REPUBLIC OF TURKEY
YILDIZ TECHNICAL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

DATA DISCRETIZATION AND BAYESIAN NETWORK MODELING: A CASE STUDY IN R

GÜNAL BİLEK

Ph.D. THESIS
DEPARTMENT OF STATISTICS
PROGRAM OF STATISTICS

ADVISER
ASSOC. PROF. DR. FİLİZ KARAMAN

İSTANBUL, 2018

REPUBLIC OF TURKEY
YILDIZ TECHNICAL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

DATA DISCRETIZATION AND BAYESIAN NETWORK MODELING: A CASE STUDY IN R

A thesis submitted by Günal BİLEK in partial fulfillment of the requirements for the degree of **DOCTOR OF PHILOSOPHY** is approved by the committee on 04.06.2018 in the Department of Statistics, Statistics Program.

Thesis Adviser

Assoc. Prof. Dr. Filiz KARAMAN

Yildiz Technical University

Approved by the Examining Committee

Assoc. Prof. Dr. Filiz KARAMAN

Yildiz Technical University _____

Prof. Dr. Bülent GÜLOĞLU

Istanbul Technical University _____

Assoc. Prof. Dr. Gülder KEMALBAY

Yildiz Technical University _____

Prof. Dr. Ali Hakan BÜYÜKLÜ

Yildiz Technical University _____

Prof. Dr. Ahmet Mete ÇİLİNGİRTÜRK

Marmara University _____

ACKNOWLEDGMENTS

First of all, I would like to thank my supervisor Assoc. Prof. Dr. Filiz KARAMAN for her excellent guidance and support throughout this thesis.

I am very grateful for the opportunity to have studied at Yildiz Technical University and would like to express sincere gratitudes and appreciation to the lecturers for teaching me the necessary skills to prepare this thesis and to the research assistants for their hospitality and friendship in the Department of Statistics.

Also, I would not like to miss the opportunity to thank every teacher and lecturer involved in my educational life both in Turkey and the UK for their support, encouragement and guidance.

Furthermore, I would like to thank the reading committee for reviewing this thesis.

In addition, I would like to thank the Ministry of Education of Turkey for funding my previous English and Master studies in the UK, which, I believe, has enhanced this thesis.

Finally, a massive thank-you goes to my family and friends, without whose unconditional love and support, this thesis would not have been possible.

June, 2018

Günal BİLEK

TABLE OF CONTENTS

Page

LIST OF SYMBOLS vii

LIST OF ABBREVIATIONS viii

LIST OF FIGURES x

LIST OF TABLES xi

ABSTRACT xii

ÖZET xiv

CHAPTER 1

INTRODUCTION 1

1.1 Motivation 1
1.2 Objectives of the Thesis 2
1.3 Structure of the Thesis 3

CHAPTER 2

DATA 5

2.1 Introduction 5
2.2 Participants 5
2.3 Variables in the Data 5
2.3.1 Psychiatric Variables 6
2.3.2 Demographic Variables 7
2.3.3 Socio-economic Variables 7
2.4 Conclusion 9

CHAPTER 3

EXPLORATORY DATA ANALYSIS 10

3.1 Introduction 10

3.2 Exploratory Data Analysis of Psychiatric Data 10
3.3 Exploratory Data Analysis of Demographic and Socio-Economic Data Sets 15
3.4 Conclusion 19

CHAPTER 4

DATA DISCRETIZATION 20

4.1 Introduction 20
4.2 What is Data Discretization? 20
4.3 Theory Underlying Data Discretization 22
4.4 Information-Preserving Discretization 23
4.4.1 Discretization of Psychiatric Variables Using the IPD 24
4.5 Discretization by Domain Knowledge 25
4.5.1 Discretization of Psychiatric Variables Using Domain Knowledge 25
4.6 Conclusion 26

CHAPTER 5

BAYESIAN NETWORKS 27

5.1 Introduction 27
5.2 Bayes' Theorem 27
5.3 Bayesian Networks (BNs) 28
5.3.1 A Classic BN example 29
5.3.2 Representation of the Joint Probability Distribution 30
5.3.3 Factorization of the Joint Probability Distribution 32
5.3.4 Conditional Independence 33
5.3.5 D-Separation 35
5.3.6 Connection Types in BNs 35
5.4 Learning Bayesian Networks 37
5.4.1 Structural learning 40
5.4.2 Parameter Learning 42
5.5 Conclusion 43

CHAPTER 6

RESULTS AND DISCUSSION 44

6.1 Introduction 44
6.2 Results 44
6.2.1 Results of $\mathcal{B}_1$ 45
6.2.2 Results of $\mathcal{B}_2$ 65
6.3 Comparison of $\mathcal{B}_1$ and $\mathcal{B}_2$ 69
6.4 Discussion and Conclusion 70

CHAPTER 7

CONCLUSION 72

APPENDIX-A

ROSENBERG SELF-ESTEEM SCALE 82

APPENDIX-B

BECK DEPRESSION INVENTORY 83

APPENDIX-C

BECK HOPELESSNESS SCALE 87

APPENDIX-D

R CODE FOR SECTION 6.2.1 89

APPENDIX-E

R CODE FOR SECTION 6.2.2 101

CURRICULUM VITAE 105

LIST OF SYMBOLS

$\mathcal{B}_1$	Bayesian network 1
$\mathcal{B}_2$	Bayesian network 2
$\mathcal{G}$	Directed acyclic graph
Ω	Discretization policy vector
$\mathcal{D}$	Data
D	Discretization degree
$\mathcal{N}$	Set of nodes
$\mathcal{A}$	Set of arcs
Θ	Set of probabilities
$\mathbf{X}$	Set of variables

LIST OF ABBREVIATIONS

BDI	Beck Depression Inventory
BHS	Beck Hopelessness Scale
BN	Bayesian Network
CI	Conditional Independence
DAG	Directed Acyclic Graph
DLC	Discretization Level Caolescence
IPD	Information-Preserving Discretization
RSES	Rosenberg Self-Esteem Scale
TMI	Total Mutual Information

LIST OF FIGURES

Page

Figure 3.1 Histogram of individual BDI scores 11
Figure 3.2 Boxplot of individual BDI scores 12
Figure 3.3 Histogram of individual BHS scores 13
Figure 3.4 Boxplot of individual BHS scores 13
Figure 3.5 Histogram of individual RSES scores 14
Figure 3.6 Boxplot of individual RSES scores 15
Figure 5.1 Classic Wet Grass BN Example 29
Figure 5.2 Serial connection 36
Figure 5.3 Diverging connection 36
Figure 5.4 Converging connection 37
Figure 6.1 DAG of $\mathcal{B}_1$ 46
Figure 6.2 Bar chart of conditional probabilities of *Mother_occupation* 47
Figure 6.3 Bar chart of conditional probabilities of *Student_income* given *Age* 48
Figure 6.4 Bar chart of conditional probabilities of *Student_job* given *Student_income* 49
Figure 6.5 Bar chart of conditional probabilities of *Family_income* given *Student_income* 50
Figure 6.6 Bar chart of conditional probabilities of *Type_of_settlement* given *Family_income* 51
Figure 6.7 Bar chart of conditional probabilities of *Father_occupation* given *Family_income* 52
Figure 6.8 Bar chart of conditional probabilities of $Father_educational_status$ given $Father_occupation$ 53
Figure 6.9 Bar chart of conditional probabilities of *Mother_educational_status* given *Father_educational_status* 54
Figure 6.10 Bar chart of conditional probabilities of $Number_of_siblings$ given *Mother_educational_status* 55
Figure 6.11 Bar chart of conditional probabilities of *Family_structure* given *Number_of_siblings* 56
Figure 6.12 Bar chart of conditional probabilities of *Smoking_status* given *Student_job* and *Gender* 57
Figure 6.13 Bar chart of conditional probabilities of *Alcohol_use* given *Smoking_status* 58
Figure 6.14 Bar chart of conditional probabilities of *Relationship_status* given *Alcohol_use* and *Student_job* 59
Figure 6.15 Bar chart of conditional probabilities of *Social_activity* given *Gender* 60

Figure 6.16 Bar chart of conditional probabilities of *Type_of_accommodation* given *Gender* 61
Figure 6.17 Bar chart of conditional probabilities of *School_type* given *Gender* 62
Figure 6.18 Bar chart of conditional probabilities of *Depression* given *Gender* 63
Figure 6.19 Bar chart of conditional probabilities of *Self_esteem* given *Depression* 64
Figure 6.20 Bar chart of conditional probabilities of *Hopelessness* given *Depression* 65
Figure 6.21 DAG of $\mathcal{B}_2$..... 66
Figure 6.22 Bar chart of conditional probabilities of *Depression* given *Social_activity*..... 67
Figure 6.23 Bar chart of the conditional probabilities of *Hopelessness* given *Depression*..... 68
Figure 6.24 Bar chart of the conditional probabilities of *Self_esteem* given *Depression*..... 69

LIST OF TABLES

Page

Table 2.1 Demographic and socio-economic data description 8

Table 3.1 Frequencies and percentages of demographic and socio-economic data sets 15

Table 6.1 Goodness-of-fit statistics of $\mathcal{B}_1$ and $\mathcal{B}_2$ 69

ABSTRACT

DATA DISCRETIZATION AND BAYESIAN NETWORK MODELING: A CASE STUDY IN R

Günal BİLEK

Department of Statistics
Ph.D. Thesis

Adviser: Assoc. Prof. Dr. Filiz KARAMAN

Bayesian networks (BNs) are a graphical representation of a probability distribution over a set of variables. It comprises a directed acyclic graph (DAG) and a set of probability distributions.

The first aim of this thesis is to review and introduce the BNs and data discretization. The second one is to illustrate how discrete BNs are constructed from the data which include continuous variables thorough discretization. The third one is to compare the effect of the two commonly-used data discretization approaches on the BNs. The last one, is to study the relationships among the Beck Depression Inventory, Beck Hopelessness Scale, Rosenberg Self-Esteem Scale scores and demographic and socio-economic variables with BN modeling.

The data of 823 university students consist of 21 continuous and discrete relevant psychiatric, demographic and socio-economic variables

The continuous variables are discretized by using the Information-Preserving Discretization (IPD) and domain knowledge available in the literature and, consequently, two discrete BNs are learned from the data sets and constructed in statistical software `R` and the results are presented via figures and probabilities.

One of the most significant results is that the structure of the two BNs do not significantly differ. The only difference is that in the first Bayesian network model, the gender of the students influences the level of depression, with female students being more likely to be more depressive, while in the second model, social activity

directly influences the level of depression and the presence of social activity decreases the risk of being more depressive. Another important result is that in each model, depression influences both the level of hopelessness and self-esteem in students; additionally, the level of depression has a positive impact on the level of hopelessness, but a negative impact on the level of self-esteem. The last remarkable result is that, based on the BIC values and specificities of the two models, the BN whose continuous data were discretized by the IPD outperforms the BN whose data were discretized by the domain knowledge.

Key Words: Bayesian networks, data discretization, Beck Depression Inventory, Beck Hopelessness Scale, Rosenberg Self-Esteem Scale

ÖZET

VERİYİ KESİKLİ HALE GETİRME VE BAYES AĞ MODELLEMESİ: R'DE BİR VAKA ÇALIŞMASI

Günal BİLEK

İstatistik Bölümü
Doktora Tezi

Tez Danışmanı: Doç. Dr. Filiz KARAMAN

Bayes ağları, bir değişkenler dizisi üzerinde bir olasılık dağılımının grafiksel bir gösterimidir ve yönlendirilmiş asiklik grafiği (DAG) ve bir dizi olasılık dağılımından oluşur.

Bu tezin amaçlarından birisi, Bayes ağları ve veri ayrıştırma teorilerini incelemek ve bu konular hakkında bilgi vermektir. İkinci amacı ise, sürekli değişkenler içeren veriyi, ayrıştırma yöntemi kullanarak kesikli hala getirmek ve kesikli Bayes ağları oluşturmaktır. Bir diğer amacı ise, veri ayrıştırmasında kullanılan iki yöntemin etkisinin Bayes ağlarının performansları üzerindeki etkilerini karşılaştırmaktır. Son amacı ise, Bayes ağ modellemesini kullanarak Beck Depresyon Ölçeği, Beck Umutsuzluk Ölçeği ve Rosenberg Benlik Saygısı Ölçeği puanları ile demografik ve sosyo-ekonomik değişkenler arasındaki ilişkinin boyutunun incelenmesidir.

Bu çalışmadaki veri seti 823 üniversite öğrencisine ait 21 sürekli ve kesikli psikiyatrik, demografik ve sosyo-ekonomik değişkenden oluşmaktadır.

Sürekli değişkenler bilgi koruyan ayrıştırma yöntemi (IPD) ve literatürde ki bilgiyi kullanarak sınıflandırma olmak üzere iki farklı yöntemle kesikli hale getirilir. Bunun ardından, 2 farklı kesikli Bayes ağ modeli istatistiksel program `R`'de oluşturulur ve sonuçlar sayı ve grafiklerle verilir.

Bu çalışmadaki önemli sonuçlardan bir tanesi, birinci modelde cinsiyetin öğrencilerin depresyon seviyeleri üzerinde önemli bir etkisinin olduğu ve kadın öğrencilerin daha yüksek depresyon seviyelerine sahip olma olasılığının daha yüksek olduğudur. İkinci modelde ise sosyal aktivite durumunun öğrencilerin depresyon seviyeleri üzerinde önemli bir etkisinin olduğudur. Ayrıca, her iki modelde de, depresyonun umutsuzluk seviyesi üzerinde pozitif, benlik saygısı üzerinde ise negatif etkisinin olduğu

tespit edilmiştir. Buna ek olarak, iki Bayes ağ modeli performans açısından BIC ve özgüllük değerleri kullanılarak karşılaştırılmış ve verisi bilgi koruma yöntemi kullanılarak kesikli hale getirilen modelin performansının daha iyi olduğu sonucuna varılmıştır.

Anahtar Kelimeler: Bayes ağları, veri ayrıştırma, Beck Depresyon Envanteri, Beck Umutsuzluk Ölçeği, Rosenberg Benlik-Saygısı Ölçeği

CHAPTER 1

INTRODUCTION

1.1 Motivation

Bayesian networks (BNs) are probabilistic graphical models for representing and reasoning about problems which include uncertainty. They consist of two parts. The first part is a graph known as the *directed acyclic graph* (DAG) consisting of a set of nodes connected by directed arcs. The nodes in the DAG represent variables, while the arcs indicate the dependencies between these variables. As a whole, the DAG shows the quantitative relationship among variables [1]. The second part consists of a set of probabilities which define the strength of the relationships in the DAG. A BN provides the opportunity to make probabilistic inference given the information available [2].

Decision-makers in all domains of life often face situations in which they have to make decisions under uncertainty. Sometimes they make decisions using models based on data sets. However, in some cases the data may not be available or, even so, the data may not be adopted to new circumstances, leading to unsuccessful predictions. In such cases, using models based on expert knowledge or models which rely on both data and expert judgement can be more successful. The BNs provide the opportunity to address these issues using a simple theorem (see Section 5.2) developed by Reverend Thomas Bayes in 1763. The theorem offers a way to update belief in an uncertain event when there is new evidence [3].

Although Bayes' theorem has been available since 1763, the BNs have been used for solving real-life problems since the 1980s as a result of the fast development in the computer technology. This is because the calculations in BNs are computationally

expensive. After the publications of algorithms [4],[5] enabling efficient belief updating for a large class of BN models, the BNs have gained popularity and been widely used in many research areas including clinical decision support [6],[7], engineering [8], ecology [9], [10], [11], [12], customer behaviors [13], [14], [15], system biology [16], [17] and so on. There are many reasons to use the BNs, some of which are as follows:

1. They are graphical models, making them capable of showing the relationships between variables distinctly and intuitively.
2. They are easy to be understood and interpreted by not only statisticians but also by non-statisticians.
3. They are directional, helping to indicate cause-effect relationship between variables.
4. They show both direct and indirect causation.
5. They can handle uncertainty.
6. They can be used for discrete, continuous or mixed data.
7. They can be constructed by using the data, expert knowledge or a combination of both.

1.2 Objectives of the Thesis

In Psychiatry, scales are used to measure the psychiatric characteristics of people. In this study, the BNs are used to investigate the relationships among the Beck Hopelessness Scale (BHS), Beck Depression Inventory (BDI) and Rosenberg Self-Esteem Scale (RSES) scores and to determine the demographic and socio-economic variables influencing these psychiatric variables.
The objectives of this thesis is as follows:

1. The first aim of this thesis is to review the BN theory. This includes explaining what a BN is, where it comes from, the reasons why it is used, how it is constructed and how it is used to make queries.

2. The second one is to review the theory of data discretization. This covers why and how a variable is discretized.

3. The third one is to illustrate what has been mentioned above (in 1 and 2) thorough a data set in `R` and interpret the results.

4. The fourth one is to determine the performance of the BNs for investigating the factors influencing university students' Beck Depression Inventory , Beck Hopelessness Scale and Rosenberg Self-Esteem Scale scores.

5. The last one is to determine which of the two commonly-used discretization methods leads to the BN that makes better predictions.

1.3 Structure of the Thesis

The rest of this thesis is structured as follows:

- In Chapter 2, the data used in this thesis are described. How the data were collected and the properties of the variables are detailed.

- In Chapter 3, exploratory data analyses are conducted in order to see how the data behave. Suitable plots are used for continuous data, whereas frequencies and percentages are used for discrete data.

- In Chapter 4, firstly the need to convert continuous variables into discrete ones is explained. Next, the theory and the commonly-used discretization techniques are given. Finally, the continuous variables are discretized by using two approaches.

- In Chapter 5, the theory of the BNs and the basic terms required to understand the BNs are detailed. In addition, learning the BNs from the data is described.

- In Chapter 6, two BNs are constructed as a result of using two different discretization approaches. The results of both models are provided via plots and figures. Then, the performances of the two models are compared. Finally, the results are compared with those of the similar studies.

- In Chapter 7, the thesis is concluded by summarizing what has been done and the significant results are compared to those in similar studies in literature.

CHAPTER 2

DATA

2.1 Introduction

This chapter is concerned with providing information about the data used in this study.

In Section 2.2, the participants of the study and the way the data set were collected are described. In Section 2.3, the types of variable in the data sets are detailed. Finally, in Section 2.4, the chapter is concluded by summarizing what has been presented.

2.2 Participants

In this study, the subjects are the students of Bitlis Eren University, located in Turkey, in the 2016-2017 academic year. The students were given a 4-page form consisting of the Turkish versions of the Beck Hopelessness Scale, Beck Depression Inventory and Rosenberg Self-Esteem Scale and also the questions about their demographic and socio-economic backgrounds . The students were asked to fill the forms on a voluntary basis. In total, 890 students from 5 schools of Bitlis Eren University were included in the study and data of 823 students' were obtained after removing the missing data.

2.3 Variables in the Data

In this study, the data are composed of three types of variables: (i) psychiatric variables, (ii) demographic variables and (iii) socio-economic variables.

2.3.1 Psychiatric Variables

The data include three psychiatric variables. In this thesis, they are called psychiatric variables because they represent the scores of three scales commonly used in Psychiatry.

Rosenberg Self-Esteem Scale (RSES) Score

Rosenberg Self-Esteem Scale (RSES), created by [20], is a self-report inventory that shows how an individual judges his or her self-worth. The RSES consists of 10 items and is rated on a 4-point Likert scale with 0 (strongly disagree) and 3 (strongly agree) for the 1st, 3th, 4th, 7th and 10th items and 0 (strongly agree) and 3 (strongly disagree) for the 2nd, 5th, 6th, 8th and 9th items. The RSES scores are calculated by summing the answers of the 10 items and they lay in the interval of 0 to 30. The RSES is given in Appendix-A.

Beck Depression Inventory (BDI) Score

Beck Depression Inventory (BDI), created by [18], is a self-report inventory used to measure the level of depression in people who are at least 13 years of age. The BDI comprises 21 multiple-choice questions with 0 being the lowest score and 3 being the highest score in each question. The total BDI score is calculated by summing the answers of the 21 questions. Possible BDI scores range from 0 to 63. The English version of the BDI is given in Appendix-B.

Beck Hopelessness Scale (BHS) Score

Beck Hopelessness Scale (BHS), developed by [19], is a self-report inventory indicating the severity of hopelessness in individuals who are aged 17-80. It consists of 20 true-false questions. 11 items are keyed true, 9 false. Then, the answers are summed up and the total BHS score is calculated. BHS scores range from 0 to 20. The English version of the BHS is given in Appendix-C.

2.3.2 Demographic Variables

Another type of variable in this study is the demographic ones. These are the variables which define the characteristics of individuals, or simply who they are, and cannot be influenced by any other variable. In the data, there are only two demographic variables which are:

1. *Age*
2. *Gender*

2.3.3 Socio-economic Variables

The last type of variable in this study is socio-economic ones. These variables represent the socio-economic status of the subjects in the society as well as their choices. The data includes 16 socio-economic variables which are as follows:

1. *Family_structure*
2. *Father_educational_status*
3. *Mother_educational_status*
4. *Family_income*
5. *Student_income*
6. *Number_of_siblings*
7. *Type_of_settlement*
8. *Father_occupation*
9. *Mother_occupation*
10. *School_type*
11. *Type_of_accommodation*
12. *Smoking_status*
13. *Alcohol_use*

14. *Social_activity*

15. *Student_job*

16. *Relationship_status*

Table 2.1 Demographic and socio-economic data description

Variable	Description	State description
Age	Age of student	20-, 20-23, 23+
Gender	Gender of student	Male, Female
Family_structure	Family structure of student he/she grew in	Nuclear, Extended, Single parent, Mother or father died
Father_educational_status	Education level of student's father	Illiterate, Primary school, Secondary school, High school, University, Master or Ph.D.
Mother_educational_status	Education level of student's mother	Illiterate, Primary school, Secondary school, High school, University, Master or Ph.D.
Family_income	Student's family's monthly income in TL*	0-1400TL, 1401-3000TL, 3001-5000TL, 5000+TL
Student_income	Student's monthly income in TL*	0-500TL, 501-750TL, 751-1000TL, 1000+TL
Number_of_siblings	Number of siblings student has	1,2,3,4,5+
Type_of_settlement	Type of settlement student were born in	Village or town, District, City, Metropolis
Father_occupation	Occupation of student's father	Unemployed, Self-employed, Public sector employee, Private sector employee
Mother_occupation	Occupation of student's mother	Unemployed, Self-employed, Public sector employee, Private sector employee
School_type	Type of school student studying at	Faculty of arts and sciences, Faculty of engineering, Faculty of economics and administrative sciences, Health school, Vocational school
Type_of_accommodation	where student is living	Government dormitory, Private dormitory, Flat
Smoking_status	Whether or not student is a smoker	Smoker, Not smoker
Alcohol_use	Whether or not student consumes alcohol	Alcohol user, Not alcohol user
Social_activity	Whether or not student has any social activity	Has social activity, No social activity
Student_job	Whether or not student has a part-time or full time job	Has job, No job
Relationship_status	Student's relationship status	Single, In a relationship

TL: Turkish Lira (currency)

Table 2.1 represents the demographic and socio-economic variables, what they show and their state descriptions.

2.4 Conclusion

This chapter defined the data used in this thesis. It started with describing the participant of the study. Next, the psychiatric variables were detailed. Finally, the demographic and socio economic variables were provided and their definitions and state descriptions were demonstrated via a table.

The next chapter conducts exploratory data analysis for the three data sets.

CHAPTER 3

EXPLORATORY DATA ANALYSIS

3.1 Introduction

The aim of this chapter is to conduct exploratory data analysis to see some of the features and properties of the data sets.

As detailed in the previous chapter, there are 2 types of data sets in this study: (i) continuous data which consist of psychiatric variables and (ii) categorical data which comprise demographic and socio-economic variables. Therefore, exploratory data analysis of each data set will be conducted separately.

In Section 3.2, exploratory data analysis for the psychiatric variables is carried out using histograms, box plots and descriptive statistics and significant features of the psychiatric data set are interpreted. In Section 3.3, exploratory data analysis is conducted for both demographic and socio-economic variables using frequencies and percentages and important figures are highlighted. Finally, in Section 3.4, the work in this chapter is summarized.

3.2 Exploratory Data Analysis of Psychiatric Data

The first data set whose properties to be to explored is the psychiatric data set. Since the psychiatric variables (BDI, BHS and RSES scores) are continuous, histograms and box plots can be useful to see the properties of the distributions.

Figure 3.1 and Figure 3.2 are the histogram and the box plot of the individual BDI scores of the students, respectively. From Figure 3.1, it is noted that the BDI scores are heavily skewed to the right and not normally distributed. The vast majority of

the BDI scores concentrate between the interval [0, 20]. From Figure 3.2, it is seen that there are many outliers in the BDI scores (represented by •). In addition, the mean, standard deviation, median, first quartile and third quartile of the BDI scores are 14.99, 10.49, 13, 8 and 20, respectively.

