4.7 TRANSFORMASI UNTUK MENDEKATI KENORMALAN

Saat asumsi kenormalan tidak dipnuhi maka kesimpulan yang kita buat berdasarkan suatu metod statistik yang mensyaratkan asumsi kenormalan menjadi tidak baik, sehingga muncul upaya untuk membuat kesimpulan menjadi lebih baik yaitu dengan mentransformasikan data shingga mendekati kenormalan.

Transformasi data ini disarankan berdasrkan pertimbanagn teori atau bentuk dta itu sendiri (dapat juga karena kedeanya). Secara teori, suatu data dapat dihitung sehingga mendekati distribusi normal dengan transformasi berikut:

Skala asli	Skala hasil transsformasi
Perhitungan, y	$\sqrt{y}$
Proporsi, $\hat{p}$	Logit ($\hat{p}$) = $\frac{1}{2}\log\left(\frac{\hat{p}}{1-p}\right)$
Korelasi, r	Fisher's z(r) = $\frac{1}{2}\log\left(\frac{1+r}{1-r}\right)$

Dalam banyak kasus, pemilihan transormasi yang tepat tidaklah mudah sehingga ada beberapa teknik untuk menentukannnya, diantaranya adalah sebagai berikut:

A. <u>*Penentuan Transformasi Hanya Berdasarkan Bentuk Data*</u>

Himpunan transformasi yang sangat berguna dalam kasus ini iebut 'himpunan kekuatan transormasi'. Kekuatan transformasi ini didiefinisikan hanya untuk variabel bernilai positif, meskipun demikian kita dapat menambahkan konstanta tunggal terhadap kasus yang memuat variabel negatif, sehingga kekuatan transformasi tetap dapat digunakan.

Misalkan x merupakan observasi acak. Himpunan kekuatan transformasi ditunjukkan oleh parameter λ. Nilai yang diberikan untuk λ menyatakan sebuah transformasi utama. Sebagai contoh, perhatikan x^{λ} dengan λ= -1. Karena x^{-1} = 1/x, pemilihan dari λ ini berkorespondebnsi secara timbal balik dengan transformasi. Kita dapat mencari himpunan transformasi dengan menerapkan rentang λ untuk kekuatan x dari yang negatif sampai yang positif. Untuk $\lambda = 0$, kita definisikan $x^0 = \ln x$. barisan tranformasi yang mungkin adalah:

$$\ldots, x^{-1} = 1/x,\ x^0 = \ln x,\ x^{1/4} = \sqrt[4]{x},\ x^{1/2} = \sqrt{x},\ x^2,\ x^3, \ldots$$

Saat data dalam bentuk diagram titik marjinal atau histogram nilainya terlalu lebar, maka kita harus menyusutkannya (anaog untuk keadaan sebaliknya). Hal ini akan dapat memperbaiki tingkat kesimetrisan data. Setelah kita mencoba bentuk transformasi yang sesuai, maka hasil transformasinya tetap harus diuji kenormalannya dengan membuatnya dalam Q-Q plot, jika masih belum normal kita dapat mencoba bentuk transformasi yang lain hingga ditemukan yang paling sesuai.

B. *Penentuan Transformasi Berdasarkan Informasi dari Data dan Faktor-Faktor di Luar Data (seperti penyederhanaan atau interpretasi yang dimudahkan)*

Metode analisis sederhana tersedia untuk memilih kkuatan transformasi. Kita awali denngan memfokuskan perhatian kita pada kasus univariat.

Box dan Cox memperhatikan himpunan kekuatan transformasi yang sedikit dimodifikasi berikut:

$$X(\lambda) = \frac{x^{\lambda} - 1}{\lambda} \qquad \lambda \neq 0 \qquad (5)$$

$$= \ln \lambda \qquad \lambda = 0$$

Dengan nilai λ kontinu untuk $x > 0$. Diberikan observasi $X_1, X_2, \ldots, X_n$, penyelesaian Box-Cox memilih pendekatan kekuatan transformasi λ dengan mencari nilai λ yang memaksimumkan fungsi

$$\ell(\lambda) = -\frac{n}{2} \ln\left[\frac{1}{n}\sum_{j=1}^{n}\left(x_j^{(\lambda)} - \overline{x}^{(\lambda)}\right)^2\right] + (\lambda - 1)\sum_{j=1}^{n} \ln x_j \qquad (6)$$

Catat bahwa $x_j^{(\lambda)}$ adalah seperti yang didefinisikan pada (5) dan

$$\overline{x^{(\lambda)}} = \frac{1}{n}\sum_{j=1}^{n} x_j^{(\lambda)} = \frac{1}{n}\sum_{j=1}^{n}\left(\frac{x_j^{(\lambda)} - 1}{\lambda}\right) \tag{7}$$

Adalah rata-rata observasi yang ditransformasi.

