Analisis Korelasi Kanonik pada Data Normal Multivariat yang Saling Berkorelasi dalam Pembentukan Model Regresi Linier Sederhana

Rosa Ayu Oktarina[1], Sigit Nugroho[2], dan Fachri Faisal[2]

[1]Alumni Jurusan Matematika Fakultas MIPA Universitas Bengkulu

[2]Staf Pengajar Jurusan Matematika Fakultas MIPA Universitas Bengkulu

ABSTRAK

Analisis Regresi adalah teknik statistika yang berguna untuk memeriksa dan memodelkan hubungan diantara variabel-variabel secara fungsional atau berbentuk fungsi. Jika dalam suatu penelitian atau kasus menggunakan peubah yang relatif banyak, maka analisis yang digunakan adalah analisis peubah ganda. Dalam penelitian ini akan dibahas suatu pendekatan dalam mencari hubungan antara variabel bebas dengan variabel tak bebas pada data normal multivariat yang saling berkorelasi dengan menggunakan analisis korelasi kanonik (*Canonical Correlation Analysis*). Dengan melihat pasangan-pasangn kombinasi linier dari masing-masing himpunan variabel. Dari hasil penelitian dengan menggunakan dua contoh kasus yang datanya dibangkitkan terlihat bahwa besar atau kecilnya sampel sangat mempengaruhi oleh nilai korelasi kanonik dan besar atau kecilnya nilai korelasi kanonik juga dipengaruhi oleh korelasi antara variabel X_i dan Y_i.

Kata Kunci :Analisis Regresi, Analisis Korelasi kanonik.

PENDAHULUAN

Analisis regresi adalah teknik statistik yang berguna untuk memeriksa dan memodelkan hubungan diantara variabel-variabel secara fungsional atau berbentuk fungsi. Penerapan analisis regresi dapat dijumpai secara luas di banyak bidang seperti teknik, ekonomi, manajemen, ilmu-ilmu biologi, ilmu-ilmu sosial, dan ilimu-ilmu pertanian.

Analisis regresi linier dikelompokkan menjadi dua yaitu: Analisis Regresi Linier Sederhana dan Analisis Regresi Berganda. Analisis Regresi Linier Sederhana bertujuan mempelajari hubungan linier antara dua variabel, variabel ini dibedakan menjadi variabel bebas (X) dan variabel tak bebas (Y). Sedangkan Analisis Regresi Berganda adalah analisis regresi yang seringkali digunakan untuk mengatasi permasalahan analisis yang melibatkan hubungan dari dua atau lebih variabel bebas.

Jika dalam suatu penelitian atau kasus menggunakan peubah yang relatif banyak, maka analisis yang akan digunakan adalah analisis peubah ganda. Beberapa hal yang mendasari penggunaan analisis ini adalah antara peubah satu dengan peubah lain ada korelasi. Salah satu jenis analisis yang masuk katagori analisis peubah ganda adalah Analisis Korelasi Kanonik *(Canonical Correlation Analysis)*. Dalam penelitian ini akan dibahas suatu pendekatan dalam mencari hubungan antara variabel bebas dan variabel tak bebas pada data normal multivariat yang saling berkorelasi, yaitu dengan menggunakan analisis korelasi kanonik (*Canonical Correlation Analysis*).

Misalkan saja terdapat *n* pengamatan bebas pada ($p+q$) variabel yang berdistribusi normal, yaitu $Y_1, Y_2, \ldots, Y_q$ yang merupakan himpunan dari variabel-variabel tak bebas *(dependent variable)* dan $X_1, X_2, \ldots, X_p$ yang merupakan himpunan dari variabel bebas *(independent variable)*. Selanjutnya dari variabel-variabel tersebut akan dilihat pasangan-pasangan kombinasi linier dari masing-masing himpunan.

Kombinasi linier dari himpunan variabel ini disebut juga sebagai variabel kanonik, selanjutnya variabel kanonik ini akan dibentuk menjadi model regresi linier sederhana dengan menggunakan metode kuadrat terkecil.

