

**PEMBANGUNAN MODEL REGRESI POISSON  
SIFAR-MELAMBUNG BERINTEGRASI DALAM  
BIOSTATISTIK**

**NUR SYABIHA BINTI ZAFKALI**

**UNIVERSITI SAINS MALAYSIA**

**2018**

**PEMBANGUNAN MODEL REGRESI POISSON  
SIFAR-MELAMBUNG BERINTEGRASI DALAM  
BIOSTATISTIK**

**oleh**

**NUR SYABIHA BINTI ZAFAKALI**

**Tesis yang diserahkan untuk  
memenuhi keperluan bagi  
Ijazah Doktor Falsafah**

**Ogos 2018**

## PENGHARGAAN

Alhamdulillah syukur ke hadrat Allah SWT kerana dengan limpah dan rahmat dari-Nya saya dapat menyiapkan tesis penyelidikan ini dengan jaya serta dalam tempoh masa yang ditetapkan. Pertama sekali, saya ingin merakamkan setinggi-tinggi penghargaan dan jutaan terima kasih kepada Prof. Madya Dr. Wan Muhamad Amir B Wan Ahmad atas segala nasihat, dorongan serta tunjuk ajar beliau semasa menyempurnakan tesis penyelidikan ini. Saya amat menghargai kesabaran dan keperihatinan Prof. Madya Dr. Wan Muhamad Amir B Wan Ahmad yang sentiasa berkongsi maklumat dan kepakaran yang dimiliki olehnya. Semangat kesabaran serta minat yang mendalam terhadap kajian ini telah banyak membantu saya untuk menyempurnakan pengajian ini. Di samping itu, saya turut ingin merakamkan setinggi-tinggi penghargaan kepada Kementerian Pengajian Tinggi (KPT) kerana memberi tajaan biasiswa MyPhD dan secara tidak langsung memudahkan lagi urusan saya untuk menyiapkan kerja-kerja tesis ini.

Selain daripada itu, saya juga ingin mengucapkan jutaan terima kasih kepada Pusat Pengajian Sains Pergigian (PPSG) dan Universiti Sains Malaysia (USM) khususnya kerana banyak memberi bantuan dana penyelidikan iaitu menerusi geran penyelidikan jangka pendek yang bernombor 304/PPSG/61313187. Malahan, saya juga ingin mengucapkan jutaan terima kasih kepada ahli keluarga terutamanya ibu tersayang iaitu Noor Aini Bt Ghazali kerana banyak membantu saya untuk meneruskan perjuangan ini dan sentiasa memberi kata-kata semangat. Dorongan yang diberikan olehnya membuatkan saya lebih bersemangat untuk menghabiskan pengajian. Selain itu, jutaan terima kasih diucapkan kepada ayah saya iaitu Zafakali Bin Wan Abdullah kerana cita-

cita dan impian beliau untuk melihat saya berjaya hingga ke peringkat ini membuatkan saya lebih fokus terhadap pengajian. Disebabkan oleh berkat doa daripada mereka saya berjaya hingga ke tahap ini. Tidak lupa juga kepada suami tercinta Luqmanul Hakim Bin Hasan di atas segala pengorbanan, dorongan dan kesabaran dalam membantu saya sepanjang tempoh pengajian. Setinggi-tinggi penghargaan dan terima kasih ditujukan kepada rakan-rakan yang turut sama memberi sokongan serta semangat untuk saya meneruskan pengajian ini. Jasa baik dari kalian amatlah saya hargai dan hanya Allah sahaja dapat membalasnya.

## ISI KANDUNGAN

Penghargaan	ii
Isi Kandungan	iv
Senarai Jadual	x
Senarai Rajah	xii
Senarai Singkatan	xiii
Abstrak	xiv
Abstract	xvi
<b>BAB 1:      PENDAHULUAN</b>	
1.1      Pengenalan Kepada Kaedah Pemodelan Regresi	1
1.2      Perkembangan Penyelidikan Berteraskan Kepada Model Regresi Poisson	3
1.3      Keperluan Kepada Pembentukan Model BRPSM	5
1.4      Persoalan Penyelidikan	6
1.5      Pernyataan Masalah Kajian	6
1.6      Objektif Kajian	8
1.7      Skop Penyelidikan	9
1.8      Kerangka Konsep Penyelidikan dan Kerangka Teori	9
1.9      Sumbangan Kajian Kepada Bidang Penyelidikan	11
1.10     Penyusunan Tesis	12
1.11     Justifikasi Penyelidikan	14

BAB 2:	SOROTAN KESUSASTERAAN	
2.1	Pengenalan kepada Jenis-jenis Taburan	16
	2.1.1 Taburan Normal	16
	2.1.2 Taburan Binomial	18
	2.2.3 Taburan Poisson	18
2.2	Sejarah Umum Model Regresi	20
2.3	Sejarah Pemodelan Regresi Binomial Negatif	21
2.4	Sejarah Pemodelan Regresi Poisson	22
2.5	Teori dan Penggunaan Model Regresi Poisson	23
2.6	Aplikasi Model Regresi Poisson Terhadap Sains Kesihatan	25
2.7	Keterbatasan Kaedah Regresi Poisson	26
2.8	Kelemahan dalam Penganggaran Apabila Kewujudan Sifar Melambung	28
2.9	Keperluan Kaedah Regresi Poisson Sifar-Melambung dalam Penyelidikan	29
2.10	Sejarah Kewujudan Sifar Melambung	29
2.11	Kesan Sifar Melambung Terhadap Data Analisis	30
2.12	Prosedur Mengenalpasti Model yang Mengalami Sifar Melambung	32
2.13	Sejarah Kepada Kaedah Butstrap	32
2.14	Kebaikan Teknik Butstrap	33
2.15	Pengembangan Teknik Butstrap Merentas Disiplin Ilmu	36
2.16	Pengenalan Set Kabur Kepada Pemodelan Tak Linear	38
2.17	Kelebihan Model Regresi Kabur	40
2.18	Keperluan Satu Model Statistik yang Lebih Teguh	41

2.19	Penyelidikan yang Lebih Merentas Disiplin Ilmu	42
<b>BAB 3:</b>	<b>METODOLOGI</b>	
3.1	Pengenalan	43
3.2	Reka Bentuk Penyelidikan	43
3.3	Saiz Sampel dan Populasi Kajian	44
3.4	Skop Kajian Penyelidikan	46
	3.4.1 Lokasi Penyelidikan	47
	3.4.2 jangkamasa Kajian	47
3.5	Kaedah A: Model Regresi Poisson	49
	3.5.1 Pengenalan kepada Model Regresi Poisson	49
	3.5.2 Bahasa Pengaturcaraan SAS Model Regresi Poisson	52
3.6	Kaedah B: Model Regresi Poisson Sifar-Melambung	56
	3.6.1 Pengenalan kepada Model Regresi Poisson Sifar-Melambung	56
	3.6.2 Penyemakan Kecukupan Model	63
	3.6.3 Bahasa Pengaturcaraan SAS Model Regresi Poisson Sifar-Melambung	70
3.7	Kaedah C: Teknik Butstrap	74
	3.7.1 Pengenalan kepada Teknik Butstrap	74
	3.7.2 Teori yang Mendasari Teknik Butstrap	78
	3.7.3 Kepentingan Teknik Butstrap	79
	3.7.4 Bahasa Pengaturcaraan SAS Teknik Butstrap	80

