KAJIAN *PRINCIPAL COMPONENT LOGISTIC REGRESSION* (PCLR) DAN *PRINCIPAL COMPONENT LOGISTIC REGRESSION STEPWISE* (PCLR$_{(S)}$) DALAM PEMILIHAN REGRESI LOGISTIK TERBAIK PADA KASUS MULTIKOLINIERITAS

SKRIPSI

oleh :
RATIH RATNAWATI
0410950040-95

PROGRAM STUDI STATISTIKA
JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS BRAWIJAYA
MALANG
2008

KAJIAN *PRINCIPAL COMPONENT LOGISTIC REGRESSION* (PCLR) DAN *PRINCIPAL COMPONENT LOGISTIC REGRESSION STEPWISE* (PCLR$_{(S)}$) DALAM PEMILIHAN REGRESI LOGISTIK TERBAIK PADA KASUS MULTIKOLINIERITAS

SKRIPSI

Sebagai salah satu syarat untuk memperoleh gelar
Sarjana Sains dalam bidang Statistika

oleh :
RATIH RATNAWATI
0410950040-95

PROGRAM STUDI STATISTIKA
JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS BRAWIJAYA
MALANG
2008

LEMBAR PENGESAHAN SKRIPSI

KAJIAN *PRINCIPAL COMPONENT LOGISTIC REGRESSION* (PCLR) DAN *PRINCIPAL COMPONENT LOGISTIC REGRESSION STEPWISE* ($PCLR_{(S)}$) DALAM PEMILIHAN REGRESI LOGISTIK TERBAIK PADA KASUS MULTIKOLINIERITAS

Oleh :
RATIH RATNAWATI
0410950040-95

Setelah dipertahankan di depan Majelis Penguji
pada tanggal 17 Juli 2008
dan dinyatakan memenuhi syarat untuk memperoleh gelar
Sarjana Sains dalam bidang Statistika

Pembimbing I	Pembimbing II
Ir. Atiek Iriany, MS.	**Ir. Soepraptini, MSc.**
NIP. 131 759 544	**NIP. 130 518 968**

Mengetahui,
Ketua Jurusan Matematika
Fakultas MIPA Universitas Brawijaya

Dr. Agus Suryanto, MSc.
NIP. 132 126 049

LEMBAR PERNYATAAN

Saya yang bertanda tangan di bawah ini :

Nama : Ratih Ratnawati
NIM : 0410950040-95
Jurusan : Matematika
Penulisan skripsi berjudul : Kajian *Principal Component Logistic Regression* (PCLR) Dan *Principal Component Logistic Regression Stepwise* ($PCLR_{(S)}$) Dalam Pemilihan Regresi Logistik Terbaik Pada Kasus Multikolinieritas

Dengan ini menyatakan bahwa :

1. **Isi dari skripsi yang saya buat adalah benar-benar karya sendiri dan tidak menjiplak karya orang lain, selain nama-nama yang termaktub di isi dan tertulis di daftar pustaka dalam skripsi ini.**
2. **Apabila di kemudian hari ternyata skripsi yang saya tulis terbukti hasil jiplakan, maka saya akan bersedia menanggung segala risiko.**

Demikian pernyataan ini saya buat dengan segala kesadaran.

Malang, 17 Juli 2008
Yang menyatakan,

Ratih Ratnawati
NIM. 0410950040-95

KAJIAN *PRINCIPAL COMPONENT LOGISTIC REGRESSION* (PCLR) DAN *PRINCIPAL COMPONENT LOGISTIC REGRESSION STEPWISE* (PCLR$_{(S)}$) DALAM PEMILIHAN REGRESI LOGISTIK TERBAIK PADA KASUS MULTIKOLINIERITAS

ABSTRAK

Regresi logistik adalah bentuk khusus analisis regresi yang memiliki variabel respon bersifat kategori dan variabel prediktor bersifat kategori, kontinyu dan campuran. Dalam regresi logistik pada kasus multikolinieritas diperlukan metode reduksi dimensi untuk mendapatkan model yang tepat. *Principal Component Logistic Regression* (PCLR) merupakan teknik *multivariate* yang dapat dipakai untuk mereduksi dimensi dari variabel-variabel prediktor. Dalam penelitian ini ingin diterapkan suatu metode pemilihan variabel untuk mengatasi masalah multikolinieritas yaitu *Principal Component Logistic Regression Stepwise* (PCLR$_{(s)}$) yang merupakan pengembangan dari PCLR dimana perbedaan mendasar dari PCLR dan PCLR$_{(s)}$ adalah pada penduga koefisien regresi ($\hat{\gamma}_{(s)}$). Tujuan dari penelitian ini adalah memperoleh model regresi logistik terbaik dengan metode PCLR dan PCLR$_{(s)}$, kemudian membandingkan kedua metode tersebut. Kriteria pemilihan model terbaik berdasarkan nilai *Pearson Residuals* dan *Percent Correct Predictions* (PCP). Dalam penelitian ini digunakan dua macam data sekunder yang terdiri dari faktor-faktor yang mempengaruhi kinerja jantung 10 tahun setelah pemeriksaan terakhir dan faktor-faktor yang mempengaruhi fertilitas induk kambing PE. Berdasarkan penelitian ini, dapat disimpulkan bahwa metode PCLR$_{(S)}$ lebih baik dibanding PCLR. Sehingga lebih disarankan untuk menggunakan PCLR$_{(S)}$ untuk membentuk model regresi logistik apabila terdapat multikolinieritas pada sebagian dan semua variabel prediktor.

Kata Kunci: regresi logistik, multikolinieritas, PCLR, PCLR$_{(S)}$.

ANALYTICAL STUDY OF PRINCIPAL COMPONENT LOGISTIC REGRESSION (PCLR) AND PRINCIPAL COMPONENT LOGISTIC REGRESSION STEPWISE ($PCLR_{(S)}$) IN PURPOSE TO CHOOSE THE BEST LOGISTIC REGRESSION FOR MULTICOLINIERITY

ABSTRACT

Logistic regression is a special form of regression analysis which is response variable have categorical character and predictor variable have categorical, mixture and continues characters. In logistic regession at multicolinierity case, it is needed a dimension reduction to find the precise model. Principal Component Logistic Regression (PCLR) is multivariate technique used to reduce the dimension from the predictor variable. This experiment is aimed to apply a Principal Component Logistic Regression Stepwise ($PCLR_{(S)}$) as the method in choosing variable to solve the multicolinierity case, which is an ascending development from PCLR, where the difference between PCLR and $PCLR_{(S)}$ is the presumption of regression coefficient ($\hat{\gamma}_{(s)}$). Find out the best logistic regression model used the comparison of both PCLR and $PCLR_{(S)}$ is the final purpose of this experiment. The criteria used in choosing the best model based on the value of pearson residuals and Percent Correct Predictions (PCP). In this case, it is used two kinds of secondary data consist of the factors affecting the heart performance after 10 years after the last inspection and the factors affecting the fertility of PE doe. Pursuant to this research, can be concluded that $PCLR_{(S)}$ is better PCLR. It is suggested to use $PCLR_{(S)}$ to form logistic regression model when there is multicolinierity in partial or entire of predictor variable.

Keyword: logistic regression, multicolinierity, PCLR, $PCLR_{(S)}$.

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT atas limpahan rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “Kajian *Principal Component Logistic Regression* (PCLR) Dan *Principal Component Logistic Regression Stepwise* ($PCLR_{(S)}$) Dalam Pemilihan Regresi Logistik Terbaik Pada Kasus Multikolinieritas” sebagai salah satu syarat untuk memperoleh gelar Sarjana Sains dalam bidang Statistika. Oleh karena itu penulis mengucapkan rasa hormat dan terima kasih kepada:

1. Ibu Ir. Atiek Iriany, MS., selaku Dosen Pembimbing I dan Ibu Ir. Soepraptini, MSc., selaku Dosen Pembimbing II yang telah memberikan pengarahan dan masukan dengan sabar kepada penulis selama penyusunan skripsi ini.
2. Ibu Suci Astutik, SSi., MSi., Ibu Dra. Ani Budi Astuti, MSi. dan Bapak Adji Achmad R.F, SSi., MSc., atas arahan serta nasehat yang telah diberikan kepada penulis selama penyusunan skripsi.
3. Dr. Agus Suryanto, MSc., selaku Ketua Jurusan Matematika Universitas Brawijaya.
4. Bapak dan Ibu Dosen Statistika atas didikan selama kuliah hingga penulis bisa menyelesaikan kuliah.
5. Bapak, Ibu dan keluarga besar di Madiun yang senantiasa mendoakan dan membantu penulis mencapai yang terbaik.
6. Teman-teman Program Studi Statistika 2004, dan Korps Sukarela yang memberikan dukungan, semangat dan bantuan.
7. Semua pihak yang tidak dapat penulis sebutkan satu per satu yang telah banyak membantu dan memberikan dorongan selama penulisan skripsi ini.

Penulis menyadari bahwa skripsi ini masih jauh dari kesempurnaan mengingat keterbatasan kemampuan penulis. Untuk itu dengan segala kerendahan hati penulis mengharap kritik dan saran. Akhirnya penulis berharap semoga skripsi ini dapat bermanfaat bagi pembaca dan penulis.

Malang, 17 Juli 2008

Penulis

DAFTAR ISI

DAFTAR GAMBAR

Halaman

DAFTAR TABEL

Halaman

Halaman

DAFTAR LAMPIRAN

BAB I
PENDAHULUAN

1.1 Latar Belakang Masalah

Banyak analisis statistika bertujuan untuk mengetahui apakah ada hubungan antara dua atau lebih variabel. Bila hubungan demikian ini dinyatakan atau dimodelkan dalam bentuk rumus matematik, maka dapat digunakan untuk keperluan peramalan dan pengklasifikasian.

Analisis regresi adalah salah satu teknik perhitungan statistika untuk meneliti dan memodelkan hubungan antara beberapa variabel. Dalam aplikasinya banyak dijumpai bentuk fungsi yang menggambarkan hubungan non linier di antara dua variabel. Terdapat banyak bentuk kurva non linier yang dapat digunakan untuk menyatakan hubungan antara dua variabel atau lebih, maka dalam menganalisis suatu hasil penelitian haruslah ditentukan terlebih dahulu bentuk kurva yang paling tepat untuk mengekspresikan data. Suatu model non linier dapat dilihat dari dua kategori yaitu non linier dalam parameter dan non linier pada variabel. Persamaan fungsi kuadratik, kubik dan bentuk polinomial ordo lebih tinggi dikatakan model non linier pada variabel, sementara persamaan fungsi eksponensial, logaritmik dikatakan model non linier dalam parameter.

Regresi logistik adalah bentuk khusus analisis regresi dengan fungsi logaritmik yang memiliki variabel respon bersifat kategori dan variabel prediktor bersifat kategori, kontinyu atau gabungan antara keduanya. Regresi ini dinamakan dengan regresi logistik karena pembentukan modelnya didasarkan pada bentuk kurva logistik. Kurva ini berbentuk landai atau kemiringannya kecil pada bagian atas dan bawah variabel prediktor, namun berbentuk curam pada bagian tengah. Persamaan modelnya menghasilkan peluang kejadian yang dipakai sebagai ukuran untuk mengklasifikasikan pengamatan.

Di dalam analisis yang melibatkan banyak variabel prediktor, sering terjadi adanya hubungan yang erat antara dua atau lebih variabel prediktor. Apabila variabel-variabel prediktor ini saling berkorelasi maka dikatakan terdapat multikolinieritas. Gejala seperti ini menimbulkan masalah dalam analisis. Dalam kasus multikolinieritas, *standard error* koefisien regresi besar sehingga interval koefisien menjadi lebar yang mengakibatkan tingkat ketelitian dalam menguji koefisien menjadi berkurang. Pada kasus multikolinieritas dengan banyak variabel prediktor, diperlukan metode reduksi dimensi untuk mendapatkan model yang tepat.

Principal Component Logistic Regression (PCLR) merupakan teknik *multivariate* yang dapat dipakai untuk mereduksi dimensi dari variabel-variabel prediktor. Prinsip utama dari model regresi logistik dengan PCLR adalah meregresikan skor komponen utama dengan variabel respon atau dengan kata lain metode ini merupakan analisis regresi dari variabel respon terhadap komponen-komponen utama yang tidak saling berkorelasi. Namun apabila tidak ada korelasi yang kuat antara variabel prediktor dan respon, PCLR tidak akan memberikan hasil yang bagus, karena variabel-variabel baru yang dihasilkan merupakan transformasi dari variabel asal yang ditata berdasarkan besar keragaman.

Dalam penelitian ini ingin dicobakan suatu metode pemilihan variabel untuk mengatasi masalah multikolinieritas, yaitu metode *Principal Component Logistic Regression Stepwise* ($PCLR_{(s)}$). Metode ini merupakan pengembangan dari PCLR. Prosedur awal metode ini sama dengan prosedur yang dilakukan pada PCLR, yaitu melakukan tranformasi variabel prediktor asli ke variabel komponen utama, kemudian melakukan pemilihan komponen utama berdasarkan metode *stepwise*.

1.2 Perumusan Masalah.

Masalah yang perlu dirumuskan dalam penelitian ini adalah:

1. Bagaimana menerapkan PCLR dan $PCLR_{(s)}$ pada regresi logistik yang tidak memenuhi asumsi non multikolinieritas ?
2. Bagaimana mendapatkan koefisien regresi logistik yang menggunakan PCLR dan $PCLR_{(s)}$?
3. Membandingkan hasil penerapan metode mana yang lebih tepat digunakan untuk mengatasi multikolinieritas pada regresi logistik ?

1.3 Batasan Masalah

Pada penelitian ini, permasalahan dibatasi pada penentuan model terbaik dengan metode PCLR dan $PCLR_{(s)}$ jika variabel prediktor mengandung mulktikolinieritas. Data yang digunakan adalah data yang disifati multikolinieritas pada sebagian dan semua variabel prediktor, memiliki variabel respon bersifat dikotomus, dan variabel prediktor kontinyu.

1.4 Tujuan Penelitian.

Tujuan dari penelitian ini adalah:

1. Mengkaji penerapan PCLR dan $PCLR_{(s)}$ pada regresi logistik dengan data yang disifati multikolinieritas pada sebagian dan semua variabel prediktor.
2. Memperoleh model regresi logistik dengan metode PCLR dan $PCLR_{(s)}$.
3. Membandingkan hasil penerapan metode PCLR dan $PCLR_{(s)}$ pada data yang disifati multikolinieritas pada sebagian dan semua variabel prediktor.

1.5 Manfaat Penelitian.

Setelah dilakukan penelitian ini, maka diharapkan dapat menjadi tambahan informasi untuk memilih metode yang lebih tepat untuk mengatasi multikolinieritas pada data yang disifati multikolinieritas pada sebagian dan semua variabel prediktor.

BAB II
TINJAUAN PUSTAKA

2.1. Regresi Logistik.

2.1.1. Gambaran Umum.

Model Regresi logistik merupakan salah satu metode yang digunakan untuk mengetahui hubungan variabel respon yang berkategori dengan satu atau lebih variabel prediktor yang bersifat kategorial atau kontinyu. Banyaknya kategori variabel respon dapat berbentuk dikotomus atau politomus (Hosmer dan Lemeshow, 1989). Model dari variabel respon biner (dikotomus) adalah sebagai berikut:

0 : Bila variabel respon tidak terjadi (gagal)

1 : Bila variabel respon terjadi (sukses)

Agresti (1990) menyatakan, jika terdapat variabel respon Y_i dan π_i adalah peluang sukses bagi variabel Y_i, maka:

$$Y_i = \begin{cases} 1, \mathrm{P}(\mathrm{Y_i} = 1) = \pi_i \\ 0, \mathrm{P}(\mathrm{Y_i} = 0) = 1 - \pi_i \end{cases}$$

di mana Y_i akan mengikuti sebaran bernoulli dengan parameter π_i, sehingga fungsi kepekatan peluangnya yaitu:

$$\begin{aligned} f(Y_i) &= \pi_i^{y_i}(1-\pi_i)^{1-y_i} \\ &= (1-\pi_i)(\pi_i/(1-\pi_i))^{y_i} \end{aligned} \qquad (2.1)$$

Menurut Agresti (1990), bentuk model regresi logistik adalah:

$$\pi_i = \frac{\exp(\beta_o + \beta_1 x_{i1} + ... + \beta_p x_{ip})}{1+\exp(\beta_o + \beta_1 x_{i1} + ... + \beta_p x_{ip})} = \frac{\exp\left[\sum_{i=1}^{n}\sum_{j=0}^{p}\beta_j x_{ij}\right]}{1+\exp\left[\sum_{i=1}^{n}\sum_{j=0}^{p}\beta_j x_{ij}\right]} \qquad (2.2)$$

di mana:

i = 1, 2,…,n

j = 1, 2,…,p

p = banyak variabel prediktor

n = ukuran contoh

$\beta_o, \beta_1, \beta_2, .., \beta_p$ = koefisien regresi setiap variabel prediktor

$x_{i1}, x_{i2}, \ldots, x_{ip}$ = pengamatan ke-i untuk variabel prediktor ke 1, 2,…,p

π_i = peluang sukses kejadian

Hosmer dan Lemeshow (1989), menyatakan bentuk transformasi logit untuk model (2.2) menjadi:

$$g(x_i) = \ln \frac{\pi_i}{1-\pi_i} \tag{2.3}$$

$$g(x_i) = \ln \left[\frac{\dfrac{\exp(\beta_o + \beta_1 x_{i1} + \ldots + \beta_p x_{ip})}{1+\exp(\beta_o + \beta_1 x_{i1} + \ldots + \beta_p x_{ip})}}{1 - \left[\dfrac{\exp(\beta_o + \beta_1 x_{i1} + \ldots + \beta_p x_{ip})}{1+\exp(\beta_o + \beta_1 x_{i1} + \ldots + \beta_p x_{ip})} \right]} \right]$$

$$= \ln(\exp(\beta_o + \beta_1 x_{i1} + \ldots + \beta_p x_{ip}))$$

$$= \beta_o + \beta_1 x_{i1} + \ldots + \beta_p x_{ip} \tag{2.4}$$

$$= \sum_{i=1}^{n} \sum_{j=0}^{p} \beta_j x_{ij}$$

$$g(x_i) = \beta \ \mathbf{X} \tag{2.5}$$

2.1.2. Pendugaan Parameter.

Pendugaan koefisien regresi logistik dilakukan dengan menggunakan metode kemungkinan maksimum. Karena metode ini memiliki banyak kelebihan dibandingkan dengan metode yang lain, yaitu dapat digunakan untuk model nonlinier seperti pada model logistik, dan hasil pendugaan akan mendekati parameter yang diduga (Hosmer dan Lemeshow, 1989).