Analisis Regresi Berganda

Hubungan fungsional antar variabel dapat ditulis sebagai
Bila variabel x lebih dari satu maka hubungan tersebut dapat ditulis sebagai

$$\hat{Y}_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + \ldots + b_k X_{k\,i} \quad i \neq k \text{ dengan } i,k = 1,2,\ldots,n. \tag{1}$$

persamaan (1) merupakan penduga dari

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{k\,i} + \varepsilon_i \tag{2}$$

dimana $\beta_0, \beta_1, \beta_2,\ dan\ \beta_k$ merupakan parameter.
Apabila dinyatakan dalam bentuk matriks persamaan (2-4) dapat dinyatakan sebagai

$$\underline{Y} = \underline{X}\underline{\beta} + \underline{\varepsilon} \tag{3}$$

Menurut Johnson dan Wichern (1998), syarat galat diasumsikan memiliki sifat :

1. $E(\varepsilon_j) = 0$
2. $Var(\varepsilon_j) = \sigma^2$ (konstan)
3. $Cov(\varepsilon_j, \varepsilon_k) = E(\varepsilon_j, \varepsilon_k) = 0,\ j \neq k.$

Metode Kuadrat Terkecil

Untuk menaksir parameter $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{k\,i} + \varepsilon_i$ maka akan digunakan metode kuadrat terkecil, dengan asumsi-asumsi sebagai berikut :

1. $E(\varepsilon_1) = 0$ untuk semua *i*
2. $E(\varepsilon_i \varepsilon_j) = \sigma^2 e$ untuk $i \neq j$; $i, j = 1,2,\ldots,n$.
3. X matriks yang mempunyai rank $k+1 < n$.

Misalkan $b' = (b_0, b_1, \cdots, b_k)$ adalah vektor baris yang merupakan taksiran untuk $\beta' = (\beta_0, \beta_1, \cdots, \beta_k)$. Dari persamaan (2-2) diperoleh $\underline{e} = \underline{y} - \underline{x}\underline{\beta}$ dimana e merupakan vektor residu. Untuk memperoleh $\underline{\hat{\beta}}$ maka $\underline{e}'\underline{e}$ diturunkan terhadap $\underline{\hat{\beta}}$ dan menyamakannya dengan 0.Sehingga di peroleh

$$\underline{\hat{\beta}} = (\underline{x}'\underline{x})^{-1}\underline{x}'\underline{y} \tag{4}$$

Untuk membuktikan bahwa $\hat{\underline{\beta}}$ adalah taksiran terbaik untuk $\underline{\beta}$ maka perlu dibuktikan bahwa $\hat{\underline{\beta}}$ mempunyai sifat tak bias dan variansi minimum.

Dapat dilihat bahwa taksiran kuadrat terkecil $\hat{\underline{\beta}}$ adalah tak bias, dengan variansi $Var\ (\hat{\underline{\beta}}) = \sigma^2 (\underline{x}'\underline{x})^{-1}$ merupakan variansi terkecil dari semua penaksir linier takbias yang dijamin oleh Teorema Gauss-Markov berikut.

Teorema Gauss-Markov

Penaksir kuadrat terkecil $\hat{\underline{\beta}} = (\underline{x}'\underline{x})^{-1}\underline{x}'\underline{y}$ mempunyai variansi terkecil dalam himpunan semua penaksir linear takbias.

Analisis Korelasi Kanonik

Analisis korelasi kanonik pertama kali diperkenalkan oleh Hotelling (1936), sebagai suatu teknik statistika peubah ganda yang menyelediki keeratan hubungan antara dua gugus peubah. Analisis korelasi kanonik digunakan untuk mencari / mengidentifikasi serta mengukur hubungan antara dua himpunan variabel (Johnson dan Wichern, 1998). Analisis korelasi kanonik memfokuskan pada korelasi antara kombinasi linier variabel pada suatu himpunan dan kombinasi linier variabel dalam himpunan lainnya.

