

Rekabentuk Ukuran Ulangan dan Proses Bercabang Dalam Biologi

Husna Binti Hasan

*Pusat Pengajian Sains Matematik,
Universiti Sains Malaysia, 11800 Minden, Penang.
E-mel: husna@cs.usm.my*

ABSTRAK

Tercetusnya konsep Biologi Bermatematik boleh melahirkan usaha kerjasama antara penyelidik bidang sains hayat, matematik dan juga penyelidik dari bidang sains lain yang berkaitan. Data yang terkumpul dari hasil penyelidikan kajian kes sains hayat perlu dianalisa dan ditafsir dengan betul supaya kesimpulannya adalah benar bagi semua kes secaman. Kewujudan pelbagai pakej statistik yang boleh mengeluarkan keputusan analisis dengan segera telah membuka peluang kepada pengguna yang mempunyai pengetahuan terhad dalam bidang statistik khasnya analisis data dan berkemungkinan menghasilkan kesimpulan yang kurang bermakna. Kertas ini membincangkan bagaimana penyelidik sains hayat dan statistik boleh bekerjasama dalam menjayakan analisis data yang melibatkan rekabentuk ukuran ulangan supaya kesimpulan yang terhasil adalah lebih bermakna, munasabah dan realistik. Masalah yang dihadapi dalam proses bercabang yang merupakan satu bidang statistik tulen juga dibangkitkan.

PENGENALAN

Matematik adalah sesuatu yang menggerunkan bagi sesetengah pelajar sains hayat (atau Biologi). Apatah lagi statistik yang merlukan latar belakang matematik untuk memahaminya. Kedua-dua subjek ini adalah antara jurang dalam pendidikan matematik bagi pelajar yang berorientasikan biologi. Walaupun terdapat kursus asas yang boleh diikuti, namun kursus-kursus statistik yang ditawarkan oleh Pusat Pengajian Sains Matematik (atau Statistik) kerap kali tidak memenuhi kehendak pelajar atau bakal penyelidik biologi kerana dikatakan terlalu *theoretical*. Pelajar pula perlu mengaplikasikan kaedah-kaedah statistik ke atas masalah atau kajian kes yang mereka minati.

Mengikuti takrif yang diberikan dalam [11],

‘Mathematical biology or biomathematics is an interdisciplinary field of academic study which aims at modeling natural, biological processes using mathematical techniques and tools. It has both practical and theoretical applications in biological research.’

Kaedah statistik merupakan salah satu teknik atau alat matematik yang digunakan dalam penyelidikan sains. Tidak dapat disangkalkan bahawa sesuatu teori menjadi lebih jelas apabila diaplikasikan kepada masalah tertentu. Bayangkan dari suatu teorem matematik yang terdiri dari hanya persamaan atau pernyataan matematik dapat dihuraikan dalam bentuk masalah harian atau kajian kes semasa.

Kebelakangan ini, aplikasi matematik dan statistik dalam bidang biologi timbul dengan kadar yang memberangsangkan. Antara sebabnya ialah lahirnya set-set data bermaklumat misalnya data *genomics* yang susah untuk difahami tanpa alat analitik. Perkembangan dan penemuan terbaru alat matematik dapat membantu penyelidik memahami konsep-konsep biologi yang kompleks. Teknologi perkomputeran yang semakin canggih dan cekap sekarang ini mampu menjalankan pengiraan dan simulasi yang sebelumnya agak mustahil. Pada masa yang sama, bidang-bidang baru dalam teori kebarangkalian dan statistik juga berkembang untuk memenuhi keperluan bidang biologi.

Kaedah statistik dan peranannya

Seperti yang dinyatakan, statistik adalah suatu bidang yang berlandaskan matematik secara dasarnya. Ianya adalah sains matematik yang mengendalikan pengumpulan, analisis, tafsiran dan persembahan data. Objektifnya adalah untuk membuat inferen mengenai suatu populasi yang diminati berdasarkan maklumat yang diperolehi dari sampel-sampel ukuran dari populasi tersebut. Kaedah-kaedah statistik yang telah dibangun boleh digunakan untuk meringkas, menghurai koleksi data, mengenal pasti corak data agar dapat dimodelkan. Ianya kemudian diguna untuk membuat inferen mengenai proses atau populasi yang dikaji. Dari suatu ujikaji atau kajian, data dikumpul dan dianalisa supaya dapat menjawab soalan penyelidik dan memberikan suatu keputusan statistik yang bererti. Daripada keputusan ini juga, suatu teori baru boleh diuji dan dibangun.

