BAB III

ESTIMASI FUNGSI TAHAN - HIDUP

Dalam metode statistik ada dua kategori untuk menyelesaikan suatu masalah, yaitu statistika parametrik dan statistika nonparametrik. Statistika parametrik digunakan bila distribusi populasi mempunyai bentuk matematika yang diketahui dan memuat beberapa parameter yang tidak diketahui. Bila dengan distribusi tertentu tidak bisa diterapkan, maka lebih baik digunakan metode yang tidak menghiraukan bentuk distribusi populasinya atau disebut dengan distribusi-bebas atau metode nonparametrik. Dua batasan mengenai statistika non-parametrik diberikan di bawah ini :

1. Suatu keluarga fungsi probabilitas dikatakan non-parametrik bila banyaknya harga-harga parameter ditentukan, dengan tidak menentukan secara tunggal suatu anggota keluarga itu.
2. Metode nonparametrik digunakan pada data dengan ukuran skala nominal, skala ordinal dan skala interval/rasio.

dimana skala nominal : jumlah yang hanya memisahkan elemen-elemen menjadi klas-klas atau golongan yang berbeda.

skala ordinal : ukuran-ukuran yang menyatakan per-bandingan, misal terbesar, ter-kecil atau sama dengan di antara ukuran-ukuran yang relevan.

skala interval : ukuran yang memberikan baik atau tidaknya ukuran yang diberikan.

Dalam pokok pembahasan ini, akan dibicarakan mengenai beberapa estimator yang digunakan untuk menaksir fungsi tahan-hidup di dalam metode statistika nonparametrik dengan menggunakan satu sampel dalam penerapannya. Kita gunakan satu sampel disini dengan maksud untuk membatasi permasalahan yang lebih kompleks. Karenanya, untuk contoh aplikasi yang menggunakan dua sampel sampai k sampel lebih baik diselesaikan dengan uji pada hipotesis.

Beberapa estimator yang akan dibicarakan yaitu :

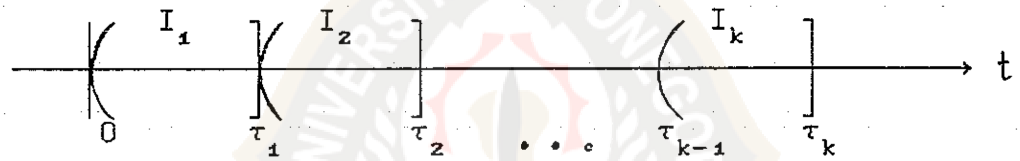
1. Estimator Produk-Limit (Kaplan-Meier)

Estimator ini merupakan lanjutan dari estimator aktuaria, yang digunakan dalam bidang demografi terutama untuk asuransi.

2. Estimator Fungsi Hazard

Dimana estimator ini merupakan estimator dalam statistika nonparametrik yang berhubungan dengan fungsi tahan-hidup dan fungsinya sama dengan estimator Produk-Limit serta banyak digunakan dalam aplikasi lain (misal dalam regresi). Estimator Produk-Limit lebih simpel (sederhana) akan lebih banyak digunakan dalam pengamatan medis karena sifatnya yang konsisten, tak bias dan normal asimtotik, maka akan diberikan dan dibicarakan secara panjang lebar beserta contoh aplikasinya dalam bidang medis. Untuk data pengamatan diambil dari tabel kehidupan, dimana data tersebut telah diperoleh dari laporan penelitian dan ukuran standart yang telah ditetapkan pada waktu dan dengan jumlah obyek tertentu. Juga memfokuskan pada data tersensor.

Ambil waktu yang dibagi menjadi barisan tetap dari interval I_1, I_2, ..., I_k. Untuk populasi manusia, panjang masing-masing interval biasanya satu tahun atau dapat juga untuk beberapa tahun. Partisi dari interval waktu dapat digambarkan sebagai berikut :



dimana

t = waktu tahan-hidup

I_k = panjang interval sampai ke-k

$k = 1, 2, \ldots$

τ_k = estimasi waktu dalam interval sampai ke-k

Dalam tabel kehidupan terdapat subyek-subyek sebagai berikut :

