

**SUATU KAEDAH ALTERNATIF BAGI
PEMODELAN LINEAR DALAM BIOSTATISTIK**

MOHAMAD SHAFIQ BIN MOHD IBRAHIM

UNIVERSITI SAINS MALAYSIA

2018

**SUATU KAEDAH ALTERNATIF BAGI
PEMODELAN LINEAR DALAM BIOSTATISTIK**

oleh

MOHAMAD SHAFIQ BIN MOHD IBRAHIM

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

September 2018

PENGHARGAAN

السَّلَامُ عَلَيْكُمْ وَرَحْمَةُ اللَّهِ وَبَرَكَاتُهُ

Alhamdulillah syukur ke hadrat Allah SWT atas segala ilham dan rahmat-Nya, dapat menyiapkan tesis penyelidikan ini dengan jayanya. Ucapan setinggi-tinggi penghargaan kepada penyelia tesis saya iaitu Prof Madya Dr. Wan Muhamad Amir Bin Wan Ahmad atas segala tunjuk ajar beliau sepanjang pembelajaran saya bermula daripada Sarjana Muda, Sarjana dan sehingga pada peringkat Doktor Falsafah. Beliau telah banyak berkorban masa dan tenaga untuk meneliti dan memperbaiki hasil kerja saya dari semasa ke semasa dalam melaksanakan tesis ini. Saya amat berbangga mempunyai penyelia seperti beliau yang tidak lekang berkongsi ilmu dalam dunia penyelidikan terutamanya dalam bidang penerbitan jurnal dan buku. Semoga Allah S.W.T. membalas segala jasa dan budi baik beliau.

Saya juga ingin merakamkan jutaan terima kasih kepada Kerajaan Malaysia di bawah Kementerian Pengajian Tinggi (KPT) kerana menaja saya sepanjang tempoh pengajian Doktor Falsafah dengan tajaan biasiswa MyPhD. Saya juga ingin mengucapkan jutaan terima kasih kepada Pusat Pengajian Sains Pergigian (PPSG) dan Universiti Sains Malaysia (USM) amnya kerana memberi bantuan dana penyelidikan menerusi geran penyelidikan jangka pendek bernombor 304/PPSG/61313187. Di samping itu, saya juga ingin berterima kasih kepada kedua ibubapa saya yang tercinta iaitu Mohd Ibrahim Bin Kadiasa dan Maharifa Beve serta ahli keluarga yang lain yang banyak membantu dari segi sokongan dan galakan sepanjang pengajian sehingga peringkat Doktor Falsafah.

Saya juga ingin merakamkan terima kasih kepada Ain Atienza Ashikin yang meluangkan masanya untuk membaca tesis ini bagi penambahbaikan bahasa yang digunakan dalam tesis ini. Akhir sekali, ucapan terima kasih juga kepada sahabat-sahabat seperjuangan saya dari USM dan UMT yang turut sama memberi pandangan dan idea bertujuan penambahbaikan tesis ini. Jasa baik kalian amat saya hargai dan terima kasih diucapkan.

ISI KANDUNGAN

Penghargaan	ii
Senarai Jadual	x
Senarai Rajah	xiii
Senarai Singkatan	xv
Abstrak	xvii
Abstract	xix

BAB 1 : PENDAHULUAN

1.1	Pengenalan Kepada Regresi	1
1.2	Latar Belakang dan Perkembangan Penyelidikan	2
1.2.1	Pendekatan Kepada Kaedah Bootstrap	4
1.2.2	Pendekatan Kepada Kaedah Bayesian Secara Plot Pengecaman	5
1.2.3	Pendekatan Kepada Regresi Kabur	5
1.3	Keperluan Pendekatan Kaedah Regresi Linear Secara Alternatif	6
1.4	Pernyataan Masalah	7
1.5	Objektif Penyelidikan Umum	8
1.5.1	Objektif Penyelidikan Khusus	8
1.6	Persoalan Penyelidikan Umum	9
1.7	Skop Penyelidikan Kajian	9
1.8	Carta Alir Penyelidikan	12
1.9	Sumbangan Penyelidikan	13
1.10	Justifikasi Penyelidikan	14

1.11	Organisasi Tesis	14
------	------------------	----

BAB 2 : SOROTAN KESUSASTERAAN

2.1	Pengenalan	17
2.2	Sejarah Awal Penggunaan Perkataan Regresi	17
2.2.1	Model Regresi Linear	22
2.2.2	Model Regresi Tak Linear	23
2.2.2	Konsep dan Teori Kaedah Regresi Linear Berganda	24
2.2.3	Kaedah Regresi Kuasa Dua Terkecil Berpemberat	28
2.3	Penerangan Ringkas Kaedah Regresi Teguh	30
2.3.1	Teori Kaedah Penganggaran MM	32
2.4	Penerangan Ringkas Kaedah Bootstrap	33
2.4.1	Konsep yang Mendasari Kaedah Bootstrap	35
2.4.2	Kaedah Regresi dengan Bootstrap	36
2.5	Penerangan Ringkas Kaedah Bayesian	37
2.5.1	Teori, Konsep dan Aplikasi Kaedah Bayesian	38
2.5.2	Kaedah Regresi Bayesian	40
2.6	Penerangan Ringkas Kaedah Regresi Kabur	42
2.6.1	Konsep dan Teori Kaedah Regresi Kabur	42
2.7	Perkaitan antara Faktor-Faktor dengan Tekanan Darah Sistolik	45
2.7.1	Berat Bayi Baru Lahir	45
2.7.2	Umur Bayi Baru Lahir	46
2.8	Perkaitan antara Faktor-Faktor dengan Kanser Oral	47
2.8.1	Umur dan Jantina	47
2.8.2	Lokasi Kanser Oral	48

2.9	Perkaitan antara Faktor-Faktor dengan Karies Awal Kanak-Kanak	48
2.10	Rumusan	50
BAB 3 : METODOLOGI		
3.1	Pengenalan	51
3.2	Reka Bentuk Penyelidikan	52
3.3	Responden Penyelidikan Kajian	52
3.4	Jangka Masa dan Lokasi Penyelidikan	52
3.5	Saiz Sampel dan Teknik Persampelan	54
3.5.1	Pengiraan Saiz Sampel Bagi Kajian Kes I	55
3.5.2	Pengiraan Saiz Sampel Bagi Kajian Kes II	55
3.5.3	Pengiraan Saiz Sampel Bagi Kajian Kes III	56
3.5.4	Teknik Persampelan	56
3.6	Proses Penganalisaan Data Serta Pemilihan Pemboleh Ubah Kajian	57
3.6.1	Tatacara Pemilihan Pemboleh Ubah bagi Kajian Kes I	58
3.6.2	Tatacara Pemilihan Pemboleh Ubah bagi Kajian Kes II	58
3.6.3	Tatacara Pemilihan Pemboleh Ubah bagi Kajian Kes III	59
3.7	Perjalanan Proses Penyelidikan Secara Keseluruhan	59
3.8	Prosedur Pelaksanaan Metodologi SAS	61
3.8.1	Prosedur Pelaksanaan Input Data Dalam Perisian SAS	61
3.8.2	Prosedur Pelaksanaan Kaedah Regresi Linear Berganda	62
3.8.3	Prosedur Pelaksanaan Kaedah Regresi Teguh (Robust)	62
3.8.4	Prosedur Pelaksanaan Kaedah Bootstrap	64
3.8.5	Prosedur Pelaksanaan Data Berpemberat	65
3.8.6	Prosedur Pelaksanaan Kaedah Bayesian	66

3.8.7	Prosedur Pelaksanaan Kaedah Regresi Kabur	67
3.8.7(a)	Prosedur Pelaksanaan Pengiraan Nilai p (Pemboleh Ubah)	69
3.8.9	Prosedur Pelaksanaan Model Regresi Linear Secara Alternatif	70
BAB 4 : KEPUTUSAN		
4.1	Pengenalan	81
4.2	Kajian Kes I: Pemodelan Aras Sistolik	87
4.2.1	Pengecaman Titik Terpencil	87
4.2.2	Pemodelan Model Regresi Linear Berganda	88
4.2.2(a)	Penganggaran Selang Model I Bagi Kajian Kes I	89
4.2.3	Pemodelan Model Regresi Linear Berganda dengan Kaedah Bootstrap	90
4.2.3(a)	Penganggaran Selang Model II Bagi Kajian Kes I	91
4.2.4	Pemodelan Model Regresi Linear Berganda Dengan Kaedah Data Berpemberat, Bootstrap dan Bayesian	93
4.2.4(a)	Penganggaran Selang Model III Bagi Kajian Kes I	96
4.2.5	Pemodelan Model Regresi Kabur dengan Kaedah Data Berpemberat dan Bootstrap	97
4.2.6	Keputusan Kajian Kes I Secara Keseluruhan	99
4.3	Kajian Kes II: Pemodelan Saiz Tumor	102
4.3.1	Pengecaman Titik Terpencil	102
4.3.2	Pemodelan Model Regresi Linear Berganda	103
4.3.2(a)	Penganggaran Selang Model I Bagi Kajian Kes II	104

