag{8b}$$

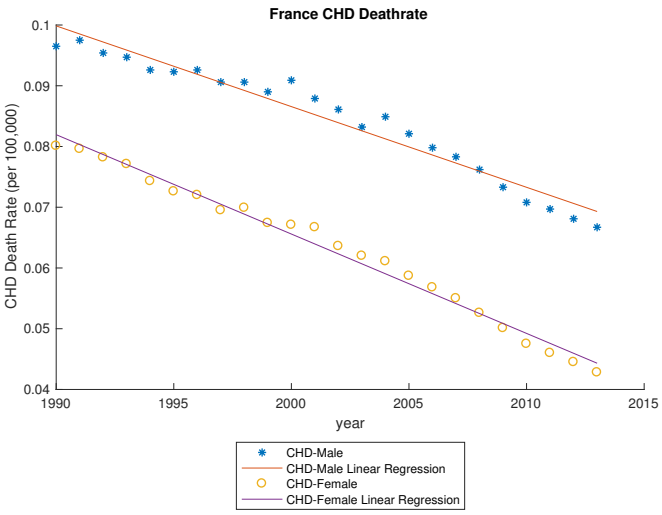
CHD death rate for male and female groups are all showing a decreasing trend with time. This generally represents a growing public awareness of health, that is also possibly related to the availability and use of statin, although no conclusive statistical analysis confirming either is known yet. The numbers 13.36542 and 12.24803, respectively for male and female categories, suggest the pre-statin rates which are higher than the following years.

Same method applied for analysing the other 12 european countries, the results of each country show in the subsection 2.2.2 to subsection 2.2.4.

2.2.2 *Mediterranean European Countries (MeEU) Block*

France

As shown in Figure 8, the Pearson's Correlation Coefficients for both males and females are larger than 0.9 which means there is a near perfect correlation between CHD death rate and time in France, for both genders.



France-male Regression Statistics	
Multiple R	0.975933915
R Square	0.952447006
Adjusted R Square	0.950285506
Standard Error	0.00214604
Observations	24

France-female Regression Statistics	
Multiple R	0.993460331
R Square	0.986963429
Adjusted R Square	0.986370857
Standard Error	0.001359098
Observations	24

Figure 8: France Linear plot and regression fit of CHD deathrate

Table 2: ANOVA test statistic of France CHD death rate

France-male	*df*	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.002029	0.002029	440.6417	4.84E-16
Residual	22	0.000101	4.61E-06		
Total	23	0.002131			

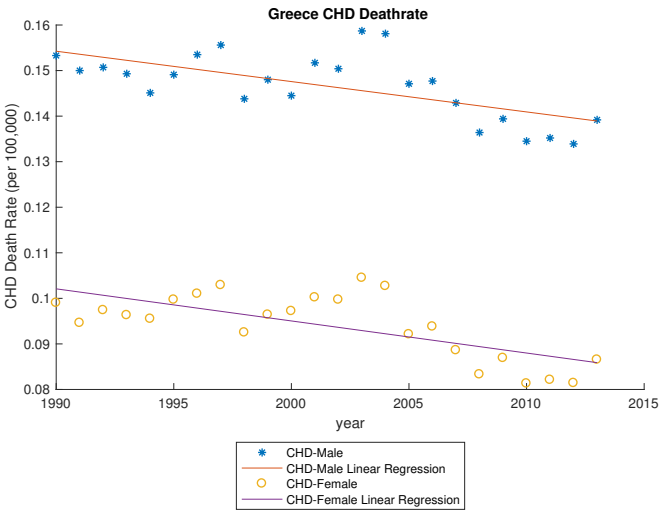
France-female	*df*	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.003077	0.003077	1665.56	3.13E-22
Residual	22	4.06E-05	1.85E-06		
Total	23	0.003117			

As can be seen from Table 2, a linear regression was performed to predict the CHD death rate as for the UK data. The significant regression equations are:

$$y_{CHD_{male}} = -0.00133\ t + 2.74338 \tag{9a}$$

$$y_{CHD_{female}} = -0.00164\ t + 3.33682 \tag{9b}$$

at $(F_{male}(1,22) = 440.642, p < 0.0001)$, $R^2_{male} = 0.952$ for males, and $(F_{female}(1,22) = 1665.560, p < 0.0001)$, with $R^2_{female} = 0.987$ for females.



Greece-male Regression Statistics	
Multiple R	0.647995
R Square	0.419897
Adjusted R Square	0.393529
Standard Error	0.005653
Observations	24

Greece-female Regression Statistics	
Multiple R	0.691093
R Square	0.477609
Adjusted R Square	0.453864
Standard Error	0.005328
Observations	24

Figure 9: Greece Linear plot and regression fit of CHD deathrate

Greece

Observed from Figure 9,both males and females' Pearson's Correlation Coefficient of CHD death rate in Greece are all larger than 0.6, which means the correlation between CHD death rate and time in Greece is not perfect but still performed a good correlation, either males or females.

Table 3: ANOVA test statistic of Greece CHD death rate

Greece-male	*df*	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.000509	0.000509	15.92433	0.000617
Residual	22	0.000703	3.2E-05		
Total	23	0.001212			

Greece-female	*df*	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.000571	0.000571	20.11407	0.000185
Residual	22	0.000625	2.84E-05		
Total	23	0.001196			

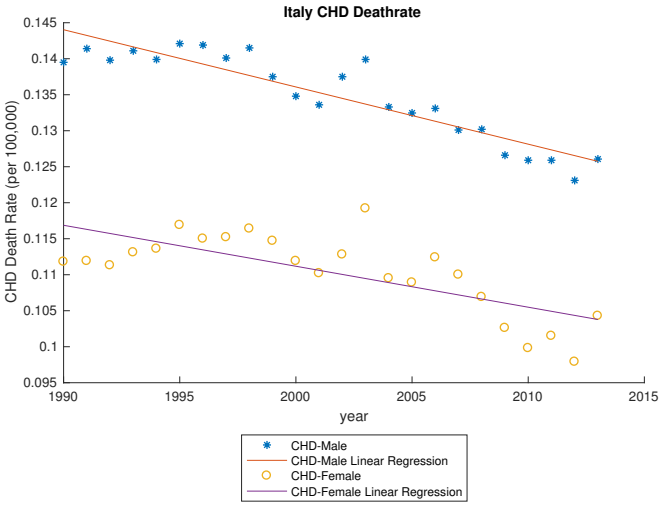
As can be seen from Table 3, a linear regression was performed, to predict CHD death rate based on time scale in Greece. The significant regression equation are:

$$y_{CHD_{male}} = -0.00067\ t + 1.47796 \tag{10a}$$

$$y_{CHD_{female}} = -0.00070\ t + 1.50436 \tag{10b}$$

at ($F_{male}(1,22) = 15.924, \rho < 0.0007$), $R^2_{male} = 0.420$ for males, and ($F_{female}(1,22) = 20.114, \rho < 0.0002$), with $R^2_{female} = 0.478$ for females.

Italy



Italy-male Regression Statistics	
Multiple R	0.91457
R Square	0.836438
Adjusted R Square	0.829003
Standard Error	0.002538
Observations	24

Italy-female Regression Statistics	
Multiple R	0.720871
R Square	0.519655
Adjusted R Square	0.497821
Standard Error	0.003953
Observations	24

Figure 10: Italy Linear plot and regression fit of CHD deathrate

Observed from Figure 10, males' Pearson's Correlation Coefficient of CHD death rate in Italy is 0.915, which performed a nearly perfect correlation between CHD death rate and time. For female, there is a slight lower Multiple R value, which is 0.721, it is not a perfect correlation, but still remain as a good one.

Table 4: ANOVA test statistic of Italy CHD death rate

Italy-male	*df*	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.000725	0.000725	112.5052	4.09E-10
Residual	22	0.000142	6.44E-06		
Total	23	0.000866			

Italy-female	*df*	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.000372	0.000372	23.80043	7.07E-05
Residual	22	0.000344	1.56E-05		
Total	23	0.000716			

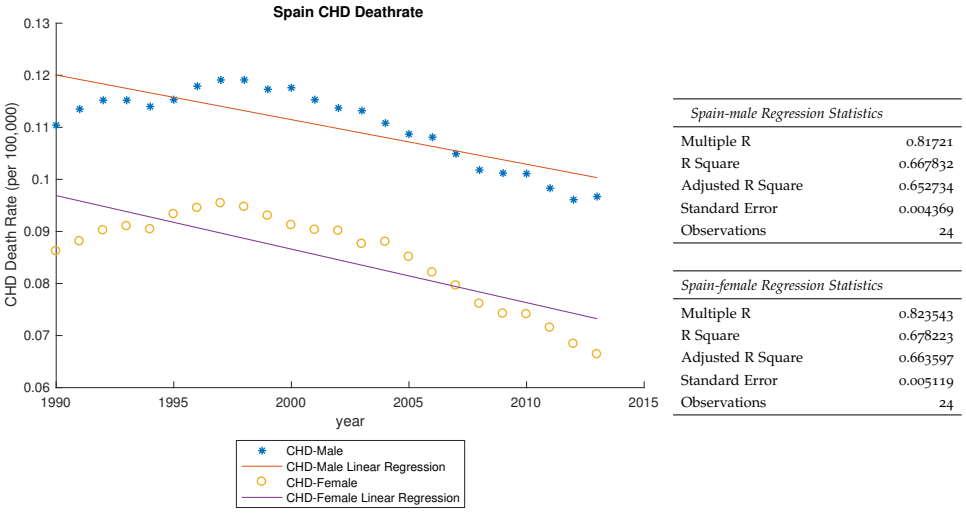
As can be seen from Table 4, a linear regression was performed, to predict CHD death rate based on time scale in Italy. The significant regression equation are:

$$y_{CHD_{male}} = -0.00079\ t + 1.72368 \tag{11a}$$

$$y_{CHD_{female}} = -0.00057\ t + 1.24842 \tag{11b}$$

at $(F_{male}(1, 22) = 112.505, \rho < 0.0001)$, $R^2_{male} = 0.836$ for males, and $(F_{female}(1, 22) = 23.800, \rho < 0.0001)$, with $R^2_{female} = 0.520$ for females.

Spain



Spain-male Regression Statistics	
Multiple R	0.81721
R Square	0.667832
Adjusted R Square	0.652734
Standard Error	0.004369
Observations	24

Spain-female Regression Statistics	
Multiple R	0.823543
R Square	0.678223
Adjusted R Square	0.663597
Standard Error	0.005119
Observations	24

Figure 11: Spain Linear plot and regression fit of CHD deathrate

Observed from Figure 11, both males and females' Pearson's Correlation Coefficient of CHD death rate in Spain are all larger than 0.8, which means there are outstanding correlation between CHD death rate and time in Spain, either males or females.

A linear regression was generated by Table 5, to predict CHD death rate based on time scale in Spain. The significant regression equation are:

$$y_{CHD_{male}} = -0.00086\ t + 1.82508 \tag{12a}$$

$$y_{CHD_{female}} = -0.00103\ t + 2.14260 \tag{12b}$$

Table 5: ANOVA test statistic of Spain CHD death rate

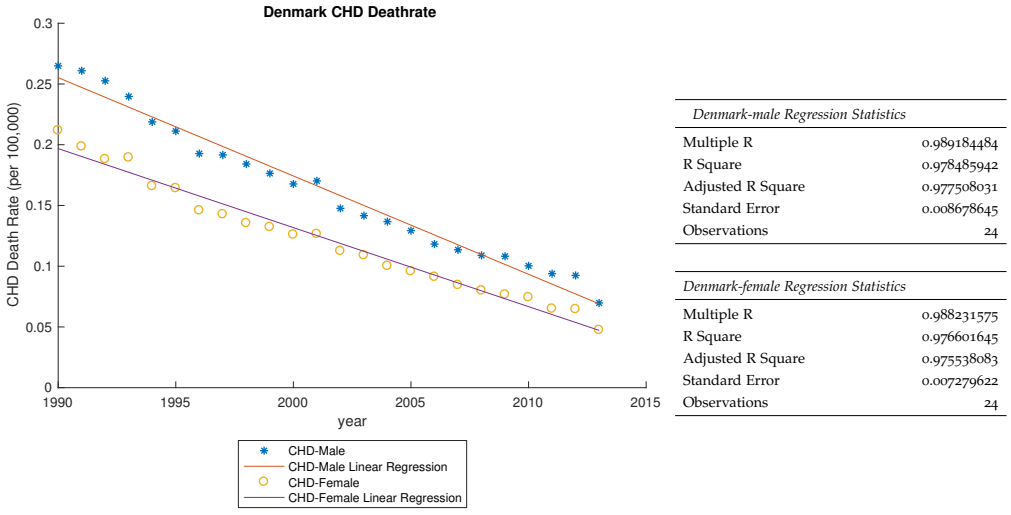
Spain-male	*df*	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.000844	0.000844	44.23163	1.1E-06
Residual	22	0.00042	1.91E-05		
Total	23	0.001264			

Spain-female	*df*	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.001215	0.001215	46.37035	7.67E-07
Residual	22	0.000577	2.62E-05		
Total	23	0.001792			

at $(F_{male}(1,22) = 44.232, \rho < 0.0001)$, $R^2_{male} = 0.668$ for males, and $(F_{female}(1,22) = 46.370, \rho < 0.0001)$, with $R^2_{female} = 0.678$ for females.

2.2.3 *Scandinavian European Countries (ScEU) Block*

Denmark



Denmark-male Regression Statistics	
Multiple R	0.989184484
R Square	0.978485942
Adjusted R Square	0.977508031
Standard Error	0.008678645
Observations	24

Denmark-female Regression Statistics	
Multiple R	0.988231575
R Square	0.976601645
Adjusted R Square	0.975538083
Standard Error	0.007279622
Observations	24

Figure 12: Denmark Linear plot and regression fit of CHD deathrate

Observed from Figure 12, both males and females' Pearson's Correlation Coefficient of CHD death rate in Denmark are all larger than 0.9, which means there are nearly perfect correlation between CHD death rate and time in Denmark, either males or females.

Table 6: ANOVA test statistic of Denmark CHD death rate

Denmark-male	*df*	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.075363	0.075363	1000.587	7.76366E-20
Residual	22	0.001657	7.53E-05		
Total	23	0.07702			

Denmark-female	*df*	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.04866	0.04866	918.237	1.95678E-19
Residual	22	0.001166	5.3E-05		
Total	23	0.049826			

A linear regression was generated from Table 6, to predict CHD death rate based on time scale in Denmark. The significant regression equation are:

$$y_{CHD_{male}} = -0.00810\ t + 16.36479 \tag{13a}$$

$$y_{CHD_{female}} = -0.00650\ t + 13.14145 \tag{13b}$$

at $(F_{male}(1, 22) = 1000.587, \rho < 0.0001)$, $R^2_{male} = 0.978$ for males, and $(F_{female}(1, 22) = 918.237, \rho < 0.0001)$, with $R^2_{female} = 0.977$ for females.

Finland

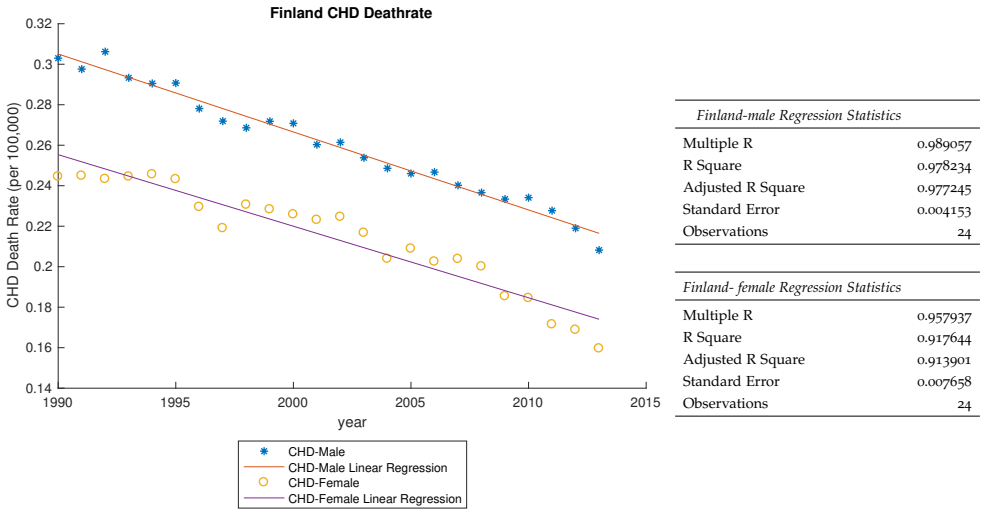


Figure 13: Finland Linear plot and regression fit of CHD deathrate

Observed from Figure 13, both males and females' Pearson's Correlation Coefficient of CHD death rate in Finland are all larger than 0.9, which means there are nearly perfect correlation between CHD death rate and time in Finland, either males or females.

Table 7: ANOVA test statistic of Finland CHD death rate

Finland-male	*df*	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.017053	0.017053	988.7721	8.82E-20
Residual	22	0.000379	1.72E-05		
Total	23	0.017433			

Finland-female	*df*	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.014376	0.014376	245.133	2.07E-13
Residual	22	0.00129	5.86E-05		
Total	23	0.015666			

A linear regression was generated from Table 7, to predict CHD death rate based on time scale in Finland. The significant regression equation are:

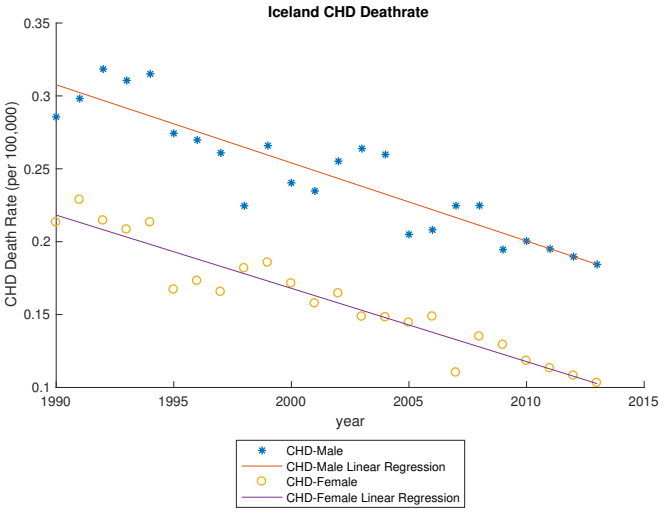
$$y_{CHD_{male}} = -0.00385\ t + 7.96821 \tag{14a}$$

$$y_{CHD_{female}} = -0.00354\ t + 7.29126 \tag{14b}$$

at $(F_{male}(1, 22) = 988.772, \rho < 0.0001)$, $R^2_{male} = 0.978$ for males, and $(F_{female}(1, 22) = 245.133, \rho < 0.0001)$, with $R^2_{female} = 0.918$ for females.

Iceland

Observed from Figure 14, both males and females' Pearson's Correlation Coefficient of CHD death rate in Iceland are all larger than 0.9, which means there are nearly perfect correlation between CHD death rate and time in Iceland, either males or females.



Iceland-male Regression Statistics	
Multiple R	0.908593
R Square	0.825542
Adjusted R Square	0.817612
Standard Error	0.017806
Observations	24

Iceland-female Regression Statistics	
Multiple R	0.955563
R Square	0.913101
Adjusted R Square	0.909151
Standard Error	0.011208
Observations	24

Figure 14: Iceland Linear plot and regression fit of CHD deathrate

Table 8: ANOVA test statistic of Iceland CHD death rate

Iceland-male	*df*	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.033005	0.033005	104.1048	8.36E-10
Residual	22	0.006975	0.000317		
Total	23	0.03998			

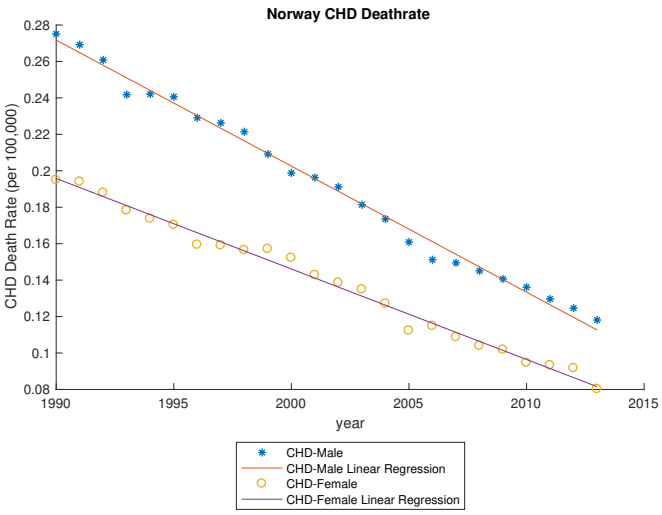
Iceland-female	*df*	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.02904	0.02904	231.1683	3.74E-13
Residual	22	0.002764	0.000126		
Total	23	0.031803			

A linear regression was generated from Table 8, to predict CHD death rate based on time scale in Iceland. The significant regression equation are:

$$y_{CHD_{male}} = -0.00536\ t + 10.96853 \tag{15a}$$

$$y_{CHD_{female}} = -0.00503\ t + 10.21824 \tag{15b}$$

at $(F_{male}(1,22) = 104.105, \rho < 0.0001)$, $R^2_{male} = 0.826$ for males, and $(F_{female}(1,22) = 231.168, \rho < 0.0001)$, with $R^2_{female} = 0.913$ for females.



Norway-male Regression Statistics	
Multiple R	0.996005
R Square	0.992025
Adjusted R Square	0.991663
Standard Error	0.004484
Observations	24

Norway-female Regression Statistics	
Multiple R	0.994751
R Square	0.98953
Adjusted R Square	0.989054
Standard Error	0.003695
Observations	24

Figure 15: Norway Linear plot and regression fit of CHD deathrate

Norway

Observed from Figure 15, both males and females' Pearson's Correlation Coefficient of CHD death rate in Norway are all larger than 0.9, which means there are nearly perfect correlation between CHD death rate and time in Norway, either males or females.

Table 9: ANOVA test statistic of Norway CHD death rate

Norway-male	*df*	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.05502	0.05502	2736.789	1.4E-24
Residual	22	0.000442	2.01E-05		
Total	23	0.055462			

Norway-female	*df*	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.028383	0.028383	2079.23	2.8E-23
Residual	22	0.0003	1.37E-05		
Total	23	0.028683			

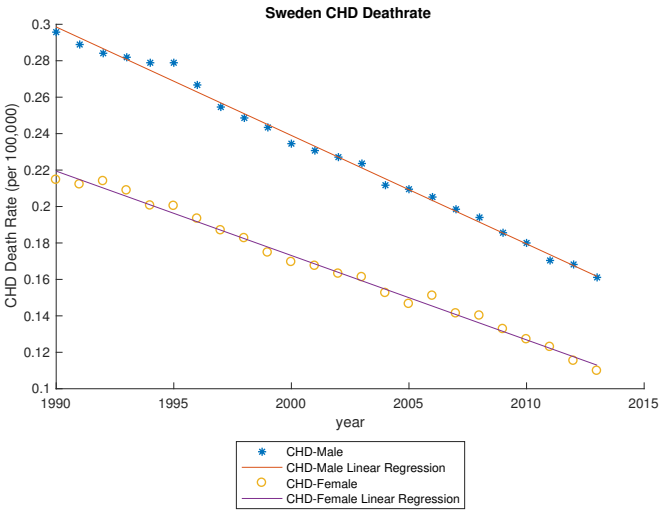
A linear regression was generated from Table 9, to predict CHD death rate based on time scale in Norway. The significant regression equation are:

$$y_{CHD_{male}} = -0.00692\ t + 14.03636 \tag{16a}$$

$$y_{CHD_{female}} = -0.00497\ t + 10.08200 \tag{16b}$$

at ($F_{male}(1,22) = 2736.789, \rho < 0.0001$), $R^2_{male} = 0.992$ for males, and ($F_{female}(1,22) = 2079.230, \rho < 0.0001$), with $R^2_{female} = 0.990$ for females.

Sweden



Sweden-male Regression Statistics	
Multiple R	0.99703
R Square	0.994069
Adjusted R Square	0.993799
Standard Error	0.003324
Observations	24

Sweden-female Regression Statistics	
Multiple R	0.996258
R Square	0.99253
Adjusted R Square	0.99219
Standard Error	0.002905
Observations	24

Figure 16: Sweden Linear plot and regression fit of CHD deathrate

Observed from Figure 16, both males and females' Pearson's Correlation Coefficient of CHD death rate in Sweden are all larger than 0.9, which means there are nearly perfect correlation between CHD death rate and time in Sweden, either males or females.

Table 10: ANOVA test statistic of Sweden CHD death rate

Sweden-male	*df*	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.040744	0.040744	3687.163	5.39E-26
Residual	22	0.000243	1.11E-05		
Total	23	0.040987			

Sweden-female	*df*	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.024673	0.024673	2922.965	6.82E-25
Residual	22	0.000186	8.44E-06		
Total	23	0.024859			

A linear regression was generated from Table 10, to predict CHD death rate based on time scale in Sweden. The significant regression equation are:

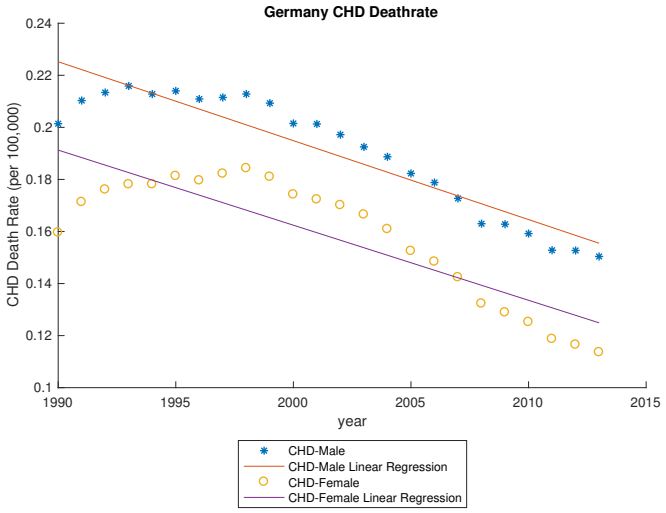
$$y_{CHD_{male}} = -0.00595\ t + 12.14356 \tag{17a}$$

$$y_{CHD_{female}} = -0.00463\ t + 9.43694 \tag{17b}$$

at $(F_{male}(1,22) = 3687.163, \rho < 0.0001)$, $R^2_{male} = 0.994$ for males, and $(F_{female}(1,22) = 2922.965, \rho < 0.0001)$, with $R^2_{female} = 0.993$ for females.

2.2.4 *Western European Countries (WeEU) Block*

Germany



Germany-male Regression Statistics	
Multiple R	0.930480705
R Square	0.865794343
Adjusted R Square	0.859694086
Standard Error	0.008630261
Observations	24

Germany-female Regression Statistics	
Multiple R	0.857787618
R Square	0.735799597
Adjusted R Square	0.723790488
Standard Error	0.012501682
Observations	24

Figure 17: Germany Linear plot and regression fit of CHD deathrate

Observed from Figure 17, males' Pearson's Correlation Coefficient of CHD death rate in Germany is 0.930, which performed a nearly perfect correlation between CHD death rate and time. For female, there is a slight lower Multiple R value, which is 0.858, it is not a perfect correlation, but still remain as an outstanding one.

Table 11: ANOVA test statistic of Germany CHD death rate

Germany-male	*df*	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.010571	0.010571	141.9275	4.57E-11
Residual	22	0.001639	7.45E-05		
Total	23	0.01221			

Germany-female	*df*	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.009576	0.009576	61.27012	8.46E-08
Residual	22	0.003438	0.000156		
Total	23	0.013014			

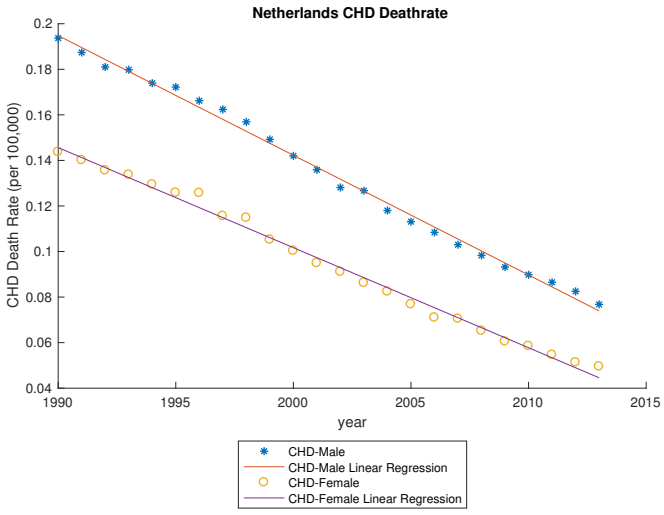
A linear regression was generated from Table 11, to predict CHD death rate based on time scale in Germany. The significant regression equation are:

$$y_{CHD_{male}} = -0.00303\ t + 6.25859 \tag{18a}$$

$$y_{CHD_{female}} = -0.00289\ t + 5.93371 \tag{18b}$$

at $(F_{male}(1, 22) = 141.928, \rho < 0.0001)$, $R^2_{male} = 0.866$ for males, and $(F_{female}(1, 22) = 61.270, \rho < 0.0001)$, with $R^2_{female} = 0.736$ for females.

Netherlands



Netherlands-male Regression Statistics	
Multiple R	0.997579907
R Square	0.99516567
Adjusted R Square	0.994945928
Standard Error	0.002647584
Observations	24

Netherlands-female Regression Statistics	
Multiple R	0.996412826
R Square	0.992838521
Adjusted R Square	0.992512999
Standard Error	0.002697626
Observations	24

Figure 18: Netherlands Linear plot and regression fit of CHD deathrate

Observed from Figure 18, both males and females' Pearson's Correlation Coefficient of CHD death rate in Netherlands are all larger than 0.9, which means there are nearly perfect correlation between CHD death rate and time in Netherlands, either males or females.

Table 12: ANOVA test statistic of Netherlands CHD death rate

Netherlands-male	*df*	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.031745	0.031745	4528.786	5.68E-27
Residual	22	0.000154	7.01E-06		
Total	23	0.0319			

Netherlands-female	*df*	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.022195	0.022195	3049.991	4.29E-25
Residual	22	0.00016	7.28E-06		
Total	23	0.022355			

A linear regression was generated from Table 12, to predict CHD death rate based on time scale in Netherlands. The significant regression equation are:

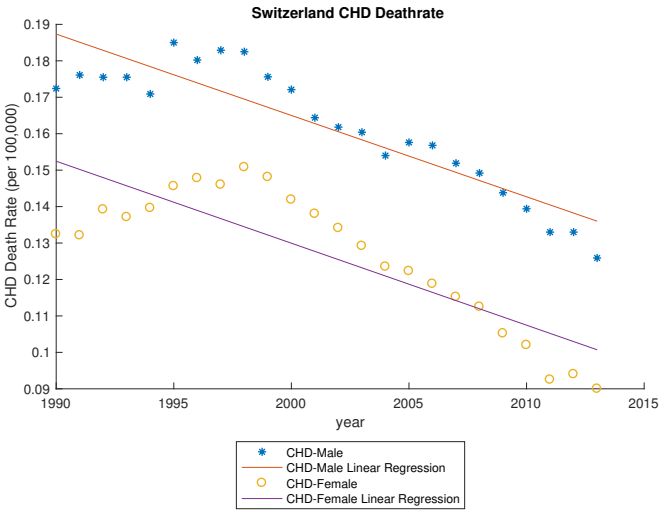
$$y_{CHD_{male}} = -0.00525\ t + 10.65029 \tag{19a}$$

$$y_{CHD_{female}} = -0.00439\ t + 8.88811 \tag{19b}$$

at $(F_{male}(1, 22) = 4528.786, \rho < 0.0001)$, $R^2_{male} = 0.995$ for males, and $(F_{female}(1, 22) = 3049.991, \rho < 0.0001)$, with $R^2_{female} = 0.993$ for females.

Switzerland

Observed from Figure 19, males' Pearson's Correlation Coefficient of CHD death rate in Germany is 0.907, which performed a nearly perfect correlation between CHD death rate and time. For female, there is a slight lower Multiple R value, which is 0.842, it is not a perfect correlation, but still remain as an outstanding one.



Switzerland-male Regression Statistics	
Multiple R	0.906883
R Square	0.822437
Adjusted R Square	0.814366
Standard Error	0.007496
Observations	24

Switzerland-femaleRegression Statistics	
Multiple R	0.842206
R Square	0.70931
Adjusted R Square	0.696097
Standard Error	0.010414
Observations	24

Figure 19: Switzerland Linear plot and regression fit of CHD deathrate

Table 13: ANOVA test statistic of Switzerland CHD death rate

Switzerland-male	*df*	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.005726	0.005726	101.8994	1.02E-09
Residual	22	0.001236	5.62E-05		
Total	23	0.006963			

Switzerland-female	*df*	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.005822	0.005822	53.68203	2.46E-07
Residual	22	0.002386	0.000108		
Total	23	0.008208			

A linear regression was generated from Table 13, to predict CHD death rate based on time scale in Switzerland. The significant regression equation are:

$$y_{CHD_{male}} = -0.00223\ t + 4.62792 \tag{20a}$$

$$y_{CHD_{female}} = -0.00225\ t + 4.62993 \tag{20b}$$

at $(F_{male}(1,22) = 101.899, \rho < 0.0001)$, $R^2_{male} = 0.822$ for males, and $(F_{female}(1,22) = 53.682, \rho < 0.0001)$, with $R^2_{female} = 0.709$ for females.

Table 14: Linear regression equations of CHD death rate for 13 European countries

CHD deathrate linear regression equations (y=mt+b)				
	Males		**Females**	
Country	**Slope (m)**	**y-intercept (b)**	**Slope (m)**	**y-intercept (b)**
United Kingdom	-0.00657	13.36542	-0.00604	12.24803
MeEU Country				
France	-0.00133	2.74338	-0.00164	3.33682
Greece	-0.00067	1.48386	-0.00071	1.51978
Italy	-0.00080	1.73646	-0.00058	1.26980
Spain	-0.00090	1.91383	-0.00111	2.31622
ScEU Country				
Denmark	-0.00810	16.38107	-0.00651	13.14806
Finland	-0.00393	8.12124	-0.00372	7.65059
Iceland	-0.00535	10.95994	-0.00506	10.27873
Norway	-0.00686	13.93186	-0.00496	10.05783
Sweden	-0.00596	12.15752	-0.00463	9.43538
WeEU Country				
Germany	-0.00309	6.37117	-0.00301	6.18520
Netherlands	-0.00525	10.65029	-0.00439	8.88811
Switzerland	-0.00223	4.62792	-0.00225	4.62993

Table 14 summarises the linear regression results for the CHD death rates against time for 13 European countries, that in turn have been subdivided into 3 blocks, based on geographical proximity that relates to food habits and life style.

Following are the key results from this section:

- All 13 European countries show decreasing CHD-vs-time trends both for males and females.
- Males have got higher CHD death rate than females in all 13 European countries.
- UK suffers from one of the highest CHD death affliction, just lower than Denmark and Norway, both for male and female groups.
- The ScEU country-block has the highest CHD death rate, whereas MeEU countries record the lowest CHD death rate.

- MeEU countries show a much more flat decreasing trend than the other European countries; the UK statistic showed between 6 to 10 times CHD deaths than the MeEU sector. Hence, the order of each block concludes as
  ScEU > UK > WeEU > MeEU

The results clearly depict that life-style and eating habit variation have the major influence on the CHD death rate; in the following works, we will quantify this impact.

2.3 LINEAR REGRESSION OF 6 LIFE-STYLE PARAMETERS

We use the same method as in section 2.2 to demonstrate the linear characteristics of our 6 life-style parameters. First we show the scatter plots for all 6 parameters respectively, after an analysis based on the statistical values, like Multiple R (Pearson's Correlation Coefficient), results of ANOVA; the linear regression functions are generated afterwards.

2.3.1 *United Kingdom*

Figure 20 shows the scatter plots of 6 parameters of UK respectively from 1990 to 2013, and linear fitted equations are generated for each of them. Each of them are discussed in details in the following.

Alcohol consumption: From the scatter plot Figure 20, a regression line with a positive gradient is found, as also observed in Table 15, although an average Pearson's Correlation Coefficient is 0.603, that is four times smaller than the 0.05 significant F value for $F(1, 22) = 12.549$. This means that this regression line is a better fit to the data than a model based only on the mean values of the predicted variable.

The regression equation of alcohol consumption in the UK is then obtained as:

$$y_{alcohol} = 0.07400\ t - 137.48345 \tag{21}$$

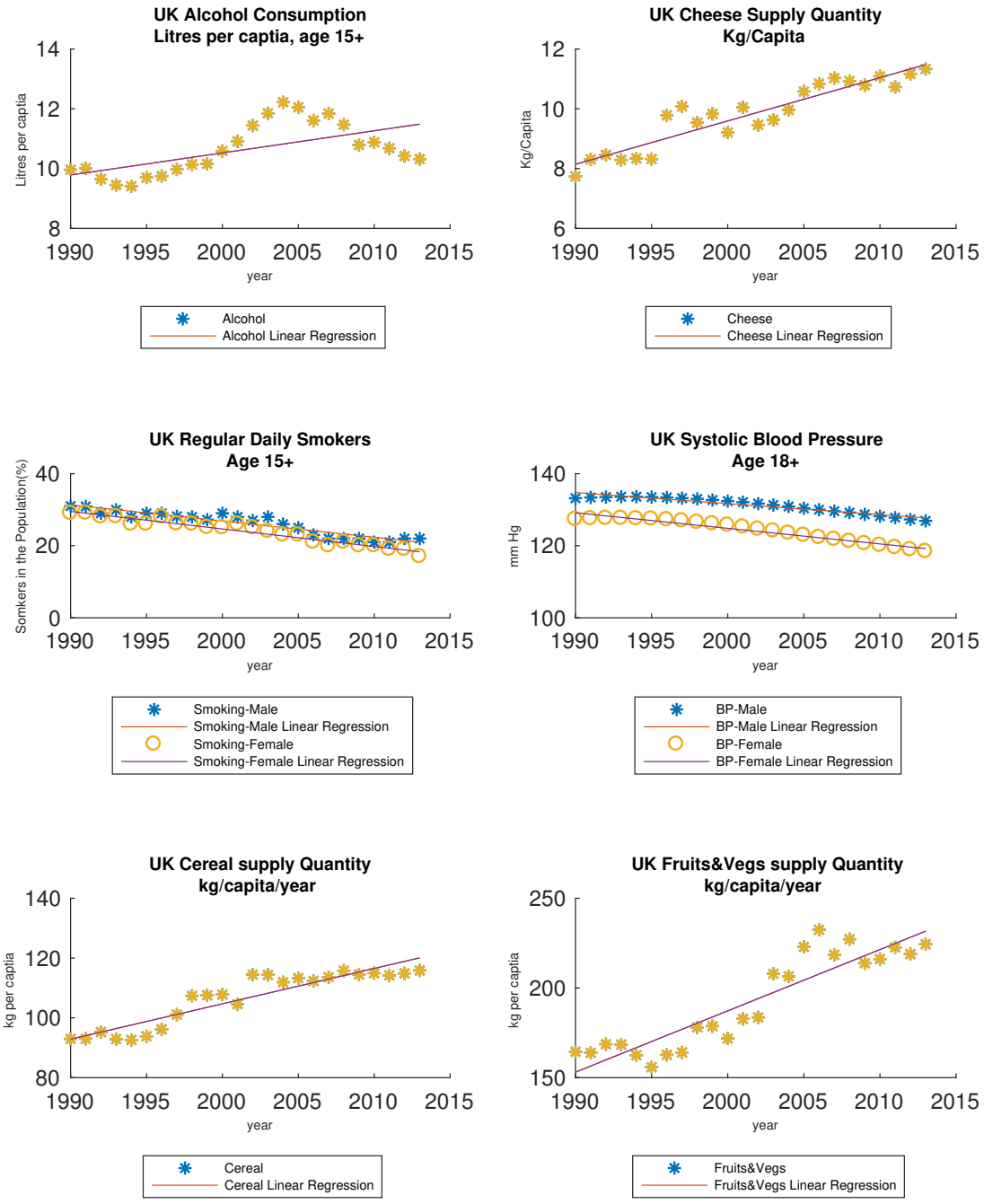


Figure 20: UK Linear plot and Linear regression of 6 parameters

Table 15: Regression and ANOVA test Statistics of UK alcohol consumption

Regression Statistics			*df*	*SS*	*MS*	*F*	*Significance F*
Multiple R	0.603	Regression	1	6.298	6.298	12.549	0.002
R Square	0.363	Residual	22	11.041	0.502		
Adjusted R Square	0.334	Total	23	17.339			
Standard Error	0.708						
Observations	24						

Cheese consumption: From the scatter plot Figure 20 accompanying Table 16, an increasing regression gradient is obtained, with a value larger than 0.9 Pearson's Correlation Coefficient. This is a nearly perfect correlation fitat $F(1,22) = 145.380$ with significant F value $\rho < 0.0001$, and $R^2 = 0.869$.

The regression equation of cheese consumption in the UK is concluded:

$$y_{cheese} = 0.14488\ t - 280.17046 \tag{22}$$

Table 16: Regression and ANOVA test Statistics of UK cheese consumption

Regression Statistics			*df*	*SS*	*MS*	*F*	*Significance F*
Multiple R	0.932	Regression	1	24.140	24.140	145.380	3.627E-11
R Square	0.869	Residual	22	3.653	0.166		
Adjusted R Square	0.863	Total	23	27.793			
Standard Error	0.407						
Observations	24						

Percentage of regular daily smokers: From the scatter plot Figure 20 accompany with Table 17, two decreasing regression lines have been found respectively for males and females in the UK, similar decreasing trending can be found for males and females, females remains about 3% less regular daily smokers than males. Both males and females have a larger than 0.9 Pearson's Correlation Coefficient, which shows nearly perfect correlation exist at $F_{male}(1,22) = 166.216$ with significance F value $\rho_{male} < 0.0001$, $R^2 = 0.883$ for males, and $F_{female}(1,22) = 351.704$ with significance F value $\rho_{female} < 0.0001$, $R^2 = 0.941$ for females

The regression equations of percentage of regular daily smokers in the UK are concluded:

$$y_{smoke_{male}} = -0.44965\ t - 926.14688 \tag{23a}$$

$$y_{smoke_{female}} = -0.48686\ t - 998.37490 \tag{23b}$$

Mean systolic blood pressure: From the scatter plot Figure 20 accompany with Table 18, two decreasing regression lines have been found respectively for males and females in the UK, similar decreasing trending can be found for males and females, females remains about 6% lower mean systolic blood pressure than males. Again, both males

Table 17: Regression and ANOVA test Statistics of UK regular daily smokers

Regression Statistics				*df*	*F*	*Significance F*
	Males	*Females*	Regression-male	1	166.216	9.905E-12
Multiple R	0.940	0.970	Residual-male	22		
R Square	0.883	0.941	Total-male	23		
Adjusted R Square	0.878	0.938	Regression-female	1	351.704	5.090E-15
Standard Error	1.183	0.880	Residual-female	22		
Observations	24	24	Total-female	23		

and females have a larger than 0.9 Pearson's Correlation Coefficient, which shows nearly perfect correlation exist at $F_{male}(1,22) = 256.233$ with significance F value $\rho_{male} < 0.0001$, $R^2 = 0.921$ for males, and $F_{female}(1,22) = 388.773$ with significance F value $\rho_{female} < 0.0001$, $R^2 = 0.946$ for females

The regression equations of mean systolic blood pressure in the UK are concluded:

$$y_{BP_{male}} = -0.30609\ t - 743.86638 \tag{24a}$$

$$y_{BP_{female}} = -0.43270\ t - 990.22368 \tag{24b}$$

Table 18: Regression and ANOVA test Statistics of UK mean systolic blood pressure

Regression Statistics				*df*	*F*	*Significance F*
	Males	*Females*	Regression-male	1	256.233	1.319E-13
Multiple R	0.960	0.973	Residual-male	22		
R Square	0.921	0.946	Total-male	23		
Adjusted R Square	0.917	0.944	Regression-female	1	388.773	1.794E-15
Standard Error	0.648	0.744	Residual-female	22		
Observations	24	24	Total-female	23		

Cereals supply quantities: From the scatter plot Figure 20 accompany with Table 19, an increasing regression line has been found, with a larger than 0.9 Pearson's Correlation Coefficient, it is a nearly perfect correlation exist at $F(1,22) = 122.392$ with significance F value $\rho < 0.0001$, and an $R^2 = 0.848$.

The regression equation of cereals supply quantities in the UK is concluded:

$$y_{cereal} = 1.18297\ t - 2261.23072 \tag{25}$$

Table 19: Regression and ANOVA test Statistics of UK cereals supply quantities

Regression Statistics			*df*	*SS*	*MS*	*F*	*Significance F*
Multiple R	0.921	Regression	1	1609	1609	122.392	1.863E-10
R Square	0.848	Residual	22	289	13		
Adjusted R Square	0.841	Total	23	1899			
Standard Error	3.626						
Observations	24						

Fruits and vegetables supply quantities: From the scatter plot Figure 20 accompany with Table 20, a steep increasing regression line has been found, with a larger than 0.9 Pearson's Correlation Coefficient, it is a nearly perfect correlation exist at $F(1, 22) = 108.069$ with significance F value $\rho < 0.0001$, and an $R^2 = 0.831$.

The regression equation of fruits and vegetables supply quantities in the UK is concluded:

$$y_{fruit} = 3.41821\ t - 6649.19466 \tag{26}$$

Table 20: Regression and ANOVA test Statistics of UK fruits and vegetables supply quantities

Regression Statistics			*df*	*SS*	*MS*	*F*	*Significance F*
Multiple R	0.912	Regression	1	13437	13437	108.069	5.932E-10
R Square	0.831	Residual	22	2735	124		
Adjusted R Square	0.823	Total	23	16172			
Standard Error	11.151						
Observations	24						

All other 12 European countries have been identically analysed; the results for each country is shown in Appendix C, includes the plots of each parameter, the regression and ANOVA test statistics table.

Table 21 and Table 22 in the following summarise the linear regression equations of 6 parameters, some key findings are observed from these results:

- Almost all negative indicators show decreasing trends, except cheese consumption. For a few countries (e.g. UK, ScEU), however, the alcohol consumption trends show an alarming increase over the years. From Table 14, we observed ScEU countries and UK suffering the highest CHD death rate which possibly indicates that alcohol may have a close influence on CHD death rates.

- Almost all two *positive* indicators show increasing trend, except Greece, Italy, Spain, Iceland and Switzerland which show slightly decreasing trends on cereals consumption. Greece, Italy, Spain, Germany and Switzerland show slightly decreasing trends on fruit and vegetable consumption. All these countries are included in the MeEU and WeEU blocks, which are the two blocks with a lower CHD death rate. This is not a result we expected, as we hypothesise these two *posotive* factors will benefit in lowering the CHD death rate. Further analysis needs to be done possibly using more detailed data mining methods to investigate these details.

Table 21: Linear regression equations of 4 negative indicators for 13 European countries

Linear regression equations of 4 negative indicators for 13 European countries (y=mx+b)												
Alcohol			**cheese**		**smoke**				**BP**			
					Males		**Females**		**Males**		**Females**	
Country	**m**	**b**	**m**	**b**	**m**	**b**	**m**	**b**	**m**	**b**	**m**	**b**
United Kingdom	0.07400	-137.48345	0.14488	-280.1705	-0.4496	926.1469	-0.4869	998.3749	-0.3061	743.8664	-0.4327	990.2237
MeEU country												
France	-0.16189	337.17674	0.13234	-241.6276	-0.2401	513.6603	0.2700	-518.2550	-0.1813	492.7561	-0.3387	799.0660
Greece	-0.08375	176.94961	0.08506	-144.5487	-0.8452	1737.6449	-0.2469	522.0402	-0.0793	285.8031	-0.2603	641.3205
Italy	-0.15679	322.77196	0.21411	-407.2223	-0.4069	846.1023	-0.0519	120.9967	-0.1807	492.4790	-0.2873	699.0318
Spain	-0.10751	226.31240	0.19370	-380.5985	-0.8970	1832.6201	-0.1872	397.2658	-0.0505	229.9197	-0.2632	647.1879
ScEU Country												
Denmark	-0.10510	221.61978	0.35225	-686.7411	-1.1580	2348.4926	-0.9194	1866.6133	-0.1961	523.9680	-0.2441	611.4234
Finland	0.05959	-110.06366	0.36756	-720.0727	-0.4859	999.2800	-0.2620	542.8759	-0.3117	755.0297	-0.3992	923.2336
Iceland	0.19193	-377.71381	0.97993	-1942.0924	-0.7493	1522.5738	-0.8076	1638.0246	-0.1112	349.1599	-0.1723	463.0380
Norway	0.09622	-186.86406	0.02014	-25.4685	-1.0117	2052.8496	-0.9141	1856.8750	-0.1676	469.0896	-0.3123	751.0735
Sweden	-0.01283	32.72225	0.16510	-312.8705	-0.6438	1305.8346	-0.5867	1193.5385	-0.2539	639.9486	-0.2969	717.1423
WeEU Country												
Germany	-0.14790	308.43477	0.23028	-441.8228	-0.3548	739.5121	-0.1507	320.5462	-0.3899	912.3860	-0.5447	1214.3829
Netherlands	-0.04145	92.60991	0.11706	-216.0024	-1.0804	2195.0918	-0.7666	1559.8812	-0.1260	383.3890	-0.1601	443.8886
Switzerland	-0.12672	264.62111	0.26380	-510.4632	-0.7284	2160.3628	-0.3990	1188.6120	-0.2250	578.9755	-0.3467	813.8525

Table 22: Linear regression equations of 2 positive indicators for 13 European countries

Linear regression equations of positive indicators (y=mt+b)				
	Cereals		Fruits and vegtables	
Country	**m**	**b**	**m**	**b**
United Kingdom	1.1830	-2261.2307	3.4182	-6649.1947
MeEU Country				
France	0.8903	-1664.8960	0.6991	-1192.3199
Greece	-0.9529	2048.1636	-3.5215	7461.9395
Italy	-0.0418	241.9576	-0.3973	1103.5726
Spain	-0.0073	115.2854	-4.1643	8592.5888
ScEU Country				
Denmark	2.0770	-4036.0693	4.6614	-9134.8632
Finland	0.9407	-1776.8261	2.3443	-4535.1422
Iceland	-0.3978	877.1454	4.7053	-9254.8926
Norway	0.3096	-497.4935	3.4085	-6636.0958
Sweden	0.5358	-973.7599	3.3423	-6506.5425
WeEU Country				
Germany	1.1431	-2185.4122	2.7021	-5255.7352
Netherlands	0.8329	-1589.8377	0.9426	-1668.2133
Switzerland	-0.2100	526.9303	-0.8796	1958.0432

2.4 SUMMARY

Chapter summary;

- For the first time, we have introduced linear least square fitting in our analysis. This will be later used to interpolate more 'synthetic' data in addition to real data which are statistically sparse in chapter 4.
- Statistical data modelling has been performed for CHD death rate and 6 life-style parameters using linear least square fitting. UK statistics have been compared against 12 European countries to compare life style trends.

In the next chapter, data mining methods will be used to rank the importance of the affecting factors and then to visualise them in lower dimensions.

3

DATA VISUALIZATION METHODS

To further analyse the nature of the multivariate relationship between all affecting variables, we have used well-established data mining and visualization tools. As generally perceived in this literature (Pal and Mitra, 2004), linear methods are on the whole not entirely reliable for visualising large datasets. So, we will take recourse to nonlinear tools as well and compare the outcomes against each other. It must be remembered that given the linear nature of relationships of all 6 parameters against time, as also between CHD and time, a linear patterning perhaps is not wholly unexpected in our case. Our following analysis will ratify this argument.

Data visualization is also known as dimensionality reduction or data projection approach, which carries a significant role of help researcher understands the significance of dataset by placing it in a pictorial or graphical format (Friedman, 1998). Such a visualization plot helps to identify the similarity of data patterns or any intrinsic structure present in the dataset. This chapter reviews some data visualization algorithms, in particular, Principal Component Analysis (PCA) will be discussed as the main visualization method, PCA, the most commonly used technique designed for visualising and thereby reducing the dimensional of the linear dataset is defined as an approach to project a high-dimensional dataset onto a low-dimensional space. Our model relies on 6 mutually correlated parameters, a simultaneous variation of which would be impossible to track and visualise at any level. This necessitates a dimensional reduction from 6 to a lower non-dimensional parametric phase-diagram, whereby these scaled non-dimensional parameters probabilistically summarise the impact of all variables pertaining to a change in statistics.

Usually, this low-dimension space is 2-D for the purpose of visualization on a scatter plot; NeuroScale (NSC) - A topographic feature extraction method enable the non-linear transformation. We also brify the other two projection algorithms: Gen-

erative Topographic Mapping (GTM), consumptionGaussian Process Latent Variable Model (GPLVM).

Two classes of data transformation can be chosen by analysing the different structure of the purposed datasets: linear transformation and non-linear transformation. PCA (Pearson, 1901) is a commonly used linear transformation mapping method. It has been widely used due to the easy and speedy apply on training. However, this method is only suitable for linear datasets. As the databases (listed in Appendix A) used in this study have been tested and analysed in chapter 2, which shows linear characterise, PCA is then to be a reasonable method to apply. The non-linear methods can be divided into two types based on their function: one type only provides the visualization, and another one enables to mapping the databases either from high dimensional space to the low dimensional space or vice verse, such as GTM (Bishop, Svensén and Williams, 1998), which based on a constrained mixture of Gaussians, provides data PDF both in high dimensional space defined by molecular descriptors and in 2D latent space. In addition, the approaches can also be categories as global and local techniques, each of which has advantages and disadvantages (Silva and Tenenbaum, 2003). Localised algorithms such as GPLVM (Lawrence, Seeger and Herbrich, 2003) are non-linear approaches relying on a version of probabilistic PCA that uses a smooth mapping from the latent space to the data space, making it difficult to accurately stabilise both local distances and dissimilarities. Another approach, NSC (Lowe and Tipping, 1996) also perform a smooth mapping from data space to latent space, but it global preserves the geometry at all scales. This means that all properties and structures are retained while local methods may not constrain points which are close in the data space to be close in the latent space identically.

3.1 PRINCIPAL COMPONENT ANALYSIS (PCA)

Principal Component Analysis (PCA) is a popular statistical technique that uses on linear data projection to map a high-dimensional dataset onto a low-dimensional subspace (Hotelling, 1933; Jolliffe, 1986; Pearson, 1901). Starting from our dataset which consists of four (only risk parameters) or six (all parameters) effective parameters, the idea here is to reduce this high-dimensional space to a more workable

two dimensional subspace for a better representation of the key properties encapsulated within the data that are otherwise difficult to observe in its high-dimensional manifestation.

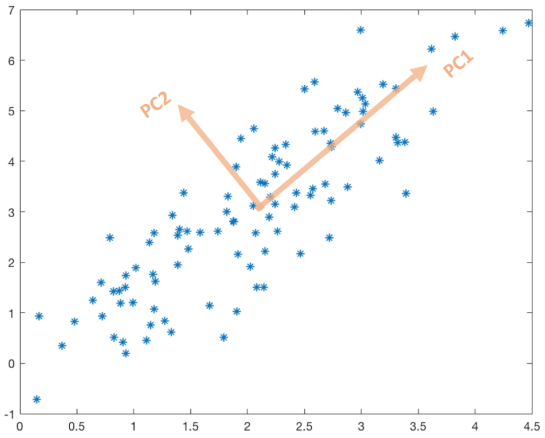


Figure 21: 100 pairs of (X_i, Y_i) randomly generated. A strong correlation is visible. PCA is tries to find the first principal component (PC1) which would explain most of the variance in the dataset. In this case it is clear that the most variance would stay present if the new random variable (PC1) would be on the direction shown with the line on the graph. This new random variable would explain most of the variation in the data set and could be used for further analysis instead the original variables.

To project a dataset of observed M-dimensional space vectors $\mathbf{v}_n$, where $n \in 1, 2, \ldots, N$, to corresponding vectors $\mathbf{z}_n$ in an K-dimensional space (normally $K = 2$ or $K = 3$). We write the vector $\mathbf{v}_n$ as a linear combination of M orthonormal vectors $\mathbf{u}_m$

$$\mathbf{v}_n = \sum_{m=1}^{M} z_m \mathbf{u}_m \tag{27}$$

According to the orthonormal property,

$$\mathbf{u}_m^T \mathbf{u}_{m'} = \delta_{mm'}, \tag{28}$$

where $\delta_{mm'}$ is a Kronecker delta[1] representation. We then have

$$z_m = \mathbf{u}_m^T \mathbf{v}. \tag{29}$$

1 **Kronecker delta** is a function of two variables, where: $\delta_{mm'} = \begin{cases} 0 & \text{if} \quad m \neq m' \\ 1 & \text{if} \quad m = m' \end{cases}$

Our purpose is to project the vectors to a K-dimensional space, so we rewrite Equation 27

$$\tilde{\mathbf{v}} = \sum_{m=1}^{K} z_m \mathbf{u}_m + \sum_{m=K+1}^{m} b_m \mathbf{u}_m, \tag{30}$$

where b_m are constants. Coefficients b_m and and vectors $\mathbf{u}_m$ are chosen to ensure the best approximation for $\mathbf{z}_n$.

$$\mathbf{v}_n - \tilde{\mathbf{v}}_n = \sum_{m=K+1}^{M} (z_{m,n} - b_m)\mathbf{u}_m, \tag{31}$$

where $t_{m,n}$ represents the mth feature of the nth data point. Minimizing the sum of squares error for the whole dataset

$$\begin{aligned} E &= \frac{1}{2}\sum_{n=1}^{N} \|\mathbf{v}_n - \tilde{\mathbf{v}}_n\|^2 \\ &= \frac{1}{2}\sum_{n=1}^{N}\sum_{m=K+1}^{M}\sum_{m'=K+1}^{M} (z_{m,n} - b_m)(z_{m',n} - b_{m'})\mathbf{u}_m^T\mathbf{u}_{m'} \\ &= \frac{1}{2}\sum_{n=1}^{N}\sum_{m=K+1}^{M} (z_{m,n} - b_m)^2. \end{aligned} \tag{32}$$

As $\mathbf{u}_m$ are orthonormal vectors, setting the derivative of E with respect to b_m to zero gives

$$b_m = \frac{1}{N}\sum_{n=1}^{N} z_{m,n} = \mathbf{u}_m^T\hat{\mathbf{v}}, \tag{33}$$

where $\hat{\mathbf{v}} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{v}_n$. Using equationsEquation 29 and (33), the error function Equation 32 takes the form

$$\begin{aligned} E &= \frac{1}{2}\sum_{m=K+1}^{M}\sum_{n=1}^{N}\left(\mathbf{u}_m^T(\mathbf{v}_n - \hat{\mathbf{v}})\right)^2 \\ &= \frac{1}{2}\sum_{m=K+1}^{M}\mathbf{u}_m^T\Sigma\mathbf{u}_m, \end{aligned} \tag{34}$$

where $\Sigma = \sum_{n=1}^{N}(\mathbf{v}_n - \hat{\mathbf{v}})(\mathbf{v}_n - \hat{\mathbf{v}})^T$ represents the covariance matrix of the data. Using Lagrange multiplier formulation, we observe that the stationary points of E re-

late to the eigenvectors of Σ showing that $\Sigma u_m = \lambda_m u_m$. The residual error equation then turns to

$$E = \frac{1}{2} \sum_{m=K+1}^{M} \lambda_m. \tag{35}$$

This means that the selection of the $M - K$ smallest eigenvalues obtains minimum error whereas the the data is mapped onto the space spanned by the first K eigenvectors corresponding to the largest eigenvalues. These eigenvectors are known as the K principal components.

3.1.1 *PCA Visualization Steps*

The following simple steps will explain how PCA can be used to visualize data in a two-dimensional or three-dimensional scatter plots.

- We have a dataset $\mathbf{V}$ with N data points in M dimensions.
- Compute the mean for each dimension i.e. $\mu_m = \frac{1}{N} \sum_{n=1}^{N} v_{nm}$
- Commpute covariance matrix $\mathbf{C} = \frac{1}{N} \sum_{n=1}^{N} (v_{nm} - \mu_m)(v_{nm'} - \mu_{m'})^T$, where m ranges over $1, 2, \ldots, M$ and for each value of m, m' ranges over $1, 2, 3, \ldots, M$. The covariance formula will yield a $M \times M$ matrix.
- Compute eigenvalues λ_k and eigenvectors ξ_m of covariance matrix $\mathbf{C}$.
- Order eigenvalues λ_k in descending order and the corresponding eigenvectors in such a way that $\lambda_1 \geqslant \lambda_2 \geqslant \lambda_3 \geqslant \cdots \geqslant \lambda_M$. For high-dimensional datasets, many of the eigenvalues λ can be neglected.
- For visualization, the dimension chosen would be two or three. Projecting data on first K eigenvectors would be obtained as $\mathbf{z}_k = (\mathbf{V} - \mu) * \xi_k$ where $\mathbf{V}$ represents data matrix, μ represents mean of each dimension and $m \in 1, \cdots, K$. Each data point will be subtracted from mean and then multiplied with first K eigenvectors. $\mathbf{Z}$ will serve as projection of data on first K-Principal Components (PC).

3.2 NEUROSCALE (NSC)

The NeuroScale approach (Lowe and Tipping, 1996), is a novel dimension-reducing, also topographic feature extraction process which employs a non-linear transformation. This model is related to Sammon's mapping (Sammon, 1969) and Multi-dimensional scaling (Kruskal, 1964), which are traditional statistical methods. The NeuroScale model uses a Radial Basis Function (RBF) network (Lowe and Tipping, 1997) for mapping from an original high-dimensional configuration space into the projected space. The architecture of this approach is shown in Figure 22.

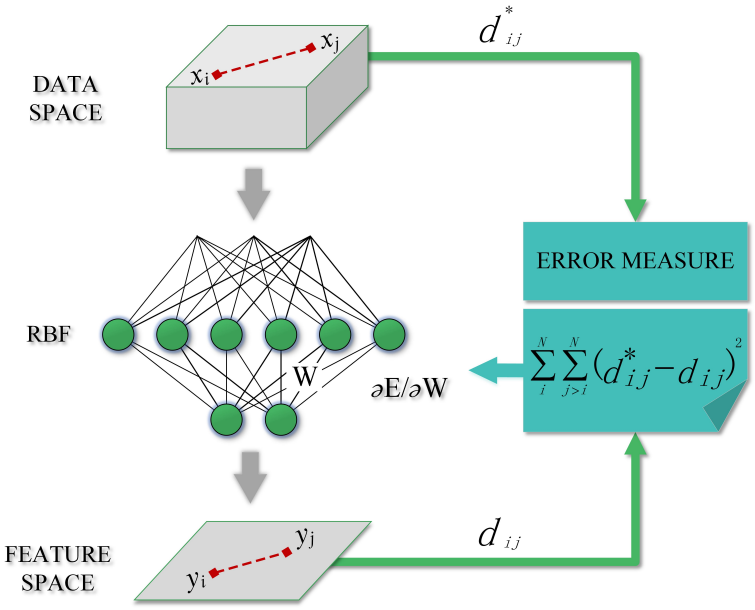


Figure 22: The NeuroScale Architecture

In this figure, x_i is the input data which intends to project into the transformed feature space y_i. This is done by a non-linear transformation using RBF networks. The advantage of this model is that when interpolations are allowed, a transformation still can be obtained. Also NeuroScale preserves the optimal topographic structure in the transformed space, the realisation of this constraint is attempt to select the inter-point distances in the projected space as closely as possible to the corresponding inter-point distances in the data space. Euclidean distance ($d^*_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|$ are the distances between data points in the original space and for projected space $d_{ij} = \|\mathbf{y}_i - \mathbf{y}_j\|$) is a common practice to approach this purpose. Since the weights in the output layer of the RBF model are used to indirectly determine the location of

the feature points, the method of initialising the weights has to be decided. The topographic transformation is resolved by optimising the network parameters, in order to minimise the error. In NeuroScale, the following *Sammon's stress metric* (Sammon, 1969) is used for achieving this

$$E = \sum_{i=1}^{N} \sum_{j>i}^{N} (d_{ij}^{*} - d_{ij})^2. \tag{36}$$

where $d_{ij}^{*} = \|\mathbf{x}_i - \mathbf{x}_j\| = \sqrt{(x_i - x_j)^T (x_i - x_j)}$ are the distance of inter-point (Euclidean distance) in the original data space, and $d_{ij} = \|\mathbf{z}_i - \mathbf{z}_j\| = \sqrt{(y_i - y_j)^T (y_i - y_j)}$ are the corresponding distance in the transformed space. The points y are predicted by RBF network in the following form

$$\mathbf{y} = \mathbf{f}(\mathbf{x}; \mathbf{W}) = \Phi(\mathbf{x})\mathbf{W} \tag{37}$$

where $\mathbf{f}$ is a $1 \times M$ projected space vector, $\Phi(\mathbf{x})$ is the nonlinear transformation $1 \times L$ vector effected by RBF with $L \times M$ weight matrix $\mathbf{W}$. The distance in the feature space thus by generated by

$$d_{ij}^2 = (\left\|\mathbf{f}(\mathbf{x_i}) - \mathbf{f}(\mathbf{x_j})\right\|)^2 \tag{38}$$

$$= \sum_{l=1}^{n} \left(\sum_{k} w_{lk} \left[\phi_k(\|\mathbf{x_i} - \mu_k\|) - \phi_k(\left\|\mathbf{x_j} - \mu_k\right\|) \right] \right)^2 \tag{39}$$

where $\phi_k()$ are the basis function from RBF network, μ_k are the centres of those functions [2], w_{lk} are the output layer weights from the basis function.

