Buletin Ilmiah Math. Stat. dan Terapannya (BIMASTER)
Volume 12, No. 4 (2023), hal 379-388.

PEMODELAN REGRESI LINEAR MENGGUNAKAN METODE *THEIL*
(STUDI KASUS: JUMLAH PENDUDUK MISKIN DAN TINGKAT PENGANGGURAN TERBUKA)

Rido Amarrullah, Shantika Martha, Wirda Andani

INTISARI

Metode Theil adalah salah satu metode regresi nonparametrik yang digunakan saat asumsi kenormalan residual tidak terpenuhi pada model regresi linear sederhana. Penelitian ini bertujuan untuk membentuk model regresi linear sederhana dengan metode Theil dari data dengan residual data tidak berdistribusi normal. Data yang digunakan pada penelitian ini adalah data semesteran jumlah penduduk miskin (Y) *dan tingkat pengangguran terbuka* (X) *di Indonesia Tahun 2012 hingga 2022 sebanyak 21 data. Terlebih dahulu dilakukan analisis regresi antara* X *dan* Y*, selanjutnya lakukan pengujian asumsi kenormalan residual data. Jika residual tidak berdistribusi normal, maka untuk mendapatkan model regresi linear sederhana dari data digunakanlah metode Theil. Setelah dilakukan analisis dengan metode Theil didapatlah model regresi linear sederhana* $\hat{Y}_i = 18932{,}085 + 1495{,}988\, X_i$*, dengan nilai koefisien determinasi* (R^2) *sebesar* $0{,}493$ $(49{,}3\%)$*. Dari model yang didapat bisa diambil kesimpulan bahwa terdapat pengaruh positif antara tingkat pengangguran terbuka terhadap jumlah penduduk miskin, sehingga setiap kali tingkat pengangguran terbuka naik, jumlah penduduk miskin juga ikut naik. Sedangkan nilai koefisien determinasi menunjukkan bahwa sebesar* $49{,}3\%$ *kemampuan variabel tingkat pengangguran terbuka menjelaskan tentang variabel jumlah penduduk miskin. Sedangkan sisanya sebesar* $50{,}7\%$ *dijelaskan oleh variabel-variabel lain diluar penelitian ini.*

Kata kunci: *Metode Theil, Regresi Parametrik dan Nonparametrik*

PENDAHULUAN

Analisis regresi bertujuan untuk menjelaskan atau memodelkan hubungan antarvariabel. Terdapat variabel Y sebagai variabel respon, *output*, tak bebas, atau variabel yang dijelaskan, dan terdapat variabel X sebagai variabel prediktor, masukan, bebas, atau variabel penjelas [1]. Tidak sedikit pengamatan-pengamatan yang biasa dilakukan memiliki sebaran data yang tidak berdistribusi normal. Sehingga pendekatan regresi parametrik yang mensyaratkan asumsi-asumsi tidak cocok bila diterapkan untuk memodelkan data tersebut. Apabila model dari pendekatan parametrik memenuhi asumsi-asumsi yang disyaratkan, maka penaksiran parametrik akan sangat efisien. Tetapi jika terdapat salah satu saja asumsi yang tidak terpenuhi, maka akan mengakibatkan interpretasi data yang menyesatkan [2].

Salah satu metode nonparametrik yang cocok digunakan untuk memodelkan regresi linear adalah metode *Theil*. Metode *Theil* yaitu metode regresi nonparametrik yang khusus digunakan sebagai pilihan terbaik dari regresi linear sederhana apabila residual tidak berdistribusi normal [3]. Oleh sebab itu, jika ingin melakukan pemodelan regresi linear sederhana pada kasus dengan residual data tidak berdistribusi normal, metode *Theil* adalah pilihan yang tepat.

Penelitian ini bertujuan untuk membentuk model regresi linear sederhana pada data yang memiliki residual tidak berdistribusi normal. Penelitian seperti ini telah dilakukan sebelumnya dengan menggunakan data kompensasi pegawai di Badan Kepegawaian kota Samarinda. Pada penelitian ini data yang digunakan memiliki residual yang tidak berdistribusi normal bahkan setelah dilakukan transformasi pada data, sehingga untuk mendapatkan model regresi dari data yang tidak berdistribusi normal digunakanlah metode *Theil* yang merupakan salah satu metode dari regresi nonparametrik. Didapatlah model regresi linear sederhana $\hat{Y}_i = 18{,}14 + 0{,}93X_i$ yang terbentuk dengan metode *Theil*. Berdasarkan model regresi yang didapat diperoleh kesimpulan bahwa terdapat pengaruh positif antara lama kerja (X) terhadap kompensasi gaji pegawai (Y) di Badan Kepegawaian kota Samarinda [3].

