

DATA ORDINAL PADA ANALISIS KONFIGURASI FREKUENSI

Resa Septiani Pontoh

Departemen Statistika, Fakultas MIPA, Universitas Padjadjaran Email

resa.pontoh@gmail.com

ABSTRAK. Analisis Konfigurasi Frekuensi atau *Configural Frequency Analysis* (CFA) adalah suatu analisis yang akan memunculkan konfigurasi dari beberapa kategori yang berasal dari beberapa variabel yang berbeda dimana secara signifikan berbeda antara yang diekspektasikan dengan yang terobservasi. Dengan kata lain, analisis ini akan sangat bermanfaat untuk mencari penyebab terjadinya penyimpangan pada model yang ada. Jika hal ini terjadi, maka *Type* dan *Antitype* akan muncul. *Type* akan dihasilkan jika konfigurasi tersebut lebih sering terjadi dari yang telah diekspektasikan sedangkan *Antitype* dihasilkan jika konfigurasi tersebut lebih jarang terjadi dari yang telah diekspektasikan. Pada banyak kasus atau data yang dianalisis menggunakan Analisis Konfigurasi Frekuensi, biasanya data penelitian yang digunakan bersifat nominal atau dianggap sebagai data nominal. Padahal, pada kenyataannya, data ordinal yang dianalisis menggunakan analisis ini akan menghasilkan kesimpulan yang berbeda jika data tersebut diasumsikan sebagai data nominal. Makalah ini akan memaparkan perbedaan yang terjadi antara CFA ordinal dan CFA nominal jika data yang digunakan adalah ordinal. Secara garis besar, diperoleh hasil bahwa perbedaan antara frekuensi observasi dan ekspektasi lebih berdekatan jika base model ordinal yang digunakan dibandingkan dengan base model nominal. Hal ini tentu saja mengakibatkan *Type* dan *Antitype* akan lebih jarang muncul jika base model yang digunakan dalam analisis tersebut menggunakan data ordinal.

Kata Kunci: *Configural Frequency Analysis; Analisis Konfigurasi Frekuensi.*

1. PENDAHULUAN

Analisis Konfigurasi Frekuensi atau *Configural Frequency Analysis* (CFA) adalah suatu analisis yang akan memunculkan konfigurasi dari beberapa kategori yang berasal dari beberapa variabel yang berbeda dimana secara signifikan berbeda antara yang diekspektasikan dengan yang terobservasi. Dengan kata lain, analisis ini akan sangat bermanfaat untuk mencari penyebab terjadinya penyimpangan pada model yang ada. Jika hal ini terjadi, maka *Type* dan *Antitype* akan muncul. *Type* akan dihasilkan jika konfigurasi tersebut lebih sering terjadi dari yang telah diekspektasikan sedangkan *Antitype* dihasilkan jika konfigurasi tersebut lebih jarang terjadi dari yang telah diekspektasikan. Analisis ini sangat bermanfaat untuk melihat konfigurasi mana yang menyimpang dari model yang ada sehingga penyimpangan tersebut dapat menjadi informasi baru bagi para peneliti.

Pada aplikasinya, CFA dapat digunakan untuk beragam bidang ilmu seperti kesehatan, psikologi, ekonomi, dsb. Salah satu ilustrasi di bidang ekonomi dan bisnis yang dapat digunakan untuk menggambarkan analisis ini adalah sebagai misal akan dilakukan pemodelan mengenai calon investor saham PT BEI. Namun, ketika diteliti ternyata berdasarkan model yang telah terbentuk, ditemukan fakta bahwa beberapa karakteristik dari para calon investor tersebut tidaklah tepat. Dengan menggunakan CFA yang berfokus pada konfigurasi dari kategori variabel yang berbeda, akan terlihat konfigurasi mana yang secara signifikan menyimpang dari yang telah diekspektasikan oleh model. Sehingga, dengan hasil tersebut, PT BEI dapat membuat suatu kebijakan mengenai masyarakat dengan karakteristik mana yang perlu lebih didukasi ataupun yang perlu lebih didekati karena memiliki peluang yang lebih besar untuk bersedia dan berminat menanamkan sahamnya.

Berkaitan dengan penelitian ini, pada tahap pengerjaan dengan menggunakan CFA, banyak kasus atau data yang dianalisis menggunakan analisis ini secara langsung mengasumsikan data penelitian yang digunakan bersifat nominal atau dianggap sebagai data nominal. Padahal, jika data ordinal dianalisis menggunakan CFA nominal, akan menghasilkan kesimpulan yang berbeda (diasumsikan sebagai data nominal).