3.3 GENERATIVE TOPOGRAPHIC MAPPING (GTM)

The Generative Topographic Mapping (GTM) is a latent variable model was first introduced by Bishop, Svensén and Williams (1998). GTM is like a non-linear version of PCA, which allows the high dimensional data to be projected by adding Gaussian noise to the original data in low dimensional latent space, using RBF to generate a nonlinear transformation between the latent space and the original data space. This

2 In this thesis, μ_k of those functions are randomly selected from the datasets in the original space

algorithm is based on a constrained mixture of Gaussians, which optimise the parameters of the model using an Expectation-Maximization (EM) algorithm.

Since GTM models are constrained by probabilistic transformations, the data here becomes even more noisy, also, with visualisation results depending strongly on the choice of parameters since the number of the RBF basis functions, the distribution of the latent space sample points which are chosen manually. For this reason, this method is not ideal for our study, the comparison of results of each method will be discussed in detail in subsection 3.5.4

3.4 GAUSSIAN PROCESS LATENT VARIABLE MODEL (GPLVM)

The Gaussian Process Latent Variable Model (GPLVM) is a probabilistic dimensionality reduction methods that use Gaussian Processes (GPs) to lower the data dimension. GPLVM is a non-linear extension of probabilistic PCA, the model is defined probabilistically and the latent variables are marginalised out and parameters are obtained by maximising the likelihood. However, GPLVM is mapping from the latent space to the data space (like GTM) whereas in PCA it acts in the opposite direction.

GPLVM uses a smooth projection from a latent space to a data space, this kind of projection does not constrain points , which means the points which are close in the latent space to the points which are also close to the data space, hence, for the reduction of dimension, GPLVM may not be able to keep the accuracy. Details of GPLVM algorithm can be found in (Lawrence, 2008).

3.5 EVALUATION OF VISUALIZATION QUALITY

Four different visualization methods have been introduced in section 3.1 to section 3.4. We have used all four methods to analyse UK datasets to evaluate the best methods for this specific case. The other European countries have been analysed using the best chosen method. Three visualization quality evaluation measures: trustworthiness, continuity, Mean Relative Rank Errors (MRRE) are first explained in the following subsection 3.5.1 to subsection 3.5.3.

3.5.1 *Trustworthiness*

The main purpose of visualising data is to reduce the dimensionality of the dataset to two or three easy visualised space. It is impossible to keep every detail without losing information during the process. One kind of error may happen, the relatively distant data points may be projected as they are in the neighbourhood. Trustworthiness is the measure of the fraction of data points distant in the original data space that become neighbours in the mapping space (Venna and Kaski, 2005).

Set n to be the data numbers, $R(i,j)$ be the rank of the data points j from the corresponding data points i with respect to the distance measure in the original data space, $U_k(i)$ denoted the data points in the k-nearest neighbourhood of the i data points in the latent visualization space but not in the original data space. Trustworthiness with k-neighbours can be measured as

$$T(k) = 1 - \frac{2}{nk(2n-3k-1)} \sum_{i=1}^{n} \sum_{j \in U_k(i)} (R(i,j) - k), \tag{40}$$

where $nk(2n-3k-1)$ is the normalising factor, ensuring the value of trustworthiness between 0 and 1; the higher the value, better is the visualization result.

3.5.2 *Continuity*

As mentioned in subsection 3.5.1, an error that may happen in the process of visualization is that the relatively distant data points may be projected as they are in the neighbourhood. Another error may trickle in the opposite way; data points that are originally in the neighbourhood can be pushed away at a distance in the visualization process. This can cause not all neighbourhood being visualised. Continuity is measured as the fraction of neighbouring data points in the original data space that becomes distant in the mapping space (Venna and Kaski, 2005).

If n represents the data size, $R^*(i,j)$ the rank of the data points, j a running index scanning the corresponding data points, i an index representing the distance measured in the latent visualization space, and $V_k(i)$ denoted the data points in the

k-nearest neighbourhood of the i data points in the original data space but not in the latent visualization space. Trustworthiness with k-neighbours can be measured as

$$C(k) = 1 - \frac{2}{nk(2n-3k-1)} \sum_{i=1}^{n} \sum_{j \in V_k(i)} (R^*(i,j) - k) \tag{41}$$

again, where $nk(2n-3k-1)$ is the normalising factor, ensuring the value of continuity keeps between 0 and 1, the higher the better visualization results.

3.5.3 *Mean Relative Rank Errors (MRRE)*

Mean relative rank errors with respect to data space (MRREd) and latent visualization space (MRREl)is another well-known quality measures work on the same principle and same notation as *trustworthiness* and *continuity* (Lee and Verleysen, 2008).

The mean relative rank errors with respect to data space (MRREd) can be calculated by

$$MRREd(k) = \frac{1}{n \sum_{k'=1}^{k} \frac{|n-2k'|}{k'}} \sum_{i=1}^{n} \sum_{j \in n_k(i)} \frac{|(R^*(i,j) - R(i,j))|}{R(i,j)} \tag{42}$$

and the mean relative rank errors with respect to latent visualization space (MRREl) can be calculated by

$$MRREl(k) = \frac{1}{n \sum_{k'=1}^{k} \frac{|n-2k'|}{k'}} \sum_{i=1}^{n} \sum_{j \in n_k^*(i)} \frac{|(R(i,j) - R^*(i,j))|}{R^*(i,j)} \tag{43}$$

where $n \sum_{k'=1}^{k} \frac{|n-2k'|}{k'}$ is the normalising factor, ensuring the value of continuity keeps between 0 and 1, the lower the better visualization results.

3.5.4 *Evaluation of visualization Quality for UK data*

By visualising the UK database using all four visualization methods (PCA, NSC, GTM and GPLVM) introduced in section 3.1 to section 3.4, a quality matrix which

showed in Table 23 can be found by applying those visualization quality evaluation measures introduced from subsection 3.5.1 to subsection 3.5.3.

Table 23: Evaluation of visualization Quality for UK

PCA					**GTM**				
	T[3]	**C**[4]	**MRREd**	**MRREl**		**T**	**C**	**MRREd**	**MRREl**
Male	0.9974	0.9969	0.0547	0.0536	Male	0.9344	0.8833	0.2064	0.1888
Female	0.9911	0.9953	0.0629	0.0540	Female	0.9083	0.8823	0.1999	0.2028
NSC					**GPLVM**				
	T	**C**	**MRREd**	**MRREl**		**T**	**C**	**MRREd**	**MRREl**
Male	0.9969	0.9964	0.0436	0.0433	Male	0.9052	0.9411	0.2147	0.2130
Female	0.9922	0.9958	0.0410	0.0390	Female	0.9094	0.9401	0.2028	0.1777

It shows the following: training for PCA visualization model based on UK datasets, the trustworthiness value is 0.9974 for males and 0.9911 for females; NSC visualization model get a similar trustworthiness value as PCA, 0.9969 for males and 0.9922 for females; GTM and GPLVM trained with comparative lower trustworthiness value, which is all around 0.9 only. A similar situation happened to the continuity value, which PCA's continuity values are 0.9969 for males and 0.9953 for females, NSC has got similar continuity values as PCA again. GTM has got the lowest continuity value, both for males and females are smaller than 0.9 whereas smaller than 0.95 on GPLVM training; MRREd and MRREl values are observed around four times higher on GTM and GPLVM training than PCA and NSC, for both males and females. The nature of both trustworthiness and continuity values are the higher the better visualization results but reversed on MRREd and MRREl values, based on these quality evaluation measures, can be concluded PCA and NSC are the better visualization methods for our study.

Furthermore, based on the comparison between these three quality evaluations between PCA and NSC, the results are not much different, combined with the linear characteristic of our databases which evaluated on **??**, PCA defining our preferred choice for our visualization methods.

The following section will emphasise the results obtained by PCA visualization models.

3 **T**: Stands for Trustworthiness.
4 **C**: Stands for Continuity.

3.6 SUMMARY

Findings from this chapter;

- 4 different data visualization methods: PCA, NSC, GTM, GPLVM are introduced in detail.
- Comparing results from the different visualization methods, our preferred option has been chosen. This has been primarily guided by the linear time profiles of all affecting parameters/variables concerned.

In the next chapter, comparisons between the PCA and NSC visualization results have been shown.

4

VISUALIZATION ANALYSIS

In this chapter, PCA and NSC visualization have been applied on real datasets (collected from WHO) and also the generated synthetic datasets for each country concerned. In all related discussions, 'synthetic' data will allude to the data that are artificially generated from the extrapolated (linear) formulae defining the parameters(s) versus time functional relationships. We then rank all affecting negative indicators in an ascending order of importance. We then artificially change the contributions of the positive indicators (cereal and fruits&vegs consumptions) to enumerate their impact on the negative indicators and eventually on CHD/CVD. The target here is to establish regression forms connecting the positive with the negative indicators, essentially to suggest how much of a change in life style could feasibly affect mortality related to atherosclerosis inflicted CHD/CVD.

4.1 GENERATION OF DATASETS

Three different datasets developed from original databases are used in this chapter:

4.1.1 *Raw real datasets*

The first set of data used for visualization are the raw real datasets, as listed in Appendix A, all obtained from open sourced repositories. From year 1990 to 2013, 24 datasets have been used involving 6 risk parameters.

4.1.2 *Pure synthetic datasets*

24 datasets effectively amount to 24 datapoints that are grossly insufficient for statistical analysis. To get around this issue, as shown in chapter 2, we extricated data

from the linear fitted regression models as discussed previously in section 2.1. Pure synthetic datasets are then generated by those linear fitted formulae.

For example, from the raw real dataset of CHD death rate of UK in Table A.1.1 and Table A.2.1, Based on linear least square analysis, we obtained the following linear fitted models for UK:

$$y_{CHD_{male}} = -0.00657\ t + 13.36542 \tag{44}$$

$$y_{CHD_{female}} = -0.00604\ t + 12.24803 \tag{45}$$

$$y_{alcohol} = 0.07400\ t - 137.48345 \tag{46}$$

$$y_{cheese} = 0.14488\ t - 280.17046 \tag{47}$$

$$y_{smoke_{male}} = -0.44965\ t - 926.14688 \tag{48}$$

$$y_{smoke_{female}} = -0.48686\ t - 998.37490 \tag{49}$$

$$y_{BP_{male}} = -0.30609\ t - 743.86638 \tag{50}$$

$$y_{BP_{female}} = -0.43270\ t - 990.22368 \tag{51}$$

$$y_{cereal} = 1.18297\ t - 2261.23072 \tag{52}$$

$$y_{fruits\&vegs} = 3.41821\ t - 6649.19466 \tag{53}$$

Here t represents the year number,; in terms of the real datasets, t corresponds to the values $t = 1990, 1991,$
$\ldots, 2013$. To get the expected CHD death rate for each year, tthe regression formulae are then used to add 100 artificial (synthetic) points corresponding to these years. In real numbers, these are for $t = 1990, 1990.01, 1990.02, \ldots, 1990.99, 1991, 1991.01, \ldots,$ $2019.99, 2020$. For even better statistical analysis, we also extended the year span to 2020. Therefore, for each of the UK datasets, we generate **3001 artificial data points** between the years 1990 to 2020 which serve as the *training dataset*. So do the same data generation on the other 12 European countries based on the linear fitted models summarised in Table 14, Table 21 and Table 22.

4.1.3 *Real-synthetic datasets*

This is the dataset we are using for the following prediction process, which is a combination of the pure synthetic datasets generated from linear fitted formulae with the raw real data discussed in subsection 4.1.2. This incorporates substitution of the regression-fit data with the real raw data, whenever the latter are available, as in subsection 4.1.1. For example, we generate a set of pure synthetic dataset for years 1990, 1990.01, 1990.02, ..., 1990.99, 1991, and the raw real data we have is for years 1990 and 1991, in this case, we exchange the synthetic data of 1990 and 1991 with the raw real data, and keeping the rest unchanged. We call this set of data real-synthetic datasets.

The next step is to enumerate whether all or some of the 6 parameters are related to the CHD death rates (for all 13 European countries considered, including the UK) as also between themselves. To find this relationship, we have used state-of-art data visualization approaches. As two parameters (*i. e.* cereals and fruits&vegs consumption) are to be considered as positive factors, therefore we train data in two stages: first, we trained the 4-dimensional datasets which include the 4 prospective negative indicators only; second, we trained augmented 6-dimensional datasets which include all the 6 parameters considered, including the positive and the negative ones. The CHD death rates across the years, for each country, are binned to mark the data points with different markers on the visualization plots for understanding any clustering structures that may appear for each of the 13 European countries.

4.2 FEATURE WEIGHTING ESTIMATION USING PCA

Feature weighting is a way to select the importance of features after simplifying and reducing the dimensional of the models. PCA is a widely used multivariate data analysis approach, and as described in subsection 3.1.1, PCA has traditionally used a linear dimension reduction approach, to extract the feature set in lower-dimensional space that can describe most of the variation within the original high-dimensional

datasets. We then use PCA weighting for each of the m^{th} feature in the first K principal components (Kim and Rattakorn, 2011) as

$$\rho_m = \sum_{k=1}^{K} |\xi_{km}| w_k \tag{54}$$

Where K is the number of principal components of interest and w_k represents the weight of the k^{th} principal component. Typically we can determine the weight w_k of the m^{th} feature as the proportion of the total variance explained by the k^{th} principal component and a ρ_m is called the weighted PC loading for the m^{th} feature.

4.3 UK VISUALIZATION

In this section, we apply the two data visualization algorithms PCA and NSC on different feature sets of the databases either from 4-dimensional datasets onto the 2-dimensional space, or from 6-dimensional datasets onto the 2-dimensional space related to the UK, respectively for males and females. Including the uses of each of the three datasets which mentioned in the beginning of this chapter. Markers on the visualization are assigned using the bins (shown in legend) of the CHD death rate.

4.3.1 *UK visualization based on real datasets*

As described in subsection 3.1.1, before we start training our datasets for PCA visualizations, we first examine the eigenvalues to determine the principal components to be considered:

Figure 23 shows the scree plot based on real datasets for four negative indicators (i.e. alcohol, cheese, smoking, SBP) for males and females.

The table shows the order of importance of each of the four components. Only the first component has eigenvalue over 1, which is 3.3204 for males and 3.3603 for females; this is about 79.55% of the variance compared to the first eigenvalue for males, and 80.51% of the variance is explained by this first eigenvalue for females. The second principal value shows up as 0.6558 for males and 0.6346 for females, with respective variance levels of 15.71% and 15.20%. Adding the successive percentages

	Eigenvalues		Percentage of Variance		Cumulative Percentage	
Component	Males	Females	Males	Females	Males	Females
1	3.3204	3.3603	0.7955	0.8051	**0.7955**	**0.8051**
2	0.6558	0.6346	0.1571	0.1520	**0.9526**	**0.9571**
3	0.1412	0.1464	0.0338	0.0351	0.9865	0.9922
4	0.0564	0.0327	0.0135	0.0078	1	1
total	4.1739	4.1739	1	1	-	-

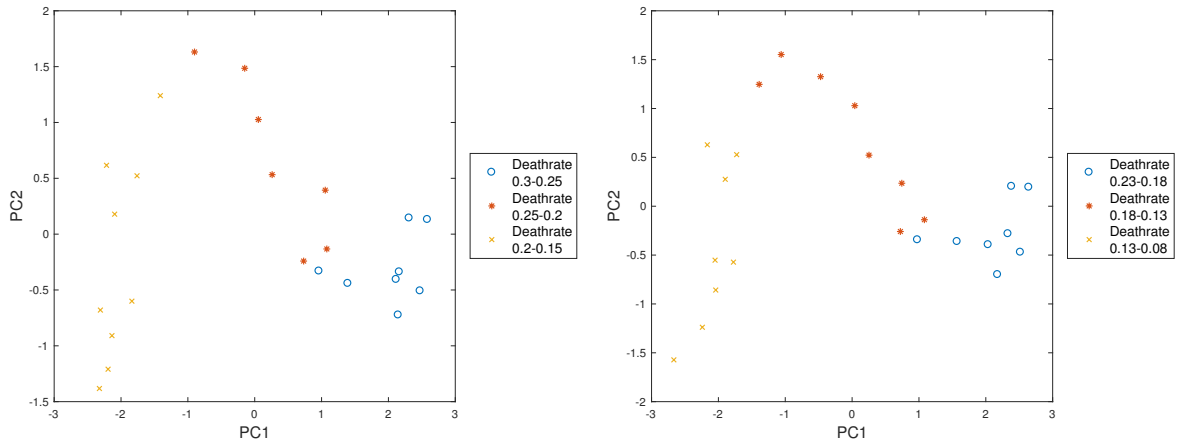
Figure 23: PCA scree plot based on real datasets of 4 negative indicators for males and females together, along with the variance explanation table for total component

of variation for PCA1 and PCA2 allows us to get an overall estimate of uncertainty combining estimates from the first two eigenvalues. Therefore, 95.26% and 95.71% of the variation respectively for males and females are explained by the first two eigenvalues together. This is an acceptably large tolerance level.

We can also determine the number of principal components by looking at the scree plots shown above. The plots show that males and females have close enough eigenvalues, indicated also by overlapping plots. With the eigenvalues ordered from largest to the smallest, the number of components is determined at the point. The first two components are large beyond which the remaining eigenvalues are all relatively small and of comparable size.

Therefore, the two largest principal components are used for the plotting of Figure 24 visualization.

In Figure 24, each marker in the plot represents one year, both males and females datasets are projected as a similar pattern. For instance, in the plot (a), the blue circle to the rightmost indicates that this particular dot has a very high value for the first principal component. Referring to Table 24, we would expect high scores for cheese, smoking and BP in this particular year. Whereas the top red star, which has a high value for the second component, refers to Table 24; we expect a high score for alcohol in this year.



(a) Males (b) Females

Figure 24: This is PCA visualization of UK real datasets based on 4 negative indicators (i.e. alcohol, cheese, smoking, SBP) for (a) males and (b) females respectively. Markers on the visualizations are assigned using the bins (shown in legends) of the death rates.

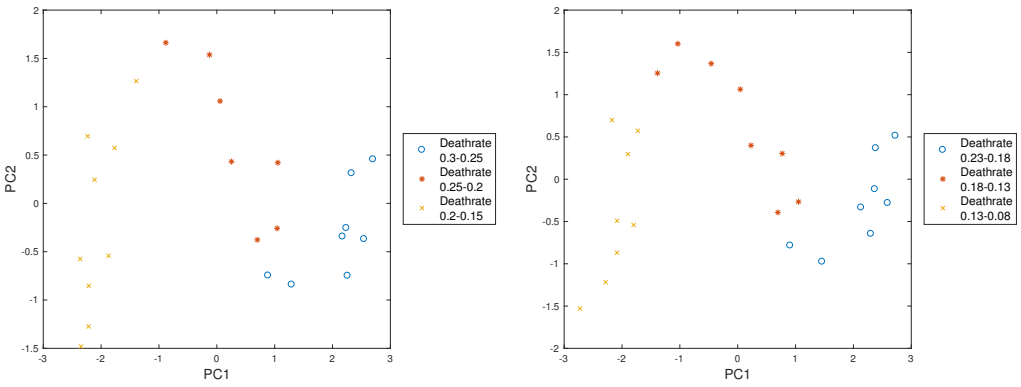
Table 24: Component Matrix of UK real data for 4 parameters

	male		female	
	PC1	PC2	PC1	PC2
Alcohol	0.3857	**0.9136**	0.3901	**0.9143**
Cheese	**0.5324**	0.1066	**0.5261**	0.1259
Smoking	**0.5360**	0.2806	**0.5335**	0.2883
SBP	**0.5296**	0.2742	**0.5352**	0.2552

Three Clusters are defined based on the range of CHD death rates, as shown in the legend. The plot shows a very clear separation for each of the cluster without any overlapping. The blue circled cluster includes the years with the death rate between $0.25 \sim 0.30$. In this cluster, high PC1 with low PC2 can be observed, indicates that in this cluster, expecting of high values of cheese, smoking and SBP. The red star cluster includes the years with the death rate between $0.20 \sim 0.25$, and sites in the high level of PC2, which means, in this cluster, alcohol score is expected to have a high value. The yellow crossed cluster includes the years with death rates between $0.15 \sim 0.20$ in the low PC1 range, indicating a low value for the alcohol score in this cluster as well. Similar trends were observed for females as well.

We also train the same dataset by NSC visualization, visualized plots shown in Figure 25 for males and females respectively.

In the projection, for both of the plots for males and females, three Clusters are defined based on the range of CHD death rates listed in the legend. These are clearly



(a) Males (b) Females

Figure 25: This is NSC visualization of UK real datasets based on negative indicators (i.e. alcohol, cheese, smoking, SBP) for (a) males and (b) females respectively. Markers on the visualizations are assigned using the bins (shown in legends) of the death rates.

separated with no overlaps. The blue circled cluster with high PC1 and low PC2 includes the years with the death rates between 0.25 ~ 0.30. The red starred cluster includes the years with the death rates between 0.20 ~ 0.25, and sites in the high level of PC2. The yellow crossed cluster includes the years with the death rates between 0.15 ~ 0.20 which occupied in the low range of PC1. A comparison of the PCA visualization results is shown in Figure 24, males and female statistics are seen to be largely commensurate with each other.

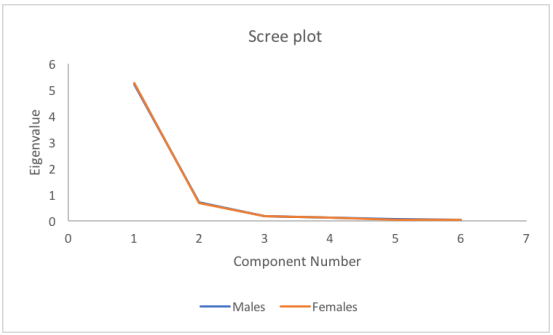
According to Table 24, the importance of the four parameters in the UK can be ranked by PC1 as the following order:

Males: Smoking > Cheese > SBP > Alcohol

Females: SBP > Smoking > Cheese > Alcohol

We now investigate possible changes wrought about by the two positive indicators. The following figure and table show the scree plot based on the real datasets of all six parameters (i.e. alcohol, cheese, smoking, SBP, cereals, fruits and vegs) for males and females.

According to Figure 26, we consider only the first two PC components, which are 5.1990 for males and 5.2519 for females as PC1. Approximately 83.04% of the variance is explained by this first eigenvalue for males, and 83.89% of the variance is explained by this first eigenvalue for females. PC2 values are shown as 0.6917 for males and 0.6610 for females, with percentage of variance 11.05% and 10.56%



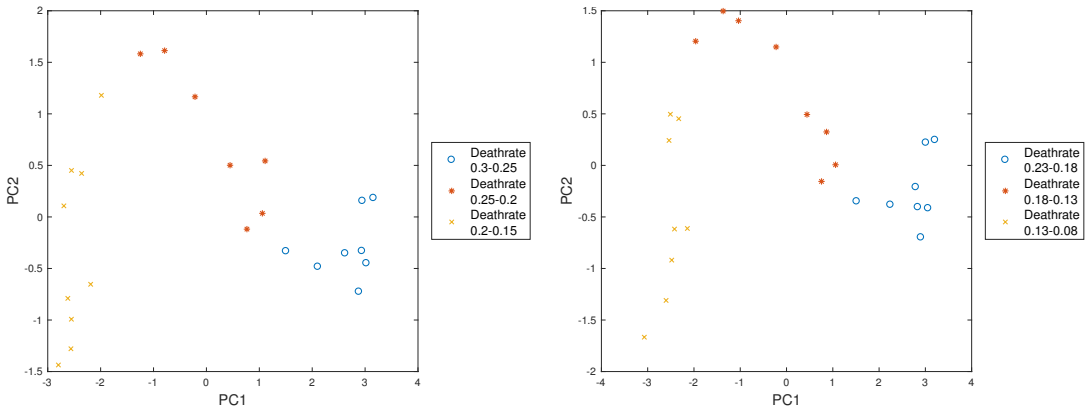
	Eigenvalues		Percentage of Variance		Cumulative Percentage	
Component	Males	Females	Males	Females	Males	Females
1	5.1990	5.2519	0.8304	0.8389	**0.8304**	**0.8389**
2	0.6917	0.6610	0.1105	0.1056	**0.9409**	**0.9444**
3	0.1746	0.1694	0.0279	0.0271		
4	0.1068	0.1055	0.0171	0.0169		
5	0.0522	0.0449	0.0083	0.0072		
6	0.0365	0.0281	0.0058	0.0045		
total	6.2609	6.2609	1	1		

Figure 26: PCA scree plot based on real datasets of all 6 paremeters for males and females together, along with the variance explanation table for total component

respectively, by adding the successive percentage of variation explained to obtain the running total. Therefore, 94.09% and 94.44% of the variation respectively for males and females are explained by the first two eigenvalues together. This is an acceptable level of tolerance

The scree plots showed again to determined at the point, first two components are high beyond which the remaining eigenvalues are all relatively small and of comparable size.

The two principal components are then plotted in Figure 27 visualization.



(a) Males (b) Females

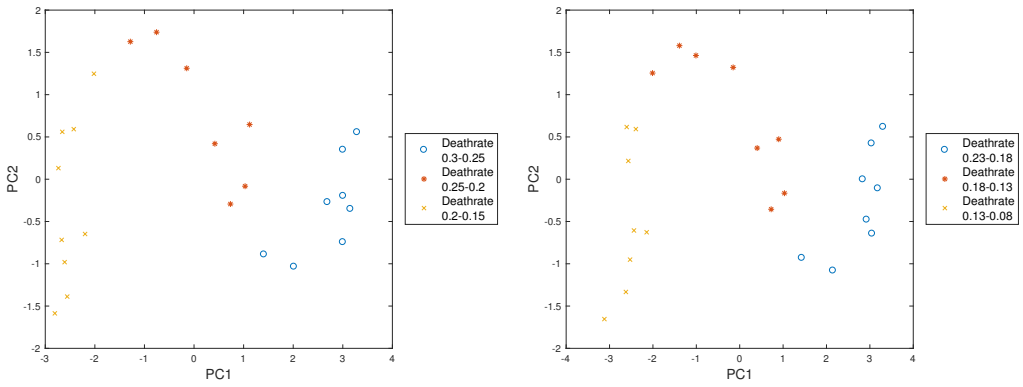
Figure 27: This is PCA visualization of generated UK real datasets based on all 6 parameters (i.e. alcohol, cheese, smoking, SBP, cereals, fruits and vegs) for (a) males and (b) females respectively. Markers on the visualizations are assigned using the bins (shown in legends) of the death rates.

Table 25: Component Matrix of UK real data for 6 parameters

	male		female	
	PC1	**PC2**	**PC1**	**PC2**
Alcohol	0.3278	**0.8256**	0.3286	**0.8402**
Cheese	**0.4185**	0.1826	**0.4143**	0.1939
Smoking	**0.4189**	0.3584	**0.4197**	0.3526
SBP	**0.4179**	0.3419	**0.4231**	0.3199
cereals	**0.4233**	0.1959	**0.4242**	0.1687
Fruits&Vegs	**0.4337**	0.0366	**0.4303**	0.0373

The performance of the visualization of 6-dimensional data are very close to the 4-dimensional projection, the three clusters show similar patterning as well. As previously analysed in Table 25, the blue circled cluster includes the years with the death rates between 0.25 ~ 0.30. In this cluster, high PC1 with low PC2 can be observed, indicating that in this cluster, high scores for cheese, smoking, SBP, cereals, fruits and vegetables are expected. The red starred cluster includes the years with the death rates between 0.20 ~ 0.25. Here large values of PC2 indicate that in this cluster, alcohol score is of a relatively high value. The yellow crossed cluster includes the years with the death rate between 0.15 ~ 0.20 represented in the low range of PC1, indicating a low value of alcohol in this cluster. Similar features were observed for females.

Again, after training the same dataset on NSC visualization, another two almost identical projections as which trained in PCA can be found in the following figure.



(a) Males (b) Females

Figure 28: This is NSC visualization of UK real datasets based on all 6 parameters (i.e. alcohol, cheese, smoking, SBP, cereals, fruits and vegs) for (a) males and (b) females respectively. Markers on the visualizations are assigned using the bins (shown in legends) of the death rates.

Summarised by Table 25, ranking of each parameters are:

Males: Fruits&Vegs > cereals > Smoking > Cheese > SBP > Alcohol

Females: Fruits&Vegs > cereals > SBP > Smoking > Cheese > Alcohol

Table 26: Correlation between 6 parameters

	Alcohol	*Cheese*	*Smoking*	*SBP*	*cereals*	*Fruits&Vegs*
Alcohol	1					
Cheese	0.5783	1				
Smoking	0.7714	0.8668	1			
SBP	0.7327	0.8598	0.8782	1		
cereals	-0.5009	-0.8993	-0.8158	-0.8987	1	
Fruits&Vegs	-0.5006	-0.8624	-0.8318	-0.9073	0.9394	1

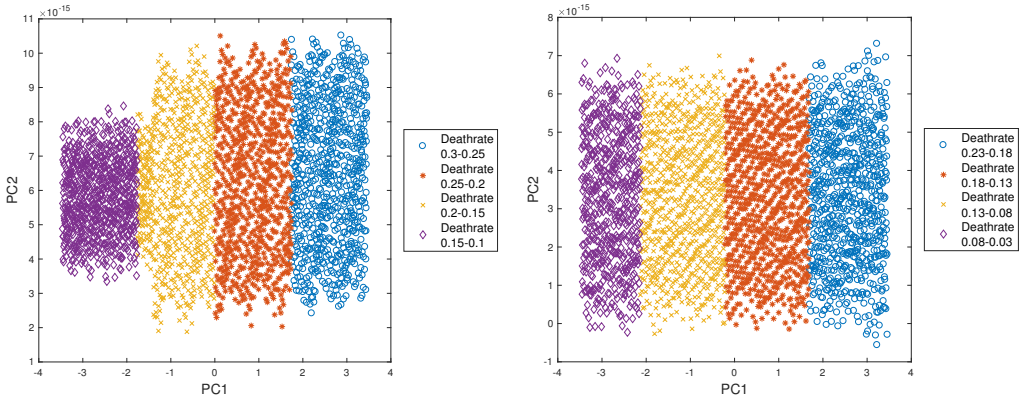
By adding the two *positive indicators*, while the weight factors change, there is no change in the ranking depicted by the four *negative indicators* only. the two positive indicators are then to be the most important parameters now. Further more, Table 26 summarise the correlation-ship between each of the 6 parameters, clearly showed, the hypothesised two positive indicators have negative correlation between both of the negative indicators. We may conclude, the two positive indicators are the most important factors influencing the CHD death rate in the UK.

To investigate the impact of positive indicators on the negative indicators, and their influence on CHD death rates, we do the following data modelling experiments.

4.3.2 *UK visualization based on synthetic datasets*

Statistical estimation relies on large datasets whereas real life datasets are limited to finite sized elements only. In our case, while we were lucky enough to avail datasets over more than 3 decades, statistically, this amounts to about 30 data points only which is a low number to base any statistical prediction on. This impacts the accuracy and validity from such enumeration. To allay this, we chose to enlarge the data bank by extrapolating *s*ynthetic (artificial) data using he regression fitted formulae together with the real data that we had. Between the years 1990 and 2020, with 0.01 interval, 3001 numbers were generated for each of the parameters for each country. Further more, by keeping the validity and factuality of the experience, another set

of databases named real-synthetic databases is generated by keeping the year of which we have got the real data in place, and also within the 0.01 interval, filling the rest of spaces with the artificial datas generated by the linear fitted regression models. This now allows us to to use a large enough statistical dataset to make our visualization based predictions. While admittedly this relied on extrapolation, but we were extremely *l*ucky in that the real data fitted almost perfectly to linear regression fits. A later target of this study will be to analyse the underlying cause of such linear behaviour, something that we are utilising presently, but are not entirely sure as to why this happens. For the rest of this study, we will concentrate on the real-synthetic databases.



(a) Pure synthetic - males

(b) Pure synthetic - females

(c) Real-synthetic - males

(d) Real-synthetic - females

Figure 29: This is PCA visualization of UK synthetic datasets based on 4 negative indicators (i.e. alcohol, cheese, smoking, SBP) for (a) pure synthetic data for males, (b) pure synthetic data for females and (c) real-synthetic data for males, (d) real-synthetic data for females respectively. Markers on the visualizations are assigned using the bins (shown in legends) of the death rates.

Figure 29 shows the PCA visualization of UK synthetic dataset based on 4 negative indicators. In which, the first row based on the pure synthetic dataset for (a) males

Table 27: Variance explanation of 4 parameters for UK real-synthetic data

	Eigenvalues		Percentage of Variance		Cumulative Percentage	
Component	Males	Females	Males	Females	Males	Females
1	3.9537	3.9542	0.9881	0.9882	**0.9881**	**0.9882**
2	0.0466	0.0462	0.0116	0.0115	**0.9997**	**0.9998**
3	0.0007	0.0008	0.0002	0.0002		
4	0.0003	0.0002	0.0001	0.0000		
total	4.0013	4.0013	1	1		

and (b) females respectively, whereas the second row based on the real-synthetic dataset. In general, the first two components (i.e. PC1 and PC2) are included in the consideration. However, in this case, we can only consider the PC1, as shown in Table 27, we can see the eigenvalue of the first component is 3.9537 for males and 3.9542 for females, they are almost 100 times larger compared with the second component, which is 0.0466 for males and 0.0462. And 98.81% and 98.82% of the variance is explained by this first eigenvalue for males and females, whereas only 1.16% and 1.15% of the variance is explained by the second eigenvalues, which are small enough to be neglected.

By neglecting the PC2, we only analysis the plots along with the x-axis, which is PC1. through Figure 29 (a) and (b), we observed four clusters are defined by the range of CHD death rate, which showed in the legend. the plots show a very clear separation for each of the cluster, and with no overlapping. As this is the pure-synthetic dataset, it only used for a visualized purpose of showing a linear trend of the datasets, there is no too much value for ranking the importance of each parameter, we then move on to the second row of Figure 29.

The plots based on real-synthetic dataset show a very clear separation for each of the clusters, and with no overlapping as well. The artificial data become a straight in the middle along with zero line of y-axis, which means they have very low values of PC2, as the PC2 is small enough to be omitted, the shape along with the y-axis is not really concerned. The blue circle cluster with the highest PC1 level, includes the years with the death rate between 0.25 ~ 0.30 for males and 0.18 ~ 0.23 for females, expecting of high values of cheese, smoking and SBP. The purple diamond cluster includes the years with the death rate between 0.10 ~ 0.15 for males and 0.03 ~ 0.08 for females, which occupied in the low range of PC1. By accompany with

Table 28: Component matrix of UK real-synthetic data for 4 parameters

Component Matrix				
	male		**female**	
	PC1	**PC2**	**PC1**	**PC2**
Alcohol	0.49408	0.86914	0.49412	0.86923
Cheese	0.50206	0.26489	0.50200	0.26940
Smoking	0.50194	0.29280	0.50190	0.29569
SBP	0.50188	0.29781	0.50194	0.29058

the component matrix table, ranking of the 4 parameters are:

Males: Cheese > Smoking > SBP > Alcohol

Females: Cheese > SBP > Smoking > Alcohol

Compared with the ranking based on real dataset, cheese becomes the most contributed negative indicator.

The same dataset applied for an NSC training and shown in the following

By neglecting the PC2, the clusters separation performed by NSC for both two datasets are similar as the PCA training.

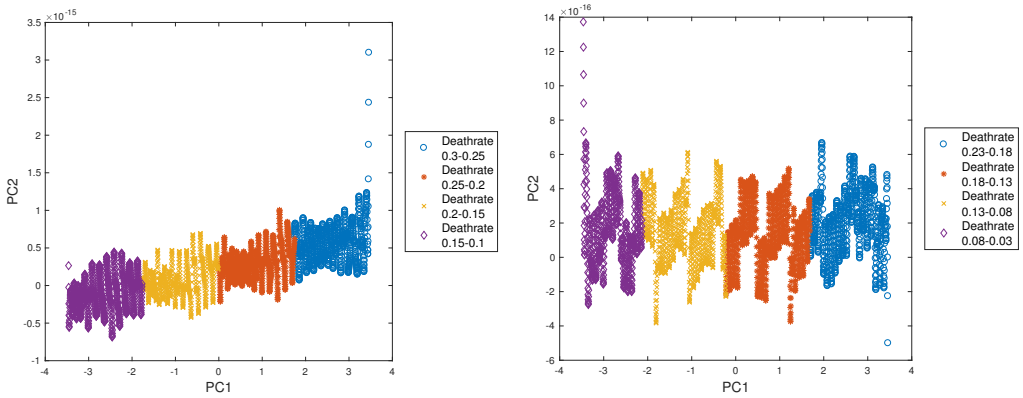
Table 29: Variance explanation of 6 parameters for UK real-synthetic data

	Eigenvalues		Percentage of Variance		Cumulative Percentage	
Component	Males	Females	Males	Females	Males	Females
1	5.9512	5.9517	0.9915	0.9916	**0.9915**	**0.9916**
2	0.0486	0.0483	0.0081	0.0080	**0.9996**	**0.9997**
3	0.0010	0.0009	0.0002	0.0002		
4	0.0006	0.0006	0.0001	0.0001		
5	0.0003	0.0003	0.0001	0.0001		
6	0.0002	0.0001	0.0000	0.0000		
total	6.0020	6.0020	1	1		

Training on the 6-dimensional synthetic datasets, the percentage of variance explained by the first component both for males and females are large enough (refer to Table 29) to represent the importance of each parameters by neglecting PC2. Results show a very similar clusters separation as in 4-dimensional visualization.

By accompanying with the component matrix table in the following, ranking of the 6 parameters are:

Males: Fruits&Vegs > cereals > Cheese > Smoking > SBP > Alcohol



(a) Pure synthetic - males

(b) Pure synthetic - females

(c) Real-synthetic - males

(d) Real-synthetic - females

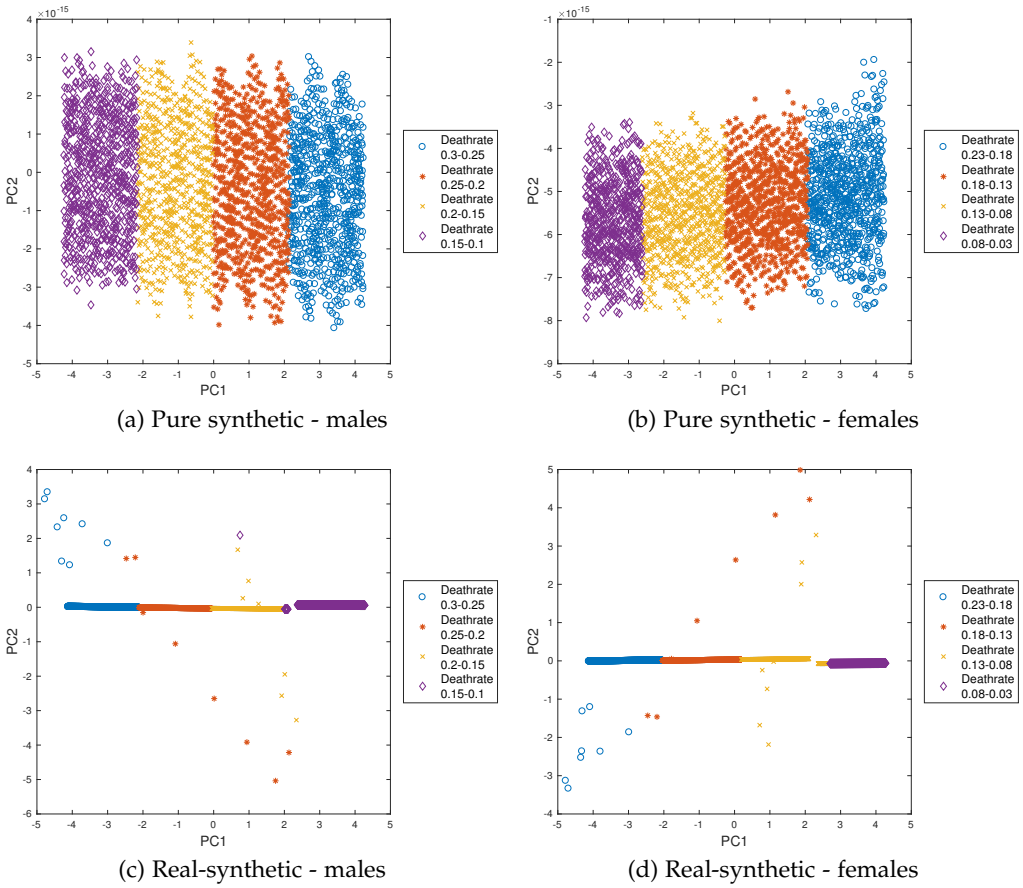
Figure 30: This is NSC visualization of UK synthetic datasets based on 4 negative indicators (i.e. alcohol, cheese, smoking, SBP) for (a) pure synthetic data for males, (b) pure synthetic data for females and (c) real-synthetic data for males, (d) real-synthetic data for females respectively. Markers on the visualizations are assigned using the bins (shown in legends) of the death rates.

Females: cereals > Fruits&Vegs > Cheese > SBP > Smoking > Alcohol

compare with the ranking ordered by the same dataset with 4 parameters only, the two positive indicators become the most important contribution of CHD death rate, and as mentioned in Table 26, the negative relationship between the two positive indicators with the other parameters, this ranking means the impact of cereals, fruits and vegetables of CHD death rate is positive whereas the other 4 parameters are negative, and also in the certain volume, they will even influence the impact of negative indicators on CHD death rate.

The same dataset applied for an NSC training and shown in the following,

By neglecting the PC2, the clusters separation performed by NSC for both two datasets are largely identical to the PCA training. Based on our linear model charac-



(a) Pure synthetic - males

(b) Pure synthetic - females

(c) Real-synthetic - males

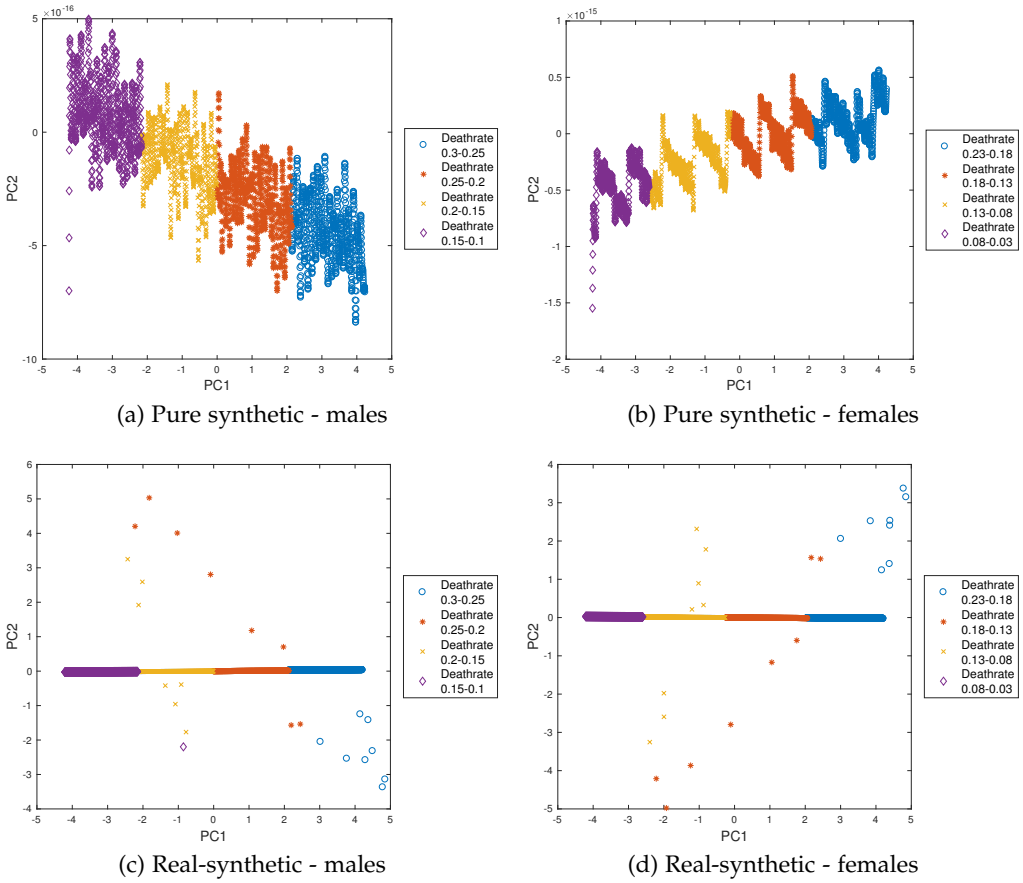
(d) Real-synthetic - females

Figure 31: This is PCA visualization of UK synthetic datasets based on all 6 parameters (i.e. alcohol, cheese, smoking, SBP, cereals, fruits and vegs) for (a) pure synthetic data for males, (b) pure synthetic data for females and (c) real-synthetic data for males, (d) real-synthetic data for females respectively. Markers on the visualizations are assigned using the bins (shown in legends) of the death rates.

Table 30: Component matrix of UK real-synthetic data for 6 parameters

Component Matrix				
	male		**female**	
	PC1	**PC2**	**PC1**	**PC2**
Alcohol	0.40166	0.90956	0.40169	0.91019
Cheese	0.40947	0.19956	0.40944	0.20217
Smoking	0.40941	0.22680	0.40940	0.22812
SBP	0.40939	0.23251	0.40943	0.22344
cereals	0.40974	0.10761	0.40974	0.11118
Fruits&Vegs	0.40975	0.12567	0.40972	0.12790

terise, it is not worse to go for NSC visualization anymore. The rest of the experiment will only train the model by PCA.



(a) Pure synthetic - males

(b) Pure synthetic - females

(c) Real-synthetic - males

(d) Real-synthetic - females

Figure 32: This is NSC visualization of UK synthetic datasets based on all 6 parameters (i.e. alcohol, cheese, smoking, SBP, cereals, fruits and vegs) for (a) pure synthetic data for males, (b) pure synthetic data for females and (c) real-synthetic data for males, (d) real-synthetic data for females respectively. Markers on the visualizations are assigned using the bins (shown in legends) of the death rates.

4.3.3 *Prediction Models*

As discussed in subsection 4.3.2, the two positive indicators are found to be the most important contribution of CHD death rate, the impact of cereals, fruits and vegetables of CHD death rate is positive whereas the other 4 parameters are negative, we hypothesise the two positive indicators parameters will influence the impact of negative indicators on CHD death rate in some certain volume, and aim to build a prediction model through this analysis.

4.3.3.1 *Parameters Ranking*

By our hypothesis, the first step is to build a set of new database by incrementally increasing the values of the positive indicators on real-synthetic datasets year 2013

onward,*i. e.* increasing the fruits and vegetables consumption by 5%, 10%, etc. while keeping the rest of the parameters unchanged. Then, by training PCA visualization on each of the new datasets, the new tables of features weighting are calculated and ordered. We intend to use these features weighting increments to find out the impact of increased positive indicators on each of the negative indicators and thereby establish functional relationships between themselves, and eventually with the CHD death rate.

The number increments chosen are in consonance with practical life style based estimation that then leads to the following:

1. set 1: Increasing the consumption of fruits and vegetables by 2%
2. set 2: Increasing the consumption of fruits and vegetables by 5%
3. set 3: Increasing the consumption of fruits and vegetables by 8%
4. set 4: Increasing the consumption of fruits and vegetables by 10%
5. set 5: Increasing the consumption of fruits and vegetables by 15%
6. set 6: Increasing the consumption of fruits and vegetables by 20%
7. set 7: Increasing the consumption of fruits and vegetables by 25%
8. set 8: Increasing the consumption of cereals by 5%, together with the increase of fruits and vegetables by 2%
9. set 9: Increasing the consumption of cereals by 5%, together with the increase of fruits and vegetables by 5%
10. set 10: Increasing the consumption of cereals by 5%, together with the increase of fruits and vegetables by 8%
11. set 11: Increasing the consumption of cereals by 5%, together with the increase of fruits and vegetables by 10%
12. set 12: Increasing the consumption of cereals by 5%, together with the increase of fruits and vegetables by 15%
13. set 13: Increasing the consumption of cereals by 5%, together with the increase of fruits and vegetables by 20%

14. set 14: Increasing the consumption of cereals by 5%, together with the increase of fruits and vegetables by 25%
15. set 15: Increasing the consumption of cereals by 10%, together with the increase of fruits and vegetables by 2%
16. set 16: Increasing the consumption of cereals by 10%, together with the increase of fruits and vegetables by 5%
17. set 17: Increasing the consumption of cereals by 10%, together with the increase of fruits and vegetables by 8%
18. set 18: Increasing the consumption of cereals by 10%, together with the increase of fruits and vegetables by 10%
19. set 19: Increasing the consumption of cereals by 10%, together with the increase of fruits and vegetables by 15%
20. set 20: Increasing the consumption of cereals by 10%, together with the increase of fruits and vegetables by 20%
21. set 21: Increasing the consumption of cereals by 10%, together with the increase of fruits and vegetables by 25%

We group these 21 new generated datasets into 3 blocks by the different change amount of two positive indicators:

- Block 01: Keeping the consumption of cereals unchanged, and increasing the consumption of fruits and vegetables by certain percentages only.
- Block 02: Increasing the consumption of cereals by 5%, together with the increase of the consumption of fruits and vegetables by certain percentages.
- Block 03: Increasing the consumption of cereals by 10%, together with the increase of the consumption of fruits and vegetables by certain percentages.

These 3 blocks therefore are visualized and analysed using the same methods which described in subsection 4.3.2. Training them by PCA visualization, both of them represent the separation of four clusters very clearly and no overlapping, and the features weighing of PC1 are calculated and tabled in the following for males and females respectively in Table 31 toTable 34.

Table 31: Features weighting of PC1: Males - Block 01

PC1 - Male - Increase consumption of Fruits&Vegs only							
	Alcohol	**Cheese**	**Smoking**	**BP**	**Cereals**	**Fruits&Vegs**	**Variance Preservation (%)**
4 parameters[a]	0.49408	**0.50206**	0.50194	0.50188			98.8093
6 parameters[b]	0.40166	**0.40947**	0.40941	0.40939	0.40974	0.40975	99.1539
fruits&vegs + 2%	0.40168	**0.40953**	0.40946	0.40944	0.40980	0.40951	99.1246
fruits&vegs + 5%	0.40185	**0.40975**	0.40967	0.40966	0.41001	0.40849	99.0006
fruits&vegs + 8%	0.40212	**0.41006**	0.40999	0.40997	0.41033	0.40696	98.8163
fruits&vegs + 10%	0.40234	**0.41030**	0.41023	0.41021	0.41057	0.40577	98.6741
fruits&vegs + 15%	0.40294	**0.41096**	0.41088	0.41087	0.41123	0.40249	98.2893
fruits&vegs + 20%	0.40357	**0.41163**	0.41156	0.41154	0.41190	0.39910	97.9006
fruits&vegs + 25%	0.40419	**0.41228**	0.41220	0.41219	0.41255	0.39580	97.5319

a This uses the real-synthetic dataset of 4 negative indicators only, with no increase of any of the positive indicators, detailed analysis discussed in subsection 4.3.2

b This uses the real-synthetic dataset of all 6 parameters, with no increase of any of the positive indicators, detailed analysis discussed in subsection 4.3.2

Table 32: Features weighting of PC1: Males - Block 02,03

PC1 - Male - Increase consumption of Cereals and Fruits&Vegetables							
	Alcohol	**Cheese**	**Smoking**	**BP**	**Cereals**	**Fruits&Vegs**	**Variance Preservation (%)**
4 parameters[a]	0.49408	**0.50206**	0.50194	0.50188			98.8093
6 parameters[b]	0.40166	**0.40947**	0.40941	0.40939	0.40974	0.40975	99.1539
cereal+5%, fruits&vegs + 2%	0.40192	**0.40990**	0.40983	0.40981	0.40765	0.41032	98.8640
cereal+5%, fruits&vegs + 5%	0.40189	**0.40991**	0.40984	0.40983	0.40826	0.40969	98.7877
cereal+5%, fruits&vegs + 8%	0.40199	**0.41006**	0.40999	0.40997	0.40891	0.40851	98.6443
cereal+5%, fruits&vegs + 10%	0.40211	**0.41020**	0.41013	0.41011	0.40935	0.40753	98.5259
cereal+5%, fruits&vegs + 15%	0.40251	**0.41064**	0.41057	0.41056	0.41042	0.40471	98.1911
cereal+5%, fruits&vegs + 20%	0.40297	**0.41115**	0.41107	0.41106	0.41142	0.40170	97.8414
cereal+5%, fruits&vegs + 25%	0.40344	**0.41166**	0.41158	0.41157	0.41231	0.39872	97.5036
cereal+10%, fruits&vegs + 2%	0.40267	**0.41073**	0.41067	0.41065	0.40321	0.41146	98.3375
cereal+10%, fruits&vegs + 5%	0.40250	**0.41061**	0.41055	0.41053	0.40411	0.41110	98.2938
cereal+10%, fruits&vegs + 8%	0.40249	**0.41064**	0.41057	0.41056	0.40501	0.41016	98.1784
cereal+10%, fruits&vegs + 10%	0.40253	**0.41071**	0.41064	0.41063	0.40559	0.40931	98.0765
cereal+10%, fruits&vegs + 15%	0.40278	**0.41101**	0.41094	0.41093	0.40697	0.40680	97.7764
cereal+10%, fruits&vegs + 20%	0.40312	**0.41139**	0.41132	0.41131	0.40820	0.40406	97.4542
cereal+10%, fruits&vegs + 25%	0.40350	**0.41180**	0.41173	0.41172	0.40929	0.40131	97.1382

a This uses the real-synthetic dataset of 4 negative indicators only, with no increase of any of the positive indicators, detailed analysis discussed in subsection 4.3.2

b This uses the real-synthetic dataset of all 6 parameters, with no increase of any of the positive indicators, detailed analysis discussed in subsection 4.3.2

Table 33: Features weighting of PC1: Females - Block 01

PC1 - Female - Increase consumption of Fruits&Vegetables only							
	Alcohol	**Cheese**	**Smoking**	**BP**	**Cereals**	**Fruits&Vegs**	**Variance Preservation (%)**
4 parameters[a]	0.49412	**0.50200**	0.50190	0.50194			98.8222
6 parameters[b]	0.40169	**0.40944**	0.40940	0.40943	0.40974	0.40972	99.1621
fruits&vegs + 2%	0.40171	**0.40950**	0.40946	0.40949	0.40979	0.40948	99.1328
fruits&vegs + 5%	0.40187	**0.40971**	0.40967	0.40970	0.41000	0.40846	99.0088
fruits&vegs + 8%	0.40214	**0.41003**	0.40999	0.41002	0.41032	0.40693	98.8246
fruits&vegs + 10%	0.40236	**0.41027**	0.41023	0.41026	0.41056	0.40574	98.6823
fruits&vegs + 15%	0.40297	**0.41093**	0.41089	0.41091	0.41122	0.40246	98.2976
fruits&vegs + 20%	0.40360	**0.41160**	0.41156	0.41159	0.41189	0.39907	97.9089
fruits&vegs + 25%	0.40421	**0.41225**	0.41220	0.41223	0.41254	0.39577	97.5402

a This uses the real-synthetic dataset of 4 negative indicators only, with no increase of any of the positive indicators, detailed analysis discussed in subsection 4.3.2

b This uses the real-synthetic dataset of all 6 parameters, with no increase of any of the positive indicators, detailed analysis discussed in subsection 4.3.2

Table 34: Features weighting of PC1: Females - Block 02,03

PC1 - Female - Increase consumption of Cereals and Fruits&Vegetables							
	Alcohol	**Cheese**	**Smoking**	**BP**	**Cereals**	**Fruits&Vegs**	**Variance Preservation (%)**
4 parameters[a]	0.49412	**0.50200**	0.50190	0.50194			98.8222
6 parameters[b]	0.40169	**0.40944**	0.40940	0.40943	0.40974	0.40972	99.1621
cereal+5%, fruits&vegs + 2%	0.40195	**0.40987**	0.40983	0.40986	0.40764	0.41029	98.8720
cereal+5%, fruits&vegs + 5%	0.40191	**0.40988**	0.40984	0.40987	0.40825	0.40967	98.7957
cereal+5%, fruits&vegs + 8%	0.40202	**0.41003**	0.40999	0.41002	0.40890	0.40848	98.6522
cereal+5%, fruits&vegs + 10%	0.40214	**0.41017**	0.41013	0.41016	0.40934	0.40750	98.5339
cereal+5%, fruits&vegs + 15%	0.40253	**0.41061**	0.41057	0.41060	0.41041	0.40468	98.1991
cereal+5%, fruits&vegs + 20%	0.40299	**0.41112**	0.41108	0.41110	0.41140	0.40167	97.8494
cereal+5%, fruits&vegs + 25%	0.40347	**0.41162**	0.41158	0.41161	0.41230	0.39870	97.5116
cereal+10%, fruits&vegs + 2%	0.40269	**0.41070**	0.41067	0.41070	0.40320	0.41143	98.3453
cereal+10%, fruits&vegs + 5%	0.40253	**0.410582**	0.41054	0.410575	0.40410	0.41108	98.3016
cereal+10%, fruits&vegs + 8%	0.40251	**0.410607**	0.41057	0.410600	0.40500	0.41013	98.1862
cereal+10%, fruits&vegs + 10%	0.40256	**0.41068**	0.41064	0.41067	0.40558	0.40929	98.0843
cereal+10%, fruits&vegs + 15%	0.40281	**0.41098**	0.41094	0.41097	0.40696	0.40678	97.7842
cereal+10%, fruits&vegs + 20%	0.40315	**0.41136**	0.41132	0.41135	0.40819	0.40403	97.4620
cereal+10%, fruits&vegs + 25%	0.40353	**0.41177**	0.41173	0.41176	0.40928	0.40128	97.1460

a This uses the real-synthetic dataset of 4 negative indicators only, with no increase of any of the positive indicators, detailed analysis discussed in subsection 4.3.2

b This uses the real-synthetic dataset of all 6 parameters, with no increase of any of the positive indicators, detailed analysis discussed in subsection 4.3.2

Observed from the weighing tables above, we can find for instance of males, if we test for 4 negative indicators only, cheese is the highest influence parameter. Once adding on the two positive of the parameters for the test, one of the positive indicators to be the greatest effect on the CHD death rate. As the two positive indicators are in the positive effective position, we now ranking the 4 negative indicators only. what can be found from our results is, whatever how much of the percentage changes on the two positive indicators, cheese consumption keeps the greatest affecting on CHD death rate always, ranking of the 4 negative indicators for males are concluded as:

$$\text{Cheese} > \text{Smoking} > \text{SBP} > \text{Alcohol}$$

And for females, if we test for 4 negative indicators only, cheese still be the greatest effect of CHD death rate, after adding up the two positive indicators for the test, cheese keeps the first place consistently among the negative indicators, and the ranking without the positive affection parameters is:

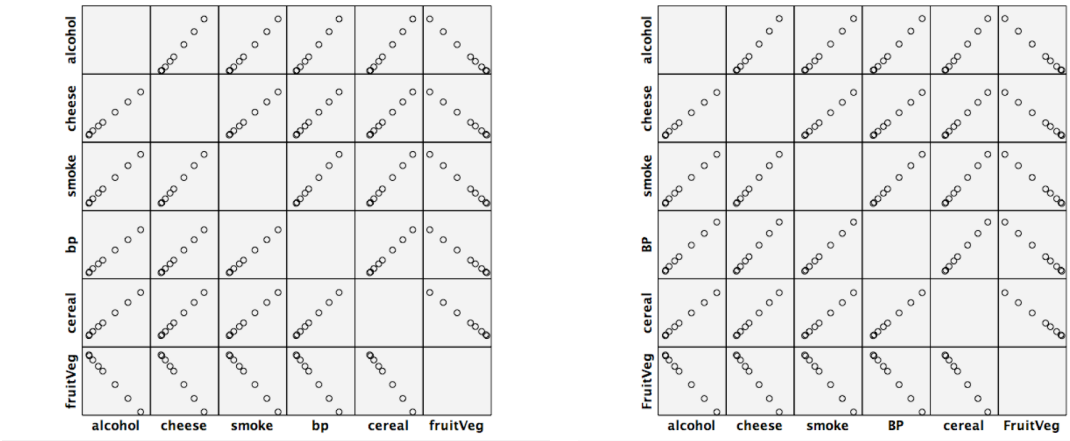
$$\text{Cheese} > \text{SBP} > \text{Smoking} > \text{Alcohol}$$

From the slight difference of the ranking, we can say that a result of gender difference can be found in our study, the impact of smoking on CHD death rate for males is greater than the impact of SBP, but SBP has a more important impact than smoking among females.

4.3.3.2 *Models*

To build up the prediction model, we have to understand the correlation-ship between each parameter for the three data blocks which introduced in subsubsection 4.3.3.1.

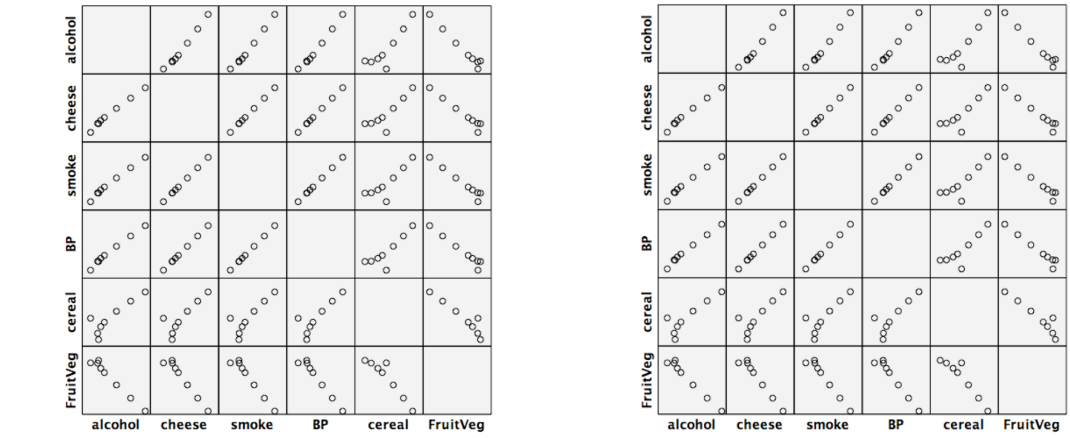
Block 01 is the datasets of Keeping the consumption of cereals unchanged, and increasing the consumption of fruits and vegetables by certain percentages, the correlation between each of the parameters are shown in the following scatter matrix, which shows a clearly linear correlation between each of the parameters. Fruits and vegetables show a negative correlation with all the other parameters, which is what are we expected, but from the scatter matrix, cereals consumption shows a positive correlation with all the negative indicators, this is opposite with our expectation, this finding should be proven in deep at our future works.



