ISM-E1004 - Business Analytics 2, Lecture, 8.1.2024-19.2.2024

Started on	Monday, 8 January 2024, 5:58 PM
State	Finished
Completed on	Thursday, 11 January 2024, 10:10 PM
Time taken	3 days 4 hours

Information

Flag question

Problem 1

A commercial aircraft engine manufacturer is predicting the number of engines it will be able to sell to aircraft manufacturers during the next supply season. When airlines do well in terms of passenger demand, they submit orders of aircraft thus the engines will also be in demand. The management team of the company has estimated three scenarios for the demand for commercial aviation: low passenger demand, medium passenger demand, and high passenger demand, and that the probabilities of these scenarios are, correspondingly, 30%, 55%, and 15%. The team has estimated that it can sell a certain number of engines in each scenario, indicated by the column labels in the table below. These estimates are conditional on the given scenario being realized. The probabilities attached to each scenario are given in the cells of the table below and they can be interpreted as conditional probabilities P(engines sold = i | passenger demand = j).

	Engines sold (i)				
	0 – 1000	**1001 – 2000**	**2001 – 3000**	**3001 – 4000**	**4001 – 5000**
j* = *'low'	0.5	0.3	0.1	0.05	0.05
j* = *'medium'	0.1	0.2	0.4	0.2	0.1
j* = *'high'	0	0.1	0.4	0.4	0.1

Question 1

Flag question Marked out of 2.00 Complete

Answer template for Problem 1

Use the Excel template to answer the following questions.

Compute the joint probability distribution, i.e., the 15 different probabilities P(passenger demand=j and engines sold=i).

What is the marginal probability distribution for the number of engines sold, i.e., what are the probabilities P(i='0 – 1000'), ..., P(i='4001 – 5000')?

problem1_Nguyen Xuan Binh_887799.xlsx

Question 2

Flag question Marked out of 1.00 Complete

Use the input area below to answer verbally the following question.

We know the probability for selling 1001–2000 engines, and we also know the probability for the scenario 'medium passenger demand'. Why is the conditional probability P(engines sold = 1001–2000 | passenger demand = medium) not equal to the multiplication of P(engines sold = 1001–2000) and P(passenger demand = medium)?

The conditional probability P(engines sold = 1001–2000 | passenger demand = medium) is not the multiplication of P(engines sold = 1001–2000) and P(passenger demand = medium) because the first thing is that the probability of selling that number of engines given that we already know the demand is medium. It's a specific scenario. The multiplication of the two separate probabilities would imply that the number of engines sold and the passenger demand are independent of each other, which they are not in this case. The number of engines sold is dependent on what the passenger demand is; hence, we cannot multiply the two probabilities to get the conditional probability.

Therefore, we can say that P(engines sold = 1001–2000 AND passenger demand = medium) = P(engines sold = 1001–2000 | passenger demand = medium) * P(passenger demand = medium), not P(engines sold = 1001–2000 AND passenger demand = medium) = P(engines sold = 1001–2000) * P(passenger demand = medium)

Information

Flag question

Problem 2

The table in the Excel answer template for this problem presents actual data on the annual returns of five financial industry portfolios. These portfolios have been structured by assigning each stock in NYSE, AMEZ, and NASDAQ to one of these portfolios based on its four-digit Standard Industrial Classification (SIC) code.

Question 3

Flag question Marked out of 4.00 Complete

Consider a probability model that has (i) a discrete sample space whose possible outcomes correspond to samples of observed annual returns (each row in the table corresponds to one outcome), and (ii) a probability measure that assigns an equal probability to each outcome. Use this probability model to compute probabilities for the following events:

a) Annual return of Hlth is above 35%
b) Annual return of HiTec is below 7%
c) Annual returns of Cnsmr and Manuf are both negative
d) Annual returns of Cnsmr and Manuf are either both positive or both negative

Answer template for Problem 2

Use the answer template to formulate answers to all parts and submit below.

HINT: All data are measured in %. You can use the Excel functions "IF", "AND", and "OR" to identify outcomes that belong to each event.

problem2_Nguyen Xuan Binh_887799.xlsx

Information

Flag question

Problem 3

Brown Automotive produces flux capacitors. Historical quality data is unflattering for Brown Automotive: it shows that seven of out ten flux capacitors coming out of the assembly line are defective and need to be reworked. In the past, the company has used an expensive but extremely reliable diagnostic procedure to test whether a flux capacitor has a defect.

To cut costs, the company's CTO is considering using a new test that is cheaper but not as reliable as the old one. Piloting data using the new test suggests that if a flux capacitor has a defect, the probability that the test declares that the product indeed has a defect is 77% (true positive rate). If the flux capacitor is properly working, the test declares it does not have a defect with a 91% probability (true negative rate).

Help the management at Brown Automotive decide whether to deploy the new test by answering the following questions.

Answer the three questions below by giving the mathematical formulas and a numerical answer. This problem does not require submitting a file, only text based answers.

