

**PEMILIHAN MODEL REGRESI MULTIVARIAT TERBAIK DENGAN
KRITERIA KULLBACK'S INFORMATION CRITERION CORRECTION (KICC)
(STUDI KASUS : TINGKAT KESEJAHTERAAN MASYARAKAT DI
PROVINSI SUMATERA UTARA)**

Daniel Limbong¹, Yudi Setyawan^{2*}

^{1,2}Jurusan Statistika, Fakultas Sains Terapan, Institut Sains dan Teknologi AKPRIND, Yogyakarta
Email: daniellimbong33@gmail.com¹, yudista2003@yahoo.com²

Abstract: *Community welfare is a multi-indicator concept that shows the size of development success in a region. Economically, the welfare of a common country is measured through an instrument of economic growth (growth), income or GRDP and a human development index (human development index). Based on these indicators can be made a model used to predict factors supporting the level of community welfare using multivariate regression analysis with KICC selection (Kullback's Information Criterion Correction). KICC uses the Kullback-Leibler distance range or so-called Kullback-Leibler Symmetric Distance, the smallest KICC value is the best model. Multivariate regression analysis is part of regression analysis involving not only one response variable but several response variables. By using KICC in the selection of the best model obtained the best model of economic growth = 6,93 - 0,0485 life expectancy + 0,1215 unemployment rate - 0,1370 per capita expenditure - 0,0317 percentage of poverty + 0,0349 labor force participation rate + 0,000679 local revenue budget, per capita GRDP = 51.7 + 0,040 life expectancy - 0.58 unemployment rate + 2.95 per capita expenditure + 0.149 percentage of poverty - 0.678 labor force participation rate + 0.01115 local revenue budget, and HDI = 1.98 + 0.623 life expectancy + 0.524 unemployment rate + 1.415 per capita expenditure - 0.2136 percentage of poverty + 0.1417 labor force participation rate - 0.00048 local revenue budget. The correlation between predictor variable of life expectancy, unemployment rate, per capita expenditure, poverty percentage, LFPR and APAD to economic growth response variable, per capita GRDP and HDI (η_A^2) is 97,556%.*

Keywords: *Welfare, Multivariate Regression Analysis, KICC.*

Abstrak: Kesejahteraan masyarakat merupakan suatu konsep multiindikator yang menunjukkan ukuran keberhasilan pembangunan di suatu wilayah. Secara ekonomi, kesejahteraan hidup suatu negara biasa diukur melalui instrumen pertumbuhan ekonomi (*growth*), pendapatan atau PDRB dan indeks pembangunan manusia (*human development index*). Berdasarkan indikator-indikator tersebut dapat dibuat suatu model yang digunakan untuk menduga faktor-faktor penunjang tingkat kesejahteraan masyarakat menggunakan analisis regresi multivariat dengan pemilihan KICC (*Kullback's Information Criterion Correction*). KICC menggunakan jumlahan jarak Kullback-Leibler atau yang biasa disebut Jarak Simetris Kullback-Leibler, nilai KICC yang terkecil merupakan model yang terbaik. Analisis regresi multivariat merupakan bagian dari analisis regresi yang melibatkan tidak hanya satu variabel respon namun beberapa variabel respon. Dengan menggunakan KICC pada pemilihan model terbaik diperoleh model terbaik pertumbuhan ekonomi = 6,93 - 0,0485 angka harapan hidup + 0,1215 tingkat pengangguran - 0,1370 pengeluaran perkapita - 0,0317 persentase kemiskinan + 0,0349 tingkat partisipasi angkatan kerja + 0,000679 anggaran pendapatan asli daerah, PDRB perkapita = 51,7 + 0,040 angka harapan hidup - 0,58 tingkat pengangguran + 2,95 pengeluaran perkapita + 0,149 persentase kemiskinan - 0,678 tingkat partisipasi angkatan kerja + 0,01115 anggaran pendapatan asli daerah, dan IPM = 1,98 + 0,623 angka harapan hidup + 0,524 tingkat pengangguran + 1,415 pengeluaran perkapita - 0,2136 persentase kemiskinan + 0,1417 tingkat partisipasi angkatan kerja - 0,00048 anggaran pendapatan asli daerah. Adapun hubungan antara variabel-variabel prediktor angka harapan hidup, tingkat pengangguran, pengeluaran perkapita, persentase kemiskinan, TPAK dan APAD terhadap variabel respon pertumbuhan ekonomi, PDRB perkapita dan IPM (η_A^2) adalah sebesar 97,556%.

Kata Kunci: *Kesejahteraan, Analisis Regresi Multivariat, KICC.*

*Corresponding author's email: yudista2003@yahoo.com

I. Pendahuluan

Analisis regresi merupakan salah satu metode dalam menggambarkan dan menentukan hubungan sebab akibat antara dua variabel ataupun lebih. Analisis regresi linear dapat dibagi menjadi dua, yaitu analisis regresi linear sederhana dan analisis regresi linear berganda (*multiple regression*). Perbedaan dari kedua analisis ini terletak pada jumlah variabel prediktornya. Dalam regresi linear sederhana, variabel prediktornya hanya terdiri dari satu variabel prediktor saja, sedangkan pada regresi linier berganda terdiri dari dua atau lebih variabel prediktor. Dalam pengaplikasiannya bisa saja ditemui analisis regresi dengan jumlah variabel respon lebih dari satu, untuk menyelesaikan permasalahan ini kita dapat gunakan analisis regresi multivariat.