(a) Males (b) Females

Figure 33: This is the correlation scatter matrix, show the correlation plots between each of the parameters based on block 01 datasets for PC1 of (a) males and (b) females.

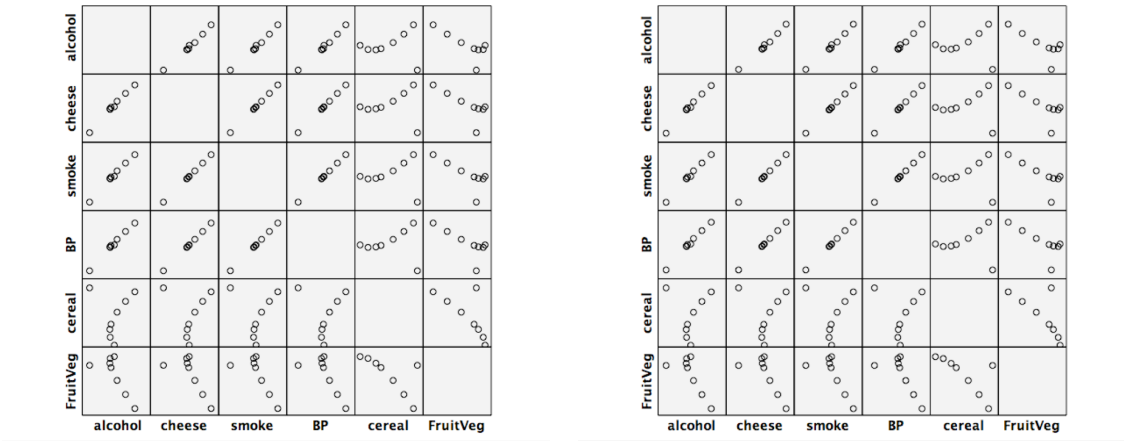
The correlation scatter matrix of block 02 datasets which are the set of increasing the consumption of cereals by 5%, together with the increase of the consumption of fruits and vegetables by certain percentages are shown in Figure 34.



(a) Males (b) Females

Figure 34: This is the correlation scatter matrix, show the correlation plots between each of the parameters based on block 02 datasets for PC1 of (a) males and (b) females.

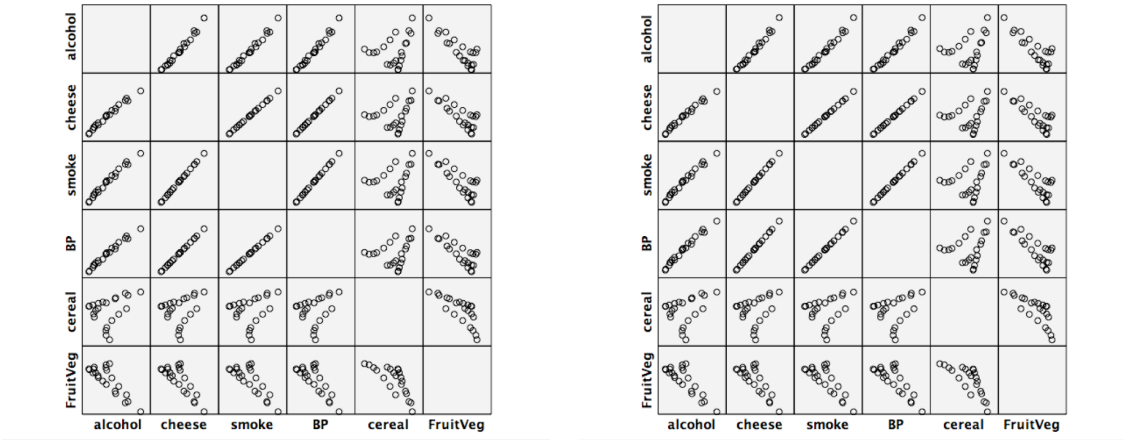
The correlation scatter matrix of block 03 datasets, which is the set of increasing cereal concentration by 10%, together with the increase of the fruits and vegetables by certain percentages are shown in Figure 35.The correlations between 4 negative indicators are all showing linear trends. Fruits and vegetables retain the negative linear correlations with all other parameters.



(a) Males (b) Females

Figure 35: This is the correlation scatter matrix, show the correlation plots between each of the parameters based on block 03 datasets for PC1 of (a) males and (b) females.

The following figure sums up the plots for all the datasets in the three blocks, the correlations between the 4 negative indicators are clearly show a perfect linear relationship, and due to the increasing percentages we have applied to the datasets, the trending lines of the correlation between the positive indicators and all other factors represented an increasing trend in the scatter matrix as well.



(a) Males (b) Females

Figure 36: This is the correlation scatter matrix, show the correlation plots between each of the parameters based on all the datasets in 3 blocks for PC1 of (a) males and (b) females.

The final step to build our model is run a multivariate regression to predict the impact of parameters from each of the other ones.

Males:

1. Set SBP as dependent variable, test the relationship between SBP and all the other parameters, we find alcohol, cereals, fruits and vegetables has a strong influence on SBP, with $R^2 = 0.999$, standard error smaller then 0.0001, the equation of this linear model is:

   $$SBP = -0.459 \times Alcohol - 0.358 \times Cereals - 0.357 \times Fruits\&Vegs + 0.886$$

2. Set smoking as dependent variable, test the relationship between smoking and all the other parameters, we find alcohol, cereals, fruits and vegetables has a strong influence on smoking, with $R^2 = 0.999$, standard error smaller then 0.0001, the equation of this linear model is:

   $$Smoking = -0.422 \times Alcohol - 0.349 \times Cereals - 0.348 \times Fruits\&Vegs + 0.864$$

3. Set cheese as dependent variable, test the relationship between cheese and all the other parameters, we find alcohol, cereals, fruits and vegetables has a strong influence on cheese, with $R^2 = 0.999$, standard error smaller then 0.0001, the equation of this linear model is:

   $$Cheese = -0.441 \times Alcohol - 0.352 \times Cereals - 0.352 \times Fruits\&Vegs + 0.875$$

4. Set alcohol as dependent variable, test the relationship between alcohol and all the other parameters, we find BP, cereals, fruits and vegetables has a strong influence on alcohol, with $R^2 = 0.999$, standard error smaller then 0.0001, the equation of this linear model is:

   $$Alcohol = -1.995 \times SBP - 0.731 \times Cereals - 0.730 \times Fruits\&Vegs + 1.817$$

Females:

1. Set SBP as dependent variable, test the relationship between SBP and all the other parameters, we find alcohol, cereals, fruits and vegetables has a strong influence on SBP, with $R^2 = 0.999$, standard error smaller then 0.0001, the equation of this linear model is:

   $$SBP = -0.454 \times Alcohol - 0.356 \times Cereals - 0.355 \times Fruits\&Vegs + 0.883$$

2. Set smoking as dependent variable, test the relationship between smoking and all the other parameters, we find alcohol, cereals, fruits and vegetables has a

strong influence on smoking, with $R^2 = 0.999$, standard error smaller then 0.0001, the equation of this linear model is:

$$Smoking = -0.448 \times Alcohol - 0.354 \times Cereals - 0.354 \times Fruits\&Vegs + 0.880$$

3. Set cheese as dependent variable, test the relationship between cheese and all the other parameters, we find alcohol, cereals, fruits and vegetables has a strong influence on cheese, with $R^2 = 0.999$, standard error smaller then 0.0001, the equation of this linear model is:

$$Cheese = -0.434 \times Alcohol - 0.351 \times Cereals - 0.351 \times Fruits\&Vegs + 0.871$$

4. Set alcohol as dependent variable, test the relationship between alcohol and all the other parameters, we find BP, cereals, fruits and vegetables has a strong influence on alcohol, with $R^2 = 0.999$, standard error smaller then 0.0001, the equation of this linear model is:

$$Alcohol = -2.042 \times SBP - 0.742 \times Cereals - 0.741 \times Fruits\&Vegs + 1.845$$

4.3.4 *Key Knowledge Base*

These models establish relations between the life-style parameters. This is a major finding that enables a doctor on how to advise a patient. For example, how much of a fruit intake could probabilistically lower their CHD death risk. We can clearly see this from our model prediction: $SBP = -0.459 \times Alcohol - 0.358 \times Cereals - 0.357 \times Fruits\&Vegs + 0.886$; this tells a doctor how an increase in fruit intake and a possible decrease in cereal consumption could impact the SBP and alcohol consumption. On the other hand, if we increase the consumption of alcohol by 20%, keeping the consumption of cereals, fruits and vegetables habit unchanged, we will have: $SBP = -0.459 \times 1.20 - 0.358 \times 1 - 0.357 \times 1 + 0.886 = 0.7218$, which means, if a consumption of alcohol increased by 20%, the chance to have higher SBP is 72.18%. However, if 25% more fruits and vegetables are consumed, the chance of high SBP reduces to 63.26%, which also means that 25% more consumption of fruits will decrease the chance of high SBP by 8.92%. Such numbers will, for the first time, define clear numerical guidance to medical practitioners on how best to advise patients. In

the next follow-up study, we expect to correlate these findings with the CHD rates, connecting all contributing factors, both positive and negative.

There same analysis process in section 4.3 is applied to all the other 12 countries, and the visualization plots and results are shown and briefed in Appendix D along with the ranking order of parameters for each country.

4.4 SUMMARY

Findings from this chapter;

- Introduced the feature weighting estimation method in the purpose of ranking and ordering the parameters we are studying with.
- Apply PCA and NSC visualization on real dataset of UK, the order for 4 negative indicators is found:

  Smoking > Cheese > SBP > Alcohol for males, and

  SBP > Smoking > Cheese > Alcohol for females
- Apply PCA and NSC visualization on synthetic dataset of UK, the order for 4 negative indicators is found:

  Cheese > Smoking > SBP > Alcohol for males, and

  Cheese > SBP > Smoking > Alcohol for females
- Apply PCA visualization and calculate the features weighting of the 21 sets of new data to generates the prediction model of each negative indicators.

In the next chapter, we formulate a new continuum model that relates to the time evolution for the purpose of prediction on the impact of all these life-style factors on CHD death rate. This approach is guided by the need to develop a probabilistic description of life style dependence on health prognosis.

5

CONTINUUM MODEL: TIME EVOLUTION OF LIFE-STYLE FACTORS

This chapter details an on-going work that uses the previous data modelling results to develop a new continuum model that relates to time evolution heuristics. The objective of this analysis is to make probabilistic prediction on the impact of all these life-style factors in affecting CHD death rates.

5.1 MATHEMATICAL MODELS

Models describe our beliefs about how the world functions, we can model the behaviour of a given population with the use of certain mathematical models. In this section, we will be comparing two classes of mathematical models and then choose the one that best fits our data modelling syntax.

The first of these is the popular Verhulst model for modelling the continuous population of a single species and the Lotka-Volterra model (Matsuda et al., 1992) for predator-prey interactions are present (Murray, 2002). We will first introduce the concept of using Ordinary Differential Equations (ODE) to model population systems before detailing these two mathematical models.

5.1.1 *Ordinary Differential Equations (ODEs)*

Given a dependent variable y that is a function of an independent variable x, for all continuous smooth functional representations of $y(x)$, we can determine the derivative of $y(x)$ with respect to x $\left(i.\ e. \frac{dy(x)}{dx}\right)$. If $\frac{dy(t)}{dt}$ is positive, the variable $y(x)$ grows as t increases and vice versa.

In mathematical biology, this concept is often used where our variables will generally represent a population (or a sub-population if the total population is split into

multiple parts) and time, such that we can observe how the population of a given organism changes over time, often incorporating the addition of various other factors. For example, let us consider a basic system with time x and a population $y = y(x)$ with a corresponding initial population of $y_0 > 0$, where the rate of growth of the population is proportional to the size of the population at any given time. This gives the following system:

$$\frac{y}{x} = ky \tag{55}$$

Where k is a constant or an arbitrary function not in terms of y. From this ODE we can deduce that, by separation of variables, the solution of y is of the form $y = Ce^{kx}$, where C is a constant. At $x = 0, y(x = 0) = y_0$, hence the constant $C = y_0$ represents our initial population and the solution can be rewritten to give

$$y = y_0 e^{kx} \tag{56}$$

Hence, when the population y (dependent variable) is plotted against time x (independent variable), y grows exponentially with a starting point of y_0. Plotting $\frac{dy}{dx}$ against y however, will instead yield a linear graph as Equation 55 implies $\frac{dy}{dx}$ and y are proportional to one another, where k is the gradient of this line. In both cases, there is a flaw in that there is no limit to the growth of the populations which could be seen as unrealistic and hence needs to be rectified.

The steady state solutions of the system, that is the point(s) at which there is no change in the value of the function (*i. e.* $\frac{dy}{dx} = 0$). Steady states are useful in that they can identify points at which there is no growth in the population or if there is no change in the rate of growth of the population.

In this study, we will be using first order ODEs of the form $\frac{dy}{dx} = f(y)$ to model the risk factors.

5.1.2 *Linear Stability Analysis*

Another important concept which used in the analysis of mathematical model in the later sections is stability of ODEs.

Given an ODE $x' = f(x)$. A fixed point is a point where $x' = 0$. This requires $f(x) = 0$. So any roots of the function $f(x)$ is a fixed point. At a fixed point where $f(x) = 0$, if $f'(x) > 0$ we have $f(x)$ is increased at x, or say $f(x + \epsilon) > 0 > f(x - \epsilon)$ for all sufficiently small and positive step ϵ. This shows that if starting with initial value $x_0 > x$, but close to x, since $f(x_0) > 0$ we will have the ODE forces the particle to increase its value of x, and move away from the fixed point. If starting with $x_0 < x$, but close to x, the ODE will now force the particle to decrease its value of x, and move away from the fixed point. Hence if $f'(x) > 0$, say that the fixed point is unstable, and vice versa (Strogatz, 2018).

Based on the description above, we can define stable and unstable fixed points as following:

$$f(x + \epsilon) < 0, f(x - \epsilon) > 0 \Rightarrow \text{stable};$$
$$f(x + \epsilon) > 0, f(x - \epsilon) < 0 \Rightarrow \text{unstable}.$$

We can now introduce the Linear Stability Analysis which is useful in identifying further features of the function, e.g. a linearly stable function implies that when perturbed, the system will revert back to its linearly stable (steady) state. Consider the constant fixed point x' of the function $f(x)$, to identify if the function is increasing or decreasing, we need to determine its rate of change, hence, we find its derivative:

$$\dot{\epsilon} = \epsilon' = f(x' + \epsilon) = f(x' + \epsilon) \tag{57}$$

Using Taylor's expansion to get:

$$\dot{\epsilon} = f(x' + \epsilon) = f(x') + \epsilon f'(x') + E(\epsilon^2) = \epsilon f'(x') + E(\epsilon^2) \tag{58}$$

where $f(x') = 0$ and $E(\epsilon^2)$ represents small distances for ϵ. Assume the distance between the fixed point x' and the value $x' + \epsilon$ is negligible as $\epsilon \approx 0$, we can eliminate $E(\epsilon^2)$ to give:

$$\dot{\epsilon} = \epsilon f'(x') \tag{59}$$

which is a linear equation in ϵ. And it can be observed ϵ is increased if $f'(x') > 0$ and decreased if $f'(x') < 0$.

Given that $f(x)$ is an ODE system of more than one dimension, we should use Jacobian matrices to determine the characteristics and stability of the system. A Jacobian matrix is a square matrix which contains various partial derivative functions (Weisstein, 2002), therefore, $n-$dimensional function system can be written as,

$$f(x) = \begin{cases} f_1(x_1, x_2 ... x_n) \\ f_2(x_1, x_2 ... x_n) \\ ... \\ f_n(x_1, x_2 ... x_n) \end{cases} \tag{60}$$

The Jacobian matrix gives,

$$J = \begin{bmatrix} \frac{\partial f_1(x)}{\partial x_1} & \frac{\partial f_1(x)}{\partial x_2} & \cdots & \frac{\partial f_1(x)}{\partial x_n} \\ \frac{\partial f_2(x)}{\partial x_1} & \frac{\partial f_2(x)}{\partial x_2} & \cdots & \frac{\partial f_2(x)}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n(x)}{\partial x_1} & \frac{\partial f_n(x)}{\partial x_2} & \cdots & \frac{\partial f_n(x)}{\partial x_n} \end{bmatrix} \tag{61}$$

Eigenvalues λ of Jacobian matrix can be caculated by using its Characteristic equation $|J - \lambda I| = 0$, Where I is the identity matrix. Eigenvalues are useful to determine the properties of the system with a given point.

In two-dimensional system, the two eigenvalues of the Jacobian matrix can either be 2 reals, 2 complexes. And the stability of a fixed point can be evaluate as following:

- Both eigenvalues are real and negative implies the fixed point is **stable**.
- At least one eigenvalue is real and positive implies the fixed point is **unstable**
- One of the eigenvalues is real and positive, the other one is real and negative implies the fixed point is a **saddle point**.
- The eigenvalues are complex-conjugate to one another implies the fixed point is a **focus point** - this is a point in which the system can circulate around it. The focus point is stable if the real part of both eigenvalues is negative and unstable if at least one of the eigenvalues has a positive real part.

In three-dimensional system, the three eigenvalues of the Jacobian matrix are either all reals or one is real and the other two are complex-conjugate, And the stability of a fixed point can be evaluate as following:

- All three eigenvalues are real and negative implies the fixed point is **stable**.
- At least one eigenvalue is real and positive implies the fixed point is **unstable**
- Either one eigenvalue is real and negative and the other two are real and positive, or one eigenvalue is real and positive and the other two are real and negative then the fixed point is a **saddle point**.
- One eigenvalue is real, and the other two are complex-conjugate to one another implies the fixed point is a **focus point**. The focus point is stable if the real part of all eigenvalues are negative and unstable if at least one of the eigenvalues has a positive real part - if the sign of the real eigenvalue differs to the sign of the real part of the complex eigenvalues, then the point is a saddle-focus.

5.1.3 *Verhulst Model (Logistic Growth Model)*

The Verhulst or Logistic Growth model (Strogatz, 2018):

$$\dot{N} = \frac{dN}{dt} = rN\left(1 - \frac{N}{K}\right), (N > 0, r > 0) \tag{62}$$

is a mathematical model that can be used to determine the population growth of a single population. Here the parameter K represents a constraint on the total food supply or any other critical resource whose availability limits the growth of the population. This is the so called *carrying capacity* of the system. r is the maximum growth rate of the population.

At $\dot{N} = 0$, there are two fixed points for Equation 62, $N = 0$ and $N = K$, whereas $N = 0$ is an unstable fixed point since once N^2 can be neglected compared with N, it becomes a linearisation status $\dot{N} \approx rN$. The other fixed point $N = K$ is stable since

$(N-K)^2$ is neglected compared with $|N-K|$, it becomes $d(N-K)/dt \approx -r(N-K)$, so that $t \to \infty$, $N \to K$, so, when $N(0) = N_0$, the solution of Equation 62 is given by:

$$N(t) = \frac{N_0 K e^{rt}}{[K + N_0(e^{rt} - 1)]} \to K, t \to \infty \tag{63}$$

The Verhulst model is normally used to represent single species population growth, e.g. to model the growth population of bacteria in humans, animals or microorganisms, given a finite amount of resources, where the circumstances of the carrying capacity K depends on the organism in question and the scenario it is in. This model effectively acts as a basis for more complex models as it provides a general overview of the way population changes work.

5.1.4 *Lotka-Volterra Model (Predator-Prey Model)*

The Lotka-Volterra Model, also known as the *Predator-Prey Model*, are a pair or a system of first-order nonlinear differential equations, frequently used to describe the dynamics of biological systems in between two species interact or interactions among multiple species, one or some represent as predator(s) and the other(s) as prey(s). The populations change through time according to the equation system (Murray, 2002):

$$\begin{cases} \dot{x} = \frac{dx}{dt} = \alpha x - \beta xy = x(\alpha - \beta y) \\ \\ \dot{y} = \frac{dy}{dt} = \gamma xy - \delta y = -y(\delta - \gamma x), \end{cases} \tag{64}$$

where x is the number of prey, y is the number of predators, t represents time, and $\dot{x}$ and $\dot{y}$ represent the instantaneous growth rates of prey and predator. $\alpha, \beta, \gamma, \delta$ are positive definite parameters that indicate the interactions of preys with predators. The equations have periodic solutions and do not have a simple expression in terms of the usual trigonometric functions, although they are tractable. To analyse the model's stability. First, we need to solve the zero solutions of the equations:

$$\begin{cases} \dot{x} = x(\alpha - \beta y) = 0 \Rightarrow y = \alpha/\beta, x = 0 \\ \\ \dot{y} = -y(\delta - \gamma x) = 0 \Rightarrow x = \delta/\gamma, y = 0. \end{cases} \tag{65}$$

This means Equation 65 yields two steady state points: $\{0, 0\}$ and $\{\frac{\delta}{\gamma}, \frac{\alpha}{\beta}\}$. The first point essentially represents the extinction of both species. If both populations are at 0, then they will continue to be so indefinitely. The second solution represents a fixed point at which both populations sustain their current, non-zero numbers, and the levels of population at which this equilibrium is achieved depend on the chosen values of the parameters $\alpha, \beta, \gamma, \delta$. So, following Equation 61, the Jacobian matrix of the system is then given by

$$J = \begin{bmatrix} \alpha - \beta y & -\beta x \\ \gamma y & \gamma x - \delta \end{bmatrix} \tag{66}$$

By substituting the values of the two steady state points above gives the following matrices:

$$J(x = 0, y = 0) = \begin{bmatrix} \alpha & 0 \\ 0 & -\delta \end{bmatrix}, J(x = \frac{\delta}{\gamma}, y = \frac{\alpha}{\beta}) = \begin{bmatrix} 0 & \frac{-\beta\delta}{\gamma} \\ \frac{\alpha\gamma}{\beta} & 0. \end{bmatrix} \tag{67}$$

Using the equation $|J(x, y) - \lambda I| = 0$, we can determine the characteristic equation and hence the eigenvalues for each matrix:

$$|J(x = 0, y = 0) - \lambda I| = \begin{vmatrix} \alpha - \lambda & 0 \\ 0 & -(\delta + \lambda) \end{vmatrix} = 0 \quad \Rightarrow \lambda_1 = \alpha, \lambda_2 = -\delta \tag{68}$$

$$|J(x = \frac{\delta}{\gamma}, y = \frac{\alpha}{\beta}) - \lambda I| = \begin{vmatrix} -\lambda & \frac{-\beta\delta}{\gamma} \\ \frac{\alpha\gamma}{\beta} & -\lambda \end{vmatrix} = 0 \quad \Rightarrow \lambda_1 = \sqrt{-\alpha\delta}, \lambda_2 = -\sqrt{-\alpha\delta} \tag{69}$$

For $J(x = 0, y = 0)$, its characteristic equations provide one positive and one negative eigenvalue which implies that the point $(0, 0)$ is a **saddle point**, which is unstable. For $J(x = \frac{\delta}{\gamma}, y = \frac{\alpha}{\beta})$, both eigenvalues are imaginary which implies that the point $\{\frac{\delta}{\gamma}, \frac{\alpha}{\beta}\}$ is a **focal point**. Effectively, this is the central point of the system with trajectories circulating around it.

In conclusion, the Lotka-Volterra model considers the interaction between two or more types of population involving the use of two or more variables, with generally

representing the predator(s) and the other(s) the prey(s). The variables also depend on each other, where changes in one variable cause changes in the other. Other variations of this model exist for other types of interactions. Hence, variations of this model are applicable to the majority of biological systems in nature as it provides a general overview for the population dynamics within such systems.

In our case, as we know, a heavy drinker is often found to be a smoker as well. This could imply that an increase in his/her cheese and alcohol consumptions may often lead to an increase in smoking and vice versa. The dormant question that then remains is to enumerate in absolute terms the extent of this dependence. Our perceived model is thus an extended combination of the Verhulst and Lotka-Volterra models.

5.2 PROPOSED MODEL

As discussed in subsection 5.1.4, a heavy drinker is often found to be a smoker as well. As previously discussed in chapter 4, cheese is the major contributor to CHD deaths, we assume that cheese is affected both by smoking and alcohol consumption, and we have already shown through data modelling in subsubsection 4.3.3.2 that SBP is related to the other negative indicators.

Drawing from the Verhulst and Lotka-Volterra models (Murray, 2002), we can then define our coupled system of ODEs combining the variation in the three negative indicators cheese, alcohol and smoking usage with each other to analyse their mutual feedback augmentation or pacification:

$$\begin{cases} \dfrac{du}{dt} = \alpha_u + \beta_u u + \gamma_{uv} uv + \gamma_{uw} uw \\ \dfrac{dv}{dt} = \alpha_v + \beta_v v - \gamma_{uv} uv \\ \dfrac{dw}{dt} = \alpha_w + \beta_w w - \gamma_{uw} uw. \end{cases} \tag{70}$$

Here u stands for *alcohol consumption*, v for *cheese consumption* while w represents *smoking*. $\alpha_u, \alpha_v, \alpha_w, \beta_u, \beta_v, \beta_w$ are parameters that we estimate from statistical ana-

lysis and visualization results from previous chapters, where α_u represents the alcohol consumption at time $t = 0$, α_v is the cheese consumption at $t = 0$, α_w is the smoking population at $t = 0$. β_u represents the growth rate of alcohol consumption, that we estimated by statistical data modelling based linear regression as in chapter 2, and also the same as β_v, β_w which stand for cheese consumption and smoking population respectively; γ_{uv} represents the relative strength of affectation (measured through correlation) of alcohol on cheese or vice versa, and γ_{uw} stands for the correlation between alcohol and smoking. These two parameters are obtained from the analysis in chapter 4.

In each of the above coupled equations, the linear parts relate to time decaying trends whereas the coupled terms represent interactions between potential *predators* with *preys* that could inject an increasing trend in a decaying profile or vice versa; in other words, the coupled terms serve as competitors to the linear terms. The model has a time conserving symmetry in that the rate of change of all three variable together $\frac{d}{dt}(u + v + w)$ is devoid of the coupling terms and is intrinsically a linear dynamics as was shown in our data analysis detailed in previous chapters.

5.2.1 *Steady State Solutions*

First, we want to analyse the proposed model using *linear stability analysis* around the steady state $\left(\frac{du}{dt} = \frac{dv}{dt} = \frac{dw}{dt} = 0\right)$. This gives

$$u_0 = \frac{-\alpha_u}{\beta_u + av_0 + bw_0} \tag{71}$$

$$v_0 = \frac{-\alpha_v}{\beta_v - au_0} \tag{72}$$

$$w_0 = \frac{-\alpha_w}{\beta_w - bu_0}, \tag{73}$$

where $a = \gamma_{uv}$ and $b = \gamma_{uw}$ and $\{u_0, v_0, w_0\}$ define the steady-state values of the variables u, v and w respectively.

Using substitution, we can show that these steady-state values can be uniquely solved as a function of the system parameters from the following cubic equation:

$$A_1 {u_0}^3 + A_2 {u_0}^2 + A_3 u_0 + A_4 = 0, \tag{74}$$

where, $A_1 = -ab\beta_u$,

$A_2 = \beta_u(a\beta_w + b\beta_v) - ab(\alpha_u + \alpha_v + \alpha_w)$,

$A_3 = -\beta_u\beta_v\beta_w + \alpha_u\beta_w(a - \beta_v) + \alpha_v\beta_w a + \alpha_w\beta_v b$,

$A_4 = -\alpha_u\beta_v\beta_w$.

In other words, if u_0 is evaluated as a function of the parameters, both v_0 and w_0 can also be estimated from a knowledge of u_0.

5.2.2 *Linear Stability Analysis*

Perturbing the model around the steady-state, we get the following matrix M:

$$\begin{bmatrix} \frac{d\widetilde{u}}{dt} \\ \frac{d\widetilde{v}}{dt} \\ \frac{d\widetilde{w}}{dt} \end{bmatrix} = \begin{bmatrix} \beta_u + av_0 + bw_0 & au_0 & bu_0 \\ -av_0 & \beta_v - au_0 & 0 \\ -bw_0 & 0 & \beta_w - bu_0 \end{bmatrix} \begin{bmatrix} \widetilde{u} \\ \widetilde{v} \\ \widetilde{w} \end{bmatrix}$$

In order to ensure convergence, we want to evaluate the combinations of a and b for which the determinant of the stability matrix M is never equal to zero. The next step will then be to estimate the 3 eigenvalues and corresponding eigenvectors such that the absolute values of the eigenvectors as functions of parameters a and b are always positive definite to ensure convergence at all points in time.

From above, we have got $b = 0$, by putting back into the original model Equation 70, and solve it, we got the solutions of positive a.

In order to estimate the appropriate growth/decay pertaining to each variable, we need to adjust the right signs before each term used. For instance, in order to find out the trend line that shows an increasing trend instead of a decreasing one, the mutual signs between the constant and linear term of the respective variable should

be opposite to each other. Applying the same logic for 'u' and 'w' and making appropriate adjustments.

To making this model converge, we suppose by change the model sign for b and a individually in the five cases listed,

- Initial model
- Plus sign for both before uv
- (+ b, - a)
- (- b, + a)
- (- b, - a)

By testing the stability for these five cases, if there is a case stable which means converges, that's the one we are looking for. And if not, go for the next step.

All works based on this part are calculated by using Mathematica. And we have tried all cases with the data range from -10 to 10, and also -100 to 100 with the step size of 0.01. From our initial evaluation, we failed to converge to a real combination of b and a values that solve the $Det(M) = 0$. At this point, we have proved that there is no combination in this range that provides linear stable solutions to our model. When we try to put the (a, b) combination back to eigenvalues, a cubic equation is generated, that we then use to test all different combinations of eigenvalues that have negative real parts (if complex), or negative overall (if real).

- All $\lambda_1, \lambda_2, \lambda_3$ are negative real solutions or all equal to zero
- one of $\lambda_1, \lambda_2, \lambda_3$ is negative or zero, and the other two are complex solutions with all negative real parts, or else equal to zero.

Now, the problem converts to solving the cubic equation. To solve this equation, we first transform our cubic equation $k_1u^3 + k_2u^2 + k_3u + k_4 = 0$ to $x^3 + px + q = 0$ by defining $u = x - \frac{k_2}{3k_1}$, where $p = \frac{-k_2^2 + 3k_1k_3}{3k_1^2}$, $q = \frac{2k_2^3 + 27k_1^2k_4 - 9k_1k_2k_3}{27k_1^3}$.

- **$p > 0$**

  $g\prime(x) = 3x^2 + p \geqslant 0$

$g(x)$ monotonically increase in the interval $(-\infty, +\infty)$, and also $g(-\infty) = -\infty, g(\infty) = \infty$. This implies that $g(x)$ has only one real solution. We then have the following conditions:

1. when $q > 0$, there is one negative real solution;
2. when $q = 0$, there is one zero solution;
3. when $q < 0$, there is one positive real solution.

Conditions one and three satisfy our requirement of the convergence of the model.

- **p<0**

  Now we find the Maximum (represented by max) and Minimum (represented by min) solutions for the $p < 0$ case:

  - $\text{Maximum}(max) = g\left(-\sqrt{\frac{|\,p\,|}{3}}\right) = \frac{2\sqrt{3}}{9}\sqrt{|\,p\,|^3} + q.$
  - $\text{Minimum}(min) = g\left(\sqrt{\frac{|\,p\,|}{3}}\right) = -\frac{2\sqrt{3}}{9}\sqrt{|\,p\,|^3} + q.$

  This leads to the following conditions:

  1. when $q > 0$;
     - if $min > 0$, there is only one real negative solution
     - if $min = 0$, there is one negative and one positive real solution
     - if $min < 0$, there is one negative and two positive real solution
  2. when $q = 0$; there is one zero, one negative real solution and one positive real solution
  3. when $q < 0$;
     - if $max > 0$, there is two real negative solution and one positive real solution
     - if $max = 0$, there is one negative and one positive real solution
     - if $max < 0$, there is one positive real solution

  We may find there is only the first condition with $m > 0$ satisfies our requirement for keeping our model converge.

Summarise above, we have three conditions at which our model converges; these are:

(i) $p > 0$, $q > 0$, there is one negative real solution

(ii) $p > 0$, $q = 0$, there is one zero solution

(iii) $p < 0$, $q > 0$, if $m > 0$, there has only one real negative solution

We solved this problem, using conditions (i)-(iii) in Mathematica. No real solutions are available within the search domain, so we need to find complex solutions with real negative parts.

The relevant condition for the model convergence (with imaginary eigenvalues) is given below:

1. $\delta > 0, \left(\frac{k_2}{k_1}\right)\left(\frac{k_3}{k_1}\right) > \left(\frac{k_3}{k_1}\right) > 0 \Rightarrow u_1 < 0 \text{ and } \Re u_{2,3} < 0, \Im u_{2,3} \neq 0$
2. $\delta = 0, \left(\frac{k_2}{k_1}\right)\left(\frac{k_3}{k_1}\right) > \left(\frac{k_3}{k_1}\right) > 0 \Rightarrow u_1 < 0 \text{ and } u_2 = u_3 < 0$
3. $\delta < 0, \frac{k_2}{k_1} > 0, \frac{k_3}{k_1} > 0, \frac{k_3}{k_1} > 0 \Rightarrow u_1 < 0, u_2 < 0, u_3 < 0$

Where $\delta = \left(\frac{k_3}{k_1}\right)^2 - 4\left(\frac{k_2}{k_1}\right)\left(\frac{k_4}{k_1}\right)$

5.3 IMPROVED MODEL

The above analysis implies that some of the parameters need to have different signs; details below:

$$\begin{cases} \frac{du}{dt} = \alpha_u + \beta_u u - \gamma_{uv} uv + \gamma_{uw} uw \\ \frac{dv}{dt} = \alpha_v + \beta_v v + \gamma_{uv} uv \\ \frac{dw}{dt} = \alpha_w + \beta_w w - \gamma_{uw} uw. \end{cases} \tag{75}$$

Here u, v, w stand for *alcohol consumption*, *cheese consumption* and *smoking*, as before.

In the following, we are using the UK male data for validate our model by substitute the parameters into the system and implement by Matlab.

The purpose of this model is to predict the future time variation of the risk factors and the affecting in CHD death rates. In our case, the prediction starts in year 2013;

therefore, the raw data of each risk factor in 2013 are used as the initial values (values at $t = 0$). These are found to be 10.32 for alcohol consumption, 11.33 for cheese consumption and 22 for smoking population, which can be found in Table A.2.1. $\alpha_u, \alpha_v, \alpha_w$ are normalised by the initial value of each factor, where $\alpha_u = 0.3721, \alpha_v = 1.4106, \alpha_w = 1.2616$. $\beta_u, \beta_v, \beta_w$ are calculated from the gradient values shown in subsection 2.3.1: $\beta_u = 0.07400, \beta_v = 0.14488, \beta_w = -0.4496$. γ_{uv} is the parameter that represents the correlation between alcohol and cheese; this is calculated from chapter 4, where we developed a multivariate regression with cheese as a dependant variable. From the relationship between cheese and all the other parameters, we find alcohol, cereals, fruits and vegetables have strong influence on cheese in subsubsection 4.3.3.2, with $R^2 = 0.999$, standard error smaller then 0.0001, the equation of this linear model is: $cheese = -0.441 \times alcohol - 0.352 \times cereals - 0.352 \times fruits\&vegs + 0.875$, therefore, from this multivariate regression, the correlation between alcohol and cheese can be found as $\gamma_{uv} = -0.441$; The same as γ_{uw}, which stands for the correlation between alcohol and smoking, that also can be found in chapter 4, where we set smoking as the dependant variable. This multivariate regression model is given by $smoking = -0.422 \times alcohol - 0.349 \times cereals - 0.348 \times fruits\&vegs + 0.864$ that has the same $R^2 = 0.999$; this gives $\gamma_{uw} = -0.422$.

Now we have all the parameters ready to solve our model system for UK males. We do this on Matlab; a representative code is given below:

```
% Implements a prey-predator derived uvw  model
%
%
% Inputs:
%   t - Time variable: not used here because equation
%       is independent of time.
%   x - Independent variables: this contains three
%       populations (U, V, and W)
% Output:
%   dx - First derivative: the rate of change of the populations

function dx = uvw(t, x, Alpha1, Alpha2, Alpha3, Beta1, Beta2, Beta3, Gamma1,
    Gamma2)
dx =[0; 0; 0];
```

```
dx(1) = Alpha1 + Beta1 * x(1) - Gamma1 * x(1) * x(2) + Gamma2 * x(1) * x(3);
dx(2) = Alpha2 + Beta2 * x(2) + Gamma1 * x(1) * x(2);
dx(3) = Alpha3 + Beta3 * x(3) - Gamma2 * x(1) * x(3);
end
```

```
% Initial proposed "UVW Model" from "prey-predator Model"
clear all;
clc;

Alpha1 = 0.372065663; % normallised 2013 data of alcohol
Alpha2 = 1.410551654; % normallised 2013 data of cheese
Alpha3 = 1.261576592; % normallised 2013 data of smoking
Beta1 = 0.07400; % linear equation of alcohol (m value)
Beta2 = 0.14488; % linear equation of cheese (m value)
Beta3 = -0.4496; % linear equation of smoking (m value)
Gamma1 = -0.441; % corrlation from visalisation alcohol vs cheese
Gamma2 = -0.422; % corrlation from visalisation alcohol vs smoking
tspan = [0 20]; % Time span
IC = [10.32 11.33 22]; % Initial conditions of 'U', 'V', 'W'

options = odeset('RelTol', 1e-4, 'NonNegative', [1 2 3]);
% 1 - Relative error tolerance of 1e-4
% 2 - To set output to non-negative.
%     Since there are three populations, the array sets [1 2 3]

[t,x] = ode45(@(t,x) uvw(t, x, Alpha1, Alpha2, Alpha3, Beta1, Beta2, Beta3,
    Gamma1, Gamma2), tspan, IC, options);

plot(t,x)
hold on
xlabel('Time (t)');
ylabel('density')
legend('U', 'V', 'W');
```

The plots clearly show that in the next 20 years, our model predicts at least two cusps in all three variables concerned, implying sudden rise/decay in the relevant

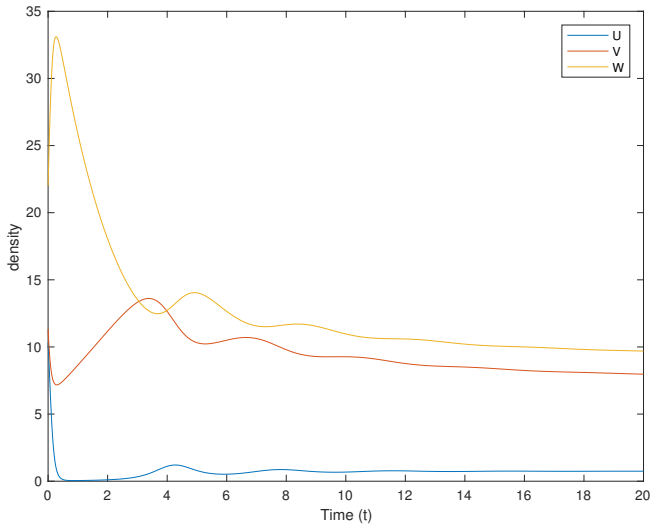


Figure 37: UVW model implemented by UK male

consumption. The first of this is expected within the next 1 to 4 years (starting from 2013, this implies a timeline of 2017 which could be tested; unfortunately, we do not have access to such recent data), there will be a very sharp decrease of smoking from about 35% to 12%, while cheese consumption is predicted to increase from 7% to 14%. This will be followed by a linear trend in smoking and cheese consumption, consistently within the 10-11% rate for smoking and around 8-10% for cheese consumption. For the alcohol consumption, a much larger fall in the intake is predicted in the first year, a result that is possibly skewed by our reliance on pure alcohol only in crunching the parameter values. This can be verified by considering all forms of alcohol that we hope to do soon. On average, alcohol consumption and smoking are expected to go down over the years while cheese consumption is likely to increase.

The next major step in this research will be to structure an 'utility function' C that will represent the equivalent of the Framingham scoring system; in other words, as u and w decrease over time with v increasing, the rate of decrease/increase in C will be proportional to the CHD death ratio. This will be pursued in a future work.

In the next chapter, we summarise this work and findings, together with a future research plan.

6

CONCLUSIONS

This thesis focuses on modelling the impact of life-style parameters on atherosclerosis, and in turn Coronary Heart Disease (CHD) for 13 European countries, principally based on data modelling and data visualization. The parameters we have chosen in this work are alcohol consumption, cheese consumption, smoking habit, high blood pressure, cereals consumption and fruits and vegetables consumption, and the 13 countries we are studying are UK, France, Greece, Italy, Spain, Denmark, Finland, Iceland, Norway, Sweden, Germany, Netherlands and Switzerland, emphasis on the UK. The real data used in this study spans the years between 1990 and 2013. Of the factors used, the first four refer to the negative indicators, as is known to all medical practitioners, while the latter two are the so called positive indicators. While the qualitative impact of these factors has long been known, this thesis, for the first time, establishes a clear numerical relation between all affecting factors. Our finding also suggests how much of a change in the positive indicators intake could reduce the impact coming from the negative indicators. The key repercussion of all of these will be in atherosclerosis afflicted CHD mortality rates, an aspect of our future study.

6.1 THESIS SUMMARY

Atherosclerosis is a low-density cholesterol promoted medical condition in which the walls of the artery thicken due to the plaques, and after medical aggravations, this problem becomes to escalate to CHD and CVD. The biological problem which motivating our work is introduced in chapter 1, CVD and CHD are the major cause of death in most of the European countries, this problem is even pronounced in European countries compared with the rest part of the world. although this is a key medical problem, not enough has been done in connection with prognosis directed theoretical analysis. The main contribution of this thesis quantifies the importance

of the health-sustaining factors and the interaction models building up between each of the factors.

Firstly, linear regression applied on to CHD death rate for all 13 European countries by statistical method of least square in chapter 2, 13 European countries all show a negative linear trend with time for both males and females, this generally represents a growth of public awareness of health and the hazard of CHD. Males in both countries face higher CHD death rate than females, there is an order ranked by the CHD death rate for each block: ScEU > UK > WeEU > MeEU. section 2.3 analysis the 6 life-style parameters using the same statistical method as for CHD death rate, almost negative indicators show a decreasing trend, except cheese consumption and few countries of alcohol consumptions, and except Greece, Italy, Spain, Iceland and Switzerland, all other countries showing the increase trend of the two positive indicators.

Next, data visualization methods are introduced in **??**, like PCA, NSC, GTM and GPLVM, by applying three visualization quality evaluation measures (i.e. trustworthiness, continuity, and mean relative rank errors) introduced from subsection 3.5.1 to subsection 3.5.3, PCA and NSC tested to be the better visualization method for this study, and due to the nature of datasets are all linear trend, PCA is chosen to be the best of the data visualization method in final. In chapter 4, three sets (i.e. raw real datasets, pure synthetic datasets and real-synthetic datasets)of UK databases are trained by PCA (also compared results from NSC training), and features weighting are estimated by using PCA, the ranking of 4 negative indicators found on real dataset are Smoking > Cheese > SBP > Alcohol for males, BP > Smoking > Cheese > Alcohol for females. And the ranking based on synthetic data are Cheese > Smoking > SBP > Alcohol for males and Cheese > SBP > Smoking > Alcohol for females. Cheese to be the most negative indicators after a following test on 21 sets of new generated datasets by increasing the positive parameters by a certain percentage on real-synthetic datasets after year 2013, which listed in subsubsection 4.3.3.1, and the features weighting based on 21 sets of new generated data are summarised in Table 31 to Table 34. The prediction models are built in subsubsection 4.3.3.2 using the multivariate regression by SPSS after analysing the correlation between each of the parameters in scatter matrix.

Last, based on previous analysis, we construct a continuum model, which is a time evolution of our risk factors: alcohol consumption, cheese consumption and smoking population in chapter 5. This model is constructed by combining the predator-prey and Lotka-Volterra models. section 5.3 describes our final model system that is studied using UK male data. Our results make some quantitative predictions that can be verified against real data, that unfortunately, we do not presently have access to.

6.2 FUTURE PLAN

Following are some key research plans for the immediate future:

1. Incorporate the nonlinear trends that have been presently overlooked. This will require more extensive applications of machine learning on data modelling, that then will fine tune the predictions from the continuum model. At this level, we hope to collaborate with NHS or equivalent agency to make more accurate predictions based on timelined, documented data.

2. Develop a more robust nonlinear equivalent of the Framingham scoring system, combining data with continuum modelling, and embedding predictive powers in the process.

3. Extend the nonlinear scoring model, defined above, to incorporate subjectively defined attributes on a patient-by-patient basis.

APPENDIX

A

DATASETS

Datasets used in this thesis which consist of two parts:

In section A.1, CHD death rate which calculated by the division of CHD death rate with all causes death rate are listed from Table A.1.1 to Table A.1.13.

In section A.2, the datasets of 6 life-style parameters are in direct use from the open data source WHO and FAO, and list from Table A.2.1 to Table A.2.13. Smoking and Systolic blood pressure have missing values and we obtain the raw real datasets of UK CHD death rate linear model fitting on the known values. The linear fitting formula is then used to predict the unknown values. The missing values are listed by red colour.

A.1 CHD DEATH RATE DATASETS

Table A.1.1: Raw real datasets of UK CHD death rate

United Kingdom **SDR,Coronary Heart Disease(CHD) Death Rate, by 100 000 inhabitants**						
	All Causes Death Rate		*CHD Death Rate*		*CHD/All Causes Death Rate*	
Year	*Male*	*Female*	*Male*	*Female*	*Male*	*Female*
1990	1055.87	659.77	309.09	145.41	0.2927	0.2204
1991	1042.74	656.32	304.95	145.77	0.2925	0.2221
1992	1009.20	637.20	292.93	140.12	0.2903	0.2199
1993	1028.00	654.59	290.70	138.78	0.2828	0.2120
1994	972.35	621.54	266.46	127.29	0.2740	0.2048
1995	979.75	629.82	259.90	123.07	0.2653	0.1954
1996	952.13	616.57	247.89	117.22	0.2604	0.1901
1997	925.10	608.79	232.72	110.77	0.2516	0.1820
1998	912.56	603.04	225.84	108.31	0.2475	0.1796
1999	902.49	602.80	214.66	101.74	0.2379	0.1688
2000	858.60	574.90	199.92	94.38	0.2328	0.1642
2001	838.63	565.76	191.15	90.56	0.2279	0.1601
2002	828.18	562.92	182.31	87.04	0.2201	0.1546
2003	817.66	567.18	173.97	83.46	0.2128	0.1471
2004	774.41	536.81	160.62	75.99	0.2074	0.1416
2005	752.40	527.44	150.44	70.79	0.1999	0.1342
2006	727.20	508.57	138.20	65.10	0.1900	0.1280
2007	711.61	501.79	131.77	61.05	0.1852	0.1217
2008	700.40	499.39	124.07	58.01	0.1771	0.1162
2009	670.69	472.64	115.60	52.22	0.1724	0.1105
2010	654.69	467.36	111.12	49.45	0.1697	0.1058
2011	630.56	451.55	100.88	44.37	0.1600	0.0983
2012	631.45	460.36	97.91	44.05	0.1551	0.0957
2013	630.83	457.16	96.48	42.53	0.1529	0.0930

Table A.1.2: Raw real datasets of Denmark CHD death rate

Denmark **SDR,Coronary Heart Disease(CHD) Death Rate, by 100 000 inhabitants**						
	All Causes Death Rate		*CHD Death Rate*		*CHD/All Causes Death Rate*	
Year	*Male*	*Female*	*Male*	*Female*	*Male*	*Female*
1990	1106.40	699.02	293.13	148.04	0.2649	0.2118
1991	1061.07	683.86	276.82	135.76	0.2609	0.1985
1992	1071.65	691.23	270.74	130	0.2526	0.1881
1993	1084.85	710.60	260.05	134.59	0.2397	0.1894
1994	1062.82	688.99	232.54	114.33	0.2188	0.1659
1995	1072.97	706.97	226.66	116.11	0.2112	0.1642
1996	1036.90	678.99	199.86	99.06	0.1927	0.1459
1997	999.26	667.02	191.46	95.16	0.1916	0.1427
1998	970.14	640.91	178.57	86.78	0.1841	0.1354
1999	955.51	660.46	168.58	87.28	0.1764	0.1322
2000	918.60	626.46	154.02	78.91	0.1677	0.1260
2001	909.60	627.60	154.83	79.38	0.1702	0.1265
2002	910.43	630.76	134.38	70.95	0.1476	0.1125
2003	898.71	608.40	127.24	66.3	0.1416	0.1090
2004	867.40	585.18	118.61	58.64	0.1367	0.1002
2005	830.53	570.37	107.36	54.63	0.1293	0.0958
2006	826.76	566.79	97.70	51.71	0.1182	0.0912
2007	807.79	564.82	91.70	47.73	0.1135	0.0845
2008	787.00	544.33	85.68	43.53	0.1089	0.0800
2009	772.74	541.11	83.60	41.43	0.1082	0.0766
2010	752.98	527.36	75.56	39.24	0.1003	0.0744
2011	714.73	501.31	67.12	32.59	0.0939	0.0650
2012	695.35	491.38	64.23	31.74	0.0924	0.0646
2013	443.17	365.78	30.90	17.36	0.0697	0.0475

Table A.1.3: Raw real datasets of France CHD death rate

France **SDR,Coronary Heart Disease(CHD) Death Rate, by 100 000 inhabitants**						
	All Causes Death Rate		*CHD Death Rate*		*CHD/All Causes Death Rate*	
Year	*Male*	*Female*	*Male*	*Female*	*Male*	*Female*
1990	944.92	493.15	91.20	39.49	0.0965	0.0801
1991	930.19	484.25	90.68	38.53	0.0975	0.0796
1992	912.70	472.46	87.07	36.94	0.0954	0.0782
1993	909.22	474.91	86.14	36.61	0.0947	0.0771
1994	877.86	454.40	81.33	33.77	0.0926	0.0743
1995	872.51	455.28	80.53	33.07	0.0923	0.0726
1996	864.58	452.44	80.02	32.56	0.0926	0.0720
1997	838.22	442.95	75.95	30.80	0.0906	0.0695
1998	862.47	460.81	78.15	32.21	0.0906	0.0699
1999	854.61	456.68	76.04	30.76	0.0890	0.0674
2000	832.20	444.37	75.65	29.80	0.0909	0.0671
2001	821.26	442.81	72.17	29.55	0.0879	0.0667
2002	812.03	444.69	69.88	28.29	0.0861	0.0636
2003	815.42	457.09	67.82	28.35	0.0832	0.0620
2004	751.55	413.21	63.78	25.25	0.0849	0.0611
2005	751.25	415.47	61.68	24.40	0.0821	0.0587
2006	715.69	391.64	57.11	22.26	0.0798	0.0568
2007	700.20	380.47	54.83	20.91	0.0783	0.0550
2008	689.67	381.81	52.52	20.07	0.0762	0.0526
2009	677.76	375.41	49.70	18.81	0.0733	0.0501
2010	667.34	368.61	47.28	17.51	0.0708	0.0475
2011	645.18	356.73	44.98	16.40	0.0697	0.0460
2012	643.00	362.26	43.82	16.11	0.0681	0.0445
2013	627.09	354.93	41.81	15.20	0.0667	0.0428

Table A.1.4: Raw real datasets of Finland CHD death rate

Finland **SDR,Coronary Heart Disease(CHD) Death Rate, by 100 000 inhabitants**						
	All Causes Death Rate		*CHD Death Rate*		*CHD/All Causes Death Rate*	
Year	*Male*	*Female*	*Male*	*Female*	*Male*	*Female*
1990	1186.05	646.48	359.35	158.08	0.3030	0.2445
1991	1140.62	621.70	339.50	152.27	0.2976	0.2449
1992	1130.47	615.52	346.10	149.75	0.3062	0.2433
1993	1112.91	628.84	326.40	153.76	0.2933	0.2445
1994	1030.98	577.26	299.47	141.78	0.2905	0.2456
1995	1046.72	578.47	304.25	140.68	0.2907	0.2432
1996	1024.93	560.90	285.07	128.74	0.2781	0.2295
1997	994.09	560.29	270.28	122.70	0.2719	0.2190
1998	993.12	540.52	266.80	124.64	0.2686	0.2306
1999	975.56	532.72	265.17	121.60	0.2718	0.2283
2000	941.29	532.64	254.88	120.28	0.2708	0.2258
2001	908.68	510.78	236.50	113.97	0.2603	0.2231
2002	895.06	513.68	234.00	115.38	0.2614	0.2246
2003	873.93	499.50	221.80	108.26	0.2538	0.2167
2004	849.56	475.95	211.21	97.05	0.2486	0.2039
2005	826.31	463.92	203.24	96.89	0.2460	0.2089
2006	812.72	446.47	200.51	90.43	0.2467	0.2025
2007	803.88	444.61	193.09	90.60	0.2402	0.2038
2008	771.18	439.29	182.45	87.90	0.2366	0.2001
2009	768.29	431.92	179.32	80.06	0.2334	0.1854
2010	754.68	429.29	176.65	79.22	0.2341	0.1845
2011	726.71	416.38	165.49	71.39	0.2277	0.1715
2012	712.12	421.80	155.94	71.21	0.2190	0.1688
2013	691.39	408.01	143.93	65.10	0.2082	0.1596

Table A.1.5: Raw real datasets of Germany CHD death rate

Germany **SDR,Coronary Heart Disease(CHD) Death Rate, by 100 000 inhabitants**						
	All Causes Death Rate		*CHD Death Rate*		*CHD/All Causes Death Rate*	
Year	*Male*	*Female*	*Male*	*Female*	*Male*	*Female*
1990	1119.22	670.13	225.41	106.96	0.2014	0.1596
1991	1100.00	654.03	231.36	112.04	0.2103	0.1713
1992	1058.99	627.08	225.96	110.41	0.2134	0.1761
1993	1060.46	626.90	228.98	111.68	0.2159	0.1781
1994	1028.49	609.05	218.85	108.46	0.2128	0.1781
1995	1012.55	595.52	216.65	107.95	0.2140	0.1813
1996	989.40	587.46	208.67	105.50	0.2109	0.1796
1997	951.81	566.10	201.31	103.16	0.2115	0.1822
1998	926.85	556.22	197.22	102.50	0.2128	0.1843
1999	902.13	544.87	188.84	98.63	0.2093	0.1810
2000	876.98	530.16	176.71	92.33	0.2015	0.1742
2001	845.89	518.47	170.24	89.33	0.2013	0.1723
2002	841.06	525.07	165.82	89.30	0.1972	0.1701
2003	840.88	529.44	161.88	88.13	0.1925	0.1665
2004	790.64	500.80	149.21	80.57	0.1887	0.1609
2005	776.25	495.69	141.49	75.60	0.1823	0.1525
2006	744.29	476.23	133.10	70.69	0.1788	0.1484
2007	730.71	465.87	126.23	66.34	0.1727	0.1424
2008	720.52	467.32	117.41	61.82	0.1630	0.1323
2009	712.17	461.51	115.92	59.47	0.1628	0.1289
2010	697.07	453.44	110.95	56.77	0.1592	0.1252
2011	674.06	442.47	103.02	52.50	0.1528	0.1187
2012	666.32	440.28	101.72	51.31	0.1527	0.1165
2013	698.12	453.07	105.02	51.45	0.1504	0.1136

Table A.1.6: Raw real datasets of Greece CHD death rate

Greece **SDR,Coronary Heart Disease(CHD) Death Rate, by 100 000 inhabitants**						
	All Causes Death Rate		*CHD Death Rate*		*CHD/All Causes Death Rate*	
Year	*Male*	*Female*	*Male*	*Female*	*Male*	*Female*
1990	886.73	612.37	135.91	60.60	0.1533	0.0990
1991	883.34	604.47	132.46	57.20	0.1500	0.0946
1992	889.58	609.54	134.02	59.37	0.1507	0.0974
1993	866.53	592.47	129.39	57.08	0.1493	0.0963
1994	859.09	583.91	124.68	55.76	0.1451	0.0955
1995	871.69	585.42	129.99	58.34	0.1491	0.0997
1996	865.69	579.38	132.87	58.54	0.1535	0.1010
1997	842.17	567.10	131.04	58.35	0.1556	0.1029
1998	849.62	581.07	122.14	53.74	0.1438	0.0925
1999	847.23	574.12	125.35	55.34	0.1480	0.0964
2000	855.46	572.18	123.62	55.61	0.1445	0.0972
2001	820.29	554.26	124.42	55.56	0.1517	0.1002
2002	806.25	557.26	121.26	55.54	0.1504	0.0997
2003	795.98	562.47	126.36	58.79	0.1587	0.1045
2004	784.91	553.61	124.06	56.84	0.1581	0.1027
2005	762.57	532.90	112.14	49.06	0.1471	0.0921
2006	728.45	515.75	107.56	48.37	0.1477	0.0938
2007	739.12	525.20	105.60	46.52	0.1429	0.0886
2008	705.75	496.13	96.29	41.32	0.1364	0.0833
2009	693.18	473.26	96.63	41.13	0.1394	0.0869
2010	662.68	464.15	89.11	37.73	0.1345	0.0813
2011	658.74	449.68	89.08	36.92	0.1352	0.0821
2012	689.16	433.44	92.31	35.29	0.1339	0.0814
2013	664.96	457.02	92.55	39.55	0.1392	0.0865

Table A.1.7: Raw real datasets of Iceland CHD death rate

Iceland **SDR,Coronary Heart Disease(CHD) Death Rate, by 100 000 inhabitants**						
	All Causes Death Rate		*CHD Death Rate*		*CHD/All Causes Death Rate*	
Year	*Male*	*Female*	*Male*	*Female*	*Male*	*Female*
1990	815.66	554.83	232.93	118.28	0.2856	0.2132
1991	893.17	523.77	266.37	119.77	0.2982	0.2287
1992	772.63	554.92	246.01	119.01	0.3184	0.2145
1993	748.90	555.88	232.63	115.81	0.3106	0.2083
1994	762.35	525.38	240.19	112.01	0.3151	0.2132
1995	820.22	578.60	224.99	96.71	0.2743	0.1671
1996	801.93	543.49	216.37	94.09	0.2698	0.1731
1997	799.54	517.06	208.58	85.55	0.2609	0.1655
1998	740.79	510.76	166.39	92.79	0.2246	0.1817
1999	752.31	528.90	200.06	98.15	0.2659	0.1856
2000	690.11	524.42	165.88	89.83	0.2404	0.1713
2001	681.24	445.50	159.95	70.20	0.2348	0.1576
2002	669.49	471.23	170.86	77.50	0.2552	0.1645
2003	632.58	475.96	166.96	70.66	0.2639	0.1485
2004	657.54	444.54	170.83	65.83	0.2598	0.1481
2005	623.82	435.36	127.88	62.85	0.2050	0.1444
2006	620.58	459.45	129.16	68.28	0.2081	0.1486
2007	614.66	442.85	138.10	48.78	0.2247	0.1102
2008	595.07	448.89	133.80	60.54	0.2248	0.1349
2009	603.57	422.50	117.48	54.60	0.1946	0.1292
2010	542.40	394.32	108.70	46.60	0.2004	0.1182
2011	519.81	378.79	101.39	42.85	0.1951	0.1131
2012	495.95	361.81	94.08	39.10	0.1897	0.1081
2013	470.70	343.17	86.78	35.35	0.1844	0.1030

Table A.1.8: Raw real datasets of Italy CHD death rate

Italy **SDR,Coronary Heart Disease(CHD) Death Rate, by 100 000 inhabitants**						
	All Causes Death Rate		*CHD Death Rate*		*CHD/All Causes Death Rate*	
Year	*Male*	*Female*	*Male*	*Female*	*Male*	*Female*
1990	967.61	564.40	134.96	63.09	0.1395	0.1118
1991	964.31	561.18	136.37	62.79	0.1414	0.1119
1992	935.85	551.67	130.80	61.39	0.1398	0.1113
1993	920.69	543.45	129.91	61.48	0.1411	0.1131
1994	907.93	534.38	127.01	60.69	0.1399	0.1136
1995	891.68	521.40	126.71	60.93	0.1421	0.1169
1996	860.82	505.90	122.16	58.19	0.1419	0.1150
1997	845.25	498.35	118.45	57.39	0.1401	0.1152
1998	842.77	497.48	119.26	57.89	0.1415	0.1164
1999	815.25	480.96	112.10	55.19	0.1375	0.1147
2000	786.17	466.82	106.00	52.23	0.1348	0.1119
2001	762.78	451.20	101.93	49.74	0.1336	0.1102
2002	745.45	441.60	102.47	49.82	0.1375	0.1128
2003	757.98	458.88	106.07	54.71	0.1399	0.1192
2004	726.61	436.10	96.83	47.75	0.1333	0.1095
2005	708.41	426.82	93.84	46.48	0.1325	0.1089
2006	668.88	401.88	89.05	45.16	0.1331	0.1124
2007	657.58	401.07	85.55	44.13	0.1301	0.1100
2008	644.67	394.91	83.96	42.23	0.1302	0.1069
2009	632.26	391.25	80.05	40.16	0.1266	0.1026
2010	610.88	376.64	76.93	37.57	0.1259	0.0998
2011	613.13	381.40	77.20	38.70	0.1259	0.1015
2012	610.80	384.72	75.20	37.65	0.1231	0.0979
2013	554.49	348.88	69.90	36.38	0.1261	0.1043

Table A.1.9: Raw real datasets of Netherlands CHD death rate

Netherlands SDR,Coronary Heart Disease(CHD) Death Rate, by 100 000 inhabitants						
	All Causes Death Rate		*CHD Death Rate*		*CHD/All Causes Death Rate*	
Year	*Male*	*Female*	*Male*	*Female*	*Male*	*Female*
1990	1001.37	571.56	193.98	82.11	0.1937	0.1437
1991	988.23	570.34	185.16	79.90	0.1874	0.1401
1992	967.36	562.96	175.05	76.36	0.1810	0.1356
1993	1007.93	585.95	181.19	78.34	0.1798	0.1337
1994	952.65	566.87	165.69	73.37	0.1739	0.1294
1995	954.52	564.80	164.37	71.06	0.1722	0.1258
1996	950.43	565.60	157.95	71.10	0.1662	0.1257
1997	911.08	556.54	147.98	64.35	0.1624	0.1156
1998	910.04	552.52	142.81	63.42	0.1569	0.1148
1999	903.03	564.13	134.73	59.34	0.1492	0.1052
2000	883.52	556.83	125.44	55.85	0.1420	0.1003
2001	860.72	551.34	116.97	52.30	0.1359	0.0949
2002	852.07	553.99	109.17	50.46	0.1281	0.0911
2003	837.24	544.01	106.08	46.87	0.1267	0.0862
2004	786.93	516.83	92.87	42.58	0.1180	0.0824
2005	767.37	506.49	86.76	38.90	0.1131	0.0768
2006	734.92	494.08	79.67	35.06	0.1084	0.0710
2007	708.11	472.38	72.91	33.29	0.1030	0.0705
2008	688.11	474.69	67.66	30.95	0.0983	0.0652
2009	671.56	455.95	62.60	27.59	0.0932	0.0605
2010	658.71	454.34	59.15	26.59	0.0898	0.0585
2011	630.95	448.88	54.59	24.53	0.0865	0.0546
2012	636.83	452.91	52.57	23.24	0.0825	0.0513
2013	622.28	443.49	47.76	21.95	0.0768	0.0495

