BAB II

DASAR TEORI

Sebelum melakukan estimasi probabilitas terhadap data air tanah PDAM Kotamadia Semarang lebih lanjut, terlebih dahulu akan diuraikan teori-teori yang digunakan dalam penulisan karya ilmiah ini. Teori-teori ini penting sebagai dasar untuk melakukan analisa data serta pengambilan kesimpulan terhadap objek penelitian.

2.1. Rancangan Blok Random Lengkap (RBRL)

Dalam banyak permasalahan percobaan, pemilihan suatu rancangan percobaan adalah sangat penting. Sehingga keragaman yang timbul dari sumber-sumber yang tidak berhubungan dapat dikontrol secara sistematis.

Jenis-jenis rancangan percobaan dalam suatu penelitian ada bermacam-macam, diantaranya : rancangan sistematis, rancangan random lengkap, rancangan blok random lengkap.

Pada penulisan karya ilmiah ini rancangan yang dipilih adalah rancangan blok random lengkap.

Rancangan blok random lengkap (RBRL) adalah suatu rancangan percobaan yang terdiri atas blok-blok dan perlakuan-perlakuan (treatments), dimana pada skema rancangan percobaannya memiliki unit perlakuan sebanyak a buah dan unit blok sebanyak b buah. Apabila unit-unit percobaan sangat heterogen maka unit-unit percobaan ini dikelompokkan ke dalam himpunan-himpunan yang lebih homogen.

Dengan demikian pada rancangan percobaan ini, unit-unit dalam bloknya lebih homogen dan berisi semua unit perlakuan . Dimana tiap perlakuan dalam blok hanya ada satu data pengamatan.

Skema Rancangan Blok Random Lengkap dapat digambarkan sebagai berikut :

Tabel 2.1 Skema Data Pengamatan Untuk Rancangan Blok Random Lengkap

Treatments (i)	Blok (j)						Jumlah Nilai Treatment ($G_{i.}$)
	1	2	.	.	.	b	
1	Y_{11}	Y_{12}	.	.	.	Y_{1b}	$G_{1.}$
2	Y_{21}	Y_{22}	.	.	.	Y_{2b}	$G_{2.}$
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
a	Y_{a1}	Y_{a2}	.	.	.	Y_{ab}	$G_{a.}$
Jumlah Nilai Blok ($G_{.j}$)	$G_{.1}$	$G_{.2}$	.	.	.	$G_{.b}$	$G_{..}$

Dalam tabel 2.1 skema rancangan percobaan diatas adalah menyatakan sebuah data pengamatan yang diambil dari perlakuan ke-i dan blok ke-j. Dimana i = 1,2,3,...,a dan j = 1,2,3,...,b. Sedang $G_{..}$ adalah nilai total jumlah nilai treatment dan jumlah nilai blok. Sehingga masing-masing blok terdapat dengan tepat sebuah data pengamatan untuk tiap perlakuan.

Karena pada rancangan percobaan itu mempunyai a buah perlakuan dan b buah blok maka data yang ada dalam tabel rancangan percobaan secara lengkap adalah sebanyak axb data pengamatan. (Montgomery, 1982)

2.1.1. Data Tidak Lengkap (Missing Data)

Pencatatan dan penyusunan data adalah suatu pekerjaan yang tidak bisa dipisahkan dalam suatu penelitian. Data harus dicatat dengan cermat dan teliti sesuai dengan hasil penelitian yang sesungguhnya terjadi dilapangan.

Namun dalam kenyataannya, walaupun penyusunan maupun pengumpulan data telah dilakukan dengan cermat dan teliti, seringkali dijumpai data pengamatan yang diperoleh terjadi missing data.

Missing data (data tidak lengkap) adalah kondisi dimana banyaknya data pengamatan dalam rancangan blok random lengkap kurang dari axb data pengamatan.

Hal ini bisa terjadi karena ada beberapa sebab, diantaranya :

1. Kesalahan mencatat nilainya
2. Kurang hati-hati saat melakukan penelitian
3. Alasan-alasan diluar kontrol
4. Data rusak

Sehingga data yang diperoleh tidak cukup valid untuk dijadikan sebagai data pengamatan.

