BAB II
LANDASAN TEORI

2.1 Himpunan

Definisi 2.1 (Munir, 2010)

Himpunan adalah sekumpulan objek yang berbeda tetapi memiliki sifat yang dapat didefinisikan dengan jelas.

Objek-objek yang terdapat di dalam suatu himpunan disebut sebagai elemen atau anggota himpunan. Himpunan dinotasikan dengan huruf kapital, seperti $A, B, C,$ dan sebagainya. Sedangkan elemen atau anggota himpunan dinotasikan dengan huruf kecil, seperti $a, b, c,$ dan sebagainya. Elemen atau anggota suatu himpunan ditulis di antara kurung kurawal dan dipisahkan dengan tanda koma. Misalkan terdapat suatu himpunan S dengan anggota himpunan a dan b, maka himpunan S dapat ditulis sebagai $S = \{a, b\}$ dan elemen-elemennya dinotasikan sebagai $a, b \in S$ dibaca a dan b elemen atau anggota dari himpunan S.

2.2 Fungsi

Definisi 2.2 (Varberg dkk., 2010)

Sebuah fungsi f adalah suatu aturan korespondensi yang menghubungkan setiap obyek x dalam satu himpunan, yang disebut daerah asal (*domain*), dengan sebuah nilai tunggal $f(x)$ dari suatu himpunan kedua. Himpunan nilai yang diperoleh secara demikian disebut daerah hasil fungsi (*range*).

Suatu fungsi dapat dinotasikan dengan sebuah huruf tunggal seperti f (atau g atau F). $f(x)$ dibaca "f dari x" atau "f pada x", menunjukkan nilai yang diberikan oleh f kepada x. Jadi, jika $f(x) = x^3 - 4$, maka

$$\begin{aligned} f(2) &= 2^3 - 4 = 4 \\ f(a) &= a^3 - 4 \\ f(a+h) &= (a+h)^3 - 4 = a^3 + 3ah^2 + 3a^2h + h^3 - 4 \end{aligned}$$

untuk sebuah persamaan berbentuk $y = f(x)$, x disebut variabel bebas dan y disebut variabel terikat.

Definisi 2.3 (Varberg dkk., 2010)

Misalkan a bilangan positif bukan 1, Maka

$$y = \log_a x \Leftrightarrow x = a^y$$

Sifat-sifat fungsi logaritma umum :

1. $\log_e x = \ln x$
2. $\log_a x = \frac{\ln x}{\ln a}$

Definisi 2.4 (Varberg dkk., 2010)

Fungsi logaritma alami dinyatakan oleh ln, didefinisikan oleh

$$\ln x = \int_1^x \frac{1}{t} dt, x > 0$$

daerah asal fungsi logaritma alami adalah himpunan bilangan real positif.

Sifat-sifat fungsi logaritma alami :

Jika a dan b bilangan-bilangan positif dan r sebuah bilangan rasional, maka

1. $\ln 1 = 0$
2. $\ln \frac{a}{b} = \ln a - \ln b$
3. $\ln ab = \ln a + \ln b$
4. $\ln a^r = r \ln a$

Definisi 2.5 (Varberg dkk., 2010)

Untuk $a > 0$ dan bilangan real bilangan x,

$$a^x = e^{x \ln a}$$

Sifat-sifat fungsi eksponen umum :

Jika $a > 0, b > 0$, dan x dan y bilangan real, maka

1. $a^x a^y = a^{x+y}$

2. $(a^x)^y = a^{xy}$
3. $\left(\frac{a}{b}\right)^x = \frac{a^x}{b^x}$
4. $\frac{a^x}{a^y} = a^{x-y}$
5. $(ab)^x = a^x b^x$

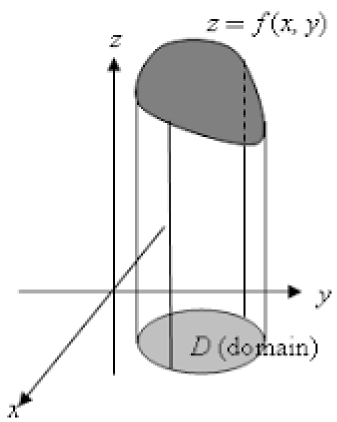
Definisi 2.6 (Varberg dkk., 2010)
Invers ln disebut fungsi eksponen alami dan dinyatakan oleh exp, jadi

$$x = \exp y \Leftrightarrow y = \ln x$$

Sifat-sifat fungsi eksponen alami:
1. Huruf e menyatakan bilangan real positif unik demikian sehingga $\ln e = 1$
2. Untuk semua nilai x (rasional maupun irrasional), $e^x = \exp x$
3. $e^{\ln x} = x, x > 0$
4. $\ln(e^y) = y$, untuk semua y

2.3 Fungsi Dua Variabel

Misalkan suatu fungsi f merupakan fungsi dengan dua variabel, maka fungsi f memadankan setiap pasangan terurut (x, y) pada suatu himpunan D dari bidang dengan bilangan real (tunggal) $f(x, y)$. Himpunan D disebut daerah asal fungsi. (Varberg dkk., 2011)



Gambar 2.1 Fungsi Dua Variabel

Grafik dari fungsi f dengan dua variabel atau grafik persamaan $z = f(x, y)$. Biasanya grafik berupa suatu permukaan seperti Gambar 2.1, karena setiap (x, y) didaerah asal hanya berpadanan dengan satu nilai z, maka setiap garis tegaklurus bidang-xy memotong permukaan pada paling banyak satu titik.

Contoh 2.1

Misalkan $f(x, y) = x^2 + 3y^2$. Tentukan nilai $f(-1,4)$.

Penyelesaian

$$f(-1,4) = (-1)^2 + 3(4)^2 = 1 + 3(16) = 49$$

2.4 Limit

Definisi 2.7 (Varberg dkk., 2010)

Untuk mengatakan bahwa $\lim_{x \to c} f(x) = L$, berarti bahwa ketika x dekat tetapi berlainan dari c, maka $f(x)$ dekat ke L.

Contoh 2.2

Carilah $\lim_{x \to 3} (4x - 5)$

Penyelesaian

Ketika x dekat 3; maka $4x - 5$ dekat terhadap $4 \times 3 - 5 = 7$. Dapat ditulis

$$\lim_{x \to 3} (4x - 5) = 7$$

Definisi 2.8 (Varberg dkk., 2010)

Untuk mengatakan bahwa $\lim_{x \to c^+} f(x) = L$ berarti bahwa ketika x dekat tetapi pada sebelah kanan c, maka $f(x)$ dekat ke-L. Demikian pula, untuk mengatakan bahwa $\lim_{x \to c^-} f(x) = L$ berarti bahwa ketika x dekat tetapi pada sebelah kiri c, maka $f(x)$ adalah dekat ke-L.

2.5 Turunan

Definisi 2.9 (Varberg dkk., 2010)

Turunan fungsi f adalah fungsi lain f' (dibaca "f aksen") yang nilainya pada sebarang bilangan c adalah

$$f'(c) = \lim_{h \to 0} \frac{f(c+h) - f(c)}{h}$$

Asalkan limit ini ada dan bukan ∞ atau $-\infty$.

Contoh 2.3

Misalkan $f(x) = 13x - 6$. Carilah $f'(4)$.

Penyelesaian

$$\begin{aligned} f'(4) &= \lim_{h \to 0} \frac{f(4+h) - f(4)}{h} \\ &= \lim_{h \to 0} \frac{13(4+h) - 6 - (13(4) - 6)}{h} \\ &= \lim_{h \to 0} \frac{13h}{h} \\ &= \lim_{h \to 0} 13 = 13 \end{aligned}$$

Sifat-sifat turunan:

1. Jika $f(x) = k$, dengan k suatu konstanta maka untuk sebarang x, $f'(x) = 0$
2. Jika $f(x) = x$, maka $f'(x) = 1$
3. Jika $f(x) = x^n$, dengan n bilangan bulat positif, maka $f'(x) = nx^{n-1}$

Jika k suatu konstanta dan fungsi f dan g adalah fungsi-fungsi yang terdiferensiasikan, maka

4. $(kf)'(x) = k \cdot f'(x)$
5. $(f+g)'(x) = f'(x) + g'(x)$
6. $(f-g)'(x) = f'(x) - g'(x)$
7. $(f \cdot g)'(x) = f(x)g'(x) + g(x)f'(x)$

8. Jika f dan g adalah fungsi-fungsi yang terdifierensiasikan dengan $g(x) \neq 0$, maka $\left(\frac{f}{g}\right)'(x) = \frac{g(x)f'(x)-f(x)g'(x)}{g^2(x)}$

2.6 Turunan Parsial

Misalkan f adalah suatu fungsi dua variabel x dan y. Jika y dijaga agar tetap konstan, asumsikan $y = y_0$, maka $f(x, y_0)$ adalah fungsi satu variabel x. Turunan dari f di $x = x_0$ disebut turunan parsial f terhadap x di (x_0, y_0) dan dinotasikan sebagai $f_x(x_0, y_0)$. (Varberg dkk., 2011) Jadi,

$$f_x(x_0, y_0) = \lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x, y_0) - f(x_0, y_0)}{\Delta x}$$

Dengan cara serupa, turunan parsial f terhadap y di (x_0, y_0) dinotasikan sebagai $f_y(x_0, y_0)$ dengan

$$f_y(x_0, y_0) = \lim_{\Delta y \to 0} \frac{f(x_0, y_0 + \Delta y) - f(x_0, y_0)}{\Delta y}$$

Contoh 2.4

Carilah $f_x(1,2)$ dan $f_y(1,2)$ jika $f(x, y) = x^2y + 3y^3$

Penyelesaian

Untuk mencari $f_x(x, y)$, maka asumsikan y sebagai suatu konstanta kemudian turunkan f terhadap x, sehingga diperoleh

$$f_x(x, y) = 2xy + 0$$

Untuk $(x, y) = (1,2)$, maka

$$f_x(1,2) = 2 \cdot 1 \cdot 2 = 4$$

Dengan cara yang serupa, diperoleh

$$f_y(x,y) = x^2 + 9y^2$$

Untuk $(x,y) = (1,2)$, maka

$$f_y(1,2) = 1^2 + 9 \cdot 2^2 = 37$$

2.7 Notasi Sigma

Definisi 2.10 (Varberg dkk., 2010)

Notasi sigma atau $\sum$ menyatakan penjumlahan semua bilangan berbentuk seperti yang ditunjukkan dengan *indeks i*, dimana *indeks i* terus meningkat dimulai dari bilangan bulat yang berada di bawah $\sum$ dan berakhir dengan bilangan bulat yang berada diatas $\sum$ seperti yang ditunjukkan sebagai berikut:

$$\sum_{i=1}^{5} a_i = a_1 + a_2 + a_3 + a_4 + a_5$$

Sifat-sifat sigma:

Jika c suatu konstanta, maka

1. $\sum_{i=1}^{n} c = nc$
2. $\sum_{i=1}^{n} ca_i = c\sum_{i=1}^{n} a_i$
3. $\sum_{i=1}^{n}(a_i + b_i) = \sum_{i=1}^{n} a_i + \sum_{i=1}^{n} b_i$
4. $\sum_{i=1}^{n}(a_i - b_i) = \sum_{i=1}^{n} a_i - \sum_{i=1}^{n} b_i$

2.8 Integral

Definisi 2.11 (Varberg dkk., 2010)

Kita sebut F suatu anti-turunan f pada interval I jika $F'(x) = f(x)$ untuk semua x dalam I.

