Distribusi Data Tidak Normal pada Pemodelan Persamaan Struktural (SEM)

Oleh Wahyu Widhiarso | Fakultas Psikologi UGM | 2010

Tulisan ini mendiskusikan masalah ketidaknormalan distribusi data yang dianalisis dengan menggunakan SEM. Apa saja yang menyebabkan data tidak terdistribusi normal? Apa dampaknya dalam pengujian SEM? dan Bagaimana penanganannya? Mari kita diskusi bersama di halaman ini.Menurut Schumacker & Lomax (2004), jika data variabel tampak berbentuk interval dan memiliki distribusi multivariat normal, maka perkiraan *maximum likelihood* (ML), eror standar, dan uji chi square akan menghasilkan nilai yang akurat dan kuat (*robust*). Akurat artinya sesuai dengan nilai kenyataannya dan kuat artinya dapat diterapkan pada data/sampel yang lain dari populasi yang sama. Namun, jika variabel tampak berbentuk ordinal kemudian distribusi datanya miring atau runcing (distribusi tidak normal), maka estimasi ML, kesalahan standar, dan uji chi-kuadrat menghasilkan nilai yang akurat dan kuat. Jika variabel-variabel yang diamati adalah tidak normal, maka varians dari variabel produk dapat sangat berbeda dari nilai-nilai yang ditunjukkan oleh model dasar, dan efek interaksi akan menghasilkan hasil estimasi yang buruk. Lomax (1989) merekomendasikan bahwa teknik estimasi bebas distribusi (distribution free) maupun estimasi yang melibatkan prosedur tertimbang (misalnya, ADF, WLS, GLS) dapat digunakan jika peneliti mendapati distribusi tidak normal. Peneliti juga dapat melakukan transformasi dengan menggunakan transformasi logit atau probit. Transformasi diperbolehkan karena dapat menghasilkan distribusi normal sesuai untuk variabel-variabel yang diamati.

Mengutip buku yang ditulis oleh Raykov & Marcoulides (2006), strategi untuk mengatasi dengan ketidaknormalan data adalah untuk membuat data tampak lebih normal dengan memperkenalkan beberapa strategi normalisasi dengan melakukan transformasi pada data mentah. Setelah data telah diubah sehingga mendekati normal, analisis teori normal dapat dilakukan. Banyak transformasi telah diusulkan dalam literatur, namun yang paling populer adalah (a) transformasi power, seperti kuadrat atau akar kuadrat maupun transformasi timbal timbal balik (*reciprocal transformations*) (b) transformasi logaritma.

Data yang berasal dari desain pengukuran dengan memberikan sedikit alternatif kategori respon dapat menggunakan metode *asymptotically distribution free* yang dapat diwakili oleh korelasi polychoric atau polyserial. Dicontohkan dalam buku Raykov & Marcoulides (2006), kuesioner dengan item, "Seberapa puaskah Anda dengan membeli mobil baru Anda",? Dengan kategori respons berlabel, "Sangat puas", "" Agak puas, dan "Tidak puas". Sejumlah besar penelitian telah menunjukkan bahwa atribut kategoris mengabaikan data yang diperoleh dari aitem seperti ini dapat menyebabkan bias pada hasil SEM yang diperoleh dengan metode standar, misalnya metode yang didasarkan pada minimisasi fungsi sesuai ML biasa. Untuk alasan ini, mereka menyarankan bahwa penggunaan koefisien korelasi-polychoric (untuk menilai derajat asosiasi antara variabel ordinal) dan koefisien korelasi-polyserial (untuk menilai derajat asosiasi antara variabel ordinal dan variabel kontinu) dapat diterapkan, atau sebagai alternatif yang laten disebutkan pendekatan pemodelan variabel di atas untuk analisis data kategorikal dapat digunakan.

Namun demikian, beberapa penelitian juga menunjukkan bahwa ketika kuesioner yang dipakai peneliti memuat lima atau lebih kategori respons, dan distribusi data bisa dilihat menyerupai normal, masalah dari pengabaian sifat kategoris respon yang mungkin relatif sedikit (Rigdon,

1998), terutama jika menggunakan pendekatan Satorra-Bentler robust ML. Oleh karena itu, sekali lagi, pemeriksaan distribusi data menjadi penting dalam pemodelan SEM.

Penanganan Data Tidak Terdistribusi Normal

Metode estimasi yang sering dipakai oleh peneliti yang menggunakan SEM adalah Maximum Likelihood (ML) yang membutuhkan asumsi data memiliki distribusi multivariat normal dengan ukuran sampel 200 atau 10 sampai 20 kali jumlah parameter bebas. Penelitian dengan menggunakan studi simulasi Hox dan Bechger (1998) merangkum beberapa hasil penelitian yang menggunakan studi simulasi telah menemukan kondisi yang dapat mengatasi ketidaknormalan data.

(a) Disarankan peneliti menggunakan ukuran sampel di atas 200. Model yang baik dan data memiliki distribusi multivariat normal biasanya tercapai pada sekitar 200 kasus, meskipun ada beberapa literatur yang menggunakan sampel yang lebih kecil dari 200.

(b) Menggunakan metode estimasi ADF. Jika data yang kontinu akan tetapi tidak normal, metode perkiraan alternatif Asymptotically Distribution Free (ADF), di dalam LISREL dinamakan dengan WLS. Estimasi ADF untuk data tidak normal memerlukan sampel yang sangat besar, biasanya lebih dari seribu kasus.

(c) Mengkoreksi nilai kai-kuadrat. Dengan data yang tidak normal peneliti dapat mengkoreksi nilai statistik kai-kuadrat untuk tingkat non-normal.

(d) Menggunakan Metode Estimasi Maximum Likelihood akan tetapi dengan memperbesar ukuran sampel. Estimasi maximum likelihood masih menghasilkan estimasi yang baik dalam banyak kasus, tapi ukuran sampel yang lebih besar diperlukan, biasanya paling sedikit 400 kasus.

(e) Mengkoreksi nilai Kai-kuadrat dengan formula Satorra-Bentler. Metode ini dipandang sebagai metode yang paling menjanjikan untuk menampung data non-normal.

Jena, 2010
Wahyu Widhiarso

REFERENSI

J.J. Hox & T.M. Bechger (1998). An introduction to structural equation modeling. *Family Science Review*, 11, 354-373.