

PENDEKATAN REGRESI LINIER MULTIVARIAT UNTUK PEMODELAN INDEKS PEMBANGUNAN MANUSIA (IPM) DAN PERSENTASE PENDUDUK MISKIN DI JAWA TENGAH

Abdul Hoyyi¹, Diah Safitri², Sugito³, Alan Prahutama⁴

^{1,2,3,4} Departemen Statistika, FSM Universitas Diponegoro

Email :

ABSTRAK

Model regresi multivariat merupakan model regresi yang dibangun dari beberapa variabel independen dan mempunyai variabel dependen lebih dari satu dengan setiap variabel dependen saling berkorelasi. Pada penelitian ini variabel dependennya adalah Indeks Pembangunan Manusia (IPM) dan jumlah penduduk miskin, sedangkan variabel independennya adalah upah minimum regional dan kepadatan penduduk. Data yang digunakan adalah data sekunder yang diperoleh dari Badan Pusat Statistika (BPS) Propinsi Jawa Tengah. Parameter pada model diestimasi dengan metode kuadrat terkecil. Berdasarkan hasil dan pembahasan, pada taraf signifikansi 5 % diperoleh hasil bahwa variabel IPM dan persentase penduduk miskin berdistribusi normal multivariat. Pengujian parameter model diperoleh bahwa koefisien variabel upah minimum regional dan kepadatan penduduk signifikan terhadap model. Pengujian asumsi normalitas, homoskedastisitas dan nonautokorelasi memberikan kesimpulan error berdistribusi normal multivariate, tidak terjadi autokorelasi dan varian dari error homogen. Hasil akhir memberikan kesimpulan bahwa variabel upah minimum regional dan kepadatan penduduk dapat menjelaskan Indeks Pembangunan Manusia dan persentase penduduk miskin sebesar 70,11 %.

Kata kunci : Regresi multivariat, IPM, BPS.

PENDAHULUAN

Menurut [8], analisis regresi adalah suatu analisis statistik yang digunakan untuk menjelaskan suatu variabel dependen (Y) menggunakan satu atau lebih variabel independen (X_1, \dots, X_k). Jika $k=1$ maka regresi yang terbentuk disebut *simple regression* (regresi sederhana), sedangkan jika $k > 1$ maka regresi yang terbentuk disebut *multiple regression* (regresi berganda). Menurut [6], model regresi multivariat merupakan model regresi yang dibangun dari beberapa variabel independen dan mempunyai variabel dependen lebih dari satu dengan setiap variabel dependen saling berkorelasi. Diberikan p variabel

dependen yaitu y_1, y_2, \dots, y_p dengan q variabel independen yaitu x_1, x_2, \dots, x_q . Vektor y sebanyak n pengamatan dapat ditulis menjadi baris berdasarkan matriks berikut :

$$Y = \begin{pmatrix} y_{11} & y_{12} & \dots & y_{1p} \\ y_{21} & y_{22} & \dots & y_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n1} & y_{n2} & \dots & y_{np} \end{pmatrix} = \begin{pmatrix} y'_1 \\ y'_2 \\ \vdots \\ y'_n \end{pmatrix}$$

Sebanyak n buah pengamatan dari x_1, x_2, \dots, x_q dapat ditulis sama dengan matriks X pada model regresi berganda seperti berikut :

$$X = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1q} \\ 1 & x_{21} & x_{22} & \dots & x_{2q} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nq} \end{pmatrix}$$

Model regresi multivariat dapat dituliskan sebagai berikut :

$$Y = XB + \varepsilon$$

Menurut [6], untuk mendapatkan estimasi yang baik, digunakan asumsi berikut :

1. $E(Y) = XB$ atau $E(\varepsilon) = \mathbf{0}$.
2. $cov(y_i) = \Sigma$ untuk semua $i = 1, 2, \dots, n$ dimana y_i' adalah baris ke- i dari matriks Y .
3. $cov(y_i, y_j) = \mathbf{0}$ untuk semua $i \neq j$.

