JURNAL GAUSSIAN, Volume 2, Nomor 1, Tahun 2013, Halaman 11-18
Online di: http://ejournal-s1.undip.ac.id/index.php/gaussian

PEMILIHAN MODEL REGRESI LINIER MULTIVARIAT TERBAIK DENGAN KRITERIA *MEAN SQUARE ERROR*

Aminuddin[1], Sudarno[2], Sugito[3]
[1]Mahasiswa JurusanStatistika FSM UNDIP
[2,3]Staff Pengajar Jurusan Statistika FSM UNDIP

ABSTRAK

Regresi linier multivariatmerupakansalahsatumetodeanalisisregresi yang melibatkanlebihdarisatuvariabelrespon, dengan model regresinyaadalah$\boldsymbol{Y} = \boldsymbol{XB} + \boldsymbol{\varepsilon}$. Penggunaan banyak variabel dalam analisis regresi linier multivariat dapat menjadi hal yang menyulitkan untuk menentukan besarnya pengaruh variabel prediktor terhadap variabel respon.Olehkarenaitu, dilakukanpenyeleksianvariabelgunamendapatkan model regresiterbaik.Prosedurseleksivariabeldengankriteria*Mean Square Error* (MSE) merupakansuatumetodeuntukmendapatkan model terbaikdengancaramencari model yang memilikinilai MSE terkecildariseluruh model yang mungkin.

Kata Kunci: Regresi Linier Multivariat, SeleksiVariabel, Kriteria MSE, Model RegresiTerbaik.

1. PENDAHULUAN

1.1 LatarBelakang

Analisisregresi linier multivariatmerupakanbagiandarianalisisregresi yang melibatkantidakhanyasatuvariabelresponnamunbeberapavariabelrespony_1, y_2, ...,y_p, dengan model regresiliniernyaadalah $\mathbf{Y} = \mathbf{XB} + \boldsymbol{\varepsilon}$. Pada model regresi linier multivariat, matriks$\mathbf{Y}$maupunmatriks$\boldsymbol{\varepsilon}$diasumsikanberdistribusi normal multivariat (Timm, 2002).Selainitu, asumsi lain yang harusterpenuhiuntuk model tersebutadalahkesamaanmatrikskovarian, danindependensi residual (Rencher, 2002).

Pada regresi linier, besarnya variansi variabel respon yang dapat dijelaskan oleh variabel prediktor tergantung pada banyaknya variabel yang terlibat di dalam model. Sembiring (2003) menyatakan bahwa hal tersebut bisa menjadi hal yang menyulitkan dalam menentukan besarnya pengaruh variabel prediktor terhadap variabel respon. Oleh karena itu, pemilihan variabel X dalam analisis regresi dilakukan untuk menyeleksi variabel yang tidak signifikan secara statistik terhadap model, biasanya melaui uji parameter dengan distribusi t. Selain dengan cara tersebut, terdapat beberapa metode pemilihan variabel prediktor misalnya dengan metode pencarian (*Forward, Stepwise, and Backward Elimination*), maupun metode *All Possible Subset* (R_p^2, MSE_p, C_p). Menurut Sembiring (2003), metode forward, stepwise, dan backward merupakan metode yang paling umum digunakan, sedangkan metode all possible subset dianggap salah satu metode yang terbaik sebab mengandung semua kombinasi model yang mungkin.

Berdasarkan hal tersebut, penulis mencoba untuk menjelaskan bagaimana pemilihan model regresi linier multivariat terbaik menggunakan kriteria pemilihan *Mean Square Error* (MSE_p). Kriteria pemilihan MSE_p sendiri baik digunakan pada data dengan jumlah sampel dan dan jumlah variabel yang kecil.

1.2 TujuanPenulisan

Tujuandaripenulisantugasakhiriniadalah:

1. Mendapatkanestimasi parameter βsehinggadapatdibentukmenjadi model regresi linier multivariat.
2. Mendapatkan model terbaikdari subset Xberdasarkankriteriapemilihan*Mean Square Error* (MSE).
3. Mendapatkanbesarnyanilaihubunganantaravariabelrespondenganvariabelprediktor.
4. Mengujiasumsi model regresi linier multivariat.

