SUATU PERBANDINGAN PENILAIAN PRESTASI PELAJAR MENGIKUT KAUM DI FAKULTI SAINS DAN TEKNOLOGI, UKM: KAEDAH MODEL LINEAR BAYESIAN DAN MODEL LINEAR KLASIK

NORHAIDAH BINTI MOHD ASRAH

PROJEK PENYELIDIKAN YANG DIKEMUKAKAN UNTUK MEMENUHI SEBAHAGIAN DARIPADA SYARAT MEMPEROLEH IJAZAH SARJANA SAINS

PUSAT PENGAJIAN SAINS MATEMATIK
UNIVERSITI KEBANGSAAN MALAYSIA
BANGI

2004

PENGHARGAAN

Alhamdulillah, bersyukur ke hadrat Allah s.w.t kerana dengan limpah kurnia dan izinNya, maka dapat saya menyiapkan dan menghasilkan penulisan projek ini dalam masa yang ditetapkan.

Setinggi-tinggi penghargaan dan ucapan terima kasih ditujukan buat Profesor Madya Dr. Kamarulzaman Ibrahim, selaku penyelia saya yang telah banyak membantu dan membimbing saya dari awal hingga ke akhir penulisan tesis ini. Beliau telah banyak memberi nasihat dan bimbingan kepada saya dalam teori Bayesian.

Ucapan terima kasih juga ditujukan kepada Puan Noorizam Daud yang telah banyak memberi tunjuk ajar dan panduan kepada saya bagaimana hendak mengendalikan program WinBugs 1.4.

Tidak lupa juga kepada rakan-rakan saya terutama sekali kepada Noor Azrin dan Sabariah kerana banyak membantu saya sepanjang menyiapkan penulisan projek ini sehingga berjaya.

Seterusnya, ucapan penghargaan khas kepada keluarga saya dan Ahmad Wira yang banyak memberi dorongan, sokongan dan bantuan semasa saya menyiapkan projek penulisan ini sehingga jayanya.

ABSTRAK

Kajian yang dijalankan ini melibatkan perbandingan prestasi pelajar di Fakulti Sains Teknologi, Universiti Kebangsaan Malaysia mengikut kaum melalui kaedah model linear klasik dan model linear Bayesian. Untuk perbandingan model linear Bayesian, kajian dijalankan dengan mengambil kira penggunaan prior bermaklumat dan prior tak bermaklumat. Kajian tertumpu kepada para pelajar semester satu sessi 2003-2004 di bawah fakulti ini seramai 1088 orang. Perbandingan prestasi pelajar bumiputera dan bukan bumiputera dibuat mengikut program dan pusat pengajian. Hasil analisis yang didapati menunjukkan bahawa tiada perbezaan yang ketara antara model linear klasik dan model linear Bayesian sekiranya prior tak bermaklumat digunakan. Apabila dibandingkan analisis antara prior bermaklumat dengan prior tak bermaklumat menggunakan model linear Bayesian, di dapati bahawa terdapat perbezaan pada nilai parameter. Model linear Bayesian yang menggunakan prior bermaklumat memberikan keputusan yang lebih baik kerana nilai anggaran varians dan selang kebarangkalian Bayesian yang lebih kecil berbanding dengan model linear Bayesian yang menggunakan prior tak bermaklumat. Program Aktuari merupakan program yang mempunyai prestasi yang lebih baik berbanding dengan program lain di fakulti ini. Pusat pengajian Sains Kimia dan Teknologi Makanan pula merupakan pusat pengajian yang terbaik berbanding dengan pusat pengajian lain. Hasil analisis juga menunjukkan bahawa prestasi pelajar bukan bumiputera adalah lebih baik berbanding dengan pelajar bumiputera. Prestasi pelajar bumiputera adalah lebih cemerlang di dalam program Bioteknologi Tumbuhan berbanding dengan prestasi pelajar bukan bumiputera yang lebih cemerlang di dalam program Aktuari. Prestasi pelajar bumiputera dan bukan bumiputera masing-masing lebih baik di dalam pusat pengajian Sains Kimia dan Teknologi Makanan.

ABSTRACT

This study was about a comparison of students achievement in Faculty of Science and Technology, Universiti Kebangsaan Malaysia by races, through classical linear model and Bayesian linear model. We compared the Bayesian linear model by using the informative and non-informative prior. The study was only focus to the first semester students in 2003-2004 session. There were about 1088 of them. Students achievement were compared between programs and schools in this faculty. The analysis show that the results from classical linear model was similar with Bayesian linear model. But, the results from Bayesian linear model with informative prior was better than Bayesian linear model with non-informative prior. It was because the estimated of variance and the range of probability interval for Bayesian linear model with informative prior was smaller. Actuarial program showed the highest achievement compared to the other programs in this faculty. Whilst school of Science Chemistry and Food Technology was the best school compared to the other schools. This study also showed that the nonbumiputera students achieved better results than bumiputera students. The bumiputera students performed very well in Biotechnology of Plantation program. While non-bumiputera students achieved better performance in Actuarial program. Lastly, both bumiputera and non-bumiputera students performed very well in school of Science Chemistry and Food Technology.