Model regresi multivariat adalah model regresi dengan lebih dari satu variabel respon yang saling berkorelasi model regresi dengan q buah variabel respon Y_1, Y_2, \dots, Y_q yang saling berkorelasi dan satu atau beberapa variabel prediktor X_1, X_2, \dots, X_p [3]. Dalam analisis regresi multivariat mempertimbangkan adanya hubungan ketergantungan diantara variabel respon Y_1, Y_2, \dots, Y_q .

Dalam mendapatkan model terbaik terdapat beberapa kriteria yang sudah umum digunakan diantaranya pemilihan model dengan *mean square error (MSE)*, *AIC (Akaike's information Criterion)*, *AICC (Akaike Information Criterion Correction)*, dan *KICC (Kullback's Information Criterion Correction)*. Dalam penelitian ini penulis menggunakan kriteria KICC dalam hal penentuan model terbaik. KICC merupakan penyempurnaan dari kriteria AIC dan AICC. KICC menggunakan jumlahan jarak Kullback-Leibler atau yang biasa disebut Jarak Simetris Kullback-Leibler.

Secara ekonomi, kesejahteraan masyarakat suatu negara biasa diukur melalui instrumen pertumbuhan ekonomi atau PDB (*growth*), pendapatan atau PDRB per kapita (*per capita income*) dan indeks pembangunan manusia (*human development index*).

Pertumbuhan ekonomi Provinsi Sumatera Utara tahun 2005-2011 masih relatif lebih tinggi bila dibandingkan dengan pertumbuhan ekonomi rata-rata provinsi di Indonesia. Tetapi pada tahun 2012 pertumbuhan ekonomi Provinsi Sumatera Utara mengalami penurunan, menjadi 6,23 persen dari sebelumnya 6,49 persen pada tahun 2011 (BPS 2012). Pada tahun 2016 tingkat pendapatan provinsi Sumatera Utara mengalami kenaikan dari tahun-tahun sebelumnya bahkan melampaui dari target yang diharapkan (Berita: Pemprov Sumut). Akan tetapi tingkat ketimpangan pendapatan perkapita masyarakat Sumatera Utara berdasarkan Indeks Gini Ratio cukup besar yaitu 0,362. Angka itu lebih tinggi dibanding Gini Ratio Nasional sebesar 0,362. Itu artinya bahwa tingkat kesejahteraan masyarakat belum merata di provinsi ini. Baik buruknya pembangunan juga dapat dilihat dari nilai Indeks Pembangunan Manusia (IPM). Di Sumatera Utara sendiri pembangunan manusianya terus mengalami kemajuan yang ditandai dengan terus meningkatnya Indeks Pembangunan Manusia (IPM) Sumatera Utara.

Berdasarkan uraian latar belakang tersebut, penulis tertarik untuk menerapkan analisis regresi multivariat untuk mengetahui faktor apa saja yang mempengaruhi kesejahteraan masyarakat provinsi Sumatera Utara dan pemilihan model terbaik dengan kriteria KICC.

II. Metodologi Penelitian

2.1 Sumber Data dan Variabel Penelitian

Data yang digunakan adalah data sekunder Data yang digunakan dalam penelitian ini adalah data sekunder, yaitu data yang diperoleh langsung dari Publikasi Provinsi Sumatera Utara dalam Angka 2016 di website Badan Pusat Statistik Provinsi Sumatera Utara. Variabel yang digunakan dalam penelitian ini adalah variabel respon Laju Pertumbuhan Ekonomi, PDRB Per Kapita dan IPM dengan variabel prediktornya angka harapan hidup, tingkat pengangguran, pengeluaran perkapita, persentase kemiskinan tingkat partisipasi angkatan kerja dan Anggaran Pendapatan Asli Daerah.

2.2 Tahapan Analisis Data

Adapun langkah-langkah analisis yang dilakukan untuk mencapai tujuan penelitian adalah sebagai berikut:

- 1) Melakukan analisis deskriptif untuk mengetahui karakteristik data pengamatan. dengan menggunakan nilai rata-rata, nilai maksimum dan nilai minimum terhadap variabel respon dan variabel prediktor. Deskriptif data dapat juga dilakukan melalui tabel, grafik, diagram, boxplot, dan scatterplot.

- 2) Pengujian korelasi antar variabel respon

Dalam analisis regresi multivariat, dibutuhkan asumsi bahwa antar variabel respon Y_1, Y_2, \dots, Y_q memiliki hubungan dependen. Untuk menguji kebebasan antar variabel dapat dilakukan dengan uji Barlett Sphericity [6] dengan hipotesis H_0 : Antar variabel respon bersifat independen dan H_1 : Antar variabel respon bersifat dependen. Statistik uji:

$$\chi_{hitung}^2 = -\left\{n - 1 - \frac{2q+5}{6}\right\} \ln|\mathbf{R}| \quad (1)$$

Dengan q adalah jumlah variabel respon, \mathbf{R} adalah matriks korelasi variabel respon dan $\ln|\mathbf{R}|$ merupakan nilai determinan matriks korelasi variabel respon. Dengan pengambilan keputusan H_0 ditolak jika $\chi_{hitung}^2 > \chi_{tabel}^2$ dengan $\chi_{tabel}^2 = \chi_{\alpha, \frac{1}{2}q(q-1)}$.

