ISSN: 2339-2541
JURNAL GAUSSIAN, Volume 4, Nomor 1, Tahun 2015, Halaman 1 - 10
Online di: http://ejournal-s1.undip.ac.id/index.php/gaussian

PERBANDINGAN REGRESI KOMPONEN UTAMA DENGAN REGRESI RIDGE PADA ANALISIS FAKTOR-FAKTOR PENDAPATAN ASLI DAERAH (PAD) PROVINSI JAWA TENGAH

Agustifa Zea Tazliqoh[1], Rita Rahmawati[2], Diah Safitri[3]
[1]Mahasiswa Jurusan Statistika FSM Universitas Diponegoro
[2,3]Staf Pengajar Jurusan Statistika FSM Universitas Diponegoro

ABSTRACT

Assuming violation multicollinearity in classical regression analysis can cause estimator resulting from classical model regression inefficient. Principal components regression and ridge regression are the methods that can be used to overcome the problem of multicollinearity. This research aimed to compare between the principal components regression with ridge regression to tackle the problem of multicollinearity in the analysis of the factors that affect revenue (PAD) of the Central Java province. The data used in this research are data revenue (PAD), and factors that affect the region, such as local tax, retribution, Gross Regional Domestic Products (GRDP) at current prices, Gross Regional Domestic Products (GRDP) at constant prices, population, regional spending. Based on the coefficient of determination value and test on individual regression coefficients, the value of variance inflation factor and correlations sufficiently high among some independent variables so we can conclude the existence of a violation of multicollinearity on analysis factors PAD. Based on standard error resulting from principal components regression and ridge regression show that principal components regression results in a standard smaller error. This shows that principal component regression is better than ridge regression in solving the problem multicollinearity on analysis of factors that affects pad province of central java.

Keywords: Multicolinearity, revenue (PAD), Principal Component Regression, Ridge Regression, standard error

1. PENDAHULUAN

Pada pembentukan model regresi terdapat kemungkinan adanya hubungan antara variabel satu dengan variabel yang lain. Adanya hubungan antar variabel bebas dalam satu regresi disebut dengan multikolinieritas. Multikolinieritas menyebabkan estimator mempunyai varian yang besar akibatnya interval estimasi cenderung lebih besar sehingga membuat variabel bebas secara statistika tidak signifikan padahal nilai koefisien determinasi (R^2) tinggi sehingga sulit mendapatkan estimasi yang tepat (Widarjono, 2007). Oleh karena itu perlu dilakukan tindakan lanjutan untuk menangani multikolinieritas.

Multikolinieritas dapat diatasi dengan beberapa metode diantaranya regresi komponen utama dan regresi ridge (Montgomery dan Peck, 1991). Regresi komponen utama merupakan metode yang menggabungkan antara regresi linier dengan analisis komponen utama. Regresi komponen utama membentuk hubungan antara variabel terikat dengan komponen utama yang dipilih dari variabel bebas (Ul-Saufie *et al.*2011). Sedangkan regresi ridge memberikan estimasi koefisien regresi yang bias dengan memodifikasi metode kuadrat terkecil untuk mendapatkan pengurangan varian dengan menambahkan suatu tetapan k dalam menstabilkan koefisien (Mardikyan dan Cetin, 2008).

Pendapatan Asli Daerah (PAD) adalah pendapatan yang diperoleh berdasarkan peraturan daerah sesuai dengan peraturan perundang-undangan untuk mengumpulkan dana bagi keperluan daerah yang bersangkutan dalam membiayai kegiatannya (BPS, 2013). PAD memiliki peran yang cukup penting dalam menentukan kemampuan daerah untuk melakukan aktivitas pemerintahan dan program-program pembangunan. Oleh karena itu pemerintah daerah diharapkan mampu meningkatkan PAD (Mardiasmo, 2002). Salah satunya yaitu dengan mengoptimalkan faktor-faktor yang mempengaruhi PAD.

Berdasarkan uraian di atas maka ingin dibandingkan antara regresi komponen utama dan regresi ridge dalam mengatasi dugaan adanya masalah multikolinieritas dalam analisis faktor-faktor PAD Provinsi Jawa Tengah.

2. TINJAUAN PUSTAKA

2.1 Pendapatan Asli Daerah

Pendapatan Asli Daerah (PAD) adalah pendapatan yang diperoleh berdasarkan peraturan daerah sesuai dengan peraturan perundang-undangan untuk mengumpulkan dana bagi keperluan daerah yang bersangkutan dalam membiayai kegiatannya. PAD terdiri dari pajak daerah, retribusi daerah, hasil perusahaan milik daerah dan pengelolaan kekayaan daerah yang dipisahkan, lain-lain pendapatan asli daerah yang sah (BPS, 2013). Menurut Santoso dan Rahayu (2005) faktor–faktor yang mempengaruhi Pendapatan Asli Daerah (PAD) antara lain pengeluaran pemerintah, Produk Domestik Regional Bruto (PDRB) dan jumlah penduduk.