3.8	Kaedah D: Model Regresi Kabur	83
3.8.1	Pengenalan kepada Model Regresi Kabur	83
3.8.2	Teori Yang Mendasari Model Regresi Kabur	84
3.8.3	Kepentingan Model Regresi Kabur	86
3.8.4	Bahasa Pengaturcaraan SAS Model Regresi Kabur	87
3.9	Kaedah E: Pengenalan Kepada Model BRPSM	87
3.9.1	Kombinasi dan Penambah Baikan Kepada Kaedah Sedia Ada	88
3.9.2	Carta Alir Pembentukan Model Berdasarkan Model Berintegrasi	89
3.9.3	Prosedur Perlaksanaan Model BRPSM	90
3.9.3(a)	Analisis Data Secara Kaedah Model Regresi Poisson Sifar-Melambung	90
3.9.3(b)	Analisis Data Secara Kaedah Regresi Poisson Sifar-Melambung dan Melibatkan Teknik Butstrap	91
3.9.3(c)	Analisis Data Secara Model Berintegrasi	92
<b>BAB 4</b>	<b>KEPUTUSAN DAN PERBINCANGAN</b>	
4.1	Pengenalan	94
4.2	Kajian Kes I: Data Thalassemia	96
4.2.1	Penentuan Saiz Sampel Dalam Kajian	96
4.2.2	Analisis Deskriptif Statistik	98



4.2.3	Kaedah A: Analisis Data Melalui Kaedah Regresi Poisson	100
4.2.4	Kaedah B: Analisis Data Melalui Kaedah Regresi Poisson dan Melibatkan Teknik Butstrap	105
4.2.5	Kaedah C: Analisis Data Melalui Kaedah Regresi Poisson Sifar-Melambung	109
4.2.6	Kaedah D: Analisis Data Melalui Kaedah Regresi Poisson Sifar-Melambung dan Melibatkan Teknik Butstrap	112
4.2.7	Kaedah E: Analisis Data Melalui Model BRPSM	116
4.2.8	Perbandingan Keputusan Secara Keseluruhan	119
4.3	Kajian Kes II: Data Karies Gigi	120
4.3.1	Penentuan Saiz Sampel Dalam Kajian	121
4.3.2	Analisis Deskriptif Statistik	122
4.3.3	Kaedah A: Analisis Data Melalui Kaedah Regresi Poisson	124
4.3.4	Kaedah B: Analisis Data Melalui Kaedah Regresi Poisson Serta Melibatkan Teknik Butstrap	129
4.3.5	Kaedah C: Analisis Data Melalui Kaedah Regresi Poisson Sifar-Melambung	132
4.3.6	Kaedah D: Analisis Data Melalui Kaedah Regresi Poisson Sifar-Melambung Serta Melibatkan Teknik Butstrap	136
4.3.7	Kaedah E: Analisis Data Melalui Model BRPSM	139
4.3.8	Perbandingan Keputusan Secara Keseluruhan	142

BAB 5	KESIMPULAN DAN CADANGAN	
5.1	Pengenalan	144
5.2	Kesimpulan	144
5.3	Cadangan	147
RUJUKAN		150
LAMPIRAN		
Lampiran A	: Pemodelan Data bagi Kajian Kes I	
Lampiran B	: Pemodelan Data bagi Kajian Kes II	
SENARAI PENERBITAN		

## SENARAI JADUAL

		<b>Halaman</b>
Jadual 3.1	Jangka masa Kajian	48
Jadual 4.1	Keterangan Setiap Pemboleh ubah bagi Kajian Kes I	97
Jadual 4.2	Analisis Deskriptif Statistik Pesakit Thalassemia	99
Jadual 4.3	Analisis Data Melalui Kaedah Regresi Poisson bagi Kajian Kes I	101
Jadual 4.4	Ujian Kesesuaian Model Regresi Poisson bagi Kajian Kes I	104
Jadual 4.5	Analisis Data Melalui Kaedah Regresi Poisson dan Melibatkan Teknik Butstrap bagi Kajian Kes I	105
Jadual 4.6	Ujian Kesesuaian Model Regresi Poisson Melibatkan Teknik Butstrap bagi Kajian Kes I	108
Jadual 4.7	Analisis Data Melalui Kaedah Regresi Poisson Sifar-Melambung bagi Kajian Kes I	109
Jadual 4.8	Ujian Kesesuaian Model Regresi Poisson Sifar-Melambung bagi Kajian Kes I	111
Jadual 4.9	Analisis Data Melalui Kaedah Regresi Poisson Sifar Melambung dan Melibatkan Teknik Butstrap bagi Kajian Kes I	112
Jadual 4.10	Ujian Kesesuaian Model Regresi Poisson Sifar Melambung Melibatkan Teknik Butstrap bagi Kajian Kes I	115
Jadual 4.11	Analisis Data Yang Melibatkan Model BRPSM bagi Kajian Kes I	116
Jadual 4.12	Ujian Kesesuaian Model Berintegrasi bagi Kajian Kes I	118

Jadual 4.13	Perbandingan Keputusan Secara Keseluruhan bagi Kajian Kes I	119
Jadual 4.14	Keterangan Setiap Pemboleh ubah bagi Kajian Kes II	121
Jadual 4.15	Analisis Deskriptif Statistik Karies Gigi	123
Jadual 4.16	Analisis Data Melalui Kaedah Regresi Poisson	125
Jadual 4.17	Ujian Kesesuaian Model Regresi Poisson bagi Kajian Kes II	128
Jadual 4.18	Analisis Data Melalui Kaedah Regresi Poisson dan Melibatkan Teknik Butstrap bagi Kajian Kes II	129
Jadual 4.19	Ujian Kesesuaian Model Regresi Poisson Melibatkan Teknik Butstrap bagi Kajian Kes II	131
Jadual 4.20	Analisis Data Melalui Kaedah Regresi Poisson Sifar-Melambung bagi Kajian Kes II	132
Jadual 4.21	Ujian Kesesuaian Model Regresi Poisson Sifar-Melambung bagi Kajian Kes II	135
Jadual 4.22	Analisis Data Melalui Kaedah Regresi Poisson Sifar-Melambung dan Melibatkan Teknik Butstrap bagi Kajian Kes II	136
Jadual 4.23	Ujian Kesesuaian Model Regresi Poisson Sifar-Melambung Melibatkan Teknik Butstrap bagi Kajian Kes II	138
Jadual 4.24	Analisis Data Yang Melibatkan Model BRPSM bagi Kajian Kes II	139
Jadual 4.25	Ujian Kesesuaian Model BRPSM bagi Kajian Kes II	141
Jadual 4.26	Perbandingan Keputusan Secara Keseluruhan bagi Kajian Kes II	142

## SENARAI RAJAH

		<b>Halaman</b>
Rajah 1.1	Kerangka Konsep secara Keseluruhan	10
Rajah 2.1	Ciri-ciri Taburan Normal	17
Rajah 3.1	Proses Analisis Model Regresi Poisson	55
Rajah 3.2	Prosedur Mengenalpasti Data yang Mengalami Sifar Melambung	63
Rajah 3.3	Pembentukan Model Regresi Poisson Sifar- Melambung	73
Rajah 3.4	Prosedur Algoritma bagi teknik butstrap	77
Rajah 3.5	Prosedur Analisis Teknik Butstrap Melalui Bahasa Pengaturcaraan SAS	82
Rajah 3.6	Carta alir bagi pembentukan Model BRPSM	89
Rajah 4.1	Perjalanan Proses Menganalisis Data Kajian Secara Berperingkat	95

## SENARAI SINGKATAN

SAS	<i>Statistical Analysis System</i>
SPSS	<i>Statistical Package Social Science</i>
ZIP	<i>Zero-Inflated Poisson</i>
MLE	<i>Maximum Likelihood Estimation</i>
IRWLS	<i>Iteratively Re-Weighted Least Squares</i>
EM	<i>Expectation Maximization</i>
GLM	<i>Generalized Linear Model</i>