- 3) Uji normal multivariat pada variabel respon

Untuk melihat atau memeriksa asumsi ini dapat dilakukan dengan menghitung jarak kuadrat pada setiap pengamatan dengan hipotesis H_0 : Data berdistribusi normal multivariat, H_1 : Data tidak berdistribusi normal multivariat, Statistik uji:

$$\mathbf{d}_i^2 = (\mathbf{Y}_i - \bar{\mathbf{Y}})^T \mathbf{S}^{-1} (\mathbf{Y}_i - \bar{\mathbf{Y}}) \quad (2)$$

Dimana: \mathbf{Y}_i = vektor objek pengamatan ke- i , $\bar{\mathbf{Y}}$ = vektor rata-rata pengamatan, \mathbf{S}^{-1} = invers matriks varian-kovarian pengamatan berukuran $q \times q$, \mathbf{d}_i^2 = jarak mahalalanobis.

Dengan pengambilan keputusan H_0 ditolak jika diperoleh kondisi $\mathbf{d}_i^2 > \chi_{tabel}^2$ terhadap lebih dari $\frac{1}{2}n$ pengamatan.

- 4) Pemilihan model dengan menggunakan metode KICC.

Menurut Hafidi dan Mkhadri (2006) kriteria *Kullback's Information Criterion Corected* (KICC) merupakan koreksi dari metode KIC dan akan menghasilkan model terbaik jika digunakan pada sampel kecil untuk pemilihan model linier multivariat [2]. KICC merupakan penyempurnaan dari kriteria AIC dan AICC. KICC menggunakan jumlahan jarak Kullback-Leibler atau yang biasa disebut Jarak Simetris Kullback-Leibler. Hafidi dan Mkhadri (2006) menyatakan besarnya nilai KICC dapat diperoleh dari persamaan berikut:

$$\text{KICC} = n \left(\ln|\hat{\Sigma}| + q \right) + \frac{d(3n-p-q-1)}{n-p-q-1} \quad (3)$$

Dengan: $d = pq + 0,5q(q+1)$

q = jumlah variabel respon

p = jumlah variabel prediktor

n = jumlah pengamatan

$\hat{\Sigma}$ = penaksir matriks varian-kovarian

- 5) Analisis Regresi Multivariat dan Estimasi Parameter

Model regresi multivariat adalah model regresi dengan lebih dari satu variabel respon yang saling berkorelasi dan satu atau lebih variabel prediktor [3]. Model regresi multivariat yang terdiri atas model linear secara simultan dapat ditunjukkan secara matriks dalam persamaan

$$\mathbf{Y}_{(n \times q)} = \mathbf{X}_{n(p+1)} \boldsymbol{\beta}_{(p+1) \times q} + \boldsymbol{\varepsilon}_{n \times q} \quad (4)$$

dengan $E(\varepsilon_{ik}) = 0$, dan $\text{cov}(\varepsilon_i, \varepsilon_k) = \sigma_i I$

Dalam model regresi multivariat persamaan (4), diestimasi parameter

$$\hat{\boldsymbol{\beta}}_k = [\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k]$$

Pemilihan Model Regresi Multivariat Terbaik Dengan Kriteria.....

Sedangkan ε_k yang merupakan matriks residual ditentukan oleh estimasi $\hat{\varepsilon}_k = Y_k - X\hat{\beta}_k$.

- 6) Pengujian signifikansi parameter model regresi multivariat
Ada dua jenis uji yang dilakukan untuk mengetahui tingkat signifikansi parameter regresi terhadap model yaitu pengujian secara keseluruhan (simultan) dan pengujian secara parsial.

1) Pengujian Simultan

Pengujian secara simultan dilakukan untuk mengetahui apakah parameter signifikan secara keseluruhan pada model. Dengan hipotesis:

$$H_0 : \beta_{11} = \beta_{12} = \dots = \beta_{p1} = \dots = \beta_{pq} = 0$$

(model tidak signifikan)

$$H_1 : \text{Paling sedikit ada satu } \beta_{jk} \neq 0, j = 1, 2, \dots, p \text{ dan } k = 1, 2, \dots, q$$

(model signifikan)

Statistik uji:

$$\Lambda = \frac{|E|}{|E+H|} = \frac{|Y^T Y - \hat{\beta}^T X^T Y|}{|Y^T Y - n\bar{y}\bar{y}^T|} \quad (5)$$

Λ adalah nilai Wilk's Lamda, \bar{y} adalah vektor rata-rata matriks Y

Kriteria uji yang diharapkan:

Jika $\Lambda_{hitung} \leq \Lambda_{tabel}$ maka H_0 ditolak artinya secara keseluruhan parameter tidak sama dengan nol sehingga model signifikan.

2) Pengujian Parsial

Pengujian parsial dilakukan untuk melihat pengaruh signifikan dari setiap variabel prediktor $X_j (k=1, 2, \dots, p)$ terhadap variabel respon $Y_k (k=1, 2, \dots, q)$ secara parsial.

