

# ANALISIS KEMALANGAN JALANRAYA MENGUNAKAN MODEL REGRASI BERGANDA

Oleh:

Syed Othmawi b. Abd. Rahman dan  
P. Madya Jamilin Jais,  
Panel Sistem Maklumat Pengurusan,  
Fakulti Sains Komputer dan Sistem Maklumat,  
Universiti Teknologi Malaysia,  
Kuala Lumpur

## ABSTRAK

Model regresi linear berganda adalah sebuah model linear yang menghubungkan di antara satu pembolehubah dengan beberapa pembolehubah yang lain. Kertas kerja ini merupakan kertas kerja asas bagi membincangkan model tersebut. Perkara-perkara yang dibincang dalam kertas kerja ini di antaranya adalah model regresi berganda secara umum, andaian-andaian yang diperlukan sewaktu menggunakan model ini, pengiraan bagi mendapatkan nilai parameter model, penilaian kebagusan model dan sebagainya. Kadar kemalangan jalanraya berbanding dengan bilangan kenderaan di atas jalan raya, jumlah penduduk dan jumlah panjang jalan (km) di Malaysia Barat telah digunakan sebagai kajian kes. Di akhir sekali, pakej SAS/STAT digunakan bagi melakukan analisis yang sama untuk tujuan perbandingan.

**Katakunci:** Regresi linear berganda, model, SAS/STAT, persamaan linear dan kadar kemalangan jalan raya di Malaysia Barat.

## ABSTRACT

Multiple linear regression is a model that incorporates the variable with several variables in a linear relationship. This paper try to introduce the basic discussions of the state model. The discussion included a general linear regression model, assumptions, calculated parameters and others. The SAS/STAT package was used in the analysis for the problem solving.

**Keywords:** Multiple linear regression, model, SAS/STAT, linear equations and rate of road accident in West Malaysia.

## 1.0 PENGENALAN

Regrasi analisis berganda adalah suatu model linear yang digunakan untuk menghubungkan satu pembolehubah dengan pembolehubah-pembolehubah yang lain. Model regresi berganda boleh ditulis di dalam format berikut:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \dots + \beta_m x_{im} + \varepsilon_i$$

ataupun boleh diringkaskan di dalam bentuk berikut:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_m X_m + \varepsilon$$

di mana  $y_i$  dan  $x_{ij}$  adalah nilai-nilai bagi pembolehubah  $Y$  dan  $X_i$  manakala  $\beta_i$  adalah parameter-parameter *konstan regresi*. Nilai  $\varepsilon$  pula merupakan nilai perbezaan di antara nilai  $y_i$  sebenar dengan nilai yang dikira menggunakan persamaan regresi yang merupakan pembolehubah rawak dengan purata 0 dan varian  $\sigma^2$ .

Pembolehubah  $X_i$  dipanggil pembolehubah bebas kerana ia boleh mengambil sebarang nilai manakala pembolehubah  $Y$  dipanggil pembolehubah bersandar kerana nilainya bergantung terus kepada nilai  $X_i$ . Model regresi yang terdiri dari satu pembolehubah bebas dipanggil *model regresi mudah* (simple regression model) manakala jika sebaliknya dipanggil *model regresi berganda* (multiple regression model). Nilai ramalan akan menjadi lebih tepat sekiranya terdapat lebih banyak pembolehubah bebas di dalam model. Contoh penggunaan model regresi berganda adalah: katakan sebuah pasaraya yang terdiri dari beberapa buah cawangan ingin mengkaji hubungan di antara jumlah jualan dibandingkan dengan kos iklan yang telah dibelanjakan, pendapatan perkapita, ketumpatan penduduk dan bilangan pasaraya lain di dalam setiap kawasan.

## 2.0 ANDAIAN-ANDAIAN YANG DIGUNAKAN DI DALAM MODEL ANALISIS REGRASI BERGANDA

Andaian-andaian yang digunakan di dalam model regresi berganda adalah seperti berikut<sup>1</sup>:

1. Nilai-nilai bagi pembolehubah bebas  $X_i$  boleh terdiri dari nilai rawak atau tidak iaitu nilai yang telah ditetapkan.
2. Pembolehubah  $X_i$  diukur tanpa terlibat dengan sebarang kesalahan.
3. Untuk setiap gabungan nilai  $X_i$ , terdapat subpopulasi nilai-nilai  $Y$  yang bertaburan normal.
4. Varian subpopulasi-subpopulasi nilai-nilai  $Y$  adalah sama.
5. Nilai  $Y$  yang didapati dari satu nilai  $X_i$  bebas dari nilai  $Y$  yang didapati dari nilai  $X_i$  yang lain.
6.  $e_i$  bertaburan normal dan bebas dengan purata 0 dan varian  $\sigma^2$ .

Jika model regresi yang sedang dikaji adalah model regresi mudah, kita boleh mewakili hubungan di antara  $X$  dan  $Y$  di atas graf dua dimensi tetapi jika model yang sedang dikaji itu merupakan model regresi berganda, kita memerlukan sebuah *hyperplane* untuk melakarkan kedudukan nilai pembolehubah-pembolehubah bebas tersebut.