Question 4

Flag question Marked out of 1.00 Complete

What is the probability that the new test declares that a randomly selected flux capacitor does not have a defect (i.e., what is the share of finished products that pass the new test)?

```
Since Python is close to text description, I can use it directly for this exercise

# Given these probabilities
P_defect = 0.7  # Probability that a flux capacitor is defective
P_not_defect = 1 - P_defect  # Probability that a flux capacitor is not defective
P_TP = 0.77  # Probability of true positive: the test declares a defective product as defective
P_TN = 0.91  # Probability of true negative: the test declares a non-defective product as not defective
P_FP = 1 - P_TN # Probability of false positive: the test incorrectly declares a non-defective product as defective
P_FN = 1 - P_TP # Probability of false negative: the test incorrectly declares a defective product as not defective

# We can use the Law of Total Probability, considering both cases where the capacitor is actually defective and non-defective

# Total probability that the test declares a flux capacitor as non-defective
P_test_not_defect = (P_defect * P_FN) + (P_not_defect * P_TN)

# The numerical answer is P_test_not_defect = 0.434
```

Question 5

Flag question Marked out of 1.00 Complete

What is the probability that a flux capacitor that did not pass the new test (and would thus be reworked) is working properly?

```
# Given these probabilities
P_defect = 0.7  # Probability that a flux capacitor is defective
P_not_defect = 1 - P_defect  # Probability that a flux capacitor is not defective
P_TP = 0.77  # Probability of true positive: the test declares a defective product as defective
P_TN = 0.91  # Probability of true negative: the test declares a non-defective product as not defective
P_FP = 1 - P_TN # Probability of false positive: the test incorrectly declares a non-defective product as defective
P_FN = 1 - P_TP # Probability of false negative: the test incorrectly declares a defective product as not defective

# We want to find the probability that a flux capacitor is actually not defective given that it did not pass the test
# This is a conditional probability that can be found using Bayes' theorem

# First, find the probability that the test declares a flux capacitor as defective (either correctly or as a FP)
P_test_defect = (P_defect * P_TP) + (P_not_defect * P_FP)

# Now, apply Bayes' theorem to find the conditional probability
# We want to find P(Not Defective | Test Defective), which is equal to
# P(Test Defective | Not Defective) * P(Not Defective) / P(Test Defective)

# Now apply Bayes' theorem to find the conditional probability
P_not_defect_given_test_defect = (P_FP * P_not_defect) / P_test_defect

# The numerical answer is P_not_defect_given_test_defect = 0.04770318021201413
```

Question 6

Flag question Marked out of 1.00 Complete

What is the probability that a flux capacitor that passed the test (and would be shipped to customers) is working properly?

```
# Given these probabilities
P_defect = 0.7  # Probability that a flux capacitor is defective
P_not_defect = 1 - P_defect  # Probability that a flux capacitor is not defective
P_TP = 0.77  # Probability of true positive: the test declares a defective product as defective
P_TN = 0.91  # Probability of true negative: the test declares a non-defective product as not defective
P_FP = 1 - P_TN # Probability of false positive: the test incorrectly declares a non-defective product as defective
P_FN = 1 - P_TP # Probability of false negative: the test incorrectly declares a defective product as not defective

# Probability that a capacitor is non-defective given that it passed the test can be found using Bayes' theorem

# First, we calculate the probability of a capacitor passing the test, which we already calculated: P_test_not_defect = 0.434 (Part a)

# Now, apply Bayes' theorem to find the probability of a capacitor being non-defective given that it passed the test
# We want to find P(Non-Defective | Test Non-Defective), which is equal to
# P(Test Non-Defective | Non-Defective) * P(Non-Defective) / P(Test Non-Defective)

P_not_defect_given_test_not_defect = (P_TN * P_not_defect) / P_test_not_defect

# The numerical answer is P_not_defect_given_test_not_defect = 0.6290322580645162
```

Information

Flag question

Problem 4

A real estate developer is considering which of two alternative investment opportunities to pursue. However, the profits depend on what happens to the real estate markets in the future. To support decision making, it has constructed six scenarios for real estate prices. The profits (in $M) from both investments under each scenario are illustrated in the table below. The probabilities for scenarios 1-6 are **0.05, 0.15, 0.30, 0.35, 0.10** and **0.05**, respectively.

Price Scenarios:	1. crash	2. down	3. stable	4. small up	5.big up	6.rocketing
Investment A	-10	0	10	20	20	50
Investment B	-5	-5	0	30	40	70

Question 7

Flag question Marked out of 4.00 Complete

Help the management team by illustrating both the **probability density function (PDF)** and the **cumulative distribution function (CDF)** of profits for both investments.

HINT: You can draw by hand, using Excel, or using Python or R. Collate both graphs on a single page and **submit as a .pdf file**. Marks will be awarded both from correctness and clarity of the graphs.

Problem4_Nguyen Xuan Binh_887799.pdf

Information

Flag question

Problem 5

The purpose of this problem is to demonstrate methods that can be useful in deriving properties for probability distributions – in particular, the use of the inverse CDF method and Monte Carlo simulation to estimate expectations, event probabilities and conditional expectations.

