

Analisis Ridge Robust Penduga Generalized M (GM) Pada Pemodelan Kalibrasi Untuk Kadar Gula Darah

Agung Tri Utomo*, Erfiani, Anwar Fitrianto

Program Studi Statistika Terapan, Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Pertanian Bogor, Indonesia

Keywords: Adjusted Boxplot, Generalized M, Pemodelan Kalibrasi, Ridge Robust

Abstract:

Calibration modeling is one of the methods used to analyze the relationship between different methods. The relationship is like the relationship between invasive and non-invasive blood sugar measurement. Problems that often arise in calibration modeling are multicollinearity and outliers. Multicollinearity problems can cause the regression confidence interval to widen, so that there is no statistically significant regression coefficient. Outliers cause statistical tests to deviate. The handling of these problems can be solved by robust ridge analysis. Ridge robust is a combined analysis of ridge regression and robust regression. Ridge regression is able to overcome the problem of multicollinearity and robust regression can overcome the problem of outliers. The estimator used is Generalized M (GM). This method will be applied to a calibration model that uses invasive and non-invasive blood sugar level data. The model used with Generalized M (GM) estimator robust regression using modulation clusters 50 to 90 in 2017 is better than the modulation group 50. up to 90 in 2019. The statistical values obtained are SSE of 0.910, RMSEadj of 0.114, and RMSEP of 0.030. Calibration models that have outliers and multicollinearity problems can be overcome by robust ridge regression. The feasibility value of the model obtained in the GM estimator robust regression is smaller than the MM estimator ridge robust regression in the calibration modeling for non-invasive blood sugar level data. That is, the best model that can be used is the robust ridge regression GM estimator.

1. Pendahuluan

Kadar gula darah merupakan salah satu faktor yang dapat menyebabkan seseorang dapat terkena diabetes. Cara untuk mengukur kadar gula darah adalah dengan metode pengukuran invasif dan non-invasif. Metode invasif dilakukan dengan cara melukai anggota tubuh sehingga menimbulkan ketidaknyamanan pada pasien (Satria dan Wildan 2013). Metode non-invasif dapat mengukur kadar gula darah tanpa melukai pasien (Rassiyanti 2020). Dibutuhkan sebuah pemodelan yang dapat merumuskan fungsi antara keluaran alat pengukuran invasif dan non-invasif. Pemodelan tersebut adalah pemodelan kalibrasi.

Model kalibrasi merupakan suatu fungsi untuk menemukan hubungan antara satuan pengukuran yang dapat diperoleh melalui proses yang relatif mudah atau murah dengan satuan pengukuran yang memerlukan waktu lama dan biaya yang mahal (Naes et al. 2002). Permasalahan yang sering terjadi pada pemodelan kalibrasi adalah multikolinearitas dan pencilan. Penelitian Rosni (2019) menemukan bahwa adanya masalah multikolinearitas dan pencilan pada data yang digunakan dalam pemodelan kalibrasi.

Masalah multikolinearitas dapat mengakibatkan nilai selang kepercayaan koefisien regresinya akan melebar, sehingga cenderung menerima H_0 atau tidak ada koefisien yang signifikan secara statistik (Jolliffe 2002). Permasalahan ini dapat diatasi dengan menggunakan regresi ridge. Analisis regresi ridge merupakan salah satu metode yang dapat

* Corresponding author.

E-mail address: agungtriotomo10@gmail.com



mengatasi masalah multikolinearitas. Analisis ini merupakan modifikasi dari metode kuadrat terkecil yang menghasilkan penduga bias dari koefisien regresi dengan penambahan nilai tetapan bias (Kutner et al. 2005).

Permasalahan yang lain pada pemodelan kalibrasi berkaitan dengan pencilan yang akan mengakibatkan pengujian statistik menjadi menyimpang. Pencilan dapat diidentifikasi dengan beberapa metode. Metode tersebut antara lain adalah adjusted boxplot.

Jika pencilan dihilangkan maka dikhawatirkan terdapat informasi penting pada pencilan tersebut, akan tetapi juga dapat menyebabkan segala uji statistik menyimpang (Hair et al. 2014). Salah satu metode yang kekar terhadap pencilan adalah regresi robust. Regresi robust memiliki beberapa metode estimasi penduga yaitu penduga M (Maximum likelihood type), penduga Penduga LTS (Least Trimmed Squares), penduga S (Scale), dan penduga MM (Method of Moment).

Penelitian yang dilakukan Samkar dan Alpu (2010) yang menggabungkan regresi ridge dan regresi robust dalam mengatasi masalah multikolinearitas dan pencilan. Analisis gabungan tersebut adalah regresi ridge robust. Penelitian tersebut dilakukan dengan berbagai metode estimasi penduga robust yaitu penduga M, penduga GM, penduga S, dan penduga MM. Penelitian Rassiyanti (2020) menggunakan regresi ridge robust penduga MM menunjukkan bahwa analisis tersebut dapat mengatasi masalah pada pemodelan kalibrasi dan menghasilkan model yang lebih baik dari analisis yang pernah digunakan pada pemodelan kalibrasi. Penduga GM merupakan pengembangan dari penduga M. Penduga GM memiliki kesamaan dengan penduga MM yaitu memiliki tingkat breakdown point dan efisiensi yang tinggi.

