

Penerapan Deteksi *Outlier* Regresi Cox Vibrasi *Bearing*

Ervietta Syiffa Amanda*, Sutawanir Darwis

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Islam Bandung, Indonesia.

*ervietta.amanda12@gmail.com

Abstract. The health of the machine is monitored prognostically as a priority to avoid unexpected downtime of operation or maintenance. The prognostic tics of health in the machine are defined as the ability to observe the future condition of a part or component of the machine and the time remaining before the failure of the machine component. This relates to predictions of the health condition of the machine especially on *bearings*. One of the analyses used for prognostic on machine health is *survival* analysis. In *survival* analysis, an *outlier* can affect the estimated parameters of the model and can change the selected model. The common approach used in detecting *outliers* is through residual analysis. There are several theories that have been discussed about *survival* analysis including Kaplan-Meier and Cox. This paper discusses the application of PH Cox regression to detect *outliers* in *bearing* vibration data. The study used vibration data from Prognostics and Health Management, FEMTO ST-Institute. Vibration data is converted into the form of kurtosis, then converted into *time to failure* by giving the limit of failure to the kurtosis using the threshold value. In this thesis, the Kaplan-Meier method is applied to predict *bearing* resistance and *bearing* failure degradation resulting from *bearing* vibration experiments. Cox Proportional Hazard regression is applied to test the effect of RPM (X1) speed and Load load (X2) on *time to failure* (Y). The dataset consists of three eskperimen conditions (1800 rpm, 4000 N), (1650 rpm, 4200 N), (1500 rpm, 5000 N) and obtained 24 times to failure, 12 of which were censored 12 others did not. Cox's regression obtained is $h(t) = \exp(-0.005212RPM+0.000786Load)h_0(t)$. Residual analysis concluded there were no pencils.

Keywords: *Survival Analysis, Kaplan-Meier, Outlier, Cox Regression, Threshold, Vibration Bearing.*

Abstrak. Kesehatan mesin di pantau secara prognostik sebagai prioritas untuk menghindari *downtime* operasi atau pemeliharaan yang tak terduga. Prognostik kesehatan pada mesin didefinisikan sebagai kemampuan untuk mengamati kondisi dimasa depan suatu bagian atau komponen mesin tersebut dan waktu yang tersisa sebelum terjadinya kegagalan pada komponen mesin. Ini berkaitan dengan prediksi kondisi kesehatan mesin khususnya pada *bearing*. Salah satu analisis yang digunakan untuk prognostik pada kesehatan mesin adalah analisis *survival*. Dalam analisis *survival*, *outlier* dapat mempengaruhi estimasi parameter model dan dapat mengubah model yang dipilih. Pendekatan umum yang digunakan dalam mendeteksi *outlier* adalah melalui analisis residual. Ada beberapa teori yang pernah membahas tentang *survival* analysis diantaranya adalah Kaplan-Meier dan Cox. Makalah ini membahas penerapan regresi PH Cox untuk mendeteksi *outlier* pada data vibrasi *bearing*. Penelitian ini menggunakan data vibrasi dari *Prognostics and Health Management, FEMTO*

ST-Institute. Data vibrasi diubah kedalam bentuk kurtosis, kemudian diubah kedalam *time to failure* dengan memberikan batas failure pada kurtosis menggunakan nilai threshold. Dalam skripsi ini, metode Kaplan-Meier diterapkan untuk memprediksi ketahanan *bearing* dan *failure degradation bearing* hasil eksperimen vibrasi *bearing*. Regresi Cox Proportional Hazard diterapkan untuk menguji efek kecepatan RPM (X1) dan beban Load (X2) terhadap *time to failure* (Y). Dataset terdiri dari tiga kondisi eksperimen yaitu (1800 rpm, 4000 N), (1650 rpm, 4200 N), (1500 rpm, 5000 N) dan diperoleh 24 *time to failure*, 12 diantaranya tersensor 12 lainnya tidak. Regresi Cox yang diperoleh adalah $h(t) = \exp(-0,005212RPM + 0,000786Load)h_0(t)$. Analisis residual menyimpulkan tidak terdapat pencilan.