## 3.0 MENDAPATKAN PERSAMAAN REGRASI

Salah satu kaedah yang digunakan untuk mendapatkan persamaan regresi adalah dengan menggunakan *kaedah kuasadua terkecil* (least squares method). Di dalam kaedah ini jumlah sisihan kuasadua kedudukan data sebenar dari *hyperplane* regresi diminimumkan. Model ini boleh didapati dengan mudah jika kita menggunakan pakej-pakej komputer seperti SAS/STAT, MINITAB, LOTUS, SPSS dan sebagainya. Dalam kertas kerja ini pengiraan bagaimana model regresi didapati dibincangkan disamping pakej SAS/STAT digunakan juga sebagai perbandingan.

Model regresi yang didapati dari pengiraan ataupun dari penggunaan pakej komputer adalah di dalam format berikut:

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + \dots + b_nx_n$$

di mana  $b_0$  merupakan pemetongan satah pada paksi  $y$  dan  $b_1, b_2, b_3, \dots, b_n$  merupakan kecerunan satah yang berkaitan dengan pembolehubah bebas  $x_1, x_2, x_3, \dots, x_n$ . Contohnya  $b_1$  adalah

jumlah pertambahan nilai  $y$  bagi setiap unit pertambahan nilai  $x_1$  dengan nilai pembolehubah-pembolehubah bebas lain tidak berubah. Dengan sebab itulah nilai  $b_i$  dipanggil *pekali-pekali regresi bahagian* (partial regression coefficients).

Anggaran parameter-parameter  $b_i$  bagi model regresi boleh didapati dengan menyelesaikan sistem persamaan berikut<sup>7</sup>:

$$b_{0n} + b_{1-x_{1j}} + b_{2-x_{2j}} + b_{3-x_{3j}} + \dots + b_{m-x_{mj}} = -y_j \quad (1)$$

$$b_{0-x_{1j}} + b_{1-x_{1j}^2} + b_{2-x_{1j}x_{2j}} + b_{3-x_{1j}x_{3j}} + \dots + b_{m-x_{1j}x_{mj}} = -x_{1j}y_j \quad (2)$$

$$b_{0-x_{2j}} + b_{1-x_{2j}x_{1j}} + b_{2-x_{2j}^2} + b_{3-x_{2j}x_{3j}} + \dots + b_{m-x_{2j}x_{mj}} = -x_{2j}y_j \quad (3)$$

.....

$$b_{0-x_{mj}} + b_{1-x_{mj}x_{1j}} + b_{2-x_{mj}x_{2j}} + b_{3-x_{mj}x_{3j}} + \dots + b_{m-x_{mj}^2} = -x_{mj}y_j \quad (m)$$

Persamaan (1) boleh ditukar ke dalam sebutan  $b_0$  menjadi seperti berikut:

$$b_0 = \bar{y} - b_1\bar{x}_1 + b_2\bar{x}_2 + b_3\bar{x}_3 + \dots + b_m\bar{x}_m \quad (m+1)$$

Dengan menggantikan nilai  $b_0$  ke dalam persamaan (2), (3), (4), ..., (m) dan setelah penyusunan semula dilakukan, kita dapati suatu sistem persamaan linear yang baru seperti berikut:

$$b_1(\bar{x}_1^2 - n\bar{x}_1^2) + b_2(\bar{x}_1x_2 - n\bar{x}_1\bar{x}_2) + b_3(\bar{x}_1x_3 - n\bar{x}_1\bar{x}_3) + \dots + b_m(\bar{x}_1x_m - n\bar{x}_1\bar{x}_m) = -\bar{x}_1\bar{y} - n\bar{x}_1\bar{y} \quad (m+2)$$

$$b_1(\bar{x}_2x_1 - n\bar{x}_2\bar{x}_1) + b_2(\bar{x}_2^2 - n\bar{x}_2^2) + b_3(\bar{x}_2x_3 - n\bar{x}_2\bar{x}_3) + \dots + b_m(\bar{x}_2x_m - n\bar{x}_2\bar{x}_m) = -\bar{x}_2\bar{y} - n\bar{x}_2\bar{y} \quad (m+3)$$

$$b_1(\bar{x}_3x_1 - n\bar{x}_3\bar{x}_1) + b_2(\bar{x}_3x_2 - n\bar{x}_3\bar{x}_2) + b_3(\bar{x}_3^2 - n\bar{x}_3^2) + \dots + b_m(\bar{x}_3x_m - n\bar{x}_3\bar{x}_m) = -\bar{x}_3\bar{y} - n\bar{x}_3\bar{y} \quad (m+4)$$

.....

$$b_1(\bar{x}_m x_1 - n\bar{x}_m \bar{x}_1) + b_2(\bar{x}_m x_2 - n\bar{x}_m \bar{x}_2) + b_3(\bar{x}_m x_m - n\bar{x}_m \bar{x}_m) + \dots + b_m(\bar{x}_m^2 - n\bar{x}_m^2) = -\bar{x}_m \bar{y} - n\bar{x}_m \bar{y} \quad (2m-1)$$

Dengan menyelesaikan sistem persamaan di atas, semua nilai  $b_i$  didapati.

