

ANALISIS REGRESI LOGISTIK BINER PADA PENYAKIT
POLYCYSTIC OVARY SYNDROME (PCOS)

Tryadvensya Purba¹, Rahmawati Pane²
Universitas Sumatera Utara
E-mail: trypurba15@gmail.com¹

INFORMASI ARTIKEL	
Submitted	: 2024-12-31
Review	: 2024-12-31
Accepted	: 2024-12-31
Published	: 2024-12-31
KATA KUNCI	

Regresi Logistik Biner,
Polycystic Ovary Syndrome.

Keywords: Binary
Logistic Regression,
Polycystic Ovary
Syndrome.

A B S T R A K

Regresi logistik biner adalah teknik analisis statistik yang berguna untuk menganalisis hubungan antara variabel bebas dengan variabel tak bebas yang bersifat biner. Dalam bidang kesehatan, regresi logistik biner dapat digunakan untuk menganalisis pengaruh gaya hidup, diagnosis penyakit, evaluasi risiko kesehatan serta memahami penyebaran penyakit diantaranya pada penyakit Polycystic Ovary Syndrome. Polycystic Ovary Syndrome adalah gangguan endokrin yang kompleks dengan banyak faktor risiko dan komplikasi yang umumnya terjadi pada wanita usia reproduktif. Tujuan penelitian ini adalah untuk mengetahui faktor-faktor yang berpengaruh secara signifikan terhadap penyakit Polycystic Ovary Syndrome (PCOS) dengan menggunakan regresi logistik biner. Penelitian ini menggunakan data pasien Polycystic Ovary Syndrome yang diperoleh dari website kaggle, dengan sampel penelitian sebanyak 300 pasien dan terdiri dari 8 variabel bebas. Hasil penelitian menunjukkan bahwa tiga variabel bebas berpengaruh signifikan terhadap penyakit Polycystic Ovary Syndrome, yaitu Indeks Massa Tubuh kategori obesitas ($X_2^{>3}$), diabetes (X_3) dan kadar Anti Müllerian Hormone (X_7).

A B S T R A C T

Binary logistic regression is a statistical analysis technique that is useful for analyzing the relationship between independent variables and dependent variables that are binary. In the health field, binary logistic regression can be used to analyze the influence of lifestyle, disease diagnosis, health risk evaluation, and understanding disease distribution, including Polycystic Ovary Syndrome (PCOS). Polycystic Ovary Syndrome is a complex endocrine disorder with many risk factors and complications that commonly occur in women of reproductive age. The purpose of this study was to determine the factors that significantly influence Polycystic Ovary Syndrome using binary logistic regression. This study used Polycystic Ovary Syndrome data obtained from the kaggle website, with a sample of 300 patients and consisting of 8 independent variables. The result showed that three independent variables

significantly influenced Polycystic Ovary Syndrome, namely body mass index for the obesity category (X_2^3), diabetes (X_3) and Anti Müllerian Hormone levels (X_7).

PENDAHULUAN

Statistika didefinisikan sebagai ilmu yang berhubungan dengan pengumpulan data, pengolahan, analisis, interpretasi dan penyajian data dengan menerapkan prinsip-prinsip probabilitas dalam pembuatan keputusan berdasarkan data (Bluman, 2018). Statistika mempunyai manfaat dalam memberikan pengetahuan dan kemampuan untuk melakukan evaluasi terhadap data.

Dalam ilmu statistika terdapat teknik analisis yang mampu menjelaskan bentuk hubungan antara variabel-variabel yang berkaitan dengan sebab-akibat. Teknik analisis ini dinamakan analisis regresi. Analisis regresi merupakan teknik analisis statistik yang digunakan untuk menaksir pola hubungan sebab-akibat antara variabel bebas dan variabel tak bebas (Khasanah, 2021). Variabel tak bebas (Y) adalah variabel yang nilainya ditentukan oleh variabel lain, sedangkan variabel bebas (X) adalah variabel yang nilainya dapat ditentukan secara bebas berdasarkan dugaan bahwa variabel tersebut memiliki pengaruh terhadap variabel tak bebas.

Analisis regresi berdasarkan bentuknya terbagi atas dua bagian, yaitu regresi linear dan regresi non-linear. Regresi non-linear adalah jenis analisis regresi dimana hubungan antara variabel bebas dan tak bebas tidak diwakili oleh persamaan linear. Analisis regresi non-linier dapat dibagi kedalam beberapa kategori berdasarkan pendekatan dan model yang digunakan. Salah satu bagian dari analisis regresi non-linier berdasarkan sifatnya adalah model regresi logistik.

Model regresi logistik mengasumsikan bentuk yang lebih rinci dari fungsi logistik dan mengestimasi parameter-parameter model berdasarkan data. Model regresi logistik disebut juga sebagai model logit yang berguna untuk memperkirakan kemungkinan dari suatu kejadian dengan data fungsi logit dari kurva logistik (Gudono, 2017). Salah satu model regresi logistik adalah regresi logistik biner.

Regresi logistik biner adalah teknik analisis statistik yang berguna untuk menganalisis hubungan antara variabel bebas dengan variabel tak bebas yang bersifat biner. Variabel bebas dapat berupa skala kategori atau skala kontinu dan variabel tak bebas berupa skala kategori. Dalam regresi logistik biner data pada variabel tak bebas bersifat biner (0 dan 1) yang menggambarkan data saling berbanding terbalik, seperti Ya dan Tidak, Benar dan Salah, Sukses dan Gagal, dan sebagainya.

Penelitian terdahulu terkait dengan model Regresi Logistik Biner dilakukan oleh Hendayanti & Nurhidayati (2020). Penelitian ini mengaplikasikan regresi logistik biner dalam penentuan ketepatan klasifikasi kemiskinan provinsi di Indonesia. Hasil yang didapat menunjukkan ketepatan klasifikasi model regresi logit akurat. Ini menunjukkan bahwa ada perbedaan dari data asli dengan model regresi logit yang diperoleh.

Selanjutnya Qudrotunanda & Chamid (2023) melakukan analisis variabel terhadap persalinan Caesar di RSUD Ploso dengan pendekatan regresi logistik biner. Hasilnya, variabel bebas yang berpengaruh secara signifikan terhadap persalinan caesar adalah variabel riwayat persalinan caesar, hipertensi, jumlah persalinan dengan kategori berisiko, kondisi ketuban tidak normal, dan posisi bayi tidak normal.