Kata kunci: Analisis *Survival*, Kaplan-Meier, *Outlier*, Regresi Cox, Threshold, Vibrasi *Bearing*.

1. Pendahuluan

Bearing adalah suatu elemen mesin yang menumpu poros berbeban, sehingga putaran atau gerakan bolak-baliknya dapat berlangsung secara halus, aman dan reliabel (Negara, 2018). Untuk memantau kesehatan mesin dipasang sensor pada rumah *bearing* untuk mengukur vibrasi horizontal maupun vertikal. Naiknya vibrasi pada *bearing* menjadi indikator kondisi mesin.

Analisis regresi yang seringkali digunakan dalam menganalisis data *survival* salah satunya adalah Regresi Proportional Hazard. Pada tahun 1972, Sir David Cox memperkenalkan sebuah metode yaitu regresi Cox Proportional Hazard yang merupakan bentuk pengembangan dari Regresi Proportional Hazard. Secara umum, konsep dasar dari Regresi Cox Proportional Hazard sama dengan Regresi Cox, yaitu suatu analisis yang dihadapkan pada situasi dimana kemungkinan kegagalan individu pada suatu waktu yang dipengaruhi oleh satu atau lebih variabel bebas (Collet, 1994).

Dalam analisis *survival*, *outlier* juga dapat mempengaruhi estimasi parameter model dan dapat mengubah model yang dipilih. Pendekatan umum yang digunakan dalam mendeteksi *outlier* adalah melalui analisis residual. Collet (2003) menyebut *outlier* dalam *survival* sebagai objek yang memiliki waktu bertahan sangat lama, *outlier* disaring menggunakan nilai residual. Residu yang dikembangkan dengan mengubah probabilitas kelangsungan hidup menggunakan logit dan probit transformasi. Residu kemudian digunakan untuk mendeteksi *outlier* di Cox Proportional Hazard.

Dalam penelitian mengenai *time to failure*, data yang diperoleh merupakan selang waktu tersensor. Karena itu, diperlukan suatu metode yang mengakomodasi data tersensor. Skripsi ini membahas prognostik kesehatan *bearing* untuk memprediksi kondisi *bearing* di masa depan dan memprediksi sisa umur pakai menggunakan metode analisis *survival*. Data yang digunakan dalam skripsi ini adalah data eksperimen vibrasi *bearing* diperoleh dari FEMTO-ST Institute. Respon Y adalah *time to failure bearing* dengan status (sensor atau tidak) dengan dua faktor eksperimen yaitu kecepatan RPM atau Revolutions Per-Minute (X1) dan beban Load (X2).

Berdasarkan latar belakang yang telah diuraikan, maka perumusan masalah dalam penelitian ini sebagai berikut: “Bagaimana menentukan model regresi Cox proportional hazard Y pada X1 dan X2 serta signifikansi kondisi eksperimen terhadap respon *time to failure bearing*? dan Bagaimana identifikasi *outlier* pada regresi Cox proportional hazard Y pada X1 dan X2?”. Selanjutnya, tujuan dalam penelitian ini diuraikan dalam pokok-pokok sbb.

1. Menentukan model regresi Cox proportional hazard Y pada X1 dan X2 serta signifikansi kondisi eksperimen terhadap respon *time to failure bearing*.
2. Mengidentifikasi *outlier* pada regresi Cox proportional hazard Y pada X1 dan X2.

2. Landasan Teori

Analisis *Survival*

Analisis *survival* digunakan berkaitan dengan lama waktu hingga terjadi peristiwa tertentu dan kemungkinan adanya data tersensor merupakan karakteristik khas yang membedakan dengan analisis lain. Misalnya peristiwa timbulnya suatu penyakit, kambuhnya penyakit, kesembuhan dan kematian (Kleinbaum dan Klein, 2012).