n_i = jumlah individu yang hidup pada permulaan interval I_i

d_i = jumlah individu yang meninggal selama interval I_i

l_i = jumlah individu yang hilang untuk diikuti lebih lanjut selama I_i

w_i = jumlah individu yang mengundurkan diri selama I_i

p_i = probabilitas tahan-hidup pada I_i

q_i = probabilitas yang meninggal pada I_i

$\hat{p}$ = proporsi tahan-hidup

$\hat{q}$ = proporsi yang meninggal

$\hat{S}(t)$ = Estimasi fungsi tahan-hidup

3.1 Metode Reduksi Sampel

Dengan metode ini kita dapat mencari estimasi dari fungsi tahan-hidup, dimana disini kita gunakan $S(\tau_k)$, yang memasukkan subyek-subyek di atas akan mempunyai resiko dalam τ (yaitu estimasi waktu dalam interval). Ambil n sebagai jumlah individu yang masih dapat mempertahankan hidup dalam data pengamatan yang tersensor sehingga diperoleh :

$$n = n_1 - \sum_{i=1}^{k} l_i - \sum_{i=1}^{k} w_i$$

dan d sebagai jumlah individu yang meninggal, yaitu :

$$d = \sum_{i=1}^{k} d_i$$

Maka diperoleh rumus untuk menghitung estimasi fungsi tahan-hidup yang menggunakan metode Reduksi Sampel dengan interval waktu konstan, yaitu :

$$\hat{S}(\tau_k) = 1 - \frac{d}{n}$$

yang merupakan estimasi dari S(t).

Sebagai ilustrasi dari metode Reduksi Sampel perhatikan data berikut :

Suatu penelitian yang diadakan untuk mendiagnosis suatu penyakit dari 126 pasien, dengan mengambil interval satu tahun dan lamanya diagnosis adalah 5 tahun. Secara lengkap disajikan dalam tabel kehidupan di bawah ini. Kita akan menaksir S (5 tahun). Data tersebut seperti pada tabel 1.2.

Tabel 1.2. Perhitungan Angka Tahan - Hidup 5 Tahun

Tahun-tahun setelah diagnosis	Kehidupan pada permulaan interval	kematian selama interval	Hilang untuk diikuti lebih lanjut selama interval	mengundurkan diri selama interval
(1)	(2)	(3)	(4)	(5)
0 - 1	126	47	4	15
1 - 2	60	5	6	11
2 - 3	38	2	-	15
3 - 4	21	2	2	7
4 - 5	10	-	-	6
Jumlah		56	12	54

Dari tabel 1.2 terlihat bahwa I_1, I_2, ... I_5 masing-masing mempunyai panjang interval 1 (satu) tahun. Kolom (2) memuat n_i, (3) memuat d_i, (4) memuat l_i dan (5) memuat w_i. Dari tabel 1.2 maka dapat diperoleh bahwa

$$n = n_1 - \sum_{i=1}^{k} l_i - \sum_{i=1}^{k} w_i$$

$$= 126 - 12 - 54$$

$$= 60$$

$$d = \sum_{i=1}^{k} d_i$$

$$= 56$$

$$\hat{S}(5 \text{ tahun}) = 1 - \frac{d}{n}$$

$$= 1 - \frac{56}{60}$$

Jadi $\hat{S}$ (5 tahun) = 0,078

3.2 Metode Aktuaria

Kita akan mengubah probabilitas tahan-hidup $S(\tau_k)$ menjadi perkalian probabilitas :

$$S(\tau_k) = P\,[T > \tau_k]$$
$$= P\,[T > \tau_1] \quad P\,[T > \tau_2 \mid T > \tau_1] \ldots$$
$$P\,[T > \tau_k \mid T > \tau_{k-1}]$$
$$= p_1 \cdot p_2 \ldots p_k$$

dimana $p_i = P\,\{\, T > \tau_i \mid T > \tau_{i-1}]$

Metode aktuaria memberikan estimasi pada masing-masing p_i secara terpisah dan kemudian menggandakannya yang akan menghasilkan estimasi $S(\tau_k)$.

Untuk estimasi pada p_i, kita dapat menggunakan $1-\frac{d_i}{n_i}$ jika tidak ada yang hilang atau mengundurkan diri dalam interval I_i. Walaupun demikian, dengan p_i dan w_i yang tidak sama dengan nol, kita anggap sebagai rata-rata sehingga individu-individu itu menjadi hilang atau mengundurkan diri selama I_i, pada resiko setengah interval.