Hipotesis:

$$H_0 : \beta_{jk} = 0$$

(parameter regresi prediktor ke-j tidak berpengaruh signifikan terhadap variabel respon ke-k).

$$H_1 : \beta_{jk} \neq 0$$

(parameter regresi prediktor ke-j berpengaruh signifikan terhadap variabel respon ke-k).

Statistik uji:

$$\Lambda = \frac{|E|}{|E+H|} = \frac{|Y^T Y - \hat{\beta}^T X^T Y|}{|Y^T Y - \hat{\beta}_j^T X_j^T Y|} \quad (6)$$

Kriteria uji yang diharapkan:

Jika $\Lambda_{hitung} \leq \Lambda_{tabel}$ maka H_0 ditolak, artinya variabel prediktor p berpengaruh signifikan terhadap variabel respon secara parsial.

7) Pengujian asumsi residual IIDN

Dalam analisis regresi multivariat ada tiga asumsi residual yang harus dipenuhi yaitu residual harus identik, independen, dan berdistribusi normal.

1) Uji Asumsi Residual Identik

Untuk menguji syarat ini dapat digunakan statistik uji Box's M [8].

Hipotesis :

$$H_0 : \sum_1 = \sum_2 = \dots = \sum_q = \sum$$

$$H_1 : \text{minimal ada satu } \sum_k \neq \sum_l, \text{ untuk } k \neq l, k, l = \{1, 2, \dots, q\}$$

Statistik uji:

$$u = -2(1 - c_1) \ln M$$

$$\text{Dengan } \ln M = \frac{1}{2} \sum_{k=1}^q v_k \ln |S_k| - \frac{1}{2} \left(\sum_{k=1}^q v_k \right) \ln |S_{pool}|$$

$$\text{Dimana } S_{pool} = \frac{\sum_{k=1}^q v_k S_k}{\sum_{k=1}^q v_k} \quad (7)$$

$$c_1 = \left[\sum_{k=1}^q \frac{1}{v_k} - \frac{1}{\sum_{k=1}^q v_k} \right] \left[\frac{2p^2 + 3p - 1}{6(p+1)(k-1)} \right]$$

Keterangan:

V_k : $n_k - 1$

k : banyak kelompok residual

p : banyak variabel prediktor

q : banyak variabel respon

S_k : matriks varian – kovarian residual ke- k

n_k : jumlah observasi untuk variabel respon ke- k

Kriteria uji yang diharapkan:

Jika $u \leq \chi_{tabel}^2$, maka H_0 tidak ditolak. Artinya matriks varian – kovarian residual adalah homogen atau dapat disimpulkan bahwa residual tersebut identik.

2) Uji Asumsi Residual Independen

Untuk menguji kebebasan antar residual digunakan uji Barlett Sphericity.

Hipotesis:

H_0 : Residual bersifat independent

H_1 : Residual bersifat dependent

Statistik uji:

$$\chi_{hitung}^2 = -\left\{n - 1 - \frac{2q+5}{6}\right\} \ln|R| \quad (8)$$

q adalah jumlah variabel respon, \mathbf{R} adalah matriks korelasi variabel respon, $\ln|R|$ adalah nilai \ln dari determinan matriks korelasi variabel respon.

Kriteria uji yang diharapkan:

Jika $\chi_{hitung}^2 \leq \chi_{tabel}^2 = \chi_{\alpha, \frac{1}{2}q(q-1)}^2$, maka H_0 tidak ditolak artinya antar residual bersifat independen.

3) Uji Asumsi Residual Berdistribusi Normal Multivariat

Untuk melihat atau memeriksa asumsi ini dapat dilakukan dengan menghitung jarak kuadrat pada setiap residual pengamatan [3].

Hipotesis:

H_0 : Residual berdistribusi normal multivariat

H_1 : Residual tidak berdistribusi normal multivariat

Statistik uji:

$$d_i^2 = (\hat{\epsilon}_i - \bar{\epsilon})^T S^{-1} (\hat{\epsilon}_i - \bar{\epsilon}), \quad i = 1, 2, \dots, n \quad (9)$$

Dengan

ϵ_i : vektor residual ke- i

$\bar{\epsilon}$: vektor rata-rata residual

S^{-1} : invers matriks varian-kovarian residual

Kriteria uji yang diharapkan:

Jika diperoleh nilai $d_i^2 \leq \chi_{q, 0.05}^2$ terhadap lebih dari $\frac{1}{2} n$ pengamatan, maka H_0 gagal ditolak artinya residual dikatakan berdistribusi normal multivariat.

III. Hasil dan Pembahasan

3.1 Statistik Deskriptif

Statistik deskriptif yang digunakan dalam adalah nilai rata-rata, nilai maksimum dan minimum. Selain itu data disajikan dalam *scatterplot* untuk melihat sebaran data secara visual. Tabel 1 berikut adalah statistik deskriptif untuk masing-masing variabel

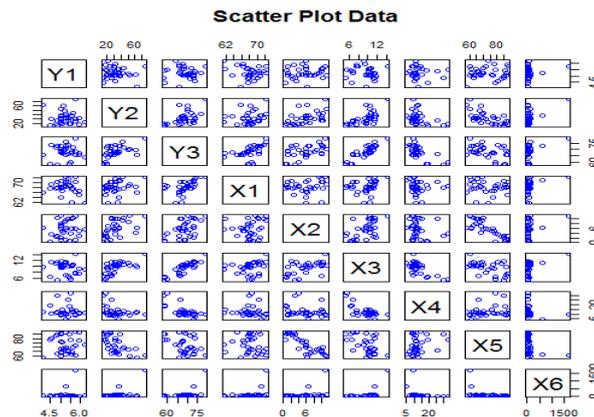
Tabel 1. Statistik Deskriptif Data

Variabel	Rata-rata	Maksimal	Minimal
Y_1	5,25	6,22	4,11
Y_2	33,16	74,47	14,99

Pemilihan Model Regresi Multivariat Terbaik Dengan Kriteria.....