Oleh karena itu, makalah ini akan memaparkan perbedaan yang terjadi antara CFA ordinal dan CFA nominal jika data yang digunakan adalah ordinal. Secara garis besar, diperoleh hasil bahwa perbedaan antara frekuensi observasi dan ekspektasi lebih berdekatan jika base model ordinal yang digunakan dibandingkan dengan base model nominal. Sehingga, *Type* dan *Antitype* akan lebih jarang muncul jika base model yang digunakan dalam analisis tersebut menggunakan data ordinal.

2. METODE PENELITIAN

Data yang digunakan pada penelitian ini berkaitan dengan data kesehatan (tidak bisa dipublikasikan) menggunakan metode *Configural Frequency Analysis* (CFA). Data yang digunakan adalah data sekunder sehingga teknik pengambilan sampel tidak terlalu menjadi permasalahan. Adapun fokus penelitian ini adalah mencari perbedaan antara CFA ordinal dan nominal jika terdapat data ordinal di dalamnya.

Seperti yang telah dijelaskan sebelumnya, CFA mempunyai keunggulan dalam menganalisis data kategori yaitu CFA tidak melihat ada atau tidaknya hubungan antara beberapa variabel namun CFA akan lebih dalam lagi menggali informasi yang ada yaitu kategori-kategori diantara variabel yang berbeda.

Secara umum, terdapat beberapa tahapan dalam CFA yaitu:

1. Pemilihan base model dan menghitung nilai ekspektasi frekuensi dari tiap konfigurasi.

Pada tahapan ini, perlu diperhatikan asumsi yang mendasar mengenai variabel yang ada apakah variabel-variabel tersebut akan terbagi menjadi prediktor dan kriteria atautkah mempunyai status yang sama. Selain itu, perlu juga diperhatikan skema pengambilan sampel bagaimana data diperoleh apakah *multinomial sampling* atautkah *product multinomial sampling*.

2. Uji kecocokan model.
Uji kecocokan model ini (*the goodness of fit*) digunakan untuk melihat apakah model sudah dapat menggambarkan fenomena pada data dengan baik.
3. Uji signifikansi konfigurasi yang menyimpang dari model.
Uji signifikansi ini digunakan untuk mencari *Type* dan *Antitype* mana yang benar-benar signifikan dengan α^* tertentu.
4. Penjelasan *Type* dan *Antitype*.
Hal ini penting dilakukan agar fenomena-fenomena yang terjadi diluar apa yang dipaparkan oleh model yang ada bisa didapatkan.

Untuk mendapatkan frekuensi ekspektasi, maka diperlukan *base model* pada analisis ini. Biasanya, model *log linear* adalah *base model* yang paling banyak digunakan karena model ini dapat dimodifikasi untuk beragam interaksi antar variabel maupun tanpa interaksi. Pada tahapan ini, base model tidak akan memaparkan interaksi apapun antar variabel. Oleh karena data yang diperoleh adalah data sekunder, maka *multinomial sampling* yang digunakan. Secara umum, base model CFA yang berasal dari log linear model adalah sebagai berikut :

$$\log E = X\lambda,$$

dimana E adalah frekuensi dari konfigurasi yang sesuai dengan model, X adalah desain matriks atau matrik indicator, sedangkan λ adalah vector dari parameter-parameter model. Hal lain yang perlu diketahui adalah parameter-parameter tersebut bukanlah fokus dari CFA karena CFA akan mencari kesenjangan atau perbedaan yang terjadi antara frekuensi observasi dan ekspektasi.

Berbeda dengan CFA pada data nominal, analisis dengan menggunakan data ordinal akan menjadi lebih sulit karena akan melibatkan satu atau lebih kolom yang memuat variabel ordinal (berlevel). Pada penelitian ini akan dimulai dengan dua variabel dengan kolom yang merefleksikan data ordinal dan baris yang merefleksikan data nominal. Secara umum, base model log linear dengan data ordinal adalah sebagai berikut:

$$\log E = \lambda_0 + \lambda_i^{\text{rows}} + \lambda_j^{\text{columns}} + (v_j - \bar{v})\lambda_j^{\text{columns}'},$$

dimana λ adalah parameter yang akan diestimasi, $\lambda_j^{\text{columns}}$ adalah parameter yang merefleksikan kolom yang mempunyai data ordinal, \bar{v} adalah rata-rata dari rank v_j . Perlu diketahui pula bahwa frekuensi ekspektasi nantinya akan diestimasi dengan proses iterasi dimulai dengan $E_{ij}^{(0)} = 1$ untuk $i=1, \dots, I$ dan $j=1, \dots, J$.