Sebelum ujikaji dijalankan, rangka pengumpulan data dirancang supaya data yang sesuai dapat dikumpulkan. Ini dikenali sebagai reka bentuk ujikaji. Data yang dikumpul kemudiannya dianalisa dengan menggunakan kaedah

statistik yang mengandungi faktor atau aras yang diambil kira atau dipersetujui semasa proses merangka ujikaji untuk menghasilkan kesimpulan yang kukuh dan memenuhi objektif kajian . Pendekatan secara statistik pada reka bentuk ujikaji adalah perlu jika ingin mendapatkan kesimpulan yang bermakna daripada data (Montgomery, 1983).

Pemilihan faktor, aras, pemboleh ubah sambutan dan reka bentuk ujikaji merupakan langkah-langkah penting dalam proses ujikaji. Dalam analisis rekabentuk ujikaji, hipotesis yang dijalankan ialah menguji kesamaan beberapa min. Tatacara yang digunakan ialah Analisis Varians (ANOVA) yang mengambil kira ralat ujikaji atau ubahan dalam data. Segala nilai ubahan yang diambil kira diringkaskan dalam Jadual ANOVA.

Reka bentuk ukuran ulangan

Kebanyakan penyelidikan sains gunaan memerlukan lebih dari satu cerapan daripada satu unit ujikaji (atau subjek) bagi penjimatan kos. Dengan itu, adalah penting untuk mendapat seberapa banyak maklumat yang boleh daripada satu unit ujikaji ini.

Salah satu rekabentuk yang boleh digunakan dalam kajian biologi ke atas data begini ialah reka bentuk ukuran ulangan. Sama seperti mana-mana ANOVA, hipotesis dalam reka bentuk ukuran ulangan ialah menguji kesamaan min. Dalam rekabentuk ukuran ulangan, data tercerap diperolehi daripada t masa yang berlainan tetapi merujuk kepada unit ujikaji (atau subjek) yang sama. Hasilnya ialah data bersandar memandangkan cerapan terkumpul bagi semua sampel atau aras rawatan adalah kepunyaan subjek yang sama.

Antara sebab lain penyelidik memerlukan lebih dari satu cerapan dari unit ujikaji adalah kerana dipercayai bahawa ubahan di kalangan atau di antara unit ujikaji adalah lebih besar jika dibandingkan dengan ubahan dalam suatu unit ujikaji (Ott, 2001). Dalam kes ini, adalah lebih baik menjadikan unit ujikaji sebagai blok dan setiap unit ujikaji diberi setiap rawatan. Kajian dengan reka bentuk ukuran ulangan mempunyai kelebihan untuk menambah baik kejituan anggaran (Agresti, 1999). Unit ujikaji atau subjek yang sama dalam setiap kumpulan boleh membantu penyelidik menghindar ralat yang berlebihan.

Jadual di bawah menunjukkan bentuk set data yang diperolehi dari suatu reka bentuk ujikaji dua faktor dengan ukuran ulangan pada satu faktor.

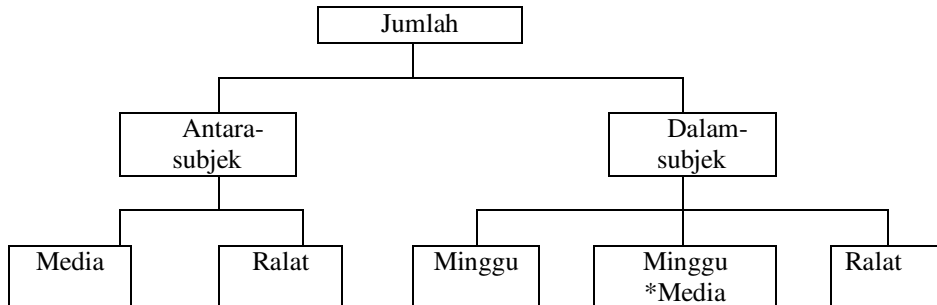
JADUAL A: Data reka bentuk ukuran ulangan