Contoh 2.5

Carilah anti-turunan umum dari $f(x) = x^2$ pada $(-\infty, \infty)$.

Penyelesaian

Akan dicari suatu fungsi F yang memenuhi $F'(x) = x^2$ untuk semua x real. Berdasarkan diferensiasi, fungsi $F(x) = x^3$ tidak akan memenuhi, karena turunan dari $F(x)$ adalah $3x^2$. Tetapi nilai $F'(x) = 3x^2$ merujuk pada $F(x) = \frac{1}{3}x^3$, karena turunan dari $F(x) = \frac{1}{3}x^2$ memenuhi $F'(x) = \frac{1}{3} \cdot 3x^2 = x^2$. Sehingga diperoleh anti-turunan umumnya yaitu $\frac{1}{3}x^3 + c$.

Sifat-sifat integral:

Misalkan f dan g mempunyai anti-turunan (integral tak tentu) dan misalkan k suatu konstanta. Maka

1. $\int kf(x)\, dx = k \int f(x) dx$
2. $\int [f(x) + g(x)] dx = \int f(x) dx + \int g(x) dx$
3. $\int [f(x) - g(x)] dx = \int f(x) dx - \int g(x)\, dx$

Definisi 2.12 (Varberg dkk., 2010)

Misalkan f suatu fungsi yang didefinisikan pada interval tertutup $[a, b]$. Jika

$$\lim_{||p|| \to 0} \sum_{i=1}^{n} f(\bar{x}_i) \Delta x_i$$

ada, asumsikan f adalah fungsi yang terintegrasikan pada $[a, b]$. Lebih lanjut $\int_a^b f(x) dx$, disebut integral tentu f dari a ke b, ditulis sebagai

$$\int_a^b f(x) dx = \lim_{||p|| \to 0} \sum_{i=1}^{n} f(\bar{x}_i) \Delta x_i$$

Sifat-sifat integral tentu:

1. $\int_a^a f(x) dx = 0$
2. $\int_a^b f(x) dx = -\int_b^a f(x) dx\, , a > b$

3. $\int_a^c f(x)dx = \int_a^b f(x)dx + \int_b^c f(x)dx$

Misalkan f dan g terintegrasikan pada $[a,b]$ dan k adalah konstanta. Maka

4. $\int_a^b kf(x)dx = k\int_a^b f(x)dx$
5. $\int_a^b [f(x)+g(x)]dx = \int_a^b f(x)dx + \int_a^b g(x)dx$
6. $\int_a^b [f(x)-g(x)]dx = \int_a^b f(x)dx - \int_a^b g(x)dx$

2.9 Peluang dan Ruang Sampel

Definisi 2.14 (Walpole d Myers, 1995)

Himpunan semua kemungkinan hasil yang memungkinkan dari suatu percobaan statistika disebut ruang sampel dan dilambangkan dengan huruf S.

Setiap kemungkinan hasil dalam suatu ruang sampel disebut unsur atau anggota ruang sampel tersebut atau biasanya disebut dengan titik sampel. Jika banyaknya unsur ruang sampel berhinggass, maka unsur-unsur tersebut dapat didaftarkan dalam bentuk himpunan. Jadi ruang sampel S dari percobaan pelemparan sekeping uang logam yaitu G (gambar) dan A (angka), atau dapat ditulis sebagai

$$S = \{G, A\}$$

Definisi 2.15 (Walpole, 2005)

Suatu kejadian adalah himpunan bagian dari ruang sampel.

Definisi 2.16 (Walpole & Myers, 1995)

Peluang suatu kejadian A adalah jumlah peluang semua titik sampel dalam A. Dimana,

$$0 \leq P(A) \leq 1, P(\emptyset) = 0, P(S) = 1.$$

Teorema 2.1 (Walpole, 2005)

Jika banyaknya ruang sampel dinotasikan dengan $n(S)$ dan banyaknya kejadian dinotasikan dengan $n(A)$. Maka peluang kejadian A adalah

$$P(A) = \frac{n(A)}{n(S)}$$

2.10 Peubah Acak

Definisi 2.17 (Walpole, 2005)

Suatu fungsi yang nilainya berupa bilangan riil yang ditentukan oleh setiap unsur dalam ruang sampel disebut peubah acak.

Definisi 2.18 (Walpole, 2005)

Bila suatu ruang sampel mengandung jumlah titik sampel yang berhingga atau suatu barisan unsur yang tidak pernah berakhir tetapi sama banyaknya dengan bilangan cacah, maka ruang itu disebut ruang sampel diskret.

Definisi 2.19 (Walpole, 2005)

Bila suatu ruang sampel mengandung tak hingga banyaknya titik sampel yang sama dengan banyaknya dengan banyaknya titik pada sebuah ruas garis, maka ruang itu disebut ruang sampel kontinu.

2.11 Populasi dan Sampel

Definisi 2.20 (Walpole, 2005)

Populasi adalah keseluruhan pengamatan yang menjadi perhatian kita.

Banyaknya pengamatan atau anggota suatu populasi disebut ukuran populasi.

Definisi 2.21 (Walpole, 2005)

Sampel adalah suatu himpunan bagian dari populasi.

2.12 Distribusi Probabilitas

Definisi 2.22 (Walpole dan Myers, 1995)

Misalkan X adalah peubah acak diskret, maka fungsi $f(x)$ disebut dengan distribusi probabilitas atau fungsi kepadatan peluang dari sebuah peubah acak X, yang mempunyai syarat-syarat sebagai berikut:

1. $f(x) \geq 0$

2. $\sum_x f(x) = 1$
3. $P(X = x) = f(x)$

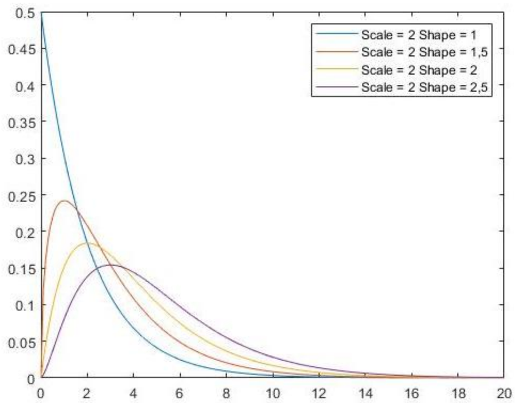
Fungsi distribusi kumulatif dari X dinotasikan dengan $F(x)$, dan didefinisikan sebagai berikut:

$$F(x) = P(X \leq x) = \sum_{t \leq x} f(t), -\infty < x < \infty$$

Definisi 2.23 (Walpole dan Myers, 1995)
Misalkan fungsi $f(x)$ adalah fungsi kepadatan peluang dari peubah acak kontinu X, yang didefinisikan pada himpunan semua bilangan real, jika
1. $f(x) \geq 0$ untuk semua $x \in (-\infty, \infty)$
2. $\int_{-\infty}^{\infty} f(x)\, dx = 1$
3. $P(a < X < b) = \int_a^b f(x)\, dx$

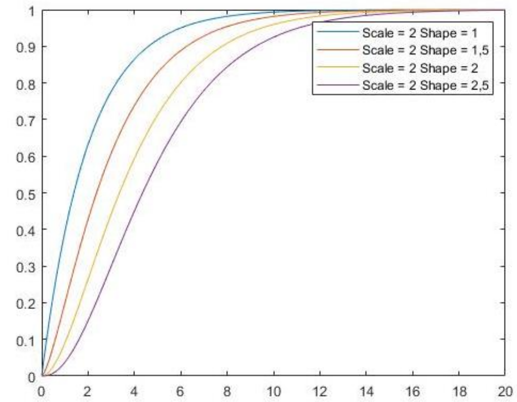
Fungsi distribusi kumulatif dari X dinotasikan dengan $F(x)$, dan didefinisikan sebagai berikut:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt, -\infty < x < \infty$$



Gambar 2.2 Grafik Fungsi Kepadatan Peluang Distribusi Gamma

Gambar 2.2 menunjukkan grafik fungsi kepadatan peluang distribusi Gamma dengan nilai parameter skala konstan pada 2 dan nilai parameter bentuk bergerak dari 1 sampai 2.5.



Gambar 2.3 Grafik Fungsi Distribusi Kumulatif Distribusi Gamma

Gambar 2.3 menunjukkan grafik fungsi distribusi kumulatif distribusi Gamma dengan nilai parameter skala konstan pada 2 dan nilai parameter bentuk bergerak dari 1 sampai 2.5.

2.13 Beberapa Distribusi Kontinu

2.13.1 Distribusi Erlang

Definisi 2.24 (Thomopoulos, 2017)

Sebuah peubah acak X dikatakan berdistribusi Erlang (α, β), jika fungsi kepadatan peluangnya berbentuk:

$$f(x) = \frac{1}{\beta^{\alpha}(\alpha - 1)!} x^{\alpha-1} \exp\left(-\left(\frac{x}{\beta}\right)\right), x \geq 0 \, , \beta > 0 \, , \alpha \in \mathbb{Z} \qquad (2.1)$$

dan fungsi distribusi kumulatif distribusi Erlang (α, β) berbentuk :

$$F(x) = \frac{1}{\beta^{m}(\mathrm{m} - 1)!} \int_{0}^{x} t^{m-1} e^{-\frac{t}{\beta}} dt$$

dengan

α = parameter bentuk

β = parameter skala

2.13.2 Distribusi Fatigue Life

Definisi 2.25

Sebuah peubah acak X dikatakan berdistribusi Fatigue Life (α, β), jika fungsi kepadatan peluangnya berbentuk:

$$f(x) = \frac{\sqrt{\frac{x}{\beta}} + \sqrt{\frac{\beta}{x}}}{2\alpha x} \phi\left(\frac{1}{\alpha}\left(\sqrt{\frac{x}{\beta}} - \sqrt{\frac{\beta}{x}}\right)\right), x > 0, \alpha > 0, \beta > 0 \qquad (2.2)$$

dan fungsi distribusi kumulatif distribusi Fatigue Life (α, β) berbentuk :

$$F(x) = \Phi\left(\frac{1}{\alpha}\left(\sqrt{\frac{x}{\beta}} - \sqrt{\frac{\beta}{x}}\right)\right)$$

dengan

α = parameter bentuk

β = parameter skala

ϕ = fungsi kepadatan peluang distribusi Normal standard

Φ = fungsi distribusi kumulatif distribusi Normal standard

Distribusi normal standard adalah distribusi Normal dengan nilai parameter $\mu = 0$ dan parameter $\sigma = 1$.

