BAB II

TINJAUAN PUSTAKA

2.1 Matriks

Matriks adalah susunan bilangan atau elemen yang tersusun menurut baris dan kolom sehingga berbentuk persegi. Matriks dinotasikan dengan huruf kapital A,B,K dan sebagianya. Banyknya baris dan kolom suatu matriks menentukan ukuran dari segi metriks tersebut yang disebut dengan ordo matriks.

Secara umum, matriks $A_{mxn} = \begin{matrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \dots & a_{3n} \\ \dots & \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & a_{m3} & \dots & a_{mn} \end{matrix}$ (2.1)

a_{ij} adalah elemen yang terdapat pada baris i dan kolom j dari A,dimana $i = 1,2,3, \dots m$ (indeks baris) dan $j = 1,2,3, \dots n$ (indeks kolom).

Matriks A dapat ditulis sebagai berikut :

$A_{mxn} = [a_{ij}]_{mxn}$ dengan $i = 1,2,3, \dots m$; $j = 1,2,3, \dots n$

2.1.Analisis Regresi Linier

Analisis regresi linier berganda adalah hubungan secara linier antara dua atau lebih variabel independen ($X_{1,}x_2$,....X_n) dengan variabel dependen(Y). Analisis ini untuk menhetahui arah hubungan antara variabel independen dengan variabel dependen apakah masing-masing variabel independen berhubungan positif atau negatif dan untuk memprediksi terjadinya satu peningkatan atau penurunan dengan data yang sering digunakan adalah berskala rasio atau interval Persamaan regresi linier berganda sebagai berikut :

$$Y= a + b_1 x_1 + b_2x_2 +....+b_nx_n + \varepsilon 1 \quad (2.2)$$

Keterangan :

Y	= Variabel dependen (nilai yang akan diprediksi)
$X_1\, dan\, X_2$	= Variabel independen
a	= Konstanta (niali pada Y apabila $X_{1,}x_2$,....X_n= 0)
b	= Koefisien regresi nilai (peningkatan atau penurunan)
ε	= terjadinya eror

Jika model regresi linier berganda dengan ρ variabel independen $b_j, j = 0,1, \dots n$ desebut dengan bilangan koefisien regresi. Dengan parameter bj mewaikili peubah pada variabel dependen Y di unit independen ke X_j ketika semua variabel independen yang tersisa X_i $(i \neq j)$ tidak berubah.

Dengan bentuk secara matriks sebagai berikut :

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & X_{11} & X_{12} & \dots & X_{1p} \\ 1 & X_{21} & X_{22} & \dots & X_{2p} \\ 1 & \vdots & \vdots & \dots & \vdots \\ 1 & X_{n1} & X_{n2} & Y_n & X_{np} \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_p \end{pmatrix} + \begin{pmatrix} \varepsilon 1 \\ \varepsilon 2 \\ \vdots \\ \varepsilon n \end{pmatrix} \text{atau } (y = xb + \varepsilon) \quad (2.3)$$

Dimana :

y = vektor kolom nx1 variabel independen Y

x = matriks nx(p+1) dari variabel dependen X

b = vektor pada kolom (p+1) untuk parameter tidak diketahui b0,b1,....bp

ε = vektor kolom n x 1 dari terjadinya eror

Adanya variabel independen ini dapat lebih menjelaskan hubungan antar variabel meski masih ada variabel yang terabaikan.

2.2.Asumsi Analisis Regresi

Pengujian asumsi analisis regresi adalah uji asumsi asumsi statistik yang harus memenuhi analisis regresi linier berganda dan beberapa asumsinya menurut *Gujarati (2003)* adalah sebagai berikut :

- Nilai residual berdistribusi normal.
- Tidak terdapat multikolinieritas yang sempurna.
- Tidak terdapat autokorelasi antara nilai residual.
- Variabel bebas berpengaruh secara signifikan terhadap variabel terikat.

2.3.1 Uji Normalitas

Menurut Sulianto (2004) cara mengetahui apakah residual telah terstandarisasi berdistribusi normal atau tidak adalah dengan cara uji Kolmogrov-Smirnov.

Dengan hipotesis :

H_0 : data berasal dari populasi berdistribusi normal

H_1 : data berasal dari tidak berdistribusi normal

Langkah-langkah dalam pengujian :

1. Melakukan persamaan regresi
2. Mencari nilai prediksinya ($\hat{Y}$)
3. Mencari nilai residualnya ($Y - \hat{Y}$)
4. Mengrutkan nilai standarisasi dari terkecil hingga terbesar
5. Mencari nilai Z_t relatif komulatif
6. Mencari nilai Z_t teoritis berdasarkan nilai tabel Z
7. Menghitung nilai Z_r dengan Z_r dan diberi klom K
8. Mencari nilai K mutlak terbesar dan diberi nama K_{hitung}
9. Membandingkan nilai K_{hitung} dengan tabel Kolmogorov-Smirnov (K_{tabel})
10. Menarik kesimpulan dengan kriteria jika $K_{hitung} < K_{tabel}$ maka residual tersebut terstandarisasi berdistribusi normal jika sig>0,05.