Kurtosis

Sudjana (2005), kurtosis adalah tingkat keruncingan distribusi relatif terhadap distribusi normal. Koefisien kurtosis dilambangkan dengan K . Bertitik tolak dari kurva model normal, runcing datarnya bentuk dapat ditentukan.

Sensor

Data dikatakan tersensor apabila data tidak dapat diamati secara lengkap karena subjek penelitian hilang atau sampai akhir penelitian belum mengalami kejadian tertentu dengan simbol 0, sedangkan data yang dapat diamati secara lengkap sampai penelitian berakhir disebut data tidak tersensor dengan simbol 1.

Threshold (batas ambang) dapat ditentukan secara subjektif untuk menentukan *time to failure* pada *bearing* dilihat dari plot kurtosis terhadap nilai *measurement point*-nya. Wang & Zhang (2008) menyarankan seperangkat aturan keputusan untuk mendeteksi kerusakan pertama atau *defect first observable*:

1. Terdapat satu titik di luar garis *action limit*.
2. Terdapat dua titik yang berturut-turut di luar garis *warning limit*.

Action limit dan *warning limit* diperoleh dengan rumus:

$$\text{Action Limit} = \mu + 3\sigma$$

$$\text{Warning Limit} = \mu + 2\sigma$$

...(1)

Estimasi Kaplan-Meier

Metode Kaplan-Meier digunakan untuk menaksir fungsi *survival* dan fungsi hazard. Metode ini disebut juga metode nonparametrik karena tidak membutuhkan asumsi distribusi dari waktu *survival*. Kaplan-Meier adalah komputasi statistika untuk menghitung peluang *survival* $S(t)$, didasarkan pada waktu kelangsungan hidup individu dan mengasumsikan bahwa data sensor independen berdasarkan waktu kelangsungan hidup.

$$\hat{S}(t) = \prod_{j=1}^k \frac{n_j - d_j}{n_j} \quad \dots(2)$$

Model Regresi Cox Proportional Hazard

Regresi Cox Proportional Hazard ialah pemodelan yang digunakan dalam analisis *survival* yang merupakan model semi parametrik. Regresi PH Cox digunakan bila outcome yang diobservasi adalah panjang waktu suatu kejadian. Pada mulanya pemodelan ini digunakan pada cabang statistika biostatistika digunakan untuk menganalisis harapan hidup seseorang. Namun seiring perkembangan zaman pemodelan ini banyak dimanfaatkan di berbagai bidang, diantaranya akademik, kedokteran, sosial, sains, rekayasa mesin, pertanian (Sari, 2011). Model umum regresi Cox:

$$h(t) = h_0(t) \exp(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_p) \quad \dots(3)$$

Pengujian Simultan

Uji simultan ini dilakukan untuk mengetahui signifikan dari model secara serentak atau overall.

Hipotesis : $H_0: \beta_1 = \dots = \beta_p = 0$, vs H_1 : minimal satu $\beta_i \neq 0$, $i = 1, 2, \dots, p$

Kriteria uji : Tolak H_0 p-value $< \alpha$

Deteksi *Outlier*

Outlier dalam kasus analisis *survival* didefinisikan sedikit berbeda dari *outlier* dalam masalah regresi linier. Hal ini karena variabel dependen dalam model analisis *survival* berisi waktu *survival* dan status (Akmal, 2008).

Bearing

Bearing adalah suatu komponen yang berfungsi untuk mengurangi gesekan pada mesin dan saling menekan antara satu dengan yang lainnya. Bila gerakan dua permukaan yang saling berhubungan terhambat, maka akan menimbulkan panas. Hambatan ini dikenal sebagai gesekan

(friction). Gesekan yang terus menerus menyebabkan panas yang makin lama semakin meningkat dan menyebabkan kerusakan pada komponen tersebut.