4.3.3	Pemodelan Model Regresi Linear Berganda dengan Kaedah Bootstrap	105
4.3.3(a)	Penganggaran Selang Model II Bagi Kajian Kes II	106
4.3.4	Pemodelan Model Regresi Linear Berganda Dengan Kaedah Data Berpemberat, Bootstrap dan Bayesian	107
4.3.4(a)	Penganggaran Selang Model III Bagi Kajian Kes II	111
4.3.5	Pemodelan Model Regresi Kabur dengan Kaedah Data Berpemberat dan Kaedah Bootstrap	113
4.3.6	Keputusan Kajian Kes II Secara Keseluruhan	117
4.4	Kajian Kes III: Pemodelan Karies Awal Kanak-Kanak	118
4.4.1	Pengecaman Titik Terpencil	118
4.4.2	Pemodelan Model Regresi Linear Berganda	119
4.4.2(a)	Penganggaran Selang Model I Bagi Kajian Kes III	120
4.4.3	Pemodelan Model Regresi Linear Berganda dengan Kaedah Bootstrap	121
4.4.3(a)	Penganggaran Selang Model II Bagi Kajian Kes III	122
4.4.4	Pemodelan Model Regresi Linear Berganda Dengan Kaedah Data Berpemberat, Bootstrap dan Bayesian	123
4.4.4(a)	Penganggaran Selang Model III Bagi Kajian Kes III	128
4.4.5	Pemodelan Model Regresi Kabur dengan Kaedah Data Berpemberat dan Bootstrap	129
4.4.6	Keputusan Kajian Kes III Secara Keseluruhan	132
BAB 5 : PERBINCANGAN DAN KESIMPULAN		
5.1	Pembaharuan Dalam Kaedah Regresi Linear Secara Alternatif	134

5.2	Kadar Kejituan Kaedah Regresi Linear Secara Alternatif	138
5.3	Ringkasan Kajian	146
5.4	Kesimpulan	147

BAB 6 : BATASAN, MANFAAT KAJIAN DAN CADANGAN KAJIAN

6.1	Batasan Kajian	149
6.2	Manfaat Daripada Hasil Penyelidikan	150
6.3	Cadangan Kajian	151

	RUJUKAN	154
--	---------	-----

LAMPIRAN

Lampiran A	Algoritma Keseluruhan Kajian Kes
Lampiran B	Output Keputusan Keseluruhan Kajian Kes
Lampiran C	Pembentangan
Lampiran D	Penerbitan Jurnal
Lampiran E	Penerbitan Buku
Lampiran F	Etika Penyelidikan

SENARAI JADUAL

		Halaman
Jadual 2.1	Notasi Umum Bagi Model Linear dan Model Tak Linear	24
Jadual 2.2	Struktur Data Bagi Regresi Linear Berganda	25
Jadual 2.3	Penerangan Berat Bagi Bayi	46
Jadual 2.4	Tekanan Darah Normal mengikut Kategori Umur bagi Bayi	46
Jadual 2.5	Insiden Kanser Oral dengan Umur dan Jantina	48
Jadual 3.1	Jangka Masa Penyelidikan	53
Jadual 3.2	Penerangan Bagi Pemboleh Ubah Dalam Kajian Kes I	58
Jadual 3.3	Penerangan Bagi Pemboleh Ubah Dalam Kajian Kes II	58
Jadual 3.4	Penerangan Bagi Pemboleh Ubah Dalam Kajian Kes III	59
Jadual 4.1a	Penerangan Data Bagi Kajian Kes I	82
Jadual 4.1b	Penerangan Data Bagi Kajian Kes II	82
Jadual 4.1c	Penerangan Data Bagi Kajian Kes III	82
Jadual 4.2a	Model I bagi Kajian Kes I	89
Jadual 4.2b	Pengiraan Purata Anggaran Selang Model I	90
Jadual 4.3a	Model II Bagi Kajian Kes I	91
Jadual 4.3b	Pengiraan Purata Anggaran Selang Model II	92
Jadual 4.4a	Model III Bagi Kajian Kes I	93
Jadual 4.4b	Pengiraan Purata Anggaran Selang Model III	96
Jadual 4.5a	Nilai a_c dan a_w	97
Jadual 4.5b	Model IV Bagi Kajian Kes I (Had Bawah)	98

Jadual 4.5c	Model IV Bagi Kajian Kes I (Had Atas)	98
Jadual 4.5d	Pengiraan Purata Anggaran Selang Model IV	99
Jadual 4.6	Keputusan Kajian Kes I Secara Keseluruhan	101
Jadual 4.7a	Model I Bagi Kajian Kes II	103
Jadual 4.7b	Pengiraan Purata Anggaran Selang Model I	105
Jadual 4.8a	Model II Bagi Kajian Kes II	105
Jadual 4.8b	Pengiraan Purata Anggaran Selang Model II	107
Jadual 4.9a	Model III Bagi Kajian Kes II	108
Jadual 4.9b	Pengiraan Purata Anggaran Selang Model III	112
Jadual 4.10a	Nilai ac dan aw	113
Jadual 4.10b	Model IV Bagi Kajian Kes II (Had Bawah)	114
Jadual 4.10c	Model IV Bagi Kajian Kes II (Had Atas)	114
Jadual 4.10d	Pengiraan Purata Anggaran Selang Model IV	115
Jadual 4.11	Keputusan Kajian Kes II Secara Keseluruhan	117
Jadual 4.12a	Model I Bagi Kajian Kes III	119
Jadual 4.12b	Pengiraan Purata Anggaran Selang Model I	121
Jadual 4.13a	Model II Bagi Kajian Kes III	121
Jadual 4.13b	Pengiraan Purata Anggaran Selang Model II	123
Jadual 4.14a	Model III Bagi Kajian Kes III	124
Jadual 4.14b	Pengiraan Purata Anggaran Selang Model III	129
Jadual 4.15a	Nilai ac dan aw	129
Jadual 4.15b	Model IV Bagi Kajian Kes III (Had Bawah)	130
Jadual 4.15c	Model IV Bagi Kajian Kes III (Had Atas)	130
Jadual 4.15d	Pengiraan Purata Anggaran Selang Model IV	131
Jadual 4.16	Keputusan Kajian Kes III Secara Keseluruhan	133

Jadual 5.1	Rumusan Kecekapan Model Bagi Kajian Kes I	139
Jadual 5.2	Rumusan Kecekapan Model Bagi Kajian Kes II	140
Jadual 5.3	Rumusan Kecekapan Model Bagi Kajian Kes III	141

SENARAI RAJAH

	Halaman
Rajah 1.1 Kerangka Konsep Penyelidikan	11
Rajah 1.2 Carta Alir Penyelidikan	12
Rajah 3.1 Proses Penganalisaan Data	57
Rajah 3.2 Lingkungan Kajian Penyelidikan Secara Keseluruhan	60
Rajah 3.3 Contoh Titik terpencil dalam Set Data Bagi Kajian Kes I	63
Rajah 3.4 Contoh Diagnosis Pemboleh Ubah Dalam Kajian Kes II	67
Rajah 3.5 Perjalanan Kajian Penyelidikan Model Regresi Linear Secara Alternatif	70
Rajah 4.1a Model Regresi Linear Berganda (Model I)	83
Rajah 4.1b Model Regresi Linear Berganda (Model II)	84
Rajah 4.1c Model Regresi Linear Berganda (Model III)	85
Rajah 4.1d Model Regresi Linear Berganda Secara Alternatif (Model IV)	86
Rajah 4.2 Set Data selepas Titik Terpencil dikeluarkan	88
Rajah 4.3a Diagnosis Pintasan bagi Pemalar	94
Rajah 4.3b Diagnosis Pemboleh ubah Berat	95
Rajah 4.3c Diagnosis Pemboleh ubah Umur	95
Rajah 4.4 Set Data Tiada Titik Terpencil	102
Rajah 4.5a Diagnosis bagi Pintasan Pemalar	109
Rajah 4.5b Diagnosis Pemboleh ubah Umur	109
Rajah 4.5c Diagnosis bagi Pemboleh ubah Jantina	110
Rajah 4.5d Diagnosis bagi Pemboleh Ubah Gusi	110
Rajah 4.5e Diagnosis bagi Pemboleh Ubah Pipi	111