Seperti dalam model regresi linier sederhana dalam kasus multivariat estimasi kuadrat terkecil juga meminimumkan jumlah kuadrat error atau dalam notasi matriks sama dengan meminimumkan $tr(Y - XB)'(Y - XB)$ karena $tr(Y - XB)'(Y - XB) = \sum_{i=1}^n \sum_{j=1}^p \hat{\varepsilon}_{ij}^2$. Dalam [3], matriks error yang didapatkan adalah :

$$\hat{\varepsilon} = Y - XB$$

$$E = \hat{\varepsilon}'\hat{\varepsilon} = (Y - XB)'(Y - XB) = Y'Y - 2B'X'Y + B'X'XB$$

Atau dapat ditulis seperti berikut :

$$E = Y'Y - B'X'XB$$

Persamaan dapat diturunkan dengan menggunakan aturan penurunan matriks berikut [2] :

1. $\frac{\partial(a'x)}{\partial x} = a$
2. $\frac{\partial(x'Ax)}{\partial x} = 2Ax$
3. $\frac{\partial(x'Ax)}{\partial x} = 2x'A$

Dalam [4], estimasi \hat{B} diperoleh sebagai berikut :

$$X'X\hat{B} = X'Y$$

$$\hat{B} = (X'X)^{-1} X'Y$$

Uji signifikansi model menurut [5] menyatakan bahwa pengujian ini dimaksudkan untuk melihat apakah seluruh parameter bernilai nol atau tidak, artinya jika semua parameter secara bersama-sama bernilai nol maka tidak ada pengaruh dari variabel independen terhadap variabel dependen. Pengujian Subset X menurut [7] , uji ini dilakukan

untuk menguji apakah subset X signifikan terhadap model.

Menurut [6], indikator yang digunakan untuk mengukur hubungan antara variabel dependen dan variabel independen pada regresi linier multivariat adalah *Eta Squared Lambda* yang dinyatakan dengan persamaan berikut :

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

Nilai η_{Λ}^2 berada pada selang 0 sampai 1. Semakin nilai η_{Λ}^2 mendekati 1 maka hubungan antara variabel dependen dan variabel independen semakin erat.

Menurut [3], sama halnya dengan asumsi eror pada regresi linier univariat, pada model regresi linier multivariat terdapat asumsi yang harus dipenuhi yaitu normalitas, homoskedastisitas, dan nonautokorelasi.

METODE PENELITIAN

Sumber Data dan Variabel Penelitian

Data yang digunakan dalam penelitian ini adalah data sekunder tentang faktor-faktor yang diduga mempengaruhi Indeks Pembangunan Manusia (IPM) dan persentase penduduk miskin di Provinsi Jawa Tengah yang diperoleh dari Badan Pusat Statistika (BPS). Dalam penelitian ini, terdapat 29 kabupaten dan 6 kota di Jawa Tengah yang dijadikan lokasi pengamatan. Data yang digunakan merupakan data pada tahun 2011 dan 2012 yang diperoleh melalui survey oleh BPS.

Penelitian ini menggunakan Indeks Pembangunan Manusia (IPM) dan persentase penduduk miskin sebagai variabel dependen. Variabel independen ditentukan dengan mengidentifikasi faktor-faktor yang diduga mempengaruhi IPM dan persentase penduduk miskin. variabel independen yang digunakan dalam penelitian ini adalah upah minimum regional dan kepadatan penduduk. Tabel 1. merupakan

ringkasan dari variabel yang digunakan dalam penelitian.

