

KAEDAH ALTERNATIF MENGUKUR ENDOGAMI

OLEH
AHMAD MAHIR BIN RAZALI

Jabatan Statistik
Fakulti Sains Matematik Dan Komputer
Universiti Kebangsaan Malaysia
43600 UKM Bangi, Selangor Darul Ehsan
Malaysia.

Abstrak. Satu kaedah alternatif yang menggunakan statistik kappa diperkenalkan untuk mengukur endogami. Kaedah ini didapati mempunyai kelebihan daripada kaedah pengukuran yang lain kerana ia bukan sahaja dapat mengukur endogami secara menyeluruh malahan dapat memberikan nilai rapat piawai bagi statistik endogami tersebut. Selain daripada itu, kaedah ini adalah lebih mudah dan hasilnya didapati tidak jauh berbeza dengan hasil yang diperolehi melalui kaedah kebolehjadian maksimum. Beberapa set data telah digunakan sebagai ilustrasi.

Abstract. An alternative method to measure endogamy via kappa statistics is presented in this paper. This method is found to be better than other measurement methods because it can measure an overall endogamy as well as it can give the standard error of the endogamy statistics. The method is very simple and the results are not much different from the maximum likelihood method. A few data sets are used as illustrations.

1. PENDAHULUAN

Endogami secara umumnya boleh ditakrifkan sebagai perkahwinan di antara anggota-anggota dalam kumpulan etnik yang sama (Sanusi [11]). Kaedah pengukuran endogami ini pernah dikaji oleh ramai ahli statistik dan juga sosiologi umpamanya McCaa & Schwartz [8] dan Rust & Seed [10]. Ahmad Mahir [3] telah mencadangkan supaya menggunakan kaedah kebolehdajadian maksimum dengan mengukur endogami melalui statistik theta. Kaedah ini didapati mempunyai kelebihan daripada log-linear atau pun kaedah pengukuran yang lain kerana ia bukan sahaja dapat mengukur endogami secara menyeluruh malahan dapat memberikan nilai ralat piawai bagi ukuran endogami tersebut (Ahmad Mahir [2]; Rust & Seed [10]). Walau bagaimanapun statistik theta ini agak sukar dihitung kerana seseorang penyelidik perlu menyelesaikan satu set persamaan yang terdiri daripada persamaan yang bukan linear. Kaedah ini juga tidak ekonomik kerana ia memerlukan banyak masa pengiraan. Di samping dua masalah tadi, kaedah ini tidak boleh digunakan untuk keadaan yang saiz jadual kontingensi terlalu besar umpamanya melebihi 7×7 kerana menghadapi masalah mendapatkan anggaran bagi parameter-parameter dalam persamaan yang tidak linear itu (Ahmad Mahir [2]). Dengan sebab itu kaedah ini kurang cekap. Oleh itu satu kaedah lain yang lebih cekap dan mudah akan dicadangkan melalui satu statistik yang diperkenalkan oleh Cohen [5] pada tahun 1960.

Statistik Cohen ini dikenali sebagai statistik ukuran persetujuan atau biasanya dinamakan statistik kappa (Cohen [5]). Rust & Seed [10] menyatakan statistik kappa boleh digunakan untuk mengukur endogami tetapi mereka tidak memberi sebarang penyelesaian mengenai ukuran endogami melalui kappa itu sendiri. Ahmad Mahir [2] telah menunjukkan cara mengira statistik kappa bagi ukuran endogami ini. Oleh itu kertas ini akan cuba memperkenalkan statistik kappa dengan menggunakan data yang digunakan oleh Ahmad Mahir [3] serta data tempatan iaitu data perkahwinan campuran bagi penduduk Bandar Melaka yang diperolehi

daripada Sanusi [11].

2. KAEDAH

2.1 Susunan Data.

Katakan data yang diperolehi diberikan dalam bentuk jadual kontingensi yang bersaiz $K \times K$ seperti yang dipaparkan dalam jadual 1. Tandakan x_{ij} sebagai bilangan kekerapan dalam sel (i, j) dengan i dan j mengambil nilai $1, 2, \dots, K$. Tandakan G_i sebagai kategori kumpulan ke- i dan G_j sebagai kategori kumpulan ke- j . N pula ialah jumlah keseluruhan kekerapan di dalam jadual kontingensi.