Y_3	68,43	78,87	58,25
X_1	68,07	72,29	61,58
X_2	5,84	11,39	0,40
X_3	9,55	14,19	5,21
X_4	12,95	32,62	4,74
X_5	72,80	88,38	57,72
X_6	123,93	1.679,24	5,94

Untuk melihat sebaran data dari semua variabel penelitian disajikan data Gambar 1 berikut



Gambar 1. Scatterplot Data Variabel Respon dan Prediktor

3.2 Uji Asumsi Kebebasan Antar Variabel Respon

Uji Hipotesis:

H_0 :Antar variabel respon Pertumbuhan Ekonomi (Y_1), PDRB Perkapita (Y_2) dan IPM (Y_3) bersifat independen.

H_1 :Antar variabel respon Pertumbuhan Ekonomi (Y_1), PDRB Perkapita (Y_2) dan IPM (Y_3) bersifat dependen.

Statistik Uji:

$$\chi_{hitung}^2 = -\left\{n - 1 - \frac{2q+5}{6}\right\} \ln|R|$$

$$\chi_{hitung}^2 = -\left\{33 - 1 - \frac{2x3+5}{6}\right\} \ln \begin{vmatrix} 1 & -0,062 & -0,002 \\ -0,062 & 1 & 0,553 \\ -0,002 & 0,553 & 1 \end{vmatrix}$$

$$\chi_{hitung}^2 = -\left\{33 - 1 - \frac{2x3+5}{6}\right\} (-0,3708)$$

$$\chi_{hitung}^2 = 11,1858$$

$$\chi_{tabel}^2 = \chi_{0.05,3}^2 = 7,815$$

Kriteria Uji:

Karena $\chi_{hitung}^2 > \chi_{tabel}^2 = 7,815$, maka H_0 ditolak, jadi dapat disimpulkan bahwa hubungan antar variabel respon Pertumbuhan Ekonomi (Y_1), PDRB Perkapita (Y_2) dan IPM (Y_3) bersifat dependen.

3.3 Uji Distribusi Normal Multivariat Variabel Respon

Dalam melakukan pengujian normal multivariat langkah pertama yang harus dilakukan adalah menghitung matriks varians kovarians variabel respon. Diperoleh matriks varians kovarians

$$s = \begin{bmatrix} 0,20 & -0,3941 & -0,0033 \\ -0,3941 & 201,5024 & 38,0933 \\ -0,0033 & 38,0933 & 23,5234 \end{bmatrix}$$

kemudian dilakukan pengujian normal multivariat

Hipotesis:

H_0 : Variabel respon berdistribusi normal multivariat

H_1 : Variabel respon tidak berdistribusi normal multivariat

Statistik uji:

$$d_i^2 = (\mathbf{Y}_i - \bar{\mathbf{Y}})^T \mathbf{S}^{-1} (\mathbf{Y}_i - \bar{\mathbf{Y}})$$

$$\chi_{tabel}^2 = \chi_{0,5,3}^2 = 2,37$$

Kriteria uji:

Berdasarkan Tabel nilai d_i^2 diperoleh nilai d_i^2 dari masing-masing pengamatan, dimana dari 33 pengamatan terdapat 18 pengamatan yang memiliki nilai $d_i^2 < \chi_{tabel}^2$ atau $d_i^2 < 2,37$, sehingga data variabel respon dikatakan normal multivariat. Berikut nilai hasil nilai jarak dari setiap pengamatan yang dirangkum dalam tabel 2

Tabel 2. Nilai d_i^2 Variabel Respon

i	d_i^2	i	d_i^2	i	d_i^2
1	4.1065	12	0.8501	23	1.6565
2	5.5915	13	0.3351	24	3.4019
3	0.4223	14	7.1716	25	5.1545
4	1.0649	15	0.4098	26	1.5757
5	3.8459	16	3.2635	27	0.9653
6	4.3607	17	1.6995	28	2.738
7	2.1303	18	0.1986	29	2.875
8	0.7122	19	13.4813	30	10.4893
9	0.566	20	2.502	31	2.0219
10	0.9481	21	1.5116	32	3.7685
11	1.0661	22	4.7589	33	0.3568

3.4 Pemilihan Model Terbaik dengan Kriteria KICC

Untuk mendapatkan nilai KICC terlebih dahulu menghitung nilai varians kovarians error ($\hat{\Sigma}$) dari masing-masing kombinasi prediktor yang telah diregresikan, hasilnya dapat dilihat pada Tabel 3 dibawah ini.