Ukuran sampel efektifnya adalah :

$$n_i' = n_i - 1/2\,(l_i + w_i)$$

$$\hat{q} = \frac{d_i}{n_i'}$$

$$\hat{p} = 1 - \hat{q}_i$$

Estimator aktuaria adalah

$$\hat{S}(\tau_k) = \prod_{i=1}^{k} p_i$$

Dengan menggunakan estimator aktuaria, lihat tabel kehidupan di bawah ini yang merupakan lanjutan dari tabel 1.2 untuk menghitung $\hat{S}$(5 tahun).

Tabel 1.3 Lanjutan Dari Perhitungan Angka Tahan-Hidup 5 Tahun

Jumlah efektif pada resiko kematian (2)-1/2(4)+(5)	Proporsi kematian (3)/(6)	Proporsi tahan hidup 1 - (7)	Proporsi Kumulatif dari tahan-hidup diagnosis melalui akhir interval π_i^k (8)
(6)	(7)	(8)	(9)
116,5	0,40	0,60	0,60
51,5	0,10	0,90	0,54
30,5	0,07	0,93	0,50
16,5	0,12	0,88	0,44
7,0	0,00	1,00	0,44

dimana kolom (6) memuat n_i', (7) memuat $\hat{q}_i$, (8) memuat $\hat{p}_i$, dan (9) adalah estimasi $S(\tau_k)$. Berikut ini akan dihitung estimasi $\hat{S}$(5 tahun). Dari ukuran sampel-sampel :

$$n_i' = n_i - 1/2\ (l_i + w_i)$$

maka dapat diperoleh :

untuk $i = 1$; $n_1' = 126 - 1/2\ (4 + 15)$

$= 116,5$

$i = 2$; $n_2' = 60 - 1/2\ (6 + 11)$

$= 51,5$

$i = 3$; $n_3' = 38 - 1/2\ (0 + 15)$

$= 30,5$

$i = 4 \quad ; n_4' = 21 - 1/2\ (2 + 7)$

$= 16{,}5$

$i = 5 \quad ; n_5' = 10 - 1/2\ (0 + 6)$

$= 7{,}0$

Mencari $\hat{q}_i = \dfrac{d_i}{n_i'}$

Untuk $i = 1 \quad ; \hat{q}_1 = \dfrac{47}{116{,}5}$

$= 0{,}40$

$i = 2 \quad ; \hat{q}_2 = \dfrac{5}{51{,}5}$

$= 0{,}10$

$i = 3 \quad ; \hat{q}_3 = \dfrac{2}{30{,}5}$

$= 0{,}07$

$i = 4 \quad ; \hat{q}_4 = \dfrac{2}{16{,}5}$

$= 0{,}12$

$i = 5 \quad ; \hat{q}_5 = \dfrac{0}{7{,}0}$

$= 0{,}00$

Mencari $\hat{p}_i = 1 - \hat{q}_i$

Untuk $i = 1 \quad ; \hat{p}_1 = 1 - 0{,}40$

$= 0{,}60$

$i = 1 \quad ; \hat{p}_2 = 1 - 0{,}10$

$= 0{,}90$

$i = 1 \quad ; \hat{p}_3 = 1 - 0{,}07$

$= 0{,}93$

$i = 1 \; ; \; \hat{p}_4 = 1 - 0{,}12$

$= 0{,}88$

$i = 1 \; ; \; \hat{p}_5 = 1 - 0{,}00$

$= 1{,}00$

mencari $\hat{S}(\tau_k) = \prod_{i=1}^{k} p_i$

Untuk $i = 1$; $\hat{S}$(1 tahun) = 0,60

$i = 2$; $\hat{S}$(2 tahun) = 0,60 * 0,90

= 0,54

$i = 3$; $\hat{S}$(3 tahun) = 0,54 * 0,93

= 0,50

$i = 4$; $\hat{S}$(4 tahun) = 0,50 * 0,88

= 0,44

$i = 5$; $\hat{S}$(5 tahun) = 0,44 * 1,00

= 0,44

Jadi $\hat{S}$(5 tahun) = 0,44

3.3 Variansi Dari $\hat{S}(\tau_k)$

Untuk menaksir variansi dari $\hat{S}(\tau_k)$, perhatikan

$$\log S(\tau_k) = \prod_{i=1}^{k} \log \hat{p}_i$$

Dengan menganggap $n_i' \hat{p}_i$ binomial (n_i', p_i) maka dengan Metode Delta diperoleh