Jadual 1. Format bagi data perkahwinan

		Wanita								Jumlah
		G_1	G_2	.	.	G_j	.	.	G_k	
Lelaki	G_1	x_{11}	x_{12}	.	.	x_{1j}	.	.	x_{1k}	x_{1+}
	G_2	x_{21}	x_{22}	.	.	x_{2j}	.	.	x_{2k}	x_{2+}
	.	.	.	.	.	.	.	.	.	.
	.	.	.	.	.	.	.	.	.	.
	G_i	x_{i1}	x_{i2}	.	.	x_{ij}	.	.	x_{ik}	x_{i+}
	.	.	.	.	.	.	.	.	.	.
	.	.	.	.	.	.	.	.	.	.
	G_k	x_{k1}	x_{k2}	.	.	x_{kj}	.	.	x_{kk}	x_{k+}
Jumlah		x_{+1}	x_{+2}	.	.	x_{+j}	.	.	x_{+k}	N

Sebenarnya x_{ij} adalah bilangan perkahwinan dengan pasangan lelaki daripada etnik G_i , sementara pasangan wanita dari-

pada etnik G_j . Apabila berlaku perkahwinan antara sesama etnik, maka $i = j$ atau $G_i = G_j$. Perkahwinan ini dinamakan perkahwinan endogami. Sebaliknya apabila $i \neq j$, maka $G_i \neq G_j$ dan perkahwinan ini dinamakan perkahwinan campuran. Data endogami adalah kekerapan yang terletak pada pepenjuru jadual. Oleh itu minat kita hanyalah kepada data di pepenjuru ini sahaja untuk mengukur endogami. Statistik kappa amat sesuai untuk tujuan tersebut kerana ia adalah suatu ukuran persetujuan yang menumpukan perhatian kepada data di pepenjuru dan juga mengambil kira data bukan dipepenjuru (Rust & Seed [10]).

2.2 Statistik Kappa.

Menurut Light [7], banyak ukuran telah diperkenalkan untuk mengukur persetujuan bagi data nominal atau kebolehpercayaan bagi data selanjar. Beberapa orang ahli statistik umpamanya Ebel [6] dan Robinson [9] telah pun membuat penyelidikan tentang ukuran persetujuan ini. Scott [12] telah memperkenalkan satu ukuran persetujuan iaitu,

$$\pi = \frac{p_o - p_e}{1 - p_e} \quad ((2.1))$$

dengan p_o adalah perkadaran yang dicerap bagi pasangan yang bersetuju, sementara itu p_e adalah jumlah kebarangkalian bagi kemasukan dalam sel (i, j) atau sel pepenjuru bagi jadual kontingensi $K \times K$. Cohen [5] telah mencadangkan satu ukuran persetujuan yang dinamakan kappa. Statistik ini ditakrifkan sebagai,

$$\kappa = \frac{p_o - p_e}{1 - p_e} \quad (2.2)$$

dengan p_o adalah nilai perkadaran yang persetujuan berlaku dan p_e adalah nilai perkadaran persetujuan dijangka berlaku secara peluang. Statistik ini hampir sama dengan yang diberikan oleh Scott [12] tetapi takrifan p_e agak berbeza sedikit. Menurut Cohen [5], pekali kappa ini bolehlah dilihat sebagai perkadaran persetujuan selepas peluang persetujuan itu disingkirkan daripada dimasukkan dalam pengiraan. Statistik kappa ini diterima umum sebagai ukuran persetujuan (lihat Bishop *et al* [4] dan Agresti [1]).

Statistik kappa ini diberikan dalam sebutan nilai kekerapan seperti berikut,

$$\kappa = \frac{N \sum_i x_{ii} - \sum_i x_{i+} x_{+i}}{N^2 - \sum_i x_{i+} x_{+i}} \quad (2.3)$$