Setelah kita memakimumkan (6) dengan memperhatikan parameter rata-rata dan variansi populasi, maka suku pertama dari (6) merupakan konstanta, yaitu logaritma natural dari fungsi likelihood normal.

Perhitungan $\ell(\lambda)$ untuk nilai λ yang banyak lebih baik dikerjakan dengan menggunakan komputer. Dengan mengggunakan komputer kita dapat membuat grafik $\ell(\lambda)$ vs λ sebaik penampilan tabel pasangan (λ, $\ell(\lambda)$), untuk mempelajari tingkah laku (sifat-sifat) yang dekat dengan nilai maksimum $\hat{\lambda}$. Singkatnya, jika $\lambda = 0$ (logaritma normal) atau $\lambda = ½$ (akar kuadrat) mendekati $\hat{\lambda}$, maka salah satunya harus dipilih karena kesederhanaan bentuk transformasinya.

Selain menggunakan metode perhitungan (6), beberapa statikawan menyarankan prosedur yang serupa dengan menentukan nilai $\hat{\lambda}$, dengan membuat variabel yang baru

$$y_j^{(\lambda)} = \frac{x_j^{\lambda} - 1}{\lambda\left[\left(\prod_{i=1}^{n} x_i\right)^{1/n}\right]^{\lambda-1}} \qquad j = 1, 2, 3, \ldots$$

Dan kemudian menghitung varians sampel saat terjadi variansi yang minimum, maka λ juga akan maksimum.

Komentar:

Kini telah diketahui bahwa transformasi yang dihasilkan dengan memaksimumkan $\ell(\lambda)$ dapat memperbaiki (mendekati) kenormalan. Bagaimanapun juga tidak ada jaminan bahwa pmilihan λ yang terbaik akan menghasilkan himpunan yang ditransformasi yang cukup sesuai dengan istribusi normal. Hasil yang diperoleh dari transformasi berdasarkan (6) harus tetap diuji untuk mengetahui ada atau tidaknya pelanggaran asumsi kenormalan sementara. Hal ini mengimplikasikan bahwa ada teknik transformasi lain dengan kekuatan yang sama yang bisa digunakan untuk mendekati kenormalan.

Setelah memahami kasus untuk univariat, sekarang kita akan mengembangkannya untuk kasus multivariat, kekuatan transormasi haruslah dipilih untuk masing-masing variabel. Misalkan $\lambda_1, \lambda_2, \lambda_3, \ldots \lambda_n$ adalah kekuatan transformasi dari p karakteristik yang diukur. Masing-masing λ_k dapat dipilih dengan memaksimumkan

$$\ell_{k(\lambda)} = -\frac{n}{2}\ln\left[\frac{1}{n}\sum_{j=1}^{n}\left(x_{kj}^{(\lambda k)} - \overline{x_k^{(\lambda k)}}\right)^2\right] + (\lambda_k - 1)\sum_{j=1}^{n}\ln x_{kj} \qquad (9)$$

Dengan $x_{k1}, x_{k2}, \ldots, x_{kn}$ merupakan n observasi pada variabel ke k, k = 1, 2, …,p . Disini

$$x^{(\lambda_k)} = \frac{1}{n}\sum_{j=1}^{n} x_{kj}^{(\lambda_k)} = \frac{1}{n}\sum_{j=1}^{n}\left(x_{kj}^{\lambda_k}\right) \qquad (10)$$

Merupakan rata-rata aritmatik dari observasi yang telah ditransformasi. Observasi multivariat ke-j yang telah ditransformasi adalah

$$x_k^{(\hat{\lambda})} = \begin{bmatrix} \dfrac{x_{1j}^{\hat{\lambda}_1} - 1}{\hat{\lambda}_1} \\ \dfrac{x_{2j}^{\hat{\lambda}_1} - 1}{\hat{\lambda}_2} \\ \vdots \\ \dfrac{x_{nj}^{\hat{\lambda}_1} - 1}{\hat{\lambda}_n} \end{bmatrix}$$

Dimana $\hat{\lambda}_1, \hat{\lambda}_2, ..., \hat{\lambda}_p$ adalah nilai yang memaksimumkan persamaan (9) secara individu.