Analisis korelasi kanonik merupakan analisis regresi ganda dengan q buah variabel tak bebas dan p buah variabel bebas. yang modelnya adalah sebagai berikut

$$\boxed{Y_1, Y_2, \ldots, Y_q \quad = \quad X_1, X_2, \ldots, X_p}$$

Matriks Matriks

Ciri data untuk korelasi kanonik adalah semua data untuk analisis korelasi kanonik bertipe matriks, yakni data interval atau data rasio. Di samping itu, analisis kanonik juga mampu menguraikan struktur hubungan di dalam gugus peubah bebas maupun di dalam gugus peubah tak bebas (Sudjana, 2002).

Dalam analisis korelasi kanonik ada beberapa asumsi yang harus dipenuhi yaitu:

1. Ada hubungan yang bersifat linier (linearitas) antar dua variabel.
2. Perlunya *Multivariate Normality* untuk menguji signifikansi setiap fungsi kanonik.
3. Tidak ada Multikolinearitas antar anggota kelompok variabel, baik variabel dependen maupun variabel independen.

Jika terjadi penyimpangan asumsi maka penanganan yang dapat dilakukan yaitu dengan melakukan transformasi data.

Proporsi Keragaman

Besarnya keragaman sampel yang diterangkan oleh peubah kanonik yang dipilih dapat dihitung dengan menggunakan pendekatan berikut :

Proporsi keragaman X yang diterangkan adalah :

$$R^2_{Z^{(1)}|U_1,U_2,\ldots,U_r} = \frac{\sum_{i=1}^{r} \sum_{k=1}^{p} r^2_{U_i, z_k^{(1)}}}{p} \qquad (5)$$

Sedangkan proporsi keragaman Y yang juga diterangkan adalah:

$$R^2_{Z^{(2)}|V_1,V_2,\ldots,V_r} = \frac{\sum_{i=1}^{r} \sum_{k=1}^{q} r^2_{V_i, z_k^{(2)}}}{q} \quad (6)$$

Besar atau kecilnya nilai proporsi keragaman menunjukkan baik atau tidaknya jumlah kanonik yang dipilih. Semakin besar nilai proporsi keragaman ini menggambarkan semakin baik peubah kanonik yang dipilih menerangkan keragaman asal.

Pendugaan Koefisien Kanonik

Misal, ingin di buat hubungan antara gugus peubah tak bebas $Y_1, Y_2, \ldots, Y_q$ yang dinotasikan dengan vektor peubah acak Y, dengan gugus peubah bebas $X_1, X_2, \ldots, X_p$ yang dinotasikan dengan vektor peubah acak X, dimana $p \le q$.
Misalkan, karakteristik dari vektor peubah acak X dan Y adalah sebagai berikut:

$$E(X^{(1)}) = \mu^{(1)}; \qquad Cov(X^{(1)}) = \Sigma_{11}$$

$$E(X^{(2)}) = \mu^{(2)}; \qquad Cov(X^{(2)}) = \Sigma_{22}$$

$$Cov(X^{(1)}, X^{(2)}) = \Sigma_{12} = \Sigma'_{21}$$

Kombinasi linear dari kedua gugus peubah tersebut dapat ditulis sebagai berikut:

$$U = \underline{a}'X^{(1)} = a_1 X_1^{(1)} + a_2 X_2^{(1)} + \cdots + a_p X_p^{(1)}$$

$$V = \underline{b}'X^{(2)} = b_1 X_1^{(2)} + b_2 X_2^{(2)} + \cdots + b_q X_q^{(2)}$$

Sehingga,

$$Var\,(U) = \underline{a}'Cov(X^{(1)})\,a = \underline{a}'\Sigma_{11}\,a$$

$$Var\,(V) = \underline{b}'Cov(X^{(2)})\,b = \underline{b}'\Sigma_{22}\,b$$

$$Cov\,(U,V) = \underline{a}'Cov\,(X^{(1)}, X^{(2)})\,b = \underline{a}'\Sigma_{12}\,b$$

Dari sini dicari koefisien vektor a dan b sehingga,

$$Corr\,(U,V) = \frac{\underline{a}'\Sigma_{12}\,b}{\sqrt{\underline{a}'\Sigma_{11}\,a}\,\sqrt{\underline{b}'\Sigma_{22}\,b}}$$ sebesar mungkin.

