Risk and Decision-Making for Data Science and AI

Semester B, 2022

Coursework

Instructions

This is an individual coursework. All submissions will be done via QMplus. Please use the Coursework link under the Assessment Tab on the module to select and submit answers. All the answers apart from Question 1 are multiple choice. For question 1, you will need to enter your answers in the blanks provided.

Submission Deadline: March 21st, 2022 11:59 PM.

Part A: Bayes Theorem

Question 1 (20 marks)

A type of cancer that affects people over the age of 50 is far more prevalent in men (2%) than in women (0.1%). There is a screening test used on people over 50 that has different accuracy rates for men and women.

For men the true positive rate is 95% and the false positive rate is 10%

For women the true positive rate is 85% and the false positive rate is 5%

Answer the following questions (enter your values to 3 significant figures):

1. What is the probability that a man will test positive? (2 marks)
2. What is the probability that a woman will test positive? (2 marks)
3. What is the probability a person will test positive? (2 marks)
4. A man tests positive. What is the probability he has the cancer? (2 marks)
5. A man tests negative. What is the probability he has the cancer? (2 marks)
6. A woman tests positive. What is the probability she has the cancer? (2 marks)
7. A woman tests negative. What is the probability she has the cancer? (2 marks)
8. A man tests positive and is subject to an additional test. Assuming that a second test is independent of the first, what is the probability they test positive in this second test? (3 marks)
9. A man tests positive in both the first and second test. What is the probability he has the cancer? (3 marks)

Part B: Utility and Decision Making

Question 2 (20 marks)

Sarah is considering playing a game in which a ball is placed under one of three cups and you win a prize of £20 if you select the cup with the ball under it. Playing the game costs £12.

i. What is the total expected utility of Sarah's decision to play the game? (8 marks)

ii. If Sarah values the excitement of playing the game as having a utility of £7. What is the total expected utility of Sarah's decision to play the game? (8 marks)

iii. Should Sarah play the game? (4 marks)

Part C: Risk Calculations

A recent outbreak of a virus has been traced to a local hotel with attendees at a particular large conference there believed to be at especially great risk. Epidemiologists have conducted a survey of recent visitors to the hotel with the following results:

	Virus	**No Virus**
Conference Attendee	180	800
Not a Conference Attendee	40	1500

Question 3 (5 marks)

What is the relative risk increase of contracting the virus for conference attendees compared to people who did not attend the conference?

Question 4 (5 marks)

What is the absolute risk increase of contracting the virus for conference attendees compared to non-attendees?

Part D: Statistical Paradoxes

Question 5 (20 marks)

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*	60	15
	Chronic	5	20
Treatment B	*Acute*	20	2
	Chronic	34	51

i. Which treatment has the better success rate overall? (5 marks)

ii. Which treatment is better for patients with the acute form of the Disease? (5 marks)

iii. Which treatment is better for patients with the chronic form of the Disease? (6 marks)

iv. What type of statistical paradox exists in this data? (4 marks)

Part E: Influence Diagram

It is March 2022. A band is scheduling their next concert for November 2022. They are deciding whether to confirm their venue as there is a long waiting list for this particular venue. If they do not book it, the concert cannot go ahead. The venue costs £40000 to reserve and this cost cannot be refunded for any reason once confirmed. However, it is uncertain whether there will be social distancing restrictions. At the time they believe the government will do one of the following: ban large gatherings altogether (.6 probability), restrict the number of people that can attend the concert (.3), or allow full concerts (.1). If the concert is allowed to go ahead, they expect to make £300000 from the ticket sales, but if there is a restriction on capacity due to government policy, they will only make a quarter of this amount in ticket sales.

Create an influence diagram in AgenaRisk based on the information above and answer the following multiple-choice questions online about it.

Question 6 (5 Marks)

Select the type of node for each variable listed:

Venue Cost

Pandemic Restrictions

Ticket Sales

Total Utility

Confirm Venue

Concert

Question 7 (10 marks)

Answer the following about the node you have made for “Ticket Sales.” (5 marks)

i. What is the node type?

ii. What is the expression that should be entered in the NPT for the “Ticket Sales” node? Drag and drop the correct expressions into the NPT provided. (5 marks)

Concert	False			True		
Pandemic Restrictions	None	Decreased Capacity	Concerts Banned	None	Decreased Capacity	Concerts Banned
Expressions						

Question 8 (10 marks)

Answer the following about the node you have made for “Concert.”

i. What is the node type? (5 marks)
ii. What is the expression that should be entered in the NPT for the “Concert” node? Drag and drop the correct values into the NPT provided. (5 marks)

Confirm Venue	No			Yes		
Pandemic Restrictions	None	Decreased Capacity	Concerts Banned	None	Decreased Capacity	Concerts Banned
False						
True						

Question 9 (5 marks)

Using the Hybrid Influence Diagram Analyser, determine the optimal decision for the band. What is the total utility of this decision?