3. Hasil Penelitian dan Pembahasan

Data *Time to failure*

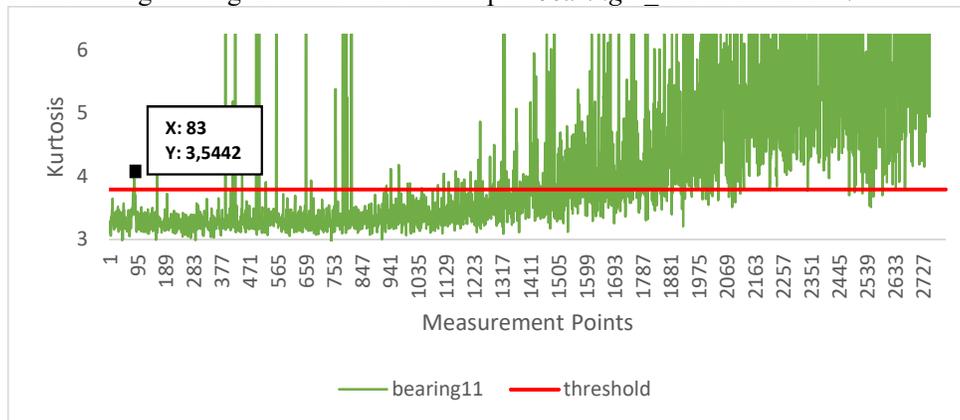
Data yang digunakan dalam skripsi ini adalah data *time to failure* vibrasi setiap *bearing* yang dirubah menjadi kurtosis. Data yang dikumpulkan terdiri dari 3 dataset, dimana dari data tersebut di threshold menggunakan rumus (1) sehingga data yang melebihi angka yang dihasilkan threshold tetapi masih dalam waktu pengamatan dianggap gagal atau failure sedangkan data yang berada dibawah angka yang dihasilkan dan melewati waktu pengamatan dianggap normal. Threshold untuk setiap *bearing* disajikan pada tabel 1.

Tabel 1 Hasil Perhitungan nilai Threshold

No	Bearing	Threshold	No	Bearing	Threshold
1	1_1 H	3,5442	13	1_6 V	2,817
2	1_2 H	7,179	14	1_7 V	2,8229
3	1_3 H	4,0384	15	2_1 H	3,6227
4	1_4 H	3,6261	16	2_2 H	2,7539
5	1_5 H	3,7701	17	2_3 H	5,9439
6	1_6 H	4,9143	18	2_1 V	10,7814
7	1_7 H	3,7476	19	2_2 V	3,4188
8	1_1 V	3,3243	20	2_3 V	3,4352
9	1_2 V	3,3695	21	3_1 H	3,6349
10	1_3 V	3,3412	22	3_2 H	3,8079
11	1_4 V	2,8238	23	3_1 V	3,3043
12	1_5 V	2,8024	24	3_2 V	3,7796

*Keterangan: H: Horizontal, V: Vertikal

Setelah nilai *threshold*, kurtosis tersebut diplot per-*bearing* dengan sumbu X yaitu nilai *t* (*Measurement Points*) dan Y kurtosis, kemudian di *threshold* dengan nilai yang sudah diperoleh masing-masing. Berikut contoh hasil plot *bearing 1_1* arah horizontal:



Gambar 1. Plot antara kurtosis dan *t* dengan *threshold* untuk *bearing 1_1* H

Dari Gambar 1 diatas terlihat untuk nilai X adalah 83 dan Y adalah 3,9168. Nilai X tersebut merupakan nilai *time to failure* dan Y merupakan *threshold*. Karena untuk setiap sinyal getaran *bearing* terjadi setiap 10 detik maka dari itu nilai X dikalikan 10, jadi *bearing 1_1* H mengalami waktu kegagalan pada saat waktu ke 830 detik.

