

# Analisi Data Tersensor Berpasangan dengan Estimasi Kaplan Meier dan Nelson Aalen

Hikmah\*<sup>1</sup>, Darma Ekawati<sup>2</sup>

<sup>1,2</sup>Universitas Sulawesi Barat

e-mail: \*[hikmah@unsulbar.ac.id](mailto:hikmah@unsulbar.ac.id), [darmaekawati@unsulbar.ac.id](mailto:darmaekawati@unsulbar.ac.id)

## Abstrak

Dalam bidang matematika, terdapat cabang statistika yang telah mengalami perkembangan yang sangat pesat, salah satunya adalah analisis kelangsungan hidup (*survival*). Masalah yang sering ditemui dalam analisis *survival* adalah adanya pengamatan yang tidak lengkap karena terdapat unit penelitian yang sampai akhir pengamatan tidak diketahui apakah mengalami kejadian atau tidak. Tujuan penelitian ini untuk membandingkan analisis data tersensor berpasangan dengan menggunakan estimasi Kaplan-Meier dan Nelson-Aalen. Kedua estimasi tersebut banyak digunakan dalam statistik nonparametrik. Dari hasil yang diperoleh, kedua estimasi tersebut menyatakan bahwa tidak ada perbedaan yang signifikan antara waktu sembuh pasangan suami istri penderita hipertensi.

**Kata Kunci:** Data Berpasangan, Kaplan-Meier, Nelson-Aalen

## 1. PENDAHULUAN

Dalam bidang matematika terdapat cabang statistika yang telah mengalami perkembangan yang sangat pesat dengan ditemukannya berbagai alat analisis yang dapat digunakan untuk menganalisis suatu kejadian/ *event*. Salah satunya adalah analisis kelangsungan hidup, yaitu prosedur statistik yang digunakan untuk menganalisis data yang berhubungan dengan waktu kelangsungan hidup yaitu dari waktu awal yang sudah ditentukan sampai waktu terjadinya suatu kejadian/ *event*. Waktu kelangsungan hidup didefinisikan sebagai variabel random tak negatif, sehingga analisis kelangsungan hidup dapat didefinisikan sebagai prosedur statistik pada variabel random taknegatif yang berfungsi untuk mengetahui ketahanan hidup objek yang diteliti. Analisis *survival* merujuk pada sekumpulan metode statistik untuk menganalisis data antar kejadian, yaitu data lama waktu sampai terjadinya suatu *event*. Istilah *survival* sendiri muncul karena pada mulanya, metode analisis ini digunakan untuk menyelidiki *event* kematian (Singh dan Mukhopadhyay, 2011).

Sesuai dengan pengertian di atas, terdapat banyak metode dalam analisis data *survival*. Pada dasarnya, semua metode tersebut digunakan untuk menganalisis dua hal yang menjadi ciri khas dari analisis data *survival*, yaitu:

- Fungsi *survival*, yaitu probabilitas suatu individu tidak mendapatkan *event* (*survive*) lebih lama dari suatu waktu.
- Fungsi *hazard*, yaitu tingkat (*rate*) terjadinya suatu *event* pada suatu waktu.

Masalah yang sering ditemui dalam analisis data *survival* adalah adanya pengamatan yang tidak lengkap karena terdapat unit penelitian yang sampai akhir pengamatan tidak diketahui apakah mengalami *event* atau tidak. Kasus tersebut sering disebut dengan observasi tersensor. Salah satu bidang kehidupan yang sering menggunakan analisis *survival* adalah uji klinis. Dalam uji klinis, informasi tentang titik waktu tertentu sangat penting untuk diketahui agar para petugas kesehatan lebih mudah dalam mengambil keputusan tentang efektivitas perawatan. Kali ini penulis akan mengestimasi fungsi *survival* dengan menggunakan estimator Kaplan-Meier dan Nelson Aalen.

Su (2011), meneliti kurva *survival* pada titik tertentu data tersensor kanan berpasangan, menggunakan studi kasus kanker payudara. Dalam artikel tersebut, dijelaskan bahwa analisis kurva *survival* tidak dilakukan untuk semua waktu, melainkan akan dipilih titik waktu tertentu saja dengan tujuan fokus

terhadap jenis pengobatan yang dianggap berhasil. Penelitian kedua mengenai aplikasi metode Kaplan Meier untuk menduga selang waktu ketahanan hidup oleh Bintang (2017). Studi kasus yang digunakan juga pasien kanker payudara pada salah satu rumah sakit di Yogyakarta.

