

## Analisis Regresi dengan Variabel Dummy pada Faktor Risiko Diabetes Mellitus Tipe 2

Putri Diana<sup>1</sup>, Dewi Murni<sup>2</sup>

<sup>1,2</sup>Prodi Matematika, Fakultas Matematika Ilmu Pengetahuan dan Alam Universitas Negeri Padang (UNP)

---

### Article Info

#### Article history:

Received July 26, 2023

Revised August 22, 2023

Accepted September 20, 2023

---

#### Keywords:

Diabetes mellitus

Dummy variabel

Multiple regression analysis

#### Kata Kunci:

Diabetes Mellitus

Variabel Dummy

Analisis Regresi Berganda

### ABSTRACT

Diabetes Mellitus is a metabolic disorder that occurs when blood glucose levels exceed normal limits, that is 126 mg/dl. If high blood sugar levels are not controlled, this can lead to complications and even death. This study aims to create a multiple linear regression model and discover the factors that affect blood sugar that it raises the risk of diabetes mellitus. This research using secondary data obtained from the Dadok Tunggul Hitam Health Center. The samples used were 286 samples. Blood sugar level is the dependent variabel, and the independent variabels are age, abdominal circumference, body mass index, and gender. By using multiple linear regression analysis using one dummy variabel, obtains models:

$$Y' = 0,000157 - 0,000001 X_1 - 0,000001 X_3 + 0,000013 D$$

The model shows that with  $\alpha=5\%$ , the variable that significantly affect blood sugar are gender, age, and body mass index.

### ABSTRAK

Diabetes Mellitus merupakan salah satu gangguan metabolik terjadi saat kadar gula darah yang melewati ambang batas normal yaitu 126 mg/dl. Apabila kadar gula darah yang tinggi tidak terkontrol maka dapat mengakibatkan munculnya berbagai komplikasi bahkan kematian. Penelitian ini bertujuan membuat model regresi linear berganda untuk mengetahui faktor yang mempengaruhi gula darah sehingga berisiko mengalami diabetes mellitus. Penelitian ini menggunakan data sekunder. Sampel yang digunakan sebanyak 286 sampel yang diperoleh dari Puskesmas Dadok tunggul Hitam. Kadar gula darah adalah variabel terikat, dan variabel bebasnya adalah umur, lingkar perut, indeks massa tubuh, dan jenis kelamin. Dengan menggunakan analisis regresi linear berganda dengan satu variabel dummy, didapatkan model:

$$Y' = 0,000157 - 0,000001 X_1 - 0,000001 X_3 + 0,000013 D$$

Model tersebut menunjukkan bahwa dengan taraf kesalahan 5%, variabel yang signifikan mempengaruhi gula darah adalah jenis kelamin, usia, dan indeks massa tubuh.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



---

### Penulis pertama

(Putri Diana)

Program studi Matematika, Departemen Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Padang, Jl. Prof. Dr. Hamka, Air Tawar barat, Padang Utara, Padang, Indonesia. Kode Pos : 25131  
Email: [diaputridian1@gmail.com](mailto:diaputridian1@gmail.com)



## 1. PENDAHULUAN

Sejalan dengan kemajuan teknologi juga terjadi perubahan pada pola hidup masyarakat. Pola konsumsi makanan tradisional yang mengandung banyak nutrisi telah bergeser kepada kebiasaan konsumsi *fast food* dan *junk food* yang minim nutrisi dan mengandung banyak gula, garam, dan lemak. Apabila konsumsi *fast food* dan *junk food* dilakukan dalam jangka waktu yang lama dan berlebihan, ternyata bisa membahayakan tubuh. Salah satunya adalah bisa memicu obesitas [1]. Obesitas adalah salah satu penyebab terjadinya resistensi insulin yang dapat memicu diabetes mellitus[2]. Salah satu indikasi orang terkena diabetes mellitus adalah dengan melakukan pengecekan kadar gula darah. Waktu yang paling tepat untuk melakukan pengecekan gula darah adalah dipagi hari atau setelah berpuasa selama 8-10 jam. Gula darah yang dicek pada waktu tersebut dikenal sebagai gula darah puasa. Diagnosa diabetes mellitus akan dilakukan saat kadar gula darah puasa melebihi 126 mg/dl [3].