#### 4.0 PENGUJIAN KEBAGUSAN MODEL

Pengujian kebagusan model biasanya dilakukan melalui dua peringkat utama iaitu peringkat menguji kebagusan model secara keseluruhan dan peringkat menguji kebagusan bahagian-bahagian tertentu model tersebut iaitu pengujian kebagusan parameter-parameter secara berasingan.

##### 4.1 Pengujian Kebagusan Keseluruhan Model

Pengujian ini dilakukan untuk mengesahkan wujudnya regresi linear di antara pembolehubah bersandar ke atas pembolehubah-pemboleubah bebas. Hipotesis yang akan diuji adalah:

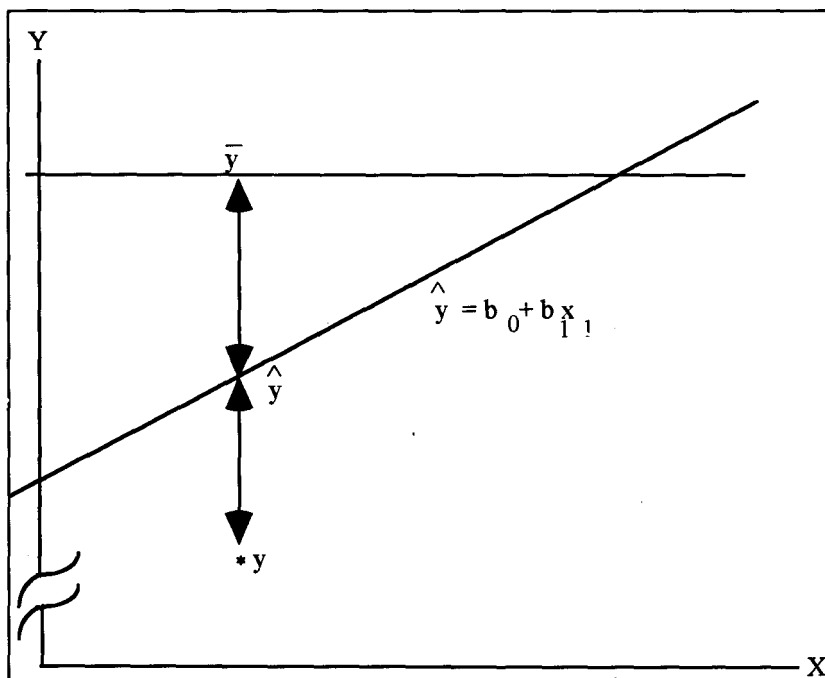
$H_0 : \beta_1 = 0$  (Tiada hubungan linear di antara pembolehubah Y dengan pembolehubah-pemboleubah bebas. Nilai  $\beta_1$  bersamaan dengan 0,  $\forall i$ )

$H_1 : \beta_1 \neq 0$  (Wujud hubungan linear di antara pembolehubah Y dengan pembolehubah-pemboleubah bebas. Terdapat satu atau lebih elemen  $\beta_1$  yang tidak bersamaan dengan 0).

Untuk memudah kefahaman mengenai analisis seterusnya perhatikan sebuah model analisis regresi mudah seperti berikut:

$$\hat{y} = b_0 + b_1 x_1$$

Graf hubungan di antara  $\hat{y}$ ,  $y$  dan  $\bar{y}$  adalah seperti berikut:



Graf 1: Lakaran Hubungan Nilai  $\hat{y}$ ,  $y$  dan  $\bar{y}$ .

Dari Rajah 1, persamaan berikut boleh dibentuk:

$$(y - \bar{y}) = (\hat{y} - \bar{y}) + (\hat{y} - y) \quad (1)$$

Dari persamaan (1)

$$\begin{aligned} \sum (y - \hat{y}) &= \sum [(y - \hat{y}) + (\hat{y} - y)] \\ \sum (y - \hat{y})^2 &= \sum [(y - \hat{y}) + (\hat{y} - y)]^2 \\ &= \sum (y - \hat{y})^2 + 2 \sum (y - \hat{y})(\hat{y} - y) + \sum (\hat{y} - y)^2 \end{aligned}$$

Gantikan  $\hat{y} = \bar{y} + \beta_1 (x - \bar{x})$  ke dalam persamaan.

$$\begin{aligned} &= \sum (y - \hat{y})^2 + 2 \sum [y - \bar{y} - \beta_1 (x - \bar{x})][\bar{y} + \beta_1 (x - \bar{x}) - y] + \sum (\hat{y} - y)^2 \\ &= \sum (y - \hat{y})^2 + 2 \sum [(y - \bar{y}) - \beta_1 (x - \bar{x})][\beta_1 (x - \bar{x})] + \sum (\hat{y} - y)^2 \\ &= \sum (y - \hat{y})^2 + 2\beta_1 [\sum (y - \bar{y})(x - \bar{x}) - \beta_1 \sum (x - \bar{x})^2] + \sum (\hat{y} - y)^2 \quad (2) \end{aligned}$$

Menggunakan konsep keadah kuasadua terkecil, nilai  $\beta_1$  dianggar menggunakan persamaan di bawah:

$$\beta_1 = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2} \quad (3)$$

Dari persamaan (3) didapati:

$$\beta_1 \sum (x - \bar{x})^2 = \sum (x - \bar{x})(y - \bar{y}) \quad (4)$$

Dengan menggantikan persamaan (4) ke dalam (2) kita dapati:

$$\sum (y - \hat{y})^2 = \sum (y - \bar{y})^2 + \sum (\hat{y} - y)^2 \quad (5)$$

Nilai  $\sum (y_i - \bar{y})^2$  dipanggil *jumlah sisihan kuasadua* (sum of squared deviations, SST). Nilai ini adalah jumlah perbezaan di antara nilai-nilai  $y$  sebenar daripada nilai-nilai anggaran yang didapati dengan mengangap tiada regrasi linear di antara pembolehubah  $Y$  dan  $X_1$ . Dengan andaian ini

model menjadi  $y = \mu + \varepsilon$  di mana nilai  $\mu$  dianggar menggunakan nilai  $\bar{y}$ .

