



REGRESI KONTINUM KEKAR DALAM PENGUKURAN KADAR GLUKOSA DARAH NON-INVASIF

NURMAI SYAROH DAULAY



**PROGRAM STUDI STATISTIKA DAN SAINS DATA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT PERTANIAN BOGOR
BOGOR
2024**

- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

PERNYATAAN MENGENAI TESIS DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa tesis dengan judul “Regresi Kontinum Kekar dalam Pengukuran Kadar Glukosa Darah Non-Invasif” adalah karya saya dengan arahan dari dosen pembimbing dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir tesis ini.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Institut Pertanian Bogor.

Bogor, Agustus 2024

Nurmai Syaroh Daulay
NIM G1501201013

- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



RINGKASAN

NURMAI SYAROH DAULAY. Regresi Kontinum Kekar dalam Pengukuran Kadar Glukosa Darah Non-Invasif. Dibimbing oleh ERFIANI dan AGUS MOHAMAD SOLEH.

Regresi Kontinum Kekar (RKK) diperkenalkan untuk mengatasi masalah multikolinearitas dengan cara memproyeksikan data asli ke dalam peubah baru (*latent*) menggunakan matriks pembobot. RKK menerapkan teknik regresi kekar (*robust*) untuk mengatasi masalah pencilan. Beberapa penelitian khususnya bidang kalibrasi sering ditemukan bahwa jumlah peubah penjelas melebihi jumlah amatan, dikenal dengan data berdimensi tinggi ($p \gg n$). Kondisi ini dapat menyebabkan struktur matriks peubah penjelas mengalami singularitas dan akan mengalami kendala dalam aspek komputasi. Tahap awal perlu dilakukan pereduksian dimensi dengan mempertahankan sebagian besar informasi yang relevan dari data asli, proses ini disebut prapemrosesan. Analisis Komponen Utama (AKU) merupakan metode prapemrosesan yang dapat mereduksi dimensi data peubah penjelas (X) dan mengidentifikasi pola penting tanpa kehilangan banyak informasi. Kelemahan metode ini sulit untuk menginterpretasi hasil dan tidak secara khusus melakukan seleksi peubah. Metode *Least Absolut Shrinkage and Selection Operator* (LASSO) dapat mengatasinya tetapi sensitif terhadap keberadaan pencilan. WLAD LASSO diperkenalkan untuk mengatasi pencilan pada peubah respon dan pencilan pada peubah penjelas dengan memberikan bobot pada nilai absolut dari sisaan regresi.

Penelitian ini bertujuan untuk mengkaji kinerja metode RKK, Regresi Kontinum Kekar seleksi LASSO (RKK-LASSO) dan Regresi Kontinum Kekar seleksi WLAD LASSO (RKK-WLAD LASSO) melalui kajian simulasi. Simulasi dilakukan pada data berdimensi rendah ($n = 75, p = 25$) dan data berdimensi tinggi ($n = 75, p = 100$), tanpa pencilan dan dengan pencilan ($outlier = \{10\%, 15\%, 20\%\}$) dan tingkat korelasi ($\rho = \{0,1; 0,5; 0,9\}$). Nilai δ optimal pada RKK antara ($0 < \delta < 1$) dalam penelitian ini adalah $\delta = 0,3$. Penduga *Minimum Regularized Covariance Determinant* (MRCD) digunakan untuk menentukan bobot pada WLAD LASSO karena penduga ini bisa digunakan untuk data berdimensi tinggi dan data berdimensi rendah. Selanjutnya metode diterapkan dalam pemodelan kalibrasi data pengukuran kadar glukosa darah non-invasif tahun 2019 modulasi 50 sampai 90. Data akan distandarisasi dan dibagi menjadi 80% data *training* dan 20% data *testing*.