2. TINJAUAN PUSTAKA

2.1 Matriks

Sebuah matriks adalah susunan segi empat siku-siku dari bilangan-bilangan. Bilangan-bilangan dalam susunan matriks tersebut dinamakan *entri* dalam matriks. Ukuran matriks dijelaskan dengan menyatakan banyaknya baris (garis horizontal) dan banyaknya kolom (garis vertikal) (Anton, 1987). Jika matriks **A** terdiri dari m baris dan n kolom, maka matriks **A** digambarkan sebagai berikut:

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{21} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}$$

Dua buah matriks dikatakan sama (simetris) apabila kedua matriks tersebut mempunyai ukuran yang sama dan entri-entri yang bersesuaian dalam kedua matriks tersebut sama, $a_{ij} = a_{ji}$ untuk seluruh i dan j.

2.2 Koefisien Korelasi

Korelasi seringkali diukur untuk mengetahui keeratan hubungan antara masing-masing variabel, cara yang dapat digunakan adalah dengan menghitung matriks korelasi antar semua variabel (Sembiring, 2003).

$$\mathbf{R} = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{21} & 1 & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & 1 \end{bmatrix}$$

dengan $r_{jk} = \dfrac{\sum_{i=1}^{n}\left(y_{ik} - \bar{y}_j\right)\left(y_{ik} - \bar{y}_k\right)}{\sqrt{\sum_{i=1}^{n}\left(y_{ik} - \bar{y}_j\right)^2}\sqrt{\sum_{j=1}^{n}\left(y_{ik} - \bar{y}_k\right)^2}}$,j = 1, 2,…, p dan k = 1, 2,…, p

Nilai r_{jk} berada antara $-1 \le r_{ik} \le +1$, ketika $r_{jk} = 0$ maka artinya tidak ada hubungan antar komponen, hubungannya sempurna bila $r_{jk} = \pm 1$; +1 artinya hubungannya searah dan -1 bila berlawanan arah (Sembiring, 2003).

2.2 Model Regresi Linier Multivariat

Model regresiliniermultivariatadalahsuatu model regresidenganlebihdarisatuvariabelrespony_1, y_2,… ,y_p yang salingberkorelasi dan satu atau lebih variabel prediktor $x_1, x_2, \dots , x_q$. Setiap variabel respon Y, diasumsikan mengikuti model regresi berikut:

$$y_1 = \beta_{01} + \beta_{11}x_1 + \beta_{21}x_2 + \dots + \beta_{p1}x_q + \varepsilon_1$$
$$y_2 = \beta_{02} + \beta_{12}x_1 + \beta_{22}x_2 + \dots + \beta_{p2}x_q + \varepsilon_2$$
$$\vdots$$
$$y_p = \beta_{0q} + \beta_{1q}x_1 + \beta_{2q}x_2 + \dots + \beta_{pq}x_q + \varepsilon_p$$

Menggunakan notasi matriks,maka sistem persamaan tersebut dapat dinyatakan sebagai:

$$\underset{(\mathbf{n\times p})}{\mathbf{Y}} = \underset{(\mathbf{n\times(q+1)})}{\mathbf{X}}\ \underset{(\mathbf{(q+1)\times p})}{\mathbf{B}} + \underset{(\mathbf{n\times p})}{\boldsymbol{\varepsilon}}$$

Untuk mendapatkan estimasi koefisienregresidenganmetodekuadratterkecil,suku random ε harusmemenuhiasumsiberikut:

1. $E(\mathbf{Y}) = \mathbf{XB}$atau$E(\boldsymbol{\varepsilon}) = \mathbf{0}$.
2. cov($\mathbf{y}_i$) = $\boldsymbol{\Sigma}$ untuk semua$i$ = 1, 2, … , $n$dimana $\mathbf{y}_\mathbf{i}^\mathbf{T}$ adalahbariske-i dari matriks **Y**.

3. $\text{cov}(\mathbf{y}_i,\mathbf{y}_j) = \mathbf{0}$untuksemua$i \neq j$.

