Pemodelan Persamaan Struktural (SEM) pada Data yang Tidak Normal

Wahyu Widhiarso
Fakultas Psikologi UGM
wahyupsy@ugm.ac.id | Tahun 2012

Pengujian hipotesis dalam pemodelan persamaan struktural (SEM) dapat dibagi menjadi dua kelas besar: (a) uji ketepatan model (*model fit*) secara keseluruhan model dan (b) uji signifikansi nilai estimasi parameter individual. Kedua jenis uji ini mengasumsikan bahwa: (a) model persamaan struktural yang pakai sebagai landasan adalah benar; dan (b) bahwa data yang digunakan untuk menguji model mengikuti distribusi normal multivariat bersama (*joint multivariate normal distribution*/JMVN) pada populasi dimana sampel diambil. Jika data sampel kita tidak memenuhi asumsi JMVN, maka nilai statistik kai-kuadrat yang dipakai dalam menunjukkan ketepatan model fit (model fit) secara keseluruhan akan meningkat dan kesalahan standar (SE) yang dipakai untuk menguji signifikansi parameter akan menurun.

Kita tahu bahwa untuk menguji ketepatan model, kita membutuhkan nilai kai-kuadrat yang tidak signifikan, karena menunjukkan perbedaan yang tidak signifikan antara model dengan data. Jika nilai kai-kuadrat yang dihasilkan terlalu besar maka kita cenderung mendapatkan hasil uji yang signifikan. Di sisi lain, ketika menguji nilai parameter (misalnya koefisien jalur), kita menghipotesiskan bahwa parameter tersebut signifikan karena menunjukkan peranan atau pengaruh yang signifikan. Jika nilai eror standar yang dihasilkan terlalu besar, maka hasil uji signifikansi parameter cenderung akan tidak signifikan.

Jika data kita tidak terdistribusi normal, maka kita lebih cenderung untuk: (a) menolak model yang belum tentu salah, padahal model tersebut harusnya tidak ditolak. (b) Memutuskan bahwa estimasi parameter di dalam model secara statistik signifikan padahal sebenarnya ini tidak signifikan (kasus eror tipe 1) .

Dalam pendekatan SEM masalah asumsi distribusi normal ini tidak hanya menjadi isu pada model analisis faktor konfirmatori saja. Pemodelan lain seperti model pertumbuhan laten, analisis jalur, atau model jenis lainnya yang cocok menggunakan program pemodelan persamaan struktural seperti LISREL, MPLUS, EQS, AMOS, dan PROC Calis dalam SAS juga membutuhkan asumsi normal.

A. Prosedur Menangani Data Tidak Normal

Bagaimana kita mengoreksi distribusi data non-normal dalam SEM? Ada empat pendekatan umum yang digunakan untuk menangani non-normal data:

1. Menggunakan estimator yang tahan terhadap data non normal misalnya, *general least square*/GLS.
2. Menyesuaikan (*adjust*) atau meskalakan penghitungan statistik kai-kuadrat dan eror standar di dalam model.
3. Menggunakan teknik bootstrapping untuk menghitung nilai kritis kai-kuadratd, nilai parameter, dan kesalahan standar.

4. Memadukan indikator dalam sebuah paketan butir

B. Estimator yang Tahan terhadap Ketidaknormalan Data

Generalized Least-Squares. Kebanyakan SEM paket perangkat lunak menawarkan metode estimasi *generalized least-squares* (GLS) selain metode utama, yaitu estimasi *maximum likelihood* (ML) untuk menghitung nilai kai-kuadrat, mengestimasi parameter, dan eror standar. Dalam konteks JMVN, ketika model yang di*fitted*-kan tidak salah maka hasil dari GLS dan ML memiliki nilai yang identik (Bollen, 1989). Penelitian terbaru menunjukkan bahwa GLS memiliki kelemahan relatif dibanding ML dalam hal berikut (Olsson , Troye , & Howell, 1999)

1. GLS lebih sering menerima model yang salah (*false*) daripada ML
2. Parameter GLS seringkali tidak akurat dibanding ML

Sebagai konsekuensi dari ketidakakuratan ini maka indeks modifikasi (MI) yang dihasilkan kurang dapat diandalkan ketika estimator GLS diterapkan. Dengan demikian penggunaan estimator GLS kurang disarankan.

Asymptotic Distribution Free. Selain GLS, ada juga estimator lain yang dapat mengakomodasi distribusi yang tidak normal yaitu *asymptotic distribution free* (ADF) (Browne, 1984). Sayangnya penggunaan ADF membutuhkan ukuran sampel yang sangat besar (lebih dari 1000 kasus). Kekurangan teknik estimasi ini adalah (a) properti asimtotik dari ADF sulit untuk direalisasikan dalam banyak jenis model; (b) memerlukan ukuran sampel yang besar; (c) membutuhkan proses komputasi berat jika model memiliki banyak variabel (Muthén, 1993). Kesimpulan kita adalah ADF mungkin secara teoritis optimal namun kurang praktis.