Berezka et al. (2022) dalam penelitiannya yang berjudul “A Binary Logistic Regression Model For Support Decision Making In Criminal Justice” membuktikan

bahwa ada hubungan antara kecenderungan narapidana untuk melakukan kejahatan berulang dengan karakteristik usia pada saat hukuman pertama, jenis pekerjaan, jumlah hukuman percobaan, jumlah kejahatan ringan dan pembebasan lebih awal. Model yang dikembangkan dapat memperkirakan peluang melakukan kejahatan berulang berdasarkan data statistik dan dinamis.

Regresi Logistik Biner dapat diterapkan dalam bidang kesehatan. Contohnya penerapan Regresi Logistik Biner pada penyakit *Polycystic Ovary Syndrome* (PCOS). *Polycystic Ovary Syndrome* adalah sebuah kelainan endokrin yang mampu mempengaruhi wanita sebesar 5-10% akibat produksi androgen yang berlebihan oleh ovarium (Octaviana & Kurniati, 2019). WHO mencatat bahwa PCOS dialami oleh lebih dari 116 juta wanita atau sekitar 3,4% di seluruh dunia.

Polycystic Ovary Syndrome (PCOS) memiliki gejala yang berbeda pada setiap orang dan dapat berubah seiring berjalannya waktu. Gejala umum yang terjadi dapat berupa periode menstruasi yang tidak beraturan, infertilitas, jerawat, penipisan rambut bahkan penambahan berat badan di sekitar perut. PCOS merupakan kondisi yang kompleks yang berpengaruh besar terhadap kesehatan fisik dan mental wanita.

Kambale et al. (2023) melakukan penelitian terhadap penyakit *Polycystic Ovary Syndrome*, dengan hasil yang didapatkan adalah peningkatan kadar hormon perangsang tiroid, Luteinizing Hormone, Follicle-Stimulating Hormone, dan prolaktin menjadi faktor penyebab *Polycystic Ovary Syndrome*.

Penelitian lainnya yang berkaitan dengan *Polycystic Ovary Syndrome* dilaksanakan oleh Carroll et al. (2012). Penelitian ini menggunakan uji chi-square, dimana hasil yang didapatkan adalah umur dan Indeks Massa Tubuh pasien berisiko tinggi terhadap kejadian *Polycystic Ovary Syndrome*.

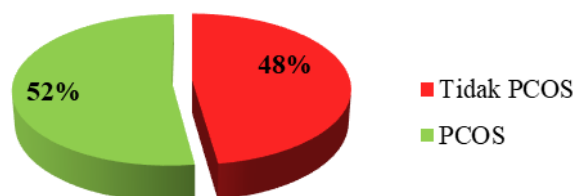
Polycystic Ovary Syndrome adalah penyakit kompleks dengan banyak faktor risiko dan komplikasi. Menjadikan *Polycystic Ovary Syndrome* sebagai bahan penelitian dapat meningkatkan pemahaman tentang penyebab, cara pengobatan dan dampaknya pada kehidupan wanita. Penggunaan metode regresi logistik biner akan dapat membantu dalam memahami kompleksitas penyakit PCOS dan dapat mengidentifikasi faktor-faktor yang berkontribusi dalam perkembangan penyakit ini. Berdasarkan pembahasan dan penelitian sebelumnya, peneliti tertarik menggunakan model Regresi Logistik Biner untuk mengetahui faktor-faktor yang mempengaruhi penyakit *Polycystic Ovary Syndrome* (PCOS).

METODE PENELITIAN

Jenis penelitian yang digunakan adalah penelitian analisis deskriptif dengan pendekatan kuantitatif menggunakan regresi logistik biner. Bantuan *software* yang digunakan adalah SPSS dan *Microsoft Excel*.

HASIL DAN PEMBAHASAN

Penelitian ini menggunakan data pasien *Polycystic Ovary Syndrome* (PCOS) yang diperoleh dari website kaggle, dengan sampel penelitian sebanyak 300 pasien. Data ini dipublikasikan oleh Prasoon Kottarathil pada tahun 2020. Variabel tak bebas (Y) pada penelitian ini adalah pasien *Polycystic Ovary Syndrome* (PCOS), dimana (Y)=1 merupakan pasien menderita *Polycystic Ovary Syndrome* (PCOS) dan (Y)=0 merupakan pasien tidak menderita *Polycystic Ovary Syndrome* (PCOS). Pengelompokan pasien *Polycystic Ovary Syndrome* dapat dilihat pada Gambar 4.1. Data keseluruhan dapat dilihat pada Lampiran 1.

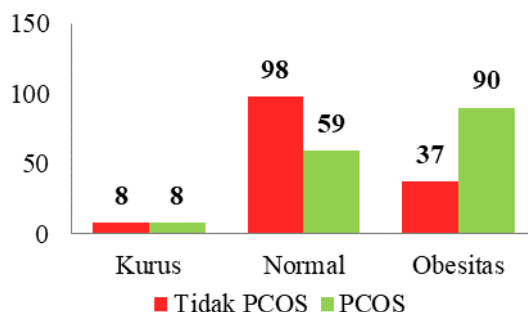
Gambar 1. Pie chart diagnosis *Polycystic Ovary Syndrome*

Berdasarkan Gambar 1. diketahui bahwa persentase wanita yang menderita penyakit Polycystic Ovary Syndrome (PCOS) ada sebesar 52% atau berjumlah 157 orang, dan wanita yang tidak menderita penyakit Polycystic Ovary Syndrome (PCOS) ada sebesar 48% sebanyak 143 orang. Polycystic Ovary Syndrome merupakan penyakit yang tidak bisa disembuhkan, akan tetapi gejala dari penyakit ini dapat dikurangi. Berikut akan dijelaskan mengenai variabel-variabel yang diduga berpengaruh terhadap Polycystic Ovary Syndrome. Variabel pertama adalah usia pasien. Dimana gangguan PCOS umumnya terjadi pada wanita yang telah mengalami proses menstruasi, pada usia 16-45 tahun.