KANDUNGAN

PENGAKUAN		**ii**
PENGHARGAAN		**iii**
ABSTRAK		**iv**
ABSTRACT		**v**
KANDUNGAN		**vi**
SEBARAI JADUAL		**xiii**
BAB I	**PENGENALAN**	**1**
1.1	Pendahuluan	1
1.2	Objektif Kajian	4
1.3	Data	5
1.4	Skop Kajian	5
1.5	Struktur Kajian	6
BAB II	**KAJIAN LAMPAU**	**7**
2.1	Pendahuluan	7
2.2	Model Bayesian	7
2.3	Kajian-kajian Lampau	8
BAB III	**PEMODELAN**	**13**
3.1	Pendahuluan	13
3.2	Data	13
3.3	Model Linear Klasik	14
3.4	Model Linear Bayesian	15
BAB IV	**ANALISIS KAJIAN**	**20**
4.1	Pendahuluan	20
4.2	Ujian Kenormalan Data	20
4.3	Perbandingan Antara Model Linear Klasik Dan Model Linear Bayesian Dengan Prior Tak Bermaklumat	22
4.4	Model Linear Bayesian Dengan Prior Bermaklumat Dan Prior Tak Bermaklumat	27

BAB V KESIMPULAN DAN KAJIAN LANJUTAN **34**
5.1 Pendahuluan 34
5.2 Perbandingan Keputusan Pelajar Bumiputera dan Bukan Bumiputera Mengikut Pusat Pengajian dan Fakulti 34
5.3 Kajian Lanjutan 38

RUJUKAN **39**

LAMPIRAN **41**

SENARAI JADUAL

4.1 Jadual pekali parameter bagi Fakulti Sains dan Teknologi mengikut jantina 21
4.2 Jadual pekali parameter bagi Fakulti Sains dan Teknologi mengikut bangsa 22
4.3 Jadual pekali parameter bagi Pusat Pengajian Sains Matematik 22
4.4 Jadual pekali parameter bagi Pusat Pengajian Fizik Gunaan 23
4.5 Jadual pekali parameter bagi Pusat Pengajian Sains Sekitaran dan Sumber Alam 24
4.6 Jadual pekali parameter bagi Pusat Pengajian Sains Kimia dan Teknologi Makanan 25
4.7 Jadual pekali parameter bagi Pusat Pengajian Biosains dan Bioteknologi 25
4.8 Jadual pekali parameter bagi setiap Pusat Pengajian 26
4.9 Jadual pekali parameter bagi Fakulti Sains dan Teknologi 26
4.10 Jadual perbandingan kaedah Bayesian dengan prior bermaklumat dan prior tak bermaklumat bagi Pusat Pengajian Sains Matematik 28
4.11 Jadual perbandingan kaedah Bayesian dengan prior bermaklumat dan prior tak bermaklumat bagi Pusat Pengajian Fizik Gunaan 28
4.12 Jadual perbandingan kaedah Bayesian dengan prior bermaklumat dan prior tak bermaklumat bagi Pusat Pengajian Sains Sekitaran dan Sumber Alam 29
4.13 Jadual perbandingan kaedah Bayesian dengan prior bermaklumat dan prior tak bermaklumat bagi Pusat Pengajian Sains Kimia dan Teknologi Makanan 30
4.14 Jadual perbandingan kaedah Bayesian dengan prior bermaklumat dan prior tak bermaklumat bagi Pusat Pengajian Biosains dan Bioteknologi 31
4.15 Jadual perbandingan kaedah Bayesian dengan prior bermaklumat dan prior tak bermaklumat bagi setiap pusat pengajian 32
4.16 Jadual perbandingan kaedah Bayesian dengan prior bermaklumat dan prior tak bermaklumat bagi Fakulti Sains dan Teknologi 32

5.1 Jadual nilai anggaran parameter bagi setiap program mengikut bangsa 36
5.2 Jadual nilai anggaran parameter bagi setiap pusat pengajian mengikut bangsa 37
5.3 Jadual nilai anggaran parameter bagi Fakulti Sains dan Teknologi mengikut bangsa 37

BAB I

PENGENALAN

1.1 Pendahuluan

Analisis regresi merupakan teknik yang paling banyak diaplikasikan dalam teknik statistik. Montgomery & Peck (1992) memperihalkan bahawa analisis regresi adalah satu teknik statistik untuk mengetahui dan memodelkan hubungan antara pemboleh ubah. Dalam kaedah Bayesian pula, anggaran untuk model linear normal dengan menggunakan pemboleh ubah univariate atau multivariate memang telah lama digunakan. Menurut Mokhtar (1994), analisis regresi merupakan salah satu kaedah statistik yang digunakan untuk mengkaji hubungan antara pemboleh ubah-pemboleh ubah. Bagi Neter *et al.* (1983) pula, analisis regresi ialah satu kaedah yang menggunakan hubungan antara dua atau lebih pemboleh ubah kuantitatif supaya satu pemboleh ubah dapat diramalkan daripada pemboleh ubah yang lain.

Penggunaan kaedah analisis regresi ini telah digunakan dalam pelbagai bidang seperti Ekonomi, Sains Fizik, Sains Kemasyarakatan, Kejuruteraan dan Teknologi dan lain-lain lagi. Analisis regresi juga digunakan untuk meramalkan nilai pemboleh ubah yang diminati (Mokhtar 1994). Model regresi juga banyak digunakan untuk beberapa tujuan tertentu. Antara contoh yang diberikan oleh Montgomery & Peck (1992) ialah seperti perihalan data, anggaran parameter,

peramalan dan anggaran dan juga untuk pengawalan. Menurut Broemeling (1985), salah satu penggunaan kaedah Bayesian ialah untuk membuat analisis peramalan. Analisis peramalan adalah salah satu aktiviti yang penting di dalam bidang perniagaan dan ekonomi di mana teknik siri masa yang canggih sering digunakan.