Weighted-Least Squares. Jika indikator di dalam model kita bersifat kategorikal, maka kita perlu mempertimbangkan untuk menggunakan metode estimasi *weighted-least squares* (WLS). Metode ini merupakan modifikasi dari teknik estimasi ADS. Kelebihan metode ini adalah tidak membutuhkan asumsi distribusi normal. Sayangnya teknik estimasi ini hanya dapat dilakukan melalui program MPLUS. Seperti apakah indikator atau data yang bersifat kategoris itu? Skor butir dari skala Likert, misalnya bergerak antara 0 hingga 4 dapat dimaknai sebagai data kontinum atau kategorikal. Skor tersebut kita interpretasikan sebagai data kontinum jika kita asumsikan skor tersebut merupakan skor interval, dan sebaliknya kita interpretasikan sebagai data kategorikal jika merupakan skor ordinal. Sampai saat ini masih terdapat perdebatan apakah skor skala Likert merupakan skor interval atau ordinal.

C. Penggunaan Penskalaan Kuat (*robust*) dan Penyesuaian Uji Kai-kuadrat

Dalam statistik dikenal istilah *robust* yang artinya kuat atau tahan. Parameter yang *robust* terhadap ketidaknormalan artinya parameter tersebut tahan terhadap data yang tidak normal sehingga hasil estimasinya tetap memiliki ketepatan yang tinggi dalam menjelaskan data.

Teknik kita maksud di sini adalah varian dari pendekatan estimasi ML yang dapat dipakai untuk memperbaiki uji ketepatan model dengan menggunakan statistik kai-kuadrat dan estimasi parameter. Pendekatan ini diperkenalkan oleh

Satorra dan Bentler (1988). Statistik ini juga dikenal sebagai statistik skala T yang dipakai untuk menguji ketepatan model secara keseluruhan. Curran, Barat, dan Finch (1996) menemukan bahwa statistik kai-kuadrat skala mengungguli penggunaan estimator ML standar ketika diterapkan pada data non-normal. Selain EQS, teknik ini digunakan oleh program LISREL dan MPLUS. Teknik ini belum tersedia pada program AMOS. MPLUS juga menawarkan statistik uji serupa yang disebut statistik *mean and variance adjusted kai-kuadrat* yang diestimasi melalui menu MLMV.

Prosedur penyesuaian (*adjusted*) terhadap skala kai-kuadrat disajikan di Bentler dan Dudgeon (1996). Sebuah studi simulasi menemukan bahwa statistik kai-kuadrat yang menggunakan penyesuaian mengungguli penggunaan teknik estimasi ML yang juga menghasilkan nilai kai-kuadrat, terutama dalam sampel dengan ukuran kecil (Fouladi, 1998).Pendekatan statistik kuat bekerja dengan cara menyesuaikan nilai kai-kuadrat dari pemodelan yang dilakukan berdasarkan data sampel yang tidak normal. Biasanya bentuk penyesuaian tersebut adalah menurunkan nilai kai-kuadrat. Semakin besar nilai ketidaknormalan multivariat di dalam data maka semakin besar penyesuaian yang dilakukan. Di sisi lain, nilai eror standar disesuaikan dengan cara ditingkatkan. Tujuannya adalah untuk mengurangi eror tipe 1 pada proses estimasi parameter individu. Meskipun demikian, nilai hasil estimasi parameter melalui penyesuaian sama dengan hasil dari estimasi ML standar.

Bentuk Penyesuaian terhadap Ketidaknormalan Data

Jika pendekatan penskalaan kuat (*robust*) berusaha menyesuaikan nilai kai-kuadrat model berdasarkan nilai normalitas multivariat data, maka ada teknik lain yang juga memilih untuk menyesuaikan nilai kritis uji. Salah satunya adalah teknik *bootstrappping*. Nilai kritis adalah nilai yang dijadikan acuan untuk menentukan hasil uji statistik kita signifikan atau tidak. Jika hasil estimasi melebihi nilai kritis, maka nilai parameter yang diestimasi tersebut signifikan.

Dengan dua asumsi, yaitu data mengikuti JMVN dan model yang dijadikan acuan tidak salah (*false*), maka nilai harapan dari uji kai-kuadrat dari model fit adalah sama dengan derajat kebebasan (*df*) model. Misalnya, jika kita hendak menguji model dengan *df* sebanyak 20, maka kita membandingkannya dengan nilai kai-kuadrat hitung dan nilai kai-kuadrat tabel dengan *df* sebesar 20 juga.

