IE 5203 Decision Analysis Lab I
Probabilistic Modeling, Inference and Decision Making with Netica

Overview of Netica Software

- Netica Application is a comprehensive tool for working with *Bayesian networks* and *standard influence diagrams*. It can build, learn, modify, transform and store networks, as well as answer queries or find optimal solutions using its inference engine. It can learn probabilistic relations from data.

- It provides easy graphical editing of Bayesian networks and influence diagrams through a GUI, and generates presentation quality graphics which can be transferred to other documents.

- It supports arc reversal operations and automatically updates the affected CPTs in the network and adds additional arcs to parent nodes.

- It can find optimal decisions for sequential decision problems (i.e. later decisions are dependent on the results of earlier ones).

- It has an extensive built-in library of probabilistic functions and other mathematical functions

- It has facilities for discretization of continuous variables.

- The Netica API can be used for development of applications that uses Bayesian networks as the underlying knowledge representation and reasoning tool.

- For more details about Netica Application and API, see http://www.norsys.com

Lab Learning Outcomes

- The Lab Exercises aim to get you familiarized with the basic functions of the Netica Application software. At the end of the lab session, you will be able to:

    1. Build Bayesian Networks using the GUI and input the relevant data.

    2. Compile the network and perform various types of probabilistic inference and experiment with the network.

    3. Observe conditional independence and dependence through experimentation.

    4. Perform sensitivity analysis on the network using mutual information.

    5. Perform basic operations on the network such as arc reversal.

    6. Build an influence diagram by extending the Bayesian network you have built.

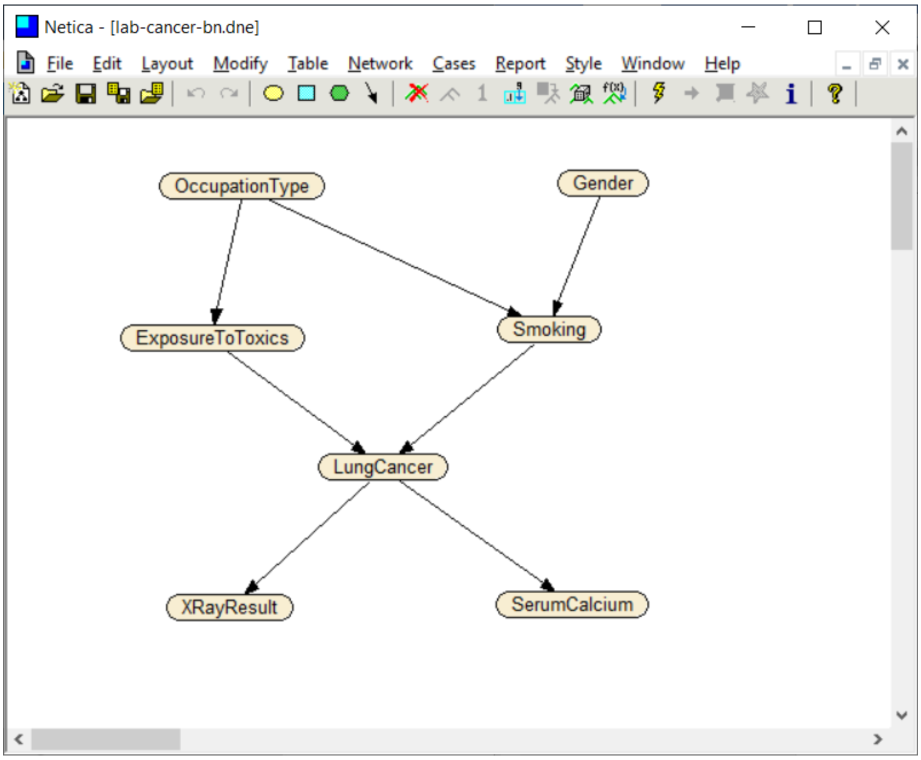
    7. Find the optimal decision policy from the influence diagram.