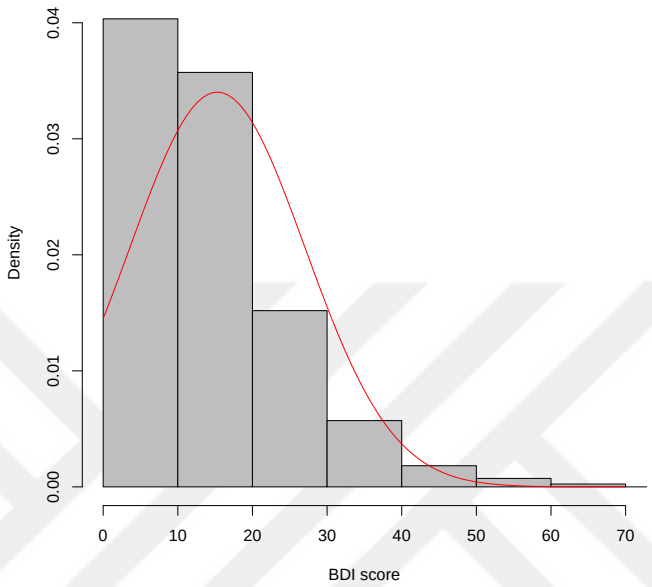


Figure 3.1 Histogram of individual BDI scores

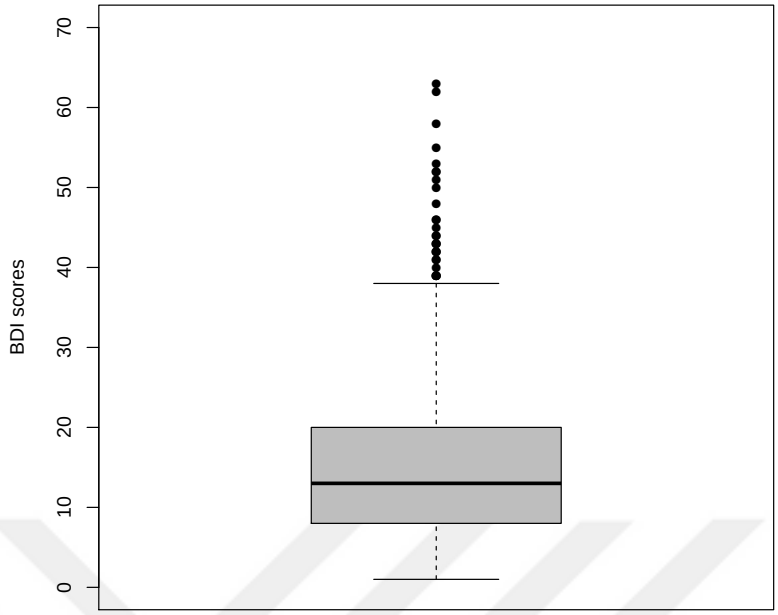


Figure 3.2 Boxplot of individual BDI scores

Figure 3.3 and Figure 3.4 represent the histogram and the box plot of the individual BHS scores of the participants, respectively. From Figure 3.3, it is seen that the BHS scores are heavily skewed to the right, making the distribution non-normal. It is notable that the concentration of the BHS scores is between 0 and 6. As Figure 3.4 shows, in contrast with the BDI scores, the BHS scores do not have too many outliers (only 1 represented by •). The mean, standard deviation, median, first quartile and third quartile of the BHS are 5.85, 4.69, 4, 2 and 9, respectively.

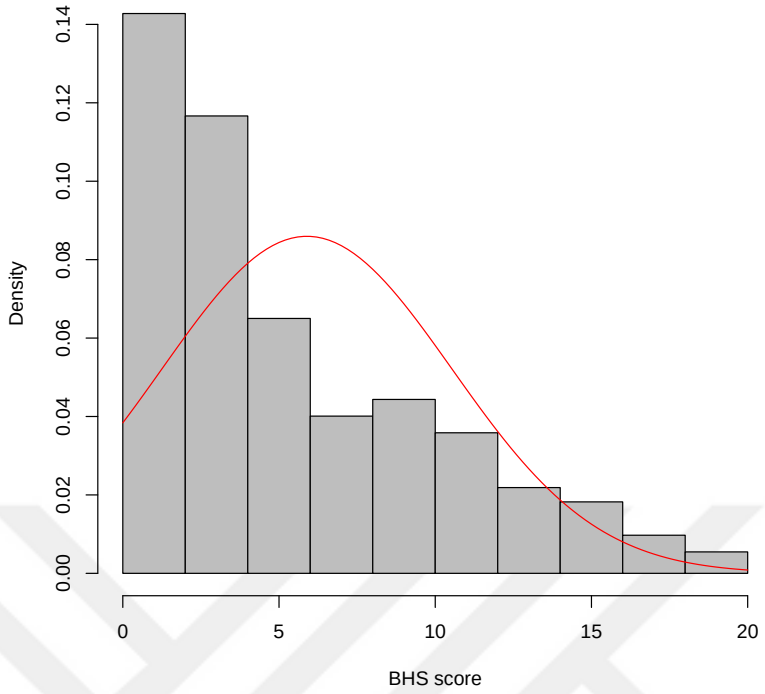


Figure 3.3 Histogram of individual BHS scores

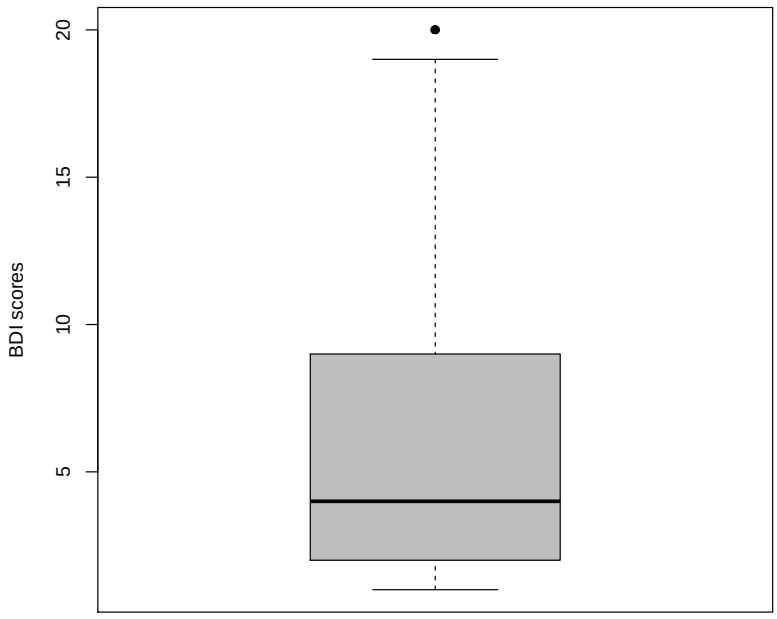


Figure 3.4 Boxplot of individual BHS scores

Figure 3.5 and Figure 3.6 indicate the histogram and the box plot of the individual RSES scores of the subjects. According to Figure 3.5, the RSES scores are skewed to the left and its distribution is not normal. Most of the RSES scores lie in the interval [18, 26]. Figure 3.6 shows that the RSES scores also have outliers which are fewer than that of the BDI scores, but more than that of the BHS scores. It is worth mentioning that the outliers of the RSES scores are below its first quartile, whereas the outliers of the BDI scores and the BHS scores are above their third quartiles. The mean, standard deviation, median, first quartile and third quartile of the RSES scores are 21.64, 4.95, 22, 19 and 25, respectively.

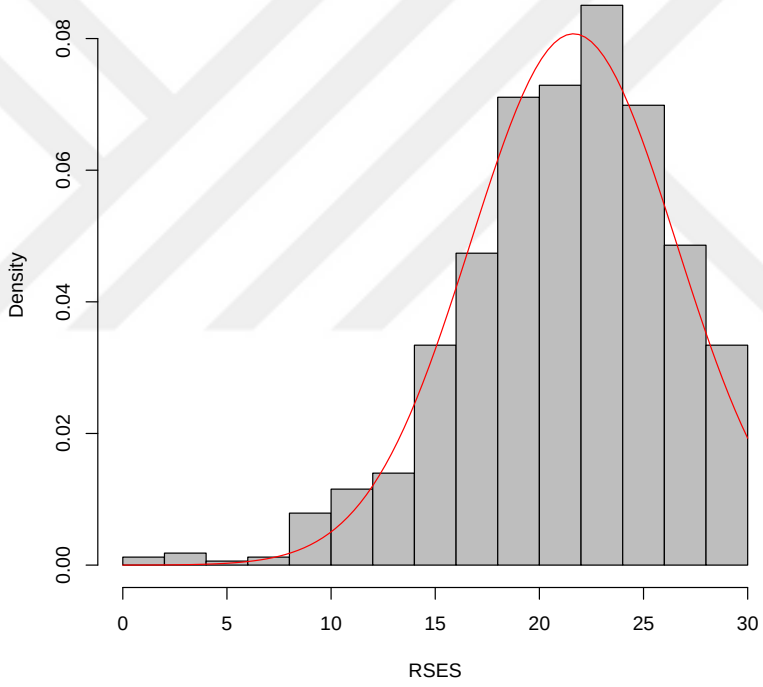


Figure 3.5 Histogram of individual RSES scores

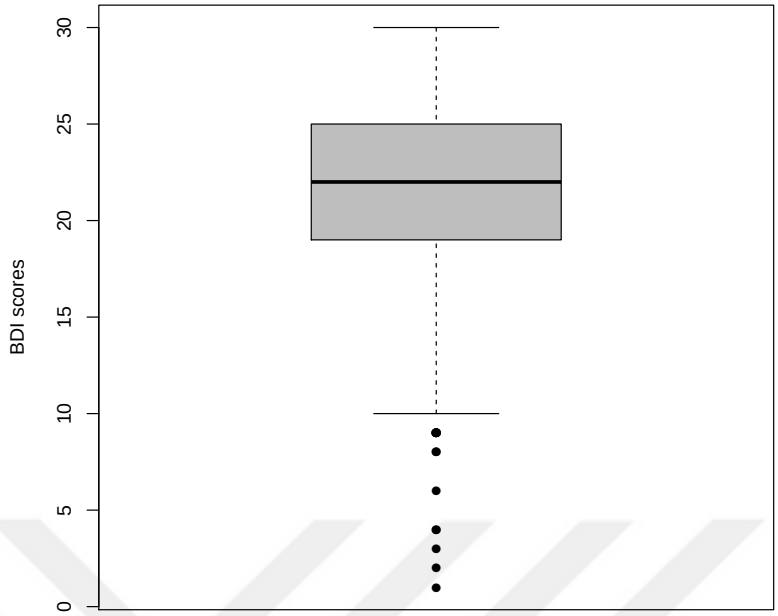


Figure 3.6 Boxplot of individual RSES scores

3.3 Exploratory Data Analysis of Demographic and Socio-Economic Data Sets

Since the demographic and socio-economic variables are categorical, a table which represents the frequencies and the percentages of the these variables can be useful to see the important properties of the demographic and socio-economic data sets. Table 3.1 indicates the frequencies and percentages of the demographic and conditional variables. The column **Variable** represents the variables of the demographic and socio-economic data sets: the first 2 variables are the demographic ones, while the last 16 variables are the socio-economic ones. The column **Level** demonstrates the levels of the corresponding variable. Finally, the column **n** indicates the frequency of the levels and the column % the percentages of the levels.

Table 3.1 Frequencies and percentages of demographic and socio-economic data sets

Variable	**Level**	**n**	%

Table 3.1(cont'd)

Age	20-	76	9.2
	20-23	567	68.9
	23+	180	21.9
Gender	Male	394	47.9
	Female	429	52.1
Family_structure	Nuclear	410	49.8
	Extended	389	47.3
	Single parent	4	0.5
	Mother or father died	20	2.4
Father_educational_status	Illiterate	87	10.6
	Primary school	361	43.9

Table 3.1 (cont'd)

	Secondary school	135	16.4
	High school	149	18.1
	University	81	9.8
	Master or Ph.D.	10	1.2
Mother_educational_status	Illiterate	405	49.2
	Primary school	289	35.1
	Secondary school	62	7.5
	High school	48	5.8
	University	15	1.8
	Master or Ph.D.	4	0.5
Family_income	0-1400 TL	384	46.7
	1401-3000 TL	277	33.7
	3001-5000 TL	111	13.5
	5000+ TL	51	6.2
Student_income	0-500 TL	667	81
	501-750 TL	84	10.2
	751-1000 TL	28	3.4
	1000+ TL	44	5.3
Number_of_siblings	1	6	0.7

Table 3.1(cont'd)

	2	29	3.5
	3	100	12.2
	4	120	14.6
	5+	568	69
Type_of_settlement	Village or town	183	22.2
	District	270	32.8
	City	251	30.5
	Metropolis	119	14.5
Father_occupation	Unemployed	230	27.9
	Self-employed	238	28.9
	Public sector employee	191	23.2
	Private sector employee	164	19.9
Mother_occupation	Unemployed	762	92.6
	Self-employed	33	4
	Public sector employee	19	2.3
	Private sector employee	9	1.1
School_type	Faculty of arts and sciences	204	24.8
	Faculty of engineering	173	21
	Faculty of economics and administrative sciences	113	13.7
	Health school	201	24.4
	Vocational school	132	16
Type_of_accommodation	Goverment dormitory	408	49.6
	Private dormitory	29	3.5
	Flat	386	46.9
Smoking_status	Smoker	216	26.2
	Not smoker	607	73.8
Alcohol_use	Alcohol user	77	9.4
	Not alcohol user	746	90.6
Social_activity	Has social activity	278	33.8
	No social activity	545	66.2

Table 3.1(cont'd)

Student_job	Has job	84	10.2
	No job	739	89.8
Relationship_status	Single	570	69.3
	In a relationship	253	30.7

TL: Turkish Lira (currency)

To summarize the properties of the demographic and socio-economic data sets, the vast majority of the students (68.9%) are aged 20-23, while the percentages of the gender are close to each other (male 47.9%, female 52.1%). The percentages of the student whose family structure are single parent and mother or father died are remarkably low (0.5% and 2.4%, respectively). The highest proportion (43.9%) of the fathers are primary school graduates, whereas the highest proportion (49.2%) of the mothers are illiterate. The proportion of the students whose family income is above 5000 TL is only 6.2%. The vast majority of the students (81%) receive 0-500 TL income a month, 3.4% and 5.3% of them have an income of 751-1000 TL and more than 1000 TL, respectively. While the percentages of the students who have 1 and 2 siblings are respectively 0.7% and 3.5%, a high percentage of them have 5 or more siblings (69%). The highest percentage of students grew up in districts (32.8%), whereas the least percentage of them grew up in metropoles (14.5%). While the percentages of the occupational status of the fathers are not significantly different from each other, the greatest proportion (92.6%) of the mothers are unemployed. While the highest percentage (24.8) of the students are from the faculty of arts and science, the least of them from the faculty of economics and administrative sciences (13.7%). The percentages of the students living in government dormitories and flats are respectively 49.6% and 46.9%, whereas those living in privates dormitories comprise only 3.5% of the total population. The highest proportion (73.8%) of the students are not smokers. Also, alcohol is not popular among the students: while 9.4% of them consume alcohol, 90.6% of them do not. The percentage of the students who participate in social activities is the approximately half of those who do not participate in social activities. The vast majority (89.8%) of the students do not prefer to work. Finally, while the proportion of the students who are single is 69.3%, 30.7% of the students have relationships.

3.4 Conclusion

In conclusion, in this chapter, exploratory data analysis was undertaken for the continuous and categorical data sets separately. Histograms, box plots and descriptive statistics were used for the continuous data, whereas a table comprising frequencies and percentages was preferred for the categorical data.

One of the most important properties of the psychiatric data set is that none of the psychiatric variables are normally distributed. While the BDI and the BHS scores are skewed to the right, the RSES scores are skewed to the left. Another one is that all of the psychiatric variables have outliers. However, the outliers of the RSES scores are below its first quartile, whereas the outliers of the BDI and the BHS scores are above their third quartiles.

One of the most important characteristics of the demographic and conditional variables is that there are huge differences between some percentages. To illustrate some of them, the vast majority of the mothers are unemployed, the percentage of students who consume alcohol is remarkably low, and the percentages of the students who have 1 and 2 siblings are notably low.

The next chapter provides information about data discretization.

CHAPTER 4

DATA DISCRETIZATION

4.1 Introduction

This chapter is concerned with introducing and reviewing the theory of data discretization.

It starts with the definition and motivation of data discretization in Section 4.2. Next, the theory of discretization is provided in Section 4.3. Afterwards, one of the discretization method to be used in this study is detailed and the continuous variables are discretized in Section 4.4 and Section 4.4.1, respectively. After that, another discretization approach to be used in this study is detailed in Section 4.5 and the continuous variables are discretized by this approach in Section 4.5.1. Finally, the chapter is concluded in Section 4.6.

4.2 What is Data Discretization?

Generally speaking, Bayesian approaches use multinomial distributions, which require all the variables to be discrete. Nevertheless, in many real world applications, the variables can be continuous.

It is possible to use continuous variables in a Bayesian model. In this case, these continuous variables are assumed to follow a known parametric distribution such as Gaussian distribution. However, this requires prior knowledge in the domain which may not be available as in our case.

In today's literature, there is no certain way to handle these continuous variables when learning a Bayesian network model. One approach to do so, is to discretize

these numeric attributes. In the process of discretization, there is no way to avoid the loss of information. However, it can be the best option in many domains including statistics, machine learning, data mining and artificial intelligence when the amount of information lost is kept minimum.

There are many reasons to apply discretization on the continuous data. The first is to reduce and simplify the data, which speeds up the analysis. The second is to obtain interpretable outputs. The third one, arguably the most important one, is that the statistical approach to be used may operate on nominal or categorical data only [21].

Discretization methods are classified according to how the discretization is conducted, such as whether it uses class information or which direction it proceeds, i.e. *top-down* versus *bottom-up*. If class information is used in the discretization process, the discretization is called *supervised*. If not, it is called *unsupervised discretization*.

If the process of discretization begins by finding one or a few points (known as *cut points*) to divide the entire attribute range and repeats this recursively on the resulting intervals, then the discretization is called *top-down* or *splitting* [22]. On the other hand, if the discretization begins with the complete list of the entire continuous values of the features as cut points and recursively remove some by uniting neighborhood values to form intervals, it is called *bottom-up* [23].

Discretization methods are also grouped in terms of *global* versus *local*. While a global discretization approach discretizes in the entire instance space, the local technique discretizes in a localized region of the instance space [24].

In addition, another category of discretization techniques is *direct* versus *incremental*. Incremental approaches start with simple discretization and continue with an improvement or refinement process, needing a stopping criterion to halt interrupt further discretization. Direct techniques split the range of k intervals simultaneously (i.e. equal-width, equal frequency), which requires additional input by the user to define the number of intervals [23], [25].

In literature, there are several discretization methods. Some of them are Error-based Discretization [26], One-Rule Discretizer (1RD) [27], Equal Frequency Discretization [28], Information-Preserving Discretization [29] and so on.

4.3 Theory Underlying Data Discretization

Let $\mathbf{x}$ be a vector of continuous values of length L. Its discretization should satisfy the conditions below:

1. Every element of $\mathbf{d}$ of length L is in the set $\{0, \ldots, D-1\}$ for some positive integer D, the degree of discretization.
2. For all $1 < i\ , j < L$; $d_i \leq d_j \iff x_i \leq x_j$.

When $\mathbf{x}$ is sorted, it is assumed for all $i < j$; $x_i < x_j$, which leads to, using the definition of discretization, for all $i < j$; $d_i < d_j$. Next, a *spanning* discretization of degree D satisfying the additional property that the first element of $\mathbf{x}$ is 0 and that the last element of $\mathbf{x}$ is $D-1$ is defined.

Additionally, a *discretization policy* of degree D is defined, that is, a real-valued vector $\mathbf{\Omega}$ of length $D+1$ which satisfies the properties below:

1. For all $i < j$; $\mathbf{\Omega}_i < \mathbf{\Omega}_j$
2. $\mathbf{\Omega}_0 = -\infty$
3. $\mathbf{\Omega}_D = \infty$

The discretization policy vector $\mathbf{\Omega}$ determines the boundaries of a set of D intervals which define the following mapping from real-valued vectors $\mathbf{x}$ to integer-valued vectors $\mathbf{d}$:

$$\Omega_i < x_i \leq \Omega_{j+1} \iff d_i = j \quad \forall i \in \{0, \ldots, L-1\}, j \in \{0, \ldots, D-1\} \tag{4.1}$$

If such a policy is used to discretize $\mathbf{x}$, then the resultant discretization $\mathbf{d}$ can be written as equal to $\mathbf{x}^{\Omega}$ [29].

4.4 Information-Preserving Discretization

Information-preserving discretization (IPD), developed by [29], is an entropy-based discretization technique . It aims to maximize the amount of information regarding the relationships between the variables. Therefore, discretization levels for every variable is combined in terms of the mutual information between the pairs of variables, even though it is possible to examine every variable independently.

The input to the discretization coalescence is an initial discretization. Simple discretization methods (e.g. interval discretization and quantile discretization) can be used as the initial discretization. The IPD algorithm includes 2 loops. The inner loop iterates over each of the variables to detech for each of the variables which single coalescence of adjacent discretization levels decreases the total mutual information score the least. The outer loop counts from the degree of the initial discretization down to one, with the loop index showing the number of discretization levels left. The total mutual information score for n discretization levels, $TMI(n)$, is defined as the sum of the pairwise mutual information between all pairs of variables when every one of them has been discretized into n discretization levels.

Suppose there are M variables to be discretized and each of them has N instances. Having applied stochastic quantile discretization of degree $D = N$ as an initial discretization, the total mutual information between variables before applying any of the discretization level coalescence (DLC) operators is $TMI(D)$. At each step of the algorithm, for every variable, the algorithm combines some adjacent pair of discretization levels into a single one, which decreases the total number of discretization levels by 1. Therefore, the algorithm runs for $D - 1$ steps and it stops when all observations have been discretized into a single level, which leads to a total mutual information score of 0, i.e $TMI(1) \equiv 0$. If the ultimate number of desired discretization levels is known in advance, then the algorithm can be stopped after computing the necessary number of coalescences instead of running for $D-1$ steps of the outer loop. If not, there are techniques for determining an appropriate value of discretization levels. To summarize one of them, in the IPD algorithm, the amount of information is lost at every step. So, for all $i < j$; $TMI(i) < TMI(j)$. Thus, it is known that $TMI(i)$ is a non-decreasing function of i with $TMI(1) \equiv 0$. Therefore,

the total mutual information obtained at each stage of the discretization level process does not change significantly until a small number of discretization levels are reached. Hence, an optimal number of discretization levels can be determined after running the algorithm.

4.4.1 Discretization of Psychiatric Variables Using the IPD

Luckily, the IPD is easily conducted in `R` using `discretize` function. After specifying how the data are initially discretized and the level of discretization, `R` produces the discretized variables.

Since the continuous variables used in this thesis represent the intensities of psychiatric properties, it is reasonable to classifty them as *low*, *normal*, and *high*. Hence, `R` discretizes them as follows (for `R` code, see Appendix-E):

1. BDI score:
   - [0, 11.5) : *low*
   - [11.5, 20.8) : *normal*
   - [20.8, 63] : *high*
2. BHS score:
   - [0, 6.89) : *low*
   - [6.89, 13) : *normal*
   - [13, 20] : *high*
3. RSES score:
   - [0, 15.8) : *low*
   - [15.8, 19.8) : *normal*
   - [19.8, 30] : *high*

4.5 Discretization by Domain Knowledge

Another way to discretize a continuous variable is to categorize it by the available domain knowledge in the literature. Although this approach has been used in many studies, from a statistical point of view, it is not advisable because it may lead to the loss of information and power [30], [31].

Although categorization is a common practice in statistics and other domain, there is no unified criteria to categorize continuous variables. It can significantly differ from one domain expert to another. However, using more than 2 categories can be desirable because this provides the opportunity to minimize the loss of information [32].

4.5.1 Discretization of Psychiatric Variables Using Domain Knowledge

In this thesis, using the available domain knowledge, the continuous variables are discretized as follows (for R code see Appendix-D):

1. BDI score: following [19], it is categorized into four levels:
   - [0, 10] :*none or minimal*
   - (10, 18] :*mild*
   - (18-29] :*moderate*
   - [30-63] :*severe*
2. BHS score: following [33], it is categorized into four levels:
   - [0, 3] :*none or minimal*
   - (3, 8] :*mild*
   - (8, 13] :*moderate*
   - [14, 20] :*severe*
3. RSES score: following [34], it is categorized into three levels:
   - [0, 15] : *low*
   - (15, 25] : *normal*
   - (25, 30] : *high*

4.6 Conclusion

In this chapter, the motivation and theory of discretization was presented.

Two methods for data discretization were described. The first one is the information-preserving discretization which aims to maximize the mutual information between the pairs of variables in the discretization process. The other approach uses the available expert knowledge in order to categorize continuous variables. In addition, the continuous variables in the data were discretized by these two approaches and the results were provided.

The next chapter reviews and introduces the theory of the Bayesian networks.

CHAPTER 5

BAYESIAN NETWORKS

5.1 Introduction

This chapter is concerned with introducing and reviewing the Bayesian networks (BNs). It starts by giving the definition of Bayes' Theorem, from which the BNs emerged, in Section 5.2. Next, the mathematical properties of the BNs and the terms used in the field of the BNs are presented in Section 5.3. Then, learning the BNs from the data is described in Section 5.4. Finally, in Section 5.5, the chapter is concluded by summarizing what has been detailed throughout the chapter.

5.2 Bayes' Theorem

The origin of the Bayesian networks comes from the interpretation of Bayes's Theorem, developed by Reverend Thomas Bayes. It is a simple and non-controversial theorem to calculate conditional probabilities [35]. It provides the opportunity to update the beliefs of an event, say X, given information about another event, say Y. While $P(X|Y)$ is called the *posterior* probability of X given Y, the probability $P(Y|X)$ is called the *likelihood* of X given Y [36]. So, the posterior probability is calculated as follows:

$$P(X|Y) = \frac{P(X \cap Y)}{P(Y)}, \qquad P(Y) \neq 0 \tag{5.1}$$

$$= \frac{P(Y|X)P(X)}{P(Y)} \tag{5.2}$$

$$= p \tag{5.3}$$

It should be emphasized that $P(X|Y) = p$ does not imply that whenever Y is true, then the probability of X is p. It implies that when Y is true, and everything else known is irrelevant for X, then the probability of X is p [37].