Berdasarkan permasalahan tersebut, pencilan pada data kadar gula darah akan diidentifikasi dengan metode adjusted boxplot dan adjusted sequential fences. Kedua metode tersebut akan disimulasikan ke beberapa sebaran data untuk melihat performa dari kedua metode pendeteksi pencilan. Kemudian melakukan pemodelan regresi ridge robust menggunakan penduga Generalized M (GM). Regresi ridge robust penduga GM akan diterapkan pada pemodelan kalibrasi dengan gugus modulasi 50 sampai 90 untuk mengatasi masalah multikolinearitas dan pencilan.

2. Tinjauan Pustaka

2.1 Adjusted Boxplot

Adjusted boxplot merupakan metode *boxplot* yang disesuaikan. Hubert dan Vandervieren (2008) pernah melakukan sebuah penelitian dengan membandingkan *boxplot* dan *adjusted boxplot*. Teknik ini diperuntukkan pada distribusi miring.

Jika $X_n = (x_1, x_2, x_3, \dots, x_n)$ merupakan data univariat, maka tahapan membuat boxplot (Hubert dan Vandervieren 2008) :

- 1) Meletakkan garis pada ketinggian median sampel Q_2 .
- 2) Menggambar sebuah kotak dari kuartil pertama Q_1 ke kuartil ketiga Q_3 . Panjang kotak ini sama dengan inter kuartil rentang $IQR = Q_3 - Q_1$, yang merupakan ukuran skala yang kuat.
- 3) Mengklasifikasikan semua titik di luar interval (batas)

$$[Q_1 - 1.5 IQR ; Q_3 + 1.5 IQR]$$

Sebagai nilai potensial pencilan dan menandainya dalam plot.

- 4) Menggambar garis *whiskers* sebagai garis yang membentang dari ujung kotak ke titik terjauh di dalam batas.

Pengukuran terhadap kemiringan sampel data univariat $\{x_1, \dots, x_n\}$ dari unimodal kontinu berdistribusi F dapat menggunakan *medcouple* (MC) yang diperkenalkan oleh Brys et al. (2004) yang didefinisikan :

$$MC = med h(x_i, x_j) \rightarrow med : x_i \leq Q_2 \leq x_j$$

Q_2 adalah sampel median di mana untuk semua $x_i \neq x_j$ fungsi kernel h sebagai berikut.

$$h(x_i, x_j) = \frac{(x_i - Q_2) - (Q_2 - x_j)}{(x_j - x_i)}$$

dalam membuat sebuah *adjusted boxplot*, dilakukan penyesuaian pada nilai potensial pencilannya dengan menggunakan rumus :

- Saat $MC \geq 0$, menghitung pengamatan di luar interval :

$$[Q_1 - 1.5e^{-4MC} IQR ; Q_3 + 1.5e^{3MC} IQR]$$
- Saat $MC < 0$, menghitung Interval menjadi :

$$[Q_1 - 1.5e^{-3MC} IQR ; Q_3 + 1.5e^{4MC} IQR]$$

2.2 Regresi Robust

Regresi *robust* merupakan metode regresi yang digunakan ketika distribusi galat tidak normal dan atau adanya beberapa pencilan yang berpengaruh pada model (Ryan 1997). Pada regresi *robust* terdapat beberapa metode-metode estimasi. Menurut Chen (2002) metode estimasi pada regresi *robust* sebagai berikut:

- a. Penduga M (*Maximum likelihood type*) yang dikenalkan oleh Huber (1981) adalah metode yang sederhana baik dalam penghitungan maupun secara teoritis. Penduga ini menganalisis data dengan mengestimasi bahwa sebagian besar yang teridentifikasi pencilan pada variabel independen.
- b. Penduga LTS (*Least Trimmed Squares*) adalah metode dengan *high breakdown point* yang dikenalkan oleh Rousseeuw (1987). *Breakdown point* adalah salah satu cara yang digunakan untuk mengukur kekekaran suatu estimator (Nurdin et al. 2014).
- c. Penduga S (*Scale*) merupakan metode dengan *high breakdown point* yang dikenalkan oleh Rousseeuw dan Yohai (1987). Dengan nilai *breakdown point* yang sama, metode ini mempunyai efisiensi yang lebih tinggi dibandingkan Penduga LTS.
- d. Penduga MM (*Method of Moment*), dikenalkan oleh Yohai (1987) metode ini menggabungkan Penduga S (Penduga dengan *high breakdown point*) dengan Penduga M.

2.3 Penduga Generalized M (GM)

Penduga GM merupakan pengembangan dari Penduga M, ketika penduga M kurang sensitif terhadap pencilan pada variabel x_i . Ide dasar yang melatarbelakangi penduga GM adalah membatasi pengaruh dari pencilan pada variabel x_i dengan menggunakan fungsi pembobot w_i , di mana w_i hanya bergantung pada x_i (Lin 2007).

