

Analisis Kovariat bagi Data Kemortalan Bayi

Noor Akma Ibrahim dan Isa Daud

Jabatan Matematik

Fakulti Sains dan Pengajian Alam Sekitar

Universiti Putra Malaysia

43400 UPM, Serdang, Selangor Darul Ehsan, Malaysia

Diterima 12 November 1994

ABSTRAK

Makalah ini memaparkan analisis untuk menilai keberkesanan kovariat ke atas kemortalan bayi dengan mempraktikkan kaedah penganalisisan kes hayat terkumpul dalam risiko bersaing. Di samping melalui model ini, kebarangkalian kegagalan di dalam sesuatu selang masa bagi sesuatu individu bersekutu dengan kovariat tertentu boleh dimanfaatkan untuk memperteguhkan kesignifikanan kovariat tersebut di dalam model.

ABSTRACT

This paper analyses the effect of various covariates on infant mortality. It uses the method for the analysis of grouped life-time in competing risks. With this approach, advantage is taken of the probability of an individual with a particular covariate failing in an interval to consolidate the significance of that covariate model.

Kata kunci: risiko bersaing, kovariat, data terkumpul, kaedah kebolehjadian, taburan eksponen

PENGENALAN

Tujuan kertas ini adalah untuk mengkaji keberkesanan pemboleh ubah penerangan, juga dikenali sebagai kovariat ke atas masa kegagalan data mandirian khususnya di sini data kemortalan bayi. Kovariat lazimnya meliputi pemboleh ubah kuantitatif dan kualitatif. Seperti mana dalam kertas ini rawatan klinikal (kovariat) ke atas kumpulan boleh digabungkan dengan mengguna pemboleh ubah penuding.

Data kemortalan bayi ini diperolehi daripada Jabatan Perangkaan Malaysia. Kira-kira 5201 kematian dilaporkan di Semenanjung Malaysia bagi tahun 1989 iaitu tahun mutakhir catatan lengkap dibuat bagi tujuan laporan demografi. Data yang ada di Jabatan Perangkaan Malaysia adalah lebih lengkap meliputi maklumat penting seperti umur ibu ketika melahirkan, turutan bayi di dalam keluarga, berat badan bayi ketika lahir, status kelahiran (normal atau tidak, komplikasi dll.) dan umur saudara bayi sebelumnya. Disebabkan kerahsiaan negara beberapa maklumat tidak dapat diperolehi. Walau bagaimanapun terdapat tiga kovariat yang boleh dianggap bersesuaian dengan kajian kami

iaitu jantina, bangsa dan tempat kediaman di samping maklumat punca kematian setiap bayi.

Kesemua maklumat dalam bentuk kod dan proses pentafsirannya berpandukan Integrasi Sistem Perangkaan Penting Semenanjung Malaysia, Sabah dan Sarawak, 1984, dibekalkan oleh Jabatan Perangkaan Malaysia. Daripada 5201 kematian, 200 cerapan dipilih secara rawak. Masa kegagalan atau masa hayat yang sebenarnya tidak diketahui disebabkan umur bayi pada kematian dicatatkan ke dalam suatu selang masa (dalam minggu selepas dilahirkan). Lazimnya apabila menguruskan data yang melibatkan bilangan cerapan besar, catatan masa hayat sering dikumpul dalam selang, berbeza dalam keadaan dalam masa hayat sebenarnya dicatatkan. Dengan demikian ianya lebih mudah dikendalikan dan dari segi persembahan lebih mudah difahami. Lagipun, mengikut Thompson (1977), data terkumpul adalah sebaik-baik cara untuk menangani cerapan yang mempunyai nilai hayat yang sama. Secara teori cerapan yang nilai hayatnya sama tidak akan terhasil jika masa hayat selanjut tetapi hakikatnya tidak begitu. Seajar dengan ini maka data kemortalan bayi ditangani sebagai data terkumpul.

Beberapa makalah telah dihasilkan berhubung dengan data terkumpul. Antaranya ialah Kimball (1969), melibatkan anggaran kebarangkalian kematian disebabkan sesuatu punca setelah punca atau punca-punca lain dihapuskan, Turnbull (1976) dengan anggaran tak berparameter, Holford (1976) membandingkan dua lengkung mandirian, Turnbull dan Weiss (1978) melakukan penyesuaian taburan Weibull, Gompertz dan kuasa eksponen kepada data hayat, Lindley (1979) menggunakan pendekatan Bayesian, Pierce *et al.* (1979) menganalisis data terkumpul berpandukan model Cox, Burrige (1981) mempertimbangkan anggaran parameter mengguna kaedah memparameterkan semula taburan normal, Pettitt (1985) melaksanakan anggaran kuasa dua dengan pemberat dan Turrero (1989) mengkaji kecekapan relatif data terkumpul berdasarkan pendekatan Bayesian.

Oleh kerana punca kematian berhubung dengan data kemortalan bayi ini terlalu banyak dan sangat terperinci, maka bagi kemudahan pengiraan, punca ini digulungkan ke dalam dua kategori sahaja iaitu penyebab pertama sebagai diketahui dan penyebab kedua sebagai tidak diketahui. Dengan ini model risiko bersaing dengan dua risiko ($k = 2$) bagi kes terkumpul dilaksanakan.

RISIKO BERSAING

Risiko bersaing merupakan suatu model mandirian digunakan untuk menganalisis data yang tertakluk kepada beberapa mekanisma kegagalan yang berbeza. Mekanisma kegagalan ini lazimnya dikenali sebagai risiko. Teori risiko bersaing adalah juga berkaitan dengan penilaian ke atas sesuatu risiko di dalam situasi yang kompleks dengan kehadiran risiko yang lain. Teori ini lazim digunakan dalam penganalisisan data klinikal termasuk data kebolehppercayaan komponen di dalam bidang industri, bidang ekonomi, demografi dan aktuari. Kesemua punca kegagalan atau risiko di dalam model

ini lazimnya diandaikan saling merdeka dan kegagalan sebenar hanya boleh disebabkan oleh satu risiko sahaja dengan kehadiran risiko-risiko yang lain.

Tiga kebarangkalian kegagalan dalam risiko bersaing, pertama, kebarangkalian hakiki iaitu kebarangkalian kegagalan secara teori dengan kehadiran satu risiko sahaja. Kedua, kebarangkalian kegagalan kasar iaitu kebarangkalian kegagalan disebabkan risiko tertentu dengan kehadiran risiko yang lain. Akhir sekali ialah kebarangkalian kegagalan kasar separa iaitu kebarangkalian disebabkan sesuatu risiko setelah sekurang-kurangnya satu risiko dikeluarkan. Gail (1975) telah membincangkan beberapa model risiko bersaing dan kebarangkalian kasar separa yang ditakrifkan dari fungsi ketumpatan sut. Di dalam kertas ini ketiga-tiga kebarangkalian akan diperluaskan berdasarkan data terkumpul.