Mokhtar (1994) menerangkan di dalam analisis regresi, pembolehubah boleh dikategorikan kepada dua jenis, iaitu pemboleh ubah bersandar dan tak bersandar. Nilai pemboleh ubah bersandar ditentukan oleh nilai pemboleh ubah-pemboleh ubah lain yang dikaitkan dengannya. Manakala nilai pemboleh ubah tak bersandar masing-masing menentukan nilai pemboleh ubah-pemboleh ubah bersandar tersebut. Pemboleh ubah bersandar juga dikenali sebagai pemboleh ubah sambutan, manakala pemboleh ubah tak bersandar dikenali sebagai pemboleh ubah penerang peramal atau peregrasi.

Manakala di dalam pentaabiran Bayesian pula, penerangan mengenai parameter kecerunan β yang tidak diketahui akan dicari dengan menggunakan taburan posterior untuk β. Taburan posterior ini di dapati dari teorem Bayes dengan menggabungkan maklumat prior dengan maklumat yang ada di dalam data itu sendiri (Iversen 1984).

Montgomery & Peck (1992) menyatakan bahawa model linear ringkas dengan satu pemboleh ubah x mempunyai hubungan dengan pemboleh ubah sambutan y dan diwakili dengan satu garis lurus. Model linear ringkas ini dapat ditunjukkan seperti di bawah :

$$y = \beta_0 + \beta_1 X + \epsilon \tag{1.1}$$

di mana pekali pintasan β_0 dan kecerunan β_1 yang juga dikenali sebagai pekali regresi adalah tetap dan tidak di ketahui. Kecerunan β_1 menunjukkan perubahan di dalam min taburan y adalah dihasilkan oleh perubahan unit di dalam x. Manakala ϵ ialah komponen ralat yang diandaikan mempunyai min sifar dan varians, σ^2 tidak diketahui. Ralat juga diandaikan tidak berkolerasi antara satu sama lain.

Model regresi yang mempunyai lebih dari satu pemboleh ubah regresi dinamakan model regresi berganda. Model linear berganda dengan k pemboleh ubah

dapat ditunjukkan seperti di bawah :

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \epsilon \tag{1.2}$$

di mana parameter β_j, j=0,1,...,k dipanggil pekali regresi. Parameter β_j mewakili jangkaan perubahan pada sambutan y per unit berubah dalam x_j apabila semua pemboleh ubah regresi x_i $(i \neq j)$ adalah tetap.

Menurut Iversen (1984), analisis Bayesian memerlukan taburan prior untuk parameter yang tidak diketahui. Maklumat prior itu terbahagi kepada kepada dua jenis, iaitu prior bermaklumat dan prior tak bermaklumat. Jika kita tidak tahu nilai yang sebenar dan tidak mempunyai cukup maklumat mengenai parameter di dalam data yang dikutip, maklumat prior ini dikenali sebagai prior tak bermaklumat. Prior tak bermaklumat boleh juga dikatakan sebagai prior maklumat tahap paling rendah mengenai parameter yang sedang dikaji. Taburan ini selalunya akan menghasilkan keputusan yang sama dengan kaedah statistik yang klasik.

Apabila kita membuat kajian yang sama seperti yang telah dilakukan sebelum ini dengan menggunakan pentaabiran Bayesian, kita boleh menggunakan taburan posterior bagi kajian sebelum ini untuk digunakan sebagai prior untuk kajian ini. Ini adalah contoh untuk taburan prior yang bermaklumat. Semakin banyak data yang kita ada, semakin kurang penting taburan prior itu kecuali kita dapat menentukan kebarangkalian prior yang sangat kecil untuk subset-subset bagi nilai-nilai parameter.

Satu kajian yang dijalankan oleh Ding & Karunamuni (2004) telah menggunakan kaedah linear Bayesian empirik. Ding & Karunamuni (2004) telah mencadangkan dua penganggar lain untuk mencari penyelesaian bagi vektor X dan Y. Vektor X (qx1) mewakili nilai sebenar bagi ciri-ciri yang diminati yang ditentukan oleh satu kaedah yang mahal dan sukar. Manakala vektor Y (px1) pula ialah ukuran yang diambil berdasarkan pada ciri-ciri yang sama pada vektor X, di mana kaedah yang digunakan sangat mudah dan murah. X dan Y juga memenuhi persamaan model linear regresi.

Penganggar-penganggar yang dicadangkan itu akan diterbitkan dengan meng-

gunakan teknik linear Bayesian empirik. Ding & Karunamuni (2004) ingin menunjukkan penganggar-penganggar yang dicadangkan adalah lebih baik dari penganggar klasik dari segi min ralat kuasa duanya yang lebih kecil. Penganggar yang dicadangkan ini juga akan dibandingkan dengan penganggar songsangannya. Kaedah simulasi juga digunakan untuk mengesahkan hasil keputusan. Pada akhir kajian, penganggar terbaik adalah merupakan penganggar yang dicadangkan oleh Ding & Karunamuni (2004) di mana penganggar ini menggunakan kaedah linear Bayesian empirik.

Walaupun model linear klasik telah meluas dan sudah lama penggunaannya, namun ini tidak bermakna yang model linear klasik lebih bagus dari model linear Bayesian. Pendekatan yang berbeza antara kedua-dua kaedah ini sudah cukup membuktikan bahawa kaedah kedua-dua model mempunyai kelebihan dan kelemahan masing-masing. Model linear klasik menolak penggunaan maklumat prior manakala model Bayesian pula menggunakan maklumat prior di dalam kaedahnya. Proses membuat keputusan menggunakan kaedah Bayesian mengambil kira akibat kesilapan semasa membuat keputusan, tetapi kaedah klasik tidak begitu (Raeside 1976).