Sehingga dapat didefenisikan bahwa pasangan pertama dari peubah kanonik adalah kombinasi linier U_1, V_1 yang memiliki ragam satu dan korelasi terbesar , pasangan kedua dari peubah kanonik adalah kombinasi linier U_2, V_2 yang memiliki ragam satu dan korelasi terbesar kedua serta tidak berkorelasi dengan peubah kanonik pertama, pasangan ke-k dari peubah kanonik adalah kombinasi linier U_k, V_k yang memiliki ragam satu dan korelasi terbesar ke-k serta tidak berkorelasi dengan peubah kanonik *1, 2, ..., k-1* (Johnson dan Wichern,1998).

Dengan demikian dapat dituliskan :
Peubah kanonik pertama :

$U_1 = \underline{a}'_1 X^{(1)} \quad Var = (U_1) = 1$

$V_1 = \underline{b}'_1 X^{(2)} \quad Var = (V_1) = 1$

Maksimum $Corr(U_1, V_{1)} = \rho_1^*$

Peubah kanonik kedua :

$U_2 = \underline{a}'_2 X^{(1)} \quad Var = (U_2) = 1$

$V_2 = \underline{b}'_2 X^{(2)} \quad Var = (V_2) = 1$

$Cov(U_1, U_2) = 0$

$Cov(V_1, V_2) = 0$

Maksimum $Corr(U_2, V_2) = \rho_2^*$

Peubah kanonik $ke-k$:

$U_k = \underline{a}'_k X^{(1)} \quad Var = (U_k) = 1$

$V_k = \underline{b}'_k X^{(2)} \quad Var = (V_k) = 1$

$Cov(U_k, U_1) = 0$

$Cov(V_1, V_k) = 0$

Maksimum $Corr(U_k, V_{k)} = \rho_k^*$

Teorema Ketaksamaan Cauchy-Schwarz

Misalkan b dan d adalah vektor $p \times 1$, dengan $(b'd)^2 \leq (b'b)(d'd)$ jika dan hanya jika $b = c\,d$ (atau $d = c\,b$) untuk beberapa nilai c konstan.

Bukti :

Dengan menggunakan ketaksamaan Cauchy-Schwarz atau metode langrange maka diperoleh $\rho_1^{*2} \geq \rho_2^{*2} \geq \cdots \geq \rho_p^{*2}$ adalah nilai eigen dari matriks $\Sigma_{22}^{-1/2}\Sigma_{21}\Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1/2}$ yang berpadanan dengan vektor eigen $f_1, f_2, \cdots, f_p$. Disamping itu, $\rho_1^{*2} \geq \rho_2^{*2} \geq \cdots \geq \rho_p^{*2}$ juga merupakan nilai eigen dari matriks $\Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1/2}\Sigma_{21}\Sigma_{11}^{-1/2}$ yang berpadanan dengan vektor eigen $e_1, e_2, \cdots, e_p$.

Sehingga vektor koefisien a dan b diperoleh sebagai berikut :

$a_1 = e_1\,\Sigma_{11}^{-1/2} \qquad b_1 = f_1\,\Sigma_{22}^{-1/2}$

$a_2 = e_2\,\Sigma_{11}^{-1/2} \qquad b_2 = f_2\,\Sigma_{22}^{-1/2}$

$\cdots\cdots \qquad \cdots\cdots$

$a_p = e_p\,\Sigma_{11}^{-1/2} \qquad b_p = f_p\,\Sigma_{22}^{-1/2}$

Penerapan Analisis Korelasi Kanonik

Pada skripsi ini contoh kasus yang digunakan dalam penerapan analisis korelasi kanonik, menggunakan data yang telah dibangkitkan dari program Minitab Release 13.20 yang dibangkitkan sebanyak 5 variabel independen dan 5 variabel dependen sebanyak 25 dan 50 sampel. Masing-masing variabel independent diberikan simbol X_1 hingga X_5 dan variabel dependen diberikan simbol Y_1 hingga Y_5 yang telah dipilih secara acak . Kemudian dilakukan analisis dengan menggunakan software SAS 6.12, diperoleh hasil sebagai berikut :