Menurut Wilcox (2005), secara umum penduga GM didefinisikan sebagai berikut :

$$\arg \min_{\beta} \sum_{i=1}^n w_i \rho \left(\frac{e_i}{v(x_i)} \right)$$

simbol $v(x_i)$ merupakan pembobot untuk variabel x_i . Penduga $\hat{\beta}$ yang diperoleh bukan merupakan skala *invariant*, maka digunakan nilai $\frac{e_i}{\hat{\sigma}}$ sebagai pengganti e_i , dengan $\hat{\sigma}$ adalah faktor skala yang juga perlu diestimasi. Dengan demikian persamaannya menjadi :

$$\sum_{i=1}^n w_i \rho \left(\frac{y_i - \sum_{j=0}^k x_{ij} \beta_j}{v(x_i) \hat{\sigma}} \right)$$

kemudian diturunkan terhadap β dan disamakan dengan nol, sehingga diperoleh persamaan :

$$\sum_{i=1}^n w_i \rho' \left(\frac{y_i - \sum_{j=0}^k x_{ij} \beta_j}{v(x_i) \hat{\sigma}} \right) = 0 \quad j = 0, 1, \dots, k$$

$$\sum_{i=1}^n w_i \psi \left(\frac{y_i - \sum_{j=0}^k x_{ij} \beta_j}{v(x_i) \hat{\sigma}} \right) x_i = 0 \quad j = 0, 1, \dots, k$$

Setelah itu, menentukan pembobot yang digunakan yaitu pembobot *Schwepe* dengan $w_i = \sqrt{1 - h_{ij}}$ dan $v(x_i) = w_i$, diperoleh :

$$w_{i,l} = w_i \frac{\psi \left(\frac{u_i}{v(x_i)} \right)}{(u_i)}$$

$$= \frac{w_i}{u_i} \psi\left(\frac{u_i}{v(x_i)}\right)$$

$$= \frac{\sqrt{1-h_{ij}}}{u_i} \psi\left(\frac{u_i}{\sqrt{1-h_{ij}}}\right)$$

Notasi $u_i = \frac{e_i}{\hat{\sigma}}$, dan $\hat{\sigma} = 1,48 M_l$, sedangkan M_l adalah median dari $(n-k)$ terbesar dari e_i dan $\psi(x) = \max\{-K, \min(K, x)\}$ adalah fungsi pengaruh Huber dengan $K = 2\sqrt{(k+1)/n}$. Dengan demikian Persamaan tersebut menjadi :

$$\sum_{i=1}^n x_{ij} w_i \left(y_i - \sum_{j=0}^k x_{ij} \beta_j \right) = 0 \quad j = 0, 1, \dots, k$$

Sisaan awal yang digunakan pada penduga GM adalah sisaan yang diperoleh dari Metode Kuadrat Terkecil (MKT). Persamaan ini dapat diselesaikan dengan MKT terboboti secara iterasi yang disebut *Iteratively Reweighted Least Square (IRLS)* hingga diperoleh $\hat{\beta}$ yang konvergen.

2.4 Regresi Ridge Robust

Menurut Samkar dan Alpu (2010), regresi *ridge robust* merupakan penggabungan dari metode regresi ridge dan regresi robust yang dilakukan untuk mengatasi masalah multikolinearitas dan pencilan. Penduga regresi *robust* yang dihasilkan akan stabil dan resistan terhadap pencilan. Rumus penduga parameter *ridge robust* adalah sebagai berikut :

$$\hat{\beta}_{RidgeRobust} = (X^T X + cI)^{-1} X^T X \hat{\beta}_{RobustGM}$$

dengan $\hat{\beta}_{RidgeRobust}$ merupakan penduga parameter regresi *ridge robust* dan $\hat{\beta}_{RobustGM}$ adalah penduga parameter regresi *robust* penduga GM.

Montgomery dan Peck (2012) menjelaskan bahwa salah satu cara untuk mendapatkan nilai c yang merupakan tetapan bias adalah dengan menggunakan metode yang diperkenalkan oleh Hoerl, Kennard dan Baldwin (HKB). Berdasarkan estimasi robust penduga GM, maka rumus untuk mencari nilai c sebagai berikut :

$$c = \frac{p \hat{\sigma}_{robustGM}^2}{\hat{\beta}'_{robustGM} \hat{\beta}_{robustGM}}$$

3. Metode Penelitian

3.1 Data

Data dari penelitian ini merupakan data kadar gula darah yang diperoleh dari hasil tes laboratorium yang dilakukan oleh Prodia untuk invasif dan dari penelitian pengembangan dan uji klinis purwarupa alat pemantauan kadar gula darah secara non-invasif oleh tim non-invasif biomarking IPB. Banyaknya responden pada data penelitian ini sebanyak 118 responden pada data penelitian tahun 2017 dan sebanyak 74 responden pada data penelitian tahun 2019. Responden pada data tahun 2017 merupakan responden yang berasal dari mahasiswa dan responden pada data tahun 2019 berasal dari masyarakat umum.

Kedua data tersebut digunakan karena desain instrumen alat non-invasif yang dikembangkan tim non-invasif biomarking IPB pada tahun 2017 dan tahun 2019 berbeda. Desain alat pada data penelitian tahun 2017 menangkap intensitas cahaya yang dilewatkan dari jari yang ditempelkan alat non-invasif, sedangkan data tahun 2019 menangkap intensitas cahaya yang diserap dari jari yang ditempelkan alat non-invasif. Peubah respon adalah data kadar gula darah invasif yang diperoleh dari pengujian labooratorium Prodia dan peubah bebas adalah hasil pengukuran alat non-nvasif berupa residu intensitas cahaya.

3.2 Tahapan Analisis Data

1. Melakukan identifikasi pencilan dengan menggunakan adjusted boxplot.
2. Melakukan pemeriksaan dengan uji multikolinearitas pada data pengamatan.
3. Melakukan pemodelan dengan regresi *ridge robust* sebagai berikut :
 - 1) Membagi data *testing* sebesar 15% dan data *training* sebesar 85%.
 - 2) Menghitung nilai $\hat{\beta}_{RobustGM}$ menggunakan regresi *robust* penduga GM dengan menggunakan persamaan :

$$\hat{\beta}_{RobustGM} = (X^T W^L X)^{-1} (X^T W^L Y)$$

dengan $\hat{\beta}_{RobustGM}$ merupakan penduga parameter *robust* penduga GM, X merupakan matriks kadar gula darah non-invasif yang berupa luasan trapesium berukuran $n \times p$, dan W^L merupakan fungsi pembobot.