Tabel 2.2 Data jumlah Pengambilan Air Tanah PDAM Kotamadia Semarang Untuk Daerah Semarang Atas dengan dua missing data

Tre(i)					Blok (j)						Jml.Nil.Tre
	Tampi	K.Lang	Campu	K.Leng	Cang.A	Gowok	Cang.B	Bubak	Blancen	Sembung	(Gi.)
Jan	48154	54458	46593	30643	66985	76282	51826	25992	X	8701	409634
Febr	58918	61118	68767	38679	58799	66647	47757	38875	2822	Y	442382
Mar	63344	71004	72075	39881	67606	77512	47809	33506	11249	13392	497378
April	63326	57693	73268	39949	84116	77022	57246	36151	21957	29035	539763
Mei	56624	61180	88387	29357	74775	76107	57740	37422	23362	24942	529896
Juni	61749	57148	51542	33880	47478	79533	46165	28699	20641	22843	449678
Juli	48239	66802	78204	40792	48946	62031	37461	19130	19310	21934	442849
Agus	38239	57988	70519	41896	51859	71082	22415	11248	20746	34694	420686
Sept	57950	55985	62612	37144	45188	65616	15865	11330	18784	34942	405416
Okt	79312	73737	78494	42451	49690	76036	16314	15012	18614	40873	490533
Nop	64816	54562	58794	32911	40076	61874	13142	10843	12825	35984	385827
Des	71264	52100	65876	31706	42512	69616	15510	9180	16359	35441	409564
J.N.B	711935	723775	815131	439289	678030	859358	429250	277388	186669	302781	5423606
(G.j)											

Dari contoh diatas terlihat ada dua missing data yaitu pada bulan Januari didesa Blancen dan pada bulan Februari didesa Sembungan.

Apabila dalam rancangan percobaan terdapat missing data seperti diatas maka untuk keperluan analisa statistik, missing data ini harus diestimasi nilainya.

2.1.2. Estimasi Data Tidak Lengkap (Estimating Missing Data)

Sebuah metode untuk mengestimasi missing data telah disusun. Bagi seorang peneliti, pengetahuan tentang metode tersebut penting karena metode estimasi missing data akan membantu disaat melakukan analisa data .

Menurut Montgomery (1982) ada dua metode pendekatan untuk mengestimasi data tidak lengkap :

1. Analisa Pendekatan (Approximate Analysis)

   Dalam metode ini perhitungan yang dipergunakan adalah dengan menggunakan formula sebagai berikut :

$$Y = \frac{rG_{.j} + tG_{i.} - G_{..}}{(r - 1)(t - 1)} \qquad \ldots(2.1)$$

   dimana :

   Y = data hilang

   r = jumlah banyaknya blok

   t = jumlah banyaknya perlakuan

$G_{.j}$ = total pengamatan dalam blok ke-j

$G_{i.}$ = total pengamatan dalam perlakuan ke-i

$G_{..}$ = jumlah total pengamatan keseluruhan

Persamaan diatas diperoleh dengan meminimumkan jumlahan error kuadrat (SSE) dari Rancangan Blok Random Lengkap. Adapun persamaan untuk SSE adalah :

$$SSE = x^2 - \frac{1}{b}(y'_{i.} + x)^2 - \frac{1}{a}(y'_{.j} + x)^2 + \frac{1}{ab}(y'_{..} + x)^2 + R$$

$$= x^2 - \frac{1}{b}(y'^2_{i.} + 2xy'_{i.} + x^2) - \frac{1}{a}(y'^2_{.j} + 2xy'_{.j} + x^2)$$

$$+ \frac{1}{ab}(y'^2_{..} + 2xy'_{..} + x^2) + R$$

dimana :

a = banyaknya blok

b = banyaknya treatments

Dengan menurunkan SSE terhadap x ($\partial SSE / \partial x$) diperoleh hasil sebagai berikut :