2.2 Analisis Regresi

Menurut Montgomery dan Peck (1991) secara umum model regresi dengan k variabel bebas dan n pengamatan dapat dituliskan sebagai berikut:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i \text{ dimana } i = 1,2,\ldots,n$$

yang dapat ditulis dalam bentuk matriks sebagai berikut:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

Dalam mengestimasi parameter model digunakan metode kuadrat terkecil dengan meminimumkan jumlah kuadarat *error* yaitu

$$S(\boldsymbol{\beta}) = \sum_{i=1}^{n} \varepsilon_i^2 = \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

Sehingga diperoleh rumus untuk mencari estimator parameter $\hat{\boldsymbol{\beta}}$ yaitu:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

Menurut Kutner *et al.*(2005) perbedaan unit satuan pada model regresi perlu dilakukan standarisasi menggunakan rumus sebagai berikut:

$$y_i^* = \frac{1}{\sqrt{n-1}}\left(\frac{y_i - \bar{y}}{s_y}\right) \qquad \text{dimana } s_y = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}}$$

$$X_{ik}^* = \frac{1}{\sqrt{n-1}}\left(\frac{X_{ik} - \overline{X_k}}{s_k}\right) \qquad s_k = \sqrt{\frac{\sum_{i=1}^{n}(X_{ik} - \overline{X_k})^2}{n-1}}$$

Sehingga diperoleh model regresi standar sebagai berikut:

$$y_i^* = \beta_1^* X_{i1}^* + \cdots + \beta_k^* X_{ik}^* + \varepsilon_i^*$$

Hubungan estimator regresi bentuk standar dengan bentuk asli adalah:

$$\beta_k = \left(\frac{s_y}{s_k}\right)\beta_k^* \text{ dan } \beta_0 = \bar{y} - \beta_1\overline{X_1} - \cdots - \beta_k\overline{X_k}$$

Menurut Montgomery dan Peck (1991) estimator kuadrat terkecil mempunyai sifat-sifat sebagai berikut:

1. $\hat{\boldsymbol{\beta}}$ adalah estimator tak bias untuk $\boldsymbol{\beta}$

    $E(\hat{\boldsymbol{\beta}}) = E[(\mathbf{X'X})^{-1}\mathbf{X'y})] = \boldsymbol{\beta}$
2. Varian dari $\hat{\boldsymbol{\beta}}$ adalah minimum yaitu:

    $Cov(\hat{\boldsymbol{\beta}}) = E([\hat{\boldsymbol{\beta}} - E(\hat{\boldsymbol{\beta}})][\hat{\boldsymbol{\beta}} - E(\hat{\boldsymbol{\beta}})]') = (\mathbf{X'X})^{-1}\sigma^2$

Dalam mengukur ketepatan model regresi digunakan beberapa uji signifikansi parameter antara lain:

1. Uji signifikansi regresi digunakan untuk menentukan apakah ada hubungan linier antara variabel terikat y dengan variabel bebas $X_1, X_2, \ldots, X_k$ dengan rumusan hipotesis sebagai berikut:

    $H_0 : \beta_1 = \beta_2 = \cdots = \beta_k = 0$

    $H_1 : \beta_j \neq 0$ untuk j = 1, 2,…,k

    Kriteria penolakan : H_0 ditolak jika $F_0 = \frac{SSR/k}{SSE/(n-k-1)} > F_{tabel} = F_{(\alpha;k;n-k-1)}$.
2. Uji koefisien regresi secara individu digunakan untuk menguji ada tidaknya pengaruh masing–masing variabel bebas terhadap model regresi linier dengan rumusan hipotesis sebagai berikut:

    $H_0 : \beta_j = 0$

    $H_1 : \beta_j \neq 0$ untuk j = 1, 2,…,k

    Kriteria penolakan : H_0 ditolak jika $|t_0| > t_{tabel} = t_{\alpha/2;\, n-k-1}$ dengan $t_0 = \frac{\hat{\beta}_j}{\sqrt{Se(\hat{\beta}_j)}}$

Sedangkan untuk mengukur kecocokan suatu model regresi menurut Gujarati (2004) dapat menggunakan koefisien determinasi (R^2) yang dapat dihitung menggunakan rumus:

$$R^2 = \frac{SSR}{SST} \text{ dengan } SST = SSR + SSE$$

2.3 Multikolinieritas

Adanya hubungan antar variabel bebas dalam satu regresi disebut dengan multikolinieritas (Widarjono, 2007). Menurut Gujarati (2004), jika terjadi multikolinieritas tidak sempurna, terdapat beberapa konsekuensi sebagai berikut:

1. Meskipun penaksir OLS mungkin bisa diperoleh, standar *error* cenderung semakin besar dengan meningkatnya tingkat korelasi antar variabel bebas
2. Karena besarnya standar *error*, selang kepercayaan untuk parameter populasi yang relevan cenderung untuk lebih besar
3. Atas dasar nomor 2, dalam kasus multikolinieritas yang tinggi menyebabkan probabilitas untuk menerima hipotesis yang salah meningkat
4. Jika multikolinieritas tinggi, mungkin akan diperoleh R^2 yang tinggi tetapi tidak satu pun atau sedikit koefisien yang ditaksir penting secara statistik.