Table A.1.10: Raw real datasets of Norway CHD death rate

Norway **SDR,Coronary Heart Disease(CHD) Death Rate, by 100 000 inhabitants**						
	All Causes Death Rate		*CHD Death Rate*		*CHD/All Causes Death Rate*	
Year	*Male*	*Female*	*Male*	*Female*	*Male*	*Female*
1990	1009.71	592.92	277.80	115.64	0.2751	0.1950
1991	967.10	574.01	260.31	111.34	0.2692	0.1940
1992	958.00	564.11	249.81	106.02	0.2608	0.1879
1993	978.34	579.97	236.56	103.39	0.2418	0.1783
1994	912.46	551.07	220.90	95.73	0.2421	0.1737
1995	929.71	549.92	223.65	93.63	0.2406	0.1703
1996	880.49	534.74	201.70	85.26	0.2291	0.1594
1997	878.45	538.81	198.78	85.70	0.2263	0.1591
1998	863.39	524.01	191.14	82.03	0.2214	0.1565
1999	864.62	532.06	180.92	83.58	0.2092	0.1571
2000	827.13	518.71	164.42	78.94	0.1988	0.1522
2001	814.62	511.57	159.88	73.00	0.1963	0.1427
2002	808.44	514.46	154.61	71.25	0.1912	0.1385
2003	761.24	491.62	138.06	66.32	0.1814	0.1349
2004	727.37	470.99	126.19	59.87	0.1735	0.1271
2005	713.88	461.48	114.86	51.81	0.1609	0.1123
2006	686.06	459.17	103.73	52.73	0.1512	0.1148
2007	690.69	457.66	103.29	49.75	0.1495	0.1087
2008	683.18	445.67	99.16	46.30	0.1451	0.1039
2009	657.05	440.70	92.35	44.85	0.1406	0.1018
2010	646.42	438.73	88.00	41.50	0.1361	0.0946
2011	635.93	427.18	82.48	39.80	0.1297	0.0932
2012	620.72	434.37	77.32	39.82	0.1246	0.0917
2013	606.30	422.26	71.59	33.87	0.1181	0.0802

Table A.1.11: Raw real datasets of Spain CHD death rate

Spain **SDR,Coronary Heart Disease(CHD) Death Rate, by 100 000 inhabitants**						
	All Causes Death Rate		*CHD Death Rate*		*CHD/All Causes Death Rate*	
Year	*Male*	*Female*	*Male*	*Female*	*Male*	*Female*
1990	961.28	556.87	106.15	48.00	0.1104	0.0862
1991	955.30	550.02	108.42	48.44	0.1135	0.0881
1992	922.78	517.31	106.26	46.67	0.1152	0.0902
1993	913.05	516.18	105.19	46.95	0.1152	0.0910
1994	892.96	500.74	101.78	45.26	0.1140	0.0904
1995	896.77	497.89	103.36	46.46	0.1153	0.0933
1996	892.10	492.23	105.20	46.54	0.1179	0.0945
1997	865.86	478.66	103.12	45.67	0.1191	0.0954
1998	872.23	480.72	103.86	45.51	0.1191	0.0947
1999	870.18	480.74	102.10	44.70	0.1173	0.0930
2000	812.13	449.16	95.49	40.98	0.1176	0.0912
2001	795.05	437.60	91.68	39.53	0.1153	0.0903
2002	787.36	435.92	89.50	39.26	0.1137	0.0901
2003	791.66	445.36	89.65	39.00	0.1132	0.0876
2004	749.96	415.76	83.09	36.57	0.1108	0.0880
2005	751.47	419.52	81.72	35.70	0.1087	0.0851
2006	704.15	391.69	76.11	32.14	0.1081	0.0821
2007	705.29	393.95	73.98	31.36	0.1049	0.0796
2008	681.05	387.14	69.32	29.47	0.1018	0.0761
2009	660.97	374.66	66.88	27.79	0.1012	0.0742
2010	641.25	361.06	64.83	26.77	0.1011	0.0741
2011	630.84	359.40	62.02	25.68	0.0983	0.0715
2012	624.65	356.64	60.00	24.40	0.0961	0.0684
2013	591.78	339.31	57.23	22.54	0.0967	0.0664

Table A.1.12: Raw real datasets of Sweden CHD death rate

Sweden **SDR,Coronary Heart Disease(CHD) Death Rate, by 100 000 inhabitants**						
	All Causes Death Rate		*CHD Death Rate*		*CHD/All Causes Death Rate*	
Year	*Male*	*Female*	*Male*	*Female*	*Male*	*Female*
1990	912.97	560.24	269.96	120.25	0.2957	0.2146
1991	902.27	549.91	260.69	116.65	0.2889	0.2121
1992	874.89	539.09	248.52	115.33	0.2841	0.2139
1993	875.76	545.68	246.88	113.94	0.2819	0.2088
1994	821.69	510.57	229.21	102.38	0.2789	0.2005
1995	826.03	511.24	230.34	102.42	0.2789	0.2003
1996	806.37	509.02	215.05	98.40	0.2667	0.1933
1997	794.60	497.17	202.33	92.90	0.2546	0.1869
1998	787.65	493.45	195.81	90.10	0.2486	0.1826
1999	777.81	499.97	189.33	87.35	0.2434	0.1747
2000	751.87	491.71	176.33	83.33	0.2345	0.1695
2001	739.35	490.59	170.59	82.14	0.2307	0.1674
2002	734.06	492.41	166.68	80.30	0.2271	0.1631
2003	718.35	475.30	160.65	76.60	0.2236	0.1612
2004	693.49	468.86	146.83	71.49	0.2117	0.1525
2005	687.85	459.84	144.08	67.35	0.2095	0.1465
2006	666.88	451.86	136.86	68.25	0.2052	0.1510
2007	652.30	451.10	129.50	63.72	0.1985	0.1413
2008	642.93	443.55	124.70	62.14	0.1940	0.1401
2009	627.63	433.61	116.48	57.54	0.1856	0.1327
2010	619.00	428.13	111.45	54.43	0.1800	0.1271
2011	604.22	420.47	102.96	51.67	0.1704	0.1229
2012	602.90	429.36	101.40	49.49	0.1682	0.1153
2013	587.40	420.74	94.61	46.18	0.1611	0.1098

Table A.1.13: Raw real datasets of Switzerland CHD death rate

Switzerland **SDR,Coronary Heart Disease(CHD) Death Rate, by 100 000 inhabitants**						
	All Causes Death Rate		*CHD Death Rate*		*CHD/All Causes Death Rate*	
Year	*Male*	*Female*	*Male*	*Female*	*Male*	*Female*
1990	943.54	534.87	162.63	70.84	0.1724	0.1324
1991	913.57	512.47	160.85	67.71	0.1761	0.1321
1992	888.19	502.40	155.87	69.92	0.1755	0.1392
1993	870.87	498.57	152.88	68.37	0.1755	0.1371
1994	847.50	486.03	144.81	67.87	0.1709	0.1396
1995	846.61	489.82	156.64	71.32	0.1850	0.1456
1996	809.52	479.37	145.84	70.86	0.1802	0.1478
1997	796.77	478.90	145.72	69.90	0.1829	0.1460
1998	791.75	460.07	144.52	69.36	0.1825	0.1508
1999	763.39	459.91	134.07	68.09	0.1756	0.1481
2000	748.99	456.69	128.91	64.79	0.1721	0.1419
2001	723.94	439.22	119.05	60.62	0.1644	0.1380
2002	692.53	435.46	112.03	58.39	0.1618	0.1341
2003	697.64	440.57	111.90	56.93	0.1604	0.1292
2004	655.57	415.51	100.96	51.31	0.1540	0.1235
2005	653.54	408.35	102.97	49.96	0.1576	0.1223
2006	624.95	399.42	97.97	47.45	0.1568	0.1188
2007	614.89	394.10	93.42	45.42	0.1519	0.1152
2008	596.04	386.18	88.92	43.44	0.1492	0.1125
2009	590.34	385.16	84.91	40.52	0.1438	0.1052
2010	576.74	376.43	80.42	38.41	0.1394	0.1020
2011	565.34	369.98	75.18	34.21	0.1330	0.0925
2012	560.79	375.74	74.57	35.32	0.1330	0.0940
2013	554.94	370.68	69.88	33.35	0.1259	0.0900

A.2 6 PARAMETERS DATASETS

Table A.2.1: UK raw real datasets of 6 life-style parameters

United Kingdom - Databases of 6 parameters								
	Alcohol Consumed	*Cheese Consumed*	*Smoking Habit*		*Mean Systolic Blood Pressure*		*Cereal Supply*	*Fruit and Veg Supply*
Year			Males	Females	Males	Females		
1990	9.96	7.74	31.00	29.00	133.20	127.30	92.97	164.43
1991	10.01	8.31	*30.90*	*29.06*	133.40	127.50	93.02	163.83
1992	9.65	8.46	29.00	28.00	133.50	127.60	95.25	168.64
1993	9.45	8.29	*29.98*	*28.07*	133.60	127.60	92.88	168.47
1994	9.41	8.34	28.00	26.00	133.60	127.50	92.61	162.46
1995	9.70	8.32	29.00	26.00	133.50	127.40	93.83	155.80
1996	9.75	9.78	29.00	28.00	133.40	127.20	96.18	162.69
1997	9.97	10.08	*28.16*	*26.09*	133.20	126.90	101.09	164.03
1998	10.14	9.54	28.00	26.00	133.00	126.50	107.39	178.02
1999	10.16	9.83	*27.25*	*25.10*	132.70	126.10	107.56	178.71
2000	10.59	9.21	29.00	25.00	132.40	125.70	107.80	171.82
2001	10.91	10.05	28.00	26.00	132.10	125.20	104.57	182.79
2002	11.44	9.46	27.00	25.00	131.70	124.60	114.47	183.63
2003	11.85	9.63	28.00	24.00	131.30	124.10	114.39	207.90
2004	12.22	9.96	26.00	23.00	130.90	123.50	111.92	206.49
2005	12.05	10.58	25.00	23.00	130.40	122.90	113.30	222.90
2006	11.61	10.84	23.00	21.00	130.00	122.30	112.34	232.44
2007	11.84	11.03	22.00	20.00	129.60	121.80	113.59	218.51
2008	11.47	10.93	22.00	21.00	129.20	121.20	115.79	227.20
2009	10.79	10.79	22.00	20.00	128.70	120.60	114.53	213.78
2010	10.88	11.09	21.00	20.00	128.20	120.10	114.95	216.12
2011	10.68	10.73	21.00	19.00	127.80	119.50	114.21	222.64
2012	10.42	11.17	22.00	19.00	127.30	118.90	114.89	218.91
2013	10.32	11.33	22.00	17.00	126.90	118.40	115.85	224.40

Table A.2.2: Denmark raw real datasets of 6 life-style parameters

Denmark - Databases of 6 parameters								
	Alcohol Consumed	*Cheese Consumed*	*Smoking Habit*		*Mean Systolic Blood Pressure*		*Cereal Supply*	*Fruit and Veg Supply*
Year			Males	Females	Males	Females		
1990	11.85	14.65	*44.07*	*36.78*	132.70	125.20	92.75	146.88
1991	11.76	12.63	*42.92*	*35.90*	132.70	125.10	95.17	149.20
1992	11.94	15.49	*41.77*	*35.03*	132.70	125.00	97.81	155.89
1993	11.89	15.90	*40.63*	*34.15*	132.80	124.90	102.17	142.10
1994	12.14	14.16	39.00	35.00	132.80	124.70	103.42	138.80
1995	12.16	17.24	38.00	33.00	132.90	124.60	109.49	139.11
1996	12.27	17.11	36.00	32.00	132.90	124.40	105.46	149.74
1997	12.20	13.80	34.00	30.00	132.80	124.20	113.46	178.71
1998	11.69	14.07	34.00	31.00	132.70	124.00	117.30	157.17
1999	11.62	14.45	35.00	27.00	132.60	123.70	115.23	187.17
2000	11.69	16.34	32.00	29.00	132.40	123.50	107.25	192.50
2001	11.56	21.49	33.50	25.50	132.20	123.20	124.96	183.31
2002	11.34	19.47	30.50	26.00	132.00	122.90	135.08	259.15
2003	11.54	19.94	31.00	25.00	131.80	122.70	134.82	237.78
2004	11.27	18.96	29.00	23.00	131.50	122.40	129.02	237.56
2005	11.28	26.24	28.00	24.00	131.30	122.10	134.27	234.06
2006	11.02	24.54	26.00	23.00	131.00	121.80	134.57	226.52
2007	10.99	21.02	28.00	21.00	130.70	121.50	129.58	208.29
2008	10.70	23.47	24.00	22.00	130.30	121.20	134.84	211.15
2009	10.08	22.84	22.00	17.00	129.90	121.00	138.81	238.86
2010	10.28	19.06	20.00	20.00	129.50	120.70	145.68	221.29
2011	10.47	16.41	17.00	18.00	129.00	120.40	137.64	220.63
2012	9.26	19.89	17.00	16.00	128.60	120.10	133.12	234.52
2013	9.50	19.83	17.00	17.00	128.20	119.90	131.33	226.86

Table A.2.3: France raw real datasets of 6 life-style parameters

France - Databases of 6 parameters								
	Alcohol Consumed	*Cheese Consumed*	*Smoking Habit*		*Mean Systolic Blood Pressure*		*Cereal Supply*	*Fruit and Veg Supply*
Year			Males	Females	Males	Females		
1990	15.73	21.62	*35.62*	*19.14*	132.20	125.40	109.49	206.92
1991	14.85	21.74	38.00	20.00	132.00	125.00	108.47	207.33
1992	14.72	21.61	*35.19*	*19.67*	131.70	124.60	105.18	210.54
1993	14.24	21.72	*34.98*	*19.93*	131.40	124.20	110.66	208.04
1994	14.11	21.87	*34.76*	*20.19*	131.20	123.70	111.22	190.68
1995	14.12	22.01	*34.55*	*20.45*	131.00	123.30	110.88	200.86
1996	13.77	22.01	35.00	21.00	130.80	122.90	111.68	194.27
1997	13.31	22.32	*34.12*	*20.97*	130.60	122.50	112.53	186.91
1998	13.27	22.84	*33.91*	*21.24*	130.50	122.20	114.22	194.05
1999	13.15	23.36	*33.69*	*21.50*	130.30	121.80	114.76	198.48
2000	13.63	23.57	33.00	21.00	130.20	121.50	115.73	206.80
2001	13.89	24.40	*33.27*	*22.02*	130.10	121.10	117.12	204.59
2002	13.78	24.54	30.60	21.50	130.00	120.80	117.73	210.27
2003	13.49	23.95	30.00	21.20	129.80	120.50	116.32	202.99
2004	13.18	23.88	*32.62*	*22.81*	129.70	120.20	118.11	219.65
2005	12.20	23.56	31.40	23.00	129.50	119.90	119.22	216.29
2006	12.40	23.96	*32.20*	*23.33*	129.40	119.60	121.76	208.48
2007	12.20	24.31	*31.98*	*23.59*	129.10	119.30	117.64	214.47
2008	11.90	24.28	*31.77*	*23.85*	128.90	119.00	125.98	216.06
2009	11.80	24.52	*31.56*	*24.12*	128.70	118.70	119.74	221.55
2010	11.70	24.45	32.40	26.60	128.40	118.40	128.50	216.02
2011	11.80	24.13	*31.13*	*24.64*	128.10	118.10	125.69	213.35
2012	11.50	23.88	*30.91*	*24.90*	127.90	117.80	128.45	206.24
2013	11.10	23.66	*30.70*	*25.16*	127.60	117.50	127.24	211.66

Table A.2.4: Finland raw real datasets of 6 life-style parameters

Finland - Databases of 6 parameters								
	Alcohol Consumed	*Cheese Consumed*	*Smoking Habit*		*Mean Systolic Blood Pressure*		*Cereal Supply*	*Fruit and Veg Supply*
Year			Males	Females	Males	Females		
1990	9.53	13.29	32.00	20.00	134.50	128.50	102.05	135.63
1991	9.22	11.72	33.00	22.00	134.30	128.20	99.56	139.24
1992	8.88	13.17	33.00	20.00	134.10	128.00	101.36	138.39
1993	8.39	12.14	30.00	19.00	133.90	127.70	90.30	134.79
1994	8.16	12.46	27.00	19.00	133.60	127.30	88.98	148.15
1995	8.31	13.19	29.00	20.00	133.30	127.00	94.58	117.54
1996	8.24	14.09	27.00	18.00	133.00	126.60	95.27	139.95
1997	8.56	13.43	30.00	20.00	132.70	126.10	104.46	144.06
1998	8.60	15.03	30.00	20.00	132.30	125.70	103.53	135.49
1999	8.62	15.91	27.00	20.00	132.00	125.30	108.53	157.68
2000	8.59	14.08	27.00	20.00	131.60	124.80	107.08	153.29
2001	8.94	15.62	29.00	20.00	131.30	124.40	108.51	160.95
2002	9.25	15.53	27.50	19.90	131.00	123.90	106.32	155.70
2003	9.31	15.89	25.70	19.30	130.70	123.50	106.58	164.41
2004	9.89	14.20	27.10	19.50	130.40	123.10	107.25	171.03
2005	9.95	14.72	26.00	18.20	130.10	122.70	109.99	173.07
2006	10.15	14.55	24.40	18.90	129.80	122.30	115.24	164.43
2007	10.45	16.06	25.80	16.60	129.50	121.90	112.47	172.62
2008	10.26	18.03	24.00	17.60	129.10	121.60	114.99	165.63
2009	9.96	17.56	21.90	16.00	128.80	121.20	110.56	174.50
2010	9.72	17.61	23.20	15.70	128.50	120.80	115.75	169.69
2011	9.81	20.95	21.90	14.80	128.20	120.40	111.26	182.65
2012	9.24	21.87	20.90	14.00	127.90	120.10	115.29	183.22
2013	8.97	23.10	19.00	13.00	127.50	119.70	115.19	184.02

Table A.2.5: Germany raw real datasets of 6 life-style parameters

Germany - Databases of 6 parameters								
	Alcohol Consumed	*Cheese Consumed*	*Smoking Habit*		*Mean Systolic Blood Pressure*		*Cereal Supply*	*Fruit and Veg Supply*
Year			Males	Females	Males	Females		
1990	14.89	17.24	*33.53*	*18.52*	135.80	129.90	93.10	189.63
1991	13.92	16.26	*33.17*	*18.50*	135.60	129.60	92.74	179.77
1992	13.84	17.07	32.40	17.80	135.50	129.20	91.41	195.62
1993	13.50	16.54	*32.46*	*18.44*	135.20	128.80	91.78	159.53
1994	13.37	17.08	*32.11*	*18.42*	135.00	128.30	94.70	166.06
1995	13.35	17.31	31.30	17.80	134.70	127.80	94.93	157.81
1996	13.12	17.86	*31.40*	*18.37*	134.30	127.30	95.02	182.10
1997	13.00	18.25	*31.04*	*18.34*	134.00	126.70	80.61	169.13
1998	12.74	18.12	*30.69*	*18.31*	133.60	126.20	96.80	169.39
1999	12.78	18.38	30.90	18.90	133.20	125.60	97.36	172.13
2000	12.91	18.58	*29.98*	*18.26*	132.80	125.00	97.67	202.93
2001	12.46	19.43	*29.63*	*18.23*	132.40	124.40	106.86	179.82
2002	12.25	19.93	*29.27*	*18.21*	132.00	123.80	114.59	186.80
2003	11.92	19.42	29.80	19.10	131.60	123.20	105.25	179.00
2004	11.83	19.80	*28.56*	*18.16*	131.10	122.70	106.86	175.32
2005	11.67	19.71	27.90	18.80	130.70	122.10	110.91	173.29
2006	11.76	19.91	*27.85*	*18.10*	130.30	121.50	110.46	172.22
2007	11.50	20.39	*27.50*	*18.08*	129.80	121.00	111.32	169.74
2008	11.36	20.35	*27.14*	*18.05*	129.40	120.50	110.55	168.94
2009	11.22	20.38	26.40	17.60	129.00	119.90	112.08	173.24
2010	11.20	20.99	*26.44*	*18.00*	128.50	119.40	112.72	168.66
2011	11.20	21.51	*26.08*	*17.97*	128.10	118.90	113.01	182.23
2012	11.18	21.70	*25.73*	*17.95*	127.60	118.40	109.89	184.83
2013	10.94	21.69	25.10	17.10	127.20	117.90	111.11	181.37

Table A.2.6: Greece raw real datasets of 6 life-style parameters

Greece - Databases of 6 parameters								
	Alcohol Consumed	*Cheese Consumed*	*Smoking Habit*		*Mean Systolic Blood Pressure*		*Cereal Supply*	*Fruit and Veg Supply*
Year			Males	Females	Males	Females		
1990	10.33	25.11	*55.77*	*31.05*	128.00	123.50	152.13	424.85
1991	10.18	24.18	60.00	32.00	128.00	123.20	150.82	442.67
1992	9.99	23.78	*54.07*	*30.92*	127.90	122.90	149.14	425.02
1993	10.67	24.22	*53.23*	*30.86*	127.80	122.60	151.20	433.15
1994	10.43	24.58	46.00	28.00	127.70	122.30	145.76	450.06
1995	10.09	24.63	49.00	29.00	127.60	122.10	147.79	418.05
1996	9.67	24.43	*50.69*	*30.67*	127.50	121.80	147.59	420.13
1997	9.50	25.78	*49.84*	*30.61*	127.50	121.50	144.92	404.01
1998	9.03	26.44	*48.99*	*30.55*	127.40	121.30	139.11	407.30
1999	9.47	26.80	*48.15*	*30.48*	127.40	121.00	142.00	456.51
2000	8.48	24.48	46.80	29.00	127.30	120.80	141.80	452.96
2001	8.62	22.59	*46.45*	*30.36*	127.30	120.50	143.62	452.70
2002	8.09	25.44	51.00	39.00	127.20	120.30	145.49	429.35
2003	9.46	26.49	*44.76*	*30.23*	127.20	120.00	144.42	439.09
2004	9.56	27.66	*43.91*	*30.17*	127.10	119.80	146.82	487.33
2005	9.95	27.51	*43.07*	*30.11*	127.00	119.50	139.33	433.74
2006	9.42	27.15	*42.22*	*30.04*	126.90	119.30	132.29	389.78
2007	9.67	30.69	*41.37*	*29.98*	126.80	119.00	127.80	399.71
2008	9.51	26.37	*40.53*	*29.92*	126.70	118.70	133.62	370.81
2009	9.08	26.31	38.00	26.10	126.60	118.50	131.06	392.16
2010	9.00	25.78	*38.83*	*29.79*	126.50	118.20	129.61	343.57
2011	8.02	25.58	*37.99*	*29.73*	126.30	117.90	130.52	364.37
2012	8.20	25.64	*37.14*	*29.66*	126.20	117.70	132.54	345.17
2013	7.46	25.47	*36.29*	*29.60*	126.00	117.40	135.34	345.91

Table A.2.7: Iceland raw real datasets of 6 life-style parameters

Iceland - Databases of 6 parameters								
	Alcohol Consumed	*Cheese Consumed*	*Smoking Habit*		*Mean Systolic Blood Pressure*		*Cereal Supply*	*Fruit and Veg Supply*
Year			Males	Females	Males	Females		
1990	5.10	9.27	30.80	29.90	128.20	120.10	83.42	112.18
1991	5.03	7.80	31.40	29.60	128.00	119.90	92.85	121.18
1992	4.64	9.16	28.10	29.70	127.80	119.80	91.59	126.17
1993	4.35	11.87	26.10	27.10	127.60	119.60	93.10	128.23
1994	4.57	12.46	27.90	25.90	127.40	119.50	82.22	127.51
1995	4.71	12.87	26.70	26.90	127.20	119.30	87.76	125.98
1996	4.83	12.89	28.20	28.00	127.00	119.10	82.81	127.99
1997	5.10	13.50	28.30	26.30	126.90	119.00	80.56	131.05
1998	5.47	13.50	24.50	25.40	126.70	118.80	77.65	133.55
1999	5.89	15.91	25.00	25.50	126.60	118.60	76.63	137.88
2000	6.17	15.31	23.30	22.50	126.50	118.40	77.36	150.74
2001	6.37	15.97	24.50	22.80	126.40	118.30	79.51	159.77
2002	6.61	23.92	22.20	21.10	126.30	118.10	70.93	167.44
2003	6.61	24.98	25.40	19.60	126.30	118.00	73.52	162.32
2004	6.79	24.96	21.50	18.90	126.20	117.80	73.18	167.93
2005	7.05	24.93	19.50	19.50	126.20	117.60	77.51	179.42
2006	7.20	25.11	21.30	17.40	126.10	117.40	77.30	209.19
2007	7.53	25.29	20.70	18.20	126.00	117.20	81.04	225.64
2008	8.49	25.47	20.30	15.30	126.00	117.10	83.76	216.13
2009	10.22	25.45	15.90	15.70	125.90	116.90	79.26	197.53
2010	8.25	25.25	14.50	13.90	125.70	116.70	79.44	192.31
2011	8.13	24.41	14.40	14.20	125.60	116.50	77.44	193.44
2012	7.81	30.62	14.90	12.80	125.50	116.30	82.86	209.16
2013	7.31	30.82	10.70	12.10	125.30	116.10	82.53	203.21

Table A.2.8: Italy raw real datasets of 6 life-style parameters

Italy - Databases of 6 parameters								
	Alcohol Consumed	*Cheese Consumed*	*Smoking Habit*		*Mean Systolic Blood Pressure*		*Cereal Supply*	*Fruit and Veg Supply*
Year			Males	Females	Males	Females		
1990	10.99	20.32	*36.66*	*17.88*	132.60	127.00	157.41	300.96
1991	10.82	19.30	*36.22*	*17.80*	132.50	126.80	160.74	306.27
1992	10.55	19.53	*35.79*	*17.73*	132.40	126.60	155.46	334.53
1993	10.27	18.80	35.60	16.60	132.30	126.40	152.53	298.33
1994	10.12	18.61	34.50	16.90	132.10	126.20	151.04	302.27
1995	9.62	18.30	34.40	17.40	132.00	126.00	159.07	297.11
1996	9.13	19.63	35.40	18.00	131.80	125.70	159.78	307.22
1997	9.12	18.65	33.60	17.50	131.70	125.50	161.13	296.21
1998	8.98	21.74	32.60	17.50	131.50	125.20	161.78	294.50
1999	8.86	22.22	32.80	17.30	131.40	125.00	162.93	321.92
2000	9.78	22.38	31.90	17.40	131.20	124.70	161.61	343.25
2001	9.69	22.21	31.60	17.10	131.10	124.40	162.96	293.97
2002	9.25	21.48	31.30	17.20	130.90	124.20	162.15	288.21
2003	9.30	21.40	31.40	17.60	130.70	123.90	161.34	308.40
2004	8.98	21.27	*30.59*	*16.82*	130.60	123.60	159.65	352.17
2005	8.65	22.17	28.70	16.40	130.40	123.30	156.41	336.96
2006	8.44	22.13	29.20	17.20	130.20	122.90	156.81	321.63
2007	8.37	22.03	28.60	16.60	130.00	122.60	156.54	319.32
2008	7.96	21.93	28.90	16.40	129.70	122.30	156.22	308.99
2009	7.25	22.33	29.90	17.10	129.50	121.90	158.18	349.21
2010	6.95	24.04	29.60	17.10	129.20	121.50	154.65	300.66
2011	6.98	23.97	28.70	16.70	128.90	121.10	155.01	288.41
2012	7.49	24.04	28.00	16.60	128.60	120.70	156.88	263.76
2013	7.35	23.11	26.60	15.90	128.30	120.40	158.17	268.66

Table A.2.9: Netherlands raw real datasets of 6 life-style parameters

Netherlands - Databases of 6 parameters								
	Alcohol Consumed	*Cheese Consumed*	*Smoking Habit*		*Mean Systolic Blood Pressure*		*Cereal Supply*	*Fruit and Veg Supply*
Year			Males	Females	Males	Females		
1990	9.91	14.73	42.30	31.30	132.40	124.80	67.54	212.24
1991	10.03	13.78	43.80	32.40	132.30	124.70	70.90	206.03
1992	10.03	13.90	42.40	30.30	132.20	124.60	69.41	256.40
1993	9.67	16.11	41.60	30.80	132.10	124.60	65.76	210.99
1994	9.68	16.20	42.50	31.90	132.00	124.50	76.45	211.98
1995	9.80	19.44	40.70	31.10	132.00	124.40	80.22	213.79
1996	9.80	16.93	40.00	31.50	131.90	124.40	83.99	205.79
1997	10.05	21.46	39.10	32.20	131.80	124.30	77.23	182.46
1998	9.93	19.24	38.50	30.70	131.70	124.10	74.94	177.55
1999	10.06	19.16	36.00	31.70	131.60	124.00	67.02	217.51
2000	10.06	20.52	35.90	29.20	131.60	123.90	66.32	219.03
2001	9.95	20.30	32.34	25.39	131.50	123.80	70.31	223.29
2002	9.68	20.39	30.88	24.45	131.40	123.60	69.69	229.92
2003	9.56	20.30	29.19	24.28	131.30	123.40	76.31	213.43
2004	9.56	21.61	28.58	22.29	131.20	123.30	76.00	224.26
2005	9.69	20.09	28.44	22.08	131.00	123.10	74.95	213.11
2006	9.79	19.17	28.77	21.74	130.90	122.90	77.10	224.25
2007	9.53	19.36	25.37	20.99	130.70	122.60	78.97	240.67
2008	9.62	16.28	25.89	20.77	130.50	122.40	79.51	228.71
2009	9.23	20.57	25.47	19.82	130.30	122.10	85.22	199.79
2010	9.33	17.31	23.05	18.82	130.00	121.90	91.71	194.57
2011	8.96	17.46	23.50	18.30	129.70	121.60	91.32	238.25
2012	9.05	16.87	20.60	16.30	129.50	121.30	91.28	234.40
2013	8.68	17.69	20.90	16.30	129.20	121.00	88.84	262.36

Table A.2.10: Norway raw real datasets of 6 life-style parameters

Norway - Databases of 6 parameters								
	Alcohol Consumed	*Cheese Consumed*	*Smoking Habit*		*Mean Systolic Blood Pressure*		*Cereal Supply*	*Fruit and Veg Supply*
Year			Males	Females	Males	Females		
1990	4.99	14.93	36.60	33.20	135.50	129.30	114.47	155.91
1991	4.90	14.48	36.50	*37.08*	135.40	129.10	118.61	150.46
1992	4.67	13.26	36.90	33.30	135.20	128.90	117.37	151.85
1993	4.55	14.79	*36.74*	*35.17*	135.00	128.70	117.52	161.44
1994	4.74	14.37	*35.70*	*34.22*	134.80	128.40	119.03	163.50
1995	4.79	14.67	33.70	32.10	134.70	128.20	115.88	157.68
1996	5.04	15.06	33.60	32.60	134.50	127.90	119.31	162.39
1997	5.28	14.88	33.90	32.50	134.30	127.60	120.29	165.31
1998	5.24	15.01	33.50	32.30	134.20	127.40	123.87	162.56
1999	5.45	14.36	32.40	32.20	134.00	127.00	125.73	170.85
2000	5.67	15.00	31.30	31.10	133.90	126.70	124.94	169.03
2001	5.49	14.91	29.50	29.70	133.70	126.40	126.24	176.75
2002	5.89	15.50	28.80	29.40	133.60	126.10	126.93	190.40
2003	6.04	15.43	27.20	25.30	133.50	125.80	125.04	199.69
2004	6.22	15.52	27.20	24.80	133.30	125.40	122.93	205.62
2005	6.37	15.22	26.00	24.00	133.20	125.10	125.36	205.89
2006	6.47	15.20	24.00	24.00	133.00	124.70	125.23	210.11
2007	6.60	15.10	21.00	23.00	132.80	124.40	125.57	220.15
2008	6.75	14.90	21.00	21.00	132.60	124.00	123.92	229.68
2009	6.68	14.79	20.00	20.00	132.40	123.70	127.00	207.88
2010	6.59	14.76	19.00	19.00	132.20	123.30	123.35	199.21
2011	6.53	14.56	17.00	18.00	131.90	123.00	123.15	215.65
2012	6.21	14.87	16.00	16.00	131.70	122.60	120.57	213.88
2013	6.21	14.59	15.00	14.00	131.50	122.30	120.00	218.10

Table A.2.11: Spain raw real datasets of 6 life-style parameters

Spain - Databases of 6 parameters								
	Alcohol Consumed	*Cheese Consumed*	*Smoking Habit*		*Mean Systolic Blood Pressure*		*Cereal Supply*	*Fruit and Veg Supply*
Year			Males	Females	Males	Females		
1990	12.91	5.24	*47.42*	*24.78*	129.40	123.70	103.36	328.30
1991	12.80	5.26	*46.54*	*24.58*	129.30	123.30	103.61	300.80
1992	12.12	5.44	*45.65*	*24.38*	129.20	123.00	100.74	308.11
1993	11.60	5.74	44.00	20.80	129.20	122.60	102.75	281.01
1994	11.28	5.57	*43.89*	*23.98*	129.10	122.30	101.43	270.38
1995	11.01	5.48	43.50	24.50	129.10	122.00	100.99	240.12
1996	10.76	5.72	*42.13*	*23.58*	129.00	121.70	102.76	259.72
1997	11.60	6.09	42.10	24.80	129.00	121.40	102.07	274.56
1998	11.51	6.10	*40.36*	*23.17*	129.00	121.10	101.22	261.88
1999	11.27	6.19	*39.48*	*22.97*	129.10	120.90	100.15	284.27
2000	11.12	6.67	38.60	22.77	129.10	120.60	98.99	271.93
2001	9.86	6.75	39.20	24.60	129.10	120.40	98.18	266.59
2002	12.26	7.07	*36.84*	*22.37*	129.10	120.10	97.74	283.21
2003	12.09	7.35	34.20	22.40	129.10	119.90	96.58	261.63
2004	11.96	7.80	*35.07*	*21.96*	129.10	119.70	96.74	257.19
2005	11.92	7.41	*34.19*	*21.76*	129.10	119.40	95.87	256.00
2006	11.86	7.99	31.60	21.50	129.00	119.20	94.45	254.63
2007	11.05	8.89	*32.43*	*21.36*	128.90	119.00	95.80	241.10
2008	10.24	8.41	*31.55*	*21.16*	128.70	118.70	100.07	242.77
2009	9.99	8.95	31.17	21.33	128.60	118.40	102.66	234.95
2010	9.78	9.10	*29.78*	*20.76*	128.40	118.20	103.05	220.17
2011	9.62	8.93	27.87	20.22	128.10	117.90	104.59	200.43
2012	9.35	8.92	*28.02*	*20.35*	127.90	117.70	105.82	192.39
2013	9.25	8.93	*27.14*	*20.15*	127.70	117.40	105.94	191.38

Table A.2.12: Sweden raw real datasets of 6 life-style parameters

Sweden - Databases of 6 parameters								
	Alcohol Consumed	*Cheese Consumed*	*Smoking Habit*		*Mean Systolic Blood Pressure*		*Cereal Supply*	*Fruit and Veg Supply*
Year			Males	Females	Males	Females		
1990	7.41	15.63	25.80	25.90	133.90	125.50	81.42	162.75
1991	7.47	15.60	25.70	24.40	133.80	125.40	82.71	149.92
1992	7.50	15.78	25.20	26.60	133.70	125.30	96.66	154.81
1993	7.51	16.72	23.30	23.40	133.60	125.10	100.04	153.16
1994	7.71	17.08	21.60	23.80	133.50	125.00	90.37	161.21
1995	7.33	16.38	22.00	23.60	133.40	124.90	96.57	146.87
1996	6.90	16.73	21.20	23.30	133.30	124.70	97.53	152.59
1997	7.11	16.02	16.50	21.90	133.10	124.50	96.79	162.69
1998	6.80	16.87	17.00	21.10	133.00	124.20	97.97	164.85
1999	6.88	17.61	19.20	19.40	132.80	124.00	102.16	173.36
2000	6.20	17.26	16.80	21.00	132.60	123.70	101.36	172.11
2001	6.60	17.08	17.90	19.90	132.30	123.50	102.74	176.90
2002	6.90	18.24	16.30	19.30	132.10	123.20	102.88	185.83
2003	6.90	18.49	16.70	18.30	131.90	122.90	102.75	193.68
2004	6.60	17.66	15.00	17.50	131.60	122.50	103.90	198.06
2005	6.50	17.75	13.90	18.00	131.30	122.20	102.92	193.83
2006	6.50	18.28	12.30	16.70	130.90	121.80	101.11	200.42
2007	6.90	17.65	12.90	15.20	130.60	121.40	102.37	203.45
2008	6.90	19.10	13.10	16.80	130.20	121.00	102.08	222.93
2009	7.30	19.14	13.50	15.00	129.70	120.50	101.38	209.40
2010	7.20	18.88	12.50	14.70	129.30	120.10	102.43	207.48
2011	7.40	19.00	12.40	13.90	128.90	119.60	98.25	210.88
2012	7.40	19.29	12.60	13.10	128.40	119.20	98.18	215.92
2013	7.30	19.82	9.80	11.70	127.90	118.70	101.42	219.35

Table A.2.13: Switzerlandraw real datasets of 6 life-style parameters

Switzerland - Databases of 6 parameters								
	Alcohol Consumed	*Cheese Consumed*	*Smoking Habit*		*Mean Systolic Blood Pressure*		*Cereal Supply*	*Fruit and Veg Supply*
Year			Males	Females	Males	Females		
1990	12.99	14.86	*875.15*	*484.53*	131.10	123.50	106.33	218.49
1991	12.90	15.30	*899.12*	*497.64*	131.00	123.30	104.73	211.01
1992	12.33	15.12	33.90	22.80	130.90	123.10	110.06	216.43
1993	12.33	15.51	*933.27*	*516.32*	130.70	122.90	104.50	209.35
1994	11.79	15.20	*951.95*	*526.55*	130.50	122.60	105.61	208.64
1995	11.45	15.09	*952.67*	*526.94*	130.30	122.30	105.26	212.33
1996	11.33	15.34	*982.33*	*543.17*	130.10	121.90	106.69	218.28
1997	11.21	15.49	39.00	28.00	129.80	121.60	108.66	206.78
1998	11.07	15.77	*996.54*	*550.94*	129.60	121.20	111.06	212.70
1999	11.05	14.80	*1019.22*	*563.35*	129.30	120.80	111.31	189.94
2000	11.26	18.73	*1030.73*	*569.65*	129.00	120.40	116.99	190.13
2001	11.12	19.51	27.00	21.00	128.80	120.00	106.78	185.32
2002	10.85	18.49	31.00	23.00	128.50	119.60	109.82	186.48
2003	10.82	18.43	*1071.80*	*592.11*	128.30	119.30	109.80	175.26
2004	10.55	18.33	24.00	20.00	128.10	118.90	107.75	172.29
2005	10.15	18.31	*1107.06*	*611.41*	127.90	118.60	105.31	159.92
2006	10.24	18.54	*1129.93*	*623.91*	127.70	118.20	105.76	162.57
2007	10.44	19.21	23.00	18.00	127.50	117.90	106.04	169.87
2008	10.29	19.65	*1153.05*	*636.56*	127.30	117.60	100.30	181.88
2009	10.15	19.62	22.00	17.00	127.10	117.20	107.08	195.56
2010	10.01	20.30	*1168.48*	*645.01*	126.80	116.90	102.94	207.01
2011	9.99	20.10	*1177.60*	*649.99*	126.60	116.60	108.75	218.80
2012	9.86	19.50	*23.02*	*17.81*	126.40	116.20	99.81	217.55
2013	9.73	19.79	*1185.91*	*654.54*	126.10	115.90	98.68	212.31

B

HISTOGRAM OF PROBABILITY DENSITY FUNCTION (PDF)

Histogram of PDFs for 12 European countries on the raw and normalised scale features

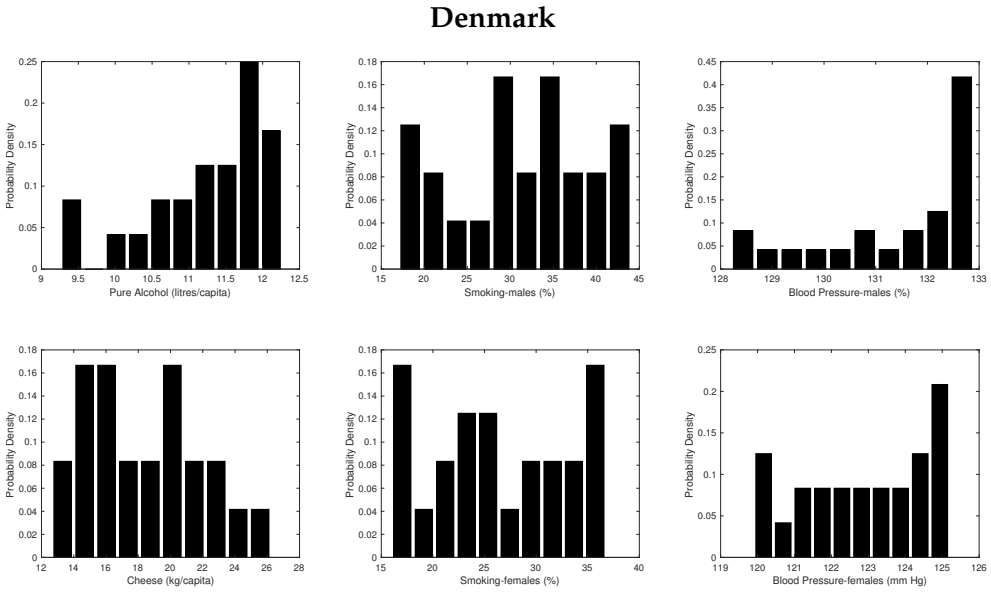


Figure B.0.1: PDF plots of the raw data features for Denmark.

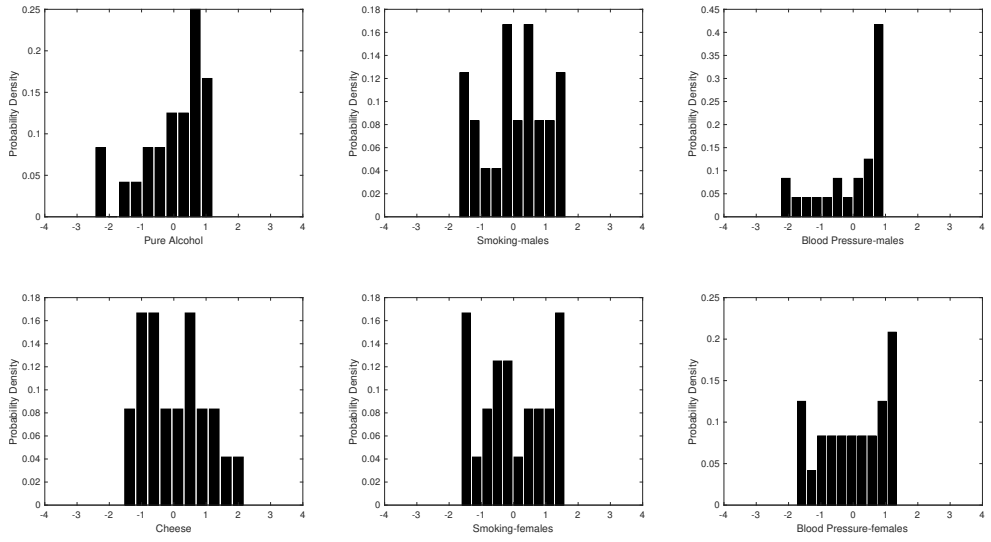


Figure B.0.2: PDF plots of the normalised data features for Denmark.

Finland

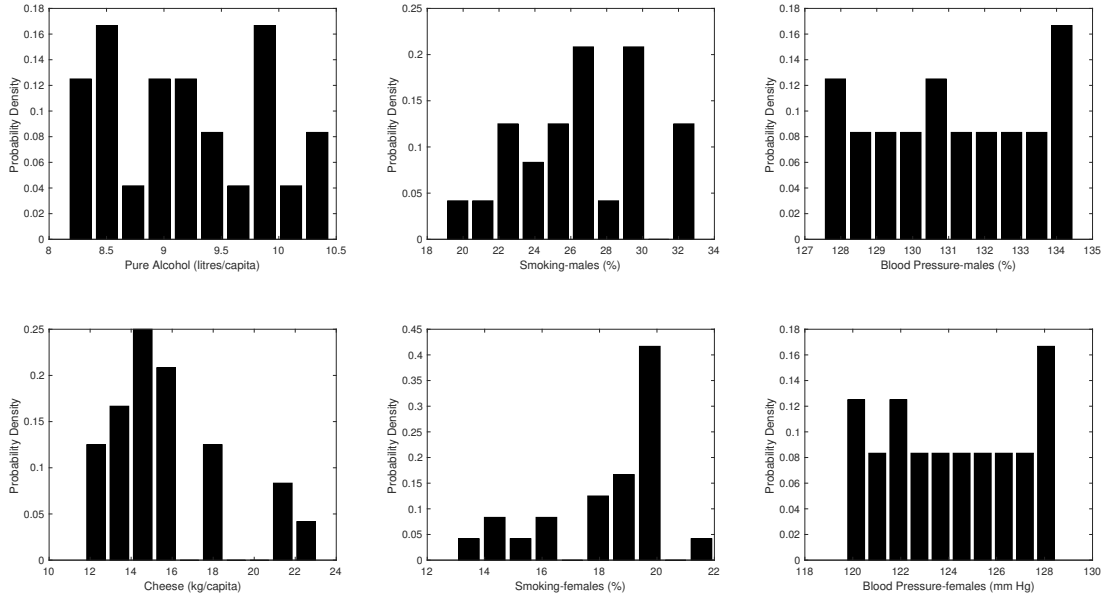


Figure B.0.3: PDF plots of the raw data features for Finland.

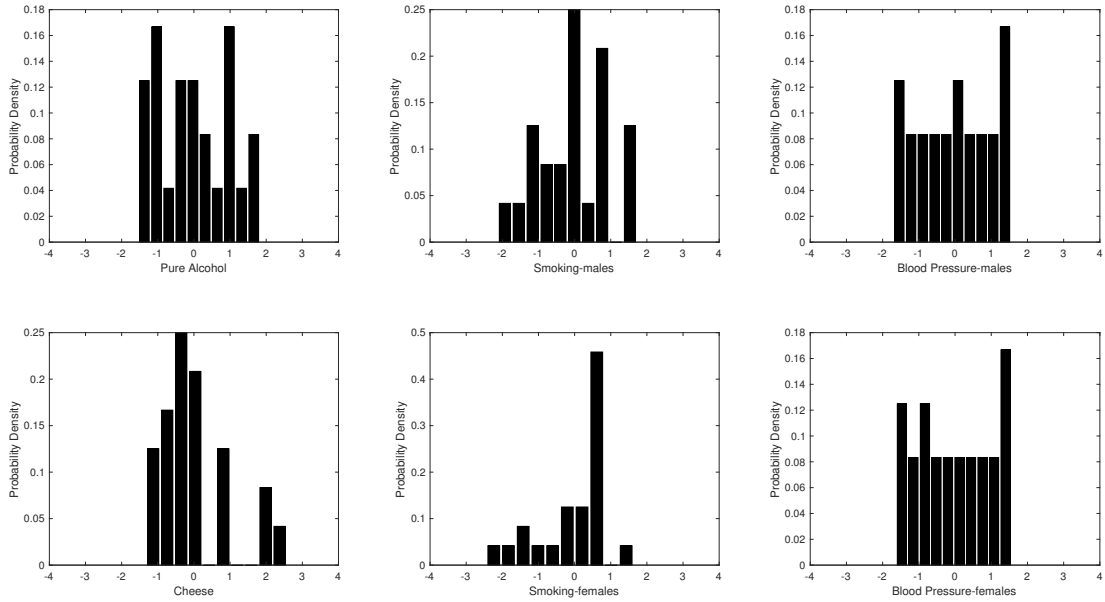


Figure B.0.4: PDF plots of the normalised data features for Finland.

France

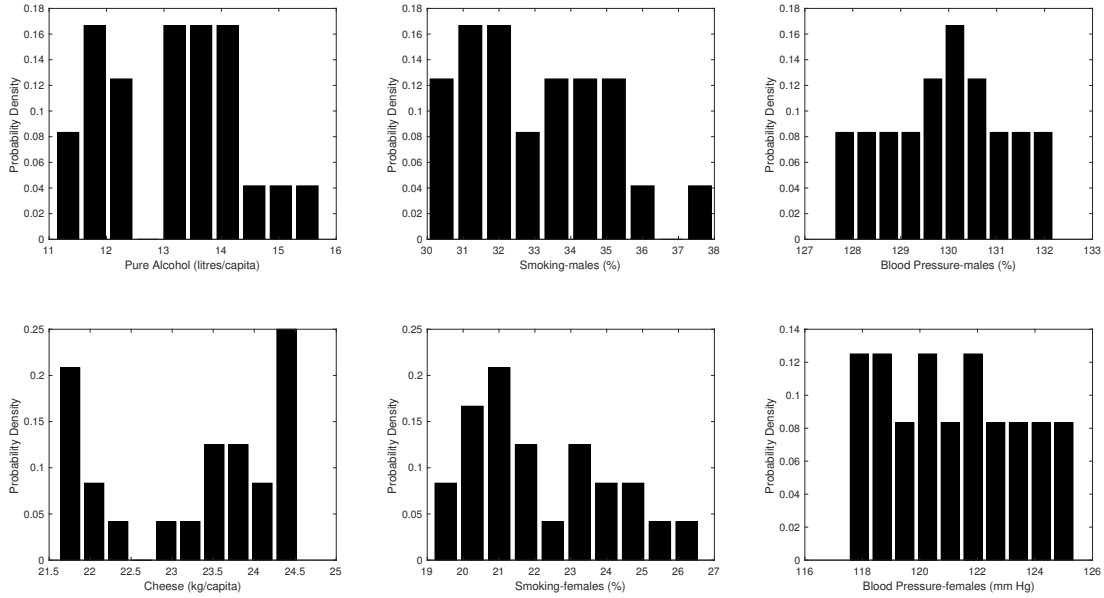


Figure B.0.5: PDF plots of the raw data features for France.

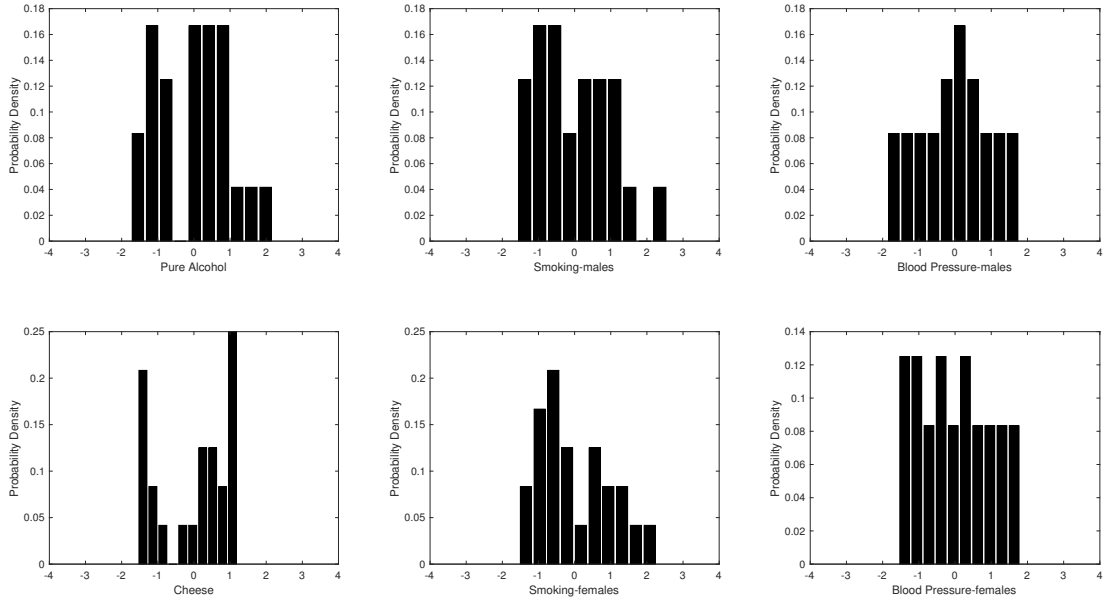


Figure B.0.6: PDF plots of the normalised data features for France.

Germany

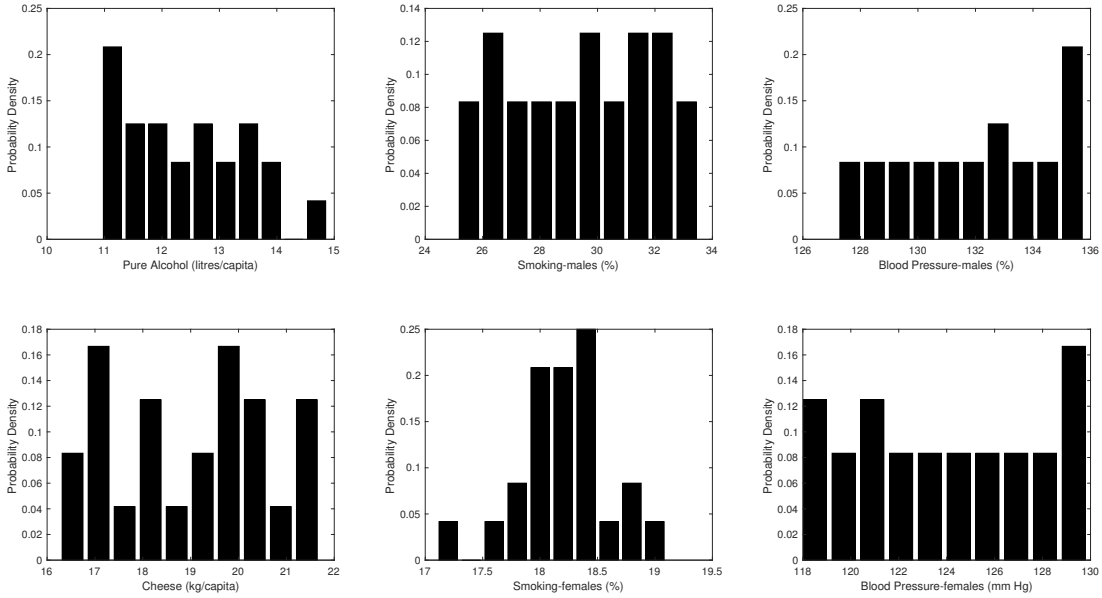


Figure B.0.7: PDF plots of the raw data features for Germany.

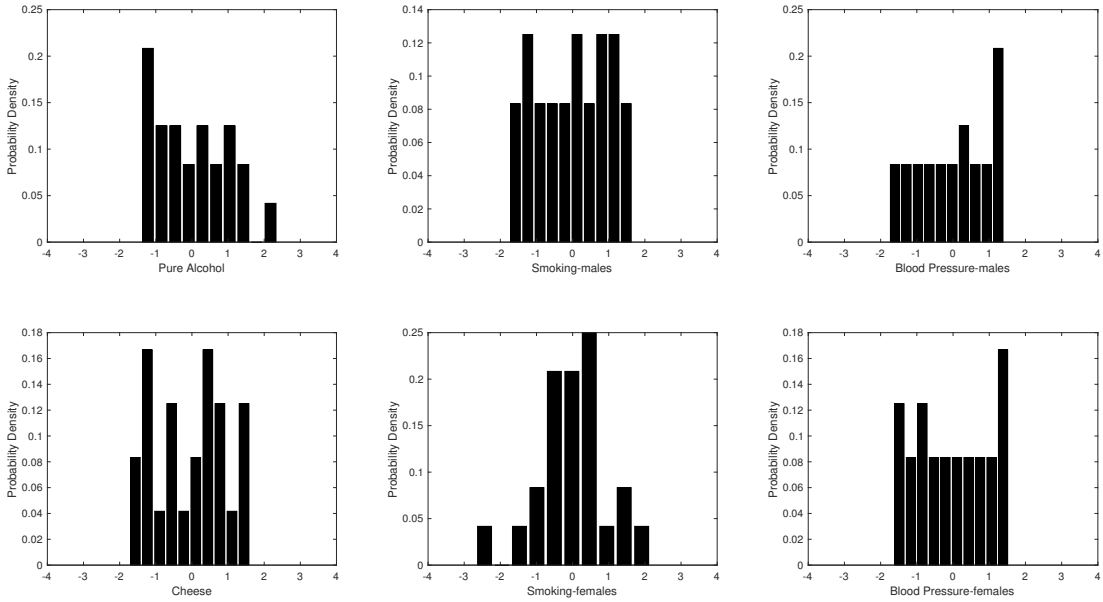


Figure B.0.8: PDF plots of the normalised data features for Germany.

Greece

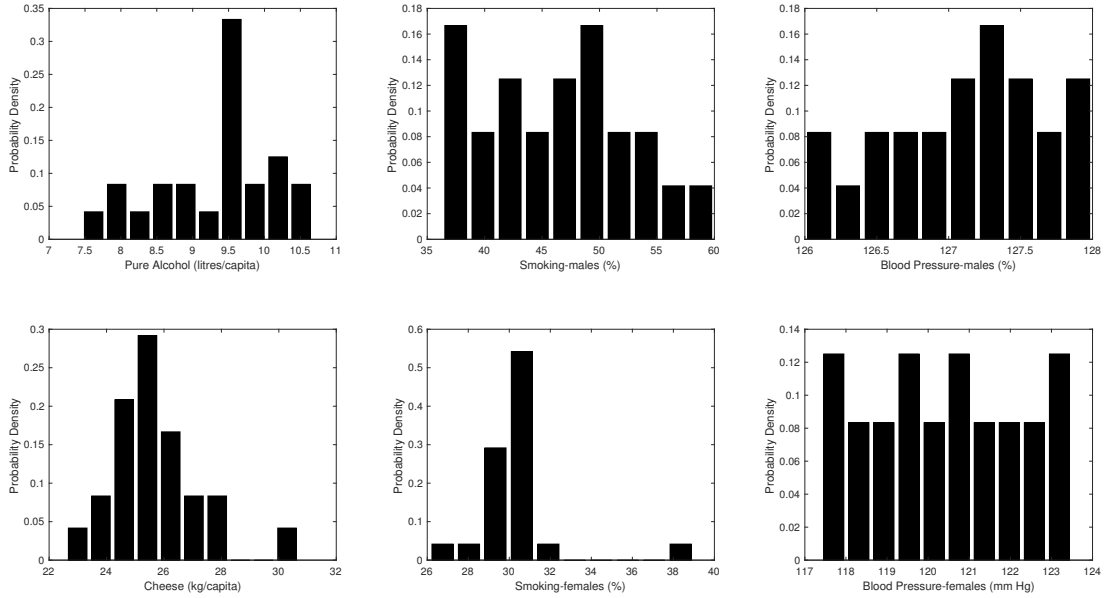


Figure B.0.9: PDF plots of the raw data features for Greece.

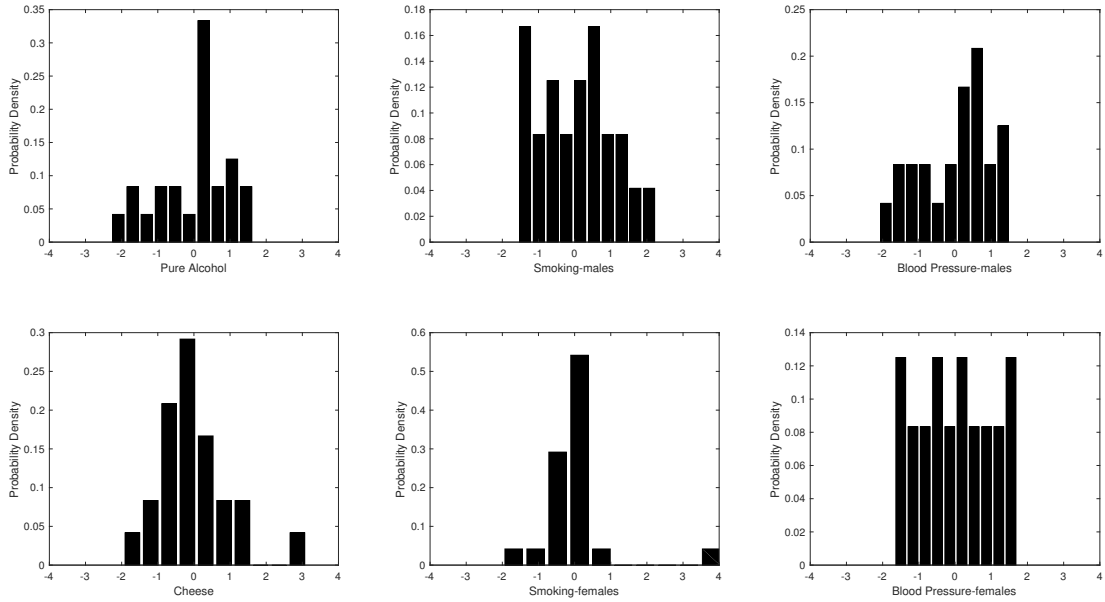


Figure B.0.10: PDF plots of the normalised data features for Greece.

Iceland

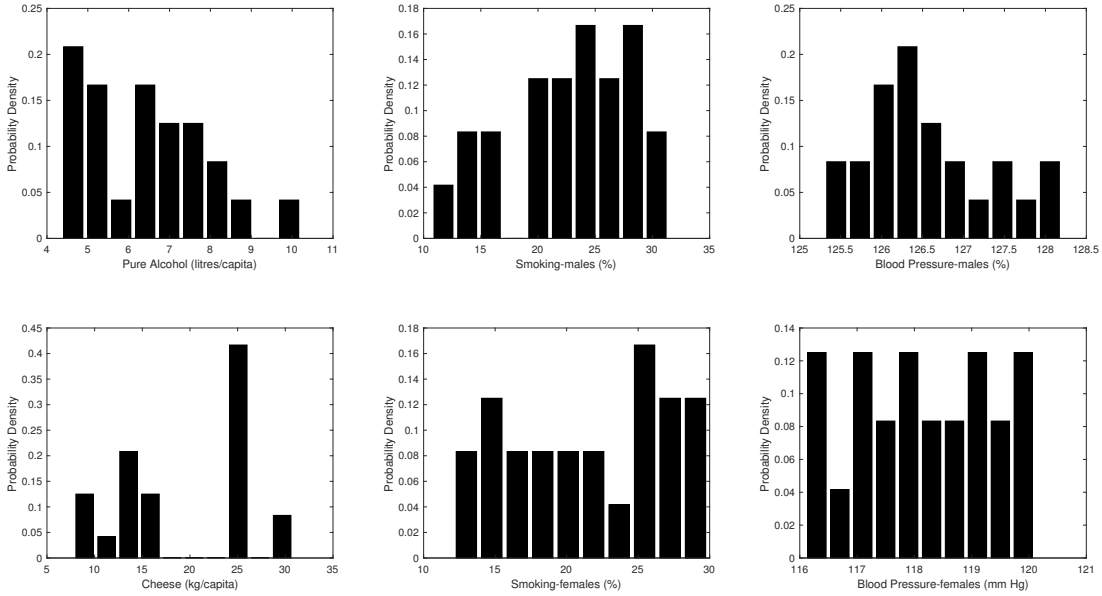


Figure B.0.11: PDF plots of the raw data features for Iceland.

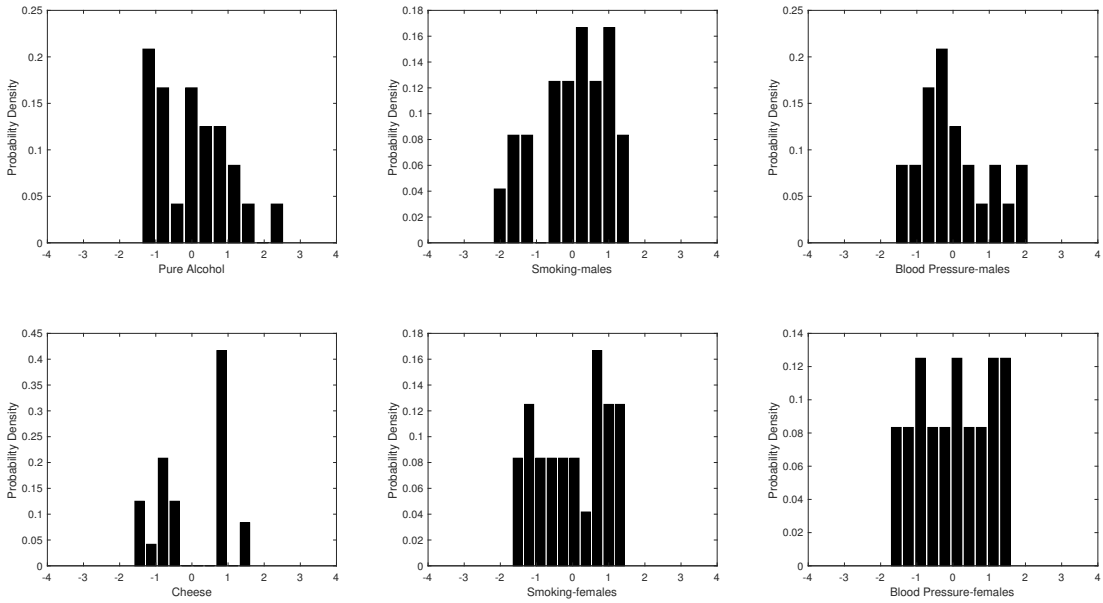


Figure B.0.12: PDF plots of the normalised data features for Iceland.

