Assignment Submission Coversheet

Faculty of Science, Engineering and Built Environment

Student ID:	218545396
Student Name:	Pooja Bhat
Campus:	☐ Burwood ☐ Waterfront ☐ Waurn Ponds ☐ Warrnambool ☒ Cloud

Assignment Title:	Assessment 2: Multivariate and Categorical Data Analysis		
Due Date:	24 May 2019 by 11.30 PM	**Assessment Item:**	Report
Course Code/Name:	S777/ Master of Data Analytics		
Unit Code/Name:	SIT743/ Multivariate and Categorical Data Analysis	**Unit Chair / Campus Coordinator:**	Sutharshan Rajasegarar
Practical Group: (if applicable)	Not applicable		

If this assignment has been completed by a group or team:

1. Each student in the group must complete and sign a separate coversheet
2. The assignment will be returned to the student in the group nominated below

Assignment to be returned to: (Student name and Student ID number)	

PLAGIARISM AND COLLUSION

Plagiarism occurs when a student passes off as the students own work, or copies without acknowledgement as to its authorship, the work of another person. Collusion occurs when a student obtains the agreement of another person for a fraudulent purpose with the intent of obtaining an advantage in submitting an assignment or other work. Work submitted may be reproduced and/or communicated for the purpose of detecting plagiarism and collusion.

DECLARATION

I certify that the attached work is entirely my own (or where submitted to meet the requirements of an approved group assignment, is the work of the group), except where work quoted or paraphrased is acknowledged in the text. I also certify that it has not been previously submitted for assessment in this or any other unit or course unless permissions for this has been granted by the Unit Chair of this unit. I agree that Deakin University may make and retain copies of this work for the purposes of marking and review, and may submit this work to an external plagiarism-detection service who may retain a copy for future plagiarism detection but will not release it or use it for any other purpose.

Signed:	POOJA BHAT	**Date:**	12.05.2019

An assignment will not be accepted for assessment if the declaration appearing above has not been signed by the author. If submitting electronically, print your full name in place of a signature.

COMMENTS

Mark Awarded:		Assessor's Signature:		Date:	

1. Australian Murray River Bayesian network

1.1) Using chain rule, p(S, F, R, N, U, T, P, C) for the above network can be written as

$$p(S)p(F)p(R|S)p(N|S,U)p(U)p(T|S)p(P|F,R,N)\ p(C|N,T)$$

1.2) The number of minimum number of parameters required to fully specify the distribution can be calculated as below.

Probability	Number of parameters	Reason
p(S)	3	Node S can have 4 states with one parameter that can be derived using sum constraint (1-p(sum of probability of other values))
p(F)	3	Node F can have 4 states with one parameter that can be derived using sum constraint (1-p(sum of probability of other values))
p(R|S)	8	Node R can take 3 states and S can take 4 states hence we can have 4 free parameters that can be derived using sum constraint.
p(N|S,U)	8	Node N can take 2 states and nodes S and U take 4 and 2 states respectively. We have 8 free parameters that can be derived using sum constraint.
p(U)	1	U can take 2 states, so we have one free parameter.
p(T|S)	8	T can take 3 states and S can take 4 states, Hence by sum constraint we have 4 free parameters
p(P|F,R,N)	24	P can take 2 states and F,R,N can take 4,3,2 states respectively. By sum constraint we will have 24 free parameters
p(C|N,T)	6	C can take 2 states and N,T can take 2 and 3 states respectively. Hence we will have 6 free parameters by sum constraint.

Table 1 : parameters calculation for question 1.1

Hence the total number of parameters required are: 3 + 3 + 8 + 8 + 1 + 8 + 24 + 6 =61

1.3) Bayesian network gives a clear representations of independence. Had we not known the independence information, or if there are no independencies among the variables assumed, the p(S, F, R, N, U, T, P, C) can be written by chain rule as below:

p(C|S, F, R, N, U, T, P)p(P|S, F, R, N, U, T)p(T|S, F, R, N, U)p(U|S, F, R, N)p(N|S, F, R)p(R|S,F)p(F|S)p(S)

Node	Number of states
S {winter, spring, summer, autumn}	4
R {high, medium, low}	3
F {Cod, Callop, Catfish, Redfin}	4
T {king fern, river cherry, maple silkwood}	3
U {high, low}	2
N {high, low}	2
C {Healthy, not healthy}	2
P {high, low}	2

Number pf parameters required:

Probability	**Parameters required**
p(C|S, F, R, N, U, T, P)	1 x 4 x 4 x 3 x 2 x 2 x 3 x 2 = 1152
p(P|S, F, R, N, U, T)	1 x 4 x 4 x 3 x 2 x 2 x 3 = 576
p(T|S, F, R, N, U)	2 x 4 x 4 x 3 x 2 x 2 = 384
p(U|S, F, R, N)	1 x 4 x 4 x 3 x 2 = 96
p(N|S, F, R)	1x 4 x 4 x 3 = 48
p(R|S,F)	2 x 4 x 4 = 32
p(F|S)	3 x 4 = 12
p(S)	4 – 1 = 3

The number of parameters that would be required to specify this distribution would be:
(1152 + 576 + 384 + 96 + 48 + 32 + 12 + 3) = 2303
So computation of 2303 parameters is much more complex then computation of 61 parameters derived with the knowledge of independence between variables. Thus it is evident that a knowledge of independence between variables simplified the computation effort.