Tabel 1. Statistik deskriptif variabel usia

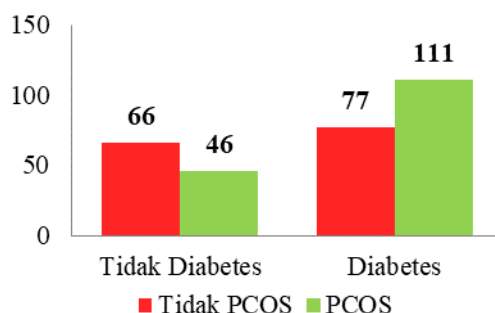
	Usia		
	Mean	Min.	Max.
Tidak Menderita PCOS	31,517	20	48
Menderita PCOS	29,949	21	47

Tabel 1. merupakan statistik deskriptif dari variabel usia. Kategori wanita yang tidak menderita PCOS, memiliki rata-rata usia sebesar 31,517, dengan rentang usia mulai dari 20 tahun – 48 tahun. Pada kategori wanita menderita PCOS, rata-rata usianya adalah 29,949, dengan rentang usia mulai dari 21 tahun – 47 tahun.



Gambar 2. Diagram variabel usia

Indeks Massa Tubuh (IMT) adalah salah satu variabel penduga dalam penelitian ini. Gambar 2. menunjukkan bahwa wanita yang menderita penyakit Polycystic Ovary Syndrome paling banyak terjadi pada wanita dengan Indeks Massa Tubuh kategori obesitas, yaitu sebanyak 90 orang. Sementara, wanita yang menderita penyakit Polycystic Ovary Syndrome pada kategori kurus ada sebanyak 8 orang dan pada kategori normal ada sebanyak 59 orang. Wanita yang tidak memiliki Polycystic Ovary Syndrome dengan IMT kategori kurus ada sebanyak 8 orang, dengan kategori normal ada sebanyak 98 orang dan dengan kategori obesitas ada sebanyak 37 orang.



Gambar 3. Diagram variabel diabetes

Gambar 3. menunjukkan banyaknya pasien berdasarkan kelompok diabetes. Wanita yang tidak menderita Polycystic Ovary Syndrome dan tidak memiliki diabetes ada sebanyak 66 orang dan yang memiliki diabetes sebanyak 46 orang. Dari 157 pasien Polycystic Ovary Syndrome, diketahui bahwa 111 diantaranya memiliki diabetes dan 77 orang lainnya tidak memiliki diabetes.

Tabel 2. Statistik deskriptif variabel Follicle-Stimulating Hormone

	<i>Follicle-Stimulating Hormone</i>		
	<i>Mean</i>	<i>Min.</i>	<i>Max.</i>
Tidak Menderita PCOS	5,196	1	13,9
Menderita PCOS	5,233	1	65,4

Tabel 2. menyajikan informasi mengenai variabel Follicle-Stimulating Hormone (FSH), dimana pada wanita yang tidak menderita Polycystic Ovary Syndrome memiliki nilai rata-rata kadar FSH sebesar 5,196 mIU/mL, dengan rentang kadar FSH berada pada 1 – 13,9 mIU/mL. Pada wanita yang menderita Polycystic Ovary Syndrome diketahui rata-rata kadar FSH yang dimiliki adalah sebesar 5,233 mIU/mL, dengan rentang kadar FSH berada pada 1 – 65,4 mIU/mL.

Tabel 3. Statistik deskriptif variabel Luteinizing Hormone

	<i>Luteinizing Hormone</i>		
	<i>Mean</i>	<i>Min.</i>	<i>Max.</i>
Tidak Menderita PCOS	2,670	0,02	10,75
Menderita PCOS	3,172	0,032	20,18

Berdasarkan Tabel 3, diketahui bahwa wanita yang tidak menderita Polycystic Ovary Syndrome memiliki rata-rata kadar Luteinizing Hormone (LH) sebesar 2,670 mIU/mL. Kadar LH terendah yang dimiliki wanita pada kategori ini adalah 0,02 mIU/mL dan kadar LH terbesar adalah 10,75 mIU/mL. Wanita yang menderita Polycystic Ovary Syndrome memiliki rata-rata kadar Luteinizing Hormone sebesar 3,172 mIU/L, dengan kadar LH terendah berada pada kadar 0,032 mIU/mL dan kadar terbesarnya sebesar 20,18 mIU/mL.

Tabel 4. Statistik deskriptif variabel Thyroid Stimulating Hormone

	<i>Thyroid Stimulating Hormone</i>		
	<i>Mean</i>	<i>Min.</i>	<i>Max.</i>
Tidak Menderita PCOS	2,765	0,22	16
Menderita PCOS	2,963	0,05	22,59

Tabel 4. merupakan statistik deskriptif dari variabel Thyroid Stimulating Hormone (TSH). Kategori wanita yang tidak menderita PCOS, memiliki rata-rata TSH sebesar 2,765 mIU/mL dengan rentang kadar TSH yang dimiliki mulai dari 0,22 – 16 mIU/mL. Pada kategori wanita menderita PCOS, rata-rata kadar TSH yang dimiliki adalah 2,963 mIU/mL, dengan rentang kadar TSH yang dimiliki mulai dari 0,05 – 22,59 mIU/mL.

Tabel 5. Statistik deskriptif Anti Müllerian Hormone

	Anti Müllerian Hormone		
	Mean	Min.	Max.
Tidak Menderita PCOS	4,586	0,2	26,8
Menderita PCOS	7,805	0,1	32

Tabel 5. menyajikan informasi mengenai variabel Anti Müllerian Hormone (AMH), dimana pada wanita yang tidak menderita Polycystic Ovary Syndrome memiliki kadar rata-rata AMH sebesar 4,586 ng/mL, dengan rentang kadar AMH berada pada 0,2 – 26,8 ng/mL. Pada wanita yang menderita Polycystic Ovary Syndrome diketahui rata-rata kadar AMH yang dimiliki adalah sebesar 7,805 ng/mL, dengan rentang kadar AMH berada pada rentang 0,1 – 32 ng/mL.

Tabel 6 Statistik deskriptif variabel Vitamin D3

	Vitamin D3		
	Mean	Min.	Max.
Tidak Menderita PCOS	30,055	9,01	87,2
Menderita PCOS	28,448	6,01	79,21

Tabel 6. merupakan statistik deskriptif dari variabel Vitamin D3 yang menampilkan nilai rata-rata, nilai minimum dan nilai maximum untuk pasien dengan kategori tidak menderita dan menderita Polycystic Ovary Syndrome. Pada kategori tidak menderita PCOS, rata-rata pasien memiliki kadar Vitamin D3 sebesar 30,055 ng/mL, dengan rentang kadar Vitamin D3 yang dimiliki sebesar 9,01 – 87,2 ng/mL. Pada pasien dengan kategori menderita PCOS, rata-rata kadar Vitamin D3 yang dimiliki adalah sebesar 28,448 ng/mL, dengan rentang kadar Vitamin D3 yang dimiliki adalah 6,01 – 79,21 ng/mL.