Italy

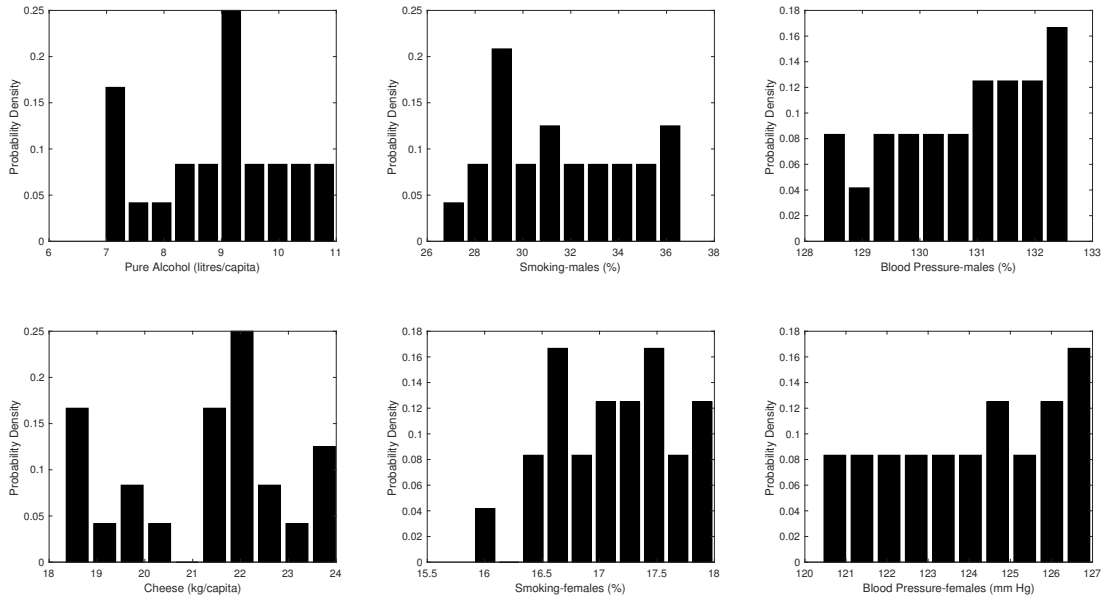


Figure B.0.13: PDF plots of the raw data features for Italy.

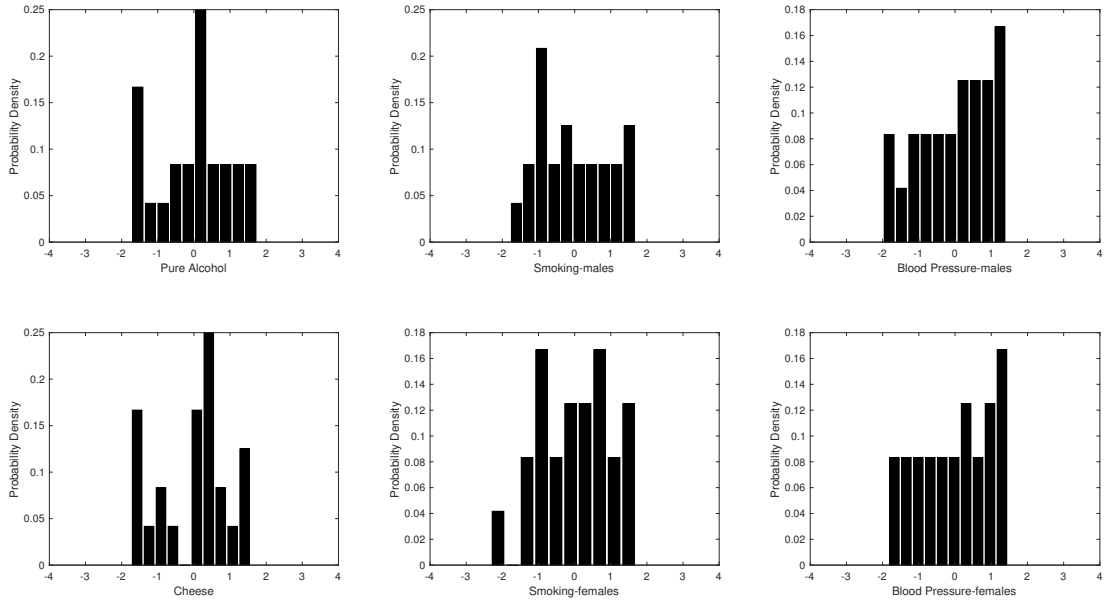


Figure B.0.14: PDF plots of the normalised data features for Italy.

Netherlands

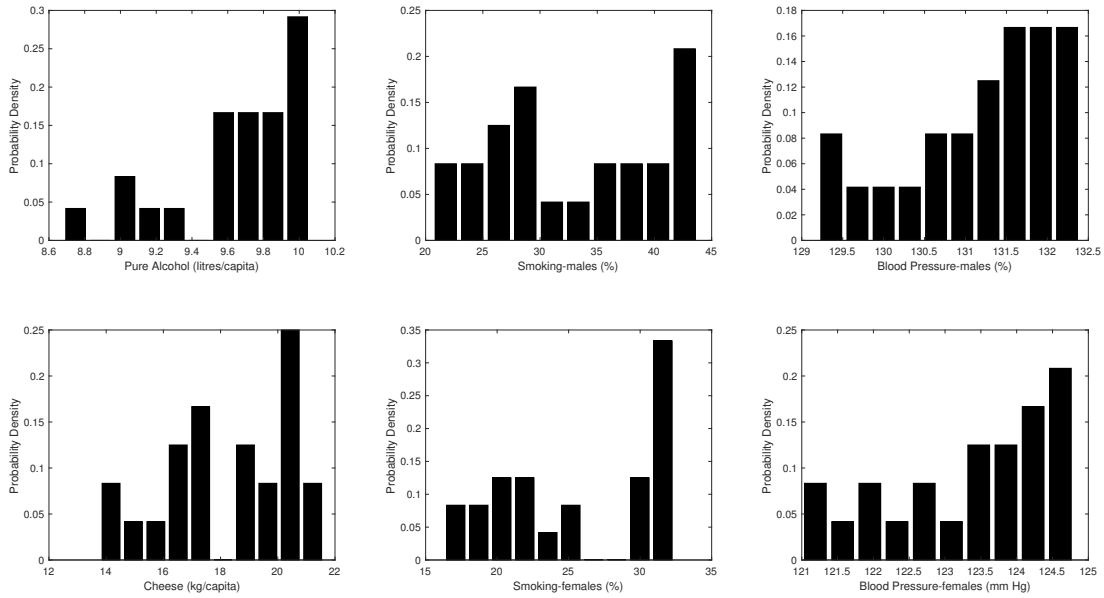


Figure B.0.15: PDF plots of the raw data features for Netherlands.

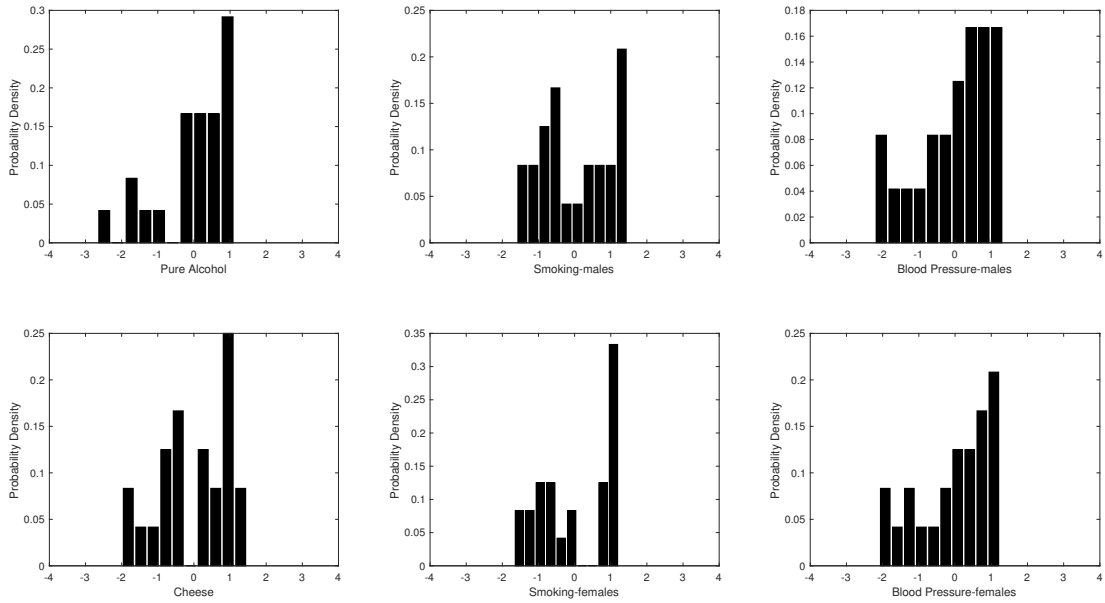


Figure B.0.16: PDF plots of the normalised data features for Netherlands.

Norway

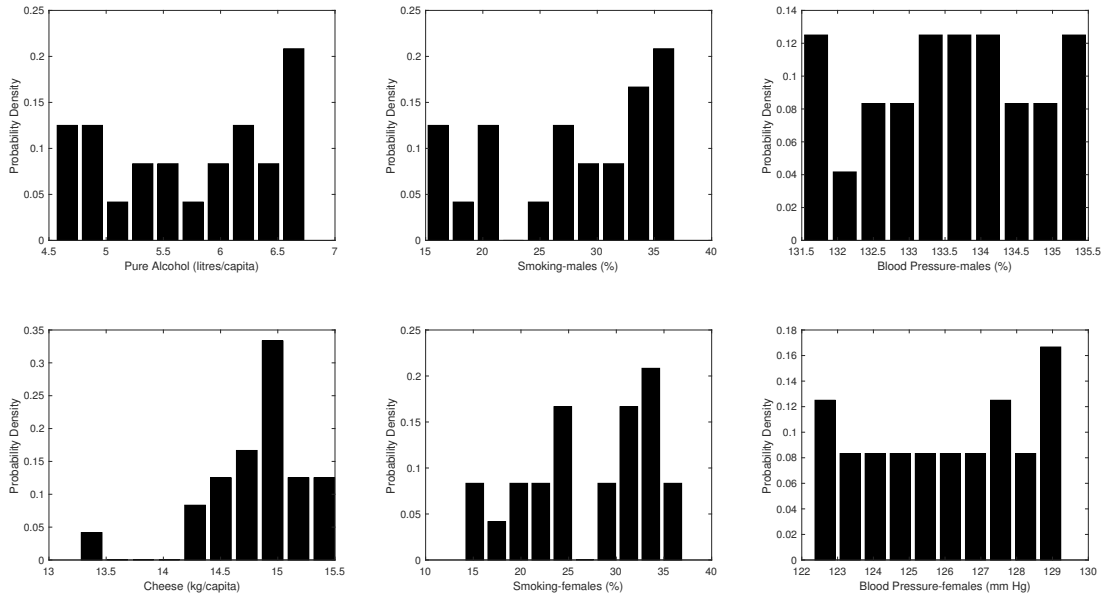


Figure B.0.17: PDF plots of the raw data features for Norway.

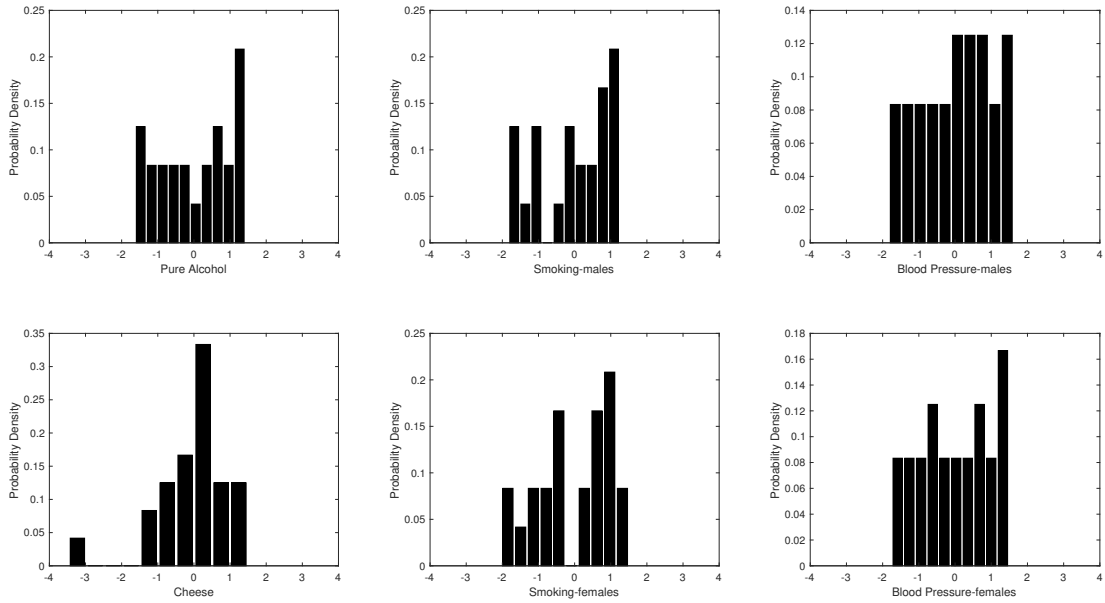


Figure B.0.18: PDF plots of the normalised data features for Norway.

Spain

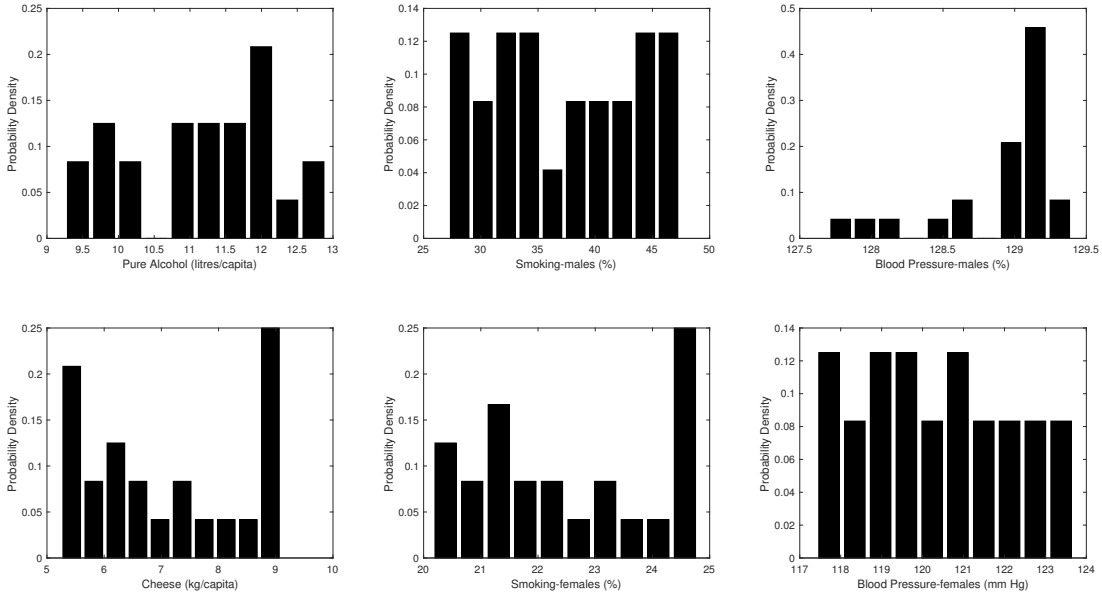


Figure B.0.19: PDF plots of the raw data features for Spain.

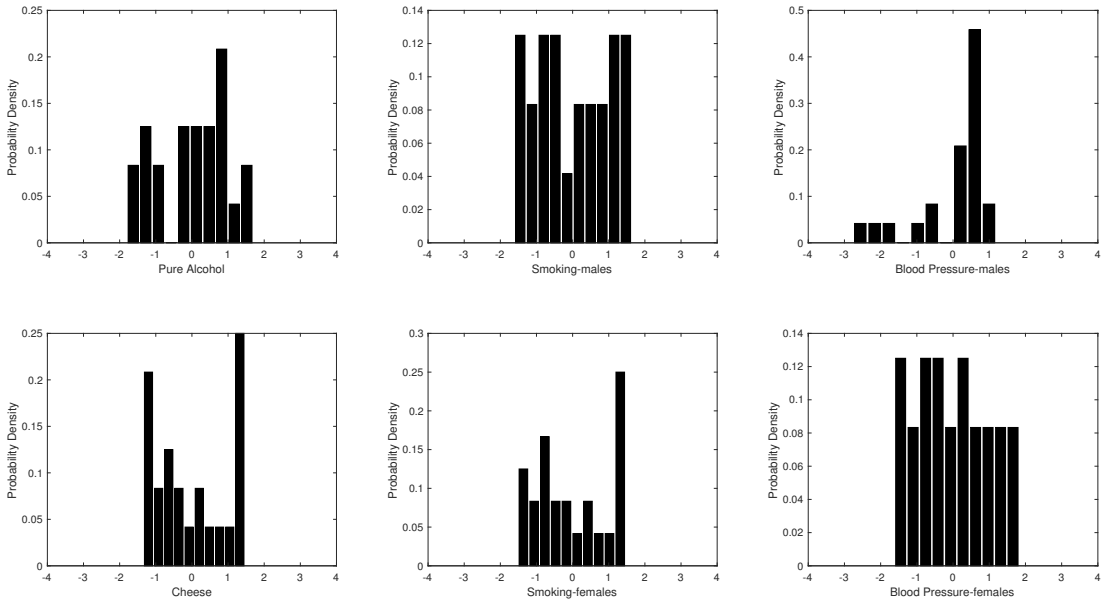


Figure B.0.20: PDF plots of the normalised data features for Spain.

Sweden

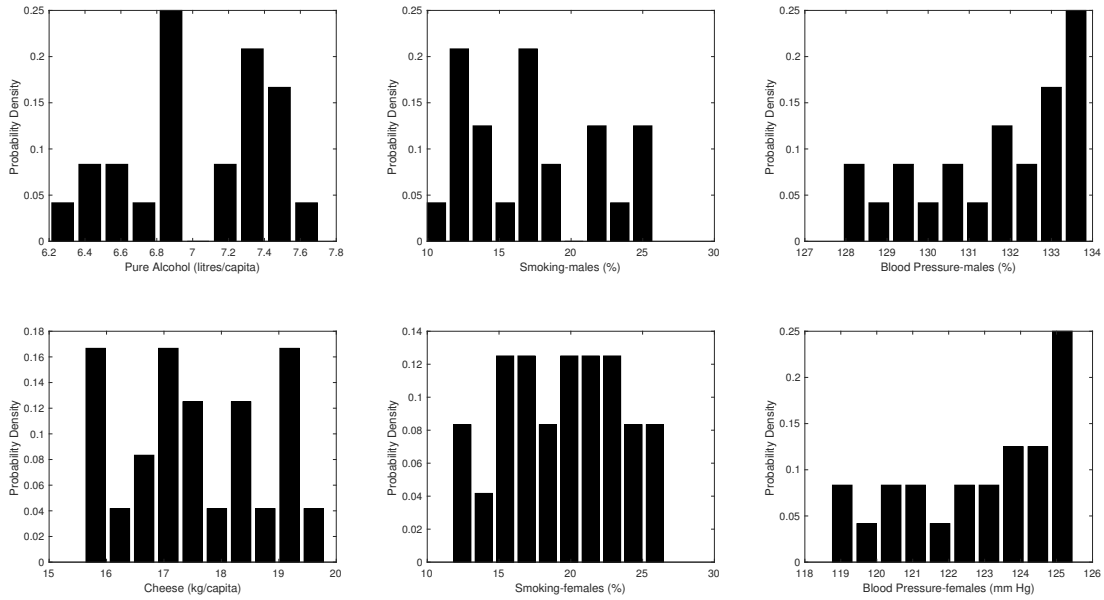


Figure B.0.21: PDF plots of the raw data features for Sweden.

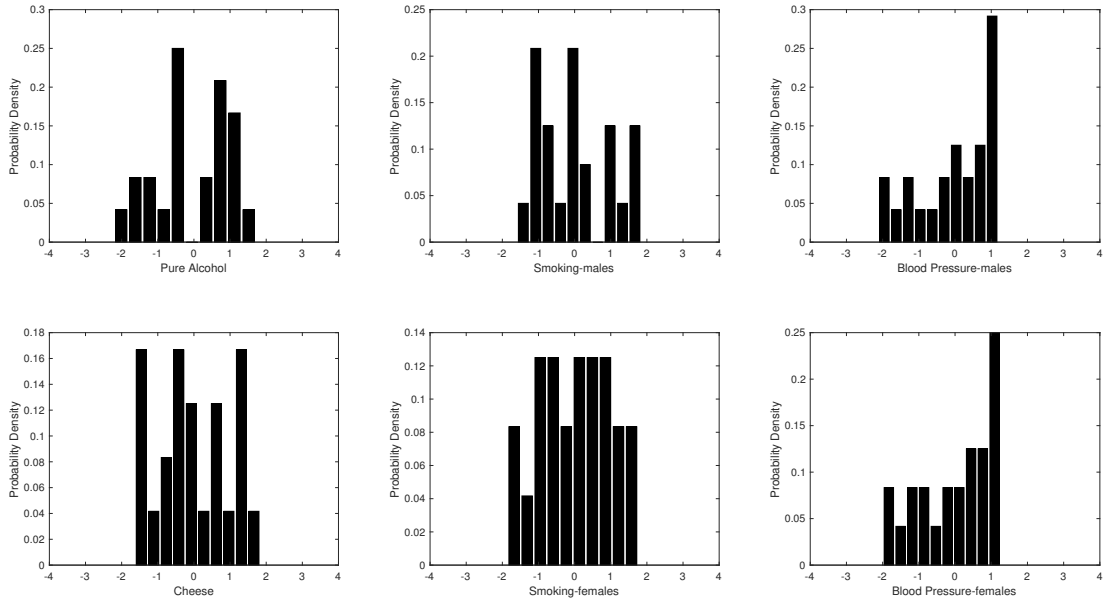


Figure B.0.22: PDF plots of the normalised data features for Sweden.

Switzerland

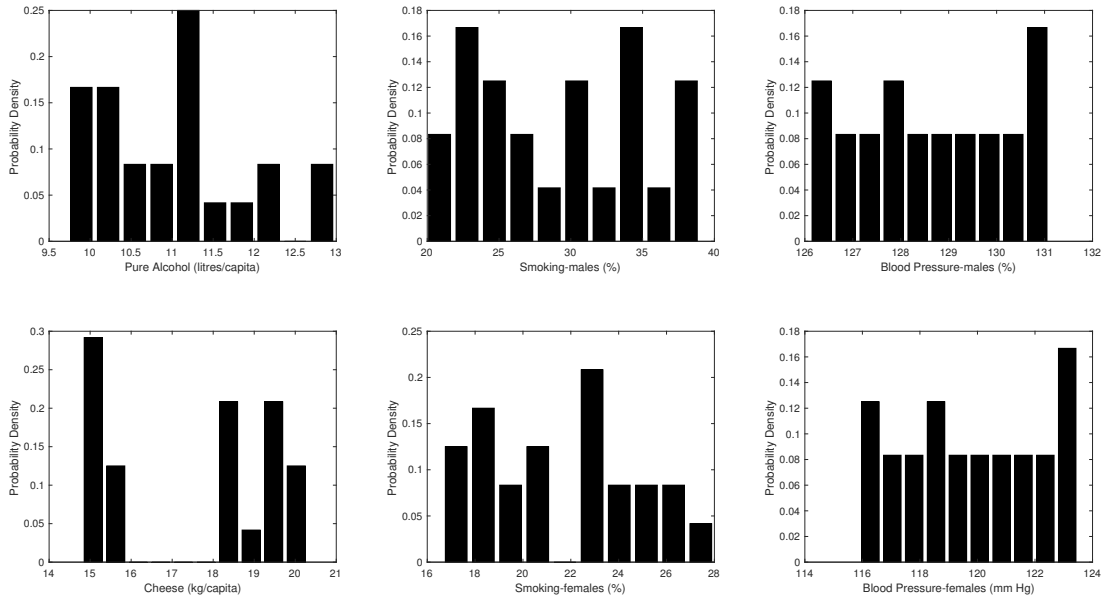


Figure B.0.23: PDF plots of the raw data features for Switzerland.

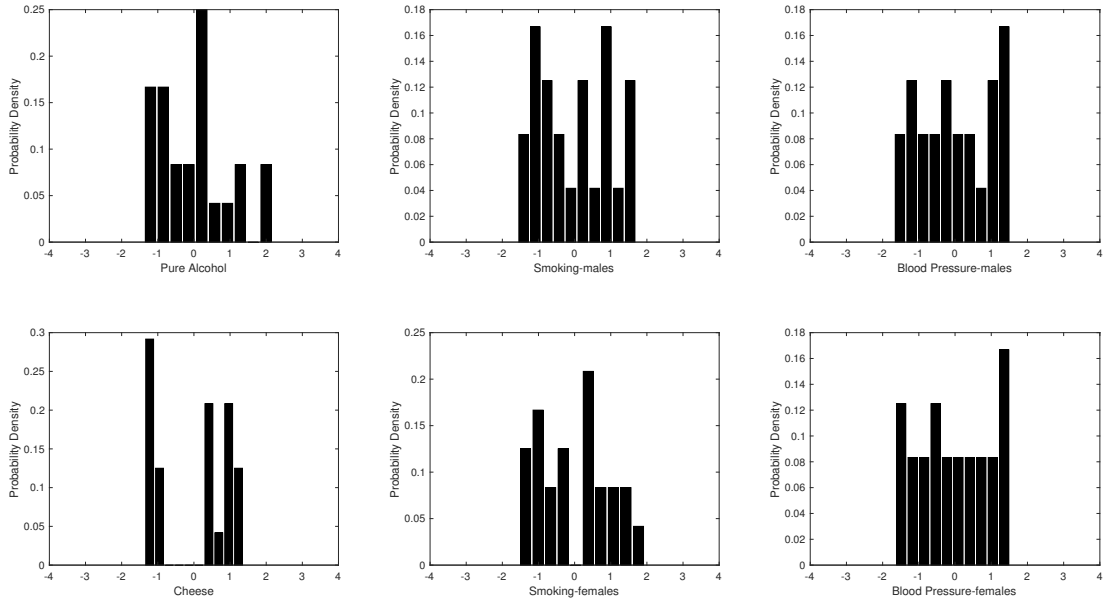


Figure B.0.24: PDF plots of the normalised data features for Switzerland.

C

LINEAR REGRESSION OF 6 LIFE-STYLE PARAMETERS FOR 12 EUROPEAN COUNTRIES

Results of linear regression of 6 life-style parameters for 12 European country

C.1 MEDITERRANEAN EUROPEAN COUNTRIES (MEEU) BLOCK

France

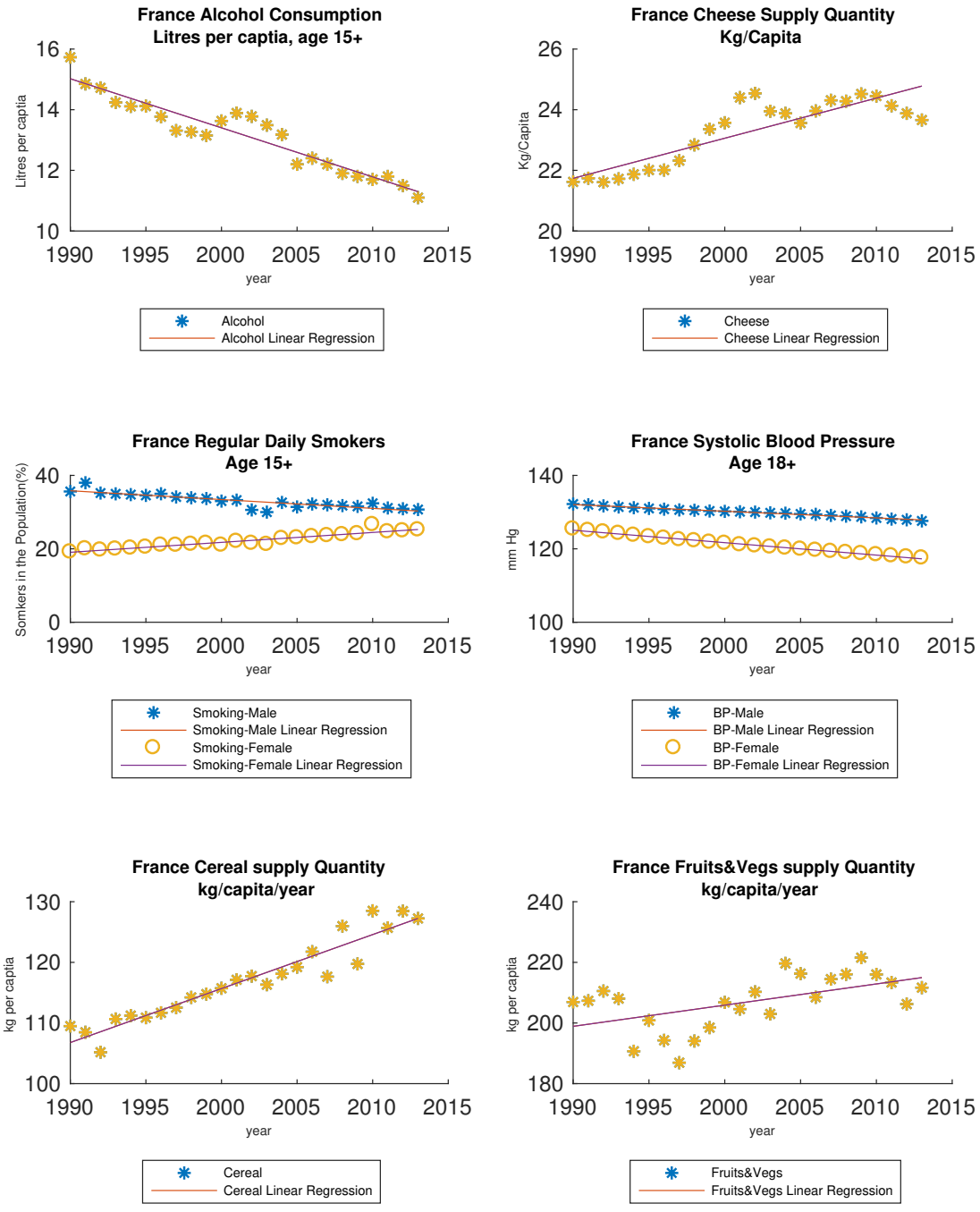


Figure C.1.1: France Linear plot and Trendline of 6 parameters

Table C.1.1: Tables A-D: Regression and ANOVA test Statistics of 6 Life-style parameters of France

Table A: Alcohol and Cheese consumptions

Regression Statistics				*df*	*F*	*Significance F*
	Alcohol	*Cheese*	Regression-Alcohol	1	204.688	1.268E-12
Multiple R	0.950	0.861	Residual-Alcohol	22		
R Square	0.903	0.742	Total-Alcohol	23		
Adjusted R2	0.899	0.730	Regression-Cheese	1	63.289	6.476E-08
Standard Error	0.384	0.564	Residual-Cheese	22		
Observations	24	24	Total-Cheese	23		

Table B: Regular daily smokers

Regression Statistics				*df*	*F*	*Significance F*
	Males	*Females*	Regression-male	1	64.620	5.449E-08
Multiple R	0.864	0.954	Residual-male	22		
R Square	0.746	0.910	Total-male	23		
Adjusted R2	0.734	0.906	Regression-female	1	222.125	5.590E-13
Standard Error	1.013	0.614	Residual-female	22		
Observations	24	24	Total-female	23		

Table C: Mean systolic blood pressure

Regression Statistics				*df*	*F*	*Significance F*
	Males	*Females*	Regression-male	1	1688.074	2.703E-22
Multiple R	0.994	0.997	Residual-male	22		
R Square	0.987	0.995	Total-male	23		
Adjusted R2	0.987	0.995	Regression-female	1	4197.831	1.304E-26
Standard Error	0.150	0.177	Residual-female	22		
Observations	24	24	Total-female	23		

Table D: Cereals, fruits and vegetables supply quantities

Regression Statistics				*df*	*F*	*Significance F*
	Cereals	*Fruits&vegs*	Regression-Cereals	1	215.210	7.679E-13
Multiple R	0.952	0.544	Residual-Cereals	22		
R Square	0.907	0.296	Total-Cereals	23		
Adjusted R2	0.903	0.264	Regression-F&V	1	9.254	0.006
Standard Error	2.058	7.794	Residual-F&V	22		
Observations	24	24	Total-F&V	23		

Greece

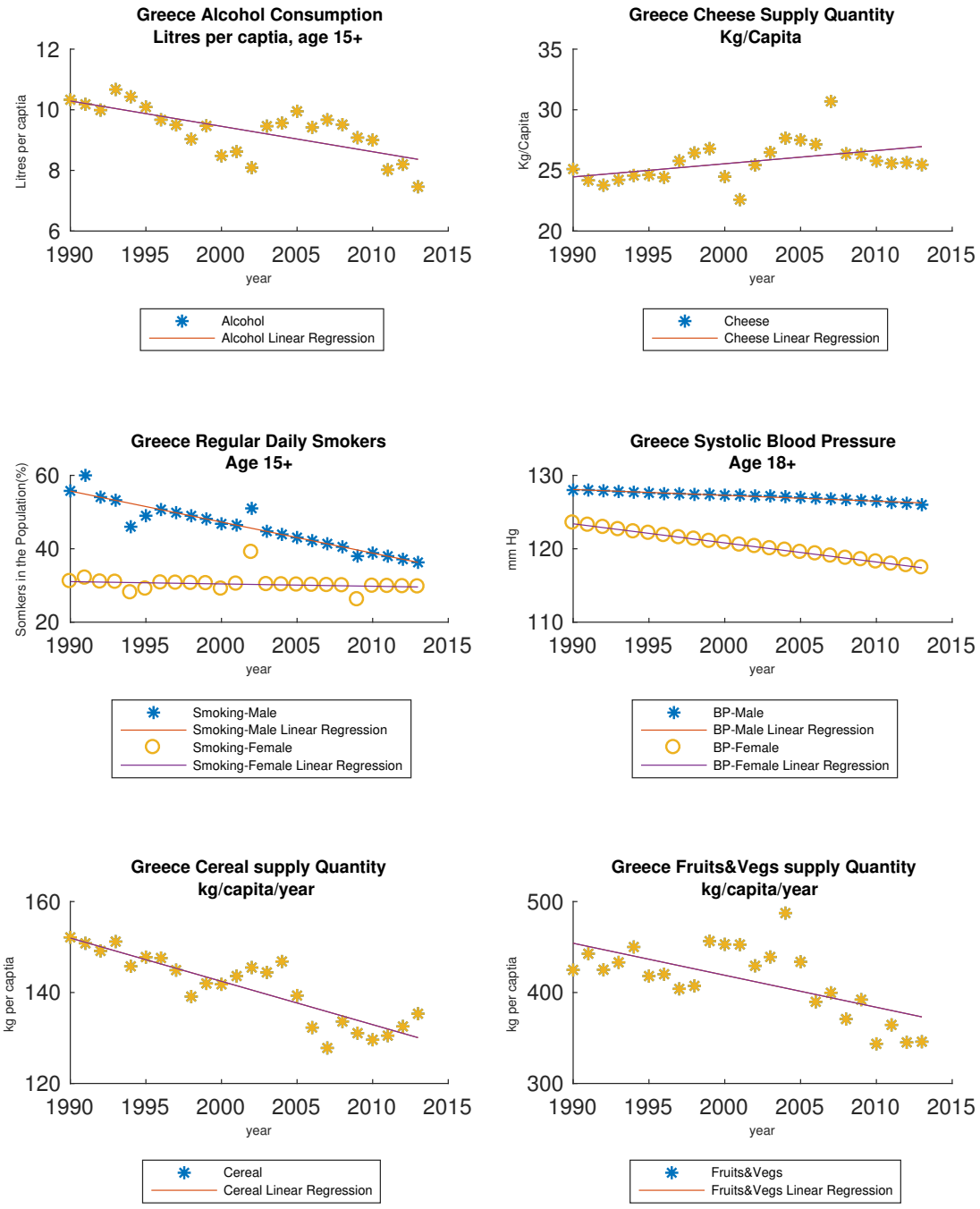


Figure C.1.2: Greece Linear plot and Trendline of 6 parameters

Table C.1.2: Tables A-D: Regression and ANOVA test Statistics of 6 Life-style parameters of Greece

Table A: Alcohol and Cheese consumptions

Regression Statistics				*df*	*F*	*Significance F*
	Alcohol	*Cheese*	Regression-Alcohol	1	22.277	1.040E-04
Multiple R	0.709	0.571	Residual-Alcohol	22		
R Square	0.503	0.326	Total-Alcohol	23		
Adjusted R2	0.481	0.295	Regression-Cheese	1	10.644	0.004
Standard Error	0.602	0.884	Residual-Cheese	22		
Observations	24	24	Total-Cheese	23		

Table B: Regular daily smokers

Regression Statistics				*df*	*F*	*Significance F*
	Males	*Females*	Regression-male	1	172.007	7.084E-12
Multiple R	0.942	0.959	Residual-male	22		
R Square	0.887	0.920	Total-male	23		
Adjusted R2	0.881	0.916	Regression-female	1	251.514	1.593E-13
Standard Error	2.185	0.528	Residual-female	22		
Observations	24	24	Total-female	23		

Table C: Mean systolic blood pressure

Regression Statistics				*df*	*F*	*Significance F*
	Males	*Females*	Regression-male	1	751.066	1.690E-18
Multiple R	0.986	1.000	Residual-male	22		
R Square	0.972	0.999	Total-male	23		
Adjusted R2	0.970	0.999	Regression-female	1	40556.402	2.001E-37
Standard Error	0.098	0.044	Residual-female	22		
Observations	24	24	Total-female	23		

Table D: Cereals, fruits and vegetables supply quantities

Regression Statistics				*df*	*F*	*Significance F*
	Cereals	*Fruits&vegs*	Regression-Cereals	1	76.273	1.335E-08
Multiple R	0.881	0.645	Residual-Cereals	22		
R Square	0.776	0.416	Total-Cereals	23		
Adjusted R2	0.766	0.389	Regression-F&V	1	15.649	6.716E-04
Standard Error	3.700	30.188	Residual-F&V	22		
Observations	24	24	Total-F&V	23		

Italy

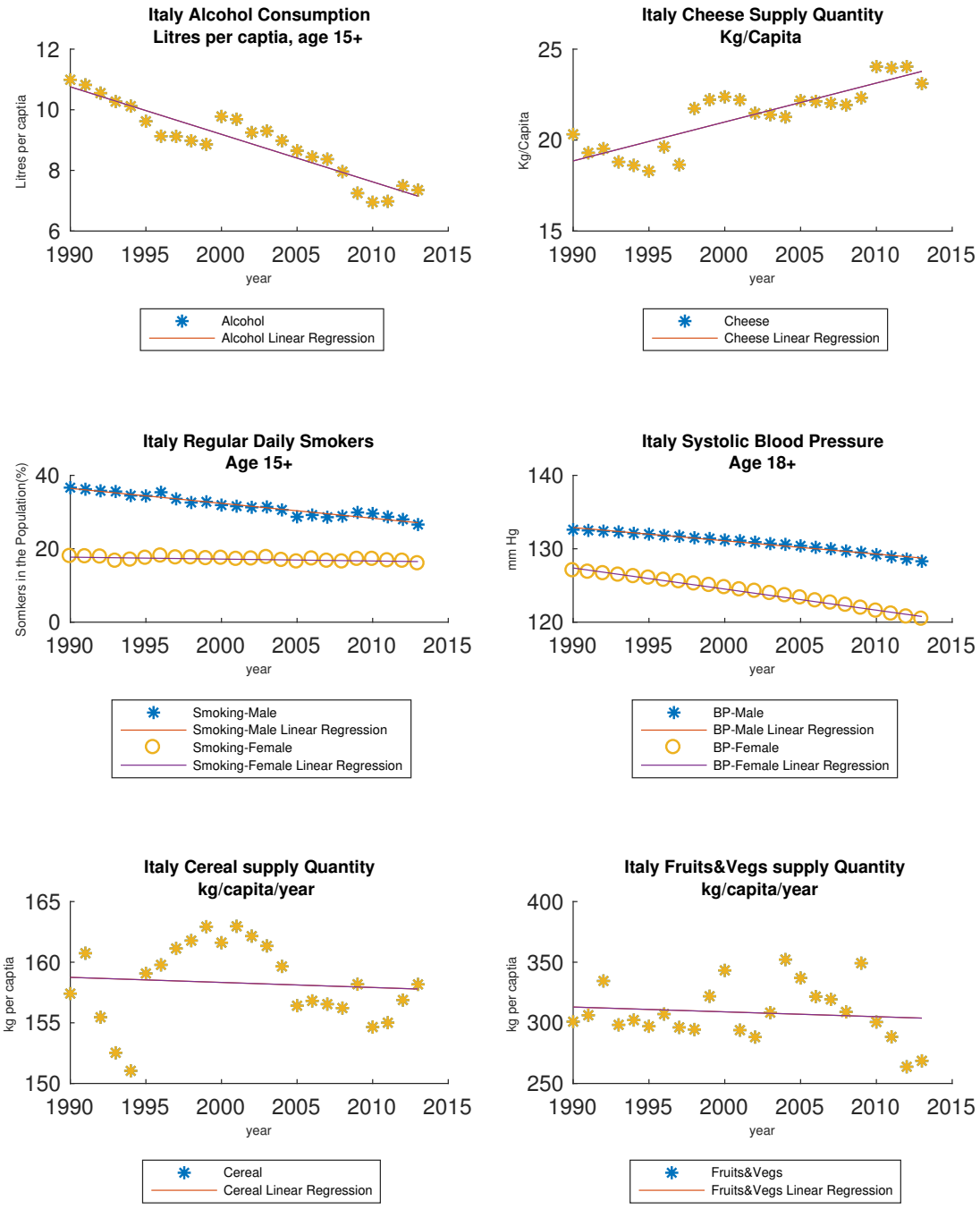


Figure C.1.3: Italy Linear plot and Trendline of 6 parameters

Table C.1.3: Tables A-D: Regression and ANOVA test Statistics of 6 Life-style parameters of Italy

Table A: Alcohol and Cheese consumptions

Regression Statistics				*df*	*F*	*Significance F*
	Alcohol	*Cheese*	Regression-Alcohol	1	149.635	2.748E-11
Multiple R	0.934	0.854	Residual-Alcohol	22		
R Square	0.872	0.729	Total-Alcohol	23		
Adjusted R2	0.866	0.717	Regression-Cheese	1	59.174	1.125E-07
Standard Error	0.435	0.944	Residual-Cheese	22		
Observations	24	24	Total-Cheese	23		

Table B: Regular daily smokers

Regression Statistics				*df*	*F*	*Significance F*
	Males	*Females*	Regression-male	1	363.652	3.597E-15
Multiple R	0.971	0.687	Residual-male	22		
R Square	0.943	0.472	Total-male	23		
Adjusted R2	0.940	0.448	Regression-female	1	19.660	2.090E-04
Standard Error	0.724	0.397	Residual-female	22		
Observations	24	24	Total-female	23		

Table C: Mean systolic blood pressure

Regression Statistics				*df*	*F*	*Significance F*
	Males	*Females*	Regression-male	1	1026.456	5.897E-20
Multiple R	0.989	0.994	Residual-male	22		
R Square	0.979	0.989	Total-male	23		
Adjusted R2	0.978	0.988	Regression-female	1	1964.077	5.208E-23
Standard Error	0.191	0.220	Residual-female	22		
Observations	24	24	Total-female	23		

Table D: Cereals, fruits and vegetables supply quantities

Regression Statistics				*df*	*F*	*Significance F*
	Cereals	*Fruits&vegs*	Regression-Cereals	1	0.183	6.726E-01
Multiple R	0.091	0.123	Residual-Cereals	22		
R Square	0.008	0.015	Total-Cereals	23		
Adjusted R2	-0.037	-0.030	Regression-F&V	1	0.335	5.684E-01
Standard Error	3.310	23.264	Residual-F&V	22		
Observations	24	24	Total-F&V	23		

Spain

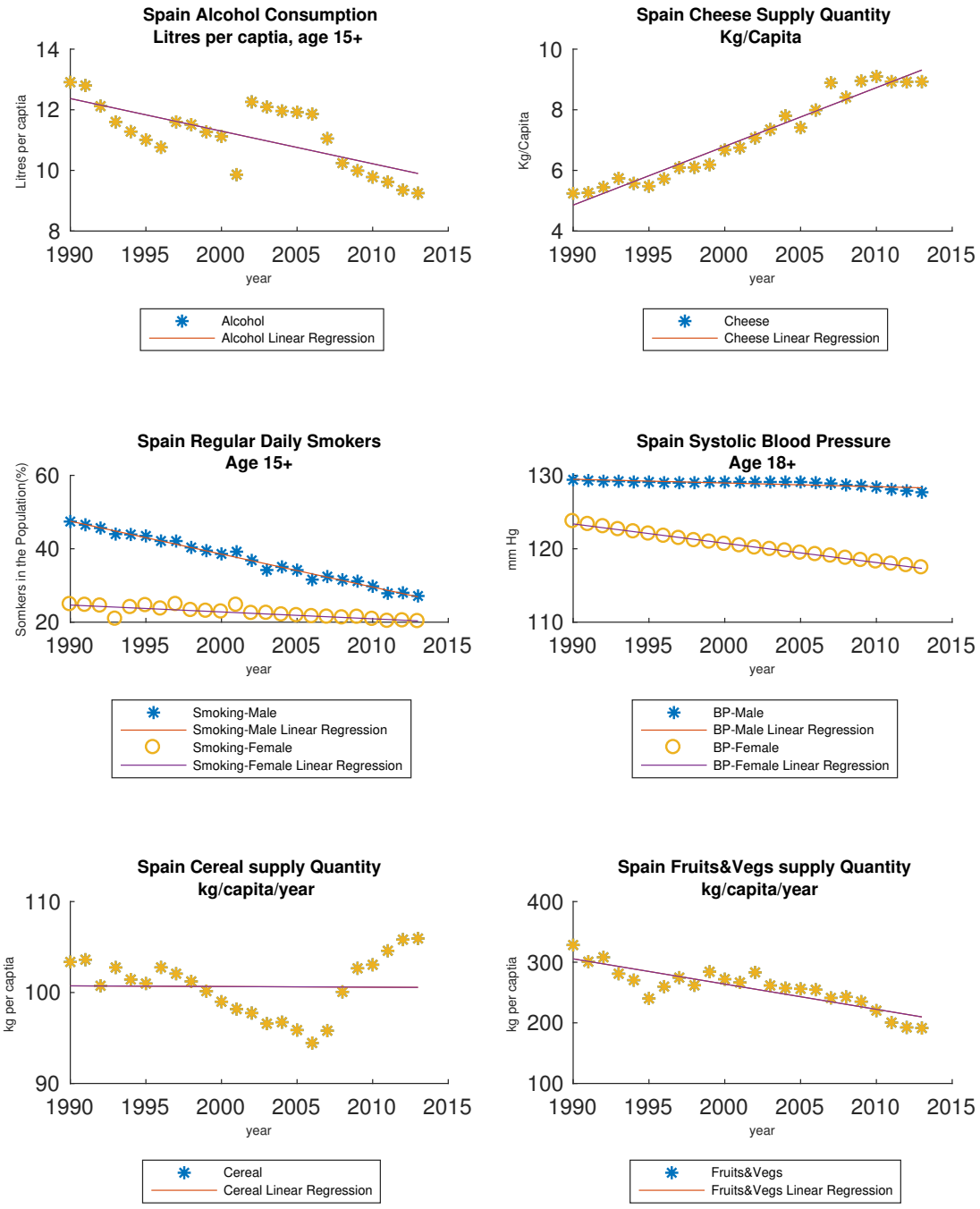


Figure C.1.4: Spain Linear plot and Trendline of 6 parameters

Table C.1.4: Tables A-D: Regression and ANOVA test Statistics of 6 Life-style parameters of Spain

Table A: Alcohol and Cheese consumptions

Regression Statistics				*df*	*F*	*Significance F*
	Alcohol	*Cheese*	Regression-Alcohol	1	22.519	9.777E-05
Multiple R	0.711	0.976	Residual-Alcohol	22		
R Square	0.506	0.953	Total-Alcohol	23		
Adjusted R2	0.483	0.951	Regression-Cheese	1	448.198	4.045E-16
Standard Error	0.768	0.310	Residual-Cheese	22		
Observations	24	24	Total-Cheese	23		

Table B: Regular daily smokers

Regression Statistics				*df*	*F*	*Significance F*
	Males	*Females*	Regression-male	1	1909.996	7.058E-23
Multiple R	0.994	0.830	Residual-male	22		
R Square	0.989	0.690	Total-male	23		
Adjusted R2	0.988	0.676	Regression-female	1	48.892	5.114E-07
Standard Error	0.696	0.908	Residual-female	22		
Observations	24	24	Total-female	23		

Table C: Mean systolic blood pressure

Regression Statistics				*df*	*F*	*Significance F*
	Males	*Females*	Regression-male	1	42.524	1.469E-06
Multiple R	0.812	0.998	Residual-male	22		
R Square	0.659	0.996	Total-male	23		
Adjusted R2	0.644	0.996	Regression-female	1	5436.453	7.682E-28
Standard Error	0.263	0.121	Residual-female	22		
Observations	24	24	Total-female	23		

Table D: Cereals, fruits and vegetables supply quantities

Regression Statistics				*df*	*F*	*Significance F*
	Cereals	*Fruits&vegs*	Regression-Cereals	1	0.006	9.411E-01
Multiple R	0.016	0.864	Residual-Cereals	22		
R Square	0.000	0.746	Total-Cereals	23		
Adjusted R2	-0.045	0.734	Regression-F&V	1	64.584	5.474E-08
Standard Error	3.321	17.572	Residual-F&V	22		
Observations	24	24	Total-F&V	23		

C.2 SCANDINAVIAN EUROPEAN COUNTRIES (SCEU) BLOCK

Denmark

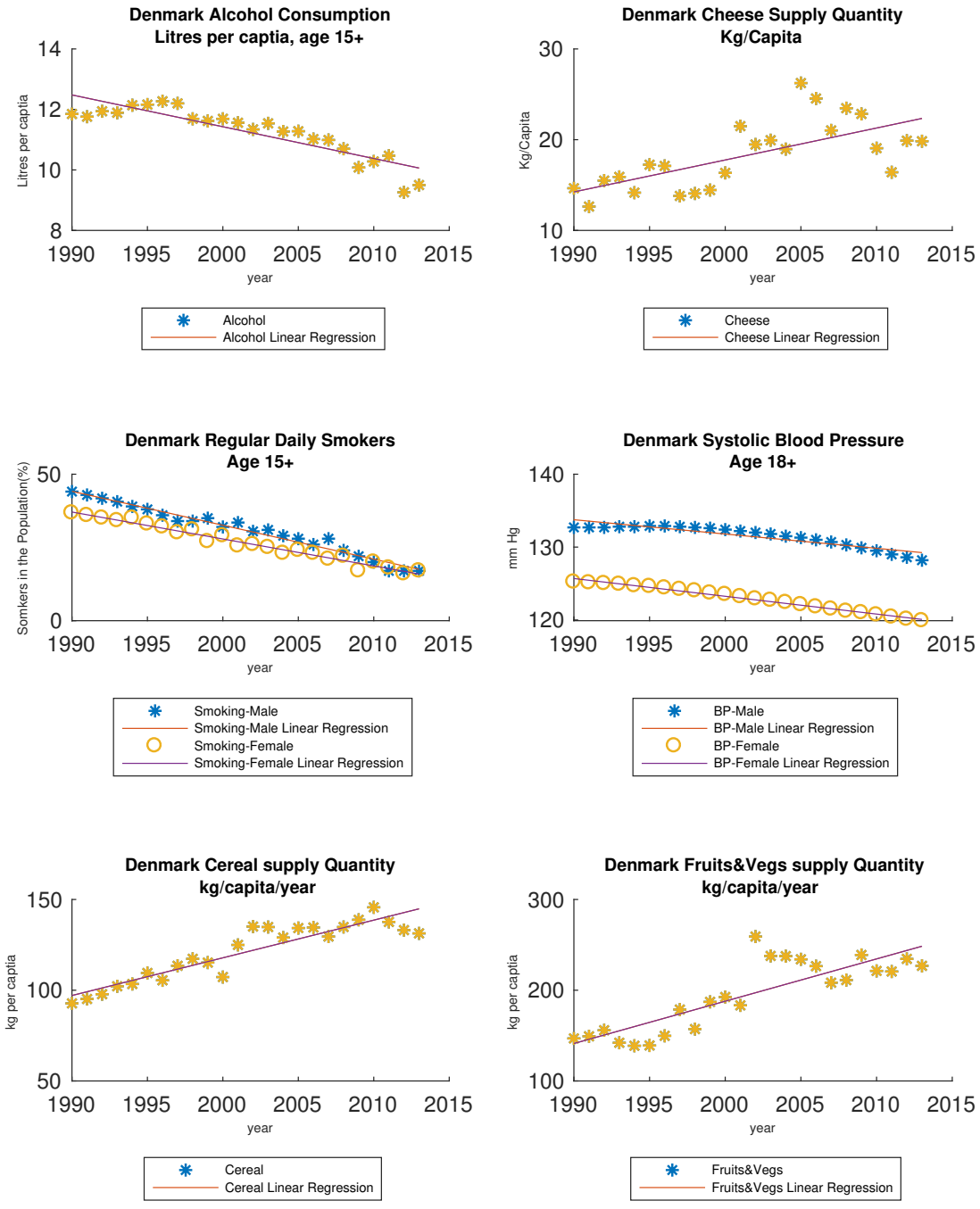


Figure C.2.1: Denmark Linear plot and Trendline of 6 parameters

Table C.2.1: Tables A-D: Regression and ANOVA test Statistics of 6 Life-style parameters of Denmark

Table A: Alcohol and Cheese consumptions

Regression Statistics				*df*	*F*	*Significance F*
	Alcohol	*Cheese*	Regression-Alcohol	1	81.372	7.607E-09
Multiple R	0.887	0.672	Residual-Alcohol	22		
R Square	0.787	0.452	Total-Alcohol	23		
Adjusted R2	0.778	0.427	Regression-Cheese	1	18.112	0.000
Standard Error	0.395	2.807	Residual-Cheese	22		
Observations	24	24	Total-Cheese	23		

Table B: Regular daily smokers

Regression Statistics				*df*	*F*	*Significance F*
	Males	*Females*	Regression-male	1	759.941	1.490E-18
Multiple R	0.986	0.986	Residual-male	22		
R Square	0.972	0.973	Total-male	23		
Adjusted R2	0.971	0.972	Regression-female	1	788.482	1.004E-18
Standard Error	1.424	1.110	Residual-female	22		
Observations	24	24	Total-female	23		

Table C: Mean systolic blood pressure

Regression Statistics				*df*	*F*	*Significance F*
	Males	*Females*	Regression-male	1	123.727	1.682E-10
Multiple R	0.921	0.993	Residual-male	22		
R Square	0.849	0.987	Total-male	23		
Adjusted R2	0.842	0.986	Regression-female	1	1641.754	3.657E-22
Standard Error	0.598	0.204	Residual-female	22		
Observations	24	24	Total-female	23		

Table D: Cereals, fruits and vegetables supply quantities

Regression Statistics				*df*	*F*	*Significance F*
	Cereals	*Fruits&vegs*	Regression-Cereals	1	118.103	2.602E-10
Multiple R	0.918	0.839	Residual-Cereals	22		
R Square	0.843	0.703	Total-Cereals	23		
Adjusted R2	0.836	0.690	Regression-F&V	1	52.125	3.104E-07
Standard Error	6.481	21.895	Residual-F&V	22		
Observations	24	24	Total-F&V	23		

Finland

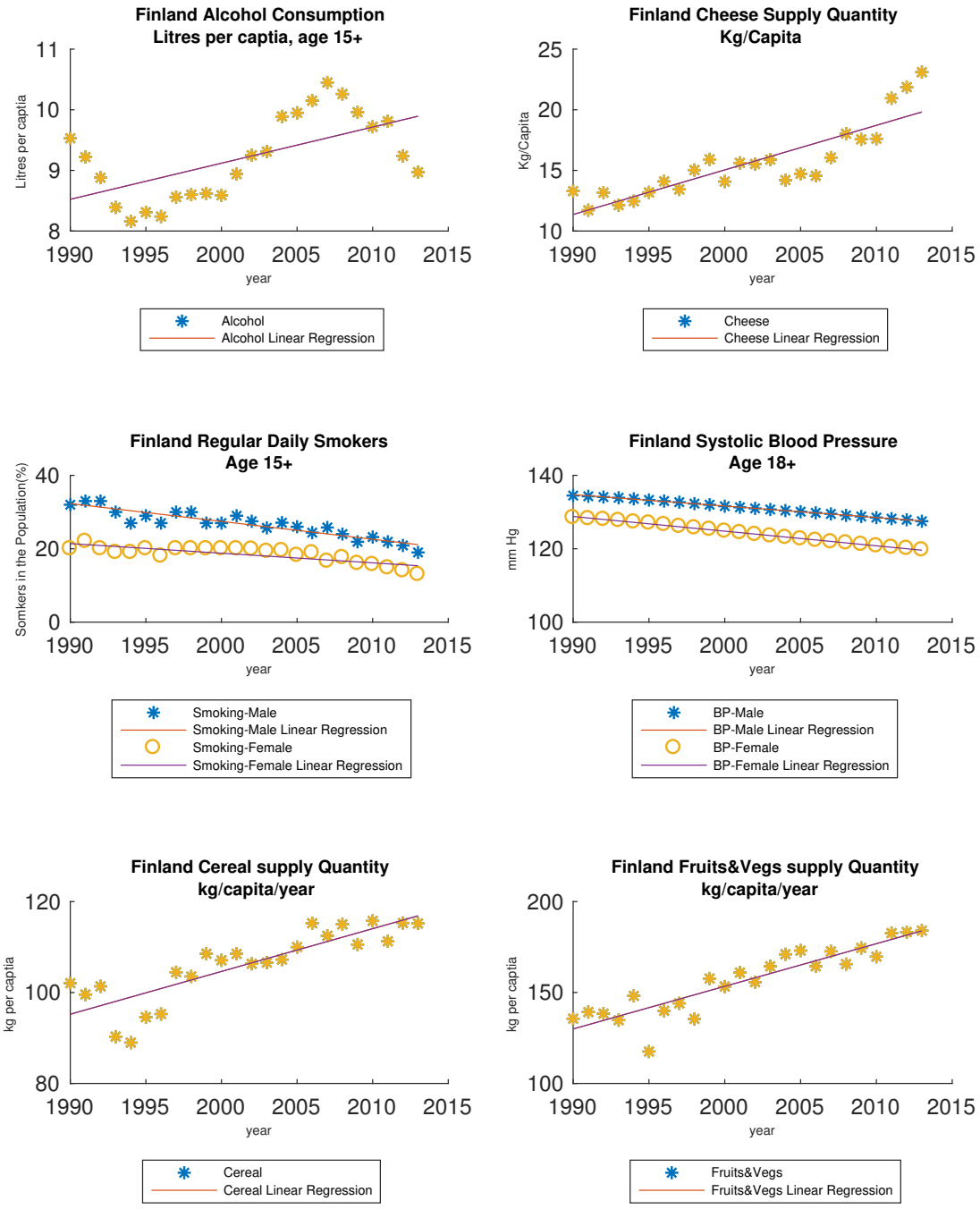


Figure C.2.2: Finland Linear plot and Trendline of 6 parameters

Table C.2.2: Tables A-D: Regression and ANOVA test Statistics of 6 Life-style parameters of Finland

Table A: Alcohol and Cheese consumptions

Regression Statistics				*df*	*F*	*Significance F*
	Alcohol	*Cheese*	Regression-Alcohol	1	12.762	0.002
Multiple R	0.606	0.869	Residual-Alcohol	22		
R Square	0.367	0.755	Total-Alcohol	23		
Adjusted R2	0.338	0.744	Regression-Cheese	1	67.816	0.000
Standard Error	0.566	1.514	Residual-Cheese	22		
Observations	24	24	Total-Cheese	23		

Table B: Regular daily smokers

Regression Statistics				*df*	*F*	*Significance F*
	Males	*Females*	Regression-male	1	128.529	1.175E-10
Multiple R	0.924	0.821	Residual-male	22		
R Square	0.854	0.673	Total-male	23		
Adjusted R2	0.847	0.658	Regression-female	1	45.333	9.109E-07
Standard Error	1.453	1.320	Residual-female	22		
Observations	24	24	Total-female	23		

Table C: Mean systolic blood pressure

Regression Statistics				*df*	*F*	*Significance F*
	Males	*Females*	Regression-male	1	15146.962	1.005E-32
Multiple R	0.999	0.999	Residual-male	22		
R Square	0.999	0.998	Total-male	23		
Adjusted R2	0.998	0.998	Regression-female	1	11883.959	1.443E-31
Standard Error	0.086	0.124	Residual-female	22		
Observations	24	24	Total-female	23		

Table D: Cereals, fruits and vegetables supply quantities

Regression Statistics				*df*	*F*	*Significance F*
	Cereals	*Fruits&vegs*	Regression-Cereals	1	55.262	1.953E-07
Multiple R	0.846	0.910	Residual-Cereals	22		
R Square	0.715	0.829	Total-Cereals	23		
Adjusted R2	0.702	0.821	Regression-F&V	1	106.478	6.799E-10
Standard Error	4.291	7.704	Residual-F&V	22		
Observations	24	24	Total-F&V	23		

Iceland

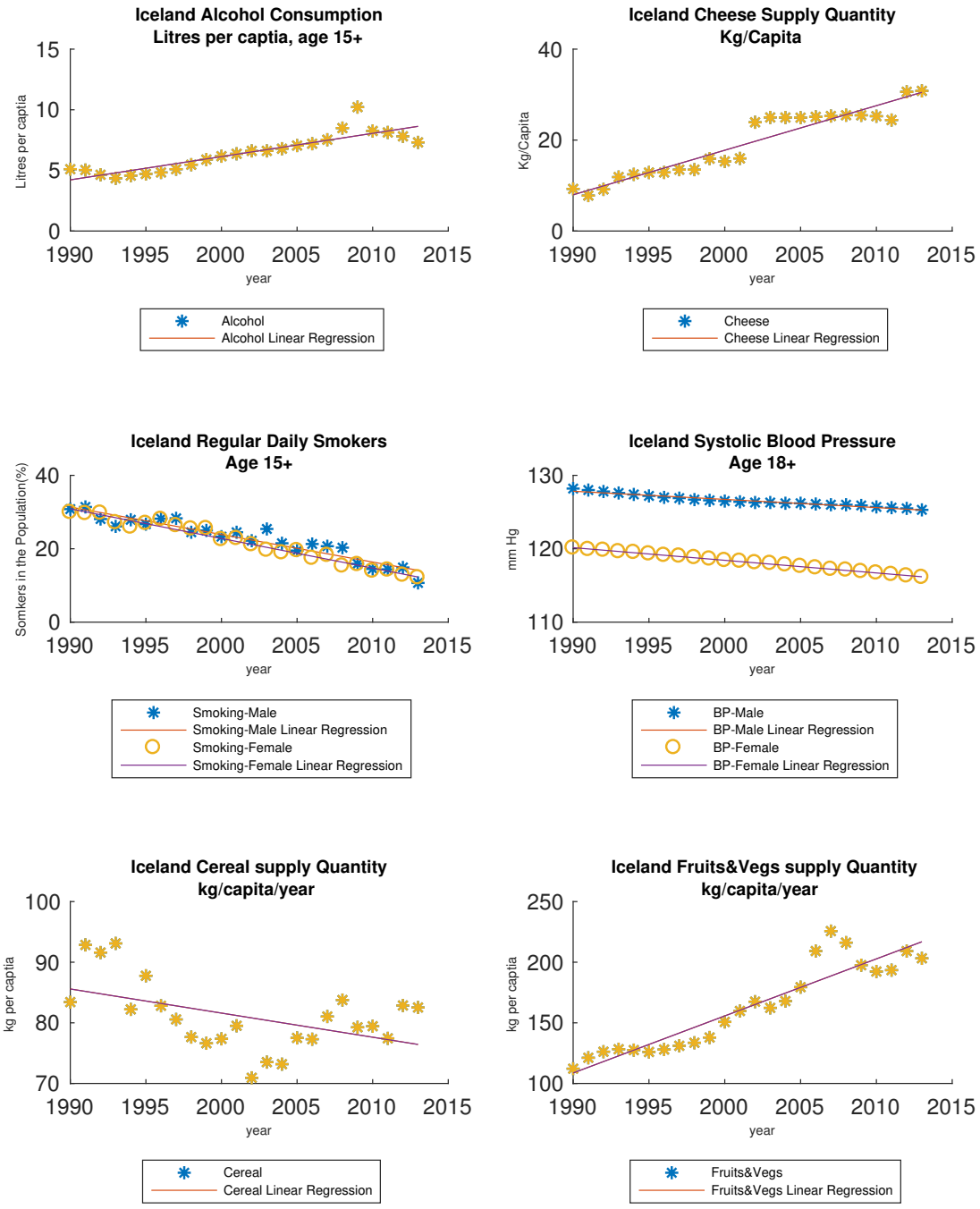


Figure C.2.3: Iceland Linear plot and Trendline of 6 parameters

Table C.2.3: Tables A-D: Regression and ANOVA test Statistics of 6 Life-style parameters of Iceland