2.3.2 Uji Multikolineritas

Multikolieritas adalah pengujian yang bertujuan untuk ditemukannya suatu korelasi pada variabel bebas (independen). Penyebab multikolinieritas menurut *Gurajati (2004)* ialah memasukkan variabel independen atau bebas yang seharusnya dikeluarkan dari model empris, selain itu *Gujarati* mengungkapkan penyebab multikolinieritas ialah siafat siafat pada variabel yang kebanyakan berubah secara bersamaan dalam waktu yang sama juga lalu kesalahan dalam menbentuk model regresi yang digunakan dalam variabel independen atau bebas dan kecilnya jumlah pengamatan yang akan dianalisis dengan model regresi. Untuk mengetahui masalah tersebut menurut *Widarjono(2007)* adalah dengan cara mengecek korelasi antara variabel independen, jika koefisien korelasi tinggi, biasanya diatas 0,9 maka dapat diartikan terdapat multikonieritas.Menurut *Johnson dan*

Wichen (2007) dengan cara menghitung korelasi koefisiennya dengan menggunakan rumus.

$$\rho jI = \frac{Sjl}{\overline{SJJ}\ \overline{SJJ}} \text{ dengan hasil Sjl} = \frac{1}{n}\ \sum_{i=1}^{n} XIJ - X\bar{J}\ (X\text{iI} - X\bar{I}) \qquad (2.4)$$

Selain cara *Widarjono, Johnson* dan *Winhen*, adapun cara untuk melihat nilai variance inflation factor (VIF) dan nilai tolerance (TOL). Yaitu mengukur besarnya multikolinieritas dan menunjukkan peningkatan regam koefisien regresi yang disebabkan karena adanya ketergantungan linier peubah prediktor tersebut peubah prediktor lain.

Dengan menggunakan rumus $$vif = \frac{1}{1-{R_j}^2} \qquad (2.5)$$

Dengan R_j merupakan koefisien determinan ke j, = 1,2,.....n. Dengan nilai multikolinieritas jika nilai VIF ≥ 10. Jika nilai Tol kurang dari 0,01 atau nilai VIF lebih dari 10 maka diartikan bahwa multikolinieritas terjadi diantara variabel independen atau bebas.

2.3.3 Uji Linieritas

Uji linieritas dilakukan untuk membuktikan apakah model linier atau tidak. Apabila terdapat kesalahan dalam menentukan model regresi maka nilai prediksi yang dihasilkan makan akan terjadi bias. Menurut *Sulianto (2008)* uji Lagrange Multipler (LM-Test) merupakan salah satu metode untuk mengukur linieritas yang dikembangkan oleh *Engle* pada tahun 1982. Prinsip metode ini ialah menbandingkan antara nilai X^2_{hitung} dengan X^2_{tabel} dengan df = (n,a) dengan langkah-langkah sebagai berikut :

1. Membuat persamaan perseginya
2. Mencari nilai prediksinya $(\hat{Y})$
3. Mencari nilai residualnya $(Y - \hat{Y})$

4. Menguadratkan semua nilai variabel bebas
5. Meregresikan kuadrat variabel bebas terhadap nilai residualnya
6. Mencari nilai koefisisen determinan R^2
7. Menghitung nilai $X^2_{hitung} = (n\ x\ R^2)$. Dimana n adalah jumlah pengamatan
8. Menarik kesimpulan uki linieritas dengankriteria hitung jika X^2_{hitung} < X^2_{tabel} dengan df=(n,a) maka model dinyatakan linier.
9. Begitu sebaliknya. Dengan Hipotesis pengujian :

H_0 : ρ = 0(tidak terdapat korelasi linier antara $X_1, X_{2,}X_{3,} \ldots X_k$ dengan Y)

H_1 : $\rho \neq 0$ (terdapat korelasi linier antara $X_1, X_{2,}X_{3,} \ldots X_k$ dengan Y)

Kriteria pengujian ini menggunakan taraf kesalahan $\propto$ dan $df = (n, a)$ yaitu jika X^2_{hitung} < X^2_{tabel} maka H_0 ditolak begitupun sebaliknya.

2.3.4 Uji Simultan

Untuk mengetahui apakah semua variabel independent atau bebas yang akan dimasukkan akan berpngaruh secara simultan atau bersama-sama terhadap variabel dependent atau terikat maka dilakukan uji statistik F menurut *Ghozali(2006:16)*. Uji ini untuk mengetahui kelayakan model *goodness of fit.* Maka nilai F_{tabel} diperoleh dari tabel distribusi F dengan tingkat signifikan α dan derajat kebebasan df = $\alpha, k - 1, (n - k)$dimana n adalah jumlah observasi dan k adalah jumlah variabel.

$$F = \frac{R^2/(k-1)}{1-R^2/(n-k)} \qquad (2.6)$$

Dengan hipotesis :

$H_0: \beta_1 = \beta_2 = \beta_3 = \cdots = \beta_i = 0$. Semua $\beta_i = 0$ untuk $i = 1,2,\dots,n$

H_1: tidak semua $\beta_i = 0. Untuk\ i = 1,2,\dots,n$

Kriteria Uji :

Jika $F_{hitung} \geq F_{tabel}$ atau nila sig $\alpha < \alpha\ maka\ H_0$ ditolak begitupun sebaliknya. Yang artinya secara simultan variabel independent atau bebas $(X_{1,}X_2,\dots.X_n)$ berpengaruh signifikan terhadap variabel independent atau bebas (Y).