Formulasi Asas

Katakan Y_i pemboleh ubah rawak mewakili masa kegagalan jika diandaikan berpunca dari risiko R_i ($i = 1, \dots, k$). Fungsi taburan Y_i ditandakan $P_i(t) = Kb(Y_i \leq t)$ dengan $Kb(\cdot)$ tandaan kebarangkalian dan fungsi ketumpatan kebarangkaliannya, $p_i(t)$, selalunya diandaikan wujud.

Lazimnya Y_i tidak tercerap secara mutlak kerana kehadiran beberapa risiko yang bersaing. Jadi, apa yang tercerap ialah masa kegagalan T bagi seseorang individu berpunca dari salah satu risiko tertentu dengan kehadiran risiko yang lain, iaitu

$$T = \text{minimum } (Y_1, \dots, Y_k)$$

Berpandukan Gail (1975), masalah risiko bersaing boleh juga dibentuk menerusi ungkapan berikut:

$$S(t_1, t_2, \dots, t_k) = Kb(Y_1 \geq t_1, Y_2 \geq t_2, \dots, Y_k \geq t_k) \quad (1)$$

dikenali sebagai fungsi mandirian bercantum memenuhi syarat $S(0, 0, \dots, 0) = 1$ dan $S(\infty, \infty, \dots, \infty) = 0$ bagi setiap $t_i \in (0, \infty)$. S adalah selanjur dari kanan dan tak menokok berekanada pada setiap masa kegagalan. Berikutan dengan ini

$$S_T(t) = Kb(T \geq t) = Kb(\text{semua } Y_i \geq t) = S(t, t, \dots, t) \quad (2)$$

Ungkapan (2) adalah fungsi mandirian bagi T sama ada masa kegagalan terpendam (pengamatan yang tidak tercerap secara mutlak) merdeka atau sebaliknya. Katakan $f_T(t)$ fungsi kebarangkalian bagi T yang dicerap, maka fungsi bahaya bersyarat bagi T ialah

$$h_T(t) = \frac{f_T(t)}{S_T(t)} \quad (3)$$

Pendekatan Berparameter

Dengan pendekatan berparameter di dalam risiko bersaing, diandaikan taburan pendasar bagi masa kegagalan diketahui. Dari itu penganggar parameter lazimnya diperolehi dengan menggunakan kaedah kebolehhadiah maksimum dengan pendekatan yang bersesuaian. Pemilihan bentuk fungsi taburan masa kegagalan akan digunakan untuk membuat pentakbiran. Ia bergantung keseluruhannya pada kefahaman penyelidik terhadap mekanisma kegagalan.

Menurut David dan Moeschberger (1978), di dalam kebanyakan analisis, taburan eksponen berjaya mewakili masa kegagalan bukan sahaja bagi kajian kebolehppercayaan komponen-komponen tetapi juga bagi mewakili hayat mandirian. Taburan eksponen mempunyai bentuk matematik yang mudah lantas fungsi bahaya dan fungsi mandirian boleh diperolehi di dalam bentuk yang mudah. Di samping itu taburan ini mempunyai dua sifat unik iaitu fungsi bahaya pemalar dan sifat tanpa ingatan. Berdasarkan ciri fungsi bahaya pemalar, taburan eksponen dipilih sebagai taburan pendasar bagi masa kegagalan dalam analisis kemortalan bayi ini.

Kes Hayat Dikumpul ke Dalam Selang

Di dalam seksyen sebelum ini telah diperkenalkan suatu pemboleh ubah rawak bukan negatif Y_i mewakili masa kegagalan secara teori bagi sesuatu individu apabila hanya risiko R_i ($i = 1, \dots, k$) yang hadir.

Katakan T_i masa kegagalan tercerap bersyaratkan punca kegagalan diketahui, R_i , maka

$$T_i = [Y_i | R_i] = \text{minimum } (Y_1, \dots, Y_k) \tag{4}$$

Kebarangkalian kegagalan berpunca dari risiko R_i boleh ditulis sebagai

$$\pi_i = Kb[Y_i = \text{minimum } (Y_1, \dots, Y_k)]$$

dengan $\pi_i > 0$ dan $\sum_{i=1}^k \pi_i = 1$

Oleh kerana di dalam penyelidikan kami saiz data besar sewajarnya data dipetakan atau dikumpulkan ke dalam beberapa selang masa. Dengan ini persembahan data menjadi lebih menarik serta senang difahami.

Misalkan data masa kegagalan dipetakan ke dalam selang $I_\alpha = (a_\alpha, a_{\alpha+1})$, $\alpha = 0, 1, \dots, r$ dengan $a_0 = 0$ dan $a_{r+1} = \infty$. Biarkan $N_{i\alpha}$ ($\alpha = 0, \dots, r$) bilangan individu yang gagal dari risiko R_i dalam I_{α_i} . Dengan ini, $N_i = \sum_{\alpha=0}^r N_{i\alpha}$ bilangan individu gagal dari risiko R_i dan $N_{.a} = \sum_{i=1}^k N_{i\alpha}$ kesemua individu yang gagal dalam I_α . Jumlah sampel semuanya adalah $n_i = \sum_{i=1}^k \sum_{\alpha=0}^r N_{i\alpha}$ Bilangan kegagalan

yang dicerap ini boleh dikelaskan ke dalam jadual dua hala mengikut selang masa dan punca kegagalan. Oleh itu $N_{i\alpha}$ bertaburan multinomial dan fungsi kebolehjadiannya ialah

$$L_G = \frac{n!}{\prod_{i=1}^k \prod_{\alpha=0}^r (n_{i\alpha})!} \prod_{i=1}^k \prod_{\alpha=0}^r \pi_{i\alpha}^{n_{i\alpha}} \quad (5)$$

dengan

$$\pi_{i\alpha} = K_b \text{ (individu gagal dari } R_i \text{ dalam } I_\alpha).$$

Tetapi

$$\pi_{i\alpha} = \pi_i [F_i(a_{\alpha+1}) - F_i(a_\alpha)],$$

dengan $\pi_i = K_b$ (individu gagal dari R_i) dan F_i fungsi taburan bagi T_i .

Bersekutu dengan masa kegagalan adalah pemboleh ubah penerangan juga dikenali sebagai kovariat yang mempunyai pertalian dengan masa kegagalan. Cox (1972) telah memperkenalkan suatu model dengan menggabungkan kovariat melalui fungsi bahayanya. Berpandukan model Cox ini, di dalam seksyen berikut diperlihatkan bagaimana pendekatan yang sama boleh digunakan bagi model risiko bersaing.

Risiko Bersaing dengan Kovariat

Andai kovariat diwakili oleh vektor $\underline{z}' = (z_1, \dots, z_p)$. Ia merupakan ukuran ke atas ciri individu seperti umur, jantina, berat badan dan sebagainya di samping ukuran ke atas rawatan seperti kumpulan rawatan, sukatan ubatan (dos) dan lain-lain lagi. Yang menjadi masalah adalah bagaimana hendak menilai hubungan di antara masa kegagalan dan kovariat ini. Berkemungkinan juga \underline{z} suatu fungsi masa.