Setelah melakukan plot antara kurtosis dengan *measurement points*, diperoleh 24 data kurtosis dan data *time to failure* dari 3 dataset yang diambil berdasarkan data kurtosis pertama yang melebihi nilai *threshold*. Data *time to failure* tersebut akan digunakan dalam analisis data selanjutnya. Berikut data *time to failure* yang akan disajikan pada Tabel 2:

Tabel 2 Data Time to failure

No	Bearing	Kurtosis	t	No	Bearing	Kurtosis	T
1	1_1 H	3,9168	830	13	1_6 V	2,8656 ⁺	10 ⁺
2	1_2 H	30,2556	380	14	1_7 V	3,1025 ⁺	10 ⁺
3	1_3 H	4,0805 ⁺	360 ⁺	15	2_1 H	3,6266	170
4	1_4 H	4,5818 ⁺	1570 ⁺	16	2_2 H	3,1582	10
5	1_5 H	4,5087 ⁺	470 ⁺	17	2_3 H	6,2055 ⁺	80 ⁺
6	1_6 H	6,8618 ⁺	200 ⁺	18	2_1 V	68,0206	1190
7	1_7 H	4,4914 ⁺	320 ⁺	19	2_2 V	3,5227	1090
8	1_1 V	3,3896	1910	20	2_3 V	3,8837 ⁺	330 ⁺
9	1_2 V	3,9368	180	21	3_1 H	4,1393	10
10	1_3 V	3,4235 ⁺	230 ⁺	22	3_2 H	4,1943	60
11	1_4 V	3,1372 ⁺	10 ⁺	23	3_1 V	3,4906	50
12	1_5 V	2,9286 ⁺	10 ⁺	24	3_2 V	4,0318	790

*Keterangan: tanda (+) menunjukkan *bearing* tersensor

Estimasi Kaplan-Meier

Berdasarkan persamaan (2), dimana t_j menyatakan *time to failure bearing* dalam detik, d_j adalah banyaknya kerusakan pada saat t_j ($j=1,2,\dots,k$) dan n_j adalah banyaknya *bearing* yang masih dapat digunakan sesaat sebelum t_j (jumlah *bearing* berisiko pada saat t_j) termasuk yang rusak atau gagal pada saat t_j , berikut hasil perhitungan estimasi Kaplan-Meier menggunakan Excel disajikan pada Tabel 3:

Tabel 3 Hasil Perhitungan $\hat{S}(t)$

j	t_j	n_j	d_j	$(n_j - d_j)/n_j$	$\hat{S}(t)$	j	t_j	n_j	d_j	$(n_j - d_j)/n_j$	$\hat{S}(t)$
0	0	24	0	1	1	10	330	10	1	0.9	0.8182
1	10	24	6	0.75	0.75	11	360	9	1	0.8889	0.8
2	50	18	1	0.9444	0.7083	12	380	8	1	0.875	0.7778
3	60	17	1	0.9412	0.8889	13	470	7	1	0.8571	0.75
4	80	16	1	0.9375	0.8824	14	790	6	1	0.8333	0.7143
5	170	15	1	0.9333	0.8750	15	830	5	1	0.8	0.6667
6	180	14	1	0.9286	0.8667	16	1090	4	1	0.75	0.6
7	200	13	1	0.9231	0.8571	17	1190	3	1	0.6667	0.5
8	230	12	1	0.9167	0.8462	18	1570	2	1	0.5	0.3333
9	320	11	1	0.9091	0.8333	19	1910	1	1	0	0

Dari Tabel 3 dapat diketahui bahwa:

- $t_0 = 0, \hat{S}(0) = 1$: Setelah 0 detik, ada sebanyak 24 *bearing* yang berisiko pada saat 0 detik dan ada sebanyak 0 *bearing* yang rusak, dan ketahanan *bearing* setelah 0 detik adalah 1 ini artinya bahwa *bearing* dapat bertahan setelah 0 detik dengan prediksi peluang sebesar 1.
- $t_1 = 10, \hat{S}(1) = 0,75$: Setelah 10 detik, ada sebanyak 24 *bearing* yang berisiko pada saat 10 detik dan ada sebanyak 6 *bearing* yang rusak, dan ketahanan *bearing* setelah 10 detik adalah 0,75 ini artinya bahwa *bearing* dapat bertahan setelah 10 detik dengan prediksi peluang sebesar 0,75.