$$\frac{\partial SSE}{\partial x} = 2x - \frac{1}{b}(2y'_{i.} + 2x^2) - \frac{1}{a}(2y'_{.j} + 2x)$$

$$+ \frac{1}{ab}(2y'_{..} + 2x)$$

$$= 2x - \frac{2}{b} y'_{i.} - \frac{2}{b} x - \frac{2}{a} y'_{.j} + \frac{2}{a} x$$

$$+ \frac{2}{ab} y'_{..} + \frac{2}{ab} x$$

$$= 2x \left(1 - \frac{1}{b} - \frac{1}{a} + \frac{1}{ab} \right) - 2 \left(\frac{y'_{i.}}{b} + \frac{y'_{.j}}{a} + \frac{y'_{..}}{ab} \right)$$

Jika $\frac{\partial SSE}{\partial x} = 0$, maka :

$$2x \left(1 - \frac{1}{b} - \frac{1}{a} + \frac{1}{ab} \right) = 2 \left(\frac{y'_{i.}}{b} + \frac{y'_{.j}}{a} + \frac{y'_{..}}{ab} \right)$$

$$x = \frac{\left(\frac{y'_{i.}}{b} + \frac{y'_{.j}}{a} + \frac{y'_{..}}{ab} \right)}{\left(1 - \frac{1}{b} - \frac{1}{a} + \frac{1}{ab} \right)}$$

$$= \frac{\left(\frac{y'_{i.}}{b} + \frac{y'_{.j}}{a} + \frac{y'_{..}}{ab} \right)}{\left(1 - \frac{1}{b} - \frac{1}{a} + \frac{1}{ab} \right)} \cdot \frac{ab}{ab}$$

$$= \frac{a y'_{i.} + b y'_{.j} + y'_{..}}{ab - b - a - 1}$$

$$= \frac{a y'_{i.} + b y'_{.j} + y'_{..}}{(a - 1)(b - 1)} \quad \ldots\ldots(2.4)$$

dengan mengganti :

$$a = r \ ; \quad b = b \ ; \quad y'_{i.} = G_{i.} \ ; \quad y'_{.j} = G_{.j} \ ; \quad y'_{..} = G_{..}$$

Maka :

$$x = \frac{rG_{i.} + tG_{.j} + G_{..}}{(r - 1)(t - 1)}$$

Terlihat pers.(2.4) sama dengan pers.(2.1) :

$$x = \frac{a y_{i.} + b y_{.j} + y_{..}}{(a - 1)(b - 1)} = \frac{r G_{i.} + t G_{.j} + G_{.}}{(r - 1)(t - 1)}$$

Maka terbukti bahwa pers. (2.1) adalah turunan dari SSE.

2. Analisa Exact (Exact Analysis)

Dalam mengestimasi missing data dengan menggunakan analisa exact ini, cara perhitungannya adalah dengan memakai test signifikan regresi general. (Montgomery,1982)

Tetapi dalam Tugas Akhir ini, yang dipakai adalah metode Analisa Pendekatan (Approximate Analysis). Hal ini disebabkan adanya kesesuaian antara formula yang dipakai untuk mengestimasi missing data dengan rancangan percobaannya.

2.1.3. Prosedur Estimasi Data Tidak Lengkap

2.1.3.1. Prosedur Estimasi Untuk Satu Data Tidak Lengkap

Untuk lebih memudahkan dalam pengestimasian data tidak lengkap (missing data) pada rancangan percobaan diatas maka dilakukan prosedur sebagai berikut :

1. Menentukan data tidak lengkap

Misal : Y_{ij} , yaitu data tidak lengkap untuk perlakuan ke-i dalam blok ke-j

$i = 1,2,3,\ldots,a$

$j = 1,2,3,\ldots,b$

2. Menghitung jumlah nilai dari perlakuan ke-i

3. Menghitung jumlah nilai dari blok ke-j

4. Menghitung total jumlah nilai

5. Memasukkan nilai-nilai diatas pada pers (2.1) :

$$Y = \frac{rG_{.j} + tG_{i.} - G_{..}}{(r - 1)(t - 1)}$$

Prosedur diatas adalah untuk kasus dimana ada satu data missing (data tidak lengkap) dalam rancangan percobaannya.