Customers arrive in a bank so that the time between two customer arrivals is exponentially distributed with a mean of 3 minutes, i.e., X ~ Exp(λ), where λ = 13.

Using the Excel template for this problem, set up a Monte Carlo simulation to generate a sample of 1000 realizations from Exp(1/3) using the inverse CDF method. The CDF of an exponentially distributed X is $F_X(x) = 1 - e^{-\lambda x}$. As the first step you need to derive the inverse of the CDF.

Question 8

Flag question Marked out of 5.00 Complete

a) Check the correctness of your implementation by estimating E[X] (should be about 3), and by comparing the histogram of the sample ("Data"-tab -> "Data Analysis" -> Histogram) to the exponential distribution's PDF (see lecture slides).

b) What is the probability that the time between two customer arrivals is less than seven minutes, i.e., P(X<7)?

c) Assume that the store owner has waited seven minutes without any customer arrivals: What is the expected time she still has to wait before the next arrival, i.e. E[X|X>7]-7?

Answer template for Problem 5

Use the Excel template to answer all the parts. Supply the answers in the green cells using formulas, i.e. do not insert hardcoded numerical values in the green cells.

problem5_Nguyen Xuan Binh_887799.xlsx

Information

Flag question

Problem 6

A hotel has 100 rooms for its guests, but it is using two different rates: (i) a full price of 190 euros and (ii) a discount price of 175 euros available for customers who pay for the room one week in advance. This helps to distinguish between leisure travelers who tend to book early, and flexibility-valuing business travelers who book late.

The average daily demand by business travelers is normally distributed with a mean value of 60 rooms and standard deviation of 30 rooms. It is assumed that all rooms offered at the discount price will be sold.

Question 9

Flag question Marked out of 6.00 Complete

Build a Monte Carlo simulation model that determines the expected revenue of a single day when the number of rooms reserved to be sold at full price is fixed to Q. Implement the model so that changing the value of Q is easy. Use the inverse CDF method to sample the normally distributed random demand by business travelers.

a) What is the expected revenue if Q is equal to the mean demand?

b) How many rooms Q* would you recommend to sell at a full price to maximize the expected revenue? What is the maximum expected revenue?

Set Q = Q* and answer the following questions:

c) What is the probability that the hotel is not fully booked? What is the expected revenue on a day when the hotel is not fully booked (expected revenue conditional on the event that the hotel is not fully booked)?

Answer template for Problem 6

Use the Excel answer template to formulate answers to all parts. The Excel file that you submit must include the the correct answers in the green cells as well as the simulations. The answers must be based on the simulation results.

problem6_Nguyen Xuan Binh_887799.xlsx

Information

Flag question

Problem 7

This problem discusses queueing models that are a common modelling application in business analytics and highly useful for the analysis of operational characteristics of services, such as mean waiting times, server utilisation, and system throughput.

Service businesses are increasingly employing automation, but it is not always clear whether automated service operations actually improve key operational metrics. Here we consider a fast-food chain that is replacing human servers with self-service kiosks in its restaurants and needs to decide what is the optimal number of kiosks to use. Each kiosk takes up space, therefore the main focus is on whether one kiosk is sufficient or should two kiosks be deployed.

The model of the system consists of the kiosk(s) and customers who are either waiting to access a kiosk or being served by a kiosk. We can assume that the inter-arrival times are exponentially distributed with mean 2.5 minutes. The supplier of the self-service kiosks advises that service times are normally distributed with mean 2 minutes and a standard deviation of 0.5 minutes.

Question 10

Flag question Marked out of 6.00 Complete

Build a simulation model using **one** self-service kiosk and use it to investigate the following questions.

a) What is the probability that the kiosk is not free when a new customer arrives?

b) What is the mean waiting time in minutes?

Expand the simulation model to include two self-service kiosks, assuming that a new arriving customer will be served by the kiosk that frees up first. Use this simulation model to answer the following question.

c) What is the mean waiting time (in minutes) with **two** kiosks?

You can formulate this problem using Jupyter Notebook and either R or Python using one of these templates:
Answer template (R) / Answer template (Python) (right click and save file)

Use the .ipynb answer template to formulate the simulation model using Jupyter Notebook and R or Python. Your answer file will be machine read so please follow the instructions in the template. The answers to the questions should be based on simulation results.

problem7_Python_Nguyen Xuan Binh_887799.ipynb

Question 11

Flag question Marked out of 2.00 Complete

In the simulation, you compared one-kiosk and two-kiosk systems. Is the system throughput increased when two kiosks are used, as compared to a one-kiosk system? Justify your answer briefly.

Yes, the system throughput increased a lot when two kiosks are used compared to a one-kiosk system. This is supported by the significant reduction in mean waiting time as observed in my simulation results (from 4.2 minutes to 0.21 minutes).

With two kiosks, the system's capacity to serve customers doubles. Since there are more service points available for customers, it is unlikely that the customers need to wait very long time. In a queueing system like this, shorter waiting times means higher throughput, as more customers are being served in a given period.

Finish review