Di dalam kajian ini, kita akan membandingkan model linear Bayesian dengan model linear klasik. Model ini akan diaplikasikan untuk menganggar min PNGK bagi pelajar-pelajar di Fakulti Sains dan Teknologi, Universiti Kebangsaan Malaysia mengikut pusat pengajian dan program dengan mengambil kira kaum. Hasil keputusan akan dibandingkan mengikut kaedah yang digunakan.

1.2 Objektif Kajian

Objektif bagi kajian ini ialah terdiri daripada :

1. Membuat perbandingan prestasi antara pelajar bumiputera dan bukan bumiputera mengikut program dan pusat pengajian menerusi model linear Bayesian dan linear klasik.

2. Membuat perbandingan keseluruhan prestasi antara pelajar bumiputera dan bukan bumiputera di Fakulti Sains dan Teknologi, Universiti Kebangsaan Malaysia menerusi model linear Bayesian dan linear klasik.

1.3 Data

Data yang diperolehi untuk kajian ialah dari jenis data sekunder. Data yang diberi mengandungi maklumat mengenai Purata Nilai Gred Kumulatif (PNGK) pelajar mengikut program yang diikuti.

Model Linear Bayesian diaplikasikan untuk mencari min dan membuat perbandingan bagi pelajar bumiputera dan bukan bumiputera mengikut program dan pusat pengajian. Analisis dijalankan dengan menggunakan perisian statistik WinBUGS 1.4.

Manakala Model Linear Klasik juga akan diaplikasikan untuk mencari min dan membuat perbandingan bagi pelajar bumiputera dan bukan bumiputera mengikut program dan pusat pengajian. Analisis untuk model ini dijalankan dengan menggunakan perisian statistik SPSS 11.0.

1.4 Skop Kajian

Kajian ini hanya tertumpu kepada pelajar-pelajar di bawah Fakulti Sains dan Teknologi, Universiti Kebangsaan Malaysia. Mereka terdiri daripada pelajar-pelajar yang mengikuti sebanyak 22 program di sini. Pemboleh ubah yang diminati untuk kajian ini ialah terdiri daripada Purata Nilai Gred Kumulatif (PNGK). Bangsa adalah terdiri daripada bumiputera dan bukan bumiputera. Mereka yang terlibat adalah merupakan pelajar prasiswazah semester pertama sesi 2003/2004. Bilangan populasi yang terlibat dalam kajian ini ialah sebanyak 1088 orang.

1.5 Struktur Kajian

Dalam bab kedua kajian ini, kajian lampau yang berkaitan dengan kajian ini akan diterangkan dengan lebih terperinci. Ini termasuklah kaedah analisis yang digunakan dan keputusan yang dihasilkan dalam kajian tersebut.

Kandungan di dalam bab tiga akan mengupas tentang data dan kaedah analisis yang digunakan dalam kajian ini. Data yang digunakan dalam kajian ini adalah dari data primer yang bersumberkan dari Penolong Pendaftar Fakulti Sains dan Teknologi. Kaedah analisis yang digunakan adalah linear Bayesian dan linear Klasik. Manakala perisian statistik yang digunakan pula ialah WinBUGS 1.4 dan SPSS 11.0.

Di dalam bab empat pula, keputusan analisis yang didapati akan dibincangkan. Ini meliputi min PNGK bagi setiap pelajar mengikut bangsa mengikut program dan pusat pengajian yang diikuti. Manakala bab terakhir, iaitu bab lima pula akan menerangkan kesimpulan secara keseluruhan mengenai kajian ini. Cadangan yang bersesuaian untuk memperbaiki dan melanjutkan kajian ini akan diutarakan pada bab ini juga.

BAB II

KAJIAN LAMPAU

2.1 Pendahuluan

Sorotan kajian lepas dibuat mengenai penggunaan kaedah Bayesian. Oleh itu, dalam bab ini beberapa kajian terdahulu yang menggunakan kaedah ini akan dinyatakan dengan lebih terperinci khususnya di dalam bidang sains sosial.

2.2 Model Bayesian

Penggunaan kaedah Bayesian semakin meningkat pada masa kini. Ini adalah kerana peningkatan teknologi di dalam bidang pengkomputeran dapat menjadikan kaedah Bayesian ini semakin mudah digunakan. Kaedah ini juga dilihat dapat memberikan penyelesaian yang berguna berbanding dengan kaedah-kaedah statistik yang lain.

Pentaabiran Bayesian menyediakan peluang untuk mengambil maklumat yang sudah tersedia sebelum sebarang data dikumpul. Parameter populasi dianggap sebagai pemboleh ubah rawak. Taburan parameter ini dikenali sebagai taburan prior. Langkah kedua dalam pentaabiran Bayesian ialah mengumpulkan data dan menggabungkan maklumat di dalam data dengan taburan prior. Hasil

yang didapati ialah taburan posterior dan dapat digunakan di dalam pentaabiran. Pengiraan taburan posterior ini akan menggunakan teorem Bayes.

Menurut (Trumbo 2000), prior tak bermaklumat tidak menyediakan sebarang maklumat. Kadangkala prior tak bermaklumat akan memberikan hasil yang sama dengan hasil yang didapati dengan kaedah statistik yang lain. Prior ini digunakan apabila kita tidak tahu nilai sebenar dan tidak mempunyai cukup maklumat mengenai parameter yang kita tidak ketahui.

2.3 Kajian-kajian Lampau

Kaedah Bayesian banyak digunakan dalam pelbagai bidang. Kaedah Bayesian banyak digunakan dalam bidang ekonomi, sains fizik, sains kemasyarakatan, kejuruteraan teknologi dan lain-lain lagi. Kaedah Bayesian terdiri daripada pelbagai jenis seperti Bayes Empirik, Bayesian Berhierarki, Model Linear Bayesian dan bermacam-macam lagi.

