PROBABILITY THEORY AND INTRODUCTORY STATISTICS

ALY6010, FALL 2019

MODULE 2 PROJECT ASSIGNMENT

PERFORMANCE ANALYSIS, PROBABILITY AND PROBABILITY TREES

SUBMITTED BY: SHIVANI ADSAR, PARINITA MAHINDRAKAR

NUID: 001399374, 001069853

CRN: 71447

SUBMITTED TO: DR. DEE CHILUIZA REYES

DATE: 09/29/2019

Introduction

The assignment aims at performing analysis using probability and probability tree. The project includes information about Quality, Speed and Cost of 50 recent projects selected by the quality control manager of a manufacturing company. Using various criteria for the given data, we can represent the probability of the events occurring and present them in tabular format, Venn diagram and probability trees. The data set is displayed in the form of percentage and probability distribution which represents the chances of the events happening based on the given criteria.

Analysis

Part 1: Using the data in excel sheets

Q1. Probability of events for the company's projects based on the attributes

Below two tables are based on the display of probability notation, count and percentage for 3 criteria. The calculations in Table 1 represent the counts, probability/percentage and probability notation based on the conditions of attributes given in the data of the company's projects. The data in Table 1 indicates count of the values where the criteria satisfies Quality score greater than 550, Speed less than 11 days and Cost is less than $180,000.

Table 1: Probability Distribution using the attributes for the company's projects

	Total PIP	50	
Conditions	Probability Notation	Count	Probability / Percentage
Event Q: Quality is satisfied if Quality Score > 550	P(Q)	20	40.00%
Event S: Speed is satisfied if Speed < 11 days	P(S)	22	44.00%
Event C: Cost is satisfied if Cost < $180,000	P(C)	17	34.00%

Table 1: Probability Distribution using the attributes for the company's projects

Tables provide an easier way of grouping information and displaying comparison between different conditions. The usage of the COUNTIF function enables us to count the cells having a single and multiple criterion. Also, calculation of percentage for the values is easier using a tabular format. It is observed that the count of events when all the three criteria are satisfied is very small that is **3**. Being the manager of the company, I would try and maximize the quality and minimize speed and cost to gain more profit.

Table 2: Probability Distribution as per the assigned scores by the manager for the company's projects

Description	Conditions	Probability Notation	Counts	Probability / Percentage
None of the three criteria is satisfied.	P(Score = 0)	P(QUSUC)'	7	14.00%
Quality criterion is satisfied but not the others	P(Score = 1)	P(S'∩C'∩Q)	11	22.00%
Speed criterion is satisfied but not the others	P(Score = 2)	P(Q'∩C'∩S)	12	24.00%
Cost criterion is satisfied but not the others	P(Score = 3)	P(Q'∩S'∩C)	7	14.00%
Quality and Speed are satisfied but the Cost is not	P(Score = 4)	P[(QUS)∩C']	3	6.00%
Quality and Cost are satisfied but the Speed is not	P(Score = 5)	P[(QUC)∩S']	3	6.00%
Speed and Cost are satisfied but the Quality is not	P(Score = 6)	P[(SUC)∩Q']	4	8.00%
All three criteria are satisfied	P(Score = 7)	P(Q∩S∩C)	3	6.00%
		TOTAL:	**50**	**100.00%**

Table 2.: Probability Distribution as per the assigned scores for the company's projects

Table 2 represents probabilities based on union, intersection and complement on the given conditions for Quality, Speed and Cost. The highest probability of **24%** is observed when the speed criteria is satisfied but not Cost and Quality. This means that 12 out of 50 projects usually accomplish completion within the given timeframe of 11 days but not in the required Speed and Cost.

Q2. Venn Diagram to represent the relationships among Quality, Speed & Cost

Venn diagrams are used to illustrate similarities and differences between various data sets. Below Venn diagram displays relationship between Cost, Quality and Speed of 50 different projects. It is observed that, 11 projects satisfy the Quality criteria, 12 projects satisfy the Speed and 7 projects satisfy Cost criteria individually. The intersection of all the three criteria shows that all the three are satisfied by only 3 projects. Moreover by taking the intersection, it is concluded that, 3 projects satisfy Quality and Speed criteria but not Cost, 3 projects satisfy the Quality and Cost criteria but not Speed and 4 projects satisfy the Speed and Cost criteria but not Quality.

