

## PERBANDINGAN TRANSFORMASI BOX-COX DAN REGRESI KUANTIL MEDIAN DALAM MENGATASI HETEROSKEDASTISITAS

Ni Wayan Yuni Cahyani<sup>§1</sup>, I Gusti Ayu Made Srinadi<sup>2</sup>, Made Susilawati<sup>3</sup>

<sup>1</sup>Jurusan Matematika, Fakultas MIPA – Universitas Udayana [Email: yuni.cahyani44@gmail.com]

<sup>2</sup>Jurusan Matematika, Fakultas MIPA – Universitas Udayana [Email: srinadigustiayumade@yahoo.com]

<sup>3</sup>Jurusan Matematika, Fakultas MIPA – Universitas Udayana [Email: susilawati.made@gmail.com]

<sup>§</sup>Corresponding Author

### ABSTRACT

*Ordinary least square (OLS) is a method that can be used to estimate the parameter in linear regression analysis. There are some assumption which should be satisfied on OLS, one of this assumption is homoscedasticity, that is the variance of error is constant. If variance of the error is unequal that so-called heteroscedasticity. The presence heteroscedasticity can cause estimation with OLS becomes inefficient. Therefore, heteroscedasticity shall be overcome. There are some method that can used to overcome heteroscedasticity, two among those are Box-Cox power transformation and median quantile regression. This research compared Box-Cox power transformation and median quantile regression to overcome heteroscedasticity. Applied Box-Cox power transformation on OLS result R<sup>2</sup> point are greater, smaller RMSE point and confidencen interval more narrow, therefore can be concluded that applied of Box-Cox power transformation on OLS better of median quantile regression to overcome heteroscedasticity.*

**Keywords:** *Heteroscedasticity, OLS, Box-Cox power transformation and Median Quantile Regression.*

### 1. PENDAHULUAN

Estimasi parameter pada analisis regresi linear dilakukan dengan menggunakan Metode Kuadrat Terkecil. Pendugaan parameter regresi dengan metode kuadrat terkecil memiliki beberapa asumsi yang harus dipenuhi agar mendapatkan penduga yang bersifat *Best Linear Unbiased Estimator* (BLUE). Salah satu asumsi tersebut adalah homoskedastisitas yang berarti nilai varians dari galat adalah sama (konstan). Apabila asumsi nilai varians dari galat tidak konstan maka dikatakan terjadi heteroskedastisitas.

Adanya heteroskedastisitas dapat menyebabkan pendugaan parameter dengan menggunakan metode kuadrat terkecil tidak dapat menghasilkan penduga yang memenuhi sifat BLUE sehingga model yang diperoleh

menjadi tidak efisien dan tidak dapat dipercaya (Gujarati [2]). Oleh karena itu, keberadaan heteroskedastisitas harus diatasi. Terdapat beberapa metode yang dapat digunakan untuk mengatasi heteroskedastisitas, antara lain Transformasi Box-Cox dan regresi kuantil median.

Penelitian tentang kemampuan transformasi Box-Cox dalam mengatasi heteroskedastisitas telah dilakukan oleh Ispriyanti [3] dan Kristina [6]. Dari penelitian tersebut disimpulkan bahwa transformasi Box-Cox dapat digunakan untuk mengatasi heteroskedastisitas dengan nilai  $\lambda$  berada pada interval  $(-2, 2)$ , dan penelitian tentang regresi kuantil median untuk mengatasi heteroskedastisitas pernah dilakukan oleh Uthami, et. al [8].

Pada penelitian ini penulis ingin membandingkan kemampuan transformasi

Box-Cox dan regresi kuantil median dalam mengatasi masalah heteroskedastisitas dan untuk melihat metode mana yang lebih baik dalam menangani heteroskedastisitas. Data yang digunakan dalam penelitian ini adalah data tentang lama pasien bertahan hidup setelah melakukan operasi liver tertentu yang diperoleh dari buku Neter, et. al [7] halaman 202 – 203.