    8. Learning probabilities from data for network with known graphical structure.

Netica Lab Exercises

1 Building a Bayesian Network in Netica

- Build the “Cancer” Bayesian Network model as discussed in Chapter 5 of the lecture notes using the Netica Application software with the following information:



Conditional Probability Tables:

OccupationType	
Office	0.25
NonOffice	0.75

Gender	
Male	0.50
Female	0.50

	ExposureToxics	
OccupationType	True	False
Office	0.05	0.95
NonOffice	0.10	0.90

		Smoking	
Gender	*OccupationType*	True	False
Male	Office	0.10	0.90
Male	NonOffice	0.20	0.80
Female	Office	0.05	0.95
Female	NonOffice	0.10	0.90

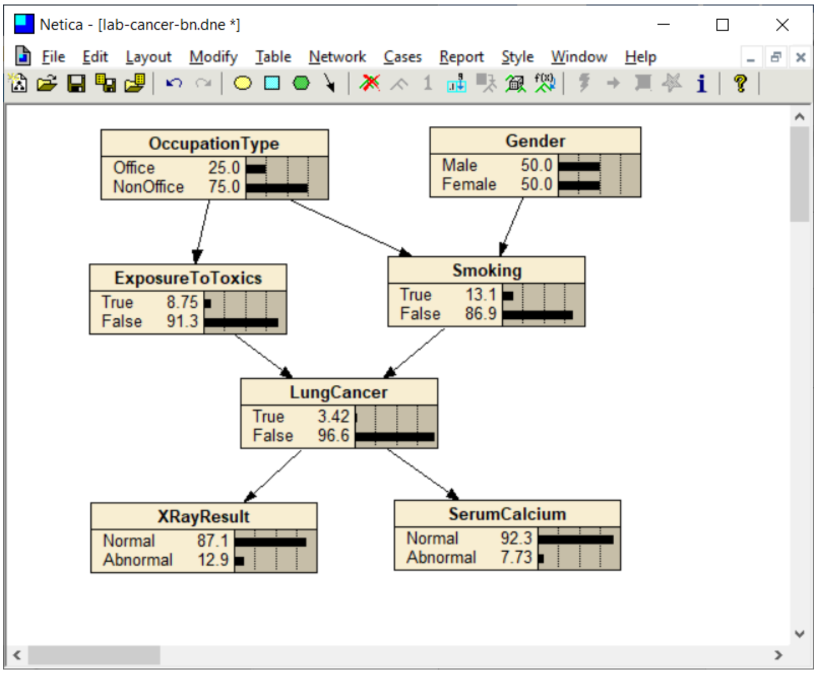
		LungCancer	
Smoking	*ExposureToxics*	True	False
True	True	0.25	0.75
True	False	0.10	0.90
False	True	0.15	0.85
False	False	0.01	0.99

	XRayResult	
LungCancer	Normal	Abnormal
True	0.05	0.95
False	0.90	0.10

	SerumCalcium	
LungCancer	Normal	Abnormal
True	0.15	0.85
False	0.95	0.05

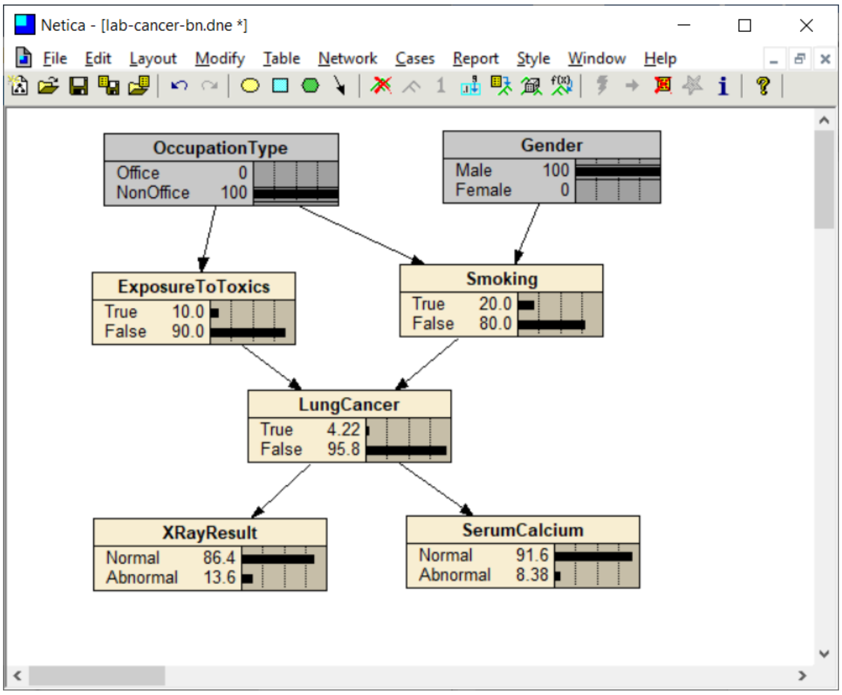
2 Probabilistic Inferences

- The network must be compiled before it is ready to perform inference.
- Run Network → Compile. The *prior marginal probabilities* for each node will be displayed. You may need to choose Network → Automatic Updating.