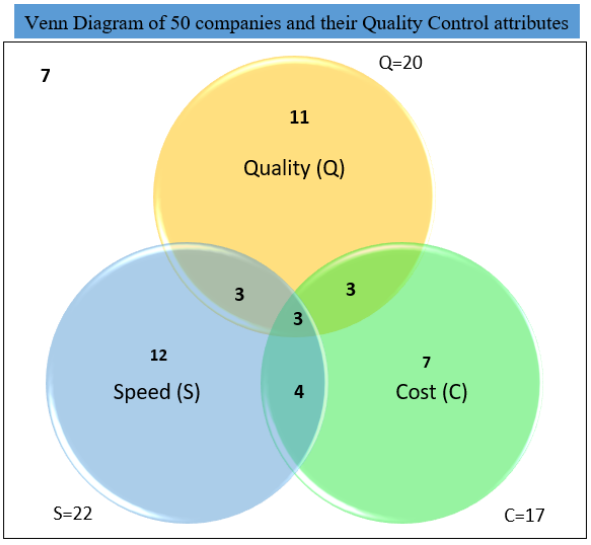

Figure 1: Venn Diagram of 50 projects

Q3. Table displaying Union, Intersection and Complement of attributes

Using the concepts of intersection, union and complement for the given data, we have calculated the probabilities for each event. It is observed from the table that the percentage of attributes who satisfied exactly one of the three criteria for the attributes is the highest ie. 60%. The percentage satisfying quality and speed out of those satisfying the cost is the lowest ie. 9.09%. This analysis helps us to understand about the attributes that should be increased so that the company is benefited. For illustration, consider that in a class of 50 students, there are 20 students who have passed the Physics exam and failed in Maths, 10 who passed the Biology exam and 20 who passed the Chemistry exam. Using the concept of intersection, union and complement of events, we can find out the number of students who failed the Maths exam.

Table 3: Analysis for the probabilities of attributes in the company's project

Description	Probability Notation	Count	Probability/Percentage
a) Of those who satisfied Speed, what percentage also satisfied Quality?	P(S∩Q)	6	27.27%
b) Of those who satisfied Quality, what percentage also satisfied Cost?	P(Q∩C)	6	30.00%
c) Of those who satisfied Speed, what percentage also satisfied Quality but did not satisfy the Cost?	P(S∩Q∩C')	3	13.64%
d) Of those who satisfied Cost, what percentage also satisfied Quality but did not satisfy theSpeed?	P(C∩Q∩S')	3	17.65%
e) Of those who did not satisfy Cost, what percentage satisfied Quality and Speed?	P(C'∩[S∩Q])	3	9.09%
f) What percentage satisfied exactly one of the three criteria?	P(Q'∩S'∩C) + P(Q'∩C'∩S) + P(S'∩C'∩Q)	30	60.00%
g) Of those who satisfied at least two of the three criteria, what percentage satisfied exactly two criteria?	P(Q∩S∩C') + P(Q∩S'∩C) + P(Q'∩S∩C)	10	76.92%
h) Of those who did not satisfy Cost, what percentage satisfied the Quality criterion?	P(Q∩C')	14	42.42%

Table 3: Analysis for the probabilities of attributes in the company's project

Part 2: Probability Trees

Q4. Probability Tree for a sample a people visiting the MD office

Probability trees are used to represent a series of independent events or conditional probabilities. Figure 2. shows the probabilities of men and women with and without health insurance who visited the MD office less than 5 times a year and more than 5 times a year.

