# STATISTICAL DISCRETE AND CONTINUOUS PROBABILITY DISTRIBUTION TABLE RE-PRODUCING USING R

NUR HAZEEMA BINTI ABDUL RAHMAN

05-4506832 pustaka.upsi.edu.my Perpustakaan Tuanku Bainun Kampus Sultan Abdul Jalil Shah PustakaTBainun ptbupsi

FACULTY OF SCIENCE AND MATHEMATICS

SULTAN IDRIS EDUCATION UNIVERSITY 2023

05-4506832 pustaka.upsi.edu.my Perpustakaan Tuanku Bainun Kampus Sultan Abdul Jalil Shah PustakaTBainun ptbupsi

## ACKNOWLEDGEMENT OF ORIGINALITY

I hereby declare that the Final Year Project entitled Statistical Discrete and Continuous Probability Distribution Table Re-producing Using R is the product of my own work and all the assistance received in completing this thesis and sources have been acknowledged.

Name : NUR HAZEEMA BINTI ABDUL RAHMAN



Date : 28 February 2023

Signature :

# ACKNOWLEDGEMENT

I would like to express my highest gratitude to God because of His' willing, I managed to complete my writing in Final Year Project or can be known as thesis. During carrying out the research, I have been received the guidance and cooperation from all parties that involved. By writing this acknowledgement, I have my opportunity to wish my sincere thank you to all of those with whom I have the pleasure to work during conducting this research.

First of all, I am pleased to give my warmest thank you to my beloved supervisor, Madam Noorazrin Binti Abdul Razak who helps me to make this work possible. Her patience, guidance and advice carried me through all the phases of writing my thesis and I am extremely grateful as well as indebted to her.

Next, I also want to manifest my gratitude to my programme mates who always give a favour and assistances in order to finalise this research especially to my closest friends who is also my housemates. They are Amirah, Amielia, Zahirah, Hafizah, Shikin, Afiqah and also Syaza. Thank you for always being there through thick and thin while completing this research. Most importantly, I would like to express my lovely thank you to my parents who never stop to wish me the best for my achievement in education.

Last but not least, I want to thank me. I want to thank me for believing in me. I want to thank me for doing all this hard work and also thank me for never quitting because I managed to achieve this education level that I never expected.

# STATISTICAL DISCRETE AND CONTINUOUS PROBABILITY DISTRIBUTION TABLE RE-PRODUCING USING R

**By**

**Nur Hazeema Binti Abdul Rahman**

## ABSTRACT

The aim of this research is to implement the statistical software which is R, in one of the Elementary Statistics topics called Probability Distribution. This study has explored the use of R and figured out that R is one of the statistical software that has its own utility to allow R user to create the function. Regarding to the objectives, R software is implemented in the topic of Probability Distribution which is Binomial distribution, Poisson distribution and Normal distribution. The researcher has been asked to explore R software and create as well as apply new functions. These functions are applied in R to produce Binomial distribution table, Poisson distribution table and Standard Normal Distribution table. The finding shows that these probability distribution tables can be produced using created functions where it shows the similar values when it compared to the actual tables of Binomial, Poisson and Standard Normal distribution. It can be clarified that the researcher managed to create functions with regard to reproduce these probability distribution tables by using R software. This research can be the usage to the other researcher or readers where they can learn on how to use R software since it is easy to learn and it does not require high programming skills. For the future research, it can be extended where other distribution tables can be produced by creating own functions in R software.

*Keywords:* R software, Binomial distribution, Poisson distribution, Normal distribution, Standard Normal distribution, Data simulation.