Using the total probability rule, Bayes Theorem can be stated as

$$P(X_k|Y) = \frac{P(X_k)P(Y|X_k)}{\sum_{i=1}^{n} P(X_i)P(Y|X_i)} \tag{5.4}$$

where X_1, X_2, ..., X_n constitute a partition of the sample space Ω and $P(X_i) \neq 0$ for $i = 1, 2, \ldots, n$ for any event Y in Ω such that $P(Y) \neq 0$ [38]

5.3 Bayesian Networks (BNs)

A Bayesian network (BN) is a graphical representation of a probability distribution over a set of variables. It comprises a directed acyclic graph (DAG) and a set of probability distributions [39].

Let $\mathcal{G} = (\mathcal{N}, \mathcal{A})$ be a DAG, where $\mathcal{N}$ is a finite set whose elements are named *nodes* (or *vertices*) and $\mathcal{A}$ is a finite set of *arcs* (or *edges*) between the nodes. Every node $n \in \mathcal{N}$ in $\mathcal{G}$ represents a random variable, X_n. The set of variables associated with $\mathcal{G}$ is denoted by $\mathbf{X} = (X_n)_{n\in\mathcal{N}}$. A BN, $\mathcal{B}$, for random variables $\mathbf{X}$ is a combination of $\mathcal{G}$ and $\mathbf{\Theta}$, denoted $\mathcal{B} = (\mathcal{G}, \mathbf{\Theta})$, where $\mathbf{\Theta}$ is the set of local distributions for all of the variables in $\mathcal{B}$ [40].

If $(X, Y) \in \mathcal{A}$, it is said that there is an arc from X and Y and X and Y are *incident* to the arc. If there is an arc from X to Y or from Y to X, then X and Y are *adjacent*. If there is an arc from Y to X, Y is called a *parent* of X and X is called a *child* of Y [41].

A node without any parents is called a *root* node and shows an input variable. A node which does not have any children is called a *leaf* node and constitutes an output variable. For each node (apart from the root node), a *conditional probability table* (CPT) is assigned. The probabilities in the CPT represent the strength and weights of causal relationships between the nodes when other nodes are in a particular state.

When the *prior* probabilities of a number of variables have been specified, the posterior probabilities for all of the nodes in the BN can be calculated using basic probability calculus and Bayes' Theorem given in Section 5.2. As new evidence is obtained in the form of observations about one or more variables, the prior probabilities for node states are updated, which is called *instantiation*.

5.3.1 A Classic BN example

To understand the fundamentals of the BNs, a classic example of the BNs adapted from [42] can be useful. Figure 5.1 shows the classic "$Weather$" example. The BN in Figure 5.1 has 4 variables:

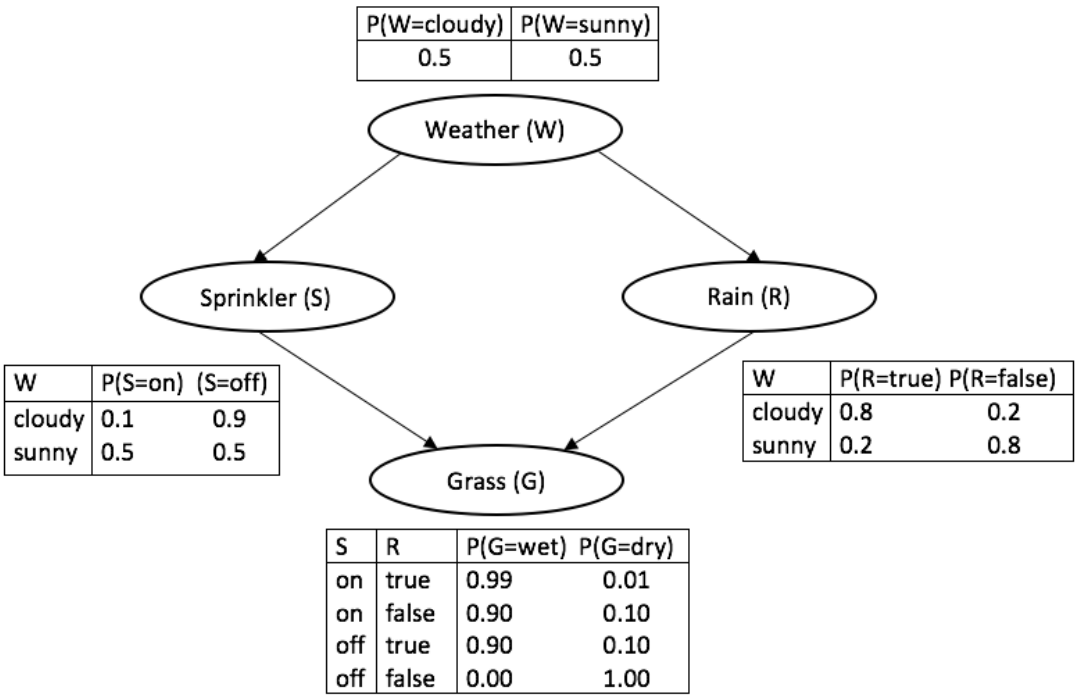


Figure 5.1 Classic Wet Grass BN Example

1. The first variable is $Weather$ (W) which shows the possible status of the sky. It is discrete and has 2 states: *cloudy* and *sunny*

2. The second variable is *Sprinkler* (S) which is discrete and has 2 states: *on* and off.

3. The third variable is *Rain* (R) which is discrete and has 2 states: *true* and $false$.

4. The last variable is *Grass* (G) which is discrete and has 2 states: *wet* and *dry*.

In Figure 5.1, the nodes named $Weather$, $Sprinkler$, $Rain$ and $Grass$ represent the variables and the arcs between the nodes show the relationships between the variables. For example, $Weather$ is the parent node of $Sprinkler$ and $Rain$ and $Sprinkler$ and $Rain$ are the child nodes of $Weather$. So, $Weather$ directly affects the current state of $Sprinkler$ and $Rain$. Similarly, $Sprinkler$ and $Rain$ are the parent nodes of $Grass$ and $Grass$ are the child node of both $Sprinkler$ and $Rain$. So, the state of $Grass$ is directly dependent on both $Sprinkler$ and $Rain$.

The tables below the nodes, one table per variable, are the CPTs which quantify the strength of dependency relations. If a variable has no parents, (e.g. $Weather$ in Figure 5.1), its local probability is unconditioned. If a variables has one or more parents (e.g. $Sprinkler$ and $Grass$), then its local distribution across the states of the variable is conditioned on all of the possible values of the parent set [43].

According to the CPTs in Figure 5.1, there is prior evidence that it is sunny in 50% of the time, i.e $P(Weather = sunny) = 0.5$. If the current state of $Weather$ is $sunny$, the probability that $Sprinkler$ will be on is 0.5 ($P(S = on|W = sunny) = 0.5$) and the probability that it will rain is 0.2 ($P(R = true|W = sunny) = 0.2$) . When it is known that the current state of $Weather$ is $cloudy$, the probability that $Sprinkler$ will be on is only 0.1 ($P(S = on|W = cloudy) = 0.1$), and the probability that it will rain is 0.8 ($P(R = true|W = cloudy) = 0.2$). Since $Grass$ has two parent nodes, the current states of both $Sprinkler$ and $Rain$ have to be known in order to make predictions about the state of $Grass$. In other words, knowing that it is not raining only is not enough to conclude that the state of $Grass$ are wet. The state of $Sprinkler$ too has to be known because there is a chance that the water on the lawn comes from the sprinkler. For example, when it is known that it is not raining but the sprinkler is on, it is highly likely that the grass will be wet ($P(G = wet|S = on, R = false) = 0.9$).

5.3.2 Representation of the Joint Probability Distribution

Representation of probabilistic knowledge of a domain $\mathbf{X} = \{X_1, \ldots, X_n\}$ can be done constructing the full joint probability distribution (FJPD) across the set of do-

main variables. Using the chain rule, the FJPD is defined as $P(\mathbf{X}) = P(X_1, \ldots, X_n)$. So, the FJPD is the probability of all combinations as defined by the values of all the variables [43], [44].

Although this view of probability theory is useful for maintaining consistency and proving mathematical theorems, it is insufficient for representing human reasoning. For instance, consider encoding of the joint distribution, $P(\mathbf{X}) = P(X_1, \ldots, X_n)$. In order to store $P(\mathbf{X})$, 2^n probabilities are required since the number of probabilities required increase exponentially. Event though, a way to store $P(\mathbf{X})$ is found, there still exists the problem of computing from it the probabilities of propositions considered interesting. For example, the computation of the marginal probability $P(X_i)$ requires summing $P(X_1, \ldots, X_n)$ over all 2^{n-1} combinations of the remaining $n-1$ variables. This is computationally-expensive and takes too much effort .

Another drawback of numerical representation of probabilistic knowledge is that it lacks *psychological meaningfulness*. The numerical representation can generate plausible probability measures for all propositional sentences, yet it often results in computations which a human reasoner would not use. Consequently, the process leading from the premises to the conclusions cannot be followed, justified, or tested by the users or even the designers of the reasoning system. For example, simple tasks such as the calculation of the influence of a piece of evidence $E = e$ on a hypothesis $H = h$ via

$$P(h|e) = \frac{P(h \cap e)}{P(e)} \tag{5.5}$$

$$= \frac{\sum_{x_i : X_i \neq H,E} P(X_1, \ldots, X_n)}{\sum_{x_i : X_i \neq E} P(X_1, \ldots, X_n)} \tag{5.6}$$

takes a very high number meaningful mathematical operations, which is not by supported familiar mental processes [45].

To overcome these problems, conditional independence assumptions can be used. The reduction in complexity is obtained by factorizing the FJPD into a set of direct probability distributions, which is detailed in the next section.

5.3.3 Factorization of the Joint Probability Distribution

As defined earlier, $P(\mathbf{X}) = P(X_1, \ldots, X_n)$ indicates the FJPD for a domain, $\mathbf{X} = \{X_1, \ldots, X_n\}$. A Bayesian network structure $\mathcal{B}$ for $\mathbf{X}$ indicates a set of conditional dependence relationships between the variables in $\mathbf{X}$. Accordingly, the structure $\mathcal{B}$ for $\mathbf{X}$ provides a compact and short graphical factorization of the joint probability distribution, $P(\mathbf{X})$ [46]. So, the factorization of the joint probability model is obtained using the algorithm below [43].

Algorithm 1 Factorization of the joint probability distribution

1: Order the variables such that X_i is a candidate parent of X_j if X_i appears before X_j in the order
2: Apply the chain rule of probability to factorize the joint probability into a set of conditional independent probabilities. Hence, every distribution in the chain corresponds to $P(X_i|pa(X_i))$, meaning that $P(X_i)$ is dependent only on the values of its local parents.
3: For each variable in the chain (obtained in the previous step), prune the parents by making conditional independence assumptions.

Applying the chain rule of probability, the factorized joint probability distribution for a Bayesian network is obtained as follows:

$$P(\mathbf{X}) = P(X_1, \ldots X_n) \tag{5.7}$$

$$= P(X_1)P(X_2|X_1) \ldots P(X_i|X_1, \ldots, X_{i-1}) \ldots P(X_n|X_1, \ldots, X_{n-1}) \tag{5.8}$$

$$= P(X_1)P(X_2|pa(X_2)) \ldots P(X_i|pa(X_i)) \ldots P(X_n|pa(X_n)) \tag{5.9}$$

$$= \prod_{i}^{n} P(X_i|pa(X_i)) \tag{5.10}$$

where n is the number of variables in the domain and $pa(X_i)$ is the parent(s) of node X_i in the BN.

The full joint probability distribution requires $O(2^n)$ probabilities, whereas the factorized joint probability distribution requires $O(n2^k)$ probabilities, where k is the maximum number of parents feeding into a node, which shows that the conditional

relationship assumptions help represent the joint probability distribution more compactly [47].

So, in the classic BN example in Section 5.3.1, the full joint probability distribution is calculated as follows:

$$P(W, S, R, G) = P(W)P(S|W)P(R|W, S)P(G|W, S, R), \tag{5.11}$$

since R is conditionally independent of S given its parent node W ($R \perp\!\!\!\perp_P S|W$) and G is conditionally independent of W given its parent nodes S and R ($G \perp\!\!\!\perp_P W|S, R$)

$$P(W, S, R, G) = P(W)P(S|W)P(R|W)P(G|S, R). \tag{5.12}$$

5.3.4 Conditional Independence

If the occurrence of event the Y does not affect the probability of the occurrence of the event X, then X and Y are *independent* [36], that is,

$$P(X|Y) = P(X). \tag{5.13}$$

This definition of independence is symmetric, which means if X is independent of Y, then Y is independent of X. This can be shown using Bayes' Theorem:

$$P(Y|X) = \frac{P(X \cap Y)}{P(X)} \tag{5.14}$$

$$= \frac{P(X|Y)P(Y)}{P(X)} \tag{5.15}$$

if X is independent of Y, then $P(X|Y) = P(X)$,

$$P(Y|X) = \frac{P(X)P(Y)}{P(X)} \tag{5.16}$$

$$= P(Y). \tag{5.17}$$

Hence, the probability that the two event will occur is calculated as follows:

$$P(X \cap Y) = P(X|Y)P(Y) \tag{5.18}$$
$$= P(X)P(Y). \tag{5.19}$$

So, the probability that both event will occur is calculated by multiplying the probabilities of the individual events.

If, when the event Z is known, knowing the event Y adds no additional information about the event X, then the events X and Y are *conditionally independent* given the event Z:

$$P(X|Y \cap Z) = P(X|Z). \tag{5.20}$$

Similar to the definition of independence, conditional independence is also symmetric. If the event X is conditionally independent of the event Y once the event Z is known, then the event Y is conditionally independent of the event X when the event Z is known. To illustrate:

$$P(Y|X \cap Z) = \frac{P(X \cap Y|Z)P(Z)}{P(X|Z)P(Z)} \tag{5.21}$$
$$= \frac{P(X|Y \cap Z)P(Y|Z)}{P(X|Z)} \tag{5.22}$$
$$= \frac{P(X|Z)P(Y|Z)}{P(X|Z)} \tag{5.23}$$
$$= P(Y|Z) \tag{5.24}$$

So, when the two events are conditionally independent, the probability that the two events will occur is calculated as follows:

$$P(X \cap Y|Z) = P(X|Z) \times P(Y|Z). \tag{5.25}$$

5.3.5 D-Separation

Let **A, B** and **C** are 3 disjoint subsets of nodes in a DAG $\mathcal{G}$. **C** is said to *d-separate* **A** from **B**, denoted $\mathbf{A} \perp\!\!\!\perp_G \mathbf{B}|\mathbf{C}$, if along every path between a node in **A** and a node in **B**, there is a node n which satisfies one of the conditions below [48]:

1. n has converging arcs and neither n nor any of its descendants are in **C**.
2. n is in **C** and does not have converging arcs.

The term d-separation is short for *directed* separation. The letter d indicates that this is not the standard use of the term *separation* found in the graph theory. In common language, d-separation can be regarded in terms of *irrelevance*, meaning that if **A** and **B** are d-separated new evidence about one of the sets **A** and **B** does not change the probability distribution of the other. If **A** and **B** are not d-separated, they are called *d-connected* [49].

D-separation is used determine whether or not two variables are conditionally independent. If **A** and **B** are d-separated given **C**, i.e. $\mathbf{A} \perp\!\!\!\perp_G \mathbf{B}|\mathbf{C}$, they are said to be conditionally independent, i.e. $\mathbf{A} \perp\!\!\!\perp_P \mathbf{B}|\mathbf{C}$ [2]

5.3.6 Connection Types in BNs

In the BNs, types of connection between the nodes define the relationships between them. They are known as *fundamental connections* and are the building blocks of the probabilistic and graphical properties of the BNs [48]. There are 3 types of connections in the BNs, which are described below.

Serial connection

A connection like $X \rightarrow Y \rightarrow Z$ (Figure 5.2) is called a *serial* connection. The information about X gives the significant information about the probability of Y and the knowledge at Y gives the significant information on Z. Nevertheless, if Y is known, the information about X adds no further information about Z as X and Z become conditionally independent. For example, knowing that the weather (X) is sunny could decrease the probability that it will rain (Y), which could rise the probability that the grass (Z) is dry. Nevertheless, when it is known that it is

raining (Y), whether or not the weather (X) is sunny does not affect the probability that the grass (Z) is dry.

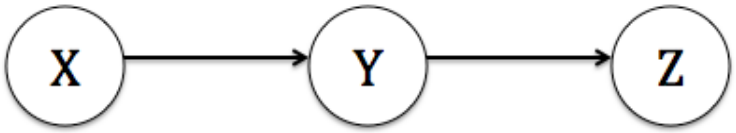


Figure 5.2 Serial connection

Diverging connection

A connection like $X \leftarrow Z \rightarrow Y$ (Figure 5.3) is a called *diverging* connection. The information about X affects the probability of Y via Z, but when the state of Z is known, the information about X does not provide any additional information about Y. For instance, when the current state of the weather (Z) is known to be cloudy, information about the current state of the grass (X) does not affect the probability that it is raining (Y).

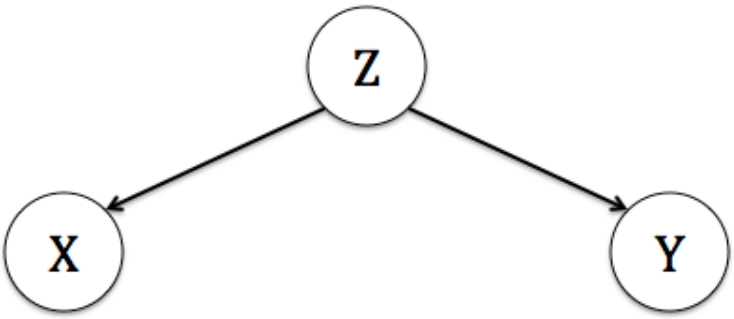


Figure 5.3 Diverging connection

Converging Connection

A connection like $X \rightarrow Z \leftarrow Y$ (Figure 5.4) is known as a *converging* connection. For the information about X to bring additional information about Z, the state of Y must be known because Z depends on both X and Y. So, the information can be provided from X to Y if and only if Y or one of its descendants is known. For example, information that the grass (Z) is wet and it is raining (Y) impacts the

probability of the sprinkler's (X) being on or off. However, if there is no information about the current state of the grass (Z), knowing the current state of rain (Y) does not influence the probability that the sprinkler is on or off.

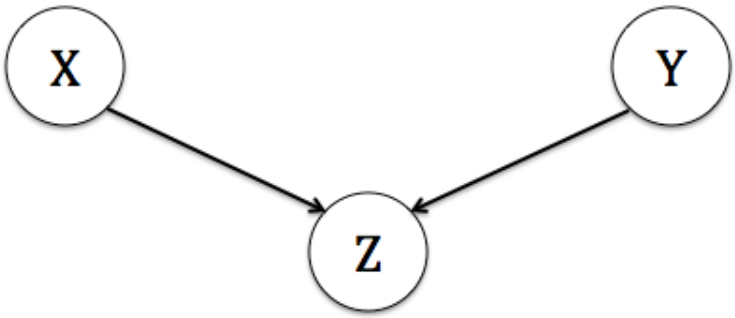


Figure 5.4 Converging connection

5.4 Learning Bayesian Networks

In the domain of Bayesian networks, model selection and estimation are known as *learning* and it consists of two steps:

1. *structure learning*, that is, the determination of the topology of the BN.
2. *parameter learning*, i.e. the estimation of conditional probabilities once the BN topology has been obtained [40].

Both of these steps above can be conducted either as *unsupervised learning*, i.e. using the evidence provided by the data, or as *supervised learning*, using the expert knowledge in the domain. However, the prior information available in the domain is generally insufficient for the expert(s) to entirely construct the BN. Furthermore, different experts can give different and sometimes conflicting assessments because of the subjective nature of the process. In addition, when the number of variables is high, determination of the DAG may be impossible. In such situations, the dependency structure and the associated conditional probability distributions can be estimated from the data [50]. It is also common to combine the supervised and unsupervised learning methods.

Let $\mathcal{D}$ be a data set, $\mathcal{B} = (\mathcal{G}, \mathbf{\Theta})$ a BN and $\mathbf{\Theta}$ the parameters of the global distribution of $\mathbf{X}$. Then, BN learning can be formalized as follows

$$\underbrace{P(\mathcal{B}|\mathcal{D}) = P(\mathcal{G}, \mathbf{\Theta}|\mathcal{D})}_{\text{learning}} = \underbrace{P(\mathcal{G}|\mathcal{D})}_{\text{structure learning}} \times \underbrace{P(\mathbf{\Theta}|\mathcal{G}, \mathcal{D})}_{\text{parameter learning}} . \tag{5.26}$$

In practice, structure learning can be undertaken by obtaining $\mathcal{G}$ which maximizes

$$P(\mathcal{G}|\mathcal{D}) \propto P(\mathcal{G})P(\mathcal{D}) = P(\mathcal{G}) \int P(\mathcal{D}|\mathcal{G}, \mathbf{\Theta})P(\mathbf{\Theta}|\mathcal{G})d\mathbf{\Theta} \tag{5.27}$$

using Bayes' theorem to decompose the posterior probability of $\mathcal{G}$ ($P(\mathcal{G}|\mathcal{D})$) into the product of the prior distribution over the possible DAGs ($P(\mathcal{G})$) and the probability of the data ($P(\mathcal{D}|\mathcal{G})$). Since it is impossible to calculate $P(\mathcal{D}|\mathcal{G})$ without estimating the parameters $\mathbf{\Theta}$ of $\mathcal{G}$, $\mathbf{\Theta}$ has to be integrated out of Equation 5.27 to make $P(\mathcal{G}|\mathcal{D})$ independent of any specific choice of $\mathbf{\Theta}$.

The prior distribution $P(\mathcal{G})$ gives the opportunity to represent any prior knowledge available about the conditional independence relationships between the variables in $\mathbf{X}$. Using $P(\mathcal{G})$, some arcs can be added in or removed from the DAG. Similarly, the directions of the arcs can be specified in the DAG. For example, demographic variables age and gender cannot be head of any arcs because these variables cannot be influenced by any variables. This information is added to the DAG using $P(\mathcal{G})$.

The most widely-used choice for $P(\mathcal{G})$ is a non-informative prior over the space of the possible DAGs, that is, allocating the same probability to every DAG. It is also possible to exclude some DAGs because of prior information, as mentioned in the previous paragraph. Using more complex priors, called *structural* priors, is also possible, but they are not preferred in practice. The first reason is that using a uniform probability distribution makes $P(\mathcal{G})$ irrelevant in maximizing $P(\mathcal{G}|\mathcal{D})$. The second reason is that the number of possible DAGs rises super-exponentially in the number of nodes. For instance, in DAG with n nodes, there are $\frac{1}{2}n(n-1)$ possible arcs given the pairs of nodes in $\mathcal{N}$. Even if the arcs' directions are disregarded, there are still $2^{O(p^2)}$ possible arcs. This makes it extremely difficult to determine a prior distribution over such a high number of DAGs for even small problems.

The calculation of $P(\mathcal{D}|\mathcal{G})$ is also a challenging task from both a computational and

algebraic point of view. Starting from the decomposition into local distributions, $P(\mathcal{D}|\mathcal{G})$ can be factorized as follows:

$$P(\mathcal{D}|\mathcal{G}) = \int \prod_{i=1}^{n} [P(X_i|\Pi_{X_i}, \mathbf{\Theta}_{X_i}) P(\mathbf{\Theta}_{X_i}|\Pi_{X_i})] d\mathbf{\Theta} \tag{5.28}$$

$$= \prod_{i=1}^{n} \left[\int P(X_i|\Pi_{X_i}, \mathbf{\Theta}_{X_i}) P(\mathbf{\Theta}_{X_i}|\Pi_{X_i}) d\mathbf{\Theta}_{X_i} \right] \tag{5.29}$$

$$= \prod_{i=1}^{n} E_{\mathbf{\Theta}_{X_i}}[P(X_i|\Pi_{X_i})]. \tag{5.30}$$

If it is possible calculate all expectations in close form, $P(\mathcal{D}|\mathcal{G})$ can be calculated in a feasible time even for big data sets. This is valid both for the multinomial distribution assumed for discrete BNs (via its conjugate Dirichlet posterior) and for the multivariate Gaussian distribution assumed for continuous BNs (via its conjugate Inverse Wishart distribution). For discrete BNs, $P(\mathcal{D}|\mathcal{G})$ can be calculated using the *Bayesian Dirichlet equivalent uniform* (BDeu or simply BDe) score from [51]. BDe assumes a flat prior over both the space of the DAGs and the parameter space of each node:

$$P(\mathcal{G}) \propto 1 \text{ and } P(\mathbf{\Theta}_{X_i}|\Pi_{X_i}) = \alpha_{ij} = \frac{\alpha}{|\mathbf{\Theta}_{X_i}|} \tag{5.31}$$

where α is the *imaginary sample size* affiliated with the Dirichlet prior, indicating the amount of weight allocated to the prior as the size of an imaginary sample supporting it.