Dari Cox (1972) fungsi ketumpatan kebarangkalian bersyaratkan kehadiran \underline{z} boleh ditulis sebagai

$$p(y|\underline{z}) = h_0(y)c(\underline{z}, \underline{\beta}) \exp \left[- H_0(y) c(\underline{z}, \underline{\beta}) \right] \quad (6)$$

dengan $\underline{\beta}$ vektor lajur bagi parameter berkaitan dengan \underline{z}' , sepadan dengan sesuatu individu dan $c(\underline{z}, \underline{\beta})$ sebarang fungsi bagi $\underline{\beta}$ dan \underline{z} sedemikian $c(\underline{z}, \underline{\beta}) = 1$ jika $\underline{z} = 0$. Fungsi bahayanya berdasarkan (3) ialah

$$\begin{aligned} h(y|\underline{z}) &= \frac{P(y|\underline{z})}{P(y|\underline{z})} \\ &= h_0(y)c(\underline{z}, \underline{\beta}) \end{aligned}$$

dengan fungsi bahaya terhimpun

$$H_0(y|z) = H_0(y) c(z, \underline{\beta})$$

dengan $H_0(y) = \int_0^y h_0(t) dt$

Taburan Eksponen Dengan Data Terkumpul

Dengan masa kegagalan tertabur secara eksponen,

$$h_0(y) = \frac{1}{\theta}$$

dan fungsi ketumpatan kebarangkalian eksponen bersekutu dengan kovariat dengan kegagalan R_i ialah

$$p_i(y) = \frac{1}{\theta_i} c(z, \underline{\beta}) \exp\left[-\frac{1}{\theta_i} y c(z, \underline{\beta})\right] \quad (7)$$

Dengan andaian fungsi bahaya malar pada setiap selang, maka fungsi kebolehdajian boleh diperolehi seperti berikut:

Dari (6) dan (7), fungsi ketumpatan kebarangkalian eksponen bersekutu dengan kovariat dengan $\lambda_i = 1/\theta_i$, $\pi_i = \lambda_i/\lambda$ dan mengambil $c(z, \underline{\beta}) = \exp(z, \underline{\beta})$ adalah

$$p_i(y) = \lambda_i \exp(z', \underline{\beta}) \left[-\lambda_i \exp(z', \underline{\beta}) y\right]$$

Bagi individu ke j ,

$$\begin{aligned} F_{ij}(y) &= \int_0^y \lambda_i \exp(z_{ij}', \underline{\beta}_i) \exp\left[-\lambda_i \exp(z_{ij}', \underline{\beta}_i) y\right] dy \\ &= 1 - \exp\left(-\lambda_i \exp(z_{ij}', \underline{\beta}_i) y\right) \end{aligned}$$

Jadi,

$$\begin{aligned} \pi_{ij\alpha} &= \Pi_i[F_{ij}(a_{\alpha+1}) - F_{ij}(a_\alpha)] \\ &= \frac{\lambda_i}{\lambda} \left[\exp\left(-\lambda_i \exp(z_{ij}', \underline{\beta}_i) a_\alpha\right) - \exp\left(-\lambda_i \exp(z_{ij}', \underline{\beta}_i) a_{\alpha+1}\right) \right] \\ i &= 1, \dots, k; \alpha = 0, \dots, r \end{aligned}$$

Untuk memudahkan tandaan, biar $g_{ij} = \exp\left(-\lambda_i \exp(z_{ij}', \underline{\beta}_i) a_\alpha\right)$ (Maka dari (5),

$$L_g \propto \prod_{i=1}^k \prod_{\alpha=0}^r \left[\prod_{j \in A_\alpha} \frac{\lambda_i}{\lambda} [g_{ij\alpha} - g_{ij(\alpha+1)}] \right]$$

$$= \prod_{i=1}^k \frac{\lambda_i}{\lambda} \prod_{\alpha=0}^r \prod_{j \in A_\alpha} [g_{ij\alpha} - g_{ij(\alpha+1)}]$$

Log kepada kebolehdajadian ialah

$$L = \log L_G$$

$$\begin{aligned} &\propto \sum_{i=1}^k \log \lambda_i - \sum_{i=1}^k \log \lambda + \sum_{i=1}^k \sum_{\alpha=0}^r \sum_{j \in A_\alpha} \log [g_{ij(\alpha+1)}] \\ &= \sum_{i=1}^k \log \lambda_i - k \log \lambda + \sum_{i=1}^k \sum_{\alpha=0}^r \sum_{j \in A_\alpha} \log [g_{ij\alpha} - g_{ij(\alpha+1)}] \end{aligned} \tag{8}$$

dengan $\lambda = \sum_{i=1}^k \lambda_i$

Satu lagi andaian diperlukan di sini dalam penganalisan seterusnya selain daripada kemerdekaan risiko, setiap selang mestilah mengandungi sekurang-kurangnya satu kegagalan. Jika tidak nilai satu akan diberikan kepada sebutan kebolehdajadian yang melibatkan selang yang tiada kegagalan atau tidak mengambil kira selang tersebut (lihat Thompson, 1977). Daripada (8) dengan mengambil terbitan separa L terhadap $\underline{\theta}_i = (\lambda_i, \beta_i)$,

$$U(\underline{\theta}_i) = \begin{pmatrix} \frac{\partial L}{\partial \lambda_i} \\ \frac{\partial L}{\partial \beta_i} \end{pmatrix} \tag{9}$$

yang mana

$$\frac{\partial L}{\partial \lambda_i} = \frac{1}{\lambda_i} - \frac{k}{\lambda} + \sum_{\alpha=0}^r \sum_{j \in A_\alpha} \left[\frac{g_{ij(a+1)} b_{ij(a+1)} - g_{ij\alpha} b_{ij\alpha}}{g_{ij\alpha} - g_{ij(\alpha+1)}} \right]$$

$$\frac{\partial L}{\partial \beta_i} = \sum_{\alpha=0}^r \sum_{j \in A_\alpha} \left[\frac{g_{ij} (-\lambda_i a_\alpha z_{ij}') - g_{ij(a+1)} (-\lambda_i a_{\alpha+1})}{g_{ij\alpha} - g_{ij(\alpha+1)}} \right]$$

dengan $b_{ij\alpha} = a_{\alpha}$ eksp $(z'_{ij} \underline{\beta}_i)$, dan

$$I(\underline{\theta}_i) = - \begin{pmatrix} \frac{\partial^2 L}{\partial \lambda_i^2} & \frac{\partial^2 L}{\partial \lambda_i \partial \underline{\beta}_i} \\ \frac{\partial^2 L}{\partial \underline{\beta}_i \partial \lambda_i} & \frac{\partial^2 L}{\partial \underline{\beta}_i \partial \underline{\beta}_i} \end{pmatrix} \quad (10)$$

Penganggar kebolehjadian maksimum $\underline{\theta}_i$ diperolehi secara berleleran daripada (10) dan (11). Ujian signifikan bagi subset daripada parameter boleh dilakukan dengan pelbagai cara umpamanya anggaran parameter dibahagikan dengan ralat piawai atau membandingkan log kepada kebolehjadian maksimum yang diperolehi. Penganggar asimptot varians-kovarians matriks bagi $\underline{\theta}_i$ didapati dengan cara biasa, dari matriks informasi, iaitu songsangan kepada $I(\underline{\theta}_i)$. Bagi menguji hipotesis nol keseluruhan, $\underline{\theta}_i = 0$, dipertimbangkan dan mengambil $U(0)' I(0)^{-1} U(0)$ sebagai ujian statistik yang secara asimptot bertaburan χ^2 dengan v darjah kebebasan dimana v bilangan parameter.