2.1 Predictive Inference

***Question* 2.1:** How likely is a male non-office worker likely to have "Lung Cancer"?



Answer: ____________________________

2.2 Diagnostic Inference

***Question* 2.2:** If a person was found to have "Lung Cancer", how likely is "Smoking" the cause?

Answer: ________________________

***Question* 2.3:** If a person was found to have "Lung Cancer", how likely is "Exposure to Toxics" the cause?

Answer: ________________________

***Question* 2.4:** If a male office worker was found to have abnormal "Serum Calcium", how likely is he to have "Lung Cancer"?

Answer: ________________________

***Question* 2.5:** If a female non-office worker was found to have abnormal "X-Ray Result", how likely is her "Exposure to Toxics"?

Answer: ________________________

2.3 Inter-Casual Reasoning

***Question* 2.6:** If a person has abnormal "Serum Calcium" what are the probabilities of the two probable causes?

Answer: Probability that "Smoking" was the cause = __________________________
Probability that "Exposure to Toxics" was the cause = __________________

***Question* 2.7:** Now, if it was confirmed that he was a "Smoker", what can you say about the probability for his "Exposure to Toxics"?

Answer: ____

2.4 Conditional Independence

***Question* 2.8:**

- Set “Lung Cancer” to True.
- Instantiate any outcomes of any of the four nodes: “Occupation Type”, “Gender”, “Exposure to Toxics”, “Smoking”, and observe the probabilities of “X-Ray Result” and “Serum Calcium”.
- What do you observe? ____
- Now, set “Lung Cancer” to False and repeat the above.
- What do you observe? ____
- What can you conclude? ____

***Question* 2.9:**

- Set “Lung Cancer” to True.
- Instantiate any outcomes of “Serum Calcium” and observe the probabilities of “X-Ray Result”.
- What do you observe? ____
- Now, set “Lung Cancer” to False and repeat the above.
- What do you observe? ____
- What can you conclude? ____

3. Value of Information Analysis: Mutual Information

Entropy

- The *amount of uncertainty* in a random variable X is given by the ***Entropy*** of its probability distribution. It is defined as:

$$H(X) = -\sum_{x} p(x)\log_2 p(x)$$

Examples:

1. The entropy of the coin tossing game with p(head)=0.5 and p(tail)=0.5 is
   a. $H(0.5, 0.5) = -0.5 \log_2 0.5 - 0.5 \log_2 0.5 = -0.5(-1) - 0.5(-1) = 1$

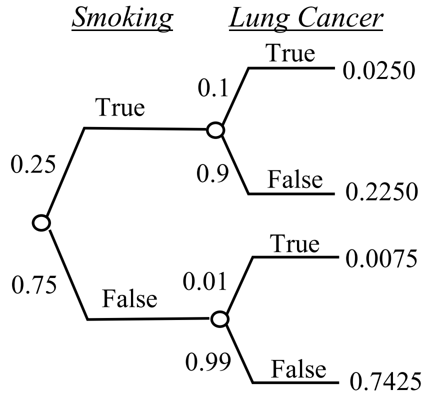
2. The entropy of the thumb tack tossing game with p(up)=0.7 and p(down)=0.3 is
   a. $H(0.7, 0.3) = -0.7 \log_2 0.7 - 0.3 \log_2 0.3 = -0.7(-0.51457) - 0.3(-1.73697) = 0.88129$

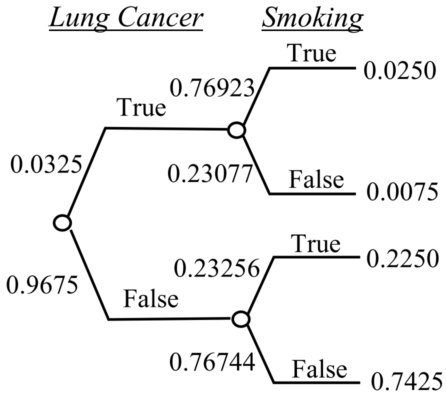
Mutual Information between Two Variables

- The ***Mutual Information*** between two random variables X and Y measures how much information one variable tells about the other. It is defined as:

$$I(X;Y) = \sum_{x}\sum_{y} p(x,y)\log_2 \frac{p(x,y)}{p(x)p(y)}$$

Example:





