

BAB 3

ANALISIS KAJIAN

3.1 Pengenalan

Sejumlah 99,877 data peminjam telah diperoleh daripada Perbadanan Tabung Pendidikan Tinggi Nasional (PTPTN) untuk dianalisis. Analisis permulaan menunjukkan data berkenaan mengandungi maklumat mengenai enam perkara atau variabel mengenai peminjam PTPTN, tidak termasuk maklumat kad pengenalan, nama dan status peminjam. Enam variabel tersebut ialah: jantina, bangsa, pendapatan keluarga, pengkhususan atau jurusan kajian, peringkat pengajian dan institusi pengajian. Berdasarkan nombor kad pengenalan (baru) peminjam, pengkaji telah mewujudkan satu lagi variabel iaitu 'umur'. Bagi variabel yang mempunyai banyak kategori maklumat, iaitu bidang kajian, peringkat pengajian dan institusi pengajian, bilangan kategori telah dikecilkan kepada beberapa jenis bagi memudahkan analisis. Berikut adalah penjelasan tentang variabel kajian dan analisis statistik yang digunakan dalam kajian ini.

3.2 Variabel Kajian

Kesemua variabel mewakili data kajian ini adalah dalam bentuk nominal/kategorikal. Data bagi beberapa variabel dikodkan semula setelah disatukan kepada bilangan kategori yang lebih kecil. Berikut adalah definisi variabel yang digunakan dalam kajian ini:

- a) *Jantina*: Lelaki dan Perempuan
- b) *Keturunan*: Melayu, Cina, India, Bumiputera Sabah, Bumiputera Sarawak, Orang Asli dan lain-lain. Data keturunan adalah seperti yang ditandakan oleh peminjam, yang mungkin tidak menepati keturunan sebenar. Misalnya, terdapat pelajar keturunan Punjabi dan India Muslim yang termasuk dalam kelompok keturunan India, keturunan Bumiputera Sabah dan keturunan Cina dan India beragama Islam yang termasuk dalam kelompok keturunan Melayu dan sebagainya.

- c) *Kaum*: Bumiputera (Melayu, Bumiputera Sabah dan Bumiputera Sarawak), Cina, India dan lain-lain.
- d) *Pengkhususan/Jurusan pengajian*: Sastera Tulen, Sastera Gunaan, Sains Tulen dan Sains Gunaan. Empat kategori tersebut disatukan daripada berbagai program pengajian yang diikuti oleh pelajar.
- e) *Peringkat pengajian*: Matrikulasi, Diploma, Intergrasi, Ijazah Pertama Sepenuh Masa dan Diploma Lepasan Ijazah. Enam kategori ini diringkaskan daripada 13 kategori asal yang mengandungi pelbagai peringkat pengajian.
- f) *Institusi pengajian*: Universiti Awam, Universiti Kolej Awam, Universiti Swasta, Universiti Kolej Swasta dan Kolej Swasta. Lima kategori tersebut diringkaskan daripada pelbagai institusi pengajian tinggi awam dan institusi pengajian tinggi swasta yang dihadiri oleh peminjam.
- g) *Pendapatan keluarga*: Pendapatan bulanan yang diperolehi oleh ibu dan bapa, dikategorikan kepada seperti berikut:
 - Kurang dari RM1,000.00
 - RM1,001.00 hingga RM3,000.00
 - RM3,001.00 hingga RM5,000.00
 - RM5,001.00 hingga RM10,000.00
 - RM10,001.00 hingga RM50,000.00
 - Melebihi RM50,000.00
- h) *Umur*: Umur peminjam pada tahun 2006. Data ini adalah berdasarkan nombor kad pengenalan peminjam. Tetapi umur peminjam bukanlah umur semasa pinjaman dipohon atau diluluskan.

3.3 Analisis Deskriptif

Analisis deskriptif (*crosstab*) dilakukan bagi mendapatkan dapatan bagi perkara kajian 2.1, 2.2 dan 2.3 di atas. Analisis telah dijalankan ke atas semua 99,877 data yang diperolehi. Pengiraan bilangan dan peratus dilakukan bagi menunjukkan taburan peminjam mengikut variabel yang dikaji, dan dijalankan ke atas data sah (*valid cases*) sahaja. Carta pai dan carta bar dihasilkan berdasarkan peratus peminjam, bagi menjelaskan secara grafik tentang taburan peminjam seperti yang dilaporkan dalam

Bab 4. Jadual kekerapan dan peratus peminjam disediakan sebagai rujukan pada bahagian Lampiran pada akhir laporan.