1.4)

a) $S \perp$ U | ∅ (S is marginally independent of U)

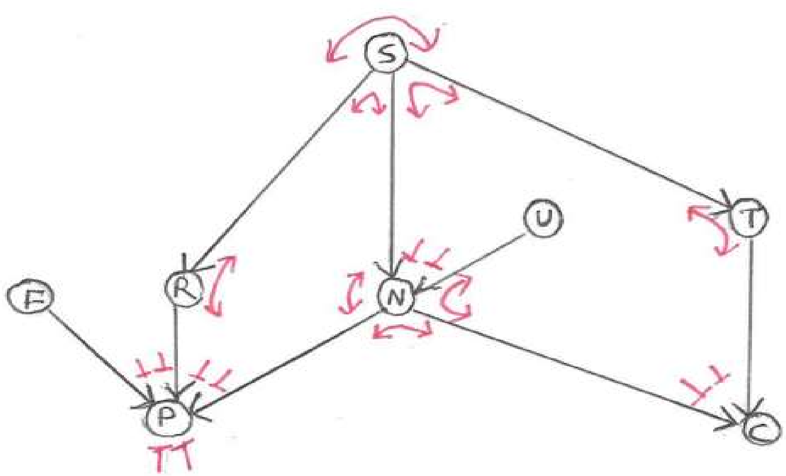


Figure 1 : Bayesian network for question 1.4 a d-separation

i) The paths that can be taken to travel from S to U are

1) S -> N <- U
2) S -> T -> C <- N <- U
3) S -> R -> P <- N <- U

ii) Path 1 is blocked at N since node N has a head to head, Path 2 is blocked at C since node C has a head to head and path 3 is blocked at P since node P has a head to head.

iii) S is d-separated by U since it is blocked on all 3 paths, hence the statement is true.

b) $F \perp$ U | {N, P} (F is conditionally independent of U given {N, P})

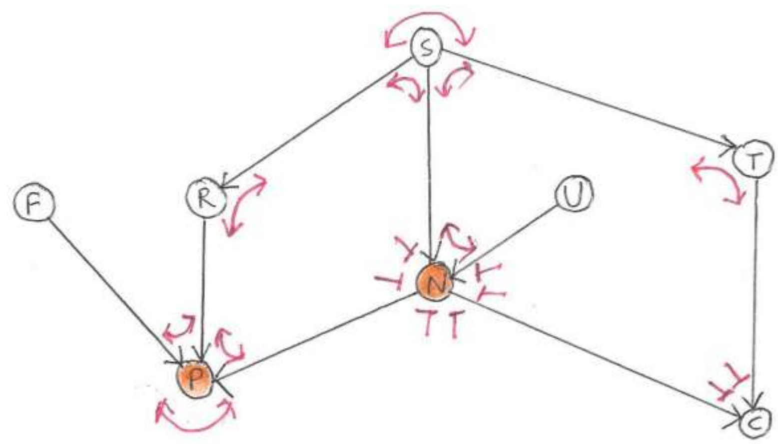


Figure 2 : Bayesian network for question 1.4 b d-separation

i) The paths that can be taken to travel from F to U are

1) F -> P <- N <- U
2) F -> P <- R <- S -> N <- U
3) F -> P <- R <- S -> T -> C<- N <- U

ii) Path 1 is blocked at N since N is an observed node and it has a head to Tail.
Path 2 is not blocked at any of the nodes as shown in figure 2
Path 3 is blocked at C and N. since N is an observed node and has a head to tail and C has a head to head.

iii) The statement is false since the 2 nodes are dependent by path 2.