$$\text{var} [\log \hat{p}_i] = \text{par} (\hat{p}_i) \left(\frac{d \log p_i}{d \, p_i} \right)^2$$

dimana

$$E\,[\hat{p}_i \mid n_i'] = p_i$$

$$\text{var}\,(\hat{p}_i \mid n_i') = \frac{p_i\ q_i}{n_i'}$$

sehingga

$$\text{var}\,[\log \hat{p}_i] = \frac{p_i\ q_i}{n_i'} \cdot \frac{1}{p_i^2}$$

$$= \frac{q_i}{n_i'\ p_i}$$

Dengan menganggap log p_i, ... , log p_k independent

maka

$$\text{var}\,[\log \hat{S}\,(\tau_k)] = \sum_{i=1}^{k} \frac{q_i}{n_i'\ p_i}$$

$$\hat{\text{var}}\,[\log \hat{S}\,(\tau_k)] = \sum_{i=1}^{k} \frac{\hat{q}_i}{n_i'\ p_i}$$

$$= \sum_{i=1}^{k} \frac{d_i/n_i'}{n_i'\left(1 - \frac{d_i}{n_i'}\right)}$$

$$= \sum_{i=1}^{k} \frac{d_i}{n_i'^{2}\left(1 - \frac{d_i}{n_i'}\right)}$$

$$= \sum_{i=1}^{k} \frac{d_i}{n_i'^{2}(n_i' - d_i)}$$

Dengan menggunakan Metode Delta, dan

$$\frac{\partial \hat{p}_i}{\partial p_i} = \frac{\hat{p}_i}{p_i}$$

$$\mathrm{var}\ (p_i' \mid n_i') = \frac{p_i\ q_i}{n_i'}$$

Maka diperoleh

$$\hat{\mathrm{var}}\ (\hat{S}\ (\tau_k) = \mathrm{var}\ (\hat{p}_i) \left(\frac{\partial \hat{p}_i}{\partial p_i} \right)^2$$

$$= \sum_{i=1}^{k} \frac{p_i\ q_i}{n_i'} \cdot \frac{\hat{p}_i^2}{p_i^2}$$

$$= \sum_{i=1}^{k} \frac{\hat{p}_i^2\ q_i}{n_i'\ p_i}$$

$$= \hat{S}^2\ (\tau_k) \sum_{i=1}^{k} \frac{q_i}{n_i'\ p_i}$$

$$\mathrm{var}\ (\hat{S}\ (\tau_k) = \hat{S}^2\ (\tau_k) \sum_{i=1}^{k} \frac{d_i}{n_i'\ (n_i' - d_i)}$$

$\hat{\mathrm{Var}}\ (\hat{S}(\tau_k))$ disebut RUMUS GREENWOOD, yang digunakan untuk mencari harga pendekatan standart error dari fungsi tahan-hidup.

Dari tabel 1.2 (tabel dari Cutter dan Edlur) diatas kita akan menghitung harga pendekatan standart error dari $\hat{S}(5)$. Gunakan rumus Greenwood :

$$\hat{\mathrm{var}}\ (\hat{S}\ (\tau_k) = S^2\ (\tau_k) \sum_{i=1}^{k} \frac{d_i}{n_i'\ (n_i' - d_i)}$$

maka diperoleh

$$\hat{\text{var}}\,(\hat{S}\,(5)) = (0,44)^2 \left[\frac{47}{116,5\,(116,5 - 47)} + \frac{5}{51,5\,(30,5 - 2)} + \frac{2}{30,5\,(51,5 - 5)} + \frac{2}{16,5\,(16,5 - 2)} + \frac{0}{7\,(7 - 0)} \right]$$

$$= 0,003608$$

Jadi

$$\hat{S}E\,(\hat{S}\,(5)) = \sqrt{\hat{\text{var}}\,(\hat{S}\,(5))}$$

$$= \sqrt{0,003608}$$

$$= 0,06$$

Sehingga pendekatan standart error = 0,06.