Contoh :

Tabel 2.3. Data Pengukuran Kekuatan Ujung Logam

Treatments (i)	Blok (j)				Jumlah Nilai Treatment ($G_{i.}$)
	1	2	3	4	
1	6,8	6,6	7,0	6,4	26,8
2	6,3	Y	5,9	7,1	19,3
3	6,4	7,3	7,7	6,7	28,1
Jumlah Nilai Blok ($G_{.j}$)	19,5	13,9	20,6	20,2	74,2

Dari tabel 2.3 diatas terlihat ada satu missing data yaitu data Y. Data ini berada pada treatment ke-2 dan blok ke-2.

Dengan memakai pers.(2.1) diatas diperoleh estimasi :

$$Y = \frac{rG_{.j} + tG_{i.} - G_{..}}{(r-1)(t-1)}$$

$$= \frac{4(13,9) + 3(19,3) - 74,2}{(4-1)(3-1)}$$

$$= 6,55$$

2.1.3.2. Prosedur Estimasi Untuk Dua Data Tidak Lengkap (Missing Data)

Kemudian apabila terdapat dua missing data, prosedurnya adalah :

1. Menentukan dua data tidak lengkap

   Misal : Y_{ij} , yaitu missing data pertama

   X_{ij} . yaitu missing data kedua

   dengan : $i = 1,2,3,\ldots,a$

   $j = 1,2,3,\ldots,b$

2. Menghitung jumlah nilai dari perlakuan
3. Menghitung jumlah nilai dari blok
4. Menghitung total jumlah nilai
5. Menghitung nilai awal estimasi :

   $$Y_o = \frac{\overline{G}_{i.} + \overline{G}_{.j}}{2}$$

6. Memasukkan nilai-nilai diatas pada pers.(2.1) :

   $$Y = \frac{rG_{.j} + tG_{i.} - G_{..}}{(r-1)(t-1)}$$

7. Ulangi prosedur diatas hingga didapat hasil yang tidak berbeda secara material.

Contoh :

Tabel 2.4. Pengukuran Kekuatan Ujung Logam

Treatments (i)	BLOK (j)				Jumlah Nilai Treatment ($G_{i.}$)
	1	2	3	4	
1	4,4	5,9	6,0	4,1	20,4
2	X	1,9	4,9	7,1	13,9
3	4,4	4,0	4,5	3,1	16,0
4	6,8	6,6	Y	6,4	19,8
5	6,3	4,9	5,9	7,1	24,2
6	6,4	7,3	7,7	6,7	28,1
Jumlah Nilai Blok ($G_{.j}$)	28,3	30,6	29,0	34,5	122,4

Dari contoh diatas terdapat dua missing data, yaitu X pada perlakuan ke-2 dalam blok ke-1 dan Y pada perlakuan ke-4 dalam blok ke-3.

Missing data diatas apabila dikerjakan akan diperoleh nilai estimasi sebagai berikut :

1. Estimasi awal Y :

$$Y_o = \frac{\bar{G}_{i.} + \bar{G}_{.j}}{2}$$

$$= \frac{19{,}8/3 - 29{,}0/5}{2}$$

$$= 6{,}2$$

dimana : $\bar{G}_{i.}$ dan $\bar{G}_{.j}$ adalah rata-rata total perlakuan dan blok