Pada fungsi logistik dengan variabel respon biner, pengamatan diasumsikan saling bebas dan nilai peluang variabel respon tidak linier terhadap parameter β untuk $x_i = (x_{i1}, x_{i2}, \ldots, x_{ip})$, maka fungsi (2.2) akan dimaksimumkan. Karena setiap pasang pengamatan bersifat bebas, maka fungsi kemungkinan maksimum merupakan perkalian dari masing-masing fungsi peluang untuk $x_{i1}, x_{i2}, \ldots, x_{ip}$ sehingga didapatkan fungsi kemungkinan maksimum:

$$L(\beta_j) = f(Y_1) \text{ x } f(Y_2) \text{ x } \ldots \text{x } f(Y_n) \tag{2.6}$$

$$= \prod_{i=1}^{n} \pi_i^{y_i} (1-\pi_i)^{1-y_i}$$

$$L(\beta_j) = \prod_{i=1}^{n} (1-\pi_i) \left(\frac{\pi_i}{1-\pi_i} \right)^{y_i}$$

$$= \prod_{i=1}^{n} \left[1 - \frac{\exp\left[\sum_{i=1}^{n}\sum_{j=0}^{p} \beta_j x_{ij}\right]}{1+\exp\left[\sum_{i=1}^{n}\sum_{j=0}^{p} \beta_j x_{ij}\right]} \right] \left[\frac{\dfrac{\exp\left[\sum_{i=1}^{n}\sum_{j=0}^{p} \beta_j x_{ij}\right]}{1+\exp\left[\sum_{i=1}^{n}\sum_{j=0}^{p} \beta_j x_{ij}\right]}}{1-\dfrac{\exp\left[\sum_{i=1}^{n}\sum_{j=0}^{p} \beta_j x_{ij}\right]}{1+\exp\left[\sum_{i=1}^{n}\sum_{j=0}^{p} \beta_j x_{ij}\right]}} \right]^{y_i}$$

$$= \prod_{i=1}^{n} \left[\frac{1}{1+\exp\left[\sum_{i=1}^{n}\sum_{j=0}^{p} \beta_j x_{ij}\right]} \right] \left[\frac{\dfrac{\exp\left[\sum_{i=1}^{n}\sum_{j=0}^{p} \beta_j x_{ij}\right]}{1+\exp\left[\sum_{i=1}^{n}\sum_{j=0}^{p} \beta_j x_{ij}\right]}}{\dfrac{1}{1+\exp\left[\sum_{i=1}^{n}\sum_{j=0}^{p} \beta_j x_{ij}\right]}} \right]^{y_i}$$

$$= \prod_{i=1}^{n} \left[\frac{1}{1+\exp\left[\sum_{i=1}^{n}\sum_{j=0}^{p} \beta_j x_{ij}\right]} \right] \left[\exp\left[\sum_{i=1}^{n}\sum_{j=0}^{p} \beta_j x_{ij}\right] \right]^{y_i} \qquad (2.7)$$

Untuk mencari nilai maksimum dari fungsi kemungkinan maka kedua ruas dialgoritmakan. Maksimum dari fungsi $L(\beta_j)$ disebut sebagai log likelihood.

$$\text{Ln}\, L(\beta_j) = \ln \prod_{i=1}^{n} \left[\frac{1}{1+\exp\left[\sum_{i=1}^{n}\sum_{j=0}^{p}\beta_j x_{ij}\right]} \right] \left[\exp\left[\sum_{i=1}^{n}\sum_{j=0}^{p}\beta_j x_{ij}\right]\right]^{y_i}$$

$$= \ln \left[\frac{1}{1+\exp\left[\sum_{i=1}^{n}\sum_{j=0}^{p}\beta_j x_{ij}\right]} \right]^{n} + \ln \left[\exp\left[\sum_{i=1}^{n}\sum_{j=0}^{p}\beta_j x_{ij}\right]\right]^{\sum_{i=1}^{n} y_i}$$

$$= n\left[\ln 1 - \ln \left[\exp\left[\sum_{i=1}^{n}\sum_{j=0}^{p}\beta_j x_{ij}\right]\right]\right]$$

$$+ \sum_{i=1}^{n} y_i \ln \left[\exp\left[\sum_{i=1}^{n}\sum_{j=0}^{p}\beta_j x_{ij}\right]\right]$$

$$= \text{-n ln} \left[1+\exp\left[\sum_{i=1}^{n}\sum_{j=0}^{p}\beta_j x_{ij}\right]\right] + \sum_{i=1}^{n} y_i \ln \left[\exp\left[\sum_{i=1}^{n}\sum_{j=0}^{p}\beta_j x_{ij}\right]\right]$$

$$= \text{-n ln} \left[1+\exp\left[\sum_{i=1}^{n}\sum_{j=0}^{p}\beta_j x_{ij}\right]\right] + \sum_{i=1}^{n} y_i \left[\sum_{i=1}^{n}\sum_{j=0}^{p}\beta_j x_{ij}\right]$$

$$= \sum_{j=0}^{p}\left[\sum_{i=1}^{n} y_i x_{ij}\right]\beta_j - n \ln \left[\exp\left[\sum_{i=1}^{n}\sum_{j=0}^{p}\beta_j x_{ij}\right]\right] \qquad (2.8)$$

Agar nilai fungsi $L(\beta_j)$ mencapai maksimum maka turunan parsial pertama terhadap β_j harus sama dengan nol.

$$\frac{\partial L[\beta_j]}{\partial \beta_j} = \sum_{i=1}^{n}\sum_{j=0}^{p} y_i x_{ij} - \sum_{i=1}^{n} n_i x_{ij} \left[\frac{\exp\left[\sum_{i=1}^{n}\sum_{j=0}^{p}\beta_j x_{ij}\right]}{1+\exp\left[\sum_{i=1}^{n}\sum_{j=0}^{p}\beta_j x_{ij}\right]} \right]$$

$$\sum_{i=1}^{n}\sum_{j=0}^{p} y_i x_{ij} - \sum_{i=1}^{n} n_i x_{ij}\left[\frac{\exp\left[\sum_{i=1}^{n}\sum_{j=0}^{p}\beta_j x_{ij}\right]}{1+\exp\left[\sum_{i=1}^{n}\sum_{j=0}^{p}\beta_j x_{ij}\right]}\right] = 0 \qquad (2.9)$$

$$\sum_{i=1}^{n}\sum_{j=0}^{p} y_i x_{ij} - \sum_{i=1}^{n} n x_{ij}\pi_i = 0 \qquad (2.10)$$

Dalam menyelesaikan persamaan (2.10) tidaklah mudah, karena β_j yang akan diduga bersifat nonlinier, sehingga diperlukan metode iterasi di mana dalam perhitungannya dibantu komputer. Adapun metode iterasi yang digunakan adalah iterasi *Newton-Raphson* (Agresti,1990).

2.1.3. Pengujian Terhadap Pendugaan Parameter.

Menurut Hosmer dan Lemeshow (1989), pengujian terhadap parameter dapat dilakukan secara parsial maupun serentak.

1. Pengujian koefisien $\hat{\beta}$ secara parsial dilandasi pada pembandingan penduga dengan ragam penduganya dengan hipotesis :

   Ho : $\beta_j = 0$ lawan

   H_1 : $\beta_j \neq 0$

   Uji koefisien regresi secara parsial digunakan untuk memeriksa peranan koefisien regresi dari masing-masing variabel prediktor secara individu dalam model. Statistik uji Wald yang digunakan dapat ditulis:

   $$W = \frac{\hat{\beta}_j}{\left[Se(\hat{\beta}_j)\right]} \qquad (2.11)$$

   di mana $\hat{\beta}_j$ adalah penduga bagi β_j dan $Se(\hat{\beta}_j)$ merupakan penduga galat baku bagi $\hat{\beta}_j$. Nilai $Se(\hat{\beta}_j)$ diperoleh dari nilai diagonal utama matriks kovarians, yaitu:

   $$\text{Cov}(\hat{\beta}_j) = [\mathbf{X}^T\mathbf{V}\mathbf{X}]^{-1} \qquad (2.12)$$

$$V = \begin{bmatrix} \hat{\pi}_1(1-\hat{\pi}_1) & 0 & \cdots & 0 \\ 0 & \hat{\pi}_2(1-\hat{\pi}_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \hat{\pi}_n(1-\hat{\pi}_n) \end{bmatrix}$$

Var ($\hat{\beta}_j$) = diagonal utama matriks $[\mathbf{X^T V X}]^{-1}$

$$Se(\hat{\beta}_j) = \sqrt{Var(\hat{\beta}_j)}$$

Hipotesis nol ditolak jika P $[|Z| > W] < \alpha$. Hal ini mengindikasikan bahwa variabel X_j berpengaruh terhadap variabel respon.

2. Pengujian $\hat{\beta}_j$ secara serentak bertujuan untuk membandingkan nilai pengamatan respon dengan nilai dugaan respon untuk kedua model. Model yang pertama adalah model penuh dan model yang kedua adalah model reduksi dengan hipotesis sebagai berikut :

   Ho : $\beta_1 = \beta_2 = ... = \beta_j$ lawan

   H_1 : Paling sedikit ada satu β_j tidak sama dengan nol.

   Uji yang digunakan adalah uji nisbah kemungkinan (*Likelihood Ratio Test*) yaitu:

   $$G = -2 (L_0 - L_p) \qquad (2.13)$$

   dengan : L_0 = log likelihood model regresi logistik tanpa variabel W_j.

   L_p = log likelihood model regresi logistik dengan variabel W_j.

   di mana

   $$\text{Log L} = \sum_{i=1}^{n}\left[y_i \ln\left(\frac{\hat{\pi}_i}{y_i}\right) + (1-y_i)\ln\left(\frac{1-\hat{\pi}_i}{1-y_i}\right)\right] \qquad (2.14)$$

   Jika Ho benar maka G berdistribusi $\chi^2_{(v)}$, dengan *v* adalah selisih banyaknya parameter yang diduga pada kedua model. Hipotesis nol akan ditolak jika nilai $P[\chi^2_{(v)} > G] < \alpha$. Hal ini mengindikasikan bahwa paling sedikit ada satu β_j yang tidak sama dengan nol.

2.1.4. Interpretasi Koefisien Regresi Logistik.

Interpretasi koefisien adalah inferensi dan pengambilan kesimpulan berdasarkan koefisien penduga. Koefisen menggambarkan perubahan atau slope pada variabel respon per unit setiap perubahan variabel prediktor (Hosmer dan Lemeshow, 1989). Interpretasi koefisien regresi logistik digunakan untuk menentukan pengaruh variabel respon yang disebabkan oleh perubahan tiap unit variabel prediktor.

Jika variabel prediktor kontinyu, interpretasi koefisien regresi logistik didasarkan pada odds ratio. Odds ratio merupakan rasio peluang suatu kejadian sukses dan peluang suatu kejadian gagal.

$$\psi = \frac{\pi(1)/[1-\pi(1)]}{\pi(0)/[1-\pi(0)]}$$

$$= \frac{\left(\frac{e^{\beta_0+\beta_j}}{1+e^{\beta_0+\beta_j}}\right)\left(\frac{1}{1+e^{\beta_0}}\right)}{\left(\frac{e^{\beta_0}}{1+e^{\beta_0}}\right)\left(\frac{1}{1+e^{\beta_0+\beta_j}}\right)}$$

$$= \frac{e^{\beta_0+\beta_j}}{e^{\beta_0}} = e^{\beta_j} \tag{2.15}$$

Odds ratio dapat bernilai antara 0 dan ~ , sehingga nilai odds ratio selalu positif dan mencerminkan bentuk asosiasi tertentu (Hosmer dan Lemeshow, 1989).

2.1.5. Koefisien Determinasi.

Draper dan Smith (1992), menyatakan bahwa koefisien determinasi merupakan besaran yang mengukur varians pengamatan Y di sekitar nilai tengah yang dapat dijelaskan oleh persamaan regresi. Pada dasarnya nilai koefisien determinasi atau statistik R^2 adalah korelasi antara Y dengan $\hat{Y}$ dan sering juga disebut koefisien korelasi ganda (*multiple correlation coefisient*). Menurut Imam (2004), dalam regresi logistik dikenal adanya *McFadden R-squared* yang merupakan proporsi varians dalam variabel dependen yang dijelaskan oleh varians dalam variabel independen. *McFadden R-squared* dapat dinyatakan sebagai:

$$McFadden\ R^2 = 1- [L_p / L_0] \tag{2.16}$$

Dimana : Lp = log likelihood model

L_0 = log likelihood jika semua slope kecuali konstanta dibatasi sampai nol.

Semakin besar nilai *McFadden* R^2 yang dihasilkan maka semakin baik model yang diperoleh.

2.2. Analisis Komponen Utama (*Principal Component Analysis*)

Metode ini pada dasarnya bertujuan menerangkan struktur varians kovarians melalui kombinasi linier dari variabel-variabel. Meskipun dari p buah variabel asal dapat diturunkan p buah komponen utama untuk menerangkan keragaman total sistem, namun seringkali varians total itu dapat diterangkan secara memuaskan oleh sejumlah kecil komponen utama, misalkan q buah komponen utama (Johnson dan Wichern, 1982).

Beberapa hal berikut dapat digunakan untuk menentukan komponen utama yang akan digunakan, yaitu menerima seluruh komponen utama, menerima komponen utama dengan akar ciri terbesar untuk meminimalkan keragaman variabel prediktor, atau menerima komponen utama yang berkorelasi tinggi dengan variabel respon untuk meminimalkan kuadrat tengah galat.

Pada matriks **X** dengan p variabel kontinyu dan n ukuran sampel di mana elemen dari matriks tersebut dapat dituliskan **X** = x_{ij} (i=1,2,…,n ; j=1,2,…,p), vektor kolom dari matriks tersebut adalah $\mathbf{X}_1,\mathbf{X}_2,\ldots,\mathbf{X}_p$. Matriks varians kovarians dari matriks tersebut dapat dinyatakan:

$$\boldsymbol{S} = \frac{1}{n-1}\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)^T(x_{ij}-\bar{x}_j) \qquad (2.17)$$

di mana: $\bar{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}$ = rata-rata dari sampel.

Sehingga elemen dari matriks varians kovarians tersebut dapat dinyatakan:

$$\boldsymbol{S} = \begin{bmatrix} s_1^2 & s_{12} & \cdots & s_{1p} \\ s_{21} & s_2^2 & \cdots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \cdots & s_p^2 \end{bmatrix} \qquad (2.18)$$

(Escabias dkk., 2005).

Pada analisis komponen utama, vektor variabel asal, yaitu $\mathbf{X} = (\mathbf{X_1,X_2,\ldots,X_p})$ ditransformasi menjadi vektor variabel baru, yaitu $\mathbf{W} = (\mathbf{W_1, W_2,\ldots,W_q})$ di mana $q<p$, dalam bentuk persamaan sebagai berikut:

$$\mathbf{W} = \mathbf{V}\ \mathbf{X} \tag{2.19}$$

Di mana **V** adalah matriks dari vektor ciri dan elemen dari matriks tersebut dapat dituliskan sebagai $\mathbf{V} = v_{jk}$ (j=1,2,…,p ; k=1,2,…,p). Ragam dari **W** dinyatakan sebagai Var(**W**) $=\mathbf{V^TSV}$.

Agar ragam **W** maksimum, digunakan batasan $\mathbf{V^TV} = 1$ dan dengan menggunakan metode pengganda Lagrange didapatkan:

$$U(\mathbf{V},\lambda) = \mathbf{V^TSV} - \lambda(\mathbf{V^TV-1}) \tag{2.20}$$

Fungsi ini akan maksimum jika turunan parsial pertama U ($\mathbf{V},\lambda$) terhadap **V** dan λ sama dengan nol, sehingga:

$$\frac{\partial U(\mathbf{V},\lambda)}{\partial\lambda} = \mathbf{V^TV} - 1=0$$

$$= \mathbf{V^TV} = 1$$

$$\frac{\partial U(\mathbf{V},\lambda)}{\partial\mathbf{V}} = 2\,\mathbf{SV} - 2\lambda\mathbf{V} = 0 \tag{2.21}$$

$$= 2(\mathbf{S} - \lambda\,\mathbf{I})\,\mathbf{V} = 0$$

$$= (\mathbf{S} - \lambda\,\mathbf{I})\,\mathbf{V} = 0 \tag{2.22}$$

Untuk mendapatkan penyelesaian dari persamaan (2.21) yang memenuhi syarat $\mathbf{V^TV} = 1$, maka $\mathbf{S} - \lambda\,\mathbf{I}$ harus bersifat singular atau dengan kata lain harus memenuhi persamaan:

$$\left|\mathbf{S} - \lambda\mathbf{I}\right| = 0 \tag{2.23}$$

Persamaan (2.21) merupakan persamaan karakteristik dari matriks varians kovarians, sehingga diperoleh akar-akar karakteristik $\lambda_1, \lambda_2 \ldots, \lambda_p$ di mana $\lambda_1 \geq \ldots \geq \lambda_p \geq 0$. Kemudian dari setiap akar ciri diperoleh vektor ciri:

$$\mathbf{V} = \begin{bmatrix} v_{11} & v_{12} & \ldots & v_{1k} \\ v_{21} & v_{22} & \ldots & v_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ v_{j1} & v_{j2} & \ldots & v_{jk} \end{bmatrix}$$

V digunakan untuk mentransformasi variabel **X** ke dalam skor komponen utama.

Jika persamaan (2.21) dikalikan dengan $\mathbf{V}^T$ dan $\mathbf{V}^T\mathbf{V} = 1$, maka akan diperoleh hasil:

$$2\,\mathbf{V}^T\mathbf{S}\mathbf{V} - 2\lambda = 0 \longrightarrow \lambda = \mathbf{V}^T\mathbf{S}\mathbf{V}$$

Sehingga ragam dari W didefinisikan sebagai:

$$\mathrm{Var}(\mathbf{W}) = \mathrm{Var}(\mathbf{V}\,\mathbf{X}) = \mathbf{V}^T\mathbf{S}\mathbf{V} = \lambda$$

Keragaman yang dapat dijelaskan oleh komponen utama ke-*i* terhadap keragaman total adalah:

$$\frac{\lambda_i}{\sum_{j=1}^{p}\lambda_i} x100\ \% \tag{2.24}$$

Sedangkan keragaman total yang dapat dijelaskan oleh *q* komponen utama pertama adalah:

$$\frac{\sum_{i=1}^{q}\lambda_i}{\sum_{j=1}^{p}\lambda_i} x100\ \% \tag{2.25}$$

di mana *q* adalah banyaknya komponen utama yang dimasukkan dalam model.

Astutik (2007) menyatakan, apabila komponen utama telah diperoleh, maka tahap selanjutnya adalah menentukan skor komponen utama. Skor komponen utama dari setiap individu akan digunakan untuk analisis lanjut. Jika nilai pengamatan dari individu ke-*i* pada variabel asal ke-*j* adalah X_{i1}, X_{i2},…,X_{ij}, maka skor komponen dari individu ke-*i* untuk komponen utama W_i yang dihasilkan dari matriks varians kovarians adalah :

$$SK_{ij} = \begin{bmatrix} v_{1j}, v_{2j}, \ldots, v_{kj} \end{bmatrix} \begin{bmatrix} X_{i1} - \overline{X}_1 \\ X_{i2} - \overline{X}_2 \\ \ldots \\ X_{ip} - \overline{X}_p \end{bmatrix} \tag{2.26}$$

atau $SK_{ij} = \mathbf{V}_j^T (\mathbf{X}_i - \overline{X})$

Jika komponen utama dihasilkan dari matriks korelasi, maka matriks data individu (**X**) digantikan oleh matriks data skor baku **X***, di mana :

$$x^*_{ij} = \frac{(x_{ij} - \bar{x}_j)}{\sqrt{Var(x_j)}} \tag{2.27}$$

Draper dan Smith (1992) menyatakan, prosedur penentuan jumlah komponen utama yang digunakan dalam membentuk model:

1. Ambil akar ciri yang lebih besar dari 1 (λ_i >1)
2. Pilih q buah komponen utama sebagai penyumbang terbesar terhadap total keragaman lebih besar dari 75%.

2.3. *Principal Component Logistic Regression* (PCLR).

Pada dasarnya PCLR merupakan teknik analisis regresi logistik yang dikombinasikan dengan teknik analisis komponen utama, di mana analisis komponen utama merupakan tahap analisis antara untuk memperoleh hasil akhir dalam analisis regresi logistik. Prinsip utama dari model PCLR ialah skor komponen utama yang diregresikan dengan variabel respon .

Dengan mensubtitusikan sifat-sifat komponen utama, model regresi logistik pada persamaan (2.2) dapat ditulis sebagai berikut:

$$\pi_i = \frac{\exp\left\{\beta_0 + \sum_{j=1}^{p}\sum_{k=1}^{p} z_{ik} v_{jk} \beta_j\right\}}{1 + \exp\left\{\beta_0 + \sum_{j=1}^{p}\sum_{k=1}^{p} z_{ik} v_{jk} \beta_j\right\}} = \frac{\exp\left\{\beta_0 + \sum_{k=1}^{p} z_{ik} \gamma_k\right\}}{1 + \exp\left\{\beta_0 + \sum_{k=1}^{p} z_{ik} \gamma_k\right\}} \tag{2.28}$$

Di mana :

z_{ik} = elemen dari matriks komponen utama.

$\gamma_k = \sum_{j=1}^{p} v_{jk} \beta_j$, *k=1,2,…,p.*

Jika model logit seperti pada persamaan (2.5) disubtitusi dengan komponen utama maka dapat dinyatakan dengan:

$$g(x_i) = \mathbf{X}\beta = \mathbf{Z}\mathbf{V}^T\beta = \mathbf{Z}\gamma$$

di mana **Z** merupakan matriks komponen utama**.**

Apabila matriks **Z** dan **V** dipecah dalam bentuk:

$$\mathbf{Z} = \left(\mathbf{Z}_{(s)} \middle| \mathbf{Z}_{(r)}\right) \text{ dan } \mathbf{V} = \left(\mathbf{V}_{(s)} \middle| \mathbf{V}_{(r)}\right)$$

di mana: *p* = banyak variabel komponen utama

s = banyak variabel komponen utama yang akan dimasukkan dalam model

r = *p-s*

Sehingga $\mathbf{Z}_{(s)} = \mathbf{X}\ \mathbf{V}_{(s)}$ dan $\mathbf{Z}_{(r)} = \mathbf{X}\ \mathbf{V}_{(r)}$. Parameter asli dapat dinyatakan dalam persamaan:

$$\beta = \mathbf{V}\gamma = \mathbf{V}_{(s)}\,\gamma_{(s)} + \mathbf{V}_{(r)}\,\gamma_{(r)}\,,$$

Di mana $\gamma = (\gamma_0\ \gamma_1 \ldots \gamma_s \mid \gamma_{s+1} \ldots \gamma_p) = (\gamma_{(s)} \mid \gamma_{(r)})$.