Jika gula darah tidak terkontrol, maka komplikasi seperti gagal ginjal, kebutaan, jantung koroner, amputasi, bahkan kematian berpeluang besar terjadi pada penderita DM. Untuk menghindari kenaikan gula darah ini, maka perlu diketahui faktor yang menyebabkan kenaikan gula darah. Hal ini penting untuk dilakukan karena kasus diabetes mellitus sering terjadi dan mengalami peningkatan setiap tahunnya. Indonesia adalah salah satu dari 10 negara dengan kasus DM terbanyak di dunia. Jumlah kasus DM pada tahun 2021 yang terdiagnosa adalah 19,5 juta [4]. Prevalensi DM di Sumatera Barat pada tahun 2018 berdasarkan diagnosa dokter adalah 1,64%. Kota Padang termasuk kota dengan prevalensi DM lebih tinggi daripada prevalensi provinsi yaitu 2,47% [5].

Dengan pengendalian faktor risiko, penyakit diabetes bisa dicegah dan tingkat fatalitasnya bisa ditekan. Faktor risiko DM terdiri dari faktor risiko yang bisa diubah seperti obesitas, pola hidup tidak sehat, dan faktor yang tidak bisa diubah seperti usia, jenis kelamin, dan riwayat keluarga menderita DM[3].

Salah satu cara untuk mengidentifikasi faktor risiko DM adalah analisis regresi linear. Regresi linear merupakan salah satu metode analisis statistika untuk melihat hubungan variabel terikat Y dengan variabel bebas X [6]. Dalam analisis regresi, variabel bebas tidak selalu bersifat kuantitatif. Dalam beberapa keadaan ditemukan variabel bebas yang bersifat kualitatif (kategorik). Untuk memasukkan variabel kualitatif ke dalam perhitungan regresi diperlukan variabel buatan dengan menggunakan nominal 0 dan 1. Variabel buatan ini disebut dengan variabel dummy [7]. Karena dalam faktor risiko diabetes terdapat faktor yang bersifat kategorik, maka dalam penelitian ini digunakan analisis regresi dengan variabel dummy pada faktor risiko penyakit diabetes mellitus.

## 2. METODE

Sumber data penelitian ini adalah data sekunder yang didapatkan dari rekam medik Puskesmas Dadok Tunggul Hitam, Kota Padang. Sampel yang digunakan pasien yang melakukan pengukuran kadar gula darah di Puskesmas Dadok Tunggul Hitam selama periode Juli-Desember tahun 2022 yang berjumlah 286 sampel. Kadar gula darah merupakan variabel terikat pada penelitian ini, sementara usia, lingkar perut, indeks massa tubuh (IMT), dan jenis kelamin sebagai variabel bebas.

Dalam penelitian ini, analisis data dilakukan dengan tahapan berikut:

1. Melakukan pengumpulan data yang diperlukan.
2. Mengidentifikasi variabel terikat, variabel bebas, dan membangun variabel dummy.

Variabel dummy adalah representasi kuantitatif dari variabel yang bersifat kualitatif. Karena pada dasarnya analisis regresi tidak bisa menggunakan data kualitatif, maka data kualitatif tersebut harus di kuantifikasi dengan menggunakan nilai 1 dan 0 [8]. Pada penelitian ini nilai D=1 untuk perempuan dan nilai D=0 untuk laki-laki.

Untuk memasukkan variabel dummy ke dalam model regresi, terdapat tiga cara yang bias dilakukan sebagai berikut:

- a. Variabel dummy sebagai komponen intersep, dengan persamaan regresi:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 D_i + \varepsilon_i \quad i = 1, 2, \dots, n \quad (1)$$

Pada persamaan di atas  $D$  merupakan variabel intersep sehingga akan terbentuk dua persamaan di mana perbedaannya terletak pada nilai intersep, yaitu :

$$Y_i = (\beta_0 + \beta_2) + \beta_1 X_{1i} + \varepsilon_i \quad \text{pada saat kasus } D \text{ bernilai } 1$$