Dari hasil output SAS 6.12, pada teladan 1 dan teladan 2, diperoleh nilai korelasi antar variabel-variabelnya. Variabel dependen maupun variabel independennya menghasilkan nilai korelasi yang cukup tinggi, hal ini menunjukkan terjadi multikolinieritas antar variabel-variabel tersebut, jika dilakukan analisis dengan menggunakan analisis regresi ganda, akan diperoleh model dugaan yang tidak mencerminkan keadaan yang sesungguhnya, sehingga perlu dilakukan analisis korelasi kanonik dalam masalah ini.

Dari hasil output SAS juga akan dihasilnya nilai korelasi dan proporsi keragaman dari masing-masing pasangan kombinasi liniernya. Kemudian akan dipilih pasangan kombinasi yang memiliki nilai proporsi keragaman yang paling tinggi, selanjutnya akan diperoleh data baru yang dapat mewakili data asalnya. Kemudian dari data tersebut akan diperoleh model dugaan sebagai berikut :

Dari data hasil pasangan kombinasi linier antara U_1 dan V_1 di atas maka dapat dibuat model dugaan dengan menggunakan metode kuadrat terkecil.

1. Untuk data simulasi pertama (*n = 25*)

Diperoleh model dugaan sebagai berikut

$$V_1 = -0.0123 + 0.7816U_1$$

dengan kuadrat galatnya adalah 12.87, sedangkan koefisien determinasinya $(R^2) = 0.5326$.

2. Untuk data simulasi kedua (*n = 50)*

Diperoleh model dugaan sebagai berikut

$$V_1 = -0.1506 + 0.7583U_1$$

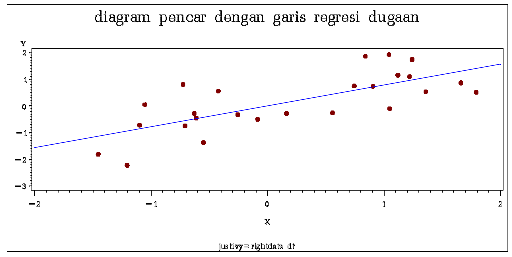
dengan jumlah kuadrat galat yang diperoleh berdasarkan output SAS 6.2 adalah 24.35, dan koefisien determinasinya $(R^2) = 0.5365$

Langkah selanjutnya adalah mengecek asumsi-asumsi yang ada pada model regresi yang diperoleh dari teladan 1 dan teladan 2. Data U_1 dan V_1 adalah merupakan data acak yang dibangkitkan dari software Minitab Release 13.20 dengan menyebar normal, dan nilai V_1 adalah bebas. Kemudian dari output SAS 6.12 (lihat lampiran 2) plot residualnya tidak membentuk model tertentu sehingga asumsi kehomogenan terpenuhi.

Untuk mengecek asumsi garis linier pada data U_1 dan V_1 dari kedua contoh teladan di atas, akan dilihat plot kenormalan dugaan garis regresinya. Apabila titik-titik dugaan yang dihasilkan telah sesuai atau mendekati garis lurus yang ditentukan berdasarkan data asal, maka titik dugaan tersebut dapat dikatakan telah mengikuti distribusi normal.

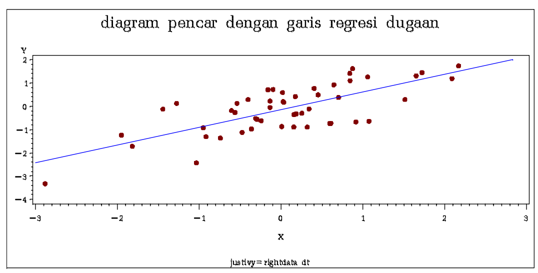
Sebaliknya, apabila titik dugaan tidak mengikuti garis lurus atau banyak titik-titik yang menyimpang, maka ada indikasi bahwa titik dugaan tidak mengikuti data normal.