**1.1 Metode Kuadrat Terkecil**

Metode kuadrat terkecil adalah metode yang bertujuan untuk meminimumkan jumlah kuadrat galat (*sum square error*). Pendugaan koefisien regresi linear berganda dengan metode kuadrat terkecil dapat dilakukan dengan menggunakan persamaan matriks (Neter, et. al [7]), sebagai berikut:

$$b = (X^T X)^{-1} X^T Y \tag{1}$$

dan untuk mencari selang kepercayaan pada metode kuadrat terkecil dapat dilakukan melalui persamaan:

$$b_i \pm t_{(\frac{\alpha}{2}, df)} se\{b_i\}. \tag{2}$$

**1.2 Transformasi Box-Cox**

Transformasi Box-Cox merupakan transformasi pangkat pada variabel respons yang dikembangkan oleh Box dan Cox, yang bertujuan untuk menormalkan data, melinearkan model regresi dan menghomogenkan varians. Box dan Cox mempertimbangkan kelas transformasi berparameter tunggal, yaitu  $\lambda$  yang dipangkatkan pada variabel respons  $Y$ , sehingga diperoleh model transformasinya  $Y^\lambda$  dengan  $\lambda$  merupakan parameter yang harus diduga. Transformasi Box-Cox hanya diberlakukan pada variabel respons  $Y$  yang bertanda positif (Draper & Smith [1]). Prosedur utama yang dilakukan pada Transformasi Box-Cox adalah menduga parameter  $\lambda$ , yang dilakukan dengan mencari nilai  $W_i$  dan  $V_i$  melalui persamaan:

$$W_i = \begin{cases} \frac{Y_i^\lambda - 1}{\lambda} & ; \text{untuk } \lambda \neq 0 \\ \ln Y & ; \text{untuk } \lambda = 0 \end{cases} \tag{3}$$

$$V_i = \begin{cases} \frac{W_i}{Y_{gso}^{\lambda-1}} & ; \text{untuk } \lambda \neq 0 \\ Y_{gso} \ln Y & ; \text{untuk } \lambda = 0 \end{cases} \tag{4}$$

dengan  $Y_{gso} = \sqrt[n]{Y_1 Y_2 \dots Y_n} = (\prod_{i=1}^n Y_i)^{\frac{1}{n}}$

Selanjutnya meregresikan antara  $V_i$  dan  $X$  sehingga diperoleh nilai JKG dan nilai  $\lambda$  yang dipilih adalah  $\lambda$  yang menghasilkan nilai JKG terkecil. Setelah itu dilakukan transformasi pada variabel respon  $Y$  dengan nilai  $\lambda$  yang telah di pilih. Pada tabel berikut disajikan nilai  $\lambda$  dan model transformasinya.

Tabel 1. Nilai  $\lambda$  dan Model Transformasinya

Nilai $\lambda$	Model tranformasi
-2	$Y^{-2} = \frac{1}{Y^2}$
-1	$Y^{-1} = \frac{1}{Y}$
-0.5	$Y^{-0.5} = \frac{1}{\sqrt{Y}}$
0	$Y^0 = \ln Y$
0.5	$Y^{0.5} = \sqrt{Y}$
1	$Y^1 = Y$
2	$Y^2 = Y^2$

Sumber: Neter, et al.(1997)

**1.3 Regresi Kuantil Median**

Regresi kuantil median adalah metode yang mendefinisikan median sebagai solusi untuk meminimumkan jumlah galat mutlak (Koenker & Basset [4]). Kuantil dapat dioperasikan dengan penyusunan atau pengurutan sampel pengamatan sehingga lebih mudah menentukan letaknya dan dapat mendefinisikan kuantil melalui alternatif yang sederhana sebagai masalah optimasi.

Untuk mendefinisikan regresi kuantil median dalam model dengan sampel acak  $\{y_1, y_2, y_3, \dots, y_n\}$  digunakan persamaan:

$$\min_{\mu \in R} \sum_{i=1}^n (y_i - \mu)^2 \tag{5}$$

Sehingga diperoleh rata-rata sampel yang merupakan pengestimasi dari rata-rata populasi  $E(Y|X)$ . Jika skalar  $\mu = X_i \beta$  maka persamaan (5) menjadi:

$$\min_{\beta \in R^p} \sum_{i=1}^n (y_i - X_i \beta)^2 \tag{6}$$

Dengan  $X_i$  merupakan variabel bebas ke- $i$ ,  $\beta$  merupakan parameter, dan  $y_i$  merupakan variabel respons ke- $i$ .