- $Y_i = \beta_0 + \beta_1 X_{1i} + \varepsilon_i$  , pada saat kasus D bernilai 0
- b. Variabel dummy sebagai komponen *slope*, dengan persamaan regresi:  
 $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 D_i X_{1i} + \varepsilon_i \quad i = 1, 2, \dots, n$  (2)  
 Dalam keadaan ini, model regresi yang diperoleh mempunyai intersep yang sama yaitu  $\beta_0$ , tetapi *slope*-nya ada dua, yaitu sebesar  $\beta_1 + \beta_2$  pada kasus  $D = 1$  dan sebesar  $\beta_1$  pada kasus  $D = 0$ .
- c. Variabel dummy sebagai komponen intersep dan *slope*, dengan persamaan regresi:  
 $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 D_i + \beta_3 D_i X_{1i} + \varepsilon_i \quad i = 1, 2, \dots, n$  (3)
3. Membuat model regresi awal dengan seluruh variabel dan estimasi parameter regresi linear berganda.  
 Dengan  $n$  observasi dan  $k$  variabel bebas, model regresi linear berganda bisa dirumuskan sebagai berikut :
- $$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + \varepsilon_i$$
- (4)
- 
- Dimana :
- 
- $Y_i$
- : Variabel terikat pada observasi ke-
- $i$
- 
- $X$
- : Variabel bebas
- 
- $\beta_0$
- : Konstanta
- 
- $\beta_k$
- : Parameter regresi pada variabel bebas ke-
- $k$
- 
- $\varepsilon_i$
- : Residual ke-
- $i$
- 
- $i$
- : Banyaknya observasi
- 
- $k$
- : Jumlah variabel prediktor
- Salah satu cara pendugaan parameter regresi yaitu dengan metode kuadrat terkecil atau *Ordinary Least Square* (OLS). Untuk mengestimasi parameter  $\beta$  dilakukan dengan meminimumkan jumlah kuadrat galat [8]. Penduga parameter dengan OLS dapat ditulis dalam bentuk matriks:
- $$\hat{\beta} = (X^T X)^{-1} X^T Y$$
- (5)
4. Melakukan uji signifikansi parameter regresi  
 Uji F adalah uji yang digunakan untuk menguji signifikansi parameter regresi secara simultan. Hipotesis yang digunakan adalah:  
 $H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$   
 $H_1 : \beta_k \neq 0$   
 Statistic uji F ini dapat diperhatikan pada tabel analisis varian berikut :

Tabel 1. Analisis Varian (ANOVA)

Sumber variansi	Df	Jumlah Kuadrat (JK)	Rataan Kuadrat (RK)	$F_{hitung}$
Regresi	K	$\hat{\beta}^T X^T Y - \frac{(\sum_{i=1}^n Y_i)^2}{n}$	$\frac{JKR}{k}$	$\frac{RKR}{RKG}$
Galat	$n-k-1$	$Y^T Y - \hat{\beta}^T X^T Y$	$\frac{JKG}{n-k-1}$	
Total	$n-1$	$Y^T Y - \frac{(\sum_{i=1}^n Y_i)^2}{n}$		

Pengambilan keputusan tolak  $H_0$  Apabila  $F_{hitung} > F_{(k, n-k-1)}$  [9].

Uji signifikansi parameter secara parsial bisa dilakukan dengan uji  $t$ . Hipotesis ujinya adalah sebagai berikut :

$$H_0: \beta_k = 0$$

$$H_1: \beta_k \neq 0$$

Statistik uji  $t$ :



$$t_{hitung} = \frac{\hat{\beta}_k}{s\{b_k\}} \quad (3)$$

Dimana:

$\hat{\beta}_k$  : Penduga parameter ke- $k$   
 $s\{b_k\}$  diperoleh dari diagonal ke- $k$  dari  $s^2(b)$   
 dengan  $s^2(b) = RKG (X^T X)^{-1}$

Apabila  $|t_{hitung}| > t_{(1-\frac{\alpha}{2}; n-k-1)}$  maka tolak  $H_0$  [10].

5. Melakukan uji asumsi klasik

Dalam analisis regresi linear berganda ada asumsi yang harus dipenuhi, diantaranya adalah kelinearan, kenormalan sisaan, rata-rata sisaan sama dengan nol, kohomogenan ragam sisaan, autokorelasi, dan multikolinearitas. Uji kelinearan dilakukan untuk memastikan adanya hubungan linear yang antara variabel terikat dan variabel bebas. Uji normalitas dilakukan untuk memastikan apakah nilai residual pada persamaan regresi sudah mengikuti distribusi normal. Untuk menguji normalitas persamaan regresi dapat dilakukan dengan uji *Anderson-Darling* [7]. Rata-rata residual dari persamaan regresi harus mendekati nol. Kehomogenan ragam sisaan adalah kesamaan varians residual pada setiap pengamatan [11]. Uji homoskedastisitas bisa diuji dengan uji *Breusch-pagan*. Autokorelasi adalah terjadinya korelasi residual antar pengamatan. Model regresi diharapkan tidak mempunyai korelasi residual antar pengamatan [12]. Untuk mendeteksi autokorelasi pada model dapat dideteksi melalui nilai Durbin-Watson [7]. Multikolinearitas adalah adanya korelasi antar variabel bebas dalam satu pengamatan. Model regresi yang baik diharapkan tidak mempunyai korelasi antar variabel bebasnya. Multikolinearitas bisa dideteksi dengan nilai VIF [13].