2.3 EstimasiKuadratTerkecil Model Regresi Multivariat

Untukmendapatkantaksirankuadratterkecildari**B**yang secararingkasditulisdalamnotasimatriks menjadi $\mathbf{Y} = \mathbf{X}\widehat{\mathbf{B}} + \boldsymbol{\varepsilon}$. Pada kasus multivariat estimasi kuadrat terkecil juga meminimumkan jumlah kuadrat residual (JKR). Oleh karena itu

$$\mathbf{E} = \hat{\boldsymbol{\varepsilon}}^{\mathbf{T}}\hat{\boldsymbol{\varepsilon}} = \left(\mathbf{Y} - \mathbf{X}\widehat{\mathbf{B}}\right)^{\mathbf{T}}\left(\mathbf{Y} - \mathbf{X}\widehat{\mathbf{B}}\right)$$
$$= \mathbf{Y}^{\mathbf{T}}\mathbf{Y} - 2\widehat{\mathbf{B}}^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\mathbf{Y} + \widehat{\mathbf{B}}^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\mathbf{X}\widehat{\mathbf{B}}$$

maka, $$\frac{\partial(\mathbf{E})}{\partial\widehat{\mathbf{B}}} = -2\mathbf{X}^{\mathbf{T}}\mathbf{Y} + 2\mathbf{X}^{\mathbf{T}}\mathbf{X}\widehat{\mathbf{B}} = \mathbf{0}.$$

Sehinggadenganmenggunakanpersamaantersebutdidapatkan $\mathbf{X}^{\mathbf{T}}\mathbf{X}\widehat{\mathbf{B}} = \mathbf{X}^{\mathbf{T}}\mathbf{Y}$.

Diketahui bahwa $\mathbf{X^TX}$ merupakan matriks kuadrat berukuran $k \times k$ yang diasumsikan sebagai *nonsingular* karena itu dapat dibalik, maka dari itu kita bisa mengalikan kedua sisinya dengan $(\mathbf{X^TX})^{-1}$ untuk memperoleh

$$\widehat{\mathbf{B}} = \left(\mathbf{X}^{\mathbf{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathbf{T}}$$

Estimator $\widehat{\mathbf{B}}$ yang diperolehbersifat tak bias dengan

1. $E\left(\widehat{\mathbf{B}}\right) = \mathbf{B}$
2. $\text{cov}\left(\widehat{\mathbf{B}}\right) = \mathbf{\Sigma(X^TX)^{-1}}$

Pada kasus multivariat, estimator tak bias dari $\mathbf{\Sigma}$ ditunjukan melalui persamaan berikut (Rencer, 2002):

$$E(\mathbf{S}) = E\left(\frac{\mathbf{Y}^{\mathbf{T}}\mathbf{Y} - \widehat{\mathbf{B}}^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\mathbf{Y}}{n-q-1}\right) = E\left(\frac{\mathbf{Y}^{\mathbf{T}}\mathbf{Y} - \hat{\mathbf{Y}}^{\mathbf{T}}\hat{\mathbf{Y}}}{n-q-1}\right) = \mathbf{\Sigma}$$

dengan penyebut $n - q - 1$,**S** merupakan estimator tak bias dari $\mathbf{\Sigma}$.

Uji Signifikansi Parameter

Pengujian hipotesis ini merupakan pengujian untuk seluruh koefisien regresi β_{jk} dalam $\mathbf{B_1}$ terhadap *Y*. Hipotesis yang digunakan adalah sebagai berikut:

H_0: $\mathbf{B_1 = 0}$

H_1: $\mathbf{B_1 \neq 0}$

dimana $\mathbf{B_1}$ mencakupseluruhbarisdarimatriks $\mathbf{B}$ kecualibarispertama:

$$\mathbf{B} = \begin{bmatrix} \boldsymbol{\beta}_\mathbf{0}^\mathbf{T} \\ \mathbf{B_1} \end{bmatrix} = \left[\begin{array}{cccc} \beta_{01} & \beta_{01} & \cdots & \beta_{0q} \\ \hline \beta_{11} & \beta_{12} & \cdots & \beta_{1q} \\ \vdots & \vdots & & \\ \beta_{p1} & \beta_{p2} & \cdots & \beta_{pq} \end{array}\right]$$

Statistik uji yang digunakan adalah Wilk's lamda,

$$\Lambda = \frac{|\mathbf{E}|}{|\mathbf{E}+\mathbf{H}|} = \frac{\left|\mathbf{Y}^{\mathbf{T}}\mathbf{Y} - \widehat{\mathbf{B}}^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\mathbf{Y}\right|}{\left|\mathbf{Y}^{\mathbf{T}}\mathbf{Y} - n\bar{\mathbf{y}}\bar{\mathbf{y}}^{\mathbf{T}}\right|}$$

Tolak H_0jika $\Lambda \leq \Lambda_{\alpha,p,q,(n-q-1)}$. Nilai $\Lambda_{\alpha,p,q,(n-q-1)}$ merupakannilaikritis dari tabel Wilk's lamda (Rencher, 2002).