2.13.3 Distribusi Frechet

Definisi 2.26

Sebuah peubah acak X dikatakan berdistribusi Frechet (α, β), jika fungsi kepadatan peluangnya berbentuk:

$$f(x) = \frac{\alpha}{\beta}\left(\frac{\beta}{x}\right)^{\alpha+1} \exp\left(-\left(\frac{\beta}{x}\right)^{\alpha}\right), x > 0, \alpha > 0, \beta > 0 \quad (2.3)$$

dan fungsi distribusi kumulatif distribusi Frechet (α, β) berbentuk :

$$F(x) = \exp\left(-\left(\frac{\beta}{x}\right)^{\alpha}\right)$$

dengan

α = parameter bentuk

β = parameter skala

2.13.4 Distribusi Gamma

Definisi 2.27 (Walpole dan Myers, 1995)

Sebuah peubah acak X dikatakan berdistribusi Gamma (α, β), jika fungsi kepadatan peluangnya berbentuk:

$$f(x) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} \exp\left(-\left(\frac{x}{\beta}\right)\right), x > 0, \alpha > 0, \beta > 0 \quad (2.4)$$

dan fungsi distribusi kumulatif distribusi Gamma (α, β) berbentuk :

$$F(x) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} \int_0^x t^{\alpha-1} e^{-\frac{t}{\beta}} dt$$

dengan

$\Gamma(\alpha)$ = fungsi gamma

α = parameter bentuk

β = parameter skala

Definisi 2.28 (Walpole dan Myers, 1995)

Fungsi gamma didefinisikan sebagai

$$\Gamma(\alpha) = \int_0^\infty e^{-x} x^{\alpha-1}\, dx \;, \alpha > 0$$

2.13.5 Distribusi Log-Logistik

Definisi 2.29

Sebuah peubah acak X dikatakan berdistribusi Log-Logistik (α, β), jika fungsi kepadatan peluangnya berbentuk:

$$f(x) = \left(\frac{\alpha}{\beta}\right)\left(\frac{x}{\beta}\right)^{\alpha-1}\left(1 + \left(\frac{x}{\beta}\right)^{\alpha}\right)^{-2} \;, x \geq 0 \;, \alpha > 0 \;, \beta > 0 \qquad (2.5)$$

dan fungsi distribusi kumulatif distribusi Log-Logistik (α, β) berbentuk :

$$F(x) = \left(1 + \left(\frac{\beta}{x}\right)^{\alpha}\right)^{-1}$$

dengan

α = parameter bentuk

β = parameter skala

2.13.6 Distribusi Pareto

Definisi 2.30

Sebuah peubah acak X dkatakan berdistribusi Pareto (α, β), jika fungsi kepadatan peluangnya berbentuk:

$$f(x) = \frac{\alpha\beta^{\alpha}}{x^{\alpha+1}} \;, \beta \leq x < +\infty \qquad (2.6)$$

dan fungsi distribusi kumulatif distribusi Pareto (α, β) berbentuk :

$$F(x) = 1 - \left(\frac{\beta}{x}\right)^{\alpha}$$

dengan

α = parameter bentuk $(\alpha > 0)$

β = parameter skala $(\beta > 0)$

2.13.7 Distribusi Pearson Tipe 5

Definisi 2.31

Sebuah peubah acak X dikatakan berdistribusi Pearson Tipe V (α, β), jika fungsi kepadatan peluangnya berbentuk:

$$f(x) = \frac{\exp\left(-\frac{\beta}{x}\right)}{\beta\Gamma(\alpha)\left(\frac{x}{\beta}\right)^{\alpha+1}}, x > 0 \tag{2.7}$$

dan fungsi distribusi kumulatif distribusi Pearson Tipe V (α, β) berbentuk :

$$F(x) = \frac{\Gamma_{\frac{\beta}{x}}(\alpha)}{\Gamma(\alpha)}$$

dengan

α = parameter bentuk $(\alpha > 0)$

β = parameter skala $(\beta > 0)$

2.13.8 Distribusi Weibull

Definisi 2.32 (Walpole dan Myers, 1995)

Sebuah peubah acak X dikatakan berdistribusi Weibull (α, β), jika fungsi kepadatan peluangnya berbentuk:

$$f(x) = \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1} \exp\left(-\left(\frac{x}{\beta}\right)^{\alpha}\right), x > 0, \alpha > 0, \beta > 0 \tag{2.8}$$

dan fungsi distribusi kumulatif distribusi Weibull (α, β) berbentuk :

$$F(x) = 1 - \exp\left(-\left(\frac{x}{\beta}\right)^{\alpha}\right)$$

dengan

α = parameter bentuk

β = parameter skala

2.14 Parameter

Definisi 2.33 (Walpole, 2005)

Sebarang nilai yang menjelaskan ciri populasi disebut parameter.

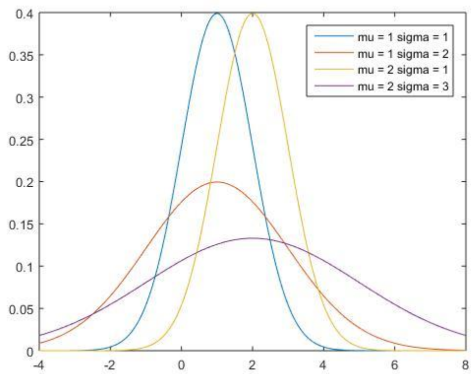
Secara umum parameter dilambangkan dengan huruf yunani dan merupakan suatu konstanta yang menjelaskan populasi. Terdapat beberapa jenis parameter yang biasa digunakan dalam ilmu peluang, seperti mean, standard deviasi, parameter skala, parameter bentuk, parameter lokasi, dan sebagainya.

Parameter skala dan parameter bentuk adalah jenis khusus parameter numerik dari kelurga parametrik dimana parameter skala menunjukkan besarnya jangkauan data. Semakin besar nilai parameter skala maka distribusi data akan semakin menyebar begitu pun sebaliknya. Berbeda dengan parameter skala, parameter bentuk menunjukkan bentuk sebaran data. Semakin besar nilai parameter bentuk maka data akan cenderung menyebar pada suatu interval tertentu begitu pun sebaliknya.

Misalkan sebuah peubah acak X dikatakan berdistribusi Normal dengan parameter μ dan σ, maka fungsi kepadatan peluangnya berbentuk:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{\left(\frac{1}{2}\right)\left(\frac{x-\mu}{\sigma}\right)^2};\ -\infty < x < \infty$$

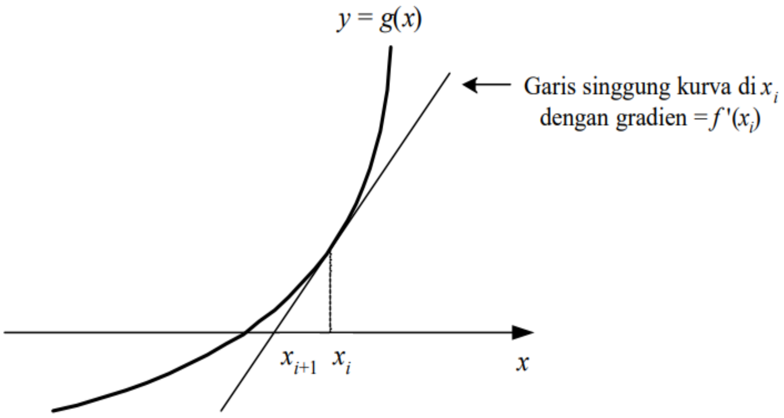
dengan parameter lokasi yaitu mean atau yang biasa ditulis dengan μ, dan paramerter skala yaitu standard deviasi atau yang biasa ditulis dengan σ. Parameter μ dari distribusi Normal menunjukkan letak puncak dari grafik fungsi kepadatan peluang. Sedangkan parameter σ dari distribusi Normal menunjukkan bentuk grafik fungsi kepadatan peluang. Semakin besar parameter σ, maka grafik yang dihasilkan akan semakin pendek dan melebar, begitu pun sebaliknya. Seperti yang terlihat pada ambar 2.2. Ketika nilai parameter $\mu = 1$, maka puncak grafik berada pada $x = 1$. Sedangkan ketika nilai parameter $\mu = 2$, maka puncak grafik berada pada $x = 2$. Kemudian berdasarkan gambar 2.2 terlihat bahwa ketika nilai parameter σ, bergerak dari 1 ke 3, maka grafik yang dihasilkan semakin pendek dan melebar.



Gambar 2.4 Grafik Fungsi Kepadatan Peluang Distribusi Normal

2.15 Metode Newton-Raphson

Metode newton-raphson adalah salah satu metode pendekatan dalam mencari akar-akar sebuah persamaan. Terdapat dua jenis pendekatan untuk menurunkan rumus newton-raphson, yaitu secara geometri dan dengan bantuan deret taylor. Penurunan rumus newton-raphson secara geometri sebagai berikut:



Gambar 2.5 Penurunan Newton-Raphson

Dari gambar diatas gradien garis singgung di x_i adalah

$$m = f'(x_i) = \frac{\Delta y}{\Delta x} = \frac{f(x_i) - 0}{x_i - x_{i+1}}$$

atau

$$f'(x_i) = \frac{f(x_i)}{x_i - x_{i+1}}$$

Sehingga diperoleh rumus newton-raphson sebagai berikut:

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)} \tag{2.9}$$

Misalkan x_i adalah nilai hampiran terhadap nilai sejati x_{i+1}, maka selisih dari x_{i+1} dan x_i disebut galat dan biasanya dinotasikan dengan ε. Jika tanda galat (positif atau negatif) tidak dipertimbangkan, maka disebut galat mutlak dan didefinisikan sebagai berikut :

$$\varepsilon = |x_{i+1} - x_i|$$

dan galat relatif didefinisikan sebagai berikut:

$$\varepsilon = \frac{|x_{i+1} - x_i|}{|x_{i+1}|}$$

Menurut Epperson (2013)

Langkah-langkah mencari akar suatu fungsi dengan metode newton-raphson adalah:

1. Menentukan nilai x_0,
2. Mentukan turunan pertama dari $f(x)$,
3. Mentukan nilai x_{i+1} dengan menggunakan rumus newton-raphson,
4. Dilakukan iterasi sebanyak i kali hingga nilai galat relatif hampiran kurang dari galat relatif yang sudah ditentukan,
5. Maka pada iterasi ke-i, diperoleh x_{i+1} akar persamaan dari $f(x)$.