Nilai  $\sum (y_i - \hat{y})^2$  dipanggil *jumlah kuasadua reja atau kesalahan* (residual or

error sum of squares, SSE). Nilai ini adalah jumlah kuasadua sisihan nilai Y sebenar daripada nilai anggaran yang diperolehi menggunakan persamaan regresi.

Nilai  $\sum (y - \hat{y})^2$  pula merupakan jumlah kuasadua perbezaan di antara anggaran wujudnya model regresi dengan tidak wujud regresi pembolehubah Y ke atas pembolehubah  $X_i$ . Nilai ini dipanggil *jumlah kuasadua regresi* (sum of squares due to regression, SSR).

Persamaan (5) telah menyatakan bahawa ketiga-tiga nilai ini mempunyai kaitan  $SST = SSR + SSE$ . Konsep yang sama juga boleh digunakan bagi model regresi berganda.

Membahagi jumlah-jumlah kuasa dua dengan darjah kebebasan (degree of freedom) masing-masing memberikan nilai purata kuasadua (mean squares) iaitu:

$MSR = SSR/m$  di mana m adalah bilangan pembolehubah bebas di dalam model dan  
 $MSE = SSE/(n-m-1)$  n adalah bilangan data.

Nilai F pula dikira dengan membahagikan MSR dengan MSE yang akan menentukan samada kita menolak atau menerima hipotesis null ( $H_0$ ).

Selain daripada nilai F, kebagusan model juga boleh ditentukan dengan melihat nilai  $R^2$ . Lebih baik *hyperplane* mewakili data, nilai  $R^2$  akan menghampiri 1.

Nilai  $R^2$  dikira menggunakan formula berikut:

$$R^2 = \frac{SSR}{SST} = \frac{\sum (\hat{y}_j - \bar{y})^2}{\sum (y_j - \bar{y})^2} = 1 - \frac{\sum (y_j - \hat{y}_j)^2}{\sum (y_j - \bar{y})^2} \quad (6)$$

Apabila nilai darjah kebebasan kecil (bilangan data yang digunakan sedikit), nilai  $R^2$  bertambah besar apabila lebih banyak pembolehubah ditambah ke dalam model. Pengukuran yang bebas dari kecenderongan ini adalah  $R^2$  yang telah diselaraskan,  $R^2(\text{sel})$ . Formula yang digunakan untuk mengira nilai baru ini adalah:

$$R^2(\text{sel}) = 1 - \frac{\sum (y_j - \hat{y}_j)^2 / (n - m - 1)}{\sum (y_j - \bar{y})^2 / (n - 1)} \quad (7)$$

Faktor  $(n-1)/(n-m-1)$  akan menghampiri 1 apabila nilai n (saiz sample) bertambah besar. Dengan ini perbezaan di antara  $R^2$  dengan  $R^2(\text{sel})$  boleh diabaikan apabila keadaan ini berlaku.

#### 4.2 Pengujian Kebagusan Setiap Parameter Model

Pengujian hipotesis null adalah pengujian di mana  $\beta_i$  bersamaan dengan nilai-nilai tertentu katakan  $\beta_{i0}$ . Untuk tujuan ini andaian bahawa  $b_i$  bertaburan normal adalah diperlukan. Pengujian statistik yang digunakan adalah:

$$t = \frac{b_i - \beta_{i0}}{s_{b_i}}$$

iaitu bertaburan Student't dengan darjah kebebasan  $n-m-1$ . Nilai  $s_{b_i}$  dikira menggunakan kaedah berikut:

Perkenalkan beberapa pembolehubah-pembolehubah baru seperti berikut untuk memudahkan pengiraan:

$$y'_j, x'_{1j}, x'_{2j}, \dots, x'_{mj}$$

di mana

$$y'_j = y_j - \bar{y}$$

$$x'_{1j} = x_{1j} - \bar{x}_1$$

$$x'_{2j} = x_{2j} - \bar{x}_2$$

.

.