Penelitian yang dilakukan Koeker dan Basset juga membahas masalah regresi tersebut,

yang kemudian berkembang menjadi median sampel yang dinyatakan dalam persamaan:

$$\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^n |y_i - X_i \beta| \quad (7)$$

Kemudian secara umum dispesifikasikan dalam fungsi kuantil bersyarat ke- $\tau$  dapat mempertimbangkan penduga bagi  $\beta(\tau)$  yaitu  $(\hat{\beta}(\tau))$  sehingga diperoleh ide bahwa masalah tersebut dapat dinyatakan pada persamaan:

$$\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^n \rho_{\tau}(y_i - Q_i(Y|X)) \quad (8)$$

Dengan  $\tau$  adalah indeks kuantil  $\epsilon(0,1)$ ,  $\rho_{\tau}$  adalah *loss function* yang asimetrik, dan  $Q_i(Y|X) = X\beta(\tau)$  merupakan fungsi kuantil ke- $\tau$  dari  $Y$  dengan syarat  $X$

Jika  $Y$  merupakan sebaran variabel acak kontinu dan  $x$  adalah salah satu vektor regresor  $X$ , maka fungsi kuantil bersyarat dalam fungsi kuantil ke- $\tau$  dapat didefinisikan sebagai:

$$Q_i(Y|X) = \inf\{y : F_y(y|X) \geq \tau\} \quad (9)$$

dengan  $F_y(y|X) =$  fungsi sebaran dari  $Y$  dengan syarat  $X$  dan fungsi kepekatannya  $f_y(y|X)$ .

Regresi kuantil median mengembangkan persamaan sederhana untuk regresor  $X$ . Dalam hal ini diasumsikan spesifikasi linear untuk kuantil bersyarat dari variabel respons  $Y$  untuk  $p$  vektor variabel bebas  $X$ .

$$Q(\tau|X_i, \beta(\tau)) = X_i \beta(\tau) \quad (10)$$

dengan  $\beta(\tau)$  adalah vektor koefisien pada kuantil ke- $\tau$ .

Untuk meminimalkan kuantil tidak bersyarat di atas dengan estimasi regresi kuantil median adalah:

$$\hat{\beta}(\tau) = \arg \min_{(\tau)} \left\{ \sum_i \rho_{\tau}(y_i - X_i \beta(\tau)) \right\} \quad (11)$$

Untuk mencari selang kepercayaan pada regresi kuantil median terdapat tiga metode yang dapat digunakan yaitu metode sparsity, metode rank dan metode bootstrap. Pada penelitian ini yang digunakan untuk membangun selang kepercayaan adalah metode bootstrap yaitu dengan melakukan proses bootstrap yang kemudian hasilnya akan dimasukkan ke dalam batas distribusi normal dengan batas yang sudah ditentukan, yaitu:

$$b_{\tau,j} \pm 1,96 \left( se(b_{\tau,j}) \right) \quad (12)$$

#### 1.4 Root Mean Square Error (RMSE)

RMSE adalah suatu indikator yang dapat digunakan untuk mengukur tingkat akurasi pendugaan suatu model. RMSE juga dapat menyatakan besarnya galat yang dihasilkan suatu model dugaan. RMSE berarti galat dari data yang diperoleh dikuadratkan, dicari rata-ratanya kemudian diakarkan. Dari sekumpulan data  $X_p (p = 1, 2, \dots, n)$  untuk mendapatkan nilai RMSE maka digunakan persamaan:

$$RMSE = \sqrt{\frac{e_1^2 + e_2^2 + e_3^2 + \dots + e_n^2}{n}} \quad (13)$$

Semakin kecil nilai RMSE berarti galat semakin kecil sehingga model yang diperoleh semakin baik.

## 2. METODE PENELITIAN

Langkah-langkah yang dilakukan pada penelitian ini adalah:

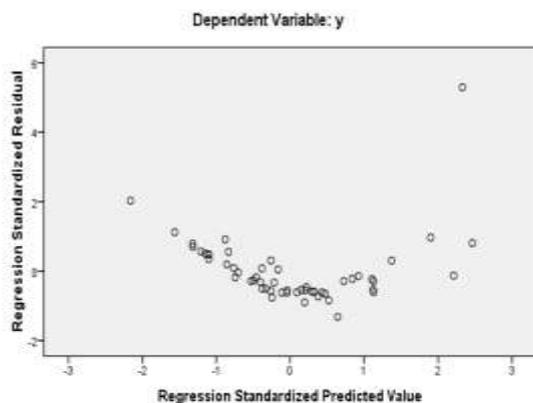
1. Mencari data yang memiliki masalah heteroskedastisitas.
2. Menerapkan Transformasi Box-Cox untuk mengatasi heteroskedastisitas dengan langkah-langkah sebagai berikut:
  - a. Menentukan nilai  $\lambda$  pada rentang  $(-2, 2)$
  - b. Melakukan transformasi terhadap variabel  $Y$  sesuai dengan nilai  $\lambda$  yang didapatkan. Nilai  $\lambda$  yang dipilih adalah nilai  $\lambda$  yang menghasilkan nilai Jumlah Kuadrat Galat (JKG) terkecil.
  - c. Melakukan uji heteroskedastisitas pada model yang diperoleh dari estimasi dengan Transformasi Box-Cox.
  - d. Mencari selang kepercayaan parameter
  - e. Mencari nilai RMSE
3. Menerapkan Regresi Kuantil Median untuk mengatasi heteroskedastisitas dengan langkah-langkah:
  - a. Melakukan estimasi parameter dengan menggunakan Regresi Kuantil Median dimana estimasi dilakukan pada setiap kuantil ke- $\tau$  untuk masing-masing koefisien regresi.

- b. Melakukan uji heteroskedastisitas pada model yang diperoleh dari estimasi dengan Regresi Kuantil Median.
  - c. Mencari selang kepercayaan parameter dengan metode Bootstrap dengan replikasi minimal 100000 kali, kemudian mencari nilai selang kepercayaan.
  - d. Mencari nilai RMSE
4. Membandingkan hasil dari Transformasi Box-Cox dan Regresi Kuantil Median berdasarkan nilai  $R^2$ , RMSE dan selang kepercayaan yang diperoleh.

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Pengujian Heteroskedastisitas Data

Heteroskedastisitas merupakan suatu keadaan apabila varians dari galat tidak konstan. Keberadaan heteroskedastisitas dapat menyebabkan estimasi dengan OLS menjadi tidak efisien. Untuk melihat adanya heteroskedastisitas dapat dilakukan dengan melihat pola antara galat dengan nilai dugaannya. Plot untuk data dapat dilihat pada gambar berikut:



Gambar 1. Plot Varians dari Galat untuk Data

Pada Gambar 1 tampak bahwa galat tidak tersebar secara merata di atas dan di bawah nol (0), yang mengindikasikan adanya masalah heteroskedastisitas pada data. Namun, untuk mendapatkan hasil yang lebih pasti dapat dilakukan pengujian secara formal.

Pengujian formal pada data menggunakan uji White yang dilakukan dengan meregresikan galat kuadrat dengan variabel bebas, variabel

bebas yang dikuadratkan dan perkalian silang antara variabel bebasnya, kemudian nilai  $R^2$  dikalikan dengan jumlah pengamatan untuk mendapatkan nilai  $\chi^2_{hitung}$ . Jika nilai  $\chi^2_{hitung} > \chi^2_{tabel}$  maka keputusannya adalah menolak hipotesis nol. Hipotesis yang digunakan adalah:

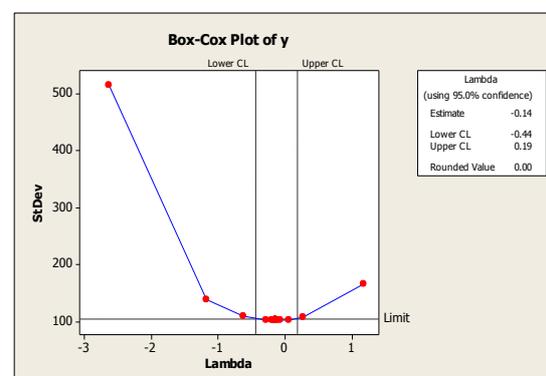
$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2 = 0$$

$$H_1 : \text{Paling tidak terdapat dua varians yang berbeda}$$

Hasil pengujian terhadap adanya heteroskedastisitas pada data diperoleh nilai  $\chi^2_{hitung} = 37,88840$ , dengan nilai  $\chi^2_{tabel} = 23,68$ . Karena  $\chi^2_{hitung} > \chi^2_{tabel}$  maka keputusannya adalah menolak hipotesis nol sehingga disimpulkan terdapat heteroskedastisitas pada data.