| | Pokok | Minggu | | | | | | | |
|----------------|----------|-----------|-----|---|---|---|---|--|-----------|
| | | 1 | 2 | 3 | 4 | 5 | 6 | | |
| | | ... | b | | | | | | |
| Media 1 | 1 | y_{111} | | | | | | | |
| | \vdots | \vdots | | | | | | | |
| | n | y_{11n} | | | | | | | |
| \vdots | | | | | | | | | |
| Media a | 1 | | | | | | | | \vdots |
| | 2 | | | | | | | | y_{abn} |
| | \vdots | | | | | | | | |
| | n | | | | | | | | |

Dalam ujikaji di atas, subjeknya ialah pokok dan data tercerap ialah peratusan akar pokok yang keluar. Tujuan ujikaji adalah untuk mengkaji samada semua aras media memberi kesan yang sama ke atas perkembangan peratusan akar yang keluar. Faktor media terdiri dari sampel tak bersandar. Dua kesan tetap bagi ujikaji ini ialah media (1, 2, 3, ..., a) yang diwakili oleh a sampel tak bersandar dan minggu yang terdiri dari b kali cerapan. Minggu merupakan faktor dalam-subjek memandangkan perbandingan kategorinya menggunakan ukuran ulangan dalam sampel. Media adalah faktor antara-subjek kerana perbandingan kategorinya menggunakan pokok berbeza dari setiap kategori.

Walaupun kedua-dua faktor (media dan minggu) adalah kesan utama, analisisnya berbeza dari ANOVA dua-hala yang biasa digunakan. Ini adalah kerana ukuran ulangan ke atas faktor dalam-pokok menghasilkan kesan ketiga, iaitu kesan rawak bagi pokok. Setiap pokok diperhatikan pada setiap kategori minggu, tetapi hanya pada 1 kategori media. Dengan itu, pokok dikatakan tersarang dalam faktor media. Maka pembahagian ubahan bagi ANOVA dua hala ini adalah seperti dalam rajah di bawah:

JADUAL B: Rajah pembahagian ubahan



Model yang digunakan bagi ujikaji dua-faktor ini ialah

$$y_{ijk} = \mu + \alpha_i + \pi_{k(i)} + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk}$$

$$i = 1, \dots, a; \quad j = 1, \dots, b; \quad k = 1, \dots, n$$

yang mana μ adalah min keseluruhan, α_i, β_j dan $\alpha\beta_{ij}$ adalah kesan tetap bagi media, minggu dan saling tindak, masing-masing. Kesan rawak yang disebabkan oleh pokok ke- k dalam media ke- i ditandai $\pi_{k(i)}$, adalah tak bersandar dan tertabur $N(0, \sigma_\pi^2)$ dan tak bersandar dengan ε_{ijk} yang juga tak bersandar dan tertabur $N(0, \sigma_\varepsilon^2)$. Jadual ANOVA yang terhasil adalah seperti dalam jadual berikut.

JADUAL C: Jadual ANOVA ujikaji reka bentuk ukuran ulangan

| Punca | SS | dk | MSE |
|-------------------|--------|-------------------|-----------------------|
| Antara-subjek | SSM | $a-1$ | $SSM/a-1$ |
| Media | SSP(M) | $a(n-1)$ | $SSP(M)/a(n-1)$ |
| Pokok-dalam-Media | | | |
| Dalam-subjek | SSW | $b-1$ | $SSW/b-1$ |
| Minggu | SSMW | $(a-1)(b-1)$ | $SSMW/(a-1)(b-1)$ |
| Media*Minggu | SSE | $(a-1)(b-1)(n-1)$ | $SSE/(a-1)(b-1)(n-1)$ |
| Ralat | | | |
| Jumlah | SST | $abn-1$ | |

Untuk menguji kesan saling tindak dan dua kesan utama iaitu

$H_0 : \theta_{MW} = 0, H_0 : \theta_W = 0$ and $H_0 : \theta_M = 0$, statistik ujian F yang digunakan ialah $F = \frac{MSMW}{MSE}$, $F = \frac{MSW}{MSE}$ dan $F = \frac{MSM}{MSP(M)}$, masing-masing.