6. Melakukan transformasi data

Apabila salah satu uji asumsi klasik tidak terpenuhi, maka diperlukan suatu transformasi data. Transformasi data dilakukan dengan merubah bentuk skala pengukuran data ke dalam bentuk lain supaya memenuhi uji asumsi. Ada dua jenis transformasi data yang mungkin dilakukan, yaitu transformasi pada  $Y$  dan transformasi pada  $X$ . Transformasi pada  $Y$  dilakukan dengan transformasi Box-Cox apabila asumsi kenormalan belum terpenuhi. Transformasi pada  $X$  dilakukan dengan transformasi Box-Tidwell apabila ada variabel yang belum memenuhi asumsi kelinearan [8]. Setelah dilakukan transformasi pada data, selanjutnya kembali ke langkah awal, yaitu pembentukan model regresi, estimasi parameter regresi, dan melakukan uji signifikansi parameter regresi.

7. Pemilihan model terbaik

Salah satu cara pemilihan model terbaik dilakukan dengan *All Possible Regresssion*. Dengan menggunakan metode *All Possible Regresssion* semua kombinasi yang mungkin bisa dipertimbangkan. Banyak kombinasi yang mungkin untuk  $k$  variabel bebas adalah sebanyak  $2^k - 1$  persamaan regresi [14]. Setiap persamaan harus dipertimbangkan dengan melihat beberapa kriteria. Kriteria dalam penilaian model adalah nilai  $R^2_{adj}$  terbesar, nilai  $S^2$  terkecil, dan nilai  $C_p$  yang mendekati banyaknya parameter dalam model [15].

8. Melakukan interpretasi model

### 3. HASIL DAN PAMBAHASAN

Berdasarkan pengolahan data diperoleh model regresi linear berganda dengan satu variabel dummy pada faktor risiko diabetes mellitus sebagai berikut :

$$Y = 75 + 1,82X_1 + 0,38X_2 - 1,59X_3 - 33,40D - 0,72X_2D + 3,21X_3D$$

Model ini memiliki nilai  $R^2_{adj}$  sebesar 10,4%. Artinya variabel-variabel bebas bisa menjelaskan variabel respon sebesar 10,4%. Sementara sisanya 89,6% dijelaskan oleh variabel lain yang tidak diperhitungkan pada penelitian ini.

Tabel 2. *Analysis of Variance*

Source	DF	SS	MS	F	P-Value
Regression	6	230544	38424	6,5	0,000

(nama penulis pertama)

Residual Error	279	1649052	5911
Total	285	1879596	

Selanjutnya dilakukan uji F untuk menguji signifikansi parameter secara simultan. Berdasarkan hasil uji *analysis of variance* pada tabel 2 diperoleh nilai  $F_{hitung}$  6,5. Jika dibandingkan dengan nilai  $F_{(6,279)} = 2,1$  pada table distribusi F, maka nilai  $F_{hitung} > F_{tabel}$ . Karena nilai  $F_{hitung} > F_{tabel}$  maka tolak  $H_0$ , jadi minimal ada satu variabel bebas yang signifikan mempengaruhi variabel terikat.

Tabel 3. *Coefficient*

Predictor	Coef	SE Coef	T	P-Value	VIF
Constant	75,010	86,340	0,87	0,386	
$X_1$	1,817	0,336	5,40	0,000	1,031
$X_2$	0,384	1,023	0,38	0,707	6,219
$X_3$	-1,595	2,715	-0,59	0,557	5,374
D	-33,400	96,830	-0,34	0,730	91,398
$X_2D$	-0,717	1,137	-0,63	0,528	104,320
$X_3D$	3,208	3,072	1,04	0,297	60,264

Selain uji F juga dilakukan uji t. tujuan dilakukan uji t adalah untuk menguji signifikansi setiap variabel bebas. Pada tabel 3 dipaparkan hasil analisis *coefficient*, pada tabel 3 tersebut diperoleh nilai  $|t_{hitung}|$  setiap variabel bebas. Jika dilihat pada table distribusi t, nilai  $t_{0,975;279}$  adalah 1,96. Jika dibandingkan nilai  $|t_{hitung}|$  setiap variabel bebas dengan nilai  $t_{tabel}$ , maka terlihat bahwa nilai  $|t_{hitung}|$  variabel  $X_2$ ,  $X_3$ , D,  $DX_2$ ,  $DX_3 < 1,96$ , artinya variabel  $X_2$ ,  $X_3$ , D,  $DX_2$ ,  $DX_3$  secara parsial tidak signifikan berpengaruh terhadap variabel terikat. Sementara nilai  $|t_{hitung}|$  variabel  $X_1 > 1,96$ , sehingga dikatakan bahwa variabel  $X_1$  berpengaruh signifikan terhadap variabel respon.