Salah satu kaedah Bayesian yang boleh digunakan untuk mengenal pasti lokasi merbahaya ialah kaedah Bayes Empirik. Mengikut kajian yang telah dijalankan oleh Noorizam & Kamarulzaman (2001), terdapat beberapa taburan yang sesuai digunakan untuk data kemalangan seperti Poisson, Binomial Negatif dan Siri Log. Terdapat percanggahan pendapat jika taburan Poisson digunakan. Andaian mengenai taburan harus dipenuhi iaitu min dan variansnya mestilah sama. Manakala sekiranya taburan Binomial Negatif digunakan, ianya lebih sesuai jika nilai varians lebih besar daripada min.

Data yang diperolehi dari kajian Noorizam & Kamarulzaman (2001) adalah mengenai bilangan kemalangan yang berlaku di lokasi-lokasi tertentu, maka tiada nilai sifar yang wujud. Dengan ini, taburan Poisson Terpangkas dan Binomial Negatif Terpangkas telah digunakan. Taburan Siri Log juga telah disuaikan terhadap data.

Parameter-parameter bagi ketiga-tiga taburan tersebut dianggarkan melalui

Kaedah Kebolehjadian Maksimum kerana mempunyai ciri-ciri saksama dan konsisten bagi parameter yang dikaji. Penganggaran kebolehjadian maksimum bagi parameter taburan Binomial Negatif Terpangkas dibuat secara lelaran.

Pada akhir kajian, taburan Poisson Terpangkas tidak begitu sesuai digunakan kerana nilai variansnya melebihi nilai min. Manakala taburan Binomial Negatif Terpangkas merupakan taburan yang sesuai bagi data yang digunakan dalam kajian ini kerana memberikan nilai statistik ujian Khi-kuasa dua yang terkecil. Pemilihan taburan data ini juga adalah konsisten dengan taburan yang digunakan oleh penyelidik-penyelidik terdahulu. Kekerapan anggaran yang dihasilkan melalui penyuaian taburan Binomial Negatif Terpangkas adalah lebih hampir dengan kekerapan kemalangan sebenar.

Kemudian, pemilihan lokasi kemalangan merbahaya dilakukan dengan mencari kebarangkalian bagi taburan ini iaitu:

$$P(R=r) = \frac{w^k}{1-w^k}\frac{(k+r-1)!}{(k-1)!r!}(1-w)^r \qquad \text{di mana}: r = 1, 2, ..., w > 0; k > 0$$

Perbandingan antara lokasi kemalangan dibuat dengan populasi rujukan menerusi kebarangkalian, $P(R > r)$. Lokasi yang merbahaya dapat dikenalpasti melalui nilai kebarangkalian yang kecil. Semakin kecil nilai kebarangkalian, semakin merbahaya lokasi tersebut.

Noorizam & Kamarulzaman (2002) telah melakukan kajian yang sama tetapi menggunakan kaedah yang berbeza. Model Bayesian berhierarki telah digunakan untuk memilih lokasi kemalangan yang merbahaya. Dengan mengandaikan bilangan kemalangan maut (x_i) dan bilangan kemalangan bukan maut (y_i), masing-masing merupakan pemboleh ubah rawak Poisson yang tak bersandar dengan parameter λ_{i1} dan λ_{i2}. Lokasi kemalangan diandaikan tak bersandar dengan min kemalangan λ_{i1} per n_i tahun untuk kemalangan maut dan min kemalangan λ_{i2} per n_i tahun untuk kemalangan bukan maut.

Taburan untuk jumlah kemalangan yang dicerap dalam n_i tahun bagi kedua-dua kategori kemalangan, iaitu (x_i) dan (y_i) bagi setiap lokasi yang bersyaratkan

λ_{i1} dan λ_{i2} diberikan seperti berikut:

$$f(x_i) = \frac{\lambda_{i1}^{xi} \exp(-\lambda_{i1})}{x_i!}$$
$$f(y_i) = \frac{\lambda_{i2}^{yi} \exp(-\lambda_{i2})}{y_i!}$$

di mana i=1,2,...,k dan $\lambda_{i1} > 0$ dan $\lambda_{i2} > 0$

Jumlah kemalangan maut dan kemalangan bukan maut yang berlaku di lokasi ke-i, $i = 1, 2, \ldots, k$ dalam tempoh n_i tahun ditakrifkan sebagai $Z_i = X_i + Y_i$. Maka, boleh ditunjukkan di sini bahawa

$$Z_i|\lambda_{i1}, \lambda_{i2} \sim Pois(\lambda_{i1} + \lambda_{i2})$$

Taburan posterior tercantum bagi $\lambda_{i1}, \lambda_{i2}$ bersyaratkan Z_i menerusi pendekatan Bayesian berhierarki ditunjukkan seperti berikut:

$$\mathrm{P}(\lambda_{i1}, \lambda_{i2}|Z_i) \propto \mathrm{P}(Z_i|\lambda_{i1}, \lambda_{i2})\mathrm{P}(\lambda_{i1})\mathrm{P}(\lambda_{i2})$$

dengan mengandaikan kedua-dua parameter bagi min λ_{i1} dan λ_{i2} tak bersandar dan masing-masing bertaburan gama.