Apa yang diperhatikan, jika penyelidik tidak mengambil kira kesan rawak bagi pokok, penyelidik akan tersilap mengambil $F = \frac{MSM}{MSE}$ untuk menjalankan pengujian hipotesis $H_0 : \theta_M = 0$. Apatah lagi dengan keupayaan sofwer statistik masa kini yang dengan mudah mengeluarkan jadual ANOVA. Walau bagaimana pun, penyelidik perlu berhati-hati kerana sesetengah sofwer menggunakan MSE dalam semua pengiraan statistik ujian F . Dalam kes ini, jika penyelidik tidak memahami statistik ujian sepenuhnya, sudah tentulah penyelidik ini akan berpandu kepada nilai F tidak betul yang dikeluarkan oleh sofwer tersebut. Dengan itu, kesimpulan yang dihasilkan adalah kurang tepat. Di sinilah perlunya pengetahuan statistik supaya penyelidik dapat menggunakan statistik ujian yang betul dan seterusnya merumus kesimpulan yang tepat.

Terdapat juga sesetengah penyelidik yang mengabaikan terus kesan rawak yang wujud sedangkan ianya benar-benar wujud. Apabila diabaikan, maka MSE akan menjadi besar. Statistik ujian untuk menguji kesan media akan terus mengambil kira MSE . Untuk memahami keperluan dan penggunaan statistik ujian yang betul, adalah suatu perkara yang kadang kala merumitkan bagi penyelidik biologi, ditambah pula kurangnya pengetahuan dalam bidang matematik khasnya statistik. Untuk mengelak daripada melakukan kesilapan, elok kiranya penyelidik biologi bekerjasama dengan penyelidik statistik dalam menjalankan kajian supaya analisis yang betul dapat dijalankan dan matlamat penyelidikan tercapai. Kes seterusnya membincangkan keperluan penyelidik biologi dalam penyelidikan matematik pula.

Proses Bercabang

Suatu populasi terdiri dari subjek yang mampu menghasilkan keturunan dari jenis yang sama. Katakan setiap subjek, pada akhir hayatnya telah melahirkan se bilangan keturunan. Dalam proses evolusi dan kelahiran, diandaikan bahawa semua subjek bertindak secara tak bersandar antara satu sama lain dan pada akhir tempoh hayat mereka menjana suatu bilangan rawak individu (atau subjek) baru berdasarkan beberapa hukum

kebarangkalian. Proses $X(t)$, iaitu saiz populasi pada generasi ke- t (t diskret di sini) dikenali proses bercabang yang merupakan kelas khas bagi rantai Markov dan mempunyai aplikasi dalam bidang biologi. Juga dikenali sebagai proses stokastik bercabang kerana ia merupakan suatu model matematik yang menerangkan pembangunan proses dalam bentuk bercabang misalnya pengembangan atau pembiakan suatu populasi.

Kajian proses bercabang bermula pada tahun 1873 oleh Sir Francis Galton yang mengkaji kepupusan suatu kaum bangsawan dan masalah ini berjaya dijawab dalam lelaran fungsi penjana oleh Rev. Henry Watson. Dalam tahun 1972, suatu kertas kerja yang telah ditulis oleh I. J. Bienayme dalam tahun 1845 yang mengkaji masalah yang serupa telah dijumpai. Sejak itu, proses bercabang masa diskret ini dikenali sebagai proses Bienayme-Galton-Watson (BGW).

Katakan pada masa kelahiran generasi ke- t , terdapat suatu imigrasi $Y(t)$ subjek ke dalam populasi tersebut. Maka, saiz populasi baru ini, $Z(t)$ dikenali sebagai proses bercabang dengan imigrasi. Dalam proses ini, setiap $Y(t)$ subjek akan memulakan suatu proses bercabang. Taburan bilangan immigrant (atau pendatang) boleh berbeza mengikut masa dan pembiakan populasi sedia ada. Faktor imigrasi agak penting dalam aplikasi proses biologi dan ekologi. Andaian ketakbersandaran proses imigrasi dan pembiakan juga adalah munasabah dalam kebanyakan bidang sains.