.

$$x'_{mj} = x_{mj} - \bar{x}_m$$

Bila andaian di dalam seksyen 2.0 digunakan, nilai  $b_i$  akan bertaburan normal dengan purata  $\beta_i$  dan variannya adalah  $c_{ii}\sigma^2$ .  $\sigma^2$  merupakan varian kepada subpopulasi nilai-nilai Y yang tidak diketahui dan dianggarkan menggunakan  $s^2_{y.12\dots k}$ . Anggaran kesalahan piawai  $s_{b_i}$  adalah:

$$s_{b_i} = s_{y.12\dots k} \sqrt{c_{ii}}$$

$s^2_{y.12\dots k}$  merupakan punca kuasadua varian yang tidak diterangkan (unexplained variance) ataupun kuasadua purata kesalahan (error mean square) yang boleh didapati dari jadual analisis varian. Nilai  $c_{ii}$  pula boleh didapati dengan menyelesaikan sistem persamaan berikut:

$$c_{11} \sum x'^2_{1j} + c_{12} \sum x'_{1j} x'_{2j} + \dots + c_{1n} \sum x'_{1j} x'_{nj} = 1$$

$$c_{21} \sum x'^2_{1j} + c_{22} \sum x'_{1j} x'_{2j} + \dots + c_{2n} \sum x'_{1j} x'_{nj} = 0$$

.

.

$$c_{m1} \sum x'^2_{1j} + c_{12} \sum x'_{1j} x'_{2j} + \dots + c_{mn} \sum x'_{1j} x'_{nj} = 0$$

$$c_{11} \sum x'_{1j} x'_{2j} + c_{12} \sum x'^2_{2j} + \dots + c_{1n} \sum x'_{2j} x'_{nj} = 0$$

$$c_{21} \sum x'_{1j} x'_{2j} + c_{22} \sum x'^2_{2j} + \dots + c_{2n} \sum x'_{2j} x'_{nj} = 1$$

.

.

$$c_{m1} \sum x'_{1j} x'_{2j} + c_{12} \sum x'^2_{2j} + \dots + c_{mn} \sum x'_{2j} x'_{nj} = 0$$

$$c_{11} \sum x'_{1j} x'_{nj} + c_{12} \sum x'_{2j} x'_{nj} + \dots + c_{1n} \sum x'^2_{nj} = 0$$

$$c_{21} \sum x'_{1j} x'_{nj} + c_{22} \sum x'_{2j} x'_{nj} + \dots + c_{2n} \sum x'^2_{nj} = 0$$

$$c_{m1} \sum x'_{1j} x'_{nj} + c_{m2} \sum x'_{2j} x'_{nj} + \dots + c_{mn} \sum x'^2_{nj} = 1$$

Di mana  $c_{ij} = c_{ji}$  dan persamaan diberikan nilai 1 apabila mengandungi ungkapan  $c_{ii} \sum x'^2_{ij}$ .

### 5.0 KAJIAN KES:

#### KEMALANGAN JALANRAYA YANG BERLAKU DI MALAYSIA BARAT

Katakan kita ingin mengkaji hubungan di antara jumlah kemalangan jalanraya dengan pembolehubah-pembolehubah bebas jumlah penduduk, jumlah kenderaan berdaftar dan jumlah panjang jalan (km). Data-data bagi Malaysia Barat dari tahun 1979 hingga tahun 1987 akan digunakan sebagai sampel. Data-data tersebut ditunjukkan dalam Jadual 1 di bawah:

Tahun	Jumlah Kemalangan (Y)	Jumlah Penduduk (X <sub>1</sub> )	Jumlah Kenderaan Berdaftar (X <sub>2</sub> )	Jumlah Panjang Jalan (X <sub>3</sub> )
1979	57,931	11,188,630	2,025,238	22,035
1980	59,084	11,442,086	2,357,386	23,113
1981	58,768	11,734,839	2,631,948	23,816
1982	68,330	12,039,195	2,930,098	24,434
1983	72,634	12,344,759	3,253,233	27,440
1984	73,727	12,817,150	3,565,205	28,582
1985	75,300	13,155,591	3,825,495	29,692
1986	71,557	13,491,549	4,005,172	29,805
1987	68,066	13,655,073	4,116,527	30,692

Jadual 1: Data-data jumlah kemalangan, Jumlah penduduk, jumlah kenderaan berdaftar dan jumlah panjang jalan. (Sumber: Statistical Report Road Accidents Malaysia, Royal Malaysia Police, 1982)

Hasil dari pengiraan kita boleh bentuk jadual di bawah untuk mewakili persamaan bagi memudahkan penyelesaian.

	X1	X2	X3	Y
X1	6464274355458.250	5359346913509.620	23248067709.667	38571329047.001
X2	5359346913509.690	4500476434466.220	19502718348.889	34259616816.334
X3	23248067709.667	19502718348.889	86646629.556	151680852.333

Persamaan di atas jika diwakilkan ke dalam bentuk matrik menjadi seperti berikut:

$$A = \begin{pmatrix} 6464274355458.250 & 5359346913509.620 & 23248067709.667 \\ 5359346913509.690 & 4500476434466.220 & 19502718348.889 \\ 23248067709.667 & 19502718348.889 & 86646629.556 \end{pmatrix}$$

$$b = \begin{pmatrix} 38571329047.001 \\ 34259616816.334 \\ 151680852.333 \end{pmatrix}$$



$$X = (x_1, x_2, x_3)^T$$

Jika dicantumkan matrik A dan b, kita dapati:

$$[A|b] = \left( \begin{array}{cccccc|c} 6464274355458.250 & 5359346913509.620 & 23248067709.667 & 38571329047.001 & & & \\ 5359346913509.690 & 4500476434466.220 & 19502718348.889 & 34259616816.334 & & & \\ 23248067709.667 & 19502718348.889 & 86646629.556 & 151680852.333 & & & \end{array} \right)$$

