

Uji Asumsi *Proportional Hazard* pada Faktor yang Mempengaruhi Waktu Tahan Hidup Pasien Kanker Paru

Elnatan Dimas A.¹, Sri Sulistijowati Handajani², dan Ririn Setiyowati³

^{1,3}Program Studi Matematika FMIPA Universitas Sebelas Maret

²Program Studi Statistika FMIPA Universitas Sebelas Maret

rr_ssh@staff.uns.ac.id, elnatandim@gmail.com

Abstract. Lung cancer is the disease that its death risk always increase, because of that the survival time of its patient is interesting to be researched. One of the method that can be used to research survival time of lung cancer patient is Cox regression. It has an assumption that called proportional hazard assumption. Proportional hazard assumption can be tested by graph method that is log-log graph, but the result is only used as temporary suspicion. In this study, the goodness of fit test is used to test the assumption by calculate the correlation between rank of survival time and Schoenfeld residual. As the result, age variable does not satisfy proportional hazard assumption.

Keyword: cox regression, proportional hazard assumption, log-log graph, goodness of fit test.

1. Pendahuluan

Regresi Cox merupakan salah satu bentuk regresi non parametrik dimana hubungan antara variabel respon dan variabel prediktornya tidak berbentuk garis lurus. Regresi Cox digunakan sebagai metode untuk menganalisis keterkaitan variabel respon yang merupakan waktu tahan hidup dan variabel prediktornya adalah faktor-faktor yang mempengaruhinya. Regresi Cox mempunyai asumsi yang harus dipenuhi yaitu rasio dari dua nilai *hazard* harus konstan terhadap waktu (Dewiyanti [1]).

Salah satu penelitian pernah dilakukan terkait pengujian asumsi *PH*, yaitu Hadiyah [2] dalam penelitiannya menggunakan plot $\ln(-\ln \hat{S}(t))$ dengan waktu tahan hidup pada setiap kovariat untuk menguji asumsi *PH*. Asumsi *PH* dipenuhi apabila kedua plot tidak saling berimpit atau berpotongan. Jika kedua berpotongan, maka asumsi *PH* tidak dipenuhi. Kelemahan metode ini adalah tidak mempertimbangkan kestabilan *hazard ratio* sehingga diperlukan metode lain untuk menguji asumsi *PH*.

Afifah [3] dalam penelitiannya menyebutkan bahwa asumsi *PH* dapat diuji dengan uji *goodness of fit* dan pendekatan variabel yang bergantung waktu menggunakan *extended Cox model*. Metode *goodness of fit* dilakukan dengan cara mencari korelasi antara residu Schoenfeld dan *rank* waktu tahan hidup. Dalam penelitiannya digunakan koefisien korelasi *rank* Pearson. *Extended Cox model* mengalikan semua kovariat dengan variabel

waktu sehingga *hazard ratio* yang tidak memuat variabel waktu t menjadi memuat variabel waktu.

Dalam penelitian ini, digunakan grafik *log-log* sebagai dugaan sementara. Selanjutnya, untuk menguji dugaan tersebut digunakan uji *goodness of fit*. Korelasi Pearson dan korelasi rank Kendall digunakan dalam uji *goodness of fit* sehingga dapat menjadi pertimbangan suatu kovariat memenuhi asumsi *PH*. Contoh kasus kanker paru digunakan karena merupakan penyakit yang beresiko besar menyebabkan kematian dan dipengaruhi banyak hal sehingga waktu tahan hidup pasien kanker paru menarik untuk diteliti. Penelitian ini bertujuan menguji asumsi *PH* faktor-faktor yang mempengaruhi waktu tahan hidup pasien kanker paru.

2. Landasan Teori

Landasan teori berisikan tentang teori yang mendasari yaitu regresi Cox, asumsi *proportional hazard*, grafik *log-log*, residu Schoenfeld, koefisien korelasi rank Kendall.

2.1. Regresi Cox. Menurut Ata dan Zoter [4], pendekatan paling sering digunakan dalam memodelkan pengaruh kovariat terhadap waktu tahan hidup adalah regresi Cox. Misalkan x_1, x_2, \dots, x_p adalah nilai dari p kovariat X_1, X_2, \dots, X_p . Berdasarkan model regresi Cox, fungsi *hazard* ditulis sebagai

$$h(t, x) = h_0(t) \exp(\sum_{i=1}^p \beta_i x_i) \quad (2.1)$$

dengan $\beta = (\beta_1, \beta_2, \dots, \beta_p)$ adalah vektor $1 \times p$ dari parameter regresi dan $h_0(t)$ adalah fungsi *baseline hazard*.