Contoh 2.6

Tentukan salah satu akar persamaan linier $x^5 + 2x^2 - 4 = 0$ dengan metode newton-raphson dengan kesalahan relatif hampiran sebesar 0.001.

Penyelesaian

Diketahui $f(x) = x^5 + 2x^2 - 4$ dan $\varepsilon = 0.001$.

Maka $f'(x) = 5x^4 + 4x$

Pilih nilai awal $x_0 = 1$, maka diperoleh

$$f(1) = 1^5 + 2(1)^2 - 4 = 1 + 2 - 4 = -1$$

Dan,

$$f'(1) = 5(1)^4 + 4(1) = 5 + 4 = 9$$

Iterasi ke-1

Hitung hampiran akar pertama.

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)} = 1 - \frac{f(1)}{f'(1)} = 1 + \frac{1}{9} = \frac{10}{9} = 1.111$$

Kemudian hitung kesalahan relatif ε_1

$$\varepsilon_1 = \left|\frac{x_1 - x_0}{x_1}\right| = \left|\frac{1.111 - 1}{1.111}\right| = \left|\frac{0.111}{1.111}\right| = 0.1$$

Karena kesalahan relatif $\varepsilon_1 = 0.1 > 0.001$, maka perhitungan dilanjutkan ke Iterasi ke-2.

Iterasi ke-2

Hitung hampiran akar selanjutnya x_2

$$f(x_1) = f(1.111) = (1.111)^5 + 2(1.111)^2 - 4 = 0.16265$$
$$f'(x_1) = f'(1.111) = 5(1.111)^4 + 4(1.111) = 12.065$$
$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)} = 1.111 - \frac{f(1.111)}{f'(1.111)} = 1.111 - \frac{0.16265}{12.065} = 1.09763$$

Kemudian hitung kesalahan relatif ε_2

$$\varepsilon_2 = \left|\frac{x_2 - x_1}{x_2}\right| = \left|\frac{1.09763 - 1.111}{1.09763}\right| = 0.01228$$

Karena kesalahan relatif $\varepsilon_2 = 0.01228 > 0.001$, maka perhitungan dilanjutkan ke Iterasi ke-3.

Iterasi ke-3

Hitung hampiran akar selanjutnya x_3

$$f(x_2) = f(1.09763) = (1.09763)^5 + 2(1.09763)^2 - 4 = 0.00283$$

$$f'(x_2) = f'(1.09763) = 5(1.09763)^4 + 4(1.09763) = 11.6482$$

$$x_3 = x_2 - \frac{f(x_2)}{f'(x_2)} = 1.09763 - \frac{f(1.09763)}{f'(1.09763)} = 1.09763 - \frac{0.00283}{11.6482}$$

$$x_3 = 1.09739$$

Kemudian hitung kesalahan relatif ε_3

$$\varepsilon_3 = \left|\frac{x_3 - x_2}{x_3}\right| = \left|\frac{1.09739 - 1.09763}{1.09739}\right| = 0.000221$$

Karena kesalahan relatif $\varepsilon_3 = 0.000221 < 0.001$, maka perhitungan dihentikan pada Iterasi ke-3. Diperoleh salah satu akar dari persamaan $x^5 + 2x^2 - 4 = 0$ adalah $x = 1.09739$.

2.16 Metode Estimasi Kemungkinan Maksimum

Metode Estimasi Kemungkinan Maksimum adalah teknik pendugaan parameter dengan memaksimumkan fungsi *Likelihood*-nya. Misalkan pada suatu kejadian di mana X hanya dapat mengambil sebagian nilai terbilang $x_1, x_2, \ldots$, dengan $P_\theta(x) = P_\theta\{X = x\}$, dan ingin menentukan nilai θ terbaik, di mana nantinya nilai θ akan digunakan untuk menghasilkan nilai x yang diamati. Hal ini menyarankan untuk mempertimbangkan untuk setiap nilai θ yang memungkinkan, berapa besar kemungkinan terjadinya x yang diamati, jika θ merupakan nilai sebenarnya. Semakin besar kemungkinannya, maka semakin banyak nilai θ yang terlibat ke dalam penjelasan yang menunjukkan bahwa θ yang dimaksud menghasilkan nilai x, dan semakin memungkinkan munculnya nilai θ. Maka dari itu, P_θ yang digunakan untuk nilai X yang tetap sebagai fungsi dari θ disebut sebagai kemungkinan dari θ yang kemudian dinotasikan sebagai $L(\theta)$ (Lehmann & Romano, 2005). Jika $X_1, X_2, \ldots, X_n$ adalah sampel random dari $f(x;\theta)$ maka

$$L(\theta) = \prod_{i=1}^{n} f(x_i|\theta) \tag{2.10}$$

Untuk mencari nilai MLE-nya, akan lebih mudah jika fungsi $L(\theta)$ diselesaikan dalam bentuk logaritma alami. Hal ini dimungkinkan karena fungsi logaritma alami naik monoton pada $(0, \infty)$, yang berarti fungsi $L(\theta)$ dan fungsi logaritma alami mempunyai ekstrem yang sama. Maka didefinisikan fungsi *log-likelihood* sebagai berikut:

$$l(\theta) = \ln\big(L(\theta)\big) \tag{2.11}$$

Langkah-langkah dalam melakukan estimasi parameter dengan Metode Estimasi Kemungkinan Maksimum sebagai berikut (Olofsson & Andersson, 2012):

1. Tentukan fungsi *likelihood*nya $\big(L(\theta)\big)$
2. Bentuk $L(\theta)$ ke dalam bentuk *log- likelihood* $\big(l(\theta)\big)$
3. Tentukan turunan dari $l(\theta)$ terhadap θ, kemudian tetapkan $\frac{\partial l(\theta)}{\partial \theta} = 0$

Contoh 2.7

Misalkan $X_1, X_2, \dots, X_n$ adalah sampel random dari distribusi poisson dengan parameter λ. Tentukan nilai estimasi λ dengan metode estimasi kemungkinan maksimum.

Penyelesaian

Diketahui fungsi kepadatan peluang dari distribusi poisson sebagai berikut:

$$f(x) = \frac{e^{-\lambda}\lambda^{x}}{x!} \; ; \; x = 0,1,2, \dots \; ; \; \lambda > 0$$

Kemudian berdasarkan Persamaan (2.10) diperoleh fungsi *likelihoodnya* sebagai berikut:

$$L(\lambda) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}$$

$$L(\lambda) = \frac{e^{-n\lambda}\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}$$

Kemudian berdasarkan Persamaan (2.11) diperoleh bentuk fungsi *log-likelihood*nya sebagai berikut:

$$l(\lambda) = \ln\left[\frac{e^{-n\lambda}\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}\right]$$

$$= \ln e^{-n\lambda} + \ln \lambda^{\sum_{i=1}^{n} x_i} - \ln \prod_{i=1}^{n} x_i!$$

$$l(\lambda) = -n\lambda + \sum_{i=1}^{n} x_i \ln \lambda - \sum_{i=1}^{n} \ln x_i!$$

Kemudian $l(\lambda)$ akan diturunkan secara parsial terhadap λ dan dijadikan sama dengan nol.

$$\frac{\partial l(\lambda)}{\partial \lambda} = -n + \frac{\sum_{i=1}^{n} x_i}{\lambda} = 0$$

$$n = \frac{\sum_{i=1}^{n} x_i}{\lambda}$$

$$\hat{\lambda} = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{X}_n \tag{2.12}$$

Jadi diperoleh nilai estimasi $\hat{\lambda} = \bar{X}_n$.

Misalkan terdapat data acak sebagai berikut:

Tabel 2.1 Data Acak

Data
9
1
10
6
1

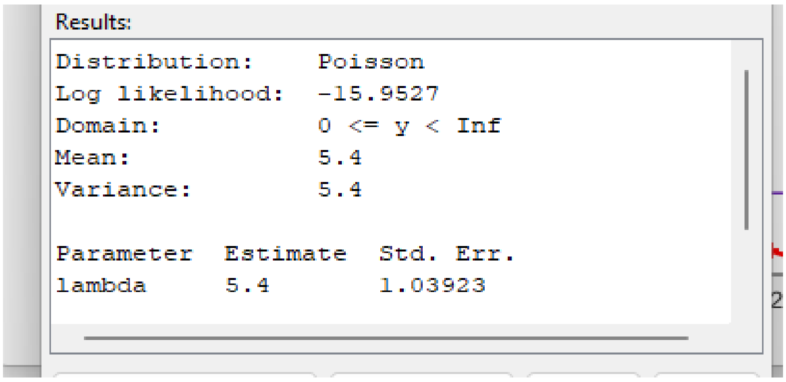
Kemudian akan dicari nilai λ. Berdasarkan Persamaan (2.12), maka diperoleh nilai λ sebesar 5.4.

$$\hat{\lambda} = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{X}_n$$

$$\hat{\lambda} = \frac{9+1+10+6+1}{5} = \frac{27}{5} = 5.4$$

Validasi dengan Matlab

```
parameter = fitdist(data,'poisson')
[h,p,kstat,cv]=kstest(data,'CDF',parameter)
```



Gambar 2.6 Estimasi Parameter dengan Matlab

Dengan menggunakan Matlab, diperoleh nilai $\hat{\lambda}$ sebesar 5.4, sama dengan perhitungan manual.

2.17 Supremum dan Infimum

Definisi 2.34 (Bartle, 2011)

Misalkan S merupakan suatu himpunan tak kosong dan merupakan himpunan bagian dari $\mathbb{R}$.

1. Hipunan S dikatakan terbatas keatas jika terdapat suatu nilai $u \in \mathbb{R}$, sedemikian sehingga $s \leq u$ untuk semua $s \in S$. Setiap nilai u yang memungkinkan disebut sebagai batas atas himpunan S.
2. Hipunan S dikatakan terbatas kebawah jika terdapat suatu nilai $w \in \mathbb{R}$, sedemikian sehingga $w \leq s$ untuk semua $s \in S$. Setiap nilai w yang memungkinkan disebut sebagai batas bawah himpunan S.

3. Sebuah himpunan dikatakan terbatas jika himpunan tersebut terbatas keatas dan kebawah. Sebuah himpunan dikatakan tidak terbatas jika himpunan tersebut tidak terbatas keatas dan kebawah.