Kebarangkalian Kegagalan Hakiki, Kasar dan Kasar separa

Daripada David dan Moeschberger (1978), di dalam selang masa $I_{\alpha} = (a_{\alpha}, a_{\alpha+1})$, $\alpha=0, 1, \dots, r$; $a_0 \geq 0, a_{r+1} \leq \infty$, nisbah

$$\frac{h_i(t)}{h(t)} = k_{i\alpha} \quad (11)$$

suatu pemalar bagi sesuatu i dan α dan merdeka dari t di mana $h(t)$ fungsi bahaya dan $h_i(t)$ fungsi bahaya dengan kehadiran risiko R_i sahaja. Jika $\pi_{i\alpha}$ menandakan kebarangkalian kegagalan dari punca R_i dalam I_{α} ,

$$\begin{aligned} \pi_{i\alpha} &= K_b (Y_i = T, T \in I_{\alpha}) \\ &= \int_{I_{\alpha}} h(t) S(t) dt \\ &= K_{i\alpha} \int_{I_{\alpha}} f(t) dt \\ &= K_{i\alpha} \pi_{\alpha} \end{aligned}$$

dengan $\pi_{\alpha} = K_b(\text{gagal dalam } I_{\alpha})$. Dari itu

$$k_{i\alpha} = \frac{\pi_{i\alpha}}{\pi_{\alpha}} \quad i = 1, \dots, k; \alpha = 0, 1, \dots, r \quad (12)$$

Daripada (11), (berdasarkan masa kegagalan Y_i)

$$\int_{I_\alpha} h_i(t) dt = k_{i\alpha} \int_{I_\alpha} h(t) dt$$

$$\log[P_i(t)]_{I_\alpha} = K_{i\alpha} \log[S(t)]_{I_\alpha}$$

$$\log\left[\frac{P_i(a_{\alpha+1})}{P_i(a_\alpha)}\right] = K_{i\alpha} \log\left[\frac{S(a_{\alpha+1})}{S(a_\alpha)}\right] \tag{13}$$

$$\frac{P_i(a_{\alpha+1})}{P_i(a_\alpha)} = \left[\frac{S(a_{\alpha+1})}{S(a_\alpha)}\right]^{K_{i\alpha}}$$

Andaian berkadaran (11) membolehkan kita membina hubungan yang mudah di antara kebarangkalian kegagalan hakiki, kasar dan kasar separa. Tambahan pula hubungan ini tidak bergantung kepada bentuk fungsi yang tertentu bagi Y_i , $i = 1, \dots, k$. Memadai dengan andaian Y_i merdeka yang fungsinya selanjur dan nisbah fungsi bahaya suatu pemalar yang berkadaran seperti (11).

Di dalam kes data terkumpul ini, ketiga-tiga kebarangkalian adalah berdasarkan selang I_α , $\alpha = 0, 1, \dots, r$ dan bersyaratkan kepada mandiri pada permulaan selang. Berikut dinyatakan rumus ketiga-tiga kebarangkalian mengikut David dan Moeschberger (1978).

Dengan kehadiran satu punca, R_i , kebarangkalian hakiki kegagalan, $q_{i\alpha}$ dalam I_α ialah

$$q_{i\alpha} = 1 - \frac{P_i(a_{\alpha+1})}{P_i(a_\alpha)} \quad i = 1, \dots, k; \alpha = 0, 1, \dots, r \tag{14}$$

Kebarangkalian kegagalan kasar, $Q_{i\alpha}$ dengan kehadiran semua k risiko ialah

$$Q_{i\alpha} = K_b(Y_i = T, T \in A_\alpha \mid T > a_\alpha)$$

$$= \frac{\pi_{i\alpha}}{S(a_\alpha)} \tag{15}$$

Daripada (13) dan (14) diperoleh

$$q_{i\alpha} = 1 - p_\alpha^{K_{i\alpha}} \quad i = 1, \dots, k; \alpha = 0, 1, \dots, r \tag{16}$$

dengan $p_\alpha = S(a_{\alpha+1})/S(a_\alpha)$, kebarangkalian seseorang individu hidup pada a_α dan mandiri hingga I_α .

Kebarangkalian kasar separa kegagalan, $Q_{g\alpha(i)}$ ialah

$$Q_{g\alpha(i)} = \frac{Q_{g\alpha}}{q_\alpha - Q_{i\alpha}} \left[1 - P_\alpha^{(q_\alpha - Q_{i\alpha})/q_\alpha} \right] \quad (17)$$

$$g = 1, \dots, k; \quad g \neq i; \quad \alpha = 0, 1, \dots, r$$

dengan $q_\alpha = 1 - p_\alpha$

Seterusnya biarkan $B_\alpha (> 0)$ sebagai bilangan yang mandiri pada masa a_α . Kebarangkalian kegagalan kasar, $Q_{i\alpha}$, disebabkan R_i dalam I_α boleh dianggarkan sebagai

$$Q_{i\alpha} = \frac{N_{i\alpha}}{B_\alpha} \quad i = 1, \dots, k; \quad \alpha = 0, 1, \dots, r \quad (18)$$

Dengan cara yang sama, p_α dianggarkan oleh

$$\hat{p}_\alpha = \frac{B_{\alpha+1}}{B_\alpha} \quad (19)$$

Oleh itu $\hat{q}_\alpha = 1 - \hat{p}_\alpha$ dan anggaran kepada kebarangkalian hakiki kegagalan ialah

$$\hat{q}_{i\alpha} = 1 - \hat{p}_\alpha^{\hat{Q}_{i\alpha}/\hat{q}_\alpha} \quad (20)$$

Anggaran kepada kebarangkalian kasar separa ialah

$$\hat{Q}_{g\alpha(i)} = \frac{\hat{Q}_{g\alpha}}{\hat{q}_\alpha - \hat{Q}_{i\alpha}} \left[1 - P_\alpha^{(\hat{q}_\alpha - \hat{Q}_{i\alpha})/\hat{q}_\alpha} \right] \quad (21)$$

$$g = 1, \dots, k; \quad g \neq i; \quad \alpha = 0, 1, \dots, r$$

Apabila $k = 2$, sebagaimana dalam penyelidikan kami boleh ditunjukkan $\hat{q}_{g\alpha} = \hat{Q}_{g\alpha(i)}$ Rumus yang dibina kami akan gunakan di dalam bahagian berikut.