Taburan posterior yang ditentukan diperolehi dengan menggunakan kaedah simulasi melalui pensampelan Gibbs. Maklumat yang didapati mengenai min posterior disusun di mana nilai yang tertinggi menggambarkan tahap merbahaya yang tinggi bagi sesebuah lokasi. Berdasarkan 10 lokasi kemalangan yang dikaji, lokasi kemalangan yang kedua mempunyai tahap merbahaya yang tertinggi diikuti dengan lokasi pertama dan ketiga.

Merujuk kepada kajian yang telah dijalankan oleh Farewell (1999), kaedah Bayesian berhierarki digunakan untuk memeringkatkan keputusan peperiksaan sekolah dengan menggunakan model multivariate berhierarki. Data yang digunakan adalah daripada keputusan peperiksaan di beberapa buah sekolah di London. Terdapat sebanyak 1978 orang pelajar dari 38 buah sekolah yang berlainan. Pemboleh ubah-pemboleh ubah yang diminati ialah seperti skor ujian membaca

London (LRT) dan ujian lisan (VR) di mana dikategorikan mengikut (1,2, atau 3 dimana nilai 1 mewakili kumpulan yang berkebolehan tinggi) yang diambil semasa setiap pelajar berusia 11 tahun. Pengambilan pelajar setiap sekolah diklasifikasikan mengikut jantina (semua perempuan, semua lelaki atau campur) dan mazhab (Church of England, Roman Katolik, sekolah biasa dan lain-lain).

Model yang digunakan dalam kajian ini ialah seperti di bawah:

$$
\begin{aligned}
Y_{ij} &\sim \text{Normal}(\mu_{ij}) \\
\mu_{ij} &= \alpha_{1j} + \alpha_{2j}\text{LRT}_{ij} + \alpha_{3j}\text{VR1}_{ij} + \beta_1\text{LRT}^2_{ij} + \beta_2\text{VR2}_{ij} + \beta_3\text{Perempuan}_{ij} \\
&\quad +\beta_4\text{Sekolah Perempuan}_j + \beta_5\text{Sekolah Lelaki}_j + \beta_6\text{Church of England}_j \\
&\quad +\beta_7\text{Roman Katolik}_j + \beta_8\text{Sekolah lain}_j \\
log\tau_{ij} &= \theta + \phi\text{LRT}_{ij}
\end{aligned}
$$

di mana i mewakili pelajar dan j ialah indeks sekolah. Farewell (1999) telah menentukan model regresi untuk varians komponen dengan memodelkan log pada τ_{ij} (songsangan variasi antara pelajar) sebagai fungsi linear untuk setiap skor LRT pelajar. Ini berbeza dengan model yang digunakan oleh Goldstein et al.'s di mana beliau menggunakan varians σ^2_{ij} yang bersandar secara linear pada LRT. Walau bagaimanapun, parameter ini akan menghasilkan anggaran yang negatif pada nilai σ^2_{ij} .

Kesan tetap β_k $(k = 1, \ldots, 8)$, θ dan ϕ adalah diandaikan mengikut taburan normal tak bersandar yang samar dengan min sifar dan kepersisan yang rendah iaitu 0.0001. Manakala kesan rawak bagi tahap sekolah, α_{jk} $(k = 1, 2, 3)$ adalah diandaikan daripada taburan populasi normal multivariat dengan min γ yang tidak diketahui dan matriks kovarians Σ. Prior tak bermaklumat bagi multivariat normal ditentukan untuk min populasi γ, sementara matrik kovarians songsangan $\text{T} = \Sigma^{-1}$ adalah diandaikan mengikut taburan Wishart. Untuk mewakili prior yang samar, darjah kebebasan untuk taburan ini dipilih sekecil yang mungkin. Skala matriks R ditentukan seperti di sebelah:

$$\begin{pmatrix} 0.1 & 0.005 & 0.005 \\ 0.005 & 0.01 & 0.005 \\ 0.005 & 0.005 & 0.01 \end{pmatrix}$$

Pekali tetap α_{j1} mengukur kesan reja untuk sekolah ke-j selepas menyesuaikan pelajar dengan kovariats tahap sekolah. Ini dapat mewakili kuantiti mana yang sesuai untuk memeringkatkan prestasi sekolah. Farewell (1999) telah menggunakan perisian BUGS untuk memeringkatkan prestasi sekolah.

Pada akhir kajian, keputusan bagi setiap min posterior dan ralat piawai bagi setiap pekali regrasi telah dibandingkan dengan kaedah anggaran yang menggunakan kebolehjadian maksimum. Keputusan yang didapati hampir sama dengan keputusan kaedah kebolehjadian maksimum.

BAB III

PEMODELAN

3.1 Pendahuluan

Dalam bab ini, kaedah model linear Bayesian dan model linear klasik digunakan untuk mencari anggaran min PNGK bagi pelajar-pelajar Fakulti Sains dan Teknologi. Kedua-dua kaedah ini akan dibandingkan keputusannya pada akhir kajian ini nanti.

3.2 Data

Data yang digunakan terdiri daripada PNGK pelajar-pelajar prasiswazah Fakulti Sains dan Teknologi, Universiti Kebangsaan Malaysia, semester pertama sessi 2003/2004. PNGK pelajar-pelajar merangkumi kesemua program bagi setiap pusat pengajian di bawah fakulti yang terdiri daripada Pusat Pengajian Sains Matematik, Pusat Pengajian Fizik Gunaan, Pusat Pengajian Sains Sekitaran dan Sumber Alam, Pusat Pengajian Sains Kimia dan Teknologi Makanan dan Pusat Pengajian Biosains dan Bioteknologi. Populasi yang diambil untuk kajian ini ialah sebanyak 1088 orang pelajar. Bilangan pelajar yang tertinggi ialah di dalam Pusat Pengajian Sains Sekitaran dan Sumber Alam manakala bilangan pelajar yang terendah pula ialah di dalam Pusat Pengajian Sains Matematik. Bilangan

pelajar adalah berbeza-beza mengikut program.