Hasil kajian simulasi menunjukkan bahwa pada data berdimensi rendah, rata-rata tingkat kesalahan RKK-LASSO dan RKK lebih rendah dibandingkan dengan RKK-WLAD LASSO. Korelasi antar peubah penjelas pada $\rho = 0,9$ menghasilkan nilai RMSE yang paling kecil dan nilai R^2 yang paling besar dalam setiap skenario tanpa pencilan dan dengan pencilan 10%, 15%. Performa RKK-LASSO, RKK dan RKK-WLAD LASSO cenderung menurun saat pencilan meningkat (RMSE pada pencilan 10%, 15%, 20% cenderung lebih tinggi). Hasil simulasi pada data berdimensi tinggi, rata-rata tingkat kesalahan metode RKK-LASSO cenderung lebih rendah daripada metode RKK-WLAD LASSO. Korelasi antar peubah penjelas pada $\rho = 0,9$ menghasilkan nilai RMSE yang paling kecil dan nilai R^2 yang paling besar dalam setiap skenario. Penggunaan RKK pada data berdimensi tinggi terbatas karena pendugaan parameter mengikuti metode kuadrat terkecil yang

@Hak Cipta Institut Pertanian Bogor



kurang efektif untuk jumlah peubah penjelas yang besar. Metode prapemrosesan, seperti seleksi peubah menggunakan metode LASSO dan WLAD LASSO perlu dilakukan.

Hasil kajian aktual data pengukuran kadar glukosa darah non-invasif menunjukkan adanya pencilan pada peubah respon sekitar 10,81% dan peubah penjelas sekitar 60,81%. Data ini mengandung multikolinearitas dan merupakan data berberdimensi rendah. Seleksi peubah dengan validasi silang, diperoleh λ optimum sebesar 0,1 dengan jumlah peubah terpilih sebanyak 3 pada LASSO dan 2 pada WLAD LASSO. Hasil menunjukkan metode RKK memiliki rata-rata RMSE terendah sebesar 0,7849 namun rata-rata RMSEP tertinggi sebesar 1,2269. Hasil ini mengindikasikan bahwa metode RKK sangat baik dalam menyesuaikan data pelatihan, tetapi kurang efektif dalam memprediksi data baru. RKK-LASSO memiliki nilai rata-rata RMSEP yang lebih rendah yaitu 0,8928 dibandingkan dengan RKK-WLAD LASSO dan RKK. Metode RKK-LASSO dengan performa terbaik dengan nilai RMSEP terendah dan memiliki pola y duga yang lebih mirip dengan pola y aktual dibandingkan dengan model pada metode RKK-WLAD LASSO dan RKK.

Kata kunci: LASSO, model kalibrasi, regresi kontinum kekar, WLAD LASSO

- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



SUMMARY

NURMAI SYAROH DAULAY. Robust Continuum Regression in Non-Invasive Measurement of Blood Glucose Levels. Supervised by ERFIANI and AGUS MOHAMAD SOLEH.

Robust Continuum Regression (RCR) was introduced to overcome the multicollinearity problem by projecting the original data into new variables (latent) using a weighting matrix. RCR applies the robust regression technique to overcome the problem of outliers. Several studies, especially in the field of calibration, often find that the number of explanatory variables exceeds the number of observations, known as high-dimensional data ($p \gg n$). This condition can cause the structure of the explanatory variable matrix to experience singularity and will experience constraints in the computational aspect. The initial stage needs to be done by reducing the dimensionality by retaining most of the relevant information from the original data; this process is called preprocessing. Principal Component Analysis (PCA) is a preprocessing method that can reduce the dimensionality of explanatory variable data (X) and identify essential patterns without losing much information. The weakness of this method is that it is difficult to interpret the results and does not explicitly select variables. The Least Absolute Shrinkage and Selection Operator (LASSO) method can overcome this but is sensitive to the presence of outliers. WLAD LASSO was introduced to overcome outliers in response variables and outliers in explanatory variables by giving weight to the absolute value of the regression residuals.

This study aims to examine the performance of the RCR method, LASSO-selected Robust Continuum Regression (RCR-LASSO), and WLAD LASSO-selected Robust Continuum Regression (RCR-WLAD LASSO). Simulations were performed on low-dimensional data ($n=75, p=25$) and high-dimensional data ($n=75, p=100$), without outliers and with outliers (outliers = {10%, 15%, 20%}) and correlation levels ($\rho = \{0,1; 0,5; 0,9\}$). The optimal δ value in RCR between ($0 < \delta < 1$) in this study is $\delta = 0,3$. The Minimum Regularized Covariance Determinant (MRCD) estimator is used to determine the weights in WLAD LASSO because this estimator can be used for high- and low-dimensional data. Furthermore, the method is applied in the calibration modeling of non-invasive blood glucose level measurement data in 2019 with a modulation of 50 to 90. The data will be standardized and divided into 80% training and 20% testing data.