Untuk mengecek asumsi garis linier pada data U_1 dan V_1 dengan jumlah *n = 25,* dari output SAS 6.12 dapat dilihat pada Gambar 2. bahwa garis dugaan merupakan garis yang lurus.



Gambar 2. Dugaan Garis Linier dengan *n = 25*

Sedangkan asumsi garis linier dengan jumlah *n = 50*, dari hasil output pada gambar 3, terlihat bahwa garis dugaannya merupakan garis yang lurus.



Gambar 3. Dugaan Garis Linier dengan *n = 50*

Dari model dugaan yang telah diperoleh tersebut untuk jumlah *n = 25* dan jumlah *n = 50* serta telah dilakukan uji asumsi pada model regresinya. Sehingga, dapat disimpulkan bahwa model dugaan yang dibuat telah memenuhi kaidah yang berlaku.

Kesimpulan Dan Saran

Dari hasil penelitian yang telah dilakukan maka dapat diambil kesimpulan sebagai berikut :

1. Dari data yang telah dibangkitkan dengan menggunakan program MINITAB Release 13.20 diperoleh besarnya nilai korelasi pada teladan 1 adalah

   $Corr\,(U_1, V_1) = 0.7334$

   $Corr\,(U_2, V_2) = 0.5927$

   $Corr\,(U_3, V_3) = 0.2583$

   $Corr\,(U_4, V_4) = 0.1625$

   $Corr\,(U_5, V_5) = 0.0480$

Besarnya nilai korelasi pada teladan 2 adalah

$Corr(U_1, V_1) = 0.7885$

$Corr(U_2, V_2) = 0.3374$

$Corr(U_3, V_3) = 0.2236$

$Corr(U_4, V_4) = 0.1479$

$Corr(U_5, V_5) = 0.0370$

2. Besar atau kecilnya sampel mempengaruhi nilai korelasi kanonik. Semakin banyak data yang digunakan sebagai sampel, maka semakin rendah nilai korelasi kanonik yang diperoleh, tergantung dengan baik atau tidaknya data dipilih.
3. Besar atau kecilnya nilai analisis korelasi kanonik sangat dipengaruhi oleh korelasi antara variabel X_i dan Y_i.

Saran

Dalam penulisan skripsi ini, penulis menggunakan data yang telah dibangkitkan dengan menggunakan software MINITAB Release 13.20. Mengingat pentingnya aplikasi analisis korelasi kanonik digunakan untuk mengetahui keeratan hubungan antara variabel, maka untuk penelitian lebih lanjut sebaiknya menggunakan data riil, sehingga penerapannya lebih nyata dan terfokus pada kasus yang dianalisis.

DAFTAR PUSTAKA

Anonim. 2007. *.Analisis Korelasi Kanonik : Analisis Pengolahan Data*. http//www.deptan.go.id/editama/statistik/web_statistik.doc.

Anonim. 2007. *Correlation Analisys With Sas : The Cancorr Procedure.* http//www.okstate.edu/sas/v7/sas pdf/stat/chap 18.pdf.

Hotelling. 1936. *Korelasi Kanonik* : *Contoh Penggunaan Korelasi Kanonik.* http//www. Youngstatistician.com.

Johnson, D. E. 1998. *Applied Multivariate Methods for Analysis*. Kansas State University, USA.

Johnson, W and Wichern, D. 1998. *Applied Multivariate Statistical Analysis*. Prentice Hall, Englewood Cliffs, New Jarsey.

Santoso, S. 2004. *Buku Latihan SPSS Statistik Multivariat*. Elex Media Komputindo, Jakarta.

Sembiring, R. K. 1995. *Analisis Regresi*. Penerbit ITB, Bandung.

Sudjana. 2002. *Teknik Analisis Regresi dan Korelasi Bagi Para Peneliti*. Edisi ke-3. Penerbit Tarsito, Bandung.

Supranto, J. 1989. *Statistika : Teori dan Aplikasi, jilid 2*. Edisi ke-5. Penerbit Erlangga.