Under these assumptions, BDe is

$$\text{BDe}(\mathcal{G}, \mathcal{D}) = \prod_{i=1}^{n} \text{BDe}(X_i, \Pi_{X_i}) \tag{5.32}$$

$$= \prod_{i=1}^{n} \prod_{j=1}^{s_i} \left\{ \frac{\Gamma(\alpha_{ij})}{\Gamma(\alpha_{ij} + m_{ij})} \prod_{k=1}^{r_i} \frac{\Gamma(\alpha_{ij} + m_{ijk})}{\Gamma(\alpha_{ijk})} \right\} \tag{5.33}$$

where r_i is the number of categories for the node X_i, s_i is the number of configurations of the categories of the parents of X_i and m_{ijk} is the number of samples that have the j-th category for node X_i and the k-th configuration for its parents.

The corresponding posterior probability for Gaussian Bayesian networks is called

Bayesian Gaussian equivalent uniform (BGeu or simply BGe) from [52]. BGe assumes a non-informative prior over both the space of the DAGs and the parameter space of each node and has only one parameter, that is, the imaginary sample size α [53].

5.4.1 Structural learning

Learning a BN starts with learning the structure of it, that is, learning a topology encoding the significant characteristics of the dependence structure between the variables in $\mathbf{X}$ of the data [54]. The goal of learning is to obtain a BN structure which estimates the data well [55]. In literature there are many methods to learn the structure of a BN. In spite of the diversity of theoretical backgrounds and terminology, these methods can be categorized into three groups: (i) *constraint-based* (ii) *score-based* and (iii) *hybrid* (a mixture of (i) and (ii)).

These three approaches work under a set of common assumptions [53]:

- There must exist a one-to-one correspondence between the nodes in the DAG and the random variables in $\mathbf{X}$, meaning that there must not be multiple nodes that are deterministic functions of a single variable.

- Every collection of the possible values of the variables in $\mathbf{X}$ must represent a valid and observable event. This assumption emphasize an exactly positive global distribution needed to have uniquely determined Markov blankets and, therefore, a uniquely identifiable model. Even if this assumption is not satisfied, constraint-based approaches still work since the presence of a perfect map is also a adequate condition for the uniqueness of the Markov blankets [45]

- All of the relationships between the variables in $\mathbf{X}$ must be conditional independencies since they are the only type of relationships which can be expressed by a BN.

- Observations must be independent of each other. If there are dependencies between the observations, they must be particularly accounted for in the definition of the BN, e.g. *dynamic* BNs.

Hill-climbing search algorithm

In literature, there are many structure learning algorithms, but there is no strong evidence that one approach is better than another [56]. Each of them has advantages and disadvantages. In this study, one of the widely-used score-based algorithms, that is, *Hill-climbing with tabu search* is used.

Hill-climbing search algorithm is a *greedy search* algorithm used to find the DAG with the highest score among the possible DAGs. This methods relies on the definition of a neighborhood space among solutions and on local moves searching for an improving solution in the neighborhood incumbent solution [57]. Algorithm 2 shows how Hill climbing search algorithm works.

Algorithm 2 Hill-climbing search algorithm

1: Select such a DAG $\mathcal{G}$ over $\mathcal{N}$ that it is generally (but not necessarily) empty
2: Calculate the score of $\mathcal{G}$ ($Score_{\mathcal{G}}$)
3: Set the maximum score ($Score_{max}$) to the score of $\mathcal{G}$, i.e. $Score_{max} = Score_{\mathcal{G}}$
4: Repeat the steps below as long as $Score_{max}$ rises:
   (a) add, delete or reverse a possible arc as long as the resulting structure is still a DAG:
      (i) calculate the score of the new network $\mathcal{G}_{new}$, denoted as $Score_{\mathcal{G}_{new}}$
      (ii) if $Score_{\mathcal{G}_{new}} > Score_{\mathcal{G}}$, set $\mathcal{G} = \mathcal{G}_{new}$ and $Score_{\mathcal{G}} = Score_{\mathcal{G}_{new}}$
   (b) Replace $Score_{max}$ with $Score_{\mathcal{G}}$
5: Return the DAG $\mathcal{G}$

The hill climbing algorithm is attractive because it is simple and only a few parameters are necessary [58]. However, it has an important drawback, that is, it may be trapped at a local optimum and, therefore, an optimum solution may never be reached [59]. To overcome this problem, a *tabu list* is used. This list is defined as a short-term memory which covers the recently visited structures so that the algorithm try new structures. Hence, the optimization does not get stuck and reaches a solution. [60]

5.4.2 Parameter Learning

Once the structure of the BN has been learned from the data, the estimation of parameters is implemented. The widely-used methods for parameter learning for complete data are *Maximum Likelihood Estimation* and *Bayesian Estimation* and for incomplete data *Expectation-Maximization*, *Robust Bayesian Estimate*, *Monte Carlo Method* and *Gaussian Approximation*.

In this study, Bayesian estimation method is used and explained in the next section.

Bayesian Estimation

In the Bayesian estimation of parameter learning, a prior distribution is assumed over the parameters of the local probability density functions (PDFs) before the data are used. The conjugacy of this prior distribution is desirable. A distribution to family is called *conjugate prior* to a data distribution if the posterior over the parameters belongs to the same family as the prior.

For multinomial distributions, the conjugate prior comes from the Dirichlet family. The Dirichlet distribution is expressed as

$$P(\theta_{ij1}, \ldots, \theta_{ijr_i}|\mathcal{G}) = Dir(\alpha_{ij1}, \ldots, \alpha_{ijr_i}) \tag{5.34}$$

$$= \Gamma(\alpha_{ij}) \prod_{k=1}^{r_i} \frac{\theta_{ijk}^{\alpha_{ijk}-1}}{\Gamma(\alpha_{ijk})} \tag{5.35}$$

where θ_{ijk} is the probability of each bin, $k = 1, \ldots, r_i$, in the local PDF of variable X_i for the parent configuration $\mathbf{pa_{ij}}$, q_i the number of configurations of the set of parents of X_i, α_{ijk} are hyperparameters and $\alpha_{ij} = \sum_{k=1}^{r_i} \alpha_{ijk}$. Under the assumption of local and global parameter independence [51], [61], [62], the distribution over the set of parameters $\mathbf{\Theta}$ of the entire BN is

$$P(\mathbf{\Theta}|\mathcal{G}) = \prod_{i=1}^{n} \prod_{j=1}^{q_i} \Gamma(\alpha_{ij}) \prod_{k=1}^{r_i} \frac{\theta_{ijk}^{\alpha_{ijk}-1}}{\Gamma(\alpha_{ijk})}. \tag{5.36}$$

Conditional on the data set $\mathcal{D}$, the posterior probability over the parameters is a member of the Dirichlet family as it is conjugate prior to the multinomial. Hence,

$$P(\theta_{ij1}, \ldots, \theta_{ijr_i}|\mathcal{G}, \mathcal{D}) = Dir(N_{ij1} + \alpha_{ij1}, \ldots, N_{ijr_i} + \alpha_{ijr_i}) \tag{5.37}$$

and

$$P(\boldsymbol{\Theta}|\mathcal{G}, \mathcal{D}) = \prod_{i=1}^{n} \prod_{j=1}^{q_i} \Gamma(\alpha_{ij} + N_{ij}) \prod_{k=1}^{r_i} \frac{\theta_k^{N_{ijk}+\alpha_{ijk}-1}}{\Gamma(N_{ijk} + \alpha_{ijk})} \tag{5.38}$$

where N_{ijk} is the number of samples in bin k of the PDF for X_i for $\mathbf{pa_{ij}}$ [39].

5.5 Conclusion

In conclusion, in this chapter the theory necessary to understand and use the Bayesian networks was presented. First, Bayes' Theorem which the BNs come from was expressed. Next, the definition and notation of the BNs were given. Then, a classic example of the BNs was used to easily define the fundamentals of the BN. Finally, learning a BN from the data was detailed. Also, the commonly-used methods to learn both the structure and parameters of the BNs were mentioned.

In the next chapter, the data described in Chapter 2 are analyzed using the necessary methods detailed in the Chapters 4 and 5.

CHAPTER 6

RESULTS AND DISCUSSION

6.1 Introduction

This chapter covers the analysis of the data set described in Chapter 2 using the methodology given in Chapters 4 and 5.

In Section 6.2, the reliabilities of the psychiatric scales are tested and two BNs are constructed. In Section 6.2.1, the results of the first model are presented and in Section 6.2.2 the results of the second model are given. In Section 6.3, the performances of the two models are compared. Finally, in Section 6.4, the results are discussed and the chapter is concluded.

6.2 Results

Before constructing a BN, firstly, to test the reliabilities of the psychiatric scales, Cronbach's alpha internal consistency coefficients [63] needs checking. The Cronbach's alpha internal consistency coefficients for BHS, BDI and RSES are 0.87, 0.89 and 0.83 respectively, indicating that the reliabilities of these scales are satisfying as they are all greater than 0.70.

In this study, all of the analyses are conducted in `R` using mainly **bnlearn** [48], **gRain** [64], **Rgraphviz** [65], and **lattice** [66] packages.

In the original data set, the psychiatric variables are continuous. However, because of the reasons mentioned in Section 4.2, they are discretized using two different approaches, which leads to two different discrete BNs. The first BN whose data were discretized by domain knowledge (see Section 4.5) is called $\mathcal{B}_1$ and the other one whose data were discretized by the IPD (see Section 4.4.1) is called $\mathcal{B}_2$.

6.2.1 Results of $\mathcal{B}_1$

The analyses start by constructing $\mathcal{B}_1$. First of all, the demographic variables *Age* and *Gender* cannot be child nodes because nothing can influence an individual's age or gender. Therefore, all the arrows towards *Age* and *Gender* are blocked, making them root nodes. This is easily done using `blacklist` function in `R`. The whole `R` code used to construct $\mathcal{B}_1$ is given in Appendix **??**.

The structure of $\mathcal{B}_1$ is given in Figure 6.1 and it shows significant several direct and indirect relationships between the variables. To interpret the results in Figure 6.1 starting with demographic and socio-economic variables (given in gray and green nodes respectively), there is only one variable that has no relations with the remaining 20 variables, which is $Mother_occupation$. This means that $Mother_occupation$ neither affects nor is affected by any of the variables used in the analyses. This is because 93% of the mothers of students are unemployed. This is illustrated in Figure 6.2.

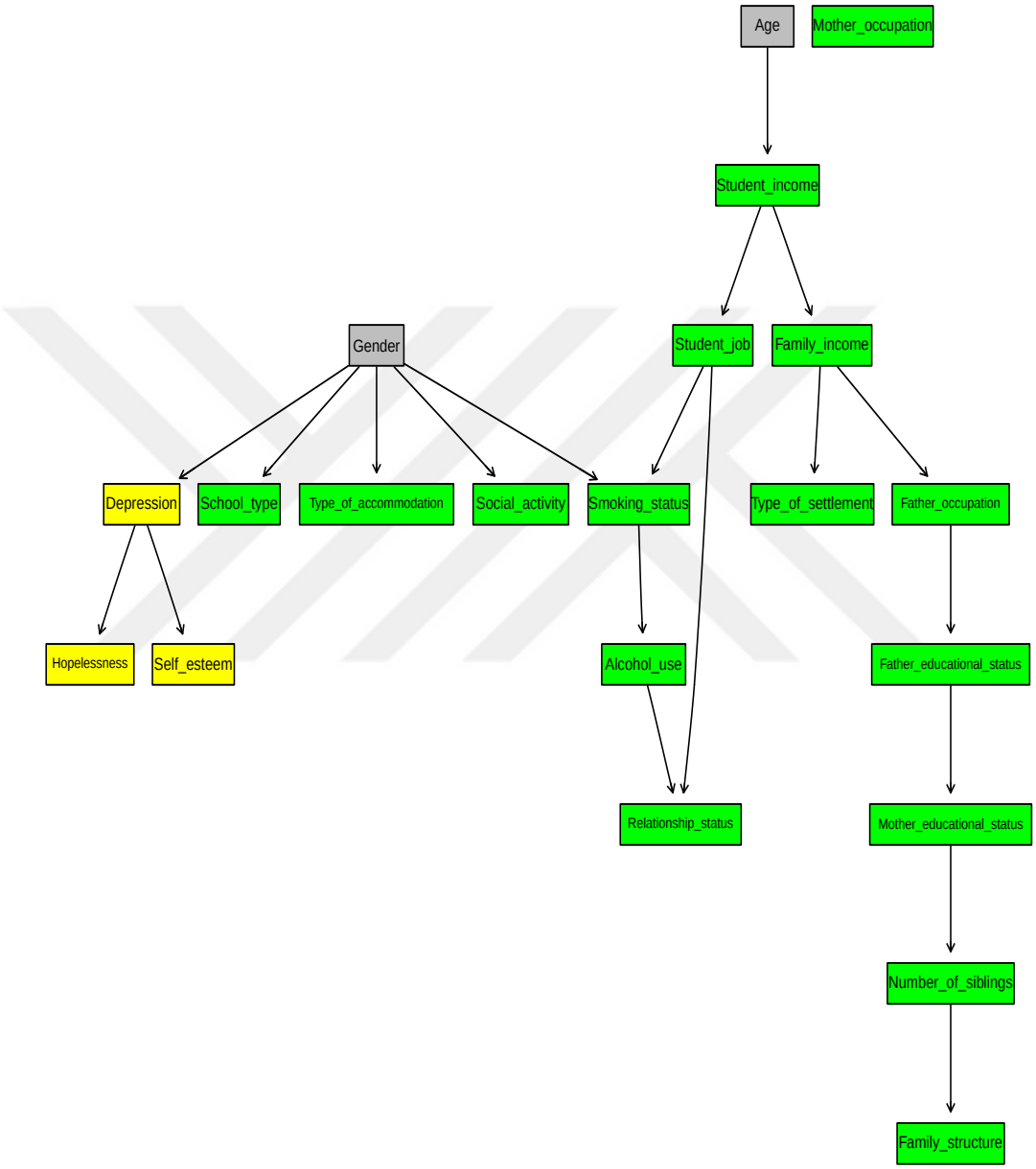


Figure 6.1 DAG of $\mathcal{B}_1$

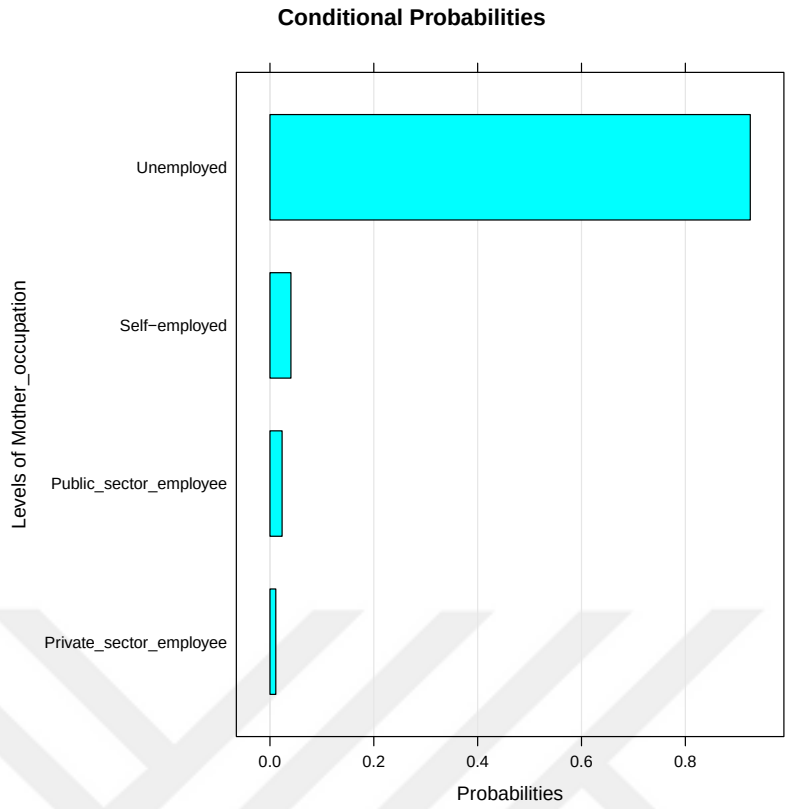


Figure 6.2 Bar chart of conditional probabilities of *Mother_occupation*

From 6.1, it is seen that *Age* directly affects *Student_income*. According to Figure 6.3, the students with an income of 0-500 TL dominates the others in each age group. While the proportions of students receiving 0-500 TL monthly are approximately the same (0.85) in age groups 20− and 20-23, this proportion is 0.68 in students aged over 23. Another remarkable point is that the percentage of students having an income over 1000 TL drastically increases as the age of student increases. This percentage is only 0.001 in students aged 20-, 0.03 in students aged 20-23 and 0.16 in students aged over 23.

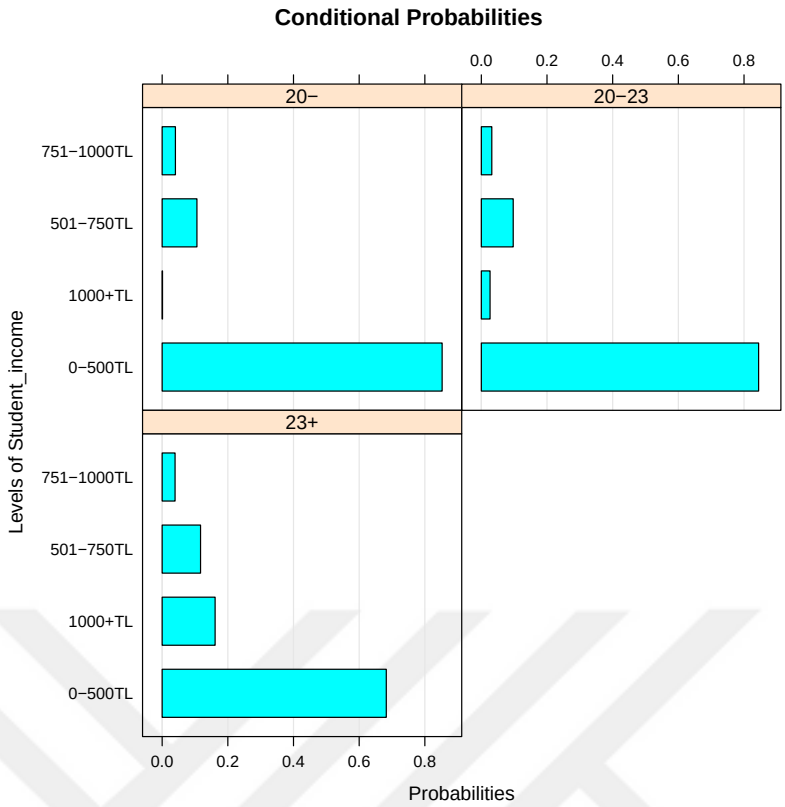


Figure 6.3 Bar chart of conditional probabilities of *Student_income* given *Age*

Figure 6.1 indicates that there is a diverging connection from *Student_income* to *Student_job* and *Family_income*, meaning that both *Student_job* and *Family_income* are directly influenced by *Student_job*.

From Figure 6.4, at first glance it is seen that the majority of students do not have any jobs. The students with incomes of 0-500 TL and 501-750 TL have approximately the same employment percentages. However, as the income of students rises to 751-1000 TL, the gap between the percentages of the students with and without jobs narrows down and the students with jobs outnumber those without job in the income group of 1000+ TL.

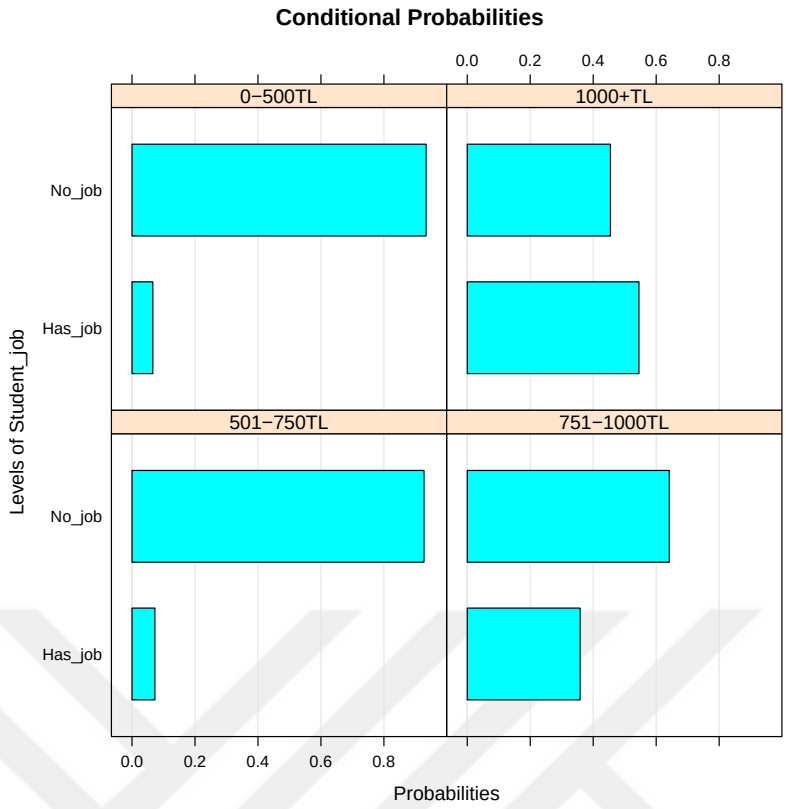


Figure 6.4 Bar chart of conditional probabilities of *Student_job* given *Student_income*

Figure 6.5 is the graph illustrating the conditional probabilities of *Family_income* given *Student_income*. It shows a positive correlation between *Student_income* and *Family_income*. For example, the proportion that a student has a family with an income of 0-1400 TL is 0.54 in the students with an income of 0-500 TL, 0.37 in the students with an income of 501-750 TL, 0.18 in the students with an income of 751-1000 TL, and 0.14 in the students with an income of over 1000 TL. Similarly, when a student is known to have an income of 0-500 TL, the probability that his/her family has an income of 5000+ TL is merely 0.04, but this probability rises to 0.13 for students whose family income is 501-750 TL, to 0.16 for students with a family income of 751-1000 TL, and to 0.41 for students with a family income of 1000+ TL. This is also the case in the other groups. These figures show that as the *Student_income* increases, so does *Family_income*.

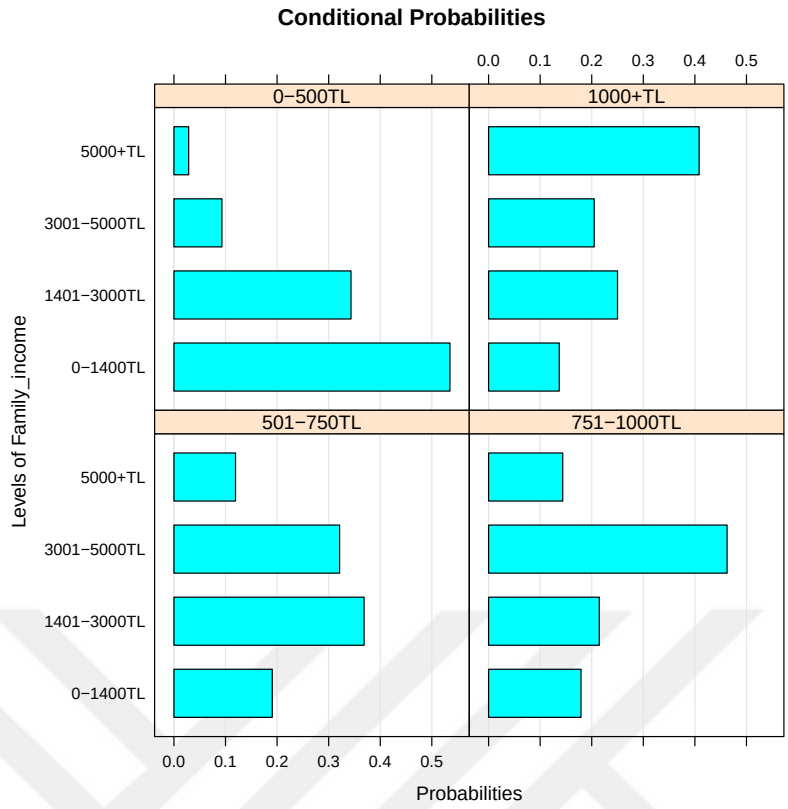


Figure 6.5 Bar chart of conditional probabilities of *Family_income* given *Student_income*

There is a converging connection from *Family_income* to *Type_of_settlement* and *Father_occupation*, meaning that where a student grew up and what his/her father does for a living depends on his/her family's income.

Figure 6.6 is the bar chart of conditional probabilities of *Type_of_settlement* given *Family_income*. The probability that a student is from a village or town is the highest (0.33) for the students with a family income of 0-1400 TL, but this probability is the lowest (0.17) for those whose family income is over 5000 TL. Another remarkable point is that the proportion of students from a metropolis is approximately the same (0.14) in the students with a family income of 0-1400 TL, 1401-3000 TL and 3001-5000 TL, but 0.29 in the students whose family income is over 5000 TL.