1.5) The R code attached in the R file has produced the below Bayesian network

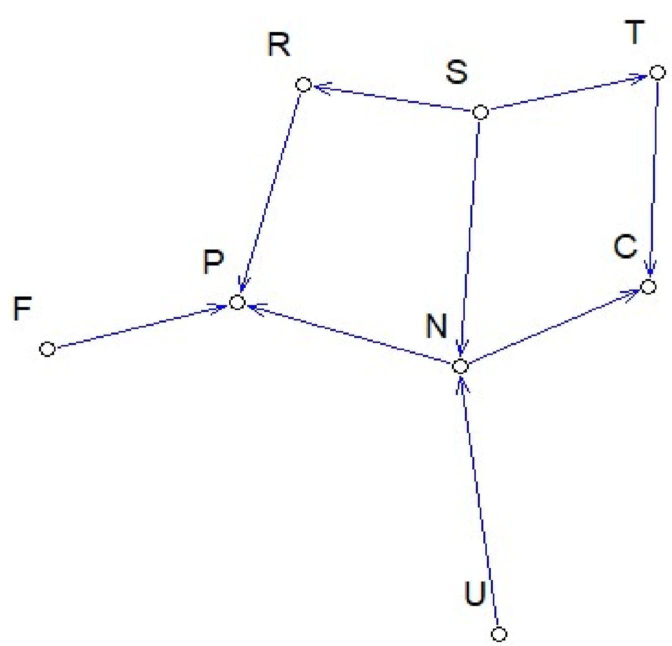


Figure 3: Bayesian network created using R-Program

The output of the d-separation test is as below:

```
> dSep(dag, first="S", second="U", cond=NULL)
[1] TRUE
> dSep(dag, first="F", second="U", cond=c("N", "P"))
[1] FALSE
```

1.6) Variable elimination:

$$P(C \mid F=\text{Cod}, T=\text{King fern}, U=\text{high}) = \frac{P(C, F=\text{Cod}, T=\text{king fern}, U=\text{high})}{P(F=\text{Cod}, T=\text{king Fern}, U=\text{high})}$$

$$P(C,F,T,U) = \sum_{RNPS} P(S,F,R,N,U,T,P,C)$$

$$= \sum_{RNPS} P(S)P(F)P(R|S)P(N|S,U)P(U)P(T|S)P(P|F,R,N)P(C|N,T)$$

$$= \sum_{RNPS} f_0(S) f_1(F) f_2(R,S) f_3(N,S,U) f_4(U) f_5(T,S) f_6(P,F,R,N) f_7(C,N,T)$$

Observe $F=\text{Cod}$

$$= \sum_{RNPS} f_0(S) f_2(R,S) f_3(N,S,U) f_4(U) f_5(T,S) f_8(P,R,N) f_7(C,N,T)$$

Observe $T=\text{king Fern}$

$$= \sum_{RNPS} f_0(S) f_2(R,S) f_3(N,S,U) f_4(U) f_9(S) f_8(P,R,N) f_{10}(C,N)$$

Observe $U=\text{high}$

$$= \sum_{RNPS} f_0(S) f_2(R,S) f_{11}(N,S) f_9(S) f_8(P,R,N) f_{10}(C,N)$$

Eliminate R

$$\sum_{NPS} f_0(S) f_{11}(N,S) f_9(S) f_{10}(C,N) \sum_R f_2(R,S) f_8(P,R,N)$$

$$\sum_{NPS} f_0(S) f_{11}(N,S) f_9(S) f_{10}(C,N) f_{12}(P,S,N)$$

Eliminate N

$$\sum_{PS} f_0(S) f_9(S) \sum_N f_{11}(N,S) f_{10}(C,N) f_{12}(P,S,N)$$

$$\sum_{P,S} f_0(S) f_9(S) f_{13}(S,C,P)$$

Eliminate P

$$\sum_S f_0(S) f_9(S) f_{14}(S,C)$$

Eliminate S $= f_{15}(C)$

Therefore,

$$P(C \mid F=\text{Cod}, T=\text{King Fern}, U=\text{high}) = \frac{f_{15}(C)}{\sum_C f_{15}(C)}$$

2.1) a) The below belief network has been produced by the R code.