Definisi 2.35 (Bartle, 2011)

Misalkan S merupakan suatu himpunan bagian dari $\mathbb{R}$.

a) Jika S terbatas keatas, maka terdapat sebuah nilai u yang merupakan supremum (batas atas terkecil) dari S jika memenuhi :
   1. u adalah batas atas dari S, dan
   2. Jika v merupakan batas atas sebarang dari S, maka $u \leq v$.

b) Jika S terbatas kebawah, maka terdapat sebuah nilai w yang merupakan infimum (batas bawah terbesar) dari S jika memenuhi :
   1. w adalah batas bawah dari S, dan
   2. Jika t merupakan batas bawah sebarang dari S, maka $w \geq t$.

2.18 Uji Kolmogorov-Smirnov

Uji Kolmogorov-Smirnov atau yang biasa dikenal dengan uji KS adalah salah satu teknik uji hipotesis statistik terkenal, yang memeriksa apakah suatu sampel mengikuti sebaran dari distribusi probabilitas tertentu dengan membandingkan fungsi empiris dan fungsi distribusi kumulatif yang diasumsikan. Hipotesis yang akan diuji dalam penelitian ini adalah :

H_0 : Data mengikuti model sebaran distribusi tersebut.

H_1 : Data tidak mengikuti model sebaran distribusi tersebut.

dengan statistik uji yang digunakan adalah :

$$D_{hitung} = sup_x|F(x) - H(x)| \tag{2.13}$$

Di mana

D_{hitung} : Jarak vertikal terjauh antara $F(x) - H(x)$.

$F(x)$: Fungsi distribusi kumulatif distribusi yang dihipotesiskan.

$H(x)$: Fungsi Empiris.

dengan

$$H(x) = \frac{frekuensi\ kumulatif\ X \leq x}{n} \tag{2.14}$$

di mana n adalah jumlah data.

Dengan kriteria uji, jika nilai $D_{hitung} > D_{tabel}$ dengan tingkat error sebesar α, maka H_0 ditolak. Jika nilai $D_{hitung} < D_{tabel}$ dengan tingkat error sebesar α, maka H_0 diterima atau dengan kata lain data mengikuti model sebaran distribusi tersebut.

Contoh 2.8

Pada Contoh 2.7 telah dilakukan estimasi parameter terhadap data acak pada Tabel 2.1. Diperoleh nilai λ sebesar 5.4. Kemudian pada Contoh 2.8 ini akan dilakukan uji hipotesis data diatas yang diasumsikan berdistribusi poisson dengan uji kolmogorv-smirnov dengan hipotesis yang digunakan sebagai berikut :

H_0 : Data mengikuti model sebaran distribusi *Poisson*.

H_1 : Data tidak mengikuti model sebaran distribusi *Poisson*.

dan dengan taraf signifikan sebesar 0.05.

Penyelesaian

Tabel 2.2 Uji Kolmogorov-Smirnov Distribusi Poisson

Data	Frekuensi	Frekuensi Kumulatif	$F(x)$	$H(x)$	$\lvert F(x) - H(x) \rvert$
1	2	2	0.0289	0.4	0.371094
6	1	3	0.701671	0.6	0.101671
9	1	4	0.95125	0.8	0.151245
10	1	5	0.977486	1	0.022514

Penjelasan Tabel 2.2

Susun data dari terkecil ke terbesar, kemudian hitung frekuensi dan frekuensi kumulatif data. Lalu hitung nilai $F(x)$ dengan menggunakan fungsi distribusi kumulatif distribusi poisson.

Diketahui fungsi kepadatan peluang dari distribusi poisson sebagai berikut:

$$f(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}\ ;\ x = (0,1,2,\dots)\ ;\ \lambda > 0$$

dan fungsi distribusi kumulatif distribusi poisson sebagai berikut:

$$F(x) = \sum_{x=0}^{n} e^{-\lambda}\frac{\lambda^{x}}{x!} = e^{-\lambda}\sum_{i=0}^{x}\frac{\lambda^{i}}{i!}$$

Untuk $\lambda = 5.4$ diperoleh

$$F(x) = e^{-5.4}\sum_{i=0}^{x}\frac{(5.4)^{i}}{i!}$$

Kemudian akan dihitung nilai $F(x)$ untuk semua data.

Untuk $x = 1$

$$F(1) = e^{-5.4}\sum_{i=0}^{1}\frac{(5.4)^{i}}{i!}$$

$$F(1) = e^{-5.4}\left[\frac{(5.4)^{0}}{0!} + \frac{(5.4)^{1}}{1!}\right]$$

$$F(1) = 0.0045[1 + 5.4] = 0.0045[6.4] = 0.0289$$

Untuk $x = 6$

$$F(6) = e^{-5.4}\sum_{i=0}^{6}\frac{(5.4)^{i}}{i!}$$

$$F(6) = e^{-5.4}\left[\frac{(5.4)^{0}}{0!} + \frac{(5.4)^{1}}{1!} + \frac{(5.4)^{2}}{2!} + \frac{(5.4)^{3}}{3!} + \frac{(5.4)^{4}}{4!} + \frac{(5.4)^{5}}{5!} + \frac{(5.4)^{6}}{6!}\right]$$

$$F(6) = 0.0045[1 + 5.4 + 14.58 + 26.244 + 35.4294 + 38.2638 + 34.4374]$$

$$F(6) = 0.0045[155.3545] = 0.70167$$

Untuk $x = 9$

$$F(9) = e^{-5.4} \sum_{i=0}^{9} \frac{(5.4)^i}{i!}$$

$$F(9) = e^{-5.4} \left[\frac{(5.4)^0}{0!} + \frac{(5.4)^1}{1!} + \frac{(5.4)^2}{2!} + \frac{(5.4)^3}{3!} + \frac{(5.4)^4}{4!} + \frac{(5.4)^5}{5!} + \frac{(5.4)^6}{6!} + \frac{(5.4)^7}{7!} + \frac{(5.4)^8}{8!} + \frac{(5.4)^9}{9!} \right]$$

$$F(9) = 0.0045[1 + 5.4 + 14.58 + 26.244 + 35.4294 + 38.2638 + 34.4374 + 26.56598 + 17.932 + 10.759]$$

$$F(9) = 0.0045[210.6118] = 0.95123$$

Untuk $x = 10$

$$F(10) = e^{-5.4} \sum_{i=0}^{10} \frac{(5.4)^i}{i!}$$

$$F(10) = e^{-5.4} \left[\frac{(5.4)^0}{0!} + \frac{(5.4)^1}{1!} + \frac{(5.4)^2}{2!} + \frac{(5.4)^3}{3!} + \frac{(5.4)^4}{4!} + \frac{(5.4)^5}{5!} + \frac{(5.4)^6}{6!} + \frac{(5.4)^7}{7!} + \frac{(5.4)^8}{8!} + \frac{(5.4)^9}{9!} + \frac{(5.4)^{10}}{10!} \right]$$

$$F(10) = 0.0045[1 + 5.4 + 14.58 + 26.244 + 35.4294 + 38.2638 + 34.4374 + 26.56598 + 17.932 + 10.759 + 5.80998]$$

$$F(10) = 0.0045[216.4217] = 0.977486$$

Kemudian akan dicari nilai $H(x)$ dengan menggunakan Persamaan (2.14).

$$H(1) = \frac{2}{5} = 0.4$$

$$H(6) = \frac{3}{5} = 0.6$$

$$H(9) = \frac{4}{5} = 0.8$$

$$H(10) = \frac{5}{5} = 1$$

Kemudian dicari nilai $|F(x) - H(x)|$.

$$|F(1) - H(1)| = |0.0289 - 0.4| = |-0.37109| = 0.37109$$
$$|F(6) - H(6)| = |0.701671 - 0.6| = |0.10167| = 0.10167$$
$$|F(9) - H(9)| = |0.95125 - 0.8| = |0.151245| = 0.151245$$
$$|F(10) - H(10)| = |0.977486 - 1| = |-0.022514| = 0.022514$$

Berdasarkan Persamaan (2.13) diperoleh nilai D_{hitung} sebesar 0.371094. Sedangkan D_{tabel} sebesar 0.56327 yang diperoleh melalui tabel Kolmogorov-Smirnov pada Lampiran 3. Karena $D_{hitung} < D_{tabel}$, maka berdasarkan kriteria uji dapat disimpulkan bahwa H_0 diterima atau dengan kata lain, data mengikuti sebaran distribusi *Poisson*.

Validasi dengan menggunakan Matlab.

```
parameter = fitdist(data,'poisson')
[h,p,kstat,cv]=kstest(data,'CDF',parameter)
```

```
h =

     0


p =

    0.3955


kstat =

    0.3711


cv =

    0.5633
```

Gambar 2.7 Uji Kolmogorov-Smirnov dengan Matlab

Dengan menggunakan Matlab, diperoleh nilai statistik uji sebesar 0.3711 dengan D_{tabel} sebesar 0.56327.

2.19 Uji Anderson-Darling

Uji Anderson-Darling adalah memeriksa apakah suatu sampel mengikuti sebaran dari distribusi probabilitas tertentu dengan berdasarkan pada nilai statistik uji Anderson-Darling. Statistik uji Anderson-Darling dikenal sebagai statistik yang kuat dengan menekankan perbedaan pada ujung kurva antara fungsi empiris dan fungsi distribusi kumulatif dari distribusi yang diasumsikan (Ang & Tang, 2007). Statistik uji Anderson-Darling dapat didefinisikan sebagai

$$A_n^2 = -n - \frac{1}{n} \sum_{i=1}^{n} (2i-1)[\log F(x_i) + \log(1 - F(x_{n+1-i})] \qquad (2.15)$$

Di mana

A_n^2 : Statistik uji Anderson-Darling.

$F(x)$: Fungsi distribusi kumulatif distribusi yang dihipotesiskan.

n : Jumlah data.

i : Frekuensi kumulatif.

Dengan kriteria uji, jika nilai $D_{hitung} > D_{tabel}$ dengan tingkat error sebesar α, maka H_0 ditolak. Jika nilai $D_{hitung} < D_{tabel}$ dengan tingkat error sebesar α, maka H_0 diterima atau dengan kata lain data mengikuti model sebaran distribusi tersebut.

Contoh 2.9

Akan dilakukan uji hipotesis data acak pada Tabel 2.1 yang diasumsikan berdistribusi normal dengan uji Anderson-Darling dengan hipotesis yang digunakan sebagai berikut :

H_0 : Data mengikuti model sebaran distribusi Normal.

H_1 : Data tidak mengikuti model sebaran distribusi Normal.

dan dengan taraf signifikan sebesar 0.05.