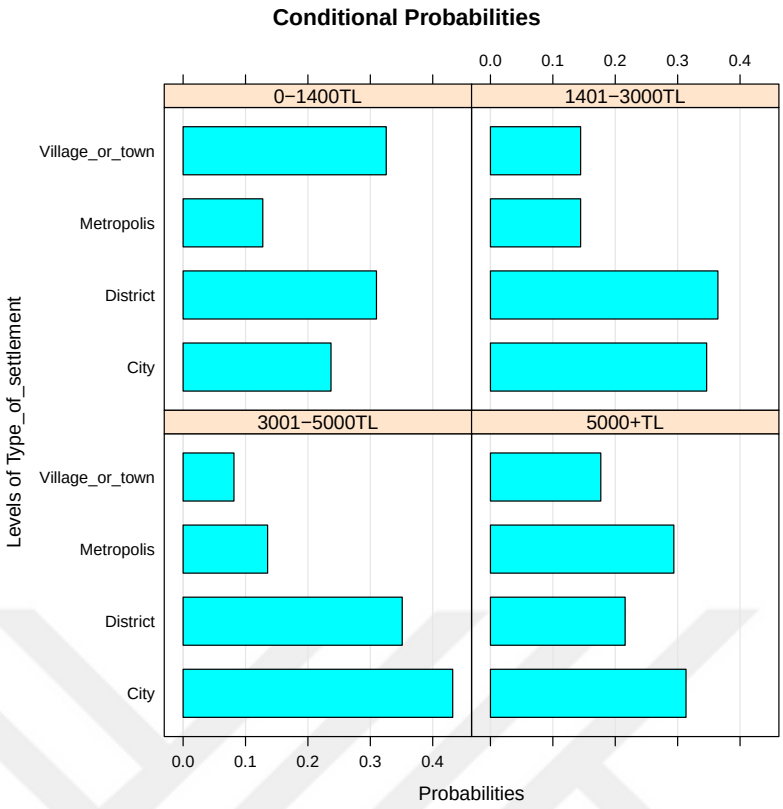


Figure 6.6 Bar chart of conditional probabilities of *Type_of_settlement* given *Family_income*

Figure 6.7 indicates the conditional probabilities of *Father_occupation* given *Family_income*. Firstly, the probability that a student's father is a private sector employee does not significantly change by the student's family income. When a student is known to have a family income of 0-1400 TL, the probability that his/her father is unemployed is the highest (0.44).

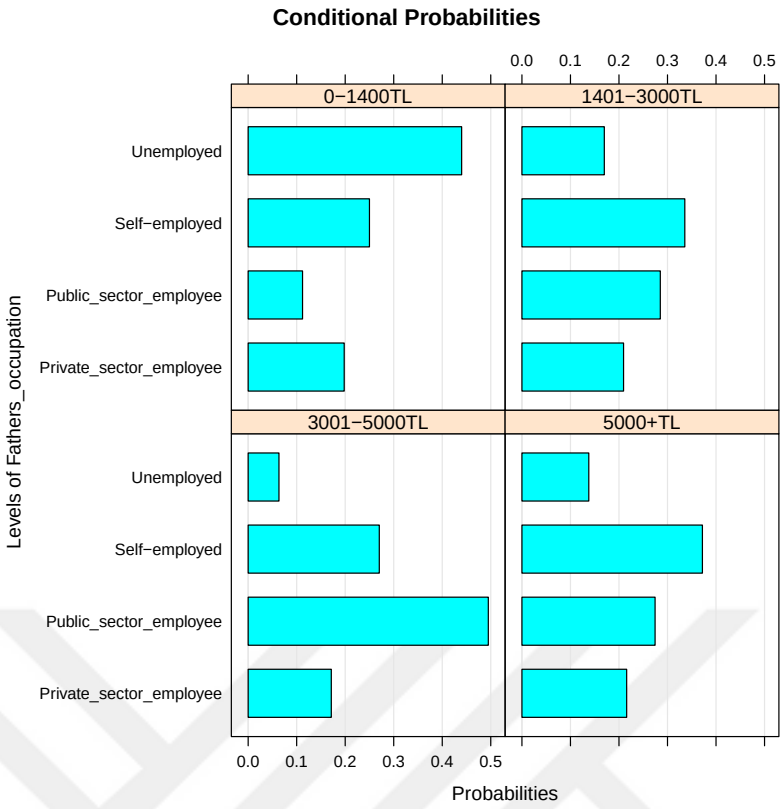


Figure 6.7 Bar chart of conditional probabilities of *Father_occupation* given *Family_income*

Figure 6.8 represents the conditional probabilities of *Father_educational_status* given *Father_occupation*. Primary school is the most common educational status among unemployed, self-employed and private sector employee fathers (0.57, 0.47 and 0.48 respectively), while master or Ph.D. is the least common one within each group. The probability that a private sector employee father is a university graduate is only 0.02, whereas the probability that a public sector employee father is a university graduate is 0.31.

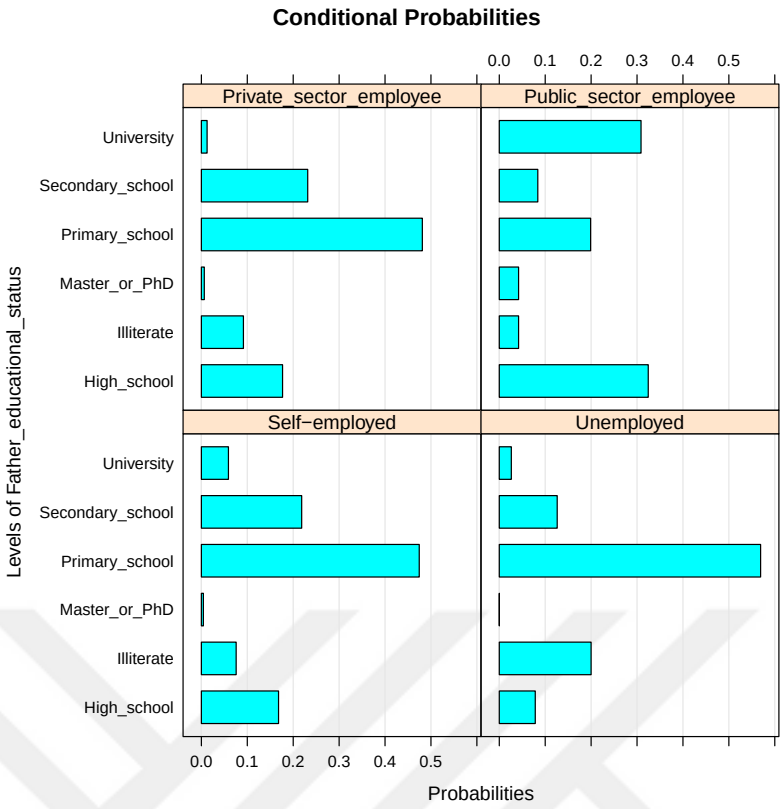


Figure 6.8 Bar chart of conditional probabilities of $Father_educational_status$ given $Father_occupation$

From Figure 6.9, it is seen that, generally speaking, the fathers of the students have wives that have lower educational levels than the fathers' themselves. To illustrate, when the father is known to be a high school graduate, the probability that the mother is a high school graduate is 0.14 whereas, in the same scenario, the probabilities that the mother is a primary school graduate and illiterate are 0.41 and 0.27 respectively. Similarly, a university graduate father's wife is a university graduate with a probability of 0.08, whereas the probabilities that she is a primary school graduate and secondary school graduate are 0.34 and 0.20 respectively.

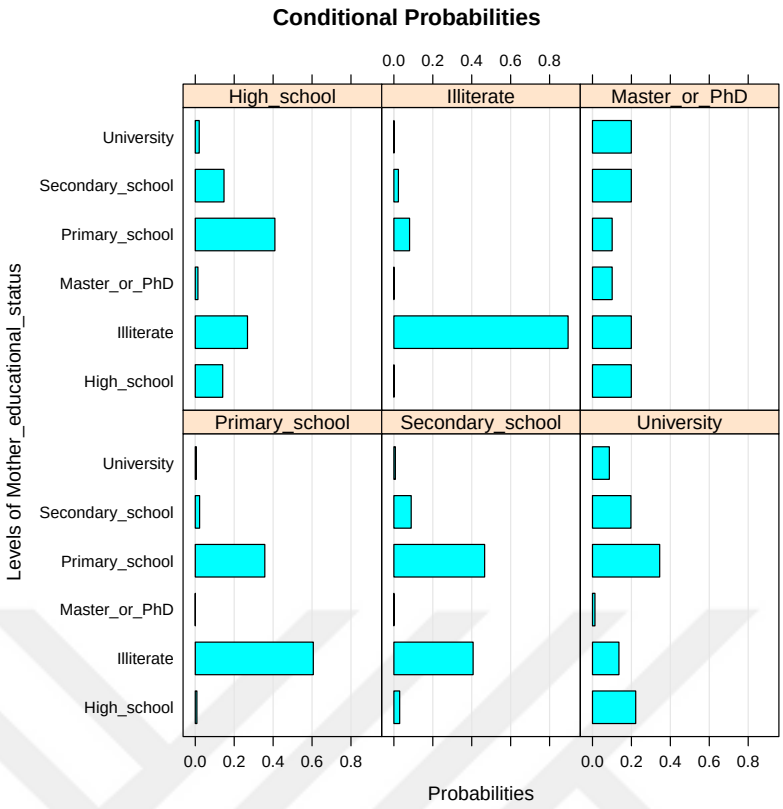


Figure 6.9 Bar chart of conditional probabilities of *Mother_educational_status* given *Father_educational_status*

Figure 6.10 presents the conditional probabilities of *Number_of_siblings* given *Mother_educational_status*. It is noted that having 5 or more siblings is the most common among students. The only exception is observed in the students whose mothers are high school graduates, in which group 3 siblings has the highest proportion (0.41). In addition, having 1 siblings has the lowest percentage in all of the students, regardless of their mothers' educational status.

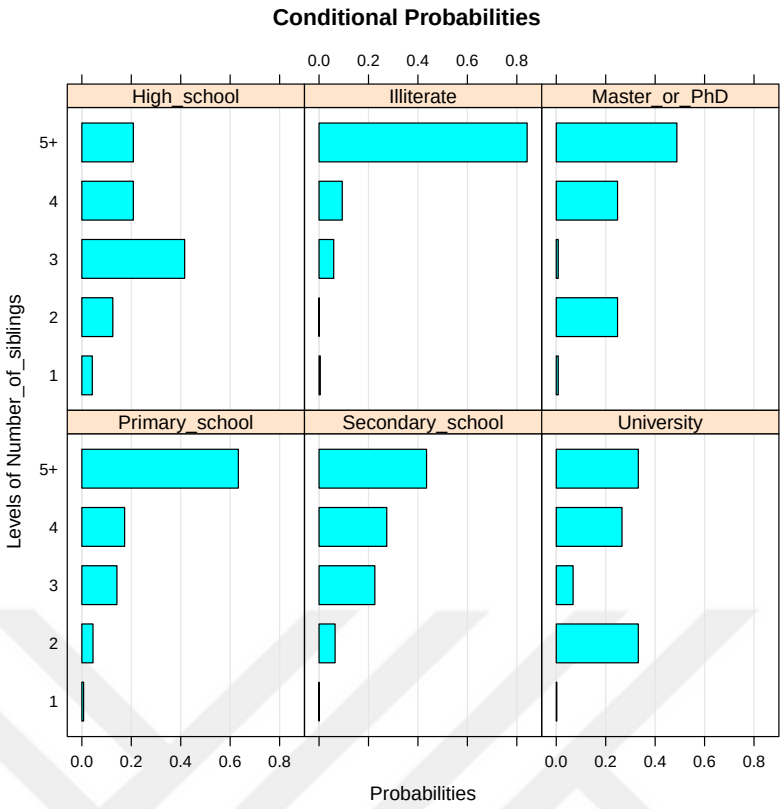


Figure 6.10 Bar chart of conditional probabilities of *Number_of_siblings* given *Mother_educational_status*

Figure 6.11 shows the conditional probabilities of *Family_structure* given *Number_of_siblings*. According to it, nuclear family is the most widespread type of family structure in students with 1,2,3, and 4 siblings. However, when a student is known to have 5 or more siblings, the type of family structure which a student belongs to with the greatest probability (0.57) is the extended family.

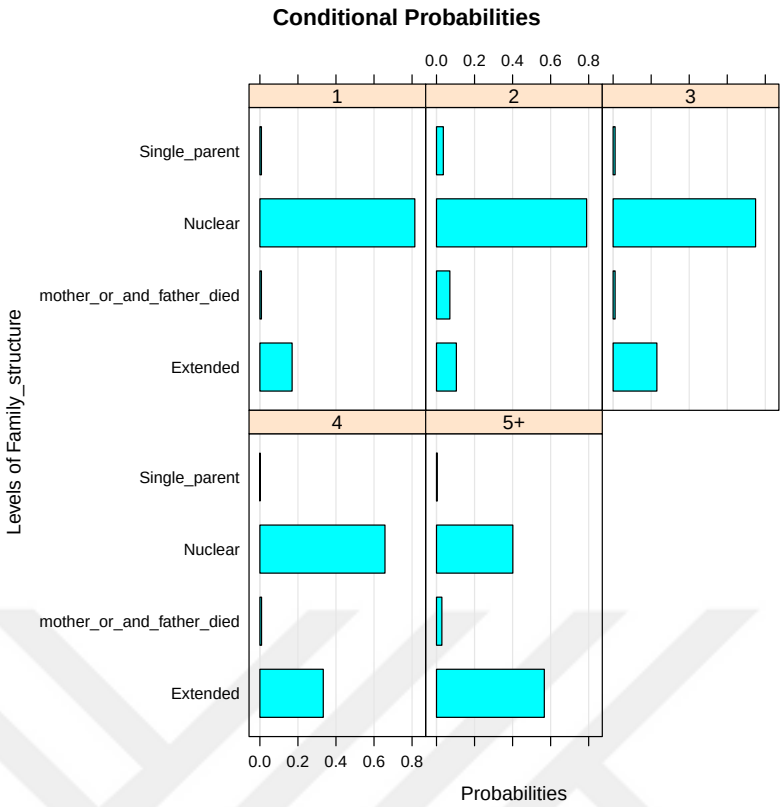


Figure 6.11 Bar chart of conditional probabilities of *Family_structure* given *Number_of_siblings*

According to Figure 6.1, there is a converging connection from *Gender* and *Student_job* to *Smoking_status*, meaning whether or not a student is a smoker is directly dependent on both his/her gender and whether or not he/she has a job. Therefore, both *Gender* and *Smoking_status* have to be known to predict the conditional probabilities of *Smoking_status*.

Figure 6.12 represents the conditional probabilities of *Smoking_status* given *Student_job* and *Gender*. While the probability that a female student who has a job is a smoker is 0.50, the probability that a female student who has no job is a smoker is only 0.11. On the other hand, the probability a male student with a job is a smoker is 0.38, whereas the probability a male student with no job is a smoker is 0.41. These numbers indicate that the female students with jobs are more likely to be a smoker, but that is not the case for the male students.

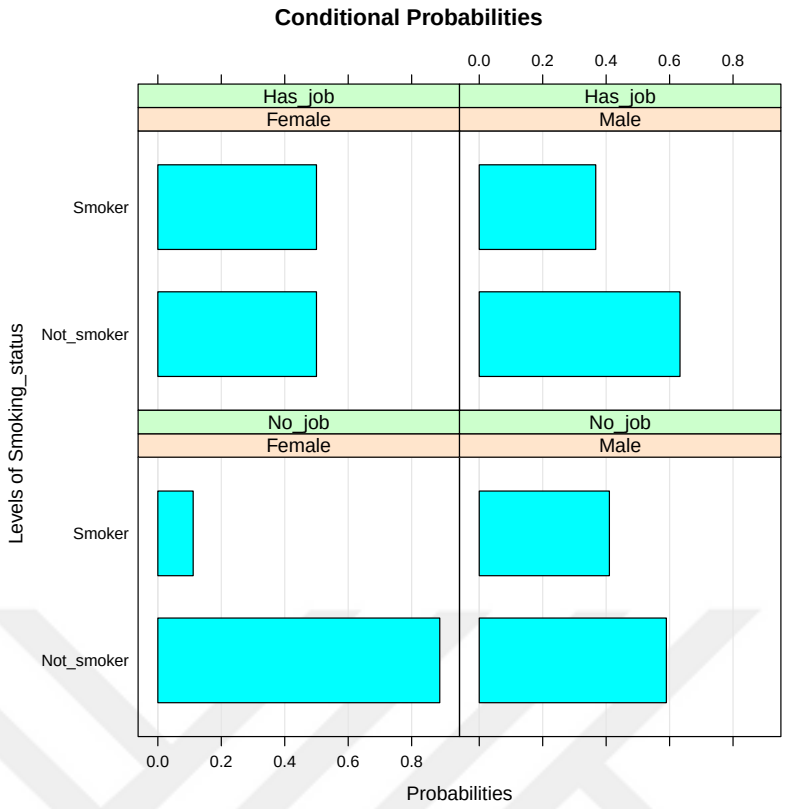


Figure 6.12 Bar chart of conditional probabilities of *Smoking_status* given *Student_job* and *Gender*

Figure 6.13 is the bar chart of conditional probabilities of *Alcohol_use* given *Smoking_status*. It shows that the probability that a student consumes alcohol is lower regardless of whether or not the student is a smoker. It also shows that *Smoking_status* has significantly positive influence on *Alcohol_use*. To illustrate, when a student is known to be a smoker, he/she has a 24% chance of consuming alcohol, while when the student is not a smoker, the probability that he/she is a alcohol user is only 2%. This indicates that smoking triggers alcohol consumption.

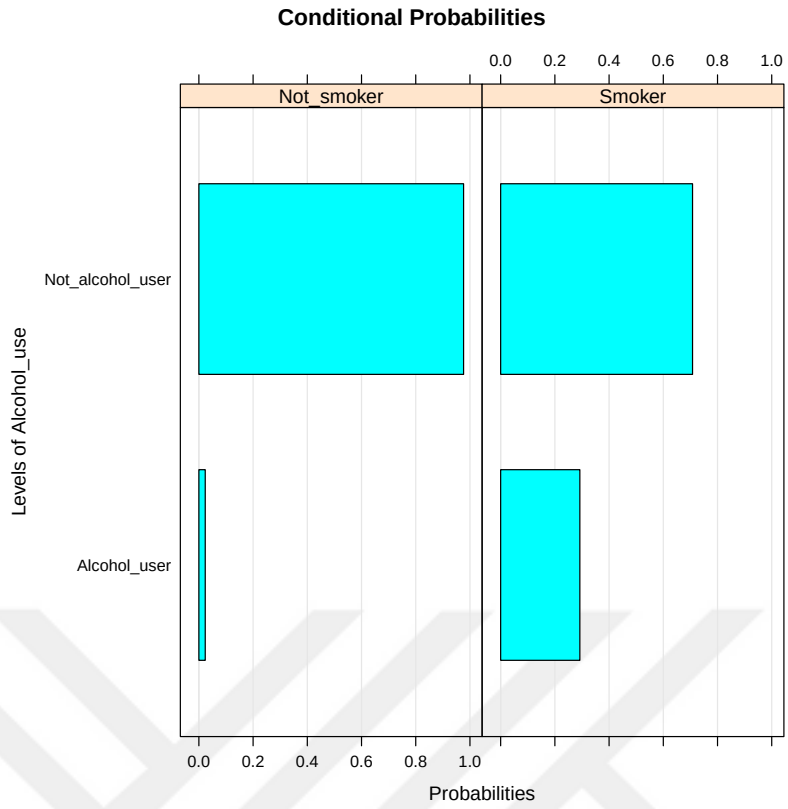


Figure 6.13 Bar chart of conditional probabilities of *Alcohol_use* given *Smoking_status*

Figure 6.1 indicates that there is a converging connection from *Alcohol_use* and *Student_job* to *Relationship_status*, indicating that whether or not a student has a relationship is directly dependent on both whether or not he/she has a job and whether or not he/she consumes alcohol. Figure 6.14 shows the conditional probabilities of *Relationship_status* given *Student_job* and *Alcohol_use*. It shows that the probability that a student with a job who consumes alcohol is in a relationship is 0.65, while a student with no job who does not consume alcohol is in relationship is 0.29, which shows that both having a job and being a alcohol user increase the chance of being in a relationship for a student.

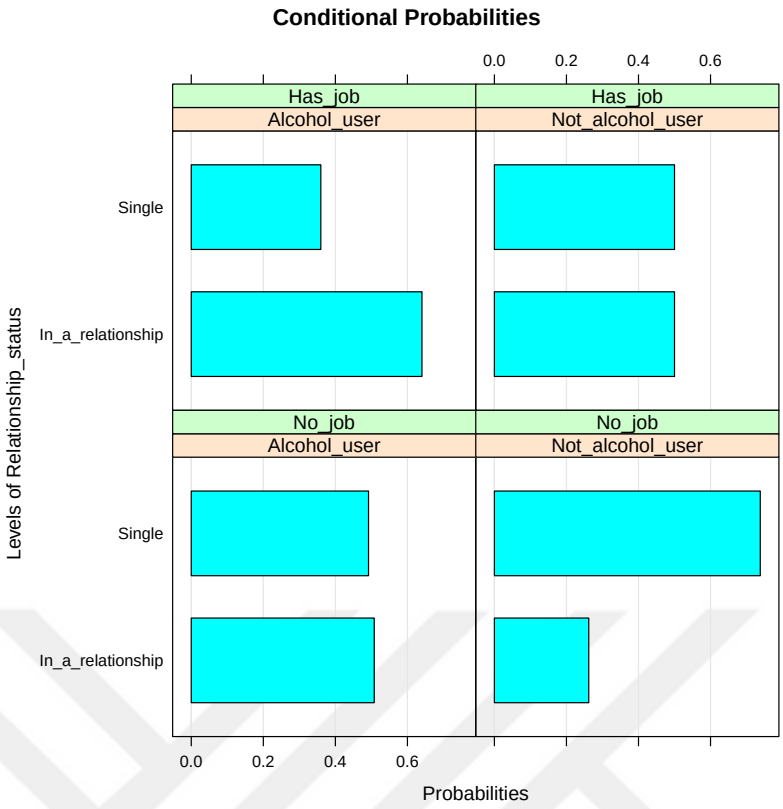


Figure 6.14 Bar chart of conditional probabilities of *Relationship_status* given *Alcohol_use* and *Student_job*

Figure 6.15 provides the conditional probabilities of *Social_activity* given *Gender*. It shows that the students are more likely not to have any social activity, irrelevant of their gender. However, male students have higher chance (41%) of having social activities than female students (28%).

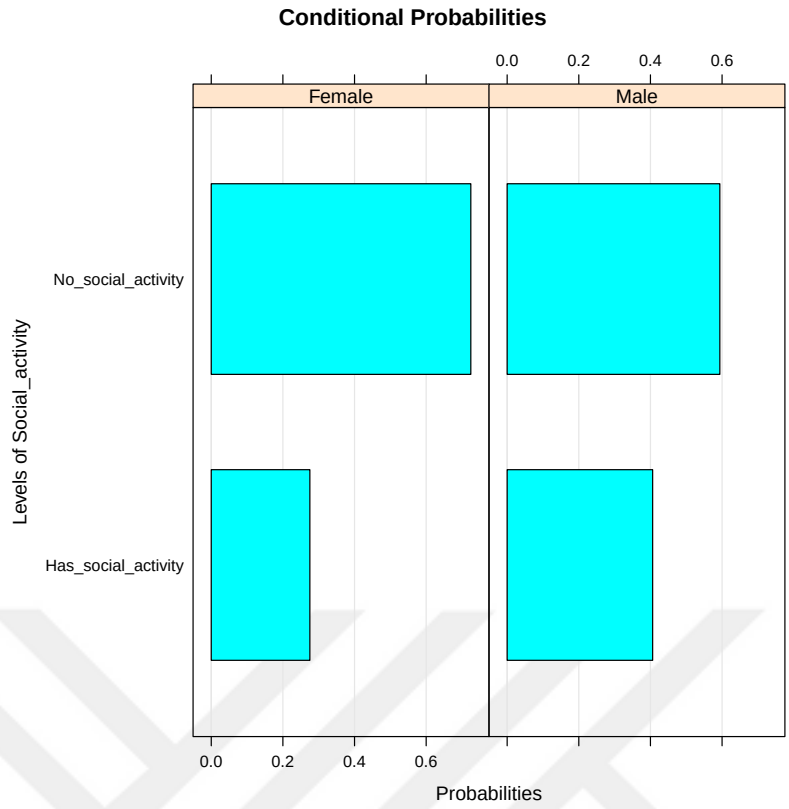


Figure 6.15 Bar chart of conditional probabilities of *Social_activity* given *Gender*

Figure 6.16 gives the conditional probabilities of *Type_of_accommodation* given *Gender*. It shows that while vast majority of females (0.65) live in governmental dormitories, the same proportion of male students live in flats. In addition, private dormitories are the least popular among both genders.