# PEMBANGUNAN MODEL REGRESI POISSON SIFAR-MELAMBUNG BERINTEGRASI DALAM BIOSTATISTIK

## ABSTRAK

Kaedah regresi Poisson merupakan satu pendekatan yang sering digunakan oleh para penyelidik dalam membuat analisis terhadap data yang berbentuk bilangan. Namun begitu, lebih nilai sifar seringkali berlaku pada data sebegini. Hal ini disebabkan oleh kewujudan sifar melambung terhadap hasil data yang dikumpulkan. Kejadian tersebut menyebabkan model regresi Poisson tidak lagi sesuai untuk diaplikasikan. Sebagai alternatifnya, kaedah regresi Poisson Sifar-Melambung telah dipilih untuk memodelkan data yang mengalami fenomena sifar berlebihan (sukatan serakan berlebihan). Penyelidikan ini turut memberi penekanan kepada pembinaan metodologi ke atas analisis data. Data yang dikumpulkan telah melibatkan dua kajian kes utama iaitu data daripada pesakit Thalassemia di kalangan kanak-kanak dan data pesakit yang mengalami masalah karies gigi. Fasa pertama dalam penyelidikan ini adalah merujuk kepada tatacara pembinaan algoritma bagi memodelkan kaedah regresi Poisson Sifar-Melambung melalui teknik butstrap (*bootstrap*) dan digabungkan dengan kaedah regresi kabur. Gabungan di antara kaedah-kaedah tersebut dirujuk sebagai Model Berintegrasi. Fasa kedua pula merupakan perbandingan bagi hasil dapatan kajian di antara Model Berintegrasi dengan kaedah yang sedia ada. Pentaabiran terhadap keseluruhan model yang diperoleh turut dilaksanakan untuk mendapatkan maklumat yang berkaitan dengan kecekapan model. Algoritma yang dibina adalah berteraskan kepada konsep penambahbaikan dan setiap perincian terhadap kaedah yang digunakan akan diterangkan secara lebih teratur pada bahagian

metodologi. Hasil utama kajian ini adalah merujuk kepada pembinaan metodologi penyelidikan yang turut membantu penyelidik untuk menganalisis data secara lebih berkesan serta memberi keputusan yang lebih tepat. Penggunaan Model Berintegrasi terhadap kajian kes yang telah dikumpulkan didapati telah menghasilkan nilai purata yang lebih cekap berbanding dengan kaedah sedia ada. Hal ini menunjukkan bahawa kaedah ini telah mempamerkan satu sifat pemodelan yang lebih baik bagi setiap set data yang dikaji serta boleh diaplikasikan dengan jayanya.



# **DEVELOPMENT OF ZERO-INFLATED POISSON REGRESSION MODEL INTEGRATED IN THE BIOSTATISTICS**

## **ABSTRACT**

The Poisson regression method is an approach often chosen by researchers in making analysis of data in the form of numbers. However, the excess zero often applies to such data. This is due to the existence of overdispersion of the data collected. This phenomenon led to the Poisson regression model is no longer suitable to be applied. Alternatively, the method of Zero-Inflated Poisson regression was selected to model the data that have excess zero phenomenon (overdispersion). This research also emphasizes the development methodology of data analysis. The data gathered involved two major cases studies of data from patients with Thalassemia among children and the patients with dental caries problems. The first phase in this research is to refer to the algorithm development procedure to model the Zero-Inflated Poisson Regression method through the bootstrap method and combined with the fuzzy regression method. The combination of these methods is referred to as the Integrated Model. The second phase is the comparison of the findings between the Integrated Model and the existing method. An overview of the overall model was also performed to obtain information related to the efficiency of the model. The algorithm was developed based on the concept of improvement and every detail of the methods used will be explained carefully on the methodology. The main outcome of this research is to refer to the development of a research methodology that also helps researchers to analyse data more effectively and to give more accurate results. The use of the Integrated Model of the two case studies has resulted in a more efficient average value

compared to the existing method. It shows that this method has demonstrated a better model for each set of data that can be studied and applied successfully.

# BAB 1

## PENDAHULUAN

### 1.1 Pengenalan Kepada Kaedah Pemodelan Regresi

Kaedah pemodelan regresi seringkali digunakan dalam bidang penyelidikan bagi mengetahui bentuk hubungan di antara pemboleh ubah bersandar dengan satu atau lebih pemboleh ubah tidak bersandar. Analisis regresi yang sering digunakan adalah analisis regresi klasik iaitu pemboleh ubah bersandar yang dimilikinya merupakan data jenis selanjar dan bertaburan normal. Namun dalam perkembangan terkini model regresi klasik ini tidak mampu mengatasi permasalahan apabila pemboleh ubah bersandar merupakan data berbentuk diskrit serta tidak bertaburan normal. Penyelesaian bagi mengatasi masalah tersebut adalah dengan menggunakan Model Linear Teritlak (MLT) atau dikenali sebagai *Generalized Linear Model (GLM)*.

Ujian penganggaran yang diterapkan oleh GLM tidak memerlukan penganggaran kenormalan daripada pemboleh ubah bersandar (De Jong & Heller, 2008). Salah satu model regresi yang tergolong dalam penerapan GLM adalah Regresi Poisson dengan setiap pemboleh ubah bersandar yang dimilikinya dianggap bertaburan Poisson. Regresi ini digunakan untuk menganalisis hubungan antara pemboleh ubah bersandar yang berbentuk diskrit dengan satu atau lebih pemboleh ubah tidak bersandar serta

nilai bagi varians dan minnya mengalami sukatan serakan yang sama (*equidispersion*) (McCullagh & Nelder, 1989).

Namun begitu, kebanyakan data kesihatan seringkali mengalami fenomena sifar yang berlebihan. Hal ini disebabkan oleh ketidakwujudan faktor yang dikaji selepas penetapan sesuatu pemboleh ubah dalam kajian. Data dikatakan mengalami fenomena sifar yang berlebihan apabila nilai varians lebih besar berbanding dengan nilai min. Istilah yang digunapakai bagi mencerminkan situasi sebegini adalah sukatan serakan yang berlebihan (*overdispersion*). Sebaliknya, apabila data mempunyai nilai varians yang lebih kecil daripada nilai min maka ianya dikenali sebagai sukatan serakan yang berkurangan (*underdispersion*). Apabila situasi ini hadir, model Regresi Poisson tidak lagi sesuai untuk diaplikasikan kerana ianya akan menghasilkan satu keputusan analisis yang kurang tepat disebabkan oleh kewujudan hubungan antara nilai varians dan min yang tidak sama.

Justeru, beberapa kaedah boleh digunakan bagi mengatasi masalah ini iaitu menerusi penggunaan model Regresi Binomial Negatif dan Regresi Poisson Sifar-Melambung. Kedua-dua model ini dapat mengatasi fenomena bagi sukatan serakan yang berlebihan mahupun berkurangan kerana ianya tidak mensyaratkan nilai varians adalah sama dengan nilai min, malah model Regresi Poisson Sifar-Melambung juga turut digunakan apabila terdapat lebih nilai sifar bagi pemboleh ubah bersandar (Jansakul, 2005). Berdasarkan kepada kepentingan kajian yang telah dibahaskan sebelum ini, satu kajian lanjutan berkaitan model Regresi Poisson dan Regresi Poisson Sifar-Melambung telah diberi perhatian serta fokus dengan lebih teliti.

Malah perbincangan yang lebih lanjut akan dibahaskan dengan lebih mendalam pada bahagian seterusnya.