2.3.5 Uji Keberartian Parsial

Uji keberartian parsial atau sering juga disebut uji t adalah untuk menunjukkan seberapa jauh satu variabel independent lainnya *Ghozali(2009:17)* nilai t tabel dengan derajat bebas yaitu df= $=\alpha, (n, -k)$, dimana n adalah jumlah observasi dan k jumlah variabel.

Hipotesis pengujiannya :

$H_0: \beta_j = 0, X_1$ tidak mempengaruhi Y

$H_1 : \beta_k \neq 0, X_1$ menpengaruhi Y

$\vdots$

$H_0: \beta_k = 0, X_k$ tidak mempengaruhi Y

$H_1 : \beta_k \neq 0, X_k$ menpengaruhi Y

Jika $F_{hitung} \geq F_{tabel}$ atau nila sig $\alpha < \alpha\ maka\ H_0$ ditolak begitupun sebaliknya. Nilai t-hitung didapatkan menggunakan rumus *Ghozali(2009):*

$$t_{hitung} = \frac{Koefisien\ Regresi(bi)}{Standar\ Deviasi\ bi} \quad (2.7)$$

Pengambilan keputusan uji hipotesis parsial didasari pada nilai probabilitas yang didapatkan jika tingkat signifikansi lebih kecil dari α maka hipotesis diterima atau dikatakan signifikan (H_1 $diterimah$) yang artinya secara parsial variabel independent atau bebas ($X_{1,}X_2, \ldots . X_n$) berpengaruh signifikan terhadap variabel dependent atau terikat (Y) begitupun sebaliknya.

2.4 Koefisien Determinan R^2

Uji koefisien determinasi (R^2) yang digunakan untuk mengukur seberapa jauh kemampuan model dalam menerangkan variasi variabel dependent atau terikat menurut *ghozali (2006).* Yang dinotasikan dengan nilai (R^2) menggunakan rumus sebagai berikut :

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})}{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2} = \frac{JKR}{JKT} = 1 - \frac{JKG}{JKT}; \text{ dimana } 0 \leq R^2 \leq 1 \qquad (2.8)$$

Bertambahnya suatu varibel prediktor pada model akan menaikkan nilai R^2 itu sendiri, sebab JKG tidak akan bertambah jika variabel prediktor lebih lebih banyak, sedangkan JKT tidak akan berubah jika data responsnya tetap sama. Karena nilai R^2 sering dibuat lebih besar dengan menambah variabel prediktornya maka ukurnnya diubah atau dimodifikasi dengan menperhitungkan seberapa banyaknya variabel prediktor dalam model. R^2 terkoreksi atau adjusted dirumuskan :

$$R^2 = 1 - \frac{JKG/(n-p)}{JKT/(n-1)} = \frac{(n-1)JKG}{(n-p)JKT} \qquad (2.8)$$

Nilai R^2 yang mendekati nol menunjukkan bahwa data tersebut tidak cocok dengan model regresi. Sebliknya jika mendekati satu menunjukkan bahwa medel regresi tersebut cocok *Johson dan Wichern,(1996).*

2.5 Koefisien Korelasi

Koefisien korelasi, dinotasikan dengan *r* digunakan untuk mengukur hubungan antara dua variabel pada analisis korelasi. Koefisien korelasi antara sampel X dan Y dinotasikan dengan r_{xy} adalah :

$$r_{xy} = \frac{S_{xy}}{\sqrt{S_{xx}}\sqrt{S_{yy}}} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{[\sum_{i=1}^{n}(X_i - \overline{X})^2 \sum_{i=1}^{n}(Y_i - \overline{Y})^2]^{1/2}} \qquad (2.9)$$

S_{xy} adalah kovariansi dari x dan y sedangkan S_x dan S_y adalah simpangan bakunya. Koefisien korelasi mengukur hubungan antara dua variabel nilainya $-1 \leq r_{xy} \leq 1$. Apabila *r* bernilai 1 atau -1 maka hubungan linier antara dua variabel tersebut sempuran atau sangat kuat. Jika korelasi bernilai positif maka kedua variabel menpunyai hubungan searah sedangkan korelasi yang benilai negatif maka kedua veriabel menpunyai hubungan yang berlawanan arah *Supranto(2008)*.

2.6 Mean Of Square (MSE)

Menurut *Ghazali (2006) Mean of sqaure eror* (MSE) adalah sala satu pengukuran kesalahan atau eror yang sering digunakan. Nilai Mse dihintung dengan cara mengkuadratkan selisih antara nilai prediksi dengan nilai sebenarnya, umumnya semakin kecil nilai MSE maka semakin tinggi tingkat keakuratan nilai suatu prediksi. Perhitungan MSE dilakukan dengan menggunakan rumus :

$$MSE = \frac{{}_{i=1}^{n} Y_i - \widehat{Y_i}^2}{n} \qquad (2.9)$$

Keterangan :