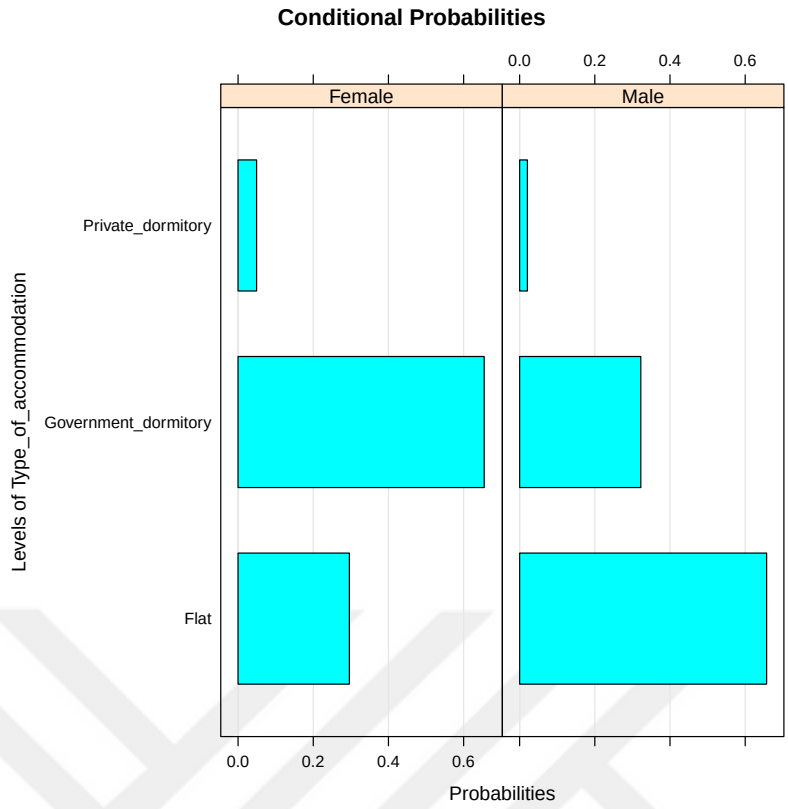


Figure 6.16 Bar chart of conditional probabilities of *Type_of_accommodation* given *Gender*

Figure 6.17 shows the conditional probabilities of *School_type* given *Gender*. The highest proportions of girls study at health school and arts and sciences school (0.30 and 0.29 respectively), while the highest proportion of males (0.36) study at engineering school which is preferred by only 8% of female students, making engineering school the least popular school among females.

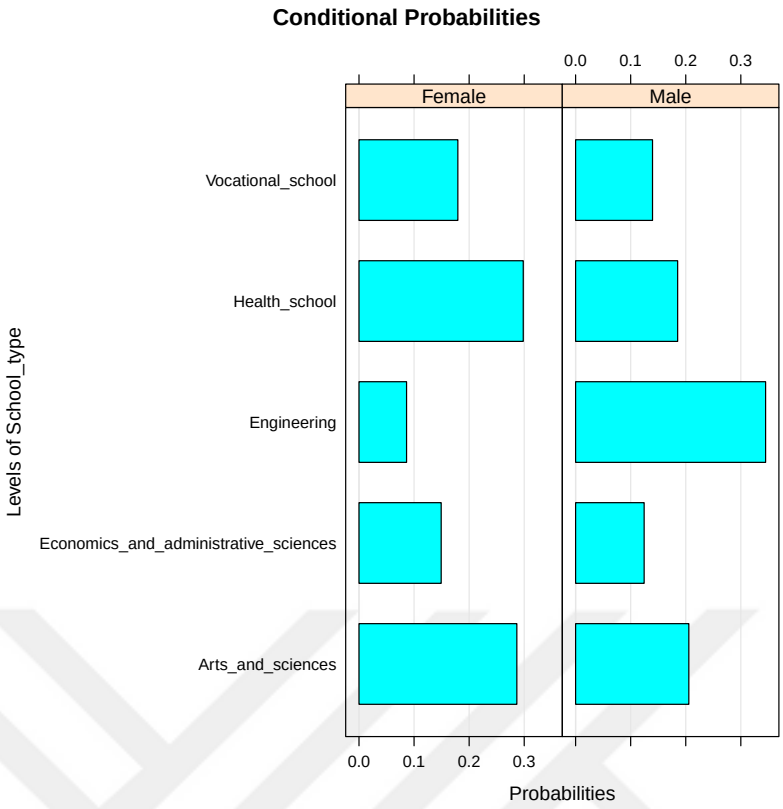


Figure 6.17 Bar chart of conditional probabilities of *School_type* given *Gender*

So far, the relationship between demographic and socio-economic variables have been investigated. Now, the focus will be on the relationship among the psychiatric variables (given in yellow nodes in Figure 6.1) and demographic and socio-economic variables (given in grey and yellow nodes in Figure 6.1, respectively). According to Figure 6.1, there is only one variable which has a direct influence on one of the psychiatric variables, that is, *Gender*. The other psychiatric variables, *Hopelessness* and *Self_esteem*, are conditionally independent of all of the demographic and socio-economic variables given *Depression*.

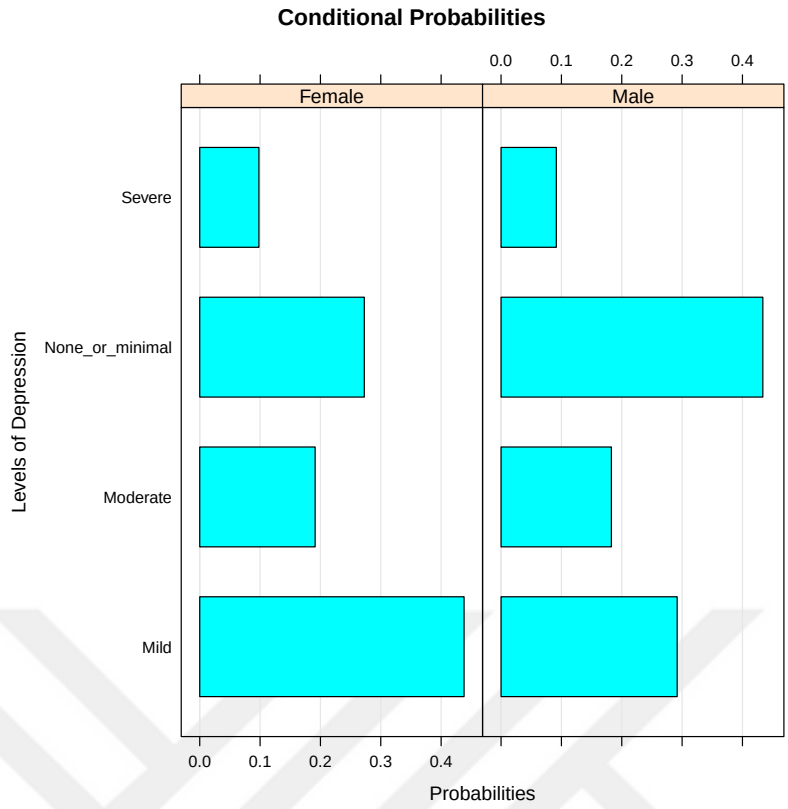


Figure 6.18 Bar chart of conditional probabilities of *Depression* given *Gender*

Figure 6.18 represents the conditional probabilities of *Depression* given *Gender*. It is notable that male and female students have approximately equal probabilities of having severe (10%) and moderate depression (19%). However, male students are much more likely to have none or minimal depression (44%) compared to their female counterparts (27%). While the probability of having mild depression is 29% for male students, it is 45% for female students. Hence, according to Figure 6.1, it can be said that female students are more likely to have higher level of depression compared to male students.

From Figure 6.1, it is seen that the level of depression affects both the levels of self-esteem and hopelessness in the students as there is a divergent connection from *Depression* to *Self_esteem* and *Hopelessness*.

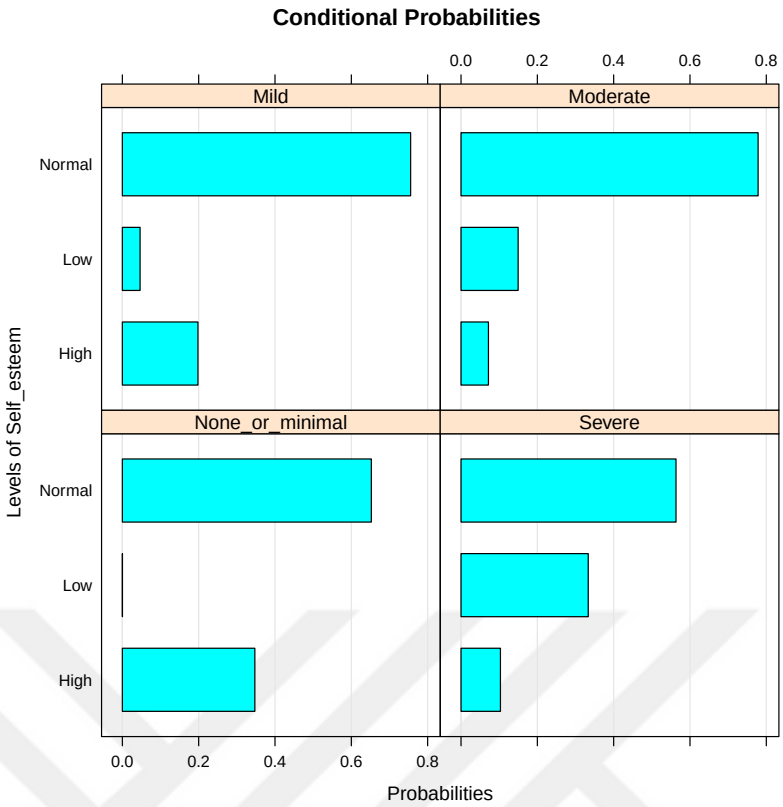


Figure 6.19 Bar chart of conditional probabilities of *Self_esteem* given *Depression*

Figure 6.19 provides the conditional probabilities of *Self_esteem* given *Depression*. Given a student has none or minimal, mild, moderate and severe depressions, the probabilities that he/she has low self-esteem are 0.01, 0.08, 0.15 and 0.33, respectively, while the probabilities that he/she has high self-esteem are 0.34, 0.22, 0.08 and 0.11, respectively. So, based on these probabilities, it can be interpreted that the level of depression has negative impact on the level of self-esteem.

Figure 6.20 is the bar chart of the conditional probabilities of *Hopelessness* given *Depression*. In contrast with *Self_esteem*, generally speaking, there is a positive relationship between *Hopelessness* and *Depression*, meaning that higher level of depression causes higher level of hopelessness. To illustrate, the probabilities of having none or minimal hopelessness, mild hopelessness, moderate hopelessness and severe hopeless are respectively 0.61, 0.29, 0.09 and 0.01 given a student has none or minimal depression, 0.41, 0.32, 0.20, and 0.07 given he/she has mild depression, 0.18, 0.35, 0.37 and 0.10 when he/she has moderate depression, 0.07, 0.17, 0.38 and 0.38 when he/she is known to have severe depression.

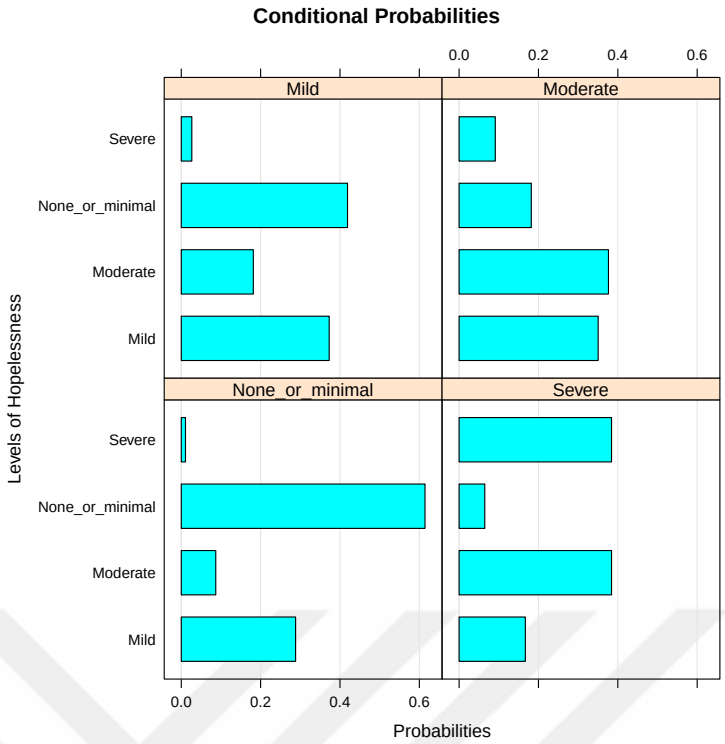


Figure 6.20 Bar chart of conditional probabilities of *Hopelessness* given *Depression*

6.2.2 Results of $\mathcal{B}_2$

In this section, a BN is built just like the one built in the previous section. The only difference is the method of discretization of the psychiatric variables. In this model, which is called $\mathcal{B}_2$, the psychiatric data discretized by the IPD method (see Section 4.4.1) are used.

Once again, an empty DAG is specified as the prior DAG and all the arrows towards *Age* and *Gender* are blocked (for `R` code, see Appendix **??**).

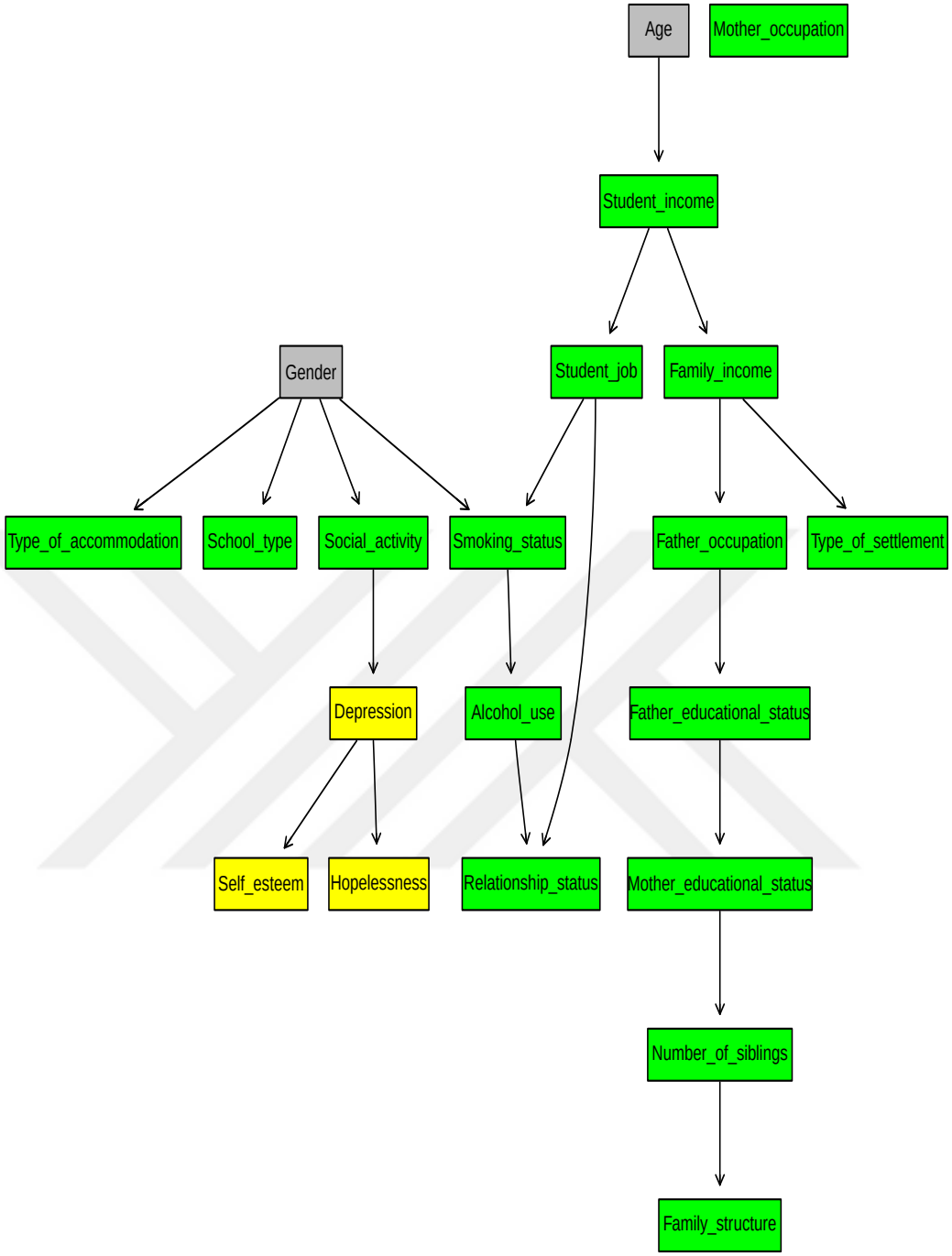


Figure 6.21 DAG of $\mathcal{B}_2$.

Since the interpretation of all of the relationships between demographic and socio-economic variables have already been given in the previous section, the focus will be on the relationships among the psychiatric variables rather than the demographic and socio-economic variables.

Figure 6.21 is the DAG of $\mathcal{B}_2$. It is clearly noted that, as expected, there is only one

difference between the DAGs of $\mathcal{B}_1$ and $\mathcal{B}_2$, that is, *Depression* is directly dependent on *Social_activity*, whereas it is directly dependent on *Gender* in $\mathcal{B}_1$.

Figure 6.22 is the bar chart of the conditional probabilities of *Depression* given *Social_activity*. It shows that the presence of social activity lowers the probability of having high level of depression. Given that a student participates in social activities, the probability that he/she has low level of depression is 56%, whereas this probability is 41% for the one who does not. In addition, 17% of students who join social activities have high level depression, whereas of the students who do not take part in social activities, 28% of them have high level of depression.

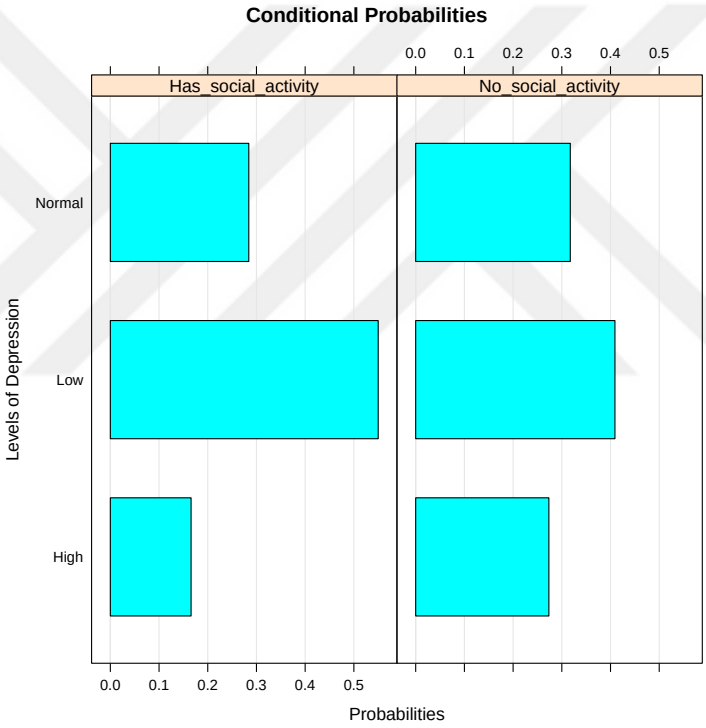


Figure 6.22 Bar chart of conditional probabilities of *Depression* given *Social_activity*.

Figure 6.23 is the bar chart of the conditional probabilities of *Hopelessness* given *Depression*. When a student has low or normal level of depression, the probability that he/she has low level of hopelessness is significantly higher (0.83 and 0.67 respectively) than the remaining probabilities. In contrast, when he/she is known to have high level of depression, the probability his/her hopelessness level is low (0.27) is lower than that of having high level of hopelessness (0.35). These results indicate that the level of depression positively correlates with the level of hopelessness.

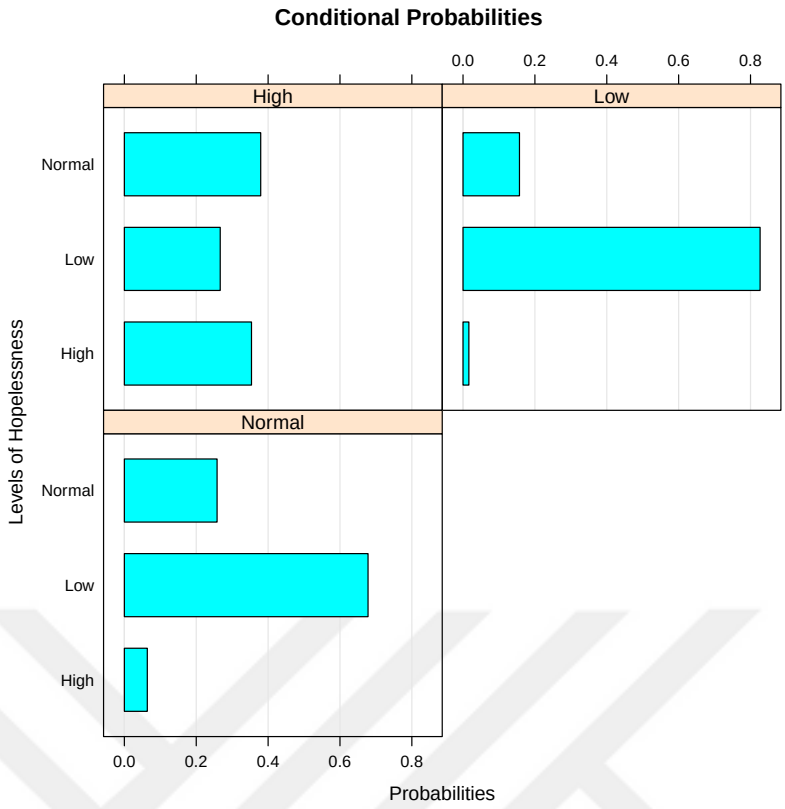


Figure 6.23 Bar chart of the conditional probabilities of *Hopelessness* given *Depression.*

Figure 6.24 is the graph of the conditional probabilities of *Self_esteem* given *Depression.* It is noted that for a randomly-chosen student, the greatest chance is that he/she has high level of self-esteem as the proportions of students with high self-esteem is higher than the other proportions for each depression level. However, as the level of depression rises, the probability that the student has low self-esteem rises, but the probabilities that he/she has normal and high level of self-esteem decreases significantly. To show that, the probabilities that a student has low and high levels of self-esteem are respectively 2% and 0.85% given he/she has low level of depression, 6% and 70% given he/she has normal level of depression and 33% and 41% given he/she has high level of depression. Hence, it is concluded that a rise in the level of depression leads to a decrease in the probability of having high level of self-esteem.

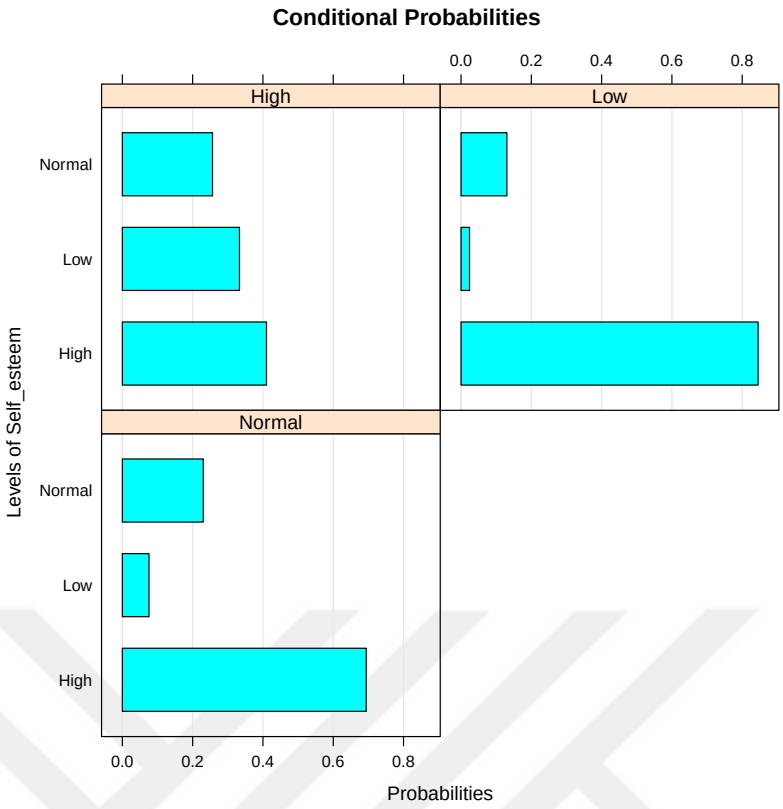


Figure 6.24 Bar chart of the conditional probabilities of *Self_esteem* given *Depression*.

6.3 Comparison of $\mathcal{B}_1$ and $\mathcal{B}_2$

To compare the relationships in $\mathcal{B}_1$ and $\mathcal{B}_2$, as predicted, they are not significantly different. The only difference is that in $\mathcal{B}_1$ *Depression* is directly influenced by *Gender*, while it is directly influenced by *Social_activity* in $\mathcal{B}_2$. However, in each model, *Hopelessness* and *Self_esteem* are directly dependent on *Depression* and conditionally independent of all the demographic and socio-economic variables. In addition, in both models, *Depression* has a positive effect on *Hopelessness*, but a negative effect on *Self_esteem*.

Table 6.1 Goodness-of-fit statistics of $\mathcal{B}_1$ and $\mathcal{B}_2$

		$\mathcal{B}_1$	$\mathcal{B}_2$
Specificity			
	$Depression$	43.62%	45.69%
	$Hopelessness$	47.14%	67.56%
	$Self_esteem$	70.6%	72.35%
BIC score		-15214.8	-14748.93

Figure 6.1 provides specificities of the target variables and BIC scores of each model.

To compare the performances of $\mathcal{B}_1$ and $\mathcal{B}_2$, firstly, $\mathcal{B}_2$'s network score is greater than that of $\mathcal{B}_1$, meaning $\mathcal{B}_2$ is more likely to produce better predictions than $\mathcal{B}_1$. This finding is supported by the specificities of *Depression*, *Hopelessness* and *Self_esteem*. As shown in Figure 6.1, $\mathcal{B}_2$ is more successful in predicting *Depression*, *Hopelessness* and *Self_esteem*.