Tabel 3. Nilai KICC Variabel Prediktor

No	Prediktor	KICC
1	X ₁	401.6189
2	X ₂	396.5525
3	X ₃	354.8281
4	X ₄	391.2557
5	X ₅	395.3149
6	X ₆	397.1396
7	X ₁ X ₂	386.3683
8	X ₁ X ₃	332.3537
9	X ₁ X ₄	373.9228
10	X ₁ X ₅	384.7090
11	X ₁ X ₆	388.2817
12	X ₂ X ₃	349.1381
13	X ₂ X ₄	373.6228
No	Prediktor	KICC
14	X ₂ X ₅	385.0383
15	X ₂ X ₆	384.6858
16	X ₃ X ₄	352.3036

[Type here]

Pemilihan Model Regresi Multivariat Terbaik Dengan Kriteria.....

17	X_3X_5	347.2692
18	X_3X_6	348.2769
19	X_4X_5	376.8314
⋮	⋮	⋮
40	$X_3X_5X_6$	341.1850
41	$X_4X_5X_6$	365.1821
42	$X_1X_2X_3X_4$	318.2085

43	$X_1X_2X_3X_5$	315.0502
44	$X_1X_2X_3X_6$	316.8464
45	$X_1X_2X_4X_5$	339.0617
46	$X_1X_2X_4X_6$	344.8407
47	$X_1X_2X_5X_6$	364.4459
⋮	⋮	⋮
63	$X_1X_2X_3X_4X_5X_6$	292.4145

Berdasarkan tabel diatas didapatkan nilai KICC terkecil pada model ke-63 sebesar 292,4145 untuk semua variabel prediktor yang ada yaitu angka harapan hidup, tingkat pengangguran, pengeluaran perkapita, persentase kemiskinan, tingkat persentase angkatan kerja (TPAK) dan anggaran pendapatan asli daerah (APAD).

3.5 Estimasi Paramater Regresi Multivariat

Berdasarkan kriteria uji KICC, variabel prediktor yang dipilih untuk faktor-faktor yang berpengaruh terhadap ketiga variabel respon adalah variabel prediktor X_1 , X_2 , X_3 , X_4 , X_5 , X_6 . Selanjutnya dilakukan estimasi parameter model. Hasil estimasi parameter disajikan dalam Tabel 4 berikut.

Tabel 4. Estimasi Parameter

	Y_1	Y_2	Y_3
β_0	6,93	51,7	1,98
β_1	-0,0485	0,040	0,623
β_2	0,1215	-0,58	0,524
β_3	-0,1370	2,95	1,415
β_4	-0,0317	0,149	-0,2136
β_5	0,0349	-0,678	0,1417
β_6	0,000679	0,01115	-0,00048

Hasil dari tabel diatas didapatkan model regresi multivariat untuk masing-masing variabel respon Pertumbuhan Ekonomi (Y_1), PDRB Perkapita (Y_2) dan IPM (Y_3). Berikut ini model hasil regresi multivariat yang didapatkan:

$$Y_1 = 6,93 - 0,0485 X_1 + 0,1215 X_2 - 0,1370 X_3 - 0,0317 X_4 + 0,0349 X_5 + 0,000679 X_6$$

$$Y_2 = 51,7 + 0,040 X_1 - 0,58 X_2 + 2,95 X_3 + 0,149 X_4 - 0,678 X_5 + 0,01115 X_6$$

$$Y_3 = 1,98 + 0,623 X_1 + 0,524 X_2 + 1,415 X_3 - 0,2136 X_4 + 0,1417 X_5 - 0,00048 X_6$$

3.6 Uji Signifikansi Parameter

3.6.1 Uji Simultan

Pengujian secara simultan atau keseluruhan dilakukan untuk mengetahui apakah parameter berpengaruh secara signifikan terhadap model secara simultan atau keseluruhan. Untuk pengujian signifikansi model secara simultan digunakan uji Wilk's Lamda.

Hipotesis:

$$H_0 : \beta_{jk} = 0 \text{ (model tidak signifikan), } j = 1, 2, \dots, p \text{ dan } k = 1, 2, \dots, q$$

$$H_1 : \text{Paling sedikit ada satu } \beta_{jk} \neq 0 \text{ (model signifikan).}$$

Statistik uji

$$\Lambda = \frac{|E|}{|E + H|} = \frac{|Y^T Y - \hat{\beta}^T X^T Y|}{|Y^T Y - n \bar{y} \bar{y}^T|} = 0,02644$$

$$\Lambda_{tabel} = \Lambda_{0,05,3,6,29} = 0,341$$

Kriteria uji

Karena nilai Wilks Lambda hitung kurang dari nilai $\Lambda_{tabel} = \Lambda_{0,05,3,6,29} = 0,341$, sehingga kesimpulannya adalah tolak H_0 , artinya secara keseluruhan atau serentak, paling tidak ada satu parameter yang berpengaruh signifikan terhadap model.

3.6.2 Uji Parsial

Untuk pengujian signifikansi model secara parsial digunakan uji Wilk's Lamda. Nilai Wilks Lambda hitung masing-masing variabel prediktor dirangkum dalam Tabel 5 uji parsial berikut.