- 3) Menghitung nilai tetapan bias c menggunakan persamaan :

$$c = \frac{p \hat{\sigma}_{robustGM}^2}{\hat{\beta}'_{robustGM} \hat{\beta}_{robustGM}}$$

- 4) Menghitung nilai $\hat{\beta}_{RidgeRobust}$ menggunakan persamaan :

$$\hat{\beta}_{RidgeRobust} = (X^T X + cI)^{-1} X^T X \hat{\beta}_{RobustGM}$$

dengan $\hat{\beta}_{RidgeRobust}$ merupakan penduga parameter *ridge robust* dan X merupakan kadar gula darah non-invasif hasil dari perhitungan luasan trapesium yang berukuran $n \times p$, nilai c merupakan nilai tetapan bias dan I merupakan matriks identitas.

- 5) Mengulang pemodelan sebanyak 100 kali pengulangan.
- 6) Menghitung nilai kelayakan model menggunakan SSE, RMSE, RMSEP dengan persamaan :

$$SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2, i = 1, 2, 3, \dots, n$$

$$RMSE_{Adj} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - k - 1}}, i = 1, 2, 3, \dots, n$$

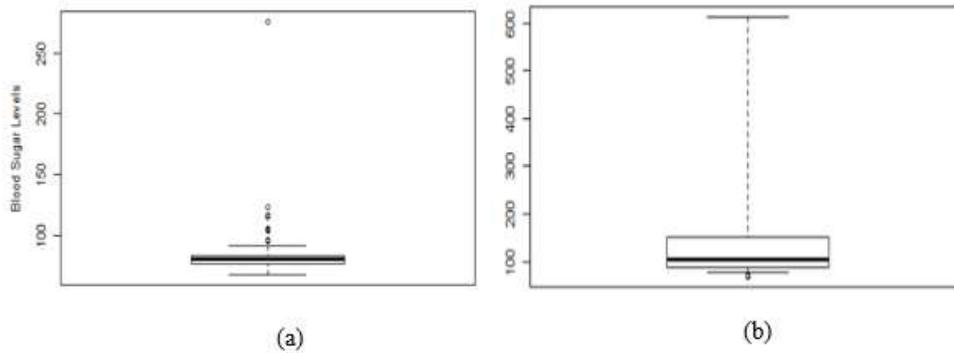
$$RMSEP = \sqrt{\frac{\sum_{i=1}^m (y_i - \hat{y}_i)^2}{m}}, i = 1, 2, 3, \dots, m$$

4. Membandingkan model berdasarkan nilai kelayakan model pada setiap model.
5. Menarik kesimpulan.

4. Hasil dan Pembahasan

4.1. Identifikasi Pencilan dengan Adjusted Boxplot

Terdapat indikasi adanya data pencilan pada data penelitian tahun 2017 dan tahun 2019. Pencilan tersebut akan diidentifikasi kembali dengan menggunakan metode *adjusted boxplot*. Berdasarkan hal tersebut, hasil dari identifikasi pencilan dapat dilihat pada Gambar 1.



Gambar 1 *Adjusted Boxplot* (a) data tahun 2017 dan (b) data tahun 2019 penelitian kadar gula darah.

Pada Gambar 1 (a) memperlihatkan hasil dari visualisasi *adjusted boxplot* pada data tahun 2017. Terdapat 12 pencilan yang terdeteksi sebagai pencilan. Pencilan tersebut merupakan responden yang memiliki kadar gula darah yang tinggi yaitu sebesar 96, 96, 96, 96, 103, 103, 104, 105, 115, 116, 123, dan 276.

Berdasarkan Gambar 1 (b), *adjusted boxplot* pada data tahun 2019 mengidentifikasi adanya pencilan pada data. Terdapat sebanyak 3 data yang teridentifikasi sebagai pencilan. Pencilan tersebut merupakan responden dengan kadar gula darah yaitu 69, 73, dan 73.

Hasil tersebut memperlihatkan bahwa *adjusted boxplot* mampu, mengidentifikasi pencilan pada data kadar gula darah. Setiap data memiliki jumlah pencilan yang berbeda. Pencilan tersebut akan menjadi sebuah informasi yang memperlihatkan adanya nilai yang berbeda pada data yang digunakan.

4.2. Uji Multikolinieritas Pada Data Penelitian Kadar Gula Darah Tahun 2017 dan Tahun 2019

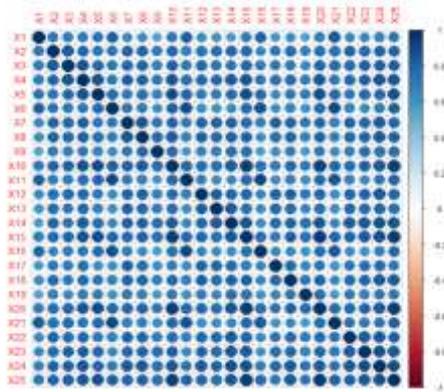
Sebelum melakukan pemodelan regresi *ridge robust*, maka akan dilakukan uji multikolinieritas. Pengujian multikolinieritas akan dilakukan pada data penelitian dengan modulasi 50 sampai dengan 90 yang dapat dilihat melalui nilai VIF. Nilai VIF pada data penelitian modulasi 50 sampai 90 dapat dilihat pada Tabel 1.