Scope: Statistics

# PENGHASILAN SEMULA JADUAL TABURAN KEBARANGKALIAN BAGI DISKRET DAN SELANJAR MENGGUNAKAN R

**Oleh**

**Nur Hazeema Binti Abdul Rahman**

## ABSTRAK

Tujuan penyelidikan ini adalah untuk melaksanakan perisian statistik iaitu R, dalam salah satu topik Statistik Permulaan yang dipanggil sebagai Taburan Kebarangkalian. Kajian ini telah meneroka penggunaan R dan mendapati bahawa R adalah salah satu perisian statistik yang mempunyai utiliti sendiri untuk membolehkan pengguna R mencipta fungsi tersebut. Selari dengan objektif, perisian R dilaksanakan dalam topik Taburan Kebarangkalian iaitu Taburan Binomial, Taburan Poisson dan Taburan Normal. Penyelidik telah diminta untuk meneroka perisian R dan mencipta serta menggunakan fungsi baharu. Fungsi-fungsi ini digunakan dalam R untuk menghasilkan jadual taburan Binomial, jadual taburan Poisson dan jadual Taburan Normal Piawai. Dapatan menunjukkan bahawa jadual taburan kebarangkalian ini boleh dihasilkan menggunakan fungsi yang telah dicipta di mana ia menunjukkan nilai yang sama apabila dibandingkan dengan jadual sebenar taburan Binomial, Poisson dan Normal Piawai. Jadi, dapat dijelaskan bahawa penyelidik berjaya mencipta fungsi berkenaan dalam menghasilkan semula jadual taburan kebarangkalian tersebut dengan menggunakan perisian R. Penyelidikan ini boleh digunakan oleh penyelidik atau pembaca lain di mana mereka boleh mempelajari cara menggunakan perisian R kerana ia mudah untuk dipelajari dan tidak memerlukan kemahiran pengaturcaraan yang tinggi. Bagi penyelidikan masa depan, ia boleh dilanjutkan di mana jadual taburan yang lain boleh dihasilkan dengan mencipta fungsi sendiri dalam perisian R.

*Kata Kunci:* Perisian R, Taburan Binomial, Taburan Poisson, Taburan Normal, Data simulasi.

Skop: Statistik

# TABLE OF CONTENTS

## LIST OF TABLES

# LIST OF FIGURES

## LIST OF APPENDICES

# CHAPTER 1

# INTRODUCTION



## 1.0 Introduction

This chapter discussed the overview of the research study which includes the background of the study and the problem statement that can be the main reason to conduct this research. This topic also derived several specific research objectives and the important of implementing this research as well as the conclusion at the end of this chapter.

## 1.1 Background of the Research

Probability distributions are a crucial foundational concept in probability which it consists of the names and shapes of probability distributions. According to Viti et al. (2015), probability distribution can be called when each event is linked to its probability where it is a function that can determines all the probabilities values of a random variable with a given range. The structure and type of the probability distribution is depending on the random variable's properties. Probability distribution can be in the form of whether it is discrete or continuous. This can lead how the distribution can be obtained or help in the part of how to calculate the possible outcome and also its probability. Throughout the sciences, probability distributions are commonly used in the real life situation such as Normal distribution which is used in the field of finance, engineering and others (Hayes, 2022). Calculation of confidence intervals around estimated values is one of the usages in probability distribution. They are used to estimate the possible of certain outcomes and to measure and predict its probability. Probability distribution functions are classified into two types: discrete and continuous (Morton and Robert, 2014).

Determination of discrete or continuous probability distribution can be calculated depends on random variables. According to Education Services Australia (2013), a random variable can be either discrete or continuous and also these random variables have distributions. A discrete distribution in statistics field can be defined as a probability distribution of the outcomes of finite variables or countable values according to Gordon (2022). It is made up of discrete random variables. There are

several type of discrete probability distribution, such as discrete Uniform, Binomial, Poisson, Multinomial and Bernoulli probability distribution.

According to Zelterman (2002), the Binomial distribution is one of the most common distributions for discrete or count data. The possible outcome whether it is a successful or unsuccessful in an experiment that is repeated several times can be thought of as a Binomial distribution. Binomial probability distribution is a type of distribution that has two possible outcomes (Glen, 2022). Binomial distribution is obtained by repeating the experiment and plotting the probability each time. As reported in the article by Barone (2022), the common discrete probability distribution is Binomial distribution where it only counts two outcomes whether it is success or fail and also the value given to represents success is 1 and failure is 0. It can be told that a Binomial distribution value is between zero and one.

Furthermore, Poisson distribution is also another distribution in discrete probability distribution. It is applied to experiments with random and independent occurrences. The Poisson distribution is a suitable model for the situations where events seem to occur at random and rare in the nature (Singh et al., 2021). According to Glen (2022), it helps to predict the probability of certain events happening after how often the event has occurred is known. What makes Poisson distribution is differ from Binomial distribution where Poisson distribution shows its probability values in a time interval specifically. Both Binomial and Poisson distributions are two statistical

methods that have been used to calculate discrete random variables and also to assess the possible of the outcomes that occur.