Tabel 5. Uji Parsial

Prediktor	Λ_{hitung}	Λ_{tabel}	Kriteria uji	Kesimpulan
X_1	0,0001119776	0,774	$\Lambda_{hitung} < \Lambda_{tabel}$	H_0 ditolak
X_2	0,00006593908	0,774	$\Lambda_{hitung} < \Lambda_{tabel}$	H_0 ditolak
X_3	0,000167123	0,774	$\Lambda_{hitung} < \Lambda_{tabel}$	H_0 ditolak
X_4	0,00006103525	0,774	$\Lambda_{hitung} < \Lambda_{tabel}$	H_0 ditolak
X_5	0,0002391941	0,774	$\Lambda_{hitung} < \Lambda_{tabel}$	H_0 ditolak
X_6	0,00007997606	0,774	$\Lambda_{hitung} < \Lambda_{tabel}$	H_0 ditolak

Hipotesis:

H_0 : $\beta_{jk} = 0$ (parameter regresi prediktor ke- j tidak berpengaruh signifikan terhadap variabel respon ke- k).

H_1 : $\beta_{jk} \neq 0$ (parameter regresi prediktor ke- j berpengaruh signifikan terhadap variabel respon ke- k).

Statistik uji:

$$\Lambda = \frac{|E|}{|E + H|} = \frac{|Y^T Y - \hat{\beta}^T X^T Y|}{|Y^T Y - \hat{\beta}_j^T X_j^T Y|}$$

$$\Lambda_{tabel} = \Lambda_{0,05,3,1,32} = 0,774$$

Kriteria Uji:

$\Lambda_{hitung} \leq \Lambda_{tabel} \rightarrow$ Tolak H_0 , Sehingga dapat disimpulkan variabel prediktor p berpengaruh secara signifikan terhadap variabel respon yang ada secara parsial.

3.7 Hubungan Antara Variabel Respon dan Variabel Prediktor

Pada regresi multivariat, nilai yang digunakan untuk mengetahui besarnya hubungan antara variabel respon dan variabel prediktor adalah *eta square lambda* (η_{Λ}^2). Nilai adalah *eta square lambda* (η_{Λ}^2) diperoleh dengan persamaan berikut:

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

$$\eta_{\Lambda}^2 = 1 - 0,026444 = 0,97556$$

Nilai η_{Λ}^2 sebesar 0,97556 dapat disimpulkan bahwa model regresi tersebut dapat menjelaskan informasi data sebesar 97,556 % sedangkan sisanya sebesar 2,244 % dijelaskan oleh variabel prediktor yang lain yang tidak masuk dalam penelitian.

3.8 Uji Asumsi Residual

3.8.1 Uji Asumsi Residual Identik

Asumsi yang harus dipenuhi dalam melakukan pemodelan regresi multivariat adalah residual memiliki matriks varians-kovarian yang homogen. Untuk menguji syarat ini digunakan uji *Box's M*.

Hipotesis:

[Type here]

Pemilihan Model Regresi Multivariat Terbaik Dengan Kriteria.....

H_0 : $\Sigma_1 = \Sigma_2 = \dots = \Sigma_q = \Sigma$ (matriks varians kovarians residual homogen/identik).

H_1 : minimal ada satu $\Sigma_k \neq \Sigma_l$, untuk $k \neq l, k, l = \{1, 2, \dots, q\}$ (Matriks varians kovarian residual heterogen).

Statistik uji:

$$u = -2(1 - c_1) \ln M = -88,02125$$

$$\chi_{tabel}^2 = \chi_{\alpha, \frac{1}{2}(k-1)p(p+1)}^2 = 58,12404$$

Kriteria uji:

Karena nilai $u <$ dari nilai $\chi_{tabel}^2 = 58,12404$, maka H_0 tidak ditolak. Sehingga dapat disimpulkan bahwa matriks varians kovarian residual homogen/identik.

3.8.2 Uji Asumsi Residual Independen

Untuk menguji kebebasan antar residual digunakan uji Bartlett Sphericity.

Hipotesis:

H_0 : Residual bersifat independen

H_1 : Residual bersifat dependen

Statistik uji:

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

$$\chi_{hitung}^2 = -\left[33 - 1 - \frac{2x3+5}{6}\right] \ln \begin{bmatrix} 1 & 0,0029 & 0,1463 \\ 0,0029 & 1 & -0,0598 \\ 0,1463 & -0,0598 & 1 \end{bmatrix}$$

$$\chi_{hitung}^2 = -\left\{33 - 1 - \frac{2x3+5}{6}\right\} (-0,0254)$$

$$\chi_{hitung}^2 = 0,7663$$

$$\chi_{tabel}^2 = 7,815$$

Kriteria pengujian:

$\chi_{hitung}^2 \leq \chi_{tabel}^2 = 7,814$, Karena nilai χ_{hitung}^2 lebih kecil dibanding nilai $\chi_{0,05,3}^2$, maka H_0 tidak ditolak, sehingga dapat disimpulkan bahwa nilai residual saling bebas (independen).