Prosedur transformasi untuk mendekati kenormalan pada observasi multivariat adalah sepadan dengan prosedur tranformasi untuk mendekati kenormalan pada tiap distribusi marjinalnya. Meskipun kenormalan marjinal tidaklah cukup menjamin kenormalan dari diatribusi gabungannya, namun dalam penerapan praktisnya hal ini sudah cukup baik. Jika tidak, kita harus memulainya dengan nilai $\hat{\lambda}_1 \hat{\lambda}_2, ..., \hat{\lambda}_p$ yang diperoleh dengan menggunakan prosedur di atas dan melakukan iterasi hingga ditemukan himpunan dari nilai λ' = [$\hat{\lambda}_1 \hat{\lambda}_2, ..., \hat{\lambda}_p$], yang secara bersama-sama memaksimumkan

$\ell(\hat{\lambda}_1 \hat{\lambda}_2, ..., \hat{\lambda}_p)=$

$$-\frac{n}{2}\ln\left|S(\lambda)\right| + (\lambda_1 - 1)\sum_{j=1}^{n} \ln x_{ij} + (\lambda_2 - 1)\sum_{j=1}^{n} \ln x_{2\lambda} + ... + (\lambda_p - 1)\ln x_{pj} \quad (11)$$

Dimana S(λ) adalah matriks kovarian yang dihitung dari

$$x_k^{(\hat{\lambda})} = \begin{bmatrix} \dfrac{x_{1j}^{\hat{\lambda}_1} - 1}{\hat{\lambda}_1} \\ \dfrac{x_{2j}^{\hat{\lambda}_1} - 1}{\hat{\lambda}_2} \\ \vdots \\ \dfrac{x_{nj}^{\hat{\lambda}_1} - 1}{\hat{\lambda}_n} \end{bmatrix} \qquad J=1, 2, \ldots, n$$

Memakimumkan persamaan (11) pada hakikatnya tidaklah lebih sulit dari memaksimumkan persamaan tersebut secara inividu yang dapat dilihat pada persamaan (9), hal ini juga tidak memberikan hasil yang benar-benar lebih baik. Terdapat metode pilihan yang berdasarkan pada persamaan (11) dan sepadan dengan memaksimumkan fungsi likelihood multivariat atas μ, $\sum$, dan λ, ingat kembali bahwa metode yang berdasarkan persamaan (9) berkorespondensi dengan memaksimumkan fungsi likelihood univariat ke-k atas μ_k, σ_{kk}, dan $\lambda_{k.}$ fungsi likelihood dibangun atas dasar anggapan bahwa terdapat beberapa $\lambda_{k.}$ yang dapat membuat observasi $x_j^{\lambda} = \frac{x_{kj}^{\lambda k} - 1}{\lambda_k}, j = 1,2,...,n$ berdistribusi normal

Contoh :

Perhatikan kembali kasus pada contoh soal nomer 1, telah kita simpulkan bahwa observasi tersbut tidak berdistribusi normal, sehingga kita dapat mencoba mentransformasikannya supaya mendekati normal.

Ita akan menentukan kekuatan transformasi (λ) dari λ = -1,0 sampai dengan λ=1,5, dan menggunakan nilai λ yang memakimumkan fungsi

$$\ell_{k(\lambda)} = -\frac{n}{2}\ln\left[\frac{1}{n}\sum_{j=1}^{n}\left(x_{kj}^{(\lambda k)} - \overline{x_k^{(\lambda k)}}\right)^2\right] + (\lambda_k - 1)\sum_{j=1}^{n}\ln x_{kj}$$

Dengan

$$X(\lambda) = \frac{x^{\lambda} - 1}{\lambda} \qquad \lambda \neq 0$$

$$= ln\,\lambda \qquad \lambda = 0$$

Dan

$$\overline{x^{(\lambda)}} = \frac{1}{n}\sum_{j=1}^{n} x_j^{(\lambda)} = \frac{1}{n}\sum_{j=1}^{n}\left(\frac{x_j^{(\lambda)} - 1}{\lambda}\right)$$

Untuk itu kita akan mencari nilai $\ell(\lambda)$ untuk λ = -1,0 sampai dengan λ = 1,5. Kita hanya akan menampilkan perhitungan untuk λ = -1,0, sedangkan untuk nilai λ yang lain dikerjakan dengan cara yang serupa dan hasilnya akan itampilkan dalam tabel.

Untuk λ = -1,0 dengan bantuan program excell diperoleh

$$\sum_{j=1}^{n}\left(x_j^{-1,0}\right)^2 = 20249{,}30$$

$$\sum_{j=1}^{42} lnx_j = -100{,}13$$

Sehingga

$$\ell(-1{,}0) = -\frac{42}{2}\ln\left[\frac{1}{42}\sum_{j=1}^{42}\left(x_{kj}^{(-1,0)} - \overline{x_k^{(-1,0)}}\right)^2\right] + (-1{,}0-1)\sum_{j=1}^{42}\ln x_j$$

= -21 ln[1/42 (20249,30)] + (-2)(-100,13)