Penyelesaian

Tabel 2.3 Uji Anderson Darling

Data	$(x_i - \mu)$	$(x_i - \mu)^2$	Frekuensi Kumulatif (i)	$F(x_i)$	$F(x_{n+1-i})$	$AD(S)_i$
1	-4.4	19.36	1	0.151844633	0.858881245	-0.768610189
1	-4.4	19.36	2	0.151844633	0.799978446	-2.096536551
6	0.6	0.36	3	0.555771679	0.555771679	-1.398814331
9	3.6	12.96	4	0.799978446	0.151844633	-0.543006713
10	4.6	21.16	5	0.858881245	0.151844633	-0.570268904

Penjelasan Tabel 2.3

Susun data dari terkecil ke terbesar, kemudian hitung frekuensi kumulatif data.

Diketahui fungsi kepadatan peluang dari distribusi normal sebagai berikut:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{\left(\frac{1}{2}\right)\left(\frac{x-\mu}{\sigma}\right)^2};\ -\infty < x < \infty$$

dan fungsi distribusi kumulatif distribusi normal sebagai berikut:

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma}} e^{\left(\frac{1}{2}\right)\left(\frac{t-\mu}{\sigma}\right)^2}\ dt$$

Hitung mean dan standard deviasi

$$\mu = \frac{1+1+6+9+10}{5} = 5.4$$

$$SD = \left[\frac{\sum(x_i - \mu)^2}{n-1}\right]^{\frac{1}{2}} = \left[\frac{73.2}{4}\right]^{\frac{1}{2}} = 4.27785$$

Hitung nilai $F(x_i)$ dengan menggunakan fungsi distribusi kumulatif distribusi normal.

$$F(1) = \int_{-\infty}^{1} \frac{1}{\sqrt{2\pi\sigma}} e^{\left(\frac{1}{2}\right)\left(\frac{t-\mu}{\sigma}\right)^2}\ dt = 0.151844633$$

$$F(6) = \int_{-\infty}^{6} \frac{1}{\sqrt{2\pi\sigma}} e^{\left(\frac{1}{2}\right)\left(\frac{t-\mu}{\sigma}\right)^2} \, dt = 0.555771679$$
$$F(9) = \int_{-\infty}^{9} \frac{1}{\sqrt{2\pi\sigma}} e^{\left(\frac{1}{2}\right)\left(\frac{t-\mu}{\sigma}\right)^2} \, dt = 0.799978446$$
$$F(10) = \int_{-\infty}^{10} \frac{1}{\sqrt{2\pi\sigma}} e^{\left(\frac{1}{2}\right)\left(\frac{t-\mu}{\sigma}\right)^2} \, dt = 0.858881245$$

Kemudian nilai $F(x_i)$ yang sudah diperoleh, disusun secara terurut dari yang terbesar ke yang terkecil untuk memperoleh nilai $F(x_{n+1-i})$. Lalu hitung nilai S dengan rumus sebagai berikut :

$$S = \sum AD(S)_i = \sum \frac{2i-1}{n} (\ln(F(x_i)) + \ln(1 - F(x_{n+1-i}))$$

Dimana n adalah banyaknya data. Sehingga diperoleh nilai $S = -5.37724$.

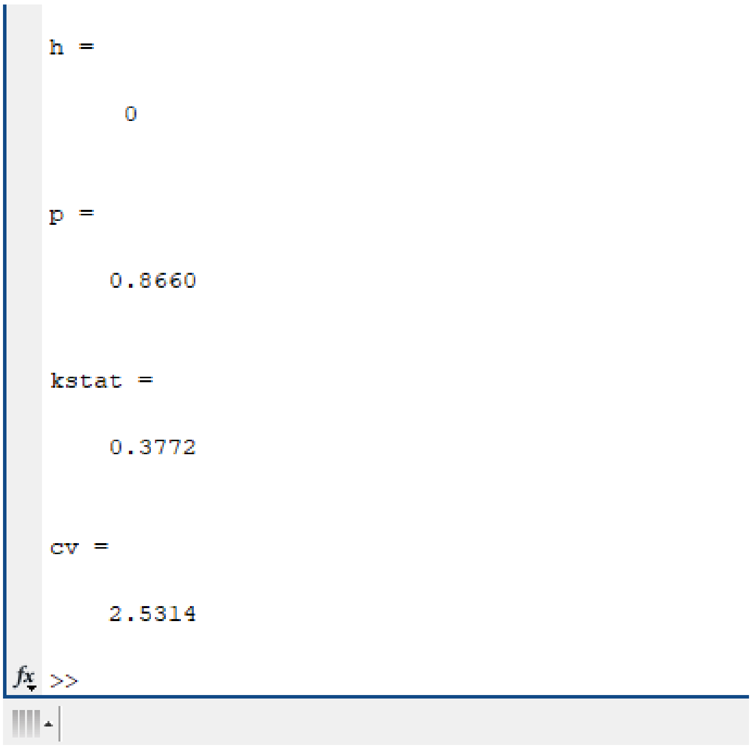
Hitung nilai statistik uji dengan rumus sebagai berikut :

$$A^2 = -n - s = -5 - (-5.37724)$$
$$A^2 = 0.377237$$

Berdasarkan Persamaan (2.15) diperoleh nilai D_{hitung} sebesar 0.377237. Sedangkan D_{tabel} sebesar 2.5 yang diperoleh melalui tabel Anderson-Darling pada Lampiran 4. Karena $D_{hitung} < D_{tabel}$, maka berdasarkan kriteria uji dapat disimpulkan bahwa H_0 diterima atau dengan kata lain, data mengikuti sebaran distribusi Normal.

Validasi dengan Matlab

```
parameter = fitdist(data,'poisson')
[h,p,kstat,cv]=adtest(data,'Distribution',parameter)
```



Gambar 2.8 Uji Anderson-Darling dengan Matlab

Dengan menggunakan Matlab, diperoleh nilai D_{hitung} sebesar 0.3772 dengan D_{tabel} sebesar 2.5314.

2.20 Normalisasi *Z-Score*

Normalisasi *Z-Score* adalah salah satu teknik transformasi data yang berdasarkan pada nilai rata-rata data dan nilai standar deviasi data (Henderi, 2021).

$$Z = \frac{x - \mu}{\sigma} \tag{2.16}$$

Di mana

Z : data yang sudah dinormalisasi

x : data yang akan dinormalisasi

μ : rata-rata data

σ : standard deviasi data

2.21 Analisis Klaster

Pengelompokan data atau *Clustering* adalah proses membagi sekelompok objek ke dalam beberapa kelompok/klaster berdasarkan kesamaan dan ketaksamaannya. Dalam mengelompokkan data, terdapat beberapa metode yang dapat digunakan. Namun, secara umum metode pengelompokkan data terbagi menjadi dua jenis, hirarki *clustering* dan non-hirarki *clustering*. Hirarki *clustering* sendiri terbagi menjadi dua jenis yaitu *Agglomerative Hierarchical Clustering* dan *Divisive Hierarchical Clustering*. *Agglomerative Hierarchical Clustering* akan menganggap setiap objek sebagai sebuah klaster, kemudian beberapa klaster digabungkan menjadi sebuah klaster baru sampai diperoleh klaster yang diinginkan. Sedangkan *Divisive Hierarchical Clustering* akan menganggap semua objek tergabung dalam sebuah klaster, kemudian dipisahkan menjadi beberapa klaster sampai diperoleh klaster yang diinginkan. Adapun contoh dari hirarki *Clustering* seperti *Single Linkage Clustering*, *Complete Linkage Clustering, Average Linkage Clustering,* dan Metode *Ward*.

2.22 Single Linkage Clustering

Single Linkage Clustering merupakan salah satu jenis dari *Agglomerative Hierarchical Clustering.* Pada metode *Single Linkage*, setiap objek akan ditempatkan ke dalam sebuah klaster terpisah. Kemudian, klaster-klaster yang memiliki jarak antar klaster terdekat akan dikelompokkan menjadi sebuah klaster baru pada setiap iterasi hingga kondisi terminasi tertentu terpenuhi. Jarak antar klaster dapat dihitung dengan menggunakan *Euclidean Distance*, dengan rumus sebagai berikut:

$$d_{(u,v)} = \left[\sum_{k=1}^{n} (x_k - y_k)^2\right]^{\frac{1}{2}} \tag{2.17}$$

Di mana

$d_{(u,v)}$: jarak antar klaster u dan klaster v

x_k : data pada klaster u

y_k : data pada klaster v

Misalkan terdapat tiga buah data yaitu U, V, dan W dengan U dan V berada dalam satu klaster (UV). Maka jarak antar data baru antara klaster (UV) ke W dapat dihitung dengan cara mencari nilai terkecil antara jarak U ke W dan jarak V ke W.

$$d_{((U,V)W)} = \min\{d_{(U,W)}, d_{(V,W)}\} \tag{2.18}$$

Di mana

$d_{(U,V)W}$: jarak antar klaster (U, V) dan klaster W

d_{UW} : jarak antar klaster U dan klaster W

d_{VW} : jarak antar klaster V dan klaster W

Langkah-langkah dalam melakukan klasterisasi dengan metode *Single Linkage Clustering* sebagai berikut:

1. Dimulai dengan mencari jarak antar data dengan menggunakan Persamaan (2.16).
2. Pada iterasi pertama, tentukan pasangan data yang memiliki jarak antar data terkecil.
3. Pasangan data yang memiliki jarak antar data terkecil akan digabungkan menjadi suatu klaster baru.
4. Ulangi langkah 2 dan langkah 3 sampai klaster yang diinginkan terbentuk.

Contoh 2.10

Diberikan data sebagai berikut. Bagaimana cara mengklasterisasi data dibawah dengan menggunakan metode *Single Linkage Clustering*?

Tabel 2.4 Data

Data	X	Y
U_1	0.40	0.53
U_2	0.22	0.38
U_3	0.35	0.32
U_4	0.26	0.19
U_5	0.08	0.41
U_6	0.45	0.30

Penyelesaian

Cari jarak antar data terlebih dahulu dengan menggunakan Persamaan (2.17).