Rajah 4.5f	Diagnosis bagi Pemboleh Ubah Lidah	111
Rajah 4.6	Set Data Tiada Titik Terpencil	118
Rajah 4.7a	Diagnosis bagi Pintasan Pemalar	125
Rajah 4.7b	Diagnosis bagi Perbelanjaan	125
Rajah 4.7c	Diagnosis bagi Pendapatan	126
Rajah 4.7d	Diagnosis bagi Skor Praktis	126
Rajah 4.7e	Diagnosis bagi Skor Sikap	127
Rajah 4.7f	Diagnosis bagi Skor Pengetahuan	127
Rajah 6.1	Cadangan Rekabentuk Penyelidikan Pemodelan Regresi	153

SENARAI SINGKATAN

MLR	Regressi Linear Berganda
BMLR	Regressi Linear Berganda dengan Kaedah Bootstrap
BWBMLR	Regressi Linear Berganda dengan Kaedah Bootstrap, Data berpemberat dan Bayesian
FBWBMLR	Regressi Linear Berganda dengan Kaedah Bootstrap, Data Berpemberat dan Regresi Kabur
SAS	Statistical Analysis System
R	Bahasa Pengaturcaraan dengan R
C++	Bahasa Pengaturcaraan Komputer
MCMC	Rantai Markov Monte Carlo
G*Power	Perisian G*Power
L1	Tatacara Norma-L1
LS	Kuasa Dua Terkecil
LTS	Kuasa Dua Trim Terkecil
M	Penganggar M
S	Penganggar Skala
MM	Penganggar MM
LMS	Median Kuasa Dua Terkecil
IID	Pemerhatian Tertabur Secara Secaman dan Tak Bersandar
TFNs	Nombor Kabur Dalam Bentuk Segitiga
USM	Universiti Sains Malaysia
PPSG	Pusat Pengajian Sains Pergigian

AC	Nilai Titik Tengah
AW	Nilai Radius

SUATU KAEDAH ALTERNATIF BAGI PEMODELAN LINEAR DALAM BIOSTATISTIK

ABSTRAK

Regresi Linear Berganda (MLR) merupakan sejenis model linear statistik yang biasa digunakan. Dengan perkembangan teknologi semasa dan peningkatan pembangunan berkaitan metodologi baru atau yang diubahsuai membawa kepada pembangunan kaedah secara alternatif bagi pengiraan model regresi linear berganda. Dalam kajian ini, model regresi linear berganda akan dikira secara komputasi statistik dengan menggunakan bahasa pengaturcaraan SAS yang mempertimbangkan kombinasi regresi teguh, bootstrap, data berpemberat, Bayesian dan kaedah regresi kabur. Pembangunan metodologi adalah berdasarkan kepada algoritma SAS (Perisian SAS 9.4) yang merupakan komputasi statistik yang lebih teguh dengan gabungan regresi teguh, bootstrap, data berpemberat, Bayesian dan kaedah regresi kabur. Tiga algoritma SAS yang berbeza iaitu (i) Regresi Linear Berganda dengan Kaedah Bootstrap (BMLR), (ii) Regresi Linear Berganda dengan Kaedah Bootstrap, Data berpemberat dan Bayesian (BWBMLR) dan (iii) Regresi Linear Berganda dengan Kaedah Bootstrap, Data Berpemberat dan Regresi Kabur (FBWMLR) akan dibandingkan secara berasingan dengan purata penganggaran selang lebar. Bagi menguji keberkesanan algoritma yang dibina, tiga kajian kes akan digunakan iaitu pemodelan aras sistolik tekanan darah, pemodelan saiz tumor dan pemodelan karies awal kanak-kanak (ECC). Purata penganggaran selang lebar semua model bagi ketiga-tiga kajian kes telah dikira dan dibandingkan. Selang lebar yang terkecil adalah menunjukkan model yang lebih baik. Hasil keputusan menunjukkan bahawa model

MLR mempunyai purata selang lebar iaitu 11.2948, 7.4816 dan 29.0407; dan model BMLR mempunyai purata selang lebar iaitu 2.5785, 3.7098 dan 19.6589. Model BWBMLR mempunyai purata selang lebar iaitu 2.3296, 3.5279 dan 16.3985; dan akhirnya, model FBWMLR mempunyai purata selang lebar iaitu 1.0066, 0.0058 dan 0.0092. Kesimpulannya, kaedah yang paling berkesan untuk mendapatkan hubungan antara pemboleh ubah bersandar dan pemboleh ubah tak bersandar adalah dengan menggunakan kaedah FBWMLR berbanding dengan kaedah MLR kerana purata penganggaran selang lebar yang terkecil.

AN ALTERNATIVE METHOD FOR LINEAR MODELING IN BIOSTATISTICS

ABSTRACT

Multiple Linear Regression (MLR) is the most common type of linear statistical model. Current technological advancement and increase in development of the new or modified methodology building leads to the development of an alternative method for multiple linear regression model calculation. In this study, multiple linear regression model will be calculated by using SAS programming language based on computational statistics which will consider combination of robust regression, bootstrap, weighted data, Bayesian and fuzzy regression method. Methodology building is based on the SAS algorithm (SAS 9.4 software) which is a robust computational statistics that consists of the combination of robust regression, bootstrap, weighted data, bayesian and fuzzy regression methods. Three different SAS algorithms that is; (i) Bootstrap Multiple Linear Regression (BMLR), (ii) Bootstrap Weighted Bayesian Multiple Linear Regression (BWBMLR) and (iii) Fuzzy Bootstrap Weighted Multiple Linear Regression (FBWMLR) will be compared separately according to their average width of prediction. To illustrate the potential of built-in algorithm, three cases of study will be used which are modeling on systolic blood pressure level, modeling on tumour size and modeling on Early Childhood Caries (ECC). The average width of prediction interval results for all cases of the models has been computed and compared. The smallest width will indicate the best fitting model. The result shows that the former MLR model has an average width of 11.2948, 7.4816 and 29.0407; and BMLR model has an average width of 2.5785, 3.7098 and 19.6589.

Where else, the BWMLR model has an average width 2.3296, 3.5279 and 16.3985; and FBWMLR model has an average width of 1.0066, 0.0058 and 0.0092. In conclusion, it is shown that the most efficient method to obtain a relationship between response and explanatory variables is to apply FBWMLR method compared to MLR method because of the small average width prediction interval.

BAB 1

PENDAHULUAN

1.1 Pengenalan Kepada Regresi

Bidang pemodelan regresi semakin mendapat tempat di kalangan penyelidik kerana bidang ini dapat membantu dalam proses pembuatan keputusan secara lebih tepat dan cekap. Hasil keputusan daripada pemodelan tersebut secara tidak langsung dapat juga digunakan sebagai panduan untuk proses peramalan. Hampir kesemua bidang sains dan bukan sains telah menggunakan kaedah regresi sebagai alat penyelidikan yang utama bagi mencapai objektif penyelidikan. Hal ini dapat dilihat terutamanya penyelidikan yang berteraskan kepada perhubungan yang melibatkan dua atau lebih pemboleh ubah (Mokthar, 1994).

Asas kepada pembentukan pemodelan regresi adalah merujuk kepada gabungan dua fenomena dapat diwakilkan ke suatu pembentukan persamaan yang melibatkan pemboleh ubah bersandar dan pemboleh ubah tak bersandar. Sifat kebergantungan sebeginilah yang telah mewujudkan satu corak perhubungan di antara pemboleh ubah, tidak kiralah sama ada hubungan tersebut linear mahupun tak linear. Secara ringkasnya, analisis regresi memerihalkan tingkah laku pemboleh ubah bersandar apabila diberikan nilai-nilai pemboleh ubah tak bersandar. Persamaan (1.1) dapat menunjukkan pemodelan model regresi linear secara umum (Draper & Smith, 1992, 1998; Neter *et al.*, 1996; Montgomery *et al.*, 2012):

$$Y = \beta_0 + \beta_1 X + \varepsilon \quad (1.1)$$

dengan

Y = Pemboleh Ubah Bersandar

X = Pemboleh Ubah Tak Bersandar

β_0 = Parameter Pintasan

β_1 = Parameter Pekali

ε = Ralat

Pemodelan regresi adalah merupakan suatu teknik analisis statistik yang digemari oleh penyelidik kerana sifatnya yang menghubungkan beberapa pemboleh ubah sehinggakan dapat meramal sesuatu keputusan yang berada dalam sesuatu julat kajian. Skop kajian penyelidikan ini merangkumi beberapa komponen yang melibatkan tatacara pembinaan kaedah regresi teguh, kaedah bootstrap, pembinaan data berpemberat, kaedah Bayesian dan kaedah regresi kabur. Turut disentuh adalah mengenai tatacara pembinaan model regresi secara alternatif dengan menggunakan bahasa pengaturcaraan SAS. Hal ini kerana perisian yang sedia ada di pasaran tidak menyediakan kemudahan kepada penyelidik bagi menganalisis data secara komprehensif seperti yang diutarakan di atas.