Ada beberapa metode yang dapat digunakan untuk mendeteksi masalah multikolinieritas diantaranya sebagai berikut:

1. Melihat nilai koefisien determinasi (R^2), multikolinieritas seringkali diduga ketika R^2 tinggi tetapi tidak satu pun atau sangat sedikit koefisien regresi secara individu penting secara statistik (Gujarati, 2004).
2. Korelasi antar variabel bebas, jika koefisien korelasi cukup tinggi maka dapat diduga ada multikolinieritas dalam model (Widarjono, 2007).

    Menurut Johnson dan Wichern (2007) koefisien korelasi dapat dihitung dengan rumus sebagai berikut:

$$\rho_{jl} = \frac{S_{jl}}{\sqrt{S_{jj}}\sqrt{S_{ll}}} \text{ dengan } S_{jl} = \frac{1}{n}\sum_{i=1}^{n}\left(x_{ij} - \bar{x}_j\right)\left(x_{il} - \bar{x}_l\right)$$

3. *Variance Inflation Factors* (VIF)

Menurut Montgomery dan Peck (1991)*Variance Inflation Factors* (VIF) dapat dihitung dengan menggunakan rumus sebagai berikut:

$$VIF = \frac{1}{1 - R_j^2} \text{ dengan } R_j \text{ merupakan koefisien determinasi ke-j, } j = 1, 2, \ldots, k$$

Jika nilai VIF lebih dari 5 atau 10 mengindikasikan adanya multikolinieritas.

Menurut Widarjono (2007) ada beberapa metode yang bisa digunakan untuk menangani kasus multikolinieritas antara lain yaitu menghilangkan variabel bebas, penambahan data, transformasi data. Sedangkan menurut Montgomery dan Peck (1991) multikolinieritas dapat juga ditangani menggunakan analisis regresi komponen utama dan regresi ridge.

2.4 Regresi Komponen Utama

Regresi komponen utama membentuk hubungan antara variabel terikat dengan komponen utama yang dipilih dari variabel bebas (Ul-Saufie *et al.* 2011).

Menurut Johnson dan Wichern (2007) Komponen utama merupakan kombinasi linier dari variabel random k (X_1, X_2, …., X_k). Analisis komponen utama tergantung pada matriks kovarian $\sum$ atau matriks korelasi $\boldsymbol{\rho}$. Misalkan vektor random $\mathbf{X}$'= [X_1, X_2, …, X_k] mempunyai matriks kovarian $\sum$ dengan nilai eigen $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k \geq 0$ maka bentuk kombinasi linier sebagai berikut:

$$W_1 = \mathbf{a}_1'\mathbf{X} = a_{11}X_1 + a_{12}X_2 + \cdots + a_{1k}X_k$$
$$W_2 = \mathbf{a}_2'\mathbf{X} = a_{21}X_1 + a_{22}X_2 + \cdots + a_{2k}X_k$$
$$\vdots$$
$$W_k = \mathbf{a}_k'\mathbf{X} = a_{k1}X_1 + a_{k2}X_2 + \cdots + a_{kk}X_k$$

dengan $Var(W_j) = \mathbf{a}_j'\sum\mathbf{a}_j$ dan $Cov(W_j,W_l) = \mathbf{a}_j'\sum\mathbf{a}_l$ dimana j,l = 1, 2,…,k

proporsi dari total varians ke-k = $\frac{\lambda_j}{\lambda_1 + \lambda_2 + \ldots + \lambda_k}$; j = 1, 2, …, k

Komponen utama dapat juga diperoleh dari variabel yang distandarkan yaitu:

$$Z_1 = \frac{(x_1 - \mu_1)}{\sqrt{\sigma_{11}}}$$
$$Z_2 = \frac{(x_2 - \mu_2)}{\sqrt{\sigma_{22}}}$$
$$\vdots$$
$$Z_k = \frac{(x_p - \mu_k)}{\sqrt{\sigma_{kk}}}$$

dimana $E(\mathbf{Z}) = \mathbf{0}$ dan $Cov(\mathbf{Z}) = (\mathbf{V}^{1/2})^{-1}\sum(\mathbf{V}^{1/2})^{-1} = \boldsymbol{\rho}$