Uji Multikolinieritas

Pada regresi logistik biner, asumsi yang harus terpenuhi adalah nonmultikolinieritas. Dalam mendeteksi multikolinieritas dapat menggunakan rumus VIF (Variance Inflation Factors) pada persamaan (2.11). Kriteria pengambilan keputusan uji multikolinieritas adalah jika nilai VIF bernilai lebih kecil dari 10, maka variabel bebas dinyatakan nonmultikolinieritas. Nilai VIF untuk masing-masing variabel bebas dicari menggunakan bantuan SPSS (Lampiran 2), yang disajikan pada Tabel 7.

Tabel 7. Hasil uji VIF

No	Variabel Bebas	Nilai VIF
1	Usia	1,079
2	Indeks Massa Tubuh	1,039
3	Diabetes	1,071
4	<i>Follicle-Stimulating Hormone</i>	1,044
5	<i>Luteinizing Hormone</i>	1,103
6	<i>Thyroid Stimulating Hormone</i>	1,019
7	<i>Anti-Müllerian Hormone</i>	1,121
8	Vitamin D3	1,028

Berdasarkan Tabel 7. diketahui bahwa 8 variabel bebas yang diteliti memiliki nilai VIF <10, maka dapat disimpulkan bahwa H₀ diterima untuk setiap variabel bebas. Asumsi nonmultikolinieritas terpenuhi pada variabel penelitian ini. Sehingga variabel-variabel bebas pada penelitian ini dapat digunakan.

Pembentukan Variabel Rancangan

Pembentukan variabel rancangan pada penelitian ini dikarenakan terdapat variabel bebas yang bersifat kategori. Tujuan pembentukan variabel rancangan adalah untuk meminimalisir bias dan kesalahan dalam pengukuran variabel. Pembentukan variabel rancangan dilakukan pada variabel Indeks Massa Tubuh (X_2) dan pada variabel Diabetes (X_3). Dimana variabel yang memiliki kategori sebanyak d, membutuhkan variabel rancangan sebanyak d-1.

Variabel X_2 memiliki 3 kategori yaitu kurus ($X_2^{(1)}$), normal ($X_2^{(2)}$) dan obesitas ($X_2^{(3)}$). Dari 3 kategori, dibutuhkan 2 variabel rancangan, yaitu $X_2^{(1)}$ dan $X_2^{(3)}$ dengan $X_2^{(2)}$ sebagai pembanding. Variabel Diabetes (X_3) terdiri dari 2 kategori yaitu tidak memiliki diabetes ($X_3^{(0)}$) dan memiliki diabetes ($X_3^{(1)}$), sehingga variabel rancangan pada variabel ini adalah ($X_3^{(1)}$) dengan ($X_3^{(0)}$) sebagai pembanding. Untuk pengkodean variabel rancangan dapat dilihat dengan jelas pada Tabel 8.

Tabel 8. Pengkodean variabel rancangan

Variabel	Nilai	Variabel Rancangan		
X_2	1	1	0	0
	2	0	0	0
	3	0	0	1
X_4	0	0	0	-
	1	1	0	-

Pembentukan variabel rancangan pada variabel X_2 yang memiliki 3 kategori dilakukan dengan bantuan SPSS pada menu transform - create dummy variable. Yang menghasilkan 2 variabel dummy yang terdiri dari $X_2^{(1)}$ untuk kategori kurus dan $X_2^{(3)}$ untuk kategori obesitas. Kemudian 2 variabel yang tercipta yang ikut serta dalam perhitungan estimasi parameter dengan variabel lainnya.

Estimasi Parameter Model

Model yang digunakan dalam menaksir model fungsi regresi logistik adalah metode Maksimum Likelihood Estimation (MLE). Fungsi distribusi mengikuti distribusi Bernoulli yang kemudian akan dicari fungsi likelihood yang dinotasikan dengan $L(\beta)$:

$$\begin{aligned}
 l(\beta) &= \prod_{i=1}^n P(Y = y_i) \\
 &= \prod_{i=1}^n \pi(X_i)^{y_i} [1 - \pi(X_i)]^{1-y_i} \\
 &= \prod_{i=1}^n \pi(X_i)^{y_i} \frac{1 - \pi(X_i)}{[1 - \pi(X_i)]^{y_i}} \\
 &= \prod_{i=1}^n (1 - \pi(X_i)) \frac{\pi(X_i)^{y_i}}{(1 - \pi(X_i))^{y_i}} \\
 &= \left\{ \prod_{i=1}^n (1 - \pi(X_i)) \right\} \left\{ \prod_{i=1}^n \left(\frac{\pi(X_i)}{1 - \pi(X_i)} \right)^{y_i} \right\}
 \end{aligned}$$

$$\begin{aligned}
&= \left\{ \prod_{i=1}^n (1 - \pi(X_i)) \right\} \left\{ \prod_{i=1}^n \exp \left(\ln \left(\frac{\pi(X_i)}{1 - \pi(X_i)} \right)^{y_i} \right) \right\} \\
&= \left\{ \prod_{i=1}^n (1 - \pi(X_i)) \right\} \exp \left\{ \sum_{i=1}^n y_i \ln \left(\frac{\pi(X_i)}{1 - \pi(X_i)} \right) \right\} \\
&= \left\{ \prod_{i=1}^n \left(1 - \frac{\exp(\beta_0 + \sum_{j=1}^p \beta_j X_{ij})}{1 + \exp(\beta_0 + \sum_{j=1}^p \beta_j X_{ij})} \right) \right\} \exp \left\{ \sum_{i=1}^n y_i \ln(\exp(\beta_0 + \sum_{j=1}^p \beta_j X_{ij})) \right\} \\
&= \left\{ \prod_{i=1}^n \left(\frac{(1 + \exp(\beta_0 + \sum_{j=1}^p \beta_j X_{ij})) - \exp(\beta_0 + \sum_{j=1}^p \beta_j X_{ij})}{1 + \exp(\beta_0 + \sum_{j=1}^p \beta_j X_{ij})} \right) \right\} \exp \left\{ \sum_{i=1}^n y_i (\beta_0 + \sum_{j=1}^p \beta_j X_{ij}) \right\} \\
&= \left\{ \prod_{i=1}^n \left(\frac{1}{1 + \exp(\beta_0 + \sum_{j=1}^p \beta_j X_{ij})} \right) \right\} \exp \left\{ \sum_{i=1}^n y_i (\beta_0 + \sum_{j=1}^p \beta_j X_{ij}) \right\} \tag{4.1}
\end{aligned}$$