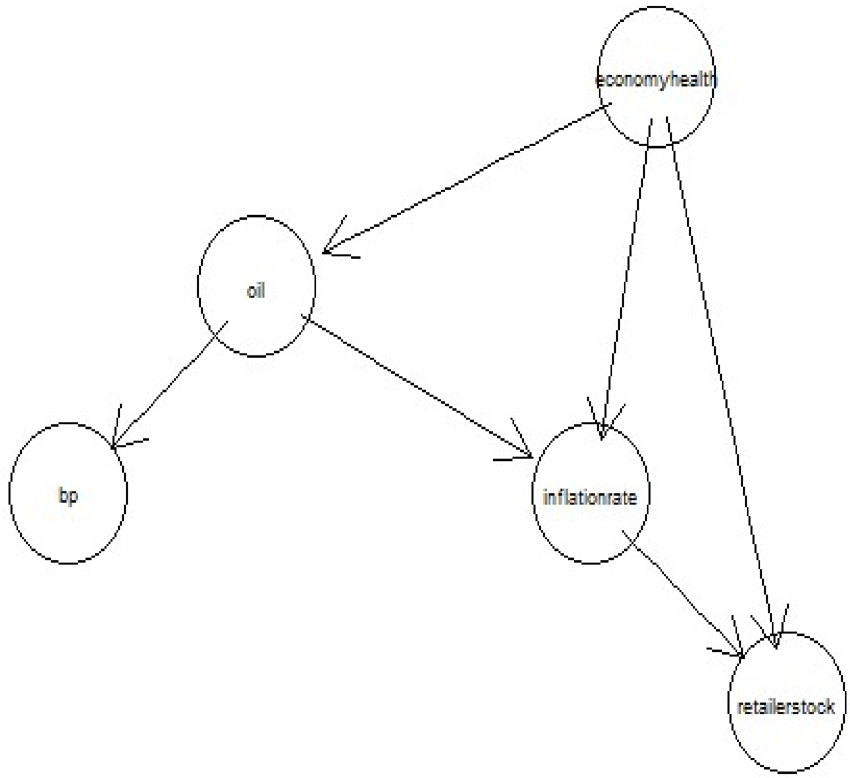


Figure 4: Belief network created using R-Program

b) The below are the probability tables generated by R program.

1) Economy health:

```
 economy - health
  low high
  0.6  0.4
```

$P(eh = high) = 0.4$

2) BP

```
              oil
bp             low  high
   low        0.2   0.7
   high       0.6   0.2
   normal     0.2   0.1
```

$p(bp = high | oil = low) = 0.6$ $p(bp = low | oil = low) = 0.2$

$p(bp = high | oil = high) = 0.2$ $p(bp = low | oil = high) = 0.7$

3) Oil

```
        economy - health
oil       low high
  low    0.3 0.15
  high   0.7 0.85
```

$p(oil = high | eh = low) = 0.7$ $p(oil = low | eh = high) = 0.15$

4) Retailer Stock

```
$`retailer - stock`
, , economy - health = low

                 inflation - rate
retailer - stock low high normal
            low  0.4  0.7   0.35
            high 0.6  0.3   0.65
, , economy - health = high

                 inflation - rate
retailer - stock low high normal
            low  0.8  0.9    0.6
            high 0.2  0.1    0.4
```

$p(rt = high \| inf = low, eh = low) = 0.6$	$p(rt = high \| inf = low, eh = high) = 0.2$
$p(rt = high \| inf = high, eh = low) = 0.3$	$p(rt = high \| inf = high, eh = high) = 0.1$
$p(rt = low \| inf = normal, eh = low) = 0.35$	$p(rt = low \| inf = normal, eh = high) = 0.6$

5) Inflation Rate

```
, , economy - health = low

                  oil
inflation - rate  low high
           low    0.15 0.4
           high   0.80 0.3
           normal 0.05 0.3
, , economy - health = high

                 oil
inflation - rate  low high
          low    0.65 0.19
          high   0.20 0.01
          normal 0.15 0.80
```

$p(inf = high \| oil = low, eh = low) = 0.8$	$p(inf = high \| oil = low, eh = high) = 0.2$
$p(inf = high \| oil = high, eh = low) = 0.3$	$p(inf = high \| oil = high, eh = high) = 0.01$
$p(inf = low \| oil = low, eh = low) = 0.15$	$p(inf = low \| oil = low, eh = high) = 0.65$
$p(inf = low \| oil = high, eh = low) = 0.4$	$p(inf = low \| oil = high, eh = high) = 0.19$

2.2)

a) Given that the BP stock price is low and the retailer stock price is high, the probability that price of oil is high is : 93.90%

```
       low       high
0.06097227 0.93902773
```

b) Given that inflation rate is high, the probability that BP stock price being normal is 15.46%.

```
bp
      low      high    normal
0.4266994 0.4186405 0.1546601
```

3)

3.1)

$$P(D=1|B=0) = \sum_{AC} \frac{P(D=1, B=0, A, C)}{P(B=0)}$$

$$= \sum_{AC} \frac{P(D=1|C)\, P(B=0)\, P(A)\, P(C|A, B=0)}{\beta}$$

The expression on top can be written as below

$$\sum_{AC} P(D=1|C) P(B=0) P(A) \sum P(C|A, B=0) =$$

$$P(A=0)P(B=0)P(C=0|A=0,B=0)\,P(D=1|C=0)$$
$$+\ P(A=0)P(B=0)P(C=1|A=0,B=0)\,P(D=1|C=1)$$
$$+\ P(A=1)P(B=0)P(C=0|A=1,B=0)\,P(D=1|C=0)$$
$$+\ P(A=1)P(B=0)P(C=1|A=1,B=0)\,P(D=1|C=1)$$