1.2 Latar Belakang dan Perkembangan Penyelidikan

Konsep asas model regresi telah diperkenalkan oleh Sir Francis Galton pada tahun 1877 dalam kertas kerja yang bertajuk "*Typical Laws of Heredity*" (Pearson, 1930; Birkes & Dodge, 1993; Stanton, 2001; Senn, 2011). Tahap penggunaan model regresi semakin meningkat dalam aktiviti penyelidikan sejurus ianya diperkenalkan. Pada tahun 1903 penemuan utama terhadap kajian yang telah dijalankan Karl Pearson telah menyokong kajian yang telah dilaksanakan oleh Galton. Bertitik tolak daripada idea

tersebut, Galton dan Pearson telah memanjangkan kajian penyelidikan mereka dengan meneruskan pemodelan pada tahap yang lebih tinggi iaitu pemodelan model regresi linear berganda dan hal ini telah membuka satu lembaran baru berkaitan perkembangan model regresi linear berganda dan seterusnya memberi impak yang tinggi terhadap pemodelan dalam pelbagai bidang sains dan bukan sains.

Terdapat juga beberapa kelemahan di dalam pemodelan yang telah diusulkan oleh Galton seperti hasil keputusan bagi sesuatu kajian menjadi tidak tepat dan menjauhi kesimpulan yang sebenar terutamanya kes yang melibatkan sampel data yang kecil. Masalah ini telah menarik perhatian Galton dan Pearson secara lebih serius untuk mengkaji permasalahan dengan lebih mendalam. Permasalahan ini juga telah menarik perhatian Goldberger (1964) dengan memperkenalkan satu kaedah yang dinamakan sebagai kaedah kuasa dua terkecil berpemberat bagi menangani masalah set data yang sedikit (lebih dikenali sebagai set data kecil) yang tidak memenuhi andaian model. Hal ini turut merangkumi anggapan kesamaan varians dan anggapan kenormalan ralat model. Kewujudan set data yang terhad menyebabkan penggunaan kaedah kuasa dua terkecil berpemberat dalam pemodelan amat digemari terutamanya bagi mengatasi masalah-masalah yang wujud dalam pemodelan.

Selanjutnya, kewujudan masalah titik terpencil juga turut menimbulkan masalah dalam menghasilkan keputusan yang lebih tepat. Sesetengah kajian penyelidikan yang dijalankan memberikan suatu nilai (dari set data yang dikaji) yang dikenali sebagai "Titik Terpencil". Kewujudan nilai sebegini menyebabkan nilai penganggaran akan tidak lagi cekap (penganggar pincang). Pada kebiasaannya penyelidik akan mengeluarkan nilai titik terpencil tersebut daripada set data kajian menerusi teknik

data screening. Walau bagaimanapun, kaedah regresi yang telah diperkenalkan oleh Huber pada tahun 1973 dapat mengatasi masalah tersebut menerusi pendekatan model regresi teguh dengan menggunakan kaedah penganggaran M. Kewujudan titik terpencil dapat diamati dengan menggunakan plot reja yang diplot hasil daripada keputusan yang diperolehi. Nilai yang diklasifikasi sebagai nilai titik terpencil dapat dilihat dengan jelas dan dapat disiasat kesahihannya sebelum tindakan selanjutnya diambil. Lanjutan daripada teknik yang telah diperkenalkan Huber, penyelidikan mengenai regresi teguh telah diperluaskan kepada teknik penganggaran MM oleh Yohai (1987). Kaedah penganggaran MM ini adalah didasarkan kepada kaedah yang menggabungkan kaedah titik musnah yang tinggi dan penganggar statistik cekap.

1.2.1 Pendekatan Kepada Kaedah Bootstrap

Penggabungan kaedah bootstrap dengan pemodelan regresi merupakan satu kaedah alternatif bagi menyelesaikan masalah kewujudan sampel yang kecil. Pada asalnya kaedah ini telah diperkenalkan pada tahun 1979 oleh penyelidik statistik Amerika Syarikat (USA) yang bernama Sir Efron. Keteguhan kaedah bootstrap (menerusi penghasilan replikasi data secara proses persampelan) dalam pemodelan regresi dapat dilihat menerusi hasil penyelidikan yang telah dijalankan oleh Efron dan Tibshirani (1993), Booth dan Sarkar (1998), Davidson dan McKinnon (2000). Secara umumnya, kaedah bootstrap ini diperkenalkan untuk menganggar suatu taburan statistik menerusi pendekatan taburan empirik bagi menambah baik ketepatan nilai penganggaran statistik. Dengan wujudnya kaedah bootstrap dalam model regresi, perkembangan penyelidikan terhadap bidang pemodelan menjadi semakin pesat (Freedman, 1981; Jeong & Maddala, 1993; Vinod, 1993; Veall, 1998; Chernick, 2011).

Senario penyelidikan pada masa kini adalah lebih kondusif kerana tatacara pembinaan model dapat dilaksanakan dengan begitu pantas dan cepat melalui penggunaan kaedah berkomputer dan juga terdapat pelbagai pilihan bahasa pengaturcaraan untuk penggunaan sains telah disediakan yang meliputi Bahasa R, SAS, C++ dan banyak lagi. Sesuai dengan perkembangan teknologi pada masa kini, penggunaan kaedah komputer dapat meningkatkan lagi tahap kejituan kaedah yang dianggarkan sehingga ianya dapat meminimumkan ralat dalam pengiraan.

1.2.2 Pendekatan Kepada Kaedah Bayesian Secara Plot Pengecaman

Kajian penyelidikan ini turut memberi fokus kepada kaedah Rantai Markov Monte Carlo (MCMC) secara aplikasi dengan menggunakan bahasa pengaturcaraan SAS. Prosedur Bayesian yang digunakan di dalam kaedah penyelidikan ini adalah diasaskan kepada kaedah persampelan Markov Monte Carlo. Menerusi prosedur ini, beberapa plot yang berkaitan dengan taburan sampel posterior akan dihasilkan mengikut setiap parameter yang dikaji. Maklumat yang berkaitan dengan kestabilan parameter, kecukupan model dan juga tahap kesesuaian parameter dihasilkan menerusi corak plot yang dihasilkan oleh Rantai Markov (corak yang menumpu). Sekiranya corak rantai Markov Monte Carlo tidak menumpu diperolehi terhadap sesuatu pemboleh ubah maka sampel tersebut tidak dapat menganggar taburan sampel tersebut dengan lebih baik (Stokes *et al.*, 2014).

1.2.3 Pendekatan Kepada Regresi Kabur

Konsep asas kaedah regresi kabur telah diperkenalkan oleh Tanaka pada tahun 1982. Idea asas beliau sering dirujuk sebagai regresi yang berkebarangkalian dapat meminimum kekaburan model dengan meminimumkan jumlah penyebaran koefisien

kabur. Menurut Shapiro (2004) dan Kahraman *et al.* (2006) model regresi statistik mempunyai banyak aplikasi, namun begitu pemodelan regresi masih juga mempunyai beberapa kekurangan terutamanya apabila berhadapan dengan set data bersaiz kecil, masalah dalam membuat andaian taburan bagi data, terdapat kekaburan dalam hubungan di antara input dan output pemboleh ubah dan akhir sekali masalah ketidaktepatan dan penyelewengan yang disebabkan oleh model linear.

1.3 Keperluan Pendekatan Model Regresi dengan Kaedah Secara Alternatif

Apabila menjalankan sesuatu kajian yang melibatkan pemodelan statistik, telah menjadi lumrah kepada setiap penyelidik untuk menghasilkan model yang terbaik dalam kajian mereka. Namun begitu, keterbatasan dalam melaksanakan sesuatu kajian sering kali berlaku terutamanya melibatkan kaedah yang sedia ada. Di samping itu, permasalahan seperti kewujudan titik terpencil, ketidakcukupan data turut berlaku sehingga menghasilkan satu keputusan yang tidak tepat. Oleh hal yang demikian, keperluan kepada penambahbaikan kaedah regresi linear berganda secara alternatif amat penting bagi menghasilkan keputusan yang lebih baik. Justeru, dalam penyelidikan ini, satu kaedah regresi linear berganda secara alternatif telah dibina untuk penambahbaikan model regresi linear berganda sedia ada. Kaedah regresi linear berganda secara alternatif ini adalah melibatkan gabungan kaedah regresi teguh, bootstrap, data berpemberat, Bayesian dan regresi kabur dengan menggunakan bahasa pengaturcaraan SAS.