Kriteria Pemilihan *Mean Square Error*

Padakasusunivariatkriteria $MSE_p = s_p^2 = SSE_p/(n-p)$. Pada kasus multivariate bentuk persamaan tersebut menjadi,

$$MSE_p = \mathbf{S}_p = \frac{\mathbf{E}_p}{n-p}\text{, dengan } \mathbf{E}_p = \mathbf{Y}^\mathbf{T}\mathbf{Y} - \hat{\mathbf{B}}_\mathbf{q}^\mathbf{T}\mathbf{X}_\mathbf{q}^\mathbf{T}\mathbf{Y}.$$

Kriteriapemilihan subset adalahdenganmemilih subset yang memilikinilai minimum dari $tr(\mathbf{S}_p)$ atau $|\mathbf{S}_p|$ (Rencher, 2002).

Pengujian Subset *X*

Uji ini dilakukan untuk menguji apakah subset *X* signifikan terhadap model. Hipotesis yang digunakan adalah:

H_0: $\mathbf{B}_d = \mathbf{O}$

H_1: $\mathbf{B}_d \neq \mathbf{O}$

Statistik uji yang digunakan adalah:

$$\Lambda = \frac{\left|\mathbf{Y}^\mathbf{T}\mathbf{Y} - \hat{\mathbf{B}}^\mathbf{T}\mathbf{X}^\mathbf{T}\mathbf{Y}\right|}{\left|\mathbf{Y}^\mathbf{T}\mathbf{Y} - \hat{\mathbf{B}}_\mathbf{r}^\mathbf{T}\mathbf{X}_\mathbf{r}^\mathbf{T}\mathbf{Y}\right|} = \frac{\Lambda_f}{\Lambda_r}$$

Kriteriapenolakan, tolak H_0jika $\Lambda \leq \Lambda_{\alpha,p,h,(n-q-1)}$, dengan*h* adalahbanyaknyavariabelprediktordalam model tereduksi(Rencher, 2002).

2.4 Hubungan Antara Variabel Respon dan Prediktor

Pada regresi linier multivariat, ukuran yang digunakan dalam mengukur hubungan antara variabel respon dan variabel prediktor adalah dengan menggunakan rasio korelasi Fisher yang disarankan oleh Wilk.

$$\eta_\Lambda^2 = 1 - \Lambda$$

Nilai η_Λ^2 terletakdiantara 0 dan 1, artinya semakin nilainya mendekati satu berarti semakin erat hubungan antara variabel respon dan variabel prediktor (Rencher, 2002).

2.5 Asumsi Regresi Linier Multivariat

Apabila model telah ditetapkan, maka selanjutnya adalah melakukan pemeriksaan asumsi. Asumsi-asumsi yang diperlukandalam regresi linier multivariatantara lain:

Independensi Residual

Residual $\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \ldots, \boldsymbol{\varepsilon}_p$ dikatakan bersifat saling bebas (independen) jika matriks korelasi antar residual membentuk matriks identitas, untuk menguji kebebasan antar residual ini dapat dilakukan uji *Bartlet Sphericity*(Morrison, 2005).

Hipotesis:

H_0: $\mathbf{P} = \mathbf{I}$ (Residual bersifat independen)

H_1: $\mathbf{P} \neq \mathbf{I}$ (Residual bersifat dependen)

Taraf signifikansi α=5 %

Statistik Uji:

$$\chi^2_{hitung} = -\left\{n - 1 - \frac{2p+5}{6}\right\}\ln|\mathbf{R}|$$

dengan *p* adalah jumlah variabel respon dan $\ln|\mathbf{R}|$ adalah nilai-nilai determinan matriks korelasi dari masing-masing residual.

Kreteria penolakan:

Tolak H_0 jika $\chi^2_{hitung} \geq \chi^2_{\alpha;\frac{1}{2}p(p-1)}$ yang berarti residualnya bersifat independen.