Which can now be written as below by substituting probability values

$$[\alpha \times \beta \times 0.1 \times (1-\gamma)] + [\alpha \times \beta \times 0.9 \times 0.7]$$
$$+ [(1-\alpha) \times \beta \times 0.2 \times (1-\gamma)] + [(1-\alpha) \times \beta \times 0.8 \times 0.7]$$

$$= \beta\,[(\alpha(1-\gamma)0.1) + (\alpha \times 0.63) + ((1-\alpha)(1-\gamma)\,0.2)$$
$$+ (0.56 \times (1-\alpha))]$$

Cancelling Beta β from numerator and denominator we get.

$$\therefore\ P(D=1|B=0) = [(0.1\alpha(1-\gamma)) + (\alpha\, 0.63) + (0.2(1-\alpha)(1-\gamma))$$
$$+ (0.56(1-\alpha))]$$

3.2) $\alpha = \frac{4}{20} = 0.2$ Since $\alpha = P(A=0)$

$\beta = \frac{3}{20} = 0.15$ Since $\beta = P(B=0)$

$\gamma = \frac{1}{20} = 0.05$ Since $\gamma = P(D=0 \text{ and } C=0)$

3.3) Substituting values $\alpha = 0.2$ and $\gamma = 0.05$ derived above we get

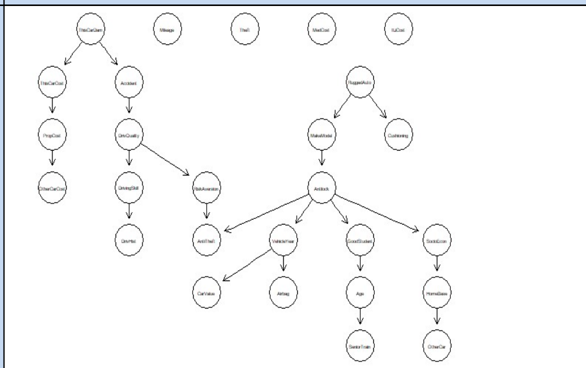
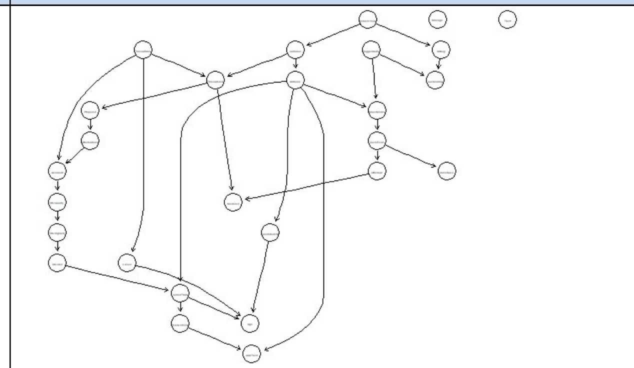
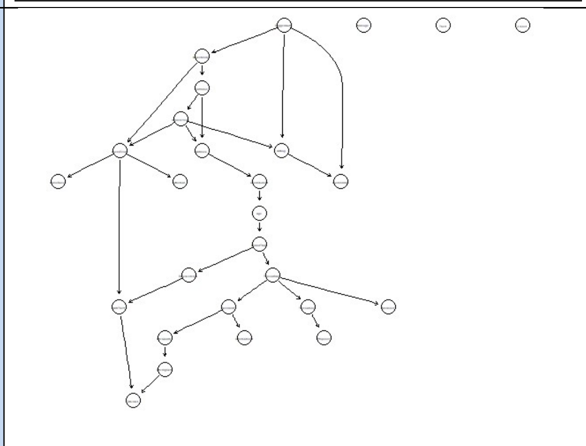
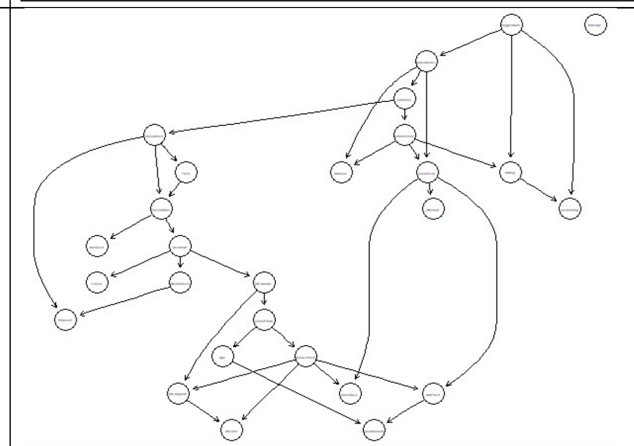
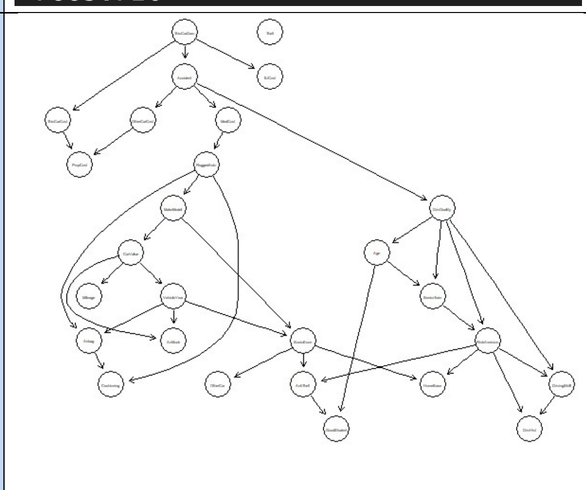
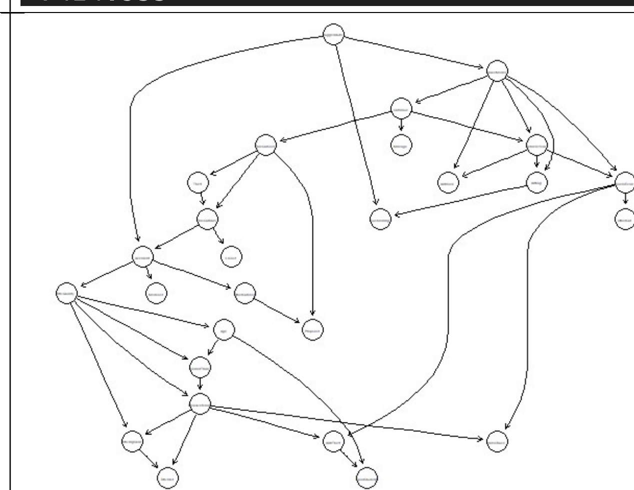
$$P(D=1|B=0) = [(0.1 \times 0.2 \times (1-0.05)) + (0.2 \times 0.63)$$
$$+ (0.2 \times (1-0.05)(1-0.2)) +$$
$$+ (0.56 \times (1-0.20))]$$