Regresi Cox Proportional Hazard

Data yang digunakan untuk analisis Cox PH sebagai berikut:

Tabel 4 Data regresi Cox *Proportional Hazard*

No	Bearing	TTF (Y)	RPM (X ₁)	LOAD (X ₂)	Status	No	Bearing	TTF (Y)	RPM (X ₁)	LOAD (X ₂)	Status
1	1_1 H	830	1800	4000	1	13	1_6 V	10	1800	4000	0
2	1_2 H	380	1800	4000	1	14	1_7 V	10	1800	4000	0
3	1_3 H	360	1800	4000	0	15	2_1 H	170	1650	4200	1
4	1_4 H	1570	1800	4000	0	16	2_2 H	10	1650	4200	1
5	1_5 H	470	1800	4000	0	17	2_3 H	80	1650	4200	0
6	1_6 H	200	1800	4000	0	18	2_1 V	1190	1650	4200	1
7	1_7 H	320	1800	4000	0	19	2_2 V	1090	1650	4200	1
8	1_1 V	1910	1800	4000	1	20	2_3 V	330	1650	4200	0
9	1_2 V	180	1800	4000	1	21	3_1 H	10	1500	5000	1
10	1_3 V	230	1800	4000	0	22	3_2 H	60	1500	5000	1
11	1_4 V	10	1800	4000	0	23	3_1 V	50	1500	5000	1
12	1_5 V	10	1800	4000	0	24	3_2 V	790	1500	5000	1

*Ket: 0 = tersensor, 1 = tidak tersensor

Model Regresi Cox *Proportional hazard*

Selanjutnya dilakukan analisis regresi PH Cox untuk melihat apakah ada kovariat yang berpengaruh secara signifikan terhadap *time to failure* vibrasi *bearing*. Pada regresi Cox, kovariat yang memiliki pengaruh signifikan terhadap variabel dependen adalah kovariat yang memiliki nilai signifikan kurang dari 0,05. Analisis ini dilakukan menggunakan software S-Plus, Gambar 2 adalah hasil output untuk regresi Cox:

```

coef exp(coef) se(coef)      z    p
RPM -0.005212    0.995 0.00785 -0.664 0.51
Load 0.000786    1.001 0.00224 0.350 0.73

Likelihood ratio test=8.02 on 2 df, p=0.0181 n= 24

```

Gambar 2 Output Regresi Cox dengan S-Plus

Dari output diatas, diperoleh model regresi PH Cox dengan menggunakan persamaan (2) sebagai berikut:

$$h(t) = \exp(-0,005212x_1 + 0,000786x_2)h_0(t)$$

Untuk koefisien X₁ (RPM) mempunyai nilai $-0,005212$ artinya bahwa semakin besar nilai RPM pada *bearing* maka waktu kegagalannya akan semakin lama. Dengan kata lain, *bearing* yang memiliki nilai RPM lebih besar memiliki peluang waktu kegagalan lebih cepat sebesar 0,005212 dibandingkan dengan *bearing* yang memiliki nilai RPM lebih kecil atau semakin tinggi nilai RPM maka semakin rendah pula peluang waktu kegagalan *bearing* tersebut.

Untuk koefisien X₂ (Load) mempunyai nilai $+0,000786$ artinya bahwa semakin besar nilai Load pada *bearing* maka waktu kegagalannya akan semakin cepat. Dengan kata lain, *bearing* yang memiliki nilai Load lebih besar memiliki peluang waktu kegagalan lebih lama sebesar 0,000786 dibandingkan dengan *bearing* yang memiliki nilai Load lebih kecil atau semakin tinggi nilai Load maka semakin besar pula peluang waktu kegagalan *bearing* tersebut.