Uji Normal Multivariat

Terdapat dua cara yang dapat dilakukan dalam mengecek asumsi normal multivariat. Pertama, memeriksa asumsi kenormalan dengan membuat plot *Chi Square* (untuk p $\geq$2). Jika hasil plot berpola linier, maka residual dapat diasumsikan berdistribusi normal multivariat.
Kemudian yang keduaadalah dengan melihat banyaknya nilai d_i^2 yang kurang dari nilai kuantil *Chi square*.Pertama-tama yang harus dilakukan adalah menghitung nilai d_i^2, i = 1, 2,…,n dan kemudian membandingkannya dengan nilai kuantil χ^2. Apabila terdapat setengah atau lebih nilai $d_i^2 \leq q_{c,p}(0.50)$, makadapatdikatakanresidualnyaberdistribusi normal(Johnson andWichern,2007).

UjiKesamaanMatriks Varians Kovarian

Kesamaan varians untuk matriks varians-kovarian residual dapat diperiksa dengan menggunakan uji Box's M.
Hipotesis:
H$_0$: $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \ldots = \boldsymbol{\Sigma}_j$ (Matriks varian kovarian homogen)
H$_1$: minimal ada satu $\boldsymbol{\Sigma}_j \neq \boldsymbol{\Sigma}_k$, untuk $j \neq k$ (Matriks varian kovarian tidak homogen)
Taraf Signifikansi α = 5%
Statistik Uji:

$$C = (1-u)M = (1-u)\left\{\left[\sum_l (n_l - 1)\right]\ln\left|\mathbf{S}_{pooled}\right| - \sum_l \left[(n_l - 1)\ln\left|\mathbf{S}_g\right|\right]\right\}$$

dengan:

$$\mathbf{S}_g = \frac{1}{g-1}\sum_{j=1}^{l}\left(\boldsymbol{\varepsilon}_j - \bar{\boldsymbol{\varepsilon}}\right)\left(\boldsymbol{\varepsilon}_j - \bar{\boldsymbol{\varepsilon}}\right)^{\mathbf{T}} \;;\mathbf{S}_{pooled} = \frac{1}{\sum_l (n_l - 1)}\left\{(n_1 - 1)\mathbf{S}_1 + \ldots + (n_g - 1)\mathbf{S}_g\right\}$$

$$u = \left[\sum_l \frac{1}{(n_l - 1)} - \frac{1}{\sum_l (n_l - 1)}\right]\left[\frac{2p^2 + 3p - 1}{6(p+1)(g-1)}\right]$$

Kriteria penolakan:
Tolak H$_0$jika $C \geq \chi_{\alpha;(p(p+1)(g-1))/2}$, yang berartimatriksvarians-kovarianbersifathomogen(Johnson and Wichern,2007).

3. METODOLOGI PENELITIAN

3.1 Jenis dan Sumber Data

Data yang digunakan bersumber dari buku Aplied Multivariate Analysis karangan Neil H. Timm. Data yang digunakan yaitu data mengenai murid taman kanak-kanak (TK) pada suatu wilayah dengan status sosial ekonomi rendah.

3.2 Metode Penelitian

Metode yang digunakan dalam penulisan tugas akhir ini adalah metode kepustakaan dan contoh kasus.

3.3 Metode Analisis

Adapun tahapan analisis yang digunakan sebagai berikut:
Tahap I : Menguji Korelasi Antar Variabel Respon
Tahap II : Memilih subset X dengan kriteria pemilihan MSE
Tahap III : Membuat Model Regresi Linier Multivariat
Tahap IV: Uji Asumsi Regresi Linier Multivariat

4. HASIL DAN PEMBAHASAN

4.1 Analisis Hubungan Variabel Respon

Berdasarkan uji *Bartlett Sphericity* hasilnya sebagai berikut:
Hipotesis
H$_0$: Antar variabel respon bersifat independen

H_1: Antar variabel respon bersifat dependen

Statistik Uji

p-value= 0.003

Kesimpulan

Menolak H_0 karena nilai *p-value* = 0.003 < α = 0.05 yang berarti antar variabel respon bersifat dependen, jadi data dapat digunakan pada analisis regresi linier multivariat.

4.2 Pemilihan Model Terbaikdengan Kriteria *Mean Square Error*

Prosedur pemilihan tersebut adalah dengan meregresikan seluruh subset X terhadap variabel respon. Jadi nantinya akan terdapat (2^5 – 1) macam model atau sebanyak 31 model regresi.