2.2. Asumsi Proportional Hazard. Pada regresi Cox dikenal *hazard ratio* yang merupakan perbandingan *hazard* dari dua individu yang mempunyai nilai kovariat berbeda. Menurut Ata dan Zoter [4], misalkan $x^* = (x_1^*, x_2^*, \dots, x_p^*)$ dan $x = (x_1, x_2, \dots, x_p)$ adalah kovariat dari dua kategori, *hazard ratio* (*HR*) dari dua kategori tersebut adalah sebagai berikut

$$HR = \exp[\sum_{i=1}^p \beta_i (x_i^* - x_i)].$$

Regresi Cox baik digunakan jika memenuhi asumsi *PH* yang menyatakan bahwa *hazard ratio* tidak berubah sepanjang waktu.

2.3. Grafik log-log. Menurut Afifah [3], asumsi *PH* dapat diuji dengan menggunakan grafik plot $\ln(-\ln S(t))$ terhadap waktu tahan hidup. Asumsi *PH* dipenuhi apabila

garis antar kategori pada kovariat sejajar (tidak berpotongan). Pendekatan ini mudah digunakan, namun tidak dapat dilihat kestabilan *hazard ratio*.

2.4. Residu Schoenfeld. Afifah [3] menyatakan bahwa residu Schoenfeld dari kovariat ke- k dari individu yang mengalami peristiwa pada waktu $t_{(j)}$ dituliskan sebagai berikut

$$RS_{kj} = x_{kj} - E(x_{kj} | R(t_{(j)})) \quad (2.2)$$

dengan RS_{kj} adalah residu Schoenfeld kovariat ke- k dan waktu ke- j , x_{kj} adalah nilai dari kovariat dari individu yang mengalami peristiwa pada waktu $t_{(j)}$ dan $E(x_{kj} | R(t_{(j)}))$ adalah ekspektasi bersyarat x_{kj} jika diketahui $R(t_{(j)})$. $R(t_{(j)})$ adalah himpunan individu yang mengalami peristiwa pada waktu $t_{(j)}$.

2.5. Koefisien Korelasi Rank Kendall. Salah satu koefisien korelasi *rank* yang dapat digunakan untuk menentukan hubungan antara residu Schoenfeld dan waktu tahan hidup adalah koefisien korelasi *rank* Kendall (Kendall's *Tau*). Menurut Conover [5], perhitungan koefisien korelasi *rank* Kendall adalah sebagai berikut

$$\tau = \frac{N_c - N_d}{n(n-1)/2} \quad (2.3)$$

dengan N_c merupakan banyaknya pasangan yang *concordant* dan N_d banyaknya pasangan yang *discordant*.

2.6. Koefisien Korelasi Pearson. Pengujian korelasi antara residu Schoenfeld dengan rank waktu tahan hidup untuk setiap variabel dapat digunakan korelasi Pearson sebagai berikut

$$r = \frac{\sum_{j=1}^n (RS_{kj} - \overline{RS}_{kj})(RT_j - \overline{RT}_j)}{\sqrt{\sum_{j=1}^n (RS_{kj} - \overline{RS}_{kj})^2} \sqrt{\sum_{j=1}^n (RT_j - \overline{RT}_j)^2}} \quad (2.4)$$

dengan RS_{kj} adalah residu Schoenfeld dan RT_j adalah rank waktu tahan hidup. (Afifah [3]).