Langkah awal yang perlu dilakukan pada penelitian ini ialah melakukan analisis korelasi pada data yang digunakan. Jika terdapat korelasi antar variabel X dan Y, maka data bisa digunakan untuk proses analisis regresi linear. Sedangkan jika tidak terdapat korelasi antar variabel X dan Y, maka data yang digunakan tidak cocok untuk dilakukan analisis regresi linear. Oleh karena itu harus dilakukan pengambilan data kembali (menggunakan data lainnya). Setelah didapat bahwa data yang digunakan terdapat korelasi antar variabel, maka langkah selanjutnya ialah membentuk model regresi dari data dengan metode *ordinary least square.* Selanjutnya dilakukan pengujian asumsi normalitas pada residual data. Jika residual berdistribusi normal, maka analisis bisa dilanjutkan dengan analisis-analisis regresi parametrik. Namun jika residual tidak berdistribusi normal, bisa dilakukan transformasi data pada variabel Y. Setelah itu lakukan pembentukan model kembali dengan data yang sudah di transformasi. Kemudian lakukan pengujian asumsi normalitas kembali pada residual data dengan model baru dari data yang sudah di transformasi. Selanjutnya lihat bagaimana hasil uji normalitas dari residual pada tiap transformasi. Jika ada salah satu residual yang berdistribusi normal setelah dilakukan transformasi berarti model yang didapat dari hasil transformasi tersebut bisa dilanjutkan dengan analisis-analisis regresi parametrik. Namun jika tidak satupun dari residual data yang sudah ditransformasi berdistribusi normal, maka selanjutnya sebagai alternatif digunakanlah analisis regresi nonparametrik. Salah satu metode regresi nonparametrik yang khusus untuk menangani data dengan residual tidak berdistribusi normal ialah metode *Theil.* Adapun untuk mendapatkan model dengan menggunakan metode *Theil* perlu dilakukan beberapa langkah berikut. Langkah awal yang harus dilakukan ialah melakukan penaksiran parameter-parameter koefisien *slope* (kemiringan) yang akan membentuk sebuah model regresi metode *Theil.* Setelah itu lakukan pengujian koefisien *slope* dari model yang diperoleh. Pengujian koefisien *slope* dilakukan dengan dasar prosedur uji *Kendall Tau.*

ANALISIS REGRESI

Analisis regresi merupakan suatu metode yang digunakan untuk menganalisis hubungan antara variabel. Hubungan tersebut dapat dinyatakan dalam bentuk persamaan yang menghubungkan variabel dependen dengan satu atau lebih variabel independen. Regresi linear sederhana menjelaskan tentang hubungan antardua variabel yang biasanya dinyatakan dalam suatu model regresi, serta merupakan teknik dalam statistika parametrik yang digunakan secara umum untuk menganalisis rata-rata *respons* dari variabel Y yang berubah sehubungan dengan besarnya pengaruh dari variabel X. Dalam regresi linear variabel Y bisa disebut sebagai variabel *respons*, *output* dan dependen. Adapun variabel X dapat disebut sebagai variabel *predictor* (digunakan untuk memprediksi nilai dari Y), *input*, *regressors*, dan independen [1].

Hubungan antar variabel pengamatan X dan Y dapat dinyatakan dalam suatu model atau pernyataan matematis. Salah satu bentuk yang paling sederhana adalah model linear, yaitu:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \qquad i = 1,2,\cdots,n \tag{1}$$

Dalam Persamaan 1 β_0 dan β_1 keduanya merupakan konstanta yang tidak diketahui nilainya. Sedangkan ε_i adalah nilai suatu variabel acak ε yang mempresentasikan faktor-faktor lain yang mempengaruhi nilai Y, biasa juga disebut sebagai variabel galat acak [4].