2. Estimasi X, pengulangan pertama :

$$X_1 = \frac{rG_{.j} + tG_{i.} - G_1}{(r - 1)(t - 1)}$$

$$= \frac{4(28,3) + 6(13,9) - (122,4+6,2)}{(4 - 1)(6 - 1)}$$

$$= 4,5$$

dimana : $G_1 = G_{..} + Y_0$

3. Estimasi Y, pengulangan pertama :

$$Y_1 = \frac{rG_{.j} + tG_{i.} - G_2}{(r - 1)(t - 1)}$$

$$= \frac{4(28,3)+(6(19,8)-(122,4+4,5)}{(4-1)(6-1)}$$

$$= 7,2$$

dimana : $G_2 = G_{..} + X_1$

4. Estimasi X, pengulangan kedua :

$$X_2 = \frac{rG_{.j} + tG_{i.} - G_3}{(r - 1)(t - 1)}$$

$$= \frac{4(28,3)+6(13,9)-(122,4+7,2)}{(4-1)(6-1)}$$

$$= 4,5$$

dimana : $G_3 = G_{..} + Y_1$

5. Estimasi Y, pengulangan kedua :

$$Y_2 = \frac{rG_{.j} + tG_{i.} - G_4}{(r - 1)(t - 1)}$$

$$= \frac{4(29,0) + 6(19,8) - (122,4 + 4,5)}{(4 - 1)(6 - 1)}$$

$$= 7,2$$

dimana : $G_4 = G_{..} + X_2$

Dari perhitungan diatas, pengulangan ke dua dengan satu tempat desimal adalah cukup. Sehingga nilai estimasi untuk X = 4,5 dan Y = 7,2.

2.2. Histogram.

Banyak fenomena atau proses yang berhubungan dengan percobaan (experimen) bersifat random, yaitu hasil yang sebenarnya tidak dapat diperkirakan (dalam tingkat tertentu). Fenomena yang demikian ditandai dengan pengamatan yang berbeda dari satu percobaan ke yang lainnya, sekalipun mungkin dilakukan dengan kondisi yang sama.

Dengan perkataan lain, umumnya terdapat daerah kisaran dari nilai-nilai yang diamati. Dalam daerah ini beberapa nilai tertentu mungkin lebih sering dijumpai pada nilai yang lain. Karakteristik dari data semacam ini dapat digambarkan secara grafis dalam bentuk histogram.

Secara umum histogram mempunyai kegunaan sebagai berikut :

1. Mempertegas dan memperjelas pencaran data yang disajikan sebagai daftar.
2. Pengganti bagi distribusi frekuensi yang berbentuk daftar. (Pasaribu, 1978)

Untuk membuat histogram, dibuat langkah-langkah sebagai berikut :

1. Menentukan garis absis (horisontal) sebagai waktu dan garis ordinat (vertikal) untukjumlah air tanah.
2. Membentuk garis absis menjadi 12 kelas interval.
3. Menentukan estimasi probabilitas, yaitu membagi jumlah air tanah pada treatment dengan total jumlah air tanah.

2.3. Fungsi Densitas

Dalam menarik kesimpulan dari suatu data pengamatan pada Rancangan Blok Random Lengkap, teori probabilitas memegang peranan yang sangat penting, khususnya fungsi densitasnya. Yang berguna dalam perhitungan nilai probabilitas dari suatu data.

Definisi 2.3.1 :

Suatu fungsi f(x) merupakan fungsi densitas dari suatu variabel random kontinu x jika :

i. $f(x) \geq 0$, untuk semua x

ii. $\int_{-\infty}^{\infty} f(x)\, dx = 1$ (2.5)

iii. $\int_{a}^{b} f(x)\, dx = P(a \leq x \leq b)$ (2.6)

Definisi 2.3.2 :

Suatu fungsi densitas dikatakan simetri disekitar μ jika untuk setiap x berlaku $F(\mu+x) = f(\mu-x)$.

2.3.1. Estimasi Fungsi Densitas

Dimuka telah dijelaskan bahwa fungsi densitas sangat bermanfaat untuk mendapatkan informasi dari suatu data. Akan tetapi fungsi densitas biasanya tidak bisa diketahui secara langsung dari data asli (raw data). Oleh karena itu perlu dilakukan suatu pendekatan terhadap fungsi densitas yang belum diketahui.

Terdapat dua macam pendekatan yaitu :

1. Pendekatan Parametrik

   Pendekatan ini dilakukan jika fungsi densitas mempunyai parameter-parameter yang belum diketahui. Oleh karena itu dilakukan estimasi terhadap parameter-parameter tersebut. Dengan kata lain pendekatan ini identik dengan estimasi parameter.