$\widehat{Y_i}$: Variabel dependent dugaan

$\bar{y}$: nilai rata rata dari variabel dependent

n : jumlah data

2.7 Metode Ordinari Least Square (OLS)

Metode Ordinari least Square atau OLS atau yang sering kita sebut metode kuadrat terkecil merupakan metode yang sering kita jumpai pada estimasi fungsi regresi yang bertujuan untuk meminimalisir kudarat kesalahan e_i sehingga nilai regresi yang didapatkan akan mendekati niali sesungguhnya. Estimasi regresi β diproleh dengan meminimlisir kuarat erornya ($\varepsilon'\varepsilon$) karena :

$$\varepsilon'\varepsilon = (e_1, e_2, \dots e_n)\begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix} \qquad (2.10)$$

$$= e_1^2, e_2^2, \dots e_n$$

$$= \sum e_i^2 \text{ dengan } \varepsilon = Y - X\hat{\beta}$$

$$\varepsilon'\varepsilon = (Y - X\hat{\beta})' (Y - X\hat{\beta})$$

$$= (Y' - X'\hat{\beta}') (Y - X\hat{\beta})$$

$$= Y'Y - Y'X\hat{\beta} + Y X'\hat{\beta}' + X'\hat{\beta}'X\hat{\beta}$$

$$= Y'Y - 2YX'^{\hat{\beta}'} + X'\hat{\beta}'X\hat{\beta}$$

$$\frac{d(\varepsilon'\varepsilon)}{d\hat{\beta}} = 0$$

$$0 - 2XY' + 2(XX')\hat{\beta} = 0$$

$$2(XX')\hat{\beta} = 2X'Y$$

Sehingga didapatkan estimasi OLS untuk β adalah $\hat{\beta} = (X'X)^{-1}X'Y$ dan pada persamaan tersebut didapatkan bahwa ε bebas satu sama lain E (ε) = 0 dan var $=\sigma^2 I$ estimator kuadrat terkecil βj terdapat suatu variansi yang minimum diantara seluruh estimator linier dan tak bias. Dengan sifat kuadrat terkecil adalah :

1. Linier dan tak bias

Jika $E(\hat{\beta}) = \beta$ maka $\hat{\beta}$ adalah estimator yang tak bias untuk β. Akan menunjukkan bahwa $\hat{\beta}$ adalah penduga linier tak bias dari β.

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

$= (X'X)^{-1}X'(X\beta + \varepsilon)$

$= (X'X)^{-1}X'X\beta + (X'X)^{-1}X'\varepsilon$

$= \beta + (X'X)^{-1}X$

Sehingga $\hat{\beta}$ adalah fungsi dari linier β dan ε $(X'X)^{-1}\, X'X = I$

$E(\hat{\beta}) = E[(X'X)^{-1}\, X'Y]$

$= (X'X)^{-1}X'E(Y)$

$= (X'X)^{-1}X'(X\beta)$

$= (X'X)^{-1}X'\ X\beta$

$= I\beta$

$= \beta$

Karena $E(\hat{\beta}) - \beta$ maka $\hat{\beta}$ adalah estimator yang tak bias untuk β

2. Varian minimum

$$\mathrm{Var}(\hat{\beta}) = \mathrm{E}\left[\left((\hat{\beta} - E(\hat{\beta})\right)\left((\hat{\beta} - E(\hat{\beta})\right)'\right]$$

$$= E[(\beta + (X'X)^{-1}\, X'\varepsilon - \beta)(\beta + (X'X)^{-1}(X')\varepsilon - \beta)']$$

$$= E[((X'X)^{-1}X'\,\varepsilon)(\, X'X)^{-1}X'\,\varepsilon)']$$

$$= E[((X'X)^{-1}X'\,\varepsilon\varepsilon'\, X(X'X)^{-1}]$$

$$= (X'X)^{-1}X'X(X'X)^{-1}E\,(\varepsilon'\varepsilon)$$

$$= (X'X)^{-1}\, I\sigma^2$$

$$= (X'X)^{-1}\, \sigma^2$$

Jadi terbukti $(\hat{\beta})$ $=(X'X)^{-1}\,\sigma^2$. Jika $\widehat{\beta}$ dan $\hat{\beta}_2$ sebuah estimator dari β diamana variansi untuk $\hat{\beta}$ lebih kecil dari varian untuk $\hat{\beta}_2$ maka estimator tersebut bervariansi minimum. Untuk menbuktikan asumsi sebuah estimator alternatif yang linear dan tidak bisa kemudian dibuktikan variansi yang lebih besar dari pada variansi estimator model regresi. Misal $\hat{\beta}_2 = [(X'X)^{-1'} + C]Y$ dimana c adalah matriks konstanta berukuran kxn yang diketahui maka :

$$\hat{\beta}_2 \;= [(X'X)^{-1'} + C]Y$$

$$= \;[(X'X)^{-1'} + C](X\beta + \varepsilon)$$

$= (X'X)^{-1'}X'(X\beta + \varepsilon) + C\;(X\beta + \varepsilon)$, nilai harapan

pada estimator $\hat{\beta}_2$ adalah :

$$E[\hat{\beta}_2] = E(X'X)^{-1}X'(X\beta + \varepsilon) + C(X\beta + \varepsilon)]$$

$$= E[(X'X)^{-1}X'\;X\beta + (X'X)^{-1'}X'\varepsilon + CX\beta + CC]$$ karena

nilai $E[\hat{\beta}_2] = \beta$. Oleh karena itu, nilai $CX = 0$ sehingga :

$$= E[\{(X'X)^{-1}X'\varepsilon + C\varepsilon\}\{(X'X)^{-1}X'\varepsilon + C\varepsilon\}']$$

$$= E[\{(X'X)^{-1}X'\varepsilon + C\varepsilon\}\{\varepsilon'X(X'X)^{-1}X'\varepsilon + \varepsilon'C'\}]$$

$$= E[\{(X'X)^{-1}X' + C\}\varepsilon\varepsilon'\{X(X'X)^{-1} + \varepsilon'C'\}]$$