Berdasarkan uraian di atas, penulis tertarik melakukan perbandingan analisis data tersensor dengan menggunakan estimasi Kaplan-Meier dan Nelson Aalen data berpasangan, dengan studi kasus pasangan suami istri penderita hipertensi. Sebuah pertimbangan tambahan ketika membandingkan tingkat ketahanan hidup adalah hal yang memungkinkan mempengaruhi waktu bertahan hidup. Hal demikian dapat timbul dikarenakan desain eksperimental. Untuk mengurangi perbedaan antara sampel, sampel yang cocok adalah sampel yang sama atau kembar dan secara acak diberikan perlakuan yang berbeda untuk mengevaluasi efeknya. Data seperti ini disebut data tersensor kanan “berpasangan”, di mana waktu bertahan hidup independen dalam kelompok tetapi dependen antara kelompok (Huster et al. 1989; Dallas dan Rao, 2000).

Menurut Bie et al. (1987) serta Borgon dan Liestøl (1990) untuk membandingkan perbedaan antara dua fungsi survival dengan data independen yang tersensor kanan, satu pendekatan yang dapat dilakukan adalah dengan menentukan interval konfidensi pada fungsi survival. Pendekatan alternatif untuk membandingkan perbedaan antara dua kelompok adalah menguji apakah perlakuan dua kelompok memiliki fungsi survival yang sama atau tidak. Pada data tersensor kanan berpasangan, Murray (2001) menggunakan tes *weighted* Kaplan-Meier untuk membandingkan perbedaan antara dua fungsi survival.

## 2. METODE PENELITIAN

Data penelitian yang digunakan adalah data waktu survive pasien hipertensi pasangan suami istri (dalam satuan hari) pada salah satu puskesmas di Kab. Sinjai, Sulawesi Selatan, Januari 2013 sampai Maret 2015 (Hikmah, 2015). Metode yang digunakan dalam penelitian ini adalah studi literatur, yaitu mempelajari buku-buku dan jurnal-jurnal yang berkaitan dengan model Kaplan-Meier, Nelson Aalen, dan modifikasinya, terutama dalam hal penerapannya dalam analisis data survival. Selanjutnya dari model Kaplan-Meier dan Nelson Aalen tersebut diperoleh formula untuk mengestimasi fungsi survival. Setelah diperoleh formula untuk mengestimasi fungsi survival, dilakukan simulasi atau studi kasus untuk membandingkan kurva survival dua data real pada titik waktu tertentu.

### 2.1 Metode Analisis Data

#### 2.1.1 Kaplan-Meier

Pada data tersensor berpasangan didefinisikan  $T_{ik}$  dan  $C_{ik}$  masing-masing sebagai waktu survival dan tersensor pada subjek ke  $k$  dan kelompok ke  $i$ , untuk  $i = 1, 2$  dan  $k = 1, 2, \dots, n$  dengan  $n$  adalah banyaknya pasangan. Catatan bahwa  $T_{ik}$  independen terhadap  $C_{ik}$ , tetapi antara  $T_{1k}$  dan  $T_{2k}$ , serta  $C_{1k}$  dan  $C_{2k}$  masing-masing dependen. Secara umum fungsi survival pada waktu  $T_{ik}$  dinotasikan dengan  $S_i(t)$ . Ketika menginputkan data survival, dalam hal ini data tersensor kanan berpasangan, maka terbentuk  $X_{ik} = \min(T_{ik}, C_{ik})$  dan  $\delta_{ik} = I(T_{ik} < C_{ik})$ , dengan  $I(A)$  adalah fungsi indikator kejadian  $A$  yang bernilai 1 jika *event* terjadi dan 0 jika tidak. Diberikan  $t_j$  adalah waktu survive untuk data gabungan dua kelompok,  $d_{ij}$  adalah banyaknya kejadian, dan  $y_{ij}$  banyaknya unit penelitian yang belum mengalami kejadian dan masih berisiko pada waktu  $t_j$ . Berikut persamaan estimator Kaplan-Meier:

$$\hat{S}_i(t) = \prod_{t_j \leq t} \left(1 - \frac{d_{ij}}{y_{ij}}\right) \quad (1)$$

dengan variansi,

$$\text{Var}(\hat{S}_i(t)) \approx (\hat{S}_i(t))^2 \sum_{t_j \leq t} \frac{d_{ij}}{y_{ij} - d_{ij}} \quad (2)$$

#### 2.1.2 Nelson Aalen

Fungsi variabel random yang cukup penting adalah fungsi hazard, yang didefinisikan sebagai