Selanjutnya dilakukan uji asumsi klasik terhadap model yang diperoleh. Berdasarkan hasil uji linearitas, terlihat bahwa terdapat hubungan linear antara variabel terikat dan variabel bebas. Linearitas dapat diuji dengan uji Ramsey-Reset. Dari hasil uji diperoleh nilai p-value  $X_1$ ,  $X_2$ , dan  $X_3$  adalah 0,647; 0,506; 0,497. Karena p-value  $> 0,05$ , maka antara variabel terikat dan variabel bebas memiliki hubungan linear.

Asumsi normalitas dilihat dari sebaran residual yang mengikuti distribusi normal dan uji Kolmogorov-Smirnov. Hasil uji Kolmogorov-Smirnov mempunyai nilai p-value sebesar 0,010. Karena  $0,010 < 0,05$ , maka tolak  $H_0$ . Jadi asumsi kenormalan belum terpenuhi.

Rata-rata residual dari persamaan regresi diatas adalah  $3,011 \times 10^{-14}$ . Jadi nilai rata-rata residual sudah mendekati nilai 0.

Untuk melihat adanya heteroskedastisitas dalam model, dapat dideteksi dengan mengamati grafik dan uji *breusch-pagan*. Hasil uji menunjukkan nilai p-value sebesar  $2,2 \times 10^{-16}$ . Karena nilai p-value  $< 0,05$  maka tolak  $H_0$ , artinya terdapat heteroskedastisitas pada model regresi. Jadi asumsi kehomogenan belum terpenuhi.

Untuk mendeteksi adanya autokorelasi pada model dilakukan uji Durbin-Watson. Nilai uji Durbin-Watson pada hasil uji adalah 1,999. Karena nilainya mendekati 2 maka diindikasikan tidak terdapat autokorelasi. Jadi asumsi autokorelasi telah terpenuhi.

Uji multikolinearitas bisa dilihat dari nilai VIF setiap variabel. Pada tabel 3 diperoleh nilai VIF setiap variabel. Suatu model dikatakan terbebas dari multikolinearitas adalah apabila nilai VIF semua variabelnya  $< 10$ . Karena pada model ini terdapat beberapa variabel dengan nilai  $VIF > 10$ , maka terdapat multikolinearitas pada model. Jadi asumsi multikolinearitas belum terpenuhi.

Pada uji multikolinearitas didapat beberapa variabel yang mempunyai nilai  $VIF > 10$ , yaitu variabel D,  $DX_2$ , dan  $DX_3$ . Oleh karena itu variabel  $DX_2$  dan  $DX_3$  akan dihilangkan dari model. Karena asumsi normalitas belum terpenuhi, maka perlu dilakukan transformasi pada Y dengan menggunakan transformasi *Box-Cox*. Setelah dilakukan transformasi  $1/Y^2$ , maka proses analisis kembali ke langkah awal.



Pembentukan model setelah transformasi pada variabel  $Y$  dengan seluruh variabel prediktor yaitu variabel jenis kelamin ( $D$ ), usia ( $X_1$ ), lingkar perut ( $X_2$ ), dan IMT ( $X_3$ ) diperoleh model :

$$Y' = 0,000170 - 0,000001 X_1 - 0,000000 X_2 - 0,000001 X_3 + 0,000013D$$

Model ini memiliki nilai  $R^2_{adj}$  sebesar 16,20%. Artinya variabel-variabel bebas telah menjelaskan variabel respon sebesar 16,20%, sementara sisanya sebesar 83,80% nilai variabel respon dijelaskan oleh variabel lain yang tidak diperhitungkan dalam penelitian ini.

Tabel 4. *Analysis of Variance*

Source	DF	SS	MS	F	P-Value
Regression	4	$8,724 \times 10^{-8}$	$2,181 \times 10^{-8}$	14,75	0,000
Residual Error	281	$4,156 \times 10^{-7}$	$1,479 \times 10^{-9}$		
Total	285	$5,029 \times 10^{-7}$			

Berdasarkan hasil *analysis of variance* pada tabel 4, diperoleh nilai  $F_{hitung}$  sebesar 14,75. Jika dilihat pada table distribusi F, nilai  $F_{(4,281)}$  dengan taraf signifikansi 5% paa tabel distribusi F adalah 2,10. Karena  $14,75 > 2,10$  maka tolak  $H_0$ , jadi minimal ada satu variabel bebas yang berpengaruh signifikan terhadap variabel terikat.