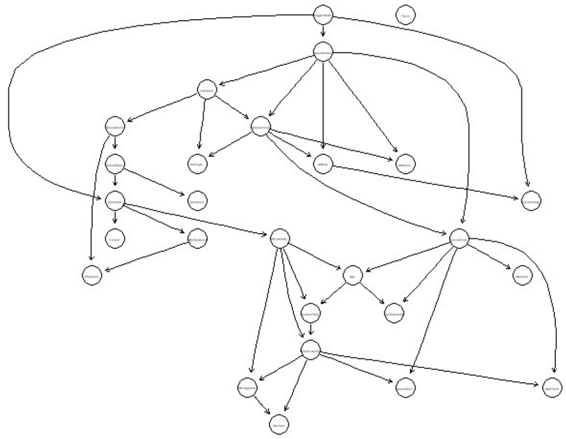
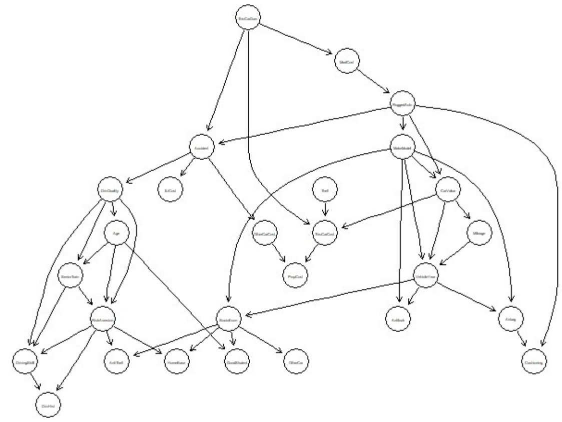
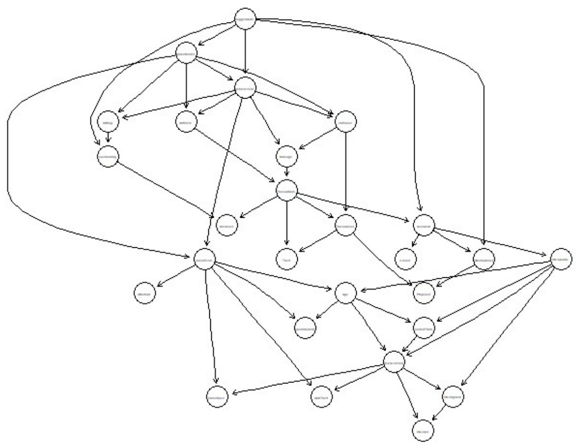
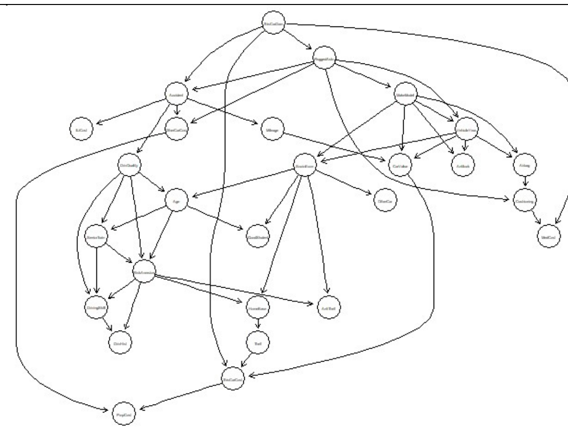
$$= 0.019 + 0.126 + 0.152 + 0.448$$
$$= 0.745$$

$$P(D=1|B=0) = 0.745$$

4) Bayesian Structure Learning

4.1)