1.4 Pernyataan Masalah

Pada bahagian ini, beberapa pernyataan masalah akan diutarakan berdasarkan kepada kajian lepas dan isu-isu yang berkaitan dengannya. Isu pertama yang akan dibentangkan dalam bahagian ini adalah berkaitan kewujudan titik terpencil yang akan mengganggu keputusan terhadap penganggaran parameter. Hal ini secara tidak langsung akan menyebabkan keputusan yang diperolehi tidak tepat. Menurut Rousseeuw dan Hubert (2011) apabila terdapat titik-titik terpencil dalam data dan hal ini akan menyebabkan keputusan yang diperolehi adalah tidak tepat.

Pernyataan masalah yang kedua adalah berkaitan kecukupan saiz sampel dalam sesuatu kajian. Pada kebiasaannya, kajian yang berdasarkan reka bentuk ujikaji melibatkan kos yang sangat tinggi terutamanya apabila melibatkan bahan-bahan kimia yang mahal dan sangat terhad. Justeru, saiz sampel yang dipungut hasil daripada kajian tersebut adakalanya mencukupi dan adakalanya tidak mencukupi disebabkan beberapa faktor kekangan seperti sumber kewangan dan lain-lain lagi. Oleh hal yang demikian, keperluan kepada teknik seperti persampelan semula iaitu kaedah bootstrap bagi mencukupkan data pada tahap yang minimum sangat diperlukan.

Isu seterusnya adalah merujuk kepada kurangnya penggunaan kaedah data berpemberat dalam analisis data. Menurut Carrol dan Ruppert (1988) penggunaan data berpemberat dapat memberikan keputusan penganggaran parameter model dengan lebih tepat. Kebanyakan perisian statistik sedia ada dipasaran tidak membekalkan analisis ini secara terus malah penyelidik terpaksa membina bahasa pengaturcaraan sendiri dengan menggunakan medium pengaturcaraan seperti C++ dan sebagainya.

Selanjutnya, isu keempat adalah merujuk kepada keterbatasan kaedah sedia ada dalam mendapatkan penganggaran parameter yang lebih tepat bagi model. Oleh itu, algoritma secara bahasa pengaturcaraan SAS melalui regresi kabur telah dibina bagi mendapatkan penganggaran parameter model tersebut. Oleh hal demikian, kajian yang dijalankan adalah mempertimbangkan kaedah regresi teguh, bootstrap, data berpemberat, Bayesian dan regresi kabur bagi menyelesaikan isu-isu permasalahan yang wujud dalam model regresi linear.

1.5 Objektif Penyelidikan Umum

Kajian ini memberi fokus kepada tatacara pembinaan model regresi linear secara alternatif yang memberi fokus kepada teknik-teknik pemodelan yang berteraskan kepada statistik komputasi dengan menggunakan perisian SAS.

1.5.1 Objektif Penyelidikan Khusus

- i. Membangunkan suatu kaedah pemodelan regresi linear yang dibina dengan menggunakan bahasa pengaturcaraan SAS yang mempertimbangkan gabungan kaedah regresi teguh, bootstrap, data berpemberat, Bayesian dan regresi kabur.
- ii. Membuat perbandingan nilai kebarangkalian (nilai p) pemboleh ubah yang signifikan antara model regresi linear yang dibina dengan model regresi linear yang sedia ada.
- iii. Menentukan tahap kecekapan antara model regresi linear yang dibina dengan model regresi linear yang sedia ada.

1.6 Persoalan Penyelidikan

Berikut adalah merupakan beberapa persoalan kajian yang mendasari penyelidikan yang dijalankan.

- i. Adakah model regresi linear yang dibina dapat menganggar parameter dengan lebih baik berbanding model regresi linear yang sedia ada?
- ii. Adakah terdapat perubahan yang signifikan dari segi nilai kebarangkalian (nilai p) secara spesifik bagi kaedah regresi linear yang dibina dengan kaedah regresi linear yang sedia ada?
- iii. Adakah terdapat perbezaan kecekapan antara model regresi linear yang dibina berbanding dengan model regresi linear yang sedia ada?

1.7 Skop Penyelidikan

Skop kajian yang mendasari penyelidikan ini adalah bersifat pemodelan regresi linear. Dalam kajian ini, satu kaedah alternatif iaitu kaedah yang telah dibina telah diperkenalkan dengan melibatkan gabungan beberapa kaedah statistik utama yang memberi fokus kepada dapatan kajian yang lebih tepat. Secara konsep ianya dapat dilihat dalam Rajah 1.1. Seterusnya, kajian ini mengetengahkan tiga fasa utama kajian iaitu Fasa Pertama adalah mengenai pembinaan algoritma bagi model regresi linear secara alternatif, diikuti dengan Fasa Kedua yang melibatkan pengujian dan pelaksanaan bagi analisis data berdasarkan algoritma yang telah dibina. Akhir sekali, Fasa Ketiga adalah membuat perbandingan antara model regresi linear yang dibina dengan model regresi linear yang sedia ada. Keterangan lanjut mengenai ketiga-tiga fasa ini adalah seperti berikut:

Fasa Pertama: Pembinaan Algoritma

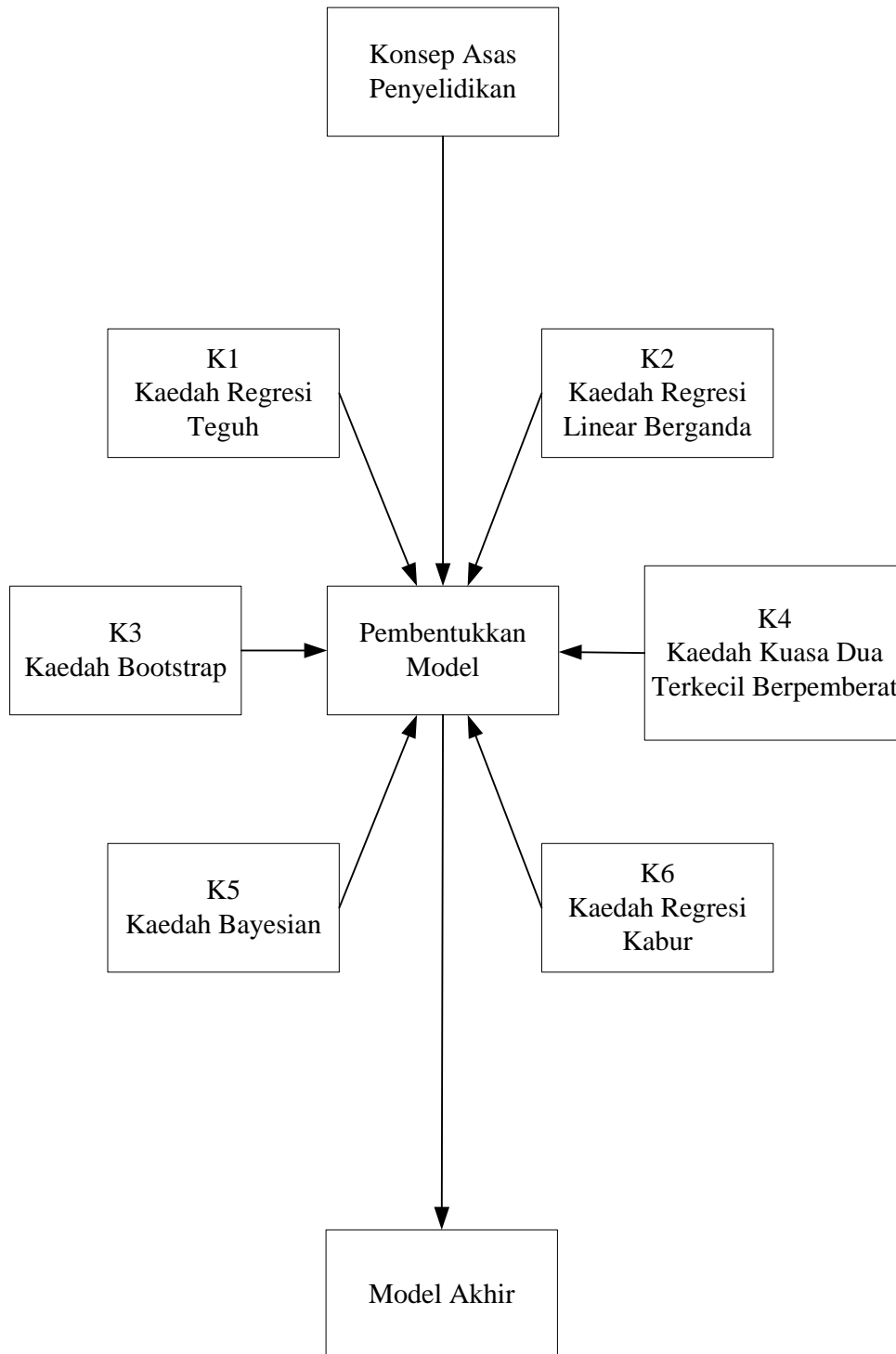
Dalam fasa ini, pembinaan algoritma dilaksanakan terlebih dahulu dengan menggunakan bahasa pengaturcaraan SAS berdasarkan kepada sorotan kajian dan kelemahan pemodelan lepas. Seterusnya, pembinaan algoritma mempertimbangkan kaedah regresi teguh, bootstrap, data berpemberat, Bayesian dan regresi kabur. Perincian mengenai pembinaan algoritma bagi setiap kaedah dapat dirujuk pada Bab 3 iaitu pada bahagian 3.8 prosedur pelaksanaan metodologi SAS.