Dalam kondisi data tidak normal, nilai kai-kuadrat dapat melambung dan melebihi *df*, misalnya aslinya 20 akan tetapi menjadi 30. Akibatnya kita peluang untuk mendapatkan model yang fit (hasil uji yang tidak signifikan) menjadi menurun. Prosedur penskalaan kuat (*robust*) dan penyesuaian kuat kai-kuadrat yang telah dijelaskan di atas bertujuan mengoreksi ketidaknormalan data dengan cara menurunkan nilai kai-kuadrat. Dalam banyak kasus nilai kai-kuadrat yang dihasilkan oleh kedua prosedur tersebut seringkali masih di atas nilai *df* model, misalnya *df*=25, padahal idealnya harus sebesar 20 sesuai dengan model awalnya.

D. Teknik *Bootstrapping*

Bootstrapping bekerja dengan cara menghitung nilai kritis uji kai-kuadrat dengan cara yang baru. Melanjutkan contoh di atas, alih-alih menghitung nilai kai-kuadrat JMVN yang diharapkan (yaitu 20), metode *bootstrapping* mungkin menghasilkan nilai *df* sebesar 27. Nilai kai-kuadrat yang dihasilkan pemodelan secara normal (yaitu 30) kemudian dibandingkan dengan nilai kritis yang dihasilkan dari proses bootstrapping (yaitu 27) dan bukan nilai DF model asli (yaitu 20). Nilai signifikansi (p) dihasilkan dari perbandingan dari nilai kai-kuadrat pencocokan model serta nilai kritis kai-kuadrat yang dihasilkan dari proses *bootstrapping*. Penjelasan ini sedikit rumit, akan tetapi anda akan menjadi memahaminya setelah membaca bagian praktek pada program AMOS.

Proses *bootstrapping* dalam menghasilkan nilai kritis

Bagaimanakah proses *bootstrapping* menghasilkan nilai kritis kai-kuadrat? Pertama, data yang dimasukkan diasumsikan merupakan data populasi. Proses *bootstrapping* kemudian menarik sampel dari data tersebut secara berulang ulang. Untuk setiap sampel yang diambil, data ditransformasi untuk mendukung asumsi bahwa pencocokan dengan model yang dilakukan adalah benar. Langkah ini diperlukan karena nilai kai-kuadrat kritis dihitung dari distribusi kai-kuadrat sentral; distribusi kai-kuadrat sentral mengasumsikan bahwa hipotesis nol tidaklah salah (*false*). Asumsi yang sama dibuat ketika kita menggunakan ML standar kai-kuadrat untuk menguji model fit: nilai kai-kuadrat yang didapatkan sama dengan *df* model ketika hipotesis nol tidak ditolak.

Selanjutnya, model dicocokkan dengan data dan nilai kai-kuadrat yang dihitung kemudian disimpan. Proses ini dilakukan secara berulang untuk setiap sampel *bootstrapping*. Di akhir pengambilan sampel *bootstrapping*, program mengumpulkan nilai kai-kuadrat dari setiap sampel dan menghitung nilai rata-ratanya. Nilai rata-rata kai-kuadrat ini menjadi nilai kritis untuk uji kai-kuadrat dari analisis yang asli.

Untuk lebih jelasnya kita bisa langsung membaca tulisan Bollen dan Stine (1993) yang diimplementasikan dalam program AMOS. AMOS memungkinkan analisis data untuk menentukan jumlah sampel bootstrapping diambil (biasanya 250-2000 sampel). Proses ini menghasilkan distribusi kai-kuadrat dari sampel *bootstrapping* seperti halnya nilai rerata kai-kuadrat dan nilai p Bollen dan Stine yang didasarkan pada perbandingan antara kai-kuadrat model asli dan nilai rerata kai-kuadrat dari sampel *bootstrapping*.

Metode *bootstrapping* juga mampu menghasilkan nilai eror standar. Nilai eror standar ini penting karena menentukan apakah sebuah parameter yang kita uji signifikan ataukah tidak. Kelemahan metode bootstrapping adalah metode ini memerlukan data yang lengkap, alias tidak ada kasus hilang didalamnya.