Pembolehubah yang diminati selain dari nilai PNGK ialah jantina dan kaum. Kaum diklasifikasikan sebagai bumiputera dan bukan bumiputera. Kaum bumiputera terdiri daripada bangsa Melayu dan pribumi Sabah dan Sarawak. Manakala kaum bukan bumiputera pula terdiri daripada bangsa Cina, India dan lain-lain.

3.3 Model Linear Klasik

Di dalam kajian ini, model yang digunakan terbahagi kepada dua bahagian, iaitu model bagi fakulti dan model bagi setiap pusat pengajian. Model bagi fakulti mengambil kira faktor kaum dan nilai PNGK mengikut pusat pengajian. Model bagi fakulti dapat diterangkan melalui persamaan di bawah ini :

$$\begin{aligned} Y_{FST} &= \beta_0 + \beta_1(\text{KAUM}) + \beta_2(\text{PPFG}) + \beta_3(\text{PPSS}) + \\ &\quad \beta_4(\text{PPSK}) + \beta_5(\text{PPBB}) + \epsilon \end{aligned} \tag{3.1}$$

di mana

$$\begin{aligned} Y_{FST} &= \text{Keputusan PNGK bagi Fakulti Sains dan Teknologi} \\ \beta_0 &= \text{Pintasan} \\ \epsilon &= \text{Ralat} \\ \text{KAUM} &= \begin{cases} 1 & \text{jika bukan bumiputera} \\ 0 & \text{jika bumiputera} \end{cases} \end{aligned}$$

Model bagi setiap pusat pengajian juga mengambil kira faktor kaum dan nilai PNGK mengikut program dalam pusat pengajian masing-masing. Model bagi pusat pengajian dapat diterangkan seperti berikut:

1. Model bagi Pusat Pengajian Sains Matematik

$$\begin{aligned} Y_{PPSM} &= \beta_0 + \beta_1(\text{KAUM}) + \beta_2(\text{STATISTIK}) + \beta_3(\text{AKTUARI}) \\ &\quad + \epsilon \end{aligned} \tag{3.2}$$

2. Model bagi Pusat Pengajian Fizik Gunaan

$$\begin{aligned} Y_{PPFG} &= \beta_0 + \beta_1(\text{KAUM}) + \beta_2(\text{SAINS BAHAN}) \\ &\quad + \beta_3(\text{SAINS NUKLEAR}) + \epsilon \end{aligned} \tag{3.3}$$

3. Model bagi Pusat Pengajian Sains Sekitaran dan Sumber Alam

$$\begin{aligned} Y_{PPSS} &= \beta_0 + \beta_1(\text{KAUM}) + \beta_2(\text{GEOLOGI}) + \beta_3(\text{SAINS LAUT}) \\ &\quad + \beta_4(\text{SAINS SEKITARAN}) + \epsilon \end{aligned} \tag{3.4}$$

4. Model bagi Pusat Pengajian Sains Kimia dan Teknologi Makanan

$$\begin{aligned} Y_{PPSK} &= \beta_0 + \beta_1(\text{KAUM}) + \beta_2(\text{TEKNOLOGI KIMIA}) + \beta_3(\text{OLEOKIMIA}) \\ &\quad + \beta_4(\text{SAINS MAKANAN \& PEMAKANAN}) \\ &\quad + \beta_5(\text{SAINS MAKANAN DGN PENGURUSAN}) + \epsilon \end{aligned} \tag{3.5}$$

5. Model bagi Pusat Pengajian Biosains dan Bioteknologi

$$\begin{aligned} Y_{PPBB} &= \beta_0 + \beta_1(\text{KAUM}) + \beta_2(\text{GENETIK MOLEKUL}) \\ &\quad + \beta_3(\text{MIKROBIOLOGI}) + \beta_4(\text{BIOTEK TUMBUHAN}) \\ &\quad + \beta_5(\text{BIOTEK DGN PENGURUSAN}) \\ &\quad + \beta_6(\text{ZOOLOGI GUNAAN}) + \beta_7(\text{BIOINFORMATIK}) \\ &\quad + \epsilon \end{aligned} \tag{3.6}$$

Model untuk keseluruhan prestasi pelajar mengikut kaum dengan mengambil kira kesemua program dan pusat pengajian adalah seperti berikut:

$$Y_{FST} = \beta_0 + \beta_1(\text{KAUM}) + \epsilon \tag{3.7}$$

3.4 Model Linear Bayesian

Menurut Gelman *et al.* (1995), di dalam model linear Bayesian, taburan y adalah normal apabila diberi X, di mana minnya adalah fungsi linear bagi X:

$$E(y_i|\beta, X) = \beta_1 x_{i1} + \ldots + \beta_k x_{ik} \tag{3.8}$$

di mana $i = 1, \ldots, n$. Nilai pembolehubah x_{i1} adalah tetap, iaitu sama dengan satu. Oleh itu, $\beta_1 x_{i1}$ adalah tetap untuk semua i. Di dalam model linear Bayesian, perkara yang paling penting dalam pembentukan model ialah:

1. Mentakrifkan pembolehubah x dan y (kemungkinan menggunakan penjelmaan) supaya jangkaan bersyarat bagi y adalah linear untuk fungsi x dan ralat menghampiri normal

2. Membina taburan prior pada parameter model, taburan parameter yang baik untuk parameter model supaya dapat membuat penganggaran dengan tepat daripada data.