Jarak U_1 ke U_2

$$d_{(U_1,U_2)} = [(0.40 - 0.22)^2 + (0.53 - 0.38)^2]^{\frac{1}{2}}$$

$$d_{(U_1,U_2)} = [(0.18)^2 + (0.15)^2]^{\frac{1}{2}}$$

$$d_{(U_1,U_2)} = [0.0324 + 0.0225]^{\frac{1}{2}} = [0.0549]^{\frac{1}{2}}$$

$$d_{(U_1,U_2)} = 0.2343 \approx 0.23$$

Jarak U_1 ke U_3

$$d_{(U_1,U_3)} = [(0.40 - 0.35)^2 + (0.53 - 0.32)^2]^{\frac{1}{2}}$$

$$d_{(U_1,U_3)} = [(0.05)^2 + (0.21)^2]^{\frac{1}{2}}$$

$$d_{(U_1,U_3)} = [0.0025 + 0.0441]^{\frac{1}{2}} = [0.0466]^{\frac{1}{2}}$$

$$d_{(U_1,U_3)} = 0.2158 \approx 0.22$$

Jarak U_1 ke U_4

$$d_{(U_1,U_4)} = [(0.40 - 0.26)^2 + (0.53 - 0.19)^2]^{\frac{1}{2}}$$

$$d_{(U_1,U_4)} = [(0.14)^2 + (0.34)^2]^{\frac{1}{2}}$$

$$d_{(U_1,U_4)} = [0.0196 + 0.1156]^{\frac{1}{2}} = [0.1352]^{\frac{1}{2}}$$

$$d_{(U_1,U_4)} = 0.3676 \approx 0.37$$

Jarak U_1 ke U_5

$$d_{(U_1,U_5)} = [(0.40 - 0.08)^2 + (0.53 - 0.41)^2]^{\frac{1}{2}}$$

$$d_{(U_1,U_5)} = [(0.32)^2 + (0.12)^2]^{\frac{1}{2}}$$

$$d_{(U_1,U_5)} = [0.1024 + 0.0144]^{\frac{1}{2}} = [0.1168]^{\frac{1}{2}}$$
$$d_{(U_1,U_5)} = 0.3417 \approx 0.34$$

Jarak U_1 ke U_6

$$d_{(U_1,U_6)} = [(0.40 - 0.45)^2 + (0.53 - 0.30)^2]^{\frac{1}{2}}$$
$$d_{(U_1,U_6)} = [(-0.05)^2 + (0.23)^2]^{\frac{1}{2}}$$
$$d_{(U_1,U_6)} = [0.0025 + 0.0529]^{\frac{1}{2}} = [0.0554]^{\frac{1}{2}}$$
$$d_{(U_1,U_6)} = 0.2353 \approx 0.24$$

Proses ini dilakukan terus hingga diperoleh semua jarak antar klaster. Jarak antar klaster yang sudah diperoleh, kemudian dibentuk ke dalam bentuk tabel untuk mempermudah perhitungan. Sehingga diperoleh tabel jarak antar klaster sebagai berikut :

Tabel 2.5 Jarak Antar Klaster

	U_1	U_2	U_3	U_4	U_5	U_6
U_1	0					
U_2	0.23	0				
U_3	0.22	0.14	0			
U_4	0.37	0.19	0.16	0		
U_5	0.34	0.14	0.28	0.28	0	
U_6	0.24	0.24	0.10	0.22	0.39	0

Iterasi 1

Kemudian berdasarkan Tabel 2.5, cari jarak antar klaster terdekat yaitu $d_{(U_3,U_6)} = 0.10$ sehingga U_3 dan U_6 akan membentuk sebuah klaster baru. Kemudian akan dicari jarak antar klaster yang baru dengan menggunakan Persamaan (2.18).

$$d_{((U_3,U_6),U_1)} = \min\{d_{(U_3,U_1)}, d_{(U_6,U_1)}\} = \min\{0.22\ , 0.24\} = 0.22$$

$$d_{((U_3,U_6),U_2)} = \min\{d_{(U_3,U_2)}, d_{(U_6,U_2)}\} = \min\{0.14\ , 0.24\} = 0.14$$

$$d_{((U_3,U_6),U_4)} = \min\{d_{(U_3,U_4)}, d_{(U_6,U_4)}\} = \min\{0.16\ , 0.22\} = 0.16$$

$$d_{((U_3,U_6),U_5)} = \min\{d_{(U_3,U_5)}, d_{(U_6,U_5)}\} = \min\{0.28\ , 0.39\} = 0.28$$

Diperoleh tabel jarak antar klaster yang baru sebagai berikut :

Tabel 2.6 Jarak Antar Klaster pada Iterasi Ke-1

	U_1	U_2	U_3, U_6	U_4	U_5
U_1	0				
U_2	0.23	0			
U_3, U_6	0.22	0.14	0		
U_4	0.37	0.19	0.16	0	
U_5	0.34	0.14	0.28	0.28	0

Jika jumlah klaster yang ingin dibentuk sebanyak 5 klaster, maka perhitungan dapat dihentikan pada Iterasi 1 dengan klaster yang terbentuk berdasarkan Tabel 2.8 yaitu klaster U_1, klaster U_2, klaster (U_3, U_6), klaster U_4, dan klaster U_5. Jika tidak, maka perhitungan akan dilanjutkan ke Iterasi 2.

Iterasi 2

Pada Tabel 2.6, cari jarak antar klaster terdekat yaitu $d_{(U_5,U_2)} = 0.14$ dan $d_{((U_3,U_6),U_2)} = 0.14$. Namun klaster baru yang akan terbentuk hanya klaster (U_2, U_5). Hal ini dikarenakan $d_{(U_5,U_2)}$ merupakan jarak antar klaster U_5 dan U_2, dimana masing-masing klaster beranggotakan satu anggota. Sedangkan $d_{((U_3,U_6),U_2)}$ merupakan jarak antar klaster (U_3, U_6) dan U_2, dimana klaster (U_3, U_6) beranggotakan dua anggota. Kemudian akan dicari jarak antar klaster yang baru dengan menggunakan Persamaan (2.18).

$$d_{((U_2,U_5),U_1)} = \min\{d_{(U_2,U_1)}, d_{(U_5,U_1)}\} = \min\{0.23\ , 0.34\} = 0.23$$

$$d_{((U_2,U_5),(U_3,U_6))} = \min\left\{d_{(U_2,(U_3,U_6))}, d_{(U_5,(U_3,U_6))}\right\} = \min\{0.14\ , 0.28\} = 0.14$$

$$d_{((U_2,U_5),U_4)} = \min\{d_{(U_2,U_4)}, d_{(U_5,U_4)}\} = \min\{0.19\ , 0.28\} = 0.19$$

Diperoleh tabel jarak antar klaster yang baru sebagai berikut :

Tabel 2.7 Jarak Antar Klaster pada Iterasi Ke-2

	U_1	U_2, U_5	U_3, U_6	U_4
U_1	0			
U_2, U_5	0.23	0		
U_3, U_6	0.22	0.14	0	
U_4	0.37	0.19	0.16	0

Jika jumlah klaster yang ingin dibentuk sebanyak 4 klaster, maka perhitungan dapat dihentikan pada Iterasi 2 dengan klaster yang terbentuk berdasarkan Tabel 2.9 yaitu klaster U_1, klaster (U_2, U_5), klaster (U_3, U_6), dan klaster U_4. Jika tidak, maka perhitungan akan dilanjutkan ke Iterasi 3.

Iterasi 3

Pada Tabel 2.7, cari jarak antar klaster terdekat yaitu $d_{((U_3,U_6),(U_2,U_5))} = 0.14$ sehingga (U_3, U_6) dan (U_2, U_5) akan membentuk sebuah klaster baru. Kemudian akan dicari jarak antar klaster yang baru dengan menggunakan Persamaan (2.18).

$$d_{((U_3,U_6,U_2,U_5),U_1)} = \min\left\{d_{((U_3,U_6),U_1)}, d_{((U_2,U_5),U_1)}\right\} = \min\{0.22\ , 0.23\ ,\} = 0.22$$
$$d_{((U_3,U_6,U_2,U_5),U_4)} = \min\left\{d_{((U_3,U_6),U_4)}, d_{((U_2,U_5),U_4)}\right\} = \min\{0.16\ , 0.19\ ,\} = 0.16$$

Diperoleh tabel jarak antar klaster yang baru sebagai berikut :

Tabel 2.8 Jarak Antar Klaster pada Iterasi Ke-3

	U_1	U_3, U_6, U_2, U_5	U_4
U_1	0		
U_3, U_6, U_2, U_5	0.22	0	
U_4	0.37	0.16	0

Jika jumlah klaster yang ingin dibentuk sebanyak 3 klaster, maka perhitungan dapat dihentikan pada Iterasi 3 dengan klaster yang terbentuk berdasarkan Tabel

2.10 yaitu klaster U_1, klaster (U_3, U_6, U_2, U_5), dan klaster U_4. Jika tidak, maka perhitungan akan dilanjutkan ke Iterasi 4.

Iterasi 4

Pada Tabel 2.8, cari jarak antar klaster terdekat yaitu $d_{((U_3 U_6 U_2 U_5), U_4)} = 0.16$ sehingga (U_3, U_6, U_2, U_5) dan U_4 akan membentuk sebuah klaster. Kemudian akan dicari jarak antar klaster yang baru dengan menggunakan Persamaan (2.18).

$$\begin{aligned} d_{((U_3,U_6,U_2,U_5,U_4),U_1)} &= \min\left\{d_{((U_3,U_6,U_2,U_5),U_1)}, d_{(U_4,U_1)}\right\} \\ &= \min\{0.22\,, 0.37\} = 0.22 \end{aligned}$$

Diperoleh tabel jarak antar klaster yang baru sebagai berikut :

Tabel 2.9 Jarak Antar Klaster pada Iterasi Ke-4

	U_1	U_3, U_6, U_4, U_2, U_5
U_1	0	
U_3, U_6, U_2, U_5, U_4	0.22	0

Jika jumlah klaster yang ingin dibentuk sebanyak 2 klaster, maka perhitungan dapat dihentikan pada Iterasi 4 dengan klaster yang terbentuk berdasarkan Tabel 2.9 yaitu klaster U_1 dan klaster $(U_3, U_6, U_2, U_5, U_4)$.

Validasi Dengan Menggunakan Matlab

Dengan menggunakan *syntax* berikut, data akan diklaster kedalam 2,3,4, dan 5 klaster.

```
A = pdist (DATA);
B = linkage (DATA,'single');
D = cluster (B,'Maxclust', c);
```

dimana c adalah banyaknya klaster yang ingin dibentuk.

Tabel 2.10 Hasil Klasterisasi dengan Matlab

Data	Jumlah Klaster			
	2	3	4	5
U_1	1	1	1	1
U_2	2	2	2	2
U_3	2	2	3	3
U_4	2	3	4	4
U_5	2	2	2	5
U_6	2	2	3	3

Berdasarkan Tabel 2.10, diketahui bahwa untuk 2 klaster yang terbentuk, U_1 berada pada klaster 1 dan $U_2 . U_3, U_4, U_5$ dan U_6 berada pada klaster 2. Untuk 3 klaster yang terbentuk, U_1 berada pada klaster 1, $U_2 . U_3, U_5$ dan U_6 berada pada klaster 2, dan U_4 berada pada klaster 3 dan seterusnya.