## **1.2 Perkembangan Penyelidikan Berteraskan Kepada Model Regresi Poisson**

Pada tahun 1822, seorang penyelidik Amerika yang dikenali sebagai Sir Francis Galton telah memperkenalkan suatu konsep statistik yang dikenali sebagai regresi malah menurut beliau ianya merupakan proses umum untuk mengaitkan satu pemboleh ubah dengan pemboleh ubah yang lain (Kalkatawi et al., 2015). Setelah beberapa tahun kemudian, konsep mengenai istilah regresi yang diperkenalkan oleh Galton ini turut dikembangkan oleh para penyelidik lain dengan menambah istilah menjadi regresi berganda untuk menggambarkan proses yang melibatkan lebih daripada satu pemboleh ubah dan mengaitkan pemboleh ubah tersebut dengan pemboleh ubah yang lain.

Istilah bagi regresi dalam pengertian moden merupakan suatu kajian terhadap kebergantungan yang melibatkan satu pemboleh ubah. Pemboleh ubah ini turut bergantung kepada pemboleh ubah yang lain dan ianya dikenali sebagai pemboleh ubah tidak bersandar serta digunakan untuk membuat penganggaran bagi sesuatu populasi dalam kajian (Gujarati et al., 2009). Manakala kajian yang telah dijalankan oleh Levin dan Rubin pada tahun 1998 pula, mereka menyatakan bahawa suatu regresi telah digunakan untuk menentukan sifat-sifat dan kekuatan hubungan antara dua pemboleh ubah dan mengaitkan nilai dari pemboleh ubah yang belum diketahui berdasarkan pemerhatian yang lepas (Lai, 2013).

Menurut kajian yang dijalankan oleh Ismail dan Jemain (2005), didapati bahawa apabila sesuatu data kajian bersifat sebagai suatu bilangan, maka pemboleh ubah bersandar yang dimilikinya adalah dalam bentuk diskrit. Maka, salah satu model yang sesuai untuk menganalisis data tersebut adalah model Regresi Poisson. Hal ini kerana model Regresi Poisson dapat menganalisis hubungan antara pemboleh ubah bersandar dalam bentuk diskrit dengan satu atau lebih pemboleh ubah tidak bersandar serta mensyaratkan nilai yang sama antara min dan varians. Namun begitu, data dalam bentuk bilangan seringkali menghasilkan nilai varians yang lebih besar daripada nilai min sehingga mengganggu hasil dapatan kajian. Sekiranya model Regresi Poisson diaplikasikan pada data tersebut ianya akan menghasilkan penganggaran parameter yang kurang tepat dan keputusan yang diperolehi tidak dapat digunakan. Bagi mengatasi permasalahan tersebut maka model Regresi Poisson Sifar-Melambung dan Regresi Binomial Negatif dapat diaplikasikan ke atas kajian kes (Liu & Power, 2012).

Model Regresi Binomial Negatif merupakan salah satu alternatif yang digunakan untuk menganalisis hubungan antara pemboleh ubah bersandar dengan melibatkan satu atau lebih pemboleh ubah tidak bersandar. Model ini telah diperkenalkan oleh Lawless pada tahun 1987, malah menurut kajian beliau juga model ini amat sesuai digunakan ke atas data yang berbentuk bilangan (Ward & Myers, 2005). Model Regresi Binomial Negatif juga dapat digunakan secara optimum samada dalam keadaan sukatan serakan sama atau terhadap data yang mengalami sukatan serakan berlebihan. Selain daripada itu, kebaikan model tersebut turut dapat mengatasi ketidaknormalan pada pemboleh ubah bersandar.

### 1.3 Keperluan Kepada Pembentukan Model BRPSM

Apabila menjalankan sesuatu kajian yang berteraskan pembinaan terhadap pemodelan statistik komputasi, adalah menjadi lumrah kepada setiap penyelidik untuk menghasilkan model yang terbaik dan lebih menekankan tatacara model yang lebih cekap. Namun begitu, keterbatasan dalam melaksanakan sesuatu kajian seringkali berlaku dan ianya sering melibatkan kaedah yang sedia ada. Di samping itu, permasalahan seperti ketidakcukupan data juga berlaku sehingga menghasilkan satu keputusan yang tidak tepat. Oleh yang demikian, keperluan kepada penambahbaikan struktur model yang sedia ada amat penting bagi menghasilkan keputusan yang lebih baik. Dalam penyelidikan yang dijalankan ini suatu kaedah berintegrasi telah dibina untuk membaik pulih struktur model yang asal dan ianya dinamakan sebagai Model Berintegrasi Regresi Poisson Sifar-Melambung atau singkatannya adalah model BRPSM. Model ini merupakan gabungan antara kaedah Regresi Poisson Sifar-Melambung yang dijalankan pensampelan semula atau dikenali sebagai teknik butstrap (*bootstrap*) dengan kaedah regresi kabur. Model BRPSM ini merujuk kepada gabungan daripada kaedah-kaedah tersebut dan merupakan satu kaedah yang lebih fleksibel untuk menganalisis setiap kajian kes yang dikumpulkan.

Model BRPSM yang dibina turut bermanfaat kepada kajian yang mempunyai saiz sampel yang kecil dan lebihan nilai sifar. Hal ini kerana para penyelidik seringkali menghadapi kekangan data terutamanya dalam bidang perubatan dan kesihatan awam sehingga mengganggu ketepatan bagi setiap penganggaran statistik (Kalktawi et al., 2015). Di samping itu, Model BRPSM ini juga adalah lebih baik berbanding

dengan kaedah sedia ada kerana model tersebut akan menghasilkan ketepatan anggaran yang lebih signifikan. Tambahan pula, model BRPSM dapat memberi keputusan yang lebih meyakinkan bagi setiap ujikaji yang digunakan. Oleh yang demikian, keperluan dalam membina Model BRPSM telah memberi dorongan untuk melaksanakan penyelidikan ini bagi memberi maklumat baru kepada para penyelidik.

#### **1.4 Persoalan Penyelidikan**

Berikut merupakan persoalan kajian penyelidikan:

- a) Adakah data yang dikumpulkan dalam setiap kajian kes mengalami masalah sifar yang berlebihan?
- b) Adakah model Regresi Poisson Sifar-Melambung dapat mengatasi masalah kewujudan sifar yang berlebihan berbanding model Regresi Poisson?
- c) Adakah data yang dijanakan menerusi teknik butstrap lebih baik berbanding dengan data yang asal?
- d) Adakah keputusan daripada Model BRPSM adalah lebih baik berbanding dengan kaedah sedia ada?

#### **1.5 Pernyataan Masalah Kajian**

Setiap penyelidikan adalah didasari oleh hipotesis penyelidikan dan hipotesis penyelidikan tersebut berkait rapat dengan pernyataan masalah yang dikaji. Masalah



utama yang akan dibincangkan dalam penyelidikan ini adalah kekurangan kaedah sedia ada (model Regresi Poisson) terhadap data yang mempunyai lebih sifar. Apabila hal ini berlaku, model Regresi Poisson tidak lagi sesuai digunakan oleh kerana bilangan sifar yang berlebihan akan menyebabkan berlakunya situasi yang dipanggil sukatan serakan berlebihan. Kewujudan lebih sifar di dalam data juga turut mengimplikasikan hal seperti memberikan keputusan yang tidak tepat sekiranya kaedah Regresi Poisson digunakan serta penggunaan keputusan yang tidak tepat akan mengakibatkan proses pentaabiran terjejas. Bagi mengatasi masalah tersebut, satu model yang telah diperkenalkan oleh Lambert pada tahun 1992 iaitu model Regresi Poisson Sifar-Melambung akan diaplikasikan. Namun begitu, penggunaan model Regresi Poisson Sifar-Melambung amatlah terhad kerana ianya memerlukan pembinaan algoritma secara khusus dan kebanyakan kajian yang telah dijalankan masih tidak dapat menangani masalah yang dihadapi. Selain daripada itu, perisian statistik yang terdapat di pasaran terutama dari segi algoritmanya masih tidak membenarkan para penyelidik untuk menjalankan analisis menerusi pemodelan Regresi Poisson Sifar-Melambung secara berintegrasi. Oleh yang demikian, kajian ini akan memberi fokus secara khusus terhadap tatacara pembinaan algoritma bagi model Regresi Poisson Sifar-Melambung dengan membuat penambahbaikan terhadap struktur pemodelannya.