Menurut analisis Bayesian, taburan prior amat diperlukan untuk parameter yang tidak diketahui. Kajian adalah sesuatu yang kumulatif, dan salah satu kekuatan kaedah Bayesian berbanding dengan kaedah klasik ialah kaedah Bayesian membolehkan kita menggunakan pengetahuan dari kajian yang terdahulu. Satu kes khas di mana taburan prior bermaklumat wujud ialah apabila kita mereplikakan atau mengkaji semula kajian lama yang menggunakan analisis Bayesian. Dalam kes ini, kita boleh mengambil taburan prior dari kajian yang lepas untuk digunakan di dalam kajian yang baru. Iversen (1984) menyatakan bahawa dengan taburan prior bermaklumat kita akan mendapat taburan posterior yang lebih memuncak dan dengan varians yang kecil. Ini bermakna kita dapat mentafsir parameter yang tidak diketahui dengan lebih tepat dan selang kebarangkaliannya akan lebih kecil berbanding dengan taburan prior tak bermaklumat.

Pollard (1986) menyatakan bahawa dalam beberapa contoh di mana maklumat prior adalah samar, kita boleh menggunakan prior tak bermaklumat untuk menjalankan kajian. Prior tak bermaklumat tidak banyak menyumbang pada taburan posterior. Pelbagai prior yang berbeza akan menghasilkan taburan posterior yang sama. Taburan posterior yang didapati dengan menggunakan prior tak bermaklumat hampir sama dengan posterior yang didapati dengan menggunakan mana-mana prior yang didominasi oleh kebolehjadian. Ini kerana posterior secara

keseluruhannya akan ditentukan oleh data itu sendiri. Tambahan pula, apabila prior tak bermaklumat digunakan, keputusan antara kaedah Bayesian dan klasik akan hampir sama.

Dalam kajian ini, persamaan model linear Bayesian yang digunakan untuk mencari nilai anggaran min PNGK untuk pelajar-pelajar Fakulti Sains dan Teknologi mengikut kaum ialah

$$y_i \sim N(\mu_i, \sigma^2) \tag{3.9}$$

$$\mu_i = \beta_0 + \Sigma\beta_i X_i + \epsilon_i \tag{3.10}$$

di mana

i = $1, \ldots, n$ n mengambil nilai bilangan pusat pengajian atau program

y_i = keputusan PNGK bagi pelajar di FST

μ_i = anggaran min PNGK bagi pelajar di FST

β_i = pekali regresi ke-i

σ^2 = varians

ϵ_i = ralat

Dalam kajian ini, kita menggunakan prior tidak bermaklumat dan bermaklumat untuk parameter σ^2. Nilai prior tak bermaklumat adalah sama ada menggunakan nilai varians yang terlalu besar atau yang terlalu kecil. Manakala untuk prior bermaklumat, nilai yang akan digunakan ialah nilai mengikut kajian yang telah dilakukan menggunakan model linear klasik. Nilai min dan varians yang didapati dari model linear klasik akan digunakan sebagai prior bermaklumat. Seterusnya, kita mengandaikan bahawa parameter σ^2 adalah bertaburan gama songsangan kerana taburan gama songsangan ini adalah merupakan prior konjugat bagi taburan normal. Taburan untuk σ^2 adalah seperti berikut:

$$\sigma^2 \sim IG(\alpha, \beta) \tag{3.11}$$

Manakala kesan tetap β_i (i=1,...,n) adalah diandaikan mengikut taburan normal tak bersandar yang samar dengan min sifar dan varians yang rendah iaitu 0.0001.

RUJUKAN

Broemeling, L. D. 1985. *Bayesian Analysis of Linear Models.* , vol. 60. Marcel Dekker,Inc.

Ding, K. & Karunamuni, R. J. 2004. A linear empirical bayes solution for the calibration problem. *Journal of Statistical Planning and Inference* hlm. 421–447.

Farewell, D. 1999. Schools: ranking school examination results using multivariate hierarchical models. (atas talian) http://www.mrc-bsu.cam.ac.uk/bugs/-documentation/exampVol2/node11.html (28 Mac).

Gelman, A., Carlin, J. B., Stern, H. S. & Rubin, D. B. 1995. *Bayesian Data Analysis*. Chapman and Hall.

Iversen, G. R. 1984. *Bayesian Statistical Inference*. 1st edn. Sage Publications.

Mokhtar, A. 1994. *Analisis Regresi*. Dewan Bahasa dan Pustaka.

Montgomery, D. C. & Peck, E. A. 1992. *Introduction to Linear Regression Analysis*. 2nd edn. John Wiley and Sons, Inc.

Neter, J., Wasserman, W. & Kutner, M. H. 1983. *Applied Linear Regression Models*. Richard D. Irwin, Inc.

Noorizam & Kamarulzaman 2001. Mengenal pasti lokasi merbahaya melalui kaedah bayes empirik. *Prosiding Simposium Kebangsaan Sains Matematik Ke-9* hlm. 217–221.

Noorizam & Kamarulzaman 2002. Model bayesian berhierarki untuk memilih lokasi kemalangan merbahaya. *Prosiding Simposium Kebangsaan Sains Matematik Ke-10* hlm. 521–526.

Pollard, W. E. 1986. *Bayesian Statistics for Evaluation Research: An Introduction.* , vol. 8. Sage Publications.

Raeside, D. E. 1976. Bayesian statistics: A guided tour*. *Medical Physics* **3** (1): 1–11.

Trumbo, B. E. 2000. A brief introduction to bayesian estimation. (atas talian) http://www.sci.csuhayward.edu/statistics/Gibbs/GibbsSess4.htm (1 Februari 2004).