Anggaran varians asimptotiknya diberikan oleh Bishop *drk* [4] sebagai,

$$\begin{aligned} \sigma_{\infty}^2[\kappa] &= \frac{1}{N} \left\{ \frac{S_o(1-S_o)}{(1-S_e)^2} + \frac{2(1-S_o)(2S_oS_e - \sum p_{ii}(p_{i+} + p_{+i}))}{(1-S_e)^3} \right\} \\ &+ \frac{1}{N} \left\{ \frac{(1-S_o)^2(\sum \sum p_{ij}(p_{j+} + p_{+i})^2 - 4S_e^2)}{(1-S_e)^4} \right\} \end{aligned} \quad (2.4)$$

dengan

$$\begin{aligned} p_{ii} &= \frac{x_{ii}}{N}, \quad p_{i+} = \frac{x_{i+}}{N}, \quad p_{+i} = \frac{x_{+i}}{N}, \\ S_o &= \sum p_{ii} \text{ dan } S_e = \sum p_{i+} p_{+i}. \end{aligned} \quad (2.5)$$

Dari segi teorinya, nilai kappa boleh berada di antara $-1 \leq \kappa \leq 1$. Walau bagaimanapun nilai kappa yang negatif jarang-jarang berlaku (Cohen [5]).

3. HASIL

Hasil daripada pengiraan statistik kappa diberikan seperti dalam jadual 2 dan jadual 3. Set data 1,2 dan 3 dalam jadual 2 adalah data perkahwinan campuran etnik Melayu, Cina, India dan Portugis di Bandar Melaka (lihat lampiran A). Data ini telah diubahsuai sedikit daripada data asal yang diperolehi daripada Sanusi [11]. Jadual asal mengandungi tujuh kumpulan etnik tetapi cuma empat kumpulan etnik sahaja yang mempunyai data kekerapan yang boleh digunakan. Jadual asal juga memberikan nama kumpulan etnik yang lain tetapi kekerapan selselnya adalah

sifar. Oleh itu dengan membuang kategori yang tiada maklumat ini tidak menjejaskan analisis. Untuk keterangan lanjut mengenai rasionalnya, sila lihat Rust & Seed [10].

Set data 4 dan 5 dalam jadual 3 pula masing-masing adalah data perkahwinan campuran bagi empat kumpulan etnik di Antequera, Mexico pada akhir abad ke 17 dan ke 18 (lihat lampiran A). Set asal data ini diperolehi daripada McCaa & Schwartz [8]. Manakala Ahmad Mahir [3] menggunakan set data ini untuk mengilustrasikan kaedah kebolehjadian maksimum melalui statistik theta bagi mengukur endogami.

JADUAL 2

Anggaran parameter, ralat piawai serta 95%
selang keyakinan bagi kadar endogami

Set data	$\hat{\kappa}$	$\hat{\sigma}(\kappa)$	95% selang keyakinan
1	0.867	0.022	(0.824, 0.910)
2	0.982	0.008	(0.966, 0.998)
3	0.967	0.011	(0.945, 0.989)

4. PERBINCANGAN

Nilai kappa yang diperolehi bagi set data 1, 2 dan 3 dalam jadual 2 adalah besar kerana ia menghampiri nilai 1. Nilai kappa yang hampir dengan satu bermakna ukuran endogami adalah sempurna sementara nilai kappa yang sifar bermaksud tiada berlakunya endogami. Sementara itu nilai di antara sifar dan satu bermaksud endogami separa. Ukuran endogami yang dihitung bagi tiga set data yang pertama menunjukkan yang ia adalah tinggi sekali atau pun kuat kerana hampir dengan nilai 1. Sembilan puluh lima peratus selang keyakinan bagi nilai kappa berkenaan menunjukkan bahawa kesemua ukuran endogami tersebut sangat bererti

kerana tiada satu pun selang tersebut yang merangkumi nilai sifar. Secara keseluruhannya endogami di kalangan kumpulan etnik utama di Bandar Melaka sangat kuat. Ini bermakna mereka masih mengamalkan cara perkahwinan sesama etnik sendiri dan tidak banyak berlaku perkahwinan campuran.