Nilai  $b_i$  boleh didapati dengan menyelesaikan sistem persamaan linear tersebut. Berbagai-bagai kaedah boleh digunakan di antaranya kaedah Penghapusan Gauss. Kaedah ini dilakukan dengan menukar matrik A menjadi matrik bersegi tiga atas di mana  $a_{ij} = 0$  untuk nilai  $i > j$ .

$$A = \left( \begin{array}{cccccc|c} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & \dots & a_{1n} & | & \dots \\ 0 & a_{22} & a_{23} & a_{24} & a_{25} & \dots & a_{2n} & | & \dots \\ 0 & 0 & a_{33} & a_{34} & a_{35} & \dots & a_{3n} & | & \dots \\ 0 & 0 & 0 & a_{44} & a_{45} & \dots & a_{4n} & | & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & | & \dots \\ 0 & 0 & 0 & 0 & 0 & \dots & a_{nn} & | & \dots \end{array} \right)$$

Penukaran ini boleh dilakukan menggunakan kaedah strategi pengantian ke belakang. Kemudiannya penyelesaian kepada setiap pembolehubah boleh didapati dengan melakukan pengiraan berikut<sup>9</sup>:

$$x_n = \frac{b_n}{a_{nm}}$$

$$x_{n-1} = \frac{(b_{n-1} - a_{n-1,n}x_n)}{a_{n-1,n-1}}$$

Dari penyelesaian tersebut didapati nilai  $b_1 = -0.0276$ ,  $b_2 = 0.0326$  dan  $b_3 = 1.8130$ . Nilai  $b_0$  yang dikira menggunakan persamaan (m+1) didapati bersamaan dengan 258159.0449.

Dengan ini persamaan regresi berganda yang di dapati untuk kes kajian adalah:

$$Y = 258159.0449 - 0.0276x_1 + 0.0326x_2 + 1.8130x_3$$

Hyperplane yang dihasilkan adalah merupakan yang terbaik bagi mewakili data kemalangan jalanraya yang telah dianalisis secara linear.

### 5.1 Pengujian kebangusan Model

Dalam kes kajian ini didapati nilai SST = 381593230, SSR=328123335 dan SSE = 53469895. Nilai MSR = SSR/3 = 328123335.16 / 3 = 109374445 manakala nilai MSE = SSE / (9-3-1) = 53469895 / 5 = 10693979. Nilai F pula bersamaan dengan MSR / MSE = 109374445 / 10693979 = 10.228. Kesemua nilai ini biasanya diletakkan di dalam jadual analisis varian seperti di bawah.

Sumber variasi	Jumlah kuasadua (SS)	Darjah kebebasan (df)	Purata kuasadua (MS)	F
Regrasi linear	328123335	3	109374445	10.228
Sisihan dari kelinearan (kesalahan)	53469895	5	10693979	
Total	381593230	8		

Jadual 2: Jadual analisis varian untuk kajian kes

Statistik pengujian adalah  $F = MSR / MSE$ , taburan F dengan m dan (n-m-1) darjah kebebasan. Katakan paras bererti yang dikehendaki adalah  $\alpha = 0.025$ . Dengan ini peraturan membuat keputusan adalah: tolak  $H_0$  sekiranya nilai F yang dikira lebih besar atau sama dengan 7.76 (rujuk jadual F dengan  $\alpha = 0.025$  dan darjah kebebasannya adalah 3 dan 5).

Nilai F yang telah dikira adalah 10.228 di mana lebih besar daripada 7.76. Dengan ini kita tolak  $H_0$ . Kesimpulannya data-data kemalangan yang telah digunakan menyokong pada paras bererti 0.025 bahawa terdapat hubungan linear di antara jumlah kemalangan yang berlaku dengan faktor jumlah penduduk ( $x_1$ ), jumlah kenderaan berdaftar ( $x_2$ ) dan jumlah panjang jalan ( $x_3$ ).

Menggunakan persamaan (6) dan (7) didapati nilai  $R^2$  dan  $R^2(\text{Sel})$  adalah 0.8599 dan 0.7758. Berdasarkan nilai  $R^2$ , regresi Y ke atas  $x_1$ ,  $x_2$  dan  $x_3$  menerangkan 86% jumlah variasi di dalam Y dan menyokong keputusan yang dibuat sebelum ini.

Nilai bagi  $s_{b_1}$ ,  $s_{b_2}$  dan  $s_{b_3}$  boleh didapati dengan menyelesaikan sistem persamaan berikut menggunakan kaedah Penghapusan Gauss:

$$\begin{aligned}
 6464274355458c_{11} + 5359346913510c_{12} + 23248067710c_{13} &= 1 \\
 6464274355458c_{21} + 5359346913510c_{22} + 23248067710c_{23} &= 0 \\
 6464274355458c_{31} + 5359346913510c_{32} + 23248067710c_{33} &= 0 \\
 \\ 
 5359346913510c_{11} + 4500476434466c_{12} + 19502718349c_{13} &= 0 \\
 5359346913510c_{21} + 4500476434466c_{22} + 19502718349c_{23} &= 1 \\
 5359346913510c_{31} + 4500476434466c_{32} + 19502718349c_{33} &= 0 \\
 \\ 
 23248067710c_{11} + 19502718349c_{12} + 86646630c_{13} &= 0 \\
 23248067710c_{21} + 19502718349c_{22} + 86646630c_{23} &= 0 \\
 23248067710c_{31} + 19502718349c_{32} + 86646630c_{33} &= 1
 \end{aligned}$$

Dengan menyelesaikan persamaan-persamaan di atas, kita dapati penyelesaian kepada  $c_{11}$ ,  $c_{22}$  dan  $c_{33}$  adalah 0.000000000122, 0.000000000255 dan 0.0000004705170.