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} \quad (3)$$

yang dapat diinterpretasikan sebagai tingkat (*rate*) terjadinya *event*. Estimator yang dapat digunakan untuk mengestimasi fungsi hazard adalah Nelson-Aalen, yang didefinisikan sebagai berikut :

$$\hat{H}(t) = \begin{cases} 0 & \text{jika } t < t_1 \\ \sum_{t_i \leq t} \frac{d_i}{Y_i} & \text{jika } t_i \leq t \end{cases} \quad (4)$$

dengan variansi,

$$\text{Var}(\hat{H}(t)) = \sum_{t_i \leq t} \frac{d_i}{Y_i^2} \quad (5)$$

Dengan  $h(t)$ : fungsi hazard

$\hat{H}(t)$  : fungsi hazard kumulatif

Estimasi Nelson-Aalen dapat digunakan untuk mengestimasi  $S(t)$  dengan menggunakan hubungan  $H(t)$  dengan  $S(t)$ , yaitu  $S(t) = \exp(-H(t))$ , Danardono (2012).

Hasil estimasi Kaplan-Meier dan Nelson Aalen akan ditampilkan dalam bentuk kurva, selanjutnya akan dilakukan uji log rank untuk mendukung kebenaran dari kesimpulan.

#### 2.4 Hipotesis

Uji dua fungsi survival data berpasangan mempunyai spesifikasi hipotesis sebagai berikut:

$H_0: S_1(t) = S_2(t)$ , melawan  $H_1: S_1(t) \neq S_2(t)$  untuk waktu tetap  $t$ , dengan  $S_1(t)$  dan  $S_2(t)$  masing-masing merupakan probabilitas survival pada kelompok pertama dan kedua.

#### 2.5 Uji Statistik

Pada penelitian ini, uji statistik yang digunakan adalah

$$Z^* = \frac{\phi(\hat{S}_2(t)) - \phi(\hat{S}_1(t))}{\sqrt{v(\phi(\hat{S}_2(t)) - \phi(\hat{S}_1(t)))}} \quad (6)$$

dengan  $\phi$  adalah fungsi terdiferensial bernilai real. Karena digunakan pendekatan normal nilai estimasi  $\hat{S}_i(t)$  ditransformasikan menjadi (1)  $\phi(\hat{S}_i(t)) = \hat{S}_2(t)$ ; (2)  $\phi(\hat{S}_i(t)) = \log(\hat{S}_i(t))$ ; (3)  $\phi(\hat{S}_i(t)) = \log(-\log(\hat{S}_i(t)))$ ; (4)  $\phi(\hat{S}_i(t)) = \log\left(\frac{\hat{S}_i(t)}{1 - \hat{S}_i(t)}\right)$ ; dan (5)  $\phi(\hat{S}_i(t)) = \arcsin(\hat{S}_2(t))$ .

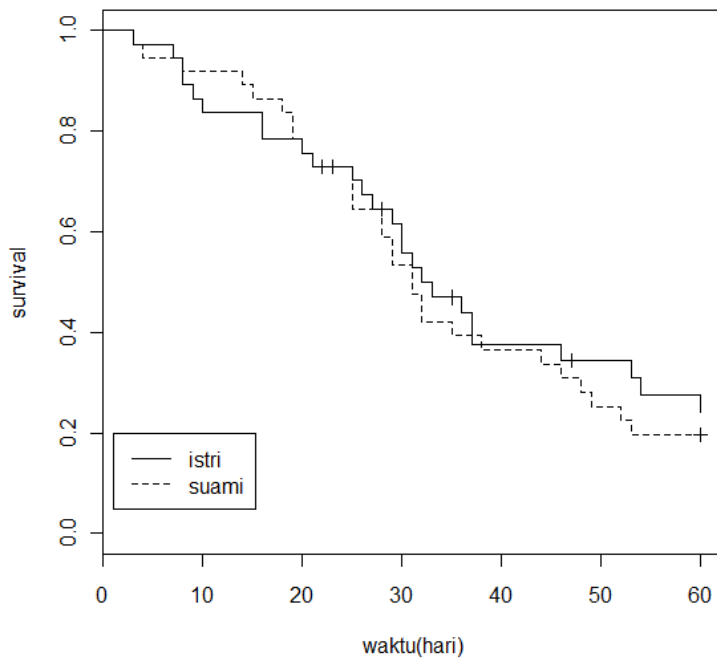
### 3. HASIL DAN PEMBAHASAN

Data hipertensi yang dianalisis terdiri dari 37 pasang suami istri. Status pasien dibedakan menjadi 2, yaitu sembuh dan tidak sembuh atau digolongkan tersensor kanan. Adanya data yang tersensor kanan disebabkan karena pasien tersebut tidak melanjutkan atau tidak melakukan kontrol/ pemeriksaan. Dari data tersebut, terdapat 8 data suami dan 11 data istri yang tersensor kanan. Informasi yang digunakan adalah identitas pasien, usia, jenis kelamin, tanggal periksa, tanggal kontrol, dan keterangan sembuh atau tidak. Adapun beberapa hal yang perlu diperhatikan kaitannya dengan data tersebut, yaitu:

- a. Kejadian yang menjadi perhatian, sembuh dari hipertensi
- b. *Origin*, saat pertama kali memeriksakan diri
- c. Unit/ satuan waktu, hari.

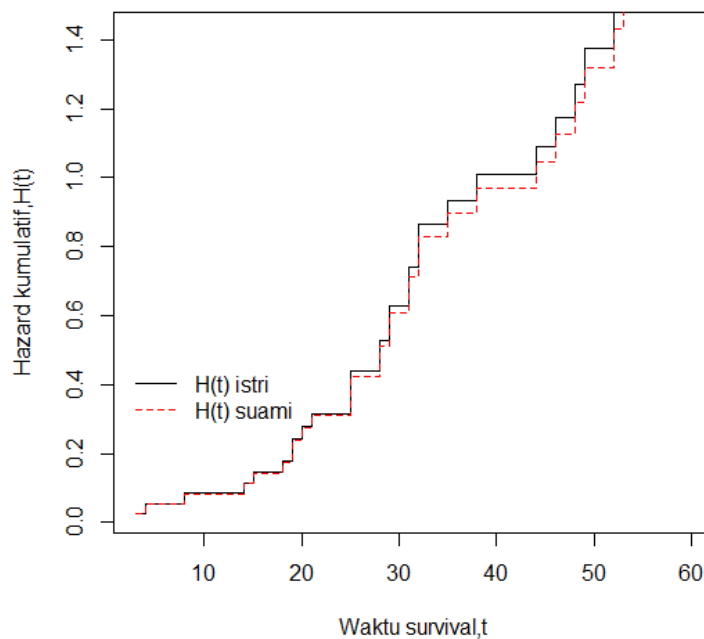
Langkah awal, melakukan estimasi fungsi survival Kaplan-Meier untuk seluruh observasi. Tujuan dari pengamatan terhadap pasien hipertensi pasangan suami istri ini adalah untuk mengetahui apakah ada perbedaan lama sembuh antara suami dan istri dari penyakit hipertensi tersebut. Selanjutnya, dilakukan

estimasi fungsi survival dengan menggunakan estimasi kaplen-Meier pasangan suami istri. Berikut hasilnya:



Gambar 1. Kurva  $\hat{S}(t)_{KM}$  Data Hipertensi Berpasangan

Berdasarkan gambar di atas tampak bahwa plot fungsi survival pasangan suami dan istri berbentuk fungsi tangga. Tangga akan turun ketika terjadi *event* saja. Hal ini menunjukkan bahwa semakin hari, peluang pasien masih hipertensi semakin kecil, dengan kata lain pasien akan sembuh. Garis pendek tegak pada grafik menunjukkan observasi yang tersensor kanan. Selain itu, pada gambar juga nampak ada garis yang berimpit, ada pula yang tidak. Artinya, tidak ada perbedaan yang signifikan waktu sembuh antara suami istri penderita hipertensi. Selanjutnya, estimasi kumulatif hazard pasangan suami istri dengan estimasi Nelson-Aalen:



Gambar 2. Kurva  $\hat{H}(t)_{NA}$  Data Hipertensi Berpasangan

Berdasarkan gambar di atas tampak bahwa plot fungsi hazard kumulatif pasangan suami dan istri berbentuk fungsi tangga. Tangga akan naik ketika terjadi *event* saja. Hal ini menunjukkan bahwa semakin bertambahnya hari maka resiko untuk sembuh dari hipertensi semakin meningkat. Selain itu, pada gambar juga nampak ada garis yang berimpit, ada pula yang tidak. Berdasarkan kurva tersebut, tidak ada perbedaan yang signifikan *rate* terjadinya *event* (sembuh) antara pasien hipertensi suami dan istri. Oleh karena itu, untuk mengetahui keadaan yang sebenarnya, dilakukan uji log rank agar diperoleh kesimpulan yang benar. Dengan menggunakan bantuan R, diperoleh:

Tabel 1. Output uji log rank

	N	Observed	Expected	$\frac{(O - E)^2}{E}$	$\frac{(O - E)^2}{V}$
Istri = 0	37	26	27,7	0,110	0,228
Suami = 1	37	29	27,3	0,112	0,228

Chisq = 0,2; dk = 1, p = 0,633

Berdasarkan output di atas diperoleh p-value = 0,633 >  $\alpha = 0,05$ , maka  $H_0$  diterima, artinya kelompok pasien suami dan istri memiliki peluang yang sama dalam hal sembuh dari hipertensi. Berikut ringkasan hasil uji statistik yang diperoleh dengan menggunakan program R. Selanjutnya dilakukan transformasi data, dan berikut ringkasan hasilnya.