Tabel 1 Nilai VIF data penelitian tahun 2017 (modulasi : 50 sampai 90).

Variabel	Nilai VIF	Variabel	Nilai VIF	Variabel	Nilai VIF
X_1	4.90	X_{11}	10.83	X_{21}	10.17
X_2	5.89	X_{12}	6.69	X_{22}	6.41
X_3	6.67	X_{13}	5.38	X_{23}	8.10
X_4	8.56	X_{14}	9.12	X_{24}	10.59
X_5	8.69	X_{15}	20.99	X_{25}	23.81
X_6	11.23	X_{16}	8.98		
X_7	7.28	X_{17}	5.02		
X_8	6.86	X_{18}	5.97		
X_9	3.25	X_{19}	3.12		
X_{10}	19.93	X_{20}	19.58		

Dalam penelitian Daoud (2017), apabila nilai VIF yang melebihi 10 maka indikasi adanya kasus multikolinieritas. Berdasarkan tabel, terdapat 8 variabel yang mengalami multikolinieritas yaitu variabel $X_6, X_{10}, X_{11}, X_{15}, X_{20}, X_{21}, X_{24},$ dan X_{25} . Adapun korelasi pada data penelitian ini yang dapat dilihat pada Gambar 2.

Gambar 2 Korelasi gugus modulasi penelitian tahun 2017.

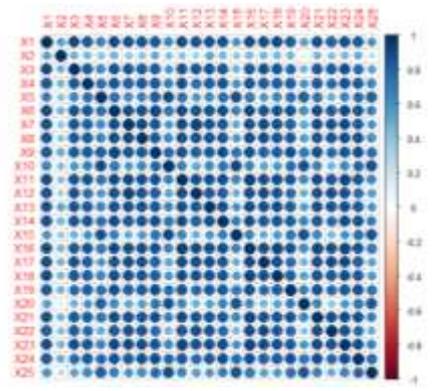


Berdasarkan Gambar 2, terlihat plot *correlogram* bahwa multikolinearitas terjadi pada variabel-variabel X_1 sampai dengan X_{25} yang ditandai dengan intensitas warna gelap hingga terang. Hampir semua dari variabel-variabel X_1 sampai dengan X_{25} mengalami masalah multikolinearitas. Selanjutnya pada data penelitian tahun 2019 dengan modulasi 50 sampai dengan 90 akan dilakukan pemeriksaan multikolinearitas dengan melihat nilai VIF pada setiap variabel. Nilai VIF dapat dilihat pada Tabel 2.

Tabel 2 Nilai VIF data penelitian tahun 2019 (modulasi : 50 sampai 90).

Variabel	Nilai VIF	Variabel	Nilai VIF	Variabel	Nilai VIF
X_1	12.74	X_{11}	14.66	X_{21}	15.31
X_2	3.33	X_{12}	12.29	X_{22}	12.63
X_3	10.13	X_{13}	10.30	X_{23}	14.84
X_4	7.86	X_{14}	10.00	X_{24}	14.52
X_5	10.91	X_{15}	12.46	X_{25}	9.48
X_6	12.69	X_{16}	20.35		
X_7	15.86	X_{17}	12.32		
X_8	11.48	X_{18}	11.28		
X_9	16.41	X_{19}	8.80		
X_{10}	16.97	X_{20}	9.65		

Berdasarkan pada Tabel 2, data penelitian tahun 2019 dengan modulasi 50 sampai dengan 90 memiliki variabel yang nilai VIF melebihi dari 10. Terdapat 10 variabel yang nilai VIF melebihi 10 yaitu variabel $X_1, X_3, X_5, X_6, X_7, X_8, X_9, X_{10}, X_{11}, X_{12}, X_{13}, X_{14}, X_{15}, X_{16}, X_{17}, X_{18}, X_{21}, X_{22}, X_{23}$, dan X_{24} . Adapun Korelasi pada data penelitian ini yang dapat dilihat pada Gambar 3.



Gambar 3 Plot korelasi gugus modulasi data penelitian tahun 2019.

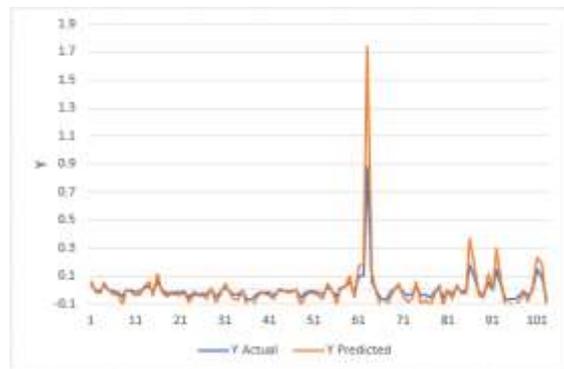
Berdasarkan Gambar 3, variabel gugus modulasi tersebut mengalami multikolinearitas yang ditandai dengan intensitas warna pada plot *correlogram*. Plot ini menunjukkan data dengan berbagai macam intensitas warna. Terlihat

bahwa variabel X_1 sampai dengan X_{25} memiliki warna yang intensitasnya agak sedikit gelap yang artinya terjadi masalah multikolinearitas.