Metode *Ordinary Least Square*

Metode *ordinary least square* merupakan salah satu teknik penaksiran yang digunakan untuk menaksir parameter β_0 dan β_1 yang tidak diketahui besarnya. Misalkan $\hat{\beta}_0$ dan $\hat{\beta}_1$ masing-masing nilai taksiran bagi parameter β_0 dan β_1, maka nilai taksiran bagi Y_i dinotasikan $\hat{Y}_i$. Sehingga diperoleh persamaan garis regresi sampel (*sample regression line*) sebagai berikut [4]:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 \mathrm{X}_i \tag{2}$$

Dengan metode *ordinary least square*, nilai taksiran bagi parameter regresi β_0 dan β_1, masing-masing adalah $\hat{\beta}_0$ dan $\hat{\beta}_1$, dengan:

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} \tag{3}$$

$$\hat{\beta}_1 = \frac{n \sum_{i=1}^{n} X_i Y_i - \sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{n \sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2} \tag{4}$$

Uji Asumsi Normalitas Residual

Penelitian tentang pemilihan uji normalitas terbaik berdasarkan jumlah data dan tingkat signifikansi menyatakan bahwa *Shapiro-Wilk* sebagai uji normalitas terbaik untuk tingkat signifikansi 5% dan 10% dengan banyaknya data berkisar 10 sampai 2000 data [5]. Uji *Shapiro-Wilk* adalah sebuah metode atau rumus perhitungan sebaran data yang dibuat oleh Shapiro dan Wilk. Metode *Shapiro-Wilk* adalah metode uji normalitas yang efektif dan valid digunakan untuk sampel berjumlah kecil.

Langkah-langkah melakukan uji *Shapiro-wilk* [5]:

1. Menentukan hipotesis dan selang kepercayaan:
   H_0 : Data berdistribusi normal
   H_1 : Data tidak berdistribusi normal
   Dengan selang kepercayaan 95% ($\alpha = 0,05$)
2. Data diurutkan dari kecil ke besar ($X_1 < X_2 < \cdots < X_n$)
3. Rumus uji *Shapiro-wilk* sebagai berikut;
$$W = \frac{\left[\sum_{i=1}^{n} a_i X_i\right]^2}{\sum_{i=1}^{n} (X_i - \bar{X})^2} \tag{5}$$
   W : nilai hitung *shapiro-wilk*
   a_i : nilai a_i pada tabel *coefficient shapiro-wilk*
   X_i : data observasi ke- i
   $\bar{X}$: nilai rata-rata dari data
4. Kriiteria uji yang digunakan sebagai berikut :
   Jika $W < W_{(n,\alpha)}$ atau *p-value* < α, maka tolak H_0
   Jika $W \geq W_{(n,\alpha)}$ atau *p-value* ≥ α, maka terima H_0
5. Kesimpulan

REGRESI NONPARAMETRIK

Pengamatan-pengamatan yang biasa dilakukan tidak selalu membentuk pola – pola seperti garis lurus, eksponensial, dan lain-lain. Sehingga beberapa asumsi-asumsi yang biasa digunakan untuk pendekatan regresi parametrik tidak terpenuhi. Sebagai alternatifnya dilakukan pendekatan dengan model regresi nonparametrik.

Regresi nonparametrik merupakan suatu analisis statistika nonparametrik dimana dalam penggunaannya tidak mensyaratkan asumsi-asumsi yang harus dipenuhi. Prosedur nonparametrik sebaiknya digunakan [6]:

1. Jika hipotesis yang diuji tidak melibatkan suatu parameter populasi
2. Jika data sudah diukur menggunakan skala yang lebih lemah dibanding yang dipersyaratkan oleh prosedur parametrik yang semestinya digunakan.
3. Jika asumsi-asumsi yang diperlukan pada suatu prosedur pengujian parametrik tidak terpenuhi.
4. Jika hasil-hasil riset harus segera disajikan dan perhitungan-perhitungan terpaksa dikerjakan secara manual.

METODE *THEIL*

Metode *Theil* merupakan salah satu metode estimasi regresi nonparametrik yang menaksir koefisien kemiringan *(slope)* garis regresi dengan cara mencari median kemiringan seluruh pasangan dari titik-titik variabel X dan Y, dengan tidak ada nilai X yang sama. Sedangkan pengujian koefisien kemiringan *(slope)* disusun dengan statistik *Kendall Tau* yang digunakan untuk mengetahui bentuk hubungan variabel-variabel pada persamaan regresi [3]. Adapun beberapa langkah untuk membentuk model regresi linear sederhana dengan metode *Theil*, sebagai berikut:

1) Penaksiran parameter koefisien *slope*

   Misalkan sebuah sampel yang terdiri atas n pasangan hasil pengamatan (X_1, Y_1), (X_2, Y_2), …,(X_n, Y_n) dengan variabel-variabel X dan Y kontinu, dimana pasangan (X_1, Y_1) merupakan hasil-hasil pengukuran terhadap observasi yang sama (ke- i). Sehingga untuk setiap pasangan (X_i, Y_i) dan (X_j, Y_j) koefisien kemiringannya adalah:

$$b_{ij} = \frac{Y_j - Y_i}{X_j - X_i}, \text{ untuk i < j dan } X_i \neq X_j \tag{6}$$