3.8.3 Uji Asumsi Residual Berdistribusi Normal Multivariat

Pengujian asumsi berdistribusi normal multivariat dilakukan dengan hipotesis berikut:

Hipotesis:

H_0 : Residual berdistribusi normal multivariat

H_1 : Residual tidak berdistribusi normal multivariat

Statistik uji:

$$d_i^2 = (\hat{\boldsymbol{\epsilon}}_i - \bar{\boldsymbol{\epsilon}})^T \mathbf{S}^{-1} (\hat{\boldsymbol{\epsilon}}_i - \bar{\boldsymbol{\epsilon}})$$

$$\chi_{tabel}^2 = \chi_{0,5,q}^2 = 2,365$$

Nilai jarak mahalalanobis residual dirangkum dalam Tabel 6 berikut:

Tabel 6. Nilai d_i^2 Residual

No	d_i^2	No	d_i^2
1	7,8684	18	1,0173
2	3,7006	19	9,7972
3	2,0676	20	2,9037
4	1,5095	21	0,2750
5	2,4835	22	8,3729

6	2,0789	23	5,0309
7	4,6576	24	1,7174
8	5,4729	25	1,7672
9	1,4071	26	1,2244
10	0,3736	27	0,1009
11	1,4374	28	0,5924
12	2,2334	29	3,3817
13	1,9923	30	0,4135
14	4,4951	31	0,2309
15	2,6353	32	8,4789
16	0,6496	33	2,0090
17	3,6241		

Kriteria uji:

Dari 33 pengamatan terdapat 19 pengamatan yang memiliki nilai $d_i^2 < \chi_{tabel}^2$ atau $d_i^2 < 2,365$, sehingga H_0 tidak ditolak dan dapat disimpulkan bahwa residual dari variabel respon Pertumbuhan Ekonomi (Y_1), PDRB Perkapita (Y_2) dan IPM (Y_3) berdistribusi normal multivariat.

3.9 Penerapan & Interpretasi Model

Jika dilihat dari setiap model untuk masing-masing variabel respon, untuk pertumbuhan ekonomi pemerintah lebih memfokuskan terhadap persentase kemiskinan, TPAK, dan APAD. Sedangkan untuk PDRB pemerintah diharapkan lebih fokus terhadap angka harapan hidup, tingkat pengangguran dan APAD. Untuk IPM sendiri pemerintah diharapkan lebih fokus pada angka harapan hidup, persentase kemiskinan, dan TPAK.

IV. Kesimpulan

4.1 Kesimpulan

Berdasarkan hasil analisis dan pembahasan yang telah dilakukan sebelumnya maka dapat diambil beberapa kesimpulan sebagai berikut:

1. Dalam pemilihan model terbaik dengan kriteria KICC, diperoleh nilai KICC terkecil yaitu pada model kombinasi X_1, X_2, X_3, X_4, X_5 , dan X_6 sebesar 292,4145.
2. Model regresi untuk tingkat kesejahteraan masyarakat di provinsi Sumatera Utara adalah:
 - Pertumbuhan Ekonomi = 6,93 - 0,0485 (angka harapan hidup) + 0,1215 (tingkat pengangguran) - 0,1370 (pengeluaran perkapita) - 0,0317 (persentase kemiskinan) + 0,0349 (TPAK) + 0,000679 (APAD).
 - PDRB Per Kapita = 51,7 + 0,040 (angka harapan hidup) - 0,58 (tingkat pengangguran) + 2,95 (pengeluaran perkapita) + 0,149 (persentase kemiskinan) - 0,678 (TPAK) + 0,01115 (APAD).
 - IPM = 1,98 + 0,623 (angka harapan hidup) + 0,524 (tingkat pengangguran) + 1,415 (pengeluaran perkapita) - 0,2136 (persentase kemiskinan) + 0,1417 (TPAK) - 0,00048 (APAD).
3. Besarnya hubungan antara variabel respon dan variabel prediktor diperoleh nilai *eta square lambda* (η_{λ}^2) sebesar 0,97556. Ini dapat diartikan bahwa variabel prediktor dalam model regresi tersebut dapat menjelaskan informasi data sebesar 97,556% sedangkan sisanya sebesar 2,44 % dijelaskan oleh variabel prediktor lain yang tidak masuk dalam penelitian.

Daftar Pustaka

[Type here]

Pemilihan Model Regresi Multivariat Terbaik Dengan Kriteria.....

- [1] BPS, 2016, *Indikator Kesejahteraan Rakyat 2016*, Badan Pusat Statistik, Jakarta.
- [2] Hafidi B dan Mkhadri A, A corrected akaike criterion based on kullback's symmetric divergence: application in time series, multiple and multivariate regression, *Jurnal Computational Statistics & Data Analyst* 50, Hal 1524-1550.
- [3] Johnson, R A and Wichern D W 2002, *Applied Multivariate Statistical Analysis Six Edition*, New Jersey.
- [4] Morrison D F, 2005, *Multivariate Statistical Methods Fourth Edition*, School University of Pennsylvania, Pennsylvania.
- [5] Mattjik A A dan Sumertajaya I M, 2011, *Sidik Peubah Ganda dengan Menggunakan SAS*, Departemen Statistika IPB, Bogor.
- [6] Morrison D F, 2005, *Multivariate Statistical Methods Fourth Edition*, School University of Pennsylvania, Pennsylvania.
- [7] Noeryanti, 2012, *Metode Statistika II*, Akprind Press, Yogyakarta.
- [8] Rencher A C, 2002, *Method of Multivariat Analysis Second Edition*, John Wiley and Sons Inc, New York.
- [9] Sarah I A, 2015, Kullback's Information Criterion Correction (KICC) untuk seleksi model regresi linear multivariat, Skripsi, UGM, Yogyakarta.
- [10] Walpole R E, 1995, *Ilmu Peluang dan Statistika Untuk Insinyur dan Ilmuwan*, Itbpress, Bandung.