

# ECS7005P - Risk and Decision Making for Big Data and AI 2023/2024

## Coursework

### Instructions:

This assignment is marked out of 100 and counts for 20% of the final mark for the module.

Please note this is an individual coursework.

**Submission Deadline:** You can submit this assignment via QM+ any time up until **14 March 2024, 11:59 PM**.

### Question 1

Dr Who works at MCU Clinic, where he attends to patients with STDs and UTIs. He will diagnose a patient with an STD correctly 70% of the time and will mistake it for UTI only 30% of the time. He will diagnose a patient with a UTI correctly 80% of the time and mistake it for an STD 20% of the time. A survey done by the Clinic shows that 65% of all patients that visit the Clinic have STD and 35% have UTI (not necessarily diagnosed with the correct disease). Dr Who attends to a new patient and diagnoses him with an STD. What is the probability that the new patient has a UTI?  
**[20 marks]**

### Question 2

Create a Bayesian Network to model question 1 and answer the same question: "Dr Who attends to a new patient and diagnoses him with an STD. What is the probability that the new patient has a UTI?". You can use any software you prefer. (If you are using Agena.ai or GeNIe please provide a picture of the model structure and a picture of the NPTs. If you are using R or Python please provide the code and the outputs.)  
**[15 marks]**

### Question 3

- i. Are the below interpretations correct or wrong and why? **[10 marks]**
  - a) Consider the relationship between outdoor temperature and ice cream sales. A negative correlation coefficient between these variables would indicate that that colder temperatures lead to lower demand for ice cream.
  - b) Suppose there's a positive correlation between the number of umbrellas sold and the incidence of flu cases in a city. This correlation suggests that buying more umbrellas can causes more people to get the flu.

- c) You set a significance level ( $\alpha$ ) of 0.05 before conducting your hypothesis test. After analysing the data, you obtain a p-value of 0.06. An interpretation would be that the observed result is statistically significant at the 0.05 level and you reject the null hypothesis accordingly.
- d) Suppose you conduct a study on the relationship between exercise and weight loss and obtain a p-value of 0.25 (significance level is 0.05). This non-significant result indicated that there is no relationship between exercise and weight loss.
- e) Suppose a study calculates a 95% confidence interval for the average weight loss due to a new diet to be [5 kg, 7 kg]. This means that if we were to repeat the study many times, approximately 95% of the resulting confidence intervals would contain the true average weight loss.
- ii. After a recession period, the Annualised Growth Rate is found to follow a normal distribution with mean -3 and variance 10. What is the probability that the growth rate will be: (You can use any software you prefer.) **[10 marks]**
- between 1 and 3
  - greater than -5

#### Question 4

A study into the effectiveness of two different treatment procedures (A and B) has been reported on patients suffering with a particular Disease which can be either chronic or acute. The results were:

Treatment Type	Disease Type	Patient Outcome	
		Success	Failure
Treatment A	Acute	16	4
	Chronic	30	50
Treatment B	Acute	62	18
	Chronic	6	14

- Which treatment has the better success rate overall? **[5 marks]**
- Which treatment is better for patients with the acute form of the Disease? **[5 marks]**
- Which treatment is better for patients with the chronic form of the Disease? **[5 marks]**
- State what is Simpson Paradox and why this is an example of it? **[5 marks]**

#### Question 5

Sandra attends a circus and is presented with the following rules of Shell Game.

- A bet cost £5.
- You can only pick 1 shell from 3 shells.

- If the ball is under your chosen shell, you win £20.
- i) What is the total expected utility of Sandra decision to play the Shell Game? **[5 marks]**
  - ii) If Sandra values the excitement of playing the game as having a utility of £7. What is the total expected utility of Sandra decision to play the Shell Game? **[5 marks]**
  - iii) Should Sandra play the Shell Game (if considering the excitement of playing)? **[5 marks]**
    - Total expected utility of not playing = 0
  - iv) Use any software to create an influence diagram based on the information above (without considering the excitement of playing). Should Sandra play the Shell Game? **[10 marks]**