Fasa Kedua: Pengujian dan Pelaksanaan Analisis Data

Pada bahagian ini, algoritma yang ditulis dengan menggunakan bahasa pengaturcaraan SAS akan diuji dengan lebih spesifik terhadap tiga kajian kes utama yang dipertimbangkan dalam kajian ini. Perisian G*Power telah digunakan untuk membuat anggaran saiz sampel minimum yang diperlukan bagi ketiga-tiga kajian kes. Bagi kes yang saiz sampelnya tidak mencapai tahap minimum, kaedah bootstrap akan digunakan untuk memastikan keperluan saiz sampel minimum dipatuhi.

Fasa Ketiga: Perbandingan Model

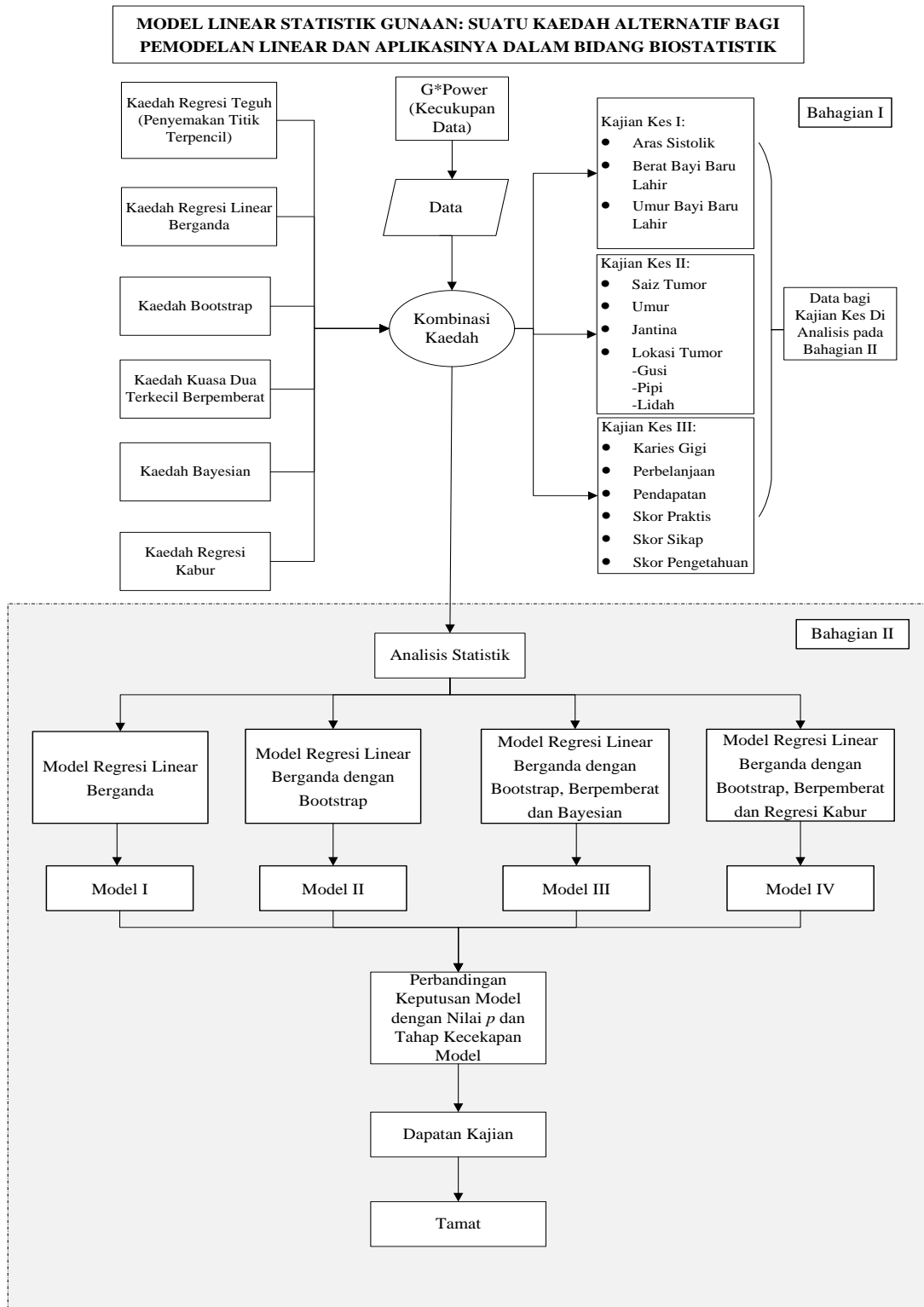
Dalam fasa ini, empat model utama dibandingkan iaitu pertamanya model regresi linear berganda dan keduanya ialah model regresi linear berganda dengan menggunakan kaedah bootstrap. Seterusnya, model ketiga ialah model regresi linear berganda dengan kaedah bootstrap, data berpemberat dan Bayesian. Akhir sekali, model yang keempat adalah model regresi linear berganda dengan kaedah bootstrap, data berpemberat dan regresi kabur. Perincian berkenaan hasil keputusan bagi setiap perbandingan model mengikut kajian kes boleh dirujuk pada Bab 5.



*K1 sehingga K6 merupakan kaedah yang digabungkan bagi membentuk model regresi secara alternatif

Rajah 1.1: Kerangka Konsep Penyelidikan

1.8 Carta Alir Penyelidikan



Rajah 1.2: Carta Alir Penyelidikan

1.9 Sumbangan Penyelidikan

Kajian penyelidikan ini memberi beberapa sumbangan utama yang dapat diterjemahkan melalui bahasa pengaturcaraan SAS dan dapatan kajian yang dilaksanakan terutamanya dalam bidang perubatan dan pergigian. Sumbangan utama dalam kajian ini adalah pembinaan tatacara model regresi linear secara alternatif melalui bahasa pengaturcaraan SAS yang melibatkan beberapa gabungan kaedah utama iaitu kaedah regresi teguh, bootstrap, data berpemberat, Bayesian dan regresi kabur.

Pembinaan model regresi linear secara alternatif ini juga dapat membuka satu lembaran baru dalam bidang penyelidikan dengan suatu pakej model regresi yang lebih lengkap berbanding model regresi linear yang sedia ada dalam mendapatkan keputusan yang lebih tepat dan jitu. Malahan, ianya juga turut dapat membantu para penyelidik dalam menyelesaikan masalah yang wujud dalam pemodelan model regresi linear terutamanya yang melibatkan masalah kecukupan data.

Selain itu, model regresi linear secara alternatif merupakan kaedah penganggaran yang lebih meyakinkan dengan penglibatan beberapa kaedah penyelidikan yang lebih teguh. Bukan itu sahaja, keputusan berdasarkan model regresi linear secara alternatif turut meyakinkan para penyelidik dalam menganalisis data dan seterusnya mendapatkan keputusan yang lebih baik berbanding model regresi linear yang sedia ada. Oleh hal demikian, kewujudan kaedah analisis yang lebih komprehensif ini dapat membantu para penyelidik terutamanya para doktor perubatan dan pergigian dalam mendapatkan suatu templat dan tatacara yang lebih lengkap bagi menganalisis data dengan berbantuan perisian SAS.

1.10 Justifikasi Penyelidikan

Berikut adalah merupakan beberapa justifikasi kajian penyelidikan yang dilaksanakan:

1. Keperluan analisis secara statistik yang lebih berfokus.
2. Menyelesaikan beberapa masalah yang ketara dalam penyelidikan terutamanya masalah kecukupan data dan kewujudan titik terpencil.
3. Mempertingkatkan tahap keupayaan kualiti analisis data dari segi penganggaran parameter.

1.11 Organisasi Tesis

Penyelidikan ini dibuat bertujuan memberi fokus kepada tatacara pembinaan model regresi linear secara alternatif yang memberi fokus kepada teknik-teknik pemodelan yang berteraskan kepada statistik komputasi dengan menggunakan perisian SAS. Penggunaan kaedah secara alternatif bagi pemodelan regresi linear diaplikasikan kepada tiga kajian kes iaitu pemodelan aras tekanan darah sistolik, pemodelan saiz tumor dan pemodelan karies awal kanak-kanak. Kajian ini mengandungi sebanyak enam bab yang merangkumi tiga fasa utama iaitu pembinaan algoritma bagi setiap model, diikuti dengan pelaksanaan analisis data dan akhirnya perbandingan keputusan model yang menggunakan kaedah regresi linear secara alternatif dengan kaedah regresi linear yang sedia ada.