Tabel 5. *Coefficient*

Predictor	Coef	SE Coef	T	P-Value	VIF
Constant	$1,704 \times 10^{-4}$	$2,149 \times 10^{-5}$	7,93	0,000	
$X_1$	$-1,110 \times 10^{-6}$	$1,700 \times 10^{-7}$	-6,56	0,000	1,022
$X_2$	$-2,200 \times 10^{-7}$	$2,200 \times 10^{-7}$	-0,97	0,332	1,164
$X_3$	$-1,200 \times 10^{-7}$	$6,300 \times 10^{-7}$	-1,89	0,060	1,172
D	$1,315 \times 10^{-5}$	$5,150 \times 10^{-6}$	2,55	0,011	1,034

Berdasarkan analisis koefisien pada tabel 5, diperoleh nilai  $t_{hitung}$  masing-masing variabel bebas. Jika dilihat pada table distribusi t, nilai  $t_{0,975;281}$  adalah 1,96. Jika dibandingkan nilai  $|t_{hitung}|$  setiap variabel prediktor dengan nilai  $t_{tabel}$ , maka terlihat bahwa nilai  $|t_{hitung}|$  variabel  $X_2$  dan  $X_3 < 1,96$ , artinya variabel  $X_2$  dan  $X_3$  secara parsial, tidak signifikan mempengaruhi variabel terikat. Sementara nilai  $|t_{hitung}|$  variabel D dan  $X_1 > 1,96$ , artinya variabel D dan  $X_1$  secara parsial berpengaruh signifikan terhadap variabel respon.

Selanjutnya dilakukan uji asumsi klasik pada persamaan regresi setelah dilakukan transformasi data.

Berdasarkan hasil uji linearitas, terlihat bahwa ada hubungan linear antara variabel bebas dan variabel terikat. Untuk menguji kelinearan dalam model bisa dilakukan dengan uji Ramsey-Reset. Dari hasil uji diperoleh nilai p-value variabel  $X_1$ ,  $X_2$ , dan  $X_3$  adalah 0,280; 0,373; 0,702. Karena p-value  $> 0,05$ , maka asumsi linearitas terpenuhi.

Asumsi normalitas dilihat dari sebaran residual yang mengikuti distribusi normal dan uji Kolmogorov-Smirnov. Nilai p-value pada hasil uji Kolmogorov-Smirnov adalah 0,15. Karena p-value  $> 0,05$ , maka terima  $H_0$ . Jadi data sudah berdistribusi normal.

Rata-rata residual dari persamaan regresi adalah  $1,59 \times 10^{-19}$ . Jadi nilai rata-rata residual sudah mendekati nilai 0.

Untuk melihat adanya heteroskedastisitas dalam model, dapat dilihat pada grafik dan juga bisa dideteksi dengan melakukan uji *breusch-pagan*. Hasil uji menunjukkan nilai p-value sebesar 0,17. Karena nilai p-value  $> 0,05$  maka terima  $H_0$ , artinya tidak terdapat heteroskedastisitas pada model regresi. Jadi asumsi kehomogenan telah terpenuhi.

Uji autokorelasi bisa dilakukan dengan uji Durbin-Watson. Nilai uji Durbin-Watson pada hasil uji adalah 2,025. Karena nilainya mendekati 2 maka diindikasikan tidak terdapat autokorelasi. Jadi asumsi autokorelasi telah terpenuhi.

Uji multikolinearitas bisa dilihat dari nilai VIF masing-masing variabel. Pada hasil uji terlihat nilai VIF variabel  $X_1$ ,  $X_2$ ,  $X_3$ , D berturut-turut adalah 1,017; 1,142; 1,151; 1,024. Karena pada model ini semua

variabel prediktor mempunyai nilai  $VIF < 10$ , maka tidak terdapat multikolinearitas pada model. Jadi asumsi multikolinearitas telah terpenuhi.

Karena semua asumsi terpenuhi, maka dilanjutkan dengan memilih model terbaik dengan cara *All Possible Regression*. Dari semua variabel bebas, diperoleh jumlah kombinasi model regresi sebanyak 15 kombinasi yang dikelompokkan ke dalam 4 kelompok.