$$= [\{(X'X)^{-1}X' + C\}E(\varepsilon\varepsilon')\{\,X(X'X)^{-1} + \varepsilon'C'\}]$$

$$= \;\sigma_\varepsilon^2[(X'X)^{-1}X' + C\}\{X(X'X)^{-1} + (X'X)^{-1} + C'\}]$$

$$=\sigma_\varepsilon^2[(X'X)^{-1}X'X(X'X)^{-1} + CX(X'X)^{-1} + (X'^{X)})^{-1}X'C'CC']$$

$$= \sigma_\varepsilon^2(X'X)^{-1} + CC']$$

$$= \sigma_{\varepsilon}^2 (X'X)^{-1} + \sigma_{\varepsilon}^2 CC'$$

$$\mathrm{Var}(\widehat{\beta_2}) = \mathrm{var}(\hat{\beta}_2) + \sigma_{\varepsilon}^2$$

Jadi terbukti $(\hat{\beta}) < \mathrm{var}(\hat{\beta}_2)$ maka $\hat{\beta}$ adalah estimator yang terbaik. Dikarenakan estimatir kuadrat terkecil memenuhi sifat linier, tak bias dan menpunyai variansi minimum.

2.8 Pricipal Component Regression (PCA)

Menurut *Jolliffe (1986) Principal component analysis (pca)* atau yang sering kita sebut dengan analisis komponen utama merupakan teknik statistik yang secara linier mentranformasikan sekelompok variabel data asli menjadi sekelompok variabel data subtansial yang tidak terkorelasi. Setelah beberapa komponen hasil dari PCA yang bebas dari multikolinieritas maka komponen tersebut menjadi variabel bebas atau independent baru yang dianaliss pengaruhnya terhadap variabel terikat atau dependent dengan menggunakan analisis regresi.

Langkah-langkah PCA adalah sebagai berikut :

a. Kaiser-Meyer-Olkin (KMO) dan Barlett Test

Mengenai layak atau tidaknya analisis faktor, maka perlu dilakukan uji Kaiser-Meyer-Olkin (KMO) dan Barlett Test. Apabila nilai KMO berkisar antara 0,5 sampai dengan 1 maka analisis faktor layak digunakan. Namun, jika nilai KMO kurang dari 0,5 maka analisis faktor tidak layak dilakukan. Sedangkan Barlett Testdigunakan untuk menguji apakah benar variabel-variabel yang dilibatkan berkorelasi.

Hipotesis:

H0: tidak ada korelasi antarvariabel bebas

H1: ada korelasi antarvariabel bebas

Kriteria uji dengan melihat p-value (signifikan): terima H0 jika sig. > 0,05 atau tolak H0 jika sig.< 0,05.

b. Anti Image Matriks

Bagian Anti Image Correlation, khususnya pada angka korelasi yang bertanda a (arah diagonal dari kiri atas ke kanan bawah). Angka MSA (Measure of Sampling Adequay) berkisar dari 0 sampai 1, dengan kriteria sebagai berikut:

- MSA = 1, variabel tersebut dapat diprediksi tanpa kesalahan oleh variabel lain.
- MSA > 0,5, variabel masih bisa diprediksi dan bisa dianalisis lebih lanjut.
- MSA < 0,5, variabel tidak bisa diprediksi dan tidak bisa dianalisis lebih lanjut, atau dikeluarkan dari variabel lainnya.

c. Communalities

Communalities menunjukkan berapa varians yang dapat dijelaskan oleh faktor yang terbentuk.

d. Total Variance Explained

Dalam analisis faktor terdapat beberapa komponen yang merupakan variabel. Setiap faktor mewakili variabel yang dianalisis. Kemampuan setiap faktor mewakili variabel yang dianalisis ditunjukkan oleh besarnya varians yang dijelaskan, yang disebut dengan eigenvalue. Eigenvalue menunjukkan kepentingan relatif masing-masing faktor dalam menghitung varians ketiga variabel yang dianalisis. Susunan eigenvalue selalu diurutkan dari yang terbesar sampai yang terkecil, dengan kriteria bahwa angka eigenvalue di bawah 1 tidak digunakan dalam menghitung jumlah faktor yang terbentuk.

e. Componen Matriks

Componen Matriks merupakan tabel yang berisikan factor loading (nilai korelasi) antara variabel-variabel analisis dengan faktor yang terbentuk.

f. Component Score Coefficient Matriks

Setelah didapatkan faktor yang terbentuk melalui proses reduksi, maka perlu dicari persamaan sehingga dapat dihitung skor setiap faktor secara manual. Persamaan yang dibuat mirip dengan regresi linear berganda, hanya dalam persamaan faktornya tidak terdapat konstanta. Setelah komponen hasil PCA yang bebas multikolinearitas diperoleh maka komponen-komponen tersebut diregresikan atau dianalisa pengaruhnya terhadap variabel tak bebas (Y) dengan menggunakan analisis regresi linear.