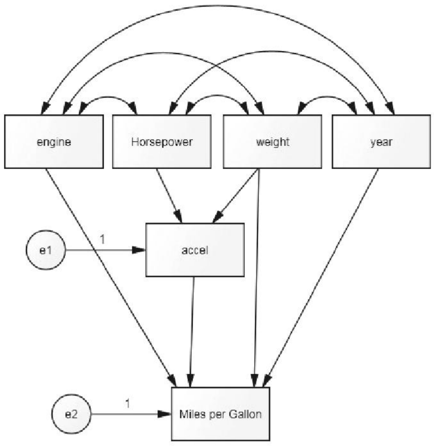
E. Praktek *Bootstrapping* di AMOS

Ada tiga langkah yang dapat diambil ketika data kita tidak terdistribusi normal:

1. Pastikan bahwa variabel di dalam model tidak berdistribusi normal multivariat bersama (*joint multivariate normal*)
2. Lakukan uji ketepatan model secara keseluruhan dengan menggunakan nilai p Bollen-Stine yang telah dikoreksi
3. Gunakan metode bootstrapping untuk menghasilkan nilai parameter, eror standar parameter, dan signifikansi parameter tersebut.

Setiap tahap dijelaskan lebih rinci pada bagian berikut.

Langkah pertama dalam menangani non-normal data sampel adalah memastikan bahwa itu adalah non-normal. Kali ini kita menggunakan file SPSS yang telah tersedia di dalam folder instalasi SPSS, yang bernama *cars.sav*. Jika kita membuang kasus yang tidak lengkap maka data kita berisi 392 kasus. Misalkan kita berencana untuk melakukan pemodelan mengenai mobil.



Gambar 1. Model yang diuji

AMOS dapat memberikan informasi mengenai nilai *skewness* dan *kurtosis* univariat dari setiap variabel di dalam model, serta *joint multivariate kurtosis.* Untuk meminta agar statistik ini dimasukkan dalam output AMOS , pilih:

Klik tab Output dan kemudian memeriksa Pengujian normalitas dan outlier kotak cek. Juga memeriksa perkiraan Standar dan Squared beberapa korelasi tab.

Assessment of normality (Group number 1)

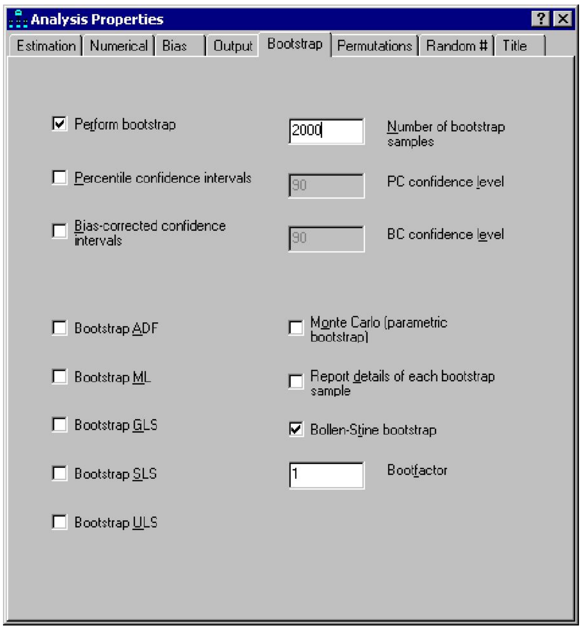
Variable	min	max	skew	c.r.	kurtosis	c.r.
engine	4.000	455.000	.697	5.632	-.764	-3.086
weight	732.000	5140.000	.479	3.873	-.744	-3.007
year	.000	82.000	-7.385	-59.692	103.789	419.459
horse	46.000	230.000	1.093	8.838	.728	2.943
accel	8.000	24.800	.267	2.161	.444	1.794
mpg	9.000	46.600	.455	3.680	-.525	-2.121
Multivariate					155.700	157.314

Gambar 2. Hasil Uji Normalitas

Nilai-nilai kritis (*c.r*) yang melebihi rentang -2.00 hingga 2,00 menunjukkan nilai distribusi yang tidak normal. AMOS juga melaporkan nilai kurtosis multivariat dan nilai rasionya. Secara praktis nilai kurtosis multivariat diharapkan sekecil mungkin (misalnya, kurang dari 1,00) sedangkan nilai kurtosis multivariat antara 1 hingga 10 menunjukkan ketidaknormalan yang moderat. Nilai yang melebihi 10 menunjukkan ketidaknormalan yang parah. Dapat disimpulkan bahwa semua variabel dalam penelitian ini terdistribusi tidak normal.

***Bootstrapping* Bollen-Stine dan Uji terkait Ketepatan Model**

Di dalam program AMOS, untuk melakukan uji Bollen-Stine, klik: VIEW – ANALYSIS PROPERTIES. Lalu pilih tab *Bootstrap* dan centang PERFORM BOOTSTRAP. Tentukan jumlah sampel bootstrapping untuk menghitung nilai p Bollen-Stine. Contoh yang ditampilkan di gambar adalah 2000 sampel.