2. Pendekatan Non Parametrik

   Pendekatan ini tidak membatasi kemungkinan bentuk dari fungsi densitas dengan asumsi fungsi f termasuk dalam keluarga fungsi densitas yang telah diketahui. Hal pokok dalam pendekatan ini adalah mengestimasi kurva kerapatannya. Pendekatan ini dilakukan ketika tidak diketahui informasi yang tepat tentang bentuk dan kelas fungsi densitas yang sebenarnya.(Wolfgang H, 1990)

2.4. Estimasi Kernel Density

Metode Estimasi Kernel Density adalah suatu metode pendekatan terhadap fungsi densitas yang belum diketahui dengan menggunakan fungsi kernel density. Dalam estimasi ini ada suatu parameter yang sangat berpengaruh terhadap keoptimalan hasil, yaitu bandwidth.

Karena metode estimasi kernel density adalah metode pendekatan terhadap fungsi densitas, maka metode ini bisa dipakai juga sebagai metode estimasi probabilitas.

Untuk itulah maka metode Estimasi Kernel Density dipakai sebagai metode untuk analisa data lebih lanjut. Metode ini didasarkan pada komponen bandwidth dan fungsi kernel.

Bandwidth adalah parameter penghalus yang mengatur derajat kehalusan untuk penghalus kernel, dilambangkan dengan h.

Menurut Wolfgang Hardle (1990) ada beberapa kemungkinan yang merupakan pengaruh pemilihan bandwidth terhadap grafik penghalusan fungsi kernel density, yaitu :

- h ——> 0 , grafik yang acak-acakan
- h kecil , fungsi agak halus
- h besar , fungsi halus
- h ——> ~ , fungsi datar (terlalu halus)

Secara umum Kernel K didefinisikan sebagai berikut :

$$K_h(x) = \frac{1}{h} K\left(\frac{x}{h}\right)$$

Jika didapat sejumlah pengamatan (observasi) $\{x_i\}_{i=1}^{\sim}$ maka estimator Kernel D ensity didefinisikan :

$$\hat{f}_h(x) = \frac{1}{n} \sum_{i=1}^{n} K_h\left(\frac{x}{h}\right)$$

$$= \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right)$$

Keterangan :

$\hat{f}_h(x)$ = estimator kernel density

K = jenis Kernel

n = banyaknya data pengamatan

x_i = data ke-i

x = nilai yang didapat dengan pendekatan

2.4.1. Pemilihan Tipe Kernel Densitas

Dalam penulisan tugas akhir ini fungsi kernel yang dipakai adalah tipe Ephanechnikov dan Quartic. Fungsi densitasnya adalah sebagai berikut :

Kernel	K(u)
Epanechnikov	$\frac{3}{4}(1-\|u\|)\ I\ (\|u\| \leq 1)$
Quartic	$\frac{15}{16}(\ 1-u^2\ )^2\ I\ (\|u\| \leq 1)$

Bentuk persamaan estimasi kernel density diatas adalah :

Tipe Kernel	Persamaan Est.Ker.Dens. K(u)
Epanechnikov	$\frac{1}{nh} \sum_{i=1}^{n} \frac{3}{4} (1-\|u\|)\ I\ (\|u\| \leq 1)$
Quartic	$\frac{1}{nh} \sum_{i=1}^{n} \frac{15}{16} (\ 1-u^2\ )^2\ I\ (\|u\| \leq 1)$

Dan untuk mendapatkan suatu nilai estimasi yang baik, harus memenuhi kriteria-kriteria :

- Estimasi memiliki kebiasan 0.

  Secara teori hal ini bisa terjadi tetapi dalam aplikasi sulit terjadi. Dengan demikian diharapkan sifat kebiasan harus mempunyai nilai kecil atau minimum.
- Varian kecil. (Wolfgang Hardle, 1990)

2.4.2. Pemilihan Bandwidth Optimal

Sebelumnya telah disinggung bahwa parameter bandwidth mempengaruhi keoptimalan hasil estimasi kernel density. Oleh karena itu pemilihan bandwidth yang optimal adalah sesuatu yang penting.