Oleh itu nilai  $s_{b_1}$  adalah seperti berikut:

$$\begin{aligned}
 s_{b_1} &= s_{y.123} \tilde{A} c_{11} \\
 &= \tilde{A} 10693978.968 \times \tilde{A} 0.000000000122 = 0.0114274236871
 \end{aligned}$$

$$s_{b_2} = s_{y.123} \tilde{A} c_{22}$$

$$= \sqrt{10693978.968} \times \sqrt{0.000000000255} = 1.9759652256549$$

$$s_{b_3} = s_{y.123} \sqrt{c_{33}}$$

$$= \sqrt{10693978.968} \times \sqrt{0.0000004705170} = 2.2431449768253$$

Pengujian kebagusan parameter  $\beta_1$  adalah seperti berikut:

Hipotesis yang hendak diuji adalah:  $H_0 : \beta_1 = 0$  dan  $H_1 : \beta_1 \neq 0$ . Ujian statistik yang akan digunakan  $t = (b_1 - \beta_{10}) / s_{b_1}$ . Paras kekesanan yang diperlukan  $\alpha = 0.1$ . Peraturan membuat keputusan adalah tolak  $H_0$  jika nilai  $t$  yang dikira  $> 2.015$  atau  $< -2.015$ , nilai genting  $\alpha = 0.1$  (pengujian dua belah) dan darjah kebebasan adalah 5. Nilai  $t$  dari pengiraan adalah  $t = (-0.0276 - 0) / 0.0114274236871 = -2.4169$ . Oleh kerana  $-2.4169 < -2.105$ , tolak  $H_0$ . Kesimpulannya adalah kita boleh membuat kesimpulan bahawa dengan kehadiran  $x_2$  dan  $x_3$ ,  $x_1$  (jumlah penduduk) adalah berhubungan linear dengan jumlah kemalangan yang berlaku dan mempengaruhi di dalam proses ramalan bilangan kemalangan yang berlaku.

Jika pengiraan yang sama dilakukan untuk  $\beta_2$  dan  $\beta_3$ , nilai  $t$  yang dikira untuk kedua-dua parameter ini adalah 1.9759 dan 0.8082. Oleh kerana kedua-dua nilai ini berada di antara nilai -2.015 dan 2.015, kita tidak boleh menolak  $H_0$ . Dengan ini kita tidak boleh membuat kesimpulan bahawa dengan kehadiran  $x_1$  dan  $x_3$  di dalam model,  $x_2$  (jumlah kenderaan berdaftar) berhubungan linear dengan jumlah kemalangan yang berlaku. Begitulah juga untuk  $x_3$ , jumlah panjang jalan.

## 6.0 ANALISIS REGRASI BERGANDA MENGGUNAKAN SAS

Proses mendapatkan model regresi lebih mudah jika pakej komputer digunakan. Dalam kajian ini pakej SAS/STAT dan prosidur REG digunakan. Aturcara SAS/STAT yang digunakan ditunjukkan dalam Lampiran A manakala outputnya ditunjukkan dalam Lampiran B.

Set data sas *kemal* digunakan untuk menyimpan data kemalangan. Lima pembolehubah yang terlibat iaitu tahun, jumlah kemalangan (*j\_kem*), jumlah penduduk (*j\_pend*), jumlah kenderaan berdaftar (*j\_kend*) dan jumlah panjang jalan (*j\_pan\_j*) ditakrifkan di dalam pernyataan input dimana menyebabkan sas membaca data ke dalam set data sas kemal. Data-data yang hendak diproses diletakkan selepas pernyataan CARDS. Kemudiannya prosedur REG dipanggil untuk melakukan analisis regresi ke atas data yang telah tersimpan di dalam sas data set kemal. Kertas kerja ke lapan dalam senarai rujukan menerangkan tentang aturcara sas dan outputnya dengan lebih lengkap.

## 7.0 KESIMPULAN

Analisis regresi berganda merupakan model regresi linear bagi menghubungkan pembolehubah bersandar dengan beberapa pembolehubah bebas. Model ini biasanya digunakan dalam proses ramalan nilai pembolehubah bersandar sekiranya diberikan nilai pembolehubah-pembolehubah bebas yang baru. Selain dari model linear, terdapat banyak model-model yang lain yang menggunakan konsep tidak linear yang boleh melakukan proses yang sama. Walaubagaimana pun oleh kerana kertas kerja ini hanya membincangkan model linear sahaja, perbincangan model-model yang tidak linear akan dilakukan di dalam kertas kerja yang akan datang.