Dalam proses iterasi, pada tahapan awal (*initial iteration step*) dimana semua frekuensi ekspektasi adalah 1, iterasi (κ) dibuat menjadi $\kappa = 0$. Kemudian satu putaran meliputi tiga langkah di bawah ini (Eye [1] ; Fienberg [2]):

$$E_{ij}^{(3\kappa+1)} = E_{ij}^{(3\kappa)} \left(\frac{N_i}{E_i^{(3\kappa)}} \right),$$

$$E_{ij}^{(3\kappa+2)} = E_{ij}^{(3\kappa+1)} \left(\frac{N_j}{E_j^{(3\kappa+1)}} \right),$$

$$E_{ij}^{(3\kappa+3)} = E_{ij}^{(3\kappa+2)} \left(\frac{\sum_k v_k N_{ik}}{\sum_k v_k E_{ik}^{(3\kappa+2)}} \right)^{v_j} \left(\frac{\sum_k (1-v_k) N_{ik}}{\sum_k (1-v_k) E_{ik}^{(3\kappa+2)}} \right)^{1-v_j},$$

dimana $i=1, \dots, I$ dan $j=1, \dots, J$. Ketiga langkah ini dilakukan berulang-ulang sampai diperoleh frekuensi ekspektasi E_{ij} yang konvergen. Analisis CFA ordinal memang sangat rumit untuk diaplikasikan dikarenakan ketika melibatkan variabel yang lebih banyak, maka rumusan nilai ekspektasi akan menjadi semakin rumit dan proses simulasi pun akan semakin lama.

Berkaitan dengan uji kecocokan model (*the goodness of fit*), perlu diketahui bahwa CFA tidak mementingkan apakah model yang digunakan cocok atau tidak tetapi CFA bahkan akan mempertimbangkan segala efek yang mungkin ada yang tentunya berakibat pada gagalannya suatu model mendeskripsikan data dengan baik (Eye [1]). Penelitian ini tidak melampirkan uji kecocokan model yang telah dilakukan. Namun, model yang ada sudah dapat menjelaskan data dengan baik pada $\alpha = 0.05$.

Setelah diperoleh nilai ekspektasi dan observasi, barulah dilakukan uji signifikansi untuk melihat apakah konfigurasi yang ada secara signifikan lebih sering maupun lebih jarang terjadi dari apa yang telah diekspektasikan (Eye et al[3]). Untuk itu perlu dicari nilai z hitung dan z criteria (z_{α^*}), dimana $\alpha^* = \alpha/k$ dengan k adalah banyaknya konfigurasi. Jika z hitung $> z_{\alpha^*}$ maka *Type* akan muncul. Jika z hitung $< -z_{\alpha^*}$ maka *Antitype* akan muncul. Pada penelitian ini, banyaknya konfigurasi adalah 15 sehingga menggunakan $\alpha^* = 0.05/15 = 0.0033$. Adapun nilai z hitung diperoleh dengan rumusan sebagai berikut:

$$z_{hitung} = \frac{N_i - E_i}{\sqrt{E_i}},$$

dimana N_i adalah frekuensi observasi dan E_i adalah frekuensi ekspektasi dari tiap konfigurasi.

Setelah semua proses tersebut dilewati, barulah *Type* dan *Antitype* yang signifikan dapat diinterpretasikan dan ditemukan fenomena-fenomena mana yang secara signifikan menyimpang dari model.

3. HASIL PENELITIAN DAN PEMBAHASAN

Penelitian ini akan menggunakan dua variabel yaitu perkembangan tiga jenis gejala penyakit (Y) yang bersifat nominal dan lama penelitian (T) yang bersifat ordinal. Data yang digunakan pada penelitian ini adalah data sekunder banyaknya keluarga di daerah X yang menderita penyakit R dengan melihat jenis gejala yang timbul Y ($Y = 1, 2, 3$) pada lima bulan pertama ($T = 1, 2, 3, 4, 5$). Instansi terkait ingin melihat gejala mana yang perlu mendapat perhatian khusus untuk segera ditangani mengingat penyebaran penyakit R yang semakin meluas.