Selain daripada itu, kebanyakan para penyelidik kurang peka terhadap penggunaan model Regresi Poisson dan Regresi Poisson Sifar-Melambung sehingga menghasilkan satu keputusan kajian yang kurang tepat. Hal ini sepertimana penggunaan model Regresi Poisson yang perlu mematuhi beberapa perkara untuk diaplikasikan iaitu data kajian harus menepati sukatan serakan yang sama. Oleh itu,

suatu ruang penambahbaikan perlu dilaksanakan bagi memperoleh hasil dapatan yang lebih optimum. Selain daripada itu, kebanyakan hasil kajian lepas amat kurang menggunakan model Regresi Poisson dalam pelaksanaan analisis statistik. Namun begitu, model ini masih boleh diperbaiki mengikut keperluan pada masa kini bergantung kepada jenis kajian yang dijalankan oleh para penyelidik. Pembentukan suatu model yang terbaik dilihat sebagai satu perkara yang begitu penting dalam konteks sains kesihatan namun kebanyakan analisis yang dijalankan tidak menjalani atau membuat sebarang perbandingan bagi memastikan tahap ketepatan setiap model.

## **1.6 Objektif Penyelidikan**

Objektif kajian yang ingin dicapai dalam penyelidikan ini adalah seperti berikut:

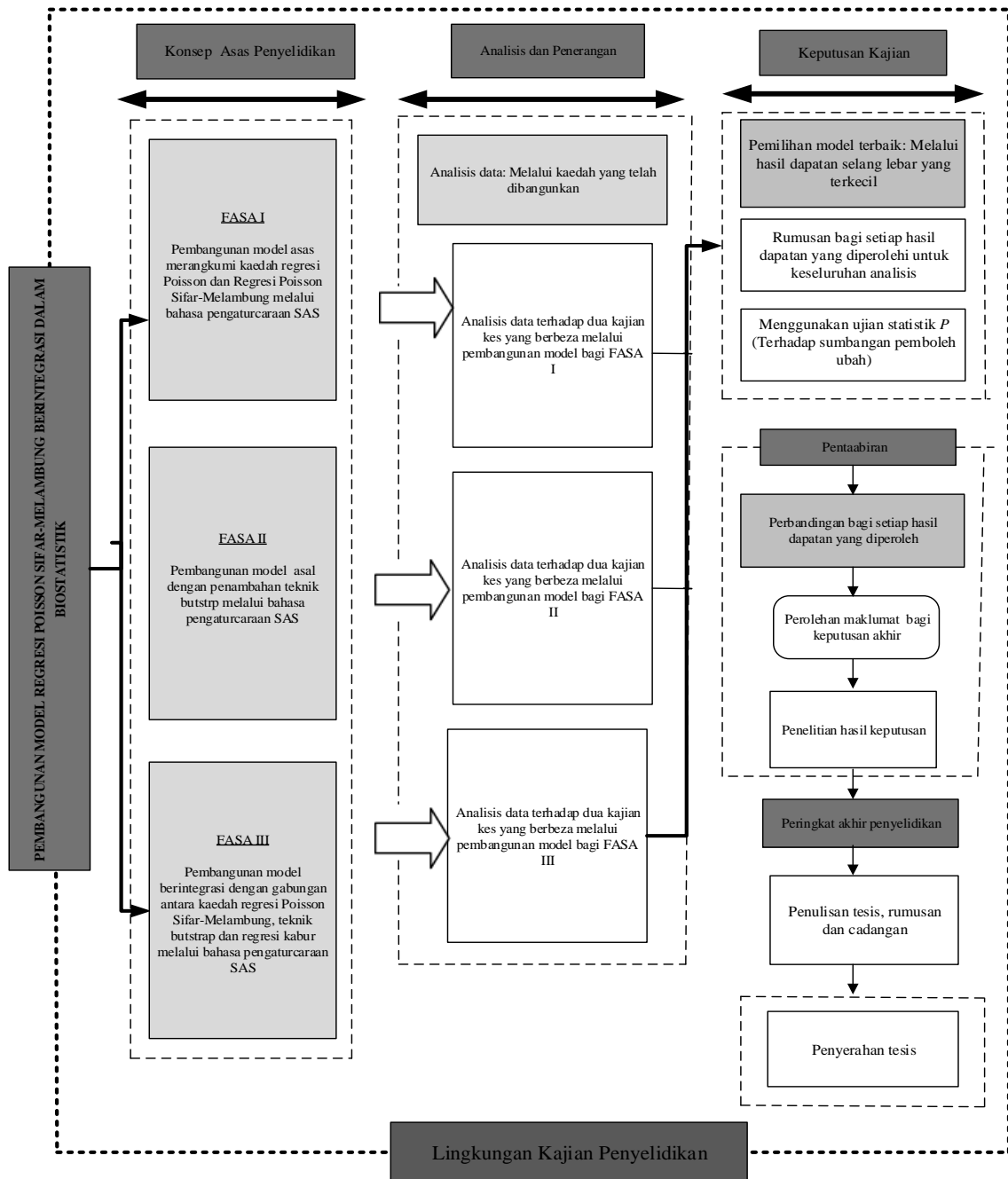
- a) Membangunkan algoritma bagi kaedah berintegrasi iaitu Model BRPSM yang berteraskan model Regresi Poisson Sifar-Melambung dengan teknik butstrap dan regresi kabur.
- b) Menganalisis data bagi setiap kajian kes yang dikumpulkan melalui kaedah Regresi Poisson, kaedah Regresi Poisson dengan teknik butstrap, kaedah Regresi Poisson Sifar-Melambung, kaedah Regresi Poisson Sifar-Melambung dengan teknik butstrap dan Model BRPSM.
- c) Membandingkan kecekapan bagi setiap anggaran yang diperoleh melalui kaedah yang dimodelkan dengan data asal dengan kaedah yang dimodelkan dengan data melibatkan teknik butstrap.

## 1.7 Skop Penyelidikan

Skop penyelidikan diperlukan bagi membatasi tahap sesuatu kajian supaya kajian yang dijalankan berada dalam satu ruang lingkup yang khusus dan lebih fokus. Skop penyelidikan yang dimaksudkan adalah merujuk kepada:

- a) Kajian ini terhad kepada pembinaan algoritma bagi model Regresi Poisson, Regresi Poisson Sifar-Melambung dan pembinaan Model BRPSM.
- b) Perbandingan keputusan hanya terhad kepada analisis data melalui pemodelan Regresi Poisson, Regresi Poisson melalui teknik butstrap, Regresi Poisson Sifar-Melambung, Regresi Poisson Sifar-Melambung melalui teknik butstrap dan Model BRPSM.
- c) Kajian ini menggunakan perisian SAS untuk menganalisis data bagi mendapatkan model Regresi Poisson, Regresi Poisson Sifar-Melambung dan model berintegrasi yang diperoleh secara Model BRPSM.