3. Hasil dan Pembahasan

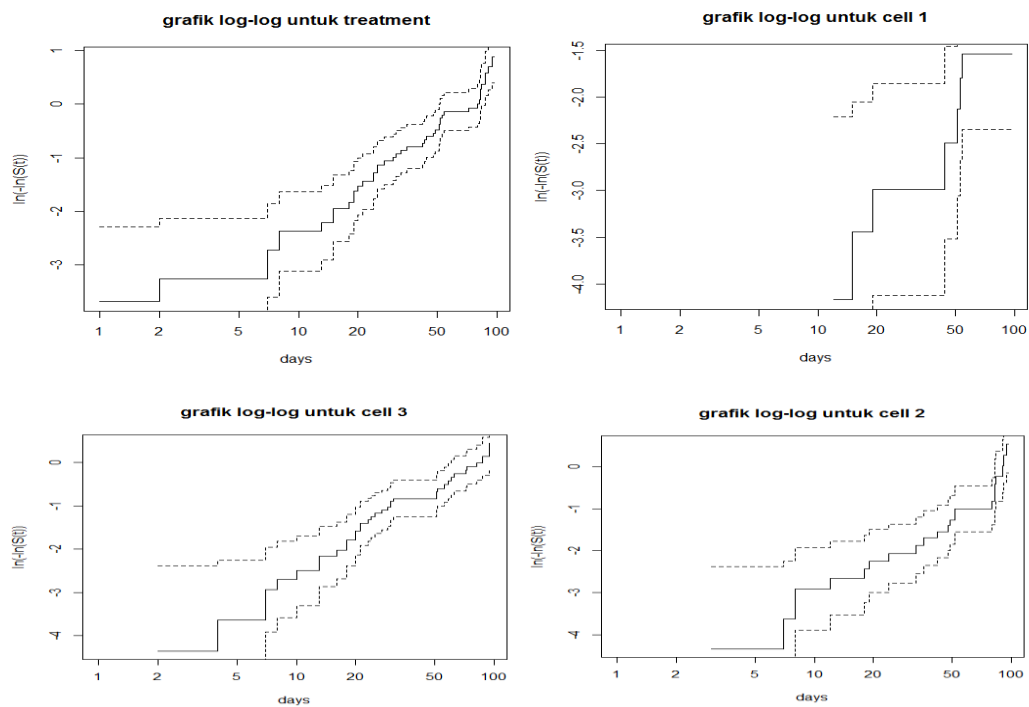
3.1. Data. Penelitian ini menggunakan data tahan hidup 80 pasien kanker paru (Kalbfleisch dan Prentice [6]). Variabel yang diamati pada data ini adalah waktu tahan

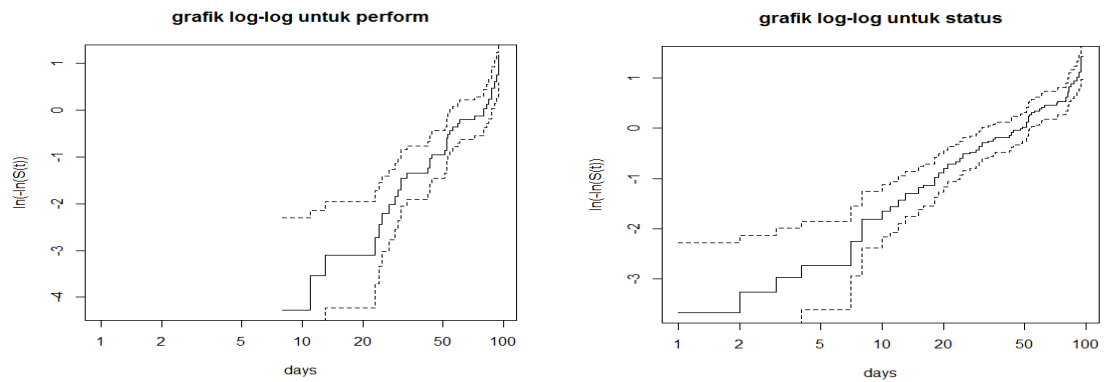
hidup (Y), status perawatan (X_1), tipe sel yaitu tipe 1 *large* (X_2), tipe 2 *adeno* (X_3), tipe 3 *small* (X_4), status performa (X_5) dan usia (X_6).

Tabel 1. Data Waktu Tahan Hidup dan Variabel

| | Y | X_1 | X_2 | X_3 | X_4 | X_5 | X_6 |
|-----|-----|-------|-------|-------|-------|-------|-------|
| 1 | 1 | 2 | 0 | 0 | 1 | 20 | 65 |
| 2 | 1 | 2 | 0 | 1 | 0 | 50 | 35 |
| 3 | 2 | 2 | 0 | 0 | 1 | 40 | 44 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 80 | 100 | 1 | 0 | 1 | 0 | 60 | 67 |

3.2. Grafik Log-log. Untuk menguji asumsi *PH* secara visual digunakan grafik *log-log* dengan hasil sebagai berikut.





Gambar 1. Grafik Log-log dari variabel yang diamati

3.3 Residu Schoenfeld. Untuk mencari residu Schoenfeld dihitung terlebih dahulu koefisien regresi Cox *PH* pada persamaan 2.1 untuk mendapatkan taksiran model. Berikut ini adalah hasil taksiran model.