Table A: Alcohol and Cheese consumptions

Regression Statistics				*df*	*F*	*Significance F*
	Alcohol	*Cheese*	Regression-Alcohol	1	86.291	4.537E-09
Multiple R	0.893	0.955	Residual-Alcohol	22		
R Square	0.797	0.912	Total-Alcohol	23		
Adjusted R2	0.788	0.908	Regression-Cheese	1	227.660	4.364E-13
Standard Error	0.701	2.202	Residual-Cheese	22		
Observations	24	24	Total-Cheese	23		

Table B: Regular daily smokers

Regression Statistics				*df*	*F*	*Significance F*
	Males	*Females*	Regression-male	1	199.729	1.619E-12
Multiple R	0.949	0.986	Residual-male	22		
R Square	0.901	0.972	Total-male	23		
Adjusted R2	0.896	0.971	Regression-female	1	763.850	1.410E-18
Standard Error	1.798	0.991	Residual-female	22		
Observations	24	24	Total-female	23		

Table C: Mean systolic blood pressure

Regression Statistics				*df*	*F*	*Significance F*
	Males	*Females*	Regression-male	1	464.564	2.774E-16
Multiple R	0.977	0.999	Residual-male	22		
R Square	0.955	0.999	Total-male	23		
Adjusted R2	0.953	0.999	Regression-female	1	15820.910	6.231E-33
Standard Error	0.175	0.046	Residual-female	22		
Observations	24	24	Total-female	23		

Table D: Cereals, fruits and vegetables supply quantities

Regression Statistics				*df*	*F*	*Significance F*
	Cereals	*Fruits&vegs*	Regression-Cereals	1	6.725	1.659E-02
Multiple R	0.484	0.931	Residual-Cereals	22		
R Square	0.234	0.867	Total-Cereals	23		
Adjusted R2	0.199	0.861	Regression-F&V	1	143.991	3.978E-11
Standard Error	5.202	13.297	Residual-F&V	22		
Observations	24	24	Total-F&V	23		

Norway

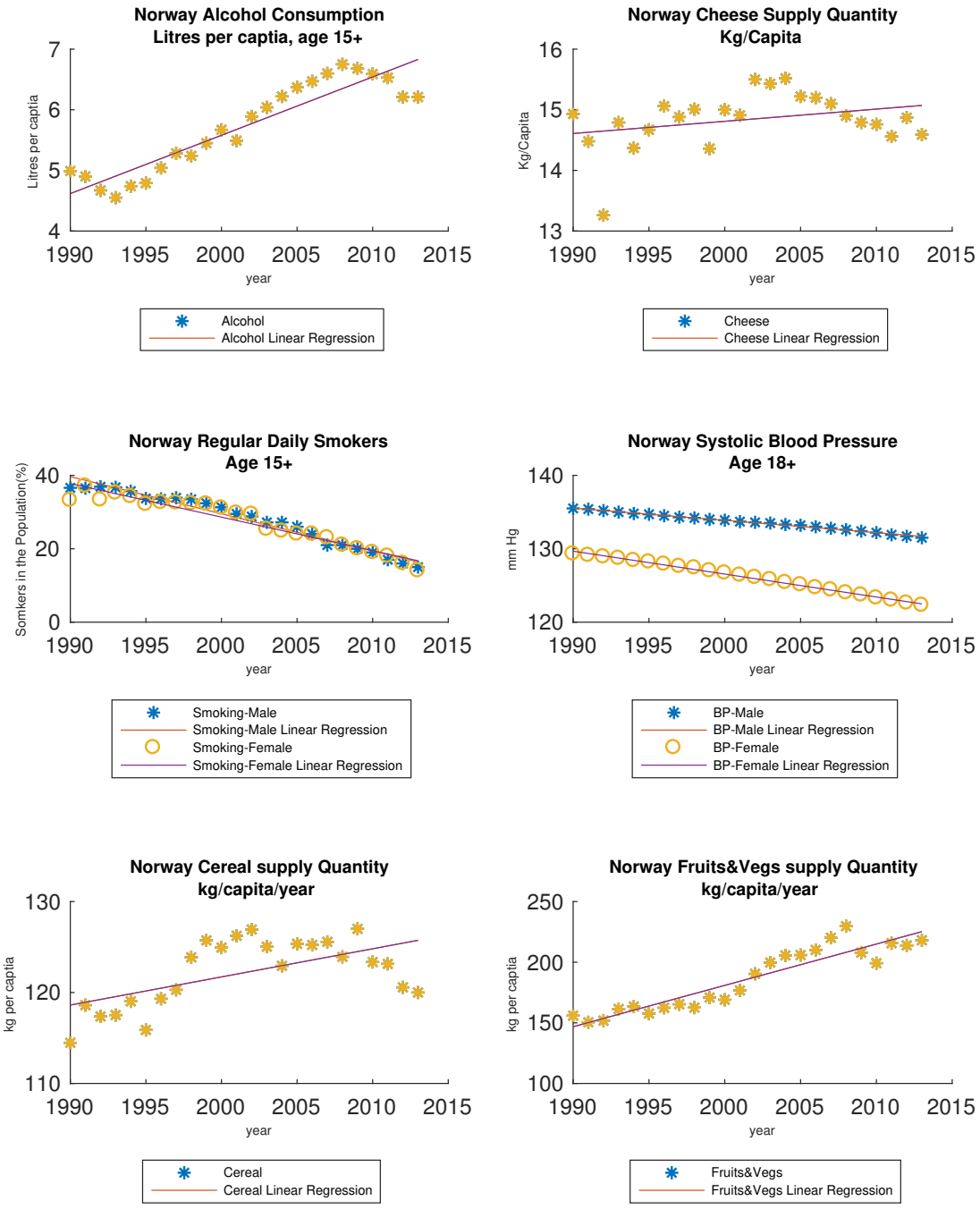


Figure C.2.4: Norway Linear plot and Trendline of 6 parameters

Table C.2.4: Tables A-D: Regression and ANOVA test Statistics of 6 Life-style parameters of Norway

Table A: Alcohol and Cheese consumptions

Regression Statistics				*df*	*F*	*Significance F*
	Alcohol	*Cheese*	Regression-Alcohol	1	124.244	1.617E-10
Multiple R	0.922	0.306	Residual-Alcohol	22		
R Square	0.850	0.094	Total-Alcohol	23		
Adjusted R2	0.843	0.053	Regression-Cheese	1	2.279	0.145
Standard Error	0.293	0.452	Residual-Cheese	22		
Observations	24	24	Total-Cheese	23		

Table B: Regular daily smokers

Regression Statistics				*df*	*F*	*Significance F*
	Males	*Females*	Regression-male	1	582.133	2.555E-17
Multiple R	0.982	0.965	Residual-male	22		
R Square	0.964	0.932	Total-male	23		
Adjusted R2	0.962	0.928	Regression-female	1	299.410	2.685E-14
Standard Error	1.422	1.792	Residual-female	22		
Observations	24	24	Total-female	23		

Table C: Mean systolic blood pressure

Regression Statistics				*df*	*F*	*Significance F*
	Males	*Females*	Regression-male	1	3682.477	5.466E-26
Multiple R	0.997	0.997	Residual-male	22		
R Square	0.994	0.995	Total-male	23		
Adjusted R2	0.994	0.995	Regression-female	1	4249.982	1.139E-26
Standard Error	0.094	0.162	Residual-female	22		
Observations	24	24	Total-female	23		

Table D: Cereals, fruits and vegetables supply quantities

Regression Statistics				*df*	*F*	*Significance F*
	Cereals	*Fruits&vegs*	Regression-Cereals	1	11.777	2.383E-03
Multiple R	0.590	0.932	Residual-Cereals	22		
R Square	0.349	0.869	Total-Cereals	23		
Adjusted R2	0.319	0.863	Regression-F&V	1	146.484	3.373E-11
Standard Error	3.059	9.550	Residual-F&V	22		
Observations	24	24	Total-F&V	23		

Sweden

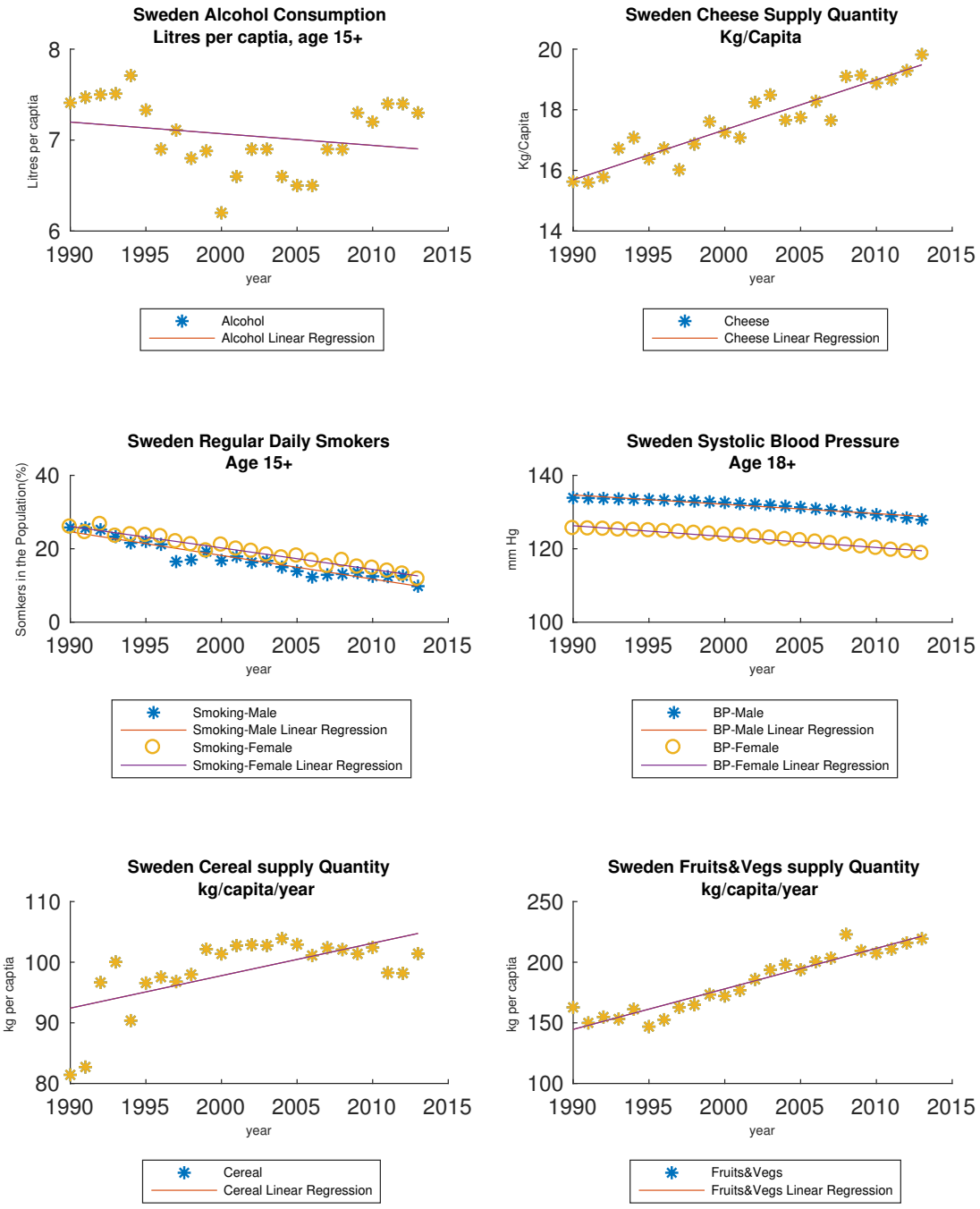


Figure C.2.5: Sweden Linear plot and Trendline of 6 parameters

Table C.2.5: Tables A-D: Regression and ANOVA test Statistics of 6 Life-style parameters of Sweden

Table A: Alcohol and Cheese consumptions

Regression Statistics				*df*	*F*	*Significance F*
	Alcohol	*Cheese*	Regression-Alcohol	1	1.232	2.789E-01
Multiple R	0.230	0.939	Residual-Alcohol	22		
R Square	0.053	0.881	Total-Alcohol	23		
Adjusted R2	0.010	0.875	Regression-Cheese	1	162.548	1.231E-11
Standard Error	0.392	0.439	Residual-Cheese	22		
Observations	24	24	Total-Cheese	23		

Table B: Regular daily smokers

Regression Statistics				*df*	*F*	*Significance F*
	Males	*Females*	Regression-male	1	232.302	3.561E-13
Multiple R	0.956	0.984	Residual-male	22		
R Square	0.913	0.969	Total-male	23		
Adjusted R2	0.910	0.968	Regression-female	1	692.444	4.027E-18
Standard Error	1.432	0.756	Residual-female	22		
Observations	24	24	Total-female	23		

Table C: Mean systolic blood pressure

Regression Statistics				*df*	*F*	*Significance F*
	Males	*Females*	Regression-male	1	340.123	7.202E-15
Multiple R	0.969	0.981	Residual-male	22		
R Square	0.939	0.963	Total-male	23		
Adjusted R2	0.936	0.961	Regression-female	1	564.850	3.518E-17
Standard Error	0.467	0.424	Residual-female	22		
Observations	24	24	Total-female	23		

Table D: Cereals, fruits and vegetables supply quantities

Regression Statistics				*df*	*F*	*Significance F*
	Cereals	*Fruits&vegs*	Regression-Cereals	1	15.035	8.128E-04
Multiple R	0.637	0.953	Residual-Cereals	22		
R Square	0.406	0.908	Total-Cereals	23		
Adjusted R2	0.379	0.904	Regression-F&V	1	217.496	6.907E-13
Standard Error	4.686	7.685	Residual-F&V	22		
Observations	24	24	Total-F&V	23		

C.3 WESTERN EUROPEAN COUNTRIES (WEEU) BLOCK

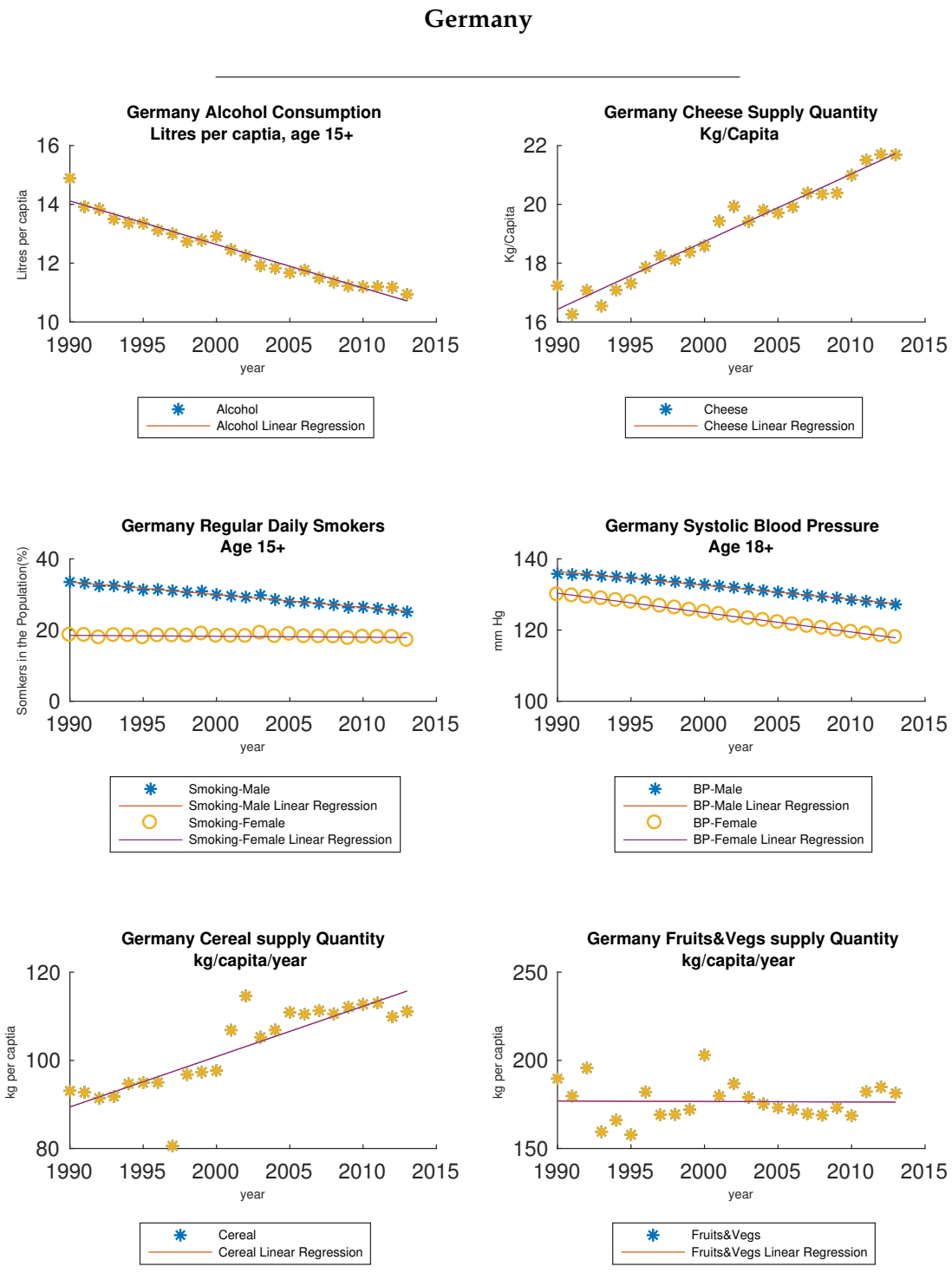


Figure C.3.1: Germany Linear plot and Trendline of 6 parameters

Table C.3.1: Tables A-D: Regression and ANOVA test Statistics of 6 Life-style parameters of Germany

Table A: Alcohol and Cheese consumptions

Regression Statistics				*df*	*F*	*Significance F*
	Alcohol	*Cheese*	Regression-Alcohol	1	467.290	2.608E-16
Multiple R	0.977	0.979	Residual-Alcohol	22		
R Square	0.955	0.959	Total-Alcohol	23		
Adjusted R2	0.953	0.957	Regression-Cheese	1	509.635	1.045E-16
Standard Error	0.232	0.346	Residual-Cheese	22		
Observations	24	24	Total-Cheese	23		

Table B: Regular daily smokers

Regression Statistics				*df*	*F*	*Significance F*
	Males	*Females*	Regression-male	1	1777.197	1.545E-22
Multiple R	0.994	0.990	Residual-male	22		
R Square	0.988	0.980	Total-male	23		
Adjusted R2	0.987	0.979	Regression-female	1	1088.123	3.143E-20
Standard Error	0.285	0.155	Residual-female	22		
Observations	24	24	Total-female	23		

Table C: Mean systolic blood pressure

Regression Statistics				*df*	*F*	*Significance F*
	Males	*Females*	Regression-male	1	2693.584	1.666E-24
Multiple R	0.996	0.999	Residual-male	22		
R Square	0.992	0.998	Total-male	23		
Adjusted R2	0.992	0.998	Regression-female	1	14594.135	1.512E-32
Standard Error	0.255	0.153	Residual-female	22		
Observations	24	24	Total-female	23		

Table D: Cereals, fruits and vegetables supply quantities

Regression Statistics				*df*	*F*	*Significance F*
	Cereals	*Fruits&vegs*	Regression-Cereals	1	59.249	1.113E-07
Multiple R	0.854	0.995	Residual-Cereals	22		
R Square	0.729	0.989	Total-Cereals	23		
Adjusted R2	0.717	0.989	Regression-F&V	1	2042.980	3.392E-23
Standard Error	5.036	2.027	Residual-F&V	22		
Observations	24	24	Total-F&V	23		

Netherlands

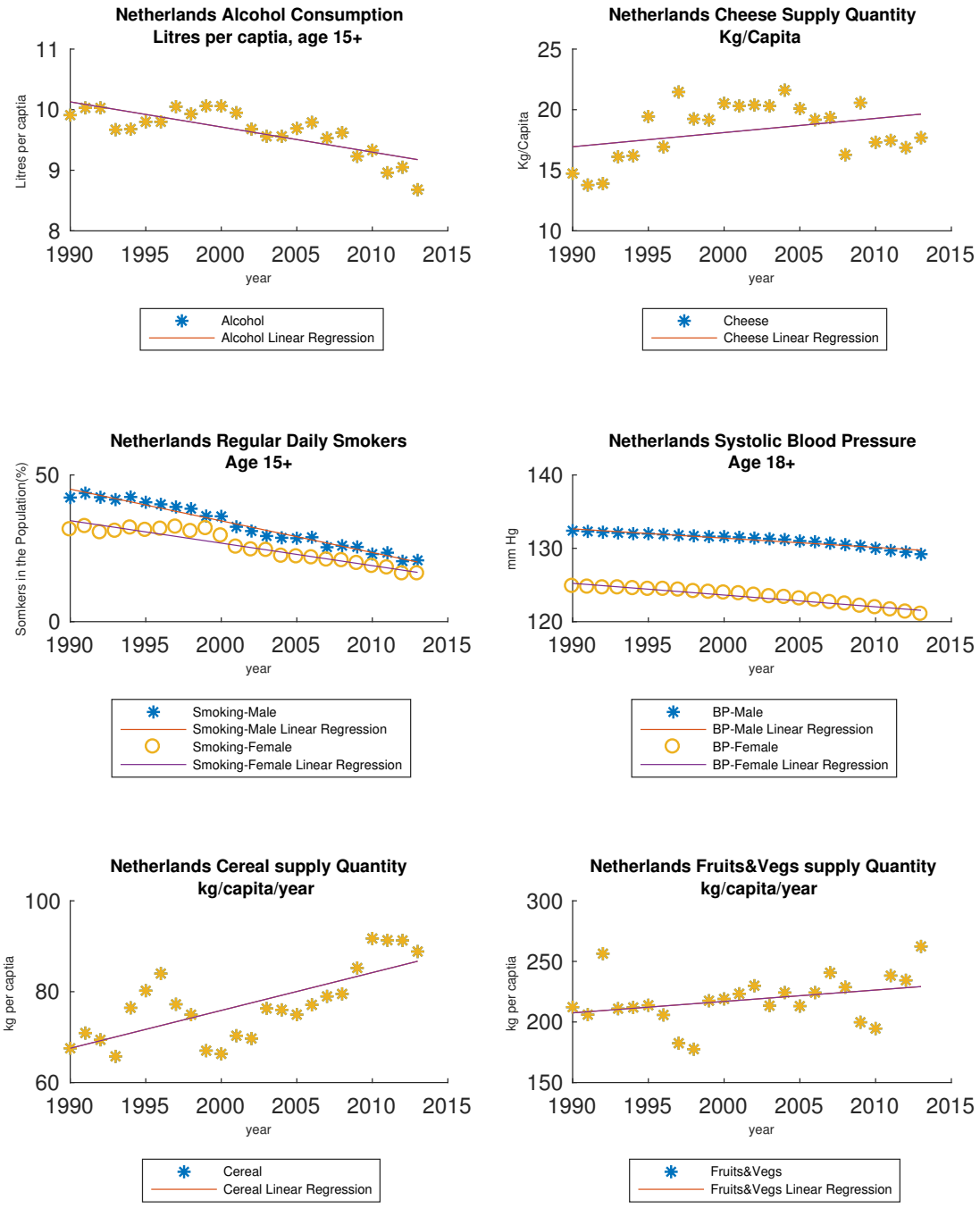


Figure C.3.2: Netherlands Linear plot and Trendline of 6 parameters

Table C.3.2: Tables A-D: Regression and ANOVA test Statistics of 6 Life-style parameters of Netherlands

Table A: Alcohol and Cheese consumptions

Regression Statistics				*df*	*F*	*Significance F*
	Alcohol	*Cheese*	Regression-Alcohol	1	36.190	4.697E-06
Multiple R	0.789	0.358	Residual-Alcohol	22		
R Square	0.622	0.128	Total-Alcohol	23		
Adjusted R2	0.605	0.088	Regression-Cheese	1	3.227	0.086
Standard Error	0.234	2.210	Residual-Cheese	22		
Observations	24	24	Total-Cheese	23		

Table B: Regular daily smokers

Regression Statistics				*df*	*F*	*Significance F*
	Males	*Females*	Regression-male	1	786.582	1.030E-18
Multiple R	0.986	0.953	Residual-male	22		
R Square	0.973	0.908	Total-male	23		
Adjusted R2	0.972	0.904	Regression-female	1	218.073	6.725E-13
Standard Error	1.306	1.760	Residual-female	22		
Observations	24	24	Total-female	23		

Table C: Mean systolic blood pressure

Regression Statistics				*df*	*F*	*Significance F*
	Males	*Females*	Regression-male	1	308.355	1.983E-14
Multiple R	0.966	0.972	Residual-male	22		
R Square	0.933	0.946	Total-male	23		
Adjusted R2	0.930	0.943	Regression-female	1	382.007	2.154E-15
Standard Error	0.243	0.278	Residual-female	22		
Observations	24	24	Total-female	23		

Table D: Cereals, fruits and vegetables supply quantities

Regression Statistics				*df*	*F*	*Significance F*
	Cereals	*Fruits&vegs*	Regression-Cereals	1	23.804	7.066E-05
Multiple R	0.721	0.332	Residual-Cereals	22		
R Square	0.520	0.110	Total-Cereals	23		
Adjusted R2	0.498	0.070	Regression-F&V	1	2.732	1.126E-01
Standard Error	5.789	19.339	Residual-F&V	22		
Observations	24	24	Total-F&V	23		

Switzerland

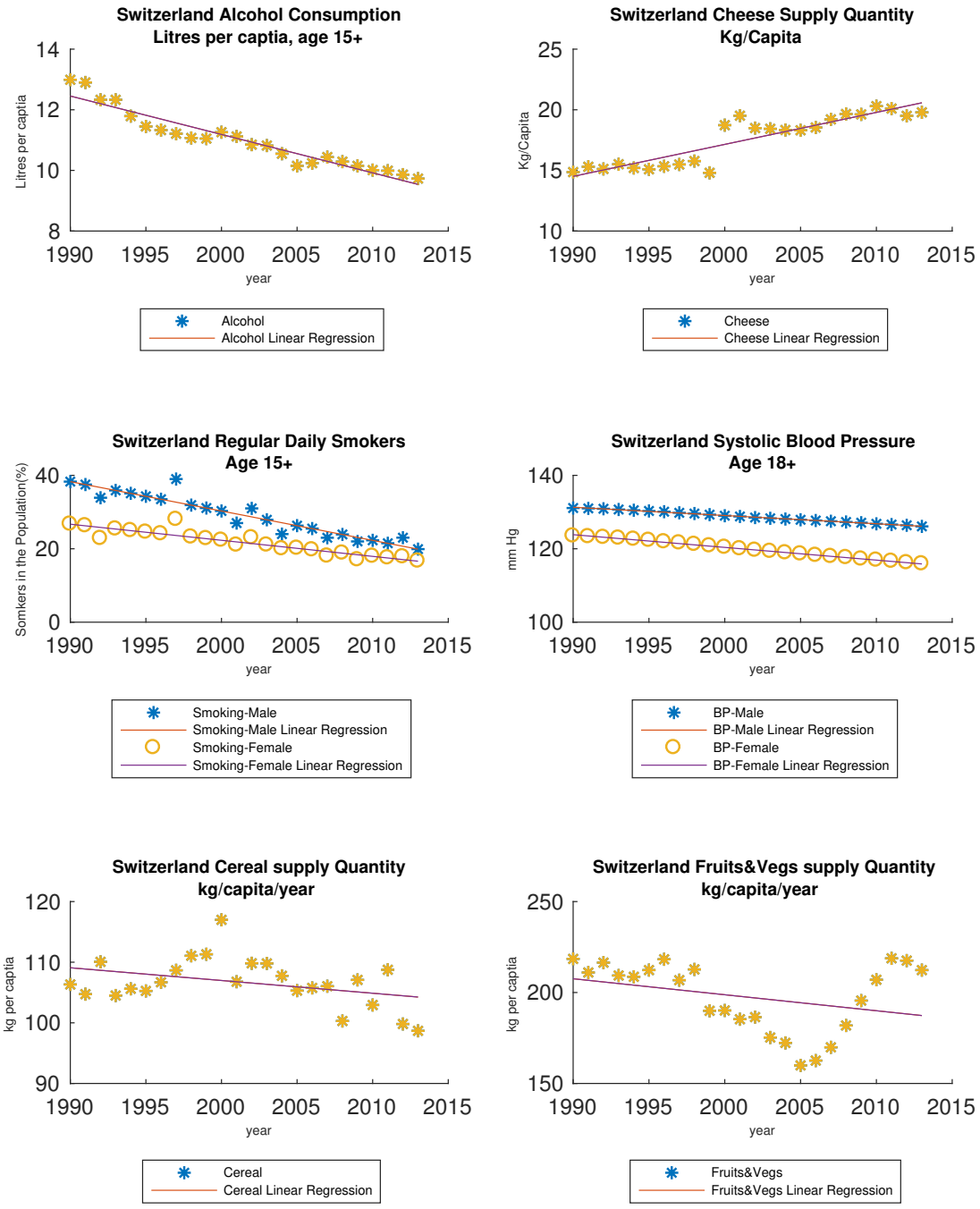


Figure C.3.3: Switzerland Linear plot and Trendline of 6 parameters

Table C.3.3: Tables A-D: Regression and ANOVA test Statistics of 6 Life-style parameters of Switzerland

Table A: Alcohol and Cheese consumptions

Regression Statistics				*df*	*F*	*Significance F*
	Alcohol	*Cheese*	Regression-Alcohol	1	236.045	3.030E-13
Multiple R	0.956	0.907	Residual-Alcohol	22		
R Square	0.915	0.823	Total-Alcohol	23		
Adjusted R2	0.911	0.815	Regression-Cheese	1	102.617	9.539E-10
Standard Error	0.280	0.883	Residual-Cheese	22		
Observations	24	24	Total-Cheese	23		

Table B: Regular daily smokers

Regression Statistics				*df*	*F*	*Significance F*
	Males	*Females*	Regression-male	1	0.002	9.615E-01
Multiple R	0.010	0.010	Residual-male	22		
R Square	0.000	0.000	Total-male	23		
Adjusted R2	-0.045	-0.045	Regression-female	1	0.002	9.614E-01
Standard Error	505.530	276.589	Residual-female	22		
Observations	24	24	Total-female	23		

Table C: Mean systolic blood pressure

Regression Statistics				*df*	*F*	*Significance F*
	Males	*Females*	Regression-male	1	8607.303	4.980E-30
Multiple R	0.999	0.999	Residual-male	22		
R Square	0.997	0.998	Total-male	23		
Adjusted R2	0.997	0.997	Regression-female	1	9016.521	2.991E-30
Standard Error	0.082	0.124	Residual-female	22		
Observations	24	24	Total-female	23		

Table D: Cereals, fruits and vegetables supply quantities

Regression Statistics				*df*	*F*	*Significance F*
	Cereals	*Fruits&vegs*	Regression-Cereals	1	3.512	7.427E-02
Multiple R	0.371	0.323	Residual-Cereals	22		
R Square	0.138	0.104	Total-Cereals	23		
Adjusted R2	0.098	0.064	Regression-F&V	1	2.560	1.238E-01
Standard Error	3.800	18.642	Residual-F&V	22		
Observations	24	24	Total-F&V	23		

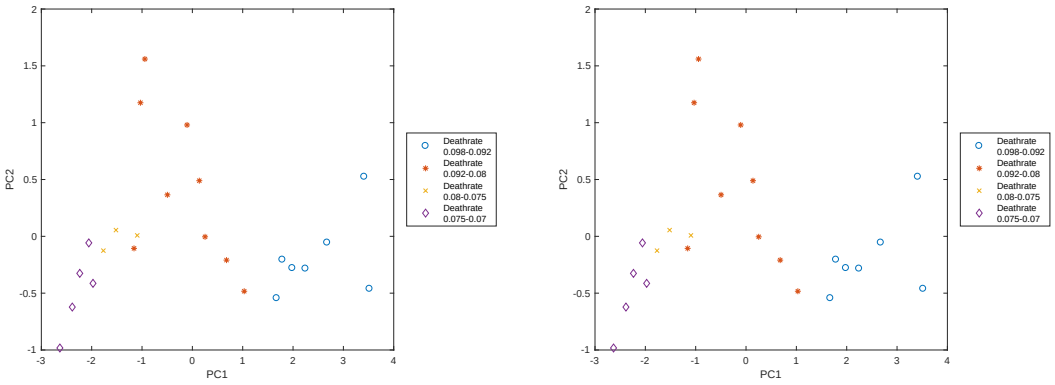
D

VISUALIZATION FOR 12 EUROPEAN COUNTRIES

Results of linear regression of 6 life-style parameters for 12 European country. Markers on the visualizations are assigned using the bins (shown in legends) of the death rates.

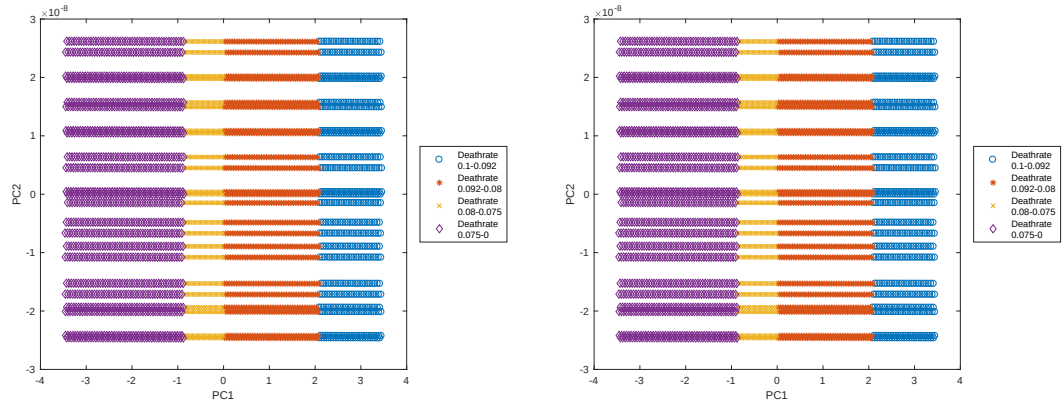
D.1 MEDITERRANEAN EUROPEAN COUNTRIES (MEEU) BLOCK

France

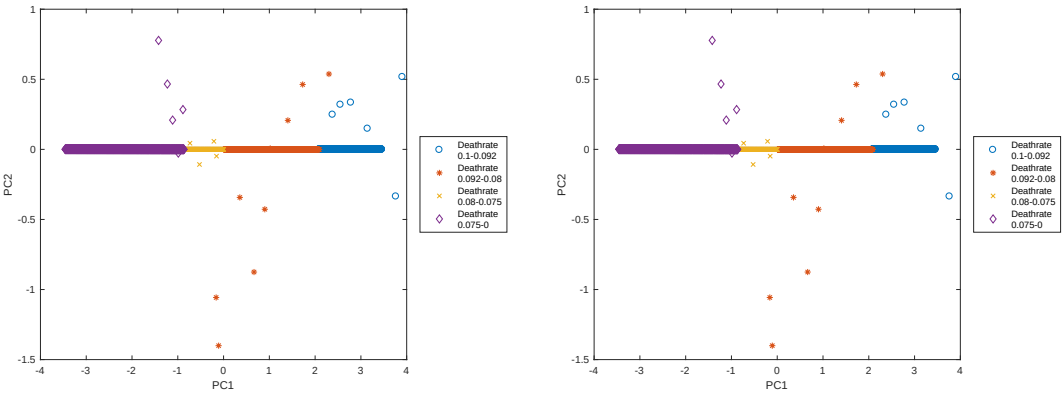


(a) Males (b) Females

Figure D.1.1: PCA visualisation of France real datasets based on 4 negative indicators (i.e. alcohol, cheese, smoking, SBP) for (a) males and (b) females respectively.

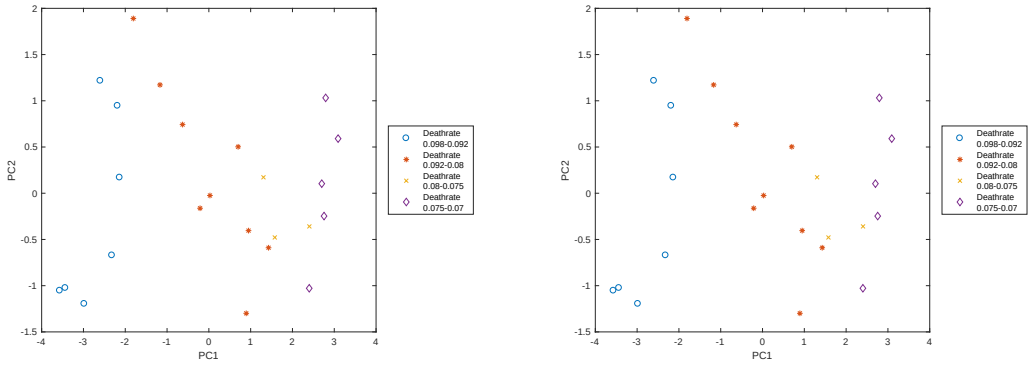


(a) 4 parameters - pure synthetic - males (b) 4 parameters - pure synthetic - females



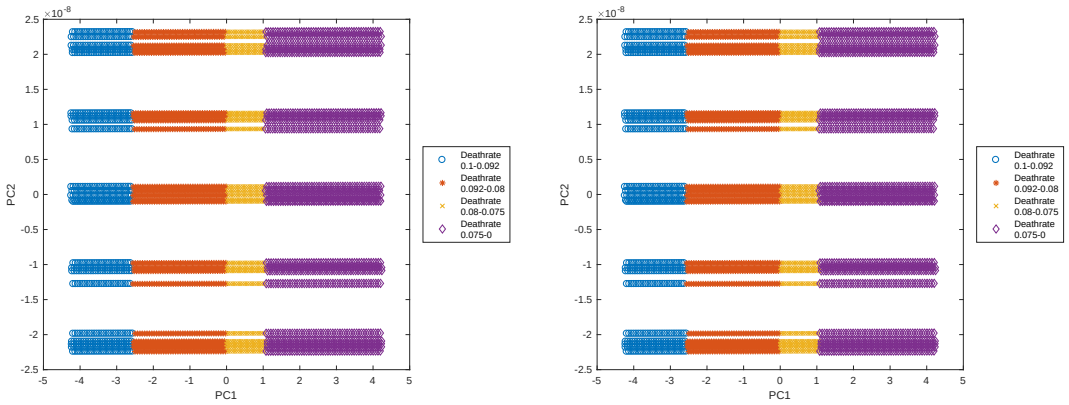
(c) 4 parameters - real-synthetic - males (d) 4 parameters - real-synthetic - females

Figure D.1.2: PCA visualisation of France synthetic datasets based on negative indicators (i.e. alcohol, cheese, smoking, SBP). Figs (a), (b) are visualised with the pure synthetic datasets; Figs (c), (d) are visualised with the real-synthetic datasets.

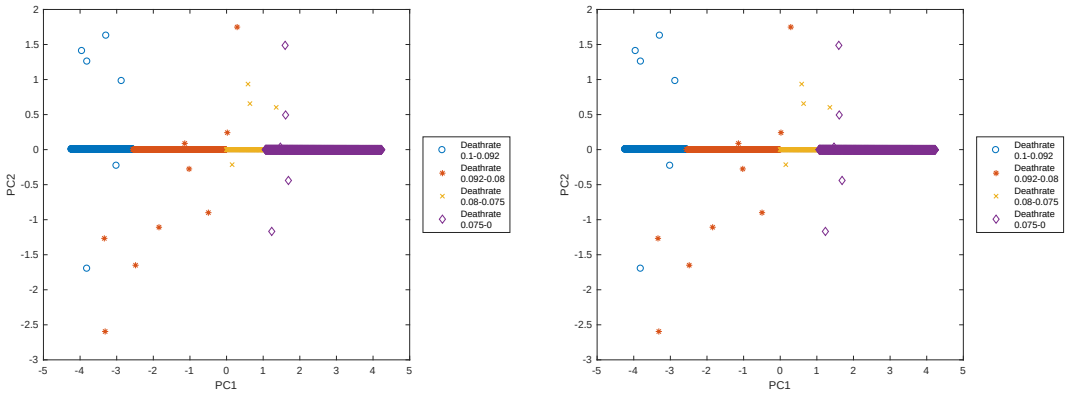


(a) Males (b) Females

Figure D.1.3: PCA visualisation of generated France real datasets based on all 6 parameters (i.e. alcohol, cheese, smoking, SBP, cereals, fruits and vegs) for (a) males and (b) females respectively.



(a) 4 parameters - pure synthetic - males (b) 4 parameters - pure synthetic - females



(c) 4 parameters - real-synthetic - males (d) 4 parameters - real-synthetic - females

Figure D.1.4: PCA visualisation of France synthetic datasets based on 6 parameters (i.e. alcohol, cheese, smoking, SBP, cereals, fruits and vegs). Figs (a), (b) are visualised with the pure synthetic datasets; Figs (c), (d) are visualised with the real-synthetic datasets.

Table D.1.1: Component Matrix of 4 parameters for France

	Component Matrix - real data				Component Matrix - synthetic data			
	Male		Female		Male		Female	
	PC1	PC2	PC1	PC2	PC1	PC2	PC1	PC2
Alcohol	0.4959	0.6279	0.5018	0.4687	0.49998	0.63277	0.50001	0.47136
Cheese	0.4872	0.5603	0.4675	0.8277	0.49993	0.56083	0.49982	0.82103
Smoking	0.4950	0.4267	0.5072	0.3085	0.49997	0.41187	0.50004	0.32080
SBP	0.5213	0.3314	0.5219	0.0091	0.50012	0.33977	0.50013	0.02853
VP (%)	87.1977	8.4131	90.4411	7.5849	99.9180	0.0528	99.9400	0.0490

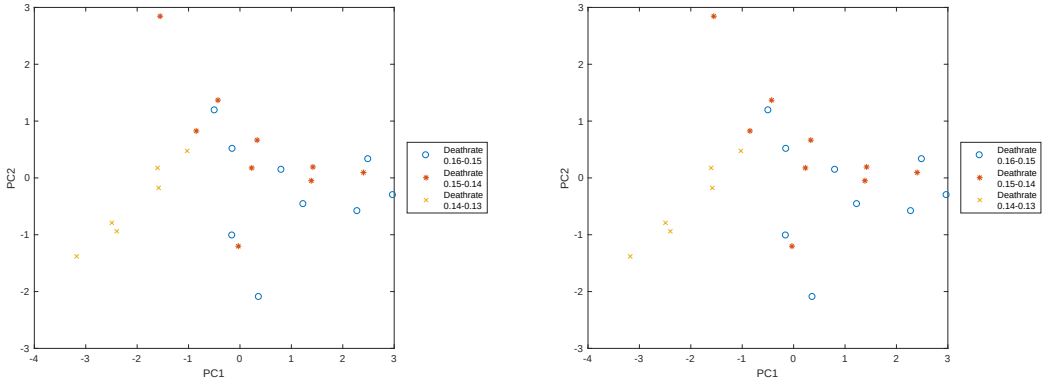
Table D.1.2: Component Matrix of 6 parameters for France

	Component Matrix - real data				Component Matrix - synthetic data			
	Male		Female		Male		Female	
	PC1	PC2	PC1	PC2	PC1	PC2	PC1	PC2
Alcohol	0.4263	0.2381	0.4242	0.2768	0.4085	0.2294	0.4085	0.2610
Cheese	0.4190	0.0864	0.4014	0.0933	0.4085	0.0829	0.4084	0.0861
Smoking	0.4144	0.1285	0.4392	0.0984	0.4084	0.2127	0.4086	0.1552
SBP	0.4506	0.1726	0.4465	0.1338	0.4086	0.1968	0.4086	0.1892
Cereals	0.4375	0.1408	0.4341	0.1749	0.4086	0.1821	0.4085	0.2106
Fruits&Vegs	0.2765	0.9326	0.2797	0.9255	0.4069	0.9074	0.4069	0.9056
VP (%)	77.8691	12.0498	81.0472	11.7643	99.7568	0.1785	99.7817	0.1728

Table D.1.3: Ranking orders of parameters for France

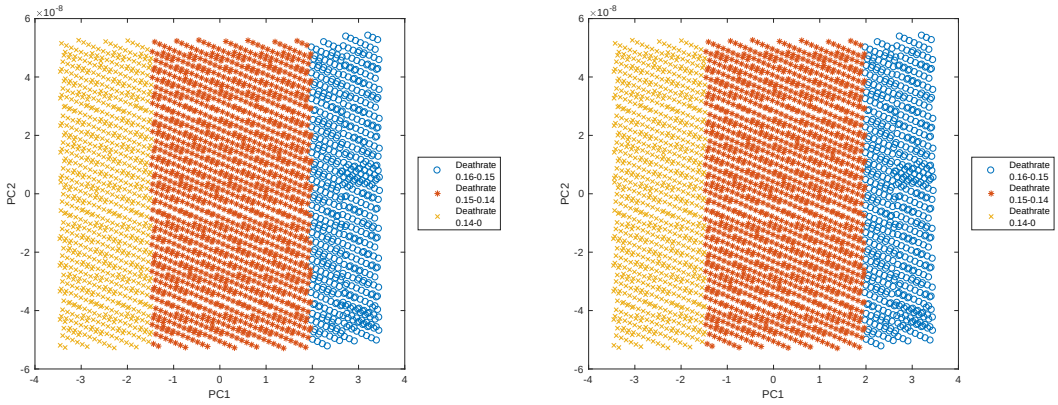
	Ranking orders of 4 negative indicators				Ranking orders of 6 parameters			
	Real Data		Synthetic Data		Real Data		Synthetic Data	
Rank	**male**	**female**	**male**	**female**	**male**	**female**	**male**	**female**
1	SBP	SBP	SBP	SBP	SBP	SBP	SBP	SBP
2	Alcohol	Smoking	Alcohol	Smoking	Cereals	Smoking	Cereals	Smoking
3	Smoking	Alcohol	Smoking	Alcohol	Alcohol	Cereals	Cheese	Cereals
4	Cheese	Cheese	Cheese	Cheese	Cheese	Alcohol	Alcohol	Alcohol
5	-	-	-	-	Smoking	Cheese	Smoking	Cheese
6	-	-	-	-	F&V	F&V	F&V	F&V

Greece

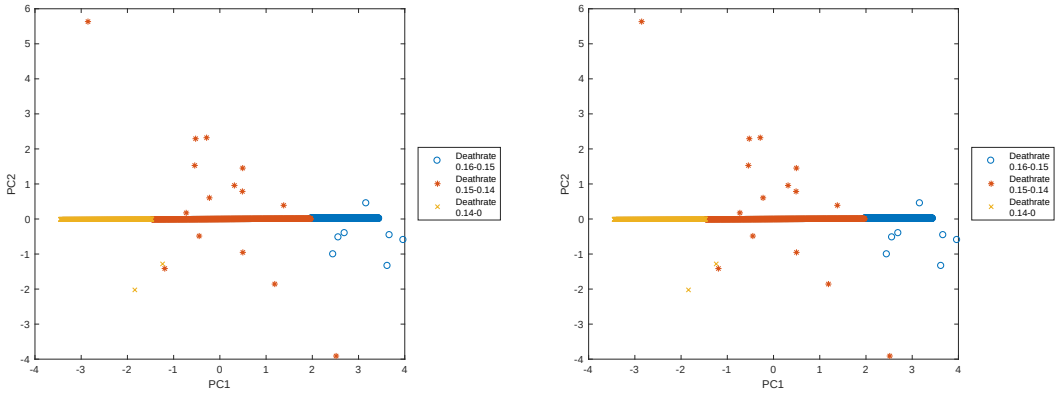


(a) Males (b) Females

Figure D.1.5: PCA visualisation of Greece real datasets based on 4 negative indicators (i.e. alcohol, cheese, smoking, SBP) for (a) males and (b) females respectively.

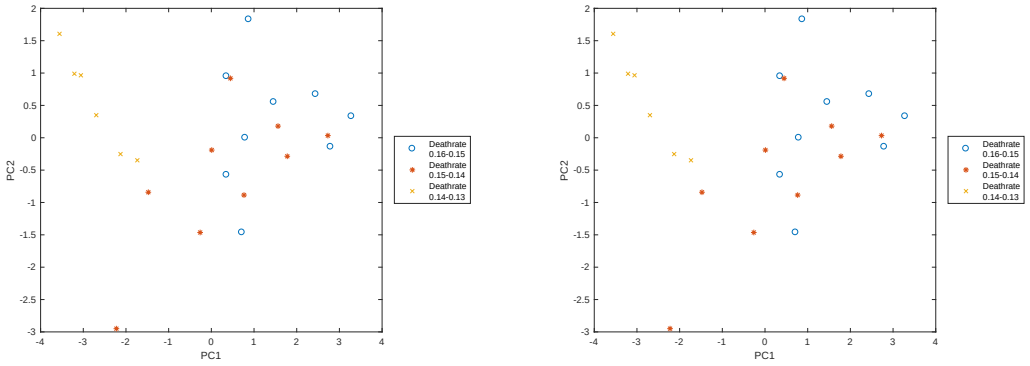


(a) 4 parameters - pure synthetic - males (b) 4 parameters - pure synthetic - females



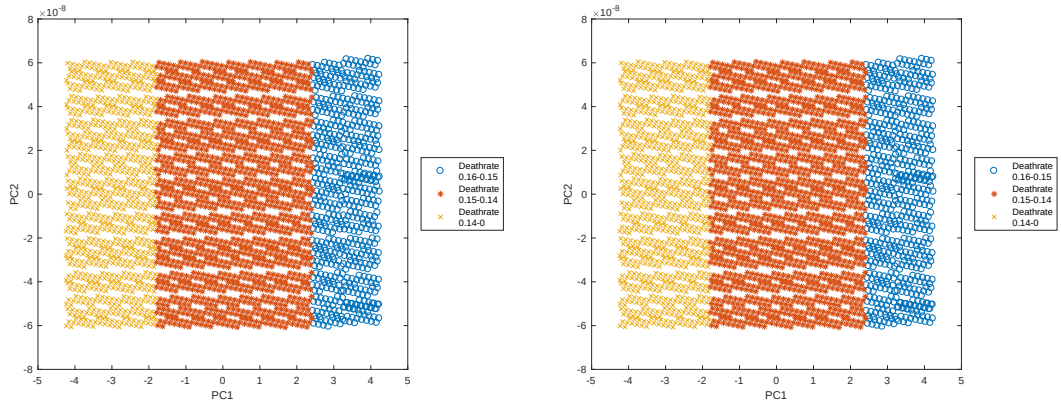
(c) 4 parameters - real-synthetic - males (d) 4 parameters - real-synthetic - females

Figure D.1.6: PCA visualisation of Greece synthetic datasets based on 4 negative indicators (i.e. alcohol, cheese, smoking, SBP). Figs (a), (b) are visualised with the pure synthetic datasets; Figs (c), (d) are visualised with the real-synthetic datasets.

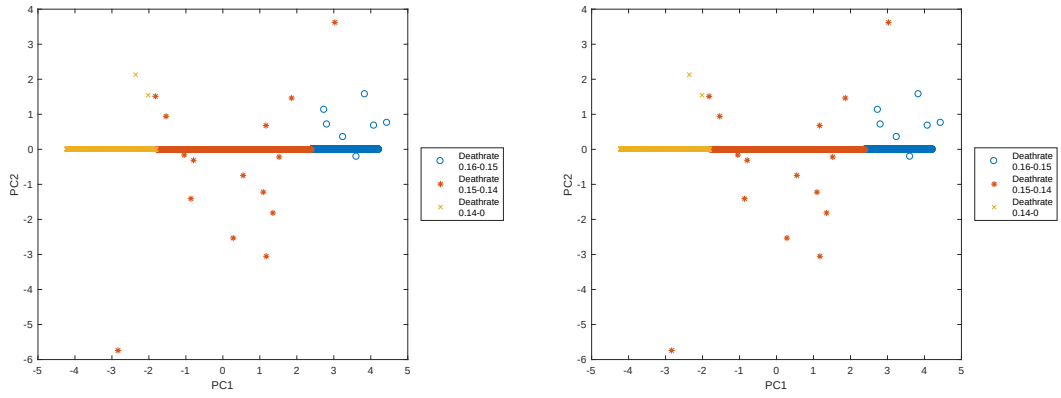


(a) Males (b) Females

Figure D.1.7: PCA visualisation of generated Greece real datasets based on all 6 parameters (i.e. alcohol, cheese, smoking, SBP, cereals, fruits and vegs) for (a) males and (b) females respectively.



(a) 4 parameters - pure synthetic - males (b) 4 parameters - pure synthetic - females



(c) 4 parameters - real-synthetic - males (d) 4 parameters - real-synthetic - females

Figure D.1.8: PCA visualisation of Greece synthetic datasets based on 6 parameters (i.e. alcohol, cheese, smoking, SBP, cereals, fruits and vegs). Figs (a), (b) are visualised with the pure synthetic datasets; Figs (c), (d) are visualised with the real-synthetic datasets.

Table D.1.4: Component Matrix of 4 parameters for Greece

	Component Matrix - real data				Component Matrix - synthetic data			
	Male		Female		Male		Female	
	PC1	PC2	PC1	PC2	PC1	PC2	PC1	PC2
Alcohol	0.4663	0.5412	0.5676	0.5092	0.50009	0.46129	0.50059	0.32075
Cheese	0.2973	0.8362	0.3974	0.4636	0.49700	0.84557	0.49773	0.84637
Smoking	0.5772	0.0670	0.1161	0.7245	0.50141	0.16928	0.49932	0.41043
SBP	0.6009	0.0581	0.7117	0.0290	0.50149	0.20874	0.50235	0.11100
VP (%)	66.9414	25.1887	46.5587	29.2119	99.2571	0.6718	98.8598	0.7039

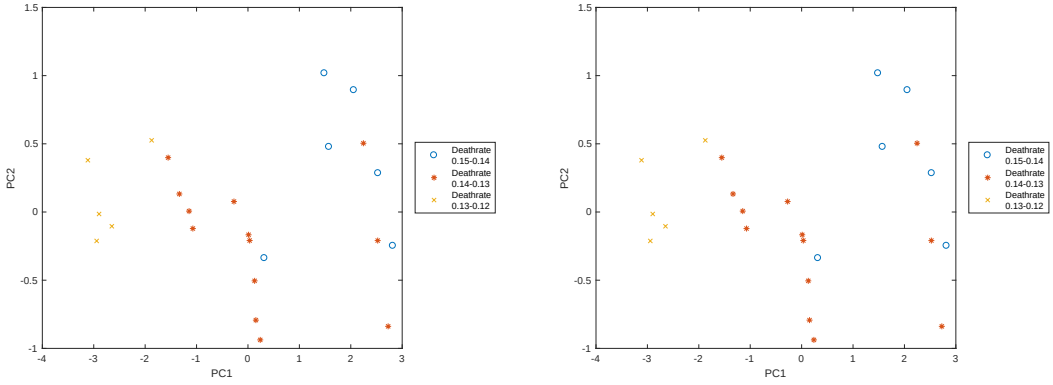
Table D.1.5: Component Matrix of 6 parameters for Greece

	Component Matrix - real data				Component Matrix - synthetic data			
	Male		Female		Male		Female	
	PC1	PC2	PC1	PC2	PC1	PC2	PC1	PC2
Alcohol	0.3515	0.5029	0.3810	0.5692	0.4082	0.3498	0.4086	0.2881
Cheese	0.2369	0.8111	0.2786	0.4487	0.4049	0.8826	0.4053	0.8862
Smoking	0.4640	0.0639	0.1504	0.6600	0.4093	0.0902	0.4070	0.2693
SBP	0.4867	0.0718	0.5301	0.0561	0.4094	0.1313	0.4099	0.0720
Cereals	0.4683	0.1668	0.5327	0.1434	0.4094	0.0346	0.4099	0.0188
Fruits&Vegs	0.3858	0.2286	0.4356	0.1241	0.4083	0.2685	0.4088	0.2315
VP (%)	67.2893	17.8744	53.8352	20.1691	99.3625	0.4860	99.0815	0.4943

Table D.1.6: Ranking orders of parameters for Greece

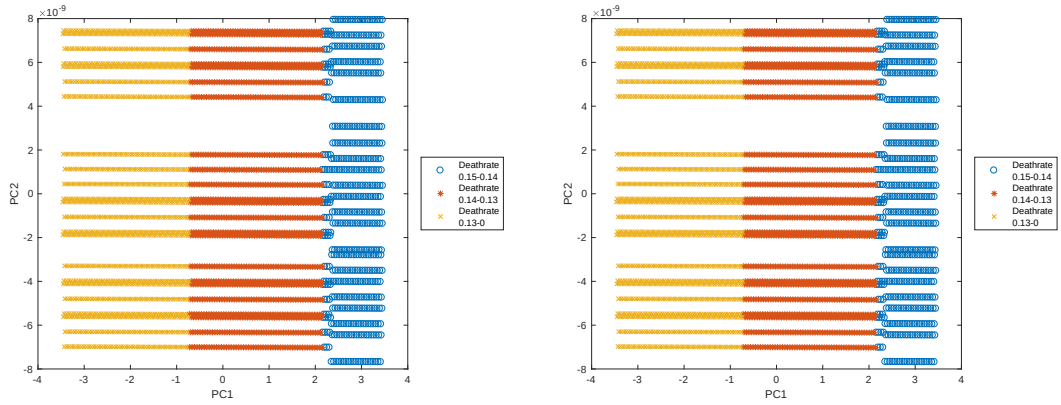
	Ranking orders of 4 negative indicators				Ranking orders of 6 parameters			
	Real Data		Synthetic Data		Real Data		Synthetic Data	
Rank	**male**	**female**	**male**	**female**	**male**	**female**	**male**	**female**
1	SBP	SBP	SBP	SBP	SBP	Cereals	Cereals	Cereals
2	Smoking	Alcohol	Smoking	Alcohol	Cereals	SBP	SBP	SBP
3	Alcohol	Cheese	Alcohol	Smoking	Smoking	F&V	Smoking	F&V
4	Cheese	Smoking	Cheese	Cheese	F&V	Alcohol	F&V	Alcohol
5	-	-	-	-	Alcohol	Cheese	Alcohol	Smoking
6	-	-	-	-	Cheese	Smoking	Cheese	Cheese

Italy

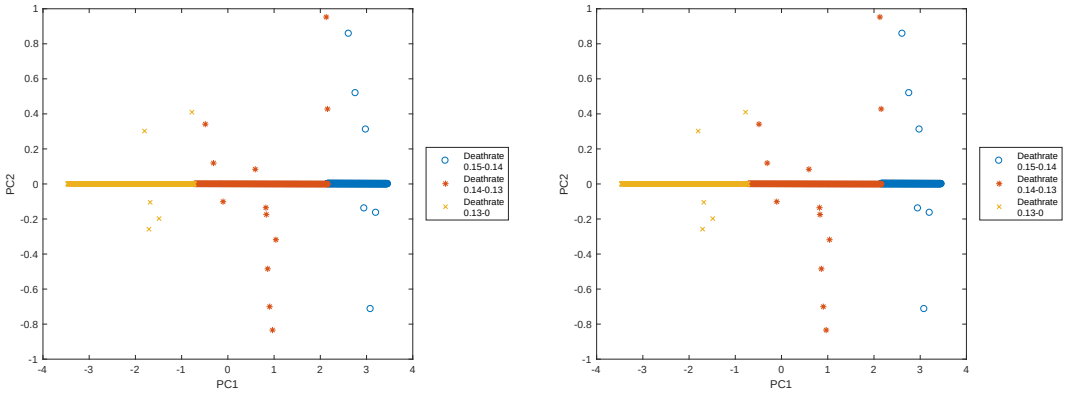


(a) Males (b) Females

Figure D.1.9: PCA visualisation of Italy real datasets based on 4 negative indicators (i.e. alcohol, cheese, smoking, SBP) for (a) males and (b) females respectively.

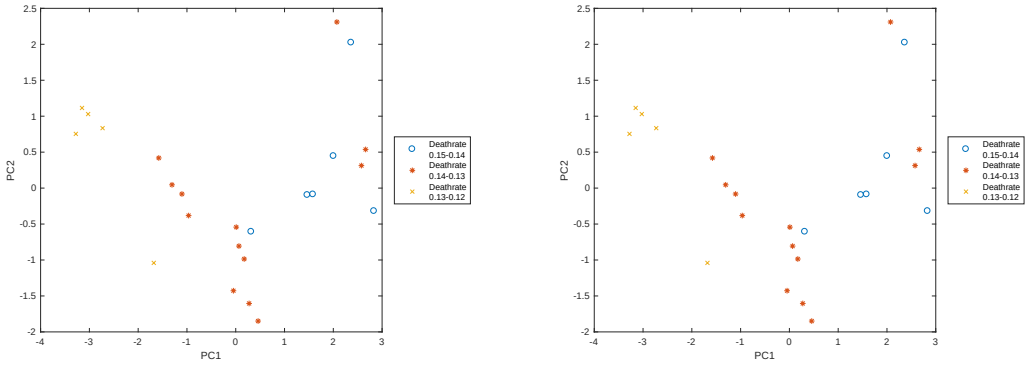


(a) 4 parameters - pure synthetic - males (b) 4 parameters - pure synthetic - females



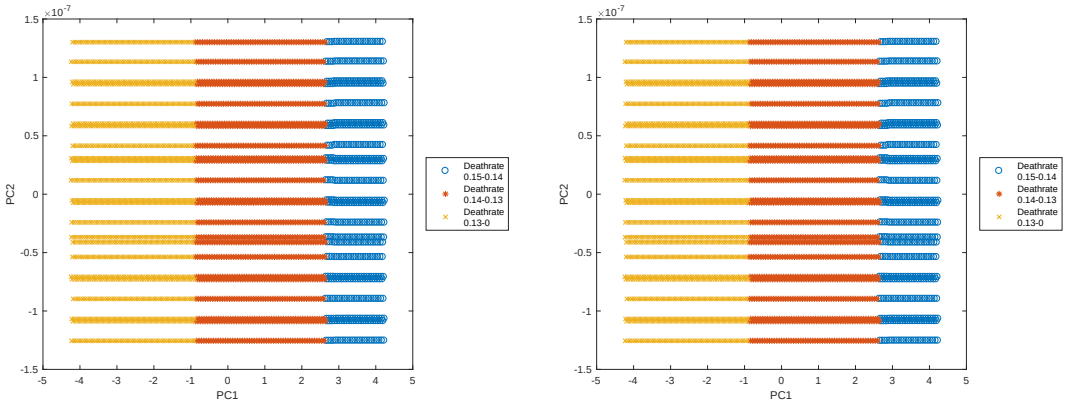
(c) 4 parameters - real-synthetic - males (d) 4 parameters - real-synthetic - females

Figure D.1.10: PCA visualisation of Italy synthetic datasets based on 4 negative indicators (i.e. alcohol, cheese, smoking, SBP). Figs (a), (b) are visualised with the pure synthetic datasets; Figs (c), (d) are visualised with the real-synthetic datasets.