## 1.8 Kerangka Konsep Penyelidikan dan Kerangka Teori



Rajah 1.1 Kerangka Konsep Secara Keseluruhan

## **1.9 Sumbangan Kajian Kepada Bidang Penyelidikan**

Penyelidikan ini memberikan beberapa sumbangan penting yang berkaitan dengan tatacara pembinaan kaedah model Regresi Poisson dan juga kaedah Regresi Poisson Sifar-Melambung serta aplikasinya terhadap bidang penyelidikan. Pembinaan metodologi ini turut melibatkan kaedah butstrap ke dalam Model BRPSM yang secara khususnya memberi fokus kepada pembangunan algoritma dalam bahasa pengaturcaraan SAS. Pembangunan algoritma ini melibatkan beberapa aspek pemodelan regresi Poisson, Regresi Poisson Sifar-Melambung dan Model BRPSM.

Seterusnya, kajian ini turut memberi sumbangan dari segi penghasilan dan penerbitan kertas-kertas kerja yang akan menjadi sumber rujukan utama kepada penyelidik lain terutamanya kepada penyelidik yang mencari Model BRPSM dalam menganalisis data yang bermasalah (sifar melambung). Selain itu juga, kaedah yang dibina ini turut memuatkan beberapa kaedah yang dapat mengoptimumkan dapatan kajian secara keseluruhan menerusi dapatan kajian itu sendiri. Sumbangan ini dapat diklasifikasikan kepada dua aspek utama iaitu, dapatan secara pembangunan kaedah penyelidikan dan dapatan secara ilmiah. Dapatan secara ilmiah melibatkan dapatan kajian secara penemuan terhadap faktor-faktor yang berkaitan dengan penyelidikan yang dijalankan. Hal ini dapat memberikan satu eksplorasi yang baru dengan menggunakan Model BRPSM. Penemuan ini bukan sahaja berguna kepada penyelidik dalam bidang statistik tetapi ianya turut memberi sumbangan kepada pelbagai lapisan penyelidik lain yang berkaitan.

Tambahan pula, penyelidikan ini juga dapat memperkayakan sorotan kesusasteraan dalam pembinaan metodologi statistik dan juga statistik aplikasi. Hal ini kerana, hasil kajian ini dapat memberi informasi yang sangat berguna kepada penyelidik mengenai kekuatan dan kelemahan model-model tersebut. Malahan, ianya turut memberi pengetahuan kepada para penyelidik lain yang menggunakan model yang sama dalam kajian mereka. Oleh itu, penyelidikan ini diharapkan dapat memberi impak yang besar terhadap pembangunan metodologi penyelidikan pada masa hadapan.

### **1.10 Penyusunan Tesis**

Tesis ini dimulai dengan pendahuluan dalam Bab 1 yang menghuraikan tentang pengenalan kepada model Regresi Poisson dan model Regresi Poisson Sifar-Melambung. Dalam bab ini turut dinyatakan latar belakang kajian secara ringkas dan perkembangan penyelidikan terhadap model Regresi Poisson dan model Regresi Poisson Sifar-Melambung. Selain daripada itu, bahagian ini turut memaparkan persoalan kajian serta pernyataan masalah, objektif utama penyelidikan, skop kajian, hipotesis dan kerangka konsep penyelidikan serta beberapa sumbangan penting terhadap kajian yang dijalankan. Malah dalam bab ini juga membincangkan tentang justifikasi penyelidikan secara terperinci mengenai model-model regresi yang dikaji.

Bab 2 pula menyentuh perkembangan penyelidikan serta menerangkan tentang sorotan kajian mengenai teori-teori dan fakta-fakta yang berkaitan dengan model Regresi Poisson dan model Regresi Poisson Sifar-Melambung berdasarkan kajian yang lepas. Malahan di bahagian ini juga turut menceritakan tentang sejarah kewujudan sifar melambung iaitu suatu fenomena yang sering berlaku sehingga

menghasilkan data yang mengalami sukatan serakan berlebihan dan kesannya terhadap data yang dianalisis. Selain daripada itu, Bab 2 turut membincangkan tentang teknik butstrap serta kelebihan penggunaan teknik tersebut di dalam sesuatu penyelidikan. Pendekatan kaedah regresi kabur terhadap pemodelan tak linear turut dibincangkan dalam bab ini di samping menerangkan tentang kelebihan kaedah tersebut dalam pembentukan model.

Bab 3 dalam kajian ini pula menerangkan tentang metodologi penyelidikan serta kaedah-kaedah yang dijalankan. Selain daripada itu, bab ini turut memberi penekanan mengenai pendahuluan kajian kes, reka bentuk penyelidikan serta skop penyelidikan yang meliputi lokasi, jangkamasa dan tahap-tahap kajian yang dijalankan. Bab ini juga memaparkan algoritma bagi pembinaan model Regresi Poisson dan model Regresi Poisson Sifar-Melambung serta algoritma mengenai teknik butstrap dan kaedah regresi kabur. Selain daripada itu, bahagian ini turut membincangkan tentang Model BRPSM serta algoritma bagi pembentukan kaedah tersebut. Perbincangan yang lebih terperinci mengenai setiap metodologi yang digunakan serta carta alir bagi setiap perjalanan pembentukan model dinyatakan di bahagian ini.

Seterusnya Bab 4 pula mempersembahkan keputusan dan perbincangan hasil daripada setiap kajian yang dijalankan berdasarkan kepada metodologi yang digunakan dalam penyelidikan. Perbincangan yang lebih terperinci mengenai setiap analisis yang dijalankan serta huraian tentang dapatan kajian telah dinyatakan dalam bab ini. Tambahan pula, dalam bab ini juga turut membincangkan tentang

perbandingan bagi setiap keputusan yang diperoleh untuk melihat model yang lebih sesuai ke atas setiap kajian kes yang dijalankan.

Manakala dalam Bab 5 yang merupakan bahagian terakhir dalam kajian ini telah menerangkan tentang kesimpulan yang dibuat berdasarkan penyelidikan yang dijalankan secara keseluruhannya serta memaparkan unjuran atau saranan kajian pada masa yang akan datang. Penekanan dan tumpuan utama diberi kepada teknik-teknik pembaharuan dalam analisis Regresi Poisson dan Regresi Poisson Sifar-Melambung serta pembinaan kaedah berintegrasi yang lebih baik.

### **1.11 Justifikasi Penyelidikan**

Setiap penyelidikan yang dijalankan mempunyai justifikasi yang kukuh. Sekiranya justifikasi sesuatu penyelidikan adalah tidak teguh, maka penyelidikan yang dijalankan boleh dipertikaikan dari segi kesahihannya. Untuk kajian ini, beberapa justifikasi telah dikenalpasti dan diringkaskan seperti berikut.

- a) Keterbatasan kaedah pemodelan Regresi Poisson dan Regresi Poisson Sifar-Melambung telah membuka ruangan kepada penyelidikan yang dijalankan ini. Keterbatasan ini merujuk kepada keputusan kajian yang kurang optimum. Kaedah yang sedia ada adalah terhad kepada aplikasi sahaja (*ready made software*). Justeru, keperluan kepada model regresi secara berintegrasi amatlah diperlukan bagi menjamin keputusan yang diperolehi adalah berkualiti.



- b) Keperluan untuk mengoptimumkan dapatan kajian secara lebih komprehensif dengan mengaplikasikan teknik tersebut ke dalam penyelidikan. Kepelbagaian teknik-teknik sebegini dapat menghasilkan dapatan kajian yang lebih bernilai dan lebih berharga di dalam proses pembuatan keputusan.
  
- c) Selain itu juga, pembinaan kaedah ini turut menyediakan satu platform metodologi penyelidikan yang dapat mengatasi masalah yang berkaitan dengan saiz sampel minimum.
  
- d) Menyediakan penyelidik dengan kepelbagaian pilihan keputusan yang lebih baik. Menerusi kajian ini juga, penyelidik dapat membuat perbandingan serta dapat memilih keputusan secara bijak berdasarkan kepada hasil dapatan yang diperolehi.