Gambar 3. Perintah Bootstrap di AMOS

Output dari *bootstrapping* Bollen - Stine ini dibagi menjadi tiga bagian. Bagian pertama berisi informasi diagnostik. Jika (a) solusi tidak ditemukan pada sampel bootstrapping tertentu atau (b) AMOS tidak dapat mencocokkan model tersebut dengan sampel bootstrapping diberikan karena munculnya matriks kovarians yang singular, maka AMOS akan menarik sampel pengganti untuk memastikan bahwa hasil akhir yang didapatkan didasarkan pada jumlah aktual sampel yang digunakan seperti yang kita minta (N=2000).

Iterations	Method 0	Method 1	Method 2
1	0	0	0
2	0	0	0
3	0	0	9
4	0	0	87
5	0	0	129
6	0	1	99
7	0	16	11
8	0	160	0
9	0	267	0
10	0	273	0
11	0	408	0
12	0	336	0
13	0	136	0
14	0	52	0
15	0	11	0
16	0	4	0
17	0	1	0
18	0	0	0
19	0	0	0
Total	0	1665	335

0 bootstrap samples were unused because of a singular covariance matrix.
0 bootstrap samples were unused because a solution was not found.
2000 usable bootstrap samples were obtained.

Gambar 4. Perbandingan Metode Proses Bootstrap

Di bagian bawah tampilan output, kita dapat melihat bahwa tidak ada sampel yang dibuang karena alasan matriks kovarians yang singular. Namun misalnya, AMOS membuang beberapa sampel bootstrapping maka kita harus memeriksa spesifikasi model yang kita kembangkan dan mengulangi analisis dari awal.

Proses Minimisasi pada AMOS

AMOS memiliki fleksibilitas karena dapat menggunakan salah satu dari dua metode yang berbeda untuk melakukan minimisasi (*minimization*) selama proses *bootstrapping*. Metode 0 dapat mengatasi masalah dengan cepat ketika model yang kita kembangkan sederhana dan menjadi lambat ketika model yang kita kembangkan cukup kompleks. Namun Metode 0 ini belum tersedia di AMOS, jadi Metode 0 selalu akan berisi nilai nol untuk semua baris. Metode 1 menggunakan algoritma yang cepat dan dapat diandalkan. AMOS pertama kali melakukan minimisasi dengan menggunakan ini.

Jika proses minimisasi pada Metode 1 terlalu sulit untuk menghasilkan sampel bootstrapping maka AMOS akan beralih ke Metode 2. Meski lebih reliabel Model 2 lebih memakan waktu daripada Metode 1. Setiap kolom metode ini berisi daftar jumlah sampel yang berhasil dibangkitkan oleh AMOS dari beberapa proses iterasi.

Sebagai contoh, pada iterasi ke 6 dari 2000 sampel bootstrapping yang dibangkitkan, pada Metode 1 iterasi ke 6, AMOS berhasil mendapatkan 1 sampel. Sebaliknya, ketika metode diganti menjadi Metode 2 didapatkan 99 sampel. berkumpul hanya dalam empat iterasi ketika AMOS beralih ke Metode 2. Metode 1 menghasilkan 1665 sampel sedangkan Metode 2 menghasilkan 335 sampel, total sampel ada 2000 seperti yang kita minta sebelumnya.

Bagian kedua dari output menampilkan nilai p untuk uji hipotesis model fit secara keseluruhan.

Bollen-Stine Bootstrap (Default model)

The model fit better in 1760 bootstrap samples.
It fit about equally well in 0 bootstrap samples.
It fit worse or failed to fit in 240 bootstrap samples.
Testing the null hypothesis that the model is correct, Bollen-Stine bootstrap p = .120

Gambar 5. Hasil Uji Kai-Kuadrat dalam Proses Bootstrap

Sebelumnya kita minta AMOS untuk membangkitkan sampel bootstrapping sebesar 2000. Dari hasil proses pembangkitan kita dapatkan bahwa ada 240 sampel yang tidak mendukung model yang kembangkan. Dari hasil ini kita dapatkan nilai p Bollen-Stine adalah 240/2000=0,120, yang memperoleh nilai p ketepatan model secara keseluruhan. Dengan menggunakan kriteria signifikansi konvensional (p=0,05) kita dapatkan bahwa nilai p yang dihasilkan oleh model kita adalah lebih dari 0,05 ($p>0,05$). Dengan demikian kita simpulkan bahwa model yang kita kembangkan didukung atau sesuai data.

Bandingkan hasil ini dengan hasil ketika kita menggunakan metode ML (lihat gambar di bawah ini). Hasil analisis menunjukkan bahwa model kita memiliki ketepatan yang lemah karena nilai p yang dihasilkan lebih kecil dari 0,05.