Tabel 2. Nilai uji statistik transformasi data berpasangan

Waktu (t)	Transformasi	Nilai uji statistik
(1)	(2)	(3)
8	$\hat{S}_i(t)$	-0,3934045
	$\log \hat{S}_i(t)$	-0,3925781
	$\log(-\log \hat{S}_i(t))$	0,3914341
	$\text{logit } \hat{S}_i(t)$	-0,3916809
	$\arcsin \hat{S}_i(t)$	-0,3939
31	$\hat{S}_i(t)$	0,4233699
	$\log \hat{S}_i(t)$	0,422266
	$\log(-\log \hat{S}_i(t))$	-0,4225333
	$\text{logit } \hat{S}_i(t)$	0,422634
	$\arcsin \hat{S}_i(t)$	0,4230233
53	$\hat{S}_i(t)$	0,9223138
	$\log \hat{S}_i(t)$	0,9086804
	$\log(-\log \hat{S}_i(t))$	-0,9153931
	$\text{logit } \hat{S}_i(t)$	0,9118655
	$\arcsin \hat{S}_i(t)$	0,9190651

Dari table di atas, nampak bahwa nilai uji statistik berada pada daerah penerimaan  $H_0$ . Dengan demikian dapat disimpulkan bahwa tidak ada perbedaan lama sembuh pasien hipertensi antara kelompok suami dan istri dan berlaku untuk semua hasil transformasi. Hal ini sesuai dengan uji statistik dengan menggunakan uji log rank.

#### 4. KESIMPULAN

Berdasarkan hasil estimasi Kaplan-Meier dan Nelson-Aalen, tidak ada perbedaan waktu sembuh hipertensi data berpasangan antara suami dan istri. Setelah dilakukan transformasi yang dilanjutkan dengan uji statistik untuk 3 (tiga) titik waktu, diperoleh bahwa tidak ada perbedaan waktu sembuh hipertensi data berpasangan antara suami dan istri. Hasil tersebut juga sama pada saat dilakukan uji log rank.

DAFTAR PUSTAKA

- Bie dkk, 1987, Confidence Intervals and Confidence Bands for The Cumulative Hazard Rate Function and Their Small Sample Properties, *Scandinavian Journal of Statistics*, 3, 221-233.
- Bintang, 2017, Aplikasi Metode Kaplan Meier untuk Menduga Selang Waktu Ketahanan Hidup, *Skripsi*, Program Studi Matematika, Universitas Sanata Dharma, Yogyakarta.
- Borgan dkk., 1990, A Note on Confidence Intervals and Bands for the Survival Function Based on Transformations, *Scandinavian Journal of Statistics*, 1, 35-41.
- Dallas, M.J., Rao, P.V., 2000, Testing Equality of Survival Function Based on Both Paired and Unpaired Censored Data, *Biometrics*, 56, 154-159.
- Danardono, 2012, *Diktat Kuliah Analisis Data Survival*, Program Studi Statistika Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Gadjah Mada, Yogyakarta.
- Hikmah, 2015, Analisis Kurva Survival pada Titik Waktu Tertentu untuk Data Tersensor Kanan Berpasangan dengan Estimasi Kaplan Meier, *Tesis*, Program Pasca Sarjana Matematika, Universitas Gadjah Mada, Yogyakarta.
- Murray, S., 2001, Using Weighted Kaplan-Meier Statistics in Nonparametric Comparison of Paired Censored Survival Outcomes, *Biometrics*, 57, 361-368.
- Singh, R. and Mukhopadhyay, K., 2011, *Survival Analysis in Clinical Trials: Basics and Must Know Areas*, College of Medicine and JNM Hospital, Kalyani, West Bengal, India.
- Su, dkk, 2010, Analyzing Survival Curves at A Fixed Point in Time for Paired and Clustered Right-Censored Data, *Computational Statistics and Data Analysis*, 55, 1617-1628