Uji Simultan

Hipotesis: $H_0: \beta_1 = \dots = \beta_p = 0$, vs H_1 : minimal satu $\beta_i \neq 0$

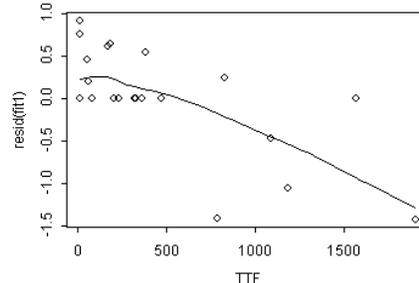
Kriteria uji: Tolak H_0 jika p-value $< 0,05$

Dapat dilihat nilai p-value pada Gambar 4.26 Output Regresi Cox dengan S-Plus adalah

0,0181, karena $0,0181 < 0,05$ maka H_0 diolak dan H_1 diterima atau dapat dikatakan minimal ada satu $\beta_i \neq 0$. Artinya secara keseluruhan model signifikan.

Deteksi *Outlier*

Dari regresi *time to failure* dicari nilai residual menggunakan aplikasi S-PLUS di plot dan diidentifikasi apakah terdapat pencilan. Berikut hasil plot yang diperoleh:



Gambar 3 Plot *Outlier*

Dari gambar 3 terlihat bahwa nilai residu berada di sekitar garis normal dan dapat dikatakan bahwa data *time to failure* tidak memiliki pencilan.

4. Kesimpulan

Berdasarkan pembahasan dalam penelitian ini, peneliti menyimpulkan beberapa hasil penelitian sebagai berikut:

1. Model regresi Cox adalah: $h(t) = \exp(-0,005212x_1 + 0,000786x_2)h_0(t)$ dan dilakukan uji simultan dan uji parsial diperoleh hasil bahwa model tersebut tidak signifikan dan kovariat X_1 (RPM) dan X_2 (Load) tidak mempengaruhi Y (*time to failure*).
2. Untuk deteksi pencilan (Gambar 4.26) terlihat bahwa residual berada disekitar garis normal artinya data *time to failure* tidak memiliki pencilan berbasis regresi Cox.

Daftar Pustaka

- [1] Akmal, N. M., Ibrahim, M., 2008. A Detection Measure of Outliers Based on Forward Search Approach for Cox-Regression Model. *Statistika, Forum Teori dan Aplikasi Statistika*, 8 (2): 103–109.
- [2] Collet, D., (1994). *Modelling Survival Data in Medical Research*. London: Chapman & Hall.
- [3] Cox, D. R. 1972. *Regression Models and Life-Tables*. *Journal of the Royal Society. Series B (Methodological)*, 34 (2):187-220..
- [4] Kleinbaum, D.G. and Klein. 2012. *Survival Analysis – A Self-Learning Text*. 3rd Edition. New York: Springer.
- [5] Negara, G. A. 2018. *Deteksi Cacat Multi Jenis pada Bantalan Tipe Double Row Menggunakan Sinyal Vibrasi*. Yogyakarta: Program Studi Teknik Mesin, Fakultas Teknik, Universitas Muhammadiyah Yogyakarta
- [6] Sari, D.R. 2011. *Analisis Survival Untuk Data Tersensor Tipe II Menggunakan Model Distribusi Log-Logistik*. Skripsi:1-13. Yogyakarta: Universitas Negeri Yogyakarta.
- [7] Sudjana, M.A. 2005. *Metode Statistika*. Bandung: PT. Tarsito Bandung.
- [8] Wenbin, W. & Zhang, W. 2008. Early defect: Application of Statistical Process Control Methods. 14(3). 229.