The simulation study results show that in low-dimensional data, the average error rate of RCR-LASSO and RCR is lower than that of RCR-WLAD LASSO. The correlation between explanatory variables at $\rho = 0,9$ produces the smallest RMSE value and the most considerable R^2 value in each scenario without outliers and with 10% and 15% outliers. The performance of RCR-LASSO, RCR, and RCR-WLAD LASSO tends to decrease as outliers increase (RMSE at outliers of 10%, 15%, and 20% tends to be higher). The simulation results on high-dimensional data, the average error rate of the RCR-LASSO method tends to be lower than the RCR-WLAD LASSO method. The correlation between explanatory variables at $\rho = 0,9$ produces the smallest RMSE value and the most considerable R^2 value in each scenario. The use of RCR on high-dimensional data is limited because parameter estimation follows the least squares method, which is less effective for many

explanatory variables. Preprocessing methods, such as variable selection using the LASSO and WLAD LASSO methods, must be carried out.

The results of the actual study of non-invasive blood glucose measurement data showed outliers in the response variables of around 10,81% and explanatory variables of around 60,81%. This data contains multicollinearity and is low-dimensional data. Variable selection with cross-validation obtained an optimum λ of 0,1 with the number of selected variables of 3 in LASSO and 2 in WLAD LASSO. The results show that the RCR method has the lowest average RMSE of 0,7849 but the highest average RMSEP of 1,2269. These results indicate that the RCR method is very good at adjusting training data but less effective in predicting new data. RCR-LASSO has a lower average RMSEP value of 0,8928 compared to RCR-WLAD LASSO and RCR. The RCR-LASSO method has the best performance with the lowest RMSEP value and has a predicted y pattern that is more similar to the actual y pattern compared to the models in the RCR-WLAD LASSO and RCR methods.

Keywords: LASSO, calibration model, robust continuum regression, WLAD LASSO

@Hak cipta milik IPB University

IPB University





@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

© Hak Cipta milik IPB, tahun 2024
Hak Cipta dilindungi Undang-Undang

Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan atau menyebutkan sumbernya. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik, atau tinjauan suatu masalah, dan pengutipan tersebut tidak merugikan kepentingan IPB.

Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apa pun tanpa izin IPB.

REGRESI KONTINUM KEKAR DALAM PENGUKURAN KADAR GLUKOSA DARAH NON-INVASIF

NURMAI SYAROH DAULAY

Tesis
sebagai salah satu syarat untuk memperoleh gelar
Magister pada
Program Studi Statistika dan Sains Data

**PROGRAM STUDI STATISTIKA DAN SAINS DATA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT PERTANIAN BOGOR
BOGOR
2024**



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

Penguji Luar Komisi Pembimbing pada Ujian Tesis:
Dr. Ir. Indahwati, M.Si.



Judul Tesis : Regresi Kontinum Kekar dalam Pengukuran Kadar Glukosa
Darah Non-Invasif
Nama : Nurmai Syaroh Dauly
NIM : G1501201013

Disetujui oleh

Pembimbing 1:
Dr. Ir. Erfiani, M.Si.



Pembimbing 2:
Dr. Agus Mohamad Soleh, S.Si., M.T.



Diketahui oleh

Ketua Program Studi:
Dr. Agus Mohamad Soleh, S.Si., M.T.
NIP 197503151999031004



Dekan Fakultas Matematika dan Ilmu
Pengetahuan Alam:
Dr. Berry Juliandi, S.Si., M.Si.
NIP 197807232007011001



Tanggal Ujian: 05 Agustus 2024

Tanggal Lulus:



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



PRAKATA

Puji dan syukur penulis panjatkan kepada Allah subhanaahu wa ta'ala atas segala karunia-Nya sehingga karya ilmiah ini berhasil diselesaikan. Tema yang dipilih dalam penelitian yang dilaksanakan sejak Februari 2022 sampai Agustus 2024, dengan judul “Regresi Kontinum Kekar dalam Pengukuran Kadar Glukosa Darah Non-Invasif”.