Tabel 6. Kombinasi Persamaan Regresi

No	Grup	Model Regresi	$R^2_{adj}$	$S^2$	$C_p$
1	A	$Y = 0,000134 - 0,000001 X_1$	13,5%	$1,5167 \times 10^{-9}$	10,933
2		$Y = 0,000111 - 0,000000 X_2$	0,2%	$1,4918 \times 10^{-9}$	56,202
3		$Y = 0,000113 - 0,000001 X_3$	0,7%	$1,4662 \times 10^{-9}$	54,575
4		$Y = 0,000075 + 0,000016 D$	2,7%	$1,476 \times 10^{-9}$	47,645
1	B	$Y = 0,000163 - 0,000001 X_1 - 0,000000 X_2$	14,0%	$1,7566 \times 10^{-9}$	10,309
2		$Y = 0,000165 - 0,000001 X_1 - 0,000001 X_3$	14,6%	$1,7097 \times 10^{-9}$	8,289
3		$Y = 0,000123 - 0,000001 X_1 + 0,000012 D$	14,8%	$1,6969 \times 10^{-9}$	7,788
4		$Y = 0,000122 - 0,000000 X_2 + 0,000001 X_3$	0,4%	$1,7240 \times 10^{-9}$	56,091
5		$Y = 0,000103 - 0,000000 X_2 + 0,000017 D$	3,1%	$1,7149 \times 10^{-9}$	47,113
6		$Y = 0,000106 - 0,000001 X_3 + 0,000018 D$	3,8%	$1,7169 \times 10^{-9}$	44,682
1	C	$Y = 0,000177 - 0,000001 X_1 - 0,000000 X_2 - 0,000001 X_3$	14,5%	$1,5080 \times 10^{-9}$	9,512
2		$Y = 0,000155 - 0,000001 X_1 - 0,000000 X_2 + 0,000012 D$	15,4%	$1,4926 \times 10^{-9}$	6,580
3		$Y = 0,000157 - 0,000001 X_1 - 0,000001 X_3 + 0,000013 D$	16,2%	$1,4788 \times 10^{-9}$	3,945
4		$Y = 0,000117 - 0,000000 X_2 - 0,000001 X_3 + 0,000018 D$	3,7%	$1,6993 \times 10^{-9}$	45,992
1	D	$Y = 0,000170 - 0,000001 X_1 - 0,000000 X_2 - 0,000001 X_3 + 0,000013 D$	16,2%	$1,4791 \times 10^{-9}$	5,000

Pada tabel 6 ditampilkan nilai  $R^2_{adj}$ ,  $S^2$ , dan nilai  $C_p$  Mallows dari semua kombinasi persamaan regresi. Nilai  $R^2_{adj}$ ,  $S^2$ , dan nilai  $C_p$  Mallows tersebut akan dijadikan sebagai dasar pertimbangan dalam pemilihan model terbaik. Dari semua persamaan regresi pada tabel 2 akan dipilih calon model terbaik pada masing-masing kelompok dengan kriteria  $R^2_{adj}$  terbesar,  $S^2$  terkecil, dan  $C_p$  Mallows yang mendekati jumlah parameter. Calon model terbaik masing-masing kelompok pada tabel 6 akan disajikan dalam tabel 7.

Tabel 7. Perbandingan Calon Model Terbaik

Grup	Predictor	Coef	SE Coef	T	P	VIF	$R^2_{adj}$	$S^2$	$C_p$
C3	Constant	0,0001569	0,0000165	9,53	0,000				
	$X_1$	-0,0000011	0,0000002	-6,54	0,000	1,021	16,2 %	$1,4788 \times 10^{-9}$	3,945
	$X_3$	-0,0000014	0,0000006	-2,42	0,016	1,014			
	D	0,0000130	0,0000052	2,52	0,012	1,033			
D1	Constant	0,0001704	0,0000215	7,93	0,000				
	$X_1$	-0,0000011	0,0000002	-6,56	0,000	1,022	16,2 %	$1,4791 \times 10^{-9}$	5,000
	$X_2$	-0,0000002	0,0000002	-0,97	0,332	1,164			
	$X_3$	-0,0000012	0,0000006	-1,89	0,060	1,172			
	D	0,0000132	0,0000052	2,55	0,011	1,034			

Tabel 7 menampilkan dua calon model terbaik dari seluruh kombinasi persamaan regresi. Model tersebut adalah model C3 dan D1. Model C3 dan D1 sama-sama memiliki nilai  $S^2$  yang kecil, nilai  $C_p$  Mallows yang mendekati banyaknya parameter, dan memiliki  $R^2_{adj}$  yang sama besar. Untuk memilih model terbaik



diantara dua model tersebut, maka dilihat signifikansi variabel bebas pada masing-masing model tersebut. Hasil uji signifikansi parameter pada model C3 menunjukkan bahwa semua variabel bebas yang berada pada model C3 signifikan karena nilai  $|t_{hitung}| > 1,96$ . Sementara pada model D1 terdapat dua variabel yang tidak signifikan karena  $|t_{hitung}| < 1,96$ . Dua variabel tersebut adalah variabel  $X_2$  dan  $X_3$ . Jadi diputuskan bahwa model terbaiknya adalah model C3 yang memuat tiga variabel bebas yaitu umur, IMT, dan jenis kelamin.