The Mutual Information between "Smoking" and "Lung Cancer" is

$$\begin{aligned} I(S, LC) = {} & 0.0250 \log_2(0.0250/(0.25\times0.0325)) + 0.2250 \log_2(0.2250/(0.25\times0.9675)) \\ & + 0.0075 \log_2(0.0075/(0.75\times0.0325)) + 0.7425 \log_2(0.7425/(0.75\times0.9675)) \\ = {} & 0.02893456 \end{aligned}$$

Exercise: Show that given any two random variables X and Y:
(a) $I(X, Y) = I(Y, X)$
(b) If X and Y are independent then $I(X, Y) = I(Y, X) = 0$.

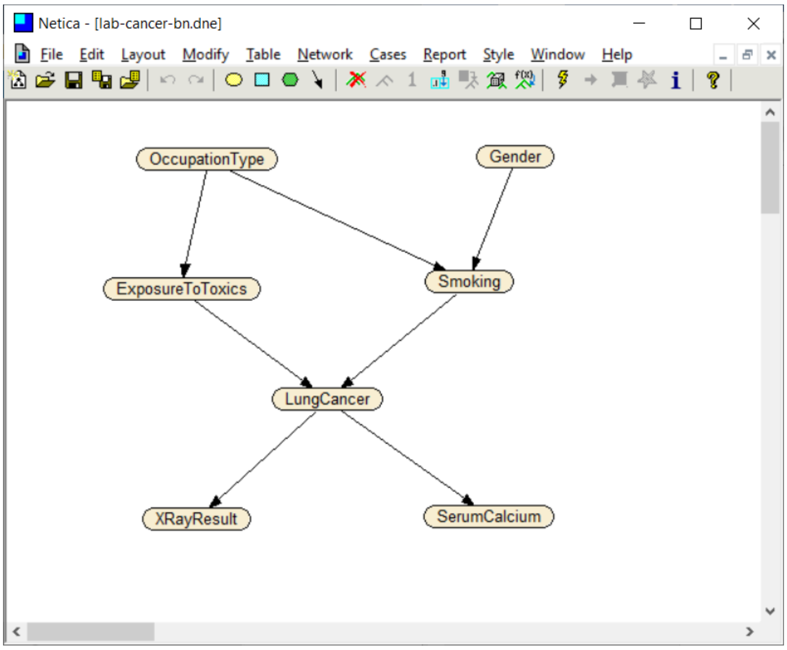
***Question* 3.1:** In our BN model, what findings or inputs will provide the best information about the presence or absence of "Lung Cancer"?

- *Hint:* Choose Network → Sensitivity to Findings.

Answers: ____

4. Basic Network Operations

Arc Reversal Operation



***Question* 4.1:** Reverse the arc between "Occupation Type" and "ExposureToToxics".
What are now the probabilities of "OccupationType" and "ExposureToToxics"?