#### 3.2 Penerapan Transformasi Box-Cox

Transformasi Box-Cox merupakan transformasi pangkat pada variabel respons yang hanya dapat dilakukan pada variabel respons yang bertanda positif. Prosedur utama yang dilakukan pada transformasi Box-Cox adalah menduga parameter  $\lambda$  dengan nilai  $\lambda$  berada pada interval nilai (-2, 2),  $\lambda$  yang digunakan adalah  $\lambda$  yang menghasilkan nilai JKG terkecil. Untuk melakukan proses transformasi digunakan program Minitab 16. Nilai  $\lambda$  yang digunakan untuk transformasi data dapat dilihat pada Gambar berikut:



Gambar 2. Plot dalam Transformasi Box-Cox

Dari gambar dapat dilihat bahwa nilai  $\lambda$  berada pada interval  $-0.44 < \lambda < 0.19$  Interval nilai  $\lambda$  melalui nilai  $\lambda = 0$  sehingga dilakukan

transformasi dengan nilai  $\lambda = 0$ , berdasarkan Tabel 1 maka model transformasi yang digunakan adalah  $\ln Y$ . Setelah dilakukan proses transformasi pada variabel respons, kemudian dilakukan estimasi dengan OLS pada data tersebut. Hasil estimasi menghasilkan model:

$$Y = E(1.125360 + 0.157788X_1 + 0.021309X_2 + 0.021816X_3)$$

Nilai  $R^2$  yang dihasilkan adalah sebesar 97.2% yang berarti semua variabel bebas yang digunakan dalam model dapat menjelaskan variabel respon sebesar 97.2%. Nilai RMSE yang diperoleh adalah sebesar 0.1038.

Kemudian untuk mengetahui apakah masalah heteroskedastisitas dapat diatasi maka dilakukan pengujian heteroskedastisitas pada hasil estimasi dengan OLS. Uji yang digunakan untuk mendeteksi heteroskedastisitas adalah uji White. Hasil pengujian terhadap adanya heteroskedastisitas dengan uji White memberikan nilai  $\chi^2_{hitung}=16.8111$  sedangkan nilai  $\chi^2_{tabel} = 16.919$ . Karena nilai  $\chi^2_{hitung} < \chi^2_{tabel}$  maka keputusannya adalah menerima hipotesis nol yang berarti tidak terdapat heteroskedastisitas. Hal ini menunjukkan bahwa transformasi Box-Cox dapat mengatasi masalah heteroskedastisitas.

### 3.3 Estimasi dengan Regresi Kuantil

Regresi kuantil median merupakan metode yang mendefinisikan median sebagai solusi untuk meminimumkan jumlah galat mutlak. Metode ini bersifat fleksibel dalam memodelkan data dengan masalah heteroskedastisitas dan dapat langsung digunakan untuk menduga parameter pada data dengan masalah heteroskedastisitas. Hasil estimasi dengan regresi kuantil median menghasilkan model:

$$Y = -481.4768 + 29.31897X_1 + 3.035971X_2 + 3.357251X_3$$

Nilai  $R^2$  yang dihasilkan sebesar 66% yang berarti semua variabel bebas yang digunakan dalam model dapat menjelaskan variabel respon sebesar 66%, dan nilai RMSE yang diperoleh adalah sebesar 63.47917.

Model yang dihasilkan dari estimasi dengan regresi kuantil median kemudian diuji dengan uji White untuk melihat apakah model mengandung heteroskedastisitas atau tidak. Hasil pengujian memberikan nilai  $\chi^2_{hitung}=30.92618$  sedangkan nilai  $\chi^2_{tabel}= 16.919$ . Karena  $\chi^2_{hitung} > \chi^2_{tabel}$  maka keputusannya adalah menolak hipotesis nol, sehingga disimpulkan terdapat heteroskedastisitas pada model yang diperoleh. Hal ini menunjukkan estimasi dengan regresi kuantil median masih mengandung masalah heteroskedastisitas.

Walaupun tidak dapat menghilangkan heteroskedastisitas tetapi regresi kuantil median tetap dapat digunakan untuk estimasi pada data dengan masalah heteroskedastisitas dan model yang dihasilkan juga efisien karena regresi kuantil median tidak memerlukan asumsi homoskedastisitas dan merupakan metode yang fleksibel terhadap kasus heteroskedastisitas.