Prinsip dari PCLR adalah membuang *r* komponen terakhir tersebut, sehingga didapatkan model:

$$Yi = \pi_{i(s)} + \varepsilon_{i(s)}\,,$$

di mana
$$\pi_{i(s)} = \frac{\exp\left(\gamma_0 + \sum_{j=1}^{s} z_{ij}\gamma_j\right)}{1 + \exp\left(\gamma_0 + \sum_{j=1}^{s} z_{ij}\gamma_j\right)}\ ,\ i = 1, 2, \ldots, n. \qquad (2.29)$$

Model di atas dapat diformulasikan kedalam bentuk logit $g(x_i)_{(s)}$ sebagai berikut :

$$g(x_i)_{(s)} = \mathbf{Z}_{(s)}\,\gamma_{(s)} = \mathbf{XV}_{(s)}\,\gamma_{(s)} = \mathbf{X}\,\beta_{(s)}$$ (Escabias dkk., 2005).

2.4. *Principal Component Logistic Regression Stepwise* (PCLR$_{(s)}$).

Metode ini adalah pengembangan dari PCLR. Prosedur awal dari metode ini sama dengan prosedur yang dilakukan pada PCLR, yaitu melakukan tranformasi variabel prediktor asli ke variabel komponen utama. Perbedaan mendasar dari PCLR dan PCLR$_{(s)}$ adalah pada penduga koefisien regresi ($\hat{\gamma}_{(s)}$), di mana 's' variabel komponen utama dari PCLR$_{(S)}$ bukanlah 's' variabel komponen pertama dari PCLR. Sehingga, $\hat{\gamma}_{(s)} = (\hat{\gamma}_{0(s)}, \hat{\gamma}_{1(s)}, \ldots, \hat{\gamma}_{s(s)}) \neq (\hat{\gamma}_0, \hat{\gamma}_1, \ldots, \hat{\gamma}_s)$. Sebagai akibat, peluang $\hat{\pi}_{i(s)}$ diduga dari model PCLR(s) berbeda dengan PCLR (Escabias dkk.,2005), sehingga:

$$\hat{\pi}_{i(s)} = \frac{\exp\left\{\hat{\gamma}_{0(s)} + \sum_{j=1}^{s} z_{ij}\hat{\gamma}_{j(s)}\right\}}{1 + \exp\left\{\hat{\gamma}_{0(s)} + \sum_{j=1}^{s} z_{ij}\hat{\gamma}_{j(s)}\right\}} \neq \frac{\exp\left\{\hat{\gamma}_0 + \sum_{j=1}^{s} z_{ij}\hat{\gamma}_j\right\}}{1 + \exp\left\{\hat{\gamma}_0 + \sum_{j=1}^{s} z_{ij}\hat{\gamma}_j\right\}} \qquad (2.30)$$

Dalam pemilihan 's' komponen utama digunakan metode *stepwise*, yaitu dengan memasukkan komponen utama secara bertahap kemudian menyeleksi komponen tersebut secara bertahap pula.

Keputusan apakah variabel tersebut dipertahankan atau dibuang tergantung pada uji nisbah kemungkinan (*Likelihood Ratio Test*). Pengujian ini dilakukan pada tiap tahap prosedur *stepwise*. Variabel prediktor yang sebelumnya pernah menjadi calon terbaik dan dimasukkan ke dalam persamaan regresi, pada tahap berikutnya mungkin dianggap berlebihan karena hubungannya dengan variabel prediktor lain yang sekarang ada di dalam model (Draper dan Smith, 1992).

Adapun langkah-langkah penentuan model terbaik dengan metode *stepwise* adalah :

1. Menentukan P_E dan P_R untuk menyeleksi variabel-variabel yang masuk ke dalam model maupun yang keluar, di mana P_E adalah taraf nyata untuk mengontrol apakah suatu variabel prediktor dimasukkan ke dalam model, sedangkan P_R adalah taraf nyata untuk mengontrol apakah suatu variabel prediktor tetap berada dalam model.
2. langkah (0) menganggap model memiliki *p* variabel prediktor yang semuanya dianggap penting sebagai variabel yang berpengaruh dalam model.
   a. langkah ini dimulai dengan membetuk model yang hanya mengandung intersep saja kemudian menghitung nilai log-likelihoodnya (*Lo*).
   b. Kemudian hitung nilai log-likelihood untuk model yang mengandung variabel W_j dan dituliskan sebagai $L_j^{(0)}$, di mana *j* adalah indeks untuk variabel yang ditambahkan ke dalam model dan (0) adalah indeks untuk langkah.
   c. Hitung nilai nisbah kemungkinan untuk model yang mengandung variabel W_j melawan model yang hanya mengandung intersep saja, $G_j^{(0)} = -2(L_0 - L_j^{(0)})$ dan nilai *p* yaitu $P_j^{(0)}$, $P_j^{(0)} = P(\chi^2_{(v)} > G_j^{(0)})$.
   d. Bandingkan nilai $P_j^{(0)}$, variabel yang paling penting adalah variabel yang memiliki nilai $P_j^{(0)}$ terkecil. Jika variabel tersebut dituliskan sebagai W_{e1} maka $P_{e1}^{(0)} = \min(P_j^{(0)})$. Indeks "ei" digunakan untuk menotasikan variabel yang menjadi calon variabel yang masuk pada langkah (1).

Variabel tersebut cukup penting untuk masuk ke dalam model jika $P_{e1}^{(0)} < P_E$, jika tidak maka proses berhenti.

3. Langkah (1), dimulai dengan model regresi logistik yang mengandung W_{e1}
   a. Hitung $L_{e1}^{(1)}$, yaitu log likelihood model ini.
   b. Tentukan apakah sisa *p*-1 variabel adalah penting sekali untuk model ketika variabel W_{e1} berada dalam model, dengan membuat *p*-1 model regresi logistik yang mengandung X_{e1} dan X_j, $j = 1,2,3,\ldots,p-1$ dan $j \neq e1$. Hitung nilai log-likelihood dan dituliskan sebagai $L_{e1j}^{(1)}$.
   c. Hitung nilai nisbah kemungkinan ini melawan model yang hanya mengandung W_{e1} ($G_j^{(1)} = -2(L_{e1}^{(1)} - L_{e1j}^{(1)})$) , hitung nilai *p* statistik ini sebagai $P_j^{(1)}$.
   d. Misalkan variabel dengan nilai $P_j^{(1)}$ terkecil pada langkah (1) W_{e2} di mana $P_{e2}^{(1)} = \min(P_J^1)$. Jika $P_{e2}^{(1)} < P_E$ lanjutkan langkah (2), jika tidak maka proses berhenti.
4. Langkah (2), dimulai dengan menduga model regresi logistik yang mengandung W_{e1} dan W_{e2}. Ini mungkin sekali W_{e1} tidak penting untuk model ketika W_{e2} berada dalam model. Secara umum langkah ini adalah memilih model yang sesuai dengan menghapus variabel yang ditambahkan pada tahap sebelumnya dan menetapkan variabel penting selanjutnya yang dimasukkan.
   - Seleksi langkah mundur.
     a. Hitung $L_{-ej}^{(2)}$ yaitu log-likelihood model dengan W_{ej} yang dibuang.
     b. Hitung nilai nisbah kemungkinan model ini melawan model penuh pada langkah (2) yaitu $G_{-ej}^{2} = -2(L_{ele2}^{(2)} - L_{-ej}^{(2)})$ dan $P_{-ej}^{(2)}$ sebagai nilai *p*.
     c. Untuk memastikan apakah suatu peubah harus dihapus dari model, pilih peubah yang menghasilkan $P_{-ej}^{(2)}$ maksimum, yang dituliskan sebagai W_{r2} , $P_{r2}^{(2)} = \text{maks}(P_{-e1}^{(2)}, P_{-e2}^{(2)})$.

d. Untuk memutuskan apakah X_{r2} harus dibuang, bandingkan $P_{r2}^{(2)}$ dengan P_R. Jika $P_{r2}^{(2)} > P_R$ maka X_{r2} keluar dari model. Jika $P_{r2}^{(2)} < P_R$ maka X_{r2} tetap berada dalam model.

- Seleksi langkah maju dengan cara :
  a. Pada tahap seleksi selanjutnya, *p*-2 model regresi logistik mengandung W_{e1}, W_{e2} dan W_{ej}, untuk j = 1,2,3,…,*p*, $j \neq$ e1, e2. Hitung nilai nisbah kemungkinan model ini melawan model yang hanya mengandung W_{e1} dan W_{e2} dan tentukan nilai $P_j^{(2)}$.
  b. Dapatkan W_{ej} variabel yang memiliki nilai $P_j^{(2)}$ minimum, $P_{e3}^{(2)} = \min(P_j^{(2)})$. Jika $P_{e3}^{(2)} < P_E$ lanjutkan ke langkah (3), jika tidak proses berhenti.

5. Langkah (3), prosedur pada langkah ini sama dengan prosedur langkah (2). Langkah ini berlanjut sampai langkah terakhir atau langkah (*S*).
6. Langkah (*S*), ini terjadi ketika : (1) semua variabel bebas masuk dalam model atau (2) semua variabel yang berada dalam model mempunyai nilai *p* untuk dipindahkan kurang dari P_R, dan variabel yang belum masuk dalam model mempunyai nilai *p* untuk masuk dalam model melebihi P_E.

2.5. Multikolinieritas.

Istilah multikolinieritas digunakan untuk menunjukkan hubungan linier antara beberapa variabel prediktor dalam model regresi. Penggunaan kata multikolinieritas dimaksudkan untuk menunjukkan adanya derajat kolinieritas yang tinggi di antara variabel-variabel prediktor. Bila variabel-variabel prediktor berkorelasi secara sempurna maka metode kuadrat terkecil tidak dapat digunakan (Sumodiningrat, 1999).

Hosmer dan Lemeshow (1989) mengatakan bahwa seperti pada regresi linier, pada model regresi logistik juga sensitif dengan adanya kolinieritas. Kolinieritas mengindikasikan bahwa terdapat hubungan ketergantungan antar dua variabel prediktor.

2.5.1. Pendeteksian multikolinieritas.

VIF (*Variansce Inflation Factor*) merupakan salah satu indikator untuk mengukur besar kolinieritas. VIF merupakan peningkatan ragam dari koefisien regresi yang disebabkan karena adanya ketergantungan linier suatu variabel prediktor dengan variabel prediktor lainnya. Nilai VIF dapat dicari dengan menggunakan rumus:

$$\mathrm{VIF} = \frac{1}{1 - R_j^{\,2}} \tag{2.31}$$

di mana $j = 1,2,..,p$.

Nilai R_j^2 adalah koefisien determinasi yang dihasilkan dengan meregresikan variabel prediktor X_j melawan variabel prediktor lain (X_j', di mana $j \neq j'$) (Hines dan Montgomery, 1990)

Jika R_j^2 mendekati 1, maka VIF akan bertambah besar. Hal ini akan terlihat jika suatu variabel prediktor mempunyai hubungan linier yang kuat dengan variabel prediktor lain. VIF menunjukkan pendekatan yang lebih akurat untuk mendeteksi adanya multikolinieritas daripada hanya dengan melihat korelasi antara dua variabel prediktor (Myers, 1990).

Nilai VIF akan bertambah besar jika terdapat korelasi yang semakin besar diantara variabel prediktor. Jika nilai VIF lebih besar daripada 10 (berarti R_j^2 lebih besar dari 0.9), maka diindikasikan terjadi multikolinieritas (Bowerman dan O'Connell, 1990).

2.5.2. Konsekuensi multikolinieritas.

Menurut Gujarati (1991), simpangan baku bagi koefisien regresi yang diduga akan menjadi lebih besar jika terjadi multikolinieritas. Hal ini dapat dilihat dari matriks $(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}$ yang dibutuhkan dalam menentukan *standart error* bagi penduga koefisien regresi melalui persamaan:

$$Se(\hat{\beta}_i) = \sqrt{C_{ij} S^2} = \sqrt{C_{ij} \frac{JKG}{n-k}} \tag{2.32}$$

di mana : Cij = unsur diagonal dari matriks $(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}$

JKG = jumlah kuadrat galat dari uji koefisien regresi secara serentak.

Multikolinieritas menyebabkan determinan matriks $(\mathbf{X}^{\mathbf{T}}\mathbf{X})$ mendekati singular ($\left|\mathbf{X}^{\mathbf{T}}\mathbf{X}\right| \longrightarrow 0$), sehingga nilai dari matriks $(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}$ yang diperoleh melalui persamaan

$$(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1} = \frac{1}{\left|\mathbf{X}^{\mathbf{T}}\mathbf{X}\right|} \, adj \left(\mathbf{X}^{\mathbf{T}}\mathbf{X}\right) \tag{2.33}$$

akan menjadi besar. Dengan demikian semakin kuat hubungan linier antara **X**, akan menyebabkan nilai $Se(\hat{\beta}_i)$ semakin besar yang berpengaruh pula terhadap nilai statistik uji *t* yang semakin kecil sehingga memungkinkan penduga koefisien regresi menjadi tidak signifikan terhadap model. Adanya multikolinieritas juga berakibat pada semakin lebarnya selang kepercayaan bagi $\hat{\beta}_i$.

Multikolinieritas juga menyebabkan pendugaan koefisien regresi menjadi tidak stabil. Hal ini disebabkan adanya simpangan baku yang besar, sehingga keakuratan pendugaan koefisien regresi akan sulit diperoleh. Sebagai contoh:

$$\mathbf{X}_{1i} = C\,\mathbf{X}_{2i},\ C \neq 0 \tag{2.34}$$

yang akan diterapkan dalam matriks ($\mathbf{X}^T\mathbf{X}$):

$$\left(\mathbf{X}^T\mathbf{X}\right) = \begin{bmatrix} n & \sum x_{1i} & \sum x_{2i} & \dots & \sum x_{ki} \\ \sum x_{1i} & \sum x_{1i}^2 & \sum x_{1i}x_{2i} & \dots & \sum x_{1i}x_{ki} \\ \sum x_{2i} & \sum x_{2i}x_{1i} & \sum x_{2i}^2 & \dots & \sum x_{2i}x_{ki} \\ \dots & \dots & \dots & \dots & \dots \\ \sum x_{ki} & \sum x_{ki}x_{1i} & \sum x_{ki}x_{2i} & \dots & \sum x_{ki}^2 \end{bmatrix} \tag{2.35}$$

Disubtitusikan persamaan (2.34) ke persamaan (2.35) diperoleh:

$$\left(\mathbf{X}^T\mathbf{X}\right) = \begin{bmatrix} n & r\sum x_{1i} & \sum x_{2i} & \dots & \sum x_{ki} \\ C\sum x_{2i} & C^2\sum x_{2i}^2 & C\sum x_{2i}^2 & \dots & C\sum x_{2i}x_{ki} \\ \sum x_{2i} & C\sum x_{1i}^2 & \sum x_{2i}^2 & \dots & \sum x_{2i}x_{ki} \\ \dots & \dots & \dots & \dots & \dots \\ \sum x_{ki} & C\sum x_{ki}x_{2i} & \sum x_{ki}x_{2i} & \dots & \sum x_{ki}^2 \end{bmatrix} \tag{2.36}$$

Berdasarkan persamaan (2.36) dapat dilihat bahwa baris kedua merupakan kombinasi linier dari baris ketiga. Berdasarkan sifat determinan matriks, jika suatu baris (kolom) dalam matriks adalah kombinasi linier dari baris (kolom) yang lain, maka determinan dari matriks tersebut bernilai nol. Dengan demikian matriks ($\mathbf{X}^T\mathbf{X}$) merupakan matriks singular, sehingga tidak memiliki matriks kebalikan (invers) yang khas dan unik. Akibatnya, penduga β tidak dapat ditentukan secara pasti, dengan kata lain tidak memiliki penyelesaian tunggal.

2.6. Uji kesesuaian model (*Goodness of fit*)

Untuk menguji apakah model yang dihasilkan sudah sesuai atau tidak, maka perlu dilakukan pengujian kesesuaian model atau *goodness of fit*. Hipotesis yang digunakan dalam pengujian ini yaitu:

Ho : Model sesuai lawan

H_1 : Model tidak sesuai

Adapun statistik uji yang digunakan untuk menguji kesesuaian model adalah uji *Pearson* dan uji *Deviance*.

1. Statistik *Pearson*

$$\chi^2 Pearson = \sum_{a=1}^{h} \chi_p^2 (y_a, \hat{\pi}_a) \tag{2.37}$$

di mana $\chi^2{}_p (y_a, \hat{\pi}_a) = \sum_{b=1}^{g} \frac{(y_{ab} - n_a \hat{\pi}_{ab})^2}{n_a \hat{\pi}_{ab}}$

2. Statistik *Deviance*.

$$D = 2\sum_{a=1}^{h} \chi^2{}_D (y_a, \hat{\pi}_a) \tag{2.38}$$

di mana $\chi^2{}_D (y_a, \hat{\pi}_a) = \sum_{b=1}^{g} y_{ab} \ln\left[\frac{y_{ab}}{n_a \hat{\pi}_{ab}}\right]$

Dengan : $\hat{\pi}_{ab}$ = peluang Y pada kategori ke-b dan X ke-*a*.

y_{ab} = pengamatan Y pada kategori ke-b dan X ke-*a*.

p = banyaknya parameter pada model.

h = banyaknya variabel prediktor X.

g = banyaknya kategori variabel respon Y.

Statistik *Pearson* dan *deviance* menyebar mengikuti sebaran khi kuadrat dengan derajat bebas *v*. Keputusan menolak Ho jika $\chi^2{}_{pearson} > \chi^2{}_{(v)}$ dan D > $\chi^2{}_{(v)}$ atau $P\left[\chi^2{}_{(v)} > \chi^2{}_{pearson}\right]$ dan $P\left[\chi^2{}_{(v)} > D\right]$ lebih kecil dari peluang yang diinginkan (α), sehingga dapat diambil kesimpulan bahwa model yang diperoleh tidak sesuai (Fahrmeir dan Gerhard, 1994).

2.7. Kriteria Pemilihan Variabel.

2.7.1. *Pearson Residuals*.

Menurut Kutner (2004), kriteria yang dapat digunakan untuk mengukur keakuratan model regresi adalah *Pearson Residuals*, di mana persamaannya dapat dinyatakan sebagai berikut:

$$\chi^2 = \sum_{i=1}^{n} \frac{\left(Y_i - \hat{\pi}_i\right)^2}{\hat{\pi}_i\left(1-\hat{\pi}_i\right)} \tag{2.39}$$

di mana : Y_i = nilai pengamatan ke-i

$\hat{\pi}_i$ = nilai duga pengamatan ke-i (i = 1,2,…,n)

Semakin kecil nilai *Pearson Residuals* yang dihasilkan maka semakin baik model yang diperoleh.

2.7.2. *Percent Correct Predictions* (PCP).

Menurut Imam (2004), hal penting untuk menilai suatu prosedur klasifikasi adalah dengan menghitung peluang kejadian yang tepat diklasifikasikan . Ukuran ini dinamakan *Percent Correct Predictions* (PCP). PCP dihitung dengan membuat tabel klasifikasi *actual* dan *predicted group* seperti pada tabel 2.1.

Tabel 2.1. Klasifikasi *actual* dan *predicted group*.

Actual Group	*Predicted Group*		Jumlah
	π_1	π_2	
π_1	n_{11}	$n_{12}= n_{1.}- n_{11}$	$n_{1.}$
π_2	$n_{21}= n_{2.}-n_{22}$	n_{22}	$n_{2.}$
Jumlah	$n_{.1}$	$n_{.2}$	n

n_{11} = banyaknya pengamatan dari π_1 tepat diklasifikasi sebagai π_1

n_{12} = banyaknya pengamatan dari π_1 salah diklasifikasi sebagai π_2

n_{22} = banyaknya pengamatan dari π_2 tepat diklasifikasi sebagai π_2

n_{21} = banyaknya pengamatan dari π_2 salah diklasifikasi sebagai π_1

$$\text{PCP} = \left[1 - \frac{n_{12}+n_{21}}{n_1+n_2}\right] x100\% \tag{2.40}$$

Semakin besar nilai PCP yang dihasilkan maka semakin baik model yang diperoleh.

BAB III
METODE PENELITIAN

3.1. Data.

Data yang digunakan dalam penelitian ini merupakan data sekunder yang terdiri dari dua data seperti disajikan pada Tabel 3.1. Data selengkapnya dapat dilihat pada Lampiran 1 dan 2.