A continuous probability distribution describes the probabilities of a continuous random variable's possible values. It is a random variable with an infinite and uncountable set of possible values. According to Turney (2022), a continuous variable can have any value between its lowest and highest values. Thus, continuous probability distributions include every number in the variable's range. For instance, Normal distribution is one of the continue probability distributions that are common in statistics. Normal distributions are belongs to the family of continuous probability distribution or probability density functions (Musselwhite & Wesolowski, 2018). Random variables in the normal distribution are continuous, meaning they can fall within any range of values. The Normal distribution can be converted to the Standard Normal distribution, *Z* where it is standardized by converting its value into *z* scores. It is a particular Normal distribution where it has the mean value is equal to zero and the value of standard deviation is one. In order to standard a value of Normal distribution, the experimental mean is subtracted from the individual value and the difference obtained is divided by the standard deviation (Bhandari, 2023). That is why Normal distributions are usually compared by putting them on the same scale to obtain the Standard Normal distribution (Brereton, 2014). Therefore, Standard Normal distribution can be obtained by standardizing the values from Normal distributions to *z*-scores.

Simulation studies are suitable tool for students, teachers, researchers or statisticians to comprehend all statistical concepts together with probability distribution. According to Zhang and Maas (2019), the tool of computer simulations is an essential part of understanding statistics. Thus, R programming is one of software that supports simulation programme that can be carried out to understand statistical concepts. A simulation study can be conducted by using R statistical programming since it is an ideal platform provided by R software (R Development Core Team, 2011) since R is a programming language and also it is flexible in the field of statistics. R programming is easy to learn because it does not require high computer skills (Smith as citied in Zhang & Maas, 2019). R is a software that provides an intensive programming environment to perform such as probability distribution calculations. The student who is conduct simulations can help them to solve simple probability problems that cannot be possible with physical experiments (Batanero & Diaz as citied in Koparan & Yilmaz, 2015).

Overall, it is a must to learn and explore on how to run R program in order to produce statistical table including Binomial, Poisson and Standard Normal distribution table. Therefore, the purpose of this research is to create a function to produce these tables by using R software.

## 1.2 Problem Statement

A probability distribution table is a table that presents the probability value that a random variable takes on specified values. Probability distribution table have its

properties that need to be satisfied which is the sum of all probabilities is equal to 1 and also the mean as well as standard deviation can be calculated. Probability distribution can be represented by using the equations and table of variable values and probabilities (Frost, 2018). These statistical tables displaying the values of the probability density function (PDF), cumulative distribution functions (CDF), or probability functions. It can be differentiated according to common probability distribution such as Binomial, Poisson and Normal distribution. Since they have different parameters of the distribution, as the result, different value of probability can be obtained and it also provided with different significance levels. Additionally, R has these in-built functions that allow R user to create own function and most importantly, probability distribution is supported by these in-built functions. R is a great functional programming language since it is an open source and also accessible for all standard working frameworks including constructing probability distribution tables.

However, probability distribution table includes discrete and continuous have been reproduced directly from the output of digital computers (Owen, 1962). This, of course, need to be checked several times to ensure the probability value is correct according to the substitution into the equation and also its function. He said that there was an error when it comes to the severe process. He also added that it did not allow the flexibility in set-up of the table that is possible with typescript, but it undoubtedly eliminated many transcription errors. Therefore, the use of R in reproducing statistical table can guarantee the value in the table mentioned. Apart from that, the students only know how to use the discrete and continuous distribution table, yet they did not know how to generate the table by using R (Frost, 2018). Then, this is an opportunity

for students to try and generate the table by creating their own function using R. This will increase their thinking skills about the probability distribution and not including its table only, but also enhance the knowledge about the distribution. Moreover, the importance of the use of discrete and continuous distribution for data analysing to the researchers are become demanding in the research field. According to Ramaekers and Janssens (2008), it is still important to know the shape of the distribution since the probability distribution of demand is an important characteristic in inventory management re-order point models. Based on their previous study, it is clear that models have been created for the demand incidence using a Poisson distribution. Not only that, normal distribution is also be recommended since the distribution is defined on both positive and negative axes and it is also symmetrical shape (Silver & Peterson as citied in Ramaekers & Janssens, 2008). Therefore, it is important of using discrete and continuous distribution to analyse data among the researchers.

Consequently, it is important for the researcher to learn on how to produce probability distribution table by using R. This is because the researcher can give acknowledgement about how R software runs, not only used for data analysing. In this context, R has its in-built function that allow R user to create and construct the function to produce something so that the objective of this research can be achieved.