2.9 Pricipal Component Regression (PCR)

Principal component regression (pcr) adalah kombinasi antara regresi dengan *Principal component analisis (pca)* kegunaan metode yang digunakan untuk mengatasi masalah multikolinieritas. Metode ini akan menghasilkan komponen-komponen utama yang tidak terkorelasi. Perbedaan diantara kedua meode ini adalah jika Principal component regression *(pcr)* untuk mengetahui ada atau tidaknya hubungan variabel dependent dan independen, sedangkan Principal componet analysis *(pca)* untuk mengsederhanakan variabel yang diamati dengan cara menyusutkan (mereduksi) dimensi. Perlu diketahui jika komponen utama diikut sertakan dalam model regresi, maka model yang dihasilkan sama seperti metode kuadrat terkecil. Sebaliknya jika beberapa komponen utama yang diikut sertakan maka akan diperoleh penduga dengan koefisien regresi yang bias namun memiliki *variance* yang minimum. Menurut *Dreper dan Smith (1992)* Principal component regression adalah analisis regresi dari variabel-variabel dependent atau terikat terhadap komponen utama yang tidak saling terkorelasi, dimana semua komponen utama merupakan kombinasi linier dari semua variabel independent atau bebas. Hal ini dilakukan untuk menghilangkan korelasi diantara variabel melalui tensformasi variabel asal ke variabel baru (merupakan kombinasi linier dari variabel-variabel asal) yang tidak saling terkorelasi dari p hasil variabel asal dapat dibentuk p hasil komponen utama, dipilih k hasil dari komponen utama $(k < p)$ yang telah menerangkan keragaman data yang cukup tinggi (antara 80% sampai 90%). Komponen utama yang dipih adalah $(k\ hasil)$dapat menggantika p hasil variabel asal tampa mengurangi informasi.

Pembentukan regresi utama melalui analisis komponen utama ada dua cara yaitu komponen utama yang berdasarkan matriks kovarians dan komponen utama yang berdasarkan matriks korelasi. Matriks korelasi dari data yang telah distandarisasi (bentuk buku Z) digunakan jika variabel yang diamati tidak memiliki satuan pengukuran yang sama. Sedangkan matrik kovarians digunakan jika semua variabel yang diamati mempunyai satuan pengukuran yang sama.

Analisis *Principal component regression (pcr)* adalah analisis regresi variabel dependent atau terikat terhadap komponen utama yang tidak saling terkorelasi, dengan regresi komponenen utama yang dinyatakan sebagai berikut :

$$Y = w_0 + w_1K_1 + w_2K_2 + \cdots + w_mK_m + \varepsilon$$

Dimana :

Y : Variabel dependent

K : Komponen utama

W : Parameter regresi utama

$K_1, K_2, K_3, \ldots K_m$ menunjukkan komponen utama yang dilibatkan dalam analisis regresi komponen utama, dengam besaran *m* lebih kecil dari banyaknya variabel bebas yaitu *p*, serta *Y* sebagai variabel terikat.

Komponen utama adalah gabungan linier variabel baku Z, sehingga diperoleh

$$K_1 = a_{11}Z_1 + a_{21}Z_2 + \cdots + a_{pi}Z_p$$

$$K_2 = a_{12}Z_1 + a_{22}Z_2 + \cdots + a_{p2}Z_p$$

$$\vdots$$

$$K_m = a_{1m}Z_1 + a_{2m}Z_2 + \cdots + a_{pm}Z_p$$

Jika $K_1, K_2, K_3, \dots K_m$ dalam persamaan didistribusikan kembali kedalam persamaan regresi komponen utama, maka persamaan tersebut diperoleh :

$$Y = W_0 + W_1(a_{11}Z_1 + a_{21}Z_2 + \cdots + a_{p1}Z_p) + W_2(a_{12}Z_1 + a_{21}Z_2 + \quad \dots + a_{p2}Z_p) + \cdots \ + W_m(a_{1m}Z_1 + a_{2m}Z_2 + \cdots + a_{pm}Z_p) + \varepsilon$$

$$= W_0 + W_1a_{11}Z_2 + W_1W_{21}Z_2 + \cdots + W_1a_{p1}Z_p + W_2a_{12}Z_1 + a_{22}Z_2$$

$$+ \cdots + W_2a_{p2}Z_p + W_ma_{1m}Z_1 + W_ma_{2m}Z_2 + W_ma_{pm}Z_p + \varepsilon$$

$$= W_0 + (W_1a_{11} + W_2a_{12} + \cdots + W_ma_{1m})Z_1 + (W_1a_{21} + W_2a_{21} + \quad \dots + W_ma_{2m})Z_2 + \cdots + (W_1a_{p1} + W_2a_{p1} + \cdots + W_ma_{pm})Z_p + \varepsilon$$

Maka dari persamaan diatas diperoleh persamaan regresi dengan komponen utama sebagai berikut :

$$Y = b_0 + b_1Z_1 + \dots + b_pZ_p$$

Dengan :

$$b_0 = W_0$$

$$b_1 = W_1a_{11} + W_2a_{12} + \cdots + W_ma_{1m}$$

$$b_2 = W_1a_{11} + W_2a_{22} + \cdots + W_ma_{2m}$$

$$\vdots$$

$$bp = W_1a_{p1} + W_2a_{22} + \cdots + W_ma_{pm}$$

Syarat *Principal component regression (pcr)* adalah sebagai berikut :

1. Menghitung nilai eigen vektor dan nilai eigen velue pada matrik korelasi atau kovarians
2. Terdapat *p* komponen utama yang orthogonal dan tidak terkorelasi
3. Dipilih komponen yang eigen velue>1 atau yang mampu menerangkan keragaman cukup tinggi (80%-90%)
4. Regresi variabel dependent atau terikat dengan komponen-komponen utama yang telah terpilih.