DATA KEMORTALAN BAYI

Data diperolehi daripada Jabatan Perangkaan Malaysia (JPM), merupakan pengumpulan rekod terkini kematian bayi bagi tahun 1989 di Semenanjung Malaysia. Bilangan kematian dicatatkan sebagai 5201 kesemuanya termasuk beberapa kovariat. Semua maklumat telah dikodkan oleh JPM. Pentafsiran kod dilakukan berpandu kepada Integrasi Sistem Perangkaan Penting Semenanjung Malaysia, Sabah dan Sarawak, 1984, diperolehi daripada jabatan tersebut. Tiga kovariat yang dipertimbangkan ialah jantina, bangsa dan tempat kediaman. Dua ratus kes dipilih secara rawak. Masa kegagalan telah dikumpulkan ke dalam selang masa dan kami hanya mempertimbangkan kematian bayi yang berumur satu tahun ke bawah. Selang masa yang digunakan adalah seperti berikut:

- < 1 minggu (S0)
- 1 - 4 minggu (S1)
- 4 - 12 minggu (S2)
- 12 - 24 minggu (S3)
- 24 - 36 minggu (S4)
- 36 - 48 minggu (S5)

Kovariat jantina ditandakan 0 jika lelaki dan 1 jika perempuan. Kovariat bangsa ditandakan 0 untuk Melayu, 1 untuk Cina, 2 untuk India dan 3 yang lainnya. Manakala kovariat kediaman terbahagi kepada enam kawasan iaitu 0 untuk kawasan metropolitan, 1 untuk bandar besar, 2 untuk bandar kecil, 3 untuk kawasan kampung, 4 tidak diketahui dan 5 kawasan Sarawak.

Anggaran dan Pentakbiran

Model risiko bersaing untuk menganalisis data kemortalan bayi ini akan berpandukan analisis bagi kes hayat terkumpul. Dua ratus sampel rawak diperolehi daripada 5201 kematian dengan tiga kovariat iaitu jantina, bangsa dan tempat kediaman. Kovariat kediaman bagi setiap data dipiawaikan untuk penyelarasan. Lagipun selain dari mengecilkan ralat pembundaran, penyelarasan ini tidak akan menjejaskan kebolehjadian dan sifat penganggar yang diperolehi. Disebabkan taburan eksponen secocok bagi mewakili data kemortalan dengan ciriannya yang mana fungsi bahaya malar, maka dengan cara automatis ia diambil sebagai taburan pendasar.

Fungsi kebolehjadian (8) bagi data terkumpul digunakan dengan sedikit olahan. Selang dibahagikan kepada $I_a = (a_a, a_{a+1})$, $a = 0, 1, \dots, 5$ dengan $a_0 = 0$ dan $a_6 = 1$ (masa hayat dalam tahun). Setiap selang mempunyai sekurang-kurangnya satu kegagalan. Fungsi yang terpilih melibatkan kovariat ialah $c(z, \beta) = 1 + z \beta$ dengan syarat $z \beta > 0$ bagi setiap nilai kovariat dan transformasi $g_i = \log(-\log(1 - I_i))$ $i = 1, \dots, k$ dipertimbangkan untuk mengatasi pembatasan ke atas parameter (lihat Kalbfleisch dan Prentice 1980: 99). Log kepada kebolehjadian bagi Y_i dan b_i dengan

$$\begin{aligned} \pi_{ij\alpha} &= \pi_i [F_{ij}(a_{\alpha+1}) - F_{ij}(a_\alpha)] \\ &= \frac{\lambda_i}{\lambda} \left[\text{eksp}(-\lambda_i a_\alpha (1 + z_{ij} \beta_i)) - \text{eksp}(-\lambda_i a_{\alpha+1} (1 + z_{ij} \beta_i)) \right] \end{aligned} \quad (22)$$

dimasukkan ke dalam (8) dan diperolehi

$$L = \log L_C(\gamma_i, \beta_i)$$

$$\begin{aligned} &\propto \sum_{i=1}^k \log(1 - e^{-c\gamma_i}) - k \sum_{i=1}^k \log(1 - e^{-c\gamma_i}) \\ &\quad \sum_{i=1}^k \sum_{\alpha=0}^r \sum_{j \in A_\alpha} \log[g_{ija} - g_{ij(\alpha+1)}] \end{aligned}$$

dengan

$$\begin{aligned} g_{ija} &= \text{eksp}\left(-\left(1 - e^{-c\gamma_i}\right) a_\alpha \left(1 + z_{ij} \beta_i\right)\right) \\ g_{ij(\alpha+1)} &= \text{eksp}\left(-\left(1 - e^{-c\gamma_i}\right) a_{\alpha+1} \left(1 + z_{ij} \beta_i\right)\right) \end{aligned}$$

Statistik skor ialah

$$\frac{\partial L}{\partial \gamma_i} = \frac{e^{d_i}}{\lambda_i} - \frac{ke^{d_i}}{\sum_{i=1}^k \lambda_i} + \sum_{i=0}^r \sum_{j \in A_\alpha} \left[\frac{g_{ij(\alpha+1)} b_{\alpha+1} - g_{ij\alpha} b_\alpha}{g_{ij\alpha} - g_{ij(\alpha+1)}} \right]$$

$$\frac{\partial L}{\partial \beta_i} = \sum_{\alpha=0}^r \sum_{j \in A_\alpha} \lambda_i \left[\frac{g_{ij(\alpha+1)} a_{\alpha+1} z_{ij} - g_{ij\alpha} a_\alpha z_{ij}}{g_{ij\alpha} - g_{ij(\alpha+1)}} \right]$$

dengan

$$\lambda_i = 1 - e^{-\gamma_i}$$

$$d_i = -e^{\gamma_i} + \gamma_i$$

$$b_\alpha = a_\alpha \left(1 + z_{ij}' \beta_i \right)$$

Penganggar kebolehdajadian maksimum diperolehi dengan penyelesaian kepada

$$U' = \left(\frac{\partial L}{\partial \gamma_i}, \frac{\partial L}{\partial \beta_i} \right) = 0$$

Kaedah kompromi Marquart terubah suai untuk mengira $(\hat{\gamma}_i, \hat{\beta}_i)$ memerlukan terbitan peringkat kedua kepada L. Matriks informasi tercerap boleh ditulis berdasarkan (10) dengan

$$\frac{\partial^2 \log L}{\partial \gamma_i^2} = \frac{(1 - e^{\gamma_i}) e^{d_i}}{\lambda_i} - \frac{e^{2d_i}}{\lambda_i^2} - \frac{k(1 - e^{\gamma_i}) e^{d_i}}{\sum_{i=1}^k \lambda_i} + \frac{ke^{2d_i}}{\sum_{i=1}^k \lambda_i^2} +$$

$$\sum_{\alpha=0}^k \sum_{j \in A_\alpha} \left[\frac{g_{ij(\alpha+1)} b_{\alpha+1}^2 \xi_{\alpha+1} - g_{ij\alpha} b_\alpha^2 \xi_\alpha}{g_{ij\alpha} - g_{ij(\alpha+1)}} - \frac{(g_{ij(\alpha+1)} b_{\alpha+1}^2 - g_{ij\alpha} b_\alpha^2)^2}{(g_{ij\alpha} - g_{ij(\alpha+1)})^2} \right]$$

