



ANALISIS RAGAM PEUBAH GANDA DATA TERDETEKSI PENCILAN UNTUK MENGETAHUI FAKTOR FAKTOR YANG MEMPENGARUHI LIPID DARAH PADA PENYANDANG DIABETES

RAHMAH HUMAEROH



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



@Hak cipta milik IPB University

IPB University



IPB University
— Bogor Indonesia —

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 “Analisis Ragam Peubah Ganda Data Terdeteksi Pencilan untuk Mengetahui Faktor Faktor yang Mempengaruhi Lipid Darah pada Penyandang Diabetes” 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

Rahmah Humaeroh
G1501211053

- 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

RAHMAH HUMAEROH. Analisis Ragam Peubah Ganda Data Terdeteksi *Pencilan* Untuk Mengetahui Faktor Faktor yang Mempengaruhi Lipid Darah Pada Penyandang Diabetes. Dibimbing oleh MUHAMMAD NUR AIDI, BAGUS SARTONO dan NAZARINA.

Analisis multivariat sangat penting dalam menangani data yang kompleks dengan melibatkan lebih dari satu peubah yang diukur secara bersamaan. Salah satu metode yang sering digunakan adalah MANOVA (*Multivariate Analysis of Variance*), yang digunakan untuk menguji perbedaan antar kelompok dalam peubah respon metrik. Ketika terdapat kovariat yang perlu diperhitungkan dalam analisis, MANCOVA (*Multivariate Analysis of Covariance*) dapat menjadi pilihan yang digunakan pada data multivariat.

Analisis data multivariat tidak lepas dari berbagai tantangan, salah satunya adalah masalah pencilan yang dapat mengganggu asumsi dasar seperti normalitas multivariat dan homokesdastisitas. Pencilan dapat mempengaruhi hasil analisis secara signifikan karena titik data yang menyimpang dari distribusi utama dapat menyebabkan distorsi dalam estimasi parameter dan kesalahan dalam interpretasi hasil. Dalam konteks multivariat, pencilan menjadi lebih kompleks karena melibatkan interaksi antar peubah. Pencilan multivariat, yang mencerminkan titik data yang tidak sesuai dengan pola distribusi beberapa peubah sekaligus, seringkali lebih sulit dideteksi dan lebih berpotensi mempengaruhi hasil analisis secara signifikan dibandingkan univariat.

Pencilan pada data multivariat dapat diatasi dengan pendekatan robust seperti R-MANCOVA yang menggunakan *Minimum Covariance Determinant* (MCD). Metode MCD bekerja dengan mengidentifikasi sub-sampel data yang relatif bebas dari pencilan dan menggunakan sub-sampel ini untuk estimasi kovariansi yang lebih stabil. Dengan mengurangi pengaruh pencilan, MCD mampu memberikan hasil analisis yang lebih *robust* dan akurat, terutama dalam konteks data multivariat yang kompleks.

Penelitian ini menggunakan nilai *Root Mean Square Error* (RMSE) untuk melihat efektivitas dan ketepatan dari metode MANOVA, MANCOVA dan R-MANCOVA. Nilai RMSE digunakan sebagai indikator kinerja model dalam mengukur kesalahan antara nilai prediksi dan nilai yang sebenarnya diamati, dengan nilai RMSE yang lebih rendah menunjukkan model yang lebih akurat. Hasil penelitian menunjukkan bahwa R-MANCOVA adalah model yang paling efektif, dengan nilai RMSE terendah dibandingkan mode lainnya yaitu 0,155 untuk HDL 0,210 untuk LDL dan 0,399 untuk TG. Analisis R-MANCOVA mengidentifikasi bahwa gangguan tidur, olahraga, dan jenis kelamin secara signifikan mempengaruhi kadar lipid darah, dengan metode robust ini dapat mengatasi pengaruh pencilan dan memberikan hasil yang lebih akurat dalam konteks analisis multivariat.

Kata kunci: MANOVA , MANCOVA, *Minimum Covariance Determinant*.



SUMMARY

RAHMAH HUMAEROH. Analysis of Multivariate Data Variance Detected Outlier to Determine Factors Influencing Blood Lipid in Diabetes Patients. Supervised by MUHAMMAD NUR AIDI, BAGUS SARTONO and NAZARINA.

Multivariate analysis is critical in handling complex data involving more than one variable being measured simultaneously. One method often used is MANOVA (Multivariate Analysis of Variance), which tests differences between groups in metric response variables. When covariates need to be considered in the analysis, MANCOVA (Multivariate Analysis of Covariance) can be an option for use on multivariate data.