Fungsi *likelihood* kemudian di maksimumkan dengan fungsi $\ln$ *likelihood* :

$$L(\beta) = \ln l(\beta)$$

$$\begin{aligned}
&= \ln \left(\left\{ \prod_{i=1}^n \left(\frac{1}{1 + \exp(\beta_0 + \sum_{j=1}^p \beta_j X_{ij})} \right) \right\} \exp \left\{ \sum_{i=1}^n y_i (\beta_0 + \sum_{j=1}^p \beta_j X_{ij}) \right\} \right) \\
&= \ln \left(\prod_{i=1}^n \left(1 + \exp(\beta_0 + \sum_{j=1}^p \beta_j X_{ij}) \right)^{-1} \exp \left\{ \sum_{i=1}^n y_i (\beta_0 + \sum_{j=1}^p \beta_j X_{ij}) \right\} \right) \\
&= - \sum_{i=1}^n \ln \{ 1 + \exp(\beta_0 + \sum_{j=1}^p \beta_j X_{ij}) \} + \sum_{i=1}^n y_i (\beta_0 + \sum_{j=1}^p \beta_j X_{ij}) \\
&= \sum_{i=1}^n y_i (\beta_0 + \sum_{j=1}^p \beta_j X_{ij}) - \sum_{i=1}^n \ln \{ 1 + \exp(\beta_0 + \sum_{j=1}^p \beta_j X_{ij}) \} \tag{4.2}
\end{aligned}$$

Selanjutnya estimasi parameter diperoleh menggunakan turunan secara parsial pada

fungsi *likelihood* $L(\beta)$ terhadap parameter β . Misalkan:

$$a = \sum_{i=1}^n y_i (\beta_0 + \sum_{j=1}^p \beta_j X_{ij})$$

$$b = \sum_{i=1}^n \ln\{1 + \exp(\beta_0 + \sum_{j=1}^p \beta_j X_{ij})\}$$

Turunan parsial pertama terhadap :

$$\beta_p$$

$$\frac{\partial a}{\partial \beta_p} = \frac{\partial}{\partial \beta_p} \sum_{i=1}^n y_i (\beta_0 + \sum_{j=1}^p \beta_j X_{ij})$$

$$= \frac{\partial}{\partial \beta_j} \sum_{i=1}^n y_i \beta_0 + \sum_{i=1}^n y_i \beta_1 X_{i1} + \dots + \sum_{i=1}^n y_i \beta_p X_{ip}$$

$$= \sum_{i=1}^n y_i X_{ip}$$

$$\frac{\partial b}{\partial \beta_p} = \frac{\partial}{\partial \beta_p} \sum_{i=1}^n \ln\{1 + \exp(\beta_0 + \sum_{j=1}^p \beta_j X_{ij})\}$$

$$= \frac{\partial}{\partial \beta_p} \sum_{i=1}^n \ln\{1 + \exp(\beta_0 + \beta_1 X_{i1} + \dots + \beta_p X_{ip})\}$$

$$= \sum_{i=1}^n X_{ip} \left\{ \frac{\exp(\beta_0 + \beta_1 X_{i1} + \dots + \beta_p X_{ip})}{1 + \exp(\beta_0 + \beta_1 X_{i1} + \dots + \beta_p X_{ip})} \right\}$$

Sehingga diperoleh persamaan:

$$\frac{\partial L(\beta)}{\partial \beta_j} = \frac{\partial a}{\partial \beta_j} - \frac{\partial b}{\partial \beta_j}$$

$$= \sum_{i=1}^n y_i X_{ip} - \sum_{i=1}^n X_{ip} \left\{ \frac{\exp(\beta_0 + \beta_1 X_{i1} + \dots + \beta_p X_{ip})}{1 + \exp(\beta_0 + \beta_1 X_{i1} + \dots + \beta_p X_{ip})} \right\} \quad (4.3)$$

Karena hasil turunan pertama masih dalam bentuk implisit, maka dibutuhkan metode

$$\beta$$

numerik seperti pendekatan iterasi *Newton Raphson*:

$$\beta^{(r+1)} = \beta^{(r)} + (H^{(r)})^{-1} g^{(r)}$$

$g^{(r)}$ adalah vektor kemiringan yang dihasilkan dari turunan parsial pertama:

$$\mathbf{g}^{(r)} = \begin{bmatrix} \frac{\partial L(\beta)}{\partial \beta_0} \\ \frac{\partial L(\beta)}{\partial \beta_1} \\ \vdots \\ \frac{\partial L(\beta)}{\partial \beta_i} \end{bmatrix} \quad (4.4)$$

Langkah berikutnya adalah menghitung turunan parsial kedua dari persamaan (4.4), dimana hasil turunan parsial kedua akan membentuk matriks *Hessian*.

In *likelihood*

Turunan parsial kedua terhadap :

$$\beta_v$$

$$\frac{\partial^2 L(\beta)}{\partial \beta_v^2} = \frac{\partial}{\partial \beta_v} \left(\sum_{i=1}^n y_i X_{ip} - \sum_{i=1}^n X_{ip} \left\{ \frac{\exp(\beta_0 + \beta_1 X_{i1} + \dots + \beta_p X_{ip})}{1 + \exp(\beta_0 + \beta_1 X_{i1} + \dots + \beta_p X_{ip})} \right\} \right)$$