**RUJUKAN**

1. Wayne W. Daniel, James C. Terrell, *Business Statistics For Management and Economics*, fifth edition, Houghton Mifflin Company, Boston
2. *Statistical Report Road Accidents Malaysia*, Royal Malaysia Police, 1982
3. *SAS/STAT User Guide*, Version 6.
4. *SAS/BASIC User Guide*, Version 6.
5. J. Supranto (1986) *Kaedah Penyelidikan Penggunaannya Dalam Pemasaran*, Dewan Bahasa dan Pustaka.
6. Richard I. Levin and David S. Rubin, *Statistics For Management*, Sixth Edition, Prentice Hall International Editions.
7. Rudolf J. Freund and Paul D. Minton (1979) *Regression Methods, A Tool for Data Analysis*, *Statistics: textbooks and monographs* Volume 30, Marcel Dekker, inc.
8. Syed Othmawi Abd. Rahman dan P.M Jamilin Jais (1994) *Regrasi Linear Mudah: Kajian Kes Kemalangan Jalan Raya di Malaysia Barat*, *Jurnal Teknologi Maklumat*, Jilid 6, Bil. 1.
9. Ab Rahman Ahmad (1990) *Kaedah Berangka Permulaan, Pengaturcaraan dengan Turbo Pascal*, Dewan Bahasa dan Pustaka.

## LAMPIRAN A

TITLES 'ANALISIS KEMALANGAN JALAN RAYA DI MALAYSIA BARAT';  
DATA KEMLGN;  
INPUT TAHUN KEMAL PENDUD KENDER JALAN;  
CARDS;

1979 57931 11188630 2025238 22035  
1980 59084 11442086 2357386 23113  
1981 58768 11734839 2631948 23816  
1982 68330 12039195 2930098 24434  
1983 72634 12344759 3253233 27440  
1984 73727 12817150 3565205 28582  
1985 75300 13155591 3825495 29692  
1986 71557 13491549 4005172 29805  
1987 68066 13655073 4116527 30692

PROC CORR DATA=KEMLGN;  
VAR KEMAL PENDUD KENDER JALAN;

PROC REG DATA=KEMLGN;  
MODEL KEMAL=PENDUD KENDER JALAN / P R;

## LAMPIRAN B

ANALISIS KEMALANGAN JALAN RAYA DI MALAYSIA BARAT 11:48 Friday, April 21, 1995 1

## Correlation Analysis

4 'VAR' Variables: KEMAL PENDUD KENDER JALAN

## Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
KEMAL	9	67266	6706.457395	605397	57931	75300
PENDUD	9	12429875	898907	111868872	11186630	13655073
KENDER	9	3190034	750340	28710302	2025238	4116527
JALAN	9	26623	3291.022439	239609	22035	30692

Pearson Correlation Coefficients / Prob &gt; |RI| under Ho: Rho=0 / N = 9

	KEMAL	PENDUD	KENDER	JALAN
KEMAL	1.00000 0.0	0.77661 0.0138	0.82671 0.0060	0.83417 0.0052
PENDUD	0.77661 0.0138	1.00000 0.0	0.99363 0.0001	0.98232 0.0001
KENDER	0.82671 0.0060	0.99363 0.0001	1.00000 0.0	0.98762 0.0001
JALAN	0.83417 0.0052	0.98232 0.0001	0.98762 0.0001	1.00000 0.0

ANALISIS KEMALANGAN JALAN RAYA DI MALAYSIA BARAT

11:48 Friday, April 21, 1995 2

Model: YJJEL1  
 Dependent Variable: KEMAL

## Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	3	328123335.16	109374445.05	10.228	0.0142
Error	5	53469896.842	10693978.968		
C Total	8	381593230			
Root MSE		3270.16498	R-square	0.8599	
Dep Mean		67266.33333	Adj R-sq	0.7758	
C.V.		4.86152			

## Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob >  T
INTERCEP	1	258159	101293.20664	2.549	0.0514
PENDUD	1	-0.027619	0.01142742	-2.417	0.0603
KEMDER	1	0.032646	0.01634702	1.997	0.1023
JALAN	1	1.8612965	2.24314498	0.808	0.4557

ANALISIS KEMALANGAN JALAN RAYA DI MALAYSIA BARAT

11:48 Friday, April 21

Obs	Dep Var KEMAL	Predict Value	Std Err Predict	Residual	Std Err residual	Student Residual	-2 -1 0 1 2	Cook's D
1	57931.0	55204.3	2838.842	2726.7	1623.255	1.680		2.158
2	59084.0	61001.7	1598.466	-1917.7	2794.493	-0.686		0.043
3	58768.0	63154.3	1666.087	-4386.0	2813.918	-1.559		0.213
4	68330.0	65601.7	2764.818	2728.3	1746.356	1.562		1.529
5	72634.0	73161.1	2507.919	-527.1	2098.647	-0.251		0.023
6	73727.0	72369.1	1612.357	1357.9	2845.045	0.477		0.018
7	75300.0	73531.5	1673.553	1768.5	2809.484	0.629		0.035
8	71557.0	70323.3	2134.954	1233.7	2477.085	0.498		0.045
9	68066.0	71053.3	2272.417	-2984.3	2351.616	-1.269		0.376

Total Residuals 0  
 Sum of Squared Residuals 53449894.842  
 Adjusted R Squared SS (Press) 304494501.68