Despite its many benefits, multivariate data analysis is not without its challenges. One such challenge is the issue of outliers, which can disrupt fundamental assumptions like multivariate normality and homoscedasticity. These outliers can have a significant impact on the analysis results, as data points that deviate from the main distribution can distort parameter estimates and lead to errors in result interpretation. In a multivariate context, outliers become even more complex due to variable interactions. Multivariate outliers, which represent data points that do not conform to the distribution pattern of several variables at once, are often more difficult to detect and have a greater potential to influence the analysis results than univariate outliers.

Outliers in multivariate data can be overcome with a robust approach such as R-MANCOVA, which uses a Minimum Covariance Determinant (MCD). The MCD method works by identifying sub-samples of data that are relatively free of outliers and using these sub-samples for more stable covariance estimates. By reducing the influence of outliers, MCD can provide more robust and accurate analysis results, especially in complex multivariate data.

This research uses the Root Mean Square Error (RMSE) value to assess the effectiveness and accuracy of the MANOVA, MANCOVA, and R-MANCOVA methods. The RMSE value serves as an indicator of model performance, measuring the error between the predicted value and the observed value. A lower RMSE value indicates a more accurate model. The results demonstrate the effectiveness of the R-MANCOVA method, with the lowest RMSE value compared to other modes, specifically 0.155 for HDL, 0.210 for LDL, and 0.399 for TG. The r-MANCOVA analysis identified sleep disorders, exercise, and gender as significant influencers of blood lipid levels. This robust method effectively overcomes the influence of outliers, providing more accurate results in the context of multivariate analysis.

Keywords: MANOVA, MANCOVA, Minimum Covariance Determinant

@Hak cipta milik IPB University

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 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.

ANALISIS RAGAM PEUBAH GANDA DATA TERDETEKSI PENCILAN UNTUK MENGETAHUI FAKTOR FAKTOR YANG MEMPENGARUHI LIPID DARAH PADA PENYANDANG DIABETES

RAHMAH HUMAEROH

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. I Made Sumertajaya, M.Si



Judul Tesis : Analisis Ragam Peubah Ganda Data Terdeteksi Pencilan untuk Mengetahui Faktor Faktor yang Mempengaruhi Lipid Darah pada Penyandang Diabetes

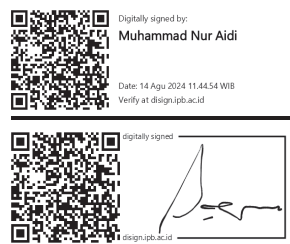
Nama : Rahmah Humaeroh
NIM : G1501211053

Disetujui oleh

Pembimbing 1:
Prof. Dr. Ir. Muhammad Nur Aidi, M.S

Pembimbing 2:
Dr. Bagus Sartono, S.Si., M.Si.

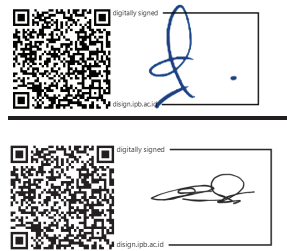
Pembimbing 3:
Dr. Nazarina M.Med., Sci



Diketahui oleh

Ketua Program Studi:
Dr. Agus M. 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:
5 Agustus 2024

Tanggal Lulus:



@Hak cipta milik IPB University

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

Alhamdulillahirabbil'allamiin. Puji dan syukur penulis panjatkan kepada Allah subhanahu wa ta'ala atas segala karunia-Nya sehingga karya ilmiah ini berhasil diselesaikan. Tema yang dipilih dalam penelitian ini ialah pemodelan pada data yang memiliki peubah kovariat serta terdeteksi pencilan, dengan judul "Analisis Ragam Peubah Ganda Data Terdeteksi Pencilan untuk Mengetahui Faktor-Faktor yang Mempengaruhi Lipid Darah Pada Penyandang Diabetes". Terima kasih, penulis ucapkan kepada semua pihak yang telah membantu dan mendukung dalam penyelesaian tesis ini, terutama kepada :

1. Prof. Dr. Ir. Muhammad Nur Aidi, M.S., Dr. Bagus Sartono, S.Si., M.Si dan Ibu Dr. Nazarina M.Med., Sci selaku komisi pembimbing yang telah membimbing, mendukung, memberikan ilmu serta saran juga masukan dalam menyelesaikan karya ilmiah ini.
2. Dr.Yenni Angraini, M.Si selaku pemimpin ujian tesis, Dr. I Made Sumertajaya, M.Si. selaku penguji luar komisi pembimbing, dan Prof. Dr-Ing. Ir Suprihatin. selaku moderator seminar yang telah membantu lancarnya proses penyelesaian tugas akhir penulis.
3. Seluruh Dosen dan Karyawan Departemen Statistika IPB yang telah banyak memberikan bantuan baik selama perkuliahan sampat terselesaikannya karya ilmiah ini.
4. Mamah, Bapak, A Karim, Teh Alinda, Hyrra, Teh Linda, A Roni dan Syafiq serta segenap keluarga yang selalu memberikan dukungan, doa, dan kasih sayang kepada penulis.
5. Sahabat-sahabat yang selalu memberikan dukungan dan mengingatkan penulis untuk segera menyelesaikan penulisan ini.
6. Teman-teman Program Studi Statistika dan Sains Data IPB yang telah memberikan semangat dan bantuan dan semua pihak yang telah banyak membantu dan tak sempat penulis sebutkan satu per satu.
7. Terakhir, penulis ucapkan kepada diri sendiri yang telah berhasil melalui proses panjang perkuliahan, penelitian dan menyelesaikan program magister ini dengan sebaik-baiknya.