   Kemudian untuk mendapatkan nilai estimasi parameter $\hat{\beta}_0$ dan $\hat{\beta}_1$ digunakanlah rumus berikut:

$$\hat{\beta}_1 = \text{Median}\left(b_{ij}\right) \tag{7}$$

$$\hat{\beta}_0 = \text{Median}\ (Y_i) - \hat{\beta}_1 \text{Median}\ (X_i) \tag{8}$$

   Dengan Median $\left(b_{ij}\right)$ sebagai nilai tengah dari koefisien *slope*, Median (Y_i) sebagai nilai tengah dari variabel Y, dan Median (X_i) sebagai nilai tengah dari variabel X.

2) Pengujian koefisien *slope*

   Langkah awal dalam melakukan uji *Kendall Tau* ialah susun pasangan–pasangan $(X_i\ ,Y_i)$ dalam sebuah kolom menurut besarnya nilai-nilai observasi X, dari nilai observasi X yang paling kecil. Disini dapat dikatakan bahwa nilai-nilai X berada dalam urutan yang wajar *(natural order)*. Selanjutnya bandingkan setiap nilai observasi Y satu demi satu dengan setiap nilai Y dibawahnya. Jika nilai Y yang dibawah lebih besar dari Y diatasnya, maka arah nilai pengamatannya wajar *(concordant)*. Dan jika nilai Y yang dibawah lebih kecil dari Y diatasnya, maka arah nilai pengamatannya tidak wajar *(discordant)*. Tetapkan P sebagai banyaknya pasangan *concordant* dan Q sebagai banyaknya pasangan *discordant*. Sedangkan S sebagai selisih antara P dan Q $(S = P - Q)$.

   Adapun langkah-langkah berikutnnya untuk melakukan pengujian koefisien *slope* (kemiringan) ialah sebagai berikut [6]:

   (i) Menentukan hipotesis dan selang kepercayaan
   $\mathrm{H_0}$: tidak terdapat hubungan antara variabel X terhadap variabel Y
   $\mathrm{H_1}$: terdapat hubungan antara variabel X terhadap variabel Y
   Dengan selang kepercayaan 95% $(\alpha = 0{,}05)$

   (ii) Statistik uji yang digunakan ialah rumus uji korelasi *Kendall Tau*, sebagai berikut:

$$\tau = \frac{P - Q}{n(n-1)/2} \tag{9}$$

   τ : nilai hitung *Kendall Tau*
   P : banyaknya pasangan *concordant*
   Q : banyaknya pasangan *discordant*
   n : jumlah data

   (iii) Kriteria uji yang digunakan sebagai berikut:
   $|\tau| > \tau_{n,\alpha/2}$, maka tolak $\mathrm{H_0}$, sedangkan
   $|\tau| \leq \tau_{n,\alpha/2}$, maka terima $\mathrm{H_0}$

   (iv) Kesimpulan

3) Interval Kepercayaan Koefisien Regresi *Slope*

   Proses ini dilakukan untuk melihat interval dari koefisien regresi *slope*, sejauh mana parameter koefisien tersebut bergerak. Adapun langkah-langkah untuk melakukan proses ini sebagai berikut:

   (i) Menghitung konstanta interval kepercayaan menggunakan rumus berikut [6]:

$$k = \frac{{}_nC_2 - S_{n,\frac{\alpha}{2}} + 2}{2} \tag{10}$$

dimana k merupakan konstanta untuk interval kepercayaan, ${}_nC_2$ merupakan banyaknya pasangan b_{ij} dan $S_{n,\frac{\alpha}{2}}$ merupakan titik kritis *Kendall Tau* untuk n (jumlah data) pada taraf α.

(ii) Menentukan β_L sebagai batas bawah untuk interval kepercayaan yang dilihat dari nilai b_{ij} ke- k yang sudah di urutkan dari kecil ke besar.

(iii) Menentukan β_U sebagai batas atas untuk interval kepercayaan yang dilihat dari nilai b_{ij} ke- k yang sudah di urutkan dari besar ke kecil.

KOEFISIEN DETERMINASI

Koefisien determinasi bertujuan untuk mengetahui sejauh mana variabel bebas X menaksir variabel dependen Y pada model probabilistik, dalam hal ini perlu diketahui beberapa jenis ukuran keragaman. Diantaranya adalah jumlah kuadrat total (JKT), yang merupakan ukuran keragaman nilai Y_i di sekitar rata-ratanya ($\bar{Y}$). Jumlah kuadrat regresi (JKR) mencerminkan hubungan antara variabel X dan Y, dan jumlah kuadrat sisaan (JKS) mencerminkan keragaman karna faktor-faktor lain selain hubungan antara X dan Y [4].