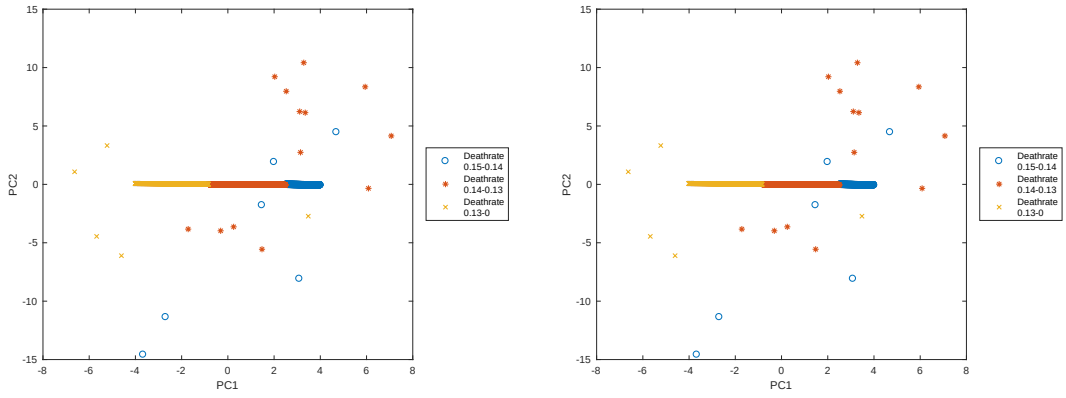


(a) Males (b) Females

Figure D.1.11: PCA visualisation of generated Italy real datasets based on all 6 parameters (i.e. alcohol, cheese, smoking, SBP, cereals, fruits and vegs) for (a) males and (b) females respectively.



(a) 4 parameters - pure synthetic - males (b) 4 parameters - pure synthetic - females



(c) 4 parameters - real-synthetic - males (d) 4 parameters - real-synthetic - females

Figure D.1.12: PCA visualisation of Italy synthetic datasets based on 6 parameters (i.e. alcohol, cheese, smoking, SBP, cereals, fruits and vegs). Figs (a), (b) are visualised with the pure synthetic datasets; Figs (c), (d) are visualised with the real-synthetic datasets.

Table D.1.7: Component Matrix of 4 parameters for Italy

	Component Matrix - real data				Component Matrix - synthetic data			
	Male		Female		Male		Female	
	PC1	PC2	PC1	PC2	PC1	PC2	PC1	PC2
Alcohol	0.4958	0.5205	0.5240	0.1862	0.49998	0.49854	0.50019	0.23961
Cheese	0.4783	0.8294	0.4928	0.4612	0.49987	0.83281	0.49998	0.46151
Smoking	0.5059	0.0846	0.4211	0.8647	0.50004	0.13458	0.49947	0.84255
SBP	0.5192	0.1845	0.5526	0.0711	0.50011	0.19943	0.50037	0.14036
VP (%)	89.9384	6.0699	79.4016	14.2441	99.9372	0.0403	99.8232	0.1360

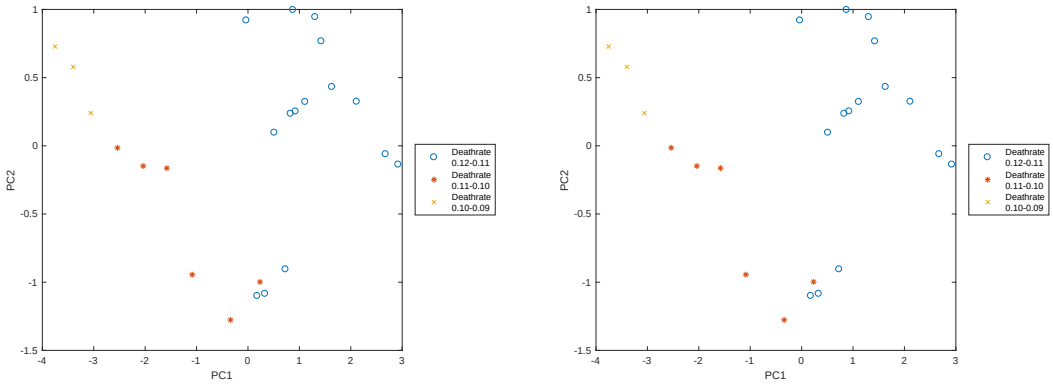
Table D.1.8: Component Matrix of 6 parameters for Italy

	Component Matrix - real data				Component Matrix - synthetic data			
	Male		Female		Male		Female	
	PC1	PC2	PC1	PC2	PC1	PC2	PC1	PC2
Alcohol	0.4942	0.0348	0.5131	0.1292	0.4248	0.1217	0.4242	0.1278
Cheese	0.4716	0.2073	0.4707	0.3517	0.4239	0.1473	0.4232	0.1530
Smoking	0.5009	0.0832	0.4314	0.3590	0.4244	0.1291	0.4251	0.0759
SBP	0.5186	0.0421	0.5416	0.1149	0.4250	0.1234	0.4244	0.1289
Cereals	0.0559	0.8088	0.1344	0.8181	0.3602	0.9218	0.3617	0.9133
Fruits&Vegs	0.1020	0.5412	0.1331	0.2194	0.3864	0.2863	0.3863	0.3220
VP (%)	60.5490	17.6613	54.2745	19.7703	90.8776	5.6579	90.9519	5.5881

Table D.1.9: Ranking orders of parameters for Italy

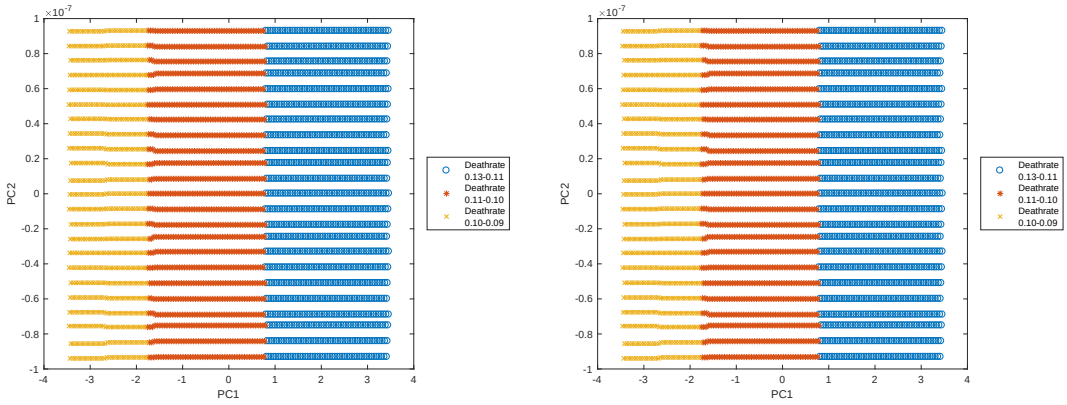
	Ranking orders of 4 negative indicators				Ranking orders of 6 parameters			
	Real Data		Synthetic Data		Real Data		Synthetic Data	
Rank	**male**	**female**	**male**	**female**	**male**	**female**	**male**	**female**
1	SBP	SBP	SBP	SBP	SBP	SBP	SBP	Smoking
2	Smoking	Alcohol	Smoking	Alcohol	Smoking	Alcohol	Alcohol	SBP
3	Alcohol	Cheese	Alcohol	Cheese	Alcohol	Cheese	Smoking	Alcohol
4	Cheese	Smoking	Cheese	Smoking	Cheese	Smoking	Cheese	Cheese
5	-	-	-	-	F&V	Cereals	F&V	F&V
6	-	-	-	-	Cereals	F&V	Cereals	Cereals

Spain

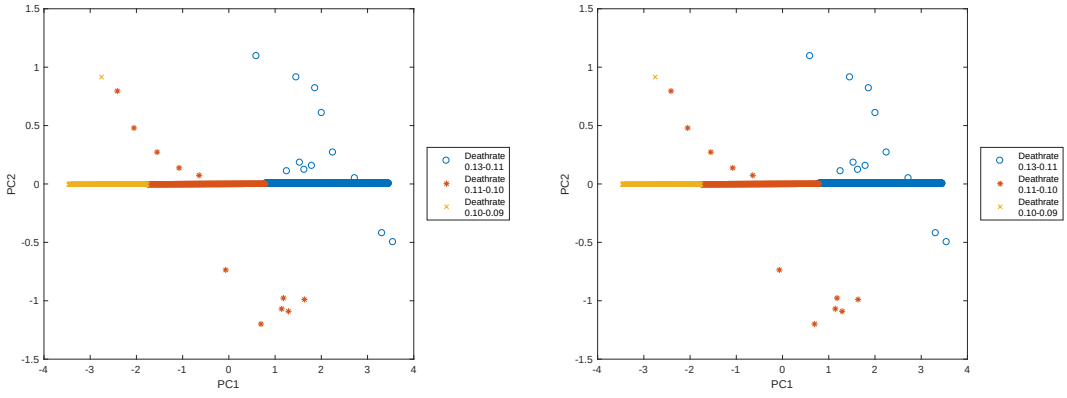


(a) Males (b) Females

Figure D.1.13: PCA visualisation of Spain real datasets based on 4 negative indicators (i.e. alcohol, cheese, smoking, SBP) for (a) males and (b) females respectively.

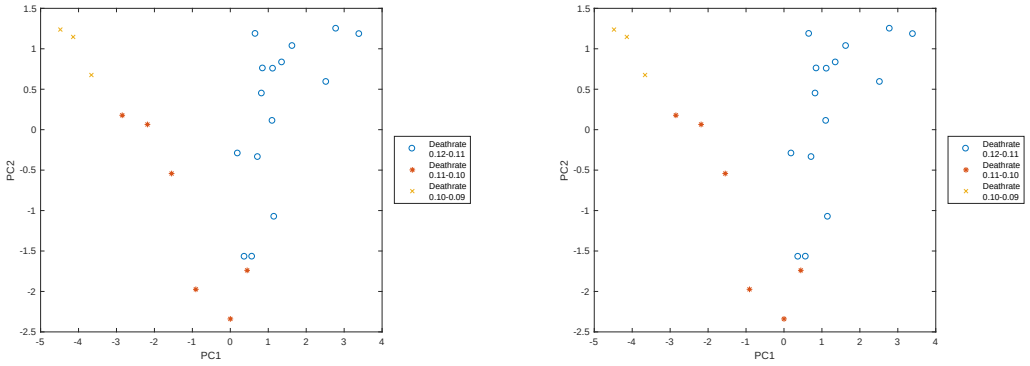


(a) 4 parameters - pure synthetic - males (b) 4 parameters - pure synthetic - females



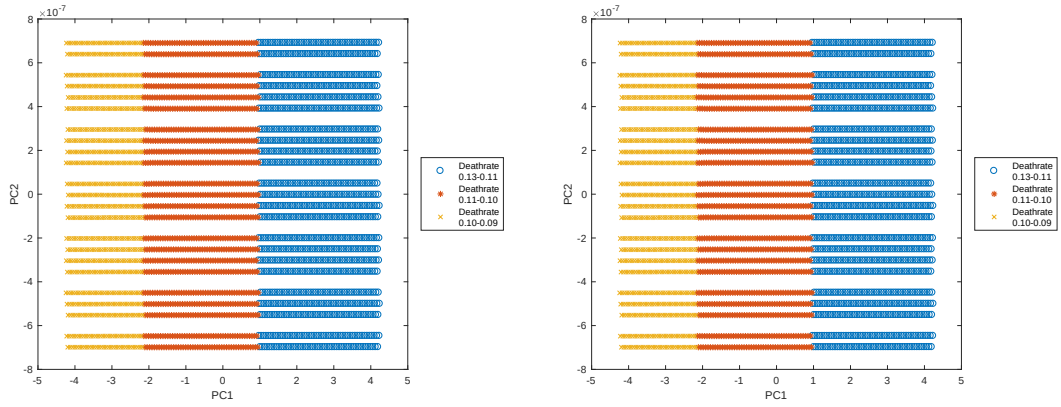
(c) 4 parameters - real-synthetic - males (d) 4 parameters - real-synthetic - females

Figure D.1.14: PCA visualisation of Spain synthetic datasets based on 4 negative indicators (i.e. alcohol, cheese, smoking, SBP). Figs (a), (b) are visualised with the pure synthetic datasets; Figs (c), (d) are visualised with the real-synthetic datasets.

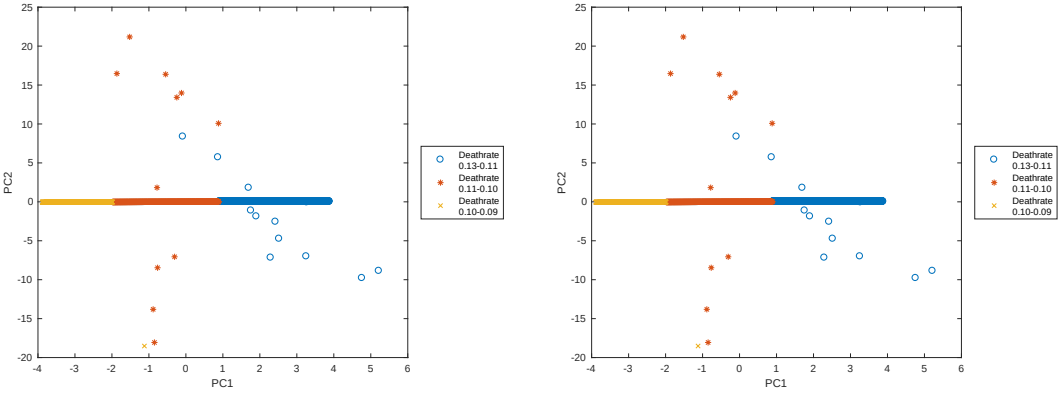


(a) Males (b) Females

Figure D.1.15: PCA visualisation of generated Spain real datasets based on all 6 parameters (i.e. alcohol, cheese, smoking, SBP, cereals, fruits and vegs) for (a) males and (b) females respectively.



(a) 4 parameters - pure synthetic - males (b) 4 parameters - pure synthetic - females



(c) 4 parameters - real-synthetic - males (d) 4 parameters - real-synthetic - females

Figure D.1.16: PCA visualisation of Spain synthetic datasets based on 6 parameters (i.e. alcohol, cheese, smoking, SBP, cereals, fruits and vegs). Figs (a), (b) are visualised with the pure synthetic datasets; Figs (c), (d) are visualised with the real-synthetic datasets.

Table D.1.10: Component Matrix of 4 parameters for Spain

	Component Matrix - real data				Component Matrix - synthetic data			
	Male		Female		Male		Female	
	PC1	PC2	PC1	PC2	PC1	PC2	PC1	PC2
Alcohol	0.4687	0.6720	0.4361	0.8626	0.49975	0.72175	0.49957	0.84123
Cheese	0.5079	0.4932	0.5308	0.2129	0.50006	0.48358	0.50022	0.23836
Smoking	0.5153	0.4458	0.4916	0.4526	0.50011	0.44822	0.49996	0.46578
SBP	0.5068	0.3262	0.5352	0.0759	0.50008	0.21054	0.50026	0.13623
VP (%)	83.8937	12.0448	82.3906	12.1639	99.8710	0.0975	99.8530	0.1111

Table D.1.11: Component Matrix of 6 parameters for Spain

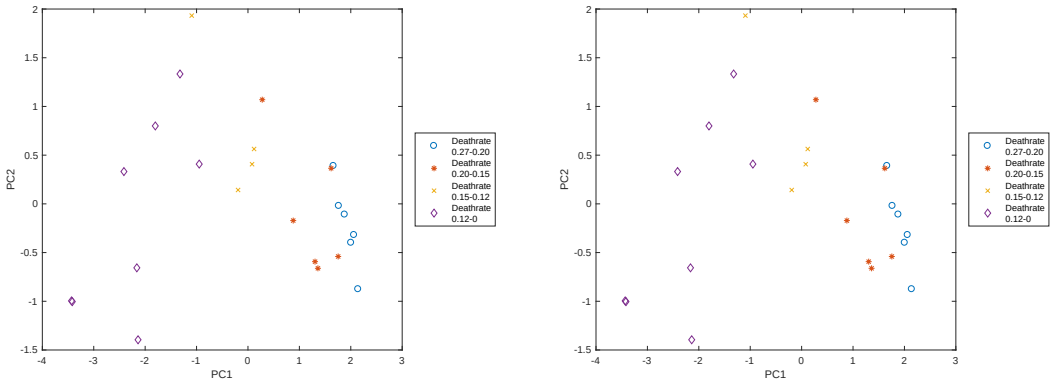
	Component Matrix - real data				Component Matrix - synthetic data			
	Male		Female		Male		Female	
	PC1	PC2	PC1	PC2	PC1	PC2	PC1	PC2
Alcohol	0.4250	0.2028	0.4106	0.3441	0.4439	0.0707	0.4434	0.0806
Cheese	0.4270	0.3464	0.4578	0.2372	0.4451	0.0239	0.4449	0.0337
Smoking	0.4369	0.3427	0.4252	0.1595	0.4452	0.0235	0.4445	0.0430
SBP	0.4588	0.1979	0.4677	0.2137	0.4442	0.0735	0.4450	0.0335
Cereals	0.1545	0.8259	0.0947	0.8607	0.1075	0.9930	0.1116	0.9930
Fruits&Vegs	0.4602	0.0151	0.4619	0.1159	0.4447	0.0488	0.4444	0.0587
VP (%)	72.1576	20.9312	69.4734	20.1863	84.0241	15.9159	84.0700	15.8414

Table D.1.12: Ranking orders of parameters for Spain

	Ranking orders of 4 negative indicators				Ranking orders of 6 parameters			
	Real Data		Synthetic Data		Real Data		Synthetic Data	
Rank	**male**	**female**	**male**	**female**	**male**	**female**	**male**	**female**
1	Smoking	SBP	Smoking	SBP	F&V	SBP	Smoking	SBP
2	Cheese	Cheese	SBP	Cheese	SBP	F&V	Cheese	Cheese
3	SBP	Smoking	Cheese	Smoking	Smoking	Cheese	F&V	Smoking
4	Alcohol	Alcohol	Alcohol	Alcohol	Cheese	Smoking	SBP	F&V
5	-	-	-	-	Alcohol	Alcohol	Alcohol	Alcohol
6	-	-	-	-	Cereals	Cereals	Cereals	Cereals

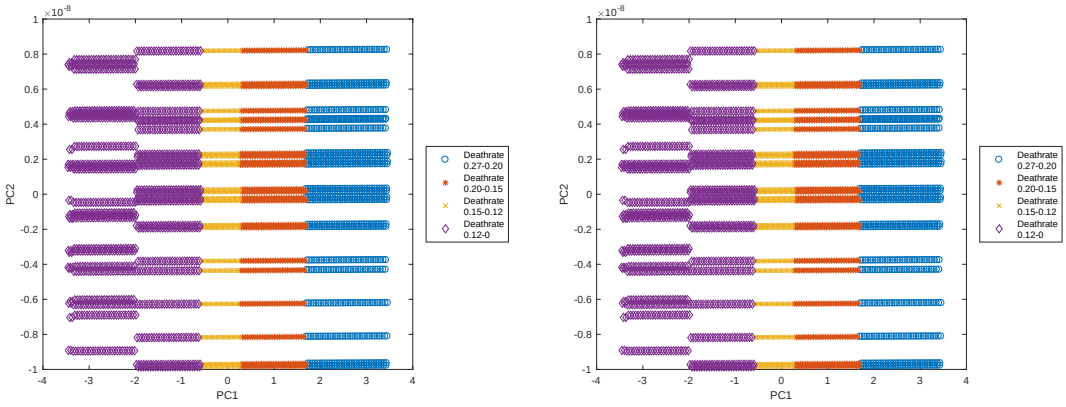
D.2 SCANDINAVIAN EUROPEAN COUNTRIES (SCEU) BLOCK

Denmark

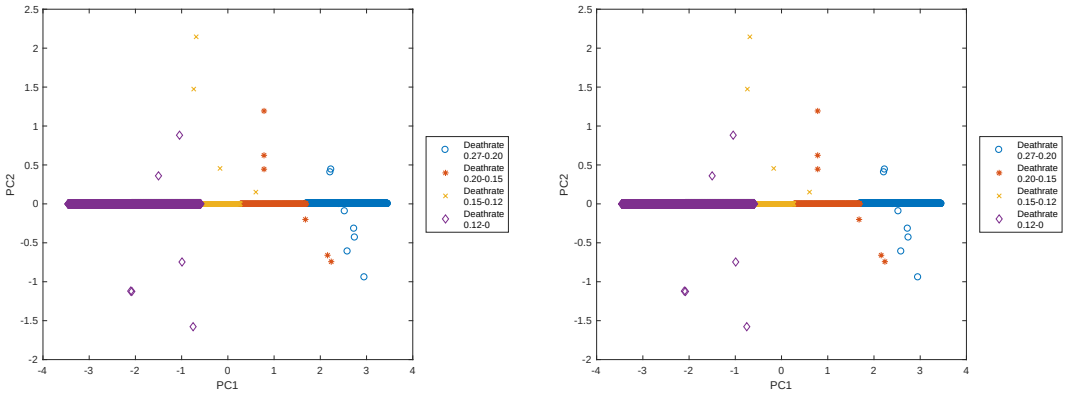


(a) Males (b) Females

Figure D.2.1: PCA visualisation of Denmark real datasets based on 4 negative indicators (i.e. alcohol, cheese, smoking, SBP) for (a) males and (b) females respectively.

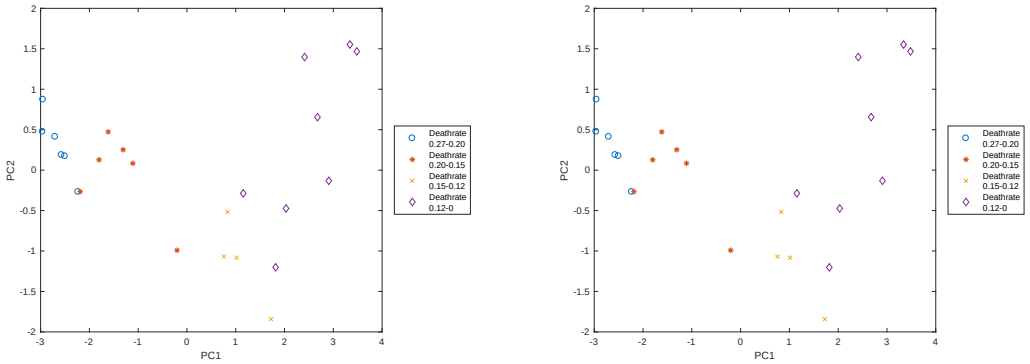


(a) 4 parameters - pure synthetic - males (b) 4 parameters - pure synthetic - females



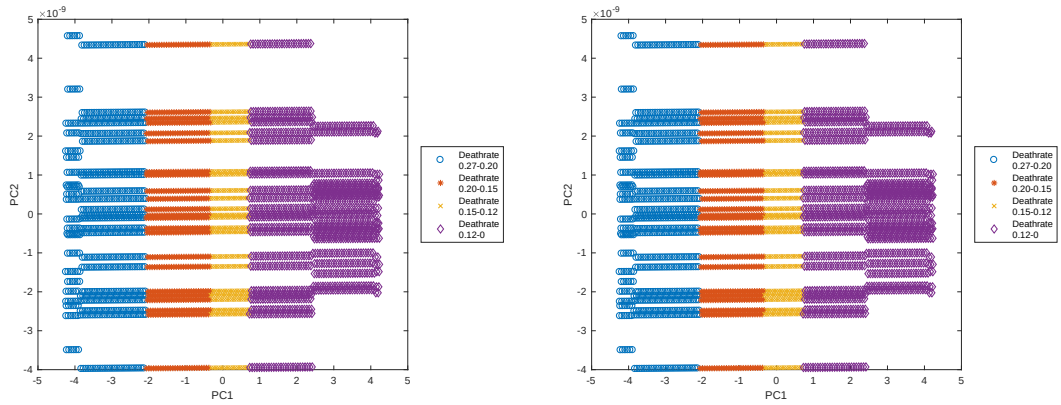
(c) 4 parameters - real-synthetic - males (d) 4 parameters - real-synthetic - females

Figure D.2.2: PCA visualisation of Denmark synthetic datasets based on 4 negative indicators (i.e. alcohol, cheese, smoking, SBP). Figs (a), (b) are visualised with the pure synthetic datasets; Figs (c), (d) are visualised with the real-synthetic datasets.

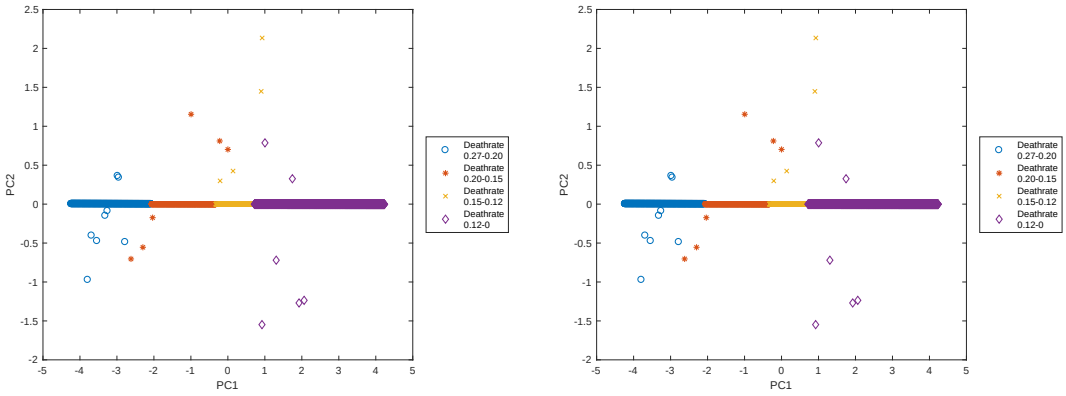


(a) Males (b) Females

Figure D.2.3: PCA visualisation of generated Denmark real datasets based on all 6 parameters (i.e. alcohol, cheese, smoking, SBP, cereals, fruits and vegs) for (a) males and (b) females respectively.



(a) 4 parameters - pure synthetic - males (b) 4 parameters - pure synthetic - females



(c) 4 parameters - real-synthetic - males (d) 4 parameters - real-synthetic - females

Figure D.2.4: PCA visualisation of Denmark synthetic datasets based on 6 parameters (i.e. alcohol, cheese, smoking, SBP, cereals, fruits and vegs). Figs (a), (b) are visualised with the pure synthetic datasets; Figs (c), (d) are visualised with the real-synthetic datasets.

Table D.2.1: Component Matrix of 4 parameters for Denmark

	Component Matrix - real data				Component Matrix - synthetic data			
	Male		Female		Male		Female	
	PC1	PC2	PC1	PC2	PC1	PC2	PC1	PC2
Alcohol	0.5291	0.2753	0.5044	0.4214	0.50019	0.33590	0.50005	0.43104
Cheese	0.3834	0.9159	0.4132	0.8837	0.49929	0.86159	0.49940	0.85193
Smoking	0.5335	0.1146	0.5341	0.1205	0.50027	0.20070	0.50027	0.19588
SBP	0.5371	0.2688	0.5381	0.1641	0.50025	0.32336	0.50028	0.22372
VP (%)	81.3458	15.5086	83.5582	13.5896	99.8275	0.1536	99.8488	0.1338

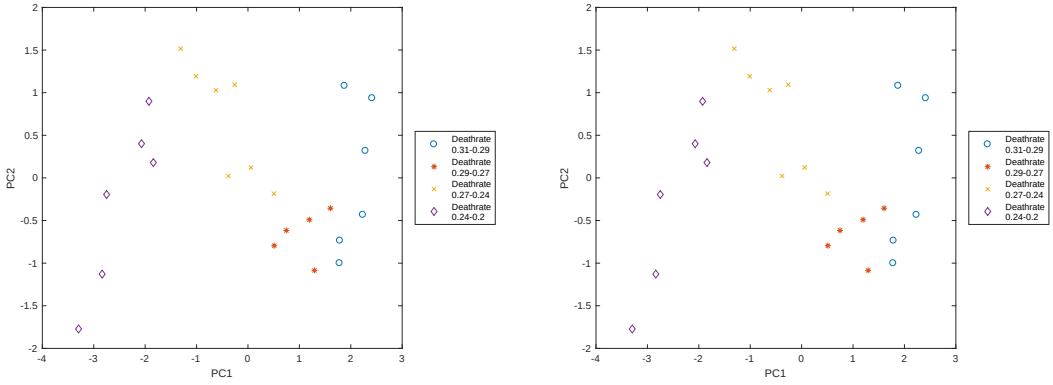
Table D.2.2: Component Matrix of 6 parameters for Denmark

	Component Matrix - real data				Component Matrix - synthetic data			
	Male		Female		Male		Female	
	PC1	PC2	PC1	PC2	PC1	PC2	PC1	PC2
Alcohol	0.4124	0.4287	0.3917	0.5358	0.4083	0.3907	0.4082	0.4433
Cheese	0.3389	0.6728	0.3456	0.7474	0.4077	0.7918	0.4077	0.8412
Smoking	0.4384	0.2112	0.4374	0.1861	0.4085	0.2344	0.4085	0.2028
SBP	0.4192	0.4235	0.4350	0.2487	0.4083	0.3775	0.4084	0.2321
Cereals	0.4259	0.2425	0.4243	0.1609	0.4084	0.0900	0.4084	0.0166
Fruits&Vegs	0.4071	0.2842	0.4081	0.1787	0.4083	0.1220	0.4083	0.0217
VP (%)	79.8687	12.0437	83.0021	9.5739	99.8367	0.1048	99.8586	0.0893

Table D.2.3: Ranking orders of parameters for Denmark

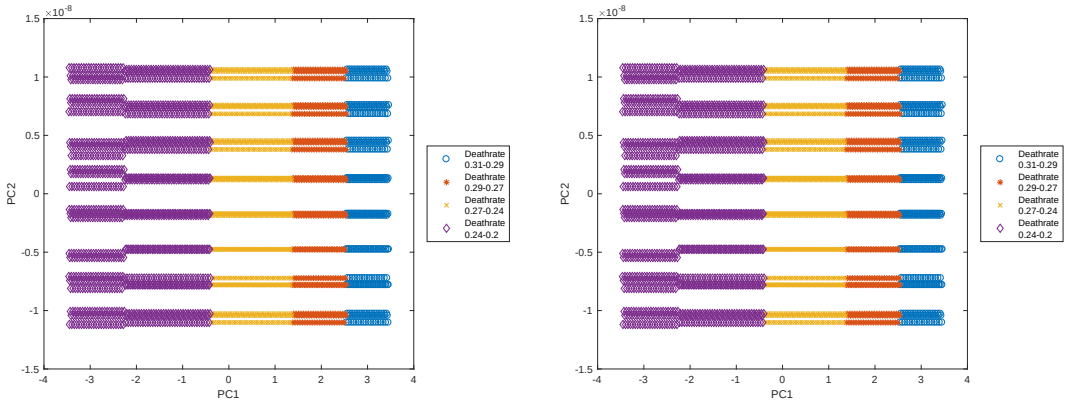
	Ranking orders of 4 negative indicators				Ranking orders of 6 parameters			
	Real Data		Synthetic Data		Real Data		Synthetic Data	
Rank	**male**	**female**	**male**	**female**	**male**	**female**	**male**	**female**
1	SBP	SBP	Smoking	SBP	Smoking	Smoking	Smoking	Smoking
2	Smoking	Smoking	SBP	Smoking	Cereals	SBP	Cereals	SBP
3	Alcohol	Alcohol	Alcohol	Alcohol	SBP	Cereals	SBP	Cereals
4	Cheese	Cheese	Cheese	Cheese	Alcohol	F&V	Alcohol	F&V
5	-	-	-	-	F&V	Alcohol	F&V	Alcohol
6	-	-	-	-	Cheese	Cheese	Cheese	Cheese

Finland

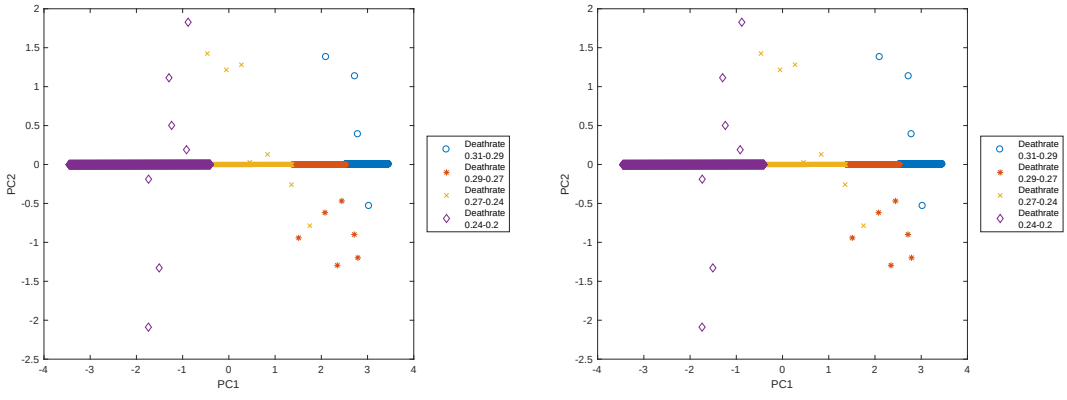


(a) Males (b) Females

Figure D.2.5: PCA visualisation of Finland real datasets based on 4 negative indicators (i.e. alcohol, cheese, smoking, SBP) for (a) males and (b) females respectively.

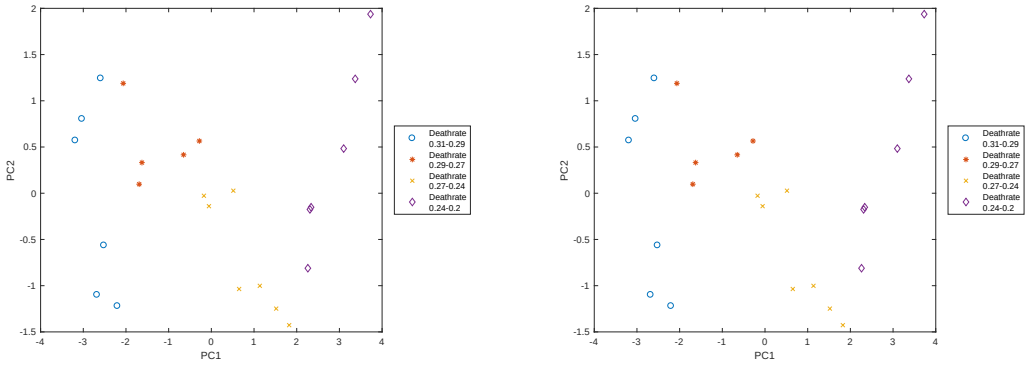


(a) 4 parameters - pure synthetic - males (b) 4 parameters - pure synthetic - females



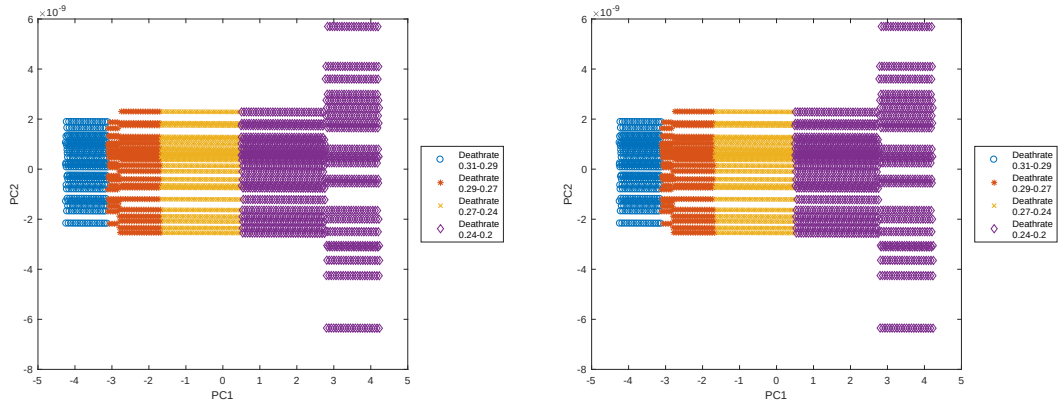
(c) 4 parameters - real-synthetic - males (d) 4 parameters - real-synthetic - females

Figure D.2.6: PCA visualisation of Finland synthetic datasets based on 4 negative indicators (i.e. alcohol, cheese, smoking, SBP). Figs (a), (b) are visualised with the pure synthetic datasets; Figs (c), (d) are visualised with the real-synthetic datasets.

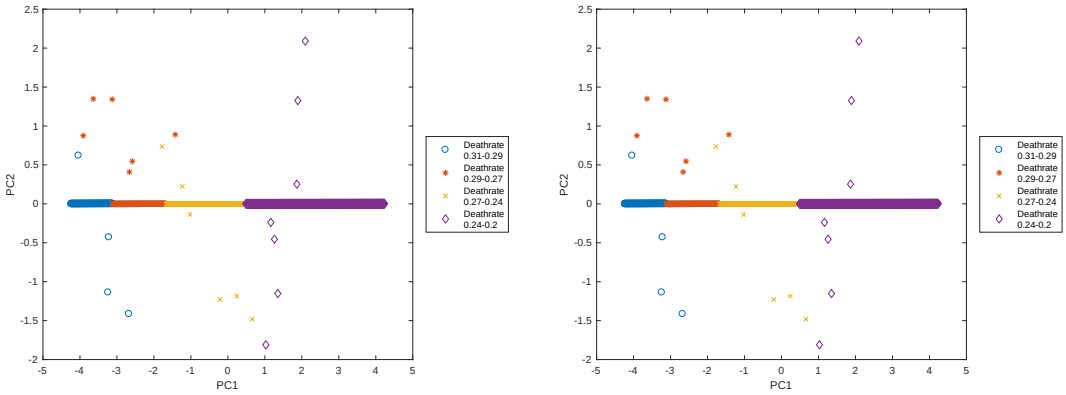


(a) Males (b) Females

Figure D.2.7: PCA visualisation of generated Finland real datasets based on all 6 parameters (i.e. alcohol, cheese, smoking, SBP, cereals, fruits and vegs) for (a) males and (b) females respectively.



(a) 4 parameters - pure synthetic - males (b) 4 parameters - pure synthetic - females



(c) 4 parameters - real-synthetic - males (d) 4 parameters - real-synthetic - females

Figure D.2.8: PCA visualisation of Finland synthetic datasets based on 6 parameters (i.e. alcohol, cheese, smoking, SBP, cereals, fruits and vegs). Figs (a), (b) are visualised with the pure synthetic datasets; Figs (c), (d) are visualised with the real-synthetic datasets.

Table D.2.4: Component Matrix of 4 parameters for Finland

	Component Matrix - real data				Component Matrix - synthetic data			
	Male		Female		Male		Female	
	PC1	PC2	PC1	PC2	PC1	PC2	PC1	PC2
Alcohol	0.3592	0.8961	0.3616	0.8721	0.49904	0.84985	0.49905	0.83818
Cheese	0.5194	0.3466	0.5343	0.3336	0.50015	0.39096	0.50025	0.37011
Smoking	0.5340	0.2770	0.5235	0.3486	0.50027	0.32981	0.50016	0.39405
SBP	0.5621	0.0108	0.5565	0.0815	0.50053	0.12702	0.50054	0.07204
VP (%)	76.9697	18.6591	75.0664	19.6155	99.7565	0.2159	99.7359	0.2274

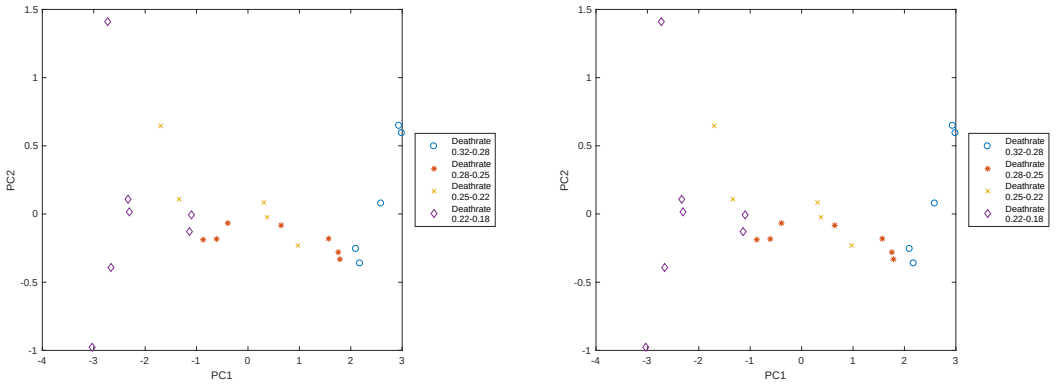
Table D.2.5: Component Matrix of 6 parameters for Finland

	Component Matrix - real data				Component Matrix - synthetic data			
	Male		Female		Male		Female	
	PC1	PC2	PC1	PC2	PC1	PC2	PC1	PC2
Alcohol	0.3151	0.7767	0.3208	0.7163	0.4074	0.8326	0.4074	0.8011
Cheese	0.4063	0.4075	0.4164	0.4069	0.4083	0.3935	0.4083	0.3833
Smoking	0.4121	0.3866	0.3897	0.4819	0.4083	0.3431	0.4082	0.4259
SBP	0.4531	0.0863	0.4549	0.0295	0.4086	0.1348	0.4086	0.0908
Cereals	0.4135	0.2710	0.4156	0.2942	0.4084	0.1077	0.4084	0.1413
Fruits&Vegs	0.4353	0.0166	0.4383	0.0425	0.4085	0.0669	0.4085	0.0408
VP (%)	78.5516	13.5247	76.6619	14.5617	99.7997	0.1460	99.7806	0.1548

Table D.2.6: Ranking orders of parameters for Finland

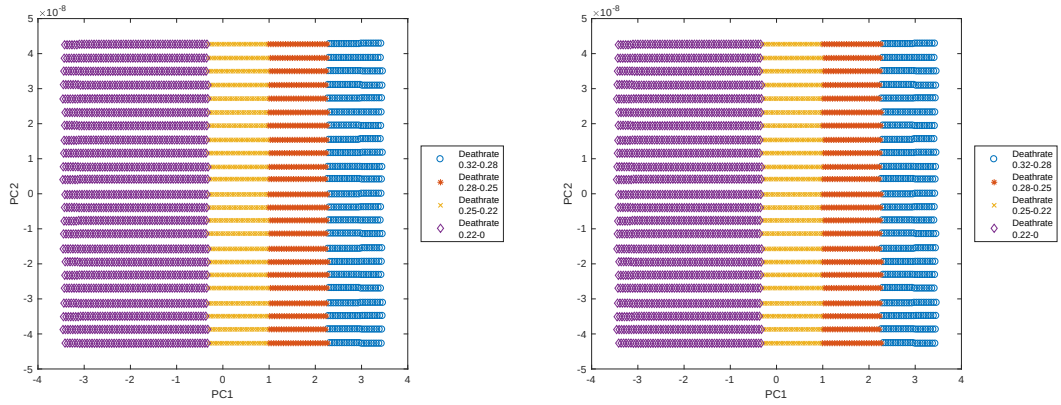
	Ranking orders of 4 negative indicators				Ranking orders of 6 parameters			
	Real Data		Synthetic Data		Real Data		Synthetic Data	
Rank	**male**	**female**	**male**	**female**	**male**	**female**	**male**	**female**
1	SBP	SBP	SBP	SBP	SBP	SBP	SBP	SBP
2	Smoking	Cheese	Smoking	Cheese	F&V	F&V	F&V	F&V
3	Cheese	Smoking	Cheese	Smoking	Cereals	Cheese	Cereals	Cereals
4	Alcohol	Alcohol	Alcohol	Alcohol	Smoking	Cereals	Smoking	Cheese
5	-	-	-	-	Cheese	Smoking	Cheese	Smoking
6	-	-	-	-	Alcohol	Alcohol	Alcohol	Alcohol

Iceland

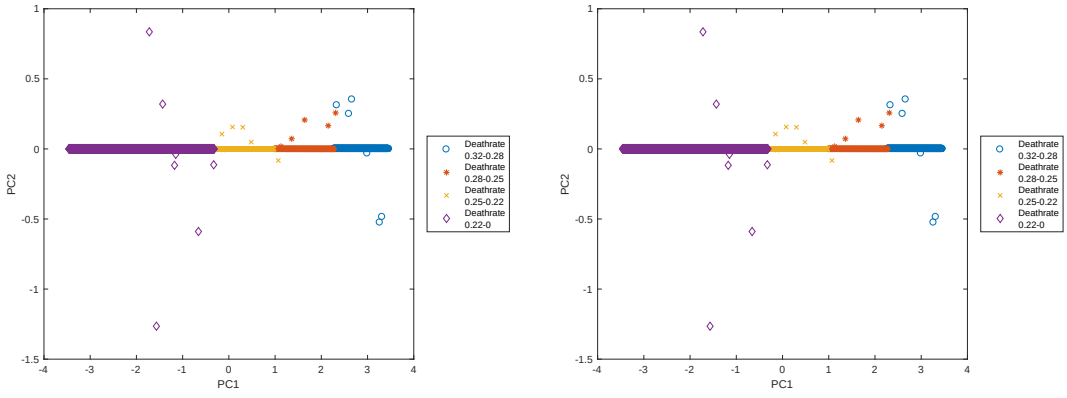


(a) Males (b) Females

Figure D.2.9: PCA visualisation of Iceland real datasets based on 4 negative indicators (i.e. alcohol, cheese, smoking, SBP) for (a) males and (b) females respectively.

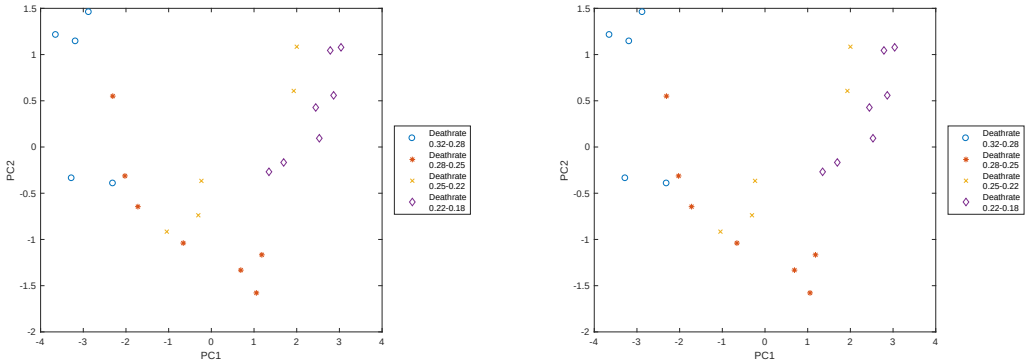


(a) 4 parameters - pure synthetic - males (b) 4 parameters - pure synthetic - females



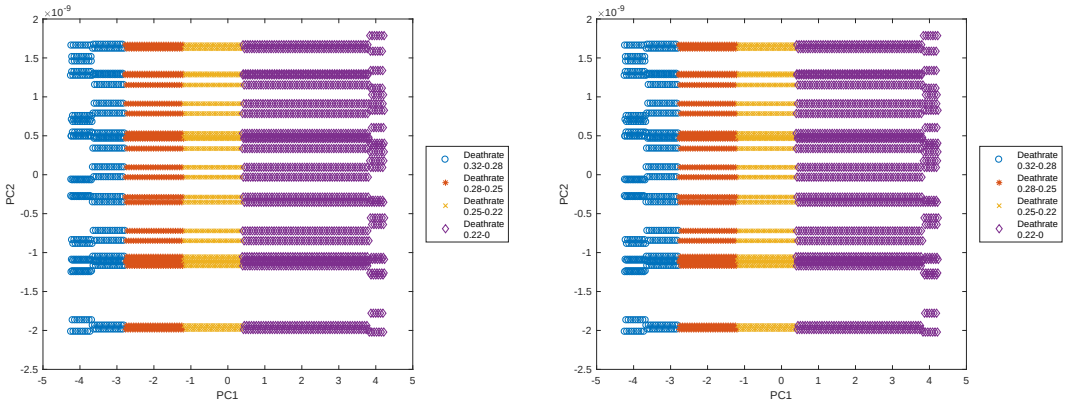
(c) 4 parameters - real-synthetic - males (d) 4 parameters - real-synthetic - females

Figure D.2.10: PCA visualisation of Iceland synthetic datasets based on 4 negative indicators (i.e. alcohol, cheese, smoking, SBP). Figs (a), (b) are visualised with the pure synthetic datasets; Figs (c), (d) are visualised with the real-synthetic datasets.

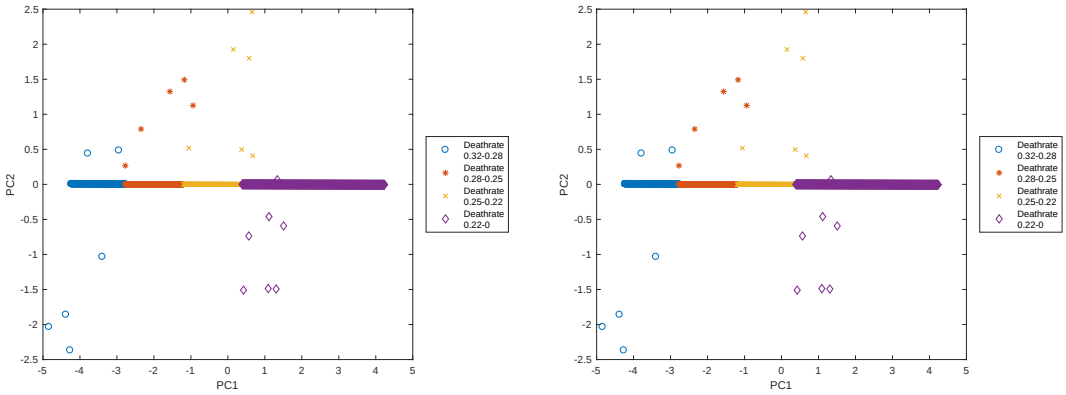


(a) Males (b) Females

Figure D.2.11: PCA visualisation of generated Iceland real datasets based on all 6 parameters (i.e. alcohol, cheese, smoking, SBP, cereals, fruits and vegs) for (a) males and (b) females respectively.



(a) 4 parameters - pure synthetic - males (b) 4 parameters - pure synthetic - females



(c) 4 parameters - real-synthetic - males (d) 4 parameters - real-synthetic - females

Figure D.2.12: PCA visualisation of Iceland synthetic datasets based on 6 parameters (i.e. alcohol, cheese, smoking, SBP, cereals, fruits and vegs). Figs (a), (b) are visualised with the pure synthetic datasets; Figs (c), (d) are visualised with the real-synthetic datasets.

Table D.2.7: Component Matrix of 4 parameters for Iceland

	Component Matrix - real data				Component Matrix - synthetic data			
	Male		Female		Male		Female	
	PC1	PC2	PC1	PC2	PC1	PC2	PC1	PC2
Alcohol	0.4834	0.8722	0.4806	0.8538	0.49991	0.86385	0.49989	0.84653
Cheese	0.5058	0.2287	0.4993	0.4535	0.50003	0.24371	0.50000	0.45421
Smoking	0.5019	0.2569	0.5102	0.1919	0.50001	0.27672	0.50006	0.20633
SBP	0.5086	0.3479	0.5094	0.1688	0.50005	0.34321	0.50005	0.18576
VP (%)	90.3762	5.0959	93.9654	4.4522	99.9432	0.0315	99.9639	0.0275

Table D.2.8: Component Matrix of 6 parameters for Iceland

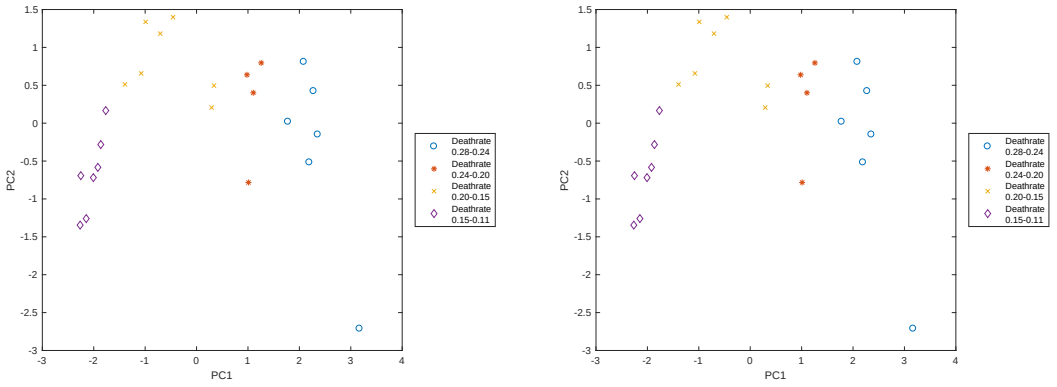
	Component Matrix - real data				Component Matrix - synthetic data			
	Male		Female		Male		Female	
	PC1	PC2	PC1	PC2	PC1	PC2	PC1	PC2
Alcohol	0.4195	0.1292	0.4171	0.0989	0.4085	0.1835	0.4085	0.1759
Cheese	0.4426	0.0152	0.4374	0.0095	0.4087	0.1253	0.4087	0.1178
Smoking	0.4232	0.2426	0.4413	0.1250	0.4085	0.2505	0.4087	0.1884
SBP	0.4437	0.0569	0.4417	0.1040	0.4087	0.0886	0.4087	0.1765
Cereals	0.2626	0.9277	0.2492	0.9537	0.4065	0.9038	0.4064	0.9099
Fruits&Vegs	0.4280	0.2459	0.4275	0.2325	0.4085	0.2513	0.4085	0.2464
VP (%)	79.6454	12.9275	82.1751	12.6835	99.7217	0.2342	99.7338	0.2355

Table D.2.9: Ranking orders of parameters for Iceland

	Ranking orders of 4 negative indicators				Ranking orders of 6 parameters			
	Real Data		Synthetic Data		Real Data		Synthetic Data	
Rank	**male**	**female**	**male**	**female**	**male**	**female**	**male**	**female**
1	SBP	Smoking	SBP	Smoking	SBP	SBP	SBP	Cheese
2	Cheese	SBP	Cheese	SBP	Cheese	Smoking	Cheese	SBP
3	Smoking	Cheese	Smoking	Cheese	F&V	Cheese	F&V	Smoking
4	Alcohol	Alcohol	Alcohol	Alcohol	Smoking	F&V	Alcohol	F&V
5	-	-	-	-	Alcohol	Alcohol	Smoking	Alcohol
6	-	-	-	-	Cereals	Cereals	Cereals	Cereals

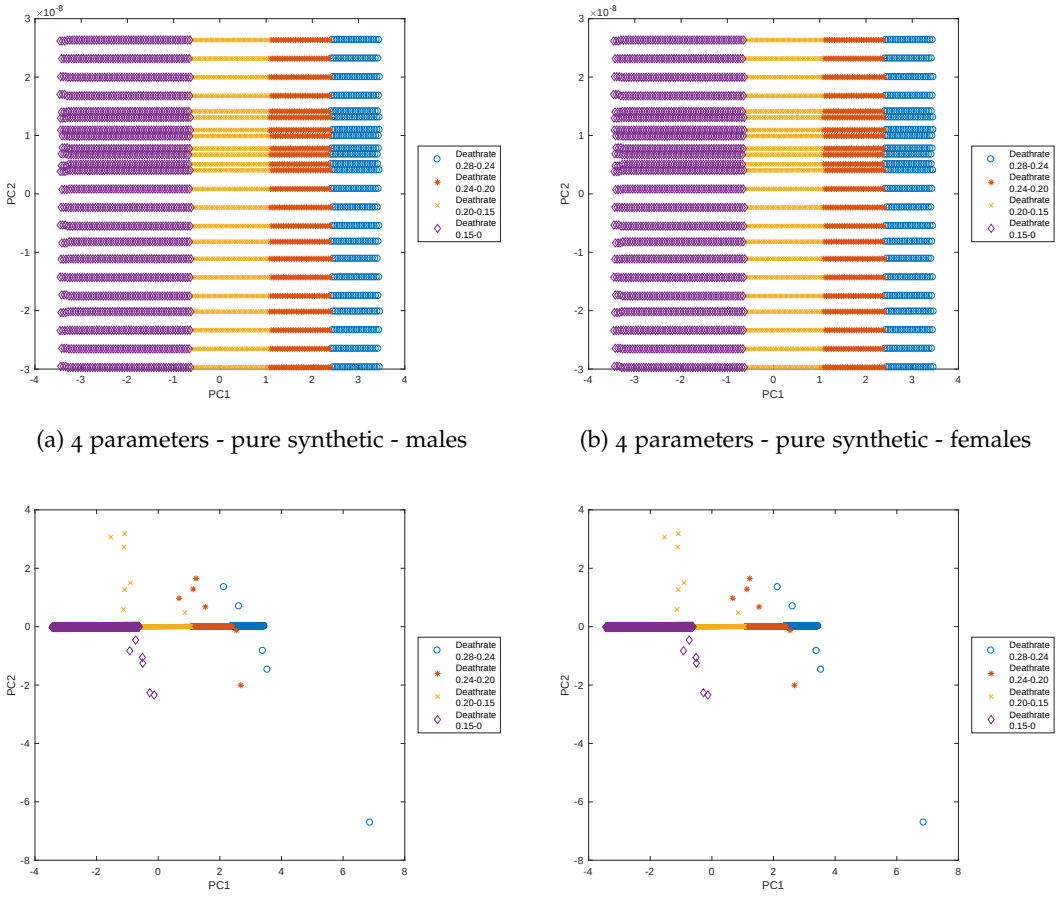
negative indicators

Norway



(a) Males (b) Females

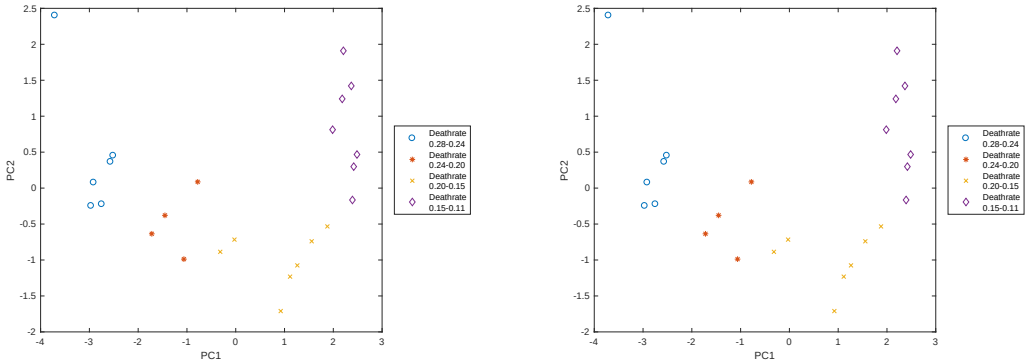
Figure D.2.13: PCA visualisation of Norway real datasets based on 4 negative indicators (i.e. alcohol, cheese, smoking, SBP) for (a) males and (b) females respectively.



(a) 4 parameters - pure synthetic - males (b) 4 parameters - pure synthetic - females

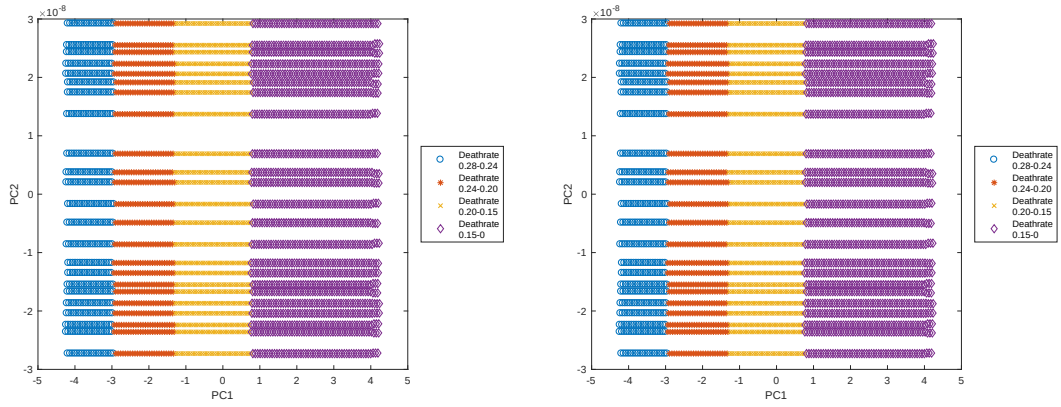
(c) 4 parameters - real-synthetic - males (d) 4 parameters - real-synthetic - females

Figure D.2.14: PCA visualisation of Norway synthetic datasets based on 4 negative indicators (i.e. alcohol, cheese, smoking, SBP). Figs (a), (b) are visualised with the pure synthetic datasets; Figs (c), (d) are visualised with the real-synthetic datasets.

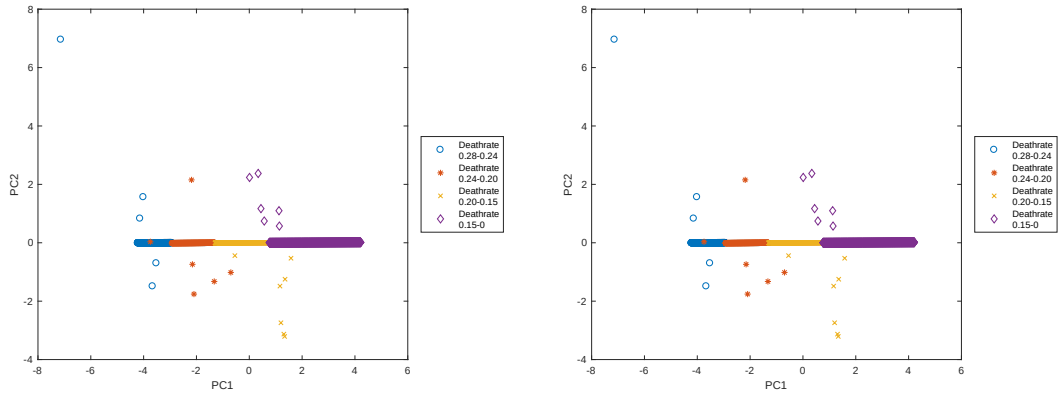


(a) Males (b) Females

Figure D.2.15: PCA visualisation of generated Norway real datasets based on all 6 parameters (i.e. alcohol, cheese, smoking, SBP, cereals, fruits and vegs) for (a) males and (b) females respectively.



(a) 4 parameters - pure synthetic - males (b) 4 parameters - pure synthetic - females



(c) 4 parameters - real-synthetic - males (d) 4 parameters - real-synthetic - females

Figure D.2.16: PCA visualisation of Norway synthetic datasets based on 6 parameters (i.e. alcohol, cheese, smoking, SBP, cereals, fruits and vegs). Figs (a), (b) are visualised with the pure synthetic datasets; Figs (c), (d) are visualised with the real-synthetic datasets.