### 3.4 Perbandingan Transformasi Box-Cox dan Regresi Kuantil Median

Penelitian ini menggunakan tiga indikator untuk mengetahui metode yang lebih baik dalam menangani masalah heteroskedastisitas yaitu nilai  $R^2$ , nilai RMSE dan selang kepercayaan. Metode yang lebih baik adalah metode yang menghasilkan nilai  $R^2$  lebih besar, nilai RMSE yang lebih kecil dan lebar selang kepercayaan yang lebih sempit. Perbandingan untuk kedua metode dapat dilihat pada Tabel berikut:

Tabel 2. Perbandingan Nilai  $R^2$  dan RMSE

$R^2$		RMSE	
OLS	Regresi Kuantil	OLS	Regresi Kuantil
97.2%	66%	0.1038	63.4791

Berdasarkan Tabel 2 terlihat bahwa penerapan transformasi Box-Cox pada OLS menghasilkan nilai  $R^2$  sebesar 97.2% dan nilai RMSE sebesar 0,103828, sedangkan estimasi dengan regresi kuantil median menghasilkan nilai  $R^2$  sebesar 66% dan nilai RMSE sebesar 63.4791. Untuk lebar selang kepercayaan dapat dilihat pada tabel berikut:

Tabel 3. Perbandingan Selang Kepercayaan

Lebar Selang Kepercayaan		
Parameter	OLS	Regresi Kuantil
$\beta_1$	0.057898	30.67955
$\beta_2$	0.00465	2.3463
$\beta_3$	0.00422	2.027028
$\beta_4$	0.103344	40.14601

Karena penerapan transformasi Box-Cox pada OLS menghasilkan nilai  $R^2$  yang lebih besar, nilai RMSE yang lebih kecil dan lebar selang kepercayaan yang lebih sempit, sehingga disimpulkan bahwa transformasi Box-Cox lebih baik digunakan untuk mengatasi heteroskedastisitas pada data.

#### 4. KESIMPULAN DAN SARAN

Dari hasil yang diperoleh diketahui bahwa penerapan transformasi Box-Cox pada OLS menghasilkan nilai  $R^2$  lebih besar, nilai RMSE lebih kecil dan lebar selang kepercayaan lebih sempit dibandingkan dengan yang dihasilkan oleh estimasi dengan regresi kuantil median. Oleh karena itu, dapat disimpulkan bahwa kemampuan transformasi Box-Cox lebih baik untuk mengatasi heteroskedastisitas.

Untuk lebih memastikan kemampuan dari metode transformasi Box-Cox dan regresi kuantil median disarankan untuk menggunakan data yang dihasilkan dari membangkitkan data dengan masalah heteroskedastisitas.

#### DAFTAR PUSTAKA

- [1] Draper, N. R. & Smith, H., 1998. *Applied regression Analysis*. Third ed. New York: John Wiley.
- [2] Gujarati, D. N., 2004. *Basic Econometrics*. 4th ed. New York: The McGrawHill Companies.
- [3] Ispriyanti, D., 2004. Pemodelan Statistika dengan Transformasi Box - Cox. *Jurnal Matematika dan Komputer*, Volume 7, pp. 8-17.
- [4] Koenker, R.&Basset, J. 1978. Regression Quantiles. *Econometrica*, 46(1), pp. 35-50.
- [5] Koenker, R & Hallock, K.F, 2001. Quantile Regression. *Journal of Economic Perspectives*, 15(4), pp. 143 - 156.
- [6] Kristina, S., 2011. Analisis Transformasi Box-Cox untuk Mengatasi Heteroskedastisitas dalam Model Regresi Linier Sederhana. *Skripsi*, Medan: Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Sumatra Utara.
- [7] Neter, J., Wasserman, W. & Kutner, M. H., 1997. *Model Linier Terapan Buku I dan II: Analisis Regresi Linier Sederhana dan Analisis Regresi Ganda*, diterjemahkan oleh Bambang Sumantri. Bogor: Jurusan Statistika FMIPA IPB.
- [8] Uthami, I. P., Sukarsa, I. dan Nila Kencana, I. E., 2013. Regresi Kuantil Median untuk Mengatasi Heteroskedastisitas pada Analisis Regresi. *E-Jurnal Matematika*, 2(1), pp. 6-16