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

Bogor, Agustus 2024

Rahmah Humaeroh



@Hak cipta milik IPB University

IPB University



IPB University
— Bogor Indonesia —

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.

Perpustakaan IPB University



DAFTAR ISI

DAFTAR TABEL	xi
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN	xi
I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan	3
1.4 Manfaat	3
1.5 Ruang Lingkup	3
II TINJAUAN PUSTAKA	4
2.1 Diabetes	4
2.2 <i>Multivariate Anlaysis of Variance (MANOVA)</i>	4
2.3 <i>Multivariate Anlaysis of Covariance (MANCOVA)</i>	4
2.4 Data Pencilan	7
2.5 Penduga <i>Minimum Covariance Determinant (MCD)</i>	7
2.6 Teorema C-steps	7
2.7 <i>Root Mean Square Error</i>	8
III METODE	10
3.1 Data	10
3.2 Metode Analisis	12
IV HASIL DAN PEMBAHASAN	14
4.1 Statistik Deskriptif	14
4.2 Pemeriksaan Asumsi Analisis Ragam dan Peragam Peubah Ganda	17
4.3 Analisis Ragam Peubah Ganda (MANOVA)	19
4.4 Analisis Ragam Peubah Ganda Lipid darah Darah pada Penyandang Diabetes dengan Kovariat	20
4.5 Analisis Ragam Peubah Ganda <i>Robust</i> Lipid Darah pada Penyandang Diabetes dengan Kovariat	21
4.6 Perbandingan RMSE MANOVA, MANCOVA dan R-MANCOVA	23
V SIMPULAN DAN SARAN	25
5.1 Simpulan	25
5.2 Saran	25
DAFTAR PUSTAKA	26
LAMPIRAN	29
RIWAYAT HIDUP	33



DAFTAR TABEL

1	Peubah yang digunakan dalam penelitian	10
2	Karakteristik penyandang diabetes	14
3	Hasil uji <i>wilks' lambda</i> faktor-faktor yang mempengaruhi lipid darah penyandang diabetes pada MANOVA	19
4	Rata rata kadar HDL, LDL, TG terhadap jenis kelamin	19
5	Hasil uji t jenis kelamin penyandang diabetes pada	20
6	Hasil uji <i>wilks' lambda</i> pada faktor-faktor yang mempengaruhi lipid darah penyandang diabetes pada MANCOVA	21
7	Hasil uji <i>wilks' lambda</i> pada faktor-faktor yang mempengaruhi lipid darah penyandang diabetes pada R-MANCOVA	22
8	Rata rata HDL, LDL, TG berdasarkan kategori gangguan tidur, olahraga, dan jenis kelamin	22
9	Perbandingan nilai RMSE MANOVA, MANCOVA dan R-MANCOVA	23

DAFTAR GAMBAR

1	Diagram alir tahapan analisis data	13
2	Boxplot HDL, LDL dan TG penyandang diabetes berdasarkan kelompok gangguan tidur	15
3	Boxplot HDL, LDL dan TG penyandang diabetes berdasarkan kelompok konsumsi obat hipoglikemi oral	16
4	Boxplot HDL, LDL dan TG penyandang diabetes berdasarkan kelompok status merokok	16
5	Korelasi dari kadar HDL, LDL dan TG	17
6	Plot chi square dengan jarak mahalanobis (a) data sebelum transformasi, (b) data setelah transformasi	18
7	Pendeteksian pencilan menggunakan mahalanobis <i>Minimum Covariance Determinant</i> dan <i>classical</i>	18

DAFTAR LAMPIRAN

1	Lampiran 1 Boxplot HDL, LDL dan TG penyandang diabetes berdasarkan kelompok	30
2	Lampiran 2 Visualiasi homogenitas masing-masing peubah respon	30
3	Lampiran 3 Jarak Mahalanobis <i>Minimum Covariance Determinant</i>	31
4	Lampiran 4 Jarak Mahalanobis	32