Menurut Widiharih (2001) komponen utama ke-i dari variabel yang distandarkan $\mathbf{Z}' = [\mathbf{Z}_1, \mathbf{Z}_2, .., \mathbf{Z}_k]$ dengan $Cov(\mathbf{Z}) = \boldsymbol{\rho}$ yaitu:

$$W_1 = \mathbf{a}_1'\mathbf{Z} = a_{11}Z_1 + a_{12}Z_2 + \cdots + a_{1k}Z_k$$
$$W_2 = \mathbf{a}_2'\mathbf{Z} = a_{21}Z_1 + a_{22}Z_2 + \cdots + a_{2k}Z_k$$
$$\vdots$$
$$W_k = \mathbf{a}_k'\mathbf{Z} = a_{k1}Z_1 + a_{k2}Z_2 + \cdots + a_{kk}Z_k$$

proporsi dari total varians ke-k = $\frac{\lambda_j}{k}$

Menurut Johnson dan Wichern (2007) jumlah komponen utama dapat ditentukan dengan melihat persentase total varian ketika j (j < k) buah komponen yang dipilih mampu menerangkan varian sekitar 80% sampai 90%. Jumlah komponen utama juga dapat diketahui dengan menggunakan *scree plot*.

Secara umum, menurut Montgomery dan Peck (1991) bentuk persamaan dari model regresi komponen utama yaitu:

$$\mathbf{y} = \mathbf{Q}\boldsymbol{\alpha} + \boldsymbol{\varepsilon} \qquad \text{dimana } \mathbf{Q} = \mathbf{XT}, \quad \boldsymbol{\alpha} = \mathbf{T}'\boldsymbol{\beta}, \quad \mathbf{T}'\mathbf{X}'\mathbf{XT} = \mathbf{Q}'\mathbf{Q} = \boldsymbol{\Lambda}$$

Estimator kuadrat terkecil dari $\boldsymbol{\alpha}$ yaitu:

$$\begin{aligned}\hat{\boldsymbol{\alpha}} &= (\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'\mathbf{y} \\ &= \boldsymbol{\Lambda}^{-1}\mathbf{Q}'\mathbf{y}\end{aligned}$$

Matriks kovarian dari $\hat{\boldsymbol{\alpha}}$ yaitu:

$$\begin{aligned}V(\hat{\boldsymbol{\alpha}}) &= \sigma^2(\mathbf{Q}'\mathbf{Q})^{-1} \\ &= \sigma^2\boldsymbol{\Lambda}^{-1}\end{aligned}$$

Menurut Tsutsumi *et al.*(1997) estimator regresi komponen utama yaitu:

$$\begin{aligned}\hat{\boldsymbol{\beta}}_s &= \mathbf{T}\hat{\boldsymbol{\alpha}} \\ &= \mathbf{T}(\mathbf{T}'\mathbf{X}'\mathbf{XT})^{-1}\mathbf{T}'\mathbf{X}'\mathbf{y}\end{aligned}$$

Estimator regresi komponen utama merupakan estimator yang bias karena:

$$\begin{aligned}E(\hat{\boldsymbol{\beta}}_s) &= \mathbf{T}(\mathbf{T}'\mathbf{X}'\mathbf{XT})^{-1}\mathbf{T}'\mathbf{X}'\boldsymbol{\beta} \\ &\neq \boldsymbol{\beta}\end{aligned}$$

2.5 Regresi Ridge

Menurut Dereny dan Rashwan (2011) teknik ridge didasarkan pada penambahan konstanta bias k pada diagonal matriks $\mathbf{X}'\mathbf{X}$ sehingga model persamaan ridge menjadi:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}_{\mathbf{R}} + \boldsymbol{\varepsilon}$$

Dalam mengestimasi parameter model menurut Tsutsumi *et al.*(1997) estimator regresi ridge diperoleh dengan meminimumkan jumlah kuadrat *error*:

$$\sum_{i=1}^{n} \varepsilon_i^2 = \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}_R)'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}_R)$$

dengan kendala $\boldsymbol{\beta}'_R\boldsymbol{\beta}_R = c^2$

dengan menggunakan metode pengali lagrange maka diperoleh:

$$L(\boldsymbol{\beta}_R, k) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}_R)'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}_R) + k(\boldsymbol{\beta}'_R\boldsymbol{\beta}_R - c^2)$$

Sehingga diperoleh estimator regresi ridge yaitu:

$$\hat{\boldsymbol{\beta}}_R = (\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\mathbf{y}$$

Adapun sifat-sifat dari regresi ridge antara lain:

1. Menurut Montgomery dan Peck (1991) estimator regresi ridge merupakan transformasi linier dari estimator metode kuadrat terkecil karena

   $$\hat{\boldsymbol{\beta}}_R = (\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}(\mathbf{X}'\mathbf{X})\hat{\boldsymbol{\beta}} = \mathbf{H}\hat{\boldsymbol{\beta}}$$

   sehingga $E(\hat{\boldsymbol{\beta}}_R) = E(\mathbf{H}\hat{\boldsymbol{\beta}}) = \mathbf{H}\boldsymbol{\beta}$ maka $\hat{\boldsymbol{\beta}}_R$ merupakan estimator yang bias.
2. Menurut Montgomery dan Peck (1991) varian dari $\hat{\boldsymbol{\beta}}_R$ yaitu:

   $$V(\hat{\boldsymbol{\beta}}_R) = \sigma^2(\boldsymbol{X}'\boldsymbol{X} + k\boldsymbol{I})^{-1}(\mathbf{X}'\mathbf{X})(\boldsymbol{X}'\boldsymbol{X} + k\boldsymbol{I})^{-1}$$
3. Menurut Hoerl dan Kennard (1970) jumlah kuadrat *error* untuk estimasi regresi ridge adalah:

   $$SSE = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_R)'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_R)$$
4. Menurut Montgomery dan Peck (1991) rata-rata jumlah kuadrat dari regresi ridge adalah:

   $$\begin{aligned}MSE(\hat{\boldsymbol{\beta}}_R) &= Var(\hat{\boldsymbol{\beta}}_R) + (bias(\hat{\boldsymbol{\beta}}_R))^2 \\ &= \sigma^2\sum_{j=1}^{p}\frac{j}{(\lambda_j + k)^2} + k^2\,\boldsymbol{\beta}'(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-2}\boldsymbol{\beta}\end{aligned}$$

Masalah yang dihadapi dalam regresi ridge adalah penentuan nilai k. Karena itu ada beberapa cara yang dapat digunakan untuk menentukan nilai k antara lain:

1. Ridge *trace*, Hoerl dan Kennard (1970) menyarankan metode grafik yang disebut *ridge trace* untuk memilih nilai parameter ridge k. Grafik plot berdasarkan nilai komponen individu $\hat{\beta}(k)$ dengan barisan dari k ($0 < k < 1$). Mallows (1973) dalam Montgomery dan Peck (1991) menyarankan nilai k yang meminimumkan nilai C_k yang dihitung dengan menggunakan rumus:

   $C_k = \frac{SSE(k)}{\hat{\sigma}^2} - n + 2 + 2Tr(\mathbf{XL})$, dengan $\mathbf{XL} = \mathbf{X}(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}' \equiv \mathbf{H}_k$
2. Hoerl, Kennard dan Baldwin (1975) dalam Dereny dan Rashwan (2011), menawarkan metode untuk memilih nilai k tunggal dari semua k_i yang dapat dihitung dengan menggunakan rumus berikut:

   $k_{HKB} = \frac{P\hat{\sigma}^2}{\hat{\boldsymbol{\beta}}'\hat{\boldsymbol{\beta}}}$

3. METODOLOGI PENELITIAN

Data yang digunakan adalah data sekunder yang diperoleh dari Badan Pusat Statistik (BPS) Provinsi Jawa Tengah Tahun 2013. Variabel yang digunakan terdiri dari variabel terikat dan variabel bebas. Variabel terikat yang digunakan adalah Pendapatan Asli Daerah (PAD) dalam rupiah (y) dan Variabel bebas yang digunakan adalah Pajak daerah dalam rupiah (X_1), Retribusi daerah dalam rupiah (X_2), PDRB atas dasar harga berlaku dalam rupiah (X_3), PDRB atas dasar harga konstan dalam rupiah (X_4), Jumlah penduduk dalam jiwa (X_5), Belanja daerah dalam rupiah (X_6).

Langkah-langkah analisis data yang dilakukan sebagai berikut:

1. Mencari data yang akan digunakan
2. Menentukan variabel terikat dan variabel bebas dari data yang diperoleh.
3. Melakukan analisis regresi untuk menentukan model regresi dengan metode kuadrat terkecil
4. Melakukan pemeriksaan asumsi nonmultikolinieritas dengan cara melihat nilai VIF, melihat nilai koefisien determinasi (R^2), melihat nilai korelasi antar variabel bebas
5. Melakukan penanganan terhadap masalah multikolinieritas apabila asumsi nonmultikolinieritas tidak terpenuhi yaitu:
   A. Regresi komponen utama
      1. Melakukan analisis komponen utama yaitu mencari nilai eigen dan vektor eigen, menghitung skor komponen utama, menentukan jumlah komponen utama yang akan digunakan
      2. Meregresikan antara skor komponen yang diperoleh dengan variabel terikat
      3. Mengembalikan persamaan regresi ke bentuk variabel standar
      4. Menghitung standar *error* untuk masing-masing koefisien regresi dan melakukan pengujian menggunakan uji t
      5. Mengembalikan persamaan regresi ke bentuk variabel asli

   B. Regresi Ridge
      1. Mentransformasikan data ke dalam bentuk baku
      2. Mencari estimator parameter regresi ridge dengan nilai k ($0< k< 1$) untuk ridge *trace*, k yang dipilih adalah nilai k yang meminimumkan Ck

3.Mencari estimator parameter regresi ridge dengan nilai k = $k_{HKB} = \frac{P\hat{\sigma}^2}{\hat{\boldsymbol{\beta}}'\hat{\boldsymbol{\beta}}}$ berdasarkan Hoerl, Kennard dan Baldwin (1975)
4.Menghitung nilai Ck
5.Menghitung standar *error* untuk masing-masing koefisien regresi baik untuk regresi ridge dengan k (ridge *trace*) maupun k_{HKB} (Hoerl, Kennard dan Baldwin) dan melakukan pengujian menggunakan uji t
6.Mengembalikan persamaan regresi ke bentuk variabel asli

6.Membandingkan hasil standar *error* yang diperoleh antara regresi komponen utama dengan regresi ridge untuk memilih metode terbaik.