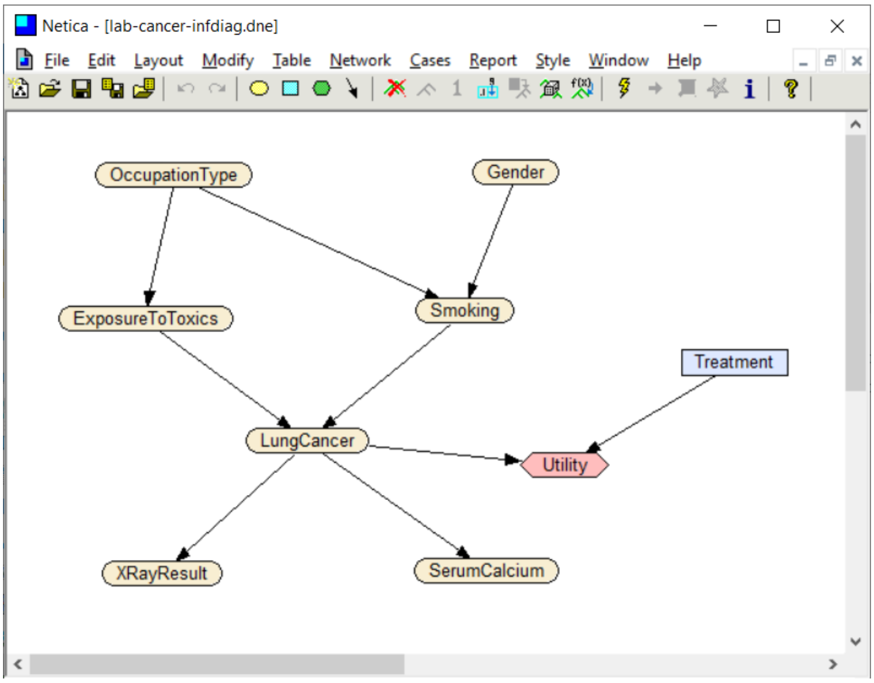
Answers: ____

***Question* 4.2:** Reverse the arc between "Smoking" and "Lung Cancer".
What are now the probabilities of "Smoking" and "LunchCancer"?

Answers: ____

5. Influence Diagram Modeling

- Extend the previous BN into an Influence Diagram by adding a decision node for "Treatment" and a value node for "Utility" using the information below:



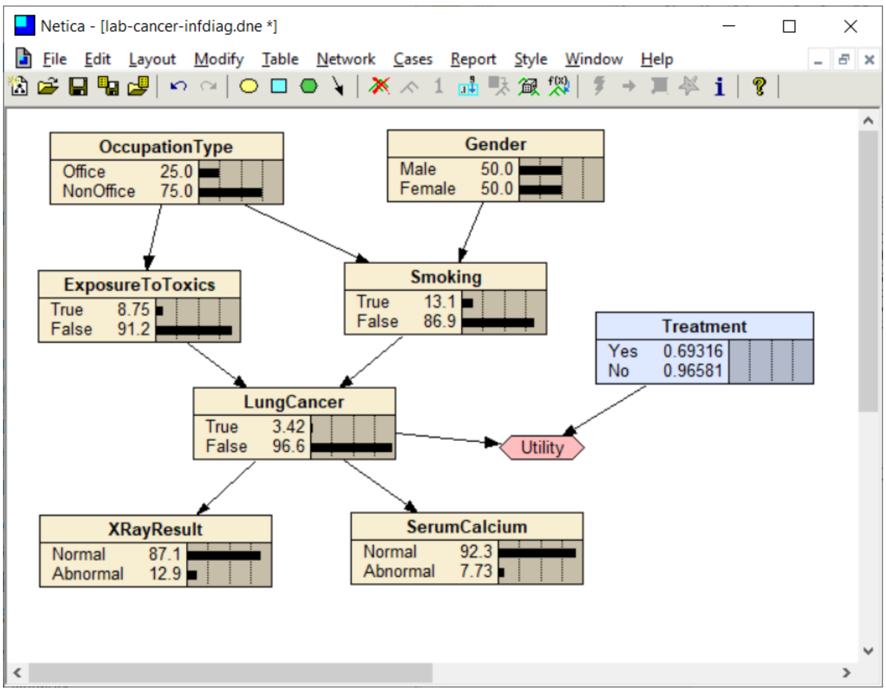
Alternatives

Treatment
Yes
No

Utility Values

Treatment	*LungCancer*	*Utility*
Yes	True	0.5
Yes	False	0.7
No	True	0
No	False	1

- Choose Network → Compile.



***Question* 5.1:** A non-office female worker has abnormal "X-Ray Result" but her "Serum Calcium" was normal. Should she go for "Treatment"?