Proses bercabang berbeza antara satu sama lain bergantung kepada suatu parameter A iaitu min bilangan keturunan oleh satu subjek, yang dikenali sebagai keterkritikan proses. Terdapat tiga jenis proses bercabang. Jika $A < 1$, $A = 1$ dan $A > 1$, proses ini dikenali proses subkritikal, kritikal dan superkritikal, masing-masing. Berbagai kemungkinan boleh berlaku pada suatu proses dalam suatu tempoh yang agak panjang. Salah satu kemungkinan ialah kepupusan proses. Dalam model kebarangkalian, jika $A \leq 1$ dan $A > 1$, kebarangkalian kepupusan ialah 1 dan < 1 , masing-masing. Walau bagaimanapun, tidak mustahil jika sesuatu proses akan wujud secara berterusan. Ini dikenali sebagai ketakpupusan proses. Kajian ke atas proses bercabang mengambil kira faktor kepupusan dan ketakpupusan ini.

Model proses bercabang berbeza antara satu sama lain dengan andaian tambahan ke atas proses pembiakan subjek baru dan juga boleh mengambil kira taburan tercantum bagi hayat hidup dan saiz keturunan. Objek utama dalam penyelidikan proses bercabang ialah proses rawak $X(t)$ (atau $Z(t)$ dalam proses dengan imigrasi), dan juga ciri lain yang menerangkan struktur pokok bagi suatu proses. Bilangan subjek dalam suatu generasi yang

mempunyai keturunan set tak kosong pada masa t dikenali sebagai proses bercabang terturun (Fleishman & Siegmund-Scgultz (1987), Yakimov (1980), Yeoh (1998)) dan bilangan pasangan yang mempunyai bilangan keturunan yang sama (Rahimov, 1995) adalah dua contoh pemboleh ubah rawak yang dikaji oleh penyelidik.

Masalah lebih pertama

Antara cirian yang dikaji ialah masalah lebih dalam proses bercabang. Kertas kerja ini membangkitkan masalah lebih pertama pada suatu aras yang diberi bagi sebilangan proses bercabang tak bersandar. Ditakrifkan

$$v(t, \theta) = \min\{k : X_k(t) > \theta(t)\}$$

bagi suatu fungsi aras $\theta(t)$ diberi, samada tetap atau menokok.

Proses $X_i(t); i = 1, 2, \dots$ boleh ditafsirkan sebagai saiz populasi yang wujud dalam bahagian yang berbeza dalam suatu kawasan yang luas. Maka, proses $v(t, \theta)$ ialah indeks proses pertama dalam famili tersebut yang mana saiznya melebihi aras $\theta(t)$. Beberapa kes boleh diambil kira untuk mengkaji masalah indeks pertama ini. Antaranya ialah apabila proses mempunyai taburan keturunan secaman, berbeza dan juga kewujudan famili imigrasi. Beberapa taburan had bagi $v(t, \theta)$ telah dibuktikan (Rahimov, 1998). Salah satu teorem yang telah dibuktikan bagi proses $X_i(t)$ dengan taburan secaman dinyatakan disini:

Teorem: Jika $A < 1$, dan $\theta(t) = \theta \in N = \{1, 2, \dots\}$, maka

$v(t, \theta) \in A[Q(t)y, E_1(x)]$, yang mana

$$E_1(x) = \begin{cases} 0 & ; x < 0 \\ 1 - e^{-q^*(\theta)x} & ; x \geq 0; \quad q^*(\theta) = \sum_{k=\theta+1}^{\infty} P_k^* \end{cases}$$

yang menggunakan Teorem Yaglom:

$$\lim_{t \rightarrow \infty} P[X(t) = k \mid X(t) > 0] = P_k^*; k = 1, 2, \dots$$

dan $Q(t)$ adalah kebarangkalian ketakpupusan.