4. HASIL DAN PEMBAHASAN

4.1 Analisis Regresi Berganda

Persamaan model regresi yang diperoleh adalah:

$\hat{y}$ = –19298026 + 0,307X_1 + 2,50X_2 – 5,03X_3 + 24,8X_4 – 60,7X_5 + 0,0748X_6

1. Uji Signifikansi Regresi, berdasarkan F_{hitung} yang diperoleh yaitu F_0 = 49,57 lebih besar dari $F_{0,05;6;28}$ = 2,44 dan nilai sig = 0 kurang dari α = 0,05 maka H_0 ditolak yang berarti bahwa model sesuai untuk menggambarkan hubungan linier antara variabel terikat dengan variabel bebas.
2. Uji koefisien regresi secara individu dapat dilihat dalam Tabel 1.

Tabel 1. Uji Signifikansi Parameter

Variabel Bebas	Koefisien	t_0	$t_{\alpha/2;n-k-1}$	Sig	Kesimpulan
Intercept	-19298026	-0,57	2,048	0,571	Tidak Signifikan
X_1	0,3070	2,88	2,048	0,008	Signifikan
X_2	2,4962	2,95	2,048	0,006	Signifikan
X_3	-5,0300	-1,64	2,048	0,111	Tidak Signifikan
X_4	24,8000	2,50	2,048	0,018	Signifikan
X_5	-60,6900	-1,26	2,048	0,220	Tidak Signifikan
X_6	0,0748	1,29	2,048	0,208	Tidak Signifikan

Nilai koefisien determinasi yang diperoleh yaitu R^2 = 91,4% yang berarti bahwa variabel bebas mempengaruhi variabel terikat sebesar 91,4% dan sisanya sebesar 8,6% dipengaruhi oleh variabel-variabel lain.

4.2 Pendeteksian Multikolinieritas

1. Melihat nilai koefisien determinasi (R^2), nilai koefisien determinasi yang dihasilkan cukup tinggi yaitu 91,4% tetapi ada beberapa variabel bebas yang secara individu tidak signifikan.
2. Korelasi antar variabel bebas, adanya korelasi yang cukup tinggi antara X_3 dengan X_4 yaitu sebesar 0,969 dan X_5 dengan X_6 yaitu sebesar 0,888.
3. *Variance Inflation Factor* (VIF), ada beberapa variabel bebas yang mempunyai nilai VIF > 10 yaitu variabel X_3 dan X_4 sebesar 26,1 dan 34.

Dari pendeteksian multikolinieritas di atas dapat disimpulkan adanya multikolinieritas pada model regresi sehingga perlu dilakukan tindakan penanganan menggunakan regresi komponen utama dan regresi ridge.

4.3 Regresi KomponenUtama

Langkah pertama yang dilakukan yaitu melakukan analisis komponen utama dengan menggunakan *software* Minitab 14 berdasarkan matriks korelasi dan diperoleh komponen utama sebagai berikut:

$W_1 = -0{,}390Z_1 - 0{,}387Z_2 - 0{,}434Z_3 - 0{,}452Z_4 - 0{,}353Z_5 - 0{,}425Z_6$

$W_2 = -0{,}385Z_1 - 0{,}4242Z_2 - 0{,}063Z_3 - 0{,}175Z_4 + 0{,}675Z_5 + 0{,}426Z_6$

$W_3 = 0{,}418Z_1 + 0{,}401Z_2 - 0{,}621Z_3 - 0{,}449Z_4 + 0{,}183Z_5 + 0{,}210Z_6$

$W_4 = 0{,}718Z_1 - 0{,}670Z_2 + 0{,}012Z_3 + 0{,}018Z_4 + 0{,}094Z_5 - 0{,}158Z_6$

$W_5 = 0{,}024Z_1 - 0{,}237Z_2 - 0{,}087Z_3 + 0{,}034Z_4 - 0{,}608Z_5 + 0{,}751Z_6$

$W_6 = 0{,}093Z_1 + 0{,}065Z_2 + 0{,}644Z_3 - 0{,}750Z_4 - 0{,}079Z_5 + 0{,}061Z_6$

Membuat model regresi menggunakan skor komponen utama dengan variabel terikat dan diperoleh model regresi sebagai berikut:

$\hat{y} = 173831738 - 59377566W_1 - 53141353W_2$

Mengembalikan model ke bentuk variabel standar dan hasilnya sebagai berikut:

$\hat{y} = 173831738 + 43592843Z_1 + 45495814Z_2 + 29088724Z_3 + 36112808Z_4 - 14915735Z_5 + 2593404Z_6$

Uji koefisien regresi secara individu menggunakan uji t seperti pada Tabel 2.