## BAB 2

### SOROTAN KESUSASTERAAN

#### 2.1 Pengenalan kepada Jenis-jenis Taburan

Data secara umumnya terdiri daripada pelbagai corak taburan antaranya taburan normal, taburan Poisson, taburan Binomial dan sebagainya. Oleh yang demikian, kaedah untuk menganalisis juga mestilah sepadan dengan corak taburan data tersebut. Berikut merupakan beberapa penerangan ringkas mengenai data yang sering digunakan oleh para penyelidik.

##### 2.1.1 Taburan Normal

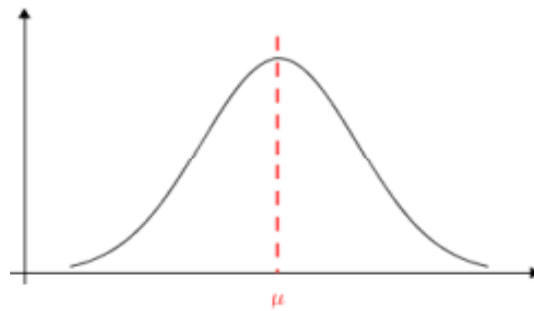
Taburan normal merupakan satu taburan yang selanjur dan berbentuk simetri serta mempunyai nilai min,  $\mu$  dan nilai sisihan piawai,  $\sigma^2$  dengan fungsi ketumpatan kebarangkaliannya adalah seperti berikut:

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}$$

bagi domainnya adalah  $x \in (-\infty, \infty)$  (Limpert, 2017). Berdasarkan kajian yang telah dijalankan oleh Feller (1968), beliau telah menggunakan simbol  $\varphi(x)$  bagi

menggantikan  $p(x)$  dalam persamaan di atas. Namun begitu, simbol ini telah ditukarkan kepada  $n(x)$  untuk kajian beliau pada tahun 1971.

De Moivre pula telah mengembangkan taburan normal sebagai suatu penghampiran kepada taburan Binomial kemudian digunakan oleh Laplace pada tahun 1783 untuk mengkaji kesilapan yang ada dalam setiap pengukuran. Seterusnya, taburan ini turut diaplikasikan oleh Gauss pada tahun 1909 dalam kajiannya untuk menganalisis data astronomi (Havil, 2003). Ciri-ciri bagi taburan normal dapat dilihat seperti dalam Rajah 2.1 yang berikut:



**Rajah 2.1** Ciri-ciri Taburan Normal

Keluk normal ini digunakan bagi menyatakan sifat-sifat yang terdapat pada sesuatu data yang ingin dikaji. Keluk normal juga berfungsi untuk memberi nilai bahagian pencerapan yang terdapat di antara dua nilai. Setiap keluk normal ini ditentukan oleh sisihan piawai,  $s$  dan kedudukannya pada paksi mengufuk ditentukan oleh min,  $\bar{x}$  (Limpert, 2017).

### 2.1.2 Taburan Binomial

Taburan Binomial merupakan satu taburan diskrit dan ianya telah dipelopori dengan meluas oleh para penyelidik sejak beratus tahun dahulu lagi. Namun begitu, terdapat beberapa andaian mengenai aplikasi taburan Binomial ke atas kajian iaitu (Lawless, 1987):

- a) Ujian yang melibatkan bilangan,  $n$ .
- b) Ujian yang dijalankan adalah tidak bersandar dan tidak bergantung dengan ujikaji yang sebelumnya.
- c) Setiap ujian terdapat dua kemungkinan samada kebarangkalian memperoleh kejayaan,  $p$  dan kebarangkalian memperoleh kegagalan,  $q = (1 - p)$ .

### 2.1.3 Taburan Poisson

Taburan Poisson merupakan taburan diskrit yang penting untuk menggambarkan fenomena fungsi-fungsi kebarangkalian seperti jumlah kemalangan yang berlaku setiap bulan, jumlah panggilan talifon yang datang dari papan suis dan jumlah pesakit yang menghidap Thalassemia di kalangan kanak-kanak. Penemuan mengenai corak taburan sebegini ini telah dikaji oleh seorang yang berbangsa Perancis dan telah menamakan corak tersebut sempena nama beliau iaitu Simeon Denis Poisson. Beliau telah menemui taburan Poisson secara kebetulan dan ketika itu para ahli matematik yang lain tidak mempedulikan catatan Simeon Denis Poisson sehingga

pada tahun 1907, seorang ahli statistik yang bernama Gosset telah menemui penggunaan sebenar taburan Poisson. Kaedah taburan Poisson tertumpu kepada kejadian yang menggambarkan keadaan pemboleh ubah bersandar yang terdiri daripada sebarang integer bukan negatif. Pemboleh ubah bersandar adalah bertaburan Poisson dengan setiap parameter,  $\mu$  mempunyai nilai set integer  $y = 0,1,2,3,\dots$ , serta kebarangkaliannya seperti mana yang berikut (Frome, 1983):

$$p(y; \mu) = \frac{\lambda^{-\mu} \mu^y}{y!}$$

dengan  $\mu$  adalah min bagi taburan Poisson, sehingga persamaan ini dinyatakan sebagaimana:

$$p(y_i; \hat{\beta}) = \frac{[t_i \mu(x_i; \hat{\beta})]^{y_i} e^{-t_i [\mu(x_i; \hat{\beta})]}}{y_i!}$$

dengan

$\mu(x_i; \hat{\beta})$  adalah min Poisson,

$\hat{\beta}$  adalah vektor parameter yang ditafsir.

Taburan Poisson seringkali digunakan untuk memodelkan peristiwa yang jarang berlaku. Ianya mempunyai beberapa ciri yang utama iaitu:

- a) Nilai varians adalah sama dengan nilai min,  $E(Y) = \text{var}(Y) = \mu$ .  
Peristiwa ini dipanggil sebagai sukatan serakan sama (*equidispersion*).
- b) Taburan Poisson lebih cenderung untuk condong ke kanan.

- c) Taburan Poisson dengan nilai min yang besar seringkali dianggarkan sebagai taburan normal.

Manakala penyelidikan yang dijalankan ini lebih tertumpu kepada penggunaan kaedah taburan Poisson dan turut mengaplikasikan model tersebut dalam setiap analisis data bagi kajian kes yang terlibat.

## **2.2 Sejarah Umum Model Regresi**

Istilah regresi pertama kali digunakan oleh Sir Francis Galton iaitu bersangkutan dengan satu kajian mengenai tinggi badan anak yang dilahirkan oleh ibu bapanya yang juga tinggi serta cenderung untuk bergerak (*regress*) kearah ketinggian mengikut populasinya. Galton memperkenalkan kata regresi (*regression*) sebagai nama proses umum bagi mengaitkan satu pemboleh ubah dengan pemboleh ubah yang lain contohnya tinggi badan anak dengan tinggi badan ibu bapa mereka. Berdasarkan kepada perkembangannya, para ahli statistik turut memperkenalkan istilah regresi berganda (*multiple regression*) untuk menggambarkan proses yang melibatkan beberapa pemboleh ubah serta mengaitkan setiap pemboleh ubah tersebut dengan pemboleh ubah yang lain (Jerome et al., 2015).

Menurut kajian yang dijalankan oleh Levin dan Rubin pada tahun 1997, regresi digunakan untuk menentukan sifat serta kekuatan bagi hubungan antara dua pemboleh ubah dan mengaitkan nilai tersebut berdasarkan kepada pemerhatian yang lepas. Analisis regresi yang sering digunakan adalah analisis regresi klasik iaitu pemboleh ubah bersandarnya merupakan data yang bersambungan dan bertaburan

normal. Namun begitu, terdapat kelemahan terhadap penggunaan model regresi klasik iaitu model ini tidak mampu untuk mengatasi permasalahan yang wujud apabila pemboleh ubah bersandarnya berbentuk diskrit serta tidak bertaburan normal.