6.4 Discussion and Conclusion

In this chapter, the data sets were analyzed in `R` and the results were presented via figures and graphs. Firstly, the reliabilities of the psychiatric scales were tested using the Cronbach's alpha internal consistency coefficients and found satisfying. Then, two BNs called $\mathcal{B}_1$ and $\mathcal{B}_2$ were learned from the data and constructed as there were two discretization methods. The same structure learning and parameter learning algorithms for the two BNs were used: hill-climbing search with tabu list to learn the DAGs and Dirichlet distribution to learn the parameters.

The BIC score of $\mathcal{B}_2$ was greater than that of $\mathcal{B}_1$, meaning that $\mathcal{B}_2$ outperforms $\mathcal{B}_1$. Another finding supporting this argument is that the specificities of the target variables of $\mathcal{B}_2$ are greater than those of $\mathcal{B}_1$.

According to $\mathcal{B}_1$, the gender of the students directly influences the probability of having a particular state of depression and female students are more likely to have a higher level depression than their male counterparts. This finding is in line with similar research papers studying the relationship between gender and depression [67], [68], [69], [70].

According to $\mathcal{B}_2$, the presence or absence of social activities has a negative impact the level of depression, which means that a student having social activities is less likely to have higher level of depression compared to those who do not have any social activities. This conclusion is supported by other studies about the relationship between social activity and depression in the literature [71], [72].

In each model, the level of depression has a positive influence on the level of hopelessness, which means that as the severity of depression rises, so does the risk of having higher level of hopelessness, which was also detected in [73], [74]. Another

important finding obtained in each model is that, in contrast with hopelessness, the level of depression has a negative effect on the level of self-esteem. In other words, higher level of depression leads to lower level of self-esteem. This result is consistent with similar studies [71], [75], [76].

CHAPTER 7

CONCLUSION

This thesis has presented data discretization and Bayesian networks thorough an application in R. This chapter concludes the study by highlighting what has been done so far.

In Chapter 1, the motivation to write this thesis was presented, explaining why the BNs are used in many domains and a brief history of the BNs. Also, the objectives of this thesis were detailed. Finally, the structure of the thesis was presented.

In Chapter 2, the data used in this study were described. The data of 823 university students consist of 3 psychiatric, 2 demographic and 16 socio-economic variables. The way the data were collected and the characteristics of each variable were detailed in this chapter.

In Chapter 3, exploratory data analysis was conducted in order to see the basic properties of the variables. Being continuous; histograms, box plots, means, standard deviations, first quartiles, third quartiles and medians were employed for the psychiatric variables. It was observed that none of the psychiatric variables are normally distributed. The BDI and BHS are skewed to the right, while the RSES scores are skewed to the left. In addition, the BDI and BHS scores have outliers above their third quartile, whereas the outliers of the RSES scores are below their first quartile. For the demographic and socio-economic variables, frequencies and percentages were used as they were discrete. There were some remarkable differences between the percentages: the notable percentage of the mothers are unemployed, the proportion of students consuming alcohol is remarkably low and the percentages of students having 1 and 2 siblings are considerably low.

In Chapter 4, the motivation of data discretization was presented. Next, the theory of data discretization and some discretization methods used in literature were detailed. After that, the two discretization approaches used in this study were presented. Next the psychiatric variables were discretized by the IPD method so that 3 states per variable were obtained. Finally, the psychiatric variables were discretized by the domain knowledge in literature, gaining 4 states for the BDI and BHS scores and 3 states for the RSES score.

In Chapter 5, the BNs were introduced and reviewed. Firstly, the Bayes' Theorem, from which the BNs emerged, was introduced. Next, the details about the terms used in the BNs were provided and a classic BN example was employed in order to show how a BN works. Afterwards, the theory of BN learning was explained. Then, the structure learning from data and then the parameter learning were described.

In Chapter 6, the analyses were conducted using related packages in `R`. Firstly, Cronbach's alpha internal consistency coefficients for the psychiatric scales were tested and found satisfying. As there were two discretization methods, two different BNs were obtained using the same structure and parameter learning approaches. The hill-climbing search with tabu list for structure learning and Dirichlet distribution for parameter learning were employed. Then, the results were presented via plots and figures and queries were made. The two BNs were compared in terms of the BIC scores and it was found $\mathcal{B}_2$ had higher BIC score than $\mathcal{B}_1$, meaning that $\mathcal{B}_2$ is likely to predict better than $\mathcal{B}_1$. Another evidence to support this argument is that $\mathcal{B}_2$ has higher specificities for the target variables. To summarize the results, in $\mathcal{B}_1$ the gender of the students directly influences their level of depression, while in $\mathcal{B}_2$, whether or not students have social activities has a direct impact on their level of depression . In both models the level of depression has a negative influence on self-esteem, but a positive impact on hopelessness.

REFERENCES

[1] Olumide, K., (2010). A Probabilistic and Graphical Analysis of Evidence in O.J. Simpson's Murder Case Using Bayesian Networks, Ph.D. Thesis, Florida State University College School of Arts and Sciences, Tallahassee.

[2] Yet, B., (2013). Bayesian Networks for Evidence Based Clinical Decision Support, Ph.D. Thesis, Queen Mary, University of London, London.

[3] Fenton, N. E. and Neil, M. (2014). "Decision Support Software for Probabilistic Risk Assessment Using Bayesian Networks", IEEE Software.

[4] Lauritzen, S. L. and Spiegelhalter, D. J. (1988). "Local Computations with Probabilities on Graphical Structures and Their Application to Expert Systems", Journal of the Royal Statistical Society, Series B (Methodological): 157-224.

[5] Pearl, J. (1986). "Fusion, Propagation, and Structuring in Belief Networks", Artificial Intelligence, 29(3): 241-288.

[6] Sesen, M. B., Nicholson, A. E., Banares-Alcantara, R., Kadir, T. and Brady, M. (2013). "Bayesian Networks for Clinical Decision Support in Lung Cancer Care", PLOS One, 8(12).

[7] Cruz-Ramírez, N., Acosta-Mesa, H. G., Carrillo-Calvet, H., Nava-Fernández, L. A. and Barrientos-Martínez, R. E. (2007). "Diagnosis of Breast Cancer Using Bayesian Networks: A Case Study", Computers in Biology and Medicine, 37(11): 1553-1564.

[8] Zweig, G. (2003). "Bayesian Network Structures and Inference Techniques for Automatic Speech Recognition", Computer Speech and Language, 17(2-3): 173-193.

[9] Marcot, B. G., Steventon, J. D., Sutherland, G. D. and McCann, R. K. (2006). "Guidelines for Developing and Updating Bayesian Belief Networks Applied to Ecological Modeling and Conservation", Canadian Journal of Forest Research, 36(12): 3063-3074.

[10] Barton, D. N., Saloranta, T., Moe, S. J., Eggestad, H. O. and Kuikka, S. (2008). "Bayesian Belief Networks As a Meta-modelling Tool in Integrated River Basin Management—Pros and Cons in Evaluating Nutrient Abatement Decisions under Uncertainty in a Norwegian River Basin", Ecological Economics, 66(1): 91-104.

[11] Aguilera, P. A., Fernández, A., Fernández, R., Rumí, R. and Salmerón, A. (2011). "Bayesian Networks in Environmental Modeling", Environmental Modeling and Software, 26(12): 1376-1388.

[12] Pollino, C. A., Woodberry, O., Nicholson, A., Korb, K. and Hart, B. T. (2007). "Parameterization and Evaluation of a Bayesian Network for Use in an Ecological Risk Assessment", Environmental Modelling and Software, 22(8): 1140-1152.

[13] Cene, E., & Karaman, F. (2015). "Analysing Organic Food Buyers' Perceptions with Bayesian Networks: A Case Study in Turkey", Journal of Applied Statistics, 42(7): 1572-1590.

[14] Renzi, M. F., Vicard, P., Guglielmetti, R. and Musella, F. (2009). "Probabilistic Expert Systems for Managing Information to Improve Services", The TQM Journal, 21(4): 429-442.

[15] Salini, S., and Kenett, R. S. (2009). "Bayesian Networks of Customer Satisfaction Survey Data", Journal of Applied Statistics, 36(11): 1177-1189.

[16] Jansen, R., Yu, H., Greenbaum, D., Kluger, Y., Krogan, N. J., Chung, S., Emili, A., Snyder, M., Greenblatt, J. F. and Gerstein, M. (2003). "A Bayesian Networks Approach for Predicting Protein-protein Interactions from Genomic Data", Science, 302(5644): 449-453.

[17] Yu, J., Smith, V. A., Wang, P. P., Hartemink, A. J. and Jarvis, E. D. (2004). "Advances to Bayesian Network Inference for Generating Causal Networks from Observational Biological Data", Bioinformatics, 20(18): 3594-3603.

[18] Beck, A. (1961). "An Inventory for Measuring Depression", Archives of General Psychiatry, 12: 63-70.

[19] Beck, A. T., Weissman, A., Lester, D. and Trexler, L. (1974). "The Measurement of Pessimism: The Hopelessness Scale", Journal of Consulting and Clinical Psychology, 42(6): 861-865.

[20] Rosenberg, M., (1965). Society and the Adolescent Self-image, Princeton University Press, Princeton.

[21] Fernandez, A. M. M. (2012). New Models and Algorithms for Semi-Naive Bayesian Classification Focused on the AODE Paradigm, Ph.D. Thesis, University of Castilla-la Mancha Computing Systems Department, Ciudad Real.

[22] Chakrabarti, S., Cox, E., Frank, E., Güting, R. H., Han, J., Jiang, X., Kamber, M., Lightstone, S., Nadeau, T., Neapolitan, R. E., Pyle, D., Refaat, M., Schneider, M., Teorey, T. and Witten, I. (2008). Data Mining: Know it All., Morgan Kaufmann, Burlington.

[23] Liu, H., Hussain, F., Tam, C. L. and Dash, M. (2002). Discretization: An Enabling Technique, Springer, Singapore.

[24] Chmielewski, M. R. and Grzymala-Busse, J. W. (1996). "Global Discretization of Continuous Attributes As Preprocessing for Machine Learning", International Journal of Approximate Reasoning, 15(4): 319-331.

[25] Cerquides, J. and De Màntaras, R. L. (1997). "Proposal and Empirical Comparison of a Parallelizable Distance-based Discretization Method", The Third International Conference on Knowledge Discovery and Data Mining, 14-17 August, 1997, California.

[26] Maass, W. (1994). Efficient Agnostic Pac-learning with Simple Hypothesis, The Seventh Annual Conference on Computational Learning Theory, 12-15 July 1994, New Jersey.

[27] Holte, R. C. (1993). "Very Simple Classification Rules Perform Well on Most Commonly Used Datasets", Machine learning, 11(1): 63-90.

[28] Dougherty, J., Kohavi, R. and Sahami, M. (1995). Supervised and Unsupervised Discretization of Continuous Features, The Twelfth International Conference on Machine Learning, 9–12 July 1995, California.

[29] Hartemink, A.J. (2001). Computational Methods for the Validation and Discovery of Genetic Regulatory Networks, Ph.D. Thesis, Massachusetts Institute of Technology Electrical Engineering and Computer Science, Boston.

[30] Altman, D. G. and Lyman, G. H. (1998). "Methodological Challenges in the Evaluation of Prognostic Factors in Breast Cancer", Breast Cancer Research and Treatment, 52 (1-3): 289-303.

[31] Royston, P., Altman, D. G. and Sauerbrei, W. (2006). "Dichotomizing Continuous Predictors in Multiple Regression: A Bad Idea", Statistics in Medicine, 25(1): 127-141.

[32] Barrio, I., Arostegui, I., Rodríguez-Álvarez, M. X. and Quintana, J. M. (2017). "A New Approach to Categorising Continuous Variables in Prediction Models: Proposal and Validation", Statistical Methods in Medical Research, 26(6): 2586-2602.

[33] Beck, A. T. and Steer, R. A. (1988). Manual for the Beck Hopelessness Scale, Psychological Corporation, San Antonio.

[34] Abraham, D. (2010). A study on the Self-esteem and Social Relations of Adolescents with Learning Disability, Master Thesis, Christ University, Bangalore.

[35] Korb, K. B. and Nicholson, A. E. (2011). Bayesian Artificial Intelligence. CRC Press, London.

[36] Nielsen, T. D. and Jensen, F. V. (2007). Bayesian Networks and Decision Graphs, Springer, New York.

[37] Inman, D. (2008). The Use of Bayesian Networks to Facilitate Implementation of Water Demand Management Strategies, Ph.D. Thesis, Cranfield University School of Applied Sciences, Bedford.

[38] Freund, J. E. and Miller, I. (2014). John E. Freund's Mathematical Statistics: With Applications, Eighth Edition, Pearson, Essex.

[39] Margaritis, D. (2003). Learning Bayesian Network Model Structure from Data, Ph.D. Thesis, Carnegie Mellon University School of Computer Science, Pittsburgh.

[40] Fuster-Parra, P., Tauler, P., Bennasar-Veny, M., Ligeza, A., López-González, A. A. and Aguiló, A. (2015). "Bayesian Network Modeling: A Case Study of an Epidemiologic System Analysis of Cardiovascular Risk", Computer Methods and Programs in Biomedicine, 126: 128-142.

[41] Neapolitan, R. E. (2004). Learning Bayesian Networks, Pearson Prentice Hall, New Jersey.

[42] Russell, S. J. and Norvig, P. (1995). Artificial Intelligence: A Modern Approach, Prentice Hall Series, New Jersey.

[43] Oteniya, L. (2008). Bayesian Belief Networks for Dementia Diagnosis and Other Applications: A comparison of Hand-crafting and Construction Using a Novel Data Driven Technique, Ph.D. Thesis, University of Stirling Department of Computing Science, Stirling.

[44] Beaumont, M. A. and Rannala, B. (2004). "The Bayesian Revolution in Genetics", Nature Reviews Genetics, 5(4): 251-261.

[45] Pearl, J. (1988). Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann Publishers, San Francisco.

[46] Dawid, A. P. (1979). "Conditional Independence in Statistical Theory", Journal of the Royal Statistical Society: Series B (Methodological), 1-31.

[47] Murphy, K. P. (2001). "An Introduction to Graphical Models", Technical Report, University of British Columbia.

[48] Scutari, M. (2010). "Learning Bayesian Networks with the bnlearn R Package", Journal of Statistical Software, 35(3).

[49] Koski, T. and Noble, M. N. (2009). Bayesian Networks, Wiley, Chichester.

[50] Castillo, E., Gutierrez, J. M. and Hadi, A. S. (1997). Expert Systems and Probabilistic Network Models. Springer, New York.

[51] Heckerman, D., Geiger, D. and Chickering, D. M. (1995). "Learning Bayesian Networks: The Combination of Knowledge and Statistical Data", Machine Learning, 20(3): 197-243.

[52] Geiger, D. and Heckerman, D. (1994). "Learning Gaussian Networks", The Tenth International Conference on Uncertainty in Artificial Intelligence, Seattle.

[53] Scutari, M. and Denis, J. B. (2014). Bayesian Networks: with Examples in R, Chapman and Hall, Boca Raton.

[54] Koski, T. J. and Noble, J. (2012). "A Review of Bayesian Networks and Structure Learning", Mathematica Applicanda, 40(1): 51-103.

[55] Teyssier, M. and Koller, D. (2012). "Ordering-based Search: A Simple and Effective Algorithm for Learning Bayesian Networks", The Twenty-first Conference on Uncertainty in Artificial Intelligence, Edinburgh.

[56] Mitchell, M., Holland, J. H. and Forrest, S. (1994). When Will a Genetic Algorithm Outperform Hill Climbing, Advances in Neural Information Processing Systems 6, Morgan Kaufman.

[57] Urcia, W. P. and Mauá, D. D. (2015). "Initialization Heuristics for Greedy Bayesian Network Structure Learning", The Third Symposium on Knowledge Discovery, Mining and Learning, 13-15 October 2015, Petrópolis.

[58] Taillard, É. (1991). "Robust Taboo Search for the Quadratic Assignment Problem", Parallel Computing, 17(4-5): 443-455.

[59] Pirim, H., Bayraktar, E. and Eksioglu, B. (2008). Tabu Search: A Comparative Study, Tabu Search. I-Tech Education and Publishing, Vienna.

[60] Hertz, A. and de Werra, D. (1987). "Using Tabu Search Techniques for Graph Coloring", Computing, 39(4): 345-351.

[61] Spiegelhalter, D. J. and Lauritzen, S. L. (1990). "Sequential Updating of Conditional Probabilities on Directed Graphical Structures", Networks, 20(5): 579-605.

[62] Cooper, G. F. and Herskovits, E. (1992). "A Bayesian Method for the Induction of Probabilistic Networks from Data", Machine Learning, 9(4): 309-347.

[63] Cronbach, L. J. (1951). "Coefficient Alpha and the Internal Structure of Tests", Psychometrika, 16(3): 297-334.

[64] Højsgaard, S. (2012). "Graphical Independence Networks with the gRain Package for R", Journal of Statistical Software, 46(10): 1-26.

[65] Gentry, J., Long, L., Gentleman, R., Falcon, S., Hahne, F., Sarkar, D. and Rgraphviz, K. H. (2009). "Provides Plotting Capabilities for R Graph Objects", R Package Version, 2(0).

[66] Sarkar, D. (2008). Lattice: Multivariate Data Visualization with R, Springer, New York.

[67] Bebbington, P. (1996). "The Origins of Sex Differences in Depressive Disorder: Bridging the Gap", International Review of Psychiatry, 8(4): 295-332.

[68] Nolen-Hoeksema, S. (1987). "Sex Differences in Unipolar Depression: Evidence and Theory", Psychological Bulletin, 101(2): 259.

[69] Nazroo, J. Y., Edwards, A. C. and Brown, G. W. (1998). "Gender Differences in the Prevalence of Depression: Artefact, Alternative Disorders, Biology or Roles?", Sociology of Health & Illness, 20(3): 312-330.

[70] Angeleri, F., Angeleri, V. A., Foschi, N., Giaquinto, S. and Nolfe, G. (1993). "The Influence of Depression, Social Activity, and Family Stress on Functional Outcome After Stroke", Stroke, 24(10): 1478-1483.

[71] Dumont, M. and Provost, M. A. (1999). "Resilience in Adolescents: Protective Role of Social Support, Coping Strategies, Self-esteem, and Social Activities on Experience of Stress and Depression", Journal of Youth and Adolescence, 28(3): 343-363.

[72] Bozo, Ö., Toksabay, N. E. and Kürüm, O. (2009). "Activities of Daily Living, Depression, and Social Support among Elderly Turkish People", The Journal of Psychology, 143(2): 193-206.

[73] Fanaj, N., Melonashi, E., Drevinja, F. and Haxhibeqiri, S. H. (2015). "Depression, Anxiety, Hopelessness and Self-esteem in a Clinical Sample of Adolescents in Kosovo", European Psychiatry, 30: 1189.

[74] Dori, G. A. and Overholser, J. C. (1999). "Depression, Hopelessness, and Self-Esteem: Accounting for Suicidality in Adolescent Psychiatric Inpatients", Suicide and Life-threatening Behavior, 29(4): 309-318.

[75] De Santis, J. P., Gonzalez-Guarda, R. M. and Vasquez, E. P. (2012). "Psychosocial and Cultural Correlates of Depression among Hispanic Men with HIV Infection: A Pilot Study", Journal of Psychiatric and Mental Health Nursing, 19(10): 860-869.

[76] Provencio-Vasquez, E., Mata, H. J., Tomaka, J. and De Santis, J. P. (2017). "Depression, Self-Esteem, and Childhood Abuse Among Hispanic Men Residing in the US–Mexico Border Region", Journal of the Association of Nurses in AIDS Care, 28(5): 795-806.

APPENDIX A

ROSENBERG SELF-ESTEEM SCALE

1. On the whole, I am satisfied with myself.
   ( )Strongly Agree ( )Agree ( )Disagree ( )Strongly Disagree
2. At times I think I am no good at all.
   ( )Strongly Agree ( )Agree ( )Disagree ( )Strongly Disagree
3. I feel that I have a number of good qualities.
   ( )Strongly Agree ( )Agree ( )Disagree ( )Strongly Disagree
4. I am able to do things as well as most other people.
   ( )Strongly Agree ( )Agree ( )Disagree ( )Strongly Disagree
5. I feel I do not have much to be proud of.
   ( )Strongly Agree ( )Agree ( )Disagree ( )Strongly Disagree
6. I certainly feel useless at times.
   ( )Strongly Agree ( )Agree ( )Disagree ( )Strongly Disagree
7. I feel that I'm a person of worth, at least on an equal plane with others.
   ( )Strongly Agree ( )Agree ( )Disagree ( )Strongly Disagree
8. I wish I could have more respect for myself.
   ( )Strongly Agree ( )Agree ( )Disagree ( )Strongly Disagree
9. All in all, I am inclined to feel that I am a failure.
   ( )Strongly Agree ( )Agree ( )Disagree ( )Strongly Disagree
10. I take a positive attitude toward myself.
    ( )Strongly Agree ( )Agree ( )Disagree ( )Strongly Disagree

APPENDIX B

BECK DEPRESSION INVENTORY

1 (0) I do not feel sad.

(1) I feel sad

(2) I am sad all the time and I can't snap out of it.

(3) I am so sad and unhappy that I can't stand it.

2 (0) I am not particularly discouraged about the future

(1) I feel discouraged about the future.

(2) I feel I have nothing to look forward to.

(3) I feel the future is hopeless and that things cannot improve.

3 (0) I do not feel like a failure.

(1) I feel I have failed more than the average person.

(2) As I look back on my life, all I can see is a lot of failures.

(3) I feel I am a complete failure as a person.

4 (0) I get as much satisfaction out of things as I used to.

(1) I don't enjoy things the way I used to.

(2) I don't get real satisfaction out of anything anymore.

(3) I am dissatisfied or bored with everything.

5 (0) I don't feel particularly guilty.

(1) I feel I may be punished.

(2) I expect to be punished.

(3) I feel I am being punished.

6 (0) I don't feel I am being punished.

(1) I feel sad

(2) I am sad all the time and I can't snap out of it.

(3) I am so sad and unhappy that I can't stand it.

7 (0) I don't feel disappointed in myself.

(1) I am disappointed in myself.

(2) I am disgusted with myself.

(3) I hate myself.

8 (0) I don't feel I am any worse than anybody else.

(1) I am critical of myself for my weaknesses or mistakes.

(2) I blame myself all the time for my faults.

(3) I blame myself for everything bad that happens.

9 (0) I don't have any thoughts of killing myself.

(1) I have thoughts of killing myself, but I would not carry them out.

(2) I would like to kill myself.

(3) I would kill myself if I had the chance.

10 (0) I don't cry any more than usual.

(1) I cry more now than I used to.

(2) I cry all the time now.

(3) I used to be able to cry, but now I can't cry even though I want to.

11 (0) I am no more irritated by things than I ever was.

(1) I am slightly more irritated now than usual.

(2) I am quite annoyed or irritated a good deal of the time.

(3) I feel irritated all the time.

12 (0) I have not lost interest in other people.

(1) I am less interested in other people than I used to be.

(2) I have lost most of my interest in other people.

(3) I have lost all of my interest in other people.

13 (0) I make decisions about as well as I ever could.

(1) I put off making decisions more than I used to.

(2) I have greater difficulty in making decisions more than I used to.

(3) I can't make decisions at all anymore.

14 (0) I don't feel that I look any worse than I used to.

(1) I am worried that I am looking old or unattractive.

(2) I feel there are permanent changes in my appearance that make me look unattractive.

(3) I believe that I look ugly.

15 (0) I can work about as well as before.

(1) It takes an extra effort to get started at doing something.

(2) I have to push myself very hard to do anything.

(3) I can't do any work at all.

16 (0) I can sleep as well as usual.

(1) I don't sleep as well as I used to.

(2) I wake up 1-2 hours earlier than usual and find it hard to get back to sleep.

(3) I wake up several hours earlier than I used to and cannot get back to sleep.

17 (0) I don't get more tired than usual.

(1) I get tired more easily than I used to.

(2) I get tired from doing almost anything.

(3) I am too tired to do anything.

18 (0) My appetite is no worse than usual.

(1) My appetite is not as good as it used to be.

(2) My appetite is much worse now.

(3) I have no appetite at all anymore.

19 (0) I haven't lost much weight, if any, lately.

(1) I have lost more than five pounds.

(2) I have lost more than ten pounds.

(3) I have lost more than fifteen pounds.

20 (0) I am no more worried about my health than usual.

(1) I am worried about physical problems like aches, pains, upset stomach, or constipation.

(2) I am very worried about physical problems and it's hard to think of much else.

(3) I am so worried about my physical problems that I cannot think of anything else.

21 (0) I have not noticed any recent change in my interest in sex.

(1) I am less interested in sex than I used to be.

(2) I have almost no interest in sex.

(3) I have lost interest in sex completely.