Dan karena begitu luasnya kemungkinan mengenai pemilihan maka perlu adanya batasan mengenai bandwidth yang akan diseleksi. Dengan batasan ini, akan diketahui bahwa bandwidth optimal berada dalam suatu interval tertentu.

Bandwidth yang terlalu besar menimbulkan kurva pendekatan terhadap fungsi densitas mengalami 'oversmoothing'. Apabila besar bandwidth tertentu mengakibatkan oversmoothing maka penghalusan tidak akan mencapai optimal untuk bandwidth yang lebih besar lagi karena justru akan terjadi oversmoothing. Oleh karena itu, keadaan ini bisa dijadikan tolok ukur sebagai batas atas bandwidth yang akan diseleksi.

Terrel (1990) memberikan suatu formula mengenai besar bandwidth yang menimbulkan oversmoothing, yaitu :

$$h_{os} = 3 \left| \frac{\| K \|_2^2}{35\, \mu_2^4 \,(K)} \right|^{\frac{1}{5}} \sigma\, n^{-\frac{1}{5}} \qquad (2.7)$$

Keterangan :

n = banyaknya data pengamatan

σ = standar deviasi

$\mu_2^4(K)$ dan $\| K \|_2^2$ dihitung menggunakan integrasi numerik dengan aturan trapesoida.

2.4.3. Pemilihan Kernel Density Terbaik

Dalam mendapatkan Kernel Density terbaik adalah dengan mengoptimalkan Kernal Density. Kondisi optimal ini bisa terjadi apabila fungsi A-MISE untuk kernel tersebut adalah minimal. Dimana masing-masing kernel mempunyai nilai A-MISE. Untuk mendapatkan kernel terbaik , maka dicari kernel yang mempunyai nilai A-MISE terkecil.

Misal $h' = h\,\partial^*$, ∂^* adalah bandwidth kanonik dari kernel kanonik, maka :

$$\text{A-MISE}\,(K,h') = (nh)^{-1} + \frac{h^4}{4}\,\|f''\|_2^2 \quad T(K)$$

dimana : T = sebuah fungsi dependent terhadap kernel K

Dengan meminimalkan K berarti akan meminimalkan T(K).

Karena dalam penulisan karya ilmiah ini ada dua kernel yang dipakai, yaitu Kernel Epanechnikov dan Kernel Quartic maka akan diuraikan proses peminimalan A-MISE untuk kedua kernel tersebut.

Peminimalan A-MISE Kernal Epanechnikov

$$K(u) = \frac{3}{4}\left(\frac{1}{15^{1/5}}\right)\left(1 - \left(\frac{u}{15^{1/5}}\right)^2\right) I\left(|u| \le 15^{1/5}\right)$$

Hasil diatas dihasilkan melalui langkah-langkah berikut :

Diketahui varian kernel $T(K) = \frac{1}{5}$

$$\partial^*_E = \left\{\frac{\|K\|_2^2}{\mu_2^2\,(K)}\right\}^{1/5}$$

$$\|K\|_2^2 = \frac{9}{16}\int_{-1}^{1} (1 - u^2)^2\, du$$

$$= \frac{9}{16}\;\frac{16}{15}\int_{-1}^{1} \frac{15}{16}(1 - 2u^2 + u^4)^2\, du$$

$$= \frac{9}{16}\;\frac{16}{15}\;\frac{15}{16}\left[\, u - \frac{2}{3}u^3 + \frac{1}{5}u^5 \,\right]_{-1}^{1}$$