2.23 Index Davies-Bouldin

Index Davies-Bouldin (IDB) merupakan salah satu metode untuk mengevaluasi klaster dalam metode pengelompokkan data dengan skema evaluasi klaster internal, dimana baik tidaknya sebuah klaster dapat dilihat dari kuantitas dan jarak antar klaster. Index Davies-Bouldin (IDB) memaksimalkan jarak antar kalster dan meminimalkan jarak antar data pada sebuah klaster. Sehingga jika nilai Index Davies-bouldin (IDB) semakin kecil maka hasil klaster yang diperoleh semakin optimal. Nilai Index Davies-bouldin dirumuskan sebagai berikut :

$$IDB = \frac{1}{k}\sum_{i=1}^{k} R_i \tag{2.19}$$

dengan

$$R_i = \max_{j=1,\dots,k, i \neq j} R_{ij}$$

dan

$$R_{ij} = \frac{SSW_i + SSW_j}{SSB_{ij}} \tag{2.20}$$

dengan

IDB : Indeks Davies-bouldin

k : Jumlah klaster

R_i : Nilai R_{ij} maksimum untuk setiap klaster

R_{ij} : Ukuran kemiripan antara klaster i dengan klaster *j*

SSW_i : Jarak rata-rata antara anggota pada klaster i terhadap pusat klaster i

SSB_{ij} : jarak pusat klaster *i* ke pusat klaster *j*

Untuk menghitung jarak pusat klaster *i* ke pusat klaster *j* dapat dihitung dengan menggunakan Persamaan (2.8) sehingga SSB_{ij} dapat dirumuskan sebagai berikut :

$$SSB_{ij} = d_{(c_i,c_j)} \tag{2.21}$$

dan SSW_i dapat dirumuskan sebagai berikut :

$$SSW_i = \left[\frac{1}{n_i}\sum_{t=1}^{n_i}\left(d_{(x_t,c_i)}\right)^2\right]^{\frac{1}{2}}, x_t \in i \tag{2.22}$$

dengan

$$c_i = \frac{1}{n_i}\sum_{t=1}^{n_i} x_t \ , x_t \in i \tag{2.23}$$

Di mana

c_i : Titik pusat klaster i

n_i : Jumlah anggota klaster i

Langkah-langkah untuk mengevaluasi hasil klaster yang terbentuk dengan Index Davies-Bouldin sebagai berikut :

1. Tentukan titik pusat untuk setiap klaster,
2. Hitung jarak rata-rata antara anggota pada klaster i terhadap pusat klaster i,
3. Hitung jarak pusat klaster i ke pusat klaster j,
4. Hitung ukuran kemiripan antara klaster i dengan klaster j,
5. Tentukan nilai R_{ij} maksimal untuk setiap klaster,
6. Hitung nilai Index Davies-Bouldin

Contoh 2.11

Berapa jumlah klaster optimum yang terbentuk untuk contoh 2.10.

Penyelesaian

Evalusi hasil klaster untuk 2 klaster dengan Index Davies-Bouldin. Berdasarkan Tabel 2.10 untuk 2 klaster sebagai berikut :

Tabel 2.11 Klaster yang Terbentuk

Data	X	Y	Klaster
U_1	0.40	0.53	1
U_2	0.22	0.38	2
U_3	0.35	0.32	2
U_4	0.26	0.19	2
U_5	0.08	0.41	2
U_6	0.45	0.30	2

Kemudian dicari titik pusat c_1 dan c_2 dengan menggunakan Persamaan (2.23).

Untuk c_1

$$\bar{x} = \frac{0.40}{1} = 0.40$$

$$\bar{y} = \frac{0.53}{1} = 0.53$$

Untuk c_2

$$\bar{x} = \frac{(0.22 + 0.35 + 0.26 + 0.08 + 0.45)}{5} = 0.272$$

$$\bar{y} = \frac{(0.38 + 0.32 + 0.19 + 0.41 + 0.30)}{5} = 0.32$$

Sehingga diperoleh titik pusat c_1 dan c_2 sebagai berikut :

Tabel 2.12 Titik Pusat

Titik Pusat	X	Y
c_1	0.4	0.53
c_2	0.272	0.320

Kemudian akan dicari jarak rata-rata antara anggota pada klaster 1 terhadap pusat klaster 1 dengan menggunakan Persamaan (2.22).

$$d_{(U_1,C_1)} = [(0.40 - 0.4)^2 + (0.53 - 0.53)^2]^{\frac{1}{2}} = 0$$

$$SSW_1 = \left[\left(d_{(U_1,C_1)}\right)^2\right]^{\frac{1}{2}} = [(0)^2]^{\frac{1}{2}} = 0$$

Kemudian akan dicari jarak rata-rata antara anggota pada klaster 2 terhadap pusat klaster 2 dengan menggunakan Persamaan (2.22).

$$d_{(U_2,C_2)} = [(0.22 - 0.272)^2 + (0.38 - 0.32)^2]^{\frac{1}{2}} = 0.0794$$

$$d_{(U_3,C_2)} = [(0.35 - 0.272)^2 + (0.32 - 0.32)^2]^{\frac{1}{2}} = 0.078$$

$$d_{(U_4,C_2)} = [(0.26 - 0.272)^2 + (0.19 - 0.32)^2]^{\frac{1}{2}} = 0.13055$$

$$d_{(U_5,C_2)} = [(0.08 - 0.272)^2 + (0.41 - 0.32)^2]^{\frac{1}{2}} = 0.21205$$

$$d_{(U_6,C_2)} = [(0.45 - 0.272)^2 + (0.30 - 0.32)^2]^{\frac{1}{2}} = 0.17912$$

$$SSW_2 = \left[\frac{\left(d_{(U_2,C_2)}\right)^2 + \left(d_{(U_3,C_2)}\right)^2 + \left(d_{(U_4,C_2)}\right)^2 + \left(d_{(U_5,C_2)}\right)^2 + \left(d_{(U_6,C_2)}\right)^2}{5}\right]^{\frac{1}{2}}$$

$$= \left[\frac{(0.0794)^2 + (0.078)^2 + (0.13055)^2 + (0.21205)^2 + (0.17912)^2}{5}\right]^{\frac{1}{2}}$$

$$= \left[\frac{0.0063 + 0.0061 + 0.017 + 0.04496 + 0.0321}{5}\right]^{\frac{1}{2}} = [0.021296]^{\frac{1}{2}}$$

$$= 0.145931$$

Kemudian hitung jarak pusat klaster 1 ke pusat klaster 2 dengan menggunakan Persamaan (2.21).

$$SSB_{12} = [(0.4 - 0.272)^2 + (0.53 - 0.32)^2]^{\frac{1}{2}}$$
$$= 0.245935$$

Kemudian hitung ukuran kemiripan antara klaster 1 dengan klaster 2 dengan menggunakan Persamaan (2.20).

$$R_{12} = \frac{(0 + 0.145931)}{0.245935} = 0.593374$$

Ukuran kemiripan yang sudah diperoleh kemudian akan dibentuk menjadi sebuah tabel. Dimana tabel yang terbentuk nantinya akan digunakan untuk menentukan ukuran kemiripan maksimal untuk setiap klaster.

Tabel 2.13 Ukuran Kemiripan

	1	2	R maksimum
1	0	0.593374	0.593374
2	0.593374	0	0.593374

Berdasarkan Tabel 2.13, dapat diketahui bahwa ukuran kemiripan klaster 1 dengan klaster 2 sebesar 0.593374. Sehingga diperoleh R maksimum untuk klaster

1 dan klaster 2 sebesar 0.593374. Dengan menggunakan Persamaan (2.19) diperoleh nilai Index Davies Bouldin,

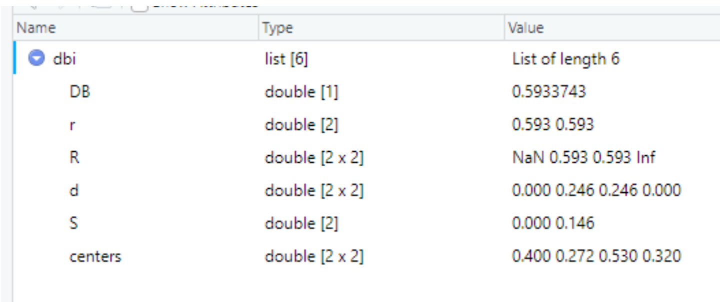
$$IDB = \frac{(0.593374 + 0.593374)}{2} = 0.593374$$

Validasi menggunakan Software RStudio

```
library(ggplot2)
library(factoextra)
library(tidyverse)
library(cluster)
library(MVN)
library(scales)
library(fpc)
library(clusterSim)
library(dendroextras)
dt=read.delim("clipboard")
dt1=scale(dt)
jarak=dist(x=dt1, method = "euclidean")
single1=hclust(jarak, method = "single")
Kelompok = cutree(single1,  k = 2)
tabel = cbind(dt1,Kelompok)
dbi = index.DB(x=dt,Kelompok,jarak,centrotypes = "centroids")
dbi$DB
```

Dimana k adalah jumlah klaster yang terbentuk.

Diperoleh nilai Index Davies-Bouldin untuk jumlah klaster sebanyak 2 klaster yaitu 0.593374.



Name	Type	Value
dbi	list [6]	List of length 6
DB	double [1]	0.5933743
r	double [2]	0.593 0.593
R	double [2 x 2]	NaN 0.593 0.593 Inf
d	double [2 x 2]	0.000 0.246 0.246 0.000
S	double [2]	0.000 0.146
centers	double [2 x 2]	0.400 0.272 0.530 0.320

Gambar 2.9 Evaluasi Klaster dengan R Studio

Kemudian dengan *syntax* yang serupa akan dicari nilai Index Davies-Bouldin untuk jumlah klaster sebanyak 3,4 dan 5 klaster. Sehingga diperoleh nilai Index Davies-Bouldin untuk jumlah klaster sebanyak 2,3,4 dan 5 klaster sebagai berikut:

Tabel 2.14 Nilai Index *Davies-Bouldin*

Jumlah Klaster	IDB
2	0.593374
3	0.82018
4	0.3720298
5	0.2401896

Berdasarkan Tabel 2.14, nilai IDB untuk 2 klaster sebesar 0.593374. Nilai IDB untuk 3 klaster sebesar 0.82018. Nilai IDB untuk 4 klaster sebesar 0.3820298 dan nilai IDB untuk 5 klaster sebesar 0.2401896. Karena nilai IDB untuk 5 klaster merupakan nilai IDB terkecil sebesar 0.2401896, maka diperoleh jumlah klaster optimum yaitu sebanyak 5 klaster.