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	18	10.061	3	.018	3.354
Saturated model	21	.000	0		
Independence model	6	2510.587	15	.000	167.372

Gambar 5. Hasil Uji Kai-Kuadrat Tanpa Bootstrap

Bagian terakhir dari output *bootstrapping* Bollen-Stine adalah penjelasan mengenai distribusi nilai kai-kuadrat yang diperoleh pada 2000 sampel *bootstrapping*.

ML discrepancy (implied vs sample) (Default model)

```
                          |--------------------
                    .003  |*********
                   1.855  |********************
                   3.706  |***************
                   5.557  |**********
                   7.409  |******
                   9.260  |*****
                  11.112  |****
N = 2000          12.963  |**
Mean = 4.943      14.814  |**
S. e. = .097      16.666  |*
                  18.517  |*
                  20.368  |*
                  22.220  |*
                  24.071  |*
                  25.923  |*
                          |--------------------
```

Gambar 6. Sebaran Kai-Kuadrat dalam Proses Bootstrap

Fitur utama dari output ini adalah nilai kai-kuadrat rata-rata dan bentuk distribusi nilai kai-kuadrat. Dari 2000 sampel yang dibangkitkan, nilai rerata kai-kuadrat yang dihasilkan adalah 4,77. Nilai ini lebih tinggi dari nilai *df* dalam kondisi normal multivariat (JMVN) yang bernilai 3 (lihat gambar 5). Rerata nilai kai-kuadrat dari sampel *bootstrapping* berfungsi sebagai nilai kritis kai-kuadrat terhadap nilai kai-kuadrat dalam kondisi normal, yaitu 10,061. Ketika kai-kuadrat model dibandingkan dengan dengan nilai kritis 4,77, maka nilai p yang dihasilkan adalah p=0,120, alias tidak signifikan. Sebaliknya ketika memperoleh membandingkan nilai kai-kuadrat 10,061 dengan nilai kritis df=3,00 maka nilai p yang dihasilkan adalah p=0,018 alias signifikan, alias model kita tidak fit.

	Probability less than the critical value				
ν	0.90	0.95	0.975	0.99	0.999
1	2.706	3.841	5.024	6.635	10.828
2	4.605	5.991	7.378	9.210	13.816
3	6.251	7.815	9.348	11.345	16.266
4	7.779	9.488	11.143	13.277	18.467
5	9.236	11.070	12.833	15.086	20.515
6	10.645	12.592	14.449	16.812	22.458
7	12.017	14.067	16.013	18.475	24.322

Gambar 7. Tabel Kai-Kuadrat

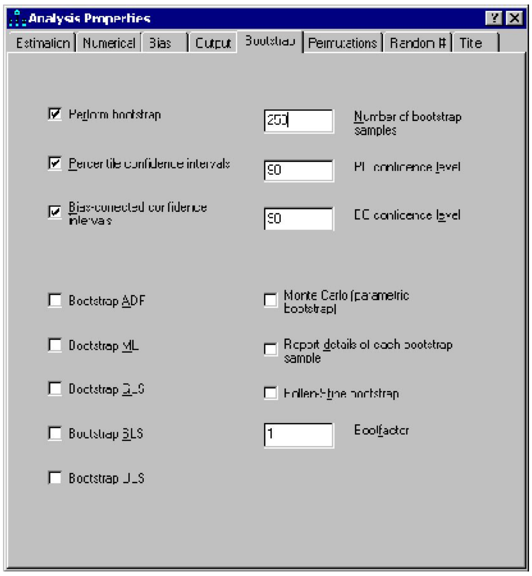
Sekilas tentang Kai-Kuadrat Tabel

Mengapa jika df kita 3 menghasilkan p yang signifikan sedangkan jika df kita 4,94 menghasilkan p yang tidak signifikan? Lihat gambar 7 pada kolom tingkat kepercayaan 0,95. Nilai kritis kai-kuadrat (kai-kuadrat tabel) dengan df=3 adalah 9,348. Jika kai-kuadrat hitung yang kita dapatkan di atas 9,348 maka hasil uji yang kita dapatkan adalah signifikan. Ingat prinsip ini: "*Nilai statistik yang bisa mengalahkan tabel maka hasil ujinya signifikan*". Apapun itu, baik uji komparasi atau korelasi, jika hasil penghitungan statistik bisa mengalahkan tabel maka hasil uji signifikan.