Terima kasih penulis ucapkan kepada Ibu Dr. Ir. Erfiani, M.Si. dan Bapak Dr. Agus Mohamad Soleh, S.Si., M.T. selaku dosen komisi pembimbing yang telah membimbing dan banyak memberi saran, masukan dan arahan dalam penulisan tesis. Ucapan terima kasih juga disampaikan kepada moderator seminar, dan penguji luar komisi pembimbing Ibu Dr. Ir. Indahwati, M.Si. Selanjutnya ucapan terima kasih juga disampaikan kepada seluruh Dosen pengajar, staf dan karyawan sekolah Pascasarjana Departemen Statistika dan Sains Data IPB yang telah banyak memberikan ilmu selama perkuliahan sampai dengan penyusunan tesis ini. Ungkapan terima kasih juga disampaikan kepada orangtua tercinta Ibu Hj. Masrohima Siregar, Mertua Bapak Selamat, Suami tercinta Syafriandi, SM., dan anak tercinta Fakhira Syakira Mahveen serta seluruh keluarga yang telah memberikan dukungan, do'a, dan kasih sayangnya. Terima kasih pula kepada seluruh civitas akademik Statistika dan Sains Data IPB serta teman-teman STA2020.

Semoga karya ilmiah ini bermanfaat bagi pihak yang membutuhkan dan bagi kemajuan ilmu pengetahuan.

Bogor, Agustus 2024

Nurmai Syaroh Daulay

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

DAFTAR ISI

DAFTAR TABEL	xii
DAFTAR GAMBAR	xii
DAFTAR LAMPIRAN	xii
I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Tujuan Penelitian	3
II TINJAUAN PUSTAKA	4
2.1 Multikolinearitas	4
2.2 Pencilan	4
2.3 LASSO	5
2.4 WLAD LASSO	6
2.5 Regresi Kontinum Kekar	7
2.6 Alat Pengukuran Kadar Glukosa Darah Non-Invasif	8
III METODOLOGI PENELITIAN	10
3.1 Data	10
3.2 Analisis Data	10
3.2.1 Data Simulasi	10
3.2.2 Data Aktual	11
IV HASIL DAN PEMBAHASAN	13
4.1 Kajian Data Simulasi	13
4.2 Hasil Simulasi RKK, RKK-LASSO dan RKK-WLAD LASSO	14
4.3 Penerapan Metode Pada Kadar Glukosa Darah Non-Invasif	20
4.3.1 Eksplorasi Data Kadar Glukosa Darah Invasif	20
4.3.2 Eksplorasi Data Kadar Glukosa Darah Non-Invasif	21
4.4 Kinerja Metode Regresi Kontinum Kekar	23
V SIMPULAN DAN SARAN	26
5.1 Simpulan	26
5.2 Saran	26
DAFTAR PUSTAKA	27
LAMPIRAN	30
RIWAYAT HIDUP	37



DAFTAR TABEL

1	Keterangan nilai VIF	4
2	Rata-rata ukuran kebaikan model untuk data berdimensi rendah	14
3	Rata-rata ukuran kebaikan model untuk data berdimensi tinggi	16
4	Nilai VIF data kadar glukosa darah non-invasif	22
	Hasil nilai kebaikan model	24

DAFTAR GAMBAR

1	Contoh boxplot (sumber : (Suri <i>et al.</i> 2019))	5
2	Grafik residu intensitas terhadap <i>time</i> domain untuk satu periode	9
3	Boxplot peubah respon berdasarkan persentase pencilan	13
4	RMSE data berdimensi rendah (a) tanpa pencilan (b) pencilan 10% (c) pencilan 15% (d) pencilan 20%	15
5	RMSE data berdimensi tinggi (a) tanpa pencilan (b) pencilan 10% (c) pencilan 15% (d) pencilan 20%	18
6	Hasil simulasi rata-rata peubah terseleksi dengan LASSO dan WLAD LASSO (a) korelasi 0,1 (b) korelasi 0,5 (c) korelasi 0,9	19
7	Boxplot kadar glukosa darah invasif	20
8	Output alat pengukuran kadar glukosa darah non-invasif	21
9	Ilustrasi perhitungan luas area trapesium	22
10	Boxplot hasil RMSEP RKK, RKK-LASSO, RKK-WLAD LASSO	23
11	Plot y aktual dengan y duga	25

DAFTAR LAMPIRAN

1	Plot sebaran data peubah penjelas (a) korelasi 0,1 (b) korelasi 0,5 (c) korelasi 0,9	31
2	Peubah terseleksi dengan LASSO dan WLAD LASSO	33
3	RMSE dan RMSEP 100 ulangan	34