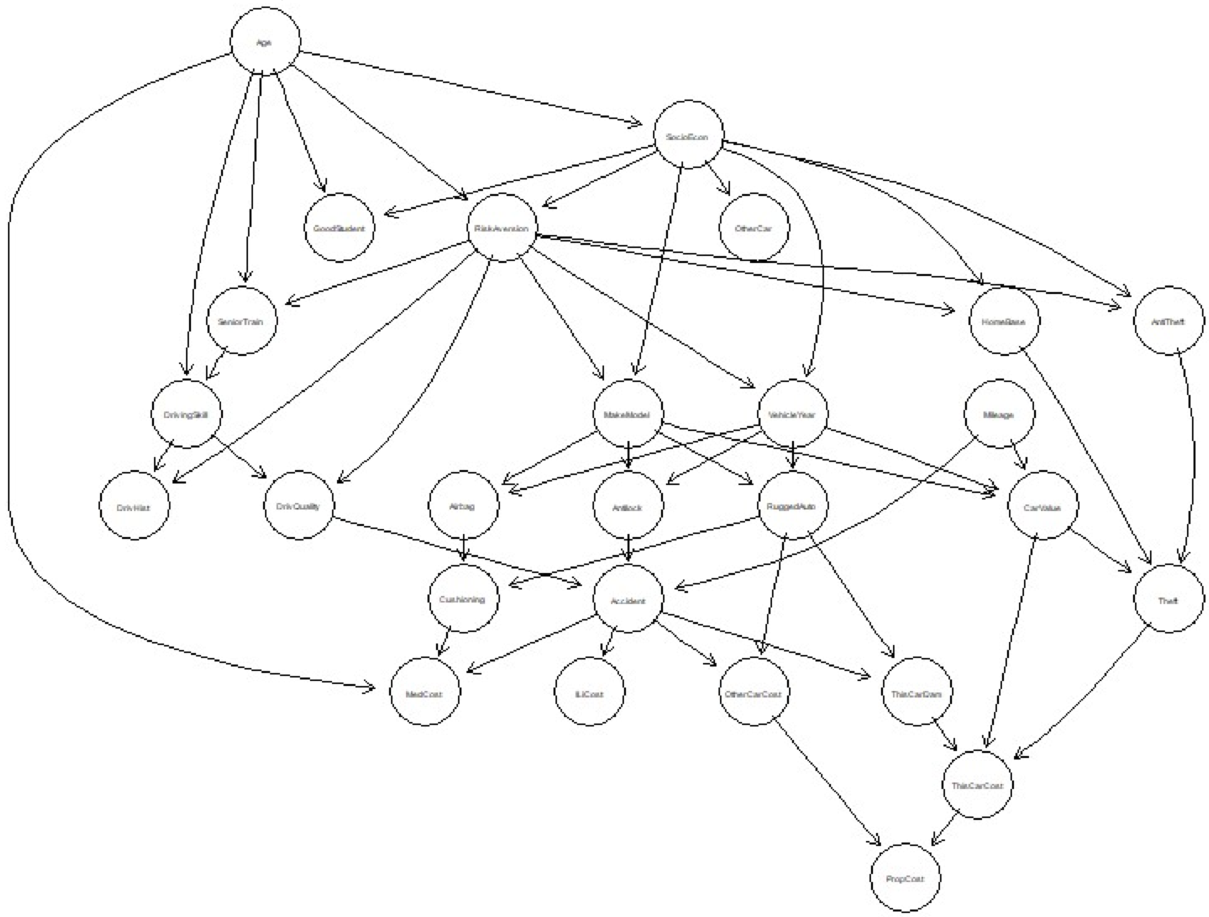
	BIC	BDE
Plot with first 100		
Score with first 100	-1869.421	-1746.589
Plot with first 500		
Score with first 500	-7883.726	-7424.533
Plot with first 1000		
Score with first 1000	-14836.5	-14218.44