Bab 1 dimuatkan tentang pengenalan berkaitan penyelidikan yang dijalankan. Malah, aspek-aspek lain yang turut dimuatkan dalam bab ini adalah seperti latar belakang dan perkembangan penyelidikan, pernyataan masalah, objektif penyelidikan, persoalan penyelidikan, skop penyelidikan, carta alir penyelidikan, sumbangan penyelidikan dan

justifikasi penyelidikan. Bab 2 menerangkan dengan lebih terperinci mengenai sorotan kesusasteraan berkaitan kaedah regresi linear berganda, kaedah bootstrap, kaedah kuasa dua terkecil berpemberat, kaedah Bayesian dan kaedah regresi kabur.

Bab 3 menerangkan tentang reka bentuk penyelidikan yang dijalankan meliputi responden dan lokasi penyelidikan serta jangka masa penyelidikan. Penentuan saiz sampel dengan menggunakan perisian G*Power turut diterangkan bagi mendapatkan saiz sampel yang minimum bagi setiap kajian kes. Turut disertakan dalam bab ini ialah proses pengumpulan data serta pemilihan pemboleh ubah bagi setiap kajian kes. Selanjutnya, tatacara pembinaan algoritma bagi setiap model dengan menggunakan bahasa pengaturcaraan SAS turut dimuatkan bagi memudahkan pemahaman penggunaan SAS bagi pelaksanaan analisis model regresi linear berganda secara alternatif.

Bab 4 memaparkan hasil keputusan kajian yang dijalankan menerusi model regresi linear yang dibina secara alternatif melalui gabungan beberapa kaedah utama iaitu kaedah regresi teguh, bootstrap, data berpemberat, Bayesian dan regresi kabur. Setiap hasil keputusan yang dijalankan akan dijelaskan secara terperinci keputusannya bagi memudahkan lagi pemahaman tentang penyelidikan yang dijalankan.

Bab 5 merupakan bab yang menekankan perbincangan dan ringkasan tentang penyelidikan yang dijalankan. Aspek-aspek yang diberi penekanan dan tumpuan utama adalah seperti kecekapan pemodelan regresi linear berganda yang dibina secara alternatif. Bab 6 merupakan bab yang terakhir yang memberi penerangan tentang batasan kajian, manfaat daripada hasil penyelidikan dan cadangan-cadangan untuk

kajian model regresi linear secara alternatif dengan menggunakan pendekatan pemodelan regresi linear yang lain untuk kajian pada masa akan datang.

BAB 2

SOROTAN KESUSASTERAAN

2.1 Pengenalan

Bab ini akan membincangkan tentang perkembangan sejarah awal model regresi linear dan juga model regresi tak linear. Selain itu, bahagian ini juga menerangkan secara ringkas berkaitan kaedah regresi linear berganda, regresi kuasa dua terkecil berpemberat, regresi teguh, bootstrap, Bayesian dan regresi kabur. Turut dimuatkan dalam bab ini adalah tentang keperluan terhadap pembentukan suatu model statistik yang lebih teguh dengan mempertimbangkan kaedah regresi linear secara alternatif dengan gabungan beberapa kaedah statistik dalam mendapatkan hasil keputusan yang lebih jitu.

2.2 Sejarah Awal Penggunaan Perkataan Regresi

Sir Francis Galton yang juga merupakan seorang sarjana British yang dilahirkan di Birmingham, England pada 16 Februari 1822 telah memperkenalkan konsep regresi dalam satu kertas kerja yang bertajuk “*Typical Laws of Heredity*” pada tahun 1877 di England (Pearson, 1930; Birkes & Dodge, 1993; Stanton, 2001; Senn, 2011). Beliau ialah seorang penjelajah dan ahli antropologi yang terkenal dengan pengajian *eugenik* (Kajian tentang perancangan penghasilan mutu kualiti terhadap manusia untuk generasi yang akan datang, melalui prinsip keturunan) dan kecerdasan manusia (Senn, 2011). Kajian beliau lebih memberi fokus kepada sebab musabab yang melibatkan perhubungan di antara pemboleh ubah.

Pada tahun 1886, Francis Galton telah menjalankan eksperimen untuk menentukan kadar pertumbuhan kacang pea terhadap saiz benihnya. Beliau telah membahagikan benih kepada tujuh kumpulan yang berasingan. Bacaan nilai diameter 100 benih yang dihasilkan oleh pokok kacang tersebut telah diambil dan nilai purata telah dicatatkan. Hasil keputusan mendapati bahawa kadar pertumbuhan benih kacang yang bersaiz kecil adalah sangat tinggi berbanding kadar pertumbuhan benih kacang yang bersaiz besar. Beliau juga mendapati kebanyakan purata diameter bagi anak kacang adalah bersifat “*regress*” dalam lingkungan purata yang biasa. Hal ini telah menarik perhatian beliau untuk mengkaji secara lebih mendalam. Justeru, satu perwakilan secara persamaan diperlukan bagi mengkaji hal ini secara lebih fokus. Pemodelan yang telah dilakukan oleh beliau mendapati bahawa terdapat dua perkara penting iaitu pertamanya diameter bagi anak kacang mempunyai hubungan linear dengan diameter bagi pokok kacang dan keduanya, purata diameter bagi anak kacang adalah bersifat “*regress*” dalam lingkungan purata yang biasa.

Selain itu, Francis Galton juga telah mengkaji hubungan antara ukuran ketinggian anak dengan ukuran ketinggian bagi ibu bapa mereka. Beliau telah menggunakan set data yang telah direkodkan sebanyak 205 set (Wachsmuth *et al.*, 2003). Berdasarkan kajian tersebut, beliau mendapati bahawa sekiranya ukuran ketinggian kedua ibu bapa mereka rendah maka anak-anak mereka adalah sedikit lebih tinggi daripada mereka. Manakala, jika ukuran ketinggian kedua-dua ibu bapa mereka tinggi, maka ukuran ketinggian anak-anak mereka adalah sedikit rendah daripada mereka. Beliau telah menamakan fenomena ini sebagai “*Regression Toward Mediocrity*” (Zimmerman, 1992). Bagi mengilustrasikan kes sebegini, beliau telah menamakan fenomena ini sebagai “*revert*” dan akhirnya ditukarkan kepada “*regress*” (Schmittlein, 1989).

Lanjutan daripada eksperimen tersebut, Francis Galton menemukan perkataan regresi bagi menjelaskan situasi yang berhubungan di antara satu sama lain. Hal ini telah membawa kepada pembangunan konsep awal regresi linear.

Regresi ditakrifkan sebagai proses pemodelan yang menghampiri kepada nilai min terhadap sesuatu pemboleh ubah (Stigler, 1997; Barnett *et al.*, 2004). Hal ini telah dibuktikan menerusi dapatan bagi kedua-dua eksperimen yang melibatkan pengujian terhadap ukuran kekacang peas dan juga ukuran ketinggian individu. Keputusan terhadap kedua-dua kajian menunjukkan proses pemodelan yang menghampiri kepada nilai min sesuatu pemboleh ubah. Dapatan hasil kajian Karl Pearson telah menyokong sepenuhnya keputusan kajian yang telah dipelopori oleh Francis Galton pada tahun 1903. Kajian Pearson dan Lee (1903) telah memberi fokus kepada pembangunan terhadap suatu perhubungan antara ketinggian anak dan bapa yang dapat diringkaskan menerusi persamaan (2.1) seperti berikut (Matthews & Farewell, 2015):

$$Y = 33.73 + 0.516X \quad (2.1)$$

dengan

33.73 adalah merupakan pintasan

Y = Ketinggian Anak

X = Ketinggian Bapa

Dalam kajian tersebut, beliau berjaya mengumpulkan data mengenai ketinggian sebanyak 1078 orang dikalangan bapa dan anak. Hasil daripada kajian tersebut, beliau telah menemukan beberapa penemuan yang sangat penting iaitu:

- i. purata ketinggian bagi bapa ialah 68 inci dan purata ketinggian bagi anak ialah 69 inci.
- ii. bapa yang mempunyai ketinggian sebanyak 72 inci pula mempunyai anak dengan purata ketinggian sebanyak 71 inci.
- iii. Manakala, bapa yang mempunyai ketinggian sebanyak 64 inci pula mempunyai anak dengan purata ketinggian sebanyak 67 inci.