Tabel 1. Variabel Penelitian

Variabel	Keterangan
Y1	Indeks Pembangunan Manusia
Y2	Jumlah penduduk miskin
X1	Upah Minimum Regional
X2	Kepadatan Penduduk

Metode Analisis

Langkah analisis pada pemodelan Indeks Pembangunan Manusia (IPM) dan persentase penduduk miskin menggunakan analisis regresi linier multivariat adalah sebagai berikut :

1. Melakukan identifikasi variabel dependen dan variabel independen yang akan digunakan.
2. Melakukan eksplorasi dan statistika deskriptif untuk mengetahui karakteristik data pada variabel dependen (Indeks Pembangunan Manusia (IPM) dan persentase penduduk miskin) dan variabel independen (upah minimum regional dan kepadatan penduduk).
3. Melakukan uji asumsi pada variabel dependen (Indeks Pembangunan Manusia (IPM) dan persentase penduduk miskin) berupa uji kenormalan multivariat dengan uji *Kolmogorov-Smirnov* serta uji kebebasan dengan menggunakan uji *Bartlett of Sphericity* antar variabel dependen untuk menentukan pemakaian analisis regresi yang digunakan.
4. Melakukan estimasi parameter model.
5. Melakukan uji signifikansi parameter model.
6. Melakukan uji subset X.
7. Uji asumsi error model yang meliputi uji normal multivariat dengan uji *Kolmogorov-Smirnov*, kehomogenan ragam dengan menggunakan *dd-plot*, uji kebebasan dengan menggunakan uji *Durbin-Watson*.

8. Menginterpretasikan model akhir yang telah didapat.

HASIL PENELITIAN

Statistika Deskriptif

Provinsi Jawa Tengah merupakan Provinsi yang terdiri dari 6 Kota dan 29 Kabupaten. Menurut [1], di Provinsi Jawa Tengah dari kurun waktu antara tahun 2011 sampai 2012 terjadi peningkatan pada nilai IPM. Pada tahun 2012, IPM di Provinsi Jawa Tengah berada pada nilai 73,36 dapat dikategorikan bahwa status pembangunan manusia adalah menengah atas. Situasi perekonomian yang semakin membaik menyebabkan berkurangnya penduduk yang hidup dibawah garis kemiskinan. Hal ini ditunjukkan dengan berkurangnya jumlah penduduk miskin dari tahun ke tahun.

Pada tahun 2011, jumlah penduduk miskin sebesar 5,26 juta atau sekitar 16,21 persen dari total penduduk Jawa Tengah. Kemudian jumlah penduduk miskin tahun 2012 menurun menjadi 4,86 juta atau sekitar 14,98 persen dari total penduduk Jawa Tengah. Banyak faktor yang mempegaruhi peningkatan IPM (Y_1) dan penurunan persentase penduduk miskin (Y_2) antara lain upah minimum regional (X_1) dan kepadatan penduduk (X_2). Pada penelitian ini IPM persentase penduduk miskin menjadi variabel dependen, sedangkan upah minimum regional dan kepadatan penduduk menjadi variabel independen. Tabel 2 berikut menyatakan statistik deskriptif dari variabel dependen dan independen.

Tabel 2. Statistik Deskriptif Data Penelitian

	Y_1	Y_2	X_1	X_2
Rata-rata	73,14	15,00	8,11	1967,70
Median	72,93	14,83	8,04	1057,50
Modus	71,86	10,04	7,95	818,00
Varian	4,56	22,24	0,35	5619483,60
Minimum	68,81	5,13	7,17	465,00
Maximum	78,60	24,21	9,92	11573,00

Pengujian Variabel Dependen Uji Normalitas Variabel Dependen

Berikut ini merupakan uji normal multivariat untuk variabel dependen, yaitu IPM (Y_1) dan persentase penduduk miskin (Y_2). Uji ini menggunakan uji formal dengan uji *Kolmorov-Smirnov* sebagai berikut :

Hipotesis :

H_0 : Data berdistribusi normal multivariat

H_1 : Data tidak berdistribusi normal multivariat

Taraf Signifikansi : $\alpha = 5\%$

Statistik uji :

$$D = 0,0691$$

Kriteria Uji :

H_0 ditolak jika $D > W_{(1-\alpha)}$ atau nilai *p-value* $< \alpha$, dengan $W_{(1-\alpha)}$ merupakan kuantil $1 - \alpha$ pada tabel *Kolmogorov-Smirnov*.