2.10 Regression Ridge

Regresi Ridge adalah salah satu metode untuk mengatasi multikolinieritas dengan cara memodifikasi metode kuadrat terkecil. Metode yang digunakan dalam mendeteksi multikolinieritas adalah *ridge trace.* Salah satu kendala dalam regresi ridge adalah menentukan nilai *c* yang tepat. *Centering* dan *rescalling* data merupakan bagian dari membekukan variabel. Bila terdapat multikolinieritas ganda yang besar maka kuadrat terkecil menghasilkan penaksir tak bias untuk koefisien regresi, tapi penaksir menpunyai variansi yang besar. Variansi yang besar menimbulkan dua kesulitas yaitu penaksir tidak stabil artinya sensitif terhadap perubahan kecil pada data yang kelihatan tidak penting dan penaksir cenderung menghasilkan koefisien yang terlalu besar. Cara mengatasinya adalah mengguanakan metode kuadrat terkecil dan menggunakan penaksir yang bias, penggunaan bias tentu dalam dugaan variansi penaksur dapat diperkecil.

Regresi ridge diperoleh dengan memasukkan suatu konstanta pembiasan dalam persamaan normal kuadrat terkecil yaitu :

$$\beta^R(c) = (Z^T Z + CI)^{-1} Z^T Y^*$$

Keterangan :

β^R = dugaan koefisien regresi ridge

Z = matriks X yang tertransformasi ($centerd\ dan\ scalled\ matrix$)

Z^T = matriks transpose dari Z

Y^* = vektor y yang telah ditransformasikan

c = Parameter ridge, umumya c terletak pada selang(0,1) atau $0 \leq c \leq 1$

I = matriks identitas yang berukuran pxp

Dalam hal ini β^R adalah vektor koefisien-koefisien ridge yaitu :

$$\beta^R(c) = \begin{bmatrix} \beta_1^R(c) \\ \beta_2^R(c) \\ \vdots \\ \beta_p^R(c) \end{bmatrix}$$

Dengan persaman regresi ridge yaitu :

$$\hat{Y}^* = \beta_1^R Z_{i1} + \beta_2^R Z_{i2} + \cdots + \beta_p^R Z_{ip}$$

Nilai VIF bagi koefisien regresi ridge β^R mirip untuk koefisien kuadrat terkecil. Dengan kata kata lain nilai VIF untuk koefisien β^R mengukur besarnya variansi bagi β^R relatif terhadap variansi koefisien yang sama jika peubah besarnya tidek terkorelasi. Ini menunjukkan bahwa nilai VIF bagi koefisien regresi ridge merupakan unsur-unsur diagonal dari matriks :

$$(Z^T Z + cI)^{-1} Z^T Z (Z^T Z + cI)^{-1}$$

2.11Variabel Dependen dan independen

Variabel adalah fenomena yang bervariasai dalam bentuk kualitas, kuantitas, mutu, dan standar. Vaiabel merupakan suatu yang diukur dan cara mengkurnya berbeda-beda Variabel juga memiliki jenis yaitu dependen (X) dan independen (Y). Variabel dependen adalah variabel terikat atau variabel yang dipengaruhi atau tergantung variabel lainnya sedangkan variabel independen adalah variabel bebas yang menjadi sebab timbulya atau berubahnya variabel independen atau bisa diartikan variabel independen adalah variabel yang mempengaruhi variabel lainnya.

Pada penelitian tersebut memiliki variabel independen (Y) dan variabel dependen (X) yang masing-masing variaebel diwakilih oleh IPM untuk variabel independen (Y) sedangkan variabel dependen (X) diwakilin oleh harapan lama sekolah (HLS), rata-rata lama sekolah (RLS), angka harapan hidup saat lahir (AHH), produk domestik regional bruto (PDRB), angka partisipasi sekolah (APS), angka melek huruf (AMH), tingkat pengangguran terbuka (TPT), tingkat partisipasi angkatan kerja (TPAK), pengeluaran perkapita non makanan (PPK), Presentase Penduduk Miskin (PPM) dan status kepemilikan rumah milik sendiri (SKTMS).

2.11.1 Indeks Pembangunan Manusia (Y)

Untuk pertama kalinya ditahun 1990 pengukuran Indeks pembangunan manusia (IPM) dilakukan oleh United Nattions Develoment Programe (UNDP). IPM dipublikasikan secara berkala dalam laporan tahunan Human Development Report (HDR). IPM menjelaskan bagaimana penduduk dapat mengakses hasil pembangunan dalam memperoleh pendapatan, kesehatan, pendidikan dan sebagainya. UNDP menbuat suatu gagagsan dimana pembangunan manusia dapat dikur melalui tiga dimensi yaitu :

- Umur panjang dan hidup yang sehat
- Pengetahuan
- tandar layak hidup

Ketiga dimensi tersebut memiliki pengertian sangat luas karena terkait banyak faktor. Pada laporan pertamanya, UNDP mengukur dimensi kesehatan dengan menggunakan angka harapan hidup waktu lahir. Selanjutnya untuk mengukur dimensi pengetahuan digunakan angka melek huruf. Adapun untuk mengukur dimensi standar hidup layak digunakan indikator Produk Domestik Bruto (PDRB) per kapita.