Bagi melihat profil peminjam (perkara kajian 2.1), peratusan peminjam mengikut kategori-kategori dalam setiap variabel kajian dibandingkan untuk mengenal pasti kategori dengan peratus terbesar peminjam. Bagi melihat profil Peminjam-Bayar dan Peminjam-Tidak Bayar (perkara kajian 2.2 dan 2.3), peratus Peminjam-Bayar dan Peminjam-Tidak Bayar dibandingkan untuk mengenal pasti yang mana lebih besar (melebihi 50 peratus), bagi setiap kategori variabel kajian.

3.4 Analisis Regresi Logistik

Bagi perkara kajian 2.4, analisis regresi logistik (*logistic regression*), bukannya analisis diskriminan (*discriminant analysis*), telah dilakukan untuk menentukan sama ada kebarangkalian calon yang membuat pinjaman PTPTN membayar balik pinjamannya dapat diramal berdasarkan variabel-variabel kajian yang ada. Analisis regresi logistik mempunyai kuasa ramalan yang agak lemah berbanding dengan analisis diskriminan. Tetapi analisis regresi logistik dipilih kerana analisis tersebut tidak memerlukan andaian *multivariate normality* (iaitu andaian terdapatnya perhubungan secara linear dan nilai varian yang sama bagi variabel yang dikaji), tetapi dapat menerima gabungan variabel selang dan variabel kategorikal sebagai variabel bebas. Oleh kerana data sedia ada didapati tidak dapat memenuhi syarat andaian *multivariate normality* yang diperlukan bagi melaksanakan analisis diskriminan bagi mengasingkan (*discriminate*) peminjam yang gagal membayar dan peminjam yang membayar pinjaman, analisis regresi logistik dilakukan.

Dalam analisis regresi logistik, transformasi logit dilakukan ke atas $p(y = 1|x)$. Transformasi logit adalah *log of the odds* bagi $y = 1$ dengan $y = 0$. Transformasi logit bersifat linear dengan parameter model, ia adalah sama dengan syarat model regresi. Oleh itu, persamaannya dibentuk seperti berikut:

$$\ln [\text{odds (membayar balik} \mid X_1, X_2, X_3, \dots, X_k) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_k X_k$$

atau

$$\ln \frac{p}{1-p} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_k X_k$$

$$\text{dimana odds (membayar balik} \mid X_1, X_2, X_3, \dots, X_k) = \frac{p}{1-p}$$

p adalah kebarangkalian peminjam membayar balik pinjaman. 1-p adalah kebarangkalian peminjam tidak membayar balik pinjaman. Kedua-dua adalah tertakluk kepada nilai variabel bebas X_1, X_2, \dots, X_k .

Bagi model ini, log bagi *odds* adalah sebagai fungsi linear variabel bebas, sama seperti dalam persamaan regresi linear. Variabel bebasnya adalah dalam bentuk variabel kategorikal dan selanjar (*continuous*). Untuk kes ini, *log of the odds* juga disebut sebagai *logits*, atau *multiple logistic regressions*. Oleh itu, interpretasi angkali (*coefficient*) bagi variabel bebas adalah berhubung dengan kesannya terhadap *log of the odds*, bukan terhadap kebarangkalian p. Bagi regresi logistik, pendekatan anggaran *maximum likelihood* digunakan untuk membuat anggaran parameter, berbanding dengan regresi berganda yang menggunakan anggaran *least-square*.

$$\ln \frac{p}{1-p} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_k X_k$$

atau

$$p = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k)}}$$

Analisis regresi logistik dilakukan dengan menggunakan program *Statistical Package for the Social Science* (SPSS) dengan memasukkan variabel bebas seperti pada Jadual 3.1.

3.5 Persediaan Data Bagi Analisis Loglinear

Data dalam kajian ini bukan bersifat selanjar (*continuous*), tetapi terdiri daripada gabungan variabel selanjar, kategorikal dan juga diskrit (*discrete*). Variabel bersandar bagi kajian ini bersifat *dummy*. Memandangkan variabel bebas bersifat gabungan selanjar, diskrit dan kategorikal, dan tidak bersifat *multivariate normal*, pendekatan *regresi logistik* digunakan. Bagi tujuan analisis ini, ramalan dengan menggunakan *odd-ratio* digunakan. Variabel bersandar dikodkan dengan kod *dummy* (1 = membayar pinjaman dan 0 = tidak membayar pinjaman).

