Homework #1

Due on February 11th, 2020
Knowledge Discovery & Data Mining
CS513B−Spring 2020
Professor Khasha Dehnad

Daniel Kadyrov

Problem 1

Jerry and Susan have a joint bank account. Jerry goes to the bank 20% of the days. Susan goes there 30% of the days. Together they are at the bank 8% of the days.

a. Susan was at the bank last Monday. What is the probability that Jerry was there too?

b. Last Friday, Susan was not at the bank. What's the probability that Jerry was there?

c. Last Wednesday at least one of them was at the bank. What is the probability that both of them were there?

Solution

	Susan at Bank	Susan not at Bank	
Jerry at Bank	**8**	12	**20**
Jerry not at Bank	22	58	80
	30	70	100

a. $P(\text{Jerry at Bank}/\text{Susan at Bank}) = 8\%/30\% = 26.67\%$

b. $P(\text{Jerry at Bank}/\text{Susan not at Bank}) = 12\%/70\% = 17.14\%$

c. $P(\text{Jerry} \cap \text{Susan at Bank}/\text{Jerry} \cup \text{Susan at Bank}) = 8\%/(1 - 58\%) = 19.05\%$

Problem 2

Harold and Sharon are studying for a test. Harold's chances of getting a "B" are 80%. Sharon's chances of getting a "B" are 90%. The probability of at least one of them getting a "B" is 91%.

a. What is the probability that only Harold gets a "B"?

b. What is the probability that only Sharon gets a "B"?

c. What is the probability that both won't get a "B"?

Solution

	Sharon "B"	Sharon not "B"	
Harold "B"	79	1	**80**
Harold not "B"	11	9	20
	90	10	100

$$\begin{aligned} P(\text{Harold and Sharon}) &= P(\text{Harold}) + P(\text{Sharon}) - P(\text{Harold} \cup \text{Sharon}) \\ &= 80\% + 90\% - 91\% = 79\% \end{aligned}$$

a. $P(\text{Only Harold "B"}) = P(\text{Harold "B"}) - P(\text{Harold "B"} \cap \text{Sharon not "B"}) = 80\% - 79\% = 1\%$

b. $P(\text{Only Sharon "B"}) = P(\text{Sharon "B"}) - P(\text{Sharon "B"} \cap \text{Harold not "B"}) = 90\% - 79\% = 11\%$

c. $P(\text{Harold} \cap \text{Sharon not "B"}) = 1 - 91\% = 9\%$

Problem 3

Jerry and Susan have a joint bank account. Jerry goes to the bank 20% of the days. Susan goes there 30% of the days. Together they are at the bank 8% of the days. Are the events "Jerry is at the bank" and "Susan is at the bank" independent?

Solution

By definition, two events are independent if and only if:

$$P(A \text{ and } B) = P(A) \times P(B) \tag{1}$$

In the situation described within the problem statement:

$$P(A \text{ and } B) = 8\% \neq 20\% \times 30\% = 60\%$$

Therefore, the statements "Jerry is at the bank" and "Susan is at the bank" are dependent.

Problem 4

You roll 2 dice.

a. Are the events "the sum is 6" and "the second die shows 5" independent?

b. Are the events "the sum is 7" and "the first die shows 5" independent?

Solution

a. The events "the sum is 6" and "the second die shows 5" are dependent.

$$P(\text{Sum is 6}) = n(\{(3,3),(2,4),(4,2),(1,5),(5,1)\})/36 = 5/36$$

$$P(\text{Second die is 5}) = n(\{(1,5),(2,5),(3,5),(4,5),(5,5),(6,5)\})/36 = 6/36$$

$$P(\text{Sum is 6} \cap \text{second die is 5}) = n(\{(1,5)\})/36 = 1/36$$

$$1/36 \neq 5/36 \times 6/36 = 30/1296 = 5/216$$

b. The events "the sum is 7" and "the first die shows 5" are dependent.

$$P(\text{Sum is 7}) = n(\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}) = 6/36$$

$$P(\text{First die is 5}) = 6/36$$

$$P(\text{Sum is 7} \cap \text{first die is 5}) = 1/36$$

$$1/36 \neq 6/36 \times 1/36 = 1/216$$

Problem 5

An oil company is considering drilling in either TX, AK and NJ. The company may operate in only one state. There is 60% chance the company will choose TX and 10% chance - NJ. There is 30% chance of finding oil in TX, 20% - in AK, and 10% - in NJ.