Tabel 3.1. Data penelitian

Data	Sumber	Penelitian	Variabel Respon	Variabel Prediktor
1	Dixon dan Massey (1983)	Faktor-faktor yang mempengaruhi kinerja jantung 10 tahun setelah pemeriksaan terakhir	Y = 0, jika terjadi serangan jantung Y = 1, jika tidak terjadi serangan jantung	X_1 = umur (tahun)
				X_2 = tekanan darah sistolik (mm merkuri)
				X_3 = tekanan darah diastolik (mm merkuri)
				X_4 = kolesterol (mg/dl)
				X_5 = tinggi (inchi)
				X_6 = berat (berat)
2	Dwi Supriyono (2008)	Faktor-faktor yang mempengaruhi fertilitas kambing lokal (PE) di kecamatan Parang Kabupaten Magetan.	Y = 0, jika terjadi kehamilan Y = 1, jika tidak terjadi kehamilan	X_1 = berat badan (kg)
				X_2 = tinggi badan (cm)
				X_3 = panjang badan (cm)
				X_4 = lingkar dada (cm)
				X_5 = lebar pinggul (cm)
				X_6 = volume tubuh (cm3)

3.2. Metode Penelitian.

3.2.1. Metode Pembentukan Model Regresi Logistik dengan PCLR:

1. Pendeteksian multikolinieritas dengan persamaan (2.31).
2. Menghitung $\mathbf{X}_1$*, $\mathbf{X}_2$*,…,$\mathbf{X}_p$* dengan cara membakukan $\mathbf{X}_1$, $\mathbf{X}_2$,…,$\mathbf{X}_p$ dengan persamaan (2.27)
3. Menghitung matriks varians kovarians **S** dengan persamaan (2.17).
4. Menghitung akar ciri (nilai eigen) dengan persamaan (2.22).
5. Menghitung vektor ciri (vektor eigen) dengan persamaan (2.23)
6. Menghitung skor variabel komponen utama dengan persamaan (2.26).
7. Memilih *q* buah komponen utama .
8. Membentuk model persamaan regresi logistik antara Y dengan *q* komponen utama.
9. Melakukan pengujian kesesuaian model dengan statistik uji *Pearson* dan *Deviance* sesuai dengan persamaan (2.37) dan (2.38).
10. Mendapatkan nilai *Pearson Residuals* dan *Percent Correct Predictions* (PCP) dengan persamaan (2.39) dan (2.40), kemudian dibandingkan dengan nilai yang didapatkan dari metode $\text{PCLR}_{(S)}$.

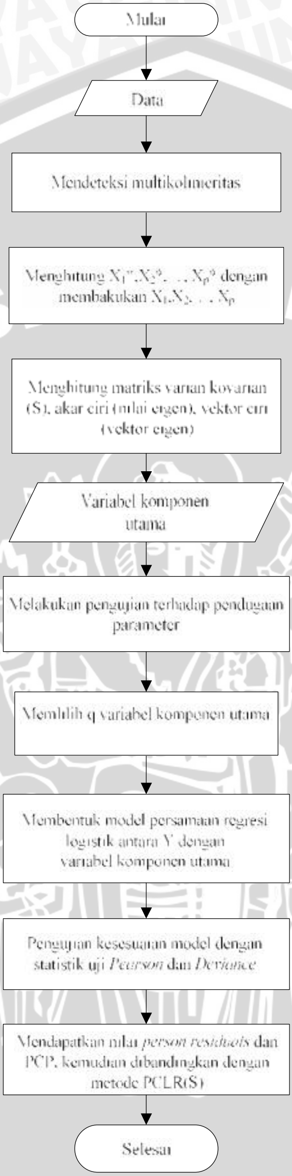
Prosedur pembentukan model regresi logistik dengan PCLR dapat dilihat pada Gambar 3.1.

3.2.2 Metode Pembentukan Model Regresi Logistik dengan $\textbf{PCLR}_{(S)}$:

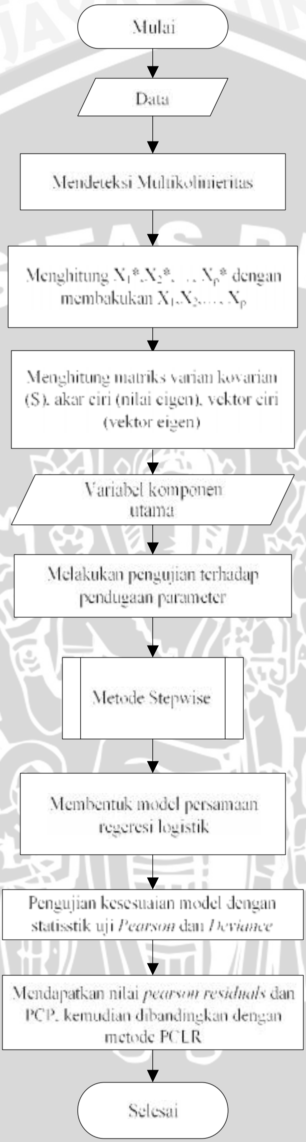
1. Pendeteksian multikolinieritas dengan persamaan (2.31).
2. Menghitung $\mathbf{X}_1$*, $\mathbf{X}_2$*,…,$\mathbf{X}_p$* dengan cara membakukan $\mathbf{X}_1$, $\mathbf{X}_2$,…,$\mathbf{X}_p$ dengan persamaan (2.27)
3. Menghitung matriks varians kovarians **S** dengan persamaan (2.17).
4. Menghitung akar ciri (nilai eigen) dengan persamaan (2.22).
5. Menghitung vektor ciri (vektor eigen) dengan persamaan (2.23)
6. Menghitung skor variabel komponen utama dengan persamaan (2.26).

7. Menggunakan metode *stepwise* untuk memilih komponen utama yang masuk dalam model dan tidak saling berkorelasi. Langkah-langkah yang ditempuh terdapat pada Gambar 3.3.
8. Membentuk model persamaan regresi logistik dengan variabel respon Y dan variabel prediktor $W_{(1)}$, $W_{(2)}$,…, $W_{(s)}$ yang didapat dari metode *stepwise*.
9. Melakukan pengujian kesesuaian model dengan statistik uji *Pearson* dan *Deviance* sesuai dengan persamaan (2.37) dan (2.38).
10. Mendapatkan nilai *Pearson Residuals* dan *Percent Correct Predictions* (PCP) dengan persamaan (2.39) dan (2.40), kemudian dibandingkan dengan nilai yang didapatkan dari metode $PCLR_{(S)}$.

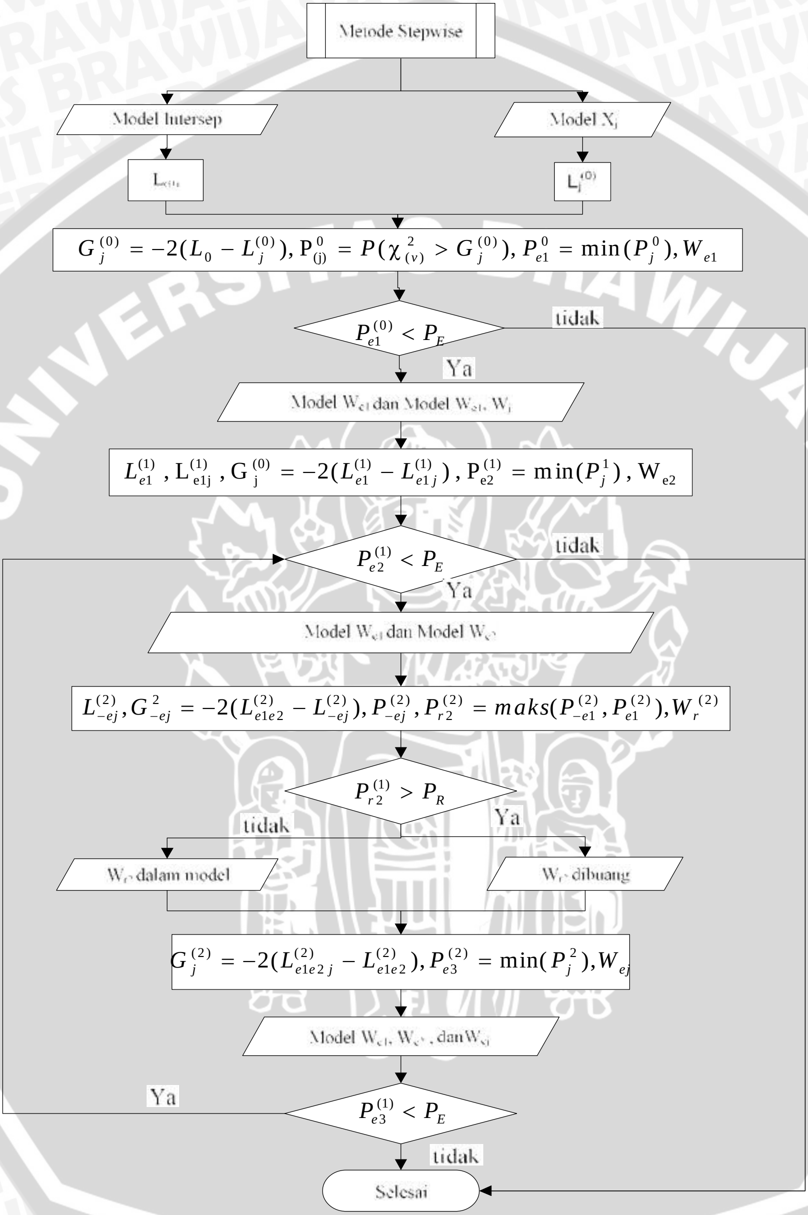
Prosedur pembentukan model regresi logistik dengan $PCLR_{(S)}$ dapat dilihat pada Gambar 3.2.



Gambar 3.1 Prosedur pembentukan model regresi logistik dengan PCLR



Gambar 3.2. Prosedur pembentukan model regresi logistik dengan $PCLR_{(S)}$



Gambar 3.3 Diagram alir metode *stepwise*.

BAB IV
HASIL DAN PEMBAHASAN.

4.1 Hasil.

4.1.1. Pendeteksian Multikolinieritas.

Seperti pada regresi linier, model regresi logistik juga sensitif terhadap adanya multikolinieritas. Oleh karena itu dilakukan pendeteksian awal dengan menghitung koefisien korelasi antar variabel prediktor sebelum variabel-variabel tersebut dianalisis dengan regresi logistik. Adanya multikolinieritas dapat dideteksi dengan menggunakan nilai VIF. Nilai VIF dari data yang digunakan dalam penelitian ini disajikan dalam Tabel 4.1, sedangkan perhitungan nilai VIF selengkapnya dapat dilihat pada Lampiran 3.

Tabel 4.1. Nilai VIF data 1 dan 2.

Data	Variabel Prediktor					
	X_1	X_2	X_3	X_4	X_5	X_6
1	2.079	16.393	18.868	1.923	2.320	2.0161
2	15.625	12.195	20.408	11.236	11.364	24.390

Pada data 1 terdapat dua variabel prediktor yang mempunyai nilai VIF lebih dari 10 yaitu pada X_2 dan X_3. Sedangkan pada data 2, semua variabel prediktor memiliki nilai VIF lebih besar dari 10. Dengan melihat nilai VIF dari kedua data yang disajikan pada Tabel 4.1. diatas, maka dapat disimpulkan bahwa data 1 mengandung multikolinieritas pada sebagian variabel prediktor, sedangkan data 2 mengandung multikolinieritas pada semua variabel prediktor.

4.1.2. Analisis Komponen Utama (AKU).

Data 1 dan 2 memiliki variabel prediktor dengan satuan pengukuran berbeda, maka variabel prediktor terlebih dulu dibakukan. Untuk proses selanjutnya, matriks masukan yang digunakan berupa matriks varians kovarians. Variabel prediktor data 1 dan 2 yang telah dibakukan dapat dilihat pada Lampiran 3 dan 4.

a. Data 1.

Konstanta transformasi AKU data 1 dapat dilihat pada Tabel 4.2, sedangkan hasil transformasi AKU dapat dilihat pada Lampiran 6. Pada Tabel 4.3 ditampilkan keragaman yang dapat dijelaskan oleh masing-masing komponen utama .

Tabel 4.2. Konstanta transformasi AKU data 1.

Variabel	W_1	W_2	W_3	W_4	W_5	W_6
X_1	-0.46044	-0.20711	0.29631	-0.71553	0.37597	-0.06296
X_2	-0.55411	0.10081	-0.33971	0.25003	0.23292	0.67129
X_3	-0.55325	0.17207	-0.34051	0.20003	-0.05277	-0.71103
X_4	-0.34285	-0.19971	0.75265	0.49088	-0.18717	0.00999
X_5	0.16082	0.65701	0.29446	0.17557	0.64328	-0.10550
X_6	-0.17739	0.66767	0.16633	-0.33720	-0.59394	0.16916

Tabel 4.3. Nilai keragaman dan total keragaman AKU data 1

Variabel	W_1	W_2	W_3	W_4	W_5	W_6
Keragaman (%)	0.400	0.247	0.158	0.107	0.081	0.006
Total keragaman (%)	0.400	0.648	0.806	0.913	0.994	1.000

Terlihat pada Tabel 4.2. dan 4.3. bahwa komponen utama pertama, kedua dan ketiga menjelaskan keragaman lebih dari 75%, yaitu sebesar 80.6% total keragaman data asli. Selanjutnya tiga skor komponen utama tersebut digunakan dalam analisis regresi logistik dengan metode PCLR, sedangkan metode $PCLR_{(S)}$ menggunakan variabel komponen utama dengan menyeleksinya secara bertahap.

b. Data 2.

Tabel 4.4. dan 4.5. menunjukkan konstanta transformasi, keragaman dan total keragaman data 2, sedangkan hasil transformasi AKU data 2 secara lengkap dapat dilihat pada Lampiran 7.

Tabel 4.4. Konstanta transformasi AKU data 2.

Variabel	W_1	W_2	W_3	W_4	W_5	W_6
X_1	-0.45776	-0.13901	-0.14332	0.54733	-0.62872	-0.23607
X_2	-0.31450	-0.20249	-0.75069	-0.53466	0.01767	-0.10186
X_3	-0.47390	-0.23302	-0.07163	0.40093	0.59201	0.45251
X_4	-0.36715	-0.39332	0.54783	-0.47339	-0.29943	0.31086
X_5	-0.32521	0.81952	-0.06404	-0.11414	-0.18464	0.41401
X_6	-0.47583	0.24302	0.32645	-0.12920	0.36079	-0.67904

Tabel 4.5. Nilai keragaman dan total keragaman AKU data 2

Variabel	W_1	W_2	W_3	W_4	W_5	W_6
Keragaman (%)	0.536	0.168	0.135	0.104	0.050	0.007
Total keragaman (%)	0.536	0.704	0.839	0.943	0.993	1.000

Dari hasil transformasi AKU data 2 didapatkan tiga komponen utama pertama yang menjelaskan lebih dari 75% yaitu sebesar 83.9% total keragaman data asli. Sama seperti data 1, tiga komponen tersebut digunakan untuk mendapatkan model dengan metode PCLR.

4.1.3. Pengujian Koefisien Regresi Logistik Secara Parsial.

Pengujian terhadap koefisien regresi logistik secara parsial digunakan untuk memeriksa apakah koefisien regresi setiap variabel prediktor secara individu harus berada dalam model. Dalam tahap ini variabel yang digunakan sebagai variabel prediktor adalah variabel komponen utama. Pengujian secara parsial ini dilakukan dengan uji Wald yang dilandasi pada hipotesis:

Ho : $\beta_j = 0$ lawan

H_1 : $\beta_j \neq 0$; j = 1,2,…,6

Hasil pengujian koefisien regresi logistik secara parsial untuk data 1 dan 2 disajikan pada Tabel 4.6. dan 4.7.

Tabel 4.6. Pengujian Koefisen Regresi Logistik Data 1 Secara Parsial.

Variabel Prediktor	Koefisien	z	P
W_1	-1.06343	-6.20	0.000
W_2	0.03098	-1.99	0.044
W_3	0.09347	-2.03	0.023
W_4	-0.02474	-4.01	0.000
W_5	0.36061	-5.81	0.000
W_6	0.67851	-5.21	0.000

Hasil uji Wald yang disajikan Tabel 4.6. menunjukkan bahwa semua variabel komponen utama berpengaruh nyata terhadap serangan jantung.

Tabel 4.7. Pengujian Koefisen Regresi Logistik Data 2 Secara Parsial.

Variabel Prediktor	Koefisien	z	P
W_1	0.19388	1.98	0.048
W_2	0.07114	2.01	0.044
W_3	-0.04895	-0.19	0.850
W_4	-0.21364	-0.71	0.478
W_5	-0.02547	2.03	0.042
W_6	-1.05349	-2.09	0.036

Berdasarkan uji Wald data 2 pada Tabel 4.7. dapat dilihat bahwa variabel prediktor W_3 dan W4 tidak berpengaruh nyata pada model, sehingga apabila dalam pemilihan model terbaik melibatkan variabel W_3 dan W_4, maka diperlukan pengujian koefisien secara simultan dengan uji nisbah kemungkinan.

4.1.4. *Principal Component Logistic Regression* (PCLR)

Tahap selanjutnya adalah meregresikan komponen-komponen utama tersebut terhadap variabel respon, kemudian melakukan pengujian kesesuaian model dengan statistik uji *Pearson* dan *Deviance*.

a. Data 1.

Hasil pendugaan koefisien regresi logistik dengan variabel prediktor W_1, W_2, dan W_3, pada data 1 dapat dilihat pada Lampiran 7. Setelah diperoleh koefisien model regresi logistik maka didapat model:

$$\pi_i = \frac{\exp(-1.53325 - 1.071341\,W_1 + 0.067109\,W_2 + 0.181730\,W_3)}{1 + \exp(-1.53325 - 1.071341\,W_1 + 0.067109\,W_2 + 0.181730\,W_3)}$$

Uji kesesuaian model dilakukan dengan hipotesis:

Ho : Model sesuai lawan

H_1 : Model tidak sesuai

Hasil uji kesesuaian model data 1 disajikan dalam Tabel 4.8.

Tabel 4.8. Hasil uji kesesuaian model data 1 (metode PCLR).

Metode	χ^2_{hitung}	$\chi^2_{(v)}$	*P*	Keterangan
Pearson	196.033	229.65	0.486	Terima Ho
Deviance	158.201	229.65	0.978	Terima Ho

Berdasarkan uji kesesesuaian model pada tabel 4.8 di atas, maka model dengan variabel prediktor W_1, W_2, dan W_3 layak digunakan.

b. Data 2

Pembentukan model regresi logistik data 2 dengan variabel prediktor W_1, W_2, dan W_3 disajikan pada Lampiran 8. Model regresi logistik data 2 dengan variabel prediktor W_1, W_2, dan W_3 adalah:

$$\pi_i = \frac{\exp(1.23935 + 0.192380\,W_1 + 0.0677836\,W_2 - 0.0338814\,W_3)}{1 + \exp(1.23935 + 0.192380\,W_1 + 0.0677836\,W_2 - 0.0338814\,W_3)}$$

Uji kesesuaian model dilakukan dengan hipotesis:

Ho : Model sesuai lawan

H_1 : Model tidak sesuai

Hasil uji kesesuaian model data 2 disajikan dalam Tabel 4.9.

Tabel 4.9. Hasil uji kesesuaian model data 2 (Metode PCLR).

Metode	χ^2_{hitung}	$\chi^2_{(v)}$	*P*	Keterangan
Pearson	100.184	119.82	0.365	Terima Ho
Deviance	100.634	119.82	0.235	Terima Ho

Berdasarkan uji kesesesuaian model pada Tabel 4.9 di atas, model dengan variabel prediktor W_1, W_2, dan W_3 layak digunakan.

4.1.5. *Principal Component Logistic Regression Stepwise* (PCLR$_{(S)}$)

Untuk memilih variabel-variabel prediktor yang masuk dalam model, maka digunakan metode *stepwise*. Metode ini dimulai dengan menetukan nilai taraf nyata untuk menyeleksi variabel-variabel yang masuk (P_E) maupun yang keluar (P_R) dari dalam model pada setiap langkah pemilihan variabel. Dalam penelitian ini nilai P_E yang digunakan adalah 0.15, sedangkan P_R menggunakan nilai 0.20. Pemilihan nilai taraf nyata pada selang 0.15 sampai 0.20 lebih direkomendasikan, karena menggunakan taraf ini akan memberikan jaminan dengan metode *stepwise* akan didapatkan variabel-variabel yang memiliki nilai koefisien regresi berbeda dengan nol (Hosmer dan Lemeshow, 1989).

a. Data 1.

Adapun langkah-langkah untuk mendapatkan variabel-variabel yang layak dalam model dimulai dari langkah (0) yaitu model yang hanya mengandung intersep saja. Langkah selanjutnya dipilih variabel yang layak masuk maupun dibuang dari model dan proses ini akan berhenti apabila sudah tidak ada lagi variabel yang dapat masuk dalam model.

Langkah (0) dimulai dengan menghitung log-likelihood model dengan intersep dan model dengan W_1, W_2, W_3, W_4, W_5, dan W_6.

Tabel 4.10 Nilai log-likelihood, *Likelihood Ratio Test* dan $P_j^{(0)}$ langkah (0) data 1.