$$= \frac{9}{16} \frac{16}{15} \frac{15}{16} \frac{16}{15}$$

$$= \frac{3}{5}$$

$$\mu_2(K) = \frac{3}{4} \int_{-1}^{1} u^2(1 - u^2) \quad du$$

$$= \frac{3}{4} \int_{-1}^{1} (u^2 - u^4) \quad du$$

$$= \frac{3}{4} [\frac{1}{3} u^3 + \frac{1}{5} u^5]_{-1}^{1}$$

$$= \frac{3}{4} (\frac{2}{3} - \frac{2}{5})$$

$$= \frac{1}{5}$$

$$\partial_E^{*2} = \{ \frac{\|K\|_2}{\mu_2^2 (K)} \}^{1/5}$$

$$= \{ \frac{3/5}{(1/5)^2} \}^{1/5}$$

$$= 15^{1/5}$$

$h' = \partial^* h$ dengan bandwidth $h = 1$ didapat $h' = 15^{1/5}$

Telah diketahui bahwa fungsi kernel untuk tipe Epanechnikov adalah

$$K_h(u) = \frac{1}{h} \quad K(\frac{u}{h})$$

$$= \frac{3}{4h} \quad (1 - (\frac{u}{h})^2) \quad I(|u| \leq h)$$

dengan substitusi h = h' didapat fungsi A-MISE minimal sebagai berikut :

$$K(u) = \frac{3}{4} \left(\frac{1}{15^{1/5}}\right) \left(1 - \left(\frac{u}{15^{1/5}}\right)^2\right) I\left(|u| \leq 15^{1/5} \right)$$

Peminimalan A-MISE Kernel Quartic

$$K(u) = \frac{15}{16} \left(\frac{1}{35^{1/5}}\right) \left(1 - \left(\frac{u}{35^{1/5}}\right)^2\right)^2 I\left(|u| \leq 35^{1/5}\right)$$

Hasil diatas dihasilkan melalui langkah-langkah berikut :

Diketahui varian kernel $T(K) = \frac{1}{5}$

$$\delta_E^* = \left\{ \frac{\|K\|_2^2}{\mu_2^2 (K)} \right\}^{1/5}$$

$$\|K\|_2^2 = \frac{225}{256} \int_{-1}^{1} ((1 - u^2)^2)^2 \, du$$

$$= \frac{225}{256} \frac{256}{284} \int_{-1}^{1} (1 - 4u^2 + 6u^4 - 10u^6 + u^8) \, du$$

$$= \frac{225}{256} \frac{256}{284} \left[u - \frac{4}{3} u^3 + \frac{6}{5} u^5 - \frac{10}{7} u^7 + \frac{1}{9} u^9 \right]_{-1}^{1}$$

$$= \frac{225}{256} \frac{256}{284} \frac{284}{315}$$

$$= \frac{5}{7}$$

$$\mu_2(K) = \frac{15}{16} \int_{-1}^{1} u^2(1 - u^2)^2 \, du$$

$$= \frac{15}{16} \int_{-1}^{1} u^2(1 - 2u^2 + u^4) \, du$$

$$= \frac{15}{16} \int_{-1}^{1} (u^2 - 2u^4 + u^6) \, du$$

$$= \frac{15}{16} \left[\frac{1}{3} u^3 - \frac{2}{5} u^5 + \frac{1}{7} u^7 \right]_{-1}^{1}$$

$$= \frac{15}{16} \quad \frac{16}{105}$$

$$= \frac{1}{7}$$

$$\delta_Q^* = \left\{ \frac{\|K\|_2^2}{\mu_2^2(K)} \right\}^{1/5}$$

$$= \left\{ \frac{5/7}{(1/7)^2} \right\}^{1/5}$$

$$= 35^{1/5}$$

$h' = \delta^* h$ dengan bandwidth $h = 1$ didapat $h' = 35^{1/5}$.

Telah diketahui bahwa fungsi kernel untuk tipe Quartic adalah

$$K_h(u) = \frac{1}{h} \, K\left(\frac{u}{h}\right)$$

$$= \frac{15}{16h} \left(1 - \left(\frac{u}{h}\right)^2\right)^2 I(\,|u| \leq h\,)$$

dengan substitusi h = h' didapat A-MISE minimal untuk tipe kernel Quartic :

$$K(u) = \frac{15}{16}\left(\frac{1}{35^{1/5}}\right) \left(1 - \left(\frac{u}{35^{1/5}}\right)^2\right)^2 I\left(|u| \leq 35^{1/5}\right)$$