Ukuran keragaman koefisien determinasi (R^2) merupakan bagian keragaman dari variabel Y yang dijelaskan oleh persamaan regresinya. Nilai R^2 dihitung dengan rumus berikut:

$$R^2 = \frac{JKR}{JKT} = \frac{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} \tag{11}$$

Dengan $\hat{Y}_i$ sebagai prediksi nilai Y dengan model yang dimiliki, Y_i nilai Y aktual (asli) dan $\bar{Y}$ ialah nilai rata-rata Y [4].

HASIL DAN PEMBAHASAN

Data yang digunakan dalam penelitian ini ialah data sekunder yang diperoleh dari publikasi BPS Indonesia, yaitu data semesteran jumlah penduduk miskin (JPM) dan data tingkat pengangguran terbuka (TPT) di Indonesia dari tahun 2012 – 2022. Jumilah data yang digunakan sebanyak 21 data, dikarenakan terdapat dua kali survei atau pengambilan data pada setiap tahunnya kecuali pada tahun 2022. Variabel yang digunakan ialah jumlah penduduk miskin sebagai variabel dependen dan variabel tingkat pengangguran terbuka sebagai variabel independen.

Tabel 1. Statistik Deskriptif JPM dan TPT

Variabel	N	Maximum	Minimum	Mean	Median	Variance
JPM (ribu jiwa)	21	29132,40	24785,87	27301,04	27727,78	1559071,69
TPT (%)	21	7,07	4,94	5,78	5,81	0,29

Berdasarkan Tabel 1 jumlah penduduk miskin terbanyak terjadi pada tahun 2012 periode I sebanyak 29132,40 ribu jiwa sedangkan paling sedikit pada tahun 2019 periode II sebanyak 24785,87 ribu jiwa. Untuk tingkat pengangguran terbuka tertinggi terjadi pada tahun 2020 periode II sebesar 7,07 % sedangkan paling rendah pada tahun 2020 periode I sebesar 4,94 %. Nilai rata-rata untuk jumlah penduduk miskin sebesar 27301,04 ribu jiwa sedangkan nilai rata-rata tingkat pengangguran terbuka sebesar 5,78 %. Nilai varians untuk jumlah penduduk miskin sebesar 1559071,69, sedangkan nilai varians untuk tingkat pengangguran terbuka sebesar 0,29.

Analisis korelasi

Analisis korelasi bertujuan untuk mengetahui ada tidaknya korelasi (hubungan) antar variabel. Jika terdapat korelasi antar variabel yang digunakan, maka data bisa digunakan untuk proses analisis regresi linear. Sedangkan jika tidak terdapat korelasi anatar variabel yang digunakan, maka data yang digunakan tidak cocok untuk dilakukan analisis regresi linear. Sehingga harus dilakukan pengambilan data kembali (menggunakan data lainnya). Hasil uji korelasi *pearson* dapat dilihat pada Tabel 2.

Tabel 2. Hasil Analisis Korelasi

	JPM	*TPT (%)*
JPM	1	0,5638
TPT (%)	0,5638	1

Berdasarkan Tabel 2 diketahui bahwa korelasi antara tingkat pengangguran terbuka (X) dengan jumlah penduduk miskin (Y) menghasilkan nilai korelasi 0,5638 lebih besar dari nilai R tabel $(R_{a,n-2})$ sebesar 0,4329. Maka dapat disimpulkan bahwa terdapat korelasi antara tingkat pengangguran terbuka (X) dengan jumlah penduduk miskin (Y). Korelasi bernilai positif $(+)$ menunjukkan hubungan positif antar variabel yang berarti semakin meningkatnya tingkat pengangguran terbuka semakin meningkat pula jumlah penduduk miskin.

Membentuk Model dengan *Ordinary Least Square (OLS)*

Sebelum melakukan uji asumsi normalitas residual perlu diketahui model regresi terlebih dahulu agar nilai residual dari model bisa diketahui. Untuk itu metode *OLS* terlebih dahulu digunakan untuk mendapatkan model dan juga nilai residual dari model. Adapun model yang terbentuk dari data yang digunakan sebagai berikut :

$$\hat{Y}_i = 19847 + 1290{,}2\ X_i \tag{12}$$

Setelah didapat model regresi dari data yang digunakan, maka didapatlah nilai residual data dari model yang dimiliki. Untuk itu bisa dilakukan uji asumsi normalitas residual pada data residual yang telah diperoleh. Hasil dari uji asumsi residual bisa dilihat pada Tabel 3.