dengan $\xi_{(\alpha+1)} = 1 - e^{\gamma_i} - b_{\alpha+1} e^{d_i}$ dan $\xi_\alpha = 1 - e^{\gamma_i} - b_\alpha e^{d_i}$,

$$\frac{\partial^2 \log L}{\partial \gamma_i \partial \beta_i} = \sum_{\alpha=0}^r \sum_{j \in A_\alpha} \frac{g_{ij(\alpha+1)} z_{ij} e^{d_i} a_{\alpha+1} \left(1 - a_{\alpha+1} (1 + z_{ij}' \beta_i) \lambda_i \right)}{g_{ij\alpha} - g_{ij(\alpha+1)}} -$$

$$\frac{g_{ij\alpha} z_{ij} e^{d_i} a_{\alpha} \left(1 - a_{\alpha} \left(1 + z_{ij} \beta'_{-i} \right) \lambda_i \right)}{g_{ij\alpha} - g_{ij(\alpha+1)}} \times$$

$$\frac{\lambda_i^{d_i} \left(1 + z_{ij} \beta'_{-i} \right) \left(g_{ij(\alpha+1)} a_{\alpha+1} z_{ij} - g_{ij\alpha} a_{\alpha} z_{ij} \right)}{\left(g_{ij\alpha} - g_{ij(\alpha+1)} \right)^2}$$

$$\frac{\left(g_{ij(\alpha+1)} a_{\alpha+1} - g_{ij\alpha} a_{\alpha} \right)}{\left(g_{ij\alpha} - g_{ij(\alpha+1)} \right)^2}$$

$$\frac{\partial^2 \log L}{\partial \beta_{-i} \partial \beta'_{-i}} = \sum_{\alpha=0}^r \sum_{j \in A_{\alpha}} \frac{\left(g_{ij\alpha} \left(\lambda_i a_{\alpha} \right)^2 z_{ij} z_{ij} - g_{ij(\alpha+1)} \left(\lambda_i a_{\alpha+1} \right)^2 z_{ij} z_{ij} \right)}{g_{ij\alpha} - g_{ij(\alpha+1)}}$$

$$- \frac{\left(A_{ij} A'_{ij} \right)}{\left(g_{ij\alpha} - g_{ij(\alpha+1)} \right)^2}$$

dengan $A_{ij} = g_{ij(\alpha+1)} \lambda_i a_{\alpha+1} z_{ij} - g_{ij\alpha} \lambda_i a_{\alpha} z_{ij}$

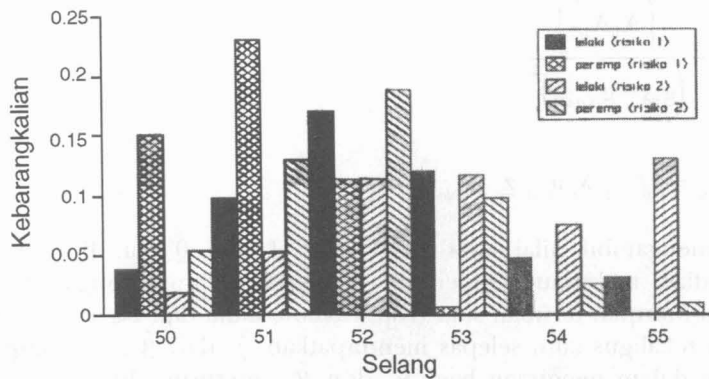
Dengan mengambil nilai awal $\gamma_1 = 1$, $\gamma_2 = 1$, $\beta_1 = 0$ dan $\beta_2 = 0$ anggaran kebolehdajian maksimum diperoleh secara berleleran dengan pendekatan kompromi Marquat terubah suai (rujuk Noor Akma dan Isa 1992). Anggaran dilakukan sekaligus iaitu selepas mendapatkan $\hat{\gamma}_1$ dan $\hat{\beta}_1$. $\hat{\gamma}_1$ disimpan untuk digunakan dalam pengiraan bagi $\hat{\gamma}_2$ dan $\hat{\beta}_2$ seterusnya hingga penumpuan. penumpuan ialah beza mutlak di antara log kebolehdajian.

Jadual 1 memberikan keputusan sepenuhnya bai risiko pertama dan kedua (di dalam kurungan) bersama ralat piawai dan nilai **P**. Kiraan bagi nilai **P** ialah membahagi nilai parameter dengan ralat piawainya dengan berpandukan fungsi Student-t (Abramowitz dan Stegun 1970). Penumpuan tercapai pada leleran ke 34 dengan nilai log kebolehdajian -436.9846. Bagi risiko pertama, kovariat jantina dan kovariat bangsa didapati bererti secara statistik sementara bagi risiko kedua hanya kovariat jantina didapati bererti.

Rajah 1 merupakan graf kebarangkalian kegagalan dengan selang masa. Kebarangkalian kegagalan ini berpandukan (22) bagi sesuatu selang dengan kovariat jantina sementara diambil min bagi kovariat bangsa dan min bagi kediaman. Graf di dalam bentuk histogram. Secara kasarnya setiap selang untuk kedua-dua risiko, kebarangkalian kegagalan bagi bayi lelaki dan perempuan terdapat perbezaan. Oleh itu dari nilai **P** dan graf didapati jantina mempunyai kesan ke atas kegagalan tanpa mengira puncanya. Kebarangkalian

JADUAL 1
 Anggaran parameter, ralat piawai dan nilai P
 bagi data kemortalan bayi

Parameter	Anggaran	Ralat piawai	P
$\hat{\gamma}_1$	1.70237	0.10421	
$\hat{\gamma}_2$	(1.72249)	(0.18335)	
$\hat{\beta}_{11}$	13.59044	2.23080	0.0046
$\hat{\beta}_{21}$	(3.57822)	(0.96810)	(0.017)
$\hat{\beta}_{12}$	6.41301	1.46097	0.011
$\hat{\beta}_{22}$	(1.12572)	(0.72302)	(0.109)
$\hat{\beta}_{13}$	0.10549	0.10493	0.19
$\hat{\beta}_{23}$	(0.18262)	(0.14027)	(0.14)

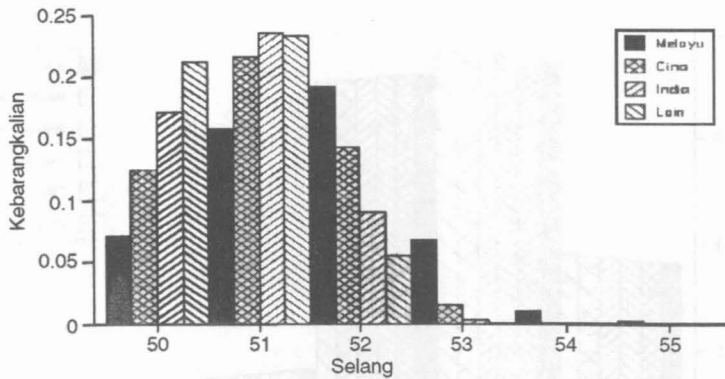


Rajah 1. Histogram kebarangkalian kegagalan bagi kovariat jantina data kemortalan bayi

kematian yang tinggi bagi jantina berlaku dalam selang S1 iaitu bayi perempuan dengan umur di antara 1 hingga 4 minggu.