Misalkan:

$$u = \exp(\beta_0 + \beta_1 X_{i1} + \dots + \beta_v X_{iv})$$

$$= \exp(\beta_0 + \sum_{i=1}^p \beta_i X_{ii})$$

$$u' = \frac{\partial}{\partial \beta_v} (\exp(\beta_0 + \beta_1 X_{i1} + \dots + \beta_p X_{ip}))$$

$$= \exp(\beta_0 + \beta_1 X_{i1} + \dots + \beta_v X_{iv}) \cdot X_{ip}$$

$$= X_{ip} (\exp(\beta_0 + \sum_{i=1}^p \beta_i X_{ii}))$$

$$v = 1 + \exp(\beta_0 + \beta_1 X_{i1} + \dots + \beta_v X_{iv})$$

$$= 1 + \exp(\beta_0 + \sum_{i=1}^p \beta_i X_{ii})$$

$$v' = \frac{\partial}{\partial \beta_p} (1 + \exp(\beta_0 + \beta_1 X_{i1} + \dots + \beta_p X_{ip}))$$

$$= \exp(\beta_0 + \beta_1 X_{i1} + \dots + \beta_p X_{ip}) \cdot X_{ip}$$

$$= X_{ip} (\exp(\beta_0 + \sum_{i=1}^p \beta_i X_{ii}))$$

Maka:

$$\frac{u}{v} = \frac{\exp(\beta_0 + \beta_1 X_{i1} + \dots + \beta_p X_{ip})}{1 + \exp(\beta_0 + \beta_1 X_{i1} + \dots + \beta_p X_{ip})}$$

Selanjutnya turunan dari adalah:

$$(u/v)$$

$$\left(\frac{u}{v}\right)' = \frac{u'v - v'u}{v^2}$$

$$= \frac{(X_{ip} \exp(\beta_0 + \sum_{j=1}^p \beta_j X_{ij})) (1 + \exp(\beta_0 + \sum_{j=1}^p \beta_j X_{ij})) - (X_{ip} \exp(\beta_0 + \sum_{j=1}^p \beta_j X_{ij})) (\exp(\beta_0 + \sum_{j=1}^p \beta_j X_{ij}))}{(1 + \exp(\beta_0 + \beta_1 X_{i1} + \dots + \beta_p X_{ip}))^2}$$

$$= \frac{(X_{ip} \exp(\beta_0 + \sum_{j=1}^p \beta_j X_{ij})) [(1 + \exp(\beta_0 + \sum_{j=1}^p \beta_j X_{ij})) - (\exp(\beta_0 + \sum_{j=1}^p \beta_j X_{ij}))]}{(1 + \exp(\beta_0 + \beta_1 X_{i1} + \dots + \beta_p X_{ip})) (1 + \exp(\beta_0 + \beta_1 X_{i1} + \dots + \beta_p X_{ip}))}$$

$$= \frac{X_{ip} (\exp(\beta_0 + \sum_{j=1}^p \beta_j X_{ij}))}{(1 + \exp(\beta_0 + \beta_1 X_{i1} + \dots + \beta_p X_{ip})) (1 + \exp(\beta_0 + \beta_1 X_{i1} + \dots + \beta_p X_{ip}))}$$

$$= X_{ip} \left(\frac{X_{ip} (\exp(\beta_0 + \sum_{j=1}^p \beta_j X_{ij}))}{(1 + \exp(\beta_0 + \beta_1 X_{i1} + \dots + \beta_p X_{ip}))} \right) \left(\frac{1}{(1 + \exp(\beta_0 + \beta_1 X_{i1} + \dots + \beta_p X_{ip}))} \right)$$

elimination adalah metode seleksi variabel yang digunakan untuk mempertahankan variabel-variabel yang paling berpengaruh dalam model regresi, dengan cara mengeleminasi variabel yang tidak signifikan secara bertahap.

Kriteria pengujian dengan metode backward elimination adalah apabila p-value $>0,05$, maka akan dilakukan eliminasi pada variabel dengan nilai signifikansi paling tinggi. Setelah mengeluarkan variabel yang tidak memenuhi kriteria pengujian, estimasi ulang model dengan variabel yang tersisa dan lakukan kembali proses eliminasi sampai semua variabel yang tersisa signifikan (p-value $<0,05$). Hasil penaksiran untuk setiap parameter β dapat dilihat pada Tabel 9. Output hasil pengolahan data dapat dilihat pada Lampiran 3.

Tabel 9. Estimasi parameter seluruh variabel

Parameter	Taksiran $\hat{\beta}$	Std. Error	df	Sig.
Usia (X_1)	-0,049	0,025	1	0,047
IMT = Kurus (1) (X_2^1)	0,736	0,590	1	0,212
IMT = Obesitas (1) (X_2^3)	1,389	0,279	1	0,000
Diabetes (X_3)	1,227	0,273	1	0,000
Follicle-Stimulating Hormone (X_4)	0,023	0,028	1	0,412
Luteinizing Hormone (X_5)	0,059	0,055	1	0,283
Thyroid Stimulating Hormone (X_6)	0,050	0,054	1	0,355
Anti Müllerian Hormone (X_7)	0,072	0,027	1	0,007
Vitamin D3 (X_8)	-0,001	0,010	1	0,948
Constant	-0,560	0,881	1	0,525

Tabel 9. menunjukkan bahwa variabel Vitamin D3 (X_8) memiliki nilai signifikansi paling tinggi diantara seluruh variabel, yaitu 0,948. Nilai signifikansi variabel Vitamin D3 tidak memenuhi kriteria pengujian, dikarenakan nilai signifikansi $>0,05$, artinya variabel Vitamin D3 tidak berpengaruh terhadap terjadinya Polycystic Ovary Syndrome (PCOS). Pada proses pengulangan selanjutnya variabel Vitamin D3 dieleminasi. Hasil pengulangan penaksiran parameter β dapat dilihat pada Tabel 10.

Tabel 10. Estimasi parameter tanpa variabel X_8

Parameter	Taksiran $\hat{\beta}$	Std. Error	df	Sig.
Usia (X ₁)	-0,049	0,025	1	0,047
IMT = Kurus (1) (X ₂ ¹)	0,734	0,589	1	0,213
IMT = Obesitas (1) (X ₂ ³)	1,389	0,279	1	0,000
Diabetes (X ₃)	1,229	0,271	1	0,000
<i>Follicle-Stimulating Hormone</i> (X ₄)	0,023	0,028	1	0,412
<i>Luteinizing Hormone</i> (X ₅)	0,059	0,055	1	0,282
<i>Thyroid Stimulating Hormone</i> (X ₆)	0,051	0,054	1	0,349
<i>Anti Müllerian Hormone</i> (X ₇)	0,072	0,027	1	0,006
<i>Constant</i>	-0,579	0,830	1	0,485

Tabel 10. menunjukkan variabel Follicle-Stimulating Hormone (X₄) memiliki nilai signifikansi yang paling tinggi yaitu 0,412, yang berarti bahwa variabel ini tidak berpengaruh terhadap terjadinya Polycystic Ovary Syndrome. Sesuai dengan kriteria pengujian, maka variabel Follicle-Stimulating Hormone dieleminasi pada proses pengulangan berikutnya. Hasil penaksiran β dapat dilihat pada Tabel 11.