Probability Tree for Survey regarding insurance coverage

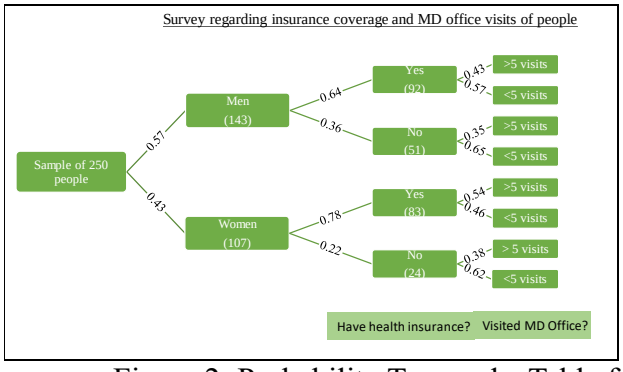

Table 4: Probability Distribution for insurance coverage

	With health insurance		Without health insurance			
	> 5 time a year	< 5 times a year	> 5 time a year	< 5 times a year	Totals	Probabilities
Men	40	52	18	33	**143**	**0.57**
Women	45	38	9	15	**107**	**0.43**
Totals	**85**	**90**	**27**	**48**	**250**	**1**
Probabilities	**0.34**	**0.36**	**0.11**	**0.19**	**1**	

Figure 2: Probability Tree and a Table for the people visiting MD office

The decision tree displays, from a sample of 250 people, there are 143 men and 107 women. Every branch denotes the probability of the event occurring. As observed, probability of people with health insurance visiting the MD office greater than 5 times a year is 34% and less than 5 times a year is 36%. The probability of people without the health insurance, visiting the MD office more than 5 times a year is 11% and less than 5 times a year is 19%.

Questions

1 Probability of a person who visited MD Office less than 5 times or is a women. = 138/250+107/250-53/250 = 0.77 or 77%

2 Probability of a man without health insurance and visited MD office more than 5 times. = 18/(18+33)= 0.35 or 35%

Q5. Probability Tree for a survey of 120 young people

Figure 3. shows a decision tree and the probability of people with and without the health insurance visiting the MD office. The advantages of using this method is, we understand the number of people who are visiting the MD office with and without health insurance. This will benefit the insurance company, in providing the health insurance to the required people based on their probabilities. The disadvantages of this method is that it may not give accurate results in determining the people who do not have health insurance and did not visit the MD office.

Probability Tree for Survey regarding insurance coverage

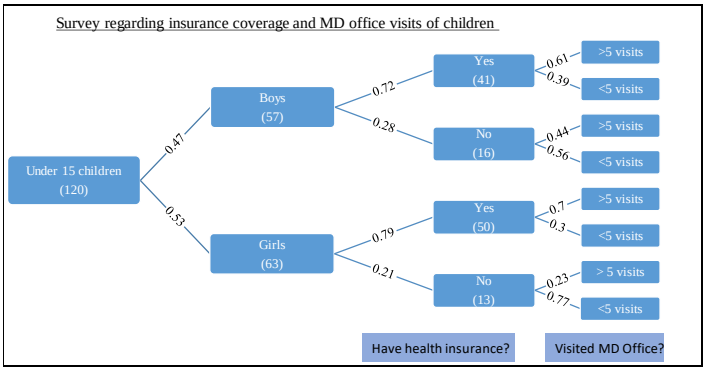

Table 5: Probability Distribution for insurance coverage

	With health insurance		Without health insurance			
	> 5 time a year	< 5 times a year	> 5 time a year	< 5 times a year	Totals	Probabilities
Boys	25	16	7	9	**57**	**0.48**
Girls	35	15	3	10	**63**	**0.53**
Totals	**60**	**31**	**10**	**19**	**120**	**1**
Probabiliti	**0.50**	**0.26**	**0.08**	**0.16**	**1**	

Figure 3: Probability Tree and Table for the survey

Conclusion

1. It is inferred from the observations of company's dataset that using the probability distribution helps us to predict the data about the chances of events happening.
2. The usage of probability trees is beneficial as it creates a comprehensive analysis of the consequences of each node and predict the outcomes stepwise.
3. The Venn diagrams use overlapping shapes to show the relationships among two or more sets of data using intersection and union. Using the Venn diagram and the probability tree, it was easier to predict the profitable attribute to the company.
4. These methods of predicting probability of events, have helped us understand the usage of methods to solve problems.

Reference

1. Jaynes, E.T., Bretthorst, G.L., 2017. *Probability theory: the logic of science. Cambridge University Press, Cambridge (UK).*

2. Bluman, A.G. (1992). *Elementary Statistics.* Retrieved from https://archive.org/details/ElementaryStatistics/page/n5