2.11.2 Angka Harapan Hidup Saat Lahir (X)

Angka harapan hidup saat lahir (AHH) adalah perkiraan rata-rata tambahan umur seseorang yang diharpkan dapat terus hidup. AHH juga didefinisikan sebagi jumlah tahun yang dijalani oleh seseorang setelah irang tersebut ulang tahun yang ke-x. Jadi pada umumnya AHH yang dimaksud adalah rata-rata jumlah tahun yang dijalani oleh seseorang sejak orang tersebut lahir. Rasio ketergantungan (Depency Ratio) adalah perbandingan antara jumlah penduduk umur 0-14 tahun, ditambah dengan jumlah penduduk 65 tahun ke atas (kedanya disebut bukan angkatan kerja) dibandingkan dengan jumlah penduduk 15-64 tahun (angkatan kerja).

2.11.3 Rata – Rata Lama Sekolah (X)

Rata-rata lama sekolah (RLS) / *mean year of school(MYS)* adalah jumlah tahun yang digunakan oleh penduduk dalam menjalani pendidikan formal. Asumsi dalam kondisi normal rata-rata lama sekolah suatu wilayah tidak akan turun. Rata-rata lama sekolah dihitung berdasarkan penduduk usia 25 tahun ke atas dengan asumsi pada umur 25 tahun proses pendidikan sudah berakhir. Perhitungan rata-rata lama sekolah pada usia 25 tahun ke atas juga mengikuti standar internasuonal yang digunakan oleh UNDP. Dalam perhitungan rata-rata lama sekolah, peduduk yang tamat SD dipehitungkan lama sekolah 6 tahun, tamat SMP dipehitungkan lama sekolah 9 tahun, penduduk tamat sekolah SMA dipehitungkan lama sekolah 12 tahun tanpa menperhitungkan apakah pernah tinggal kelas atau tidak.

2.11.4 Angka Partisipasi Sekolah (X)

Angka partsipasi sekolah (APS) adalah perbandingan jumlah murid kelompok usia sekolah tertentu yang bersekolah pada berbagai jenjang pendidikan dengan penduduk kelompok usia sekolah yang bersesuaian . Semakin tinggi APS maka semakin banyak usia sekolah yang bersekolah di suatu daerah.

2.11.5 Angka Melek Huruf (X)

Angka melek huruf (AMH) adalah proporsi penduduk usia 15 tahun keatas yang memiliki kemampuan menbaca dan menulis kalimat sederhana dalam huruf latin, huruf arab dan huruf lainnya (seperti hrurf jawa, kanji dll). Tingkat melek huruf yang tinggi menunjukkan adanya sebuah sistem pendidikan dasar yang efektif atau program keasaraan yang memungkinkan sebagian besar penduduk yang memperoleh kemampuan menggunakan kata-kata tertulis dalam kehidupan sehari-hari.

2.11.6 Produk Domestik Regional Bruto (X)

Produk domestik regional bruto merupakan produk domestik ditambah dengan pendapatan dari faktor-faktor produksi yang diterimah dari luar daerah atau negeri dikurangi dengan pendapatan dari faktor produksi yang dibayarkan keluar daerah atau negeri. Jadi produk regional merupakan produk yang ditimbulkan oleh faktor produksi yang dimiliki presiden.

2.11.7 Tingkat Pengangguran Terbuka (X)

Tingkat pengangguran terbuka (TPT) adalah presentase jumlah pengangguran terhadap jumlah angkatan kerja dengan usia 15 tahun ke atas. TPT yang tinggi menunjukkan bahwa banyaknya angkatan kerja yang tidak terserap pada pasar kerja.

2.11.8 Tingkat Partisipasi Angkatan Kerja (X)

Tingkat partisipasi angkatan kerja (TPAK) adalah ukuran proporsi penduduk usia kerja dengan usia 15 tahun ke atas disuatu daerah yang bergerak aktif di pasar tenaga kerja dan menjadi faktor penting dalam menggerakkan pembangunan.

2.11.9 Presentasi Pengeluaran Perkapita Non - Makanan (X)

Pengeluaran perkapita non makanan adalah biaya yang dikeluarkan untuk komsumsi non makanan bagi anggota rumah tangga yang dijadikan sebagai tolak ukur dalam kesejateraan ekonomi penduduk. Semakin besar proporsi pengeluaran non-makanan maka semakin meningkatnya kesejateraan masyarakat. Sebaliknya semakin kecil pengeluaran non-makanan maka kesejateraan masyarakan semakin menurun.

2.11.10 Status Kepemilikan Rumah Sendiri (X)

Status Kepemilikan Rumah Milik Sendiri adalah kepemilikan rumah atau bangunan tempat tinggal milik sendiri menurut jenis bukti kepemilikan tanah tempat tinggal milik sendiri dikali dengan 100(persen), dibagi dengan jumlah rumah tangga dengan status kepemilikan rumah atau bangunan milik sendiri.

2.11.11 Presentase Penduduk Miskin (X)

Presntase penduduk miskin (headcount index/P0) adalah penduduk miskin yang berada dibawa garis kemiskinan. Headcount Index secara sederhana mengukur proporsi yang diketegorikan miskin. Angka yang ditunjukkan oleh headcount index menunjukkan proporsi penduduk miskin disuatu wilayah. Presentese penduduk miskin yang tinggi menunjukkan bahwa tingkat kemiskinan disuatu wilaya juga tinggi