Tabel 3. Uji Asumsi Normalitas Residual dengan Metode *Shapiro-Wilk*

Data	*N*	*P-Value*
Residual	21	0,02218

Berdasarkan Tabel 3 diketahui bahwa nilai *p-value* dari data residual sebesar 0,02218. Maka, dapat disimpulkan bahwa residual dari data tidak berdistribusi normal pada *alpha* 5% dikarenakan nilai *p-value* lebih kecil dari *alpha* (0,02218 < 0,05). Dengan begitu model regresi yang diperoleh tidak bisa dilanjutkan dengan analisis-analisis regresi parametrik.

Transformasi data

Setelah residual dari model pertama yang didapat tidak berdistribusi normal, maka selanjutnya lakukan transformasi pada variabel Y. Seperti sebelumnya untuk melakukan uji asumsi residual diperlukan model dan data residual dari data yang telah ditransformasi. Ringkasan hasil uji asumsi residual dan model dari data yang telah ditransformasi bisa dilihat pada Tabel 4.

Tabel 4. Ringkasan Hasil Uji Asumsi Residual dan Model dari Data Transformasi

Transformasi	*N*	Model	*P-Value*
$1/Y$	21	$1/\hat{Y}_i = 4{,}705\text{e}-05 + 1{,}791\text{e}-06\ X_i$	0,02332
$\sqrt{Y}$	21	$\sqrt{\hat{Y}_i} = 142{,}452 + 3{,}936\ X_i$	0,02224
$\log Y$	21	$\log \hat{Y}_i = 4{,}315216 + 0{,}020862\ X_i$	0,02256

Berdasarkan Tabel 4 bisa dilihat model yang didapat dari masing-masing transformasi. Begitu juga dengan nilai *p-value* dari data residual masing-masing model dengan data yang telah ditransformasi. Terlihat bahwa semua nilai *p-value* lebih kecil dari *alpha*. Maka, dapat disimpulkan bahwa residual dari data yang telah ditransformasi juga tidak berdistribusi normal. Dengan begitu tidak satupun model yang didapat dari data yang telah ditransformasi bisa dilanjutkan dengan analisis-analisis regresi parametrik. Oleh sebab itu, sebagai alternatif untuk mendapatkan model regresi linear dari data dengan residual tidak berdistribusi normal digunakanlah analisis regresi nonparametrik, yaitu metode *Theil.*

Penaksiran parameter koefisien regresi *slope* (kemiringan)

Adapun langkah-langkah untuk menaksir parameter koefisien *slope* sebagai berikut.

(1) Memastikan tidak ada data yang bernilai sama pada variabel X. Maka, jika pada data terdapat adanya X yang bernilai sama, diambil nilai rata-rata dari nilai Y pada observasi tersebut. Seperti halnya pada kasus kali ini yang terdapat nilai X yang sama yaitu pada data observasi X ke-9 (X_9) dan data observasi X ke-12 (X_{12}). Sehingga untuk Y diambil nilai rata-rata dari data observasi Y ke-9 (Y_9) dan data observasi Y ke-12 (Y_{12}).
Sehingga :
$$\frac{Y_9 + Y_{12}}{2} = \frac{28005{,}39 + 26582{,}99}{2} = 27294{,}19$$
Didapatlah data baru untuk Y yaitu 27294,19.

(2) Mengurutkan data berdasarkan nilai X terkecil ke terbesar, sedangkan Y mengikuti. Dengan tidak ada X yang bernilai sama.

Tabel 5. Data Baru dengan Tidak Ada Nilai X yang Sama

No	X	Y
1	4,94	26424,02
2	4,98	25144,72
3	5,10	25949,80
4	5,23	24785,87
5	5,30	25674,58
6	5,33	27771,22
7	5,50	27294,19
8	5,61	27764,32
9	5,70	28280,01
10	5,81	28592,79

No	X	Y
11	5,83	26161,16
12	5,88	28066,60
13	5,94	27727,78
14	6,13	28594,64
15	6,17	28553,93
16	6,18	28513,57
17	6,26	27542,77
18	6,37	29132,40
19	6,49	26503,65
20	7,07	27549,69