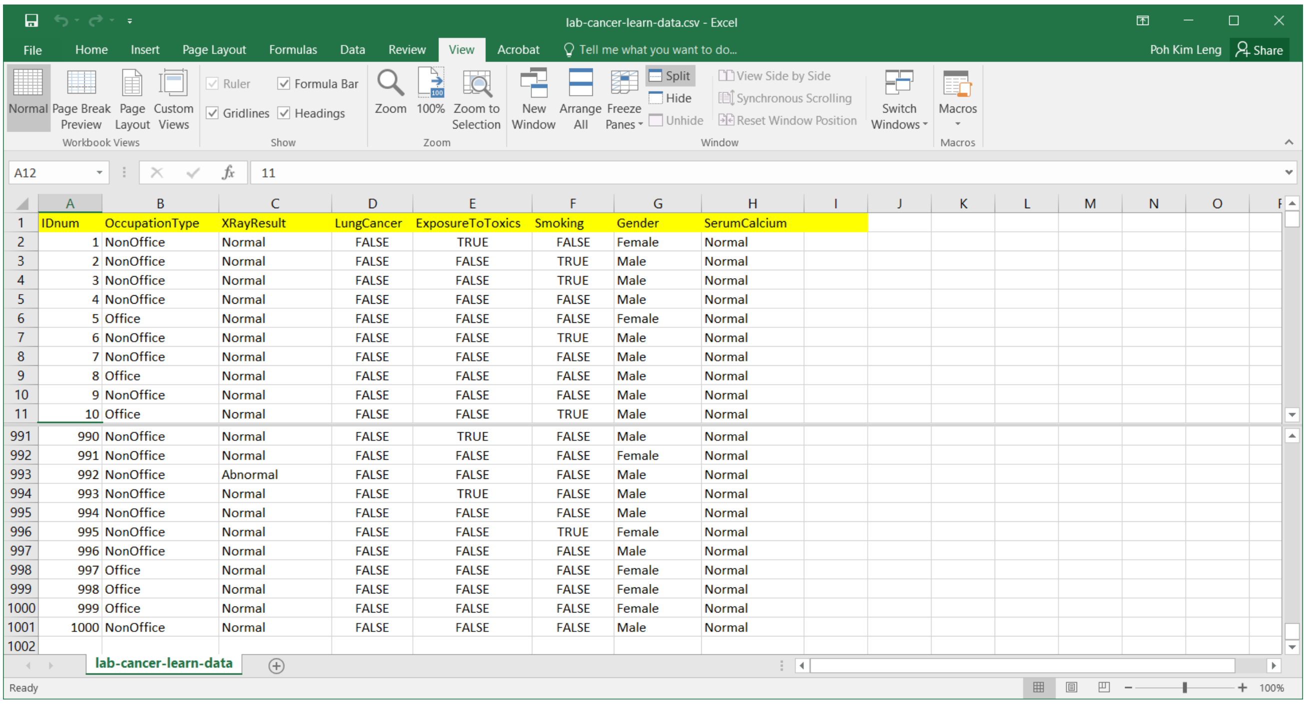
Answer: ________________________________

***Question* 5.2:** A non-office female worker is abnormal in both "X-Ray Result" and "Serum Calcium". Should she go for "Treatment"?

Answer: ________________________________

6. Learning Probabilities from Data for BN with known graphical structure

- Suppose that the structure of the BN is known but the conditional probability tables are unknown. A data file of 1,000 cases with values of the seven variables is given:



	A	B	C	D	E	F	G	H
1	IDnum	OccupationType	XRayResult	LungCancer	ExposureToToxics	Smoking	Gender	SerumCalcium
2	1	NonOffice	Normal	FALSE	TRUE	FALSE	Female	Normal
3	2	NonOffice	Normal	FALSE	FALSE	TRUE	Male	Normal
4	3	NonOffice	Normal	FALSE	FALSE	TRUE	Male	Normal
5	4	NonOffice	Normal	FALSE	FALSE	FALSE	Male	Normal
6	5	Office	Normal	FALSE	FALSE	FALSE	Female	Normal
7	6	NonOffice	Normal	FALSE	FALSE	TRUE	Male	Normal
8	7	NonOffice	Normal	FALSE	FALSE	FALSE	Male	Normal
9	8	Office	Normal	FALSE	FALSE	FALSE	Male	Normal
10	9	NonOffice	Normal	FALSE	FALSE	FALSE	Male	Normal
11	10	Office	Normal	FALSE	FALSE	TRUE	Male	Normal
991	990	NonOffice	Normal	FALSE	TRUE	FALSE	Male	Normal
992	991	NonOffice	Normal	FALSE	FALSE	FALSE	Female	Normal
993	992	NonOffice	Abnormal	FALSE	FALSE	FALSE	Male	Normal
994	993	NonOffice	Normal	FALSE	TRUE	FALSE	Male	Normal
995	994	NonOffice	Normal	FALSE	FALSE	FALSE	Male	Normal
996	995	NonOffice	Normal	FALSE	FALSE	TRUE	Female	Normal
997	996	NonOffice	Normal	FALSE	FALSE	FALSE	Male	Normal
998	997	Office	Normal	FALSE	FALSE	FALSE	Female	Normal
999	998	Office	Normal	FALSE	FALSE	FALSE	Female	Normal
1000	999	Office	Normal	FALSE	FALSE	FALSE	Female	Normal
1001	1000	NonOffice	Normal	FALSE	FALSE	FALSE	Male	Normal
1002								

***Question* 6.1:** Learn the conditional probabilities of the BN using the data file.