Berdasarkan analisis data didapatkan model regresi terbaik dari faktor risiko diabetes adalah:

$$Y' = 0,000157 - 0,000001 X_1 - 0,000001 X_3 + 0,000013 D$$

Untuk D=1 (perempuan) model regresinya adalah:

$$Y' = 0,000170 - 0,000001 X_1 - 0,000001 X_3$$

Sedangkan untuk D=0 (laki-laki) model regresinya adalah:

$$Y' = 0,000157 - 0,000001 X_1 - 0,000001 X_3$$

Dari model tersebut terlihat bahwa konstanta pada model untuk D=1 (perempuan) lebih besar daripada nilai konstanta pada model untuk D=0 (laki-laki). Dengan demikian bisa dikatakan bahwa perempuan cenderung mempunyai gula darah yang lebih tinggi dibanding laki-laki sehingga perempuan mempunyai risiko lebih tinggi terkena diabetes mellitus. Variabel yang signifikan memengaruhi tingginya kadar gula darah yang menyebabkan seseorang berisiko menderita diabetes mellitus adalah usia dan IMT. Variabel usia dan IMT mempunyai hubungan negatif dengan gula darah, yang dapat diartikan apabila terjadi peningkatan umur dan IMT maka kadar gula darah akan menurun. Hal ini tidak sesuai dengan teori yang dipaparkan sebelumnya yang menyatakan bahwa usia dan IMT berpengaruh positif terhadap gula darah. Hal ini bisa disebabkan oleh beberapa faktor. Salah satunya karena ada faktor-faktor lain yang tidak diikutsertakan ke dalam penelitian ini seperti faktor genetik, aktifitas fisik, pola makan, dan banyak faktor lainnya. Meningkatnya kasus diabetes mellitus pada usia muda beberapa tahun terakhir ini juga bisa menyebabkan hasil penelitian bertolak belakang dengan teori yang dikemukakan.

#### 4. KESIMPULAN

Bentuk model regresi linear berganda dengan variabel dummy yang diperoleh pada faktor-faktor risiko diabetes mellitus adalah sebagai berikut:

$$Y' = 0,000157 - 0,000001 X_1 - 0,000001 X_3 + 0,000013 D$$

Dari model tersebut terlihat bahwa faktor yang signifikan membuat seseorang berisiko mengalami peningkatan gula darah yang berujung pada diabetes mellitus usia, IMT, dan jenis kelamin. Sedangkan lingkaran perut tidak mempunyai pengaruh yang signifikan.

#### REFERENSI

- [1] Y. H. Tarihoran dan D. F. Silaban, "Hubungan Pola Makan dengan Kejadian Diabetes Mellitus di Puskesmas Namorambe Kabupaten Deli Serdang," *Penelit. Keperawatan Med.*, vol. 4, 2022.
- [2] R. Yosmar, D. Almasdy, dan F. Rahma, "Jurnal Sains Farmasi Dan Klinis," *Survei risiko penyakit diabetes melitus terhadap Kesehatan. Masy. Kota Padang*, vol. 5, hal. 134–141, 2018.
- [3] Infodatin, *Tetap Produktif, Cegah, dan Atasi Diabetes Mellitus*. Kemenkes RI, 2020.
- [4] Internasional Diabetes Federation, *IDF Diabetes Atlas*, 10 ed. 2021. doi: 10.1016/j.diabres.2013.10.013.
- [5] Riskesdas, *Laporan Nasional Riskesdas 2018*. Jakarta: Balitbangkes Kemenkes RI, 2019.
- [6] N. Draper dan H. Smith, *Analisis Regresi Terapan*, Dua. Jakarta: Gramedia Pustaka Utama, 1992.
- [7] D. N. Gujarati, *Basic Econometrics*, Fourth. New York: McGraw-Hill, 2004.
- [8] R. K. Sembiring, *Analisis Regresi*. Bandung: ITB, 1995.
- [9] Sugiarto, *Tahap Awal + Aplikasi Analisis Regresi*. Yogyakarta: Andi Offset, 1992.
- [10] R. Kurniawan dan B. Yuniarto, *Analisis Regresi Dasar dan Penerapannya dengan R*. Jakarta: Kencana, 2016.
- [11] M. F. Qudratullah, *Analisis Regresi Terapan: Teori, Contoh, Kasus, dan Aplikasi dengan SPSS*. Yogyakarta: Andi Offset, 2014.
- [12] I. Ghozali, *Analisis Multivariat dengan IBM SPSS 25*. Semarang: Universitas Diponegoro, 2016.
- [13] D. C. Montgomery, G. C. Runger, dan N. F. Hubele, *Engineering Statistic*. United States: John Wiley & Sons Inc., 2010.
- [14] R. . Walpole, *Probability and Statistic for Engineers and Cientits*. Bonston: Pearson Education, 2012.
- [15] N. Amalita dan Y. Kurniawati, "Model Regresi Dummy dalam Memprediksi Performansi Akademik Mahasiswa Jurusan Matematika FMIPA UNP," *Pros. Semirata FMIPA Univ. Lampung*, 2013.