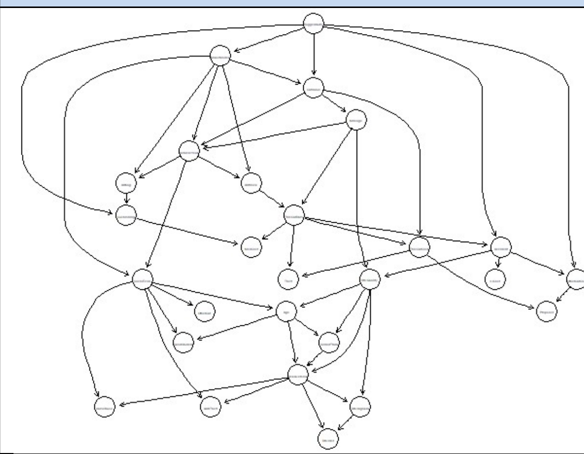
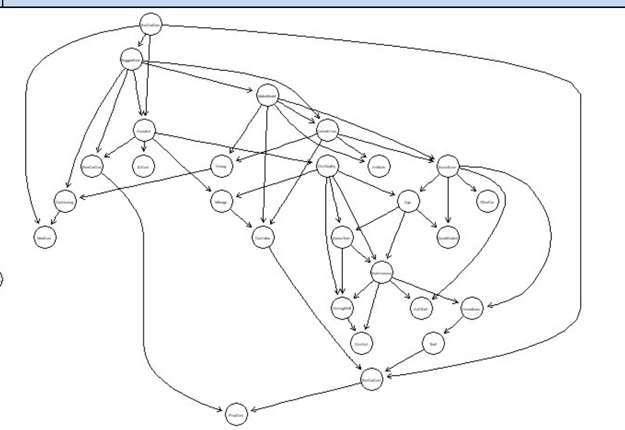
Plot with first 5000		
Score with first 5000	-68329.68	-67558.05
Plot with first 15000		
Score with first 15000	-199756.7	-198383.8

4.2) There are 27 variables in the input dataset, so it is a somewhat complex structure. We tried fitting BIC and BDE models both of which chose simpler structures for smaller samples. As the samples grew, the scoring techniques tried to bring the structure as close to the data as possible. In all cases, we can see that the BIC scores are lower than BDE, which means that for all the samples BIC was out-performing BDE for all sample sizes. As the number of samples increase, structures learning increases with both the BIC and BDE scoring techniques. The scores are also getting better as the sample size increases, since it has more data to learn from. We can also see that there are no edges in the structures generated BIC until 5000 samples, however BDE has all edges mapped by 1000 samples. Also, we can notice that the average Markov blanket size increases as the number of samples used for training both BIC and BDE increase. Also, for BIC the penalization coefficients increase as the number of samples used increase.

4.3) Plot and score with full data set

Below is the output of the true network structure.



	BIC	**BDE**
Plot with full dataset		
Score with all data	-266113	-264021.8

In comparison to the original network, we can see that both BIC and BDE networks have some incorrect directed edges along with some incorrectly identified parent and child dependencies. For example we can see Age has 5 child nodes and no parent nodes in the true network structure, however both BIC and BDE have shown 2 parent nodes and 3 child nodes for Age.

5. Bayesian usage examples

1. Bayesian networks are used for criminal profiling of unknown offenders using a prior dataset of homicide behavior. Such a dataset is created by investigators after documenting the examined characteristics and psychological behaviors of convicted offenders in addition to forensic evidence obtained from crime scenes. Bayesian Network considers probabilistic relationships between all known variables from past criminal knowledge which is then used to train and infer insights that can be used in identifying psycho-behavior in suspected criminals and helps narrow down list of suspects in unsolved criminal investigations. The Bayesian Network comprises of a directed acyclic graph which when combined with conditional probability tables can provide a joint probability distribution. It is found that with these methods, on an average about 80% of characteristics in unknown offenders are predicted correctly.

   References:
   ‘Constructing Bayesian networks for criminal profiling from limited data’, 2007, ‘*Knowledge-Based Systems’*, retrieved 21 March 2007,
   <http://lisc.mae.cornell.edu/LISCpapers/KNOSYSprofiling08.pdf>

2. Bayes linear classifier models are used in morbidity probably in patients after a heart surgery. This method was developed and tested using a dataset that was curated by using patient records from 1090 patients who underwent artery bypass grafting University Hospital of Siena (Italy) over years 2002-2004. A collection of 78 variables were considered as likely risk predictors before doing a feature selection of about 8 variables. The dataset was divided into a training set and a test set. Bayesian method were employed to predict a binary variable that represented the morbidity outcome. This method assumes the input classifiers to have a normal distribution with equal covariance matrices and may sometimes see loss of performance due to non-normality or non-homoscedasticity however it has been very effective in predicting morbidity in patients post heart surgery.

   References:
   ‘Bayesian Approach in Medicine and Health Management‘, 2012, retrieved 15 May 2013, <https://www.intechopen.com/books/current-topics-in-public-health/bayesian-approach-in-medicine-and-health-management>