3.6 Penentuan Kod *Dummy* dan Pembentukan Variabel Bebas

Kajian ini menggunakan variabel bebas umur peminjam yang bersifat selanjar (*continuous*) dan dilakukan kod *dummy* 1 dan 0. Bagi variabel umur, jantina dan jenis institusi pengajian tinggi, nilai 1 merujuk kepada peratus tertinggi kumpulan membayar balik pinjaman manakala nilai 0 merujuk kepada kumpulan peratus yang rendah membayar balik pinjaman. Variabel keturunan (berbentuk nominal dan terdiri daripada tiga kategori: Bumiputera, Cina dan India) dikelompokkan kepada dua variabel (Keturunan1 dan Keturunan2). Bagi variabel Keturunan1, nilai 1 merujuk kepada keturunan Cina manakala dua kumpulan lain diberi nilai 0. Bagi variabel Keturunan2, nilai 1 merujuk kepada keturunan India dan dua keturunan lain diberi nilai 0. Bagi variabel Keturunan1, keturunan Cina diberi nilai 1 merupakan peratus tertinggi bagi variabel keturunan yang membayar balik pinjaman, manakala bagi variabel Keturunan2, kaum India merupakan peratus kedua tertinggi yang membayar balik pinjaman diberi nilai 1. Variabel jurusan yang juga berbentuk nominal terdiri dari empat jurusan (Sastera Tulen, Sastera Gunaan, Sains Tulen dan Sains Gunaan) dikelompokkan kepada tiga variabel (Jurusan1, Jurusan2 dan Jurusan3). Pembentukan kod *dummy* jurusan ini juga berpandu kepada peratus membayar balik pinjaman mengikut Jurusan. Jadual 3.1 menunjukkan cara kod *dummy* dibentuk.

Jadual 3.1: Pembentukan Kod *Dummy* Variabel Bersandar Kategorikal

Variabel bersandar dan kategori	Peratus membayar pinjaman	Peratus tidak bayar pinjaman	Kod <i>dummy</i>	
			Nilai 1	Nilai 0
Jantina: a) Lelaki b) Perempuan	Perempuan = 62.8 Lelaki = 37.2	Perempuan = 41.2 Lelaki = 58.8	Perempuan	Lelaki
Umur: a) 30 tahun ke bawah b) Melebihi 30 tahun	Umur 30 tahun dan ke bawah = 43.29 Umur melebihi 30 tahun = 75.69	Umur 30 tahun dan ke bawah = 56.71 Umur melebihi 30 tahun = 24.31	Umur melebihi 30 tahun	Umur 30 tahun dan ke bawah
Keturunan (2 variabel): Keturunan1/Keturunan2: a. Melayu b. Cina c. India	Melayu= 41.6 Cina = 68.0 India = 54.7	Melayu = 58.4 Cina = 32.0 India = 45.3	Keturunan1:	
			Cina = 1	India = 0 Melayu = 0
			Keturunan2:	
			India = 1	Cina = 0 Melayu = 0
Institut Pengajian Tinggi: a. IPT Awam (Universiti Awam dan Universiti Kolej Awam) b. IPT Swasta (Universiti Swasta, Universiti Kolej Swasta dan Kolej Swasta)	IPT Awam= 49.5 IPT Swasta= 37.8	IPT Awam= 50.5 IPT Swasta= 62.2	IPT Awam	IPT Swasta
Pendapatan Keluarga: a. Pendapatan RM5,000 dan kebawah b. Pendapatan melebihi RM5,000	Pendapatan RM5,000 kebawah = 32.7 Pendapatan Melebihi RM5000= 53.7	Pendapatan RM5,000 kebawah = 62.3 Pendapatan Melebihi RM5000= 46.3	Pendapatan Melebihi RM5000	Pendapatan RM5,000 kebawah
Penghusususan (3 variabel): Khusus1/Khusus2/Khusus3: a. Sastera Gunaan b. Sains Gunaan c. Sains Tulen Oleh kerana saiz sampel kecil, kategori Sastera Tulen tidak dapat digunakan dalam analisis ini)	Sastera Gunaan= 45.3 Sains Gunaan= 43.3 Sains Tulen= 68.0 Sastera Tulen= 65.0	Sastera Gunaan= 54.7 Sains Gunaan= 56.7 Sains Tulen= 32.0 Sastera Tulen= 35.0	Khusus1:	
			Sastera Gunaan =1	Sains Gunaan=0 Sains Tulen=0 Sastera Tulen=0
			Khusus2:	
			Sains Gunaan =1	Sastera Gunaan=0 Sastera Tulen=0 Sains Tulen=0
		Khusus3:		
		Sains Tulen=1	Sastera Gunaan=0 Sains Gunaan=0 Sastera Tulen=0	