Tabel 2. Uji koefisien regresi secara individu regresi komponen utama

Variabel Bebas	Koefisien	Standar Error	t_0	$t_{\alpha/2;n-k-1}$	Kesimpulan
Z_1	43592843	0,02673	1630999170	2,048	Signifikan
Z_2	45495814	0,02892	1573263457	2,048	Signifikan
Z_3	29088724	0,01345	2163162648	2,048	Signifikan
Z_4	36112808	0,01730	2087710053	2,048	Signifikan
Z_5	-14915735	0,04360	-342127154	2,048	Signifikan
Z_6	2593404	0,02954	87789149,7	2,048	Signifikan

Persamaan regresi yang diperoleh dari variabel standar dikembalikan ke bentuk variabel asli sehingga diperoleh model regresi komponen utama sebagai berikut:

$\hat{y} = 30561720 + 0{,}3389X_1 + 2{,}6799X_2 + 2{,}2294X_3 + 7{,}7746X_4 - 36{,}5240X_5 + 0{,}0065X_6$

4.4 Regresi Ridge

Dalam analisis regresi ridge digunakan *software* MATLAB R2010a dengan nilai k yang diperoleh dari ridge trace dan metode yang ditawarkan oleh Hoerl, Kennard dan Baldwin (1975). Pada regresi ridge dengan nilai k yang diperoleh dari *ridge trace* yaitu k=0,015 menghasilkan model regresi ridge sebagai berikut:

$\hat{y} = 0{,}3135X_1^* + 0{,}3318X_2^* - 0{,}2014X_3^* + 0{,}5160X_4^* - 0{,}1755X_5^* + 0{,}2014X_6^*$

Sedangkan pada regresi ridge dengan nilai k yang diperoleh dari Hoerl, Kennard dan Baldwin (1975) yaitu $k_{HKB} = 0{,}0157$ menghasilkan model regresi ridge sebagai berikut:

$\hat{y} = 0{,}3142X_1^* + 0{,}3326X_2^* - 0{,}1946X_3^* + 0{,}5084X_4^* - 0{,}1750X_5^* + 0{,}2006X_6^*$

Uji koefisien regresi ridge secara individu menggunakan uji t yang dapat dilihat pada Tabel 3 dan Tabel 4.

Tabel 3. Uji Koefisien Regresi Ridge Secara Individu dengan Ridge *Trace*

Variabel Bebas	Koefisien	Standar Error	t_0	$t_{\alpha/2;n-k-1}$	Kesimpulan
X_1^*	0,3135	0,0872	3,5961	2,048	Signifikan
X_2^*	0,3318	0,0911	3,6420	2,048	Signifikan
X_3^*	– 0,2014	0,1543	–1,3055	2,048	Tidak Signifikan
X_4^*	0,5160	0,1738	2,9629	2,048	Signifikan
X_5^*	– 0,1755	0,1153	–1,5218	2,048	Tidak Signifikan
X_6^*	0,2014	0,1356	1,4847	2,048	Tidak Signifikan

Tabel 4. Uji Koefisien Regresi Ridge Secara Individu dengan Hoerl, Kennard dan Baldwin (1975)

Variabel Bebas	Koefisien	Standar Error	t_0	$t_{\alpha/2;n-k-1}$	Kesimpulan
X_1^*	0,3142	0,0852	3,6878	2,048	Signifikan
X_2^*	0,3326	0,0893	3,7245	2,048	Signifikan
X_3^*	– 0,1946	0,1484	– 1,3113	2,048	Tidak Signifikan
X_4^*	0,5084	0,1669	3,0461	2,048	Signifikan
X_5^*	– 0,1750	0,1124	– 1,5569	2,048	Tidak Signifikan
X_6^*	0,2006	0,1323	1,5163	2,048	Tidak Signifikan

Mengembalikan persamaan yang diperoleh ke bentuk variabel asli sehingga diperoleh model regresi ridge sebagai berikut:

1. Model regresi ridge dengan nilai k yang diperoleh dari *ridge trace*

$$\hat{y} = -14769416 + 0{,}3433X_1 + 2{,}7530X_2 - 2{,}1743X_3 + 15{,}6477X_4 - 60{,}5340X_5 + 0{,}0708X_6$$

2. Model regresi ridge dengan nilai k yang diperoleh dari Hoerl, Kennard dan Baldwin (1975)

$$\hat{y} = -14592681 + 0{,}3441X_1 + 2{,}7597X_2 - 2{,}1009X_3 + 15{,}4172X_4 - 60{,}3616X_5 + 0{,}0705X_6$$

4.5 Pemilihan Metode Terbaik

Pemilihan metode terbaik untuk mengatasi masalah multikolinieritas antara regresi komponen utama dan regresi ridge didasarkan pada standar *error* yang dapat dilihat pada Tabel 5.