Gejala penyakit (Y) bersifat nominal karena tidak ada perbedaan tingkatan pada ketiga jenis tipe gejala tersebut sedangkan waktu penelitian (T) dianggap mempunyai perbedaan tingkatan atau level karena semakin lama waktu penelitian, kemungkinan penyakit berkembang semakin besar. Oleh karena itu, waktu penelitian dikategorikan mempunyai sifat ordinal.

Adapun beberapa penjelasan kode pada tabel 1 dan 2 adalah sebagai berikut:

- 11 adalah responden yang diteliti pada bulan pertama dengan gejala yang muncul adalah jenis 1
- 12 adalah responden yang diteliti pada bulan pertama dengan gejala yang muncul adalah jenis 2
- 13 adalah responden yang diteliti pada bulan pertama dengan gejala yang muncul adalah jenis 3
- 21 adalah responden yang diteliti pada bulan kedua dengan gejala yang muncul adalah jenis 1
- 22 adalah responden yang diteliti pada bulan kedua dengan gejala yang muncul adalah jenis 2.
- 23 adalah responden yang diteliti pada bulan kedua dengan gejala yang muncul adalah jenis 3
- 31 adalah responden yang diteliti pada bulan ketiga dengan gejala yang muncul adalah jenis 1
- 32 adalah responden yang diteliti pada bulan ketiga dengan gejala yang muncul adalah jenis 2
- 33 adalah responden yang diteliti pada bulan ketiga dengan gejala yang muncul adalah jenis 3
- 41 adalah responden yang diteliti pada bulan keempat dengan gejala yang muncul adalah jenis 1
- 42 adalah responden yang diteliti pada bulan keempat dengan gejala yang muncul adalah jenis 2
- 43 adalah responden yang diteliti pada bulan keempat dengan gejala yang muncul adalah jenis 3
- 51 adalah responden yang diteliti pada bulan kelima dengan gejala yang muncul adalah jenis 1
- 52 adalah responden yang diteliti pada bulan kelima dengan gejala yang muncul adalah jenis 2

- 53 adalah responden yang diteliti pada bulan kelima dengan gejala yang muncul adalah jenis 3

Dengan menggunakan data ini, akan dibandingkan perbedaan hasil menggunakan CFA dengan base model yang diperuntukkan untuk data nominal (CFA Nominal) dijelaskan pada tabel 1 dan ordinal (CFA Ordinal) dijelaskan pada tabel 2. Secara kasat mata terlihat sangat jelas perbedaan yang terjadi dimana tabel 1 memperlihatkan adanya satu *Type* dan dua *Antitype*, sedangkan tabel 2 hanya satu *Type* yang muncul. Oleh karena itu perlu dicari perbedaan lain yang mungkin saja terjadi.

Tabel 1 menunjukkan bahwa *Type* muncul pada konfigurasi 11 sedangkan *Antitype* muncul pada konfigurasi 22 dan 52. Informasi yang dapat dijelaskan pada tabel 1 adalah sebagai berikut:

1. *Type* pada konfigurasi 11 menjelaskan bahwa pada bulan pertama, jenis gejala tipe 1 terjadi lebih sering daripada yang telah diekspektasikan oleh model.
2. *Antitype* pada konfigurasi 22 menjelaskan bahwa pada bulan kedua, jenis gejala tipe 2 terjadi lebih jarang daripada yang telah diekspektasikan oleh model.
3. *Antitype* pada konfigurasi 52 menjelaskan bahwa pada bulan kelima, jenis gejala tipe 2 terjadi lebih jarang daripada yang telah diekspektasikan oleh model.

Hasil tersebut mengindikasikan bahwa instansi tersebut perlu lebih memfokuskan program yang telah berjalan pada responden dengan gejala penyakit jenis 1 yang terjadi pada bulan pertama. *Antitype* pada konfigurasi 22 dan 52 mengindikasikan bahwa pada bulan kedua dan kelima dengan jenis gejala tipe 2 tidak terlalu mengkhawatirkan dalam hal banyaknya responden karena jumlahnya lebih sedikit dari yang telah diekspektasikan. Namun, terdapat indikasi bahwa diantara kelima waktu pengamatan kecuali pada bulan kedua dan kelima, gejala penyakit tipe 1 akan lebih sering terjadi dari yang telah diekspektasikan oleh model yang ada.