Kembali kepada tabel kai-kuadrat. Proses *bootstrapping* yang kita lakukan menghasilkan nilai kritis sebesar 4,943, kita bulatkan saja menjadi 5. Nilai kai-kuadrat tabel dengan df=5 adalah 11,070. Jika dibandingkan dengan nilai kai-kuadrat model kita yang besarnya adalah 10,061, dapat disimpulkan bahwa nilai kai-hitung yang kita dapatkan tidak bisa mengalahkan tabel. Kesimpulannya tidak ada perbedaan yang signifikan antara model dengan data kita. Dengan kata lain model kita fit.

Bootstrap pada Parameter

Setelah kita mendapatkan model memiliki indeks ketepatan model yang memuaskan secara keseluruhan, pertanyaan perlu dijawab adalah: Apakah koefisien jalur yang dihasilkan secara statistik signifikan? AMOS menyediakan berbagai pilihan *bootstrapping* untuk menjawab pertanyaan ini. Sayangnya, kita tidak dapat memperoleh hasil estimasi parameter *bootstrapping* dan eror standar terkait dengan nilai p Bollen-Stine. Oleh karena itu kita harus mengulangi dari awal, masuk lagi ke tab BOOTSTRAP pada menu ANALYSIS PROPERTIES.



Gambar 8. Perintah untuk Menjalankan Bootstrap pada Parameter

Kali ini berikan tanda centang hanya pada PERFORM BOOTSTRAP, PERCENTILE CONFIDENCE INTERVAL (PC) dan BIAS CORRECTED CONFIDENCE INTERVAL (BC). Jumlah sampel *bootstrapping* yang kita inginkan adalah 250 yang sesuai dengan rekomendasi dari Nevitt dan Hancock

(1998). Mereka menemukan bahwa ukuran sampel proses *bootstrapping* yang besar tidak membawa banyak perbaikan. Namun demikian jika kita berencana untuk menafsirkan nilai probabilitas (p) maka kita perlu ukuran sampel yang lebih besar (misalnya 2000) untuk memastikan perkiraan probabilitas stabil.

Hasil Estimasi dengan Prosedur Biasa

Hasil analisis yang perlu diperhatikan ada pada Gambar 9. Setiap parameter yang diestimasi di dalam model akan diestimasi pula oleh proses bootstrapping, misalnya koefisien regresi (path), varians, kovarians, rerata dan intesep. Kita akan membandingkan keluaran analisis dari proses normal (*maximum likelihood*) dan hasil *bootstrapping*.

Covariances: (Group number 1 - Default model)

			Estimate	S.E.	C.R.	P	Label
horse	<-->	engine	3585.440	271.730	13.195	***	
horse	<-->	weight	27860.389	2164.965	12.869	***	
year	<-->	weight	-504.570	229.771	-2.196	.028	
weight	<-->	engine	83310.047	6172.877	13.496	***	
horse	<-->	year	-55.358	10.619	-5.213	***	
year	<-->	engine	-103.767	28.601	-3.628	***	

Gambar 9. Hasil Estimasi Parameter tanpa *Bootstrapping*

Semua nilai kovarians yang dihasilkan adalah signifikan. Tanda *** menunjukkan signifikan pada taraf di bawah 1%. Nilai C.R yang merupakan singkatan dari critical ratio atau nilai kritis, didapatkan dari nilai estimasi dibagi dengan nilai eror standarnya (S.E). Nilai kritis yang di luar rentang -1,96 s/d 1,96 menghasilkan nilai p yang signifikan.

Hasil Estimasi dengan Prosedur Bootstrap

Pada bagian *bootstrapping* dijelaskan *output* yang berisi rata-rata estimasi parameter dari berbagai sampel *bootstrapping*. Perbedaan antara hasil estimasi normal (berbasis *maximum likelihood*) dan hasil estimasi berbasis *bootstrapping* ditampilkan dalam kolom BIAS (Gambar 10). Nilai bias yang besar, seperti yang terjadi di sini, menunjukkan perbedaan besar antara hasil analisis *bootstrapping* dengan hasil analisis yang mengasumsikan data terdistribusi normal.

Parameter			SE	SE-SE	Mean	Bias	SE-Bias
horse	<-->	engine	247.590	11.073	3539.587	-45.854	15.659
horse	<-->	weight	1796.482	80.341	27586.892	-273.497	113.620
year	<-->	weight	403.171	18.030	-542.551	-37.982	25.499
weight	<-->	engine	4531.894	202.672	82420.186	-889.861	286.622
horse	<-->	year	7.254	.324	-55.117	.241	.459
year	<-->	engine	36.935	1.652	-106.141	-2.374	2.336

Gambar 10. Hasil Estimasi Parameter dengan *Bootstrapping*

Kita dapat menggunakan kolom *Mean* dan SE untuk menghitung nilai rasio kritis berdasarkan hasil *bootstrapping*. Sebagai contoh, perhatikan hasil uji hipotesis nol bahwa kovarians antara WEIGHT dan YEAR di dalam gambar adalah nol. Perkiraan nilai parameter rata-rata dari 250 sampel *bootstrapping* adalah -542.551 dengan eror standar estimasi (SE) sebesar 403.171. Dari kedua nilai ini kita dapat menghitung nilai kritis estimasi parameter dengan cara membagi nilai estimasi parameter *bootstrapping* dengan eror standardnya (-542.551/403.171). Nilai kritis yang dihasilkan adalah -1.345. Nilai ini berada dalam rentang -1,96 s/d 1,96, sehingga kita simpulkan secara statistik tidak signifikan.