(3) Menghitung koefisien *slope* (b_{ij} , untuk i < j dan $X_i \neq X_j$), untuk setiap pasangan (X_i, Y_i) dan (X_j, Y_j) menggunakan Persamaan 6 untuk setiap pasangan b_{ij}.
Sehingga:

Untuk pasangan (X_1, Y_1) dan (X_2, Y_2)

$$b_{1,2} = \frac{Y_2 - Y_1}{X_2 - X_1} = \frac{25144{,}72 - 26424{,}02}{4{,}98 - 4{,}94} = -31982{,}5$$

Untuk pasangan (X_2, Y_2) dan (X_3, Y_3)

$$b_{2,3} = \frac{Y_3 - Y_2}{X_3 - X_2} = \frac{25949{,}8 - 25144{,}72}{5{,}1 - 4{,}98} = 6709$$

dan seterusnya sehingga didapat hasil b_{ij} untuk setiap pasangan.

(4) Setelah didapat nilai b_{ij} untuk setiap pasangan, selanjutnya lakukan penaksiran parameter-parameter koefisien *slope*. Pertama lakukan penaksiran parameter β_1 yang akan dinotasikan sebagai $\hat{\beta}_1$. Dengan menggunakan Persamaan 7.

$$\hat{\beta}_1 = Median\ b_{ij} = 1495{,}988$$

(5) Setelah nilai $\hat{\beta}_1$ didapat, maka selanjutnya lakukan penaksiran parameter β_0 yang akan dinotasikan sebagai $\hat{\beta}_0$. Dengan menggunakan Persamaan 8.

$$\begin{aligned}\hat{\beta}_0 &= Median\ Y_i - \hat{\beta}_1\ Median\ X_i \\ &= 27638{,}735 - 1495{,}988\ (5{,}82) = 18932{,}085\end{aligned}$$

(6) Maka didapatlah model regresi linear nonparametrik dengan metode *Theil* sebagai berikut:

$$\hat{Y}_i = 18932{,}085 + 1495{,}988\ X_i \tag{13}$$

Model pada Persamaan 13 menunjukkan bahwa terdapat pengaruh positif $(+)$ antara variabel X terhadap Y.

Pengujian koefisien *slope* (kemiringan)

Adapun langkah-langkah pengujiannya sebagai berikut:

(1) Menentukan hipotesis dan selang kepercayaan

H_0 : tidak terdapat hubungan antara variabel X terhadap variabel Y

H_1 : terdapat hubungan antara variabel X terhadap variabel Y

Dengan selang kepercayaan 95% $(\alpha = 0{,}05)$

(2) Melakukan pengujian dengan statistik uji pada Persamaan 9.

Sebelum masuk ke Persamaan 10 terlebih dahulu perlu diketahui nilai P dan Q. Oleh karena itu terlebih dahulu hitung nilai P dan Q. Dengan cara membandingkan nilai Y dengan nilai-nilai Y dibawahnya. Sehingga didapat $P = 132$ dan $Q = 58$.

$$\tau = \frac{P - Q}{n(n-1)/2} = \frac{132 - 58}{20(19)/2} = 0{,}389$$

(3) Selanjutnya bandingkan nilai τ_{hit} dengan nilai τ_{tab} $\left(\tau_{20,0.025} = 0{,}326\right)$, dengan kriteria uji sebagai berikut:

$|\tau| > \tau_{n,\alpha/2}$, maka tolak H_0, sedangkan

$|\tau| \leq \tau_{n,\alpha/2}$, maka terima H_0

Maka, dari hasil perhitungan didapat bahwa $|\tau| > \tau_{n,\alpha/2}$ $(0{,}389 > 0{,}326)$. Oleh karena itu tolak H_0.

(4) Karena $|\tau| > \tau_{n,\alpha/2}$, maka dapat disimpulkan bahwa terdapat pengaruh antara tingkat pengangguran terbuka terhadap jumlah penduduk miskin

Menghitung interval kepercayaan koefisien regresi *slope*

Adapun langkah-langkah untuk melakukannya ialah:

(1) Menghitung konstanta interval kepercayaan dengan ${}_{20}C_2 = 190$ dan $S_{20,0,025} = 62$ menggunakan Persamaan 10.

$$k = \frac{{}_nC_2 - S_{n,\frac{\alpha}{2}} + 2}{2} = \frac{190 - 62 + 2}{2} = 65$$