Variabel	Log-likelihood	G	$P_j^{(0)}$
C	-112.467		
C, W_1	**-79.560**	**65.814**	**0.000**
C, W_2	-112.440	0.053	0.817
C, W_3	-112.321	0.292	0.589
C, W_4	-112.461	0.012	0.911
C, W_5	-111.302	2.330	0.127
C, W_6	-110.844	3.245	0.072

Pada langkah (0) dipilih variabel prediktor sebagai calon peubah yang akan dimasukkan pada langkah (1) yaitu variabel dengan nilai $P_j^{(0)}$ terkecil pada Tabel 4.10, yaitu variabel W_1. Karena nilai $P_j^{(0)}<0.15$, maka proses dilanjutkan ke langkah (1).

Langkah (1) dimulai dengan menghitung nilai log-likelihood model yang mengandung intersep dan W_1 dengan W_2, W_3, W_4, W_5, dan W_6.

Tabel 4.11. Nilai log-likelihood, *Likelihood Ratio Test* dan $P_j^{(1)}$ langkah (1) data 1.

Variabel	Log-likelihood	G	$P_j^{(1)}$
C, W_1	-79.560		
C, W_1, W_2	-79.457	0.206	0.65433
C, W_1, W_3	-79.187	0.742	0.39071
C, W_1, W_4	-79.517	0.086	0.77398
C, W_1, W_5	-78.543	2.034	0.15442
C, W_1, W_6	**-78.039**	**3.042**	**0.08520**

Variabel pada langkah (1) dengan nilai $P_j^{(1)}$ terkecil (0.0852) sebagai calon variabel yang masuk pada model adalah variabel W_6. Karena nilai $P_j^{(1)}$ <0.15, maka proses akan dilanjutkan ke langkah (2).

Langkah (2) dimulai dengan *backward,* yaitu membuang variabel yang tidak penting dalam model. Kemudian dilanjutkan dengan *forward* untuk memilih variabel yang masuk dalam model.

Tabel 4.12. Nilai log-likelihood, *Likelihood Ratio Test* dan $P_j^{(2)}$ langkah (2) data 1.

	Variabel yang dibuang	Variabel dalam model	Log-likelihood	G	$P_j^{(2)}$
Backward		C, W_1, W_6	-78.039		
	W_1	C, W_6	-110.844	65.610	0.0000
	W_6	**C, W_1**	**-79.560**	**3.042**	**0.0926**
Forward	Variabel	Log-likelihood	G	$P_j^{(2)}$	
	C, W_1, W_6	-78.039			
	C, W_1, W_6, W_2	-77.959	0.160	0.69938	
	C, W_1, W_6, W_3	-77.751	0.576	0.45217	
	C, W_1, W_6, W_4	-77.972	0.134	0.71671	
	C, W_1, W_6, W_5	**-77.008**	**2.062**	**0.14080**	

Pada langkah (2) ini variabel W_1 dan W_6 tidak ada yang keluar dari model, karena $P_j^{(2)}$ maksimal pada langkah *backward* lebih kecil dari P_R. Untuk langkah *forward* dipilih variabel dengan nilai $P_j^{(2)}$ terkecil (0.14080), agar menjadi variabel pada langkah (3), yaitu variabel W_5.

Langkah (3) dimulai dengan *backward*, kemudian dilanjutkan dengan *forward* sama halnya dengan langkah (2).

Tabel 4.13. Nilai log-likelihood, *Likelihood Ratio Test* dan $P_j^{(3)}$ langkah 3 data 1.

	Variabel yang dibuang	Variabel dalam model	Log-likelihood	G	$P_j^{(3)}$
Backward		C, W_1, W_6, W_5	-77.008		
	W_1	C, W_6, W_5	-109.680	65.344	0.0000
	W_6	C, W_1, W_5	-78.543	3.070	0.0803
	W_5	**C, W_1,W_6**	**-78.039**	**2.062**	**0.1408**
Forward	Variabel	Log-likelihood	G	$P_j^{(3)}$	
	C, W_1, W_6, W_5	-77.008			
	C, W_1, W_6, W_5, W_2	-76.925	0.332	0.6859	
	C, W_1, W_6, W_5, W_3	-76.671	0.674	0.4134	
	C, W_1, W_6, W_5, W_4	**-76.923**	**0.170**	**0.4134**	

Pada langkah *backward* tidak ada variabel yang dibuang dari model, sedangkan untuk *forward* calon variabel yang masuk adalah W_4. Akan tetapi $P_j^{(3)}$ untuk variabel W_4 lebih besar dari 0.15, maka W_4 tidak layak masuk dalam model dan proses berhenti. Dari proses pembentukan model dengan metode $PCLR_{(S)}$ pada data 1 ini didapatkan variabel komponen utama yang layak berada dalam model adalah W_1, W_5 dan W_6, sehingga 3 variabel tersebut yang digunakan sebagai variabel prediktor untuk mendapatkan model regresi logistik dengan metode $PCLR_{(S)}$.

Setelah didapatkan variabel prediktor yang layak masuk kedalam model, yaitu varibel W_1, W_5 dan W_6, maka perlu dilakukan uji kesesuaian model dengan hipotesis:

Ho : Model sesuai lawan

H_1 : Model tidak sesuai

Adapun statistik uji yang digunakan untuk menguji kesesuaian model adalah uji *Pearson* dan *Deviance*. Hasil uji kesesuaian model untuk data 1 disajikan dalam Tabel 4.14.

Tabel 4.14. Hasil uji kesesuaian model data 1 (metode $PCLR_{(S)}$).

Metode	χ^2_{hitung}	$\chi^2_{(v)}$	*P*	Keterangan
Pearson	99.996	229.65	0.370	Terima Ho
Deviance	100.932	229.65	0.345	Terima Ho

Berdasarkan uji kesesesuaian model pada Tabel 4.14, maka model dengan variabel prediktor W_1, W_5, dan W_6 layak digunakan.

b. Data 2

Sama seperti perlakuan pada data 1, Langkah (0) dimulai dengan menghitung nilai log-likelihood model dengan intersep dan model dengan W_1, W_2, W_3, W_4, W_5 dan W_6..

Tabel 4.15. Nilai log-likelihood, *Likelihood Ratio Test* dan $P_j^{(0)}$ langkah (0) data 2.

Variabel	Log-likelihood	G	$P_j^{(0)}$
C	-53.928		
C, W_1	-52.864	2.128	0.145
C, W_2	-53.887	0.080	0.777
C, W_3	-53.910	0.036	0.850
C, W_4	-53.675	0.504	0.478
C, W_5	-53.926	0.003	0.954
C, W_6	**-51.623**	**4.609**	**0.032**

Pada langkah (0) dipilih variabel prediktor sebagai calon peubah yang akan dimasukkan pada langkah (1) yaitu variabel dengan nilai

$P_j^{(0)}$ terkecil pada Tabel 4.15, yaitu variabel W_6. Karena nilai $P_6^{(0)}<0.15$, maka proses dilanjutkan ke langkah (1).

Langkah (1) dimulai dengan menghitung nilai log-likelihood model yang mengandung intersep dan W_6 dengan W_1, W_2, W_3, W_4, W_5.

Tabel 4.16. Nilai log-likelihood, *Likelihood Ratio Test* dan $P_j^{(1)}$ langkah (1) data 2.

Variabel	Log-likelihood	G	$P_j^{(1)}$
C, W_6	-51.623		
C, W_6, W_1	**-50.466**	**2.314**	**0.14566**
C, W_6, W_2	-51.533	0.180	0.67480
C, W_6, W_3	-51.611	0.024	0.87950
C, W_6, W_4	-51.305	0.636	0.42846
C, W_6, W_5	-51.623	0.000	0.99999

Variabel pada langkah (1) dengan nilai $P_j^{(1)}$ terkecil (0.14566) sebagai calon variabel yang masuk pada model adalah W_1. Karena nilai $P_j^{(1)} < 0.15$, maka proses akan dilanjutkan ke langkah (2).

Langkah (2) dimulai dengan *backward*, yaitu membuang variabel yang tidak penting dalam model. Kemudian dilanjutkan dengan *forward* untuk memilih variabel yang masuk dalam model.

Tabel 4.17. Nilai log-likelihood, *Likelihood Ratio Test* dan $P_j^{(2)}$ langkah (2) data 2.

	Variabel yang dibuang	Variabel dalam model	Log-likelihood	G	$P_j^{(2)}$
Backward		C, W_6, W_1	-50.466		
	W_6	C, W_1	-52.864	4.796	0.0307
	W_1	**C, W_6, W_1**	**-51.623**	**2.314**	**0.1309**
Forward	Variabel	Log-likelihood	G	$P_j^{(2)}$	
	C, W_6, W_1	-50.466			
	C, W_6, W_1, W_2	-50.410	0.112	0.74296	
	C, W_6, W_1, W_3	-50.465	0.002	0.93670	
	C, W_6, W_1, W_4	**-50.152**	**0.628**	**0.43162**	
	C, W_6, W_1, W_5	-50.466	0.000	0.99999	

Pada langkah (2) ini, variabel W_6 dan W_1 tidak ada yang keluar dari model, karena $P_j^{(2)}$ maksimal pada langkah *backward* lebih kecil dari P_R. Untuk langkah *forward* dipilih variabel dengan nilai $P_j^{(2)}$

terkecil (0.43162) , akan tetapi $P_j^{(2)}$ untuk variabel W_4 lebih besar dari 0.15, maka W_4 tidak layak masuk dalam model dan proses berhenti. Sehingga didapatkan dua variabel prediktor yang layak masuk dalam model, yaitu variabel W_1 dan W_6 , yang selanjutnya dua variabel tersebut digunakan untuk mendapatkan model regresi logistik.

Pada pembentukan model regresi logistik dengan variabel prediktor W_1 dan W_6 perlu adanya uji kesesuaian model. Sama seperti data 1, uji kesesuaian model yang digunakan adalah statistik uji *Pearson* dan *Devianc*e, dengan hipotesis :

Ho : Model sesuai lawan

H_1 : Model tidak sesuai

Hasil uji kesesuaian model untuk data 2 disajikan dalam Tabel 4.18 berikut:

Tabel 4.18. Hasil uji kesesuaian model data 2 (metode $PCLR_{(S)}$).

Metode	χ^2_{hitung}	$\chi^2_{(v)}$	*P*	Keterangan
Pearson	100.083	119.82	0.395	Terima Ho
Deviance	100.932	119.82	0.372	Terima Ho

Berdasarkan uji kesesesuaian model pada Tabel 4.18 di atas, maka model dengan variabel prediktor W_1 dan W_6 layak digunakan.

4.2. Pembahasan.

Dari metode PCLR untuk data 1 dengan dua variabel prediktor mengandung multikolinieritas digunakan tiga variabel komponen utama sebagai variabel prediktor, yaitu W_1, W_2 dan W_3. Namun dengan metode $PCLR_{(S)}$, variabel komponen utama yang digunakan sebagai variabel prediktor pada pembentukan model regresi logistik adalah W_1, W_5 dan W_6. Dalam Tabel 4.19. berikut akan diperlihatkan perbedaan pemilihan komponen utama yang masuk ke dalam model dari kedua metode dan nilai *McFadden R-squared* yang dihasilkan.

Tabel 4.19. Perbandingan variabel prediktor yang masuk ke dalam model (data 1).

No	Metode	Variabel prediktor yang terlibat	*McFadden R-squared*

1.	PCLR	W_1, W_2, W_3	0.605111
2.	$PCLR_{(S)}$	W_1, W_5, W_6	0.635818

Penerapan metode PCLR pada data 2 di mana semua variabel prediktornya mengandung multikolinieritas menghasilkan tiga variabel komponen utama sebagai variabel prediktor (W_1, W_2 dan W_3), sedangkan dengan metode $PCLR_{(S)}$ didapatkan dua variabel komponen utama yang digunakan sebagai variabel prediktor pada pembentukan model regresi logistik, yaitu W_1 dan W_6. Adapun ringkasan hasil yang diperoleh dengan menggunakan metode PCLR dan $PCLR_{(S)}$ disajikan pada Tabel 4.20.

Tabel 4.20. Perbandingan variabel prediktor yang masuk ke dalam model (data 2).

No	Metode	Variabel prediktor yang terlibat	*McFadden R-squared*
1.	PCLR	W_1, W_2, W_3	0.615635
2.	$PCLR_{(S)}$	W_1, W_6	0.654166

Dari Tabel 4.19. dan Tabel 4.20. di atas terlihat bahwa model yang diperoleh dengan menggunakan metode $PCLR_{(S)}$ menghasilkan nilai *McFadden R-squared* lebih besar dibandingkan metode PCLR, sehingga besarnya keragaman variabel respon yang dijelaskan oleh metode $PCLR_{(S)}$ lebih besar daripada metode PCLR. Pada data 2 terlihat bahwa banyaknya variabel komponen utama yang digunakan dalam pembentukan model dengan metode $PCLR_{(S)}$ lebih sedikit dibanding metode PCLR, namun hal tersebut tidak memperkecil besarnya keragaman variabel respon yang mampu dijelaskan oleh model.

Keragaman variabel prediktor yang dapat dijelaskan oleh masing-masing variabel komponen utama pada data 1 dan 2 dengan metode PCLR dan $PCLR_{(S)}$ disajikan pada tabel 4.21.

Tabel 4.21. Keragaman variabel prediktor yang dapat dijelaskan variabel komponen utama pada data 1 dan 2 dengan metode PCLR dan $PCLR_{(S)}$.

	Metode	Variabel Prediktor	Keragaman
Data 1	PCLR	W_1, W_2, W_3	80.60%

	$PCLR_{(S)}$	W_1, W_5, W_6	48.70%
Data 2	PCLR	W_1, W_2, W_3	83.90%
	$PCLR_{(S)}$	W_1, W_6	54.30%

Tabel 4.21. di atas menunjukkan bahwa keragaman variabel prediktor yang dijelaskan variabel komponen utama dengan metode PCLR lebih besar daripada $PCLR_{(S)}$, baik diterapkan pada data 1 ataupun 2. Namun, hal tersebut bukan merupakan kriteria untuk memilih metode yang lebih tepat untuk mendapatkan model regresi logistik.

Berdasarkan penerapan kedua metode (PCLR dan $PCLR_{(S)}$) pada dua data, yaitu data dengan sebagian variabel prediktor mengandung multikolinieritas dan data dengan semua variabel prediktor mengandung multikolinieritas dengan menggunakan indikator pembanding keakuratan model yaitu *Pearson Residuals* dan *Percent Correct Predictions* (PCP), hasil analisis secara lengkap disajikan pada Tabel 4.22.

Tabel 4.22. Perbandingan Nilai *Pearson Residuals* dan *Percent Correct Predictions* (PCP) data 1 dan 2.

	Ukuran Keakuratan	PCLR	$PCLR_{(S)}$
Data 1	*Pearson Residuals*	3.66378	3.00851
	PCP	83.00%	84.50%
Data 2	*Pearson Residuals*	0.14132	0.03534
	PCP	77.00%	78.00%

Dari Tabel 4.22 dapat dilihat bahwa nilai nilai *pearson residuals* metode $PCLR_{(S)}$ lebih kecil dibandingkan metode PCLR, sedangkan nilai PCP metode $PCLR_{(S)}$ lebih besar dibandingkan metode PCLR, baik untuk data 1 maupun data 2. Dengan melihat nilai *Percent Correct Prediction* (PCP) data 1 dengan model $PCLR_{(S)}$ pada Lampiran 12, dapat dikatakan bahwa model ini mampu mengklasifikasikan secara tepat 84.5% data dari variabel prediktor, atau dapat dikatakan kemampuan pengklasifikasian model ini 1.5% lebih bagus dibandingkan model PCLR. Sedangkan pada data 2, model $PCLR_{(S)}$ mampu mengklasifikasikan secara tepat 80% data dari variabel prediktor, atau dapat dikatakan kemampuan pengklasifikasian model ini 3% lebih bagus dibandingkan model PCLR seperti yang terlihat dari nilai PCP pada Lampiran 12 .Hal ini mengindikasikan bahwa model $PCLR_{(S)}$ lebih baik dibandingkan model PCLR, sehingga untuk kedua data tersebut akan lebih baik jika menggunakan metode $PCLR_{(S)}$ untuk mendapatkan model persamaan regresi logistik. Meskipun pada tabel 4.21 disebutkan bahwa dengan metode $PCLR_{(S)}$ keragaman variabel prediktor yang mampu dijelaskan oleh variabel komponen utama lebih kecil, namun dengan metode $PCLR_{(S)}$ model yang didapatkan lebih akurat dengan indikator pembanding *pearson residuals* dan *percent correct prediction* (PCP).

Berdasarkan penelitian ini, dapat disimpulkan bahwa metode PCLR lebih bagus dibanding $PCLR_{(S)}$, baik untuk data dengan dua variabel prediktor yang mengandung multikolinieritas (data 1) maupun data dengan semua variabel prediktor mengandung multikolinieritas (data 2). Hal ini terjadi karena pemilihan variabel komponen utama pada $PCLR_{(S)}$ didasarkan pada keeratan hubungan variabel komponen utama tersebut dengan variabel respon, sehingga meminimalkan kuadrat tengah galat. Sedangkan untuk model PCLR pemilihan variabel komponen utama yang masuk dalam model didasarkan pada akar ciri terbesar. Sehingga lebih disarankan untuk menggunakan $PCLR_{(S)}$ untuk membentuk model regresi logistik apabila terdapat multikolinieritas pada sebagian dan semua variabel prediktor.

Model regresi logistik data 1 dengan metode $PCLR_{(S)}$ adalah sebagai berikut :

$$\pi_i = \frac{\exp(-1.47819 - 1.08931\, W_1 + 0.395269\, W_5 - 0.863340\, W_6)}{1 + \exp(-1.47819 - 1.08931\, W_1 + 0.395269\, W_5 - 0.863340\, W_6)}$$

Model di atas dapat ditransformasi menjadi:

$$\pi_i = \frac{\exp(-1.47819 + 0.37774\, X_1 + 0.11611\, X_2 + 1.19571\, X_3 + 0.29092\, X_4 + 0.17017\, X_5 - 0.18755\, X_6)}{1 + \exp(-1.47819 + 0.37774\, X_1 + 0.11611\, X_2 + 1.19571\, X_3 + 0.29092\, X_4 + 0.17017\, X_5 - 0.18755\, X_6)}$$

Interpretasi model regresi di atas didasarkan pada nilai *odds ratio* sebagaimana disajikan pada Tabel 4.23 berikut :

Tabel 4.23. Nilai *odds ratio* data 1.

Variabel	*odds ratio*
X_1	1.45898
X_2	1.12312
X_3	3.30587
X_4	1.33766
X_5	1.18551

X_6	0.82899

Untuk variabel prediktor X_1 didapatkan *odds ratio* sebesar 1.45898, artinya untuk setiap bertambahnya umur 1 tahun akan meningkatkan resiko terjadi serangan jantung sebesar 1.45898 kali. Untuk variabel prediktor X_2 dengan nilai *odds ratio* sebesar 1.12312 mengindikasikan adanya peningkatan resiko terjadi serangan jantung sebesar 1.12312 kali setiap penambahan 1 millimeter merkuri tekanan darah sistolik, sedangkan diastolik setiap penambahan 1 milimeter merkuri dapat meningkatkan resiko serangan jantung sebesar 3.30587 kali. Untuk variabel X_4 yaitu kolesterol, setiap penambahan 1 mg/dl dapat meningkatkan resiko serangan jantung sebesar 1.33766 kali. Setiap selisih 1 inchi tinggi badan penderita penyakit jantung menyebabkan resiko terjadi serangan jantung meningkat sebesar 1.18551 kali. Berbeda dengan variabel prediktor lain, X_6 yaitu berat badan mengindikasikan bahwa untuk setiap penurunan berat badan sebesar 1 pon, dapat mengurangi resiko serangan jantung sebesar 0.82899 kali.

Untuk data 2, model regresi terbaik adalah model yang terbentuk dengan metode $PCLR_{(S)}$. Model regresi logistik data 2 dengan metode $PCLR_{(S)}$ adalah sebagai berikut :

$$\pi_i = \frac{\exp(1.30905 + 0.205902W_1 - 1.10912W_6)}{1+\exp(1.30905 + 0.205902W_1 - 1.10912W_6)}$$

Model di atas dapat ditransformasi menjadi :

$$\pi_i = \frac{\exp(1.30905 + 0.16758X_1 + 0.048214X_2 - 0.59947X_3 - 0.42038X_4 - 0.52614X_5 - 0.65517X_6)}{1+\exp(1.30905 + 0.16758X_1 + 0.048214X_2 - 0.59947X_3 - 0.42038X_4 - 0.52614X_5 - 0.65517X_6)}$$

Untuk masing-masing variabel prediktor didapatkan nilai *odds ratio* seperti pada tabel 4.24 berikut :

Tabel 4.24. Nilai *odds ratio* data 2.