Keputusan :

H_0 diterima karena $D(0,0691) < W_{(0,95)}(0,160)$

Kesimpulan :

Pada taraf signifikansi $\alpha = 0,05$ diperoleh bahwa variabel IPM (Y_1) dan persentase penduduk miskin (Y_2) berdistribusi normal multivariat.

Pengujian *Independency* antar Variabel

Pengujian kebebasan antar variabel IPM(Y_1) dan persentase penduduk miskin (Y_2) dapat menggunakan uji *Bartlett of Sphericity* berikut:

Hipotesis :

H_0 : $\mathbf{R} = \mathbf{I}$ (Antar variabel saling bebas)

H_1 : $\mathbf{R} \neq \mathbf{I}$ (Antar variabel tidak saling bebas)

Taraf Signifikansi : $\alpha = 5\%$

Statistik Uji : (sesuai Lampiran 3)

$$\chi^2_{(hit)} = -\left\{n - 1 - \frac{2p + 5}{6}\right\} \ln|\mathbf{P}|$$

$$= -\left\{70 - 1 - \frac{2(2)+5}{6}\right\} \ln \begin{vmatrix} 1 & -0,64 \\ -0,64 & 1 \end{vmatrix}$$

$$= -67,5 \ln 0,6$$

$$= 35,39$$

Kriteria Uji :

H_0 ditolak jika $\chi^2_{(hit)} \geq \chi^2_{(\alpha; \frac{1}{2}m(m-1))}$

Keputusan : H_0 ditolak karena $\chi^2_{(hit)} = 35,39 \geq \chi^2_{(0,05;1)} = 0,004$

Kesimpulan :

Berdasarkan taraf signifikansi $\alpha = 5\%$ diperoleh bahwa antara variabel IPM (Y_1) dan persentase penduduk miskin (Y_2) tidak saling bebas.

Berdasarkan uji normal multivariat dan *independency* antar variabel dependen maka regresi linier multivariat dapat digunakan untuk memodelkan variabel IPM (Y_1) dan persentase penduduk miskin (Y_2).

Model Awal Regresi Linier Multivariat

Berdasarkan output dengan didapatkan model awal sebagai berikut :

$$\hat{Y}_1 = 60,1645 + 1,4680 X_1 + 0,0005 X_2$$

$$\hat{Y}_2 = 50,2346 - 4,1924 X_1 - 0,0006 X_2$$

Diperoleh matriks \mathbf{E} adalah sebagai berikut :

$$\mathbf{E} = \mathbf{Y}'\mathbf{Y} - \hat{\mathbf{B}}'\mathbf{X}'\mathbf{Y}$$

$$|\mathbf{E}| = 85427,2$$

Diperoleh matriks \mathbf{H} adalah sebagai berikut :

$$\mathbf{H} = \hat{\mathbf{B}}'\mathbf{X}'\mathbf{Y} - n\bar{y}\bar{y}'$$

Diperoleh matriks $\mathbf{E}+\mathbf{H}$ adalah sebagai berikut :

$$\mathbf{E} + \mathbf{H} = \mathbf{Y}'\mathbf{Y} - n\bar{y}\bar{y}'$$

$$|\mathbf{E} + \mathbf{H}| = 285826,4$$

Diperoleh nilai wilk's lambda sebagai berikut :

$$\Lambda = \frac{|\mathbf{E}|}{|\mathbf{E} + \mathbf{H}|} = 0,2989$$

Uji Signifikansi Model

Hipotesis :

$$H_0 : \mathbf{B}_1 = \mathbf{0}$$

$$H_1 : \mathbf{B}_1 \neq \mathbf{0}$$

Taraf Signifikansi : $\alpha = 5\%$

Statistik Uji :

$$\Lambda = \frac{|\mathbf{E}|}{|\mathbf{E} + \mathbf{H}|} = 0,2989$$

Kriteria Uji :

$$H_0 \text{ ditolak jika } \Lambda \leq \Lambda_{\alpha,p,q,n-q-1}$$

Keputusan :

$$H_0 \text{ ditolak karena } 0,2989 <$$

$$\Lambda_{0,05;2;2,67} (0,8652)$$

Kesimpulan :

Pada taraf signifikansi $\alpha = 0,05$ diperoleh bahwa ada pengaruh nyata atau secara bersama sama variabel independen mempengaruhi model.