JADUAL 3

Anggaran parameter, ralat piawai serta 95%
selang keyakinan bagi kadar endogami

Set data	$\hat{\kappa}$	$\sigma(\hat{\kappa})$	95% selang keyakinan
4	0.531	0.023	(0.487, 0.575)
5	0.408	0.024	(0.361, 0.455)

Nilai endogami bagi set data 4 dan 5 dalam jadual 3 adalah lebih kecil sedikit daripada yang diperolehi bagi set data 1, 2 dan 3. Ini tidak menghairankan kerana set data 4 dan 5 adalah daripada sampel yang berlainan. Walau bagaimanapun kedua-dua ukuran endogami ini masih lagi bererti kerana sembilan puluh lima peratus selang keyakinan bagi nilai kappa berkenaan menunjukkan bahawa kesemua ukuran endogami tersebut tidak merangkumi nilai sifar. Secara keseluruhannya kedua-dua set tersebut menunjukkan yang endogami dikalangan etnik di Antè-quera, Mexico pada abad ke 17 dan 18 masih lagi kuat. Kesimpulan ini sama seperti yang dibuat oleh Ahmad Mahir [3].

Ahmad Mahir [3] telah memberikan nilai endogami bagi kedua-dua set data ini melalui ukuran theta. Nilai theta bagi set data 4 adalah 0.528 dan set data 5 pula adalah 0.393. Jika dibandingkan dengan apa yang diperolehi melalui statistik kappa iaitu masing-masing 0.531 dan 0.408, didapati perbezaan tersebut tidak begitu ketara.

4. KESIMPULAN

Kaedah alternatif untuk mengukur endogami melalui statistik kappa didapati mempunyai kelebihan daripada kaedah pengukuran yang lain kerana ia bukan sahaja dapat mengukur endogami secara menyeluruh malahan dapat memberikan nilai ralat piawai bagi statistik endogami tersebut. Selain daripada itu, kaedah ini adalah lebih mudah dan hasilnya didapati tidak jauh berbeza dengan hasil yang diperolehi dengan kaedah kebolehjadian maksimum melalui statistik theta.

RUJUKAN

- [1] A. Agresti, *Categorical Data Analysis*, Wiley, New York, 1990.
- [2] B. R. Ahmad Mahir, *Statistical Issues on Interracial Marriages. PhD dissertation*, Medical University of South Carolina, Charleston, South Carolina, 1991.
- [3] B. R. Ahmad Mahir, *Endogamy Measure in Interracial Marriages*, Sains Malaysiana **21**(4) (1992), 47-56.
- [4] Y.M.M. Bishop, S.E. Fienberg & P.W. Holland, *Discrete Multivariate Analysis: Theory and Practice*, MIT Press, Cambridge, 1975.
- [5] J. Cohen, *A Coefficient of Agreement for Nominal Scales*, Educ. Psychol. Meas. **20** (1960), 37-46.
- [6] R. L. Ebel, *Estimation of the Reliability of Ratings*, Psychometrika **16** (1951), 407-424.
- [7] R.J. Light, *Measures of Response Agreement for Qualitative Data: Some Generalizations and Alternatives*, Psychological Bulletin **76**(5) (1971), 365-377.
- [8] R. McCaa & S. B. Schwartz, *Measuring Marriage Patterns: Percentages, Cohen's Kappa, and Log-Linear Models*, Comparative Studies in Society and History **25** (1983), 711-720.
- [9] W.S. Robinson, *The Statistical Measurement of Agreement*, American Sociological Review **22** (1957), 17-27.
- [10] P.F. Rust & P. Seed, *Equality Endogamy: Statistical Approaches*, Social Science Research **14** (1985), 57-79.
- [11] Sanusi Osman, *Ikatan Etnik dan Kelas Di Malaysia*, Penerbit Universiti Kebangsaan Malaysia, Bangi (1989), 72-83.
- [12] W.A. Scott, *Reliability of Content Analysis: The Case of Nominal Scale Coding*, Public Opinion Quarterly **XIX** (1955), 321-325.

LAMPIRAN A

SET DATA 1.

Corak perkahwinan dikalangan responden lelaki.

		Wanita				
		Melayu	Cina	India	Portugis	Jumlah
L e l a k i	Melayu	119	8	1	0	128
	Cina	8	168	2	0	178
	India	3	5	43	0	51
	Portugis	0	6	1	28	35
	Jumlah	130	187	47	28	392

Dipetik dan diubahsuai daripada Sanusi [11]

SET DATA 2.

Corak perkahwinan dikalangan ibu responden.