Variabel	*odds ratio*
X_1	1.18244
X_2	1.04994
X_3	0.54910
X_4	0.65680
X_5	0.59088

X_6	0.51914

Nilai *odds ratio* untuk variabel X_1 sebesar 1.18244 mengindikasikan bahwa untuk setiap penambahan 1 kg berat badan kambing menyebabkan kehamilan meningkat sebesar 1.18244 kali. Untuk variabel X_2, setiap penambahan 1 cm tinggi badan kambing dapat meningkatkan kemungknan terjadi kehamilan sebesar 1.04994 kali, tetapi apabila panjang badan bertambah sebesar 1 cm, maka kemungkinan terjadinya kehamilan akan turun sebesar 0.54910 kali. Dari nilai *odds ratio* X_4 yaitu lingkar dada, diindikasikan apabila terjadi penambahan 1 cm, maka kemungkinan terjadi kehamilan kambing akan menurun sebesar 0.65680 kali. Sama seperti lingkar dada, X_5 yaitu lebar pinggul, apabila terjadi penambahan sebesar 1 cm, maka kemungkinan kehamilan turun sebesar 0.59088 kali. Volume tubuh kambing memiliki pengaruh yang tidak baik terhadap fertilitas kambing, karena setiap penambahan 1 cm^3 dapat menurunkan kemungkinan terjadinya kehamilan sebesar 0.51914 kali.

BAB V
KESIMPULAN DAN SARAN

5.1 Kesimpulan.

Berdasarkan hasil analisis dan pembahasan pada bab sebelumnya, beberapa kesimpulan yang dapat diambil dari penelitian ini adalah :

1. Model regresi logistik data 1 yang didapatkan dari metode PCLR adalah sebagai berikut :

$$\pi_i = \frac{\exp(-1.53325 - 1.071341\,W_1 + 0.067109\,W_2 + 0.181730\,W_3)}{1+\exp(-1.53325 - 1.071341\,W_1 + 0.067109\,W_2 + 0.181730\,W_3)}$$

Sedangkan model yang didapatkan dari metode $PCLR_{(S)}$ adalah:

$$\pi_i = \frac{\exp(-1.47819 - 1.08931\,W_1 + 0.395269\,W_5 - 0.863340\,W_6)}{1+\exp(-1.47819 - 1.08931\,W_1 + 0.395269\,W_5 - 0.863340\,W_6)}$$

Untuk data 2, model regresi logistik yang didapatkan dengan menerapkan metode PCLR adalah sebagai berikut:

$$\pi_i = \frac{\exp(1.23935 + 0.192380\,W_1 + 0.0677836\,W_2 - 0.0338814\,W_3)}{1+\exp(1.23935 + 0.192380\,W_1 + 0.0677836\,W_2 - 0.0338814\,W_3)}$$

Jika digunakan metode $PCLR_{(S)}$ untuk data 2, didapatkan model:

$$\pi_i = \frac{\exp(1.30905 + 0.205902\,W_1 - 1.10912\,W_6)}{1+\exp(1.30905 + 0.205902\,W_1 - 1.10912\,W_6)}$$

2. Bentuk transformasi model regresi logistik data 1 adalah:

a. Metode PCLR:

$$\pi_i = \frac{\exp(-1.53325 + 0.53324\,X_1 + 0.50309\,X_2 + 0.54239\,X_3}{1+\exp(-1.53325 + 0.53324\,X_1 + 0.50309\,X_2 + 0.54239\,X_3}$$

$$\frac{0.46868\,X_4 - 0.21639\,X_5 + 0.16403\,X_6)}{0.46868\,X_4 - 0.21639\,X_5 + 0.16403\,X_6)}$$

b. Metode $PCLR_{(S)}$

$$\pi_i = \frac{\exp(-1.47819 + 0.37774\,X_1 + 0.11611\,X_2 + 1.19571\,X_3}{1+\exp(-1.47819 + 0.37774\,X_1 + 0.11611\,X_2 + 1.19571\,X_3}$$

$$\frac{0.29092\,X_4 + 0.17017\,X_5 - 0.18755\,X_6)}{0.29092\,X_4 + 0.17017\,X_5 - 0.18755\,X_6)}$$

Untuk data 2 diperoleh model transformasi :

a. Metode PCLR

$$\pi_i = \frac{\exp(1.23935 - 0.09263\,X_1 + 0.0759\,X_2 - 0.10454\,X_3}{1+\exp(1.23935 - 0.09263\,X_1 + 0.0759\,X_2 - 0.10454\,X_3}$$

$$\frac{-0.11585\,X_4 - 0.00484\,X_5 - 0.08611\,X_6)}{-0.11585\,X_4 - 0.00484\,X_5 - 0.08611\,X_6)}$$

b. Metode $PCLR_{(S)}$

$$\pi_i = \frac{\exp(1.30905 + 0.16758X_1 + 0.048214X_2 - 0.59947X_3 - 0.42038X_4 - 0.52614X_5 - 0.65517X_6)}{1 + \exp(1.30905 + 0.16758X_1 + 0.048214X_2 - 0.59947X_3 - 0.42038X_4 - 0.52614X_5 - 0.65517X_6)}$$

3. Pada data dengan sebagian variabel prediktor mengandung multikolinieritas metode yang tepat untuk mendapatkan model terbaik adalah *Principal Component Logistic Regression Stepwise* ($PCLR_{(S)}$). Jika metode $PCLR_{(S)}$ tersebut diterapkan pada data dengan semua variabel prediktornya mengandung multikolinieritas, model yang didapatkan lebih bagus dibanding metode PCLR dengan dengan menggunakan indikator pembanding keakuratan model yaitu *Pearson residuals* dan *Percent Correct Predictions* (PCP), meskipun variabel komponen utama yang digunakan sebagai variabel prediktor pada metode $PCLR_{(S)}$ menjelaskan keragaman variabel prediktor lebih sedikit dibanding metode PCLR.

5.2 Saran.

1. Apabila peneliti ingin mengatasi multikolinieritas dengan tetap mempertahankan sebesar mungkin keragaman variabel prediktor, maka sebaiknya menggunakan metode PCLR, namun jika peneliti ingin mengatasi multikolinieritas dengan meminimumkan ragam galat maka lebih disarankan menggunakan metode $PCLR_{(S)}$, dengan resiko keragaman variabel prediktor lebih kecil. Model yang didapatkan dengan metode $PCLR_{(S)}$ lebih bagus daripada PCLR, namun penerapannya tetap diserahkan kepada peneliti.
2. Penelitian selanjutnya dapat dikembangkan untuk pemilihan model regresi logistik dengan variabel prediktor diskrit dan campuran, sehingga analisis komponen utama yang digunakan untuk mentransformasi variabel prediktor adalah analisis komponen utama nonlinier.
3. Untuk selanjutnya dapat dilakukan analisis yang sama dengan variabel respon bersifat politomus, baik politomus ordinal maupun politomus nominal.

DAFTAR PUSTAKA

Agresti, A. 1990. **Categorical Data Analysis**. John Wiley and Sons. New York.

Astutik, S. 2007. **Diktat Analisis Multivariate**. Jurusan Matematika FMIPA Universitas Brawijaya. Malang.

Bowerman, L.B. dan O'Connel, T.R. 1990. **Linier Statistical Models An Applied Approach**. Second Edition. PWS Kent Publishing Company. Boston.

Dixon, W. J. dan F. J. Massey. 1983. **Introduction to Statistical Analysis**. Mc Graw-Hill Book Company. Singapore.

Escabis, M., A. M. Aguilera, dan M. J. Valderrama. 2003. **Principal Component Estimation Of Functional Logistic Regression** : Discussion Of Two Different Approaches. http://hera.ugr.es/doi/15057823.pdf , tanggal akses : 21 Januari 2008.

Escabis, M., A. M. Aguilera, dan M. J. Valderrama. 2005. **Modelling Environmental Data by Functional Component Logistic Regression.** http://hera.ugr.es/doi/15771842.pdf , tanggal akses : 30 Januari 2008.

Fahrmeir , L. dan T. Gerhard. 1994. **Multivariate Statistical Modelling Based on Generalized Linier Models** . John Willey dan Sons , New York.

Gujarati, D. 1991. **Ekonometrika Dasar**. Erlangga. Jakarta.

Hines, W.W. dan Montgomery, D.C. 1990. **Probability and Statistics in Enginering and Management Science**. Third Edition. John Willey and Sons. New York.

Hosmer, D.W. dan S. Lemeshow. 1989. **Applied Logistic Regression**. John Wiley and Sons Inc. New York.

Imam, Kamarul. 1994. **Introduksi Regresi Linier Logit**. http://www2.chass.ncsu.edu/garson/PA765/logistic.htm , tanggal akses: 15 Maret 2008.

Johnson, R.A. dan D. W. Wichern. 1982. **Applied Multivariate Analysis**. Prentice Hall Inc. New Jersey.

Kutner, M. H, C. J. Nachtsheim, dan J. Neter. 2004. **Applied Linear Regression Models**. 4th Edition. Mc. Graw-Hill Companies, Inc. New York.

Miller, W. G. 2007. **Statistics and Measurement Using OpenStat**. http://www.statpages.org/miller/openstat/.

Myers, H.R. 1990. **Classical and Modern Regression with Application**. Second Edition. Duxbury Press. Belmont California.

Sembiring, R.K.1995. **Analisis Regresi**. Penerbit ITB. Jakarta.

Sumodiningrat, G. 1999. **Pengantar Ekonometrika**. BPFE. Yogyakarta.

Supranto, J. 1992. **Metode Statistika Teori dan Aplikasi.** Erlangga. Jakarta.

Walpole, R.E. 1980. **Pengantar Statistika** (diterjemahkan oleh: Bambang Sumantri). Gramedia. Jakarta.

Lampiran 1. Data 1 yaitu faktor-faktor yang mempengaruhi serangan jantung pasien laki-laki 10 tahun setelah pemeriksaan terakhir

No	Y	X1	X2	X3	X4	X5	X6
1	1	44	124	80	254	70	190
2	0	35	110	70	240	73	216
3	0	41	114	80	279	68	178
4	0	31	100	80	284	68	149
5	1	61	190	110	315	68	182
6	0	61	130	88	250	70	185
7	0	44	130	94	298	68	161
8	1	58	110	74	384	67	175
9	0	52	120	80	310	66	144
10	0	52	120	80	337	67	130
11	1	52	130	80	367	69	162
12	0	40	120	90	273	68	175
13	0	49	130	75	273	66	155
14	0	34	120	80	314	74	156
15	0	37	115	70	243	65	151
16	1	63	140	90	341	74	168
17	0	28	138	80	245	70	185
18	0	40	115	82	302	69	225
19	1	51	148	110	302	69	247
20	0	33	120	70	386	66	146
21	1	37	110	70	312	71	170
22	0	33	132	90	302	69	161
23	0	41	112	80	394	69	167
24	0	38	114	70	358	69	198
25	1	52	100	78	336	70	162
26	0	31	114	80	251	71	150
27	1	44	110	80	322	68	196
28	0	31	108	70	281	67	130
29	1	40	110	74	336	68	166
30	0	36	136	80	314	73	178
31	0	42	124	82	383	69	187
32	0	28	120	82	360	67	148

Lampiran 1. (Lanjutan)

33	0	40	150	85	369	71	180
34	1	40	150	100	333	70	172
35	0	35	100	70	253	68	141
36	0	32	120	80	268	68	176
37	0	31	110	80	257	71	154
38	1	52	130	90	474	69	145
39	1	45	110	80	391	69	159
40	0	39	106	80	248	67	181
41	1	40	130	90	520	68	169
42	1	48	110	70	285	66	160
43	0	29	110	70	352	66	149
44	1	56	141	100	428	65	171
45	0	53	90	55	334	68	166
46	0	47	90	60	278	69	121
47	0	30	114	76	264	73	178
48	1	64	140	90	243	71	171
49	0	31	130	88	348	72	181
50	0	35	120	88	290	70	162
…	…	…	…	…	…	…	…
196	0	29	110	70	238	72	143
197	0	49	112	78	283	64	149
198	0	49	100	70	264	70	166
199	1	50	128	92	264	70	176
200	0	31	105	68	193	67	141

Keterangan :
X_1 = Umur (tahun)
X_2 = Tekanan darah sistolik (millimeter merkuri)
X_3 = Tekanan darah diastolik (millimeter merkuri)
X_4 = Kolesterol (mg/dl)
X_5 = tinggi (inchi)
X_6 = Berat (pond)

Lampiran 2. Data 2 yaitu faktor-faktor yang mempengaruhi fertilitas kambing PE di Kecamatan Parang Kabupaten Magetan.

No	Y	X_1	X_2	X_3	X_4	X_5	X_6
1	1	30	68	58	69	25	100050
2	1	30	68	59	67	24	102778
3	1	33	68.5	59	71	26	117292
4	1	33	67	60	70	27	117600
5	1	33	68	60	72	26	99360
6	1	33	67.5	56	69	25	92736
7	1	35	68	58	66	32	103356
8	1	32	69	65	75	24	126750
9	1	30	68	60	68	23	114240
10	1	34	69	58	70	29	97440
11	0	32	68	59	69	27	118059
12	1	33	67	61	71	26	110050
13	1	35	69.5	65	76	23	109440
14	1	31	68	66	69	21	82593
15	0	30	71	68	70	30	103250
16	1	31	70	59	71	26	97980
17	1	33	68	62	68	26	119340
18	1	35	70	66	72	29	114696
19	1	32	71	64	73	24	117676
20	0	32	67.5	66	68	25	108324
21	1	34	69	66	70	30	121520
22	1	35	68	67	68	31	96288
23	0	31	70	64	70	25	105560
24	0	33	70	63	72	24	104400
25	0	31	68	62	69	23	104328
26	1	32	69	62	69.5	25	104250
27	1	32	68.5	63	69	25	104328
28	1	31	69	60	70	23	112840
29	1	32	70	62	73	24	126144
30	1	30	68	61	72	24	116064

Lampiran 2. (Lanjutan)

31	1	34	68.5	65	72	27	115200
32	1	27	65	58	74	25	156288
33	1	27	65	59	72	26	112320
34	1	27	65	59	70	28	103040
35	1	27	66	60	74	28	141636
36	1	27	67	60	72	23	128304
37	1	27	66	56	73	24	115778
38	0	28	64	58	75	27	112125
39	1	29	60	65	73	26	101178
40	1	28	62	60	69	28	140760
41	1	28	61.5	58	72	24	110448
42	0	28	62	59	72	29	116064
43	1	28	68	62	74	25	141636
44	1	29	60	60	70	24	107520
45	1	29	65	57	71	21	117150
46	1	29	64	59	73	25	144540
47	1	29	65	60	74	23	153698
48	0	29	65	65	69.5	27	111200
49	1	29	66	59	71	27	107352
50	1	29	68	62	71	26	101246
…..	…..	…..	…..	….	….	….	….
96	1	41	70	69	74	29	148074
97	1	41	69	72	76	25	136800
98	0	41	69	73	75	26	142350
99	0	42	70	71	73	28	145124
100	1	42	70	72	78	27	151632

Keterangan:
X_1 = Berat badan (kg)
X_2 = Tinggi badan (cm)
X_3 = panjang badan (cm)
X_4 = lingkar dada (cm)
X_5 = lebar pinggul (cm)
X_6 = volume tubuh (cm^2)

Lampiran 3. Pendeteksian multikolinieritas data 1 dan data 2.

Data 1

Regression Analysis: X1 versus X2, X3, X4, X5, X6

```
The regression equation is
X1 = 54.4 + 0.187 X2 + 0.072 X3 + 0.0561 X4 - 0.885 X5
+ 0.0254 X6

Predictor      Coef   SE Coef      T      P
Constant      54.37     22.19   2.45  0.015
X2          0.18721   0.06904   2.71  0.007
X3           0.0725    0.1214   0.60  0.551
X4          0.05607   0.01121   5.00  0.000
X5          -0.8854    0.3171  -2.79  0.006
X6          0.02542   0.03312   0.77  0.444

S = 9.73900   R-Sq = 51.9%   R-Sq(adj) = 50.1%
```

Regression Analysis: X2 versus X1, X3, X4, X5, X6

```
The regression equation is
X2 = - 0.8 + 0.195 X1 + 1.29 X3 + 0.0042 X4 + 0.224 X5
- 0.0456 X6

Predictor       Coef   SE Coef      T      P
Constant       -0.81     23.00  -0.04  0.972
X1           0.19506   0.07193   2.71  0.007
X3           1.29135   0.08236  15.68  0.000
X4           0.00417   0.01215   0.34  0.732
X5            0.2235    0.3297   0.68  0.499
X6          -0.04562   0.03370  -1.35  0.177

S = 9.94095   R-Sq = 93.9%   R-Sq(adj) = 93.0%
```

Regression Analysis: X3 versus X1, X2, X4, X5, X6

```
The regression equation is
X3 = 42.0 + 0.0253 X1 + 0.433 X2 + 0.00336 X4 - 0.425
X5 + 0.0849 X6

Predictor       Coef    SE Coef      T      P
Constant       41.95      12.97   3.23  0.001
X1           0.02531    0.04239   0.60  0.551
X2           0.43282    0.02761  15.68  0.000
X4          0.003355   0.007034   0.48  0.634
X5           -0.4251     0.1887  -2.25  0.025
X6           0.08489    0.01863   4.56  0.000

S = 5.75522   R-Sq = 94.7%   R-Sq(adj) = 93.8%
```

Regression Analysis: X4 versus X1, X2, X3, X5, X6

```
The regression equation is
X4 = 211 + 2.04 X1 + 0.146 X2 + 0.349 X3 - 0.81 X5 -
0.019 X6

Predictor     Coef  SE Coef      T      P
Constant     211.2    135.0   1.56  0.119
X1          2.0375   0.4073   5.00  0.000
X2          0.1456   0.4239   0.34  0.732
X3          0.3492   0.7320   0.48  0.634
X5          -0.815    1.949  -0.42  0.676
X6         -0.0189   0.2000  -0.09  0.925

S = 58.7088   R-Sq = 48.0%   R-Sq(adj) = 45.9%
```

Regression Analysis: X5 versus X1, X2, X3, X4, X6

```
The regression equation is
X5 = 66.4 - 0.0436 X1 + 0.0106 X2 - 0.0600 X3 -
0.00110 X4 + 0.0480 X6

Predictor       Coef   SE Coef      T      P
Constant      66.432     1.505  44.15  0.000
X1          -0.04363   0.01563  -2.79  0.006
X2           0.01057   0.01560   0.68  0.499
X3          -0.05998   0.02662  -2.25  0.025
X4         -0.001105  0.002643  -0.42  0.676
X6          0.047952  0.006510   7.37  0.000

S = 2.16188   R-Sq = 56.9%   R-Sq(adj) = 55.0%
```

Regression Analysis: X6 versus X1, X2, X3, X4, X5

```
The regression equation is
X6 = - 220 + 0.119 X1 - 0.205 X2 + 1.14 X3 - 0.0024 X4
+ 4.56 X5

Predictor      Coef  SE Coef      T      P
Constant    -219.66    46.14  -4.76  0.000
X1           0.1190   0.1551   0.77  0.444
X2          -0.2051   0.1515  -1.35  0.177
X3           1.1385   0.2499   4.56  0.000
X4         -0.00244  0.02577  -0.09  0.925
X5           4.5579   0.6188   7.37  0.000

S = 21.0771   R-Sq = 50.4%   R-Sq(adj) = 48.6%
```

Lampiran 3. (Lanjutan).

Data 2

Regression Analysis: X1 versus X2, X3, X4, X5, X6

The regression equation is
X1 = - 23.0 + 0.158 X2 + 0.567 X3 + 0.057 X4 + 0.144 X5 + 0.000022 X6

Predictor	Coef	SE Coef	T	P
Constant	-22.98	12.05	-1.91	0.060
X2	0.1575	0.1337	1.18	0.242
X3	0.56740	0.09169	6.19	0.000
X4	0.0572	0.1481	0.39	0.700
X5	0.1438	0.1432	1.00	0.318
X6	0.00002226	0.00002611	0.85	0.396

S = 3.58019   R-Sq = 93.6%   R-Sq(adj) = 91.2%

Regression Analysis: X2 versus X1, X3, X4, X5, X6

The regression equation is
X2 = 46.0 + 0.0924 X1 + 0.158 X3 + 0.085 X4 + 0.139 X5 - 0.000014 X6

Predictor	Coef	SE Coef	T	P
Constant	45.976	8.123	5.66	0.000
X1	0.09241	0.07842	1.18	0.242
X3	0.15772	0.08170	1.93	0.057
X4	0.0846	0.1132	0.75	0.456
X5	0.1386	0.1093	1.27	0.208
X6	-0.00001398	0.00002002	-0.70	0.487

S = 2.74206   R-Sq = 91.8%   R-Sq(adj) = 90.5%

Regression Analysis: X3 versus X1, X2, X4, X5, X6

The regression equation is
X3 = 16.4 + 0.510 X1 + 0.242 X2 + 0.127 X4 - 0.136 X5 + 0.000065 X6

Predictor	Coef	SE Coef	T	P
Constant	16.43	11.52	1.43	0.157
X1	0.51019	0.08244	6.19	0.000
X2	0.2418	0.1252	1.93	0.057
X4	0.1271	0.1399	0.91	0.366
X5	-0.1359	0.1358	-1.00	0.319
X6	0.00006526	0.00002392	2.73	0.008

S = 3.39488   R-Sq = 95.1%   R-Sq(adj) = 94.9%

Lampiran 3. (Lanjutan).