(2) Setelah didapat konstanta ialah 65, maka didapat $\beta_L = 398{,}453$ sebagai batas bawah interval yang dilihat dari nilai b_{ij} ke – 65 yang sudah diurutkan berdasarkan nilai terkecil ke terbesar.
(3) Sedangkan $\beta_U = 2505{,}984$ sebagai batas atas interval yang dilihat dari nilai b_{ij} ke – 65 yang sudah diurutkan berdasarkan nilai terbesar ke terkecil.
(4) Maka, didapat interval koefisien regresi *slope* berada diantara :

$$398{,}453 < \hat{\beta}_1 < 2505{,}984$$

Menghitung Nilai Koefisien Determinasi

Koefisien determinasi untuk menunjukkan besarnya keragaman variabel Y yang bisa dijelaskan oleh variabel X. Dengan $\sum_{i=1}^{20}\left(\widehat{Y}_i - \bar{Y}\right)^2 = 30169773{,}62$ dan $\sum_{i=1}^{20}(Y_i - \bar{Y})^2 = 14876407$. Hitung *R Square* menggunakan Persamaan 11.

$$R^2 = \frac{JKR}{JKT} = \frac{\sum_{i=1}^{20}\left(\widehat{Y}_i - \bar{Y}\right)^2}{\sum_{i=1}^{20}(Y_i - \bar{Y})^2} = \frac{30169773{,}62}{14876407} = 0{,}493$$

Berdasarkan perhitungan diatas didapat R^2 sebesar $0{,}493$, nilai ini menunjukkan bahwa pengaruh tingkat pengangguran terbuka terhadap jumlah penduduk miskin sebesar $49{,}3\%$, sedangkan sisanya sebesar $50{,}7\%$, dipengaruhi variabel yang tidak diteliti.

PENUTUP

Berdasarkan hasil penelitian yang telah dilakukan, maka didapat model regresi linear dari data yang residualnya tidak berdistribusi normal dengan metode *Theil*, yaitu:

$$\widehat{Y}_i = 18932{,}085 + 1495{,}988\, X_i$$

Model tersebut menjelaskan bahwa terdapat pengaruh positif $(+)$ antara variabel X terhadap Y. Hal ini menunjukkan bahwa disaat variabel X meningkat, maka variabel Y ikut meningkat. Sehingga jika tingkat pengangguran terbuka meningkat sebesar 1%, maka jumlah penduduk miskin akan meningkat sebanyak 1495,988 ribu jiwa.

Hasil penelitian ini juga menunjukkan bahwa terdapat pengaruh yang signifikan antara tingkat pengangguran terbuka terhadap jumlah penduduk miskin di Indonesia tahun 2012 hingga 2022. Dengan nilai *R Square* sebesar 49,3% yang berarti bahwa sebesar itulah kemampuan tingkat pengangguran terbuka menjelaskan tentang jumlah penduduk miskin dan juga berarti masih terdapat variabel-variabel lain yang juga mempengaruhi jumlah penduduk miskin diluar penelitian ini.

DAFTAR PUSTAKA

[1]. Kurniawan, R., dan Yuniarto, B., 2016. *Analisis Regresi: Dasar dan Penerapannya dengan R*, KENCANA. Jakarta.

[2]. Zahro, J., Caraka, R. E., dan Herliansyah, R., 2018. *Aplikasi Generalized Linear Model pada R*, Innosain, Yogyakarta.

[3]. Putra, T. D., Fathurahman, M., dan Yuniarti, D., 2013, Pemodelan Regresi Linier Menggunakan Metoda Theil pada Kasus Kompensasi Pegawai di Badan Kepegawaian Daerah Kota Samarinda. *EKSPONENSIAL*, 4(1):33-38.

[4]. Kusnandar, D., Debataraja, N. N., Mara, M. N., dan Satyahadewi, N., 2019. *Metode Statistika Serta Aplikasinya dengan MINITAB, Excel dan R*, UNTAN Press. Pontianak.

[5]. Razali, N. M., dan Wah, Y. B., 2011. Power Comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling Tests, *Journal of Statistical Modeling and Analytics*, 2(1):21-33.

[6] Sukarna, Zaki, A., dan Thomas, Y., 2015, Regresi Linier Nonparametrik dengan Metode Theil dengan Studi Kasus Inflasi dan Uang Beredar di Indonesia. *JMathCoS*, 1(1):33-39.

RIDO AMARRULLAH	:	Jurusan Matematika, FMIPA UNTAN, Pontianak, rido@student.untan.ac.id
SHANTIKA MARTHA	:	Jurusan Matematika, FMIPA UNTAN, Pontianak, shantika.martha@math.untan.ac.id
WIRDA ANDANI	:	Jurusan Matematika, FMIPA UNTAN, Pontianak, wirda.andani@math.untan.ac.id