Rajah 2 memperlihatkan graf kebarangkalian kegagalan dengan selang bagi kovariat bangsa dalam risiko pertama sementara kovariat yang lain dianggap malar. Berpandukan graf, secara kasar boleh dikatakan terdapat perbezaan kebarangkalian kegagalan bagi kovariat bangsa walaupun di dalam selang kedua bezanya tidak begitu ketara. Nilai P yang telah dikira boleh diteguhkan lagi dengan histogram ini. Jadinya bangsa iaitu Melayu, Cina, India dan lain-lain mempunyai kesan terhadap kemortalan bayi. Dalam selang S0 (< 1 minggu), bangsa yang digulungkan di dalam kategori lain-lain mempunyai kebarangkalian kegagalan yang paling tinggi. Dalam selang S2 (4 - 12 minggu) dan S4 (24 - 36 minggu) bangsa Melayu mempunyai kebarangkalian kegagalan yang tinggi

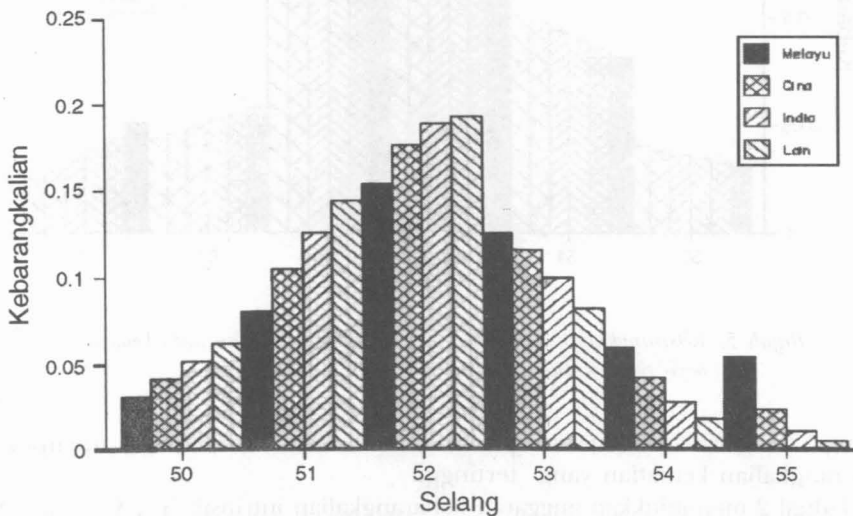
Analisis Kovariat bagi Data Kemortalan Bayi



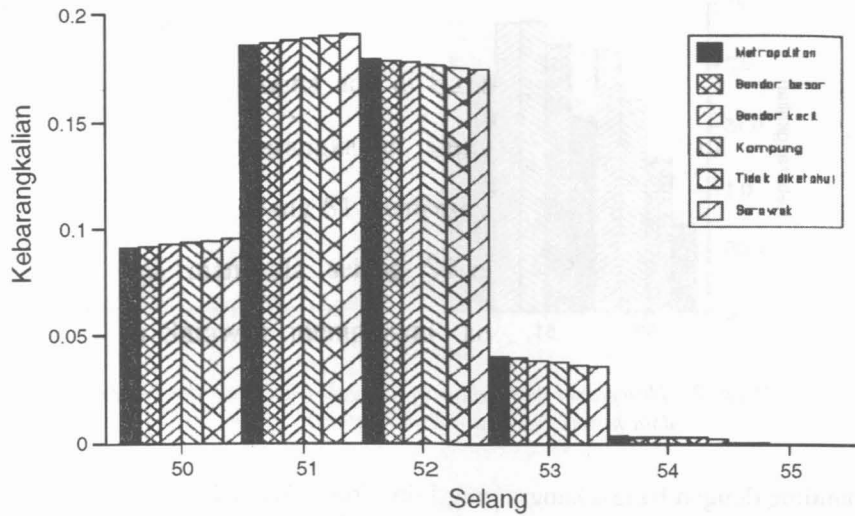
Rajah 2. Histogram kebarangkalian kegagalan bagi kovariat bangsa data kemortalan bayi - risiko pertama

berbanding dengan bangsa-bangsa yang lain. Rajah 3 adalah graf kebarangkalian bagi kovariat bangsa untuk punca yang tidak diketahui. Kebarangkalian kegagalan bagi setiap bangsa di dalam selang yang sama tidak terdapat banyak perbezaan. Kebarangkalian kematian yang tinggi bagi kesemua bangsa berlaku di dalam selang S2 iaitu umur bayi di antara 4 hingga 12 minggu.

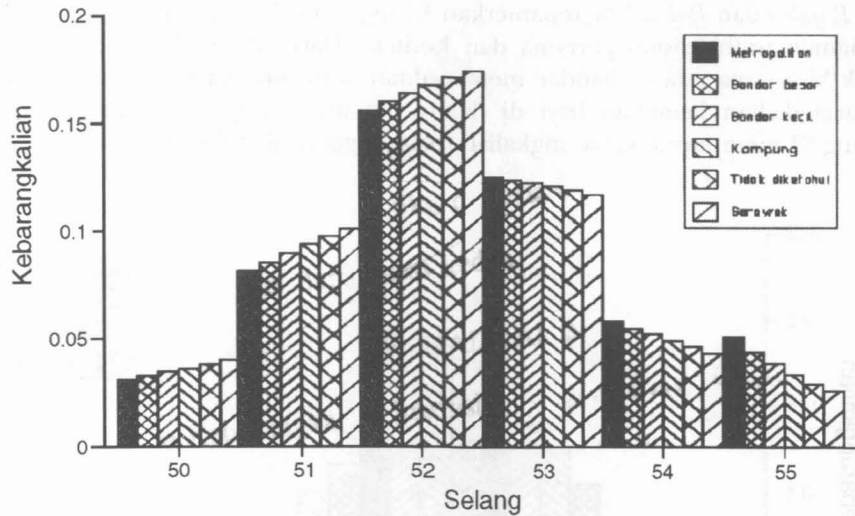
Rajah 4 dan Rajah 5 mempamerkan histogram kebarangkalian bagi kovariat kediaman untuk risiko pertama dan kedua. Daripada graf boleh dikatakan tidak kira sama ada di bandar metropolitan atau diperkampungan terpencil, kebarangkalian kematian bayi di dalam sesuatu selang lebih kurang sama. Selang S1 mempunyai kebarangkalian yang tinggi bagi risiko pertama manakala



Rajah 3. Histogram kebarangkalian kegagalan bagi kovariat bangsa data kemortalan bayi - risiko kedua



Rajah 4. Histogram kebarangkalian kegagalan bagi kovariat kediaman data kemortalan bayi - risiko pertama



Rajah 5. Kebarangkalian kegagalan bagi kovariat kediaman data kemortalan bayi - risiko kedua

selang S2 bagi risiko kedua dengan kawasan kediaman di Sarawak mempunyai kebarangkalian kematian yang tertinggi.