Regression Analysis: X4 versus X1, X2, X3, X5, X6

The regression equation is
X4 = 58.5 + 0.0277 X1 + 0.0699 X2 + 0.0685 X3 - 0.275 X5 + 0.000084 X6

Predictor	Coef	SE Coef	T	P
Constant	58.540	6.048	9.68	0.000
X1	0.02772	0.07171	0.39	0.700
X2	0.06987	0.09343	0.75	0.456
X3	0.06846	0.07536	0.91	0.366
X5	-0.27485	0.09608	-2.86	0.005
X6	0.00008429	0.00001603	5.26	0.000

S = 2.49127   R-Sq = 91.1%   R-Sq(adj) = 90.8%

Regression Analysis: X5 versus X1, X2, X3, X4, X6

The regression equation is
X5 = 29.3 + 0.0738 X1 + 0.121 X2 - 0.0776 X3 - 0.291 X4 + 0.000100 X6

Predictor	Coef	SE Coef	T	P
Constant	29.307	8.263	3.55	0.001
X1	0.07382	0.07350	1.00	0.318
X2	0.12129	0.09567	1.27	0.208
X3	-0.07759	0.07752	-1.00	0.319
X4	-0.2914	0.1019	-2.86	0.005
X6	0.00009999	0.00001569	6.37	0.000

S = 2.56498   R-Sq = 91.2%   R-Sq(adj) = 90.9%

Regression Analysis: X6 versus X1, X2, X3, X4, X5

The regression equation is
X6 = - 208447 + 345 X1 - 369 X2 + 1124 X3 + 2696 X4 + 3017 X5

Predictor	Coef	SE Coef	T	P
Constant	-208447	43284	-4.82	0.000
X1	344.7	404.3	0.85	0.396
X2	-369.0	528.6	-0.70	0.487
X3	1124.1	412.0	2.73	0.008
X4	2695.9	512.8	5.26	0.000
X5	3016.9	473.4	6.37	0.000

S = 14089.1   R-Sq = 95.9%   R-Sq(adj) = 94.9%

Lampiran 3. (Lanjutan).

$$\text{VIF} = \frac{1}{1 - R_j^{\,2}}$$

Data 1		Data 2	
Variabel Prediktor	VIF	Variabel Prediktor	VIF
X_1	2.079	X_1	15.625
X_2	16.393	X_2	12.195
X_3	18.868	X_3	20.408
X_4	1.923	X_4	11.236
X_5	2.320	X_5	11.364
X_6	2.016	X_6	24.390

Lampiran 4. Data 1 setelah dibakukan.

No	X1*	X2*	X3*	X4*	X5*	X6*
1	0.12361	0.12886	-0.15909	-0.48607	0.57066	0.99554
2	-0.64895	-0.70443	-1.15969	-0.70474	1.77205	2.0382
3	-0.13391	-0.46634	-0.15909	-0.09559	-0.23027	0.51431
4	-0.9923	-1.29964	-0.15909	-0.01749	-0.23027	-0.64866
5	1.58288	4.05723	2.84269	0.4667	-0.23027	0.67472
6	1.58288	0.48599	0.64138	-0.54855	0.57066	0.79503
7	0.12361	0.48599	1.24174	0.20118	-0.23027	-0.16743
8	1.32536	-0.70443	-0.75945	1.54442	-0.63073	0.39401
9	0.81032	-0.10922	-0.15909	0.38861	-1.03119	-0.84917
10	0.81032	-0.10922	-0.15909	0.81032	-0.63073	-1.4106
11	0.81032	0.48599	-0.15909	1.2789	0.1702	-0.12732
12	-0.21975	-0.10922	0.8415	-0.1893	-0.23027	0.39401
13	0.55281	0.48599	-0.65939	-0.1893	-1.03119	-0.40804
14	-0.73478	-0.10922	-0.15909	0.45108	2.17251	-0.36794
15	-0.47727	-0.40682	-1.15969	-0.65788	-1.43166	-0.56845
16	1.75456	1.08119	0.8415	0.8728	2.17251	0.11329
17	-1.24982	0.96215	-0.15909	-0.62664	0.57066	0.79503
18	-0.21975	-0.40682	0.04102	0.26365	0.1702	2.39912
19	0.72448	1.55736	2.84269	0.26365	0.1702	3.28138
20	-0.82062	-0.10922	-1.15969	1.57566	-1.03119	-0.76896
21	-0.47727	-0.70443	-1.15969	0.41984	0.97112	0.19349
22	-0.82062	0.60503	0.8415	0.26365	0.1702	-0.16743
23	-0.13391	-0.58539	-0.15909	1.70062	0.1702	0.07319
24	-0.39143	-0.46634	-1.15969	1.13833	0.1702	1.31636
25	0.81032	-1.29964	-0.35921	0.7947	0.57066	-0.12732
26	-0.9923	-0.46634	-0.15909	-0.53293	0.97112	-0.60855
27	0.12361	-0.70443	-0.15909	0.57604	-0.23027	1.23616
28	-0.9923	-0.82347	-1.15969	-0.06435	-0.63073	-1.4106
29	-0.21975	-0.70443	-0.75945	0.7947	-0.23027	0.03308
30	-0.56311	0.84311	-0.15909	0.45108	1.77205	0.51431
31	-0.04807	0.12886	0.04102	1.52881	0.1702	0.87523
32	-1.24982	-0.10922	0.04102	1.16956	-0.63073	-0.68876
33	-0.21975	1.6764	0.3412	1.31014	0.97112	0.59452
34	-0.21975	1.6764	1.84209	0.74785	0.57066	0.2737

Lampiran 4. (Lanjutan)

35	-0.64895	-1.29964	-1.15969	-0.50169	-0.23027	-0.96947
36	-0.90646	-0.10922	-0.15909	-0.2674	-0.23027	0.43411
37	-0.9923	-0.70443	-0.15909	-0.43921	0.97112	-0.44814
38	0.81032	0.48599	0.8415	2.95015	0.1702	-0.80907
39	0.20945	-0.70443	-0.15909	1.65376	0.1702	-0.24763
40	-0.30559	-0.94251	-0.15909	-0.57978	-0.63073	0.63462
41	-0.21975	0.48599	0.8415	3.66863	-0.23027	0.15339
42	0.46697	-0.70443	-1.15969	-0.00187	-1.03119	-0.20753
43	-1.16398	-0.70443	-1.15969	1.04461	-1.03119	-0.64866
44	1.15368	1.14071	1.84209	2.23167	-1.43166	0.2336
45	0.89616	-1.89484	-2.66058	0.76347	-0.23027	0.03308
….	….	….	….	….	….	….
196	-1.16398	-0.70443	-1.15969	-0.73598	1.37159	-0.88927
197	0.55281	-0.58539	-0.35921	-0.03311	-1.83212	-0.64866
198	0.55281	-1.29964	-1.15969	-0.32988	0.57066	0.03308
199	0.63864	0.36695	1.04162	-0.32988	0.57066	0.43411
200	-0.9923	-1.00203	-1.35981	-1.43884	-0.63073	-0.96947

Lampiran 5. Data 2 setelah dibakukan.

No	X1*	X2*	X3*	X4*	X5*	X6*
1	-0.89399	0.24301	-1.14818	-0.87796	-0.48681	-1.06328
2	-0.89399	0.24301	-0.95489	-1.52116	-0.79887	-0.94228
3	-0.30841	0.4106	-0.95489	-0.23477	-0.17475	-0.29853
4	-0.30841	-0.09218	-0.76159	-0.55637	0.13731	-0.28487
5	-0.30841	0.24301	-0.76159	0.08683	-0.17475	-1.09388
6	-0.30841	0.07542	-1.53477	-0.87796	-0.48681	-1.38768
7	0.08198	0.24301	-1.14818	-1.84276	1.69759	-0.91664
8	-0.5036	0.57819	0.20489	1.05163	-0.79887	0.12097
9	-0.89399	0.24301	-0.76159	-1.19956	-1.11092	-0.4339
10	-0.11321	0.57819	-1.14818	-0.55637	0.76142	-1.17904
11	-0.5036	0.24301	-0.95489	-0.87796	0.13731	-0.26451
12	-0.30841	-0.09218	-0.56829	-0.23477	-0.17475	-0.61974
13	0.08198	0.74578	0.20489	1.37323	-1.11092	-0.6468
14	-0.6988	0.24301	0.39819	-0.87796	-1.73504	-1.83756
15	-0.89399	1.24856	0.78478	-0.55637	1.07348	-0.92135
16	-0.6988	0.91338	-0.95489	-0.23477	-0.17475	-1.15509
17	-0.30841	0.24301	-0.375	-1.19956	-0.17475	-0.20769
18	0.08198	0.91338	0.39819	0.08683	0.76142	-0.41367
19	-0.5036	1.24856	0.0116	0.40843	-0.79887	-0.2815
20	-0.5036	0.07542	0.39819	-1.19956	-0.48681	-0.69629
21	-0.11321	0.57819	0.39819	-0.55637	1.07348	-0.111
22	0.08198	0.24301	0.59149	-1.19956	1.38553	-1.23014
23	-0.6988	0.91338	0.0116	-0.55637	-0.48681	-0.81889
24	-0.30841	0.91338	-0.1817	0.08683	-0.79887	-0.87034
25	-0.6988	0.24301	-0.375	-0.87796	-1.11092	-0.87353
26	-0.5036	0.57819	-0.375	-0.71716	-0.48681	-0.87699
27	-0.5036	0.4106	-0.1817	-0.87796	-0.48681	-0.87353
28	-0.6988	0.57819	-0.76159	-0.55637	-1.11092	-0.49599
29	-0.5036	0.91338	-0.375	0.40843	-0.79887	0.09409
30	-0.89399	0.24301	-0.56829	0.08683	-0.79887	-0.353
31	-0.11321	0.4106	0.20489	0.08683	0.13731	-0.39132
32	-1.47958	-0.76254	-1.14818	0.73003	-0.48681	1.43109
33	-1.47958	-0.76254	-0.95489	0.08683	-0.17475	-0.51906

Lampiran 5. (Lanjutan)

34	-1.47958	-0.76254	-0.95489	-0.55637	0.44936	-0.93066
35	-1.47958	-0.42736	-0.76159	0.73003	0.44936	0.78122
36	-1.47958	-0.09218	-0.76159	0.08683	-1.11092	0.1899
37	-1.47958	-0.42736	-1.53477	0.40843	-0.79887	-0.36568
38	-1.28438	-1.09773	-1.14818	1.05163	0.13731	-0.52771
39	-1.08919	-2.43846	0.20489	0.40843	-0.17475	-1.01325
40	-1.28438	-1.7681	-0.76159	-0.87796	0.44936	0.74237
….	….	….	….	….	…..	…..
95	1.64354	0.57819	1.36467	1.05163	1.07348	1.58465
96	1.25315	0.91338	0.97808	0.73003	0.76142	1.06677
97	1.25315	0.57819	1.55797	1.37323	-0.48681	0.56673
98	1.25315	0.57819	1.75127	1.05163	-0.17475	0.81289
99	1.44834	0.91338	1.36467	0.40843	0.44936	0.93593
100	1.44834	0.91338	1.55797	2.01642	0.13731	1.22458

Lampiran 6. Hasil transformasi AKU terhadap data 1.

No	W_1	W_2	W_3	W_4	W_5	W_6
1	0.04153	1.09671	0.01481	-0.56215	-0.04833	0.2952
2	1.49577	2.52969	0.7723	-0.66586	-0.2856	0.54335
3	0.31259	0.16454	0.11871	-0.31338	-0.58628	-0.08116
4	1.34908	-0.53375	0.01278	0.52296	-0.42698	-0.78245
5	-4.86641	0.77631	-1.48155	0.4116	0.75391	0.7458
6	-1.21414	0.84682	-0.02706	-1.31995	0.67204	-0.16064
7	-1.0895	-0.06619	-0.49553	0.39623	0.00781	-0.56647
8	-0.50058	-0.93596	1.93283	-0.76185	-0.55454	0.13229
9	-0.373	-1.52829	0.17897	-0.34289	0.05588	-0.04219
10	-0.35359	-1.72426	0.52092	0.12374	0.56801	-0.1752
11	-0.94288	-0.37482	1.12068	0.21047	0.37198	0.36162
12	-0.34588	0.32889	-0.4593	0.03204	-0.49917	-0.56876
13	-0.18756	-1.0911	-0.29076	-0.54232	-0.02974	0.79815
14	0.74686	1.20541	0.79158	1.19355	1.23834	-0.20087
15	1.18293	-1.33048	-0.6196	-0.3748	-0.6732	0.62982
16	-1.84248	1.21909	1.18153	-0.04512	2.03398	-0.18431
17	0.29593	1.35937	-0.81438	0.62753	-0.22521	0.90572
18	-0.18469	1.67256	0.70672	-0.58595	-1.54434	0.10209
19	-3.41434	2.74614	-0.488	-0.50755	-1.40367	-0.48163
20	0.51032	-1.5462	0.9432	1.17961	-0.77433	0.79735
21	1.22959	0.51167	1.12691	0.24474	0.14888	0.31621
22	-0.45629	0.32313	-0.51453	1.12254	-0.05243	-0.18416
23	-0.09462	-0.2376	1.55561	0.75763	-0.43059	-0.26001
24	0.48382	0.59789	1.56315	0.07629	-1.08001	0.75225
25	0.38767	-0.22945	1.54891	-0.44339	0.31487	-0.74184
26	1.25016	0.46929	-0.2978	0.6757	0.6126	-0.34818
27	-0.03237	0.43503	0.90146	-0.47089	-1.09935	-0.12838
28	1.72564	-1.4204	-0.08819	0.60549	-0.05957	0.16153
29	0.59631	-0.44409	0.96862	0.16772	-0.52315	0.11877
30	-0.08079	1.5918	0.54776	0.94102	0.74309	0.61911
31	-0.724	0.42087	1.27436	0.56003	-0.68673	0.20573
32	0.23305	-0.85293	0.23278	1.57081	-0.71307	-0.0621
33	-1.41496	1.04655	0.62013	1.25778	0.31622	0.90779
34	-2.06005	0.93979	-0.48543	1.31986	0.27521	-0.17703
35	1.96748	-0.89454	0.03745	-0.05237	0.03608	-0.15173
36	0.54355	0.34131	-0.37417	0.2714	-0.71376	0.19193

Lampiran 6. (Lanjutan)

37	1.32149	0.53367	-0.1197	0.60809	0.44433	-0.47993
38	-1.94852	-0.99159	1.92443	1.46088	0.41128	-0.44847
39	-0.11378	-0.52556	1.60916	0.56736	-0.12991	-0.41627
40	0.73575	0.06602	-0.23273	-0.65815	-1.00017	-0.33224
41	-1.9557	-0.54224	2.20214	2.15577	-0.93973	-0.17138
42	0.68854	-1.18297	0.43298	-0.85422	-0.46706	0.39596
43	1.15895	-1.3487	0.66398	0.97522	-1.01412	0.43446
44	-3.21922	-1.03732	0.62402	0.59354	-0.87516	-0.40382
45	1.80464	-1.1161	2.32751	-1.32401	-0.27471	0.60085
….	….	….	….	….	….	….
195	0.78352	0.86728	-1.53969	-0.13291	-0.71926	-0.1263
196	2.19852	0.4249	-0.00867	0.60416	1.00774	0.12249
197	0.10035	-1.86551	-0.18732	-0.73296	-0.69667	-0.08913
198	1.30621	0.01784	0.92545	-1.02536	0.37551	-0.14057
199	-0.94579	0.8146	-0.29814	-0.36499	0.44163	-0.52456
200	2.32829	-0.90381	-0.92051	-0.30264	-0.09534	0.24485

Lampiran 7. Hasil transformasi AKU terhadap data 2.

No	W_1	W_2	W_3	W_4	W_5	W_6
1	1.86352	0.03059	-0.76896	-0.47102	-0.14422	-0.08573
2	2.05198	0.0122	-1.07568	-0.06906	0.26409	-0.40955
3	0.74963	0.05882	-0.41051	-0.60153	-0.36929	-0.34373
4	0.82624	0.50113	-0.23863	-0.14036	-0.22013	-0.1851
5	0.97112	-0.27206	-0.38201	-0.4839	-0.64107	0.40086
6	1.98574	-0.00562	-0.80528	-0.17399	-0.86125	-0.16155
7	0.99083	2.10019	-1.5294	0.2516	-0.81936	0.18874
8	-0.23227	-1.13373	0.29022	-0.9249	0.3244	0.06673
9	1.70188	-0.29153	-0.7274	-0.17386	0.52326	-0.69652
10	0.93178	0.72252	-1.07403	-0.50265	-0.99772	0.39116
11	1.01017	0.63687	-0.61797	-0.35429	-0.10228	-0.37443
12	0.87739	-0.00752	-0.16563	-0.13621	-0.26519	0.10054
13	-0.2043	-1.81788	0.02601	-0.71142	-0.35649	0.40355
14	1.81571	-1.568	-1.08052	0.49832	0.59966	0.57694
15	-0.06177	0.56325	-1.53969	-0.58234	0.68471	1.33609
16	1.1778	-0.19688	-1.01161	-0.97335	-0.424	0.27886
17	0.83853	0.35917	-0.82513	0.16556	0.29272	-0.42585
18	-0.59614	0.20019	-0.86217	-0.3584	-0.11551	0.69092
19	0.07615	-1.06925	-0.68292	-1.00434	0.2692	-0.01567
20	0.94816	-0.13441	-0.86625	0.55707	0.75155	0.18977
21	-0.41073	0.87746	-0.85612	-0.05627	0.24547	0.49486
22	0.18091	1.10991	-1.38401	0.72074	-0.03754	1.25959
23	0.77937	-0.46964	-1.12729	-0.44143	0.42339	0.25874
24	0.58208	-1.00008	-0.81384	-0.56747	-0.09003	0.1848
25	1.52044	-0.64208	-0.7504	-0.00747	0.37449	-0.16916
26	1.06532	-0.28969	-0.98301	-0.22675	0.09306	0.06134
27	1.08381	-0.23671	-0.95801	0.01602	0.25393	0.11354
28	1.30051	-0.6546	-0.6749	-0.5427	0.19146	-0.53464
29	0.18606	-0.82002	-0.281	-1.02865	0.16992	-0.41151
30	0.998	-0.56713	-0.0301	-0.7514	0.22407	-0.13491
31	-0.06474	-0.13187	-0.39566	-0.20557	0.0072	0.42718
32	0.67055	0.28934	1.76502	-1.33738	0.62463	-1.03898

Lampiran 7. (Lanjutan)

33	1.64156	0.27909	0.7422	-0.73906	0.17045	0.30197
34	1.87059	0.94352	0.21551	-0.45264	0.09931	0.63991
35	0.38671	0.74067	1.2136	-1.38449	0.4521	-0.06932
36	1.30623	-0.49662	0.51651	-1.00472	0.72537	-0.54783
37	1.82286	-0.25434	0.79834	-1.25159	-0.09265	-0.25714
38	1.29763	0.23904	1.48543	-1.02174	-0.42226	0.63754
39	1.55739	0.04733	1.87613	0.74726	0.3074	1.34088
40	1.32787	1.60802	1.29852	0.20541	0.77316	-0.4523
….	….	….	…..	…..	….	….
95	-3.07013	0.18766	0.25732	0.31248	-0.15658	-0.13407
96	-2.34765	0.00905	-0.23591	0.01937	-0.16701	-0.1285
97	-2.10932	-1.45566	0.24323	0.33367	0.02784	0.19078
98	-2.30147	-1.05866	0.11358	0.49599	0.26977	0.14031
99	-2.33839	-0.26922	-0.49049	0.48597	0.04585	-0.13995
100	-3.05623	-1.1323	0.49078	-0.19941	-0.15945	0.12217

Lampiran 8. Prosedur metode PCLR data 1.

Binary Logistic Regression: Y versus W1, W2, W3

Link Function: Logit

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	95% CI Upper
Constant	-1.53325	0.226987	-6.75	0.000			
W1	-1.07134	0.172715	-6.20	0.000	0.34	0.24	0.48
W2	0.0671097	0.161690	0.42	0.678	1.07	0.78	1.47
W3	0.181730	0.216131	0.84	0.400	1.20	0.79	1.83

Log-Likelihood = -79.101
Test that all slopes are zero: G = 66.733, DF = 3, P-Value = 0.000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	196.033	196	0.486
Deviance	158.201	196	0.978

Dependent Variable: Y
Method: ML - Binary Logit
Date: 05/23/08 Time: 10:30
Sample: 1 200
Included observations: 200
Convergence achieved after 5 iterations
Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-1.533253	0.226988	-6.754771	0.0000
W1	-1.071341	0.172717	-6.202873	0.0000
W2	0.067109	0.161691	0.415047	0.6781
W3	0.181730	0.216132	0.840832	0.4004

Mean dependent var	0.250000	S.D. dependent var	0.434099
S.E. of regression	0.355303	Akaike info criterion	0.831007
Sum squared resid	24.74304	Schwarz criterion	0.896973
Log likelihood	-79.10068	Hannan-Quinn criter.	0.857702
Restr. log likelihood	-200.3112	Avg. log likelihood	-0.395503
LR statistic (3 df)	66.73269	McFadden R-squared	0.605111
Probability(LR stat)	2.13E-14		

Obs with Dep=0	150	Total obs	200
Obs with Dep=1	50		

Lampiran 9. Prosedur metode PCLR data 2.