(Pearson & Lee, 1903)

Hal ini menunjukkan bahawa ukuran ketinggian di kalangan bapa dan ukuran ketinggian di kalangan anak adalah bergerak secara hampir selari (*regress*) secara keseluruhan bagi populasi tersebut. Seterusnya, hasil daripada idea Sir Francis Galton telah membuka satu lembaran yang sangat meluas dalam penggunaan regresi dalam pelbagai bidang seperti Sains Fizik, Sains Hayat, Ekonomi, Perniagaan dan Pengurusan, Sains Kemasyarakatan, Kejuruteraan dan Sains Teknologi dan lain-lain lagi. Menurut Montgomery dan Peck (2012), analisis regresi merupakan satu kaedah statistik bagi mengetahui dan memodelkan hubungan antara pemboleh ubah. Bagi Neter *et al.* (1996), analisis regresi merupakan suatu kaedah statistik yang digunakan bagi mengkaji hubungan antara dua atau lebih pemboleh ubah yang bersifat kuantitatif supaya satu pemboleh ubah dapat diramal berdasarkan daripada pemboleh ubah yang lain. Bagi Mokhtar (1994), analisis regresi merupakan suatu kaedah statistik yang digunakan bagi mengkaji hubungan antara dua pemboleh ubah atau lebih.

Mokhtar (1994) telah memperihalkan tentang hubungan antara dua atau lebih pemboleh ubah dengan menggunakan analisis regresi bagi meramalkan nilai pemboleh ubah. Dalam analisis regresi ini, pemboleh ubah dapat dikategorikan kepada dua jenis

iaitu pemboleh ubah bersandar dan pemboleh ubah tak bersandar. Lazimnya, pemboleh ubah bersandar juga dikenali sebagai pemboleh ubah sambutan, manakala pemboleh ubah tak bersandar pula dikenali sebagai pemboleh ubah penerang, peramal atau peregrasi (Mokthar, 1994).

Secara umum, model regresi terbahagi kepada dua model, iaitu model regresi linear dan model regresi tak linear. Model regresi linear dalam bahasa yang mudah ialah apabila dua pemboleh ubah yang mempunyai hubungan dalam satu garis yang lurus, manakala model regresi tak linear ialah apabila dua pemboleh ubah yang tidak mempunyai hubungan dalam satu garis yang lurus. Model regresi linear merupakan suatu model regresi yang sering digunakan oleh penyelidik dalam pelbagai bidang. Hal ini kerana, dengan penggunaan model regresi linear ianya menyediakan kemudahan yang lebih fleksibel berbanding model regresi tak linear dalam merangka kerja-kerja penyelidikan yang bersesuaian dengan penyelidikan.

Walau bagaimanapun, model regresi linear juga mempunyai kekurangan dan kadangkala ia juga tidak sesuai diaplikasikan dalam sesetengah keadaan. Kekurangan yang dimaksudkan adalah merujuk kepada fungsi ketaklinearan antara pemboleh ubah sambutan dan pemboleh ubah peramal. Keterbatasan ini seterusnya mewujudkan kepada kaedah pemodelan regresi tak linear. Penerangan lebih lanjut berkenaan model regresi linear dan model regresi tak linear ditunjukkan pada Bahagian 2.2.1 dan 2.2.2.

2.2.1 Model Regresi Linear

Menurut Montogemery dan Peck (2012) persamaan bagi model regresi linear adalah seperti berikut:

$$y = \beta_0 + \beta_1 x + \varepsilon \quad (2.2)$$

Model regresi linear tidak terhad kepada peringkat pertama sahaja seperti yang kita lihat dalam persamaan (2.2) malah ianya juga boleh dikembangkan kepada peringkat kedua, ketiga dan juga peringkat yang seterusnya (Salleh *et al.*, 2014; Ahmad *et al.*, 2016c). Selain itu, model regresi linear juga dapat ditulis seperti persamaan (2.3) yang mana z diwakili fungsi regresi bagi x termasuk transformasi seperti $\exp(x)$, \sqrt{x} dan $\sin(x)$. Kesemua model ini adalah model regresi linear kerana ianya adalah linear dalam parameter yang tidak diketahui β_1 walaupun mempunyai pemboleh ubah yang tak linear.

$$y = \beta_0 + \beta_1 z + \varepsilon \quad (2.3)$$

Seterusnya, persamaan model regresi linear secara umum dapat ditulis seperti dalam persamaan (2.4).

$$y = x' \beta + \varepsilon = f(x, \beta) + \varepsilon \quad (2.4)$$

dengan $X' = [1, X_1, X_2, \dots, X_k]$. Oleh sebab nilai bagi jangkaan daripada ralat model adalah sifar, maka nilai jangkaan bagi pemboleh ubah sambutan adalah $E(y) = E[f(x, \beta) + \varepsilon] = f(x, \beta)$. Kebiasaannya, $f(x, \beta)$ adalah merujuk sebagai fungsi jangkaan bagi model. Oleh itu, dapat dilihat dimana fungsi jangkaan hanyalah berperanan sebagai fungsi linear bagi parameter yang tidak diketahui.

2.2.2 Model Regresi Tak Linear

Model regresi linear tidak dapat diaplikasikan dalam semua keadaan bagi menyelesaikan sesuatu masalah dalam pemodelan. Oleh yang demikian, model regresi tak linear dapat digunakan bagi mendapatkan penyelesaian sesuatu masalah terutamanya apabila terdapat parameter yang tak linear dalam model. Sesuatu model dapat ditakrifkan sebagai model regresi tak linear apabila terdapat parameter yang tidak diketahui yang tak linear dalam model. Menurut Draper dan Smith (1992, 1998), model regresi tak linear dalam bentuk umum dapat ditakrifkan seperti dalam persamaan (2.5).

$$y = f(\xi, \theta) + \varepsilon \quad (2.5)$$

yang mana θ ialah $p \times 1$ bagi vektor parameter yang tidak diketahui dan ε adalah korelasi ralat secara rawak dengan $E(\varepsilon) = 0$ dan $Var(\varepsilon) = \sigma^2$. Seterusnya, ralat dalam model ini dianggap bertaburan secara normal seperti dalam model regresi linear. Oleh itu, $f(\xi, \theta)$ dipanggil fungsi jangkaan dengan $E(y) = E[f(\xi, \beta) + \varepsilon] = f(\xi, \beta)$ sebagai model regresi tak linear. Dalam sesuatu model regresi dikatakan sebagai regresi tak linear apabila terdapat satu atau lebih daripada satu parameter dalam model tersebut. Persamaan (2.6) dapat menunjukkan gambaran yang lebih jelas berkenaan model regresi tak linear.

$$y = \theta + \theta^2 \xi + \varepsilon \quad (2.6)$$

Selanjutnya, apabila sesuatu pemboleh ubah bersandar (y) dan pemboleh ubah tak bersandar (x) mempunyai hubungan antara satu sama lain melalui satu garis yang lurus ianya dipanggil model regresi linear. Manakala, sekiranya pemboleh ubah bersandar (y) dan pemboleh ubah tak bersandar (x) tidak mempunyai hubungan antara satu sama

lain dalam satu garis yang lurus ianya dipanggil sebagai model regresi tak linear (Neter *et al.*, 1996). Mengikut Draper dan Smith (1992, 1998) telah menyenaraikan beberapa persamaan dan perbezaan yang dapat diperhatikan antara model linear dan model tak linear seperti dalam Jadual 2.1 di bawah:

Jadual 2.1: Notasi Umum Bagi Model Linear dan Model Tak Linear

	Linear	Tak Linear
Pemboleh Ubah Sambutan	Y	Y
Pemboleh Ubah Peramal	X_1, X_2, \dots, X_k	$\xi_1, \xi_2, \dots, \xi_k$
Parameter	$\beta_0, \beta_1, \dots, \beta_p$	$\theta_0, \theta_1, \dots, \theta_p$

(Sumber: Draper & Smith, 1998)

2.2.3 Konsep dan Teori Kaedah Regresi Linear Berganda

Kaedah regresi linear berganda merupakan suatu model statistik yang dibentuk bagi mengkaji hubungan antara pemboleh ubah bersandar dan pemboleh ubah tak bersandar (dua atau lebih). Pada amnya, model regresi linear berganda dapat ditakrifkan seperti dalam persamaan (2.7) dengan $(\beta_j), j = 0, 1, \dots, p$ ialah parameter tak diketahui dan ε iaitu ralat rawak yang mempunyai nilai yang tak tercerap seperti di bawah (Ahmad & Shafiq, 2013; Ahmad *et al.*, 2014, 2016a, 2016b; Shafiq *et al.*, 2017):

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \varepsilon = \sum_{i=1}^p \beta_i x_i + \varepsilon \quad (2.7)$$

Pada kebiasaannya, persamaan (2.7) diubah suai sedikit dengan mengambil $x_0 = 1$ apabila β_0 dapat dijadikan sebagai parameter pintasan. Dalam menganggar parameter $(\beta_j), j = 0, 1, \dots, p$ cerapan sebanyak n diambil daripada setiap pemboleh ubah dan dapat dituliskan seperti yang terdapat dalam persamaan (2.8).