4.3. Pemodelan Regresi Ridge Robust Penduga Generalized M (GM)

Analisis data yang diterapkan pada gugus data penelitian adalah menggunakan analisis regresi *ridge robust* penduga GM. data kemudian dibagi menjadi 2 bagian. Masing-masing bagian memiliki proporsi yaitu data *training* sebesar 85% dan data *testing* sebesar 15%. Analisis data akan diulang sebanyak 100 kali untuk mendapatkan hasil yang optimal.

Berdasarkan pemeriksaan yang telah dilakukan, data penelitian tahun 2017 dengan modulasi 50 sampai 90 memiliki masalah multikolinearitas. Analisis regresi *ridge robust* penduga GM dapat mengatasi permasalahan multikolinearitas dan pencilan. Penduga parameter yang diperoleh dari *robust* penduga GM akan digunakan untuk menduga parameter *ridge* dengan menambahkan nilai tetapan bias. Tetapan bias yang diperoleh dari 100 kali ulangan kemudian dirata-ratakan dan mendapatkan nilai 0.302. Nilai tetapan bias ini akan digunakan untuk menstabilkan koefisien penduga. Plot hasil nilai Y prediksi terhadap nilai Y aktual untuk model kalibrasi dengan gugus data modulasi 50 sampai dengan 90 pada tahun 2017 dapat dilihat pada Gambar 4.10.



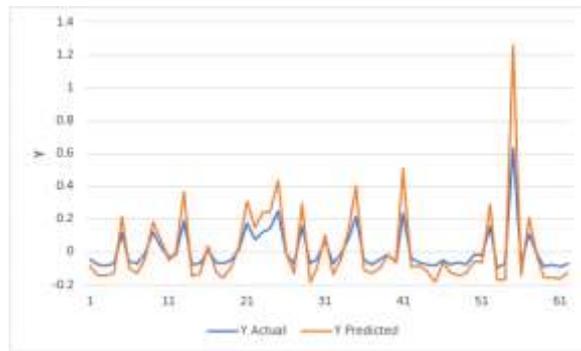
Gambar 4 Pola prediksi Y aktual terhadap Y prediksi data penelitian 2017.

Berdasarkan Gambar 4, pola antara Y aktual dan Y prediksi memiliki pola yang cenderung sama pada gugus modulasi 50 sampai dengan 90. Berdasarkan Tabel 3, nilai statistik SSE dan $RMSE_{adj}$ pada data *training* yaitu sebesar 0.910 dan 0.114. Pada data *testing*, nilai statistik RMSEP yang dihasilkan sebesar 0.030.

Tabel 3 Hasil pemodelan data *training* dan *testing* regresi *ridge robust* penduga GM pada data penelitian 2017.

Statistik	Data <i>training</i>		Data <i>testing</i>
	SSE	$RMSE_{adj}$	RMSEP
<i>Ridge robust</i> penduga GM	0.910	0.114	0.030

Berdasarkan hasil pemeriksaan data penelitian 2019, gugus data yang digunakan memiliki masalah yang sama dengan data sebelumnya yaitu multikolinearitas. Pemodelan untuk mengatasi masalah tersebut adalah pemodelan *ridge robust* dengan penduga GM. parameter penduga *robust* GM akan digunakan untuk menduga parameter *ridge* dengan menambahkan nilai tetapan bias. Analisis ini akan diulang sebanyak 100 kali. Nilai rata-rata tetapan bias dari analisis dengan menggunakan *ridge robust* penduga GM yaitu sebesar 0.075. Plot hasil dari Y prediksi dengan Y aktual untuk modulasi gugus data pada tahun 2019 dapat dilihat pada Gambar 4.



Gambar 5 Pola prediksi Y aktual terhadap Y prediksi data penelitian 2019.

Berdasarkan Gambar 5, pola yang terbentuk antara Y aktual dengan Y prediksi terlihat hampir membentuk pola yang sama. Pada Tabel 4 nilai statistik SSE dan $RMSE_{adj}$ pada data *training* memiliki nilai sebesar 0.553 dan 0.129. Pada data *testing*, nilai statistik RMSEP yang dihasilkan sebesar 0.103

Tabel 4 Hasil pemodelan data *training* dan *testing* regresi *ridge robust* penduga GM pada data penelitian 2019.

Statistik	Data <i>training</i>		Data <i>testing</i>
	SSE	$RMSE_{adj}$	RMSEP
<i>Ridge robust</i> penduga GM	0.553	0.129	0.103

Perbandingan model pada penelitian ini menggunakan ukuran kebaikan model SSE, $RMSE_{adj}$, dan RMSEP. Nilai tersebut digunakan untuk mengetahui kebaikan model kalibrasi yang diperoleh menggunakan regresi *ridge robust*. $RMSE_{adj}$ diperoleh dari data *training*, sedangkan RMSEP diperoleh dari data *testing*. Jika nilai SSE, $RMSE_{adj}$, dan RMSEP semakin kecil maka model yang diperoleh juga akan semakin baik. Berikut perbandingan nilai kebaikan model dari regresi *ridge robust* penduga MM dan GM yang disajikan pada Tabel 5.

Tabel 5 Nilai perbandingan ukuran kebaikan model regresi *ridge robust* GM dan MM.