Table D.2.10: Component Matrix of 4 parameters for Norway

	Component Matrix - real data				Component Matrix - synthetic data			
	Male		Female		Male		Female	
	PC1	PC2	PC1	PC2	PC1	PC2	PC1	PC2
Alcohol	0.5571	0.0152	0.5589	0.0229	0.50169	0.23048	0.50171	0.22790
Cheese	0.2612	0.9455	0.2577	0.9440	0.49556	0.86655	0.49554	0.86628
Smoking	0.5563	0.2478	0.5508	0.2645	0.50133	0.32012	0.50128	0.32650
SBP	0.5585	0.2106	0.5638	0.1958	0.50140	0.30577	0.50144	0.30168
VP (%)	74.9757	22.1385	74.7169	22.3678	99.1009	0.8822	99.0958	0.8869

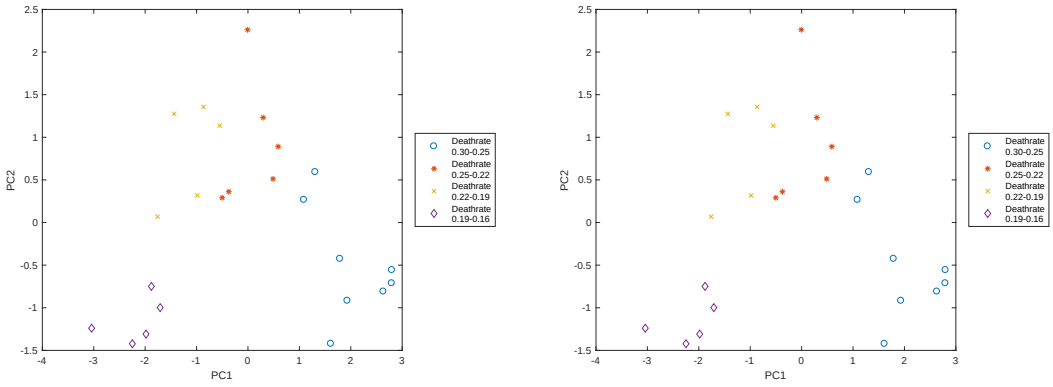
Table D.2.11: Component Matrix of 6 parameters for Norway

	Component Matrix - real data				Component Matrix - synthetic data			
	Male		Female		Male		Female	
	PC1	PC2	PC1	PC2	PC1	PC2	PC1	PC2
Alcohol	0.4669	0.0149	0.4697	0.0053	0.4095	0.1668	0.4095	0.1661
Cheese	0.2269	0.8035	0.2260	0.7740	0.4035	0.9034	0.4035	0.9022
Smoking	0.4459	0.3221	0.4377	0.3585	0.4092	0.2512	0.4091	0.2589
SBP	0.4522	0.2593	0.4586	0.2449	0.4093	0.2377	0.4093	0.2360
Cereals	0.3405	0.4113	0.3342	0.4468	0.4086	0.0509	0.4086	0.0445
Fruits&Vegs	0.4596	0.1186	0.4634	0.1134	0.4094	0.1842	0.4094	0.1839
VP (%)	72.8903	16.4426	72.4556	16.9768	99.2470	0.6188	99.2411	0.6212

Table D.2.12: Ranking orders of parameters for Norway

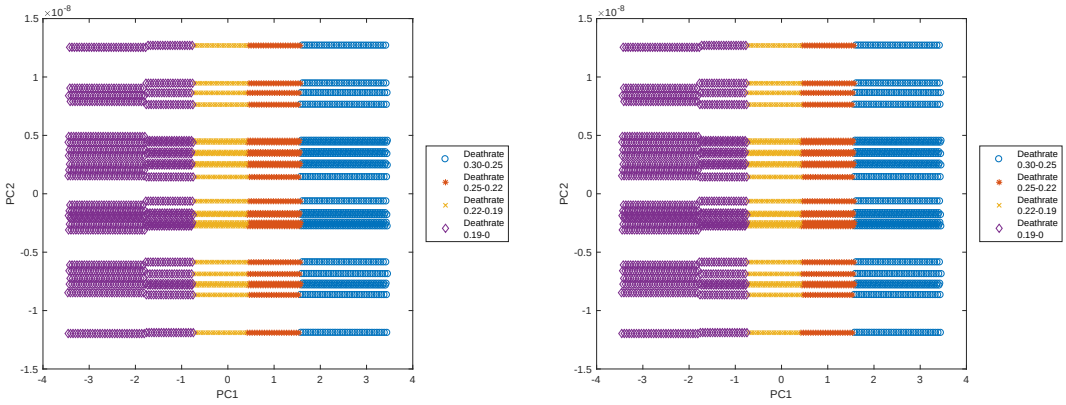
	Ranking orders of 4 negative indicators				Ranking orders of 6 parameters			
	Real Data		Synthetic Data		Real Data		Synthetic Data	
Rank	**male**	**female**	**male**	**female**	**male**	**female**	**male**	**female**
1	SBP	SBP	Alcohol	Alcohol	Alcohol	Alcohol	Alcohol	Alcohol
2	Alcohol	Alcohol	SBP	SBP	F&V	F&V	F&V	F&V
3	Smoking	Smoking	Smoking	Smoking	SBP	SBP	SBP	SBP
4	Cheese	Cheese	Cheese	Cheese	Smoking	Smoking	Smoking	Smoking
5	-	-	-	-	Cereals	Cereals	Cereals	Cereals
6	-	-	-	-	Cheese	Cheese	Cheese	Cheese

Sweden

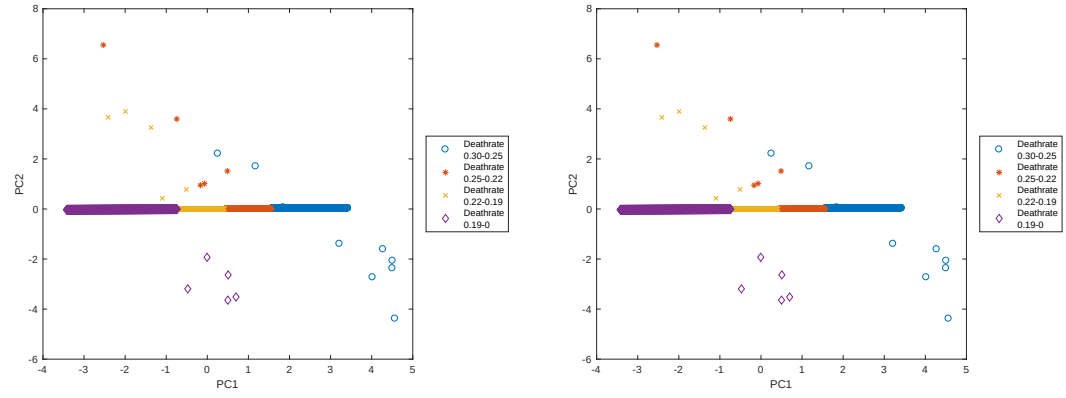


(a) Males (b) Females

Figure D.2.17: PCA visualisation of Sweden real datasets based on 4 negative indicators (i.e. alcohol, cheese, smoking, SBP) for (a) males and (b) females respectively.

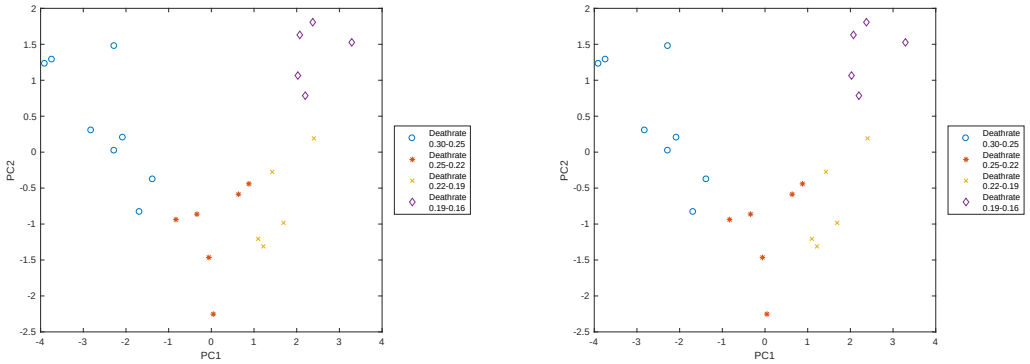


(a) 4 parameters - pure synthetic - males (b) 4 parameters - pure synthetic - females



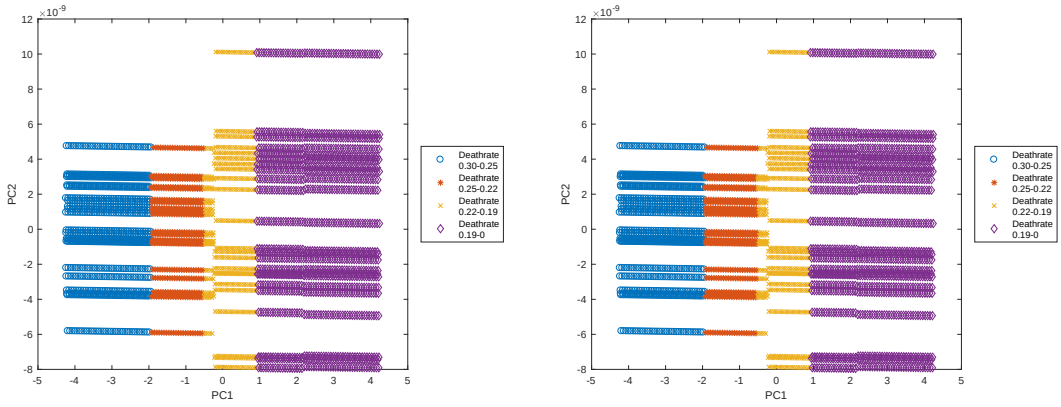
(c) 4 parameters - real-synthetic - males (d) 4 parameters - real-synthetic - females

Figure D.2.18: PCA visualisation of Sweden synthetic datasets based on 4 negative indicators (i.e. alcohol, cheese, smoking, SBP). Figs (a), (b) are visualised with the pure synthetic datasets; Figs (c), (d) are visualised with the real-synthetic datasets.

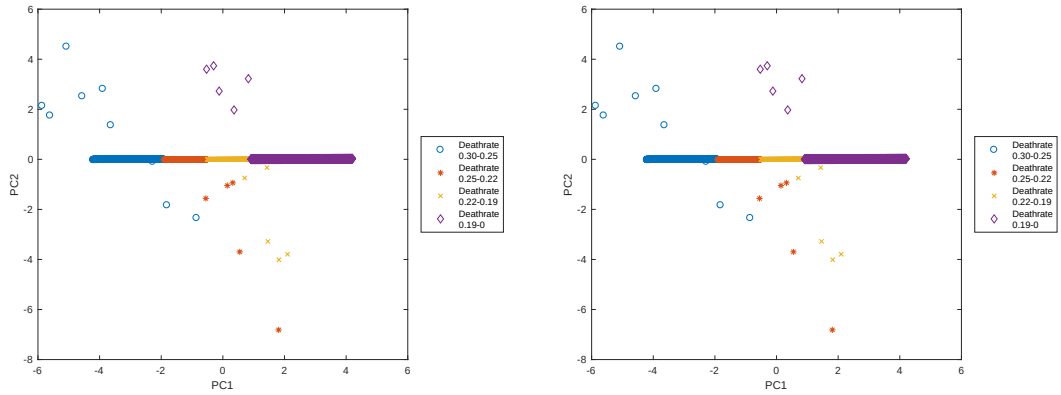


(a) Males (b) Females

Figure D.2.19: PCA visualisation of generated Sweden real datasets based on all 6 parameters (i.e. alcohol, cheese, smoking, SBP, cereals, fruits and vegs) for (a) males and (b) females respectively.



(a) 4 parameters - pure synthetic - males (b) 4 parameters - pure synthetic - females



(c) 4 parameters - real-synthetic - males (d) 4 parameters - real-synthetic - females

Figure D.2.20: PCA visualisation of Sweden synthetic datasets based on 6 parameters (i.e. alcohol, cheese, smoking, SBP, cereals, fruits and vegs). Figs (a), (b) are visualised with the pure synthetic datasets; Figs (c), (d) are visualised with the real-synthetic datasets.

Table D.2.13: Component Matrix of 4 parameters for Sweden

	Component Matrix - real data				Component Matrix - synthetic data			
	Male		Female		Male		Female	
	PC1	PC2	PC1	PC2	PC1	PC2	PC1	PC2
Alcohol	0.1695	0.9457	0.1218	0.9885	0.49161	0.86813	0.49135	0.87070
Cheese	0.5671	0.1561	0.5667	0.0751	0.50262	0.29938	0.50280	0.28567
Smoking	0.5761	0.1256	0.5789	0.0046	0.50331	0.22144	0.50304	0.26533
SBP	0.5637	0.2558	0.5734	0.1311	0.50237	0.32815	0.50271	0.29980
VP (%)	70.8055	25.5592	72.7835	24.4752	98.3438	1.6350	98.3183	1.6664

Table D.2.14: Component Matrix of 6 parameters for Sweden

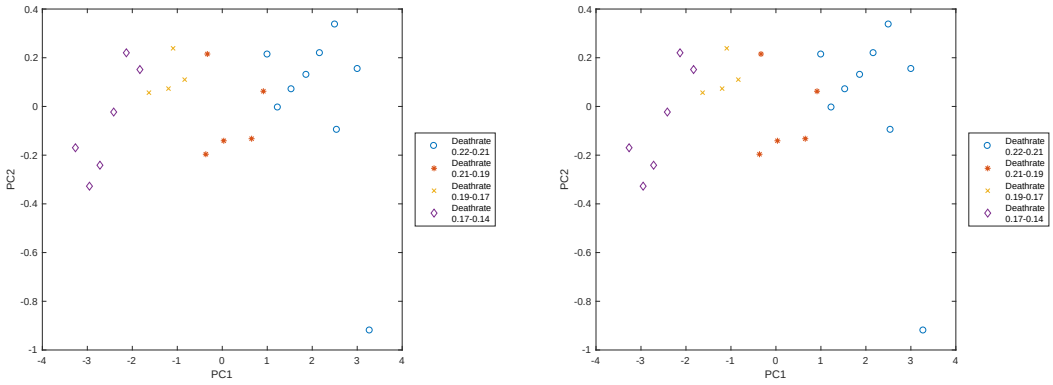
	Component Matrix - real data				Component Matrix - synthetic data			
	Male		Female		Male		Female	
	PC1	PC2	PC1	PC2	PC1	PC2	PC1	PC2
Alcohol	0.1686	0.7956	0.1442	0.8115	0.3991	0.9003	0.3989	0.9012
Cheese	0.4590	0.1905	0.4623	0.1424	0.4100	0.2307	0.4101	0.2222
Smoking	0.4700	0.0352	0.4707	0.1092	0.4104	0.1548	0.4102	0.2024
SBP	0.4463	0.3141	0.4602	0.2291	0.4098	0.2587	0.4100	0.2362
Cereals	0.3596	0.4452	0.3451	0.4881	0.4101	0.0198	0.4101	0.0120
Fruits&Vegs	0.4599	0.1809	0.4616	0.1360	0.4101	0.2122	0.4101	0.2038
VP (%)	70.5006	20.7978	71.3603	20.6100	98.7472	1.1583	98.7353	1.1736

Table D.2.15: Ranking orders of parameters for Sweden

	Ranking orders of 4 negative indicators				Ranking orders of 6 parameters			
	Real Data		Synthetic Data		Real Data		Synthetic Data	
Rank	**male**	**female**	**male**	**female**	**male**	**female**	**male**	**female**
1	Smoking	Smoking	Smoking	Smoking	Smoking	Smoking	Smoking	Smoking
2	Cheese	SBP	Cheese	Cheese	F&V	Cheese	Cereals	F&V
3	SBP	Cheese	SBP	SBP	Cheese	F&V	F&V	Cereals
4	Alcohol	Alcohol	Alcohol	Alcohol	SBP	SBP	Cheese	Cheese
5	-	-	-	-	Cereals	Cereals	SBP	SBP
6	-	-	-	-	Alcohol	Alcohol	Alcohol	Alcohol

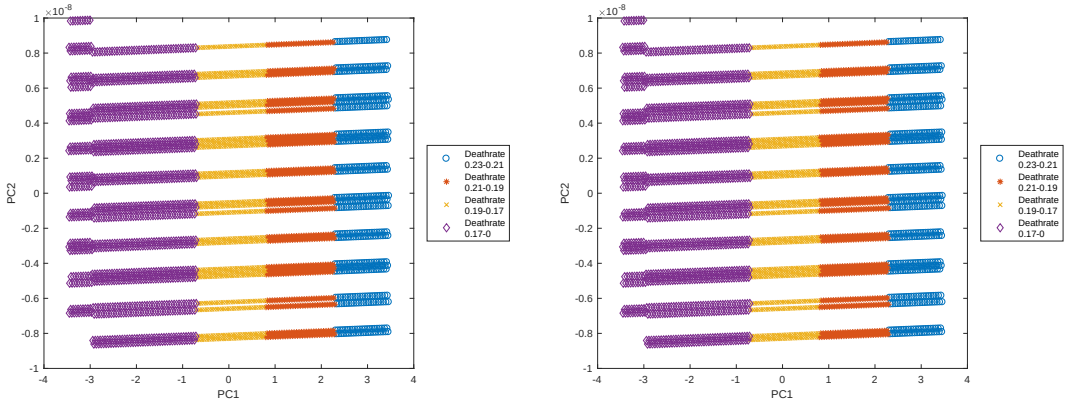
D.3 WESTERN EUROPEAN COUNTRIES (WEEU) BLOCK

Germany

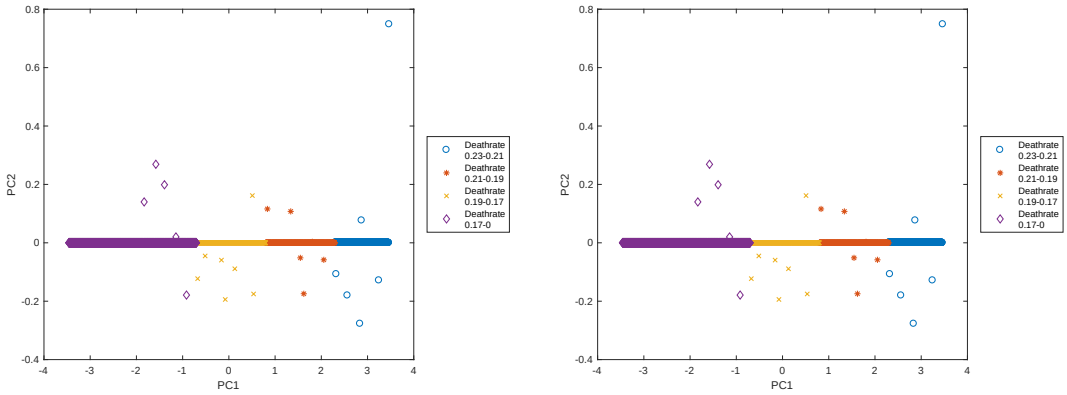


(a) Males (b) Females

Figure D.3.1: PCA visualisation of Germany real datasets based on 4 negative indicators (i.e. alcohol, cheese, smoking, SBP) for (a) males and (b) females respectively.

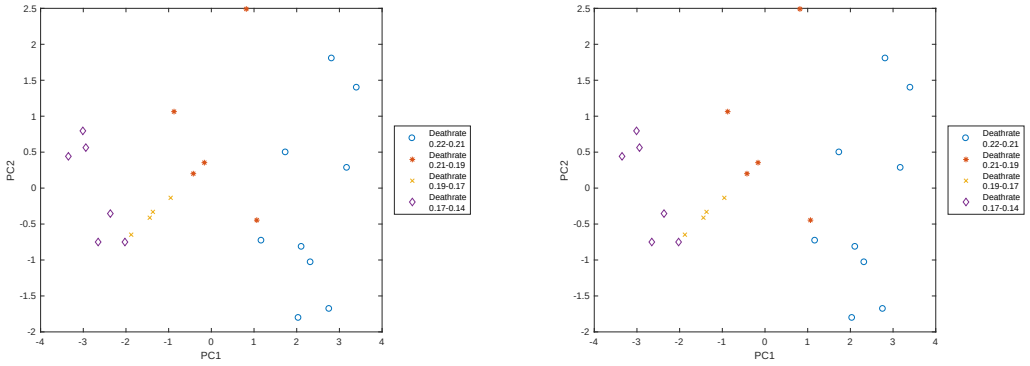


(a) 4 parameters - pure synthetic - males (b) 4 parameters - pure synthetic - females



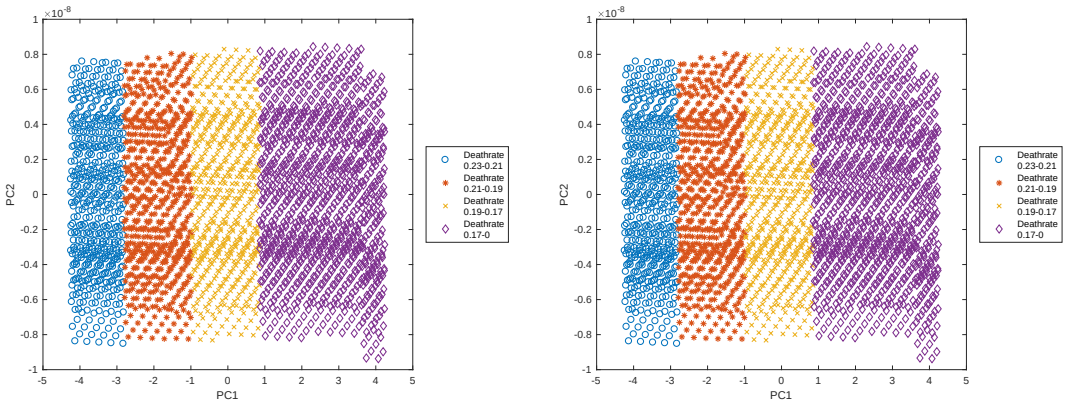
(c) 4 parameters - real-synthetic - males (d) 4 parameters - real-synthetic - females

Figure D.3.2: PCA visualisation of Germany synthetic datasets based on 4 negative indicators (i.e. alcohol, cheese, smoking, SBP). Figs (a), (b) are visualised with the pure synthetic datasets; Figs (c), (d) are visualised with the real-synthetic datasets.

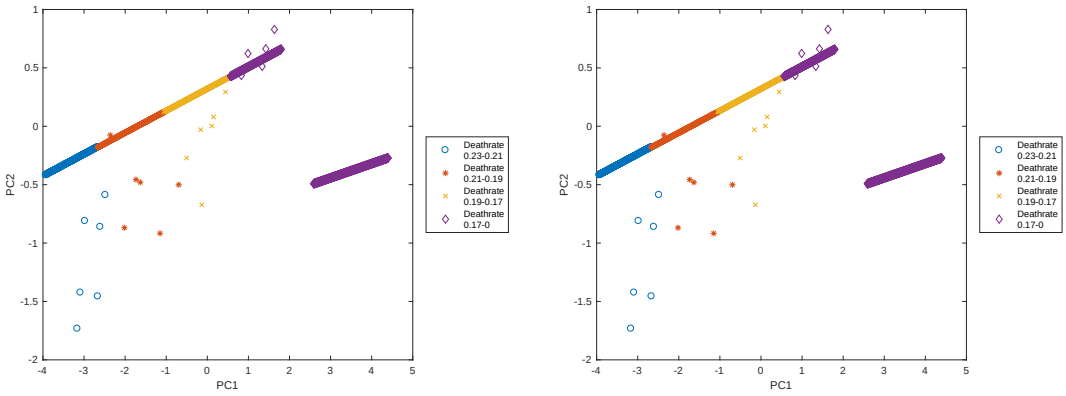


(a) Males (b) Females

Figure D.3.3: PCA visualisation of generated Germany real datasets based on all 6 parameters (i.e. alcohol, cheese, smoking, SBP, cereals, fruits and vegs) for (a) males and (b) females respectively.



(a) 4 parameters - pure synthetic - males (b) 4 parameters - pure synthetic - females



(c) 4 parameters - real-synthetic - males (d) 4 parameters - real-synthetic - females

Figure D.3.4: PCA visualisation of Germany synthetic datasets based on 6 parameters (i.e. alcohol, cheese, smoking, SBP, cereals, fruits and vegs). Figs (a), (b) are visualised with the pure synthetic datasets; Figs (c), (d) are visualised with the real-synthetic datasets.

Table D.3.1: Component Matrix of 4 parameters for Germany

	Component Matrix - real data				Component Matrix - synthetic data			
	Male		Female		Male		Female	
	PC1	PC2	PC1	PC2	PC1	PC2	PC1	PC2
Alcohol	0.4951	0.7763	0.5413	0.2262	0.49997	0.77379	0.50018	0.30272
Cheese	0.4976	0.6094	0.5478	0.1598	0.49999	0.61276	0.50020	0.28438
Smoking	0.50363	0.0003	0.3145	0.9477	0.50002	0.00045	0.49939	0.86623
SBP	0.50357	0.1614	0.5550	0.1588	0.50002	0.16055	0.50024	0.27771
VP (%)	97.5801	1.6433	79.1032	19.1204	99.9872	0.0088	99.8630	0.1273

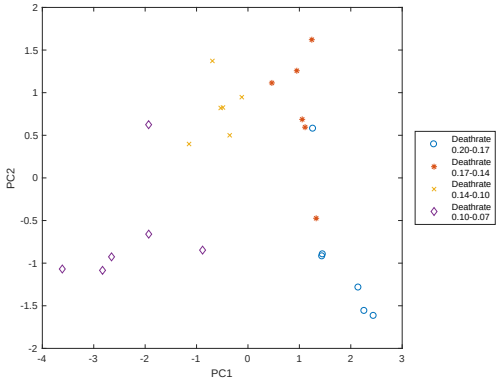
Table D.3.2: Component Matrix of 6 parameters for Germany

	Component Matrix - real data				Component Matrix - synthetic data			
	Male		Female		Male		Female	
	PC1	PC2	PC1	PC2	PC1	PC2	PC1	PC2
Alcohol	0.4495	0.1314	0.4856	0.1636	0.4086	0.2032	0.4085	0.2888
Cheese	0.4528	0.0929	0.4921	0.0621	0.4086	0.1582	0.4086	0.2546
Smoking	0.4566	0.0187	0.2563	0.2230	0.4086	0.1787	0.4083	0.3730
SBP	0.4570	0.0053	0.4976	0.0315	0.4086	0.1733	0.4086	0.2652
Cereals	0.4191	0.0591	0.4567	0.0066	0.4084	0.1952	0.4084	0.3035
Fruits&Vegs	0.0016	0.9850	0.0046	0.9585	0.4066	0.9130	0.4071	0.7417
VP (%)	78.1599	17.1458	65.5996	17.3671	99.7571	0.2107	99.7140	0.2402

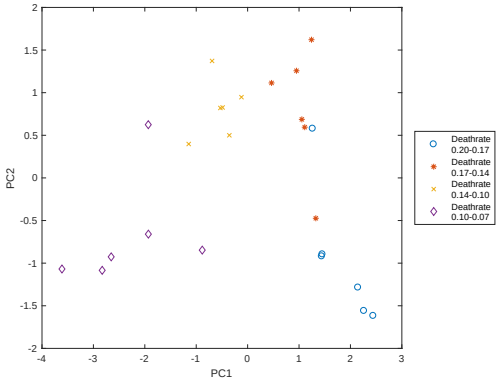
Table D.3.3: Ranking orders of parameters for Germany

	Ranking orders of 4 negative indicators				Ranking orders of 6 parameters			
	Real Data		Synthetic Data		Real Data		Synthetic Data	
Rank	**male**	**female**	**male**	**female**	**male**	**female**	**male**	**female**
1	Smoking	SBP	Smoking	SBP	SBP	SBP	SBP	SBP
2	SBP	Cheese	SBP	Cheese	Smoking	Cheese	Cheese	Cheese
3	Cheese	Alcohol	Cheese	Alcohol	Cheese	Alcohol	Smoking	Alcohol
4	Alcohol	Smoking	Alcohol	Smoking	Alcohol	Cereals	Alcohol	Cereals
5	-	-	-	-	Cereals	Smoking	Cereals	Smoking
6	-	-	-	-	F&V	F&V	F&V	F&V

Netherlands

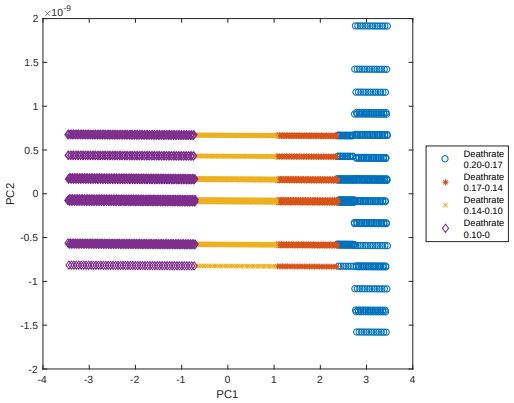


(a) Males

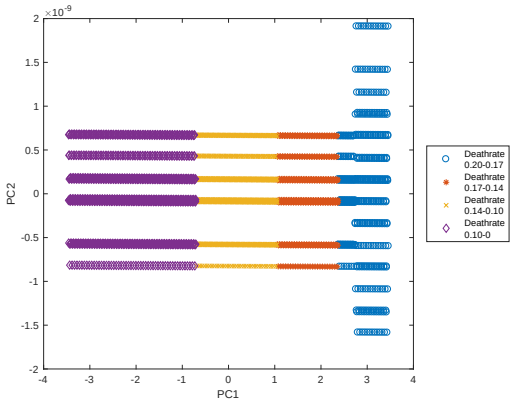


(b) Females

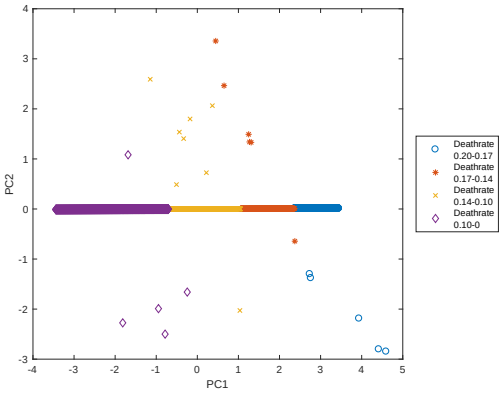
Figure D.3.5: PCA visualisation of Netherlands real datasets based on 4 negative indicators (i.e. alcohol, cheese, smoking, SBP) for (a) males and (b) females respectively.



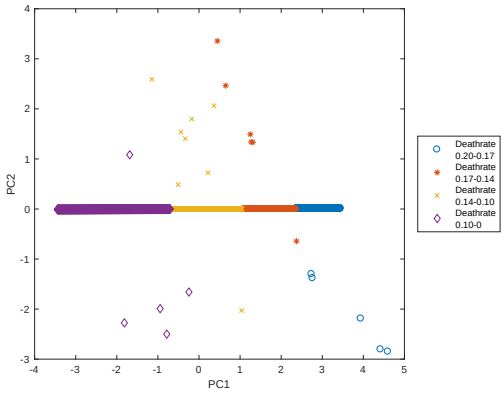
(a) 4 parameters - pure synthetic - males



(b) 4 parameters - pure synthetic - females

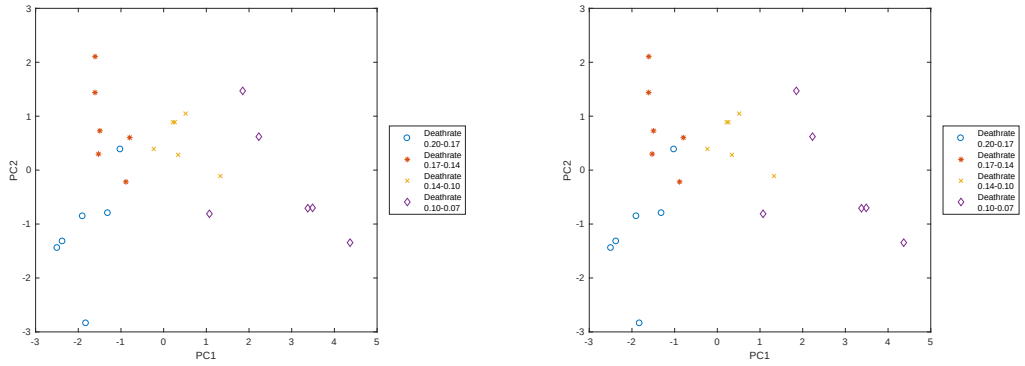


(c) 4 parameters - real-synthetic - males



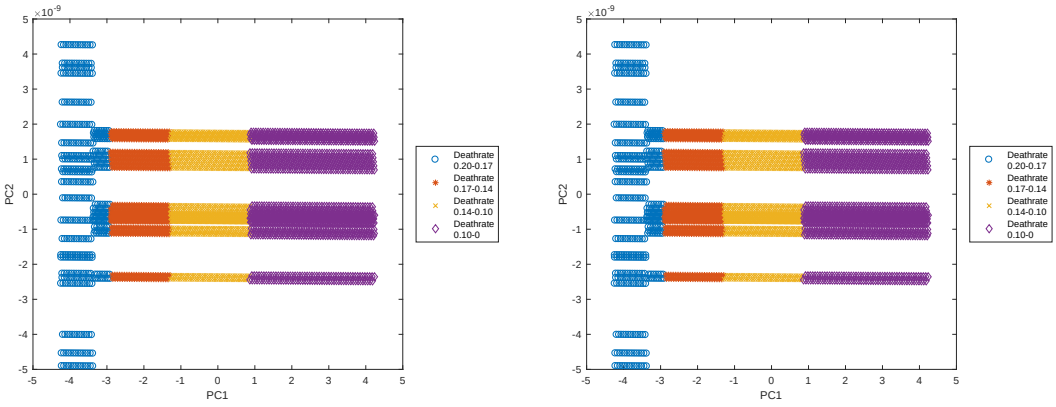
(d) 4 parameters - real-synthetic - females

Figure D.3.6: PCA visualisation of Netherlands synthetic datasets based on 4 negative indicators (i.e. alcohol, cheese, smoking, SBP). Figs (a), (b) are visualised with the pure synthetic datasets; Figs (c), (d) are visualised with the real-synthetic datasets.

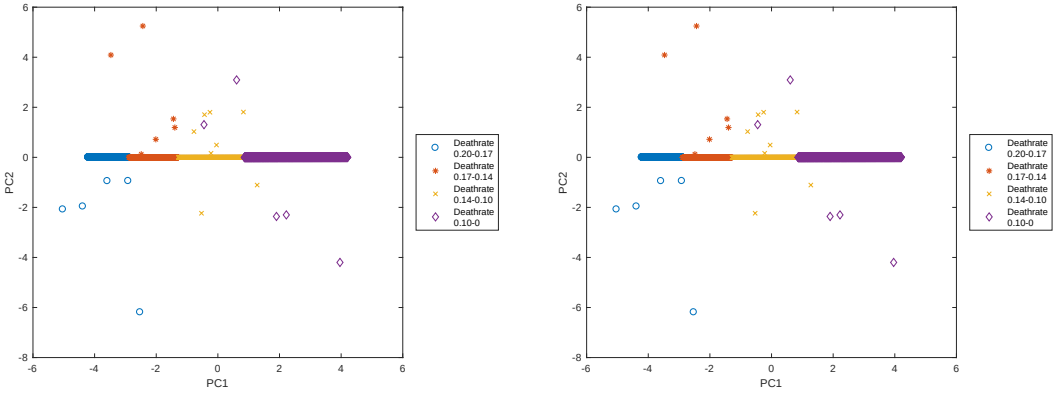


(a) Males (b) Females

Figure D.3.7: PCA visualisation of generated Netherlands real datasets based on all 6 parameters (i.e. alcohol, cheese, smoking, SBP, cereals, fruits and vegs) for (a) males and (b) females respectively.



(a) 4 parameters - pure synthetic - males (b) 4 parameters - pure synthetic - females



(c) 4 parameters - real-synthetic - males (d) 4 parameters - real-synthetic - females

Figure D.3.8: PCA visualisation of Netherlands synthetic datasets based on 6 parameters (i.e. alcohol, cheese, smoking, SBP, cereals, fruits and vegs). Figs (a), (b) are visualised with the pure synthetic datasets; Figs (c), (d) are visualised with the real-synthetic datasets.

Table D.3.4: Component Matrix of 4 parameters for Netherlands

	Component Matrix - real data				Component Matrix - synthetic data			
	Male		Female		Male		Female	
	PC1	PC2	PC1	PC2	PC1	PC2	PC1	PC2
Alcohol	0.5424	0.2878	0.5518	0.2159	0.50072	0.37868	0.50085	0.35927
Cheese	0.1616	0.9481	0.1206	0.9759	0.49633	0.86095	0.49619	0.86457
Smoking	0.5783	0.0988	0.5800	0.0230	0.50158	0.19824	0.50148	0.24061
SBP	0.5876	0.0922	0.5871	0.0203	0.50136	0.27578	0.50146	0.25603
VP (%)	69.8662	25.4741	70.1532	25.0417	99.2101	0.7559	99.1910	0.7744

Table D.3.5: Component Matrix of 6 parameters for Netherlands

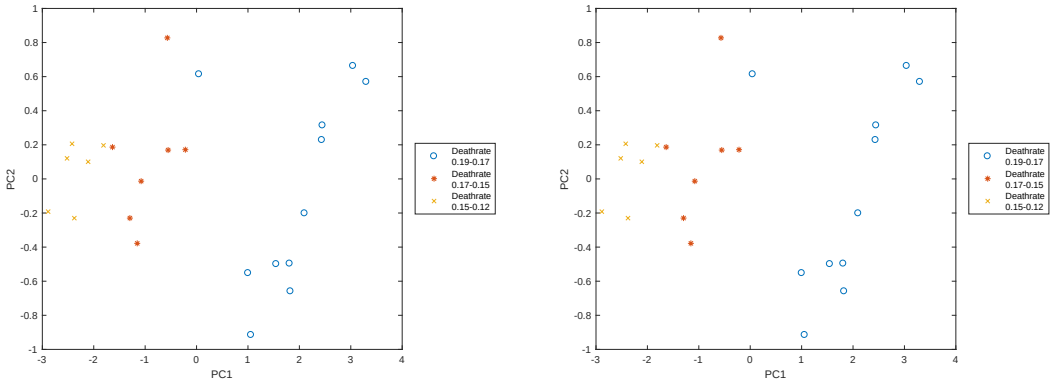
	Component Matrix - real data				Component Matrix - synthetic data			
	Male		Female		Male		Female	
	PC1	PC2	PC1	PC2	PC1	PC2	PC1	PC2
Alcohol	0.4905	0.1110	0.4916	0.0567	0.4103	0.0675	0.4103	0.0609
Cheese	0.0762	0.7931	0.0510	0.7924	0.4041	0.7160	0.4040	0.7263
Smoking	0.4891	0.1668	0.4955	0.0765	0.4106	0.0329	0.4106	0.0041
SBP	0.5143	0.0480	0.5082	0.0877	0.4106	0.0039	0.4106	0.0012
Cereals	0.4410	0.0500	0.4352	0.1176	0.4096	0.0169	0.4096	0.0230
Fruits&Vegs	0.2352	0.5710	0.2502	0.5844	0.4043	0.6938	0.4044	0.6843
VP (%)	60.5034	20.7440	61.0968	20.1344	98.7472	0.8531	98.7461	0.8520

Table D.3.6: Ranking orders of parameters for Netherlands

	Ranking orders of 4 negative indicators				Ranking orders of 6 parameters			
	Real Data		Synthetic Data		Real Data		Synthetic Data	
Rank	**male**	**female**	**male**	**female**	**male**	**female**	**male**	**female**
1	SBP	SBP	Smoking	Smoking	SBP	SBP	Smoking	Smoking
2	Smoking	Smoking	SBP	SBP	Alcohol	Smoking	SBP	SBP
3	Alcohol	Alcohol	Alcohol	Alcohol	Smoking	Alcohol	Alcohol	Alcohol
4	Cheese	Cheese	Cheese	Cheese	Cereals	Cereals	Cereals	Cereals
5	-	-	-	-	F&V	F&V	F&V	F&V
6	-	-	-	-	Cheese	Cheese	Cheese	Cheese

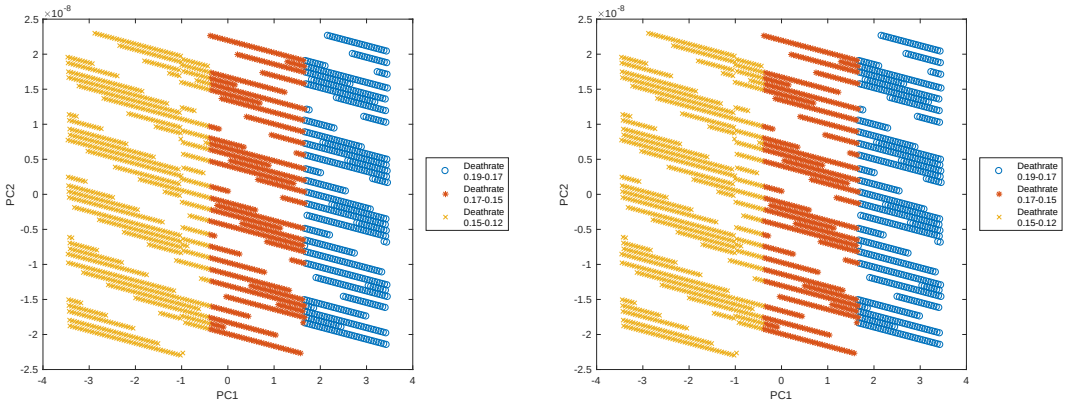
negative indicators

Switzerland

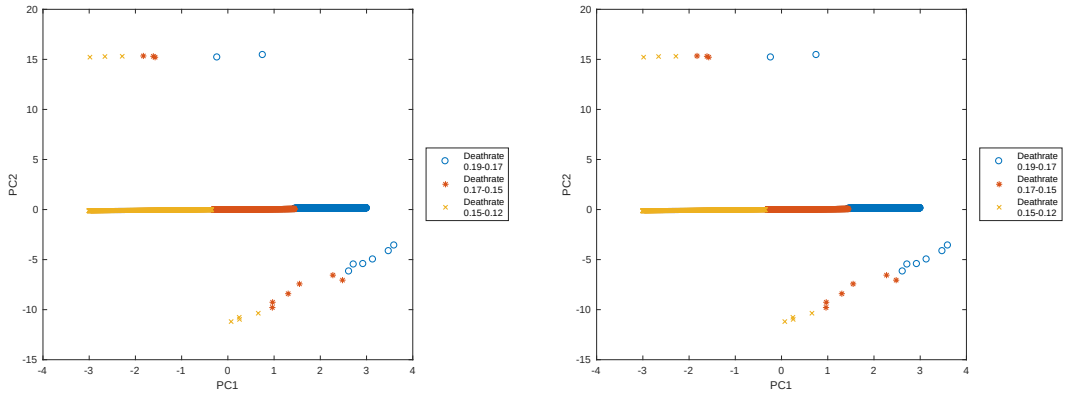


(a) Males (b) Females

Figure D.3.9: PCA visualisation of Switzerland real datasets based on 4 negative indicators (i.e. alcohol, cheese, smoking, SBP) for (a) males and (b) females respectively.

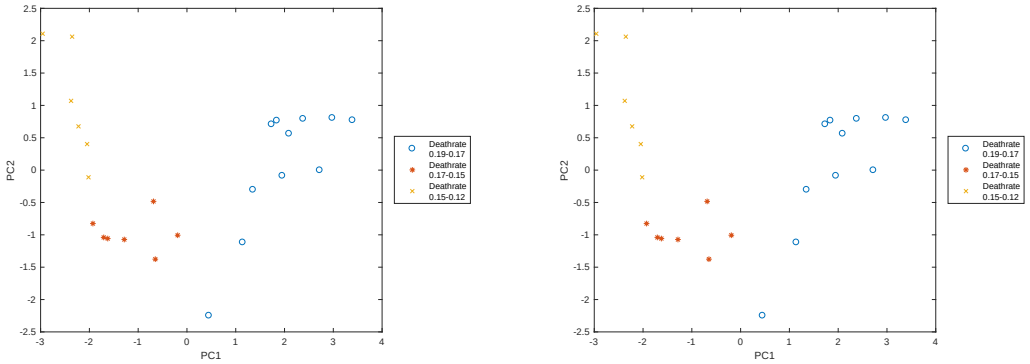


(a) 4 parameters - pure synthetic - males (b) 4 parameters - pure synthetic - females



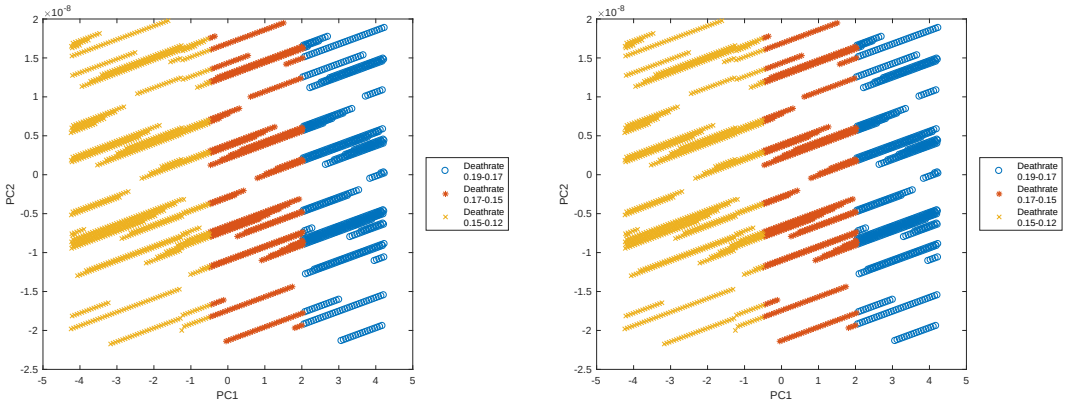
(c) 4 parameters - real-synthetic - males (d) 4 parameters - real-synthetic - females

Figure D.3.10: PCA visualisation of Switzerland synthetic datasets based on 4 negative indicators (i.e. alcohol, cheese, smoking, SBP). Figs (a), (b) are visualised with the pure synthetic datasets; Figs (c), (d) are visualised with the real-synthetic datasets.

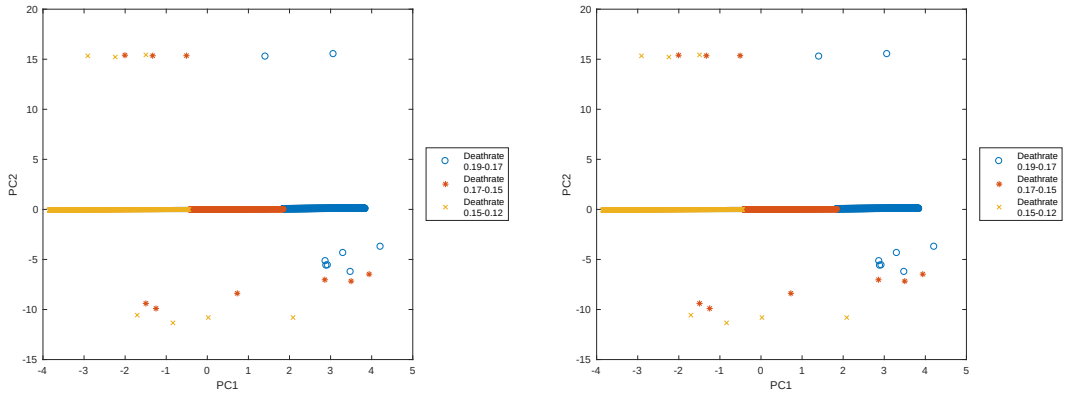


(a) Males (b) Females

Figure D.3.11: PCA visualisation of generated Switzerland real datasets based on all 6 parameters (i.e. alcohol, cheese, smoking, SBP, cereals, fruits and vegs) for (a) males and (b) females respectively.



(a) 4 parameters - pure synthetic - males (b) 4 parameters - pure synthetic - females



(c) 4 parameters - real-synthetic - males (d) 4 parameters - real-synthetic - females

Figure D.3.12: PCA visualisation of Switzerland synthetic datasets based on 6 parameters (i.e. alcohol, cheese, smoking, SBP, cereals, fruits and vegs). Figs (a), (b) are visualised with the pure synthetic datasets; Figs (c), (d) are visualised with the real-synthetic datasets.

Table D.3.7: Component Matrix of 4 parameters for Switzerland

	Component Matrix - real data				Component Matrix - synthetic data			
	Male		Female		Male		Female	
	PC1	PC2	PC1	PC2	PC1	PC2	PC1	PC2
Alcohol	0.4917	0.6908	0.4924	0.7175	0.57280	0.07396	0.57277	0.07428
Cheese	0.4893	0.7060	0.4910	0.6579	0.57304	0.06698	0.57301	0.06730
Smoking	0.5047	0.1057	0.4994	0.1910	0.12350	0.99233	0.12376	0.99230
SBP	0.5139	0.1150	0.5167	0.1261	0.57296	0.07296	0.57295	0.07277
VP (%)	92.8259	4.7645	91.7954	4.8183	75.7546	24.2146	75.7578	24.2112

Table D.3.8: Component Matrix of 6 parameters for Switzerland

	Component Matrix - real data				Component Matrix - synthetic data			
	Male		Female		Male		Female	
	PC1	PC2	PC1	PC2	PC1	PC2	PC1	PC2
Alcohol	0.4693	0.0223	0.4706	0.0343	0.4480	0.0376	0.4480	0.0377
Cheese	0.4674	0.0433	0.4696	0.0544	0.4481	0.0306	0.4481	0.0308
Smoking	0.4830	0.0247	0.4763	0.0451	0.0809	0.9964	0.0810	0.9963
SBP	0.4901	0.0666	0.4938	0.0633	0.4481	0.0366	0.4481	0.0362
Cereals	0.1873	0.7416	0.1920	0.7317	0.4424	0.0567	0.4425	0.0569
Fruits&Vegs	0.2296	0.6653	0.2248	0.6741	0.4420	0.0192	0.4420	0.0195
VP (%)	66.9281	18.9776	66.1438	19.0355	82.7061	16.2437	82.7080	16.2421

Table D.3.9: Ranking orders of parameters for Switzerland

	Ranking orders of 4 negative indicators				Ranking orders of 6 parameters			
	Real Data		Synthetic Data		Real Data		Synthetic Data	
Rank	**male**	**female**	**male**	**female**	**male**	**female**	**male**	**female**
1	SBP	SBP	Cheese	Cheese	SBP	SBP	SBP	SBP
2	Smoking	Smoking	SBP	SBP	Smoking	Smoking	Cheese	Cheese
3	Alcohol	Alcohol	Alcohol	Alcohol	Alcohol	Alcohol	Alcohol	Alcohol
4	Cheese	Cheese	Smoking	Smoking	Cheese	Cheese	Cereals	Cereals
5	-	-	-	-	F&V	F&V	F&V	F&V
6	-	-	-	-	Cereals	Cereals	Smoking	Smoking

BIBLIOGRAPHY

Anderson, P. and B. Baumberg (2006). *Alcohol in Europe*. Luxembourg: European Commission.

Antikainen, Riitta et al. (1998). 'Systolic blood pressure, isolated systolic hypertension and risk of coronary heart disease, strokes, cardiovascular disease and all-cause mortality in the middle-aged population'. In: *Journal of hypertension* 16.5, pp. 577–583.

Aune, Dagfinn et al. (2016). 'Whole grain consumption and risk of cardiovascular disease, cancer, and all cause and cause specific mortality: systematic review and dose-response meta-analysis of prospective studies'. In: *bmj* 353, p. i2716.

Awojoyogbe, O.B. et al. (2011). 'Mathematical models of real geometrical factors in restricted blood vessels for the analysis of CAD (Coronary Artery Diseases) using legendre, boubaker and bessel polynomials'. In: *Journal of medical systems* 35.6, pp. 1513–1520.

Berglee, Royal (2012). *World Regional Geography: People, Places and Globalization*. The Saylor Foundation.

Bien, T.H. and R. Burge (1990). 'Smoking and Drinking: A Review of the Literature'. In: *Substance Use and Misuse* 25.12, pp. 1429–1454.

Bishop, Christopher M et al. (1998). 'GTM: The generative topographic mapping'. In: *Neural computation* 10.1, pp. 215–234.

Boeing, Heiner et al. (2012). 'Critical review: vegetables and fruit in the prevention of chronic diseases'. In: *European Journal of Nutrition* 51.6, pp. 637–663. ISSN: 1436-6215. DOI: 10.1007/s00394-012-0380-y. URL: https://doi.org/10.1007/s00394-012-0380-y.

Dauchet, Luc et al. (2006). 'Fruit and vegetable consumption and risk of coronary heart disease: a meta-analysis of cohort studies'. In: *The Journal of nutrition* 136.10, pp. 2588–2593.

De Oliveira Otto, M.C. et al. (2012). 'Dietary intake of saturated fat by food source and incident cardiovascular disease: The multi-ethnic study of atherosclerosis'. In: *American Journal of Clinical Nutrition* 96.2, pp. 397–404.

Endo, A. (1992). 'The discovery and development of HMG-CoA reductase inhibitors'. In: *Journal of lipid research* 33.11, pp. 1569–1582.

Endo, Akira (2010). 'A historical perspective on the discovery of statins'. In: *Proceedings of the Japan Academy, Series B* 86.5, pp. 484–493.

Friedman, Jerome H (1998). 'Data mining and statistics: What's the connection?' In: *Computing Science and Statistics* 29.1, pp. 3–9.

George, S. and J. Johnson (2010). *Atherosclerosis: Molecular and Cellular Mechanisms*. Wiley-Blackwell.

Gerhard-Herman, M. (2002). 'Atherosclerosis in Women: The Role of Gender'. In: *Cardiology Rounds*, 6.7.

Glantz, Stanton A and William W Parmley (1991). 'Passive smoking and heart disease. Epidemiology, physiology, and biochemistry.' In: *Circulation* 83.1, pp. 1–12.

— (1995). 'Passive smoking and heart disease: mechanisms and risk'. In: *Jama* 273.13, pp. 1047–1053.

Gunzerath L.and Faden, V. et al. (2004). 'National Institute on Alcohol Abuse and Alcoholism report on moderate drinking'. In: *Alcoholism: Clinical and Experimental Research* 28.6, pp. 829–847.

Haidong, W. et al. (2016). 'Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015'. In: *The Lancet* 388.10053, pp. 1545 –1602. ISSN: 0140-6736. DOI: http://dx.doi.org/10.1016/S0140-6736(16)31678-6. URL: http://www.sciencedirect.com/science/article/pii/S0140673616316786.

Han, H. et al. (2013). 'Association of genetic polymorphisms in ADH and ALDH2 with risk of coronary artery disease and myocardial infarction: A meta-analysis'. In: *Gene* 526.2, pp. 134–141.

Hansson, Göran K. (2005). 'Inflammation, Atherosclerosis, and Coronary Artery Disease'. In: *New England Journal of Medicine* 352.16, pp. 1685–1695.

He, Jiang et al. (1999). 'Passive smoking and the risk of coronary heart disease?a meta-analysis of epidemiologic studies'. In: *New England Journal of Medicine* 340.12, pp. 920–926.

Hooper, L. et al. (2012). 'Reduced or modified dietary fat for preventing cardiovascular disease'. In: *Cochrane database of systematic reviews* 5.

Hotelling, H. (1933). 'Analysis of a Complex of Statistical Variables into Principal Components'. In: *Journal of Educational Psychology* 24.6, pp. 417–441.

Ikehara, S. et al. (2013). 'Alcohol consumption and risk of stroke and coronary heart disease among Japanese women: The Japan Public Health Center-based prospective study'. In: *Preventive medicine*.

Isles, CG et al. (1992). 'Relation between coronary risk and coronary mortality in women of the Renfrew and Paisley survey: comparison with men'. In: *Lancet* 339.8795, pp. 702–706.

Jolliffe, Ian T (1986). 'Principal Component Analysis and Factor Analysis'. In: *Principal component analysis*. Springer, pp. 115–128.

Kalin, M.F. and B Zumoff (1990). 'Sex hormones and coronary disease: A review of the clinical studies'. In: *Steroids* 55.8, pp. 330–352.

Kannel, WB et al. (1980). 'Perspectives on systolic hypertension. The Framingham study.' In: *Circulation* 61.6, pp. 1179–1182.

Kim, Seoung Bum and Panaya Rattakorn (2011). 'Unsupervised feature selection using weighted principal components'. In: *Expert systems with applications* 38.5, pp. 5704–5710.

Kim, Youngyo and Youjin Je (2016). 'Dietary fibre intake and mortality from cardiovascular disease and all cancers: A meta-analysis of prospective cohort studies'. In: *Archives of Cardiovascular Diseases* 109.1, pp. 39 –54. ISSN: 1875-2136. DOI: http://dx.doi.org/10.1016/j.acvd.2015.09.005. URL: http://www.sciencedirect.com/science/article/pii/S1875213615001783.

King, Russell et al. (2014). *The Mediterranean: environment and society*. Routledge.

Klatsky, A.L. (1999). 'Moderate drinking and reduced risk of heart disease'. In: *Alcohol Research and Health* 23.1, pp. 15–22.

Kritz, Harald et al. (1995). 'Passive smoking and cardiovascular risk'. In: *Archives of internal medicine* 155.18, pp. 1942–1948.

Kruskal, J. B. (1964). 'Nonmetric multidimensional scaling: A numerical method'. In: *Psychometrika* 29.2, pp. 115–129.

Law, Malcolm R et al. (1997). 'Environmental tobacco smoke exposure and ischaemic heart disease: an evaluation of the evidence'. In: *Bmj* 315.7114, pp. 973–980.

Lawrence, N. D. (2008). *Large Scale Learning with the Gaussian Process Latent Variable Model*. Tech. rep. ftp://ftp.dcs.shef.ac.uk/home/neil/gplvmSparse.pdf. United Kingdom: University of Sheffield.

Lawrence, N. D. et al. (2003). 'Fast Sparse Gaussian Process Methods: The Informative Vector Machine'. In: *Advances in Neural Information Processing Systems 15*. MIT Press, pp. 609–616.

Lawson, CL and RJ Hanson (1995). *Solving Least Square Problems. Classics in Applied Mathematics SIAM*.

Lee, J. A. and M. Verleysen (2008). 'Rank-based quality assessment of nonlinear dimensionality reduction.' In: *ESANN*, pp. 49–54.

Lee, James (1986). 'An Insight on the Use of Multiple Logistic Regression Analysis to Estimate Association between Risk Factor and Disease Occurrence'. In: 15.1, pp. 22–29.

Lerner, Debra J and William B Kannel (1986). 'Patterns of coronary heart disease morbidity and mortality in the sexes: A 26-year follow-up of the Framingham population'. In: *American Heart Journal* 111.2, pp. 383 –390.

Liu, Simin et al. (2002). 'A prospective study of dietary fiber intake and risk of cardiovascular disease among women'. In: *Journal of the American College of Cardiology* 39.1, pp. 49 –56. ISSN: 0735-1097. DOI: http://dx.doi.org/10.1016/S0735-1097(01)01695-3. URL: http://www.sciencedirect.com/science/article/pii/S0735109701016953.

López-Candales (2002). 'Cardiovascular diseases: A review of the Hispanic perspective. Awareness is the first step to action'. In: *Journal of medicine* 33.1-4, pp. 227–245.

Lowe, D. and M. E. Tipping (1996). 'NeuroScale: Novel Topographic Feature Extraction using RBF Networks'. In: *NIPS*, pp. 543–549.

Lowe, David and Michael E. Tipping (1997). 'NeuroScale: Novel Topographic Feature Extraction using RBF Networks'. In: *Advances in Neural Information Processing*

Systems 9. Ed. by M. C. Mozer et al. MIT Press, pp. 543–549. URL: http://papers.nips.cc/paper/1323-neuroscale-novel-topographic-feature-extraction-using-rbf-networks.pdf.

Matsuda, Hirotsugu et al. (1992). 'Statistical mechanics of population: the lattice Lotka-Volterra model'. In: *Progress of theoretical Physics* 88.6, pp. 1035–1049.

Møller, L. et al. (2010). *European status report on alcohol and health 2010*. Copenhagen, Denmark: WHO, Regional Office for Europe.

Mozaffarian, D. et al. (2010). 'Effects on coronary heart disease of increasing polyunsaturated fat in place of saturated fat: A systematic review and meta-analysis of randomized controlled trials'. In: *PLoS Medicine* 7.3, pp. 1–10.

Murray, J.D. (2002). *Mathematical biology*. Third. Springer-Verlag New York.

Nhs.uk (2013). *Atherosclerosis - NHS Choices*. URL: http://www.nhs.uk/conditions/atherosclerosis/Pages/Introduction.aspx.

Nichols, M. et al. (2012). *European Cardiovascular Disease Statistics*. European Heart Network and European Society of Cardiology.

PHE (2017). *The Eatwell Guide - GOV.UK*. URL: https://www.gov.uk/government/publications/the-eatwell-guide.

Pal, Sankar K and Pabitra Mitra (2004). *Pattern recognition algorithms for data mining*. CRC press.

Parish, S. et al. (2000). 'Cigarette smoking, tar yields and non-fatal myocardial infarct: 14 000 cases and 32 000 controls in the United Kingdom'. In: *Tobacco: The Growing Epidemic: Proceedings of the Tenth World Conference on Tobacco or Health, 24–28 August 1997, Beijing, China*. Ed. by Rushan Lu et al. London: Springer London, pp. 111–111. ISBN: 978-1-4471-0769-9. DOI: 10.1007/978-1-4471-0769-9_40. URL: https://doi.org/10.1007/978-1-4471-0769-9_40.

Pearson, K. (1901). 'On lines and planes of closest fit to systems of points in space'. In: *Philosophical Magazine* 2, pp. 559–572.

Phillips-fit (2013). *Reduced blood cholesterol levels*. URL: http://www.phillips-fit.co.uk/personaltraining/reduced-blood-cholesterol-levels.

Plackett, Ronald L (1950). 'Some theorems in least squares'. In: *Biometrika* 37.1/2, pp. 149–157.

Ross, R. (1993). 'The pathogenesis of atherosclerosis: A perspective for the 1990s'. In: *Nature* 362.6423, pp. 801–809.

Rossi, Pietro (2015). *The boundaries of Europe: from the fall of the ancient world to the age of decolonisation*. Vol. 1. Walter de Gruyter GmbH & Co KG.

Russell, Jesse and Ronald Cohn (2012). *Atherosclerosis*. Bookvika.

Sammon, J. (1969). 'A nonlinear mapping for data structure analysis'. In: *IEEE Transactions on Computers* 18, pp. 401–409.

Silva, Vin De and Joshua B. Tenenbaum (2003). 'Global versus local methods in nonlinear dimensionality reduction'. In: *Advances in Neural Information Processing Systems 15*. MIT Press, pp. 705–712.

Slijkhuis, W. et al. (2009). 'A historical perspective towards a non-invasive treatment for patients with atherosclerosis'. In: *Netherlands Heart Journal* 17.4, pp. 140–144.

Steenland, Kyle (1992). 'Passive smoking and the risk of heart disease'. In: *Jama* 267.1, pp. 94–99.

Steffen, Lyn M et al. (2003). 'Associations of whole-grain, refined-grain, and fruit and vegetable consumption with risks of all-cause mortality and incident coronary artery disease and ischemic stroke: the Atherosclerosis Risk in Communities (ARIC) Study'. In: *The American journal of clinical nutrition* 78.3, pp. 383–390.

Strogatz, Steven H (2018). *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering*. CRC Press.

Thompson, R.C. et al. (2013). 'Atherosclerosis across 4000 years of human history: The Horus study of four ancient populations'. In: *The Lancet* 381.9873, pp. 1211–1222.

Venna, J. and S. Kaski (2005). 'Local multidimensional scaling with controlled tradeoff between trustworthiness and continuity'. In: URL: http://eprints.pascal-network.org/archive/00001233/.

Vos, Theo et al. (2016). 'Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the Global Burden of Disease Study 2015'. In: *The Lancet* 388.10053, pp. 1459 –1544. ISSN: 0140-6736. DOI: http://dx.doi.org/10.1016/S0140-6736(16)31012-1. URL: http://www.sciencedirect.com/science/article/pii/S0140673616310121.

WHO (2017). *European Health for All Database (HFA-DB)*. URL: http://data.euro.who.int/hfadb/.

Weisstein, Eric W (2002). *CRC concise encyclopedia of mathematics*. CRC press.

Wells, A Judson (1994). 'Passive smoking as a cause of heart disease'. In: *Journal of the American College of Cardiology* 24.2, pp. 546–554.

Wilhelmsen, L. et al. (1973). 'Multivariate analysis of risk factors for coronary heart disease'. In: *Circulation* 48.5, pp. 950–958.

Wilhelmsen, Lars (1988). 'Coronary heart disease: Epidemiology of smoking and intervention studies of smoking'. In: *American Heart Journal* 115.1, Part 2, pp. 242 –249.

Wilkins, E et al. (2017). 'European cardiovascular disease statistics 2017'. In: *European Heart Network: Brussels, Belgium*.

Wilson, Peter W.F. and Christopher J. O'Donnell (2018). '1 - Epidemiology of Chronic Coronary Artery Disease'. In: *Chronic Coronary Artery Disease*. Ed. by James A. de Lemos and Torbjørn Omland. Elsevier, pp. 1 –15. ISBN: 978-0-323-42880-4. DOI: https://doi.org/10.1016/B978-0-323-42880-4.00001-7. URL: http://www.sciencedirect.com/science/article/pii/B9780323428804000017.