AMOS tidak menampilkan nilai signifikansi (p) untuk uji seperti ini. Oleh karena itu kita menghitungnya secara manual. Cara lainnya adalah kita melihat pada bagian uji hipotesis persentil-terkoreksi (*bias-corrected percentile-corrected*) yang bisa dilihat pada Gambar 11. Pada bagian ini nilai P dikeluarkan oleh program AMOS. Terlihat bahwa nilai signifikansi kovarian antara WEIGHT dan YEAR adalah p=0,311 alias tidak signifikan dan konsisten dengan hasil pada BOOTSTRAP STANDAR ERROR (Gambar 10).

Parameter	Estimate	Lower	Upper	P
engine <--> year	-103.767	-165.971	-29.732	.018
engine <--> weight	83310.047	75164.687	92290.261	.006
engine <--> horse	3585.440	3183.222	4121.071	.005
year <--> horse	-55.358	-71.897	-44.760	.003
horse <--> weight	27860.389	24869.698	31760.497	.007
year <--> weight	-504.570	-1137.799	313.071	.311

Gambar 11. Hasil Estimasi Parameter tanpa *Bootstrapping*

Dua metode yang dipakai di sini yaitu *percentile confidence interval* dan *bias corrected confidence interval* menghasilkan kesimpulan yang sama. Tidak ada metode yang terbaik dan digunakan dalam semua situasi analisis data. Namun yang direkomendasikan adalah kita diharapkan melaporkan nilai hasil dari proses *bootstrapping* dalam bentuk interval kepercayaan (*lower* and *upper*) karena akan memberikan informasi yang lebih akurat.

Hal ini dimungkinkan untuk memiliki interval kepercayaan atas dan bawah bias-dikoreksi yang tidak termasuk nol , namun memiliki nilai p yang tidak signifikan secara statistik. Hal ini dikarenakan nilai p dihitung lepas dari interval kepercayaan. Pada awal analisis di program AMOS kita bisa memilih persentase interval kepercayaan, dari 90% sampai 95%. Perubahan persentase interval kepercayaan, misalnya dari 90% menjadi 95 akan mengubah nilai interval konfidensi akan tetapi tidak mengubah nilai p yang dihasilkan oleh PC dan BC.

Referensi

Bentler, P. M., & Dudgeon, P. (1996). Covariance structure analysis: Statistical practice, theory, and directions. *Annual Review of Psychology, 47,* 563-592.

Bollen, K. A. (1989). Structural equations with latent variables. New York, NY: John Wiley and Sons.

Bollen, K. A., & Stine, R. A. (1993). Bootstrapping goodness-of-fit measures in structural equation models. In K. A. Bollen and J. S. Long (Eds.) Testing structural equation models. Newbury Park, CA: Sage Publications.

Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology, 37,* 62-83.

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods, 1,* 16-29.

Fouladi, R. T. (1998). Covariance structure analysis techniques under conditions of multivariate normality and nonnormality - Modified and bootstrap test statistics. Paper presented at the American Educational Research Association Annual Meeting, April 11-17, 1998, San Diego, CA.

Muthén, B. O. (1993). Goodness of fit with categorical and other nonnormal variables. In K. A. Bollen and J. S. Long (Eds.) Testing structural equation models. Newbury Park, CA: Sage Publications.

Nevitt, J., & Hancock, G. R. (1998). Relative performance of rescaling and resampling approaches to model chi-square and parameter standard error estimation in structural equation modeling. Paper presented at the American Educational Research Association Annual Meeting, April 11-17, 1998, San Diego, CA.

Olsson, U. H., Troye, S. V., & Howell, R. D. (1999). Theoretic fit and empirical fit: The performance of maximum likelihood versus generalized least squares estimation in structural equation models. Multivariate Behavioral Research, 34(1), 31-59.

Satorra, A., & Bentler, P. M. (1988). Scaling corrections for chi-square statistics in covariance structure analysis. 1988 Proceedings of the Business and Economics Statistics Section of the American Statistical Association, 308-313.