Teorem tersebut telah dibuktikan menggunakan teknik jelmaan Laplace. Penghasilan teorem ini adalah satu kejayaan bagi bagi seorang penyelidik matematik atau statistik tulen kerana suatu teorem baru telah dibangunkan. Seperti yang telah dinyatakan di awal kertas kerja ini, suatu teorem akan lebih jelas dan bermakna jika dapat diaplikasikan kepada masalah nyata atau sebenar agar dapat diketahui sejauh mana ketepatan teorem yang telah dibuktikan. Penyelidik matematik kebiasaannya menggunakan contoh matematik untuk mengukuhkan teorem yang telah dibuktikan. Seperti dalam kes ini jika taburan keturunan adalah taburan geometri dengan $p > 1/2$ dan $\theta \in N$, maka $v(t)(1-A)A^t$ menumpu dalam taburan kepada suatu pemboleh ubah rawak yang mempunyai fungsi ketumpatan kebarangkalian eksponen $A^\theta e^{-A^\theta x}$ yang mana $A = q/p$, $q = 1-p$ dan $x > 0$. Bagi kes ini, $E[v(t)] \sim \left(\frac{p}{2p-1}\right)\left(\frac{p}{1-p}\right)^{\theta+t}$, $t \rightarrow \infty$.

Memandangkan masalah proses bercabang ini boleh diaplikasikan dalam proses biologi, adalah baik jika data nyata/sebenar dapat diperolehi dan dianalisa supaya ketepatan teorem dapat dilihat. Untuk mendapat data sebenar ini, sudah tentulah penyelidik statistik memerlukan bantuan kepakaran penyelidik biologi. Disebabkan kerumitan mendapat data yang boleh memenuhi kehendak masalah, penyelidik matematik biasanya kebiasaannya mengguna data simulasi sebagai ganti kepada data sebenar.

KESIMPULAN

Daripada kedua-dua kes yang telah dinyatakan, jelas kelihatan bahawa penyelidik statistik dan biologi saling merlukan antara satu sama lain supaya hasil penyelidikan akan lebih bererti dan bermakna. Dalam kes reka bentuk ukuran ulangan, jika penyelidik biologi tidak mempunyai pengetahuan yang cukup, berkemungkinan berlakunya kesilapan dalam analisis dan seterusnya kesimpulan yang diperolehi adalah tidak kukuh. Dalam kes proses bercabang, penyelidik matematik atau statistik tulen amat memerlukan kepakaran penyelidik biologi dalam mendapatkan data sebenar bagi proses yang dapat memenuhi kehendak masalah supaya keputusan yang telah dibuktikan dapat ditentu sahkan dan diguna pakai.

RUJUKAN

- [1] Agresti A. & Finlay B. 1999, *Statistical Method for the Social Sciences*, 3rd ed. Prentice Hall.

- [2] Athreya K. B. & Ney P. E. 1972, *Branching Processes*, New York: Springer Verlag.
- [3] Fleishmann K & Siegmund-Schultze R. 1977, The structure of reduced critical G-W processes, *Math. Nachr.* **79**, 233-244.
- [4] Jagers P. 1975, *Branching Processes with Biological Applications*, London: Wiley.
- [5] Montgomery D C. 1983, *Design & Analysis of Experiments*, 2nd ed., John Wiley & sons.
- [6] Ott R. L. & Longnecker M. 2001, *An Introduction to Statistical Method & data Analysis*, 5th ed. Duxbury.
- [7] Rahimov I. 1995 *Random Sums and Branching Processes*, New York: Springer-Verlag.
- [8] Rahimov I. & Hasan H. 1998, Exceedance problems concerning a family of branching processes, *Pakistan J. Stats.* **14**(1), 37-47.
- [9] Yakimov A. 1980, Reduced Branching Processes, *Theory Prob. & Appl.* **25**, No. 3, 593-596.
- [10] Yeoh H. B. 1998, *Teorem Penghad bagi Proses Galton-Watson Terturun dengan Imigrasi*, Msc. Thesis. USM.
- [11] <http://en.wikipedia.org/wiki/Mathematical%5Fbiology>