Tabel 11. Estimasi parameter tanpa variabel X_8 dan X_4

Parameter	Taksiran $\hat{\beta}$	Std. Error	df	Sig.
Usia (X ₁)	-0,046	0,024	1	0,059
IMT = Kurus (1) (X ₂ ¹)	0,721	0,590	1	0,222
IMT = Obesitas (1) (X ₂ ³)	1,380	0,278	1	0,000
Diabetes (X ₃)	1,199	0,269	1	0,000
<i>Luteinizing Hormone</i> (X ₅)	0,064	0,055	1	0,245
<i>Thyroid Stimulating Hormone</i> (X ₆)	0,048	0,054	1	0,377

<i>Anti Müllerian Hormone</i>	0,073	0,027	1	0,006
(X_7)				
<i>Constant</i>	-0,542	0,826	1	0,512

Hasil iterasi ke-3 menampilkan bahwa variabel Thyroid Stimulating Hormone tidak memiliki pengaruh terhadap terjadinya Polycystic Ovary Syndrome, karena nilai signifikansi yang lebih besar dari 0,05 yaitu 0,377. Sehingga variabel Thyroid Stimulating Hormone dieleminasi, yang selanjutnya proses iterasi dilakukan tanpa melibatkan variabel tersebut. Hasil penaksiran parameter β dapat dilihat pada Tabel 12.

Tabel 12. Estimasi parameter tanpa variabel X_8, X_4 dan X_6

Parameter	Taksiran $\hat{\beta}$	Std. Error	df	Sig.
Usia	-0,044	0,024	1	0,067
(X_1)				
IMT = Kurus (1)	0,731	0,590	1	0,215
(X_2^1)				
IMT = Obesitas (1)	1,373	0,277	1	0,000
(X_2^3)				
Diabetes	1,196	0,268	1	0,000
(X_3)				
<i>Luteinizing Hormone</i>	0,063	0,055	1	0,254
(X_5)				
<i>Anti Müllerian Hormone</i>	0,073	0,027	1	0,006
(X_7)				
<i>Constant</i>	-0,446	0,818	1	0,585

Tabel 12 menunjukkan bahwa variabel Luteinizing Hormone (X_5) memiliki nilai signifikansi paling tinggi diantara seluruh variabel, yaitu 0,254. Nilai signifikansi variabel Luteinizing Hormone tidak memenuhi kriteria pengujian, dikarenakan nilai signifikansi yang lebih besar dari 0,05. Ini mengartikan bahwa variabel Luteinizing Hormone tidak berpengaruh terhadap terjadinya Polycystic Ovary Syndrome (PCOS). Pada proses pengulangan selanjutnya variabel X_5 dieleminasi. Hasil pengulangan penaksiran parameter β dapat dilihat pada Tabel 13.

Tabel 13 Estimasi parameter tanpa variabel X_8, X_4, X_6 dan X_5

Parameter	Taksiran $\hat{\beta}$	Std. Error	df	Sig.
Usia	-0,047	0,024	1	0,050
(X_1)				
IMT = Kurus (1)	0,693	0,589	1	0,239
(X_2^1)				
IMT = Obesitas (1)	1,329	0,274	1	0,000
(X_2^3)				
Diabetes	1,192	0,267	1	0,000
(X_3)				

<i>Anti Müllerian Hormone</i>	0,079	0,026	1	0,002
(X_7)				
<i>Constant</i>	-0,193	0,787	1	0,806

Selanjutnya pada iterasi ke-5, menghasilkan estimasi parameter dimana variabel Indeks Massa Tubuh dengan kategori kurus tidak memenuhi kriteria pengujian. Nilai signifikansi variabel yang lebih besar dari 0,05 yaitu 0,259 menunjukkan bahwa variabel tidak memenuhi kriteria pengujian, maka pada proses iterasi selanjutnya, variabel Indeks Massa Tubuh dengan kategori kurus dieleminasi. Hasil iterasi ke-6 dapat dilihat pada Tabel 14.

Tabel 14. Estimasi parameter tanpa variabel X_8 , X_4 , X_6 , X_5 dan X_2^1

Parameter	Taksiran $\hat{\beta}$	Std. Error	df	Sig.
Usia (X_1)	-0,046	0,024	1	0,054
IMT = Obesitas (1) (X_2^3)	1,267	0,268	1	0,000
Diabetes (X_3)	1,162	0,265	1	0,000
<i>Anti Müllerian Hormone</i> (X_7)	0,081	0,026	1	0,002
<i>Constant</i>	-0,145	0,787	1	0,854

Tabel 14 menunjukkan variabel Usia memiliki nilai signifikansi yang paling tinggi diantara variabel lainnya yaitu 0,054. Sesuai kriteria pengujian, apabila nilai sig. variabel lebih besar dari 0,05 maka variabel tersebut akan di eliminasi pada proses iterasi selanjutnya. Pada proses iterasi ke-7, variabel usia di eliminasi dan menghasilkan estimasi parameter yang dapat dilihat pada Tabel 15.

Tabel 15. Estimasi parameter tanpa variabel X_8 , X_4 , X_6 , X_5 , X_2^1 dan X_1

Parameter	Taksiran $\hat{\beta}$	Std. Error	df	Sig.
IMT = Obesitas (1) (X_2^3)	1,263	0,266	1	0,000
Diabetes (X_3)	1,119	0,261	1	0,000
<i>Anti Müllerian Hormone</i> (X_7)	0,087	0,025	1	0,001
<i>Constant</i>	-1,588	0,271	1	0,000

Tabel 15. menunjukkan bahwa variabel yang tersisa yaitu variabel Indeks Massa Tubuh kategori obesitas, variabel diabetes dan variabel Anti Müllerian Hormone memiliki nilai signifikansi yang lebih kecil dari 0,05. Oleh karena 3 variabel yang tersisa memiliki nilai signifikansi lebih kecil dari 0,05 maka disimpulkan bahwa variabel-variabel tersebut memiliki pengaruh terhadap penyakit Polycystic Ovary Syndrome. Selanjutnya variabel-variabel tersebut akan dimasukkan kedalam model regresi logistik biner yang akan dilanjutkan dengan melakukan uji signifikansi parameter.