Jadual 2 menunjukkan anggaran kebarangkalian intrinsik \hat{q}_{ix} , \hat{Q}_{ix} dan $\hat{Q}_{(g)}$ bagi kedua-dua risiko (risiko kedua di dalam kurungan). Di sini ketiga-tiga kebarangkalian ini dikira tanpa melibatkan kovariat. Disebabkan data dianalisis

JADUAL 2
 Anggaran kebarangkalian intrinsik, $\hat{Q}_{i\alpha}$, $\hat{q}_{i\alpha}$, $\hat{Q}_{1\alpha(2)}$, $\hat{Q}_{2\alpha(1)}$
 bagi data kemortalan bayi

α	$\hat{Q}_{i\alpha}$	$\hat{q}_{i\alpha}$	$\hat{Q}_{1\alpha(2)}$	$\hat{Q}_{2\alpha(1)}$
0	0.46 (0.045)	0.47299 (0.06073)	0.47299	0.06073
1	0.253 (0.071)	0.26289 (0.08186)	0.26289	0.08186
2	0.194 (0.328)	0.24003 (0.37154)	0.24003	0.37154
3	0.156 (0.281)	0.18575 (0.30918)	0.18575	0.30918
4	0.222 (0.556)	0.34930 (0.65850)	0.34930	0.65850
5	0.50 (0.50)	1 (1)	1	1

berdasarkan kehadiran dua risiko sahaja maka $\hat{Q}_{i\alpha} = \hat{Q}_{i\alpha(g)}$, $i \neq g$. Bagi risiko pertama jika digrafkan kebarangkalian kasar kegagalan dengan masa, didapati bentuknya menyerupai graf fungsi bahaya dengan- bagi kemortalan manusia iaitu berbentuk "bath tub". Ini tidak berlaku kepada kebarangkalian kasar bagi risiko kedua. Kebarangkalian kasar separa berpunca dari risiko pertama setelah risiko kedua dihapuskan memberikan graf "bath tub" juga dengan kebarangkalian kegagalan pada selang akhir sebagai satu. Tidak banyak boleh diperkatakan jika digrafkan kebarangkalian kasar separa kegagalan berpunca dari risiko kedua.

KESIMPULAN

Data kemortalan bayi dianalisis berpandukan model risiko bersaing ($k = 2$) dengan masa kegagalan dikumpulkan ke dalam beberapa selang masa. Taburan eksponen telah digunakan sebagai taburan pendasar berdasarkan kesesuaiannya mewakili data hayat serta keunikannya mempunyai fungsi bahaya yang malar. Penjelmaan ke atas parameter eksponen adalah perlu untuk mendapatkan suatu model yang stabil. Penganggaran parameter pada keseluruhannya dapat dilaksanakan melalui kaedah kebolehdjian maksimum dengan pendekatan leleran kompromi Marquat terubah suai.

Melalui pengiraan nilai **P** beberapa kovariat di dalam model ini didapati signifikan terhadap masa kegagalan. Dari risiko pertama (punca diketahui) kovariat jantina dan bangsa didapati signifikan iaitu di dalam selang S_1 (1 - 4 minggu) bayi perempuan dan bayi dari bangsa lain-lain di dalam selang S_0 (< 1 minggu) mempunyai kemortalan yang tinggi. Sementara itu kovariat yang signifikan dari risiko kedua (punca tidak diketahui) ialah jantina. Boleh dikatakan

kovariat jantina mempunyai kaitan yang signifikan ke atas kematian seseorang bayi. Secara kasar kebarangkalian kegagalan, berpandukan $\pi_{ij\alpha}$ dapat mengukuhkan lagi kesignifikan kovariat di dalam model bagi analisis data terkumpul. Adalah wajar kovariat seperti ini diselidiki dengan lebih mendalam.

Tidak banyak boleh diperkatakan berkenaan kebarangkalian kegagalan risiko bersaing bagi data terkumpul. Walau bagaimanapun jika digrafkan $\hat{Q}_{i\alpha} = \hat{Q}_{i\alpha(g)}$ dari risiko pertama bagi keseluruhan selang, ia memperlihatkan suatu bentuk graf yang mirip graf bagi kemortalan makhluk. Mungkin kebarangkalian kegagalan ini perlu diteliti dengan lebih mendalam lagi dengan melibatkan kovariatnya sekali.

Untuk kajian selanjutnya beberapa kovariat lain seperti berat badan bayi ketika dilahirkan dan peringkat imunisasi perlu diambil kira agar keputusan yang lebih tepat dan mantap dapat dibuat.

RUJUKAN

- ABRAMOWITZ, M dan I.A. STEGUN. 1970. *Handbook of Mathematical Functions*. New York: Dover.
- BURRIDGE, J. 1981. A note on maximum likelihood estimation for regression models using grouped data. *Journal of Royal Statistical Society* **43(1)**: 41-45.
- COX, D.R. 1972. Regression models and life tables (with discussion). *Journal of Royal Statistical Society* **B34**: 187-220.
- DAVID, H.A. dan M.L. MOESCHBERGER. 1978. *The Theory of Competing Risks*. London: Griffin.
- GAIL, M. 1975. A review and critique of some models used in competing risks analysis. *Biometrics* **31**: 209-211.
- HOLFORD, T.R. 1976. Life tables with concomitant information. *Biometrics* **32**: 587-597.
- KALBFLEISCH, J.D. dan R.L. PRENTICE. 1980. *The Statistical Analysis of Failure Time Data*. New York: Wiley.
- KIMBALL, A.W. 1969. Models for the estimation of competing risks from censored data. *Biometrics* **25**: 329-337.
- LINDLEY, D.V. 1979. Analysis of life tables with grouping and withdrawals. *Biometrics* **35**: 605-612.
- NOOR AKMA IBRAHIM dan ISA DAUD. 1992. Diagnostik pengaruh bagi model risiko bersaing dengan tapisan sebagai satu risiko. *Pertanika* **15**: 159-169.
- PETTITT, A.N. 1985. Reweighted least squares estimation with grouped data: An application of EM algorithm. *Journal of Royal Statistical Society* **B47**: 253-260.
- PIERCE, D.A., W.H. STEWART dan K.J. KOPECKY. 1979. Distribution-free regression analysis of grouped survival data. *Biometrics* **35**: 785-793.
- THOMPSON, W.A. Jr. 1977. On the Treatment of grouped observations in life studies." *Biometrics* **33**: 463-470.

- TURNBULL, B.W. 1976. The empirical distribution function with arbitrarily grouped, censored and truncated data. *Biometrics* **3**: 290-295.
- TURNBULL, B.W. dan L. WEISS. 1978. A likelihood ratio statistics for testing goodness of fit with randomly censored data. *Biometrics* **34**: 367-375.
- TURRERO, A. 1989. On the relative efficiency of grouped and censored survival data. *Biometrika* **76**: 125-131.