Binary Logistic Regression: Y versus W1, W2, W3

```
Link Function: Logit

Logistic Regression Table

                                                 Odds     95% CI
Predictor  Coef     SE Coef     Z      P     Ratio  Lower Upper
Constant 1.23935  0.243850  5.08  0.000
W1        0.19238  0.132703  1.45  0.147 1.21   0.93  1.57
W2      0.067784  0.246234  0.28  0.783 1.07   0.66  1.73
W3     -0.033881  0.268453 -0.13  0.900  0.97  0.57  1.64

Log-Likelihood = -52.817
Test that all slopes are zero: G = 2.221, DF = 3, P-Value =
0.528
Goodness-of-Fit Tests

Method          Chi-Square  DF      P
Pearson            100.184  96  0.365
Deviance           105.635  96  0.235
```

Dependent Variable: Y
Method: ML - Binary Logit
Date: 05/24/08   Time: 09:03
Sample: 1 100
Included observations: 100
Convergence achieved after 3 iterations
Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z Statistic	Prob.
C	1.239352	0.243847	5.082492	0.0000
W1	0.192380	0.132701	1.449716	0.1471
W2	0.067785	0.246233	0.275287	0.7831
W3	-0.033881	0.268450	-0.126209	0.8996

Mean dependent var	0.770000	S.D. dependent var	0.422953
S.E. of regression	0.424347	Akaike info criterion	1.136348
Sum squared resid	17.28674	Schwarz criterion	1.240555
Log likelihood	-52.81739	Hannan-Quinn criter.	1.178522
Restr. log likelihood	-137.4149	Avg. log likelihood	-0.528174
LR statistic (3 df)	2.220497	McFadden R-squared	0.615635
Probability(LR stat)	0.527922		
Obs with Dep=0	23	Total obs	100
Obs with Dep=1	77		

Lampiran 10. Prosedur metode $PCLR_{(S)}$ data 1.

Binary Logistic Regression: Y versus W1

Link Function: Logit

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	95% CI Upper
Constant	-1.50766	0.222426	-6.78	0.000			
W1	-1.06343	0.171436	-6.20	0.000	0.35	0.25	0.48

Log-Likelihood = -79.560
Test that all slopes are zero: G = 65.814, DF = 1, P-Value = 0.000

Binary Logistic Regression: Y versus W2

Link Function: Logit

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	95% CI Upper
Constant	-1.09897	0.163350	-6.73	0.000			
W2	0.030983	0.134179	-1.99	0.044	1.03	0.79	1.34

Log-Likelihood = -112.440
Test that all slopes are zero: G = 0.053, DF = 1, P-Value = 0.817

Binary Logistic Regression: Y versus W3

Link Function: Logit

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	95% CI Upper
Constant	-1.10057	0.163579	-6.73	0.000			
W3	0.093466	0.173257	-2.03	0.023	1.10	0.78	1.54

Log-Likelihood = -112.321
Test that all slopes are zero: G = 0.292, DF = 1, P-Value = 0.589

Lampiran 10. (Lanjutan)

Binary Logistic Regression: Y versus W4

```
Link Function: Logit

Logistic Regression Table
                                                  Odds   95% CI
Predictor  Coef      SE Coef    Z      P    Ratio Lower Upper
Constant -1.09870  0.163311 -6.73  0.000
W4       -0.02474  0.221704 -4.01  0.000  0.98  0.63 1.51

Log-Likelihood = -112.461
Test that all slopes are zero: G = 0.012, DF = 1, P-Value =
0.911
```

Binary Logistic Regression: Y versus W5

```
Link Function: Logit

Logistic Regression Table
                                                  Odds     95% CI
Predictor   Coef    SE Coef    Z     P    Ratio  Lower  Upper
Constant  -1.11405 0.165486 -6.73 0.000
W5        0.360605 0.237884 -5.81 0.000  1.43   0.90   2.29

Log-Likelihood = -111.302
Test that all slopes are zero: G = 2.330, DF = 1, P-Value =
0.127
```

Binary Logistic Regression: Y versus W6

```
Link Function: Logit

Logistic Regression Table
                                                  Odds     95% CI
Predictor   Coef     SE Coef    Z      P    Ratio  Lower  Upper
Constant  -1.11907 0.166157 -6.74  0.000
W6        0.678509 0.377412 -5.21  0.000  1.97  0.94   4.13

Log-Likelihood = -110.844
Test that all slopes are zero: G = 3.245, DF = 1, P-Value =
0.072
```

Lampiran 10. (Lanjutan)

Binary Logistic Regression: Y versus W1, W2

Link Function: Logit

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	Upper
Constant	-1.51019	0.222848	-6.78	0.000			
W1	-1.07025	0.173293	-6.18	0.000	0.34	0.24	0.48
W2	0.073169	0.161128	0.45	0.650	1.08	0.78	1.48

Log-Likelihood = -79.457

Test that all slopes are zero: G = 66.021, DF = 2, P-Value = 0.000

Binary Logistic Regression: Y versus W1, W3

Link Function: Logit

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	Upper
Constant	-1.53168	0.226752	-6.75	0.000			
W1	-1.06635	0.171350	-6.22	0.000	0.34	0.25	0.48
W3	0.18548	0.215578	0.86	0.390	1.20	0.79	1.84

Log-Likelihood = -79.187

Test that all slopes are zero: G = 66.560, DF = 2, P-Value = 0.000

Binary Logistic Regression: Y versus W1, W4

Link Function: Logit

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	Upper
Constant	-1.51285	0.223847	-6.76	0.000			
W1	-1.06478	0.171515	-6.21	0.000	0.34	0.25	0.48
W4	-0.07315	0.247967	-0.30	0.768	0.93	0.57	1.51

Log-Likelihood = -79.517

Test that all slopes are zero: G = 65.900, DF = 2, P-Value = 0.000

Lampiran 10. (Lanjutan)

Binary Logistic Regression: Y versus W1, W5

Link Function: Logit

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	95% CI Upper
Constant	-1.51089	0.223683	-6.75	0.000			
W1	-1.06009	0.169941	-6.24	0.000	0.35	0.25	0.48
W5	0.38607	0.272342	1.42	0.156	1.47	0.86	2.51

Log-Likelihood = -78.543
Test that all slopes are zero: G = 67.848, DF = 2, P-Value = 0.000

Binary Logistic Regression: Y versus W1, W6

Link Function: Logit

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	95% CI Upper
Constant	-1.48241	0.221641	-6.69	0.000			
W1	-1.09139	0.178528	-6.11	0.000	0.34	0.24	0.48
W6	0.84692	0.498378	1.70	0.089	2.33	0.88	6.20

Log-Likelihood = -78.039
Test that all slopes are zero: G = 68.856, DF = 2, P-Value = 0.000

Binary Logistic Regression: Y versus W1, W2, W6

Link Function: Logit

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	95% CI Upper
Constant	-1.48475	0.222073	-6.69	0.000			
W1	-1.09554	0.179446	-6.11	0.000	0.33	0.24	0.48
W2	0.065306	0.163394	0.40	0.689	1.07	0.77	1.47
W6	0.837855	0.496990	1.69	0.092	2.31	0.87	6.12

Log-Likelihood = -77.959
Test that all slopes are zero: G = 69.016, DF = 3, P-Value = 0.000

Binary Logistic Regression: Y versus W1, W3, W6

Link Function: Logit

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	95% CI Upper
Constant	-1.50508	0.225858	-6.66	0.000			
W1	-1.08933	0.177943	-6.12	0.000	0.34	0.24	0.48
W3	0.163761	0.216492	0.76	0.449	1.18	0.77	1.80
W6	0.819465	0.496624	1.65	0.099	2.27	0.86	6.01

Log-Likelihood = -77.751
Test that all slopes are zero: G = 69.432, DF = 3, P-Value = 0.000

Binary Logistic Regression: Y versus W1, W4, W6

Link Function: Logit

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	95% CI Upper
Constant	-1.48886	0.223220	-6.67	0.000			
W1	-1.09395	0.178730	-6.12	0.000	0.33	0.24	0.48
W4	-0.09135	0.249793	-0.37	0.715	0.91	0.56	1.49
W6	0.855025	0.499378	1.71	0.087	2.35	0.88	6.26

Log-Likelihood = -77.972
Test that all slopes are zero: G = 68.990, DF = 3, P-Value = 0.000

Binary Logistic Regression: Y versus W1, W5, W6

Link Function: Logit

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	95% CI Upper
Constant	-1.47819	0.222734	-6.64	0.000			
W1	-1.08931	0.176841	-6.16	0.000	0.34	0.24	0.48
W5	0.395269	0.277340	1.43	0.154	1.48	0.86	2.56
W6	0.863340	0.505853	1.71	0.088	2.37	0.88	6.39

Log-Likelihood = -77.008
Test that all slopes are zero: G = 70.918, DF = 3, P-Value = 0.000

Lampiran 10. (Lanjutan)

Binary Logistic Regression: Y versus W1, W2, W5, W6

Link Function: Logit

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	Upper
Constant	-1.48099	0.223189	-6.64	0.000			
W1	-1.09264	0.177590	-6.15	0.000	0.34	0.24	0.47
W2	0.066907	0.164266	0.41	0.684	1.07	0.77	1.48
W5	0.396647	0.278016	1.43	0.154	1.49	0.86	2.56
W6	0.851594	0.504587	1.69	0.091	2.34	0.87	6.30

Log-Likelihood = -76.925
Test that all slopes are zero: G = 71.084, DF = 4, P-Value = 0.000

Binary Logistic Regression: Y versus W1, W3, W5, W6

Link Function: Logit

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	Upper
Constant	-1.50457	0.227589	-6.61	0.000			
W1	-1.09043	0.176620	-6.17	0.000	0.34	0.24	0.48
W3	0.178236	0.218018	0.82	0.414	1.20	0.78	1.83
W5	0.401316	0.274920	1.46	0.144	1.49	0.87	2.56
W6	0.838854	0.504079	1.66	0.096	2.31	0.86	6.21

Log-Likelihood = -76.671
Test that all slopes are zero: G = 71.591, DF = 4, P-Value = 0.000

Binary Logistic Regression: Y versus W1, W4, W5, W6

Link Function: Logit

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	Upper
Constant	-1.48591	0.224533	-6.62	0.000			
W1	-1.09284	0.177214	-6.17	0.000	0.34	0.24	0.47
W4	-0.103912	0.251880	-0.41	0.680	0.90	0.55	1.48
W5	0.399517	0.277971	1.44	0.151	1.49	0.86	2.57
W6	0.871552	0.506658	1.72	0.085	2.39	0.89	6.45

Log-Likelihood = -76.923
Test that all slopes are zero: G = 71.088, DF = 4, P-Value = 0.000

Lampiran 11. Prosedur metode $PCLR_{(S)}$ data 2.

Binary Logistic Regression: Y versus W1

Link Function: Logit

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	Upper
Constant	1.23859	0.243698	5.08	0.000			
W1	0.19388	0.132641	1.98	0.048	1.21	0.94	1.57

Log-Likelihood = -52.864
Test that all slopes are zero: G = 2.128, DF = 1, P-Value = 0.145

Binary Logistic Regression: Y versus W2

Link Function: Logit

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	Upper
Constant	1.20954	0.237878	5.08	0.000			
W2	0.071143	0.251331	2.01	0.044	1.07	0.66	1.76

Log-Likelihood = -53.887
Test that all slopes are zero: G = 0.080, DF = 1, P-Value = 0.777

Binary Logistic Regression: Y versus W3

Link Function: Logit

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	Upper
Constant	1.20885	0.237735	5.08	0.000			
W3	-0.048948	0.258840	-0.19	0.850	0.95	0.57	1.58

Log-Likelihood = -53.910
Test that all slopes are zero: G = 0.036, DF = 1, P-Value = 0.850

Lampiran 11. (Lanjutan)

Binary Logistic Regression: Y versus W4

Link Function: Logit

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	Upper
Constant	1.21588	0.239169	5.08	0.000			
W4	-0.21364	0.300827	-0.71	0.478	0.81	0.45	1.46

Log-Likelihood = -53.675
Test that all slopes are zero: G = 0.504, DF = 1, P-Value = 0.478

Binary Logistic Regression: Y versus W5

Link Function: Logit

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	Upper
Constant	1.20836	0.237635	5.08	0.000			
W5	-0.025466	0.438523	2.03	0.042	0.97	0.41	2.30

Log-Likelihood = -53.926
Test that all slopes are zero: G = 0.003, DF = 1, P-Value = 0.954

Binary Logistic Regression: Y versus W6

Link Function: Logit

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	Upper
Constant	1.27194	0.250005	5.09	0.000			
W6	-1.05349	0.503008	-2.09	0.036	0.35	0.13	0.93

Log-Likelihood = -51.623
Test that all slopes are zero: G = 4.609, DF = 1, P-Value = 0.032

Binary Logistic Regression: Y versus W1, W6

Link Function: Logit

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	95% CI Upper
Constant	1.30905	0.257498	5.08	0.000			
W1	0.20590	0.135256	1.52	0.128	1.23	0.94	1.60
W6	-1.10912	0.524625	-2.11	0.035	0.33	0.12	0.92

Log-Likelihood = -50.466
Test that all slopes are zero: G = 6.923, DF = 2, P-Value = 0.031

Binary Logistic Regression: Y versus W2, W6

Link Function: Logit
Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	95% CI Upper
Constant	1.27912	0.251925	5.08	0.000			
W2	0.108886	0.256386	0.42	0.671	1.12	0.67	1.84
W6	-1.07833	0.510043	-2.11	0.034	0.34	0.13	0.92

Log-Likelihood = -51.533
Test that all slopes are zero: G = 4.790, DF = 2, P-Value = 0.091

Binary Logistic Regression: Y versus W3, W6

Link Function: Logit
Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	95% CI Upper
Constant	1.27208	0.250042	5.09	0.000			
W3	-0.04113	0.266072	-0.15	0.877	0.96	0.57	1.62
W6	-1.05137	0.502877	-2.09	0.037	0.35	0.13	0.94

Log-Likelihood = -51.611
Test that all slopes are zero: G = 4.632, DF = 2, P-Value = 0.099

Binary Logistic Regression: Y versus W4, W6

Link Function: Logit

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	Upper
Constant	1.28507	0.252815	5.08	0.000			
W4	-0.25892	0.324767	-0.80	0.425	0.77	0.41	1.46
W6	-1.07950	0.509556	-2.12	0.034	0.34	0.13	0.92

Log-Likelihood = -51.305
Test that all slopes are zero: G = 5.245, DF = 2, P-Value = 0.073

Binary Logistic Regression: Y versus W5, W6

Link Function: Logit

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	Upper
Constant	1.27211	0.250098	5.09	0.000			
W5	0.01366	0.465595	0.03	0.977	1.01	0.41	2.52
W6	-1.05464	0.504805	-2.09	0.037	0.35	0.13	0.94

Log-Likelihood = -51.623
Test that all slopes are zero: G = 4.609, DF = 2, P-Value = 0.100

Binary Logistic Regression: Y versus W1, W2, W6

Link Function: Logit

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	Upper
Constant	1.31325	0.258678	5.08	0.000			
W1	0.20308	0.135386	1.50	0.134	1.23	0.94	1.60
W2	0.08513	0.254854	0.33	0.738	1.09	0.66	1.79
W6	-1.11832	0.525695	-2.13	0.033	0.33	0.12	0.92

Log-Likelihood = -50.410
Test that all slopes are zero: G = 7.035, DF = 3, P-Value = 0.071

Binary Logistic Regression: Y versus W1, W3, W6

Link Function: Logit

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	Upper
Constant	1.30870	0.257529	5.08	0.000			
W1	0.20513	0.135862	1.51	0.131	1.23	0.94	1.60
W3	-0.01440	0.275261	-0.05	0.958	0.99	0.57	1.69
W6	-1.10763	0.525092	-2.11	0.035	0.33	0.12	0.92

Log-Likelihood = -50.465
Test that all slopes are zero: G = 6.926, DF = 3, P-Value = 0.074

Binary Logistic Regression: Y versus W1, W4, W6

Link Function: Logit

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	Upper
Constant	1.32138	0.259986	5.08	0.000			
W1	0.20546	0.135247	1.52	0.129	1.23	0.94	1.60
W4	-0.26280	0.331953	-0.79	0.429	0.77	0.40	1.47
W6	-1.13011	0.528064	-2.14	0.032	0.32	0.11	0.91

Log-Likelihood = -50.152
Test that all slopes are zero: G = 7.551, DF = 3, P-Value = 0.056

Binary Logistic Regression: Y versus W1, W5, W6

Link Function: Logit

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	Upper
Constant	1.30922	0.257604	5.08	0.000			
W1	0.20586	0.135242	1.52	0.128	1.23	0.94	1.60
W5	0.01311	0.470863	0.03	0.978	1.01	0.40	2.55
W6	-1.11042	0.527019	-2.11	0.035	0.33	0.12	0.93

Log-Likelihood = -50.466
Test that all slopes are zero: G = 6.924, DF = 3, P-Value = 0.074

Lampiran 12. Nilai PCP

Nilai PCP data 1 metode PCLR

Classification Table[a,b]

		Predicted		
		Y		Percentage Correct
		.00	1.00	
	.00	150	0	100.0
	1.00	50	0	.0
				75.0

Nilai PCP data 1 metode PCLR$_{(S)}$

Classification Table[a]

		Predicted		
		Y		Percentage Correct
		.00	1.00	
	.00	144	6	96.0
	1.00	25	25	50.0
				84.5

Nilai PCP data 2 metode PCLR

Classification Table[a]

		Predicted		
		Y		Percentage Correct
		.00	1.00	
	.00	0	23	.0
	1.00	0	77	100.0
				77.0

Nilai PCP data 2 metode PCLR$_{(S)}$

Classification Table[a]

		Predicted		
		Y		Percentage Correct
		.00	1.00	
	.00	2	21	8.7
	1.00	1	76	98.7
				78.0

Lampiran 13. Nilai *Pearson Residuals* data 1.

No	Metode	
	PCLR	$PCLR_{(S)}$
1	2.11854	1.90377
2	-0.24346	-0.25267
3	-0.39941	-0.34637
4	-0.22179	-0.15016
5	0.177	0.09233
6	-0.91366	-0.98576
7	-0.79433	-0.67795
8	1.42513	1.68026
9	-0.5478	-0.58094
10	-0.55558	-0.60057
11	1.18803	0.99594
12	-0.54224	-0.40866
13	-0.4823	-0.7421
14	-0.34843	-0.37238
15	-0.22286	-0.28807
16	0.6917	0.55608
17	-0.38538	-0.57475
18	-0.57848	-0.40671
19	0.32951	0.52982
20	-0.36562	-0.43785
21	3.69033	3.46555
22	-0.57229	-0.55965
23	-0.55847	-0.41276
24	-0.4216	-0.41011
25	2.31925	3.34759
26	-0.23513	-0.23475
…	…	…
197	-0.40657	-0.37912
198	-0.25117	-0.23764
199	1.2966	1.43781
200	-0.1191	-0.14656
Jumlah	3.66378	3.00851

Lampiran 14. Nilai *Pearson Residuals* data 2.

No	Metode	
	PCLR	$PCLR_{(S)}$
1	0.44353	0.40905
2	0.43357	0.33524
3	0.49622	0.3976
4	0.48667	0.43075
5	0.49148	0.58731
6	0.43861	0.3873
7	0.44394	0.52108
8	0.57465	0.55233
9	0.45575	0.29641
10	0.47143	0.58653
11	-2.11467	-2.62784
12	0.49331	0.50203
13	0.58393	0.66385
14	0.4679	0.59363
15	-1.93268	-0.9114
16	0.47548	0.53733
17	0.48361	0.37644
18	0.55782	0.81059
19	0.54753	0.51117
20	-2.05649	-1.90961
21	0.53558	0.71333
22	0.49751	1.02569
23	-2.00937	-1.80628
24	-1.92622	-1.84405
25	-2.13161	-2.4716
26	0.4824	0.48182
…	…	…
97	0.69536	0.7178
98	-1.43404	-1.40463
99	-1.4828	-1.6346
100	0.75653	0.76175
Jumlah	0.14132	0.03534