KESIMPULAN

Berdasarkan hasil analisis dan pembahasan yang sudah dilaksanakan, maka diperoleh kesimpulan sebagai berikut:

1. Dari 8 variabel bebas yang di analisa sebagai faktor-faktor yang mempengaruhi penyakit *Polycystic Ovary Syndrome* (PCOS), terdapat 3 variabel yang berpengaruh signifikan, yaitu Indeks Massa Tubuh kategori obesitas (X_2^3), diabetes (X_3) dan kadar Anti Müllerian Hormone (X_7) dengan model sebagai berikut:

$$\pi(X_i) = \frac{e^{(-1,588 + 1,263 X_2^3 + 1,119 X_3 + 0,087 X_7)}}{1 + e^{(-1,588 + 1,263 X_2^3 + 1,119 X_3 + 0,087 X_7)}}$$

2. Dengan menggunakan metode regresi logistik biner, terdapat 3 variabel yang berpengaruh signifikan terhadap penyakit *Polycystic Ovary Syndrome* yaitu:

- a. Variabel Indeks Massa Tubuh Kategori Obesitas (X_2^3)

Dengan nilai odds ratio sebesar 3,536 menunjukkan bahwa wanita yang memiliki riwayat obesitas memiliki peluang 3,536 kali lebih besar untuk mengalami *Polycystic Ovary Syndrome* (PCOS) dibandingkan wanita dengan Indeks Massa Tubuh yang normal.

- b. Variabel Diabetes (X_3)

Dengan nilai odds ratio sebesar 3,061, menunjukkan bahwa wanita yang terkena diabetes memiliki peluang 3,061 kali lebih besar untuk mengalami *Polycystic Ovary Syndrome* (PCOS) dibandingkan wanita yang tidak menderita diabetes.

- c. Variabel Anti Müllerian Hormone (X_7)

Dengan nilai odds ratio sebesar 1,090, menunjukkan setiap peningkatan 1% kadar AMH memiliki peluang 1,090 kali lebih besar untuk mengalami *Polycystic Ovary Syndrome* (PCOS) dibandingkan wanita dengan kadar AMH yang lebih rendah.

DAFTAR PUSTAKA

- Agresti, A. (1990). *Categorical Data Analysis*. New York: John Wiley and Sons.
- Berezka, K. M., Kovalchuk, O. Y., Banakh, S. V., Zlyvko, S. V., & Hrechaniuk, R. (2022). A Binary Logistic Regression Model For Support Decision Makin In Criminal Justice. *Folia Oeconomica Stetinensia*, 22(1), 1-17.
- Bluman, A. G. (2018). *Elementary statistics: A step by step approach* (10th ed.) McGraw-Hill Education.
- Carroll, J., Saxena, R., & Welt, C. K. (2012). Environmental and Genetic Factors Influence Age at Menarche In Women with *Polycystic Ovary Syndrome*. *Journal of Pediatric Endocrinology and Metabolism*, 25(5-6), 459-466.
- Ghozali, I. (2016). *Aplikasi Analisis Multivariate dengan Program IBM SPSS 23*.
- Gudono. (2017). *Analisis Data Multivariat*. Yogyakarta: BPFE.
- Hendayanti, N. P. N., & Nurhidayati, M. (2020). Regresi Logistik Biner dalam Penentuan Ketepatan Klasifikasi Tingkat Kedalaman Kemiskinan Provinsi-Provinsi di Indonesia. *Sainstek: Jurnal Sains dan Teknologi*, 12(2), 63-70.
- Hosmer, D. W., Lemeshow, S. (2000). *Applied Logistic Regression Second Edition*. New York: Wiley.
- Kambale, T., Sawaimul, K. D., & Prakash, S. (2023). A Study of Hormonal and Anthropometric Parameters in *Polycystic Ovary Syndrome*. *Annals of African Medicine*, 22(1), 112-116.
- Kaggle. (2020). *Polycystic Ovary Syndrome (PCOS)*. Dikutip dari <https://www.kaggle.com/datasets/prasoonkottarathil/polycystic-ovary-syndrome->

pcos/data.

- Kementerian Kesehatan. (2019). Tabel Batas Ambang Indeks Massa Tubuh (IMT). P2PTM Kemkes.
- Khasanah, U. (2021). Analisis Regresi. Yogyakarta: UAD PRESS.
- Kurniawan, R., Yuniarto, B., (2016). Analisis Regresi Dasar dan Penerapannya Dengan R. Jakarta: Kencana.
- Octaviana, A. D., & Kurniati, K. L. (2019). Suplementasi Vitamin D pada Wanita dengan Polycystic Ovarian Syndrome (PCOS). Jurnal Ilmiah Kesehatan Sandi Husada, 8(2), 169-177.
- Okta, P. P. (2020). Faktor-Faktor yang Mempengaruhi Kejadian Sindrom Ovarium Polikistik di RSUP Dr. M. Djamil Padang Tahun 2015-2019 (Doctoral dissertation, Universitas Andalas).
- Qudrotunanda, M. I., & Chamid, M. S. (2023). Analisis Variabel yang Mempengaruhi Persalinan Caesar di RSUD Ploso dengan Pendekatan Model Regresi Logistik Biner. Jurnal Sains dan Seni ITS, 11(6), D335-D342.
- Qomusuddin, I. F., & Romlah, S. (2021). Analisis Data Kuantitatif Dengan Program IBM SPSS Statistic 20.0. Sleman: Deepublish.
- Risdiyarningsih, V., Kurniawati, E. Y., & Darmawati, D. (2023). Faktor Risiko Terjadinya Sindrom Ovarium Polikistik (SOPK). Jurnal Ilmu Kebidanan, 9(2), 107-111.
- Sugeng, B. (2020). Fundamental Metodologi Penelitian Kuantitatif (Eksplanatif). Sugeng: Deepublish.
- Uyanto, S. S. (2009). Pedoman Analisis Data Dengan SPSS. Yogyakarta: Graha Ilmu, 282.
- World Health Organization (WHO). (2023). Polycystic Ovary Syndrome. World Health Organization.