## 1.3 Research Objectives

1.3.1 To explore R software.

1.3.2 To build and apply new functions to produce Binomial, Poisson and Standard Normal probability distribution tables by using R software.

## 1.4 Conceptual Framework of Research

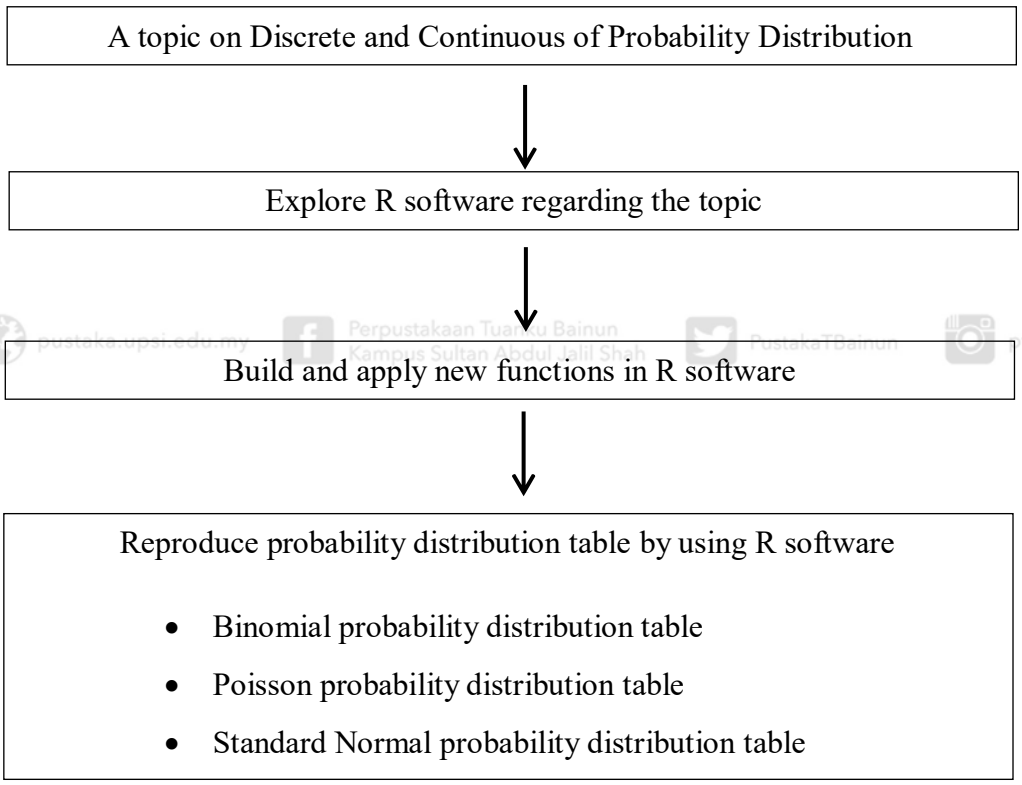


Figure 1.1 Research Frameworks

Based on Figure 1.1, the exploration and also its application of R software will focus on a topic of statistical probability distribution table which includes discrete and

continuous probability distribution. For discrete probability distribution, it involves Binomial and Poisson probability distribution table. Meanwhile, continuous probability distribution involves Standard Normal probability distribution table only. Next step is to create and build new functions to produce these distribution tables by using R software. The output of R will be compared to the actual distribution table to ensure the function created is efficient. Lastly, discussions and conclusion of the research will be discussed at the end of this research.

## 1.5 Significance of Research

Understanding how R programming works is the crucial thing in increasing the level of sophistication of using R programming. According to the first objective of this study, it motivates the researcher to study R regarding to the topic of probability distribution. For this reason, it can encourage the readers to use R by itself and see how R software works. This will increase readers' programming skills indirectly and it also can be an advantage for the reader to have knowledge about R software. Regarding to the second objective, it helps the reader to try and build a new function by itself in order to reproduce another probability distribution table such as $t$ distribution. It also can be the reference for other researcher that put an interest in the topic of probability distribution by using other statistical software such as Minitab and others.

## 1.6 Summary

From this chapter, it has explained about the background of study and also its problem statement that encourage the researcher to conduct this research. The problem statement inspires the researcher to explore more on how R runs and increase the programming skills. Research objective and problem statement will be derived from this study to build a new function in order to reproduce probability distribution table. So, the next chapter will discuss about the introduction of R and three types of common probability distribution which are Binomial, Poisson and Normal distribution.