Statistik	Modulasi	Tahun	Data <i>training</i>		Data <i>testing</i>
			SSE	$RMSE_{adj}$	RMSEP
<i>Ridge robust</i> penduga GM	50 - 90	2017	0.910	0.114	0.030
		2019	0.553	0.129	0.103
<i>Ridge robust</i> penduga MM	50 - 90	2017	0.918	0.115	0.034
		2019	0.744	0.148	0.116

Tabel 5 memperlihatkan bahwa hasil pemodelan data penelitian tahun 2017 dan tahun 2019 dengan gugus modulasi 50 sampai dengan 90 pada model kalibrasi menggunakan regresi *ridge robust* penduga GM memiliki nilai SSE, $RMSE_{adj}$, dan RMSEP yang lebih kecil dibandingkan dengan regresi *ridge robust* penduga MM. Hal ini menyimpulkan bahwa regresi *ridge robust* penduga GM lebih layak untuk diterapkan dalam model kalibrasi dibandingkan dengan regresi *ridge robust* penduga MM. Data penelitian tahun 2017 merupakan gugus data yang lebih baik dengan nilai kelayakan model yang lebih kecil.

Perbandingan model tanpa pencilan dilakukan dengan model regresi *ridge robust* penduga GM. Model yang akan dibandingkan berdasarkan jumlah pencilan pada *adjusted boxplot*. Jumlah pencilan tersebut akan dihilangkan. Hasil dari model tersebut dapat dilihat pada Tabel 6.

Tabel 6 Perbandingan model regresi *ridge robust* penduga GM dengan metode *adjusted boxplot*.

Tahun	Nilai Kelayakan Model	
	RMSE _{adj}	RMSEP
2017	0.108	0.073
2019	0.126	0.111

Tabel 6 menunjukkan bahwa nilai kelayakan model tanpa mengikutsertakan pencilan mendapatkan nilai yang cenderung kecil. Perbandingan nilai kelayakan model tanpa pencilan (Tabel 6) dan model dengan pencilan (Tabel 5) pada model regresi *ridge robust* memiliki perbedaan. Pada nilai RMSE_{adj} model tanpa pencilan (Tabel 6) lebih kecil dibandingkan dengan model dengan pencilan (Tabel 5) dan nilai RMSEP model dengan pencilan (Tabel 5) lebih kecil dibandingkan dengan model tanpa pencilan (Tabel 6). Hal ini menunjukkan bahwa model tanpa pencilan lebih baik dari pada data dengan pencilan pada model *training* dan tidak lebih baik pada model *testing*. Disarankan untuk lebih baik memasukkan pencilan pada regresi *ridge robust* agar hasil yang diperoleh lebih maksimal.

5. Kesimpulan

Pada data kadar gula darah, *adjusted boxplot* mampu mengidentifikasi data pencilan pada data tahun 2017 sebanyak 12 data dan pada data tahun 2019 sebanyak 3 data. Pemodelan kalibrasi yang memiliki masalah pencilan dan multikolinieritas dapat diatasi dengan regresi *ridge robust*. Nilai kelayakan model yang didapatkan pada regresi *ridge robust* penduga GM lebih kecil dibandingkan dengan regresi *ridge robust* penduga MM pada pemodelan kalibrasi untuk data kadar gula darah non-invasif. Artinya, model terbaik yang dapat digunakan adalah regresi *ridge robust* penduga GM.

References

- [ADA] American Diabetes Association (USA). 2020. Classification and diagnosis of diabetes: standard of medical in diabetes. *Diabetes Care*. 43(1): 514-531
- Avan. 2017. Pendugaan konsentrasi glukosa darah menggunakan jaringan syaraf tiruan pada alat non-invasive. [tesis]. Bogor (ID): Sekolah Pascasarjana IPB.
- Bowley AL. 1920. *Elements of Statistics*. Ed ke-4. London (UK): P.S. King & Son, Ltd.
- Brys G, Hubert M, Struyf A. 2004. A robust measure of skewness. *Journal of Computational and Graphical Statistics*. 13(4): 996-1017.
- Chandraningtyas S, Safitri D, Ispriyanti D. 2013. Regresi robust MM-estimator untuk penanganan pencilan pada regresi linier berganda. *Jurnal Gaussian*. 2(4): 395-404.
- Chen C. 2002. Robust regression and outlier detection with the ROBUSTREG procedure. Statistics and Data Analysis. SUGI Paper 265-27. North Carolina: SAS Institute.
- Cousineau D, Chartier S. 2010. Outlier detection and treatment : a review. *International Journal of Psychological Research*. 3(1): 59-68
- Daoud IJ. 2017. Multicolinierity and regression analysis. *Journal of Physics: Conference Series* 949.
- Dereny M, Rashwan NI. 2011. Solving multicolinierity problem using ridge regression model. *International Journal Contemporary Mathematical Sciences*. 6(12): 585-600.
- Gujarati DN. 2003. *Basic Econometrics*. Ed ke-4. New York (NY) : McGraw-Hill Companies, Inc.
- Gujarati DN, Porter D. 2009. *Basic Econometrics* . Ed ke-5. New York (NY) : McGraw-Hill/Irwin.
- Hair JF, Black WC, Babin BJ, Anderson RE. 2014. *Multivariate Data Analysis*. Ed ke-7. Edinburgh Gate (UK): Pearson.
- Herianti. 2020. Pemodelan kalibrasi kekar (pengukuran glukosa darah non-invasif). [tesis]. Bogor (ID): Sekolah Pascasarjana IPB.
- Hruschka WR. 1987. Data analysis : wavelength selection methods. *American of Association of Cereal Chemists*. 35-55.
- Huber PJ. 1981. *Robust Statistics*. New York (NY): John Wiley And Sons. Inc.
- Huber PJ, Ronchetti EM. 2009. *Robust Statistics*. New York (NY): John Wiley And Sons. Inc.
- Hubert M, Vandervieren E. 2008. An adjusted boxplot for skewed distribution. *Computational Statistics and Data*

Analysis. 52: 5186 – 5201.