No	Prediktor	MSEp	No	Prediktor	MSEp	No	Prediktor	MSEp
1	1	376548.72	11	24	280631.88	21	145	283469.73
2	2	382457.92	12	25	299422.55	22	234	253905.42
3	3	324843.6	**13**	**34**	**240388.51**	23	235	264546.88
4	4	259613.32	14	35	264565.25	24	245	287686.87
5	5	295520.92	15	45	267908.73	25	345	245605.33
6	12	361970.23	16	123	297945.88	26	1234	267555.32
7	13	308696.65	17	124	295770.72	27	1235	276161.91
8	14	272944.5	18	125	311694.83	28	1245	305144.61
9	15	303813.93	19	134	252764.28	29	1345	260163.05
10	23	300706.2	20	135	272932.59	30	2345	259720.50
						31	12345	275684.87

Berdasarkan tersebut dapat dilihat bahwa model dengan variabel prediktor x_3 dan x_4 memiliki nilai MSE yang paling minimum. Oleh karena itu, dapat dikatakan bahwa model regresi linier multivariat dengan variabel prediktornya adalah x_3 dan x_4 merupakan model yang terbaik.Sehingga, model regresi terbaiknya sebagai berikut:

$$\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \end{bmatrix} = \begin{bmatrix} 10.2847 \\ 34.8560 \\ 9.2686 \end{bmatrix} + \begin{bmatrix} -0.3660 & 1.1660 \\ 0.1736 & 1.1086 \\ 0.1866 & 0.0620 \end{bmatrix} \begin{bmatrix} NS \\ NA \end{bmatrix}$$

4.3 Pengujian Kecocokan Model Regresi Terbaik

Hipotesis

H_0: $\mathbf{B_1} = \mathbf{0}$ (parameter tidak signifikan secara serentak terhadap model)

H_1: $\mathbf{B_1} \neq \mathbf{0}$ (parameter signifikan secara serentak terhadap model)

Statistik Uji

$\Lambda_h = 0.4787$

Kesimpulan

Menerima H_1karena $\Lambda_h = 0.4787 < \Lambda_{0.05,3,2,34} = 0.481$, artinyaparameter signifikansecara serentak terhadap model. Dengan kata lain, model regresi linier multivariat terbaik cocok untuk dipergunakan.

4.4 Pengujian Subset *X*

Pengujian ini dilakukan dengan cara membandingkan model lengkap dengan model tereduksinya.

Hipotesis

H_0: $\mathbf{B}_d = \mathbf{O}$ (Model linier pada x_3 dan x_4)

H_1: $\mathbf{B}_d \neq \mathbf{O}$ (Model tidak linier pada x_3 dan x_4)

Statistik Uji

$$\Lambda_h = \frac{\Lambda_f}{\Lambda_r} = 0.8765$$

Kesimpulan

Menerima H_0karenadiketahuibahwa $\Lambda_h = 0.8765 > \Lambda_{0.05,3,2,31} = 0.657$. Artinya model linier pada x_3 dan x_4, sehingga model tereduksi dengan variabel x_3 dan x_4 dapat diterima.

Setelah seluruh pengujian atas model terbaik dilakukan, sehingga dapat ditentukan besarnya hubungan antara variabel respon dan variabel prediktor. Nilai η^2_Λ =1 – 0.4787 = 0.5213. Artinya, model dapat menjelaskan informasi data sebesar 52.13 %.

4.5 Pengujian Asumsi Regresi Linier Multivariat

Asumsi yang harus dipenuhi dalam regresi linier multivariat antara lain:

Asumsi Residual Berdistribusi Normal Multivariat

Hipotesis

H_0 : Residual data berdistribusi normal multivariat

H_1 : Residual data tidak berdistribusi normal multivariat

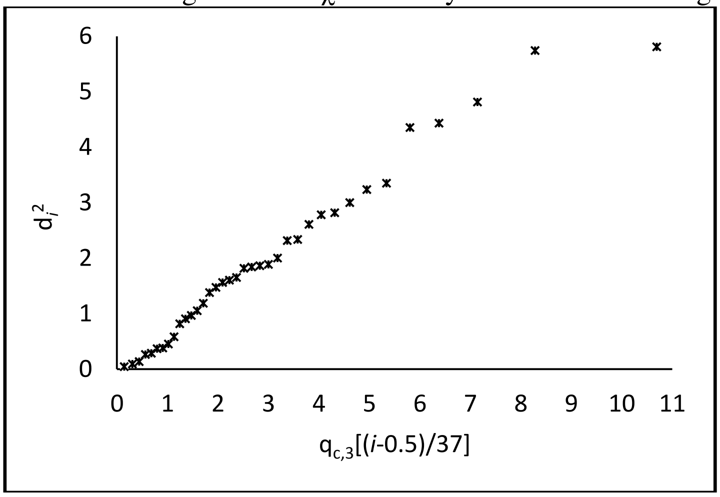
Statistik uji

p-value = 0.3868

Kesimpulan

H_0 diterima sebab *p-value* = 0.3868 > α = 0.05 yang berarti residual dari data berdistribusi normal multivariat.