Penyelesaian bagi mengatasi masalah tersebut adalah melalui penggunaan Model Linear Teritlak (MLT) atau dikenali sebagai *Generalized Linear Model* (GLM). Ujian penganggaran yang diterapkan kepada model ini apabila pemboleh ubah bersandar tidak disyaratkan supaya bertaburan normal serta tidak mengharuskan kehomogenan daripada setiap variansnya (De Jong & Heller, 2008). Manakala model-model regresi yang termasuk dalam penerapan GLM ini adalah Regresi Binomial Negatif dan Regresi Poisson.

### **2.3 Sejarah Pemodelan Regresi Binomial Negatif**

Model Regresi Binomial Negatif telah diperkenalkan oleh Lawless pada tahun 1987 dan model ini digunakan terhadap sesuatu data jenis bilangan serta berbentuk diskrit (Ward & Myers, 2005). Manakala taburan Binomial Negatif pula merupakan gabungan daripada Poisson-gamma malah fungsi utamanya digunakan bagi meluaskan taburan Poisson (McCullagh & Nelder, 1989). Model Binomial Negatif mampu mengatasi ketidaknormalan pada pemboleh ubah bersandar. Selain daripada itu, model ini juga sesuai digunakan terhadap kajian yang mengalami sifar melambung kerana ianya tidak mengharuskan nilai min sama dengan nilai varians seperti model regresi Poisson. Bagi mengatasi sifar melambung, model Binomial Negatif memperkenalkan parameter serakan yang membenarkan variansnya melebihi nilai min. Varians bagi model Binomial Negatif adalah:

$$\text{Var}(Y_i | X_i) = \mu_i + \alpha \mu_i^2$$

dengan  $\alpha$  adalah sifar melambung dalam model Binomial Negatif. Jika nilai  $\alpha = 0$ , maka sifar melambung tidak wujud dan Regresi Binomial Negatif diturunkan kepada Regresi Poisson. Namun begitu, jika nilai bagi  $\alpha$  melebihi sifar ( $\alpha > 0$ ) maka fenomena sifar melambung disahkan wujud dan model Regresi Poisson tidak lagi efektif digunakan ke atas kajian tersebut. Kes ini boleh diuji melalui ujian nisbah kebolehdjian (Lee & Nelder, 1996). Berdasarkan kajian yang dijalankan oleh Asrul dan Naing pada tahun 2012, mereka turut mencadangkan penggunaan model Regresi Poisson Sifar-Melambung bagi mengatasi lebih nilai sifar dan model tersebut turut diaplikasikan dalam penyelidikan ini.

#### **2.4 Sejarah Pemodelan Regresi Poisson**

Model Regresi Poisson merupakan satu pemodelan statistik yang diaplikasikan dalam sesuatu kajian bagi mengaitkan hubungan antara pemboleh ubah bersandar dengan pemboleh ubah tidak bersandar. Model ini tergolong dalam kumpulan GLM dan merupakan lanjutan bagi model regresi klasik. Hal ini seterusnya membolehkan nilai min bagi sesuatu populasi bergantung kepada pemboleh ubah tidak bersandar dan melalui satu fungsi pautan tidak linear (Silva & Rodrigues, 2014). Penggunaan model Regresi Poisson sangat luas malahan kajian terhadap model ini sentiasa berterusan. Ianya telah banyak dijalankan dalam pelbagai bidang termasuk ekonomi, epidemiologi, sosiologi dan perubatan. Pemodelan Regresi Poisson sangat berguna



apabila sesuatu kajian memiliki hasil dalam bentuk bilangan serta mengalami peristiwa yang jarang berlaku (Kutner et al., 2004).

Data bilangan yang digunakan ke atas model Regresi Poisson merupakan integer dan bukan negatif. Manakala taburan bagi jenis bilangan tersebut adalah dalam bentuk diskrit dan tidak berterusan. Oleh itu, model Regresi Poisson lebih cenderung untuk condong ke kanan sehingga mempunyai nilai yang positif (Atkins & Gallop, 2007). Berdasarkan kajian yang dijalankan oleh Frome dan Morris pada tahun 1989, Regresi Poisson turut memberi satu nilai anggaran bagi kebolehdajian maksimum, ujian hipotesis dan diagnostik regresi. Nilai anggaran ini telah digunakan secara meluas dalam kajian ‘*cohort*’ serta melibatkan penggunaan epidemiologi lain yang turut mengaitkan kadar bagi sesuatu peristiwa. Manakala kajian yang dijalankan oleh Frome et al. pada tahun 1973 turut menjelaskan bahawa kesetaraan bagi kebolehdajian maksimum adalah fungsi umum bagi regresi tak linear yang bertaburan Poisson.

## **2.5 Teori dan Penggunaan Model Regresi Poisson**

Kebelakangan ini, kaedah Regresi Poisson seringkali menjadi pilihan para penyelidik yang menjalankan kajian terhadap pemodelan yang melibatkan data berbentuk bilangan serta nilai pemboleh ubah bersandar bukan negatif. Hal ini antaranya bilangan kemalangan yang berlaku pada setiap tahun (Wiklund et al., 2011). Terdahulu, Fridstrom dan Ingebrigtsen (1991) telah menjalankan satu penyelidikan dengan membina model berdasarkan kaedah regresi Poisson. Pemodelan tersebut melibatkan pemboleh ubah bersandar yang terdiri daripada empat musim manakala

perubahan waktu siang dan malam pula dijadikan sebagai pemboleh ubah tidak bersandar. Dua penyelidik daripada Sweden telah mengadaptasikan teori Regresi Poisson ke dalam karya mereka iaitu Brude (1995) dan Forsman (2008). Melalui pemodelan Brude, beliau telah membuat anggapan mengenai bilangan kematian yang berlaku setiap tahun yang turut mengikuti taburan Poisson. Jumlah daripada kematian yang dijangkakan turut dimodelkan mengikut kaedah Regresi Poisson manakala pemboleh ubah bersandar melibatkan bilangan kematian yang telah berlaku. Pemboleh ubah tidak bersandar pula mengambil kira faktor lain seperti struktur jalan raya yang baik, pemandu yang berhemah dan sebagainya (Brude, 1995). Manakala kajian yang dijalankan oleh Forsman (2008) pula telah melanjutkan teori penggunaan Regresi Poisson dan ianya dinamakan sebagai model korelasi bersiri (*serial correlation model*). Daripada kajian tersebut dapat disimpulkan bahawa model korelasi bersiri lebih sesuai digunakan untuk memodelkan setiap bilangan kematian yang berlaku.

Walau bagaimanapun, model ini hanya dianggap sebagai pelengkap kepada teori Regresi Poisson kerana tiada kaedah yang lebih tepat untuk menganggar setiap parameter dalam kajian (Forsman, 2008). Menurut Forsman lagi, parameter serakan dapat digunakan bagi menggambarkan faktor skala antara varians dan min. Hal ini dilakukan bagi menyemak kemungkinan samada berlakunya situasi lebihan serakan (*overdispersion*). Walau bagaimanapun, sebelum andaian terhadap kejadian lebihan serakan dijalankan, faktor lain yang lebih sesuai untuk menganalisis kaedah Regresi Poisson turut dinilai (Olsson, 2002). Selain daripada itu, model Regresi Poisson turut digunakan sebagai satu kaedah untuk menyelesaikan kebanyakan masalah umum seperti mengkaji setiap kadar jenayah yang berlaku (Osgood, 2000). Malahan,