- Ismah, Wigena AH, Djuraidah A. 2009. Pendekatan regresi kuadrat terkecil partial robust dalam model kalibrasi. *Forum Statistika dan Komputasi*. 14(1): 34-41.
- Jayanti PGK, Anisa R, Aidi MN, Erfiani. 2018. Penerapan teknik prapemrosesan smoothing spline pada data hasil pengukuran alat pemantau glukosa darah non-invasif. *Jurnal Xplore*. 2(2): 15-35.
- Jolliffe IT. 2002. *Principal Component Analysis*. Ed ke-2. New York (NY): Springer-Verlag.
- Kim T, White H. 2004. On more robust estimation of skewness and kurtosis: simulation and application to the S & P500 indeks. *Finance Research Letters*. 1(1): 56-73.
- Kutner MH, Nachtsheim CJ, John N, Li WG. 2005. *Applied Linier Statistics Model*. Ed ke-5. New York (US): Mc-Grawhill.
- Lin TI, Lee JC, Hsieh WJ. 2007. Robust mixture modeling using the skew t-distribution. *Statistics and Computing*. 17(2): 81-92.
- Mardikyan S, Cetin E. 2008. Efficient choice of biasing constant for ridge regression. *International Journal Contemporary Mathematical Sciences*. 3(11): 527-536.
- Montgomery DC, Peck EA. 2012. *Introduction to Linear Regression Analysis*. Ed ke-5. New York (US): John Willey and Sons.
- Naes T, Issakson T, Fearn T, Davies T. 2002. *Multivariate Calibration and Classification*. United Kingdom (GB): NIR Publications Chichester.
- Nalwan A. 2004. *Pengolahan Gambar Secara Digital*. Jakarta (ID): PT Elex Media Komputindo.
- Nugroho AB, Rintyarna BS, Athoillah DK. 2021. Analisis spektrum tegangan pada alat pendeteksi kadar gula darah menggunakan *near*. *Jurnal Teknik Elektro dan Komputasi (ELKOM)*. 3(1): 1-13.
- Nurdin N, Raupong, Islamiyati A. 2014. Penggunaan regresi robust pada data yang mengandung pencilan dengan metode momen. *Jurnal Matematika, Statistika, dan Komputasi*. 10(2): 114-123.
- Pambudi N. 2012. Tingkat efisiensi estimasi-M terhadap estimasi-GM dalam regresi robust. [skripsi]. Surakarta (ID): Universitas Sebelas Maret Surakarta.
- Peck R, Devore JL. 2012. *The Exploration and Analysis of Data*. Ed ke-7. Boston (USA): Brooks.
- Rassiyanti L. 2020. Analisis regresi ridge robust pada pengukuran kadar glukosa darah non-invasif. [tesis]. Bogor (ID): Sekolah Pascasarjana IPB.
- Rosni. 2019. Pendekatan regresi komponen utama dan regresi kuadrat terkecil parsial untuk menduga kadar glukosa darah non-invasif. [tesis]. Bogor (ID): Sekolah Pascasarjana IPB.
- Rousseeuw PJ, Leroy AM. 1987. *Robust Regression and Outlier Detection*. New York (US): John Willey and Sons.
- Ryan TP. 1997. *Modern Regression Methods*. New York (US): John Wiley And Sons. Inc.
- Said AZ, Fitrianto A, Erfiani. 2022. Pendeteksian pencilan pada data berdistribusi miring univariat menggunakan pendekatan modifikasi sequential fences. [tesis]. Bogor (ID): Sekolah Pascasarjana IPB.
- Samkar H, Alpu O. 2010. Ridge regression based on some robust estimators. *Journal of Modern Applied Statistical Methods*. 9(2): 495-501.
- Satria E, Wildian. 2013. Rancangn bangun alat ukur kadar gula darah non-invasif berbasis mikrokontroler AT89S51 dengan mengukur tingkat kekeruhan spesimen urine menggunakan sensor fotodioda. *Jurnal Fisika UNAND*. 2(1): 40-47.
- Schwertman NC, Silva RD. 2007. Identifying outliers with sequential fences. *Computational Statistics and Data Analysis*. 51(8) : 3800-3810.
- Susanti Y, Liana T, Pratiwi H, Sulistijowati SH. 2014. M estimation, S estimation dan MM estimation in robust regression. *International Journal of Pure and Applied Mathematics*. 3(91): 349-360.
- Tiro MA, Ahsan M. 2015. *Penyajian Informatif : Tabel, Grafik, dan Statistik*. Makassar (ID): Andira Publisher.
- Wahyono T. 2004. *Sistem Informasi*. Yogyakarta: Graha Ilmu.
- Wilcox RR. 2005. *Introduction to Robust Estimation and Hypothesis*. San Diego: Academic Press.
- Wong HS, Fitrianto A. 2019. Adjusted sequential fences for detecting uniovariate outliers in skewed distribution. *ASM Science Journal*, 12(5): 107-115.
- Yohai VJ. 1987. High breakdown-point and high-efficiency robust estimates for regression. *The Annals of Statistics*. 15(20): 642-656.