Wanita

L e l a k i		Melayu	Cina	India	Portugis	Jumlah
	Melayu	128	1	1	0	130
	Cina	1	200	0	1	202
	India	0	0	53	1	54
	Portugis	0	0	0	33	33
	Jumlah	129	201	54	35	419

Dipetik dan diubahsuai daripada Sanusi [11]

SET DATA 3.

Corak perkahwinan dikalangan bapa responden.

Wanita

L
e
l
a
k
i

	Melayu	Cina	India	Portugis	Jumlah
Melayu	122	2	5	0	129
Cina	0	201	0	0	201
India	1	0	53	0	54
Portugis	0	0	1	34	35
Jumlah	123	203	59	34	419

Dipetik dan diubahsuai daripada Sanusi [11]

SET DATA 4.

Corak perkahwinan di kalangan etnik,
Antequera, Mexico, 1693-1700.

Wanita

L e l a k i		Spanish	Mestizo	Mulatto	Indian	Jumlah
	Spanish	219	39	21	10	289
	Mestizo	30	48	33	17	128
	Mulatto	16	30	132	36	216
	Indian	3	23	8	111	145
	Jumlah	268	140	194	174	776

Dipetik daripada McCaa & Schwartz [8]

SET DATA 5.

Corak perkahwinan di kalangan etnik,
Antequera, Mexico, 1793-1797.

Wanita

L e l a k i		Spanish	Mestizo	Mulatto	Indian	Jumlah
	Spanish	163	49	12	9	233
	Mestizo	35	55	28	20	138
	Mulatto	15	20	13	8	56
	Indian	33	45	26	183	287
	Jumlah	246	169	79	220	714

Dipetik daripada McCaa & Schwartz [8]

LAMPIRAN B

Contoh pengiraan anggaran parameter, ralat piawai serta 95% selang keyakinan kadar endogami bagi set data 1:

$$N = 392$$

$$\sum x_{ii} = \{119 + 168 + 43 + 28\} = 358$$

$$S_o = \sum p_{ii} = \sum x_{ii}/N = \{119 + 168 + 43 + 28\}/392 = 0.9133$$

$$\sum x_{i+}x_{+i} = \{(128)(130) + (178)(187) + (51)(47) + (35)(28)\} = 53,303$$

$$S_e = \{(128)(130) + (178)(187) + (51)(47) + (35)(28)\}/(392)(392) = 0.3469$$

$$\begin{aligned} \sum p_{ii}(p_{i+} + p_{+i}) = & \{(119)(128 + 130) + (168)(178 + 187) + (43)(51 + 47) \\ & + (28)(35 + 28)\}/(392)(392) = 0.6378 \end{aligned}$$

$$\begin{aligned} \sum \sum p_{ij}(p_{j+} + p_{+i})^2 \\ = & \{(119)(128 + 130)^2 + (8)(128 + 187)^2 + (1)(128 + 47)^2 \\ & + (0)(128 + 28)^2 + (8)(178 + 130)^2 + (168)(178 + 187)^2 \\ & + (2)(178 + 47)^2 + (0)(178 + 28)^2 + (3)(51 + 130)^2 \\ & + (5)(51 + 187)^2 + (43)(51 + 47)^2 + (0)(51 + 28)^2 \\ & + (0)(35 + 130)^2 + (6)(35 + 187)^2 + (1)(35 + 47)^2 \\ & + (28)(35 + 28)^2\}/(392)(392)(392) = 0.5511 \end{aligned}$$

$$\hat{\kappa} = \frac{(392)(358) - (53,303)}{(392)^2 - (53,303)} = 0.867$$

$$\begin{aligned} \sigma^2[\hat{\kappa}] = & \frac{1}{392} \left\{ \frac{(0.9133)(0.0867)}{(0.6531)^2} + \frac{2(0.0867)(2(0.9133)(0.3469) - 0.6378)}{(0.6531)^3} \right\} \\ & + \frac{1}{392} \left\{ \frac{0.0867^2(0.5511 - 4(0.3469)^2)}{(0.6531)^4} \right\} = 0.00047178 \end{aligned}$$

$$\therefore \sigma[\hat{\kappa}] = 0.0217$$

$$\begin{aligned} \hat{\kappa} \pm 1.96\sigma[\hat{\kappa}] &= 0.867 \pm 1.96(0.0217) \\ &= (0.824, 0.910) \end{aligned}$$