1. What's the probability of finding oil?
2. The company decided to drill and found oil. What is the probability that they drilled in TX?

Solution

	TX	AK	NJ	
Oil	18	6	1	25
No Oil	42	24	9	75
	60	30	**10**	100

$$P(\text{Oil}|\text{TX}) = 30\% = P(\text{Oil} \cap \text{TX})/P(\text{TX}) = P(\text{Oil} \cap \text{TX})/60\%$$
$$P(\text{Oil} \cap \text{TX}) = 30\% \times 60\% = 18\%$$
$$P(\text{Oil}|\text{NJ}) = 10\% = P(\text{Oil} \cap \text{NJ})/P(\text{NJ}) = P(\text{Oil} \cap \text{NJ})/10\%$$
$$P(\text{Oil} \cap \text{NJ}) = 10\% \times 10\% = 1\%$$
$$P(\text{Oil}|\text{AK}) = 30\% = P(\text{Oil} \cap \text{AK})/P(\text{AK}) = P(\text{Oil} \cap \text{AK})/20\%$$
$$P(\text{Oil} \cap \text{AK}) = 30\% \times 20\% = 6\%$$

1. $P(\text{Oil}) = 25\%$
2. $P(\text{TX}/\text{Oil}) = 18\%/25\% = 0.72\%$

Problem 6

The following tables show the survival status of individual passengers on the Titanic:

	1st	2nd	3rd	Crew	Sub Total
Adult	197	94	151	212	654
Child	6	24	27		57
Sub Total	203	118	178	212	711

Table 1: Survived

	1st	2nd	3rd	Crew	Sub Total
Adult	122	167	476	673	1438
Child			52		52
Sub Total	122	167	528	673	1490

Table 2: Not Survived

	1st	2nd	3rd	Crew	Sub Total
Adult	319	261	627	885	2092
Child	6	24	79		109
Sub Total	325	285	706	885	2201

Table 3: Total

Use this information to answer the following questions:

1. What is the probability that a passenger did not survive?
2. What is the probability that a passenger was staying in the first class?
3. Given that a passenger survived, what is the probability that the passenger was staying in the first class?
4. Are survival and staying in the first class independent?
5. Given that a passenger survived, what is the probability that the passenger was staying in the first class and the passenger was a child?
6. Given that a passenger survived, what is the probability that the passenger was an adult?
7. Given that a passenger survived, are age and staying in the first class independent?

Solution

Assuming that passenger includes 1st, 2nd, 3rd, and crew members since the prompt states "survival status of individual passengers" when describing the table.

1. $P(\text{Not Survived}) = 1490/2201 = 67.69\%$
2. $P(\text{First Class}) = 325/2201 = 14.77\%$
3. $P(\text{Survived} \cap \text{First Class}) = 203/711 = 28.55\%$
4. Surviving and staying in first class are dependent.

$$P(\text{Survived}) = 711/2201 = 32.30\%$$

$$P(\text{First Class}) = 325/2201 = 14.77\%$$

$$P(\text{Survived} \cap \text{First Class}) = 203/711 = 28.55\% \neq 711/2201 \times 325/2201 = 4.77\%$$

5. $P(\text{Survived} \cap \text{First Class} \cap \text{Child}) = 6/711 = 0.84\%$
6. $P(\text{Survived} \cap \text{Adult}) = 654/711 = 91.98\%$
7. Given that a passenger survived, age and staying in the first class are dependent.

$$P(\text{Survived} \cap \text{Adult}) = 654/711 = 91.98\%$$

$$P(\text{Survived} \cap \text{First Class}) = 203/711 = 28.55\%$$

$$P(\text{Survived} \cap \text{Adult} \cap \text{First Class}) = 197/711 = 27.70\%$$

$$27.70\% \neq P(\text{Survived} \cap \text{First Class}) \times P(\text{Survived} \cap \text{Adult} \cap \text{First Class}) = 26.26\%$$