APPENDIX C

BECK HOPELESSNESS SCALE

1. I look forward to the future with hope and enthusiasm.
   ( )True ( )False

2. I might as well give up because there's nothing I can do to make things better for myself.
   ( )True ( )False

3. When things are going badly, I am helped by knowing that they can't stay that way for ever.
   ( )True ( )False

4. I can't imagine what my life would be like in ten years.
   ( )True ( )False

5. I have enough time to accomplish the things I most want to do.
   ( )True ( )False

6. In the future I expect to succeed in what concerns me most.
   ( )True ( )False

7. My future seems dark to me.
   ( )True ( )False

8. I happen to be particularly lucky and I expect to get more of the good things in life than the average person.
   ( )True ( )False

9. I just don't get the breaks, and there's no reason to believe that I will in the future.

   ( )True ( )False

10. My past experiences have prepared me well for my future.

    ( )True ( )False

11. All I can see ahead of me is unpleasantness rather than pleasantness.

    ( )True ( )False

12. I don't expect to get what I really want.

    ( )True ( )False

13. When I look ahead to the future I expect I will be happier than I am now.

    ( )True ( )False

14. Things just won't work out the way I want them to.

    ( )True ( )False

15. I have great faith in the future.

    ( )True ( )False

16. I never get what I want, so it's foolish to want anything.

    ( )True ( )False

17. It is very unlikely that I will get any real satisfaction in the future.

    ( )True ( )False

18. The future seems vague and uncertain to me.

    ( )True ( )False

19. I can look forward to more good times than bad times.

    ( )True ( )False

20. There's no use in really trying to get something I want because I probably won't get it.

    ( )True ( )False

APPENDIX D

R CODE FOR SECTION 6.2.1

```
#read and attach data
data <- read.table("data.txt", header=T)
attach(data)
#load required packages
library(bnlearn)
library(psych)
library(Rgraphviz)
source("https://bioconductor.org/biocLite.R")
biocLite("RBGL")
library(gRain)
library(lattice)

# Calculate Croanbach alpha coefficient for BDI
alpha_bdi <- data.frame(Q2_1, Q2_2, Q2_3, Q2_4, Q2_5, Q2_6, Q2_7,
                        Q2_8, Q2_9, Q2_10, Q2_11, Q2_12, Q2_13,
                        Q2_14, Q2_15, Q2_16, Q2_17, Q2_18, Q2_19,
                        Q2_20, Q2_21)


alpha(alpha_bdi)

# calculate Croanbach alpha coefficient for BHS
alpha_bhs <-  data.frame(Q3_1, Q3_2, Q3_3, Q3_4, Q3_5, Q3_6, Q3_7,
                        Q3_8, Q3_9, Q3_10, Q3_11, Q3_12, Q3_13,
                        Q3_14, Q3_15, Q3_16, Q3_17, Q3_18, Q3_19,
```

```
Q3_20, Q3_21)

alpha(alpha_bhs)

# Calculate Croanbach alpha coefficient for RSES
alpha_se <- data.frame(Q1, Q2, Q3, Q4, Q5, Q6, Q7, Q8, Q9, Q10)

alpha(alpha_se)

#discretize psychiatric variables by domain knowledge
Depression <- ifelse(BDI_score>0 & BDI_score<=10,
"None_or_minimal",ifelse(BDI_score>10 & BDI_score<=18,"mild",
ifelse(BDI_score>18 & BDI_score<=29,"mild","severe")))

Hopelessness <- ifelse(BHS_score>0 & BDI_score<=3,"None_or_minimal",
ifelse(BDI_score>3 & BDI_score<=8,"mild",
ifelse(BDI_score>8 & BDI_score<=13,"moderate","severe")))

Self_esteem <- ifelse(Self_esteem_score>0 & Self_esteem_score<=15,
"Low",ifelse(Self_esteem_score>15 & BDI_score<=25,"Normal","High"))

data1 <- data.frame(Depression, Hopelessness, Self_esteem, Age,
      Gender, School_type,Family_structure, Social_activity
      Father_educational_status, Mother_educational_status,
      Family_income, Student_income, Number_of_siblings,
      Type_of_settlement,Father_occupation, Mother_occupation,
      Type_of_accommodation, Smoking_status, Alcohol_use,
      Student_job, Relationship_status)

#create ban list
banlist <- matrix(c("Self_esteem","Age","Depression","Age",
      "Hopelessness","Age", "School_type","Age","Family_structure",
```

```
      "Age","Number_of_siblings","Age","Father_educational_status",
      "Age","Mother_educational_status", "Age","Family_income",
      "Age","Student_income","Age","Smoking_status","Age",
      "Type_of_settlement","Age","Father_occupation","Age",
      "Mother_occupation","Age","Type_of_accommodation","Age",
      "Alcohol_use","Age","Student_job","Age","Relationship_status",
      "Age","Social_activity","Age","Self_esteem", "Gender",
      "Depression","Gender","Hopelessness","Gender","School_type",
      "Gender","Family_structure","Gender",
      "Father_educational_status","Gender",
      "Mother_educational_status","Gender","Family_income","Gender",
      "Student_income","Gender","Number_of_siblings","Gender",
      "Type_of_settlement","Gender","Father_occupation","Gender",
      "Mother_occupation","Gender","Type_of_accommodation","Gender",
      "Smoking_status","Gender","Alcohol_use","Gender","Student_job",
      "Gender","Relationship_status","Gender","Social_activity",
      "Gender","Age","Gender", "Gender","Age"),
      ncol = 2, byrow = TRUE)

#Build bayesian network and calculate network score
bn1 <- tabu(data1, blacklist = banlist)
fit1 <- bn.fit(bn1, data1, method = "bayes")
score(bn1, data1)

#Plot DAG1
highlight1 <- list(nodes = nodes(bn1), arcs = arcs(bn1), col="black")
g <- graphviz.plot(bn1, highlight = highlight1, shape = "rectangle")

nodeRenderInfo(g1) <- list(textCol = c("Depression" = "black",
      "Hopelessness" = "black","Self_esteem" = "black"),
       fill = c("Age"="grey", "Mother_educational_status" = "green",
       "Mother_occupation" = "green",
```

```
        "Father_educational_status" = "green",
        "Number_of_siblings"="green",  "Father_occupation" = "green",
        "Family_structure" = "green", "Family_income" = "green",
        "Student_income" = "green","Type_of_settlement" = "green",
        "Student_job" = "green","Relationship_status" = "green",
        "Alcohol_use"="green","Smoking_status" = "green",
        "Gender"="grey", "School_type"="green",
        "Type_of_accommodation" = "green",
        "Social_activity" = "green","Depression" = "yellow",
        "Hopelessness" = "yellow","Self_esteem"="yellow"),
        fontsize = c("Age"="15","Mother_educational_status" = 14.5,
        "Mothers_occupation" = 15,"Father_educational_status" = 14.5,
        "Number_of_siblings"=15,"Father_occupation" = 14.5,
        "Family_structure"=15,"Family_income" = 15,
        "Student_income"=15,"Type_of_settlement" = 15,
         "Student_job" = 15,"Relationship_status" = 14.5,
         "Alcohol_use"=15, "Smoking_status" = 15,
         "Gender"=15, "School_type"=15,
         "Type_of_accommodation" = 14.5, "Social_activity"=15,
         "Depression"=15, "Hopelessness"=14.5, "Self_esteem"=15))
g1 <- renderGraph(g1)

#Make queries and plot conditional probabilities
q <- compile(as.grain(fit1))

#Student_income
q_Gender <- setEvidence(q,nodes = "Age",states = "20-")
querygrain(q_Gender,nodes = "Students_income")
q_Gender <- setEvidence(q, nodes = "Age",states = "20-23")
querygrain(q_Gender, nodes = "Students_income")
q_Gender <- setEvidence(q, nodes = "Age",states = "23+")
querygrain(q_Gender,nodes = "Students_income")
```

```
bn.fit.barchart(fit1$Student_income,
ylab = "Levels of Student_income")

#Student_job
q_Student_income <- setEvidence(q,nodes = "Student_income",
states = "0-500TL")
querygrain(q_Student_income, nodes = "Student_job")$Student_job
q_Student_income <- setEvidence(q,nodes = "Student_income",
states = "501-750TL")
querygrain(q_Student_income,nodes = "Student_job")$Student_job
q_Student_income <- setEvidence(q,nodes = "Student_income",
states = "751-1000TL")
querygrain(q_Student_income,nodes = "Student_job")$Student_job
q_Student_income <- setEvidence(q,nodes = "Student_income",
states = "1000+TL")
querygrain(q_Student_income,nodes = "Student_job")$Student_job

bn.fit.barchart(fit1$Student_job, ylab = "Levels of Student_job")

#Family_income
q_Student_income <- setEvidence(q, nodes = "Student_income",
states = "0-500TL")
querygrain(q_Student_income, nodes = "Family_income")
q_Student_income <- setEvidence(q, nodes = "Student_income",
states = "501-750TL")
querygrain(q_Student_income,nodes = "Family_income")
q_Student_income <- setEvidence(q, nodes = "Student_income",
states = "751-1000TL")
querygrain(q_Student_income,nodes = "Family_income")
q_Student_income <- setEvidence(q, nodes = "Student_income",
states = "1000+TL")
```

```
querygrain(q_Student_income,nodes = "Family_income")

bn.fit.barchart(fit1$Family_income,
ylab = "Levels of Family_income")

#Type_of_settlement
q_Family_income<- setEvidence(q, nodes = "Family_income",
states = "0-1400TL")
querygrain(q_Family_income, nodes = "Type_of_settlement")
q_Family_income<- setEvidence(q, nodes = "Family_income",
states = "1401-3000TL")
querygrain(q_Family_income, nodes = "Type_of_settlement")
q_Family_income<- setEvidence(q, nodes = "Family_income",
states = "3001-5000TL")
querygrain(q_Family_income, nodes = "Type_of_settlement")
q_Family_income<- setEvidence(q, nodes = "Family_income",
states = "5000+TL")
querygrain(q_Family_income, nodes = "Type_of_settlement")
bn.fit.barchart(fit1$Type_of_settlement,
ylab = "Levels of Type_of_settlement")

#Father_occupation
q_Family_income<- setEvidence(q, nodes = "Family_income",
states = "0-1400TL")
querygrain(q_Family_income, nodes = "Fathers_occupation")
q_Family_income<- setEvidence(q, nodes = "Family_income",
states = "1401-3000TL")
querygrain(q_Family_income, nodes = "Father_occupation")
q_Family_income<- setEvidence(q, nodes = "Family_income",
states = "3001-5000TL")
querygrain(q_Family_income, nodes = "Father_occupation")
q_Family_income<- setEvidence(q, nodes = "Family_income",
```

```
states = "5000+TL")
querygrain(q_Family_income, nodes = "Father_occupation")
bn.fit.barchart(fit1$Father_occupation,
ylab = "Levels of Father_occupation")

#Father_educational_status
q_Fathers_occupation<- setEvidence(q, nodes = "Father_occupation",
states = "Private_sector_employee")
querygrain(q_Father_occupation, nodes = "Father_educational_status")
q_Father_occupation<- setEvidence(q, nodes = "Father_occupation",
states = "Public_sector_employee")
querygrain(q_Father_occupation, nodes = "Father_educational_status")
q_Father_occupation<- setEvidence(q, nodes = "Father_occupation",
states = "Self-employed")
querygrain(q_Father_occupation, nodes = "Father_educational_status")
q_Father_occupation<- setEvidence(q, nodes = "Father_occupation",
states = "Unemployed")
querygrain(q_Father_occupation, nodes = "Father_educational_status")
bn.fit.barchart(fit1$Father_educational_status,
ylab = "Levels of Father_educational_status")

#Mother_educational_status
q_Father_educational_status<- setEvidence(q,
nodes = "Father_educational_status", states = "Illiterate")
querygrain(q_Father_educational_status,
nodes = "Mother_educational_status")
q_Father_educational_status<- setEvidence(q,
nodes = "Father_educational_status", states = "Primary_school")
querygrain(q_Father_educational_status,
nodes = "Mother_educational_status")
q_Father_educational_status<- setEvidence(q,
nodes = "Father_educational_status", states = "Secondary_school")
```

```
querygrain(q_Father_educational_status,
nodes = "Mother_educational_status")
q_Father_educational_status<- setEvidence(q,
nodes = "Father_educational_status", states = "High_school")
querygrain(q_Father_educational_status,
nodes = "Mother_educational_status")
q_Father_educational_status<- setEvidence(q,
nodes = "Father_educational_status", states = "University")
querygrain(q_Father_educational_status,
nodes = "Mother_educational_status")
q_Father_educational_status<- setEvidence(q,
nodes = "Father_educational_status", states = "Master_or_Ph.D.")
querygrain(q_Father_educational_status,
nodes = "Mother_educational_status")
bn.fit.barchart(fit1$Mother_educational_status,
ylab = "Levels of Mother_educational_status")

#Number_of_siblings
q_Mother_educational_status<- setEvidence(q,
nodes = "Mother_educational_status", states = "Illiterate")
querygrain(q_Mother_educational_status, nodes = "Number_of_siblings")
q_Mother_educational_status<- setEvidence(q,
nodes = "Mother_educational_status", states = "Primary_school")
querygrain(q_Mother_educational_status, nodes = "Number_of_siblings")
q_Mother_educational_status<- setEvidence(q,
nodes = "Mother_educational_status", states = "Secondary_school")
querygrain(q_Mother_educational_status, nodes = "Number_of_siblings")
q_Mother_educational_status<- setEvidence(q,
nodes = "Mother_educational_status", states = "High_school")
querygrain(q_Mother_educational_status, nodes = "Number_of_siblings")
q_Mother_educational_status<- setEvidence(q,
nodes = "Mothers_educational_status", states = "University")
```

```
querygrain(q_Mother_educational_status,
nodes = "Number_of_siblings")
q_Mother_educational_status<- setEvidence(q,
nodes = "Mother_educational_status", states = "Master_or_Ph.D.")
querygrain(q_Mother_educational_status,
nodes = "Number_of_siblings")

bn.fit.barchart(fit1$Number_of_siblings,
ylab = "Levels of Number_of_siblings")

#Family_structure
q_Number_of_siblings<- setEvidence(q,
nodes = "Number_of_siblings", states = "1")
querygrain(q_Number_of_siblings, nodes = "Family_structure")
q_Number_of_siblings<- setEvidence(q,
nodes = "Number_of_siblings", states = "2")
querygrain(q_Number_of_siblings, nodes = "Family_structure")
q_Number_of_siblings<- setEvidence(q, nodes = "Number_of_siblings",
states = "3")
querygrain(q_Number_of_siblings, nodes = "Family_structure")
q_Number_of_siblings<- setEvidence(q, nodes = "Number_of_siblings",
states = "4")
querygrain(q_Number_of_siblings, nodes = "Family_structure")
q_Number_of_siblings<- setEvidence(q, nodes = "Number_of_siblings",
states = "5+")
querygrain(q_Number_of_siblings, nodes = "Family_structure")

bn.fit.barchart(fit1$Family_structure,
ylab = "Levels of Family_structure")

#Smoking_status
bn.fit.barchart(fit1$Smoking_status,
```

```
ylab = "Levels of Smoking_status")

#Relationship_status
bn.fit.barchart(fit1$Relationship_status,
ylab = "Levels of Relationship_status")

#Alcohol_use
q_Smoking_status<- setEvidence(q, nodes = "Smoking_status",
 states = "Smoker")
querygrain(q_Smoking_status, nodes = "Alcohol_use")
q_Smoking_status<- setEvidence(q, nodes = "Smoking_status",
states = "Not_smoker")
querygrain(q_Smoking_status, nodes = "Alcohol_use")

bn.fit.barchart(fit1$Alcohol_use, ylab = "Levels of Alcohol_use")

#Social_activity
q_Gender<- setEvidence(q, nodes = "Gender", states = "Male")
querygrain(q_Gender, nodes = "Social_activity")
q_Gender<- setEvidence(q, nodes = "Gender", states = "Female")
querygrain(q_Gender, nodes = "Social_activity")

bn.fit.barchart(fit1$Social_activity,
ylab = "Levels of Social_activity")

#Type_of_accommodation
q_Gender<- setEvidence(q, nodes = "Gender", states = "Male")
querygrain(q_Gender, nodes = "Type_of_accommodation")
q_Gender<- setEvidence(q, nodes = "Gender", states = "Female")
querygrain(q_Gender, nodes = "Type_of_accommodation")
bn.fit.barchart(fit1$Type_of_accommodation,
ylab = "Levels of Type_of_accommodation")
```

```
#School_type
q_Gender<- setEvidence(q, nodes = "Gender", states = "Male")
querygrain(q_Gender, nodes = "School_type")
q_Gender<- setEvidence(q, nodes = "Gender", states = "Female")
querygrain(q_Gender, nodes = "School_type")
bn.fit.barchart(fit1$School_type, ylab = "Levels of School_type")

#Depression
querygrain(q,nodes = "Depression")$Depression
qGender <- setEvidence(q,nodes = "Gender",states = "Male")
querygrain(qGender,nodes = "Depression")
qGender <- setEvidence(q,nodes = "Gender",states = "Female")
querygrain(qGender,nodes = "Depression")

bn.fit.barchart(fit1$Depression, ylab = "Levels of Depression")

#Self_esteem
querygrain(q,nodes = "Self_esteem")$Self_esteem
qDepression <- setEvidence(q,nodes = "Depression",
states = "None_or_minimal")
querygrain(qDepression,nodes = "Self_esteem")
qDepression <- setEvidence(q,nodes = "Depression",
states = "Mild")
querygrain(qDepression,nodes = "Self_esteem")
qDepression <- setEvidence(q,nodes = "Depression",
states = "Moderate")
querygrain(qDepression,nodes = "Self_esteem")
qDepression <- setEvidence(q,nodes = "Depression",
states = "Severe")
querygrain(qDepression,nodes = "Self_esteem")
```

```
bn.fit.barchart(fit1$Self_esteem, ylab = "Levels of Self_esteem")

#Hopelessness
querygrain(q,nodes = "Hopelessness")$Hopelessness
qDepression <- setEvidence(q,nodes = "Depression",
states = "None_or_minimal")
querygrain(qDepression,nodes = "Hopelessness")
qDepression <- setEvidence(q,nodes = "Depression",states = "Mild")
querygrain(qDepression,nodes = "Hopelessness")
qDepression <- setEvidence(q,nodes = "Depression",
states = "Moderate")
querygrain(qDepression,nodes = "Hopelessness")
qDepression <- setEvidence(q,nodes = "Depression",states = "Severe")
querygrain(qDepression,nodes = "Hopelessness")

bn.fit.barchart(fit1$Hopelessness, ylab = "Levels of Hopelessness")

#Calculate Specificities
pred.Depression1 <- predict(fit1, data=data1, node="Depression")
c1 <- ifelse(Depression==pred.Depression1,1,0)
sum(c1)/length(c1)

pred.Hopelessness1 <- predict(fit1, data=data1, node="Hopelessness")
c2 <- ifelse(Hopelessness==pred.Hopelessness1,1,0)
sum(c2)/length(c2)

pred.Selfesteem1 <- predict(fit1, data=data1, node="Self_esteem")
c3 <- ifelse(Self_esteem==pred.Selfesteem1,1,0)
sum(c3)/length(c3)
```

APPENDIX E

R CODE FOR SECTION 6.2.2

```
#Discretize psychiatric variables by the IPD method
d.frame <- data.frame(Self_esteem_score, BDI_score, BHS_score)
discrete.data <- discretize(d.frame, method = "hartemink",breaks = 3,
ibreaks = 100, idisc = "interval")

Depression <-ifelse(discrete.data$BDI_score=="(0.938,11.5]","Low"
,ifelse(discrete.data$BDI_score=="(11.5,20.8]","Normal","High"))

Hopelessness <- ifelse(discrete.data$BHS_score=="(0.981,6.89]","Low"
,ifelse(discrete.data$BHS_score=="(6.89,13]" ,"Normal","High"))

Self_esteem <- ifelse(discrete.data$Self_esteem_score=="(0.971,15.8]"
,"Low",ifelse(discrete.data$Self_esteem_score=="(15.8,19.8]",
"Normal","High"))

#prepare data
data2 <- data.frame(Depression, Hopelessness, Self_esteem, Age,
      Gender, School_type,Family_structure, Social_activity
      Fathers_educational_status, Mothers_educational_status,
      Family_income,Students_income, Number_of_siblings,
      Type_of_settlement,Fathers_occupation, Mothers_occupation,
      Type_of_accommodation, Smoking_status, Alcohol_use,
      Students_job, Relationship_status)
```

```
bn2 <- tabu(data2, blacklist = banlist)
fit2 <- bn.fit(bn2, data2, method="bayes")
score(bn2, data2, type="bic")

highlight2 <- list(nodes = nodes(bn2),arcs = arcs(bn2),col="black")
g2 <- graphviz.plot(bn2, highlight = highlight2, shape="rectangle")

#Plot DAG2
highlight1 <- list(nodes = nodes(bn1), arcs = arcs(bn1), col="black")
g <- graphviz.plot(bn1, highlight = highlight1, shape = "rectangle")

nodeRenderInfo(g2) <- list(textCol = c("Depression" = "black",
      "Hopelessness" = "black","Self_esteem" = "black"),
       fill = c("Age"="grey", "Mother_educational_status" = "green",
       "Mother_occupation" = "green",
       "Father_educational_status" = "green",
       "Number_of_siblings"="green",  "Father_occupation" = "green",
       "Family_structure" = "green", "Family_income" = "green",
       "Student_income" = "green","Type_of_settlement" = "green",
       "Student_job" = "green","Relationship_status" = "green",
       "Alcohol_use"="green","Smoking_status" = "green",
       "Gender"="grey", "School_type"="green",
       "Type_of_accommodation" = "green",
       "Social_activity" = "green","Depression" = "yellow",
       "Hopelessness" = "yellow","Self_esteem"="yellow"),
       fontsize = c("Age"="15","Mother_educational_status" = 14.5,
       "Mother_occupation" = 15,"Father_educational_status" = 14.5,
       "Number_of_siblings"=15,"Father_occupation" = 14.5,
       "Family_structure"=15,"Family_income" = 15,
       "Student_income"=15,"Type_of_settlement" = 15,
        "Student_job" = 15,"Relationship_status" = 14.5,
        "Alcohol_use"=15, "Smoking_status" = 15,
```

```
        "Gender"=15, "School_type"=15,
        "Type_of_accommodation" = 14.5, "Social_activity"=15,
        "Depression"=15, "Hopelessness"=14.5, "Self_esteem"=15))
g2 <- renderGraph(g2)
renderGraph(g2)
dev.off()

# Queries
q <- compile(as.grain(fit2))

#Depression
querygrain(q,nodes = "Depression")$Social_activity
qSocial_activity <- setEvidence(q,nodes = "Social_activity",
states = "Has_social_activity")
querygrain(qSocial_activity,nodes = "Depression")
qSocial_activity <- setEvidence(q,nodes = "Social_activity",
states = "No_social_activity")

bn.fit.barchart(fit2$Depression, ylab = "Levels of Depression")

#Hopelessness
querygrain(q,nodes = "Hopelessness")$Hopelessness
qDepression <- setEvidence(q,nodes = "Depression",
states = "Low")
querygrain(qDepression,nodes = "Hopelessness")
qDepression <- setEvidence(q,nodes = "Depression",
states = "Normal")
querygrain(qDepression,nodes = "Hopelessness")
qDepression <- setEvidence(q,nodes = "Depression",
states = "High")
querygrain(qDepression,nodes = "Hopelessness")
```

```
#Self_esteem
pdf("/Users/gunalbilek/Dropbox/Entropy/Selfesteem2.pdf")
bn.fit.barchart(fit2$Self_esteem, ylab = "Levels of Self_esteem")
dev.off()

querygrain(q,nodes = "Self_esteem")$Self_esteem
qDepression <- setEvidence(q,nodes = "Depression",states = "Low")
querygrain(qDepression,nodes = "Self_esteem")
qDepression <- setEvidence(q,nodes = "Depression",states = "Normal")
querygrain(qDepression,nodes = "Self_esteem")
qDepression <- setEvidence(q,nodes = "Depression",states = "High")
querygrain(qDepression,nodes = "Self_esteem")

#Calculate specifities
pred.Depression2 <- predict(fit2, data=data2, node="Depression")
c1 <- ifelse(Depression==pred.Depression2, 1, 0)
sum(c1)/length(c1)

pred.Hopelessness2 <- predict(fit2, data=data2, node="Hopelessness")
c2 <- ifelse(Hopelessness==pred.Hopelessness2, 1, 0)
sum(c2)/length(c2)

pred.Selfesteem2 <- predict(fit2, data=data2, node="Self_esteem")
c3 <- ifelse(Self_esteem==pred.Selfesteem2, 1, 0)
sum(c3)/length(c3)
```

CURRICULUM VITAE

PERSONEL INFORMATION

Name Surname : Günal BİLEK

Date of birth and place : 31.03.1989, Istanbul

Foreign Languages : English

E-mail : gunal-34@hotmail.com

EDUCATION

Year	Department	University	Date of Graduation
Master	School of Mathematics and Statistics	University of Sheffield	2013
Undergraduate	Department of Statistics	Mugla University	2009

WORK EXPERIENCE

Degree	Institute	Enrollment
2014 - present	Bitlis Eren University	Research Assisstant
2013 - 2014	Ministry of National Education	Strategy Development Officer

PUBLISHMENTS

Papers

1. Bilek, G., and Karaman, F. (2018). "An Investigation into the Relationship among Psychiatric, Demographic and Socio-Economic Variables with Bayesian Network Modeling", Entropy, 20(3): 189.

2. Sarı, S. A., Bilek, G., and Çelik, E. (2018). "Test anxiety and self-esteem in senior high school students: a cross-sectional study", Nordic Journal of Psychiatry, 72(2): 84-88.

3. Özdemir, A. A., Gündemir, Y. E., Küçük, M., Sarıcı, D. Y., Elgörmüş, Y., Çağ, Y., and Bilek, G. (2018). "Vitamin D Deficiency in Pregnant Women and Their Infants", Journal of Clinical Research in Pediatric Endocrinology, 10(1): 44.

CONFERENCE PAPERS

1. Bilek, G. (2015). "How to Estimate Health State Utilities in Health Economics", 12th Applied Statistics 2015, Bled, Slovenia.