Hasil tersebut sejalan dengan munculnya *Type* pada konfigurasi 11 yang juga menjelaskan bahwa ternyata kejadian gejala 1 lebih sering muncul pada bulan pertama. Dengan demikian instansi terkait perlu lebih memfokuskan program yang telah berjalan pada gejala penyakit tipe 1 yang terjadi pada bulan pertama.

Berbeda dengan apa yang termuat pada tabel 1, tabel 2 menunjukkan bahwa hanya muncul satu *Type* yaitu pada konfigurasi 31 yang mengindikasikan bahwa pada bulan ketiga, gejala tipe 1 lebih sering muncul dari yang telah diekspektasikan model. Oleh karena itu, instansi terkait perlu lebih memfokuskan program yang telah berjalan pada responden dengan gejala penyakit jenis 1 yang terjadi pada bulan ketiga.

Berdasarkan interpretasi hasil tersebut, terlihat dengan jelas terdapat perbedaan hasil yang cukup signifikan diantara kedua tipe analisis. Dari kedua tabel tersebut terlihat bahwa secara garis besar, perbedaan antara frekuensi observasi dan ekspektasi lebih berdekatan jika data ordinal yang digunakan pada base model dibandingkan dengan jika data nominal yang

digunakan. Sebagai missal pada konfigurasi 11, CFA nominal memperlihatkan perbedaan sebesar Hal ini tentu saja mengakibatkan *Type* dan *Antitype* akan lebih jarang muncul jika base model yang digunakan dalam analisis tersebut menggunakan data ordinal. Terlihat pada CFA nominal muncul satu *Type* dan dua *Antitype* sedangkan untuk CFA ordinal hanya muncul 1 *Type* yaitu pada konfigurasi 31.

Tabel 1. CFA Nominal

TY	Observasi	Ekspektasi	<i>Type/Antitype</i>
11	11	6.258	T
12	4	5.781	
13	7	9.671	
21	8	7.67	
22	2	5.132	A
23	4	2.858	
31	6	5.411	
32	8	3.865	
33	4	2.671	
41	9	5.867	
42	14	10.571	
43	25	19.65	
51	7	9.57	
52	5	8.639	A
53	14	10.586	

Tabel 2. CFA Ordinal

TY	Observasi	Ekspektasi	<i>Type/Antitype</i>
11	11	10.873	
12	4	5.446	
13	7	8.045	
21	8	8.076	
22	2	3.587	
23	4	7.665	
31	6	4.67	T

32	8	7.89	
33	4	3.787	
41	9	10.891	
42	14	12.788	
43	25	23.546	
51	7	8.57	
52	9	8.06	
53	14	13.673	

4. SIMPULAN

Berdasarkan analisis yang telah dilakukan dengan membandingkan hasil yang diperoleh pada tabel 1 dan tabel 2, penelitian ini memperlihatkan bahwa ketika analisis frekuensi konfigurasi mengabaikan struktur data ordinal dengan mengasumsikan bahwa data berskala nominal akan dihasilkan kesimpulan yang berbeda. Terlihat dengan jelas pada kedua tabel, ketika base model mengakomodasi adanya data ordinal, frekuensi ekspektasi dan observasi mempunyai nilai yang lebih berdekatan. Hal ini diikuti pula dengan munculnya *Type* dan *Antitype* dengan jumlah yang lebih sedikit. Selain itu, *Type* dan *Antitype* muncul pada konfigurasi yang berbeda diantara CFA nominal dan ordinal. Kaitannya dengan permasalahan yang perlu dikaji, akurasi pada CFA baik nominal maupun ordinal belum terlihat dengan baik. Oleh karena itu, penelitian tersebut dapat dilakukan agar hasil yang tersaji menjadi lebih akurat.

DAFTAR PUSTAKA

- [1] Eye, A. 2002. *Configural Frequency Analysis: Methods, models, and applications*. Mahwah, N.J: Lawrence Erlbaum Associates.
- [2] Eye, A., Indurkha, A., & Kreppner, K. 2000. *CFA as a tool for person-oriented research Unidimensional and within-individual analyses of nominal level and ordinal data*. *Psychologische Beitrage*, 42, 383 – 401.
- [3] Fienberg, S.E. 1980. *The Analysis of cross-classified categorical data (2nd ed)*. Cambridge, MA: MIT Press.