Secara grafis pun residual dari data dapat dikatakan berdistribusi normal, sebab plot antara nilai d_i^2 yang telah diurutkan dengan kuantil χ^2 bentuknya mendekati bentuk garis lurus.



Gambar 1. Grafik Chi-Square Untuk Cek Asumsi Normal Multivariat

Asumsi Independensi Residual

Hipotesis

H_0: **P** = **I** (Residual bersifat independen)

H_1: **P** ≠ **I** (Residual bersifatdependen)

Statistik Uji

$\chi^2_{hitung} = 3.6873$;*p-value* = 0.051

Kesimpulan

Berdasarkantabel χ^2, diperoleh $\chi^2_{0.05;3} = 7.81$ dannilai*p-value* = 0.051, nilai *ChiSquare test* =3.6873< $\chi^2_{0.05;3} = 7.81$ dan *p-value* = 0.051 > α = 0.05 maka H_0 diterima,artinya residual bersifatindependen.

Asumsi HomogenitasMatriks Varian Kovarian
Hipotesis

$H_0 : \Sigma_1 = \Sigma_2 = \Sigma_3$ (matriksvariankovarianhomogen)

H_1 : minimal adasatu $\Sigma_j \neq \Sigma_k$ (matriksvariankovariantidakhomogen)

Statistik uji

Box's M = 10.226, *p-value* = 0.712

Kesimpulan

Berdasarkantabel χ^2 diperoleh $\chi^2_{0.05;12} = 21.03$ dannilai*p-value* = 0.712, karena nilai uji Box M = 10.226 < $\chi^2_{0.05;12} = 21.03$ dan*p-value* = 0.712 > α = 0.05 maka H_0 diterima yang berarti matriks varian kovarian adalah homogen.

5. KESIMPULAN

Kesimpulan yang dapat diambil dari pembahasan tugas akhir ini adalah:
Model dengan nilai MSE terkecil ditetapkan sebagai model terbaik. Hasil yang didapatkan adalah memilih model yang memiliki subset dengan 2 variabel yaitu variabel x_3 dan x_4.Hasil estimasi parameter $\widehat{\mathbf{B}}$ untuk model regresi terbaik adalah:

$$\widehat{\mathbf{B}} = \begin{bmatrix} 10.2847 & 34.8560 & 9.2688 \\ -0.3660 & 0.1736 & 0.1866 \\ 1.1660 & 1.1086 & 0.0620 \end{bmatrix}$$

Model regresi linier multivariat terbaik yang didapat adalah:

$$\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \end{bmatrix} = \begin{bmatrix} 10.2847 \\ 34.8560 \\ 9.2686 \end{bmatrix} + \begin{bmatrix} -.3660 & 1.1660 \\ 0.1736 & 1.1086 \\ 0.1866 & 0.0620 \end{bmatrix} \begin{bmatrix} NS \\ NA \end{bmatrix}$$

Besarnya hubungan antara variabel respon dengan variabel prediktor yaitu sebesar 52.13 %, sisanya dipengaruhi oleh hal lain diluar model.

6. DAFTAR PUSTAKA

Anton, H., 1987, Aljabar Linier Elementer, Edisi Kelima, Alih Bahasa oleh Pantur Silaban dan I. Nyoman Susila, Jakarta, Erlangga.

Johnson, R.A and Wichern, D.W., 2007, Applied Multivariate Statistical Analysis, Sixth Edition, New Jersey, Prentice Hall International Inc.

Morrison, D.F., 2005, Multivariat Statistical Methods, Fourth Edition, The Wharton School University of Pennsylvania.

Rencher, A.C., 2002, Methods of Multivariate Analysis, Second Edition, New York, John Wiley & Sons Inc.

Sembiring, R.K., 2003, Analisis Regresi, Edisi Kedua, Bandung, Penerbit ITB.

Timm, N.H., 2002, Applied Multivariate Analysis, New York, Springer-Verlag New York Inc.