Uji Subset X

Hipotesis :

$$H_0 : \mathbf{B}_d = \mathbf{0}$$

$$H_1 : \mathbf{B}_d \neq \mathbf{0}$$

Taraf Signifikansi : $\alpha = 5\%$

Statistik uji :

Tabel 3. Pengujian Subset X Model

Variabel	$\Lambda_{\text{(hitung)}}$	Λ_{Tabel}	Keputusan
X_1	0,5888	0,9114	H_0 ditolak
X_2	0,5058	0,9114	H_0 ditolak

Kriteria Uji :

H_0 ditolak jika $\Lambda \leq \Lambda_{\alpha,p,h,(n-q-1)}$ dimana h merupakan banyaknya variabel independen yang direduksi pada model tereduksi.

Kesimpulan :

Pada taraf signifikansi $\alpha = 5\%$ diperoleh bahwa koefisien untuk variabel upah minimum regional dan kepadatan penduduk signifikan terhadap model.

Uji Asumsi Eror

Berikut merupakan asumsi yang harus terpenuhi pada model regresi linier multivariat

Normalitas

Pengujian normal multivariat dengan menggunakan *Kolmogorov-Smirnov* seperti berikut :

Hipotesis :

H_0 : Error berdistribusi normal multivariat

H_1 : Error tidak berdistribusi normal multivariat

Taraf Signifikansi : $\alpha = 5\%$

Statistik uji :

$$D = 0,1226 \text{ atau } p\text{-value} = 0,2239$$

Kriteria Uji :

H_0 ditolak jika $D > W_{(1-\alpha)}$ atau nilai $p\text{-value} < \alpha$, dengan $W_{(1-\alpha)}$ merupakan kuantil $1 - \alpha$ pada tabel *Kolmogorov-Smirnov*.

Keputusan :

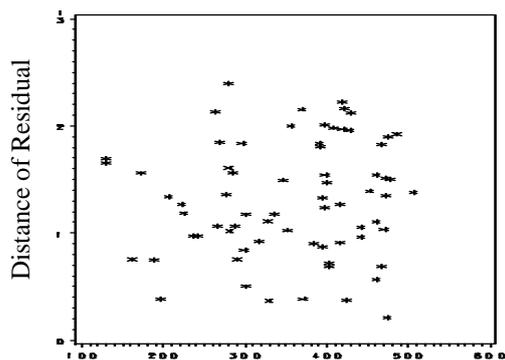
H_0 diterima karena $D(0,1226) < W_{(0,95)}(0,160)$ atau $p\text{-value} (0,2239) > \alpha(0,05)$.

Kesimpulan :

Pada taraf signifikansi $\alpha = 0,05$ didapatkan bahwa eror berdistribusi normal multivariat.

Homoskedastisitas

Berdasarkan dd-plot yang terbentuk pada **Gambar 1**, didapatkan bahwa error untuk model tidak membentuk trend sehingga dapat disimpulkan bahwa varian dari eror homogen.



Gambar 1. DD-Plot Error

Nonautokorelasi

Untuk menguji kebebasan antar eror pengamatan maka dapat digunakan uji Durbin-Watson seperti berikut :

Hipotesis :

H_0 : Eror saling bebas

H_1 : Eror tidak saling bebas

Taraf Signifikansi : $\alpha = 5\%$

Statistik uji:

Tabel 4. Uji Nonautokorelasi

	d_{hitung}	d_{tabel}	Keputusan
Y_1	1,8413	$D_l = 1,554$ $D_u = 1,672$ $4 - D_u = 2,328$	H_0 diterima
Y_2	2,1074	$D_l = 1,554$ $D_u = 1,672$ $4 - D_u = 2,328$	H_0 diterima

Kriteria Uji : H_0 diterima jika $D_u < d < 4 - D_u$

Kesimpulan :

Pada taraf signifikansi $\alpha = 0,05$ didapatkan bahwa eror untuk model saling bebas.