Tabel 2. Hasil Taksiran Model Regresi Cox

| Parameter | Koefisien |
|-----------|-----------|
| X_1 | -1.10975 |
| X_2 | -0.00823 |
| X_3 | -0.01848 |
| X_4 | -0.23596 |
| X_5 | -0.06485 |
| X_6 | -0.03029 |

Dari taksiran model, didapatkan residu Schoenfeld dengan persamaan 2.2 yang dapat dilihat pada Tabel 3.

Tabel 3. Hasil Residu Schoenfeld

| Y | X_1 | X_2 | X_3 | X_4 | X_5 | X_6 |
|-----|--------|----------|---------|---------|---------|----------|
| 1 | 0.4545 | 7.5721 | -0.1055 | -0.2241 | -0.4676 | -19.7731 |
| 1 | 0.4545 | -22.4278 | -0.1055 | -0.2241 | -0.4675 | 10.2268 |
| ... | | | | | | |
| 100 | 0.7433 | 2.8888 | 0.0000 | -0.2226 | 0.2226 | 2.9795 |

3.4. Korelasi Rank Pearson. Setelah diperoleh hasil residu Schoenfeld pada Tabel 3, dihitung korelasi antara waktu tahan hidup dan residu Schoenfeld. Salah satu korelasi

yang dapat digunakan yaitu rank Pearson. Berikut adalah hasil korelasi rank Pearson setiap variabel dengan menggunakan persamaan 2.4.

Tabel 4. Nilai Korelasi Rank Pearson

| Parameter | Korelasi Pearson |
|-----------|------------------|
| X_1 | 0.15567 |
| X_2 | 0.09562 |
| X_3 | 0.12351 |
| X_4 | 0.07425 |
| X_5 | -0.08514 |
| X_6 | 0.7575 |

3.5. Korelasi Rank Kendall. Korelasi rank lain yang digunakan untuk menguji korelasi rank waktu tahan hidup dan residu Schoenfeld adalah korelasi rank Kendall. Berikut hasil korelasi rank Kendall dengan persamaan 2.3 dari setiap variabel.

Tabel 5. Nilai Korelasi Rank Kendall

| Parameter | Korelasi rank Kendall |
|-----------|-----------------------|
| X_1 | 0.062658 |
| X_2 | 0.08924 |
| X_3 | 0.225 |
| X_4 | -0.18101 |
| X_5 | 0.08418 |
| X_6 | 0.7785 |

Tabel 4 dan Tabel 5 menunjukkan baik korelasi rank Pearson maupun Kendall memperlihatkan hasil yang sama yaitu variabel usia yang memiliki korelasi yang besar. Melalui hasil tersebut asumsi *PH* tidak dipenuhi oleh variabel usia, sedangkan untuk variabel lain yaitu status perawatan, tipe sel 1, tipe sel 2, tipe sel 3, dan status performa memenuhi asumsi *PH*.

4. Kesimpulan

Asumsi *PH* dapat diuji dengan grafik *log-log* saja, namun untuk mendapatkan hasil yang lebih akurat dapat dibandingkan dengan uji *goodness of fit*. Pada penerapan kasus didapatkan hasil variabel yang memenuhi asumsi *PH* melalui grafik *log-log* adalah semua

variabel, namun setelah diuji dengan *goodness of fit* ternyata terdapat variabel yang tidak memenuhi asumsi *PH* yaitu variabel usia.

Daftar Pustaka

- [1] Dewiyanti, L. Perluasan Regresi Cox dengan Penambahan Peubah Terikat Waktu. *E-Jurnal Matematika Vol. 3*, Universitas Udayana, 2014.
- [2] Hadiyah. *Analisis Survival Faktor-faktor yang Mempengaruhi Ketahanan Hidup Penderita Hipertensi dengan Terapi Tablet Captopril*, Institut Sepuluh November, Surabaya, 2010.
- [3] Afifah, N. Uji Proportional Hazard pada Data Penderita Kanker Serviks di RSUD dr. Soetomo Surabaya. *Jurnal Sains dan Seni ITS*. 2016.
- [4] Ata, N., Zoter, M. T. Cox Regression Models with Nonproportional Hazards Applied to Lung Cancer Survival Data. *Hacettepe Journal of Mathematics and Statistics*. 2007.
- [5] Conover, W. J. *Practical Nonparametric Statistic*. John Wiley and Sons. New York. 1999.
- [6] Kalbfleisch, J. D., Prentice, R., L., *The Statistical Analysis Failure Time Data*, John Wiley and Sons, New York, 2011.