Tabel 5. Hasil Perbandingan PCR dan Regresi Ridge

Variabel Bebas	Estimasi Parameter			Standar *Error*		
	Ridge Trace	HKB	PCR	Ridge Trace	HKB	PCR
X_1	0,3433	0,3441	0,3389	0,0855	0,0852	0,02673
X_2	2,7530	2,7597	2,6799	0,0897	0,0893	0,02892
X_3	-2,1743	-2,1009	2,2294	0,1515	0,1484	0,01345
X_4	15,6477	15,4172	7,7746	0,1707	0,1669	0,01730
X_5	-60,5340	-60,3616	-36,5240	0,1133	0,1124	0,04360
X_6	0,0708	0,0705	0,0065	0,1334	0,1323	0,02954

Berdasarkan Tabel 5 dapat dilihat bahwa regresi komponen menghasilkan standar *error* yang lebih kecil oleh karena itu dapat disimpulkan bahwa metode regresi komponen utama lebih baik dibandingkan dengan metode regresi ridge dalam mengatasi masalah multikolinieritas pada analisis faktor-faktor Pendapatan Asli Daerah (PAD) Provinsi Jawa Tengah.

5 KESIMPULAN

Berdasarkan analisis dan pembahasan yang telah dipaparkan sebelumnya, maka dapat diambil kesimpulan sebagai berikut:

1. Pada analisis Pendapatan Asli Daerah (PAD) dengan menggunakan metode kuadrat terkecil terdapat masalah multikolinieritas sehingga dilakukan penanganan dengan menggunakan regresi komponen utama dan regresi ridge.
2. Berdasarkan perhitungan nilai standar *error* menunjukkan bahwa penanganan multikolinieritas menggunakan regresi komponen utama lebih baik dibandingkan dengan regresi ridge dalam analisis faktor-faktor Pendapatan Asli Daerah (PAD) Provinsi Jawa Tengah.

6. DAFTAR PUSTAKA

BPS. 2013. *Jawa Tengah dalam Angka 2013*. BPS Provinsi Jawa Tengah

BPS. 2013. *Statistika Keuangan Pemerintah Provinsi dan Kabupaten/Kota Jawa Tengah 2011-2012*. BPS Provinsi Jawa Tengah

Dereny, M. El., Rashwan, N. I. 2011. *Solving Multicollinearity Problem Using Ridge Regression Models*. Int. J. Contemp. Math. Sciences Vol. 6 No.12 Hal. 585 – 600

Gujarati, D. 2004. *Ekonometrika Dasar*. Sumarno Zain Penerjemah. Jakarta: Erlangga. Terjemahan dari: *Basic Econometrics*

Hoerl, A. E., Kennard, R. W. 1970. *Ridge Regression: Biased Estimation for Nonorthogonal Problem*. Technometrics, Vol. 12 No. 1 Hal. 55 - 67

Johnson, R. A., Wichern, D. W. 2007. *Applied Multivariate Statistical Analysis*. Sixth edition, Prentice Hall. New Jersey

Kutner, M. H., Nachtsheim, C.J., Neter, J., Li, W. 2005. *Applied Linear Statistical Models*. Fifth Edition.New York: McGraw-Hill

Mardiasmo. 2002. Otonomi dan Manajemen Keuangan Daerah. Yogyakarta: ANDI

Mardikyan, S., Cetin, E. 2008. *Efficient Choice of Biasing Constant for Ridge Regression*. Int. J. Contemp. Math. Sciences Vol. 3 No.11 Hal. 527 - 536

Montgomery, D. C., Peck, E. A. 1991. *Introduction Linear Regression Analysis*. Second Edition. New York: John Wiley & Sons

Santoso, P. B., Rahayu, R. F. 2005. *Analisis Pendapatan Asli Daerah(PAD) dan Faktor-Faktor yang Mempengaruhinya dalam Upaya Pelaksanaan Otonomi Daerah di Kabupaten Kediri*. Dinamika Pembangunan Vol. 2 No.1 Hal. 9 - 18

Tsutsumi, M., Shimizu, E., Matsuba, Y. 1997. *A Comparative Study on Counter-Measures for Multicollinearity in Regression Analysis*. Journal of the Eastern Asia Society for Transportation Studies Vol. 2 No. 6

Ul-Sulfie, A., Yahya, A. S., Ramli, N. A. 2011. *Improving Multiple Linear Regression Model Using Principal Component Analysis for Predicting PM10 Concentration in Seberang Prai, Pulau Pinang*. International Journal of Environmental Sciences Vol. 2 No. 2

Widarjono, A. 2007. *Ekonometrika Teori dan Aplikasi untuk Ekonomi dan Bisnis*. Yogyakarta: Ekonisia FE UII

Widiharih, T. 2001. *Penanganan Multikolinieritas (Kekolinieran Ganda) dengan Anlisis Regresi Komponen Utama*. Jurnal Matematika dan Komputer Vol. 4 No. 2 Hal. 71 - 81