Model Akhir

Model regresi linier multivariat terbaik adalah sebagai berikut :

$$(\hat{Y}_1 \ \hat{Y}_2) = (1 \ X_1 \ X_2) \begin{pmatrix} 60,1645 & 50,2346 \\ 1,4680 & -4,1924 \\ 0,0005 & -0,0006 \end{pmatrix}$$

atau

$$\hat{Y}_1 = 60,1645 + 1,4680 X_1 + 0,0005 X_2$$

$$\hat{Y}_2 = 50,2346 - 4,1924 X_1 - 0,0006 X_2$$

Interpretasi model :

Apabila upah minimum regional naik 1 satuan dengan menganggap kepadatan penduduk tetap, maka IPM naik sebanyak 1,4680 % dan apabila kepadatan penduduk naik sebesar 1 satuan dengan menganggap upah minimum regional tetap, maka IPM akan bertambah 0,0005 %, sedangkan untuk persentase kemiskinan apabila upah minimum regional naik 1 satuan dengan menganggap kepadatan penduduk tetap, akan mengurangi persentase penduduk miskin sebesar 4,1924 %, dan apabila kepadatan penduduk naik sebesar 1 satuan dengan menganggap upah minimum regional tetap, akan mengurangi persentase penduduk miskin sebesar 0,0005 %.

Pada model regresi linier multivariat, indikator yang digunakan untuk mengukur hubungan antara variabel dependen dan variabel independen adalah

Eta Squared Lambda :

$$\eta_{\Lambda}^2 = 1 - \Lambda = 70,11 \%$$

Variabel upah minimum regional dan kepadatan penduduk menjelaskan Indeks Pembangunan Manusia dan persentase penduduk miskin sebesar 70,11 %.

KESIMPULAN

Berdasarkan hasil dan pembahasan pada bab sebelumnya, dapat diperoleh beberapa kesimpulan, diantaranya :

1. Model regresi linier multivariat terbaik yang didapatkan sebagai berikut:

$$\hat{Y}_1 = 60,1645 + 1,4680 X_1 + 0,0005 X_2$$

$$\hat{Y}_2 = 50,2346 - 4,1924 X_1 - 0,0006 X_2$$

2. Pada model yang didapatkan, variabel upah minimum regional dan kepadatan penduduk dapat

menjelaskan Indeks Pembangunan Manusia dan persentase penduduk miskin sebesar 70,11 %.

DAFTAR PUSTAKA

- [1] BPS. 2014. *Jawa Tengah Dalam Angka 2014*. Semarang : BPS.
- [2] Gujarati, D. 2004. *Basic Econometrics, 4th Edition*. New York : McGraw Hill Book Co.
- [3] Johnson, R. A and Wichern, D. W. 2007. *Applied Multivariate Statistical Analysis, 6th Edition*. USA : Pearson Prentice Hall.
- [4] Khatree, R. and Naik, D.N. 1999. *Applied Multivariate Statistics with SAS Software, 2nd Edition*. Cary, NC : SAS Institute Inc
- [5] Rencher, A.C. 1998. *Multivariate Statistical Interference and Applications*. Canada: John Wiley & Sons, Inc.
- [6] Rencher, A.C. 2002. *Methods of Multivariate Analysis, 2nd Edition*. Canada : John Wiley & Sons, Inc.
- [7] Rencher, A.C. and Christensen, W.F. 2012. *Methods of Multivariate Analysis, 3rd Edition*. Canada : John Wiley & Sons, Inc.
- [8] Rosadi, D. 2012. *Ekonometrika & Analisis Runtun Waktu Terapan dengan Eviews*. Yogyakarta : ANDI.