= -21 ln (482,13) + 200,26

= -21 (6, 1782) + 200,26

= 70,52

Berikut disajikan nilai ℓ(λ) untuk λ = -1,0 sampai dengan λ = 1,5

λ	ℓ(λ)	λ	ℓ(λ)	λ	ℓ(λ)
-1.0	70.52	-0.1	103.35	0.8	101.33
-0.9	75.65	0.0	104.83	0.9	99.34
-0.8	80.46	0.1	105.84	1.0	97.10
-0.7	84.94	0.2	106.39	1.1	94.64
-0.6	89.06	0.3	106.51	1.2	91.96
-0.5	92.79	0.4	106.20	1.3	89.10
-0.4	96.10	0.5	105.50	1.4	86.07
-0.3	98.97	0.6	104.43	1.5	82.88
-0.2	101.39	0.7	103.03		

Berdasarkan hasil di atas, dapat kita lihat bahwa pada saat λ ekitar 0.3 nilai $\ell(\lambda)$ menjadi maksimum. Kita akan memilih menggunakan λ = 0.25 untuk mentransformasikan data pada soal. Sehingga data observasi x_j, j= 1, 2, …, 42 akan kita ubah menjadi x^{λ}_j dan diharapkan akan berdistribusi normal.

Berikut isajikan nilai dari x_j dan $x^{\lambda}_{j\ ,\ j\ =\ 1,}$ 2, …, 42 ; λ = 0.25

No. oven	Radiasi (x_j)	$x_j^{\lambda} = \frac{x_j^{0.25} - 1}{0.25}$	N0. Oven	Radiasi (x_j)	$x_j^{\lambda} = \frac{x_j^{0.25} - 1}{0.25}$
1	0.15	-1.51	22	0.05	-2.11
2	0.09	-1.81	23	0.03	-2.34
3	0.18	-1.39	24	0.05	-2.11
4	0.10	-1.75	25	0.15	-1.51
5	0.05	-2.11	26	0.10	-1.75
6	0.12	-1.65	27	0.15	-1.51
7	0.08	-1.87	28	0.09	-1.81
8	0.05	-2.11	29	0.08	-1.87
9	0.08	-1.87	30	0.18	-1.39
10	0.10	-1.75	31	0.10	-1.75
11	0.07	-1.94	32	0.20	-1.33
12	0.02	-2.50	33	0.11	-1.70

13	0.01	-2.74	34	0.30	-1.04
14	0.10	-1.75	35	0.02	-2.50
15	0.10	-1.75	36	0.20	-1.33
16	0.10	-1.75	37	0.20	-1.33
17	0.02	-2.50	38	0.30	-1.04
18	0.10	-1.75	39	0.30	-1.04
19	0.01	-2.74	40	0.40	-0.82
20	0.40	-0.82	41	0.30	-1.04
21	0.10	-1.75	42	0.05	-2.11

IV. kesimpulan

Pada pembahasan teknik-teknik statistik multivariat, akan banyak diasumsikan bahwa stiap vektor observasi X_j berdistribusi normal multivariat. Kesimpulan yang diambil berdasarkan teknik multivariat yang mengasumsikan kenormalan pada datanya akan menjadi tidak kokoh (diragukan keberlakuannya) saat asumsi kenormalan multivariat pada data tidak terpenuhi.

Oleh karena itu, menjadi penting bagi kita untuk memplajari tentang distribusi normal multivariat ini, yaitu mengenai sifat-sifat yang dimilikinya agar kita dapat mengetahui data yang kita olah memenuhi distribusi normal atau tidak.

Menaksir asumsi-asumsi kenormalan multivariat dapat dilakukan melalui pengujian pada distribusi marjinal univariat dan bivariatnya, karena pada umumnya data yang memenuhi asumsi kenormalan pada distribusi marjinal univariat atau bivariatnya akan berbentuk normal pada multivariatnya. Adapun pengujian kenormalan pada data univariat diantaranya dapat dilakukan dengan menggunakan pendekatan kenormalan untuk distribusi sampling proporsi, Q-Q plot dan menggunakan bantuan program komputer melalui uji Shapiro-Wilks, sedangkan pengujian kenormalan pada data bivariat diantaranya dapat dilakukan dengan menghitung titik-titik dalam kontur dan membandingkannya dengan teori peluang dan menggunakan Chi-square plot.

Saat asumsi kenormalan tidak dipenuhi maka kesimpulan yang kita buat berdasarkan sesuatu metode statistika yang mensyaratkan asumsi kenormalan menjadi tidak baik, sehingga muncul upaya untuk membuat keimpulan menjadi lebih baik yaitu dengan mentransformasikan data sehingga mendekati kenormalan. Transformasi data untuk mendekati kenormalan ini dapat ditentukan berdasarkan bentuk data serta faktor-faktor di luar data.